The tepary bean genome provides insight into evolution and domestication under heat stress

Samira Mafi Moghaddam¹,2, Atena Oladzad³, Chushin Koh⁴,⁵, Larissa Ramsay⁴, John P. Hart⁶, Sujan Mamidi⁷, Genevieve Hoopes¹, Avinash Sreedasyam⁷, Andrew Wiersma¹,², Dongyan Zhao¹, Jane Grimwood⁷,⁸, John P. Hamilton¹, Jeremy Jenkins⁷,⁸, Brieanne Vaillancourt¹, Joshua C. Wood¹, Jeremy Schmutz⁷,⁸, Sateesh Kagale⁹, Timothy Porch⁶, Kirstin E. Bett⁴, C. Robin Buell¹,²,¹⁰ & Phillip E. McClean³

Tepary bean (Phaseolus acutifolis A. Gray), native to the Sonoran Desert, is highly adapted to heat and drought. It is a sister species of common bean (Phaseolus vulgaris L.), the most important legume protein source for direct human consumption, and whose production is threatened by climate change. Here, we report on the tepary genome including exploration of possible mechanisms for resilience to moderate heat stress and a reduced disease resistance gene repertoire, consistent with adaptation to arid and hot environments. Extensive collinearity and shared gene content among these Phaseolus species will facilitate engineering climate adaptation in common bean, a key food security crop, and accelerate tepary bean improvement.
Heat and drought threaten global food security due to their impact on crop productivity. While genetic adaptations to these stresses can be found in plant species that evolved under extreme climatic conditions, few domesticated crops show economically sustainable food production when exposed to abiotic stress. To increase the productivity of a modern crop for a warmer and drier climate, it is important to discover the variation in either the crop species or close relatives for which genetic variation for fitness under abiotic stress has evolved.

Tepary bean (Phaseolus acutifolius A. Gray) recently gained attention in modern crop improvement efforts as a source of genetic traits for biotic and abiotic stress resistance for its sister species, common bean (P. vulgaris L.). Common bean, the most consumed grain legume worldwide, is adapted to a wide range of climates, but is sensitive to drought and extreme temperatures, which result in significant yield reduction under those environmental conditions. The close phylogenetic relationship between tepary bean and common bean enables tepary bean to serve as a key source of heat and drought tolerance. While hybridization between these two species is difficult, persistent breeding efforts have successfully transferred genes conferring common bacterial blight resistance, bruchid seed pest resistance, and drought tolerance from tepary to common bean.

Considering the rapid rate of climate change documented worldwide, the adaptation of tepary bean to high ambient temperatures and to the dry, arid conditions of its native Sonoran Desert habitat positions it as an alternative protein source in areas threatened with a hotter and drier future. Tepary bean is currently produced on a limited scale, primarily by subsistence farmers in arid regions. Researchers have posited its limited adoption to inadequate marketing and reduced appeal owing to its small seed size, prostrate plant habit, and culinary characteristics. Recent tepary bean breeding efforts improved yield, adaptation, seed size, seed quality, common bacterial blight resistance, and rust resistance, highlighting its potential for improvement via focused breeding efforts. Indeed, tepary bean could serve as an alternative pulse crop for marginal environments and as a genetic resource for abiotic stress tolerance to breed more resilient common beans under a changing climate.

In this work, we describe reference genome sequences for the cultivated landrace Frijol Bayo (P. acutifolius var. acutifolius; PI 692269; G40001-Seq) and the wild accession W6 15578 (P. acutifolius var. acutifolius; PI 638833). Gene expression responses of Frijol Bayo relative to a heat-tolerant common bean variety (Amadeus-77) under moderate night-time temperature stress reveal differential thermotolerance responses in tepary compared to the common bean. Landrace and wild gene family composition differ with those in common bean, with a special reference to the disease resistance genes associated with effector-driven resistance to pathogens. Loci associated with the domestication of P. acutifolius and P. vulgaris have limited overlap. These genome resources and associated analyses provide insights into innovations in adaptation to diverse environments and provide a roadmap for improvements of both Phaseolus species needed for the climate stress environments of the twenty-first century.

**Results**

**De novo genome assembly and annotation.** Frijol Bayo is an indeterminate, photoperiod insensitive, and broadly adapted white-seeded landrace used to introgress common bacterial blight resistance from tepary to common bean (Fig. 1a, b). W6 15578 is an indeterminate, photoperiod sensitive wild accession with small, mottled seeds, and pods that shatter at maturity and used to improve abiotic stress tolerance in common bean. Reference genome assemblies for tepary bean were generated using Pacific Biosciences (PACBIO) long-read sequencing for Frijol Bayo with the MECAT assembler, and Illumina paired-end, mate-pair, and linked reads with NRGene’s DeNovoMAGIC23M,15 platform for W6 15578 and scaffolded into pseudochromosomes using proximity-by-igation and anchoring to genetic maps (Table 1). Both assemblies provided robust representation of the gene space as reflected in the detection of 92.5% (Frijol Bayo) and 92.9% (W6 15578) of the BUSCO Embryophyta orthologs (Supplementary Data 1). Based on k-mer analysis, the estimated genome sizes for Frijol Bayo and W6 15578 are 684 Mb and 676 Mb, respectively (Supplementary Fig. 1), in agreement with flow cytometry which estimated both Frijol Bayo and W6 15578 at 682 Mb. Repetitive sequence content in Frijol Bayo was comparable to that of P. vulgaris G19833 (53.0% vs 55.1%, Supplementary Data 2) whereas W6 15578 had a higher repetitive sequence content (64.2%) than Frijol Bayo and P. vulgaris G19833; the majority of this increase was attributable to unclassified repetitive sequences. It is unlikely that the wild W6 15578 genome has substantially more repetitive sequences than the domesticated landrace Frijol Bayo. P. vulgaris G19833 and Frijol Bayo were sequenced and assembled with long read PACBIO reads while W6 15578 was assembled with linked short reads and a different assembler, thus these differences in genome size and repetitive sequence content may be technical in nature. The Frijol Bayo and W6 15578 genomes showed no large rearrangements or gaps outside of the heterochromatic, pericentromeric regions (Fig. 1c and Supplementary Figs. 2 and 3). Annotation of the Frijol Bayo and W6 15578 genomes resulted in a similar number of high-confidence genes for Frijol Bayo (27,538) and W6 15578 (27,095) and robust representation of the BUSCO orthologs with 94.6% and 95.7% complete BUSCO orthologs, respectively (Table 1 and Supplementary Data 1).

Tepary bean is a sister species to the common bean. Nucleotide substitutions at synonymous sites (Ks) are selectively neutral and are used as a molecular clock to date the divergence time of closely related species. Individual peaks in Ks distributions of orthologous genes between two species follow a Gaussian (normal) distribution and can be detected using mixture model analysis. Synonymous substitutions between pairs of orthologous genes in cultivated (Frijol Bayo) and wild (W6 15578) tepary bean, and other legume crops, including P. vulgaris (common bean), Glycine max (soybean), and Vigna unguiculata (cowpea), were used to estimate the divergence time between these genomes. Mixture model analysis of Ks distribution of orthologous genes between common bean and soybean revealed a single peak with a mean Ks value of 0.32 (Fig. 1d and Supplementary Data 3). Using the widely reported age of divergence between common bean and soybean of ~19.2 Mya, and the mean Ks value of the corresponding peak (0.32), we extrapolated the rate of silent mutations for leguminous species to be $8.3 \times 10^{-9}$ substitutions/synonymous site/year (see "Methods" section). The age of divergence between Frijol Bayo and W6 15578 was estimated to be 0.77 million years based on this calibrated synonymous substitution rate. The geometric mean of the peak observed in the Ks distribution ($K_s = 0.013$) derived from a mixture model analysis of Ks distributions revealed the presence of a major peak at $K_s = 0.013$ (Fig. 1d and Supplementary Data 3). Based on similar comparisons, the divergence of tepary bean (Frijol Bayol) from other warm-season legume crops, such as common bean, cowpea, and soybean, was estimated to have occurred 2.1, 10.1, and 19.1 Mya, respectively (Fig. 1d and Supplementary Data 3). Similarly, the ages of divergence between cowpea and common bean, and cowpea and soybean, were estimated to be 10.3 and 21.2 Mya, respectively.
Both Frijol Bayo and W6 15578 shared substantial orthology with *P. vulgaris* as reflected in the species tree and shared gene families (Fig. 1e and Supplementary Fig. 4). As expected, with a limited divergence time from *P. vulgaris*, gene family expansion and contraction frequency were low, 241 genes expanded and 554 contracted in Frijol Bayo relative to the cultivated *P. vulgaris* G19833 reference genome (Fig. 1e). Gene ontology enrichment of genes expanded in Frijol Bayo and W6 15578, relative to *P. vulgaris* G19833, included genes involved in chitin binding, chitinase activity, cell wall macromolecule catabolic/metabolic processes, and amino sugar metabolic processes (Supplementary Fig. 5).
Synteny analysis between *P. vulgaris* G19833 and the *P. acutifolius* Frijol Bayo and W6 15578 genomes revealed 403 collinear blocks ranging from 5 to 2,684 orthologous pairs with 90.7% of W6 15578 and 91.3% of the Frijol Bayo orthologous pairs present in collinear blocks with *P. vulgaris* (Fig. 1c). This high level of collinearity between *P. vulgaris* and *P. acutifolius* is likely due to the short divergence time of 2.1 Mya. Breaks in synteny relative to *P. vulgaris* G19833 include intrachromosomal rearrangements on chromosomes 2, 3, and 9 previously identified by Gujarati-Verma et al.\(^9\) as well as on most other chromosomes in the pericentromeric regions (Fig. 1c and Supplementary Fig. 3). Introggression between common and tepary bean is challenging to hybrid sterility\(^13,20\). While we did not detect large-scale interchromosomal structural rearrangements, the small structural variations, as well as the larger internal rearrangements present on chromosomes 2 and 9 could contribute to difficulties in creating fertile hybrids\(^20\).

### Tolerance in desert-adapted tepary bean to moderate elevated temperature

Access to the tepary genome sequence will enable the discovery of genetic factors associated with climate resilience. While high night-time temperature is known to decrease crop yield in common bean when it occurs during micro- and megasporogenesis, tepary bean tolerates elevated temperature. To rule of possible reasons of differential heat tolerance, Frijol Bayo and Amadeus-77, a Middle American small red common bean which shows heat tolerance relative to other common beans, were exposed to moderate night-time temperature stress (32 °C/27 °C, day/night), and gene expression profiles were compared during the first 24 h (hrs) of night-time stress induction. Not only did the absolute numbers of differentially expressed genes differ between Frijol Bayo and Amadeus-77 across the time course, but the overall temporal pattern and magnitude of genes impacted differed between these sister species (Supplementary Fig. 6), suggesting substantial differences in gene regulatory responses to elevated temperature.

The heat shock response is evolutionarily conserved across all kingdoms of life, and entails a suite of physiological, molecular, and genomic responses that enable the repair of cellular machinery. Activation of transcriptional cascades following heat stress is common in a wide number of plant species as exemplified in *Arabidopsis thaliana* where 21 heat shock transcription factors (HSFs) and >2500 genes were differentially expressed following heat stress\(^21\). Within 1 h of the heat stress treatment, 15 genes encoding heat shock-related proteins (HSPs), including the master regulator HSF2A, were upregulated in Amadeus-77 while one was downregulated (Fig. 2a; Supplementary Figs. 7 and 8). Throughout the 24-h time course, upregulation rather than downregulation of HSPs was observed in Amadeus-77. This contrasted with Frijol Bayo, in which only three heat shock-related genes were downregulated at 1 hr. Even after 6 and 12 h of heat stress, the predominant response in Frijol Bayo was down- rather than upregulation of heat shock-related genes. The highest number of upregulated HSPs in Frijol Bayo was seven at 12 h after stress while the majority of genes were downregulated. The substantial induction of the heat shock response in common but not tepary bean to an elevation of day- (32 °C) and night-time (27 °C) temperature is consistent with the evolution and adaptation of tepary bean to the hot Sonoran Desert environment. In a second heat stress experiment, exposure of Frijol Bayo to prolonged and elevated temperatures (36 °C day/32 °C night) in which leaves emerged and developed under heat stress conditions, upregulation of 15 HSPs was observed in tepary bean (Supplementary Data 4 and 5) suggesting that relative to common bean, tepary bean has an elevated temperature threshold for induction of the canonical heat shock response.

### Cell cycle and protective molecules are induced following exposure to moderate heat stress

A moderate night-time temperature (27 °C) is sufficient to impact yield in common bean but not tepary bean. Interestingly, while Frijol Bayo did not respond with a canonical heat shock response following exposure to moderate heat stress (27 °C night-time temperature), it is clear Frijol Bayo perceived the heat stress within 1 h based on differential gene expression. We observed enrichment of gene ontology (GO) terms associated with phosphorelay signal transduction system (GO:0000160), negative regulation of cellular process (GO:0048523), cell cycle arrest (GO:0007050), and negative regulation of cell cycle (GO:0045786) (Fig. 2b and Supplementary Data 6, 7). Eight differentially expressed cell cycle pathway genes were identified at 1 h in Frijol Bayo; two upregulated genes (*Phacu.CVR.003G292700* and *Phacu.CVR.003G347000*) which encode a CDK (cyclin-dependent kinase) inhibitor and negative regulator of cell division\(^22\) (Fig. 2b and Supplementary Fig. 9). Six orthologs of E2F target genes that control transcription of downstream cell cycle genes were downregulated at 1 h in Frijol Bayo including *Phacu.CVR.006G085000* (CDC6A, CELL DIVISION CONTROL 6), *Phacu.CVR.009G156300* (CDT1a), *Phacu.CVR.001G020700* (ORC2, ORIGIN RECOGNITION COMPLEX 2), *Phacu.CVR.002G305100* (FAS2), *Phacu.CVR.003G169400* (MCM2, MINICHROMOSOME MAINTENANCE 2), *Phacu.CVR.011G048900* (MCM4, MINICHROMOSOME MAINTENANCE 4), and *Phacu.CVR.009G085300* (ETG1, E2F TARGET GENE 1). This combination of up- and downregulated genes in Frijol Bayo is suggestive of a G1-S phase transition arrest in the cell cycle that may result when low-energy homeostasis induces KIP-RELATED PROTEINS\(^23\). During G1,

---

**Table 1** Genome assembly and annotation metrics of *Phaseolus* species.

|                | *Phaseolus acutifolius* Frijol Bayo, Cultivated | *Phaseolus acutifolius* W6 15578, Wild | *Phaseolus vulgaris* Chaucha chuga v2.1 |
|----------------|-----------------------------------------------|-------------------------------------|----------------------------------------|
| Total assembly length (bp) | 512,626,114 | 661,870,301 | 537,218,636 |
| Total contig number | 667 | 29,934 | 1044 |
| Contig N50 size (L50 number) (bp) | 6,178,816 (25) | 47,040 (3,492) | 1,885,876 (73) |
| Total Scaffold number | 155 | 609 | 478 |
| Scaffold L50 size (N50 number) (bp) | 45.2 Mb (5) | 42.6 Mb (8) | 49.7 Mb (5) |
| Percent assembly anchored to 11 pseudomolecules | 99.00% | 70.54% | 95.67% |
| Number of high-confidence genes | 27,538 | 27,095 | 27,433 |
| Number of high-confidence gene models | 50,635 | 40,833 | 36,995 |

*Synteny analysis between *P. vulgaris* G19833 and the *P. acutifolius* Frijol Bayo and W6 15578 genomes revealed 403 collinear blocks ranging from 5 to 2,684 orthologous pairs with 90.7% of W6 15578 and 91.3% of the Frijol Bayo orthologous pairs present in collinear blocks with *P. vulgaris* (Fig. 1c). This high level of collinearity between *P. vulgaris* and *P. acutifolius* is likely due to the short divergence time of 2.1 Mya. Breaks in synteny relative to *P. vulgaris* G19833 include intrachromosomal rearrangements on chromosomes 2, 3, and 9 previously identified by Gujarati-Verma et al.\(^9\) as well as on most other chromosomes in the pericentromeric regions (Fig. 1c and Supplementary Fig. 3). Introggression between common and tepary bean is challenging to hybrid sterility\(^13,20\). While we did not detect large-scale interchromosomal structural rearrangements, the small structural variations, as well as the larger internal rearrangements present on chromosomes 2 and 9 could contribute to difficulties in creating fertile hybrids\(^20\).*
the cell commits to cell division or exits the cell cycle. Such control enables the cells to respond to environmental cues, such as nutrition availability or abiotic stresses, by inhibiting cell growth and development in order to conserve energy. By contrast, only one cell cycle gene Phvul.011G017500, a ortholog of Arabidopsis Recognition Complex subunit 3 (ORC3) that initiates DNA replication at S phase, was upregulated in Amadeus-77 at 1 h.

At 3 h after stress induction in Frijol Bayo, GO terms related to cell redox homeostasis (GO:0045454), cellular homeostasis (GO:0019725), and homeostatic process (GO:0042592) were enriched (Supplementary Data 7) including seven genes from the glutaredoxin family (GRXs) and two genes from the thioredoxin (TRXs) family; all were upregulated. Glutaredoxins (GRXs) are small ubiquitous glutathione (GSH)-dependent oxidoreductases involved in oxidative stress responses. All of the differentially expressed GRXs are of the CC-type. This included a homolog (Phacu.CVR.011G148000) of Arabidopsis GRXC9 that is induced by salicylic acid (SA) and is part of the early SA-induced non-expressor of pathogenesis-related (PR) genes 1 (NPR1)-independent pathway in which GRXC9/ROXY19 interacts with class II TGA transcription factors to induce genes with antioxidant and detoxifying activities. In Arabidopsis, 10 other CC-type GRXs also interact with TGAs. Among these are ROXY 10, and Roxy 21 for which we found homologs in Frijol Bayo. Redox genes that were upregulated at 6 h and peaked at 12 h after stress induction included four genes continuously upregulated from 3 to 12 h after stress induction; these genes were homologs of Arabidopsis ACHT4, ROXY 21, and ROXY 10.

Four trehalose synthesis-related genes were upregulated at 3 h of heat stress in Frijol Bayo: Phacu.CVR.003G017200 [homolog of Arabidopsis TPS6], Phacu.CVR.003G017200 [homolog of Arabidopsis TPS11], Phacu.CVR.009G053300, and Phacu.CVR.002G288900 (both homologs to Arabidopsis TPS6-6-PHOSPHATASE SYNTHASE (TPS6), Phacu.CVR.003G183300 (homolog of Arabidopsis TPS11), Phacu.CVR.009G053300, and Phacu.CVR.002G288900 (both homologs to Arabidopsis TPS6-6-PHOSPHATASE PHOSPHATA E J (TPPI))]. In contrast, trehalose biosynthesis
membranes and proteins from degradation resulting from heat, downregulated at 1 h post-stress. Previous studies have demonstrated stress. Furthermore, supplemental trehalose can rescue tolerance mutants had reduced survival rates following heat stress and trehalose accumulates in response to heat stress and trehalose may act instead as a regulatory molecule for osmoprotection. In Arabidopsis, organisms with high trehalose concentration (including some species that survive in extreme habitats). In Arabidopsis, biochemical mechanisms of thermotolerance in stress will further our understanding of the physiological and biochemical mechanisms of thermotolerance in Phaseolus vulgaris.

Expansion of the NB-ARC gene family in Phaseolus vulgaris. Plants evolved a modular-domain protein family as a critical component of resistance against biotic pathogens. Central to this family is a canonical NB-ARC domain bordered by C-terminal leucine-rich repeat domains (NLR proteins), and often an N-terminal coiled-coil or TIR domain. Biotic resistance can be activated when a pathogen-specific effector or an effector/host protein complex molecule interacts with one of the canonical domains. Furthermore, other NLR proteins contain a non-canonical integrated domain (ID) that interacts with the effector. The implication of this discovery is that a much greater repertoire of domains than the canonical NLR domains, can sense the presence of an effector in the plant cell. Given that tepary originated from an arid environment, and that biotic pathogens often favor humid conditions to thrive, a comparison of the NB-ARC repertoire in tepary and common bean, a temperate sister species that have higher water requirements for growth, may shed light on the evolution of this protein family under an abiotic stress environment.

Because of the importance of the NB-ARC domain in activating the defense response, genome-wide searches for potential NLR resistance genes are based on the presence of this domain. The Frijol Bayo genome has 200 genes with an NB-ARC domain (Supplementary Data 8 and Fig. 3), of which 155 encode NLR proteins. By comparison, the common bean genome (Chaucha Chuga, G19833) encodes 330 NB-ARC domain proteins (Supplementary Data 9), and of these, 248 encode NLR proteins. To assess whether the sizes of tepary and common bean NB-ARC families evolved differently than other gene families, we compared the size of 80 transcription factor (TF) families in these two genomes. The total number of TFs across families in tepary (n = 1818; Supplementary Data 10) and common bean (n = 1881; Supplementary Data 11) were similar, and a high correlation (r = 0.99) was observed for the number of members of the 80 transcription factor families between the two genomes (Supplementary Data 12). Therefore, unlike the size of transcription factor families which are conserved within the Phaseolus genera, the NB-ARC family differs in size between these two Phaseolus species suggesting unique evolutionary dynamics for this gene family.

Multiple domain structures are possible because of the modular nature of NB-ARC proteins. The Frijol Bayo genome encodes 20 different domain structures, while the G19833 common bean genome encodes 23 (Supplementary Data 13). The primary difference between the two genomes is the greater number of common bean proteins encoding an NB-ARC/LRR domain structure. For Frijol Bayo, 34 NB-ARC genes encode proteins with an ID, compared to 40 in common bean (Supplementary Data 14). Two of the tepary and common bean ID, zf-BED (P102892) and zf-RVT (P139666), are frequently found in NLR genes from other species, and were noted in a previous assembly of common bean. Of the 58 IDs identified collectively between tepary and common bean, only eight were shared between the two species. This points to the unique evolution of tepary NB-ARC genes relative to the common bean. Further, of the 58 IDs found in Phaseolus spp., only nine were integrated into NB-ARC proteins in other plant genomes. This discovery supports previous observations that many IDs are associated with a large array of cellular functions that have fused with NB-ARC domains to develop potential sensors of pathogen effectors.

Early genetic experiments demonstrated that resistance genes are tightly linked within the genome, while subsequent research in many plant species demonstrated that physical clusters of NB-ARC genes often contain functional NLR genes. Clustering of NLR genes is also a feature of the Phaseolus genus where a total of 21 clusters were observed in tepary and 25 in common bean with 63% and 73%, respectively, of the NB-ARC genes located in...
clusters (Supplementary Data 15). In most cases, each tepary NB-ARC cluster has a corresponding cluster in common bean with genes of similar domain structure: IA = NL; 2B = TNL; 3A = RNL. While some of the clusters are of equal size in the two genomes (2B, 3A), other clusters (4A, 10A, 11C) have greatly expanded in common bean relative to tepary bean. As observed previously for common bean chromosomes Pv0449 and Pv1150, many clusters are located on the ends of the chromosomes. The only exceptions are NL clusters 2A and 10D that are located in the recombination poor heterochromatic region. For common bean, many genetically mapped resistance factors are associated within clusters51, yet the only functionally proven R gene in common bean with a NB-ARC domain (Phvul.005G031200) resides as a singleton52.

Phylogenetic analyses uncovered several evolutionary patterns in the NB-ARC gene families in the Frijol Bayo and G19833 genomes (Fig. 3). Many of the Frijol Bayo genes (Fig. 3; red diamonds) are paired with a G19833 ortholog. Most of these are also best reciprocal BLASTP hits (Supplementary Data 16), especially those with high identity between the two orthologs. As often seen in plant species, the TNL and non-TNL NB-ARC genes form two distinct phylogenetic clusters. The Pv04 and Pv11 CNL clusters group together at a higher taxonomic level suggesting an ancestral relationship. Two examples of NB-ARC gene expansion in Frijol Bayo were noted, clusters 8B and 10B with the NNL or CNNL domain structure, and five genes (between the NL 1A and NL 7A) that have short (<88 amino acids) NB-ARC sequences. The greatest expansion in G19833 occurred at the distal ends of Pv04 and Pv11, respectively. The first 21 common bean NB-ARC genes in the CNL4A-E group are specific to cluster 4A. The same Frijol Bayo cluster only consists of six NB-ARC genes. On Pv11, clusters 11G and 11H are entirely (11G) or almost entirely (11H) represented by G19833 genes at the distal end of the long arm of Pv11. This expansion supports the suggestion that the ends of chromosomes are hot spots for resistance gene evolution via expansion50. Contrary to these clusters, the RNL cluster 3B is nearly a perfect one-for-one match between the two species.

**Population structure, domestication, and adaptation.** A total of 28,750 physically mapped single nucleotide polymorphisms (SNPs; MAF ≥ 0.05) were identified among 425 wild, weedy, and cultivated tepary genotypes. These represent the Tepary Diversity Panel, a collection of genotypes representing the entire geographic distribution of the wild and cultivated tepary beans (Fig. 4a). A principal component analysis (PCA) using this variant data set (Fig. 4b) showed that much of the variation was accounted for by PC1 (25.6%) while the first three PCs account for 45.2% of the variation. The neighbor-joining (NJ) tree analysis with n = 2317 SNPs with LD values <0.1 clearly separated the cultivated and wild/weedy genotypes (Fig. 4c). Cultivated genotypes from two geographically distinct regions, the Sonoran...
Desert region and the Pacific Coast of Central America, were closely related phylogenetically. Climatically, these two regions experience high, dry desert heat (Sonora) or high, humid coastal heat (Central American Pacific coast) suggesting that within the narrow diversity within the cultivated population, genetic variability exists for adaptation to these two distinct heat stress environments.

A subset of 55 geographically diverse genotypes representing the geographical range in the Tepary Diversity Panel was selected for whole-genome resequencing (Supplementary Data 17). A STRUCTURE analysis suggested three subpopulations (Fig. 4d). The narrow genetic base of cultivated tepary was best demonstrated by the observation that among 3,377,439 SNPs, 2,247,877 (66.6%) SNPs were polymorphic among the wild yet monomorphic among the cultivated genotypes (Supplementary Table 1). Overall, genes, 5003 (16%) were monomorphic among both wild and cultivated genotypes.

Population genomic analyses revealed signatures of domestication and adaptation in tepary bean. As demonstrated in the phylogenetic analyses, the wild population was more diverse [$\pi$ (per bp) $= 0.0018$; $\Theta = 0.00013$] than the cultivated population [$\pi$ (per bp) $= 0.0007$; $\Theta = 0.0005$]. Domestication signatures were identified\(^{16}\) using the $\pi_{wild}/\pi_{domesticated}$ ratio over 100 kb/10 kb sliding windows (Fig. 4e) and focusing on the windows in the 0.1% and 1.0% tails of the bootstrap distribution. Chromosomes Pa07 and Pa11 were essentially monomorphic between the wild and cultivated populations, while the two populations were highly polymorphic for chromosome Pa05. Chromosomes Pa02 and Pa03 were polymorphic across the euchromatic tails, and the remaining chromosomes exhibited high $\pi$ ratios isolated to narrow genomic regions. A total of 100 genomic blocks, consisting of neighboring windows in the 0.1% bootstrap tail, contained 591 genes. A reciprocal BLASTP analysis of the 591 tepary beans isolated to narrow genomic regions. A total of 100 genomic blocks, consisting of neighboring windows in the 0.1% bootstrap tail, contained 591 genes. A reciprocal BLASTP analysis of the 591 tepary bean domestication genes found that only 32 were homologs of the common bean Middle American domestication genes ($n = 1835$\(^{18}\)), and four were homologs of the common bean Andean domestication genes ($n = 748$\(^{18}\)). The comparison found 20 of the tepary/Middle American shared domestication genes were located on chromosome nine of each genome. All 20 of these shared homologs were all located in shared domestication blocks of the two genomes. Overall, this demonstrates that domestication involved unique sets of genes in each of these two Phaseolus genomes.

The total length of the cultivated tepary genome blocks with low diversity, and presumed to be associated with domestication, was 8.50 Mb. Furthermore, a total of 242 out of 591 genes located within these domestication blocks were monomorphic relative to their wild orthologs (Supplementary Data 18). Collectively, those blocks containing monomorphic genes covered 1.75 Mb. Compared with the P. vulgaris G19833 genome, a greater proportion of the wild tepary bean genome became monomorphic during domestication. Gene ontology enrichment analyses of the 242 genes with low diversity among cultivated genotypes identified genes associated with ion transport enriched (Supplementary Fig. 10). Ion transport regulators were also associated with the domestication of salt-tolerant weedy rice\(^{23}\) and maize\(^{24}\).

Adaptation occurs when individuals from the initial domestication environment are grown in other environments and diversity is selected from standing variation or occurs via de novo mutation. To identify such regions, SNPs in which one allele was fixed in all wild genotypes and the alternate allele was fixed in all landraces were investigated. All of the fixed SNPs in the 0.1% bootstrap tail were located on Pa05 while the second region of fixed SNPs in the 1% tail was located on Pa08 (Fig. 4f). Of the 98 Pa05 genes associated with these fixed SNPs, GO enrichment revealed an association with inositol phosphatase activity (Supplementary Fig. 11). The role of inositol phosphatase in plant adaptation under abiotic stresses is well documented\(^{35}\). Inositol phosphatase regulates/catalyzes the network of inositol phosphate and phosphoinositide signaling pathways which results in altering other stress response pathways including calcium release and abscisic acid signaling which have important roles in plant growth in stressful environments.

**Discussion**

Tepary bean is an orphan crop distinguished from other pulse crops by its abiotic stress tolerance, primarily thought to result from its origin and evolution in the hot and dry Sonoran Desert. Access to the genome of tepary bean revealed exceptional synteny and orthology with common bean yet highlighted mechanisms of resilience. Under a moderate night-time heat temperature which is limiting to yield in common bean but not tepary bean, gene expression analyses suggest that tepary bean arrests the cell cycle and synthesizes compounds with redox potential as well as the disaccharide trehalose that has been associated with abiotic stress tolerance. However, tepary bean is not immune to the impacts of heat stress, and exposure of tepary bean to prolonged and elevated night-time heat stress resulted in upregulation of genes encoding the canonical heat shock response. Evolution in arid conditions most likely resulted in reduced biotic pressure and as a consequence, the classical disease resistance genes family evolved differentially in common and tepary bean.

Currently, the common bean is the most important legume source of plant protein for direct human consumption. However, common bean production is increasingly limited geographically due to its sensitivity to high temperatures and its susceptibility to diseases due to climate change. In fact, it is predicted that by 2050, not only will the majority of common bean production areas in southeastern Africa be unsuitable, but the nutritional content of common bean will also be reduced\(^{36}\). Interestingly, the protein and nutritional composition of tepary are similar to its sister species, P. vulgaris\(^{37}\). The close sister-species relationship between P. vulgaris and P. acutifolius provides a powerful framework for the improvement of tepary agronomic and disease resistance traits through introgression from common bean, and the improvement of common bean abiotic stress tolerance via introgressions from tepary bean. For example, work nearly a century ago in common bean showed that the presence of recessive alleles at the Fin locus conferred determinate growth habit\(^{38}\). Fin encodes an ortholog of the Arabidopsis thaliana TERMINAL FLOWER 1 (PvTFL1) for which known mutations lead to a determinate growth habit\(^{39,40}\). Phac.001G234100, a P. acutifolius homolog of PvTFL1, is 99% identical to the P. vulgaris indeterminate allele, providing a target to alter the growth habit of tepary via mutagenesis or gene editing. Tepary bean also provides a resource for allele and trait mining across the Phaseolus genus as exemplified by the P (Pigment) locus, a basic helix-loop-helix (bHLH) transcription factor, shown to control seed coat color in common bean\(^{41}\). In white-seeded Frijol Bayo, a ~56 kb inversion involving six genes disrupts the putative tepary P gene whereas the dark-seeded W6 15578 P ortholog lacks this inversion (Supplementary Fig. 12). Given the expansion of crop production into areas marginalized by warming and drying climates, leveraging knowledge and genetic resources and applying comparative genomics across these Phaseolus species provides a powerful approach to accelerate the improvement of climate-adapted bean varieties to address climate change.

**Methods**

**Germplasm description**. Frijol Bayo, was collected in a market in Cordoba, Veracruz, Mexico in 1951 by Howard Scott Gentry. It is a white-seeded cultivated tepary bean that is photoperiod insensitive, has an indeterminate growth habit, has a hundred seed weight of 12.3 g, and exhibits broad adaptation. Frijol Bayo shows resistance to common bacterial blight\(^{42}\) caused by Xanthomonas axonopodis, is
drought tolerant\(^2\), heat tolerant\(^6\), and is resistant to leaf hoppers (Empoasca kramerii)\(^3\). Frijol Bayo, and another accession, G40020, were used as parents in the early challenge introgression of the common bean highblight resistance from tepary bean to common bean\(^1\), using embryo rescue, which subsequently resulted in the release of the VAX 1 to VAX 6 lines\(^4\). This broad and durable common bacterial blight resistance has been extensively introgressed into both common bean gene pools, aided by the use of the S191 and BC420 markers\(^5\). A core segregating line population was recently developed from the cross Frijol Bayo and G40022. These lines facilitate the investigation of the genetic architecture of seed traits, agronomic characteristics, and disease response in tepary bean\(^6\).

Priorization of the Frijol Bayo was completed over three generations of single-seed descent in a greenhouse at the USDA-ARS-TARS in Mayaguez, Puerto Rico, from 2015 to 2017. The original G40001 accession was received from the International Center for Tropical Agriculture Research. The purified G40001 seed lot was further increased in a greenhouse in the Fall of 2019. A total of 5000 bulked seed from the 2019 increase were provided to the USDA GRIN and assigned a unique accession name and number (G40001-Seq; PI 692629, https://www.ars-grin.gov/) to differentiate it from the original accession. Purified G40001 were dark adapted on two occasions in a dark, air-conditioned room for one day, and young trifoliolate leaf tissue was then harvested and immediately frozen in liquid nitrogen, and subsequently stored at \(-80^\circ\text{C}\) for high molecular weight DNA extraction at the University of Arizona.

W6 15578 is a wild P. acutifolius accession obtained from the Western Regional Plant Introduction Station of the USDA. Originally collected by O.W. Norvell in Mexico in 1955, it now has the Plant Introduction number PI 638833 (https://www.ars-grin.gov/). It is photoperiod sensitive, has an indeterminate growth habit, and has large seeds, and pods shatter at maturity. It was used as a donor parent in the development of an interspecific hybrid population to improve abiotic stress tolerance in common bean\(^8\). W6 15578 was one of the parents, along with PI 430219, of the tepary bean mapping population, BR-06, developed and genotyped at the University of Saskatchewan\(^9\).

**Genome sequencing and assembly.** *P. acutifolius* Frijol Bayo was sequenced using a whole-genome shotgun sequencing strategy and standard sequencing protocols. Illumina fragment reads were generated using the Illumina X10 and HiSeq platforms and PacBio reads were generated using the RS platform. reads (101.28× sequence coverage) using the MECAT assembler (v1.0)\(^14\) and a total of 149 homozygous SNPs and 10,468 homozygous INDELs were corrected in the release consensus. To annotate the Frijol Bayo genome, species-specific custom repeat library (CRL) was created using RepeatModeler (v1.0.8; http://repeattaskmaker.org/), and putative protein-coding genes identified as repeats by RepeatModeler were identified and excised from the CRL by searching against a curated library of plant protein-coding gene sequences using ProtExcluder\(^2\) (v1.1). Output from ProtExcluder was then combined with the RepBase\(^76\) (v20150807) Viridiplantae repeats to create a final CRL and the final pseudomolecules were masked using RepeatMasker (v4.0.6; http://repeattaskmaker.org/) with the options: -s \(-nolow\_noisy\_is\_gf\). Reads from a single Oxford Nanopore library were used to confirm order where discrepancies occurred with the order of the cultivated assembly. Final assembly statistics are in Supplementary Table 7.

**Genome size estimates and quality assessments.** Processed Illumina short-read data from Frijol Bayo and W6 15578 were provided to Jellyfish v2.2.67 to generate k-mer \((K = 21)\) frequency distribution with \(C – m\) \(\min\) – \(\min\) quality = 25 parameters. The output histograms were analyzed for findGSE (v0.1.0)\(^10\) to estimate the genome size, heterozygosity, and repeat fraction. The presence of conserved plant orthologs was evaluated using BUSCO\(^84\) (v3.0.2b) in Embryophyta database 9 and an \(E\)-value of \(1e^{-05}\).

**Genome annotation.** To annotate the Frijol Bayo genome, species-specific custom repeat library (CRL) was created using RepeatModeler (v1.0.8; http://repeattaskmaker.org/), and putative protein-coding genes identified as repeats by RepeatModeler were identified and excised from the CRL by searching against a curated library of plant protein-coding gene sequences using ProtExcluder\(^2\) (v1.1). Output from ProtExcluder was then combined with the RepBase\(^76\) (v20150807) Viridiplantae repeats to create a final CRL and the final pseudomolecules were masked using RepeatMasker (v4.0.6; http://repeattaskmaker.org/) with the options: -s \(-nolow\_noisy\_is\_gf\). Reads from a single Oxford Nanopore library were used to confirm order where discrepancies occurred with the order of the cultivated assembly. Final assembly statistics are in Supplementary Table 7.

**Genome size estimates and quality assessments.** Processed Illumina short-read data from Frijol Bayo and W6 15578 were provided to Jellyfish v2.2.67 to generate k-mer \((K = 21)\) frequency distribution with \(C – m\) \(\min\) – \(\min\) quality = 25 parameters. The output histograms were analyzed for findGSE (v0.1.0)\(^10\) to estimate the genome size, heterozygosity, and repeat fraction. The presence of conserved plant orthologs was evaluated using BUSCO\(^84\) (v3.0.2b) in Embryophyta database 9 and an \(E\)-value of \(1e^{-05}\).

**Genome annotation.** To annotate the Frijol Bayo genome, species-specific custom repeat library (CRL) was created using RepeatModeler (v1.0.8; http://repeattaskmaker.org/), and putative protein-coding genes identified as repeats by RepeatModeler were identified and excised from the CRL by searching against a curated library of plant protein-coding gene sequences using ProtExcluder\(^2\) (v1.1). Output from ProtExcluder was then combined with the RepBase\(^76\) (v20150807) Viridiplantae repeats to create a final CRL and the final pseudomolecules were masked using RepeatMasker (v4.0.6; http://repeattaskmaker.org/) with the options: -s \(-nolow\_noisy\_is\_gf\). Reads from a single Oxford Nanopore library were used to confirm order where discrepancies occurred with the order of the cultivated assembly. Final assembly statistics are in Supplementary Table 7.
planted in SUREMIX (SURE, Galesburg, MI, USA) in gallon size pots. Plants were grown under a 12 h photoperiod, 60% relative humidity, 29 °C day/20 °C night temperature maintained at maximum temperature of 33 °C, and minimum temperature over 1 h) in a controlled growth chamber with light intensity reaching full capacity of 500 μmol m⁻² s⁻¹ over one hour and dimming to complete darkness over one hour to mimic sunrise and sunset, respectively. A total of 12 developmental tissues were collected with sampling occurring between 10:00 and 12:00 h using the time point variation due to flowering. Each biological replicate consisted of a pool of samples from three plants with three biological replicates for each developmental stage. Total RNA was extracted from all tissues using the Spectrum Plant Total RNA Kit (Sigma-Aldrich, St. Louis, MO). Quality and quantity assessments using a Qubit fluorometer (Invitrogen INC. Carlsbad, CA) and Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA). Libraries were prepared using the KAPA RNA HyperPrep Kit (KK8540; KAPA Biosystems, Roche Sequencing Solutions, Pleasanton, CA, USA) using NEBNext oligos. Sequencing was performed on an Illumina HiSeq 4000: one biological replicate of all 12 tissues was sequenced using the 150 nt pair-end read configuration to achieve ~50 million read pairs for use in genome annotation. The other two biological replicates were sequenced to 50 nt in single-end configuration to generate greater than 25–30 million reads.

To determine expression abundances in the Frijol Bayo gene atlas, raw RNA-seq read quality was assessed using FastQC (v 0.11.5; https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and adapters and low-quality bases were removed using Cutadapt(77) (v 1.16) with the following parameters: q -20 n -2 -trim-n --minimum-length 31. For expression abundance estimates, paired-end reads were trimmed to 50 nt, and all samples were subsampled at 20 million reads using seqtk (v 1.0.182) (https://github.com/ih3/seqtk/). Reads from all three replicates were aligned to the Frijol Bayo reference genome using HISAT2 (v 2.1.0). Expression abundances in the form of Fragments per Kilobase of transcript per Million mapped reads (FPKM) were quantified using Cufflinks(80) (v 2.2.1). Pairwise Pearson’s correlation was calculated for all samples using the “cor” function in R (v 3.5.1; https://www.project.org/). P-values were calculated using R package “psych.” Principal Component Analysis (PCA) was conducted using the “prcomp” function in R (v 3.5.1; Supplementary Figure 15).

To understand gene expression under moderate heat stress, Frijol Bayo and Adameus-77 seeds were surface sterilized using 2% sodium hypochlorite for 10 min, rinsed, and soaked in Petri dishes at 29 °C/20 °C day/night temperature for 24 h (Frijol Bayo) or 72 h (Adameus-77), and then planted in SUREMIX (SURE, Galesburg, MI, USA) in 4 x 4 x 6 inch pots in a growth chamber. Plants were grown under controlled growth chamber conditions with a 12 h photoperiod, 60% relative humidity, 29 °C day/20 °C night temperature for control, and 32 °C day/27 °C night temperature for heat stress (ramping up to maximum temperature and ramping down to minimum temperature over one hour), light intensity reaching full capacity of 500 μmol m⁻² s⁻¹ over 1 h and dimming to complete darkness over 1 h to mimic sunrise and sunset, respectively. At 11 days before anthesis, leaves were sampled at five time points for 24 h starting at programmed sunset: 1, 3, 6, 12, and 24 h following programmed anthesis. To induce heat stress, from half an hour before sunset, the temperature was ramped up and reached the night-time relative humidity, 29 °C day/20 °C night temperature until 11 days prior to anthesis at which they were shifted to a 36 °C day/32 °C night temperature regime. At flowering, leaflets of 7th to 8th trifoliate, which either completely developed or developed under heat stress subsequent to its emergence, were sampled for RNA. Total RNA was isolated as described for the developmental gene atlas. Illumina TruSeq stranded mRNA libraries were generated and sequenced on an Illumina HiSeq 4000 in single-end mode (50 nt reads). Raw reads were assessed for quality using FastQC (v 0.11.8; https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and adapters and low-quality bases were removed using Cutadapt(77) (v 2.2.1) and read counts per gene were calculated using HTSeq(92) (v 0.11.2). Read counts were then used for differential gene expression analysis using the DESeq2 package(91) (v 1.22.2) within R (v 3.5.0). Within DESeq2, the contrasts function was used to assess whether log2 fold-change was equal compared to the log2 fold-change along with an adjusted P-value cutoff of 0.01 was used for extracting differentially expressed genes. Furthermore, the “lfcshrink” function within DESeq2 was used to help restrain the high log2 fold changes of genes that had low expression values.

Evolutionary analyses. Orthologous pairs between P. vulgaris GI9833, and cultivated (Frijol Bayo) and wild (W6 15578) P. acutifolius were identified by reciprocal best BLASTP hits (e-value cutoff = 10⁻10). MCScanX(82) (v2) with default options was used to scan for collinear blocks between the assemblies. The collinear blocks were visualized using Synvisio (https://synvisio.github.io/). To determine the orthologous and paralogous groups in the Phaseoleae tribe, OrthoFinder(299) (v2.2.7) was run with default search parameters and the representative proteins from A. thaliana (TAIR10)85, Medicago truncatula (Mt4.0v1)100, Cajanus cajan (v1)91, Glycine max (Wm82.a2.v1)101, Vigna unguiculata (v1.1)103, P. acutifolius (v2.1), P. acutifolius Frijol Bayo, and P. acutifolius W6 15578. The R package ‘Upset’ (v1.4.0) was used to visualize the results. An ultrametric tree was constructed using the OrthoFinder rooted species tree with the ‘make_ultrametric.py’ python script in OrthoFinder. The root age was estimated to be 120 million years ago based on data from two separate studies102,103. CAFE(104) (v4.1.2; –filter p 0.01 -r 10000) was used with the ultrametric tree and the identified gene families from the root to characterize the expanded and contracted families. The Biocore R package ‘topGO’ (v2.36.0; https://bioconductor.org/packages/release/bioc/html/topGO.html) was performed to extract a Fisher’s exact test with the classic method for gene ontology (GO) enrichment analyses. The p-values were adjusted using the false discovery rate (FDR) correction and GO terms with adjusted P-values < 0.05 were used.

Protein sequences of P. acutifolius (Frijol Bayo and W6 15578) were aligned (BLAST09; with E-value cutoff of 1E-10) MCScanX(82) (v2) with default options was used to scan for collinear blocks between the assemblies. The collinear blocks were visualized using Synvisio (https://synvisio.github.io/). To determine the orthologous and paralogous groups in the Phaseoleae tribe, OrthoFinder(299) (v2.2.7) was run with default search parameters and the representative proteins from A. thaliana (TAIR10)85, Medicago truncatula (Mt4.0v1)100, Cajanus cajan (v1)91, Glycine max (Wm82.a2.v1)101, Vigna unguiculata (v1.1)103, P. acutifolius (v2.1), P. acutifolius Frijol Bayo, and P. acutifolius W6 15578. The R package ‘Upset’ (v1.4.0) was used to visualize the results. An ultrametric tree was constructed using the OrthoFinder rooted species tree with the ‘make_ultrametric.py’ python script in OrthoFinder. The root age was estimated to be 120 million years ago based on data from two separate studies102,103. CAFE(104) (v4.1.2; –filter p 0.01 -r 10000) was used with the ultrametric tree and the identified gene families from the root to characterize the expanded and contracted families. The Biocore R package ‘topGO’ (v2.36.0; https://bioconductor.org/packages/release/bioc/html/topGO.html) was performed to extract a Fisher’s exact test with the classic method for gene ontology (GO) enrichment analyses. The p-values were adjusted using the false discovery rate (FDR) correction and GO terms with adjusted P-values < 0.05 were used.

Protein sequences of P. acutifolius (Frijol Bayo and W6 15578) were aligned (BLAST09; with E-value cutoff of 1E-10) with protein sequences of P. vulgaris GI9833 (Phytozone; Prulgaris_442_v1.2.protein.fa), G. max (EnsemblPlants; Glycine_max_v2.1.pep.all.fa) and V. unguiculata (Phytozone; Vignuculата_469_v1.1.protein.fa). Putative orthologs between different species were identified using the reciprocal best BLAST hit method. Estimation of the level of synonymous substitution (Ks) between orthologous pairs was performed using pairwise_kaks.PLS (v 1.8; https://github.com/cybersiddhu/boperlive/blob/master/scripts/utilities/pairwise_kaks.PLS). This script takes an input of cDNA length, verifies that they do not exceed the expanded and contracted families. The Biocore R space project, processes the alignment back into cDNA, and estimates the Ks (non-synonymous) and Ks (synonymous) substitutions based on the maximum likelihood method of Yang with the PAML package100,101. Ks values less than 0.001 are affected by rounding errors that result in spurious peaks. So we used Ks values greater than 0.001 for generating histograms. Gaussian mixture models were fitted by the R package Mclust(v5.4.6) and the number of Gaussian components (G), the mean of each component, and fractions of data were recorded. The best fitting model to the data was determined and tested using the Bayesian Information Criterion and p2 tests, respectively. In the case of skewed peaks resulting from overlapping components with a Gaussian mixture model, the formula: λ = λ/2 ⋅ 7, where 7 is the age of divergence (19.2 Mya) and λ is the geometric mean of synonymous distances between orthologous gene pairs of soybean and common bean (Ks = 0.32). Using
Domestication (a reduction/loss of diversity) and adaptation (acquiring new alleles or allele frequency change from the neutral genome) analyses were performed on SNPs called from the alignments of whole-genome data of each genome type (average coverage of 20×) against the tepary Frijol Bayo landrace reference genome. The diversity parameters were calculated over 100–10 kb sliding windows. π (the average pairwise nucleotide differences in a sample), θ (number of segregating sites in a sample), and unique polymorphisms in both genotypes were calculated to identify genomic regions associated with domestication. The distribution of the SNPs in which one of the alleles was present in the wild subpopulation and a different allele was present in the landrace subpopulation was calculated to detect the adaptation genomic regions. The genomic regions that were significantly different were then investigated to identify candidate genes (using the gene models from the Frijol Bayo reference genome). The predicted effect of polymorphism on a gene was determined using snpEff ([4.0](https://snpeff.sourceforge.net/)), and the associated gene ontology for a gene was determined using topGO ([v2.38.1](https://bioconductor.org/packages/release/bioc/html/topGO.html)).

**Reporting summary** Further information on research design is available in the Nature Reporting Summary linked to this article.

**Data availability** Data supporting the findings of this work are available within the paper and the Supplementary Information files. A reporting summary for this article is available as a Supplementary Information file. The data sets generated and analyzed during this study are available from the corresponding author upon request. Raw sequence reads are available in the National Center for Biotechnology Information Sequence Read Archive under Project ID PRJNA670318. Large associated data sets include (i) genome assembly, gene annotation (genes, gene models, transcripts, peptides), and associated gift file; (ii) gene expression matrices and differentially expressed genes; and (iii) orthologous groups are available in the Dryad Digital Repository [https://doi.org/10.5061/dryad.6q573n5w2](https://doi.org/10.5061/dryad.6q573n5w2). The genome assemblies with their annotation are also available at Phytozome [https://phytozome-next.jgi.doe.gov/pv3;0;info/PacutifoliusWLD_v2.0](https://phytozome-next.jgi.doe.gov/pv3;0;info/PacutifoliusWLD_v2.0); External data sets used in this study include: G. max proteome (EnsemblPlants; Glycine_max_v2.1.pep.all.fa; https://plantsensembl.org/); PFAM (v32; http://pfam.xfam.org/); Phytozone genome/proteomes (A. thaliana proteome, TAIR10); Cajanus cajan (v1); Glycine max (Wrm82.a2.v1); Medicago truncatula (M40v1); P. vulgaris V2.1; Vigna unguiculata (v1.1); https://phytozone.jgi.doe.gov/pv3;0;info/Pacutifolius_v1;0; http://phytozone-next.ugd.oeo.gov/info/PacutifoliusWLD_v2.0); Plant Transcription Factor Database v4.0 (planttfdb.cbi.pku.edu.cn/prediction.php). The germplasm is available from the U.S. Department of Agriculture Germplasm Resources Information Network (https://www.ars-grin.gov/). Source data are provided with this paper.

Received: 23 March 2020; Accepted: 7 January 2021; Published online: 11 May 2021

**References**

1. Binkley, A. The amount of blossom and pod drop on six varieties of garden beans. *Proc. Am. Soc. Hortic. Sci.* 29, 489–492 (1932).
2. Monterroso, V. A. & Wien, H. C. Flower and pod abscission due to heat stress in beans. *J. Am. Soc. Hortic. Sci.* 115, 631–634 (1990).
3. Gross, Y. & Kigel, J. Differential sensitivity to high temperature of stages in the reproductive development of common bean (*Phaseolus vulgaris* L.). *Field Crops Res.* 36, 201–212 (1994).
4. Porch, T. G. & Jahn, M. Effects of high-temperature stress on microsporogenesis in heat-sensitive and heat-tolerant genotypes of *Phaseolus vulgaris*. *Plant Cell Environ.* 24, 723–731 (2001).
5. Parker, J. P. & Michaels, T. E. Simple genetic control of hybrid plant development in interspecific crosses between *Phaseolus vulgaris* L. and *P. acutifolius* A. Gray. *Plant Breed.* 97, 315–323 (1986).
6. Singh, S. P. & Muñoz, C. G. Resistance to common bacterial blight among *Phaseolus* Species and common bean improvement. *Crop Sci.* 39, 80–89 (1999).
7. Kuosova, P. M. & Myers, J. R. Seed storage proteins LR2 and its variants from the apocarp of wild tepary bean G40199 confers resistance to acanthoscellides obtectus when expressed in common beans. *Afr. Crop Sci. J.* 9, 255–265 (2011).
8. Stover, J. R., Gurusamy, V., Porch, T. G. & Bett, K. E. Successful introgression of abiotic stress tolerance from wild tepary bean to common bean. *Crop Sci.* 57, 1160–1171 (2017).
9. Thomas et al. Teparies as a source of useful traits for improving common beans. *Desert Plants* 5, 43–48 (1983).
null
101. Vanshney, R. K. et al. Draft genome sequence of pigeonpea (Cajanus cajan), an orphan legume crop of resource-poor farmers. Nat. Biotechnol. 30, 83–89 (2011).

102. Schmitz, J. et al. Genome sequence of the palaeopolyploid soybean. Nature 463, 178–183 (2010).

103. Lonardi, S. et al. The genome of cowpea (Vigna unguiculata [L.] Walp.). Plant J. 98, 767–782 (2019).

104. Conway, J. R., Lex, A. & Gehlenborg, N. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics 33, 2938–2940 (2017).

105. Hohmann, N., Wolf, E. M., Lysak, M. A. & Koch, M. A. A time-calibrated road map of brassicaceae species radiation and evolutionary history. Plant Cell 27, 2770–2784 (2015).

106. Magallón, S., Gómez-Acvedo, S., Sánchez-Reyes, L. L. & Hernández-Fernández, T. A. Metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. N. Phytol. 207, 437–453 (2015).

107. Han, M. V., Thomas, G. W. C., Lugo-Martínez, J. & Hahn, M. W. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAPE. Mol. Biol. Evol. 30, 1987–1997 (2013).

108. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. J. Mol. Biol. 215, 403–410 (1990).

109. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586–1591 (2007).

110. Lupas, A., Van Dyke, M. & Stock, J. Predicting coiled coils from protein sequences. Science 252, 1162–1164 (1991).

111. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35, 1547–1549 (2018).

112. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797 (2004).

113. Jun, J., Zhang, H., Kong, L., Gao & Luo, J. PlantTFDB 3.0: a portal for the visualization and evolutionary study of plant transcription factors. Nucleic Acids Res. 42, D1182–D1187 (2014).

114. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly 6, 80–92 (2012).
