Forage genomics accelerate the germplasm resource innovation

Wu, Fan; Yan, Qi; Xu, Pan; Duan, Zhen; Zhang, Jiyu*
State Key Laboratory of Herbage Improvement and Grassland Agro-Ecosystems; College of Pastoral Agriculture Science and Technology, Lanzhou University; Lanzhou 730000, China; * Corresponding author
Key words: genome assembly; genome evolution; stress tolerance; genetics resource

Abstract
To achieve sustainability and food security we need expand the germplasm base and access novel genetic diversity to accelerate breeding. For developing new forage cultivars, the availability of a high-quality genome facilitates accurate characterization of new germplasm, and an understanding of the genetics underlying important traits. Here, we sequenced and assembled three high-quality chromosome-level forage genomes. The contig-level assembly of *Cleistogenes songorica* (2n = 4x = 40) comprised 540.12 Mb of the genome, with a contig N50 of 21.28 Mb. Complete assemblies of all telomeres, and of ten chromosomes were derived. The chromosome-scale genome size of elephant grass (2n = 4x = 28) was 1.97 Gb and heterozygosity rate was 1.5%. The chromosome-scale genome size of *Melilotus albus* (2n = 2x = 16) was 1.04 Gb, containing 71.42% repetitive elements. This study provides implementation pathways to study genome evolution, adaptation to stress and genetic basis of unique or complex traits in three species. The genomic resources that we developed in this study offer valuable information that will facilitate efficient germplasm exploration and genetic improvement of the three species for pasture uses.

Introduction
In the face of global environmental variability, food security is critical to the feeding of upwards of 10 billion people by 2050 (Tester and Langridge, 2010). There is growing recognition that improving the environmental performance of agriculture and livestock systems and establishing sustainable levels of the consumption of crop and animal-sourced foods, are essential for the sustainability of the global food system(Foley et al., 2011; Herrero et al., 2013). Global climate change is predicted to greatly increase the prevalence and severity of drought, salt and other stress (Dai, 2012). Legume plants play an important role in delivering sustainability(Stagnari et al., 2017), which are as legumes perform well in conservation and intercropping systems, important in low productivity farming systems(Stagnari et al., 2017). And a lot of native grass have an ability to survive extreme environment. *Cleistogenes songorica* is an important perennial forage, and ecologically significant C_4 grass in temperate saline, semi-arid, and desert areas in central Asia where average annual rainfall is below 110 mm. With a strong root system, *C. songorica* has found application in desert ecosystem and grassland restoration by stabilizing soil structure and reducing soil erosion (Niu and Nan, 2017). Elephant grass is an excellent fodder crop with a yield of up to 150 tons green matter per hectare each year and is capable of withstanding repeated cuttings (four to six cuts per year), resisting high temperatures, drought stress, low soil fertility and biotic stress (Kebede et al., 2017; Liu et al., 2008). Elephant grass, as a lignocellulosic plant, has high potential for bioenergy and paper production (Daud et al., 2014). *Melilotus* have adapted to extremely varied environments including drought, cold and saline conditions, and are widely used as fodder, green manure, and soil conservation crops(Stickler and Johnson, 1959; Zhang et al., 2018). Sweet clover was considered the king of green manures and grazing legumes in the South and Midwest of the USA in the first half of the 20^{th} century(Clark, 2007). The paucity of genetic information in *C. songorica*, elephant grass and sweet clover has hindered the accurate characterization of new germplasm for developing forage cultivars. Genome sequencing of *C. songorica*, elephant grass and sweet clover will facilitate an understanding of the genetics underlying stress tolerance, and adaptability.

Methods
For *M. albus* and *C. songgorica*, both Illumina short-read sequencing and PacBio single-molecule real-time (SMRT) sequencing platforms were used to independently sequence genomes. For elephant grass, Illumina Navoseq 6000 platform and Nanopore PromethION platform were used to genome sequence. *De novo* assembly of the PacBio long reads was conducted using Falcon v 0.3.0. (Chin et al., 2016). The Nanopore reads was assembled with SMARTdenvo v1.0.0. To obtain the chromosome-level genome assembly, we used Lachesis software to cluster, order and orient the contigs by the Hi-C data (Burton et al., 2013). Structural predictions and *de novo* approaches were used to annotate the three genomes repetitive sequences. A combination of *de novo* predictions, homologue-based predictions and RNA-seq-based predictions were adapted to predict the protein-coding genes of the three genomes. Protein-coding gene functional annotations were based on homologous alignment with BLAST (e-value < 1e-5) against Nr (https://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/), KOG

(ftp://ftp.ncbi.nih.gov/pub/COG/KOG/), KEGG (https://www.kegg.jp/), SwissProt (http://www.uniprot.org/) and TrEMBL (https://www.ebi.ac.uk/uniprot).

Rice, *Arabidopsis* and *M. truncatula* genes were used as queries, and coding and protein sequences of rice, *Arabidopsis* and *M. truncatula* were downloaded from RiceData (https://www.ricedata.cn/gene/), The *Arabidopsis* Information Resource (https://www.arabidopsis.org/index.jsp) and Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html#), respectively. The corresponding genes in the three species were identified based on a BLAST search against the three genomes with a cut-off e-value < 1e-5. All the identified sequences with redundant sequences removed were submitted to Pfam (http://pfam.xfam.org/search#tabview=tab1) for annotations and subjected to a NCBI Conserved Domain Search (https://www.ncbi.nlm.nih.gov/cdd/) to further confirm their identities. Genes without the conserved domain were discarded. Gene expression were verified by real time quantitative PCR. *CsAP2_9* and miR172l were transformed into rice by *Agrobacterium tumefaciens*-mediated transformation method, respectively (Hiei and Komari). *MaBGLU1* was transformed into *Escherichia coli* to explore its heterologous expression activity. The *MaBGLU1* was transformed into roots of *M. albus*, the transformed roots were checked by RFP visualization. More details about materials and methods can refer to three papers (Wu et al., 2022; Yan et al., 2021; Zhang et al., 2021).

Results and Discussion

Genome Sequencing and Assembly

We sequenced and assembled three high-quality chromosome-level forage genomes. The contig-level assembly of *C. songorica* (2n = 4x = 40), elephant grass (2n = 4x = 28) and *M. albus* (2n = 2x = 16) comprised 540.12 Mb, 1.97 Gp and 1.05 Gb of the genome, with a contig N50 of 21.28 Mb, 1.83 Mb and 7.49 Mb, respectively. To anchor and orient the contigs onto chromosomes we constructed a Hi-C library. 528.52 Mb of contigs (35 contigs, 97.85% coverage) were anchored to *C. songorica* 20 pseudochromosomes, ten of them having no gaps. 96.65% (1.90 Gp) of the total contig bases were anchored and oriented to the elephant grass 14 chromosomes. A total of 239 contigs covering 1.04 Gb (99.40%) of the assembled genome were anchored to *M. albus* 8 pseudochromosomes.We identified 54,383 (89.48% annotated), 65,927 (98.03% annotated) and 41,910 (98.84% annotated) protein-coding genes in *C. songorica*, elephant grass and *M. albus* genome, respectively. The *C. songorica*, elephant grass and *M. albus* genome contained 41.99%, 66.32% and 71.42% repetitive sequences, respectively. The expansion of the *M. albus* genome mainly caused by the proliferation of repetitive sequences (Wu et al., 2022). GC content of the *C. songorica*, elephant grass and *M. albus* was 45.02%, 46.95% and 35.77%. The GC content of the *C. songorica* and elephant grass near to the upper limit of the range (33.6% to 48.9%) in monocots (Šmarda et al., 2014). High GC content has been reported associated with plant adaptation to abiotic stress (Costa et al., 2017).

Genetic determination of dimorphic flowers in C. songorica

C. songorica serves as a natural forage source in harsh environments largely because of its dimorphic flowering mechanism, which allows it to survive and reproduce under extreme conditions. It develops two types of inflorescences in a single plant, enabling open pollination (chasmogamy, CH) on the top panicle, and self-pollination (cleistogamy, CL) on spike flowers embedded in the leaf sheath at each node (Figure 1a) (Wu et al., 2018). The interaction of miR172s and *AP2* genes is known to promote floral opening in barley, as mutations in *cleistogamy 1* (*cly1*) cause failure in lodicule development (Nair et al., 2010; Ning et al., 2013). Nine Cs*AP2* were identified and were putatively targeted by miR172s (Figure 1c). miR172l is specifically targeted to two paralogs Cs*AP2_8* and Cs*AP2_9* (Figure 1d). Transcripts of Cs*AP2_9* were more abundant in CL than in CH, with the Cs*AP2_8* showing inverse patterns (Figure 1b). These results support the hypothesis that Cs*AP2*_8 and Cs*AP2*_9, by interacting with miR172l, contribute to the regulation of cleistogamy in *C. songorica*. We performed over-expression of Cs*AP2_9* and miR172l in transgenic rice, driven by the constitutive CaMV *35S* promoter. Compared with wild type, transgenic plants exhibited floral defects, with Cs*AP2_9* transgenic lines showing abnormal palea, and smaller and thinner lodicules; and miR172l lines showing longer filaments, and reduced anther numbers (Figure 1e-h). These results validate the functions of Cs*AP2_9* and miR172l in regulating lodicule, and filament and anther development, respectively.

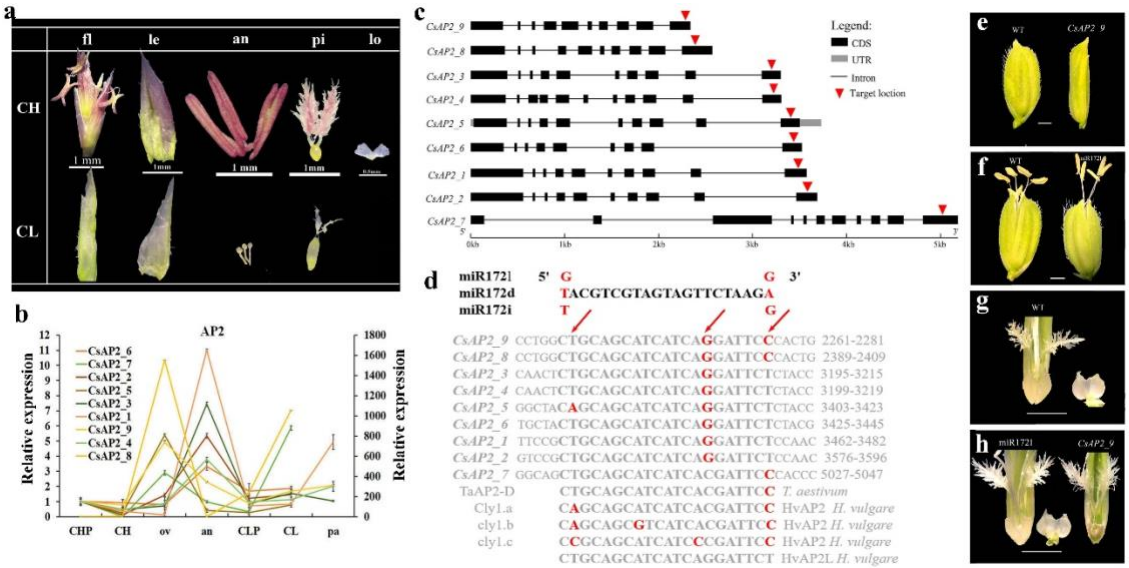

Figure 1. Genome-wide identification of flowering genes. (a) Morphology of chasmogamy (CH) and cleistogamy (CL) flowers. (b) Relative expression levels of AP2 in different tissues. CHP, CH flower primordium; ov, ovary; an, anther; pa, palea; CLP, CL flower primordium. (c) miR172 target locations in *CsAP2* genes. (d) Nucleotide variants in miR172 and its *AP2* TF target site sequences. Nucleotide variants are marked in red in bold. (e) *CsAP2_9* overexpression confers an abnormal palea in a spikelet in the T_0 generation of transgenic plants. (f) miR172l overexpression confers five anthers with longer filaments in the T_0 generation of transgenic plants, compared to six anthers in wild type. (g,h) Spikelets with dissected stigma and lodicules, showing degenerated lodicules in flowers in the T_0 generation of *CsAP2_9* overexpressed transgenic plants. Scale bars represent 1 mm

The expansion of genes involved in the plant hormone signal transduction pathway in elephant grass

Among the plant hormones, auxin has been shown to mediate cell enlargement, and adventitious root development. Cytokinine play essential roles in many aspects of plant development through regulation of cell division (Artner and Benkova, 2019). GAs are one of the plant hormones that stimulate plant development, including stem elongation and fertility. BRs are growth-promoting steroid hormones that regulate cell division and cell elongation. The hormone signal transduction pathway, which included some key genes, plays an important role in the process. Overexpression *SAUR* genes is sufficient to induce cell elongation and growth (Stortenbeker and Bemer, 2018). The *gh3* mutants exhibit reduced lateral root number, and auxin-deficient traits in *Arabidopsis* (Zhang et al., 2007). We found that some of key gene families invove in hormones signal transduction pathway were expanded in elephant grass (Figure 2a). The analysis of transcript levels showed that a lot of genes were highly expressed in tissues of elephant grass (Figure 2b), expecially, two *AUX1*, two *TIR1*, two *CRE1* and two BSK were highly expressed in the stem tip (Figure 2b), which may contribute to the fast growth of elephant grass.

a

Gene family	Elephant grass	Pearl millet	Foxtail millet	Rice
G3H	29	12	11	9
BIN2	17	6	7	9
SAUR	16	5	7	5
BRI1	13	6	9	6
TIR1	12	5	5	5
AUX1	10	5	4	5
BSK	10	3	5	5

Gene family	Elephant grass	Pearl millet	Foxtail millet	Rice
CRE1	10	5	4	4
ARF	6	3	3	3
AUX/IAA	6	1	2	3
AHP	4	2	2	2
B-ARR	4	2	2	2
DELLA	3	1	1	1
GID1	2	1	1	1

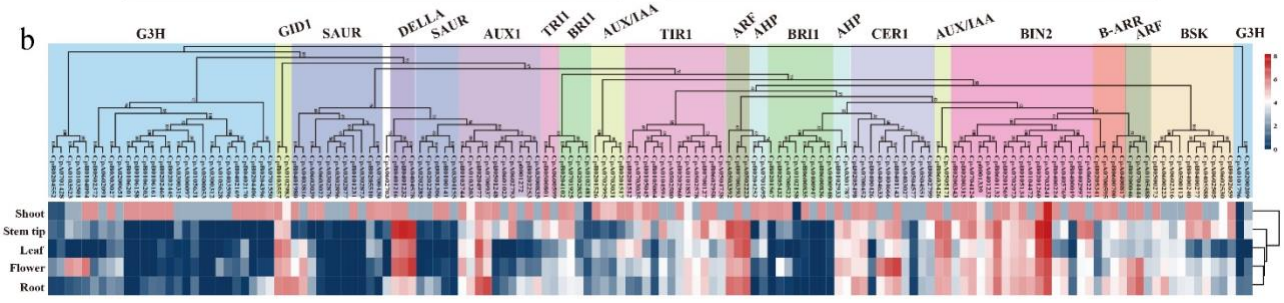

Figure 2. (a) Gene copy number of key gene families involved in hormone signal transduction pathway in elephant, foxtail millet, rice and pearl millet. (b) Phylogenetic tree and heatmap (value = Log_2(FPKM+1)) of plant hormone signal transduction-related genes. AHPs, arabidopsis histidine phosphotransfer proteins; AUX1, auxin influx carrier; ARF, auxin response factor; Aux/IAAs, auxin/indole-3-acetic acid; B-ARRs, B-type Arabidopsis response regulators; BRI1, Brassinosteroid insensitive 1;

BSKs, BR-signaling kinases; BIN2, Brassinosteroid insensitive 2; CRE1, cytokinin receptor 1; TIR1, transport inhibitor response 1; GH3, gretchenhagen-3; GID1, GA-insensitive dwarf1; SAUR, small auxin upregulated

The MaBGLU1 contributes to scopoletin biosynthesis in M. albus

Six *BGLU* genes were identified which arranged in tandem within a 290 kb region in chromosome 3 of *M. albus* (Figure 3a). They cluster into a same clade with *AtBGLU21–23* (regulate scopoletin biosynthesis) (Ahn et al., 2010) (Figure 3c). Therefore, we speculated that all six *MaBGLUs* are involved in converting scopolin to scopoletin in sweet clover. *MaBGLU1* showed the highest expression in all tissues among the six genes (Figure 3b). Heterogeneous expression of *MaBGLU1* in *Escherichia coli* generated 506 amino acid peptide. We then set out to test the activity of *MaBGLU1* products by conducting an enzymatic activity assay of *MaBGLU1* proteins using scopolin as a substrate. Scopoletin and scopolin were both detected (Figure 3d), the result demonstrated that *MaBGLU1* is a key enzyme in the scopolin biosynthesis. We also overexpressed the *MaBGLU1* gene in *M. albus* via a hairy root transformation system. We were able to generate transgenic lines emitting red fluorescent protein (RFP) signals (Figure 3e). qRT-PCR analyses showed an obvious increase (8.9 times) in relative expression levels in the positive transgenic lines (Figures 3f). The scopolin and scopoletin levels were increased in these lines. These results demonstrated the role of the *MaBGLU1* gene in regulating scopoletin content.

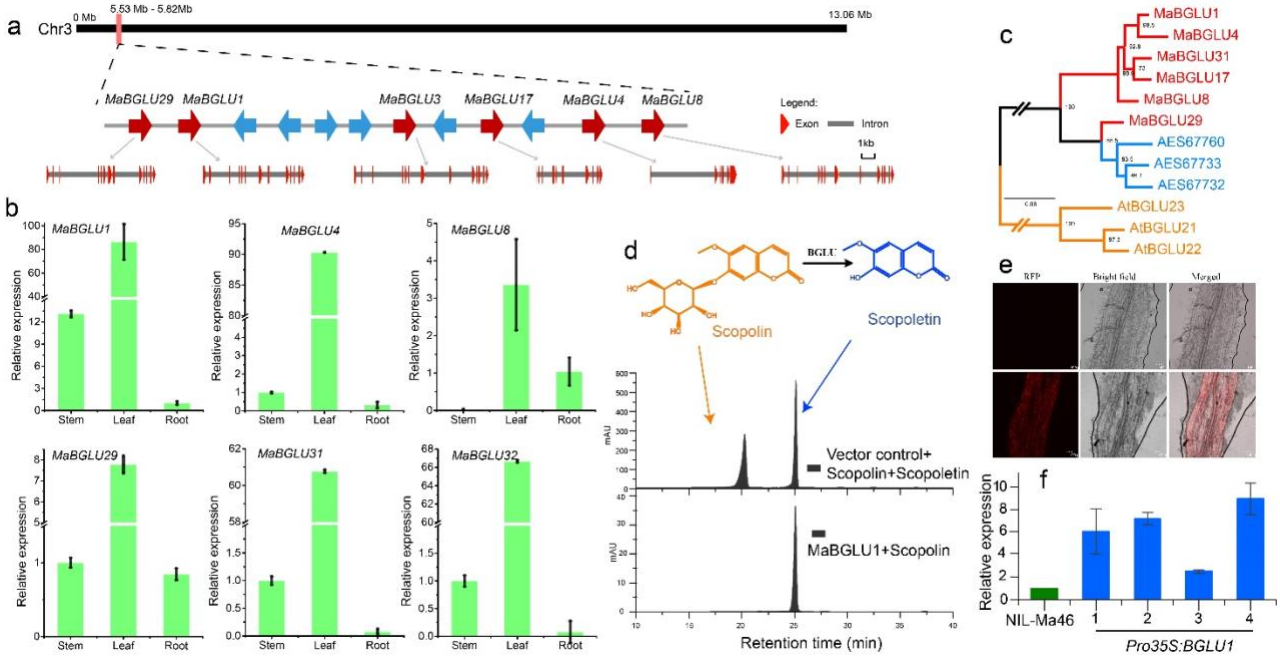

Figure 3. Functional analysis of *MaBGLU1*. (a) Molecular structures and chromosomal locations of six MaBGLUs. (b) Phylogenetic analysis of the BGLU subfamily in *M. albus*, *M. truncatula* and *Arabidopsis*. (c) Relative expression of six MaBGLUs. (d) LC-MS/MS analysis of scopolin and scopoletin in the presence of recombinant MaBGLU1 and the empty vector control. (e) Observation and photo taking of RFP in control and OE-BGLU1 roots under a laser scanning confocal microscope. (f) The relative expression level of *MaBGLU1* in NIL-Ma46 and OE-BGLU1 (*Pro35S:BGLU1*) root.

Conclusions

Genome-wide gene characterizations in *C. songorica*, elephant grass and *M. albus* provides rich gene resources for genetic breeding. Interaction of *CsAP2_9* and miR172l contribute to the regulation of cleistogamy of *C. songorica*. The expansion of genes involved in the plant hormone signal transduction pathway which may contribute to the fast growth of elephant grass. *MaBGLU1* regulate scopoletin (a kind of coumarin) biosynthesis, which is the target gene for cultivating low coumarin sweet clover. These function verified genes provide opportunities for germplasm resource innovation.

Acknowledgements

This work was supported by the Inner Mongolia Seed Industry Science and Technology Innovation Major Demonstration Project (2022JBGS0040), the Fundamental Research Funds for the Central Universities (lzujbky-2022-ey17)

References

Artner, C., Benkova, E. 2019. Ethylene and cytokinin: partners in root growth regulation. Mol. Plant, 12(10): 1312-1314.

Ahn, Y.O., Shimizu, B., Sakata, K., *et al.* 2010. Scopolin-hydrolyzing beta-glucosidases in roots of *Arabidopsis*. *Plant Cell Physiol*. 51: 132-143.

Burton, J.N., Adey, A., Patwardhan, R.P., et al. 2013. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat. Biotechnol.* 31: 1119-1125.

Clark, A. 2007. Managing cover crops profitably, 3rd ed. Sustainable Agriculture Network, Beltsville, MD.

Costa, M.D., Artur, M.A., Maia, J., et al. 2017. A footprint of desiccation tolerance in the genome of Xerophyta viscosa. *Nat. Plants*, 3: 17038.

Chin, C.S., Peluso, P., Sedlazeck, F.J., *et al*. 2016. Phased diploid genome assembly with single molecule real-time sequencing. *Nat. Methods*, 13: 1050.

Dai, A. 2012. Increasing drought under global warming in observations and models. *Nat. Clim. Change*, 3: 52-58.

Daud, Z., Hatta, M., Kassim, A., et al. 2014. Analysis of Napier grass (*Pennisetum purpureum*) as a potential alternative fibre in paper industry. *Mater. Res. Innov.*, 18: 18-20.

Foley, J.A., Ramankutty, N., Brauman, K.A.,et al., 2011. Solutions for a cultivated planet. *Nature*, 478: 337-342.

Herrero, M., Havlík, P., Valin, H., et al. 2013. Biomass use, production, feed efficiencies, and greenhouse gas emissions from global livestock systems. *PNAS.*, 110: 20888-20893.

Kebede, G., Feyissa, F., Assefa, G., et al. 2017. Agronomic performance, dry matter yield stability and herbage quality of Napier grass (*Pennisetum purpureum* (L.) Schumach) accessions in different agro-ecological zones of Ethiopia. *J. Agr. Crop Res*., 5(4): 49-65

Liu, X., Shen, Y., He, Y., et al. 2008. Tolerance to copper stress in Elephantgrass (*Pennisetum purpureum*) under soil culture. *XXI International Grassland Congress & the VIII International Rangeland Congress*.

Nair, S.K., Ning, W., Turuspekov, Y., et al. 2010. Cleistogamous flowering in barley arises from the suppression of microRNA-guided HvAP2 mRNA cleavage. *PNAS.*, 107: 490-495.

Niu, X., Nan, Z. 2017. Roots of Cleistogenes songorica improved soil aggregate cohesion and enhance soil water erosion resistance in rainfall simulation experiments. *Water Air Soil Poll.*, 228: 109.

Petr, S., Petr, B., Lucie, H., et al. 2014. Ecological and evolutionary significance of genomic GC content diversity in monocots. *PANS.*, E4096-E4102.

Ning, S., Wang, N., Sakuma, S., et al. 2013. Structure, transcription and post-transcriptional regulation of the bread wheat orthologs of the barley cleistogamy gene *Cly1*. *Theor. Appl. Genet.*, 126: 1273–1283.

Stagnari, F., Maggio, A., Galieni, A., et al. 2017. Multiple benefits of legumes for agriculture sustainability: an overview. *Chem. Biol. Technol. Agr.*, 4: 2.

Stickler, F.C., Johnson, I.J., 1959. Dry matter and nitrogen production of legumes and legume associations in the fall of the seeding year[1]. *Agron. J*., 51, 135-137.

Stortenbeker, N., Bemer, M., 2018. The SAUR gene family: the plant's toolbox for adaptation of growth and development. *J. Exp. Bot*., 70: 17–27.

Šmarda P, Bureš P, Horová L, et al. 2014. Ecological and evolutionary significance of genomic GC content diversity in monocots. *PNAS*, 2014, 111(39): E4096-E4102.

Tester, M., Langridge, P., 2010. Breeding technologies to increase crop production in a changing world. *Science.*, 327, 818-822.

Wu, F., Duan, Z., Xu, P., et al. 2022. Genome and systems biology of *Melilotus albus* provides insights into coumarins biosynthesis. *Plant Biotechnol. Jl.*, 20, 592-609.

Wu F, Zhang D, Muvunyi B P, et al. 2018. Analysis of microRNA reveals cleistogamous and chasmogamous floret divergence in dimorphic plant. *Sci. Rep*., 8(1): 6287.

Yan, Q., Wu, F., Xu, P., et al., 2021. The elephant grass (*Cenchrus purpureus*) genome provides insights into anthocyanidin accumulation and fast growth. *Mol. Ecol. Res*., 21: 526-542.

Zhang, J., Wu, F., Yan, Q., et al. 2021. The genome of *Cleistogenes songorica* provides a blueprint for functional dissection of dimorphic flower differentiation and drought adaptability. *Plant Biotechnol. J.*, 19: 532-547.

Zhang, Z., Li, Q., Li, Z., et al. 2007. Dual regulation role of GH3.5 in salicylic acid and auxin signaling during Arabidopsis-pseudomonas syringae interaction. *Plant Physiol.*, 145: 450-464.