AgroScience Today

Popular Article

e-ISSN: 2583-0147

Volume 6 Issue 12
Page: 1181 - 1184

The Grasspea Genome Sequence - A Blueprint for Grasspea Improvement

Neetu Singh Kushwah[1]*, Antra Das[1], Virendra Pratap Singh Rathor [2]

[1]Senior Scientist, ICAR-Indian Institute of Pulses Research, Kanpur, India.

[2]Associate Professor, Maharana Pratap College of Pharmacy, Kanpur, India.

*Corresponding author's e-mail: *neeturajawat@gmail.com*

Published on: December 31, 2025

ABSTRACT

Lathyrus sativus **commonly known as grasspea is a nutrient dense legume crop with tolerance for various abiotic stresses like drought and flood and has potential for climate smart agriculture. It is also a reservoir of nutrients and pharmacological compound that can contribute to human health. Grasspea has received little attention from breeders and researchers in the past due to the presence of neurotoxic compound in its seeds and other plant parts that causes a disease called neurolathyrism. However, in the current climate change scenario, interest on grasspea research is renewed due its hardiness to both drought and flooding. Several genomic resources in grasspea have been developed from the last 10 years. As per NCBI data, 246 transcriptome sequences of grasspea have been published. Two genome assemblies and one reference genome sequence have recently been published. This article discusses about currently sequenced grasspea genomes and how it can be utilized for improving the grasspea.**

INTRODUCTION

Grasspea (*Lathyrus sativus* L.) (Family Fabaceae) is a diploid (2n=2x=14) pulse crop, cultivated predominantly in South Asia and Africa. Grasspea has long history of cultivation and preferred among resource poor farmers due to its hardiness and low-cost of cultivation and high protein content. It is also a reservoir of compound that can contribute to human health. Notably, it is the only known dietary source of l-homoarginine, which offers remarkable benefits for the treatment of cardiovascular diseases. Besides these attractive features, grasspea also contains antinutritional compound called β-ODAP, which causes paralysis when grasspea continuously consumed as a staple food. Generation of genetic and genomic resources is the prerequisite to accelerate modern crop breeding program in grasspea. However, the large genome size (~6.52 Gbp) and the relatively limited availability of genetic resources, such as high-density genetic and molecular maps, have made it difficult to scaffold the grasspea genome information to the pseudochromosome level. Recent advancements in long-read sequencing enable whole-genome shotgun approaches for assembling multi-gigabase genomes. By utilising this advancement in sequencing technology, the two draft genome sequence of grasspea (*Lathyrus sativus* L.) were released in year 2023 as a result of endeavour from National Agri-Food Biotechnology Institute, Mohali, India and John Inn Centre, UK. Later, chromosome scale reference genome of grasspea (*Lathyrus sativus)* were released. In the following sections, we will discuss the various grass pea genome assemblies.

National Agri-Food Biotechnology Institute, India chose Pusa-24 as a choice cultivar for sequencing and used two next generation sequencing platforms Illumina HiSeq 2500 and PacBio Sequel (I) to assemble the grasspea genome sequence into seven chromosome-sized scaffolds (Rajarammohan et al. 2023) (Table 1). The assembled genome of Lathyrus covered 3.80 Gb of the genome, representing 57.4% of the estimated genome size (6.62 Gb) of the Lathyrus. The assembly exhibits a BUSCO completeness score of 98.35% and scaffolded to chromosome level using *Pisum sativum* cv. Caméor v1a assembly. Repetitive and transposable elements covered the bulk of the genome (83.31%). The LTR retrotransposons (*Gypsy* and *Copia*) (37.58%) were more abundant as compared to DNA transposons (hobo-Activator, PiggyBac, Tourist/Harbinger etc.). Gene prediction tools identified 50,106 protein coding genes, of which 45,632 were located on the chromosome-sized scaffolds (96.21%). Orthology analysis indicated 13840 genes that are specific to Lathyrus. However, their assembly did not adequately represent rDNA and satellite DNA.

In the same year, John Inn Centre, UK has released draft genome assembly of European grasspea genotype Ls007 (Table 1). They used next generation sequencing platforms Oxford Nanopore Technology (ONT) and Illumina HiSeq. The assembled genome of grasspea is 6.2 Gbp (6.5 Gbp estimated genome size). Of which 42.7% of total assembly were scaffolded into 7 chromosome size scaffolds, and 2 sub-chromosome-scale scaffolds using Hi-C data. The assembly exhibits a BUSCO completeness score of 88.5%. In the assembled sequence, 30,167 high-confidence protein-coding genes and 15,307 low-confidence protein-coding genes were identified. Repetitive and transposable elements covered the bulk of the genome (80.61%). LTR-retrotransposons dominant the grasspea genome account for 57% of the genome and Ty3/Gypsy Ogre elements (37.3%) representing the majority of the population of LTR- retrotransposons.

Satellite repeats, estimated at 8% of the genome, are second in terms of genome abundance. In the genome assembly they have annotated the key genes involved in ODAP biosynthesis pathway viz., LsCAS present on contig ctg2942, LsAAE3 present on contig ctg4766 and LsBOS is present on contig ctg14433.

Vigouroux et al. (2024) made further advancements in the genome assembly of grasspea and developed a reference genome for the species (Table 1). They utilized the Pacific Biosciences HiFi long-read platform to sequence the European grasspea cultivar Ls007, generated a 5.96 Gbp assembly. Of this assembly, 84.40% was scaffolded into a chromosome-scale framework using Hi-C data. Scaffolding with Hi-C data is considered robust because it creates chromosome-scale assemblies by using the 3D structure of the genome to order and orient contigs. This leads to more contiguous and complete genome assemblies compared to hybrid-based assembly approach. The assembly achieved a BUSCO completeness score of 99.3%. They used fluorescence in situ hybridization (FISH) technique to assign pseudomolecules to specific chromosomes. This reference assembly represent significant improvement over previous grasspea assemblies in terms of completeness, contiguity, and assembly correctness. The previous assembly in grasspea cv. Pusa-24 was scaffolded by aligning its contigs to the *Pisum sativum* cv. Caméor v1a assembly (Table 1). This method implies that regions of the grasspea genome that are not sufficiently similar to the pea genome such as those that are missing in the pea genome or expanded in the grasspea genome could not be scaffolded. On the other hand, Ls007 assembly developed by Edwards et al. (2023) was only partially scaffolded due to its high level of fragmentation (Table 1).

Table 1. Comparative analysis of different genome assemblies of grasspea (*Lathyrus sativus* L.)

	Vigouroux et al. (2024)	Edwards et al. (2023)	Rajarammohan et al. (2023)
Submitted GenBank assembly	GCA_036972225.1	GCA_963859935.3	GCA_026873245.1
Taxon	*Lathyrus sativus* (white pea)	*Lathyrus sativus* (white pea)	*Lathyrus sativus* (white pea)
Cultivar	LS007	LS007	PUSA-24
WGS project	JAVSPV01	CAWUDS03	JAPMLZ01
Assembly type	haploid	haploid	Haploid
Submitter	John Innes Centre	John Innes Centre	National Agri-Food Biotechnology Institute
Date	Sep, 2024	Feb, 2023	Jan, 2023
Genome size	5.96Gb	6.2 Gb	3.8 Gb
Total ungapped length	5.96 Gb	6.2 Gb	3.8 Gb
Number of chromosomes	7	7	Nil
Number of organelles	1	Nil	1
Number of scaffolds	7,641	7,646	80,708

Scaffold N50	700.4 Mb	700.4 Mb	78.3 kb
Scaffold L50	4	4	14,791
Number of contigs	13,491	13,593	80,892
Contig N50	3.3 Mb	3.3 Mb	78 kb
Contig L50	494	498	14,851
GC percent	38	38	38.5
Genome coverage	23.6x	24.0x	118.0x
Assembly level	Chromosome	Chromosome	Scaffold
Sequencing technology	PacBio HiFi	PromethION nanopore + Illumina PE	Illumina; PacBio Sequel
Assembly method	hifiasm	hifiasm	MaSuRCA v. 4.0.3
Genes	31,719	31,719	50,106
Protein-coding	31,719	31,719	50,106
Illumina QV	42.4	41.9	18.0
Illumina kmer completeness	90.0	77.6	48.6
BUSCO-Complete-ness, Viridiplantae	99.1%	89.8%	98.3%
BUSCO- Complete-ness, Fabales	97.4%	82.6%	96.0%

CONCLUSION

Whole genome sequence information of grasspea opened a wide platform for the identification of genes and specific metabolic pathways responsible for β-ODAP production. This will aid in developing strategies to reduce or remove the neurotoxin through advanced breeding tool like gene editing. Additionally, the genome sequence information will facilitate the identification of genes/pathway that could be used to improve the crop or to understand its remarkable drought tolerance and pharmacologically active compounds that can contribute to human health.

REFERENCES

Edwards, A., *et al.* (2023). Genomics and biochemical analyses reveal a metabolon key to β-L-ODAP biosynthesis in *Lathyrus sativus*. *Nature Communications, 14*, Article 876. https://doi.org/10.1038/s41467-023-00876-0

Rajarammohan, S., *et al.* (2023). Genome sequencing and assembly of *Lathyrus sativus*—A nutrient-rich hardy legume crop. *Scientific Data, 10*, Article 32. https://doi.org/10.1038/s41597-023-01932-1

Vigouroux, M., *et al.* (2024). A chromosome-scale reference genome of grasspea (*Lathyrus sativus*). *Scientific Data, 11*, Article 1035. https://doi.org/10.1038/s41597-024-03868-y.