



# Prospects of next generation sequencing in lentil breeding

Jitendra Kumar<sup>1</sup> · Debjyoti Sen Gupta<sup>1</sup>

Received: 6 June 2020 / Accepted: 3 October 2020 / Published online: 10 October 2020  
© Springer Nature B.V. 2020

## Abstract

Lentil is an important food legume crop that has large and complex genome. During past years, considerable attention has been given on the use of next generation sequencing for enriching the genomic resources including identification of SSR and SNP markers, development of unigenes, transcripts, and identification of candidate genes for biotic and abiotic stresses, analysis of genetic diversity and identification of genes/ QTLs for agronomically important traits. However, in other crops including pulses, next generation sequencing has revolutionized the genomic research and helped in genomic assisted breeding rapidly and cost effectively. The present review discuss current status and future prospects of the use NGS based breeding in lentil.

**Keywords** DNA sequencing · Functional markers · Unigenes · Transcriptome · Genetic gain · Pan-genome · Lentil

## Introduction

Lentil (*Lens culinaris* Medik.) is a diploid ( $2n=14$ ) self-pollinated crop and has a genome size of 4 Gbp [1]. It is a cool season crop that is cultivated globally in more than 52 countries. It is one of the healthiest pulse crops because its seeds are rich in minerals, fibers, and carbohydrates besides proteins (22–35%). During the past years significant increase has been observed in the productivity of lentil due to cultivation of high yielding varieties developed through traditional breeding. However, poor competitive ability to weeds, higher flower drop rate, pod shedding, and several biotic and abiotic stresses are still causes of reduced productivity in lentil crops [2]. Moreover, environmental conditions have high impact on the expression of agronomically important quantitative traits leading to poor genetic gain in lentil [3]. This is challenging to overcome under current climatic conditions. However, genomics has provided a potential way to increase the genetic gain in lentil and hence significant progress has been made in the past years for developing genomic resources in lentil. These genomic resources include availability of genome sequences, transcriptome sequences, molecular markers (i.e. SNPs and

SSRs), mapping populations, markers linked to genes/QTLs controlling important traits [4–10]. Further, advanced next generation sequencing technologies have accelerated the development of genomic resources rapidly and cost effectively and their use enriched genomic resources in different crops including orphan legume crops. Thus genomically poor legume crops have become rich with genomic resources due to the availability of draft genome sequences [11, 12] and NGS based approaches. For example, QTL-Seq made possible to identify genes/QTLs rapidly in different legume crops including chickpea, pigeonpea and groundnut [13–17]. Therefore, several workers have discussed the prospects of next generation sequencing for discovering genes for agronomically important traits and their use in breeding legume and food crops [18–21]. Three food legume (chickpea, pigeonpea and groundnut) crops where NGS has been used for different purposes have been covered in these reviews for discussing NGS based breeding [19, 20]. In lentil, which is an important food legume crop, considerable attention has been given on the use of next generation sequencing for enriching the genomic resources during the past years and used for identification of SSR and SNP markers, development of unigenes, transcripts, and identification of candidate genes for biotic and abiotic stresses, analysis of genetic diversity and identification of genes/QTLs for agronomically important traits [10, 22]. Next generation sequencing has revolutionized the genomic research and helped in genomic assisted breeding rapidly and cost effectively in different

✉ Jitendra Kumar  
jitendra.kumar@icar.gov.in

<sup>1</sup> Division of Crop Improvement, ICAR-Indian Institute of Pulses Research, Kalyanpur, Kanpur 208024, India

crops including legume crops in the past years. In this review, we discuss prospects of NGS based breeding in lentil.

## DNA sequencing methods

Different methods have been developed for analysing genomic sequences. These can be categorized broadly into following two groups.

### Conventional DNA sequencing

This is first generation DNA sequencing technology, which is based on Sanger's sequencing method. It is widely used in crop plants for development of molecular markers and sequencing the plant genome. It emerged in 1977 and used for three decades [23]. This first generation sequencing technology has been used in the genome sequencing of model plant species including *Arabidopsis thaliana*, *Oryza sativa* (rice), *Carica papaya* (papaya) and *Zea mays* (maize) [18]. In lentil, first generation DNA sequencing has been used to develop ESTs, genic and genomic SSR and SNP markers [4]. The kompetitive allele specific PCR (KASP) methodology has been used to detect the SNPs from available EST-database using this sequencing technology [2, 24].

### Advanced DNA sequencing

Advances in DNA sequencing led to next-generation sequencing (NGS) technologies that allow sequencing millions to billions of DNA nucleotides in parallel in a high throughput manner so that multiple samples can be sequenced at low cost [25]. Additionally, cloning of DNA fragments is not required [26]. NGS technologies are further divided into second and third generation and are discussed below in the light of lentil crop.

### Second-generation sequencing

Sequencing technologies of this generation use template libraries, which are prepared by ligating the DNA fragments with linker and/or adapters. Thus DNA fragments are not cloned into the host cells before taking the sequencing [27]. In crop plants, Roche 454 pyrosequencing, Illumina (Solexa) HiSeq and MiSeq sequencing, Sequencing by Oligonucleotide Ligation and Detection (SOLiD), DNA nanoball sequencing by BGI Retrovoluty and Ion torrent are commonly used platforms of second generation sequencing (SGS) and each one has advantages and disadvantages that have been discussed earlier in different reviews [27, 28]. The draft genome sequencing of at least 421 plant species including food and horticulture crops have been published using the second generation

sequencing platforms as observed on March 2020 at [www.plabipd.de](http://www.plabipd.de) website [18, 29, 30]. In lentil, different second generation technologies have been used to develop the draft genome sequence and sequencing of transcriptomes [2, 10, 22, 31–33].

### Third-generation sequencing

In the past years, the draft genome sequences of several crop species have been developed by assembling the short DNA fragments into contigs and scaffolds that are generated by using second generation technologies. However, polyploid nature, presence of repetitive DNA sequence and large genome size of many crop species make difficult to assemble long chromosome sequences by using short DNA fragments and hence these fragments are not properly mapped to their chromosomal locations [34–36]. However, the advancement in DNA sequencing led to the availability of third generation sequencing technologies. These emerging technologies sequence single molecule and are able to generate long DNA sequences and scaffolds that can cover the complete chromosome [37]. The optical mapping [38], chromosome conformation capture [39] and DNA dilution-based technologies [40, 41] are such third generation sequencing technologies. These are used by several sequencing platforms: (i) single-molecule real-time (SMRT) sequencing by Pacific Biosciences, (ii) helicosequencing by the Genetic Analysis System, (iii) nanopore sequencing by Oxford Nanopore Technologies (MinION and PromethION) and (iv) NGS by electron microscopy. These platforms are replacing the SGS platforms rapidly and in coming years these could replace SGS due to their cost-effectiveness and rapidness [42]. Although, in lentil, a draft genome sequence has been developed based on short DNA reads generated using second generation sequencing methods [31], it has many large gaps. These gaps can be filled by generating longer DNA sequence rapidly and more cost effectively using third generation sequencing (TGS) technologies. Moreover, these technologies can help to increase the accuracy for discovering SNPs, and reduce the chance of finding the false SNP and sequencing/genotyping of many markers in a single step in lentil [43].

### Current application of NGS

Next-generation sequencing technologies have been deployed for enriching genomic resources. In lentil, NGS based SNP markers developed in lentil have been used for different purposes [44–48]. These have been discussed below.

## Identification of unigenes, transcripts and functional markers

In lentil, second generation sequencing platforms have been used to accelerate the development of molecular markers, unigenes and transcripts (Table 1). Initially, second generation sequencing technologies have been used to develop simple sequence repeat (SSR)-containing ESTs, which were identified from consensus sequences [22]. In this study, 2393 EST-SSR markers were developed and a subset comprising 192 EST-SSR markers were validated across genotypes of cultivated and wild species [22]. Subsequently, these markers have been used to study the genetic diversity and identification of QTLs for agronomically important traits in several studies [7, 8]. Further second generation sequencing based transcriptome analysis led to identification of SSRs and SNPs from functional genomic regions of lentil [2, 6, 10, 33] and used to develop a SNP genotyping platform carrying SNP array chips for genotyping—the lentil genotypes for genetic diversity analysis and molecular mapping of QTLs for agronomically important traits [2, 49].

## Phylo-genetic relationship studies

In lentil, NGS based SNPs allowed to identify the genetic relationship between different species and their time of divergence. The SNPs from coding regions recognize genetic variation between different species due to the changes in the coding regions of the genome. The rate of substituted synonymous SNPs between *L. culinaris* and *L. ervoides* species of lentil estimated their separation from each other 677,000 years ago, while *L. culinaris* was separated from *M. truncatula* 38 million years ago (MYA) that was also similar to other earlier studies [2, 53]. However, advanced next generation sequencing platforms may further increase the amount of genome sequence data of multiple genotypes of *L. culinaris* and other available *Lens* species in near future that can help to elucidate the nature of the domestication process for this crop [52].

## Identification of genetic diversity

NGS technologies have accelerated the development of SNPs widely in lentil that has been used to assess the genetic diversity and determine the genetic relationship among different species. In a study, 384 SNP markers distributed genome-wide, which were developed earlier using NGS, have been used to genotype 505 cultivars and landraces and clustered lentil cultivars according to their geographical origin. However, land races were not grouped clearly and landraces belonging to the Mediterranean region especially from Greece and Turkey, showed high level of genetic diversity [44]. In another study, 1194 SNPs have been used to estimate genetic diversity among 352 accessions belonging to 54 diverse countries. These genotypes grouped broadly according to their geographical origin including Mediterranean Basin (also including the Nile valley from Egypt to Ethiopia), subtropical Asia and northern temperate [46]. In a recent study, genotyping by sequencing (GBS) following NGS has been used to study the association of Mediterranean gene pool with specific geographic or phenotypic patterns. For this, 6693 SNPs characterized 349 accessions of Mediterranean gene pool according to geographic patterns and phenotypic traits. This study suggested that cultivation of lentil in Mediterranean countries has been introduced by post-domestication routes and selection of improved types shaped the structure of lentil population [52]. In these studies, considerable genetic diversity has been observed among the genotypes of the cultivated gene pool and hence NGS based SNP genotyping can be used to mine diverse genotypes for hybridization in the breeding program for genetic improvement [46]. In another study, genetic diversity was assessed among 467 wild and cultivated accessions of lentil belonging 10 diverse geographical regions using 422,101 high-confidence SNP markers. In this study, these germplasm accessions grouped into four clusters, which were poorly correlated with geographical origin. However, accessions belonging to *L. nigricans* showed the greatest allelic diversity compared to all other species/subspecies [54].

**Table 1** Genetic resources developed in lentil through NGS

NGS platforms	SSR	SNP	Unigenes/ Transcripts	References
454 pyrosequencing	–	44,879	–	[2]
Illumina HiSeq2500 platform	9949	8260	–	[6]
Illumina HiSeq 2000 platform	141,050	194,178	–	[10]
Roche 454 GS-FLX Titanium technology	2415	–	25,592	[22]
Ion Proton sequencer	–	–	317,412	[32]
Illumina HiSeq 2500 platform	–	6306	48,453	[33]
Illumina paired-end sequencing,	8722	–	20 009/42 196	[50]
Illumina HiSeq 2000 sequencing system	–	50,960	–	[51]
Illumina HiSeq 2500 system (Illumina)	–	6693	–	[52]

## Linkage maps and mapping of genes/QTLs

High density molecular maps developed in lentil using SNP markers, which were generated using NGS. In a study, 376 SNPs selected from 50,960 SNPs developed through NGS based transcriptome analysis used as markers along with other SSR and ISSR markers for making the dense molecular map. This map covered 432.8 cM distance having average distance of 1.11 cM between two markers [51]. Using same NGS approach, 6306 high quality polymorphic SNPs used to construct a high density inter-specific linkage map covering 5782.19 cM distance. This molecular map helped to identify the major QTLs controlling seed coat spotting (Scp), flower color (FC), stem pigmentation (SP), time to flowering, seed size, and ascochyta blight resistance (Table 2). A major QTL for flowering time explained 55.73% of phenotypic variance and 3 major QTLs explained 35.48% of phenotypic variance for seed size. Among these QTLs, one has been associated with stem pigmentation. These QTLs were present in the chromosomal region covering 20.8 Mb, which was annotated with 366 genes [33].

## Identification of candidate genes and pathways

Identification of candidate genes for tolerance to several biotic and abiotic stresses and other agronomic traits could be accelerated with the use of NGS technologies [6, 10, 32]. These technologies have been used to sequence genes expressed at a specific time and growth condition/stage. The gene sequences have been annotated with sequences of known function leading to the identification of candidate genes or pathways associated with a phenotype. Genes expressed by plants at different stages under a particular stress have been identified easily and rapidly through NGS based transcriptome analysis. In lentil, NGS based transcripts sequencing identifies up- and down-regulated genes and pathways expressed at seedling stage under drought and heat conditions [6, 10]. In these studies, the genes encoding plasmodesmata callose-binding protein 3 (PDCB), phosphatidylinositol/phosphatidylcholine transfer protein SFH13,

CDP-diacylglycerol–glycerol-3-phosphate 3-phosphatidyltransferase 1 chloroplastic, probable glycerol-3-phosphate acyltransferase 2 (GPAT2), O-acyltransferase, phosphatidylcholine diacylglycerol choline phosphotransferase have been identified through transcriptome analysis. These genes showed up-regulation under heat stress conditions and one gene encoding pyruvate phosphate dikinase was related to shikimate pathway, which produces secondary metabolites responsible for heat tolerance in plants [10]. However, in lentil, heat stress usually affects the terminal stage of growth and development. Therefore, using the same approach, up-regulated and down-regulated genes and pathways have been identified at reproductive stage under heat stress conditions in the field (J. Kumar, personal communication). These genes/pathways were different from the genes/pathways expressed at seedling stage under heat stress conditions in the aforementioned studies (J. Kumar, personal communication). Similarly, candidate genes associated with different functional groups including molecular process, cellular and biological process expressed under drought conditions have also been identified through transcriptome analysis in lentil [6]. In lentil, a reference set of unigenes comprising of 58,986 contigs and scaffolds were used to compare with databases related to genes/proteins available in public domain and identified the candidate genes associated with mechanisms related to boron toxicity tolerance and time of flowering [50]. In another study, comparative mapping of flanking sequences of SNPs markers associated with boron toxicity led to identification of candidate genes for boron transporter-like protein and MIP family [49]. NGS based transcriptome analysis have also been used to uncover the generic basis of disease resistance and identified candidate genes involved in defense-response in lentil. For example, differential gene expression studies between resistant and susceptible genotypes using RNA-seq approach [32, 55] as well as MACE technique [56] have allowed to identify a number of key genes that played an important role in providing of resistance against ascochyta blight. These genes were involved in different defence response functions (see Table 3). Khorramdelazad et al. [32] observed that the resistant genotype has ability to detect and respond signalling to disease infection much earlier and faster than susceptible genotype and structural defence-related genes are over-expressed in lentil. In lentil, NGS has been used to identify the genes encoding disease resistance proteins in the host and later virulence proteins (i.e. effectors) in the pathogen. A complex interaction between resistance and effector genes has been observed for developing anthracnose disease in lentil [57]. In this study, 26 resistance genes including *suppressor of npr1-1*, *constitutive 1 (NBS-LRR)* and *dirigent* (resistance response protein) have been identified in the host after the infection with an isolate of the virulent race 0 of *Colletotrichum lentis* [57]. Transcriptome analysis identified

**Table 2** Identification of QTLs for economically important traits using NGS in lentil. [source 33]

Trait	QTL Name	LG group	PVE
Seed coat spotting	Scp	LG 6	85.07
Flower color	FC	LG6	84.2
Stem pigmentation	SP	LG 1	33.96
Time to flowering	TF	LG 6	55.73
Seed Size	Q1, Q2, Q3	LG 1, LG5a, LG5b	35.48
Ascochyta blight	AS-Q1, AS-Q2, AS-Q3	LG 6	28.46

**Table 3** Key candidate genes responsible for defence-response to *A. lentis* (Ascochyta blight) in lentil [32]

Name of the genes	Functional group	Type of genotype having up-regulated transcripts
LRR receptor-like kinase ( <i>LRR-RK</i> )	Recognition: Receiving & signalling molecules	Resistant
NB-ARC domain disease resistance protein ( <i>NB-ARC</i> )	Recognition: Receiving & signalling molecules	Susceptible
Delta (12)-fatty-acid desaturase ( <i>Delta (12)-FAD</i> )	Early structural response: Papilla formation	Susceptible
Xyloglucan endotransglucosylase/ hydrolase ( <i>XTH</i> )	Early structural response: Papilla formation	Resistant
PR protein 10 – Bet V I type ( <i>PR10</i> )	Early Biochemical response: Pathogenesis-related & anti-fungal compounds	Resistant
Auxin-repressed protein ( <i>ARP</i> )	Early Biochemical response: Pathogenesis-related & anti-fungal compounds	Susceptible
Polygalacturonase inhibitor ( <i>PGIP</i> )	Early Biochemical response: Pathogenesis-related & anti-fungal compounds	Susceptible
Superoxide dismutase ( <i>SOD</i> )	Early Biochemical response: Pathogenesis-related & anti-fungal compounds	Susceptible
Heterogeneous nuclear ribonucleoprotein RNA binding motif ( <i>RBP-hnRNP</i> )	Regulation of defence-related transcripts: DNA & RNA binding proteins	Susceptible
Gibberellin signalling DELLA protein ( <i>DELLA</i> )	Hypersensitive response: Cell death & SAR	Susceptible
Gibberellin receptor ( <i>GIDI</i> )	Hypersensitive response: Cell death & SAR	

candidate resistance genes encoding calcium transporting ATPase and glutamate receptor 3.2 and another candidate gene with unknown function for the susceptibility to stemphyllium blight disease in lentil. These candidate genes have been validated through bulk segregant analysis in a mapping population used previously for identification of QTL for this disease [58]. In another study, transcriptome analysis of *Ascochyta lentils* infected plants identified 18 candidate genes and also reported critical role of lignin biosynthesis and jasmonic acid (JA) pathways in resistance reaction while gibberellins synthesis has been observed to be related with susceptibility to pathogen [56]. These candidate genes can be used for different purposes including identification of genes for pathway-specific expression analysis, genetic modification approaches, development of resources for genotypic analysis, and assistance in the annotation of a future lentil genome sequence and also can be useful for developing the diagnostic functional markers for breeding.

## Prospects

### Accelerating genetic gain in breeding populations

Advancement of next generation sequencing (NGS) led to generation of genome sequences and discovery of SNPs rapidly and cost-effectively. It accelerated sequencing/re-sequencing of the entire genome or a part of it in a large number of genotypes for identification of polymorphism in a crop species. Therefore, in the past years, high throughput genome-wide SNP genotyping platforms such

as genotype-by-sequencing (GBS) helped to select the high breeding value genotypes on the basis of its genotypic constitution in breeding populations [59] and used earlier in several crops including legumes for genomic selection (GS) [60]. It has been found that GBS covers much greater fraction of the genome and capture population/family specific genetic variation than other SNPs genotyping methods currently used in crop plants [61]. Therefore, NGS based GBS has been identified as ideal for GS due to its flexibility, low cost and genomic estimated breeding value (GEBV) prediction accuracy. The prediction accuracy (0.1 to 0.2) of GEBV has been observed higher through NGS based genotyping compared to other established marker platforms [59]. In soybean, prediction accuracy for grain yield, assessed using cross validation, was estimated to be 0.64, indicating good potential for using GS for grain yield in soybean [62]. Therefore, NGS based GS helps to enhance genetic gain in the following ways.

- NGS has led to genotyping of a large number of accessions, which helps to uncover a wide genetic variation available in the genome. This increases the power of GS. More than 1000 SNPs can be screened in single run for a sample through NGS and hence GS can be performed for minor QTLs of the traits having low heritability [63]. Consequently, higher genetic gain is possible through NGS based GS.
- Selection based on marker/genotype profile helped to identify the individuals in a breeding population with high breeding values by deploying higher selection intensity and accuracy for quantitative traits. This can give

higher genetic gains per year compared to phenotypic selection [64, 65].

- Genomic selection also helps to get better genetic gain for those traits that have a long generation time or are difficult to evaluate (i.e., insect resistance, bread making quality, and others). It also makes GS cheaper. Genotyping and environment interactions highly influence quantitatively inherited traits. Therefore, considering genomic loci interacting with environments during genomic selections can enhance the prediction accuracy and genetic gain. It predicts breeding values of lines on the basis of genome wide marker profile without phenotyping in the field and helps to select improved breeding lines across target environments. Thus discovering large number of SNPs and their use in genotyping across large breeding population through NGS especially through GBS can replace the phenotypic selection with GS in coming years [61]. In chickpea, accuracy of GS was predicted for yield related traits from 0.138 (seed yield) to 0.912 (100 seed weight) and inclusion of genotype  $\times$  environment (G  $\times$  E) interaction in GS models improved prediction accuracy [66, 67].

In lentil, different GS models and prediction scenarios have been evaluated for designing GS strategies for breeding. Models included G  $\times$  E interactions and multiple traits showed higher prediction accuracy for a low heritability trait. Moreover, prediction accuracies within-population and across-environments was observed moderate to high in lentil [68]. This study suggested that GS can accelerate genetic gain within population and across environments, if applied to larger population size in lentil [68]. Therefore, use of NGS can help to accelerate genetic gain rapidly and cost effectively through GS in lentil by genotyping the large breeding population with large number of markers.

### Towards the development of pan-genome and super pan-genome

A pan-genome refers to the total set of genes for individuals of a species [69]. These genes can be grouped as core and dispensable genes. The core genes are conserved across all individuals and hence these are usually housekeeping genes that are responsible for essential cellular functions [70]. The pan-genome with core genes is also known as closed pan-genome. The dispensable genes in the pan genome are present either in a specific individual or few individuals, but not in all individuals. These genes are functionally associated with various adaptive traits like tolerance to biotic and abiotic stresses, receptor and antioxidant activity, gene regulation, and signal transduction [71–76]. Therefore, these genes contribute more to the diversity of a species and evolve faster than core genes [70]. The non-synonymous

and synonymous substitutions across dispensable genes have been observed higher in soybean [72]. In lentil, substitutions of synonymous and non synonymous SNPs in the coding regions cause genetic variation [2]. The concept of pan-genome is based on capturing the genetic variation, particularly structural variation, available in the gene content of individuals belonging to the same species [69]. It became feasible due to advancement in NGS technology that allowed sequencing/resequencing of multiple accessions belonging to one or more species. These structural variations (SVs) include presence/absence variations, copy number variations (CNVs) and other form of variations such as inversion, trans-versions and inter/intra chromosomal translocations [77–80]. Pan-genome analysis involves genomic sequences of multiple accessions belonging to a cultivated species, while genomic sequences of accessions belonging to each species available within a genus are included in the super pan-genome analysis [81, 82]. As wild relatives have many specific traits, super-pangenome analysis can give a unique opportunity of exploiting the available structural genomic variations of a genus in the genetic improvement by associating them with a trait of interest through genome wide association analysis [82]. The pan genomics have been used in several crops including soybean [72], *B. oleracea* [74], *B. napus* [76], maize [83], rice [84], wheat [85], sesame [86], sunflower [87] and tomato [88] for genetic diversity analysis. In lentil, NGS technology has been used to sequence RNAs of many accessions belonging to cultivated and wild species and a partial genome analysis identified SNPs/InDels in the genome. However, SNPs/InDels do not alone to contribute of all genetic diversity available within species [89, 90]. As reference genome sequence is now available in lentil [31] and structural variations in the chromosomes due to translocations have been reported within and between the species earlier [91], re-sequencing of diverse accessions can help to identify prevalence of structural variations at genomic level in lentil as observed in soybean [92, 93] and pigeonpea, a pulse crop [94]. The following factors determine the identification of dispensable genes in the pan-genome and they will be useful to consider during the development of a pan- genome in lentil.

### Size of pan-genome

It is determined by the number of sequenced individuals involved in the pan-genome because use of large number of sequenced individuals in the pan genome increases percentage of dispensable genes and decreases the percentage of core genes. For example, in rice, a pan genome with 48,098 genes developed from 3010 accessions possess 41% dispensable genes, while another pan-genome of 40,362 genes had 8% dispensable in three accessions [73, 84].

## Type of accessions

It has been observed that the use of closely related accessions does not complete the size of the pan-genome. Therefore it is better to use diverse accessions and use the accessions of wild species along with accessions of cultivated species can help to develop a larger pan-genome with a higher percentage of dispensable genes rather than using accessions of cultivated species. For example, in rice, a pan-genome with 66 accessions of cultivated (*Oryza sativa*) and wild species (*Oryza rufipogon*) contained 42,580 genes, in which 38% were dispensable [95], while another pan-genome with accessions of cultivated species contained 40,362 genes involving 7.83% as dispensable genes [73]. Moreover, use of accessions of wild species in the development of a pan-genome can help to identify the genes lost during the domestication of a crop as suggested by Khan et al. [82].

## Accelerating the use of gene-based markers in breeding

Before the advent of NGS, re-sequencing of unigene-derived amplicons or ESTs through conventional sequencing has led to development of gene-based SNP markers, which were validated through PCR [96, 97]. However, next generation sequencing based transcriptome analysis has identified candidate genes expressed under several biotic and abiotic stresses in lentil [6, 10]. Moreover, these functional unigene sequences also carried the SNPs and SSRs, which became useful resources for developing gene-based/functional markers [2, 6, 10, 22, 50–52, 98]. The SNPs identified through transcriptome analysis have been used to develop array for genotyping and identified their association with traits of interest [2, 49]. Moreover, through NGS it is possible to re-sequence many candidate genes in large number of genotypes in a single run at lower cost, which can be used to associate the SNPs of candidate genes with target traits [49]. The candidate gene sequences can also be used to develop PCR based markers for plant breeders for easily genotyping the large breeding populations as done in other crops [99]. Also, use of functional markers in marker trait association mapping can help in development of the perfect markers for breeding. In lentil, comparative genomic mapping of flanking sequences of SNPs associated with target trait with the genome sequences of other species (*Medicago truncatula*, soybean and *Arabidopsis thaliana*) led to identification of candidate genes that are functionally associated with boron toxicity tolerance. These candidate gene sequences can potentially be used for developing a gene based diagnostic marker in lentil [50].

## Capturing exome variation

A draft genome sequence of lentil is available for reference [31]. Capturing the genetic variation across whole genome through whole genome sequencing of many accessions is challenging due to difficulties in assembling the large (4063 Mb) and complex lentil genome that is caused by gene duplication, chromosomal arrangements and repetitive elements [100]. However, capturing genetic variation available in the coding regions of the genome is useful for making genetic improvement because it carries genes controlling many traits of agronomic importance. Therefore, coding regions are more important than non coding regions for breeding of those crops that has large and complex genome [101]. In lentil, 3.2% genome (i.e. 130 Mb) is made of genic region [100], which can be common targets for lentil research. Therefore, in lentil, efforts have been made to capture the genetic variation available in these genic regions and exome capture arrays comprising of 85 Mb have been developed in lentil [100]. Such exome capture arrays can help to sequence only the protein coding regions of the genome rather than whole genome and hence it is cost effective sequencing method [102]. Developed exome arrays have been used to identify genetic variation among 38 diverse accessions of lentil including 16 accessions of wild species [100]. They also suggested use of exome capture arrays in downstream research including identification of genetic relationships within the *Lens* genus, gene discovery, development of genic markers for tracking, identification and selection of beneficial traits in breeding population, identification of genes for adaptive traits in wild relatives and use in DNA barcoding in future [100].

## Concluding remarks

In lentil, NGS technologies have been used to identify SSRs, SNPs and candidate genes expressed under specific environmental conditions. The SSR and SNPs have been used to develop markers to describe genetic variation of germplasm collections and their association with different phenotypic traits, including seed quality, disease resistance, and micronutrient concentration [44–48]. However, use of NGS technology in lentil breeding program is not widely used compared to other crops despite low genotyping cost for a large number of individuals. In other crops, NGS technologies have been used to develop pan genome and super pan-genome by re-sequencing or sequencing large number of accessions. In lentil, these efforts have not been made due to genome complexity that impairs assembling and scaffolding of short reads generated on the basis of second generation sequencing. Third-generation single-molecule sequencing technologies have emerged that reduced the

cost of sequencing and can be used for sequencing the long DNA fragments leading to easy assembling and scaffolding of complex genome. Hence use of third generation sequencing can overcome problems associated with large genome size of lentil and in coming years, use of NGS will boost genetic gain in lentil.

**Funding** This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

### Compliance with ethical standards

**Conflict of interest** All authors declare that they have no conflict of interest.

**Ethical approval** This article does not contain any studies with human participants or animals performed by any of the authors.

### References

- Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. *Plant Mol Biol Rep* 9:210–220. <https://doi.org/10.1007/BF02672069>
- Sharpe AG, Ramsay L, Sanderson LA, Fedoruk MJ, Clarke WE, Li R, Kagale S, Vijayan P, Vandenberg A, Bett KE (2013) Ancient orphan crop joins modern era: gene-based SNP discovery and mapping in lentil. *BMC Genomics* 14:192. <https://doi.org/10.1186/1471-2164-14-192>
- Kumar S, Ali M (2006) GE interaction and its breeding implications in pulses. *Botanica* 56:31–36
- Kumar S, Rajendran K, Kumar J, Hamwiah A, Baum M (2015) Current knowledge in lentil genomics and its application for crop improvement. *Front Plant Sci* 6:78. <https://doi.org/10.3389/fpls.2015.00078>
- Bett K, Ramsay L, Sharpe A, Cook D, Penmetsa RV, Verma N (2014) Lentil genome sequencing: establishing a comprehensive platform for molecular breeding. In: *Proceedings of international food legumes research conference (IFLRC-VI) and ICCLG-VII*. Crop Development Center, SK, Saskatoon, p. 19
- Singh D, Singh CK, Taunk J, Tomar RS, Chaturvedi AK, Gaikwad K, Pal M (2017) Transcriptome analysis of lentil (*Lens culinaris* Medikus) in response to seedling drought stress. *BMC Genomics* 18:206. <https://doi.org/10.1186/s12864-017-3596-7>
- Kumar J, Gupta S, Biradar RS, Gupta P, Dubey S, Singh NP (2018) Association of functional markers with flowering time in lentil. *J Appl Genet* 59:9–21. <https://doi.org/10.1007/s13353-017-0419-0>
- Kumar J, Gupta S, Gupta DS, Singh NP (2018) Identification of QTLs for agronomic traits using association mapping in lentil. *Euphytica* 214:75. <https://doi.org/10.1007/s10681-018-2155-x>
- Kumar J, Basu PS, Gupta S, Dubey S, Gupta DS, Singh NP (2018) Physiological and molecular characterisation for high temperature stress in *Lens culinaris*. *Func Plant Biol* 45:474–487. <https://doi.org/10.1071/FP17211>
- Singh D, Singh CK, Taunk J, Jadon V, Pal M, Gaikwad K (2019) Genome wide transcriptome analysis reveals vital role of heat responsive genes in regulatory mechanisms of lentil (*Lens culinaris* Medikus). *Sci Rep* 9:1–9. <https://doi.org/10.1038/s41598-019-49496-0>
- Varshney RK, Chen W, Li Y, Bharti AK, Saxena RK, Schlueter JA, Donoghue MT, Azam S, Fan G, Whaley AM, Farmer AD (2012) Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nat Biotechnol* 30:83. <https://doi.org/10.1038/nbt.2022>
- Varshney RK, Song C, Saxena RK, Azam S, Yu S, Sharpe AG, Cannon S, Baek J, Rosen BD, Tar'an B, Millan T (2013) Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nat Biotechnol* 31:240. <https://doi.org/10.1038/nbt.2491>
- Das S, Upadhyaya HD, Bajaj D, Kujur A, Badoni S, Kumar V, Tripathi S, Gowda CL, Sharma S, Singh S, Tyagi AK (2015) Deploying QTL-seq for rapid delineation of a potential candidate gene underlying major trait-associated QTL in chickpea. *DNA Res* 22:193–203. <https://doi.org/10.1093/dnares/dsv004>
- Deokar A, Sagi M, Daba K, Tar'an B (2018) QTL sequencing strategy to map genomic regions associated with resistance to ascochyta blight in chickpea. *Plant Biotechnol J* 17:275–288. <https://doi.org/10.1111/pbi.12964>
- Singh VK, Khan AW, Saxena RK, Kumar V, Kale SM, Sinha P, Chitikineni A, Pazhamala LT, Garg V, Sharma M, Sameer Kumar CV (2016) Next-generation sequencing for identification of candidate genes for *Fusarium* wilt and sterility mosaic disease in pigeonpea (*Cajanus cajan*). *Plant Biotechnol J* 14:1183–1194. <https://doi.org/10.1111/pbi.12470>
- Pandey MK, Khan AW, Singh VK, Vishwakarma MK, Shashidhar Y, Kumar V, Garg V, Bhat RS, Chitikineni A, Janila P, Guo B (2017) QTL-seq approach identified genomic regions and diagnostic markers for rust and late leaf spot resistance in groundnut (*Arachis hypogaea* L.). *Plant Biotechnol J* 15:927–941. <https://doi.org/10.1111/pbi.12686>
- Clevenger J, Chu Y, Chavarro C, Botton S, Culbreath A, Isleib TG, Holbrook CC, Ozias-Akins P (2018) Mapping late leaf spot resistance in peanut (*Arachis hypogaea*) using QTL-seq reveals markers for marker-assisted selection. *Front Plant Sci* 9:83. <https://doi.org/10.3389/fpls.2018.00083>
- Unamba CI, Nag A, Sharma RK (2015) Next generation sequencing technologies: the doorway to the unexplored genomics of non-model plants. *Front Plant Sci* 6:1074
- Varshney RK, Pandey MK, Bohra A, Singh VK, Thudi M, Saxena RK (2019) Toward the sequence-based breeding in legumes in the post-genome sequencing era. *Theor Appl Genet* 132:797–816. <https://doi.org/10.1007/s00122-018-3252-x>
- Afzal M, Alghamdi SS, Migdadi HH, Khan MA, Mirza SB, El-Harty E (2019) Legume genomics and transcriptomics: from classic breeding to modern technologies. *Saudi J Biol Sci* 27(1):543–555. <https://doi.org/10.1016/j.sjbs.2019.11.018>
- Le Nguyen K, Grondin A, Courtois B, Gantet P (2019) Next-generation sequencing accelerates crop gene discovery. *Trend Plant Sci* 24(3):263–274. <https://doi.org/10.1016/j.tplants.2018.11.008>
- Kaur S, Cogan NO, Pembleton LW, Shinozuka M, Savin KW, Materne M, Forster JW (2011) Transcriptome sequencing of lentil based on second-generation technology permits large-scale unigenic assembly and SSR marker discovery. *BMC Genomics* 12:265. <https://doi.org/10.1186/1471-2164-12-265>
- Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74:5463–5467. <https://doi.org/10.1073/pnas.74.12.5463>
- Fedoruk MJ, Vandenberg A, Bett KE (2013) Quantitative trait loci analysis of seed quality characteristics in lentil using single nucleotide polymorphism markers. *Plant Genome*. <https://doi.org/10.3835/plantgenome2013.05.0012>
- Mardis ER (2011) A decade's perspective on DNA sequencing technology. *Nature* 470:198–203. <https://doi.org/10.1038/nature09796>



26. Metzker ML (2010) Sequencing technologies —the next generation. *Nat Rev Genet* 11:31–46. <https://doi.org/10.1038/nrg2626>
27. Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M (2012) Comparison of next-generation sequencing systems. *BioMed Res Int*. <https://doi.org/10.1155/2012/251364>
28. Lam HY, Clark MJ, Chen R, Chen R, Natsoulis G, O’huallachain M, Dewey FE, Habegger L, Ashley EA, Gerstein MB, Butte AJ (2012) Performance comparison of whole-genome sequencing platforms. *Nat Biotechnol* 30:78–82. <https://doi.org/10.1038/nbt.2065>
29. Thottathil GP, Jayasekaran K, Othman AS (2016) Sequencing crop genomes: a gateway to improve tropical agriculture. *Trop Life Sci Res* 27:93
30. Chen F, Song Y, Li X, Chen J, Mo L, Zhang X, Lin Z, Zhang L (2019) Genome sequences of horticultural plants: past, present, and future. *Hortic Res* 6:1–23. <https://doi.org/10.1038/s41438-019-0195-6>
31. Bett K (2016) Lentil 1.0 and Beyond. PAG XXIV: plant and animal genomics conference, 8–13 January 2016, San Diego, California, USA
32. Khorramdelazad M, Bar I, Whatmore P, Smetham G, Bhaaskaria V, Yang Y, Bai SH, Mantri N, Zhou Y, Ford R (2018) Transcriptome profiling of lentil (*Lens culinaris*) through the first 24 hours of *Ascochyta lentis* infection reveals key defence response genes. *BMC Genomics* 19:108. <https://doi.org/10.1186/s12864-018-4488-1>
33. Polanco C, de Miera LE, González AI, García P, Fratini R, Vaquero F, Vences FJ, de la Vega MP (2019) Construction of a high-density interspecific (*Lens culinaris* x *L. odemensis*) genetic map based on functional markers for mapping morphological and agronomical traits, and QTLs affecting resistance to *Ascochyta* in lentil. *PLoS ONE* 14:e0214409. <https://doi.org/10.1371/journal.pone.0214409>
34. Neale DB, Wegrzyn JL, Stevens KA, Zimin AV, Puiu D, Crepeau MW, Cardeno C, Koriabine M, Holtz-Morris AE, Liechty JD, Martínez-García PJ (2014) Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biol* 15:R59. <https://doi.org/10.1186/gb-2014-15-3-r59>
35. Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin YC, Scofield DG, Vezzi F, Delhomme N, Giacomello S, Alexeyenko A, Vicedomini R (2013) The Norway spruce genome sequence and conifer genome evolution. *Nature* 497:579–584. <https://doi.org/10.1038/nature12211>
36. Salman-Minkov A, Sabath N, Mayrose I (2016) Whole-genome duplication as a key factor in crop domestication. *Nat Plants* 2:1–4. <https://doi.org/10.1038/nplants.2016.115>
37. Koren S, Phillippy AM (2015) One chromosome, one contig: complete microbial genomes from long-read sequencing and assembly. *Curr Opin Microbiol* 23:110. <https://doi.org/10.1016/j.mib.2014.11.014>
38. Schwartz DC, Li X, Hernandez LI, Ramnarain SP, Huff EJ, Wang YK (1993) Ordered restriction maps of *Saccharomyces cerevisiae* chromosomes constructed by optical mapping. *Science* 262:110–114. <https://doi.org/10.1126/science.8211116>
39. Burton JN, Adey A, Patwardhan RP, Qiu R, Kitzman JO, Shendure J (2013) Chromosome-scale scaffolding of *de novo* genome assemblies based on chromatin interactions. *Nat Biotechnol* 31:1119. <https://doi.org/10.1038/nbt.2727>
40. Amini S, Pushkarev D, Christiansen L, Kostem E, Royce T, Turk C, Pignatelli N, Adey A, Kitzman JO, Vijayan K, Ronaghi M (2014) Haplotype-resolved whole-genome sequencing by contiguity-preserving transposition and combinatorial indexing. *Nat Genet* 46:1343. <https://doi.org/10.1038/ng.3119>
41. Zheng GX, Lau BT, Schnall-Levin M, Jarosz M, Bell JM, Hindson CM, Kyriazopoulou-Panagiotopoulou S, Masquelier DA, Merrill L, Terry JM, Mudivarti PA (2016) Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. *Nat Biotechnol* 34:303. <https://doi.org/10.1038/nbt.3432>
42. Peterson TW, Nam SJ, Darby A (2010) Next gen sequencing survey. North America equity research. JP Morgan Chase & Co, New York
43. Stapley J, Reger J, Feulner PG, Smadja C, Galindo J, Ekblom R, Bennisson C, Ball AD, Beckerman AP, Slate J (2010) Adaptation genomics: the next generation. *Trends Ecol Evol* 25:705–712. <https://doi.org/10.1016/j.tree.2010.09.002>
44. Lombardi M, Materne M, Cogan NO, Rodda M, Daetwyler HD, Slater AT, Forster JW, Kaur S (2014) Assessment of genetic variation within a global collection of lentil (*Lens culinaris* Medik.) cultivars and landraces using SNP markers. *BMC Genet* 15:150. <https://doi.org/10.1186/s12863-014-0150-3>
45. Sudheesh S, Rodda MS, Davidson J, Javid M, Stephens A, Slater AT, Cogan NO, Forster JW, Kaur S (2016) SNP-based linkage mapping for validation of QTLs for resistance to ascochyta blight in lentil. *Front Plant Sci* 7:1604. <https://doi.org/10.3389/fpls.2016.01604>
46. Khazaei H, Caron CT, Fedoruk M, Diapari M, Vandenberg A, Coyne CJ, McGee R, Bett KE (2016) Genetic diversity of cultivated lentil (*Lens culinaris* Medik.) and its relation to the world’s agro-ecological zones. *Front Plant Sci* 7:1093. <https://doi.org/10.3389/fpls.2016.01093>
47. Khazaei H, Podder R, Caron CT, Kundu SS, Diapari M, Vandenberg A, Bett KE (2017) Marker–trait association analysis of iron and zinc concentration in lentil (*Lens culinaris* Medik.). *Plant Genome*. <https://doi.org/10.3835/plantgenome2017.02.0007>
48. Khazaei H, Fedoruk M, Caron CT, Vandenberg A, Bett KE (2018) Single nucleotide polymorphism markers associated with seed quality characteristics of cultivated lentil. *Plant Genome*. <https://doi.org/10.3835/plantgenome2017.06.0051>
49. Kaur S, Cogan NO, Stephens A, Noy D, Butsch M, Forster JW, Materne M (2014) EST-SNP discovery and dense genetic mapping in lentil (*Lens culinaris* Medik.) enable candidate gene selection for boron tolerance. *Theor Appl Genet* 127:703–713. <https://doi.org/10.1007/s00122-013-2252-0>
50. Sudheesh S, Verma P, Forster JW, Cogan NO, Kaur S (2016) Generation and characterization of a reference transcriptome for lentil (*Lens culinaris* Medik.). *Int J Mol Sci* 17:1887. <https://doi.org/10.3390/ijms17111887>
51. Temel HY, Gol D, Kahriman A, Tanyolac MB (2015) Single nucleotide polymorphism discovery through Illumina-based transcriptome sequencing and mapping in lentil. *Turk J Agric For* 39:470–488. <https://doi.org/10.3906/tar-1409-70>
52. Pavan S, Bardaro N, Fanelli V, Marcotrigiano AR, Mangini G, Taranto F, Catalano D, Montemurro C, De Giovanni C, Lotti C, Ricciardi L (2019) Genotyping by sequencing of cultivated lentil (*Lens culinaris* Medik.) highlights population structure in the Mediterranean gene pool associated with geographic patterns and phenotypic variables. *Front Genet* 10:872. <https://doi.org/10.1186/s12864-018-4488-1>
53. Lavin M, Herendeen PS, Wojciechowski MF (2005) Evolutionary rates analysis of Leguminosae implicates a rapid diversification of the major family lineages during tertiary. *Syst Biol* 54:575–594. <https://doi.org/10.1080/10635150590947131>
54. Dissanayake R, Braich S, Cogan NO, Smith K, Kaur S (2020) Characterization of genetic and allelic diversity amongst cultivated and wild lentil accessions for germplasm enhancement. *Front Genet*. <https://doi.org/10.3389/fgene.2020.00546>
55. Sari E, Bhadauria V, Ramsay L, Borhan MH, Lichtenzweig J, Bett KE, Vandenberg A, Banniza S (2018) Defense responses of lentil (*Lens culinaris*) genotypes carrying non-allelic ascochyta blight resistance genes to *Ascochyta lentis* infection. *PLoS ONE* 13:e0204124. <https://doi.org/10.1371/journal.pone.0204124>

56. García-García P, Vaquero F, Vences FJ, de Miera LE, Polanco C, González AI, Horres R, Krezdorn N, Rotter B, Winter P, de la Vega MP (2020) Transcriptome profiling of lentil in response to *Ascochyta lentis* infection. *Span J Agric Res* 17:0703. <https://doi.org/10.5424/sjar/2019174-14982>
57. Bhadauria V, Vijayan P, Wei Y, Banniza S (2017) Transcriptome analysis reveals a complex interplay between resistance and effector genes during the compatible lentil-*Colletotrichum lentis* interaction. *Sci Rep* 7:1–3. <https://doi.org/10.1038/srep42338>
58. Cao Z, Li L, Kapoor K, Banniza S (2019) Using a transcriptome sequencing approach to explore candidate resistance genes against stemphylium blight in the wild lentil species *Lens ervoides*. *BMC Plant Biol* 19:1–6. <https://doi.org/10.1186/s12870-019-2013-6>
59. Poland JA, Brown PJ, Sorrells ME, Jannink JL (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0032253>
60. Bassi FM, Bentley AR, Charmet G, Ortiz R, Crossa J (2016) Breeding schemes for the implementation of genomic selection in wheat (*Triticum* spp). *Plant Sci* 242:23–36. <https://doi.org/10.1016/j.plantsci.2015.08.021>
61. Bhat JA, Ali S, Salgotra RK, Mir ZA, Dutta S, Jadon V, Tyagi A, Mushtaq M, Jain N, Singh PK, Singh GP (2016) Genomic selection in the era of next generation sequencing for complex traits in plant breeding. *Front Genet* 7:221. <https://doi.org/10.3389/fgene.2016.00221>
62. Jarquín D, Kocak K, Posadas L, Hyma K, Jedlicka J, Graef G, Lorenz A (2014) Genotyping by sequencing for genomic prediction in a soybean breeding population. *BMC Genomics* 15:740. <https://doi.org/10.1186/1471-2164-15-740>
63. Xu Y, Crouch JH (2008) Marker-assisted selection in plant breeding: from publications to practice. *Crop Sci* 48:391–407. <https://doi.org/10.2135/cropsci2007.04.0191>
64. Zhong S, Dekkers JC, Fernando RL, Jannink JL (2009) Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. *Genetics* 182:355–364. <https://doi.org/10.1534/genetics.108.098277>
65. Heffner EL, Lorenz AJ, Jannink JL, Sorrells ME (2010) Plant breeding with genomic selection: gain per unit time and cost. *Crop Sci* 50:1681–1690. <https://doi.org/10.2135/cropsci2009.11.0662>
66. Roorkiwal M, Rathore A, Das RR, Singh MK, Jain A, Srinivasan S, Gaur PM, Chellapilla B, Tripathi S, Li Y, Hickey JM (2016) Genome-enabled prediction models for yield related traits in chickpea. *Front Plant Sci* 7:1666. <https://doi.org/10.3389/fpls.2016.01666>
67. Roorkiwal M, Jarquín D, Singh MK, Gaur PM, Bharadwaj C, Rathore A, Howard R, Srinivasan S, Jain A, Garg V, Kale S (2018) Genomic-enabled prediction models using multi-environment trials to estimate the effect of genotype × environment interaction on prediction accuracy in chickpea. *Sci Rep* 8:1–1
68. Haile TA, Heidecker T, Wright D, Neupane S, Ramsay L, Vandenberg A, Bett KE (2019) Genomic selection for lentil breeding: empirical evidence. *Plant Genome* 13:e20002. <https://doi.org/10.1002/tpg2.20002>
69. Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS, DeBoy RT (2005) Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome”. *Proc Natl Acad Sci USA* 102:13950–13955. <https://doi.org/10.1073/pnas.0506758102>
70. Tao Y, Zhao X, Mace E, Henry R, Jordan D (2019) Exploring and exploiting pan-genomics for crop improvement. *Mol Plant* 12:156–169. <https://doi.org/10.1016/j.molp.2018.12.016>
71. McHale LK, Haun WJ, Xu WW, Bhaskar PB, Anderson JE, Hyten DL, Gerhardt DJ, Jeddeloh JA, Stupar RM (2012) Structural variants in the soybean genome localize to clusters of biotic stress-response genes. *Plant Physiol* 159:1295–1308. <https://doi.org/10.1104/pp.112.194605>
72. Li YH, Zhou G, Ma J, Jiang W, Jin LG, Zhang Z, Guo Y, Zhang J, Sui Y, Zheng L, Zhang SS (2014) De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nat Biotechnol* 32:1045–1052. <https://doi.org/10.1038/nbt.2979>
73. Schatz MC, Maron LG, Stein JC, Wences AH, Gurtowski J, Biggers E, Lee H, Kramer M, Antoniou E, Ghiban E, Wright MH (2014) Whole genome *de novo* assemblies of three divergent strains of rice, *Oryza sativa*, document novel gene space of *aus* and *indica*. *Genome Biol* 15:506. <https://doi.org/10.1186/s13059-014-0506-z>
74. Golicz AA, Bayer PE, Barker GC, Edger PP, Kim H, Martinez PA, Chan CK, Severn-Ellis A, McCombie WR, Parkin IA, Paterson AH (2016) The pangenome of an agronomically important crop plant *Brassica oleracea*. *Nat Commun* 7:1–8. <https://doi.org/10.1038/ncomms13390>
75. Gordon SP, Contreras-Moreira B, Woods DP, Des Marais DL, Burgess D, Shu S, Stritt C, Roulin AC, Schackwitz W, Tyler L, Martin J (2017) Extensive gene content variation in the Brachypodium distachyon pan-genome correlates with population structure. *Nat Commun* 8:1–3. <https://doi.org/10.1038/s41467-017-02292-8>
76. Hurgobin B, Golicz AA, Bayer PE, Chan CK, Tirnaz S, Dolatabadian A, Schiessl SV, Samans B, Montenegro JD, Parkin IA, Pires JC (2018) Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid *Brassica napus*. *Plant Biotechnol J* 16:1265–1274. <https://doi.org/10.1111/pbi.12867>
77. Feuk L, Marshall CR, Wintle RF, Scherer SW (2006) Structural variants: changing the landscape of chromosomes and design of disease studies. *Human Mol Genet* 15:R57–R66. <https://doi.org/10.1093/hmg/ddl057>
78. Qi X, Li MW, Xie M, Liu X, Ni M, Shao G, Song C, Yim AK, Tao Y, Wong FL, Isobe S (2014) Identification of a novel salt tolerance gene in wild soybean by whole-genome sequencing. *Nat Commun* 5:4340. <https://doi.org/10.1038/ncomms5340>
79. Wang Y, Xiong G, Hu J, Jiang L, Yu H, Xu J, Fang Y, Zeng L, Xu E, Xu J, Ye W (2015) Copy number variation at the GL7 locus contributes to grain size diversity in rice. *Nat Genet* 47:944–948. <https://doi.org/10.1038/ng.3346>
80. Cook DE, Lee TG, Guo X, Melito S, Wang K, Bayless AM, Wang J, Hughes TJ, Willis DK, Clemente TE, Diers BW (2012) Copy number variation of multiple genes at *Rhg1* mediates nematode resistance in soybean. *Science* 338:1206–1209. <https://doi.org/10.1126/science.1228746>
81. Hu H, Yuan Y, Bayer PE, Fernandez CT, Scheben A, Golicz AA, Edwards D (2020) Legume pangenome construction using an iterative mapping and assembly approach. In: *Legume genomics*. Humana, New York, pp 35–47
82. Khan AW, Garg V, Roorkiwal M, Golicz AA, Edwards D, Varshney RK (2020) Super-pangenome by integrating the wild side of a species for accelerated crop improvement. *Trend Plant Sci* 25:148–158. <https://doi.org/10.1016/j.tplants.2019.10.012>
83. Hirsch CN, Foerster JM, Johnson JM, Sekhon RS, Muttoni G, Vaillancourt B, Peñagaricano F, Lindquist E, Pedraza MA, Barry K, de Leon N (2014) Insights into the maize pan-genome and pan-transcriptome. *Plant Cell* 26:121–135. <https://doi.org/10.1105/tpc.113.119982>
84. Wang W, Mauleon R, Hu Z, Chebotarov D, Tai S, Wu Z, Li M, Zheng T, Fuentes RR, Zhang F, Mansueto L (2018) Genomic

- variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* 557:43–49. <https://doi.org/10.1038/s41586-018-0063-9>
85. Montenegro JD, Golicz AA, Bayer PE, Hurgobin B, Lee H, Chan CK, Visendi P, Lai K, Doležel J, Batley J, Edwards D (2017) The pangenome of hexaploid bread wheat. *The Plant J* 90:1007–1013. <https://doi.org/10.1111/tpj.13515>
  86. Yu J, Golicz AA, Lu K, Dossa K, Zhang Y, Chen J, Wang L, You J, Fan D, Edwards D, Zhang X (2019) Insight into the evolution and functional characteristics of the pan-genome assembly from sesame landraces and modern cultivars. *Plant Biotechnol J* 17:881–892. <https://doi.org/10.1111/pbi.13022>
  87. Hübner S, Bercovich N, Todesco M, Mandel JR, Odenheimer J, Ziegler E, Lee JS, Baute GJ, Owens GL, Grassa CJ, Ebert DP (2019) Sunflower pan-genome analysis shows that hybridization altered gene content and disease resistance. *Nat Plants* 5:54–62. <https://doi.org/10.1038/s41477-018-0329-0>
  88. Gao L, Gonda I, Sun H, Ma Q, Bao K, Tieman DM, Burzynski-Chang EA, Fish TL, Stromberg KA, Sacks GL, Thannhauser TW (2019) The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. *Nat Genet* 5:1044–1051. <https://doi.org/10.1038/s41588-019-0410-2>
  89. Springer NM, Ying K, Fu Y, Ji T, Yeh CT, Jia Y, Wu W, Richmond T, Kitzman J, Rosenbaum H, Iniguez AL (2009) Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. *PLoS Genet*. <https://doi.org/10.1371/journal.pgen.1000734>
  90. Saxena RK, Von Wettberg E, Upadhyaya HD, Sanchez V, Songok S, Saxena K, Kimurto P, Varshney RK (2014) Genetic diversity and demographic history of *Cajanus* spp. illustrated from genome-wide SNPs. *PLoS ONE* 9:e88568. <https://doi.org/10.1371/journal.pone.0088568>
  91. Jha TB, Mahanti A, Ghorai A (2015) Karyotype analysis of Indian lentils through EMA based Giemsa staining. *Caryologia* 68:280–288. <https://doi.org/10.1080/00087114.2015.1109921>
  92. Lam HM, Xu X, Liu X, Chen W, Yang G, Wong FL, Li MW, He W, Qin N, Wang B, Li J (2010) Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat Genet* 42:1053–1059. <https://doi.org/10.1038/ng.715>
  93. Zhou Z, Jiang Y, Wang Z, Gou Z, Lyu J, Li W, Yu Y, Shu L, Zhao Y, Ma Y, Fang C (2015) Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat Biotechnol* 33(4):408–414. <https://doi.org/10.1038/nbt.3096>
  94. Varshney RK, Saxena RK, Upadhyaya HD, Khan AW, Yu Y, Kim C, Rathore A, Kim D, Kim J, An S, Kumar V (2017) Whole-genome resequencing of 292 pigeonpea accessions identifies genomic regions associated with domestication and agronomic traits. *Nat Genet* 49:1082–1088. <https://doi.org/10.1038/ng.3872>
  95. Zhao Q, Feng Q, Lu H, Li Y, Wang A, Tian Q, Zhan Q, Lu Y, Zhang L, Huang T, Wang Y (2018) Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. *Nat Genet* 50:278–284. <https://doi.org/10.1038/s41588-018-0041-z>
  96. Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS (2005) The effects of artificial selection on the maize genome. *Science* 308:1310–1314. <https://doi.org/10.1126/science.1107891>
  97. Batley J, Barker G, O'Sullivan H, Edwards KJ, Edwards D (2003) Mining for single nucleotide polymorphisms and insertions/deletions in maize expressed sequence tag data. *Plant Physiol* 132:84–91. <https://doi.org/10.1104/pp.102.019422>
  98. Verma P, Shah N, Bhatia S (2013) Development of an expressed gene catalogue and molecular markers from the de novo assembly of short sequence reads of the lentil (*Lens culinaris* Medik.) transcriptome. *Plant Biotechnol J* 11:894–905. <https://doi.org/10.1111/pbi.12082>
  99. Yang H, Tao Y, Zheng Z, Li C, Sweetingham MW, Howieson JG (2012) Application of next-generation sequencing for rapid marker development in molecular plant breeding: a case study on anthracnose disease resistance in *Lupinus angustifolius* L. *BMC Genomics* 13:318. <https://doi.org/10.1186/1471-2164-13-318>
  100. Ogutcen E, Ramsay L, von Wettberg EB, Bett KE (2018) Capturing variation in *Lens* (Fabaceae): development and utility of an exome capture array for lentil. *Appl Plant Sci* 6:e01165. <https://doi.org/10.1002/aps3.1165>
  101. Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, Shendure J (2011) Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet* 12:745–755. <https://doi.org/10.1038/nrg3031>
  102. Hodges E, Xuan Z, Balija V, Kramer M, Molla MN, Smith SW, Middle CM, Rodesch MJ, Albert TJ, Hannon GJ, McCombie WR (2007) Genome-wide *in situ* exon capture for selective resequencing. *Nat Genet* 39:1522–1527. <https://doi.org/10.1038/ng.2007.42>

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