Genomic resources for lupins are coming of age

Lars G. Kamphuis1,2,3 | Gagan Garg1 | Rhonda Foley1 | Karam B. Singh1,2,3

1CSIRO Agriculture and Food, Floreat, Western Australia, Australia
2Centre for Crop and Disease Management, Curtin University, Bentley, Western Australia, Australia
3The UWA Institute of Agriculture, The University of Western Australia, Crawley, Western Australia, Australia

Correspondence
Karam Singh, karam, CSIRO Agriculture and Food, Floreat 6014, Western Australia, Australia.
Email: karam.singh@curtin.edu.au

Funding information
Commonwealth Scientific and Industrial Research Organisation; Curtin University of Technology; Grains Research and Development Corporation, Grant/Award Number: 9176622; University of Western Australia

Abstract
Lupins are underutilised pulse crops subject to increasing interest for human consumption of the high-protein grain. They are also valued as a source of animal nutrition and make an excellent break crop in agricultural production systems. Like other orphan legumes, the genomic revolution has made it cost-effective to also apply modern genetic and genomic approaches in lupins. These have predominantly been conducted in the two major domesticated lupin species, namely, narrow-leafed lupin (NLL; *Lupinus angustifolius*) and white lupin (*Lupinus albus*), with transcriptome studies also emerging in other domesticated and undomesticated species. This review provides an overview of the current lupin genomic resources developed including two reference genomes for NLL and white lupin, several transcriptome resources and the development of pan-genomes for NLL and white lupin, and describes how these offer great potential to increase grain yield and quality for these recently domesticated pulse crops. Furthermore, we highlight the importance of lupins to further our understanding of many aspects of fundamental legume biology. Combined, this will aid breeders and growers to improve lupin crops to help meet the increasing demand for plant protein in more sustainable cropping systems.

KEYWORDS
*Lupinus albus*, *Lupinus angustifolius*, narrow-leafed lupin, pan-genome, plant protein, synteny, TILLING, white lupin

1 | INTRODUCTION

The genomic revolution over the last two to three decades has improved our understanding of many aspects of plant biology as well as accelerating on-going efforts to improve a wide range of crops. It has led to a suite of genomic resources from complete reference genomes for many crops, a growing number of comprehensive pan-genomes and valuable gene expression atlases based on transcriptomic data. Complimentary proteomic and metabolomic resources for lupins are also emerging, but are not as extensive as the genomic and transcriptomic resources. These genomic advances have been pivotal in greatly increasing our knowledge of plant biology, initially in the model plants such as *Arabidopsis*, and major crops such as maize and rice, but increasingly in many of the world’s other crops. In parallel, these genomic resources have paved the way for more and more sophisticated plant-breeding tools starting with the generation and deployment of various types of molecular markers such as simple sequence repeats (SSRs) and now for many crops the use of massive single nucleotide polymorphism (SNP)-based chip and/or direct sequencing approaches to analyse specific breeding populations, in combination with high-throughput phenotyping methods.

The focus of this review is on the relatively recent development and application of genomic resources for lupins, which are members of the legume family. Legumes are an important group of plants that belong to the *Fabacea* (or *Leguminosae*) family and are distributed throughout the world. Legumes constitute the third largest family of flowering plants based on total number of known species (Christenhusz & Byng, 2016) and are considered the second most
important family of crop plants from an economic perspective (Smykal, Coyne, Ambrose, Maxted, & Schaefer, 2014). Thus, the legume family has about 19,000 known species, and these make up about 7% of all flowering plant species. The legume family is varied with a range of plant types, and many legumes have been used throughout the world as either grain or forage crops. This has occurred for a combination of reasons including their unique ability to fix nitrogen thereby reducing the need for fertilizers and enriching the soil for subsequent nonlegume crops as well as providing disease breaks through crop rotations. Grain legumes are also valued for their nutritional benefits for both human and livestock consumption (Duranti & Gius, 1997; Proksina & Pilvere, 2019; Rubio & Molina, 2016; Santa & Rekkel, 2020), where they can provide valuable levels of major nutrients such as protein, dietary fibre and carbohydrates. Recently, interest in grain legumes has grown due to the relatively high levels of protein in their grains and the rapidly growing demand for plant protein in human diets (Lonnie & Johnstone, 2020; Murphy-Bokern et al., 2017; Sa et al., 2020).

Our knowledge of the legume family has benefited greatly through the advent of genomic approaches. Initially, these were focused on a couple of model legumes, Medicago truncatula and Lotus japonicus, as well as soybean, the major grain legume crop. However, as genomic approaches have become more and more accessible, in large part through the development of high-throughput and low cost, next generation sequencing technologies, legume scientists have been able to deploy these technologies to an increasing number of legume crops. Many of these crops had previously been considered as ‘orphan’ crops in terms of the genomic resources available and progress in crop improvement for these crops lagged behind those with good genomic resources. Consequently, the development of advanced genomic resources has greatly facilitated research and breeding efforts for these crops like chickpea, peanut and pigeon pea (Pandey et al., 2016; Sahruzaini et al., 2020; Varshney et al., 2019), while there is growing interest in using proteomic and metabolomic approaches for legume crop improvement (Ramalingam et al., 2015). There has also been an expansion in the use of comparative genomic approaches that have helped to give valuable insight into many aspects of legume biology such as syntenic relationships among various legume species and a deeper understanding of legume evolution. For example, a big recent push has involved the large scale surveying of sequence diversity in some legume species through low coverage sequencing of a range of accessions including many wild relatives and in some cases to use this information to form pan-genomes that help illustrate the conserved and variable gene sets for the species.

Lupins are important members of the legume family with several interesting properties. Lupins are important ecological ‘engineers’ and are often among the first plant species to colonise new areas such as damaged sites. They are also able to colonise impoverished and low nutrient soils due to their nitrogen fixing ability and take up phosphorous efficiently from soils. Lupins belong to the genus Lupinus in the genistoids clade of legumes, which diverged early in papilionoid legume evolution (Lavin et al., 2005), and so genomic advances in lupin species provide valuable information on how this somewhat distant branch of the legume phylogenetic tree has evolved. Lupin grain is also receiving considerable interest for its potential to help with major human health issues around diabetes, obesity and cardiovascular diseases (Lee et al., 2006). There are around 800 known lupin species that are widely distributed geographically. Thus, they are found extensively in North and South America, where they can inhabit a wide range of habitats, but are also in the Mediterranean region (Drummond et al., 2012). Particularly in S. America, wild lupin species have been found to have one of the highest levels of diversification of any plant species, and genomic advances are shedding light on how this is achieved (Nevado et al., 2016). From a crop perspective, only a few lupin species have been domesticated, and today, the most widely cultivated species are Lupinus angustifolius and Lupinus albus, whereas Lupinus luteus and Lupinus mutabilis are niche crops. Although overall lupin production has fluctuated over the last 20 years, over a million tonnes are produced every year with Australia being the largest producer, followed by Poland and the Russian Federation.

This review on lupin genomics focuses primarily on L. angustifolius and L. albus, both of whose genomes have been recently sequenced. The review also touches on some of the other genomic resources that have/are being developed for lupins (Figure 1) and how these are and will be used for furthering our understanding of lupin biology and evolution as well as drive step changes in lupin crop improvement to help meet the growing demand for high quality plant protein.

2 | OVERVIEW OF LUPIN TRANSCRIPTOME RESOURCES AND THEIR USE IN PREBREEDING/BREEDING ACTIVITIES

Some of the earliest genomic resources developed in lupin species were the generation of expressed sequenced tag (EST) libraries (Fischer et al., 2015; Foley et al., 2011; Kroc et al., 2014; Nelson et al., 2006, 2010; Parra-González et al., 2012; Phan et al., 2007; Tian et al., 2009). These EST datasets were predominantly used to convert expressed sequences into gene-based molecular markers. This was achieved through identifying either sequence or length-based polymorphisms between expressed sequence reads of different lupin lines and converting these into polymerase chain reaction (PCR)-based molecular markers (reviewed in detail in Kamphuis et al., 2020). They have been successfully employed to construct the first genetic maps in white lupin (L. albus) (Croxford et al., 2008; Phan et al., 2007) and yellow lupin (L. luteus) (Iqbal et al., 2019; Lichtin et al., 2020) as well as improve the density of genetic maps in narrow-leaved lupin (NLL) (Fischer et al., 2015; Nelson et al., 2006, 2010). Length-based polymorphic markers derived from these ESTs have also been demonstrated to be transferable across lupin species, where Parra-Gonzalez and colleagues (Parra-González et al., 2012) showed that EST-SSR markers from yellow lupin also generated PCR amplicons in L. hispanicus and L. mutabilis. This allowed them to assess the level of genetic diversity between different lupin species (Parra-González et al., 2012). Similarly, genetic maps constructed for white lupin (Phan et al., 2007) and NLL (Nelson et al., 2006, 2010) have
EST-derived molecular markers that generated amplicons in both species and even some that amplify across a range of legume species.

With a new age of sequencing technology, EST libraries have nowadays been replaced with RNA sequencing (RNAseq) data generated using next generation sequencers. In NLL, the first comprehensive RNA sequencing study of five different tissue types (root, stem, leaf, flower and seed) for three different accessions (Tanjil, Unicrop and P27255) was employed by Kamphuis et al. (2015). The RNA sequencing data for these three lines together with RNAseq data for breeding line ‘83A:476 generated through the 1,000 plants initiative (Cannon et al., 2015) were utilised to predict in silico gene-based molecular markers evenly distributed across the NLL genome (Kamphuis et al., 2015). The four lupin lines are the parents of two recombinant inbred line populations in NLL, a wide cross (83A:476 × P27255) and a narrow cross derived from two cultivars (Tanjil × Unicrop), and a set of 239 insertion/deletion markers and 96 SNP markers were generated of which 179 and 88, respectively, were incorporated in the reference genetic map (Hane et al., 2017; Kamphuis et al., 2015). A recent study by Książkiewicz and associates (Książkiewicz et al., 2017) utilised RNAseq data to generate a total of 3,597 transcriptome-derived markers to significantly improve the density of the reference genetic map for white lupin.

The generation of dense reference genetic maps in both NLL and white lupin has aided the identification of candidate genes for domestication traits such as vernalisation requirement and alkaloid content (Hane et al., 2017; Kamphuis et al., 2015; Kroc et al., 2014, 2019; Książkiewicz et al., 2017) as well as genes associated with resistance to key diseases of lupins such as anthracnose caused by Colletotrichum lupini and Phomopsis stem blight caused by Diaporthe toxica (Fischer et al., 2015; Hane et al., 2017; Książkiewicz et al., 2017). Similarly, genetic maps developed in yellow lupin and quantitative trait locus (QTL) analysis for flowering time and anthracnose resistance have identified candidate genes for these traits (Iqbal et al., 2019; Lichtin et al., 2020). More recently, massive analysis of cDNA ends (MACE) was employed for 126 NLL lines surveyed in the field for several agronomic traits to conduct a genome wide association study (GWAS) to identify genes associated with these traits (Plewinski et al., 2020). Strong marker trait associations were identified for the start and end to flowering, maturity, plant height, yield and total grain weight, where the deletion in the major flowering time locus LanFtc1 (Taylor et al., 2019) was strongly associated with both the start and end to flowering as well as maturity (Plewinski et al., 2020). The MACE technique was also used for expression QTL mapping in NLL, constituting the first e-QTL survey in lupins (Plewinski et al., 2019). This study highlighted numerous genes co-regulated at major domesticated loci, including one conferring low alkaloid content.

EST and RNA sequencing libraries both investigate global gene expression in a species, and different treatments can be investigated to determine changes in global gene expression. The latter is referred to as differential gene expression analysis, and these have become popular over the last decade in lupin species with Table 1 summarising RNA-sequencing studies generated to date. These transcriptome studies have also proven extremely useful for the annotation of the two reference genome assemblies for NLL (L. angustifolius) and white lupin (L. albus) (Hane et al., 2017; Hufnagel, Marques et al., 2020; Xu et al., 2020) as discussed in more detail later in this review. The following sections will focus on some specific RNA sequencing studies for narrow-leafed, white, yellow and some wild lupin species.
| Authors                     | Year | Species                  | Sequencing dataset type | Transcriptome dataset                                                | BioProject ID         | Short read archive       |
|-----------------------------|------|--------------------------|--------------------------|---------------------------------------------------------------------|-----------------------|--------------------------|
| Parra-Gonzalez et al.       | 2012 | *Lupinus luteus*         | RNA                      | Young leaves, buds, flowers and seeds                               | PRJNA170318           | SRX159101                |
| O’Rourke et al.             | 2013 | *Lupinus albus*          | RNA                      | Leaves, roots and cluster roots                                     | PRJNA144943           | SRX087928-SRX087939      |
| Wang et al.                 | 2014 | *L. albus*               | RNA                      | Root tips, juvenile cluster roots and mature cluster roots           | -                     | -                        |
| Secco et al.                | 2014 | *L. albus*               | RNA                      | Root tips, immature cluster roots, mature cluster roots             | PRJNA241146           | SRA145661                |
| Cannon et al.               | 2015 | *Lupinus angustifolius*  | RNA                      | Leaves                                                              | PRJEB8056             | ERX651085                |
| Cannon et al.               | 2015 | *Lupinus polyphyllus*    | RNA                      | Leaves                                                              | PRJEB8056             | ERX651026                |
| Kamphuis et al.             | 2015 | *L. angustifolius*       | RNA                      | Root, stem, leaf, flower and seed                                   | PRJNA248164           | SRX547928-SRX547934      |
| Foley et al.                | 2015 | Several *Lupinus* spp.   | RNA                      | Seed                                                                | PRJNA271721           | SRX832018-SRX8320        |
| Fischer et al.              | 2015 | *L. angustifolius*       | RNA                      | Infected leaf tissue with Colletotrichum lupini                     | -                     | -                        |
| Nevado et al.               | 2016 | Several *Lupinus* spp.   | RNA                      | Stem and leaves                                                     | PRJNA318864           | SRX1721920-SRX1721977    |
| Glazinska et al.            | 2017 | *L. luteus*              | RNA                      | Flowers, pods and flower pedicels                                   | PRJNA285604           | SRX1069734               |
| Yang et al.                 | 2017 | *L. angustifolius*       | RNA                      | Small pods, large pods without seeds, seeds, flowers, pedicels, leaves, stems and roots | PRJNA386115           | SRX3415722               |
| Książkiewicz et al.         | 2017 | *L. albus*               | RNA                      | Young leaves, floral buds and developing pods                       | PRJNA380248           | SRX2663946-SRX2663947    |
| Kroc et al.                 | 2019 | *L. angustifolius*       | RNA                      | Leaves                                                              | PRJNA389154           | SRR5723679-SRR5723682    |
| Venuti et al.               | 2019 | *L. albus*               | RNA                      | Roots under Fe and P deficiency                                     | PRJNA445290           | SRX3832719-SRX3832730    |
| DeBoer et al.               | 2019 | *L. angustifolius*       | Small RNA                | Leaf, stem, flower, root and developing seed stages                  | PRJNA299755           | SRX2635181-SRX2635189    |
| Glazinska et al.            | 2019 | *L. luteus*              | Small RNA                | Flowers, flower pedicels and pods                                   | PRJNA419564           | SRX3417635-SRX3417645-SRX3417655-SRX3417661-SRX3417664 |
| Glazinska et al.            | 2019 | *L. luteus*              | Degradome                | Flowers, flower pedicels and pods                                   | PRJNA419564           | SRX3417662-SRX3417663    |
| Glazinska et al.            | 2019 | *L. luteus*              | RNA                      | Flowers, flower pedicels and pods                                   | PRJNA419564           | SRX3417646-SRX3417655-SRX3417665-SRX3417672 |
| Keller et al.               | 2019 | Several *Lupinus* spp.   | RNA                      | Root nodules                                                        | PRJNA394259           | SRX3005828-SRX3005845    |
| Hufnagel et al.             | 2020 | *L. albus*               | RNA                      | Leaf, different types of roots, flower, seeds, pods                 | PRJNA575804           | SRX7341624-SRX7341633    |
| Hufnagel et al.             | 2020 | *L. albus*               | RNA                      | Different sections of cluster and lateral roots                     | PRJNA593912           | SRX7286280-SRX7286a315   |
Several RNAseq studies have been conducted in NLL. While the first RNAseq study for this species was focussed on identifying as many differently expressed genes from five different tissue types (Kamphuis et al., 2015) that aided the annotation of the reference genome (Hane et al., 2017), several other studies focussed on specific areas, such as seed biology, alkaloid regulation, phenology and yield. Two studies, conducted in the reference genotype, Tanjil, have focussed on the lupin grain to identify small RNAs that regulate aspects of seed biology (DeBoer et al., 2019) and identification of the key seed storage proteins (Foley et al., 2015). Conglutins are a major protein class of seed storage proteins, which provide the energy for germinating seedlings. Members of this conglutin family have been demonstrated to have both beneficial effects for example for the ability to reduce glycaemia and enhance satiety, as well as detrimental effects as some members can cause allergic reactions, when used as a human food. Using RNAseq data, the authors demonstrated that Tanjil contains at least 16 members of the conglutin family, confirming earlier findings (Foley et al., 2011), and they used RNAseq data from five different lupin species to identify their homologous genes. Interestingly, the different members of the conglutin families showed distinct expression patterns between lupin species and high plasticity for conglutin gene expression, thus providing opportunities to remove detrimental and increase preferential conglutins to improve the lupin grain as a human food. The first insight into how these conglutins might be differentially regulated came from a small RNA study in 2019, where the authors identified key small RNAs highly abundant during seed development, and the majority of these small RNAs were predicted to target transcription factors and hormone signalling pathway genes (DeBoer et al., 2019).

Several other studies generated or utilised RNAseq datasets of NLL to identify key genes in the production of quinolizidine alkaloids (QAs). QAs are produced by lupins to protect themselves from insect predation. These secondary metabolites also accumulate in the lupin grain and are toxic to animals and humans and must remain below an industry threshold of 0.02% (Frick et al., 2017). Modern NLL varieties have a low-alkaloid content in the grain but are not alkaloid free. The low alkaloid content in NLL varieties is controlled by a major locus termed iucundus, and Kroc et al. (2019) use transcriptome data guided approach from bitter and sweet accessions to identify candidate genes for this locus. This resulted in the identification of 12 candidate genes that were differentially expressed between bitter and sweet accessions that were associated with QA biosynthesis. Of these, two genes, an APETALA2/ethylene response transcription factor (ERF) and a 4-hydroxytetrahydrodipicolinate synthase (DHDPS) gene were closely linked to iucundus. The DHDPS is involved in lysine biosynthesis, an essential building block to generate QAs, whereas the ERF gene could be the key regulator of the pathway like ERF in tobacco that regulates production of the alkaloid nicotine (Liu et al., 2019).

A similar approach comparing bitter and sweet NLL RNAseq data to identify key genes involved in QA biosynthesis was used in two earlier studies (Frick et al., 2018; Yang et al., 2017). As genes involved in secondary metabolite production are often co-expressed, these studies compared the expression profiles of known QA biosynthesis genes such as a lysine decarboxylase gene (LaLDC), which identified 33 genes with similar expression profiles to the known QA biosynthesis genes in the study by Yang and associates (Yang et al., 2017). One included a gene with homology to a copper amine oxidase (LaCAO), and this gene was shown to have strong conservation with other CAO genes and was able to oxidise cadaverine into 1-piperideine. The first two key enzymes in the QA biosynthesis pathway are thus now known with LaLDC, converting lysine into cadaverine (Bunsupa et al., 2012) and LaCAO oxidising cadaverine into 1-piperideine (Yang et al., 2017). A separate study by Frick et al. (2018) also identified LaCAO as co-expressed with LaLDC in a bitter accession. This study utilised the transcriptome datasets generated by Kamphuis et al. (2015) to identify additional candidate genes involved in QA production, where three major latex-like proteins (LaMLP1, LaMLP2 and LaMLP4) were co-expressed with LaLDC and LaODC (Frick et al., 2018). Other candidate genes identified that did not show co-expression with known QA biosynthetic genes were a fourth major latex-like protein (LaMLP3) and a berberine bridge-like enzyme (LaBBE-like) gene. The expression of these genes in different tissues as well as various stages of seed development using qRT-PCR demonstrated an unchanged expression level during seed development, supporting our current understanding that in NLL, the majority of the QAs are transported to the grain rather than produced within the seed.

The expression pattern of the functionally characterised QA biosynthesis genes and QA content in the mature grain was subsequently investigated following application of biotic (aphid predation and wounding) and abiotic (drought and high temperature) stress. Both drought and high temperature stress as well as their interaction could affect alkaloid content in the NLL grain, albeit in a genotype specific manner (Frick et al., 2018). In contrast aphid predation using the green peach aphid (Myzus persicae) did not exhibit changes in QA biosynthetic gene expression or grain QA content (Frick et al., 2019). In conclusion, a number of genes that contribute to QA biosynthesis and regulation have been identified (Frick et al., 2018, 2019; Kroc et al., 2019; Yang et al., 2017).

White lupins have been intensively studied as a model system to investigate cluster root formation and their associated efficient uptake of phosphorous (Cheng, Bucciarelli, Shen, Allan, & Vance, 2011). Phosphorous is taken up by plants in the form of orthophosphate and is an essential plant nutrient for growth and development. Since 2013, four
transcriptome sequencing studies have been conducted to investigate the regulatory network of orthophosphate uptake and the associated cluster root formation (Cheng et al., 2011; O’Rourke et al., 2013; Secco et al., 2014; Venuti et al., 2019; Wang et al., 2014). The study by O’Rourke and colleagues compared expression profiles of leaves, normal roots and cluster roots in the presence and absence of phosphorous, which identified putative transcription factors associated with cluster root formation. Comparisons of the white lupin data with Arabidopsis (Arabidopsis thaliana) and potato (Solanum tuberosum) phosphorous deficient RNAseq data identified 12 conserved transcripts differentially expressed across all species, which were recommended as biomarkers for phosphorous deficiency.

The second RNAseq study in 2014 identified the role of the plant signalling hormone ethylene in driving cluster root maturation, whereas abscisic and jasmonic acid-related transcripts were highly abundant in mature cluster roots. In contrast, transcripts associated with auxin and cytokinin receptors were most abundant during cluster root initiation (Wang et al., 2014). This is similar to the findings from O’Rourke that showed ethylene biosynthesis and auxin metabolism, and sensing genes were induced in phosphate deficient cluster roots, and cytokinin degradation was also important for cluster root growth (O’Rourke et al., 2013). This was confirmed by exogenous application of auxin to white lupin, which promoted cluster root formation, whereas application of exogenous cytokinin inhibited cluster root formation (Meng et al., 2013; O’Rourke et al., 2013). RNA interference (RNAi) construct targeting a cytokinin oxidase, which was highly upregulated in phosphate deficient cluster roots, was also used to demonstrate the importance of cytokinin degradation.

The third RNAseq study (Secco et al., 2014) further highlighted the importance of auxin and specific transcription factors in the initiation and formation of cluster roots. It also identified transcripts associated with the TCA cycle and glycolysis leading to the production of organic acids, but their data suggested that the root tip is not producing and exporting malate and citrate. These two compounds are known to accumulate in mature cluster roots to facilitate orthophosphate mobilisation, and these transcripts encoding malate dehydrogenase were also previously shown to be highly upregulated in juvenile clusters in contrast to mature cluster roots, whereas a phosphoenolpyruvate carboxylase was highly expressed in mature clusters suggesting a burst of malate to mobilise phosphorus (Wang et al., 2014).

The fourth RNAseq study compared samples grown under phosphorous and iron deficient conditions and showed expression features in common between the two conditions (Venuti et al., 2019). Upregulation of TCA cycle and glycolysis genes was observed under both conditions, whereas iron-deficient conditions only resulted in the upregulation of genes associated with the phenylpropanoid pathway and glycolytic reactions mainly by phosphorous deficient conditions. Combined, these studies have improved the white lupin gene expression atlas and identified the signalling network involved in cluster root initiation, development and maturation, which will aid further research to improve phosphorous uptake in plants.

5 L. LUTEUS TRANSCRIPTOME STUDIES WITH A FOCUS ON FLORAL DEVELOPMENT AND ORGAN ABSCISSION

In the last decade, several RNA sequencing studies have been conducted in yellow lupin, with the first study by Parra-Gonzalez and associates (Parra-González et al., 2012) aimed to generate gene-based molecular markers from Roche 454-sequencing data. The more recent studies used a combination of small RNA, degradome and RNA sequencing to investigate floral development and organ abscission in L. luteus (Glazinska et al., 2017; Glazinska et al., 2019). Yellow lupin produces inflorescences with large quantities of flowers with up to 40 per inflorescence on the main stem. While it produces a large number of flowers per inflorescence, 9/10th of the flowers shed and only a few set pods. Pod abscission therefore has a major influence on the amount of seed produced and ultimately the yield of a lupin crop. How this low frequency of pod set per inflorescence is controlled is poorly understood, and Glazinska et al. (2017); Glazinska et al. (2019) gained more insight in the molecular mechanisms through transcriptome studies. Comparing RNAseq data from pedicels with active and inactive absicising zones revealed gene expression differences associated with cell wall metabolism, metabolic pathways and hormone signalling. Moreover, auxin and ethylene were shown to be important for the regulation of abscission with decreased expression levels observed for genes involved in auxin biosynthesis and transport as well as ethylene response factors and biosynthesis genes (Glazinska et al., 2017). Furthermore, gibberellin was shown to play a role in retaining pods as gibberellin catabolism genes were downregulated in abscised pods and upregulated in retained pods.

In 2019, Glazinska and colleagues further investigated floral and pod development through small RNA, transcriptome and degradome analysis and identified 394 known and 28 novel mRNA in addition to 316 phased siRNAs. Of these, 59 mRNA displayed differential expression patterns at different stages of flower development, and their predicted targets were enriched in genes associated with carbohydrate metabolism and hormone signalling. The combined transcriptome resources for yellow lupin have been collated and are maintained in the L. luteus database termed LuluDB (http://luluseqdb. umk.pl/basic/web/index.php), which allows one to search the different small RNAs identified and download the raw and assembled transcriptome datasets and has a BLAST interface to perform sequence homology searches against these datasets (Glazinska et al., 2020). Together these yellow lupin resources lay the foundation to build a sound understanding of flower development and abscission in lupins to further improve lupin yield.

6 LUPIN GENOMES

An essential step in the development of genomic resources in lupin species was the generation of reference genomes, which were needed to facilitate a range of research and molecular breeding activities. This includes being able to greatly expedite the functional analysis of genes
to help with gene discovery and gene regulation. As detailed recently by Singh et al. (2020), some of the first efforts to gain some understanding of lupin genomes came from the construction and characterisation of bacterial artificial chromosome (BAC) libraries in NLL. These libraries, which typically have an insert size in the range of 100 Kb or more, have been very useful in the initial stages of genomic research for a range of plant species. For example, they have been used in BAC-end sequencing projects that have helped with the generation of genetic and physical maps (Varshney et al., 2010). The first lupin BAC library was constructed for the NLL cultivar Sonet (Kasprzak et al., 2006), and this library was used in situ hybridisation (BAC-FISH) experiments for chromosome identification in NLL (Lesniewska et al., 2011) and as tools for comparative chromosome mapping among lupin species (Suska & Naganowska, 2020). More recently, BAC-FISH and a new approach in Oligo-FISH were used to detect chromosomal rearrangements between smooth and rough-seeded lupin species (Bielski et al., 2020). A second BAC library was constructed for NLL cultivar Tanjil, and 9,600 randomly selected BAC clones from this library were sequenced to give 13,985 bp of BAC-end sequences with an average size of 683 bp (Gao et al., 2011). These BAC-end sequences, which made up approximately 1% of the NLL genome, gave valuable, early insight into the composition of the NLL genome.

The first draft genome assembly for a lupin species was generated for NLL cultivar Tanjil, by Yang and colleagues and published in 2013. This was a highly fragmented genome of 598 Mb captured in 234,534 scaffolds using short read Illumina libraries with 500- and 800-bp insert sizes. Annotation of this assembly was not supported with any lupin transcriptome data and relied on gene identification based on sequence homology to the model plant A. thaliana and other legume species' annotated gene sets and predicted 57,807 genes with an average transcript length of 2,038 bp and coding sequence of 1,033 bp. In 2017, an improved assembly was published that was assembled using both paired end (170-, 500- and 800-bp insert size) and mate pair (2-, 5-, 10-, 20- and 40-Kb insert size) data (Hane et al., 2017) as well as BAC-end sequencing data from Gao et al. (2011). This scaffold assembly was 609 Mb and was captured in 14,379 scaffolds (Hane et al., 2017). The NLL genome generated by Hane et al. (2017) is housed on the lupin genome portal (http://lupinexpress.org) along with a number of other NLL genomic resources.

Hane et al. (2017) used a combination of transcriptome data derived from different tissue types (Kamphuis et al., 2015) for Tanjil and other NLL accessions in combination with annotated gene sets for other legumes to predict the genes in the assembly. This resulted in a total of 33,076 genes with an average transcript length of 3,673 bp and an average coding sequence of 1,289 bp, more in line with the predicted number of genes in other annotated legume genomes. A current project led by our lab has utilised long read sequencing technology (Pacific Biosciences) and Hi-C data to improve the Tanjil reference genome. This has resulted in a substantial further improvement, for example, 90% of the new assembly is captured in under 30 scaffolds compared with 1,629 scaffolds in the previous assembly (Garg, Kamphuis, Foley and Singh unpublished results).

The published NLL genome assembly uncovered some interesting findings. Firstly, Hane et al. (2017) demonstrated that NLL had undergone a whole genome triplication (WGT) event approximately 25 million years ago. Secondly, following the genome annotation, they demonstrated that NLL, which does not undergo mycorrhizal symbiosis, one of the major symbiotic relationships associated with most plants, has lost all genes specifically associated with mycorrhizal interactions. NLL like other legumes has an important symbiotic relationship with Rhizobia involved in nitrogen fixation, and Hane et al. (2017) found that all the genes associated with both mycorrhizal and Rhizobia interactions were present in the NLL genome, as well as those genes uniquely associated with the Rhizobia interactions. Hane et al. (2017) also re-sequenced an early domesticated NLL cultivar and a representative wild accession, both of whom were important parents for NLL RIL populations. These studies helped contribute to the discovery of many SNPS and indels and revealed that while the wild accession was quite divergent across the entire genome from Tanjil, the two domesticated lines showed much lower levels of diversity.

Finally, the study helped with the generation of an improved dense reference genetic map that facilitated the assigning of scaffolds to pseudochromosomes and aided the fine mapping and identification of candidate genes for various domestication traits such as flowering time and seed shattering as well as disease resistance to anthracnose, an important issue for NLL and other lupins (Hane et al., 2017).

In 2020, the genome of white lupin (L. albus) cultivar ‘Amiga’ was sequenced and assembled by two different research groups in France and China (Hufnagel, Marques, et al., 2020; Xu et al., 2020). The French group utilised a combination of PacBio long reads, Illumina short reads and a Bionano optical map to generate an assembly of 451 Mb captured in 89 contigs. RNAseq data from 10 different tissue types and the proteome of M. truncatula were used to annotate the genome, resulting in the identification of 38,258 protein coding genes (Hufnagel, Marques, et al., 2020). This assembly is housed on the white lupin genome portal (https://www.whitelupin.fr/). The Chinese team used a combination of long read PacBio, Illumina short read and Hi-C data to generate a genome assembly of 559 Mb captured in 1,580 scaffolds. Comparisons of these two assemblies showed good chromosomal synteny with some small inversions. This second assembly was annotated using protein coding sequences of A. thaliana and several legume species as well as RNAseq data from root, leaf and stem tissue resulting in 48,719 predicted coding sequences (Xu et al., 2020).

Like the findings for NLL by Hane et al. (2017), white lupin has also undergone a WGT event, and the authors reconstructed the three subgenomes. Through the WGT, the genes involved in phosphorus use efficiency and cluster root formation were shown to have duplicated and diverged (Xu et al., 2020). In addition to the reference genome, 15 modern accessions, landraces and wild accessions were re-sequenced by the French team, and comparison of wild and modern cultivar root formation revealed that modern accessions establish lateral cluster roots a lot earlier than their wild counterparts.
Furthermore, proteomics analysis of the white lupin seed protein composition showed different profiles for the seed storage protein beta conglutin between wild and modern white lupin varieties. Lastly, the genome assembly was used to identify candidate genes in the region underlying the QTL locus termed pauper, which controls alkaloid accumulation. Among the 66 candidates in the region of interest are transcription factors and genes encoding acyltransferases and cinnamoyl-CoA reductases (Hufnagel, Marques, et al., 2020). These findings support earlier research, which demonstrated a high association between a marker anchored in the acyltransferase-like (LaAT) gene and a low alkaloid pauper phenotype in white lupin (Rychel & Ksiazkiewicz, 2019).

While the white lupin assembly by Hufnagel and associates is less fragmented, the assembly by Xu and colleagues is larger in size and has a larger number of predicted coding sequences. This led us to compare the different NLL and white lupin genome assemblies through a Benchmarking Universal Single-Copy Ortholog (BUSCO) analysis (Seppey, Manni, & Zdobnov, 2019). BUSCO analysis allows one to assess the completeness of a genome assembly using the concept of single-copy orthologous genes that should be highly conserved across related species. The most recent BUSCO version (v4.1.3) for plantae contains 425 conserved BUSCO genes, hereafter referred to as BUSCOs. These 425 BUSCOs were determined based on the consensus of 20 OrthoDB species with M. truncatula and G. max the only legume representatives. The BUSCO analysis on the lupin genomes showed that four genes were missing in the NLL assembly (Hane et al., 2017) and 4 and 11 in the white lupin assemblies by Hufnagel and Xu, respectively (Table 2). Of these, two BUSCOs (31855at33090 and 40630at33090) are absent in both the NLL and white lupin assemblies. To determine if these are also absent in other legume genomes, we conducted BUSCO analysis on several published legume genomes including alfalfa (Shen, Du, Chen, Lu, & Zhu, 2020), adzuki bean (Yang et al., 2015), barrel medic (Young et al., 2011), chickpea (desi and kabuli; Parween et al., 2015; Varshney et al., 2013), common bean (Schmutz et al., 2014), cowpea (Lonardi et al., 2019), lotus (Sato et al., 2008), mungbean (Kang et al., 2014), pea (Kreplak et al., 2019), peanut (Bertioli et al., 2019), pigeonpea (Varshney et al., 2012), soybean (Schmutz et al., 2010) and subclover (Hirakawa et al., 2016) and found both BUSCOs to be absent across all the legume genomes with the exception of BUSCO 31855at33090 being present in the chickpea genomes (desi and kabuli) and BUSCO 40630at33090 being present in the L. japonicus genome.

7 | DIVERSITY STUDIES IN NLL AND WHITE LUPIN

Domestication of grain legumes such as lupins has undergone a series of population bottlenecks as various domestication alleles undergo extreme selection pressure. In crops like chickpea and NLL, this has resulted in a low level of genetic diversity in cultivated varieties. A genetic diversity analysis of 248 wild and 95 domesticated NLL lines using Diversity Arrays Technology (DArT) genotyping indeed demonstrated this domestication bottleneck (Berger et al., 2012), and it was thus proposed that there was a need to determine the useful genetic diversity for NLL (Berger et al., 2013). Using high resolution DArT sequencing on 142 wild NLL accessions from the Mediterranean region, it was highlighted that germplasm from the western Mediterranean was more diverse and that eastern Mediterranean material showed phenological adaptation to warmer, lower rainfall environments (Mousavi-Derazmahalleh, Bayer, et al., 2018). More detailed investigation of the diversity in comparison with current NLL varieties and breeding lines showed that western Mediterranean germplasm provided the founder populations for the domestication of NLL (Mousavi-Derazmahalleh et al., 2018). Indeed in the early 'Phase I' of

Table 2: Overview of the published genome assemblies for narrow-leaved lupin (Lupinus angustifolius) and white lupin (Lupinus albus) and their BUSCO statistics using version 4.1.3 for plantae

|                        | Narrow-leaved lupin | White lupin                  |
|------------------------|---------------------|------------------------------|
|                        | Yang et al., 2013   | Hane et al., 2017            | Hufnagel, Marques, et al., 2020 | Xu et al., 2020 |
| Genome size            | 523 Mb              | 609 Mb                       | 451 Mb                        | 559 Mb          |
| Number of scaffolds    | 71,995              | 14,379                       | 89                            | 1,580           |
| N50                    | 10,140              | 232                          | 12                            | 14              |
| L50                    | 15,485 bp           | 703,185 bp                   | 17.35 Mb                      | 18.66 Mb        |
| N90                    | 35,018              | 1,629                        | 23                            | 33              |
| L90                    | 4.01 Kb             | 35.35 Kb                     | 14.55 Mb                      | 1.67 Mb         |
| GC content (%)         | 33.25               | 33.46                        | 33.79                         | 36.82           |
| Annotated protein-coding sequences | 57,806             | 33,076                       | 38,258                        | 47,603          |
| Mean coding sequence length | 1,033 bp           | 1,289 bp                     | 1,110 bp                      | 1,194 bp        |
| BUSCOs complete        | 402                 | 422                          | 418                           | 412             |
| BUSCOs fragmented      | 18                  | 1                            | 3                             | 2               |
| BUSCOs missing         | 5                   | 2                            | 4                             | 11              |

Abbreviation: BUSCO, Benchmarking Universal Single-Copy Ortholog.
the Australian lupin breeding programme wild accessions, seven accessions from the western Mediterranean were used and one from the eastern Mediterranean (Cowling, 2020). These studies also highlighted that there is much higher linkage disequilibrium in NLL cultivars relative to their wild counterparts. Therefore, breeding in novel beneficial loci can be hampered by unwanted linkage to undesirable loci.

To further understand the genetic diversity of NLL, a pan-genome is also being developed by our group that should capture the genetic diversity for the species. Here, genetically diverse wild accessions as well as domesticated European and Australian varieties have been selected based on the previous diversity studies for NLL (Mousavi-Derazmahalleh, Bayer, et al., 2018; Mousavi-Derazmahalleh, Nevada, et al., 2018). Similarly, in white lupin, the first pan-genome for a lupin species was recently published, where 38 white lupin accessions representing cultivars, landraces and wild accessions were de novo assembled, aligned to the reference and a pan-genome assembled (Hufnagel, Soriano, et al., 2020). A total of 38,443 genes were annotated of which 78.5% (32,068 genes) were present in all 39 white lupin lines and considered core genes. A further 6,046 were absent in a single accession and 8,776 genes absent in 2–38 accessions. In the white lupin pan-genome 1,195 presence absence variants (PAVs) in protein coding genes were identified of which 1,132 genes were present in the reference genome and 63 in newly identified genes in the white lupin pan-genome. The wild white lupin accessions had a significantly higher number of newly identified genomes compared with the domesticated varieties and landraces. The protein coding sequences with the PAVs were functionally analysed, and this revealed an enrichment in gene families coding encoding membrane receptors and transporters as well as genes associated with the regulation of biotic and abiotic stress responses. A 686-bp deletion in the third intron of the major flowering time gene LaFTa1 was identified, which may contribute to later flowering time in accessions with this deletion (Hufnagel, Soriano, et al., 2020). A mapping population derived from the early flowering Kiev Mutant and the late flowering P27174 identified four QTLs associated with flowering time control including one that flowering Kiev Mutant and the late flowering P27174 identified four QTLs associated with flowering time control including one that

helped to find a single origin for L. mutabilis and pinpointed the area and timing of domestication (Atchison et al., 2016). In the case of lupin diversification, an initial landmark study used nuclear DNA sequences from Andean Lupinus species, involving 53 accessions from 36 species to show that these Andean lupin species exhibited very recent and rapid diversification (Hughes & Eastwood, 2006). Using a similar approach, Drummond (2008) used three regions of rapidly evolving chloroplast DNA to study North American lupin species and showed that these were also undergoing rapid diversification. A follow-up study used further DNA analysis to provide additional support for the rapid rates of diversification for Andean lupin species and to support that drivers for this high rate of diversification were a shift from annual to perennial life cycle and adaptation to higher altitude environments (Drummond et al., 2012). Overall, these initial studies showed that these diversification rates were among the highest documented for plants and were still accelerating (Drummond, 2008; Drummond et al., 2012; Hughes & Eastwood, 2006).

The mechanisms underlying these rapid diversifications in some lupin species were explored using transcriptome sequencing for 55 New World lupin species (Nevado et al., 2016) that included annual and perennial species from both North America and South America. In the rapidly diversifying perennial species, the study found major increases in the frequency of adaptive evolution in both coding sequences and regulatory regions compared with what was found for the annual lupin species that were undergoing low rates of diversification. These results indicated that both processes, changes in protein sequence and changes in gene expression, were underlying the rapid diversification observed with some new world lupins.

9 | OTHER DEVELOPING GENETIC AND GENOMIC RESOURCES

Several additional resources and tools are being developed to aid lupin crop improvement. The published NLL and white lupin genomes have been integrated in the Legume Information System (LIS; https://legumeinfo.org/), where their genomes can be browsed and syntenic comparisons made to other legume species in the database. This will aid the gene discovery for key traits through syntenic comparisons with the model and more intensively studied legume crops such as soybean.

A current project is focusing on developing a NLL Targeting Induced Local Lesions in Genomes (TILLING) population in the genetic background of cultivar Tanjil (Foley et al., 2020). Seeds of cultivar Tanjil have been treated with ethyl methanesulfonate (EMS), which causes random C to T base transitions throughout the genome with the aim to generate random mutations in genes, with cost-effective re-sequencing allowing the identification of these mutations across the genome. Typically, large quantities of seeds are treated and then seeds are nurtured to maturity to collect M1 seeds for a several thousand individuals. To date, DNA from 384 M1 lines have been isolated and sequenced with 4,326 premature ‘stop’ mutations identified across 3,676 annotated genes. These 384 TILLING lines thus capture

8 | GENOMIC STUDIES ARE SHEDDING LIGHT ON DOMESTICATION, DIVERSIFICATION AND EVOLUTION FOR NEW WORLD LUPINS

Increasingly more sophisticated DNA-based approaches are being used to address important issues around domestication, diversification and evolutionary drivers using new world lupins. For example, a nextRADseq approach was used to determine the origins of domestication for L. mutabilis, an important crop in parts of S. America. Data from the nextRADseq for 200 accessions on Andean lupin species
Approximately 10% of the predicted genes in the NLL genome, and one could thus extrapolate that a population of ~4,000 M₂ would capture ‘stop’ mutations in essentially all predicted genes that do not lead to a lethal phenotype. The generation of premature stops in genes would allow the identification of mutants that would improve NLL as a crop. For example, one could identify TILLING lines with reduced alkaloid content, reduced allergenecity proteins, altered protein composition, altered oil content, among others. This could make the NLL grain more attractive as a source for either high end animal and aquaculture feed or human consumption. The TILLING population will also provide a powerful tool to confirm the role of a given gene in a biological process. The generation of additional M₂ TILLING lines is underway and ~600 M₂ lines have been obtained to date some of which have shown interesting phenotypes in seed biology and vegetative growth (Foley et al., 2020).

A co-investment by Agriculture Victoria Research (AVR) and the Grains Research and Development Corporation (GRDC) is focused on developing a SNP genotyping platform for chickpea, lentil, field pea, faba bean and NLL. Here, an imputation enabled Infinium XT SNP chip is under development and early stages of validation, which aims to capture the genetic diversity within internationally available datasets and germplasm for these pulse species. The chip is anticipated to capture ~30,000 SNPs for these pulse species of which 5,425 derived for the current NLL genome that would allow imputation of the SNP data to a whole genome level (Kaur et al., 2020; Sukhijwan Kaur, pers comm). The SNP chip will allow pre-breeding research to connect research findings across projects and link these to outcomes for the breeding programme. Combined, these advances will aid the NLL breeding programme in understanding the genetic basis of important traits and accelerate their incorporation into the development of new varieties.

10 | ADVANCES IN LUPIN GENOMIC RESOURCES OPEN UP POWERFUL NEW BREEDING APPROACHES

A (the) major target for many crop breeding programs is to increase the overall grain yield of the target crop(s). However, a host of genetic and environmental factors contributes in either a positive or negative manner to grain yield. Thus, many of the individual traits that contribute to grain yield are complex, quantitative traits, whereby several genes or quantitative trait loci (QTLs) underlie such traits, with each gene/QTL often only making a relatively small contribution to the trait. With the advent of adequate genomic resources for a crop, coupled with increasingly powerful phenotyping methods, progress in tackling such complex traits has improved significantly. Some of these approaches, such as genotyping-by sequencing (GBS) and genomic selection offer great potential for lupin crop breeding as demonstrated by some pioneering studies.

GBS is a powerful marker-assisted selection tool for crop breeding (He et al., 2014) that involves sequencing of many samples, which can be multiplexed to allow for both marker discovery and genotyping for a crop species. As mentioned briefly in the lupin transcriptome section, GBS approaches have been used in white lupin to identify loci associated with vernalisation responsiveness, anthracnose resistance and seed alkaloid content (Książkiewicz et al., 2017). An alternative high-throughput genotyping approach in DArT sequencing has been utilised in NLL (Mousavi-Derazmahalleh, Bayer, et al., 2018; Mousavi-Derazmahalleh, Nevado, et al., 2018), and such genotyping approaches can be utilised to conduct GWAS or genomic prediction/selection.

Genomic selection is another advanced breeding method that combines genotyping and phenotyping data from a germplasm sample of interest such as a collection of inbred lines or landraces in the case of crops, into a prediction model. The prediction model is initially based on the genotypic and phenotypic data of a training population, which allow the prediction of individuals from a breeding programme that will perform better under certain situations, for example, drought conditions (reviewed in Bhat et al., 2016; Voss-Fels et al., 2019). Once the model has been validated, it can be applied to much larger germplasm populations without requiring the large amount of corresponding phenotyping and selection of a more classical breeding approach and therefore lends itself particularly well for tackling complex, polygenic traits such as yield.

In the case of lupins and genomic selection, good progress has been made in white lupin (Annichiarico et al., 2019). Annichiarico and colleagues have examined the utility of genomic selection to predict grain yield and several other agronomic traits in various white lupin germplasm accessions. They evaluated two major statistical models for genome selection in white lupin, the ridge regression BLUP method (Meuwissen et al., 2001) and the Bayesian Lasso method (Park & Caella, 2008). They found both models had similar predictive ability when applied to the prediction of grain yield and other traits such as winter survival and the onset of flowering, based primarily on the analysis of 83 white lupin landraces from diverse cropping regions.

11 | CONCLUSION AND FUTURE DIRECTION

This review has highlighted the development of a growing suite of genomic resources for two key lupin crops, NLL and white lupin, and these advances offer great promise to accelerate yield gains for these important crop species while helping to further our understanding of many aspects of lupin and indeed legume biology. Figure 1 summarises the major genomic resources that have or are in the process of being developed for both crops. Other tools that will help in this regard are the development and utilisation of efficient gene editing processes, the widespread implementation of speed breeding approaches for both crops and the development of proteomic and metabolomic resources; for example, a pan-proteome of the lupin grain would be a valuable resource.

Collectively, these resources will help researchers and breeders tackle some of the pressing issues facing these crops. For example,
while domestication of NLL has had some success, the genetic base of domesticated accessions has been shown to be quite narrow (Berger et al., 2013) and restricts efforts to improve the crop. However, the development of an NLL pan-genome with over 40 wild accessions will provide valuable insight into the core and variable gene set for the crop and facilitate efforts to increase diversity in the breeding of new cultivars. Efforts are also underway to develop genomic resources for the other two domesticated lupin species, namely, yellow lupin and pearl lupin. These combined genomic and related resources will help researchers and breeders to improve lupin crops to help meet the rapidly growing demand for plant protein and the establishment of more sustainable cropping systems, where lupin crops can be important contributors.

DATA AVAILABILITY STATEMENT
The genome information for narrow-leaved lupin (NLL) can be obtained from NCBI under BioProject PRJNA299755 and http://www.lupinexpress.org. The genome information for white lupin can be obtained from NCBI under BioProjects PRJNA592024 and PRJNA575804 as well as https://www.whitelupin.fr/. The transcriptome resources described in this review can be found under the BioProject IDs and short read archives (SRAs) listed in Table 2.

ACKNOWLEDGEMENTS
The authors are grateful to CSIRO, Curtin University, The University of Western Australia (UWA) and the Grains Research and Development Corporation (GRDC, Project code 9176622) for the financial support for the authors' lupin genomic research. The authors thank Drs. Jonathan Anderson and Ling-ling Gao for helpful comments on the manuscript. The authors regret that not all relevant lupin genomic work could be cited due to space limitations.

CONFLICT OF INTEREST
The authors declare no conflict of interest. All authors have read and approved the final version of the manuscript.

AUTHOR CONTRIBUTIONS
All authors contributed conceptually to the idea and content of this invited review, and all authors contributed to the writing of the review.

POST-SCRIPT NOTE
After this manuscript was in the review process, a paper describing an improved ‘third generation’ NLL reference genome was published by Wang et al 2020 (The Plant Journal, early view).

ETHICS STATEMENT
There were no ethical issues related to this review.

ORCID
Lars G. Kamphuis, https://orcid.org/0000-0002-9042-0513
Gagan Garg, https://orcid.org/0000-0002-9523-6360
Karam B. Singh, https://orcid.org/0000-0002-2777-7448

REFERENCES
Annicchiarico, P., Nazzicari, N., & Ferrari, B. (2020). Genetic and genomic resources in white lupin and the application of genomic selection. In K. B. Singh, L. G. Kamphuis, & M. N. Nelson (Eds.), Compendium of Plant Genomes, The Lupin Genome. Cham, Switzerland: Springer Nature. https://doi.org/10.1007/978-3-030-21270-4_10

Annicchiarico, P., Nazzicari, N., Ferrari, B., Harzic, N., Carroni, A. M., Romani, M., & Pacetti, L. (2019). Genomic prediction of grain yield in contrasting environments for white lupin genetic resources. Molecular Breeding, 39, 142. https://doi.org/10.1007/s11032-019-1048-6

Atchison, G. W., Nevado, B., Eastwood, R. J., Contreras-Oritz, N., Reynel, C., Madrinan, S., Filetov, D. A., & Hughes, C. E. (2016). Lost crops of the Incas: Origins of domestication of the Andean pulse crop tarwi, Lupinus mutabilis. American Journal of Botany, 103, 1592–1606. https://doi.org/10.3732/ajb.1600171

Berger, J., Buirschell, B., Luckett, D., & Nelson, M. (2012). Domestication bottlenecks limit genetic diversity and constrain adaptation in narrow-leaved lupin (Lupinus angustifolius L.). Theoretical and Applied Genetics, 124, 637–652. https://doi.org/10.1007/s00122-011-1736-z

Berger, J. D., Clements, J. C., Nelson, M. N., Kamphuis, L. G., Singh, K. B., & Buirschell, B. (2013). The essential role of genetic resources in narrow-leaved lupin improvement. Crop and Pasture Science, 64, 361–373. https://doi.org/10.1071/CP13092

Bertioli, D. J., Jenkins, J., Clevenger, J., Dudchenko, O., Gao, D., Seijo, G., Leal-Bertioli, S. C. M., Ren, L., Farmer, A. D., Pandey, M. K., Samoluk, S. S., Abernathy, B., Agarwal, G., Ballén-Taborda, C., Cameron, C., Campbell, J., Chavarro, C., Chitikineni, A., Chu, Y. … Schmutz, J. (2019). The genome sequence of segmental allotetraploid peanut Arachis hypogaea. Nature Genetics, 51, 877–884. https://doi.org/10.1038/s41588-019-0405-z

Bhat, J. A., Ali, S., Salgota, R. K., Mir, Z. A., Dutta, S., Jadon, V., Tyagi, A., Mushtaq, M., Jain, N., Singh, P. K., Singh, G. P., & Prabhu, K. V. (2016). Genomic selection in the era of next generation sequencing for complex traits in plant breeding. Frontiers in Genetcs, 7, 221.

Bieliski, W., Ksiazkiewicz, M., Simonikova, D., Hribova, E., Susek, K., & Naganowska, K. (2020). The puzzling fate of a lupin chromosome revealed by reciprocal oligo-FISH and BAC-FISH mapping. Genes, 11, 1489. https://doi.org/10.3390/genes11121489

Bunsupa, S., Katayama, K., Ikereu, E., Oikawa, A., Toyooka, K., Saito, K., & Yamazaki, M. (2012). Lysine decarboxylase catalyzes the first step of amino acid metabolism and plays a role in the evolution of plant-shrinking bacteria. Proceedings of the National Academy of Sciences, 109, 13721–13726. https://doi.org/10.1073/pnas.1211065109

Christenhusz, M., & Byng, J. W. (2016). The number of known plant species in the world and its annual increase. Phytotaxa, 261, 201–207. https://doi.org/10.11646/phytotaxa.261.3.1

Cannon, S. B., McKain, M. R., Harkess, A., Nelson, M. N., Dash, S., Deyholos, M. K., Peng, Y., Joyce, B., Stewart, Jr, C. N., Rolf, M., Kutchan, T., Tan, X., Chen, C., Zhang, Y., Carpenter, E., Wong, G. K.-S., Doyle, J. J., & Leebens-Mack, J. (2015). Multiple polyploidy events in the early radiation of nodulating and nonnodulating legumes. Molecular Biology and Evolution, 32, 193–210. https://doi.org/10.1093/molbev/msu296

Cheng, L., Bucciarelli, B., Shen, J., Allan, D., & Vance, C. P. (2011). Update on lupin cluster roots: Update on white lupin cluster root acclimation to phosphorus deficiency. Plant Physiology, 156, 1025–1032. https://doi.org/10.1104/pp.111.175174

Christenhusz, M., & Byng, J. W. (2016). The number of known plant species in the world and its annual increase. Phytotaxa, 261, 201–207. https://doi.org/10.11646/phytotaxa.261.3.1

Cowling, W. A. (2020). Genetic diversity in narrow-leaved lupin breeding after the domestication bottleneck. In K. B. Singh, L. G. Kamphuis, & M. N. Nelson (Eds.), Compendium of Plant Genomes, The Lupin Genome. Cham, Switzerland: Springer Nature. https://doi.org/10.1007/978-3-030-21270-4_1

Croxford, A. E., Rogers, T., Caligari, P. D., & Wilkinson, M. J. (2008). High-resolution melt analysis to identify and map sequence-tagged site anchor points onto linkage maps: A white lupin (Lupinus albus) map as
Gao, L. L., Hane, J. K., Kamphuis, L. G., Foley, R., Shi, B. J., Atkins, C. A., & Glazinska, P., Wojciechowski, W., Kulasek, M., Glinkowski, W., Wysocka, M., & Kosi

12

Frick, K. M., Foley, R. C., Kamphuis, L. G., Siddique, K. H. M., Garg, G., & Fischer, K., Dieterich, R., Nelson, M. N., Kamphuis, L. G., Singh, K. B., Foley, R., Kamphuis, L. G., & Singh, K. B. (2020). Cutting-edge techniques for lupin (Lupinus angustifolius), an emerging healthy food: Insights into plant–microbe interactions and legume evolution. Plant Biotechnology Journal, 15, 318–330. https://doi.org/10.1111/pbi.12615

He, J., Zhou, X., Laroche, A., Lu, Z.-X., Liu, H., & Li, Z. (2014). Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Frontiers in Plant Science, 5, 484.

Hirakawa, H., Kaur, P., Shirasawa, K., Nichols, P., Nagano, S., Appels, R., Erskine, W., & Isoe, S. N. (2016). Draft genome sequence of subterranean clover, a reference for genus Trifolium. Scientific Reports, 6, 30358. https://doi.org/10.1038/srep30358

Hufnagel, B., Marques, A., Soriano, A., Marqués, L., Divol, F., Doumas, P., Sallet, E., Mancinotti, D., Carrère, S., Marande, W., Arribat, S., Keller, J., Huneau, C., Blein, T., Aimé, D., Laguerre, M., Taylor, J., Schubert, V., Nelson, M., ... Pérét, B. (2020). High-quality genome sequence of white lupin provides insight into soil exploration and seed quality. Nature Communications, 11, 492. https://doi.org/10.1038/s41467-019-14197-9

Hufnagel, B., Soriano, A., Taylor, J., Divol, F., Kroc, M., Sanders, H., Nelson, M., & Pérét, B. (2020). Pangeneome of white lupin provides insights into the diversity of the species. bioRXiv. https://doi.org/10.1101.2020.1105.1161.163378

Hughes, C., & Eastwood, R. (2006). Island radiation on a continental scale: Exceptional rates of plant diversification after uplift of the Andes. Proceedings of the National Academy of Sciences of the United States of America, 103, 10334–10339. https://doi.org/10.1073/pnas.0601928103

Iqbal, M. M., Huynh, M., Udall, J. A., Kilian, A., Adhikari, K. N., Berger, J. D., Erskine, W., & Nelson, M. N. (2019). The first genetic map for yellow lupin enables genetic dissection of adaptation traits in an orphan grain legume crop. BMC Genetics, 20, 68. https://doi.org/10.1016/j.d唐朝.12863-019-0767-3

Kamphuis, L. G., Foley, R., Frick, K. M., Garg, G., & Singh, K. B. (2020). Transcriptome resources paving the way for lupin crop improvement. In K. Singh, L. G. Kamphuis, & M. N. Nelson (Eds.), Compendium of Plant Genomes, The Lupin Genome. Cham, Switzerland: Springer Nature. 10.1007/978-3-030-21720-4_5

Kamphuis, L. G., Hane, J. K., Nelson, M. N., Gao, L. Atkins, C. A., & Singh, K. B. (2015). Transcriptome sequencing of different narrow-leaved lupin tissue types provides a comprehensive uni-gene assembly and extensive gene-based molecular markers. Plant Biotechnology Journal, 13, 14–25. https://doi.org/10.1111/pbi.12229

Kang, Y. J., Kim, S. K., Kim, M. Y., Lestari, P., Kim, K. H., Ha, B.-K., Jun, T. H., Hwang, W. J., Lee, T., Lee, J., Shim, S., Yoon, M. M., Kang, Y. E., Han, K. S., Tsaprayoon, P., Yoon, N., Somta, P., Tanya, P., Kim, K. S., ... Lee, S.-H. (2014). Genome sequence of mungbean and insights into evolution within Vigna species. Nature Communications, 5, 5443. https://doi.org/10.1038/ncomms5443

Kasprowak, Z., Safar, J., Janda, J., Dolezel, J., Wolkow, B., & Naganoewska, B. (2006). The bacterial artificial chromosome (BAC) library of the narrow-leaved lupin (Lupinus angustifolius L.). Cellular and Molecular Biology Letters, 11, 396–407. https://doi.org/10.1024/s11658-006-0033-3

Kaur, S., Keeble-Gagnere, G., Pasam, R.K., & Hayden, M. (2020) Pulse pipeline integration to become more efficient. In: GRDC Groundcover. Canberra, Australia: GRDC.
bean (Vigna angularis) provides insight into high starch and low fat accumulation and domestication. *Proceedings. National Academy of Sciences. United States of America, 112*, 13213–13218. https://doi.org/10.1073/pnas.1420949112

Yang, T., Nagy, I., Mancinotti, D., Otterbach, S. L., Andersen, T. B., Motawia, M. S., Asp, T., & Geu-Flores, F. (2017). Transcript profiling of a bitter variety of narrow-leaved lupin to discover alkaloid biosynthetic genes. *Journal of Experimental Botany, 68*, 5527–5537. https://doi.org/10.1093/jxb/erx362

Young, N., Debellé, F., Oldroyd, G., Geurts, R., Cannon, S. B., Udvardi, M. K., Benedito, V. A., Mayer, K. F. X., Gouzy, J., Schoof, H., Van de Peer, Y., Proost, S., Cook, D. R., Meyers, B. C., Spannagl, M., Cheung, F., De Mita, S., Krishnakumar, V., Gundlach, H., ... Roe, B. A. (2011). The Medicago genome provides insight into the evolution of rhizobial symbioses. *Nature, 480*, 520–524.

How to cite this article: Kamphuis LG, Garg G, Foley R, Singh KB. Genomic resources for lupins are coming of age. *Legume Science*. 2021:1–15. https://doi.org/10.1002/leg3.77