PCP
PLANT & CELL PHYSIOLOGY

The *Vigna* Genome Server, '*Vig*GS': A Genomic Knowledge Base of the Genus *Vigna* Based on High-Quality, Annotated Genome Sequence of the Azuki Bean, *Vigna angularis* (Willd.) Ohwi & Ohashi

Hiroaki Sakai[1,*], Ken Naito[2], Yu Takahashi[2], Toshiyuki Sato[3], Toshiya Yamamoto[4], Isamu Muto[4], Takeshi Itoh[1] and Norihiko Tomooka[2]

[1]Agrogenomics Research Center, National Institute of Agrobiological Sciences, 2-1-2 Kannondai, Tsukuba, Ibaraki, 305-8602 Japan
[2]Genetic Resources Center, National Institute of Agrobiological Sciences, 2-1-2 Kannondai, Tsukuba, Ibaraki, 305-8602 Japan
[3]Mizuho Information & Research Institute, Inc., 2-3 Kanda-Nishikicho, Chiyoda-ku, Tokyo, 101-8443 Japan
[4]BITS Co., Ltd., 5-201 Kandamatsunaga-cho, Chiyoda-ku, Tokyo, 101-0023 Japan
*Corresponding author: E-mail, hirsakai@affrc.go.jp; Fax, +81-29-838-7065.
(Received August 31, 2015; Accepted November 20, 2015)

The genus *Vigna* includes legume crops such as cowpea, mungbean and azuki bean, as well as >100 wild species. A number of the wild species are highly tolerant to severe environmental conditions including high-salinity, acid or alkaline soil; drought; flooding; and pests and diseases. These features of the genus *Vigna* make it a good target for investigation of genetic diversity in adaptation to stressful environments; however, a lack of genomic information has hindered such research in this genus. Here, we present a genome database of the genus *Vigna*, *Vigna* Genome Server ('*Vig*GS', http://viggs.dna.affrc.go.jp), based on the recently sequenced azuki bean genome, which incorporates annotated exon–intron structures, along with evidence for transcripts and proteins, visualized in GBrowse. *Vig*GS also facilitates user construction of multiple alignments between azuki bean genes and those of six related dicot species. In addition, the database displays sequence polymorphisms between azuki bean and its wild relatives and enables users to design primer sequences targeting any variant site. *Vig*GS offers a simple keyword search in addition to sequence similarity searches using BLAST and BLAT. To incorporate up to date genomic information, *Vig*GS automatically receives newly deposited mRNA sequences of pre-set species from the public database once a week. Users can refer to not only gene structures mapped on the azuki bean genome on GBrowse but also relevant literature of the genes. *Vig*GS will contribute to genomic research into plant biotic and abiotic stresses and to the future development of new stress-tolerant crops.

Keywords: Genome database • Legume • *Vigna*.

Abbreviations: CAGE, cap analysis of gene expression; CDS, coding sequence; CTSS, CAGE transcription start site; EST, expressed sequence tag; FPKM, fragments per kilobase of transcript per million fragments mapped; GFF, general feature format; GO, Gene Ontology; MCL, Markov clustering; SMRT, single-molecule real-time; SNP, single nucleotide polymorphism; SSR, simple sequence repeat; VCF, variant call format.

Introduction

The genus *Vigna* consists of >100 species that are distributed throughout the world, mainly in warm temperate and tropical regions (Schrire 2005). *Vigna* is phylogenetically closely related to agriculturally important genera such as *Cajanus*, *Glycine* and *Phaseolus* (Gepts et al. 2005). Species within the genus *Vigna* include nine crop plants: cowpea [*Vigna unguiculata* (L.) Walp.], mungbean [*Vigna radiata* (L.) R. Wilczek], azuki bean [*Vigna angularis* (Willd.) Ohwi & Ohashi], bambara groundnut [*Vigna subterranea* (L.) Verdcourt], black gram [*Vigna mungo* (L.) Hepper], creole bean (*Vigna reflexo-pilosa* Hayata), moth bean [*Vigna aconitifolia* (Jacq.) Maréchal], rice bean [*Vigna umbellata* (Thunb.) Ohwi & Ohashi] and tuber cowpea [*Vigna vexillata* (L.) A. Rich]. One of the most attractive characteristics of *Vigna*, however, is the vast genetic diversity of the wild species of the genus. A number of the wild species are highly tolerant to severe environmental conditions including high-salinity, acid or alkaline soil; drought; flooding; and pests and diseases (Chankaew et al. 2014, Tomooka et al. 2014, Yoshida et al. 2015). Furthermore, tolerant and susceptible species have been identified and are cross-compatible, facilitating conventional forward genetic studies to identify causative genes (Chankaew et al. 2014, Tomooka et al. 2014). Therefore, the genus *Vigna* is a potentially valuable new model system for studying plant adaptation to stressful environments. Recently, wild relatives of crop species have received a great deal of attention in research communities as potential genetic resources for crop improvement (McCouch et al. 2013). In fact, new breeding concepts such as 'neo-domestication' and 'reverse-breeding', attempting to harness the desirable traits of wild species for crop improvement, have been proposed (Tomooka et al. 2014, Palmgren et al. 2015). Understanding the molecular mechanisms of how wild *Vigna* species have

Plant Cell Physiol. 57(1): e2(1–9) (2016) doi:10.1093/pcp/pcv189, Advance Access publication on 7 December 2015, available online at www.pcp.oxfordjournals.org

adapted to a wide variety of stress environments will advance the breeding of stress-tolerant varieties, not only of *Vigna* crops, but also of related important legume crops such as soybean and common bean.

High-quality annotated reference genome sequences are essential resources to conduct molecular genetic and functional genomic studies efficiently, and genome databases are important tools to facilitate full use of these genomic resources. For some plant species with their genomes sequenced, especially for major crop species such as rice, wheat, maize and soybean, a number of genome databases and other genomics tools have been developed (Gonzales et al. 2005, Grant et al. 2010, Goodstein et al. 2012, Li et al. 2012, Sakai et al. 2013, Kersey et al. 2014, Yonemaru et al. 2014, Bolser et al. 2015, Krishnakumar et al. 2015a, Krishnakumar et al. 2015b). However, little genomic information is available for *Vigna*, which has hindered research on this genus. Recently, the draft genome sequences of two *Vigna* species were released (Kang et al. 2014, Kang et al. 2015). However, only about half of these genome assemblies were anchored to chromosomes. Thus, to establish a knowledge base for *Vigna*, based on a high-quality reference genome sequence, we sequenced the genome of the azuki bean (*V. angularis*) cultivar 'Shumari', using single-molecule real-time (SMRT) sequencing technology (Eid et al. 2009, Sakai et al. 2015). The reconstructed pseudomolecules covered 85.6% of the estimated genome size, with gaps comprising only 1.7% (Sakai et al. 2015). Here, we present the first genome database for genus *Vigna*, the *Vigna* Genome Server ('*Vig*GS'), which is based on a high-quality, annotated azuki bean reference genome sequence.

Database Contents and Data Analysis

Gene annotation

For the first version of the azuki bean genome assembly (Vangularis_v1), gene structures were predicted by combining gene expression analysis and ab initio gene prediction (Sakai et al. 2015), resulting in 33,735 loci mapped to the Vangularis_v1 assembly. Of the 33,735 loci, 32,044 (95.0%) loci were predicted based on RNA sequencing (RNA-Seq) data. To compare the azuki bean gene content with those of other eudicot species, we obtained protein sequences of *Arabidopsis thaliana* (L.) Heynh. from TAIR10 (Lamesch et al. 2012), those of *Glycine max* (L.) Merr., *Medicago truncatula* Gaertn., *Phaseolus vulgaris* L. and *Populus trichocarpa* Torr. & A. Gray from Phytozome 10 (Goodstein et al. 2012), and those of *Vigna radiata* from http://plantgenomics.snu.ac.kr/, and constructed gene families using mclblastline implemented in the mcl (The Markov Cluster Algorithm) package (Enright et al. 2002). In total, 256,456 protein sequences encoded by 11,820 gene families were identified. Of the 31,241 azuki bean protein-coding genes, 26,840 (85.9%) had homologs in rosids or fabids (**Fig. 1**). Interestingly, the number of conserved genes in the *P. vulgaris* genome was almost the same as that of the azuki bean (**Fig. 1**). In addition, *G. max*, which has experienced a recent whole-genome duplication event (Schmutz et al. 2010), possessed approximately twice as many conserved genes as *V. angularis*. These observations suggest that a reservoir of approximately 27,000 functionally conserved genes in the haploid genome of a common ancestor of the Fabaceae species was maintained after diversification of these species. The exception is *M. truncatula* whose genome has undergone a large number of gene duplications (Young et al. 2011). The relatively small number of conserved genes in the *V. radiata* genome may be due to an incomplete genome assembly. There were 4,401 azuki bean genes that had no homologs in the other six eudicot genomes. We mapped protein sequences of these 4,401 genes to the genomes of *V. radiata*, *P. vulgaris* and *G. max* using Exonerate (Slater and Birney 2005) and found that 2,981 genes were not mapped to any of the three genomes with thresholds of $\geq$50% amino acid identity and $\leq$50% sequence coverage. Out of the 2,981 genes, 1,532 genes had no homologs in the nr database of NCBI with $\geq$50% amino acid identity and $\geq$50% sequence coverage, and showed an expression level of $\geq$1 fragment per kilobase of transcript per million fragments mapped (FPKM) in one or more of the eight tissues examined (see below), suggesting that these 1,532 genes are missing in the genomes of the other legume species and are specific to the azuki bean genome. The remaining 1,393 genes might be

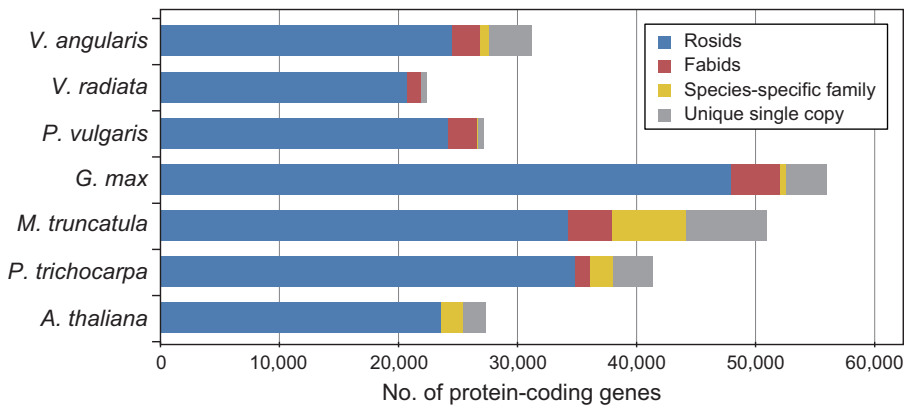

Fig. 1 Comparative gene analysis among seven dicot species. Colored bars represent the numbers of genes conserved in rosids (blue) and fabids (red), genes comprising species-specific families (yellow) and unique single-copy genes (gray).

artifacts or expressed in tissues other than the eight examined. The genome assembly and annotation have been deposited in DDBJ under accession Nos. AP015034–AP017294.

Gene expression data

We sequenced transcriptomes of eight tissues of the azuki bean by RNA-Seq to predict gene structures (Sakai et al. 2015). The expression level of each gene is shown at single nucleotide resolution on the GBrowse map (Fig. 2). In addition, FPKM values of each gene are shown in the GBrowse details page (Fig. 2). In total, 26,699 of the 33,735 genes showed an expression level of ≥1 FPKM in one or more tissues. We also obtained CAGE (Cap Analysis of Gene Expression) (Shiraki et al. 2003) data from the same transcriptome samples. Sequenced reads were mapped to the azuki bean genome using Bowtie2 (Langmead and Salzberg 2012) and aggregated as CAGE transcription start site (CTSS) data. Next, we clustered CTSSs using a simple distance-based clustering method and obtained a set

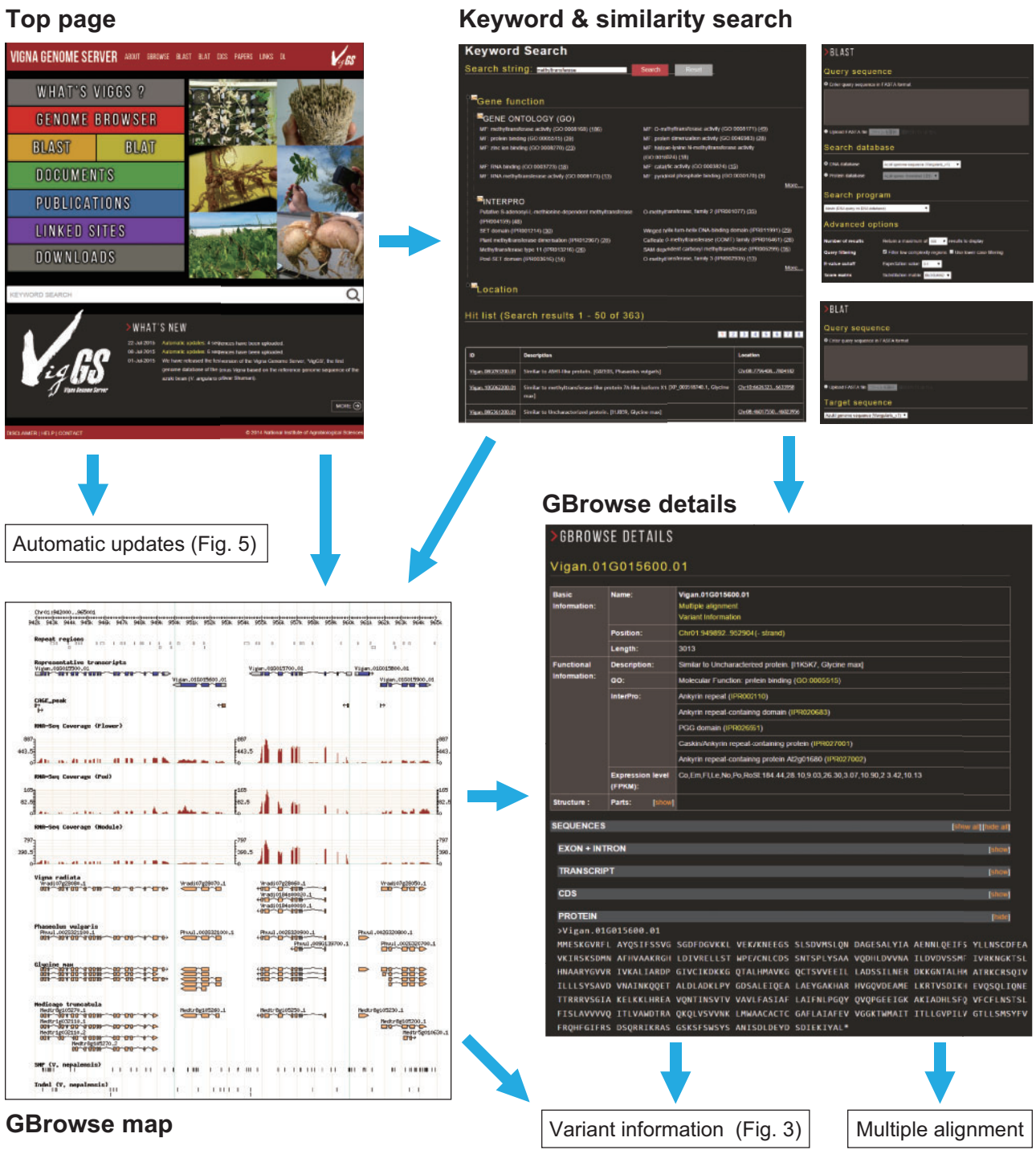

Fig. 2 Overall framework of the *Vig*GS.

of consensus clusters using the CAGEr package (Haberle et al. 2015). In this way, we detected 33,170 consensus clusters of which 30,185 (91.0%) were located within 3 kb upstream of, or inside, genes. Locations of the CAGE clusters are shown on the 'CAGE_peak' track in the GBrowse map.

Variant information

Variant data detected between azuki bean and its close wild relative, *V. nepalensis* Tateishi & Maxted, were incorporated into *Vig*GS. *Vigna nepalensis* has very similar morphological characteristics to *V. angularis* var. *nipponensis*, the wild ancestor of the azuki bean (Tateishi and Maxted 2002). Since *V. nepalensis* is cross-compatible with azuki bean, genetic maps of azuki bean have been constructed using the interspecific populations derived from the two species (Han et al. 2005, Isemura et al. 2007). We sequenced the *V. nepalensis* genome on the Illumina HiSeq platform to design single nucleotide polymorphism (SNP) markers used to genotype an F_2 population derived from an azuki bean and *V. nepalensis* cross (Sakai et al. 2015). Next, we constructed a genetic map and reconstructed chromosome sequences by anchoring the assembled scaffolds to the genetic map. To implement the variant information in *Vig*GS, we reanalyzed the *V. nepalensis* genome sequences. After trimming off low-quality and adaptor sequences using Trimmomatic (Bolger et al. 2014), we mapped sequence reads using BWA-MEM (Li and Durbin 2009). Next, we discarded PCR duplicates using Picard (MarkDuplicates) (http://broadinstitute.github.io/picard/) and carried out local realignment around indels with GATK (IndelRealigner) (McKenna et al. 2010). Variants were detected by GATK (HaplotypeCaller) and filtered by VariantFiltration in GATK with options 'DP < 5 || DP > 100 || QD < 2.0 || FS > 30.0 || MQ < 30.0'. We detected simple sequence repeats (SSRs) in the azuki bean genome with SSIRT (Temnykh et al. 2001) and indels in SSRs were annotated as 'SSR-associated indels' in the 'Variant information' page (**Fig. 3**). As a result, a total of 4,613,994 SNPs, 20,768 SSR-associated indels and 784,266 other indels were detected. We annotated the variants with SnpEff (Cingolani et al. 2012) to infer the functional effects exerted by the variants. We focused on variants predicted to lead to high impacts on gene function, such as alterations in splice sites, frameshifts and loss or gain of start/stop codons. Among the 5,419,028 variants, 7,683 resided in 4,078 genes and were predicted to be high impact, as defined by SnpEff including 'EXON_DELETED', 'FRAME_SHIFT', 'RARE_AMINO_ACID', 'SPLICE_SITE_ACCEPTOR', 'SPLICE_SITE_DONOR', 'STOP_LOST', 'START_LOST' and 'STOP_GAINED.' Since the variants causing such high impacts have the potential to result in proteins of altered length, we compared predicted protein length for the 4,078 genes with orthologs from four legume species, *V. radiata*, *P. vulgaris*, *G. max* and *M. truncatula* predicted using the OrthoMCL pipeline (Li et al. 2003). We found that 269 genes had significantly shorter or longer protein sequences (a difference of >2 SDs longer or shorter than the average length of orthologous genes), suggesting that these mutations resulted in severe functional alteration or even non-functional products. For example, Vigan.10G017700.01 harbored a single base pair deletion resulting in a frameshift in the coding sequence (CDS; **Fig. 4**), with the resulting CDS shortened to 220 amino acids, whereas the orthologous genes in *P. vulgaris* and *G. max* had a longer CDS of 361 and 364 amino acids, respectively. The inferred amino acid sequence of the *V. nepalensis* gene was fully aligned with those of the orthologous genes, suggesting that the deletion occurred in the azuki bean lineage, possibly during the domestication process. In *Vig*GS, functional effects predicted by SnpEff are available in the 'Variant information' page (**Fig. 3**) and users can compare protein sequences of the six dicot species by browsing comparative protein mapping data in the GBrowse map or by constructing multiple alignments (described below).

Comparative protein mapping data

We mapped the protein sequences of *G. max*, *P. vulgaris*, *M. truncatula* and *V. radiata* to the azuki bean genome using Exonerate and selected the best mapping locus for each sequence with thresholds of ≥70% amino acid identity and ≥90% sequence coverage. The mapped sequences and structures are shown for each species on the GBrowse map, which links to the GBrowse details page and the source databases.

Other data

We predicted 1,291 tRNA genes on the azuki bean genome using tRNAscan-SE (Lowe and Eddy 1997). We also mapped 297 SSR markers by conducting e-PCR (Schuler 1997) using the primer sequences (Han et al. 2005, Chaitieng et al. 2006, Isemura et al. 2012, Kongjaimun et al. 2012, Tian et al. 2013).

Database Functions

Visualizing gene structures and annotations

We implemented 21 tracks in GBrowse, including gene structures mapped to the Vangularis_v1 assembly, gene expression profiles for eight tissues, transcription start sites predicted from CAGE data sets, tRNA genes, repeat-masked regions, SNPs and indels detected between azuki bean and *V. nepalensis*, SSR markers, comparative protein mapping data and cowpea expressed sequnce tag (EST) mapping data. Detailed information for each feature of each track is shown in the customized GBrowse details page, where items in the original GFF (General Feature Format) file are displayed in the order and category defined by the administrator (**Fig. 2**). In addition, for azuki bean genes, transcripts, CDS and protein sequences are displayed.

Data search

*Vig*GS offers a keyword search and sequence similarity searches using BLAST (Altschul et al. 1990) or BLAT (Kent 2002). After an initial keyword search, users can further filter resulting gene hits using additional keywords. In addition, a list of functional domains, represented as Gene Ontology (GO) and InterPro domain IDs, is shown, and users can select any domain to filter gene hits or can select genes on a specific chromosome, scaffold or genomic region.

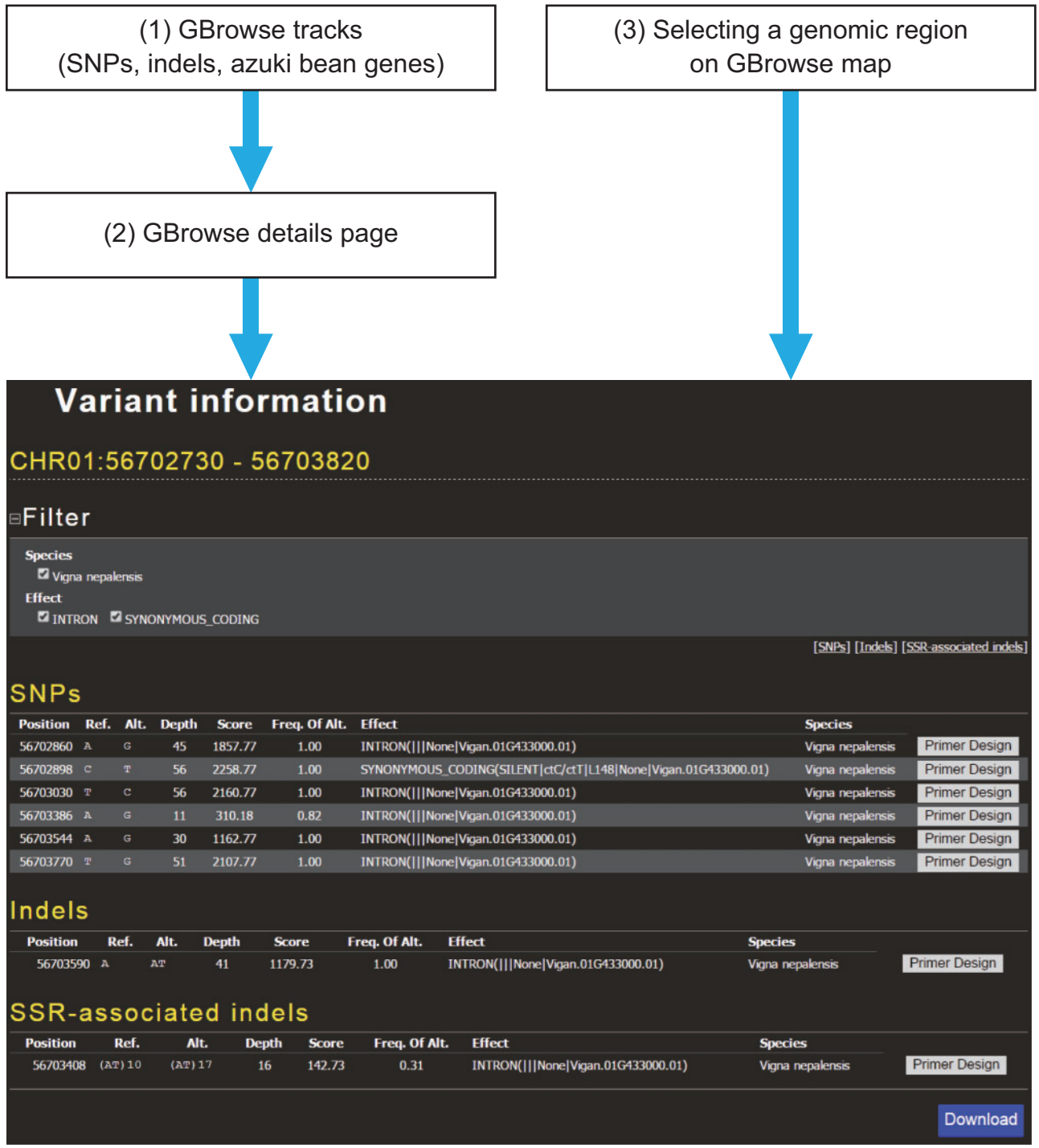

Fig. 3 Visualization of variant information.

*Vig*GS allows users to search any DNA or protein sequences against the azuki bean genome or gene sequences using BLAST or BLAT. Where query sequences are searched against the genome sequence, the results link to the GBrowse map to enable users to identify genomic regions to which query sequences align, which is especially useful when transcript sequences, such as ESTs or cDNAs, are aligned using BLAT.

Multiple gene alignment

Predicting orthologous genes among species of interest and constructing multiple gene alignments are fundamental steps in comparative genomic and molecular phylogenomic analyses, yet they are laborious and time consuming. In *Vig*GS, orthologous genes in six dicot species (indicated above), as well as paralogous azuki bean genes, were preliminarily predicted using the OrthoMCL pipeline, facilitating relatively simple

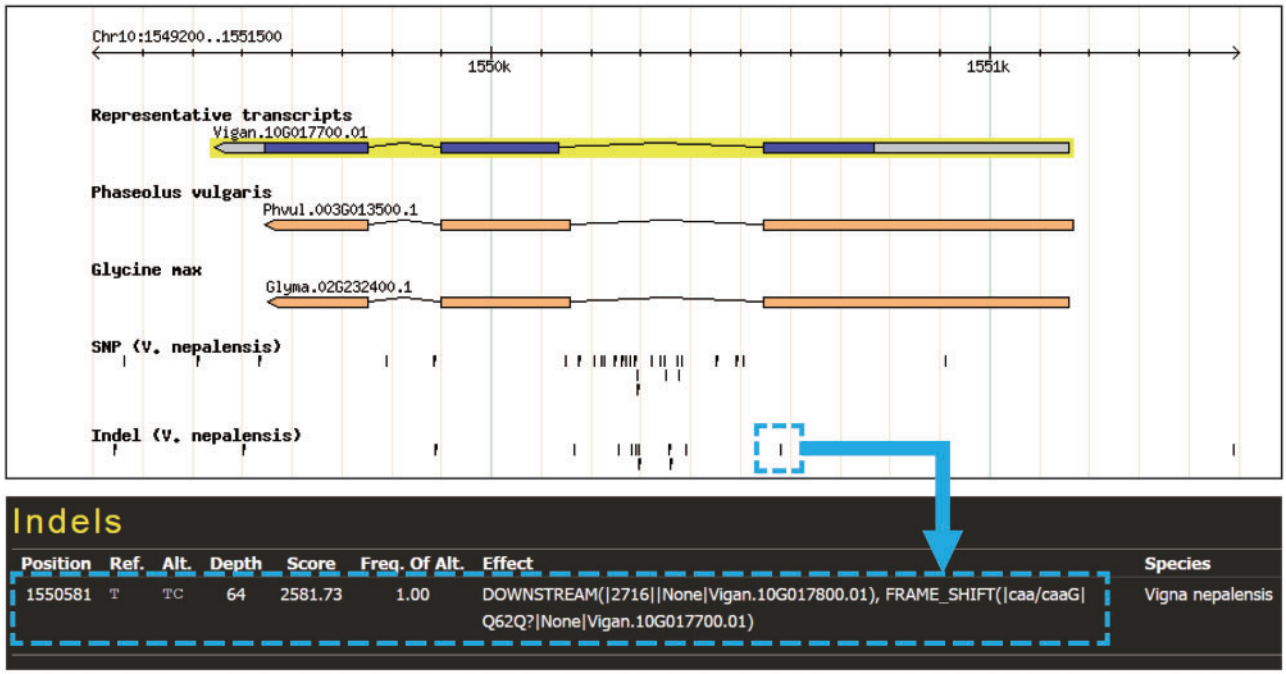

Fig. 4 GBrowse map of the Vigan.10G017700.01 gene with comparative protein mapping and variant data. A single base pair deletion in the azuki bean Vigan.10G017700.01 gene resulted in a frameshift and shorter CDS compared with *P. vulgaris* and *G. max*. Boxes shown on the comparative protein mapping tracks indicate amino acid-coding sequences.

construction of multiple gene alignments. The GBrowse azuki bean gene details pages provide a link to another page for multiple gene alignment construction (**Fig. 2**). Users select which species are to be aligned and carry out multiple alignments with ClustalW (Chenna et al. 2003) or MAFFT (Katoh and Standley 2013) using protein, CDS or transcript sequences. If more than one orthologous or paralogous gene is predicted in a species, users can select the longest sequence as representative or include every sequence for alignment. Query sequences can be downloaded in multi-FASTA format and alignment results can be downloaded in appropriate formats according to the program used.

Exploring variant information

There are three ways to access variant information in *Vig*GS: (i) from GBrowse tracks; (ii) from GBrowse details pages; and (iii) by selecting a genomic region on the GBrowse map (**Figs. 2**, **3**). Each SNP or indel shown on the GBrowse tracks links to a GBrowse details page, where a series of variant data are provided. In addition, in the 'Basic Information' section of the GBrowse details page, there is a link to the variant information page. GBrowse azuki bean gene details pages also offer links to pages displaying variants detected within the selected gene (**Fig. 2**). When a user clicks and drags a genomic region on the GBrowse map, a pop-up menu is displayed with a link to the variant information page, where variants are shown by their type: SNPs, indels and SSR-associated indels (**Fig. 3**). Variant information can be downloaded in VCF (Variant Call Format). For each variant, a series of details, including alternative allele, depth, score, frequency of alternative allele and SnpEff variant annotation, are provided. Users can filter variants by the effect assigned by SnpEff. In addition, users can design primer sequences targeting any variant site using Primer3 (Untergasser et al. 2012), and the specificity of primer sequences can be examined by performing BLASTN searches against the azuki bean genome. Users can download sequences of all primer pairs or only specific pairs in FASTA format.

Automatic updates

Transcript and protein sequences are deposited in public databases daily after the initial annotation of a genome. Such newly deposited sequence information is valuable to improve the accuracy of genome annotation (Yandell and Ence 2012). In addition, literature relating to new sequences provides additional functional information about annotated genes. Therefore, genome annotation ideally requires constant updating to incorporate such newly deposited sequence information. In reality, however, in many cases genome annotation is rarely updated, since updating can be more labor-intensive than annotating from scratch (Yandell and Ence 2012). To provide users with up to date sequence and literature data for legume genes, we have developed a unique system in which our database incorporates newly deposited mRNA sequences automatically. Currently *Vig*GS downloads mRNA sequences of *G. max*, *P. vulgaris*, *M. truncatula* and any *Vigna* species once a week from NCBI GenBank. The obtained sequences are mapped to the azuki bean genome by BLAT or Exonerate (if CDS are indicated in the sequence record). Mapped sequences are uploaded to GBrowse and displayed on the 'Automatic Updates' track, with links to the original GenBank and PubMed records in the

GBrowse details page. In addition, exon–intron structures of the mapped sequences are compared with annotated azuki bean genes. If all the intron positions are consistent with those of any of the annotated genes, such sequences are reported as 'Known variant' or otherwise reported as 'New variant' if at least one of the introns is unique. Mapped sequences without any overlapping annotated genes are classified into the category 'New Locus'. Results of automatic updates are reported in the 'What's new' window on the top page, which links to a table showing the accession numbers, species names, descriptions, PubMed IDs, mapping results and links to GBrowse map for the submitted sequences (**Fig. 5**). Users can filter sequences by accession number, description, species name, map position and status.

System Architecture and Software

*Vig*GS was developed with the Mojolicious web application framework for Perl and implemented on a Linux server with CentOS release 6.4 (Final), Apache web server, and Nginx as a reverse proxy. All data based on the azuki bean genome were installed in a MySQL database and visualized on GBrowse 2.55.

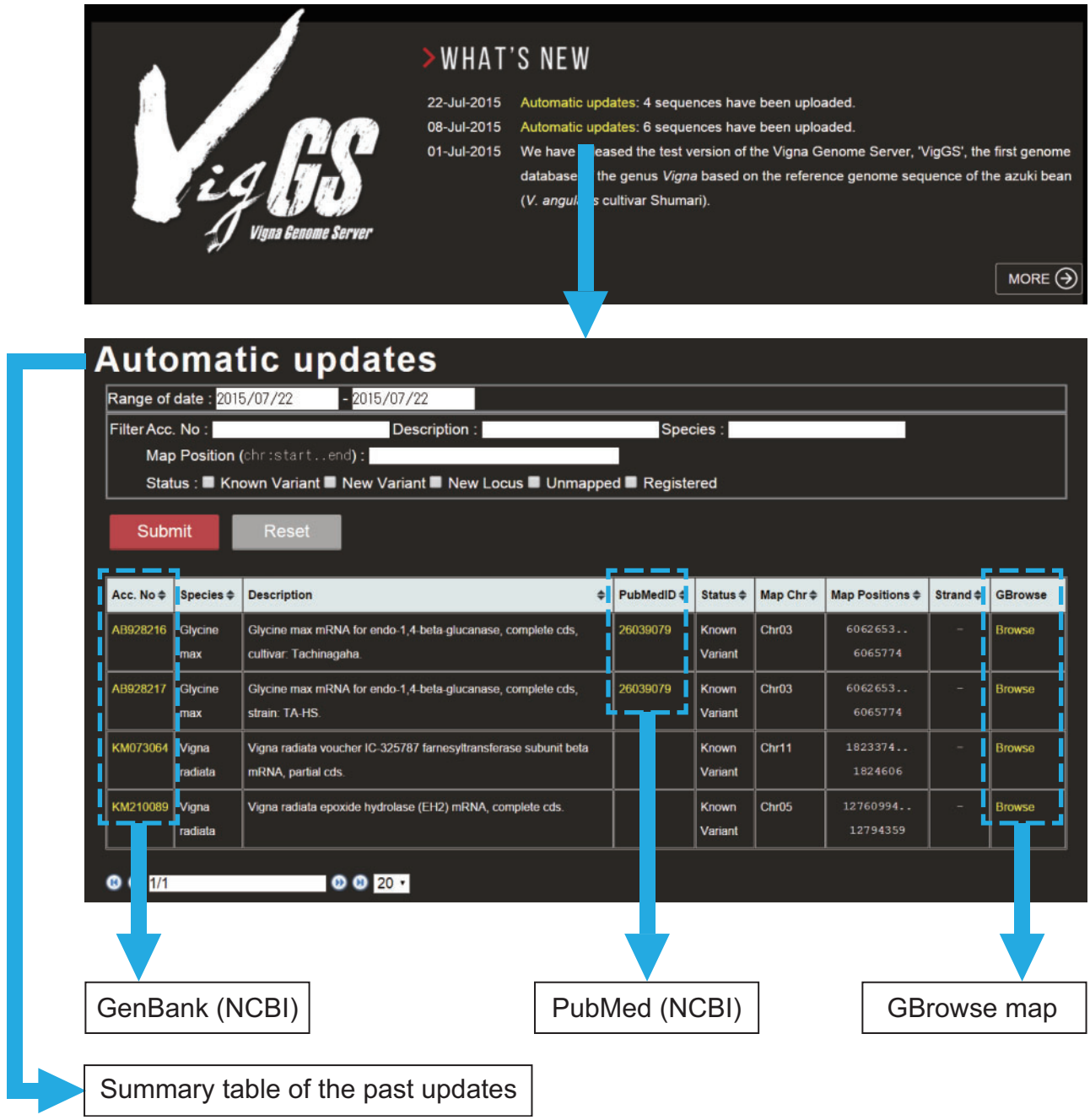

Fig. 5 Automatic updates. The number of uploaded sequences is reported once a week in the 'What's new' section of the top page. The title of the information page ('Automatic updates') links to a list of past updates arranged by date.

Table 1 List of the 13 wild *Vigna* species for which genome sequencing is in process

Subgenus	Section	Species	Natural distribution	Habitat and ecology
Ceratotropis	Aconitifoliae	*V. aconitifolia* (Jacq.) Maréchal	South Asia	Dry, sandy soil
		V. aridicola N. Tomooka & Maxted	South Asia	Open, dry, sandy soil grasslands
		V. indica T.M. Dixit, K.V. Bhat & S.R. Yadav	South Asia	Grasslands of drier parts
		V. stipulacea Kuntze	South Asia	Open, clay soil
		V. trilobata (L.) Verdcourt	South Asia and Myanmar	Confined to sandy lowland soils
	Ceratotropis	*V. mungo* (L.) Hepper	South Asia	Open, wet, disturbed habitats
	Angulares	*V. exilis* Tateishi & Maxted	Southeast Asia (Thailand)	Only on open limestone outcrops (Thailand)
		V. minima (Roxb.) Ohwi & Ohashi	Southeast and East Asia, Papua New Guinea	Deciduous forest floor or open wet habitats such as paddy bunds
		V. nakashimae (Ohwi) Ohwi & Ohashi	East Asia	Open, wet, disturbed habitats of northern East Asia
		V. riukiuensis (Ohwi) Ohwi & Ohashi	East Asia	Open, coastal grasslands subject to grazing, cutting or high wind
Plectrotropis	Plectrotropis	*V. vexillata* (L.) A. Rich.	Pan tropical	Grassland, savannah, swamp, lake shore, etc.
Vigna	Vigna	*V. luteola* (Jacq.) Benth	Pan tropical	Grasslands, lake edges, riverbanks, swamp edges, etc.
		V. marina (Burm.) Merr.	Pan tropical	Seashores, etc.

Modified from Maxted et al. (2004), Tomooka et al. (2006), Dixit et al. (2011) and Tomooka et al. (2011).

Conclusion and Future Prospects

*Vig*GS is the first genome database of the genus *Vigna* and an important milestone toward the comprehensive understanding of plant adaptation to stress environments. Currently genome sequencing of 13 wild *Vigna* species is in process (**Table 1**). After annotating the genomes of wild species, we will conduct comparative protein mapping based on the genome of each species and genome alignment for each pair of closely related species; the data will be implemented in *Vig*GS. Such comparative genomic data will provide the basic information on the gain and loss of genes and the genomic rearrangements which occurred in the genomes of wild species, which will further shed light on the genetic consequences of the adaptive evolution of the *Vigna* species. *Vig*GS will play a central role as a knowledge base for genomic and genetic studies of *Vigna* and other legume species.

Funding

This research project was funded by the Scientific Technique Research Promotion Program for Agriculture, Forestry, Fisheries, and Food industry.

Acknowledgements

We would like to thank Ryo Aono of YUMA DESIGN Inc. and Hiroaki Takahashi for designing a unique logo for *Vig*GS.

Disclosures

The authors have no conflicts of interest to declare.

References

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

Bolger, A.M., Lohse, M. and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120.

Bolser, D.M., Kerhornou, A., Walts, B. and Kersey, P. (2015) Triticeae resources in Ensembl Plants. *Plant Cell Physiol.* 56: e3.

Chaitieng, B., Kaga, A., Tomooka, N., Isemura, T., Kuroda, Y. and Vaughan, D.A. (2006) Development of a black gram [*Vigna mungo* (L.) Hepper] linkage map and its comparison with an azuki bean [*Vigna angularis* (Willd.) Ohwi and Ohashi] linkage map. *Theor. Appl. Genet.* 113: 1261–1269.

Chankaew, S., Isemura, T., Naito, K., Ogiso-Tanaka, E., Tomooka, N., Somta, P., et al. (2014) QTL mapping for salt tolerance and domestication-related traits in *Vigna marina* subsp. *oblonga*, a halophytic species. *Theor. Appl. Genet.* 127: 691–702.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G., et al. (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.* 31: 3497–3500.

Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., et al. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *DrosophilaDrosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 6: 80–92.

Dixit, T.M., Sutar, S.P., Yadav, S.R., Bhat, K.V. and Rao, S.R. (2011) *Vigna indica*, a new name for *Vigna trilobata* var. *pusilla* and a note on section Aconitifoliae in India. Rheedea 21: 1–7.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., et al. (2009) Real-time DNA sequencing from single polymerase molecules. *Science* 323: 133–138.

Enright, A.J., Van Dongen, S. and Ouzounis, C.A. (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* 30: 1575–1584.

Gepts, P., Beavis, W.D., Brummer, E.C., Shoemaker, R.C., Stalker, H.T., Weeden, N.F., et al. (2005) Legumes as a model plant family. Genomics for food and feed report of the Cross-Legume Advances Through Genomics Conference. *Plant Physiol.* 137: 1228–1235.

Gonzales, M.D., Archuleta, E., Farmer, A., Gajendran, K., Grant, D., Shoemaker, R., et al. (2005) The Legume Information System (LIS): an integrated information resource for comparative legume biology. *Nucleic Acids Res.* 33: D660–D665.

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., et al. (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40: D1178–1186.

Grant, D., Nelson, R.T., Cannon, S.B. and Shoemaker, R.C. (2010) SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Res.* 38: D843–D846.

Haberle, V., Forrest, A.R., Hayashizaki, Y., Carninci, P. and Lenhard, B. (2015) CAGEr: precise TSS data retrieval and high-resolution promoterome mining for integrative analyses. *Nucleic Acids Res.* 43: e51.

Han, O.K., Kaga, A., Isemura, T., Wang, X.W., Tomooka, N. and Vaughan, D.A. (2005) A genetic linkage map for azuki bean [*Vigna angularis* (Willd.) Ohwi & Ohashi]. *Theor. Appl. Genet.* 111: 1278–1287.

Isemura, T., Kaga, A., Konishi, S., Ando, T., Tomooka, N., Han, O.K., et al. (2007) Genome dissection of traits related to domestication in azuki bean (*Vigna angularis*) and comparison with other warm-season legumes. *Ann. Bot.* 100: 1053–1071.

Isemura, T., Kaga, A., Tabata, S., Somta, P., Srinives, P., Shimizu, T., et al. (2012) Construction of a genetic linkage map and genetic analysis of domestication related traits in mungbean (*Vigna radiata*). *PLoS One* 7: e41304.

Kang, Y.J., Kim, S.K., Kim, M.Y., Lestari, P., Kim, K.H., Ha, B.K., et al. (2014) Genome sequence of mungbean and insights into evolution within *Vigna* species. *Nat. Commun.* 5: 5443.

Kang, Y.J., Satyawan, D., Shim, S., Lee, T., Lee, J., Hwang, W.J., et al. (2015) Draft genome sequence of adzuki bean, *Vigna angularis*. *Sci. Rep.* 5: 8069.

Katoh, K. and Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30: 772–780.

Kent, W.J. (2002) BLAT—the BLAST-like alignment tool. *Genome Res.* 12: 656–664.

Kersey, P.J., Allen, J.E., Christensen, M., Davis, P., Falin, L.J., Grabmueller, C., et al. (2014) Ensembl Genomes 2013: scaling up access to genome-wide data. *Nucleic Acids Res.* 42: D546–D552.

Kongjaimun, A., Kaga, A., Tomooka, N., Somta, P., Shimizu, T., Shu, Y., et al. (2012) An SSR-based linkage map of yardlong bean (*Vigna unguiculata* (L.) Walp. subsp. *unguiculata* Sesquipedalis Group) and QTL analysis of pod length. *Genome* 55: 81–92.

Krishnakumar, V., Choi, Y., Beck, E., Wu, Q., Luo, A., Sylvester, A., et al. (2015a) A maize database resource that captures tissue-specific and subcellular-localized gene expression, via fluorescent tags and confocal imaging (Maize Cell Genomics Database). *Plant Cell Physiol.* 56: e12.

Krishnakumar, V., Kim, M., Rosen, B.D., Karamycheva, S., Bidwell, S.L., Tang, H., et al. (2015b) MTGD: the *Medicago truncatula* genome database. *Plant Cell Physiol.* 56: e1.

Lamesch, P., Berardini, T.Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., et al. (2012) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* 40: D1202–D1210.

Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9: 357–359.

Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25: 1754–1760.

Li, J., Dai, X., Liu, T. and Zhao, P.X. (2012) LegumeIP: an integrative database for comparative genomics and transcriptomics of model legumes. *Nucleic Acids Res.* 40: D1221–1229.

Li, L., Stoeckert, C.J., Jr. and Roos, D.S. (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13: 2178–2189.

Lowe, T.M. and Eddy, S.R. (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25: 955–964.

Maxted, N., Mabuza Dlamini, P., Moss, H., Padulosi, S., Jarvis, A. and Guarino, L. (2004) African Vigna: An Ecogeographic Study. p. 454. International Plant Genetic Resources Institute (IPGRI).

McCouch, S., Baute, G.J., Bradeen, J., Bramel, P., Bretting, P.K., Buckler, E., et al. (2013) Agriculture: feeding the future. *Nature* 499: 23–24.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20: 1297–1303.

Palmgren, M.G., Edenbrandt, A.K., Vedel, S.E., Andersen, M.M., Landes, X., Osterberg, J.T., et al. (2015) Are we ready for back-to-nature crop breeding? *Trends Plant Sci.* 20: 155–164.

Sakai, H., Lee, S.S., Tanaka, T., Numa, H., Kim, J., Kawahara, Y., et al. (2013) Rice Annotation Project Database (RAP-DB): an integrative and interactive database for rice genomics. *Plant Cell Physiol.* 54: e6.

Sakai, H., Naito, K., Ogiso-Tanaka, E., Takahashi, Y., Iseki, K., Muto, C., et al. (2015) The power of single molecule real-time sequencing technology in the *de novo* assembly of a eukaryotic genome. *Sci Rep.* 5: 16780.

Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., et al. (2010) Genome sequence of the palaeopolyploid soybean. *Nature* 463: 178–183.

Schrire, B.D. (2005) Tribe Phaseoleae. *In* Legumes of the World. Edited by Lewis, G., Schrire, B., Mackinder, B. and Lock, M. pp. 393–431. Royal Botanic Gardens, Kew.

Schuler, G.D. (1997) Sequence mapping by electronic PCR. *Genome Res.* 7: 541–550.

Shiraki, T., Kondo, S., Katayama, S., Waki, K., Kasukawa, T., Kawaji, H., et al. (2003) Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc. Natl. Acad. Sci. USA* 100: 15776–15781.

Slater, G.S. and Birney, E. (2005) Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6: 31.

Tateishi, Y. and Maxted, N. (2002) New species and combinations in *Vigna* subgenus *Ceratotropis* (Piper) Verdc. (*Leguminosae, Phaseoleae*). *Kew Bull.* 57: 625–633.

Temnykh, S., DeClerck, G., Lukashova, A., Lipovich, L., Cartinhour, S. and McCouch, S. (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. *Genome Res.* 11: 1441–1452.

Tian, J., Isemura, T., Kaga, A., Vaughan, D.A. and Tomooka, N. (2013) Genetic diversity of the rice bean (*Vigna umbellata*) genepool as assessed by SSR markers. *Genome* 56: 717–727.

Tomooka, N., Kaga, A. and Vaughan, D.A. (2006) The Asian *Vigna* (*Vigna* Subgenus *Ceratotropis*) Biodiversity and Evolution. In Plant Genome: Biodiversity and Evolution. Edited by Sharma, A.K. and Sharma, A. pp. 87–126. CRC Press.

Tomooka, N., Kaga, A., Isemura, T. and Vaughan, D.A. (2011) Vigna. In Wild Crop Relatives: Genomic and Breeding Resources. Edited by Kole, C. pp. 291–311. Springer-Verlag Berlin Heidelberg.

Tomooka, N., Naito, K., Kaga, A., Sakai, H., Isemura, T., Ogiso-Tanaka, E., et al. (2014) Evolution, domestication and neo-domestication of the genus *Vigna*. *Plant Genetic Resources* 12: S168–S171.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., et al. (2012) Primer3—new capabilities and interfaces. *Nucleic Acids Res.* 40: e115.

Yandell, M. and Ence, D. (2012) A beginner's guide to eukaryotic genome annotation. *Nat. Rev. Genet.* 13: 329–342.

Yonemaru, J., Ebana, K. and Yano, M. (2014) HapRice, an SNP haplotype database and a web tool for rice. *Plant Cell Physiol.* 55: e9.

Yoshida, Y., Marubodee, R., Ogiso-Tanaka, E., Iseki, K., Isemura, T., Takahashi, Y., et al. (2015) Salt tolerance in wild relatives of adzuki bean, *Vigna angularis* (Willd.) Ohwi et Ohashi. *Genet. Resour. Crop Evol.*

Young, N.D., Debelle, F., Oldroyd, G.E., Geurts, R., Cannon, S.B., Udvardi, M.K., et al. (2011) The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature* 480: 520–524.