

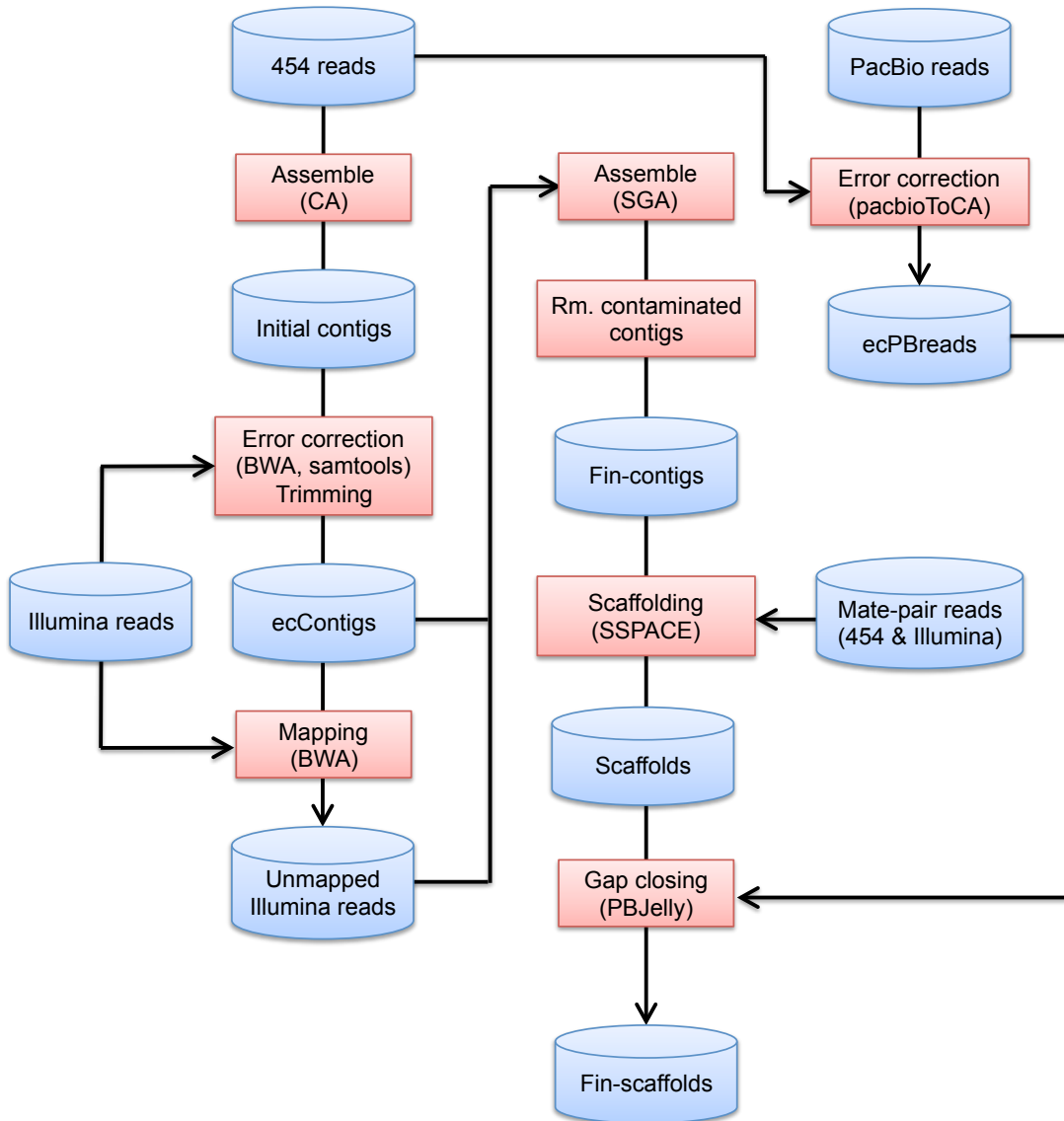
## **The power of single molecule real-time sequencing technology in the *de novo* assembly of a eukaryotic genome**

Hiroaki Sakai<sup>1</sup>, Ken Naito<sup>2\*</sup>, Eri Ogiso-Tanaka<sup>2</sup>, Yu Takahashi<sup>2</sup>, Kohtaro Iseki<sup>2</sup>, Chiaki Muto<sup>2</sup>, Kazuhito Satou<sup>3</sup>, Kuniko Teruya<sup>3</sup>, Akino Shiroma<sup>3</sup>, Makiko Shimoji<sup>3</sup>, Takashi Hirano<sup>3</sup>, Takeshi Itoh<sup>1</sup>, Akito Kaga<sup>2</sup>, and Norihiko Tomooka<sup>2</sup>

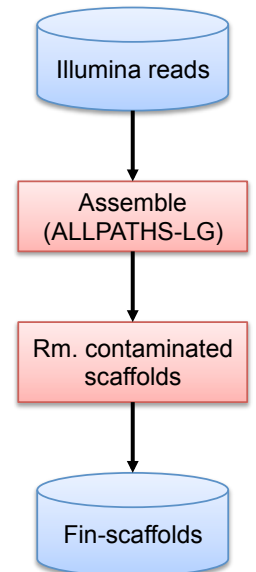
<sup>1</sup>Agrogenomics Research Center, National Institute of Agrobiological Sciences, 2-1-2 Kannondai, Tsukuba, Ibaraki, 305-8602, Japan. <sup>2</sup>Genetic Resources Center, National Institute of Agrobiological Sciences, 2-1-2 Kannondai, Tsukuba, Ibaraki, 305-8602, Japan. <sup>3</sup>Okinawa Institute of Advanced Sciences, Uruma, 904-2234, Japan.

Correspondence should be addressed to Ken Naito (knaito@affrc.go.jp).

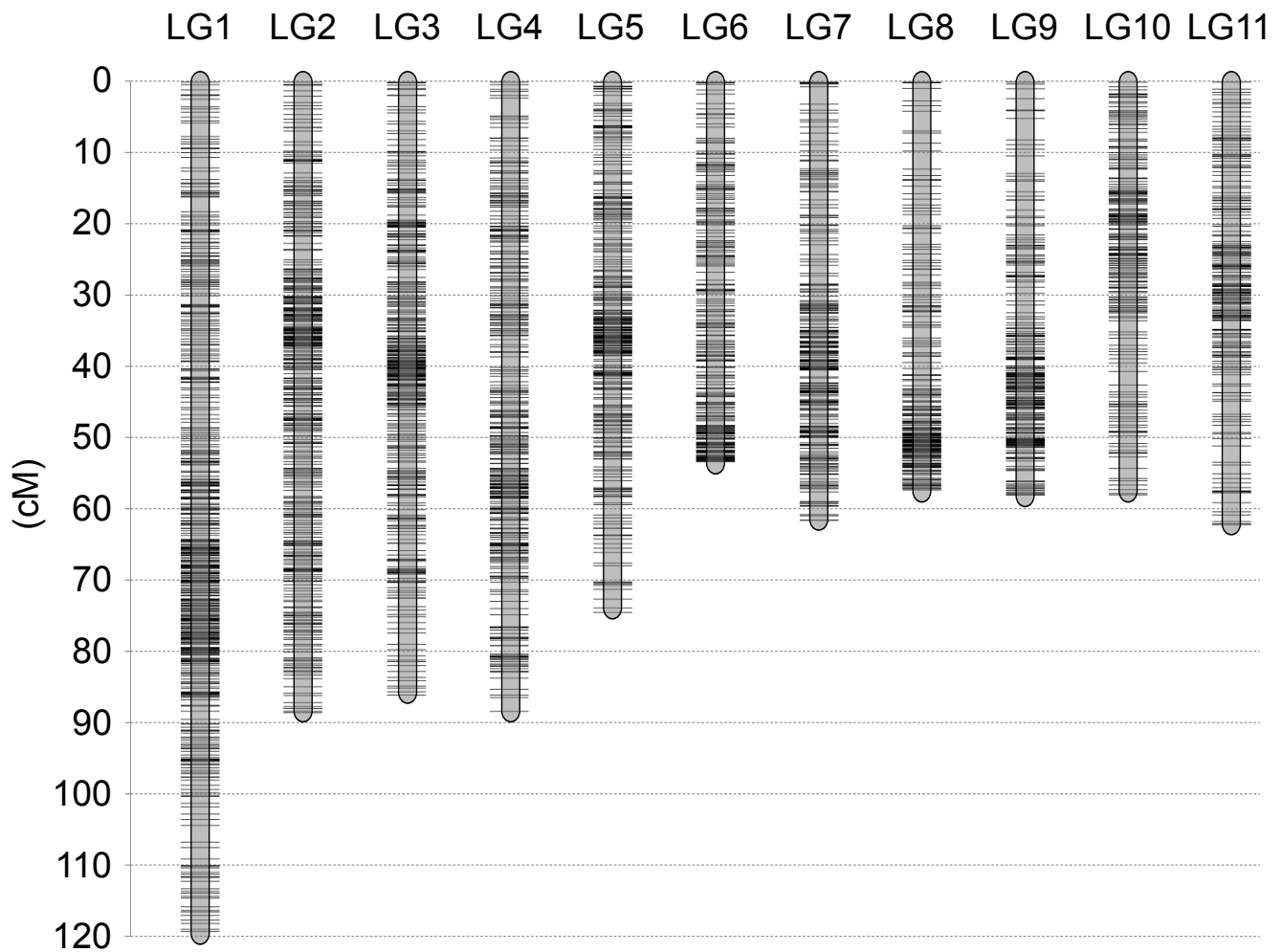
## Assembly\_1



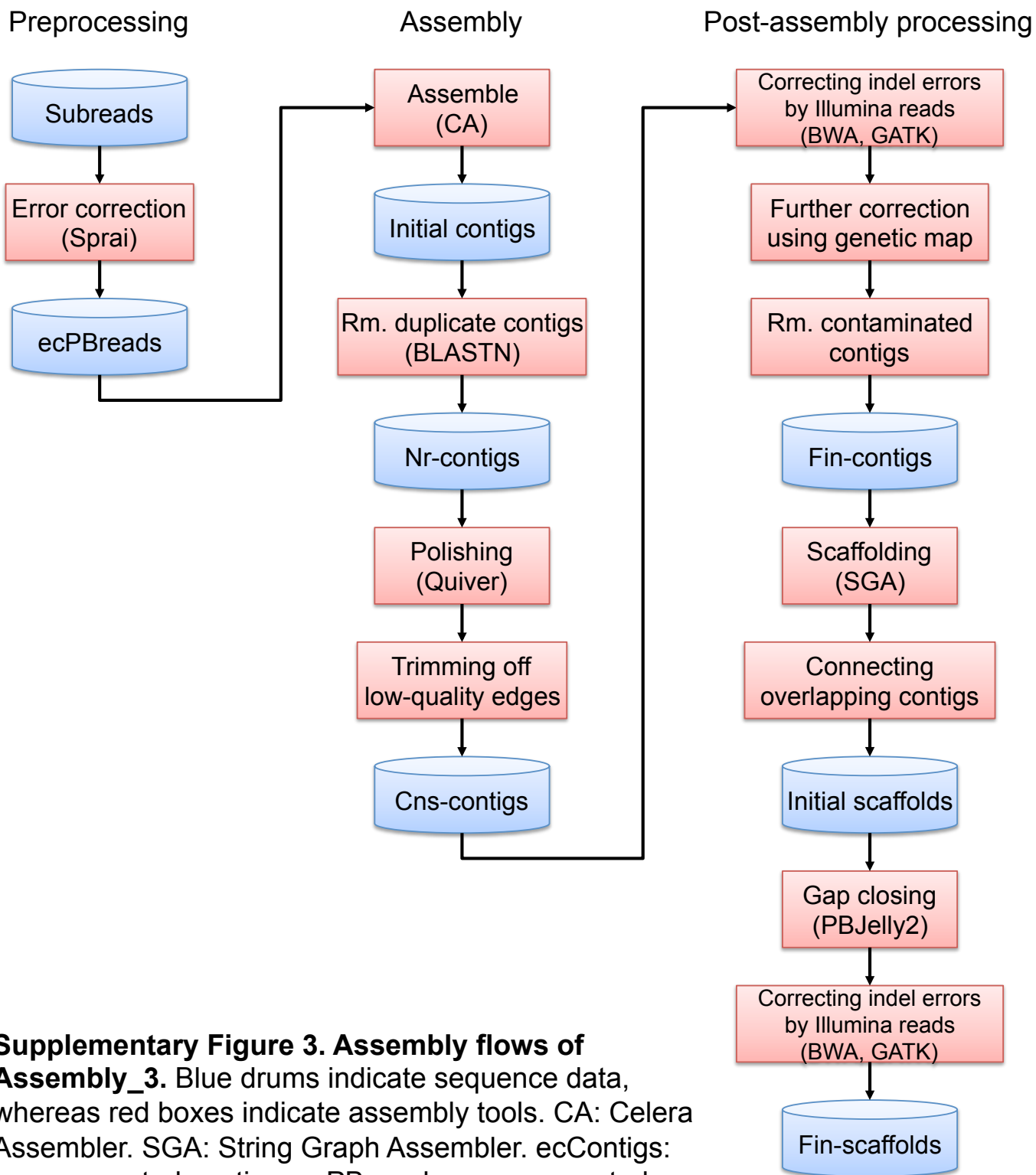
## Assembly\_2



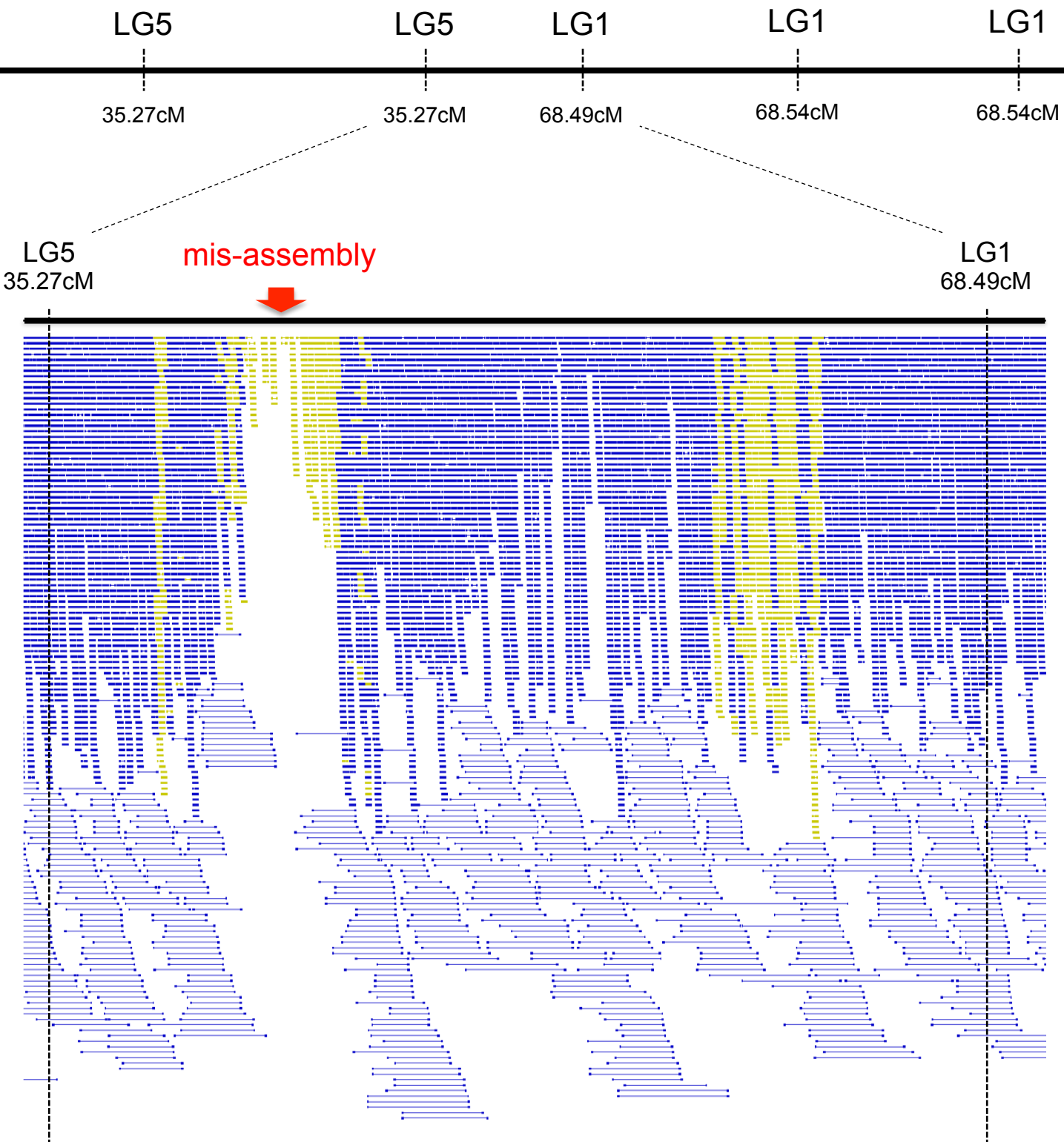
**Supplementary Figure 1. Assembly flows of Assembly\_1 and Assembly\_2.** Blue drums indicate sequence data, whereas red boxes indicate assembly tools. CA: Celera Assembler. SGA: String Graph Assembler. ecContigs: error-corrected contigs. ecPBreads: error-corrected PacBio reads. Rm: remove. Fin-contigs: final contigs. Fin-scaffolds: final scaffolds.



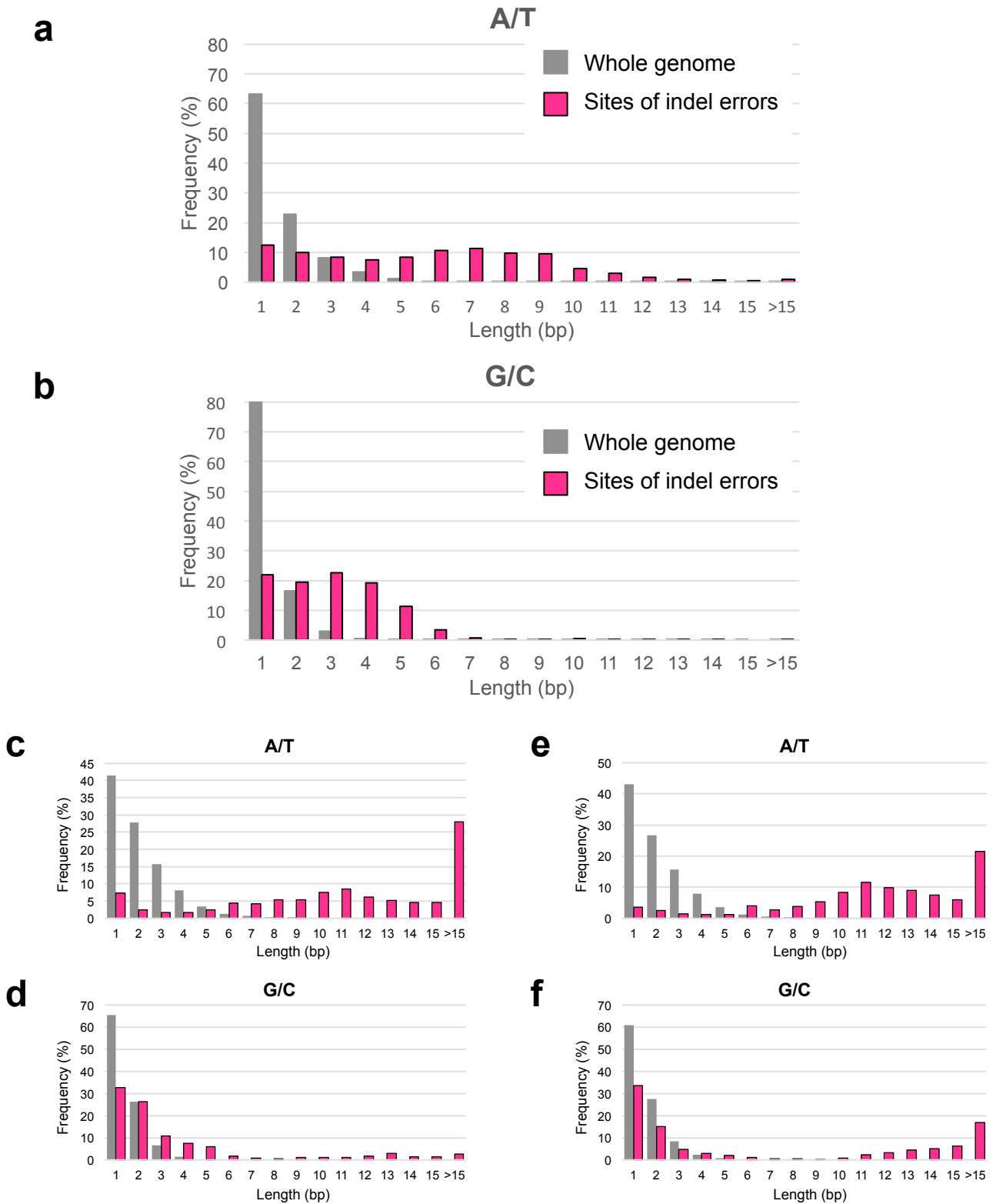
**Supplementary Figure 2. A high-density genetic linkage map of the azuki bean.** In total, 4,912 marker loci were genotyped in 995 F<sub>2</sub> plants derived from azuki bean x *V. nepalensis*.



**Supplementary Figure 3. Assembly flows of Assembly\_3.** Blue drums indicate sequence data, whereas red boxes indicate assembly tools. CA: Celera Assembler. SGA: String Graph Assembler. ecContigs: error-corrected contigs. ecPB reads: error-corrected PacBio reads. Rm: remove. Nr-contigs: non-redundant contigs. Cns-contigs: consensus contigs. Fin-contigs: final contigs. Fin-scaffolds: final scaffolds.



**Supplementary Figure 4. A selection of misassembly sites.** In this contig, DNA markers on LG1 and LG5 were present. Blue thick lines indicate the uniquely mapped paired-end reads with 250 bp inserts 250 bp, while the blue dots bridged with thin lines indicate those of 3 kb inserts. The yellow lines indicate the reads with multiple hits.

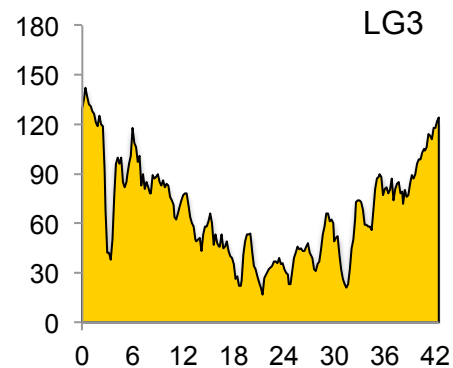
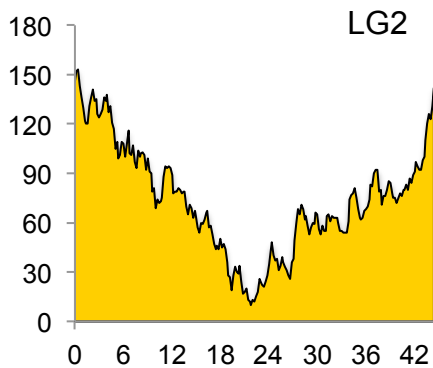
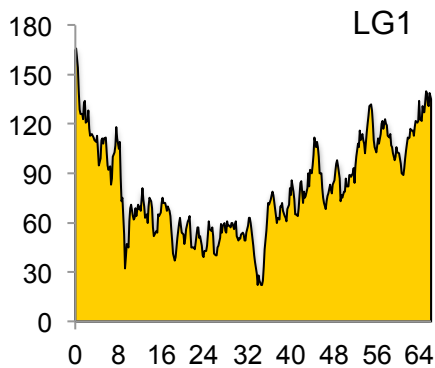


**Supplementary Figure 5. Frequency of homopolymer length at indel-errors in PacBio assemblies. (a)** Homopolymers of A/T or **(b)** Homopolymers of G/C in the azuki bean genome. Gray bars indicate frequency of homopolymers in the whole assembled sequences as a control. **(c)(d)** *Arabidopsis thaliana*. **(e)(f)** *Drosophila melanogaster*.

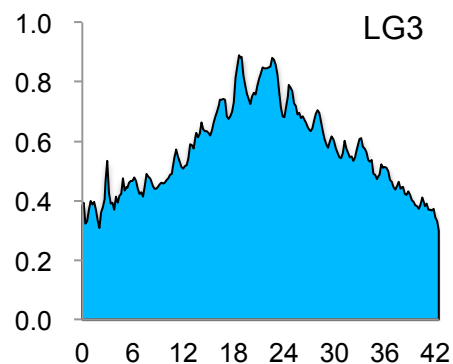
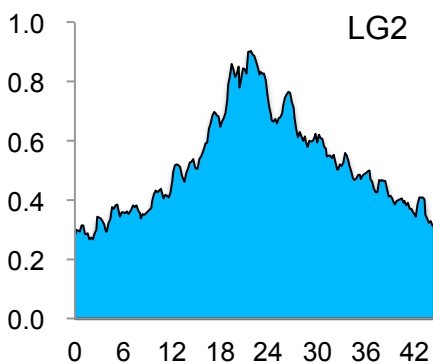
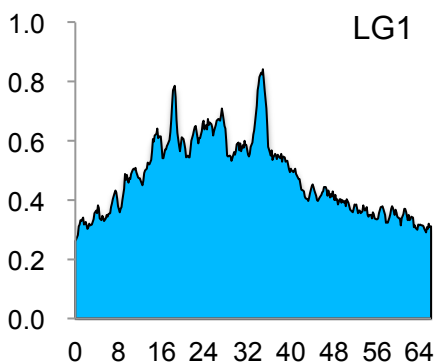
Supplementary Table 3. Summary of pseudomolecules in the Assembly\_3.

	No. of scaffolds	Total scaffolds length	Pseudomolecule length
Chr01	35	65,919,377	67,115,795
Chr02	28	45,096,540	45,515,668
Chr03	27	41,962,598	43,462,759
Chr04	25	53,470,133	54,073,736
Chr05	25	36,897,451	37,380,367
Chr06	23	38,789,640	38,860,970
Chr07	20	32,020,224	33,495,452
Chr08	32	46,426,616	46,951,433
Chr09	26	35,726,726	37,388,052
Chr10	17	28,785,286	28,886,040
Chr11	21	37,398,860	38,115,440
Unmapped	2250	51,515,385	-

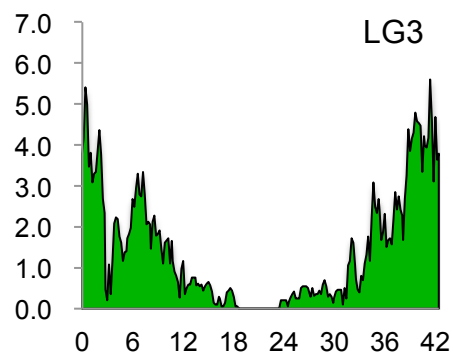
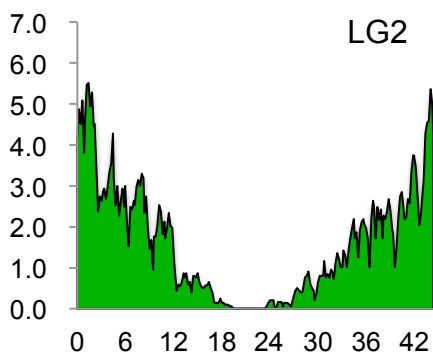
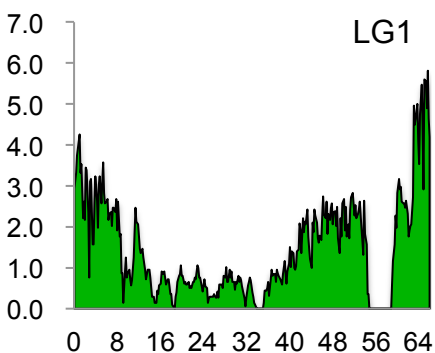
## Gene Density (Genes/Mb)



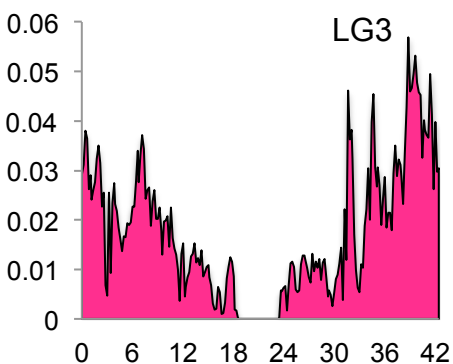
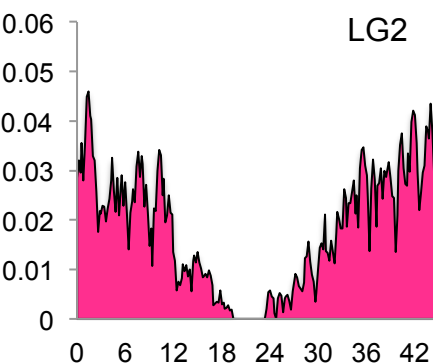
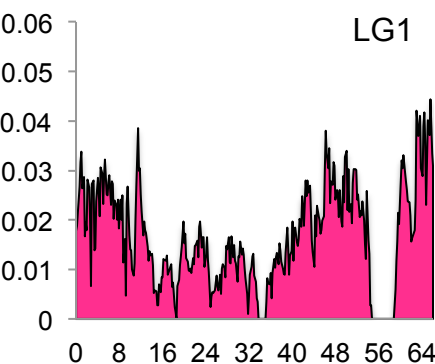
## Repeat Rate/Mb



## Recombination Frequency (cM/Mb)



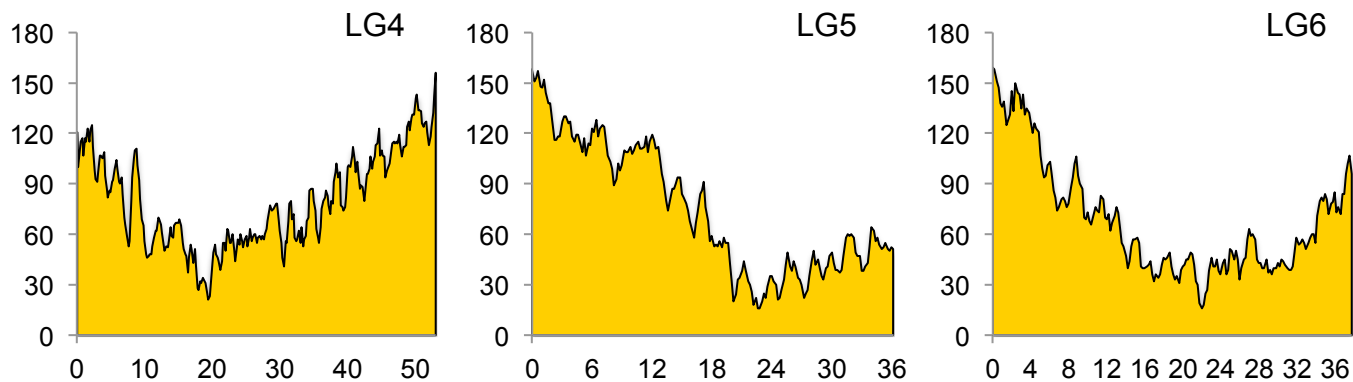
## Recombination Frequency (cM/Gene)



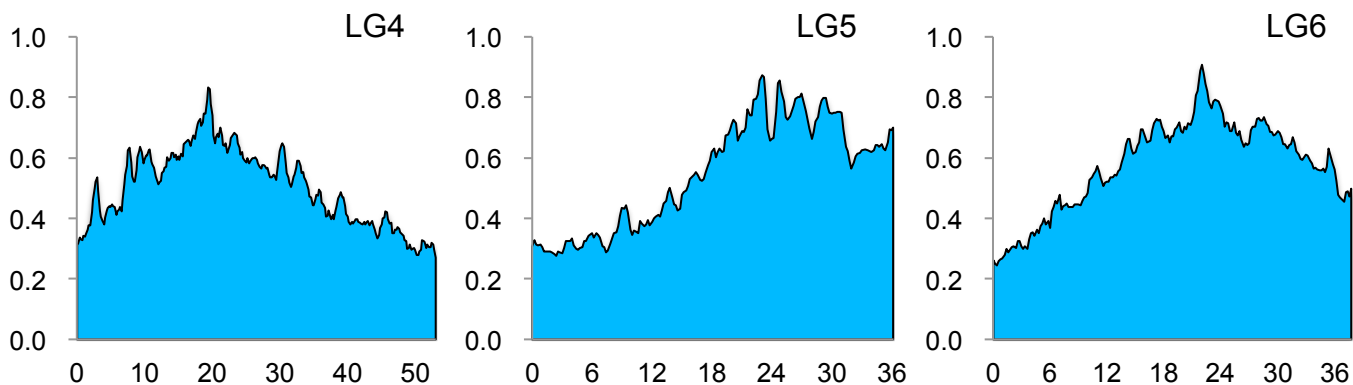
Supplementary Figure 6. An overview of the azuki bean genome.



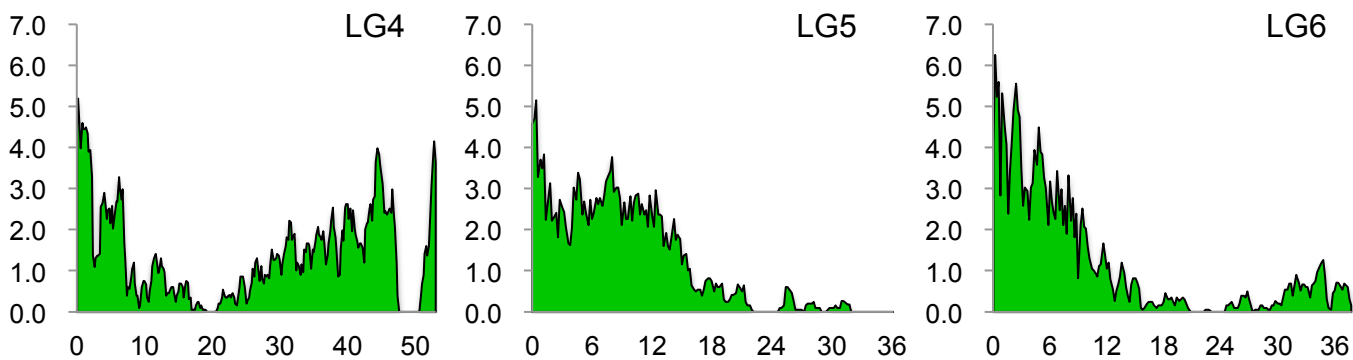
## Gene Density (Genes/Mb)



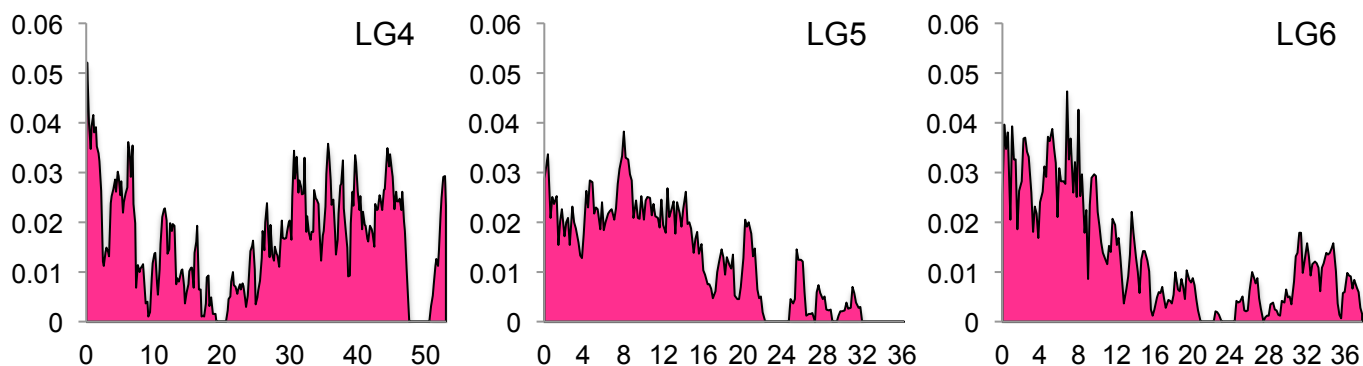
## Repeat Rate/Mb



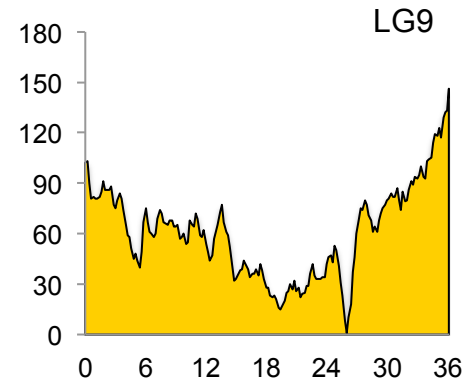
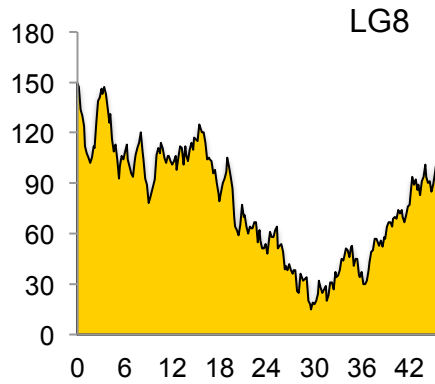
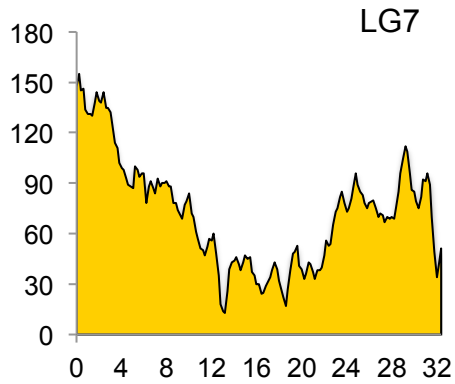
## Recombination Frequency (cM/Mb)



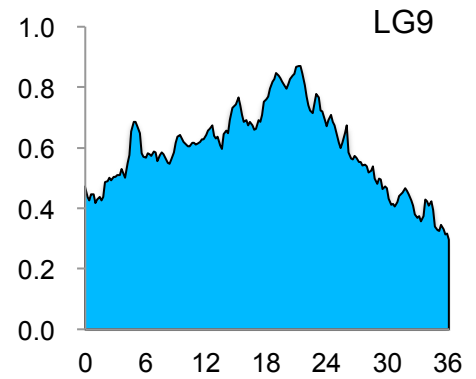
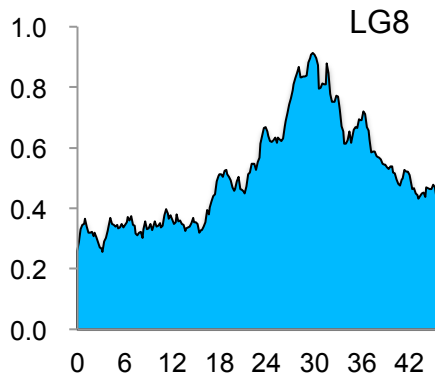
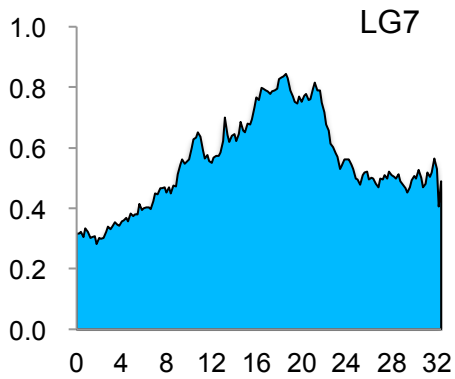
## Recombination Frequency (cM/Gene)



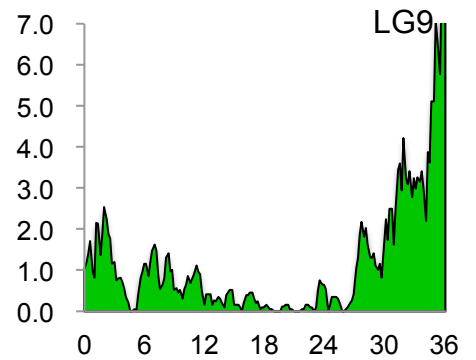
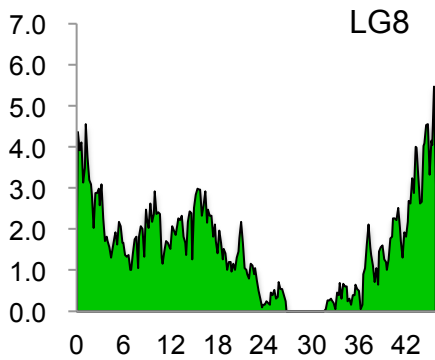
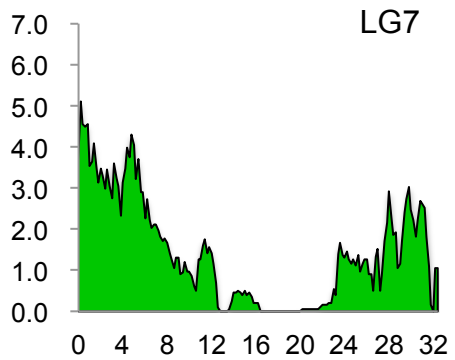
## Gene Density (Genes/Mb)



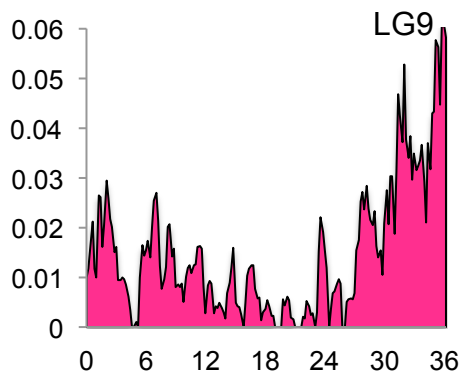
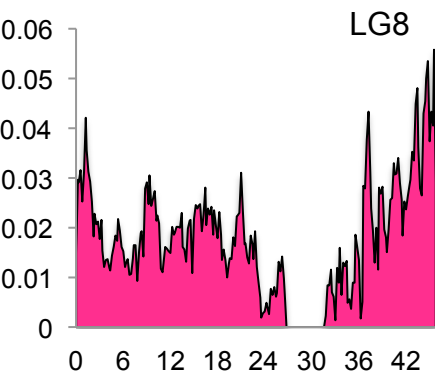
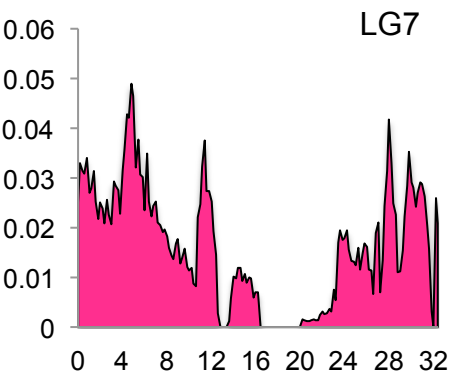
## Repeat Rate/Mb



## Recombination Frequency (cM/Mb)

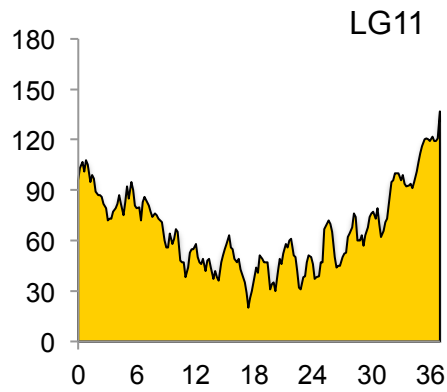
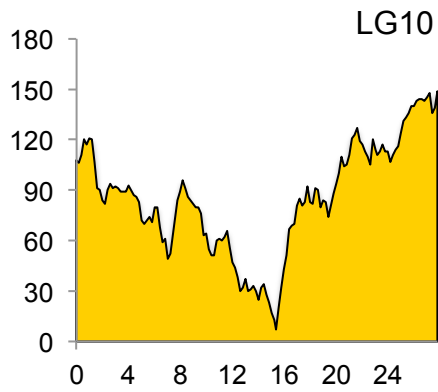


## Recombination Frequency (cM/Gene)

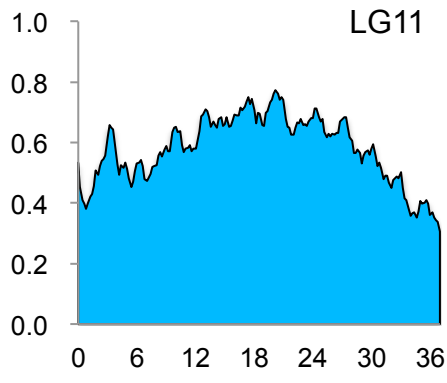
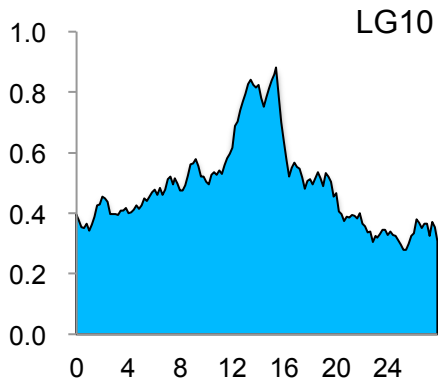


Supplementary Figure 6. An overview of the azuki bean genome. Continued.

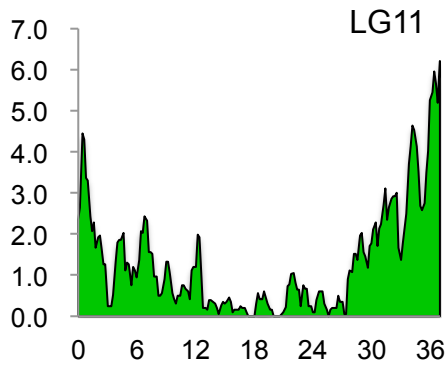
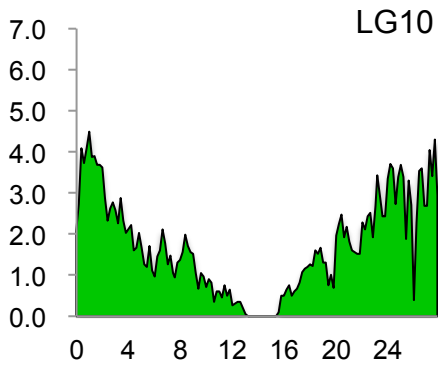
## Gene Density (Genes/Mb)



## Repeat Rate/Mb



## Recombination Frequency (cM/Mb)



## Recombination Frequency (cM/Gene)

