

SUPPLEMENTAL INFORMATION

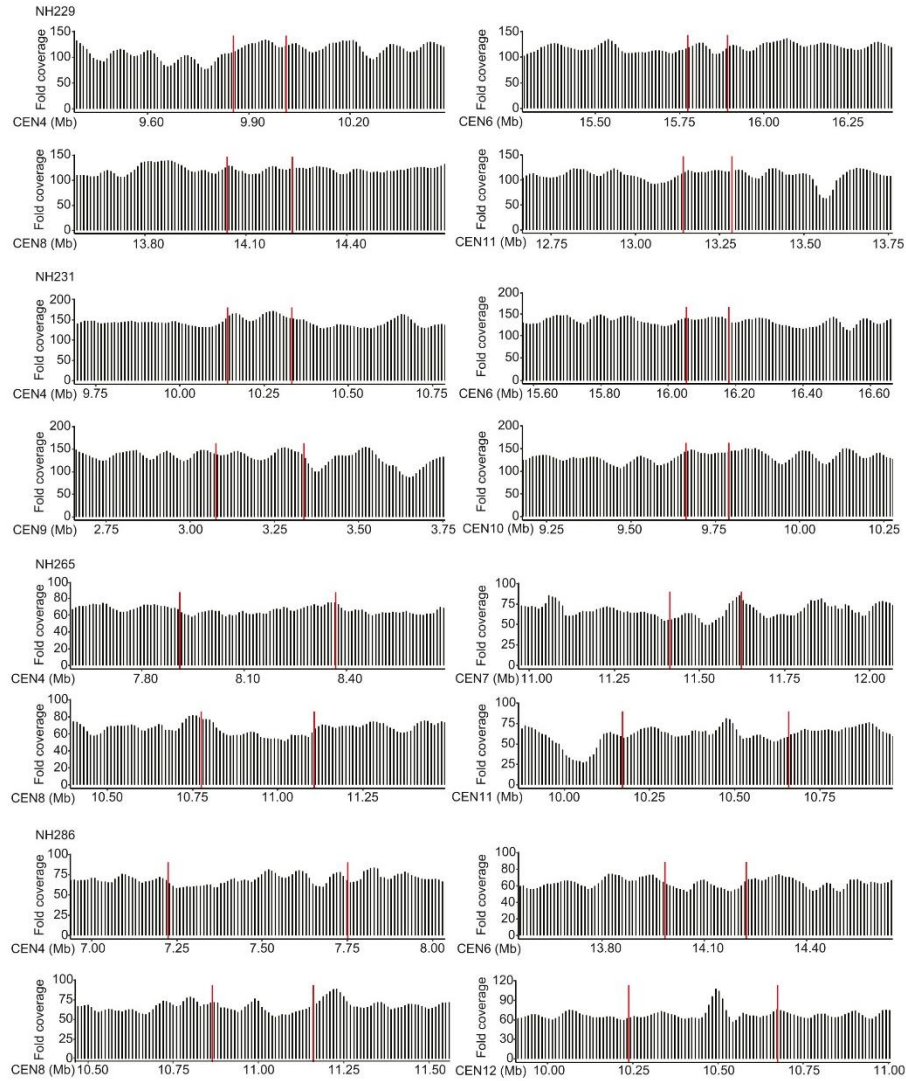


Figure S1. Assembly quality assessment using Nanopore long-read coverage depth for complete CoERs. The average coverage depth of Nanopore long reads from four assemblies, i.e., NH229, NH231, NH265, and NH286. The x-axis represents the physical position on the chromosomes, and the y-axis represents average of the ONT long-read coverage depth in a 50 kb sliding window with a 10 kb step size. The red lines indicate the boundaries of the complete CoERs.

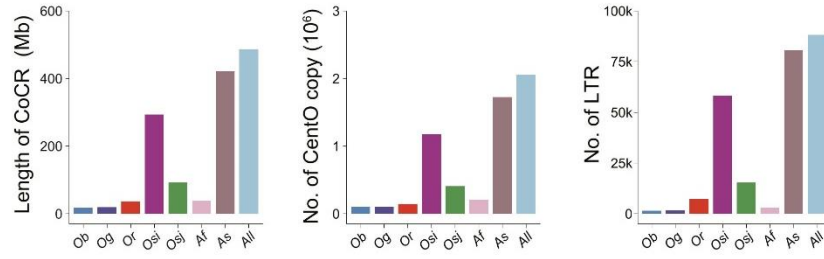


Figure S2. Comparison of the total CoERs length, CentO repeats copy, and LTR numbers across different subpopulations in the centromere map. *Ob*, *Og*, *Or*, *Osi*, *Osj*, *As*, *Af* and *All* refer to *O. barthii*, *O. glaberrima*, *O. rufipogon*, *O. sativa indica*, *O. sativa japonica*, African rice, Asian rice, and the 251 accessions, respectively.

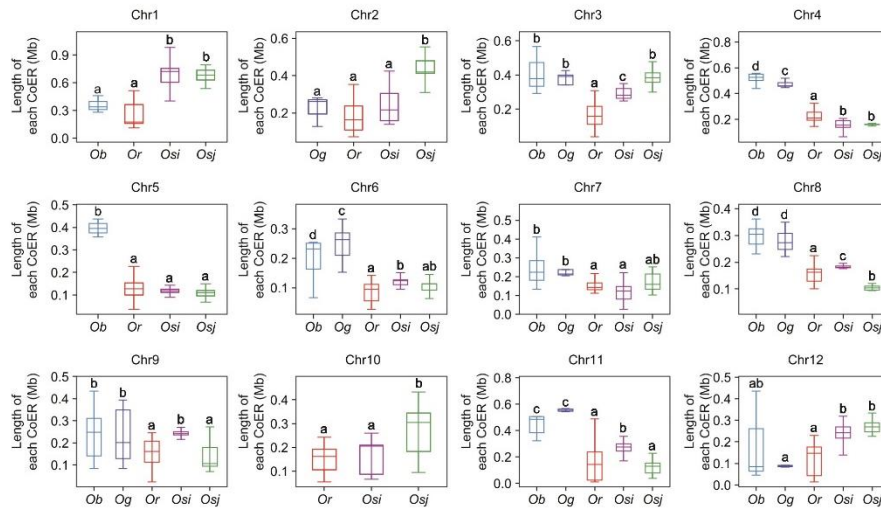


Figure S3. CoERs lengths across different subpopulations on different chromosomes in the centromere map. In the box plot, significance is represented in alphabetical notation ($P < 0.05$ by Tukey, the CoER samples on each subpopulation less than 3 were filtered). *Ob*, *Og*, *Or*, *Osi*, and *Osj* refer to *O. barthii*, *O. glaberrima*, *O. rufipogon*, *O. sativa indica*, and *O. sativa japonica*, respectively.

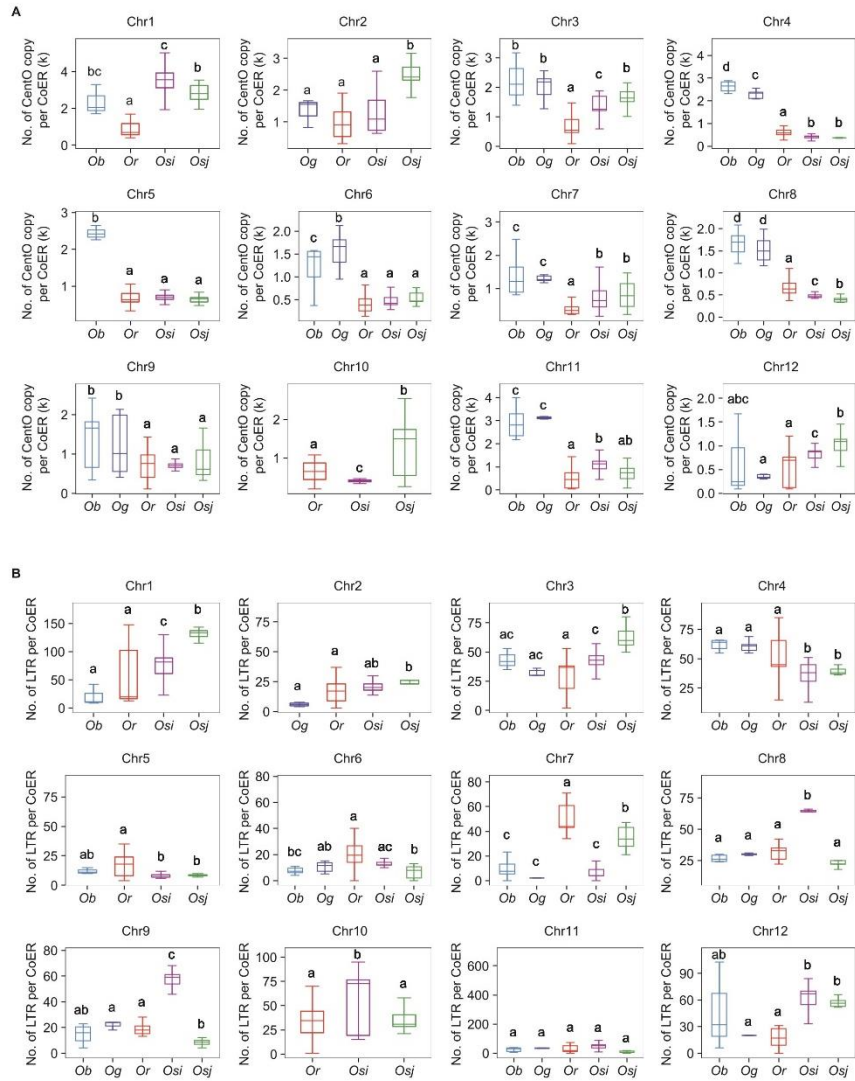


Figure S4. Number of CentO copies (a), and LTRs (b) across different subpopulations on different chromosomes in the centromere map. In the box plot, Tukey's multiple comparisons test of the number of CoER element with CentO satellite (A) and long terminal repeats (LTRs) (B) across different subpopulations. Significance is represented in alphabetical notation ($P < 0.05$ by Tukey, the CoER samples on each subpopulation less than 3 were filtered). Ob, Og, Or, Osi, and Osj refer to *O. barthii*, *O. glaberrima*, *O. rufipogon*, *O. sativa indica*, and *O. sativa japonica*, respectively.

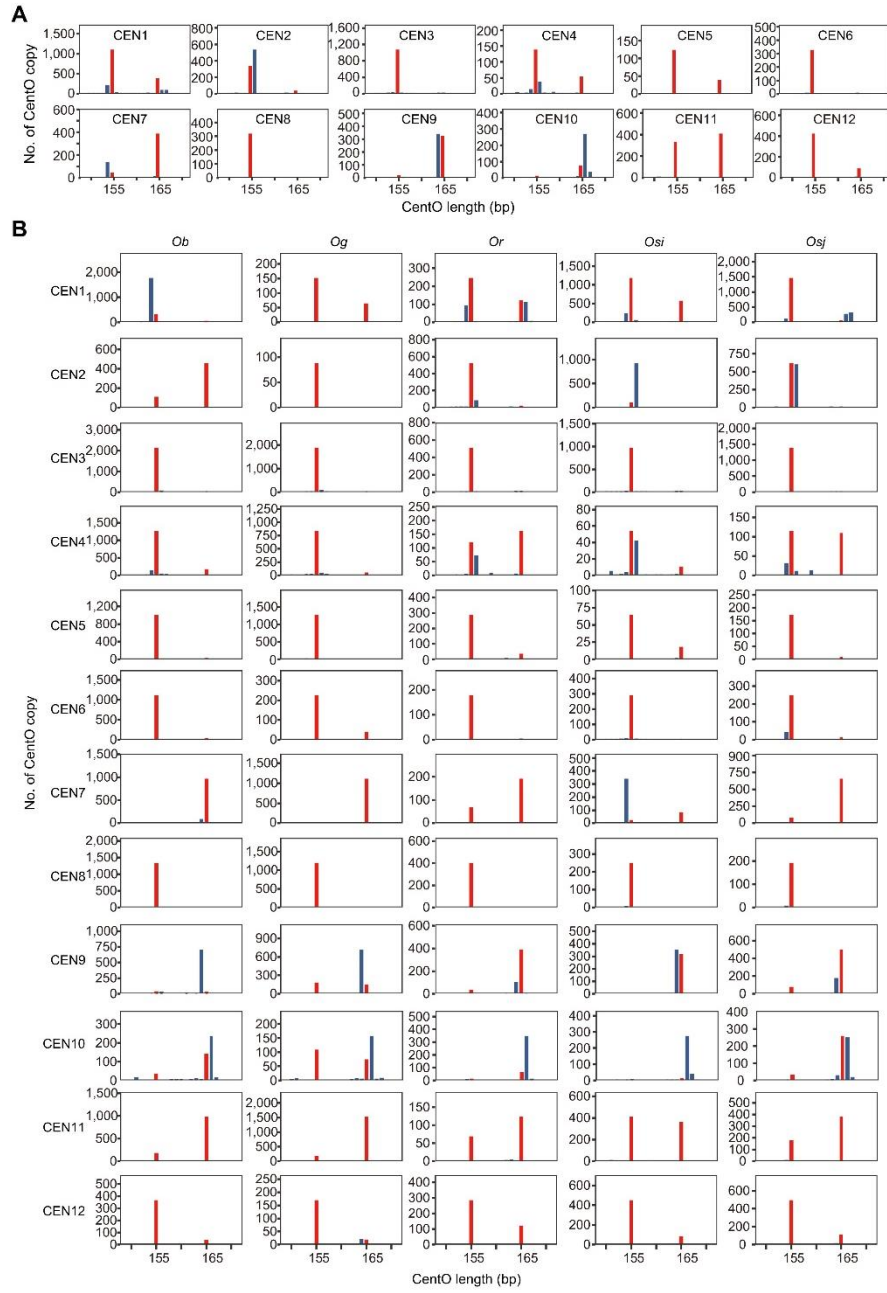


Figure S5. The distribution and diversity of different CentO monomer lengths on different chromosomes (A), and different subpopulations (B) in the centromere map, with red lines highlighting the 155 bp and 165 bp CentO repeats. *Ob*, *Og*, *Or*, *Osi*, and *Osj* refer to *O. barthii*, *O. glaberrima*, *O. rufipogon*, *O. sativa indica*, and *O. sativa japonica*, respectively.

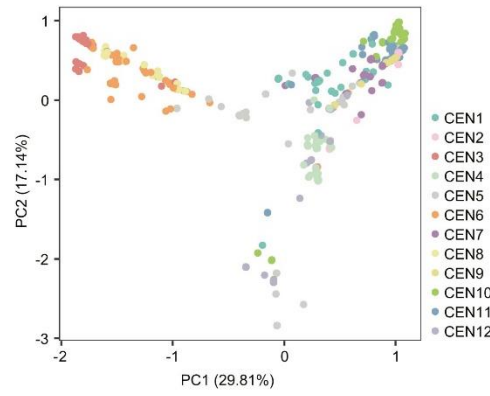


Figure S6. PCA of CentO consensus sequence from each CoERs. The color-coded representation indicates different CoERs.

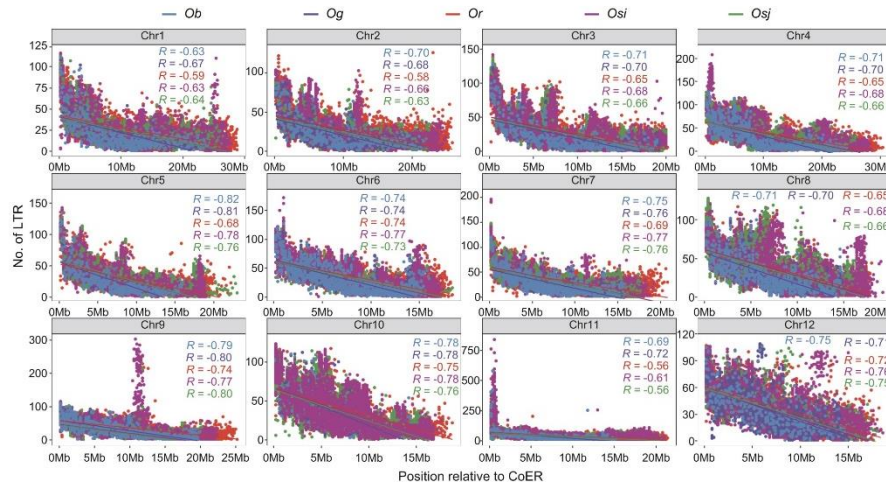


Figure S7. The linear regression of the number of LTRs (y-axis) and the distance from the CoERs (x-axis) across different subpopulations on each chromosome. The linear regression curves show correlation between the number of LTRs (y-axis) and the distance from the CoER (x-axis) (R values are shown in the plot, P values are less than $2.2e-16$ by spearman correlation analysis). Different colors representing different subpopulations. *Ob*, *Og*, *Or*, *Osi*, and *Osj* refer to *O. barthii*, *O. glaberrima*, *O. rufipogon*, *O. sativa indica*, and *O. sativa japonica*, respectively.

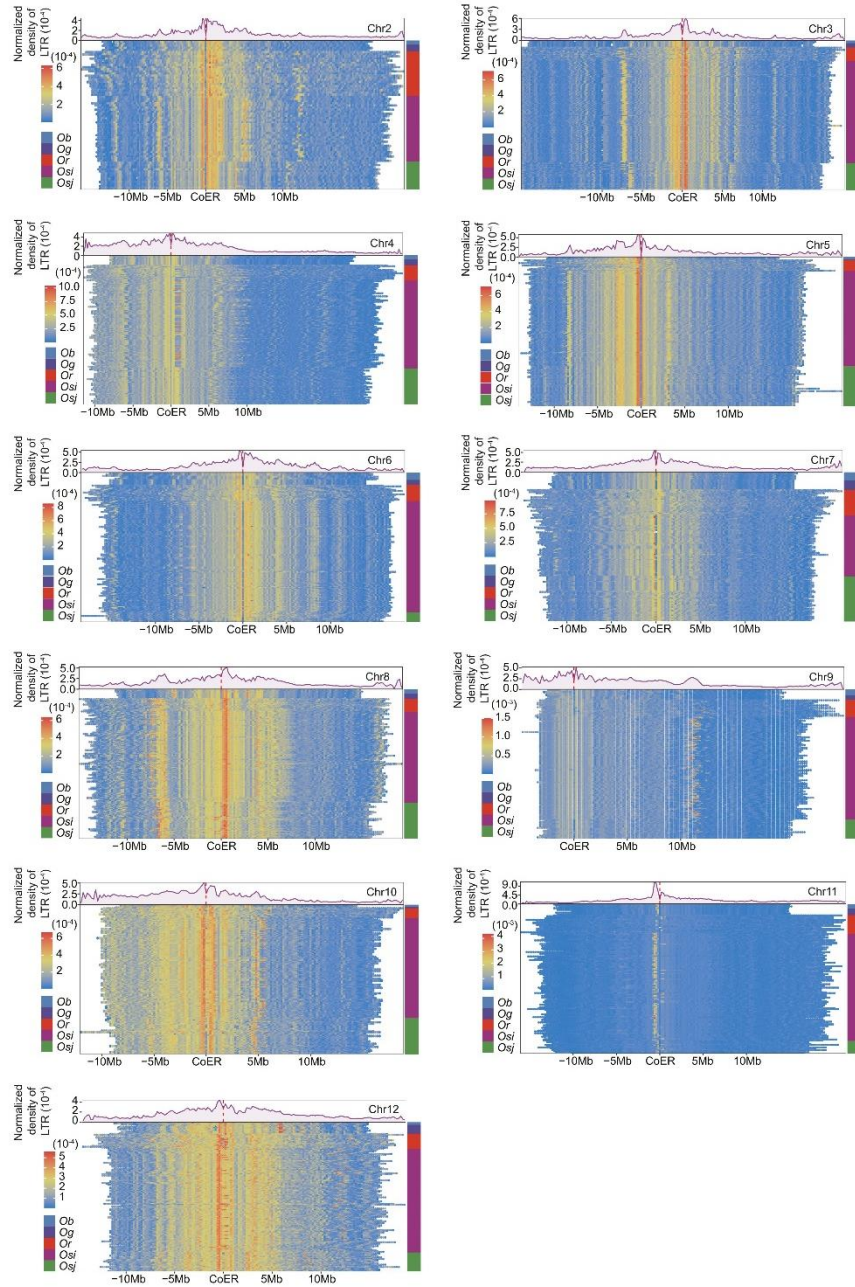


Figure S8. The genome-wide distribution of LTRs from complete CoER samples of different chromosomes. Upper panel, density map of LTR numbers in CoERs start sites and genomic regions on each chromosome. Lower panel, heatmap of LTR numbers distribution in each CoERs start sites and genomic regions in each chromosome. The dashed lines indicate the CoERs. The window size is 200 kb. The heatmap shows the LTR density between different genome regions of different chromosomes, ranging from blue (lowest) to red (highest). Accessions in different subpopulations are indicated by different colors.

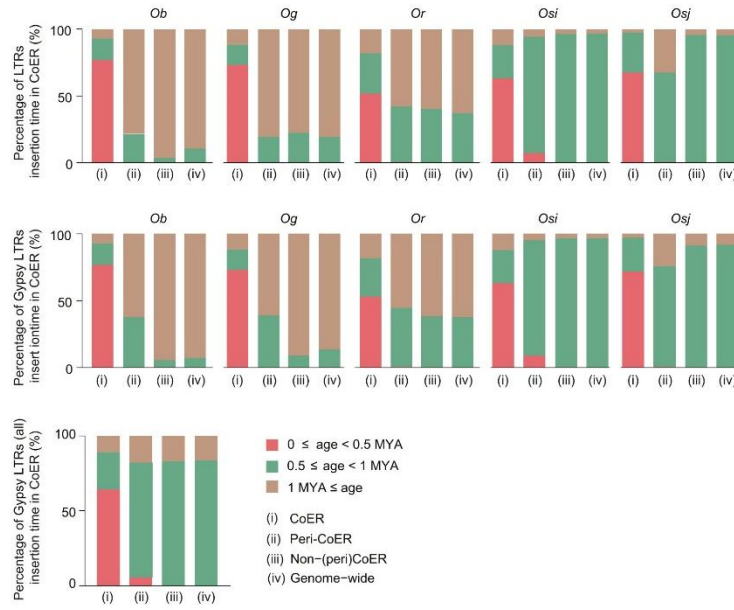


Figure S9. The ratio of LTRs and Gypsy LTRs insertion times in different chromosomal regions across different subpopulations from the complete CoERs dataset. The regions are defined as the core centromere region (CoER) (i), the 1 Mb flanking region of the CoER on either side (peri-CoER) (ii), the rest of the genome (non-(peri) CoER) (iii), and genome-wide (iv). *Ob*, *Og*, *Or*, *Osi*, and *Osj* refer to *O. barthii*, *O. glaberrima*, *O. rufipogon*, *O. sativa indica*, and *O. sativa japonica*, respectively.

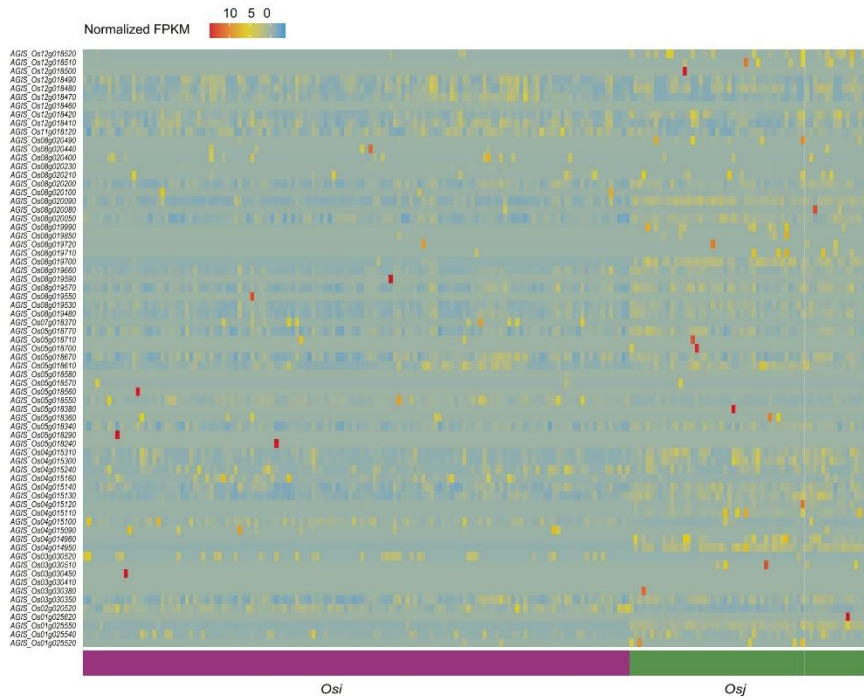


Figure S10. Expression level of the centromeric genes in *Osi* and *Osj* accessions.
The FPKM value is represented by different colors, with red indicating high values.

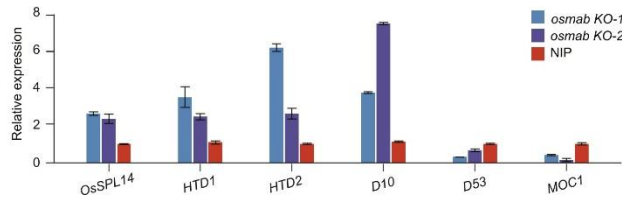


Figure S11. Expression of tiller-related genes by RT-qPCR in *KO-1*, *KO-2*, and WT.
Error bars represent mean $\pm$ SE.

Table S1. Summary of quality assessments from 19 African rice assemblies.

Table S2. Location and evaluation of genome sequence quality for centromere spanning regions across 251-rice genomes.

Table S3. Number of complete CoERs on each chromosome.

Table S4. Summary statistics of the rice complete CoERs dataset in the centromere map.

Table S5. Significant SV eQTL signaling sites were located in centromeric region.

Table S6. Haplotype analysis of the gene expression level and agronomic traits data for significant eQTL signals.

Table S7. TEs sequence annotation of the 19 accessions.

Table S8. Evaluation of strategies for CoERs definition.

Table S9. List of primers for analysis.