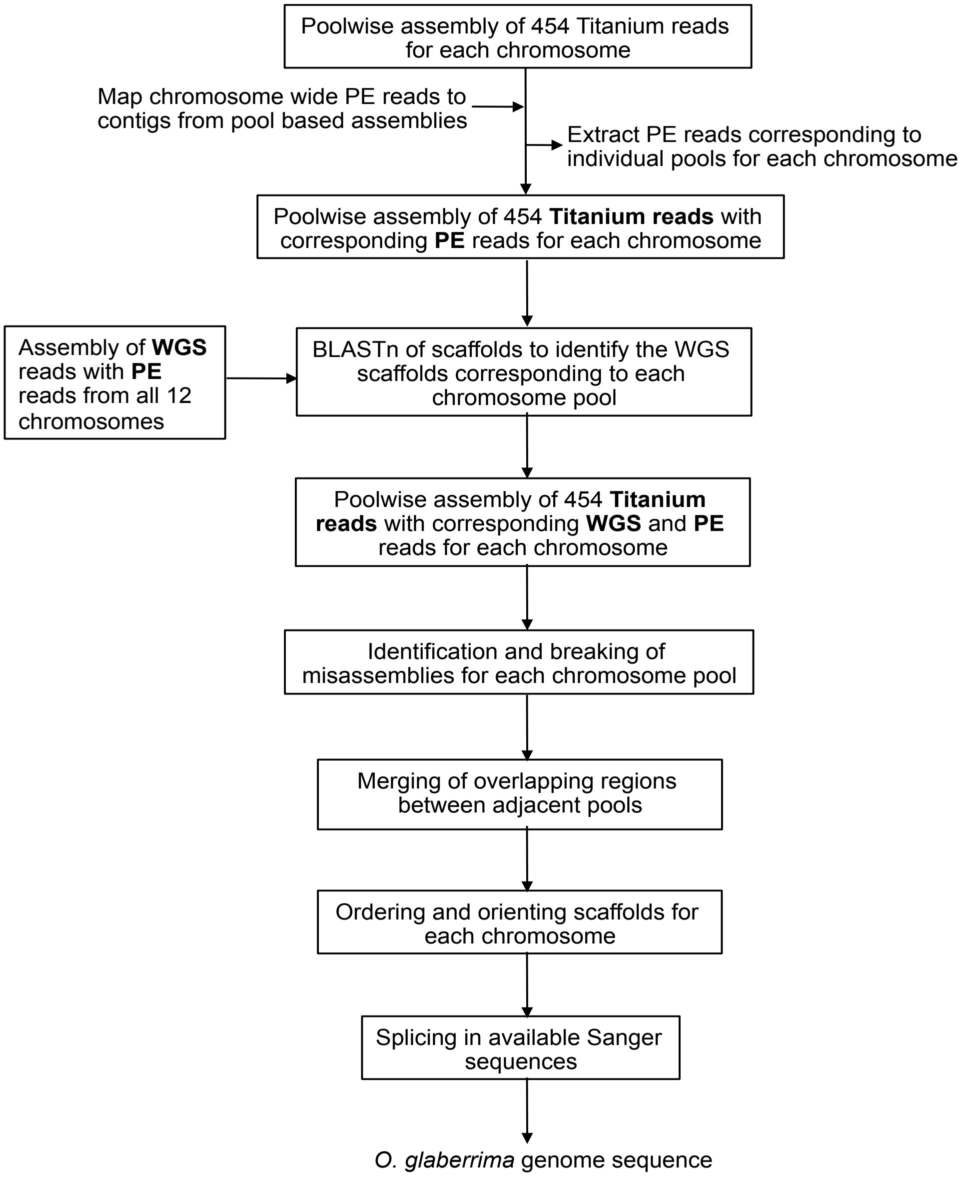
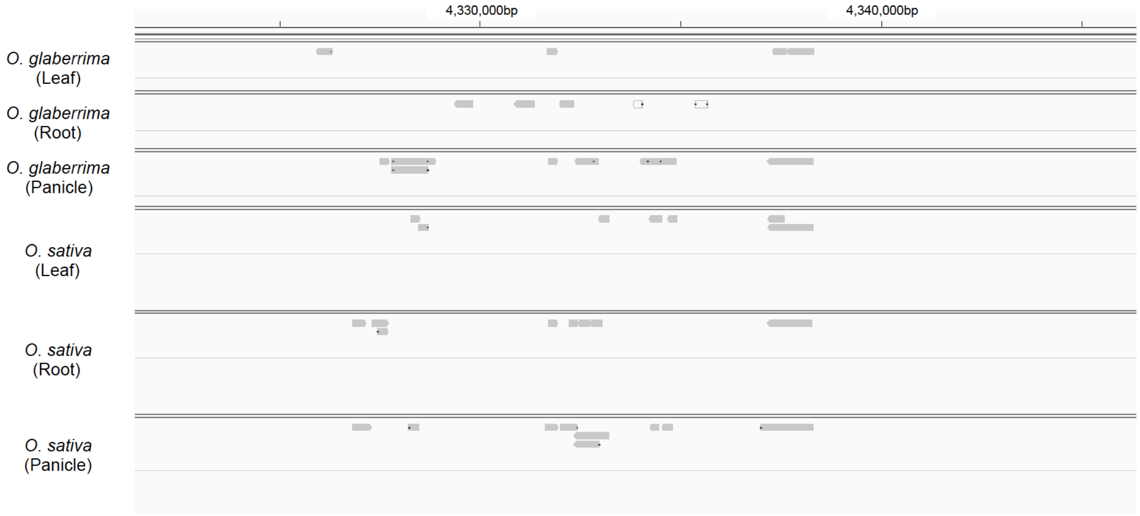
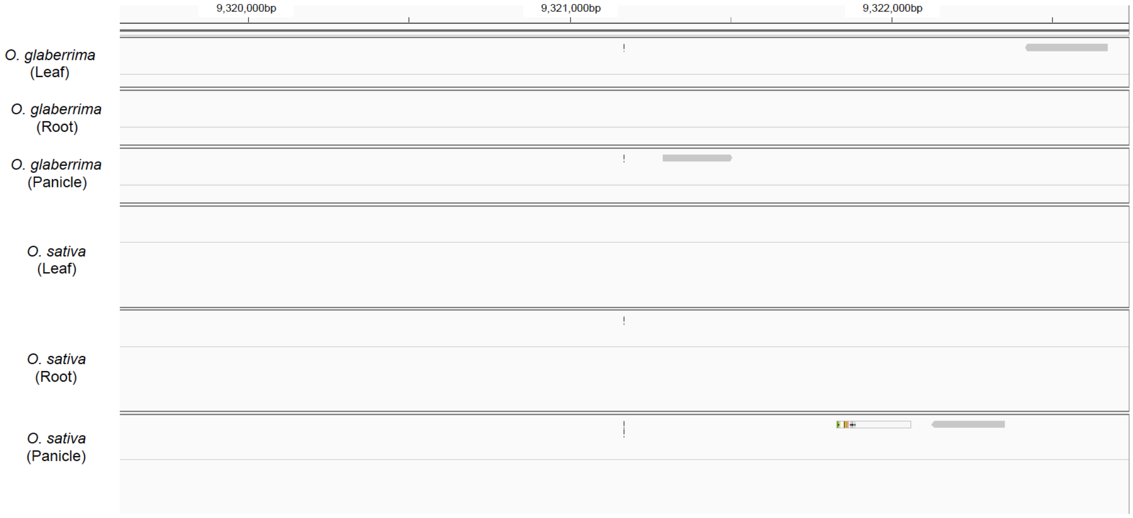


Supplementary Figure 1 *O. glaberrima* genome assembly flowchart.



Supplementary Figure 2A Expression pattern of the *GI* gene in three tissues (leaf, root, mixed stage panicle) of *O. glaberrima* and *O. sativa*. *GI* was expressed in *O. glaberrima* and *O. sativa*. The raw RNA sequencing (RNA-Seq) reads from both species were assembled independently and then mapped to the *O. sativa* ssp. *japonica* reference sequence. Results of these alignments were visualized using integrated genomics viewer (IGV). Each gray bar represents one assembled RNA-Seq fragment.

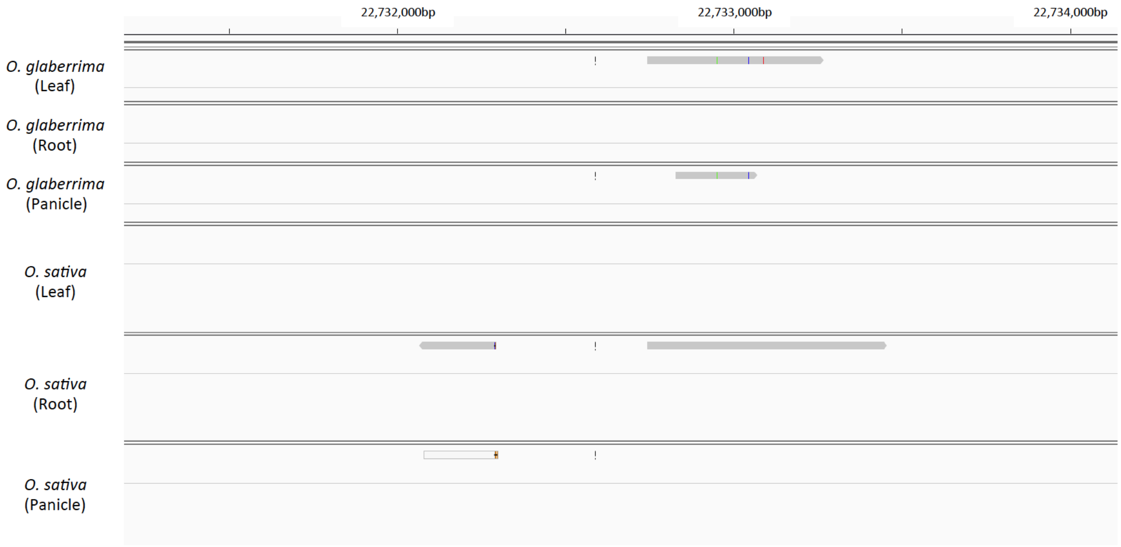


Supplementary Figure 2B Expression pattern of *ARP6* gene in three tissues (leaf, root, mixed stage panicle) in *O. glaberrima* and *O. sativa*. *ARP6* was expressed in *O. glaberrima* and *O. sativa*. The raw RNA sequencing (RNA-Seq) reads from both species were assembled independently and then mapped to the *O. sativa* ssp. *japonica* reference sequence. Results of these alignments were visualized using integrated genomics viewer (IGV). Each gray bar represents one assembled RNA-Seq fragment.

Supplementary Figure 2C Expression pattern of *RFT1* gene in three tissues (leaf, root, mixed stage panicle) in *O. glaberrima* and *O. sativa*. *RFT1* was expressed in *O. glaberrima* and *O. sativa*. The raw RNA sequencing (RNA-Seq) reads from both species were assembled independently and then mapped to the *O. sativa* ssp. *japonica* reference sequence. Results of these alignments were visualized using integrated genomics viewer (IGV). Each gray bar represents one assembled RNA-Seq fragment.

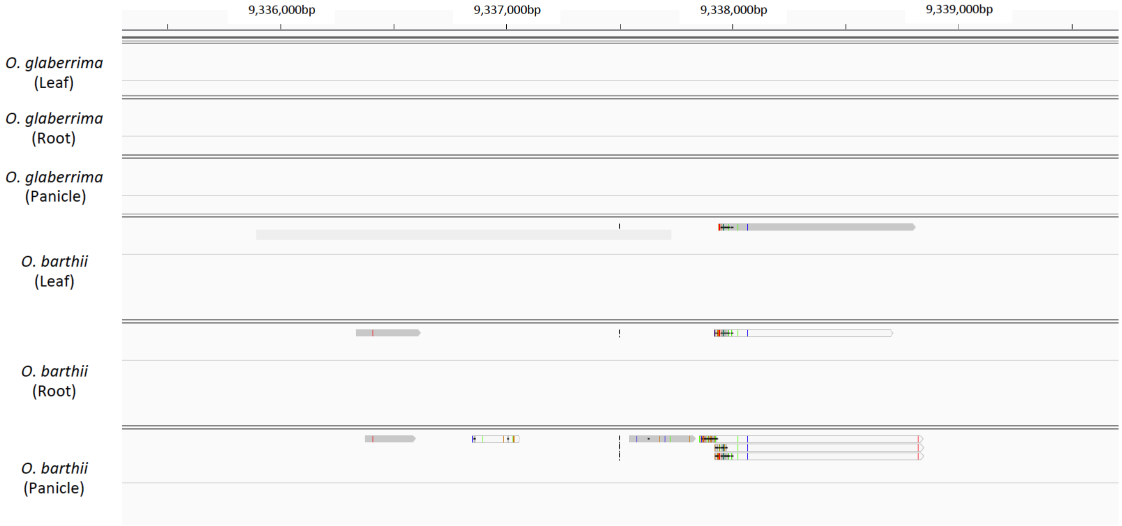
Supplementary Figure 2D Expression pattern of *SPL3* gene in three tissues (leaf, root, mixed stage panicle) in *O. glaberrima* and *O. sativa*. *SPL3* was expressed in *O. glaberrima* and *O. sativa*. The raw RNA sequencing (RNA-Seq) reads from both species were assembled independently and then mapped to the *O. sativa* ssp. *japonica* reference sequence. Results of these alignments were visualized using integrated genomics viewer (IGV). Each gray bar represents one assembled RNA-Seq fragment.

Supplementary Figure 2E Expression pattern of *PIF3* gene in three tissues (leaf, root, mixed stage panicle) in *O. glaberrima* and *O. sativa*. *PIF3* was expressed in *O. glaberrima* and *O. sativa*. The raw RNA sequencing (RNA-Seq) reads from both species were assembled independently and then mapped to the *O. sativa* ssp. *japonica* reference sequence. Results of these alignments were visualized using integrated genomics viewer (IGV). Each gray bar represents one assembled RNA-Seq fragment.

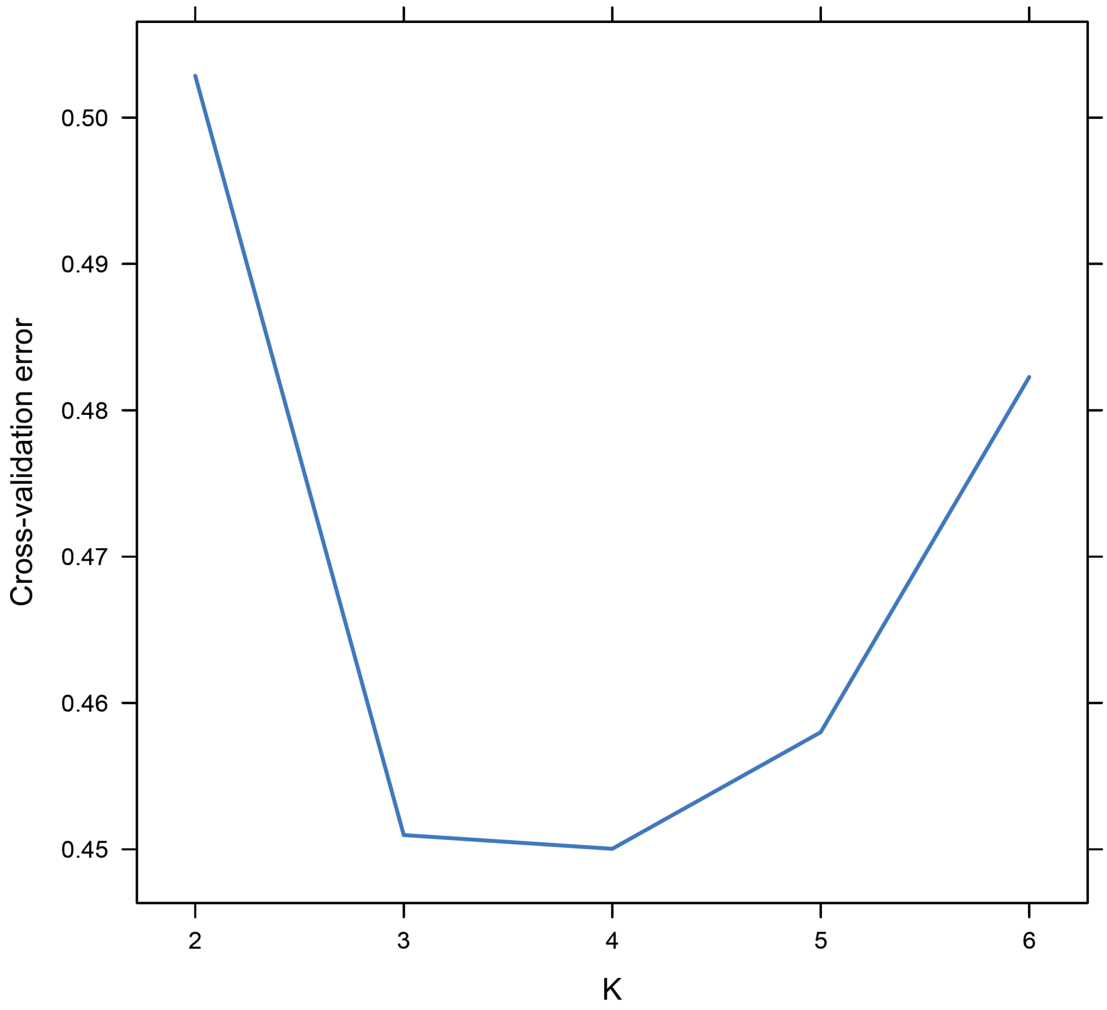


Supplementary Figure 2F Expression pattern of *WRKY77* gene in three tissues (leaf, root, mixed stage panicle) in *O. glaberrima* and *O. sativa*. *WRKY77* was expressed in *O. glaberrima* and *O. sativa*. The raw RNA sequencing (RNA-Seq) reads from both species were assembled independently and then mapped to the *O. sativa* ssp. *japonica* reference sequence. Results of these alignments were visualized using integrated genomics viewer (IGV). Each gray bar represents one assembled RNA-Seq fragment.

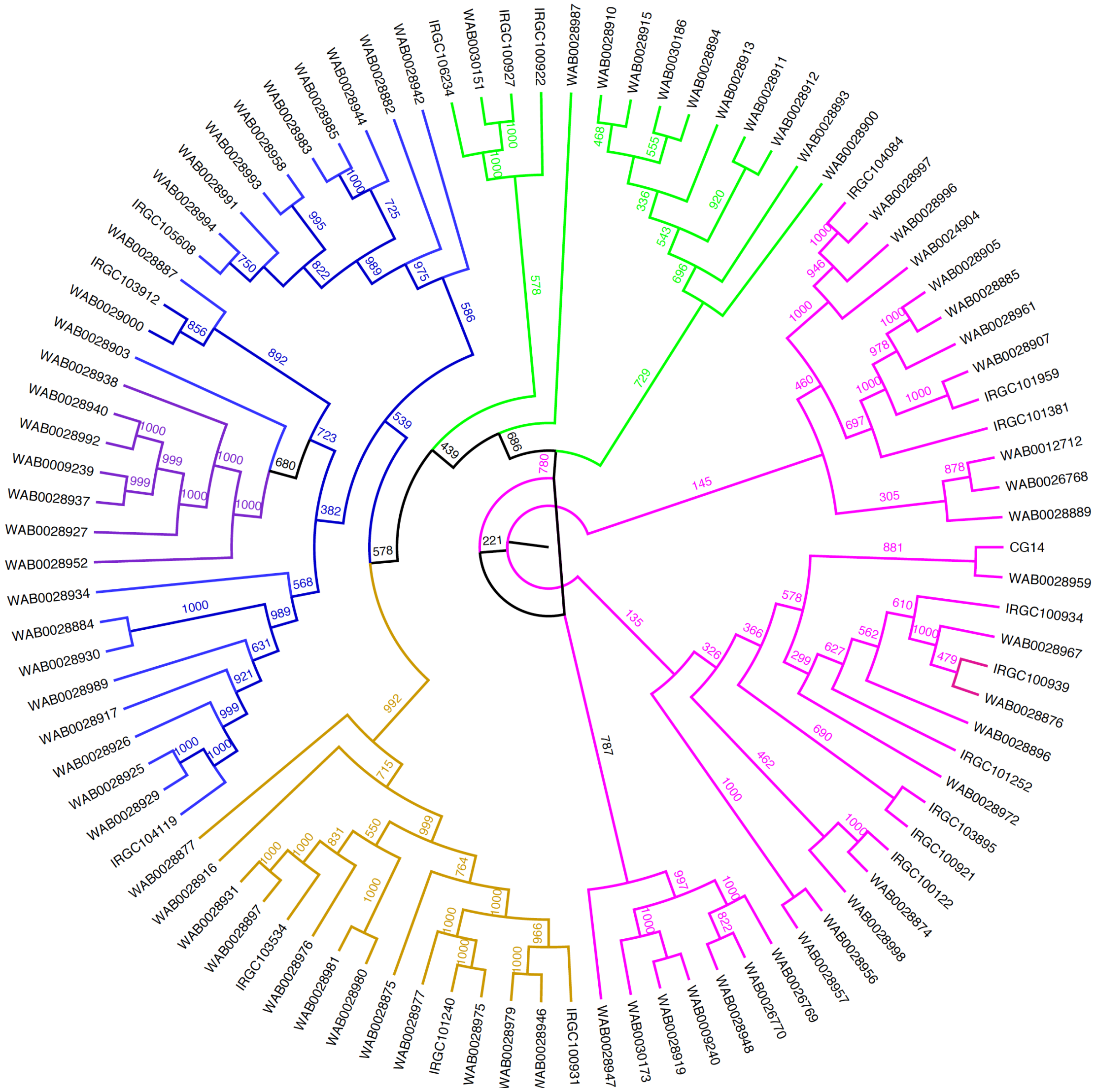
Supplementary Figure 2G Expression pattern of *WRKY* gene in three tissues (leaf, root, mixed stage panicle) in *O. glaberrima* and *O. sativa*. *WRKY* was expressed in *O. glaberrima* and *O. sativa*. The raw RNA sequencing (RNA-Seq) reads from both species were assembled independently and then mapped to the *O. sativa* ssp. *japonica* reference sequence. Results of these alignments were visualized using integrated genomics viewer (IGV). Each gray bar represents one assembled RNA-Seq fragment.



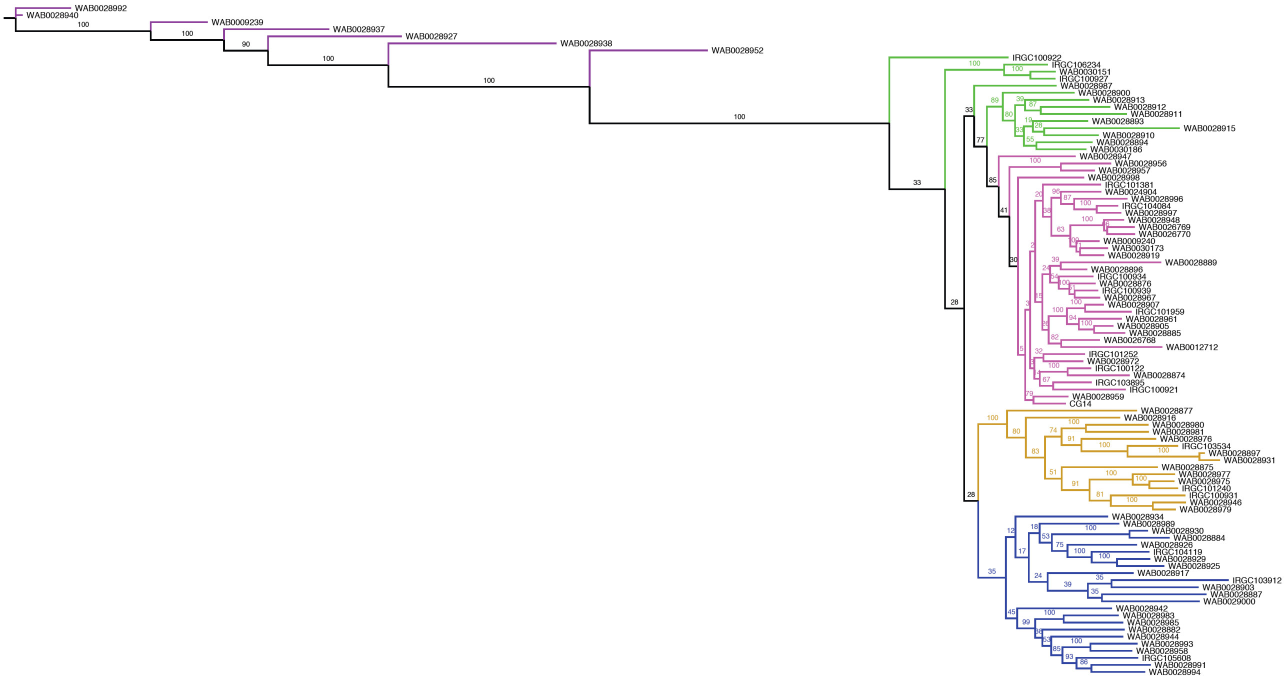
Supplementary Figure 3 Expression pattern of *Hd1* gene in three tissues (leaf, root, mixed stage panicle) in *O. glaberrima* and *O. barthii*. *Hd1* was expressed in *O. barthii* but not in *O. glaberrima*. The raw RNA sequencing (RNA-Seq) reads from both species were assembled independently and then mapped to the *O. sativa* ssp. *japonica* reference sequence. Results of these alignments were visualized using integrated genomics viewer (IGV). Each gray bar represents one assembled RNA-Seq fragment.



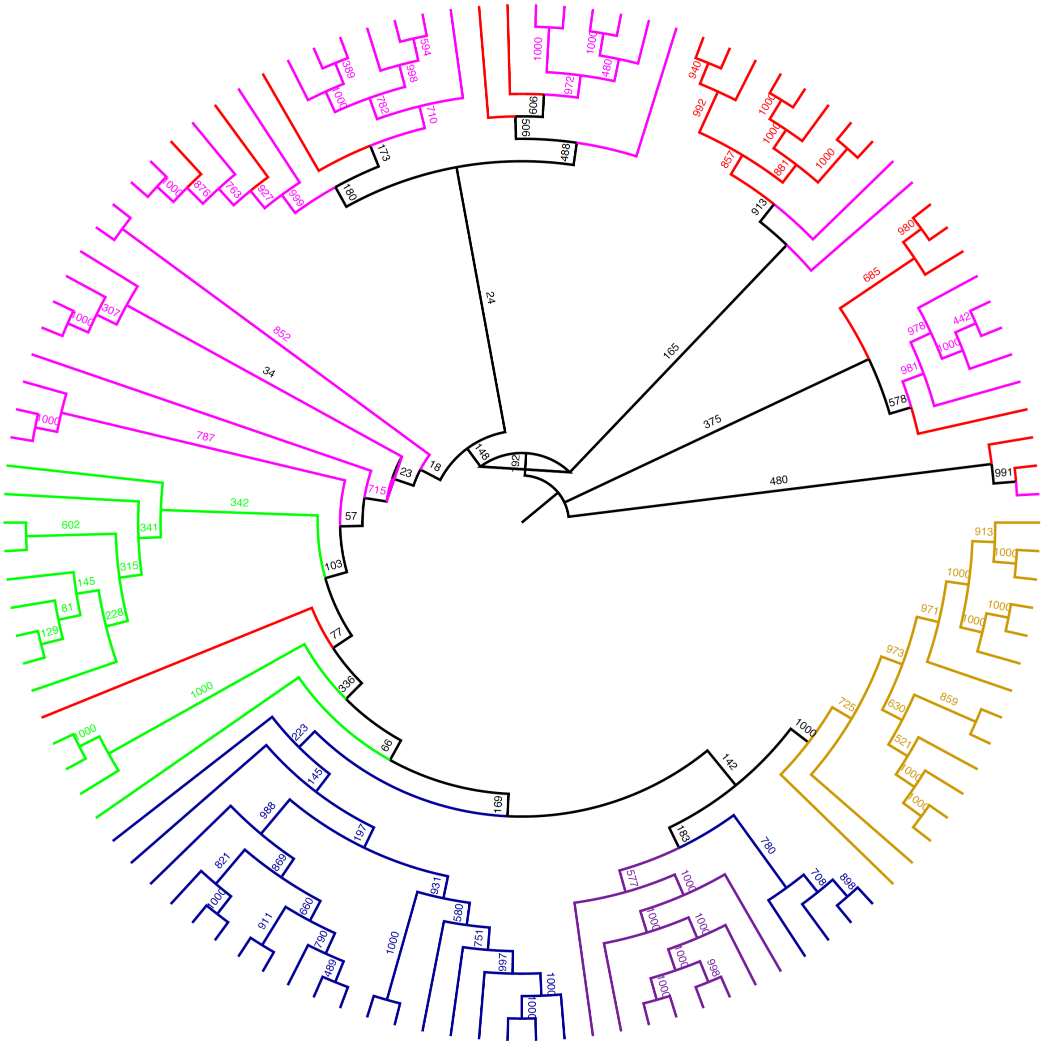
Supplementary Figure 4 Cross-validation error of each K value when estimating the population structure of 94 *O. barthii* accessions. Cross-validation error is smallest when K equals to 4. The 4 subgroups of *O. barthii* population were identified according to the result when K equals to 4.



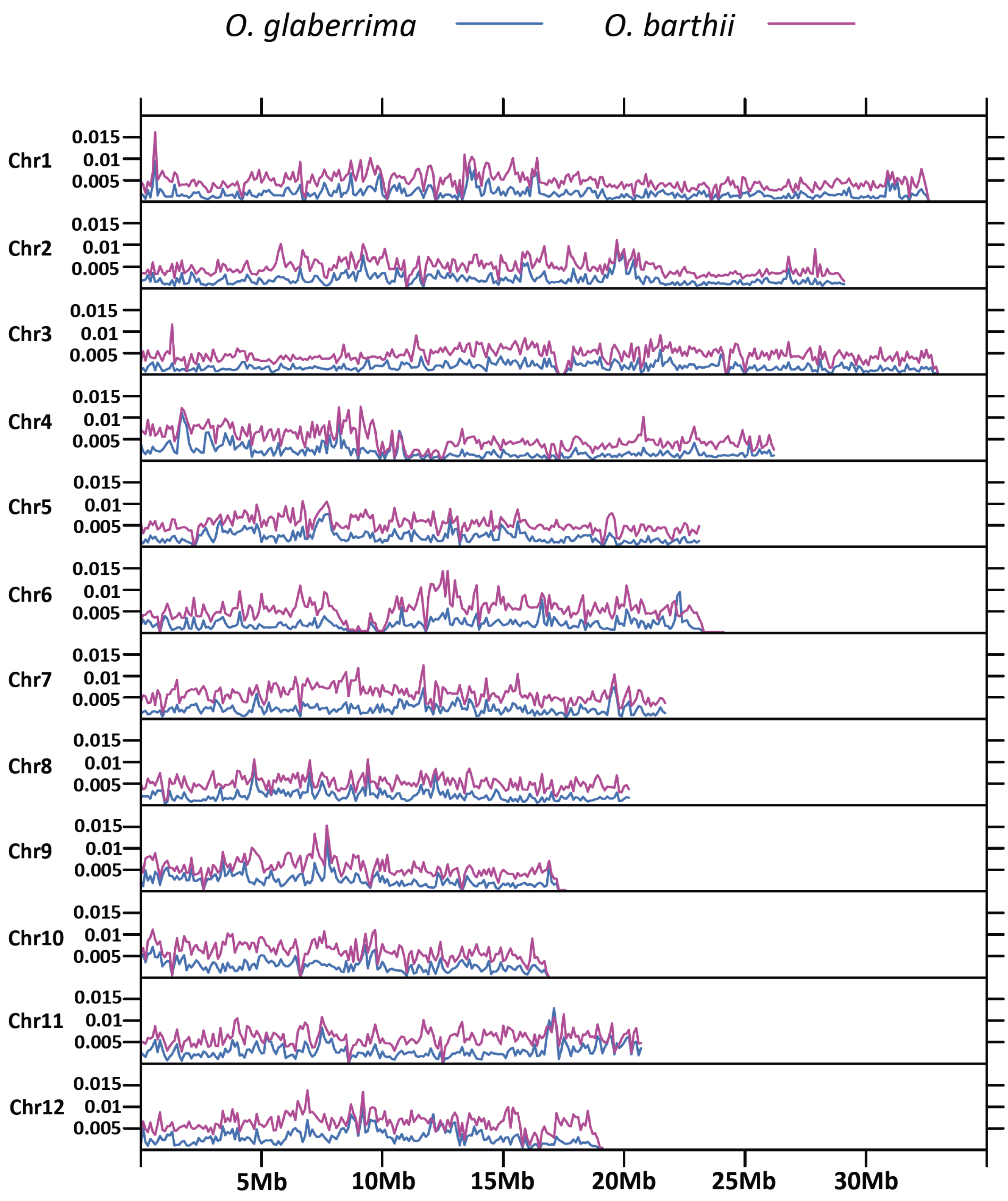
Supplementary Figure 5 Neighbor-joining phylogenetic tree of 94 *O. barthii* accessions. The bootstrap value is labeled on each branch (from 1000 replicates). The *O. barthii* population is partitioned into four subgroups: OB-I (Orange), OB-II (Blue), OB-III (Purple), OB-V (Pink), as well as an admixed group: OB-IV (Green).



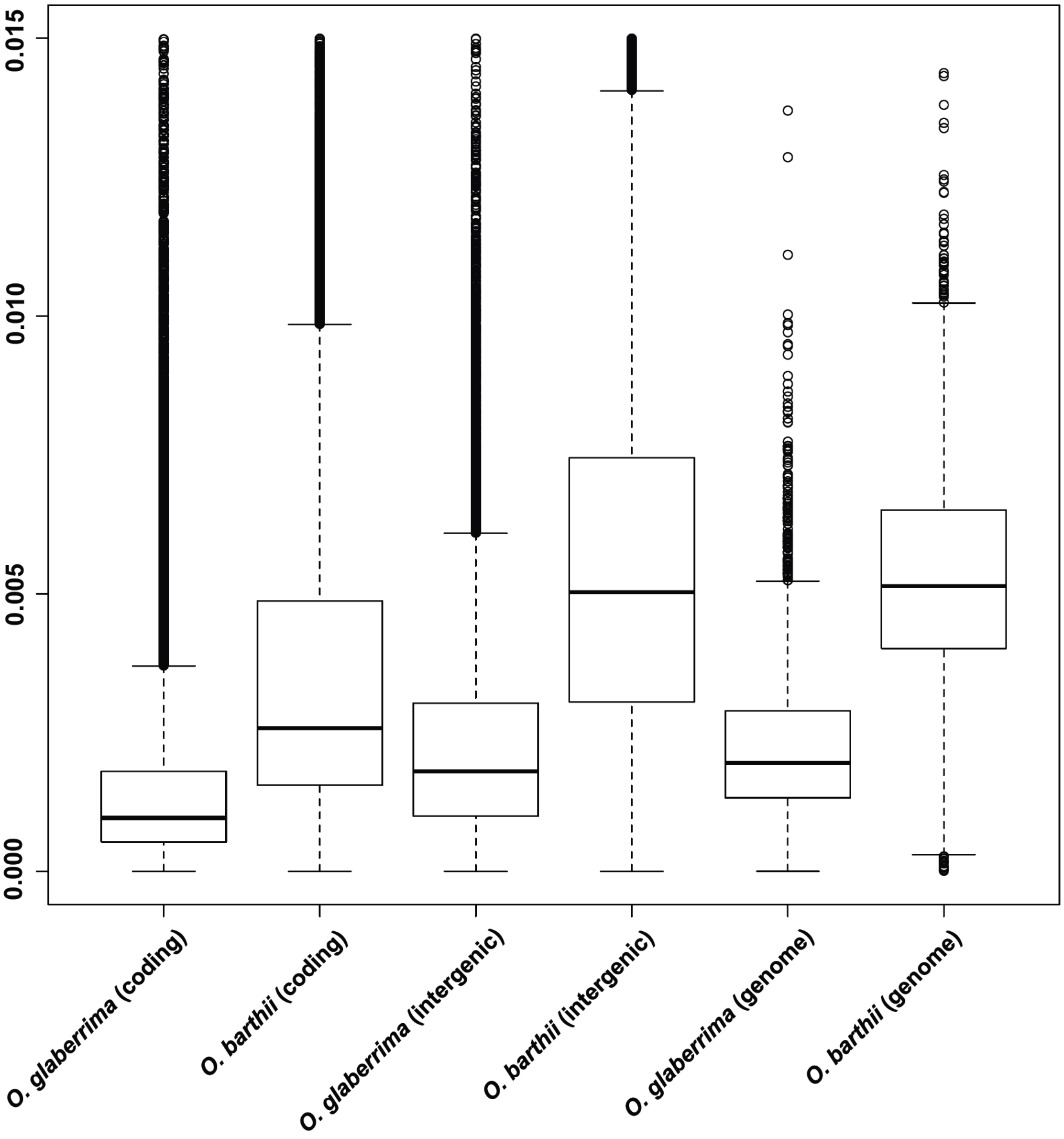
Supplementary Figure 6 Maximum likelihood (ML) tree of 94 *O. barthii* accessions. The bootstrap value is labeled on each branch (from 100 replicates). The *O. barthii* population is partitioned into four subgroups: OB-I (Orange), OB-II (Blue), OB-III (Purple), OB-V (Pink), as well as an admixed group: OB-IV (Green). OB-III is genetically distinct from the rest of the groups.



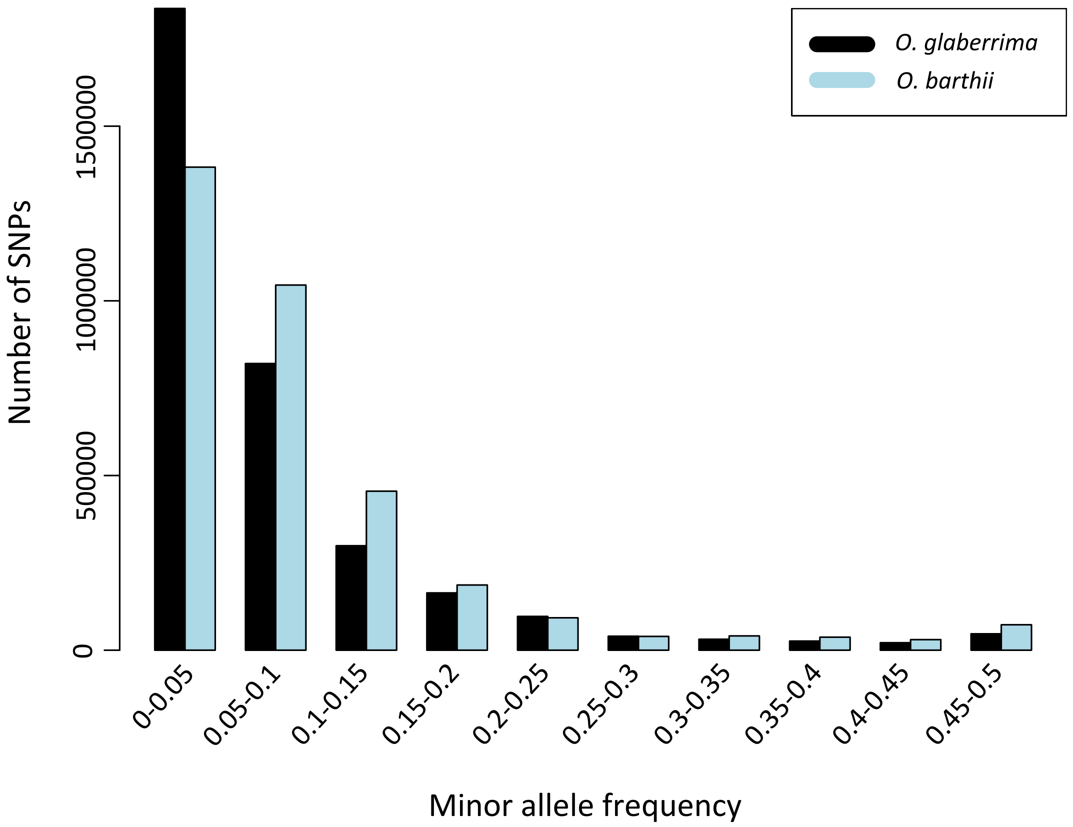
Supplementary Figure 7 Neighbor-joining phylogenetic tree of 20 *O. glaberrima* and 94 *O. barthii* accessions. Bootstrap value (from 1000 replicates) is labeled on each branch. Five *O. barthii* subgroups are labeled as orange (OB-I), blue (OB-II), purple (OB-III), green (OB-IV), and pink (OB-V). *O. glaberrima* accessions are labelled in red.



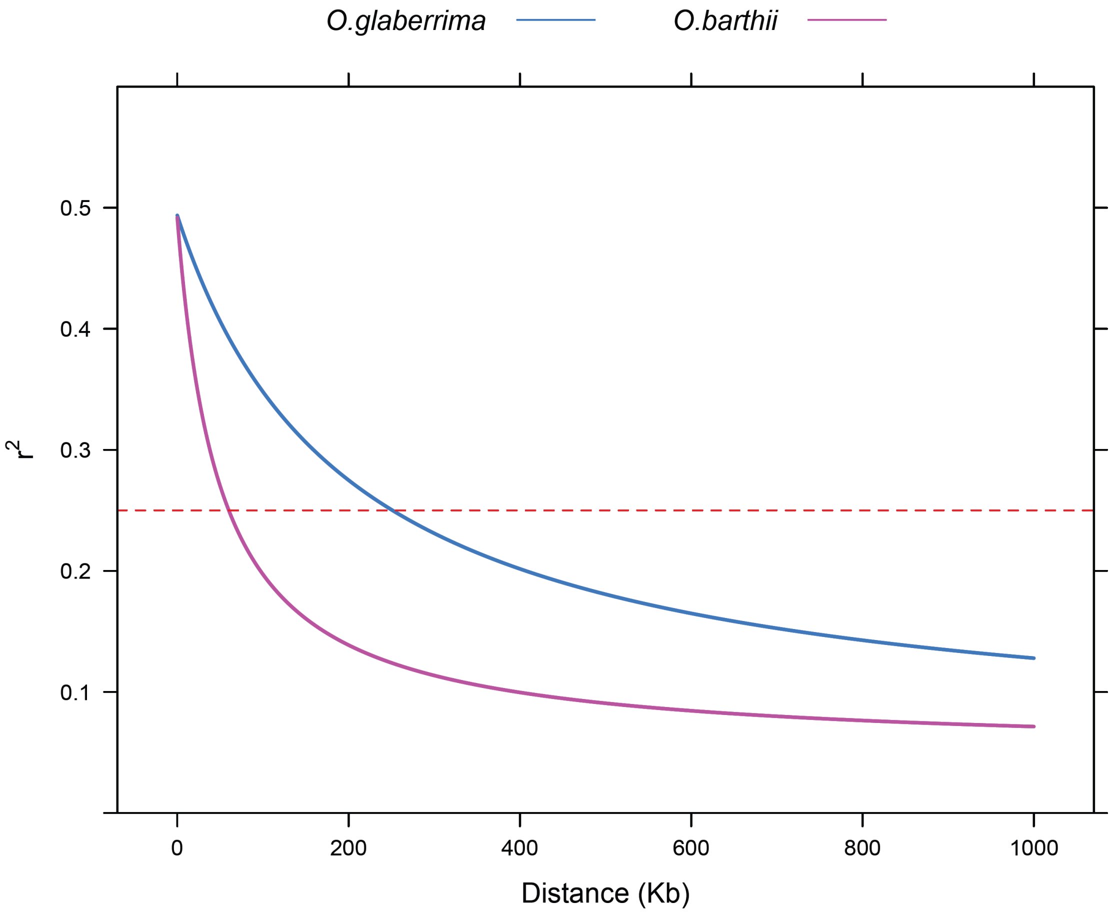
Supplementary Figure 8 Genome-wide distribution of genetic diversity (π) along the 12 chromosomes of *O. glaberrima* and *O. barthii*.



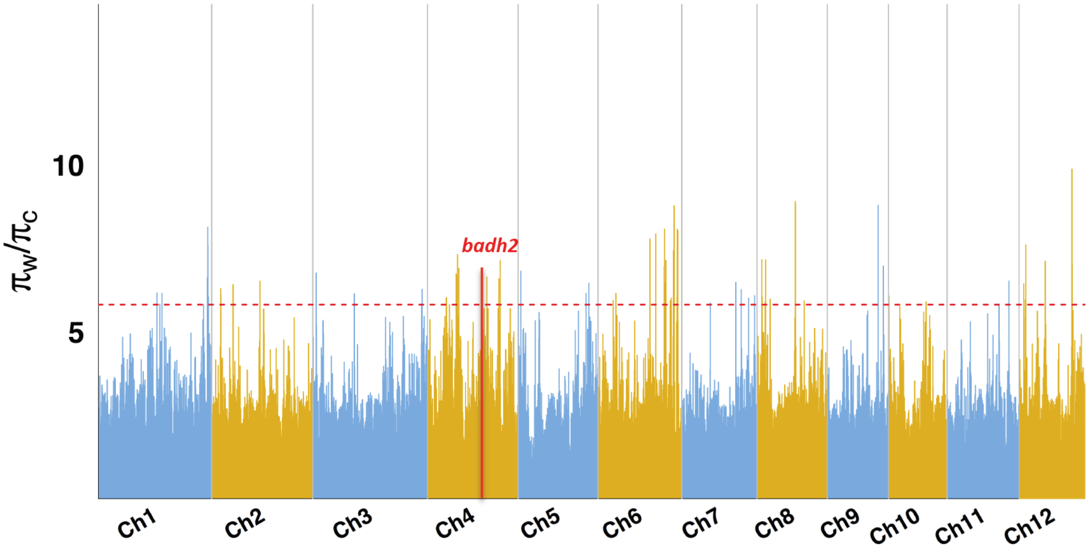
Supplementary Figure 9 Boxplot of coding, intergenic and genome-wide nucleotide diversity (π) between *O. glaberrima* and *O. barthii*.



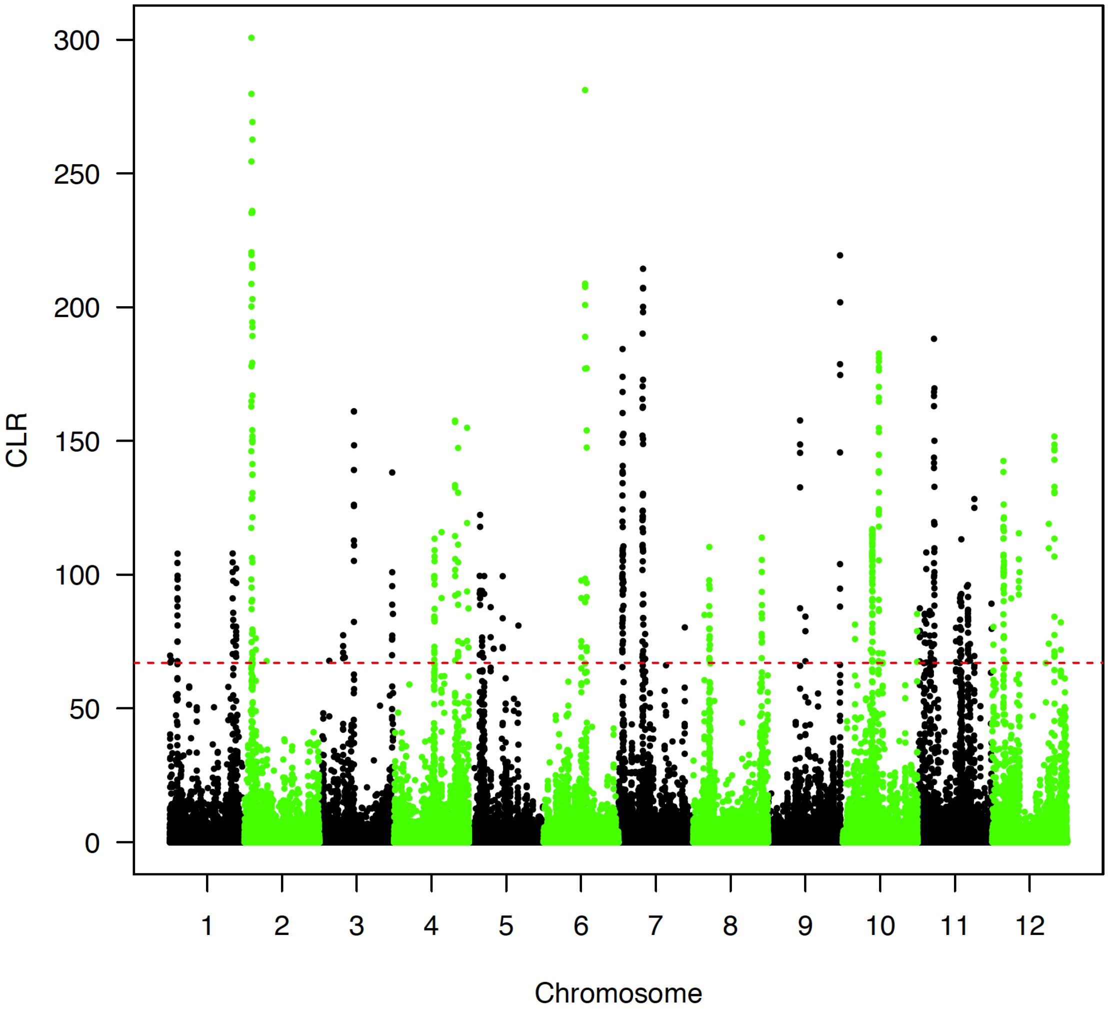
Supplementary Figure 10 Minor allele frequency (MAF) spectrum of *O. glaberrima* and *O. barthii*.



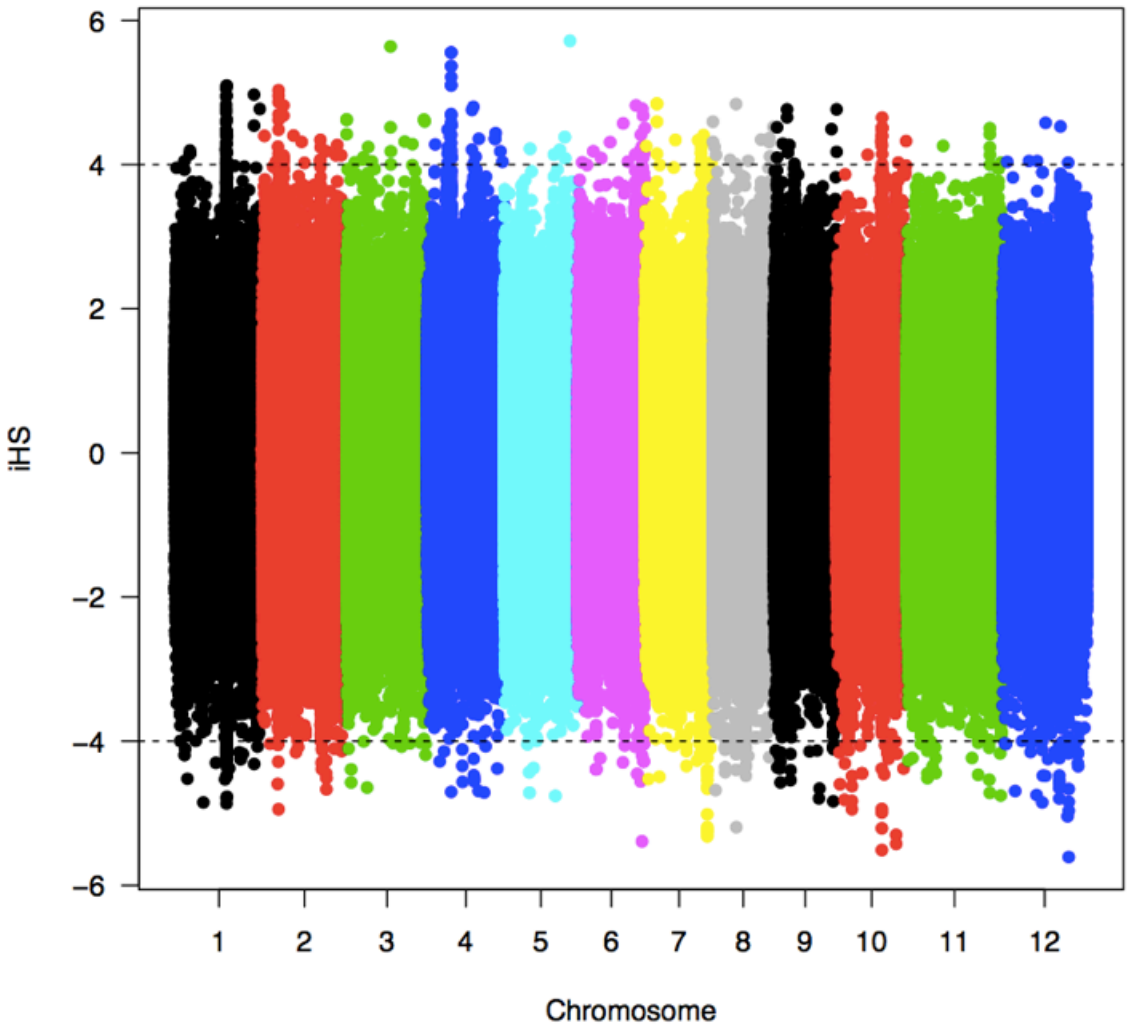
Supplementary Figure 11 Linkage disequilibrium (LD) decay of *O. glaberrima* and *O. barthii*. The red dotted line indicates the half of the largest value of r^2.



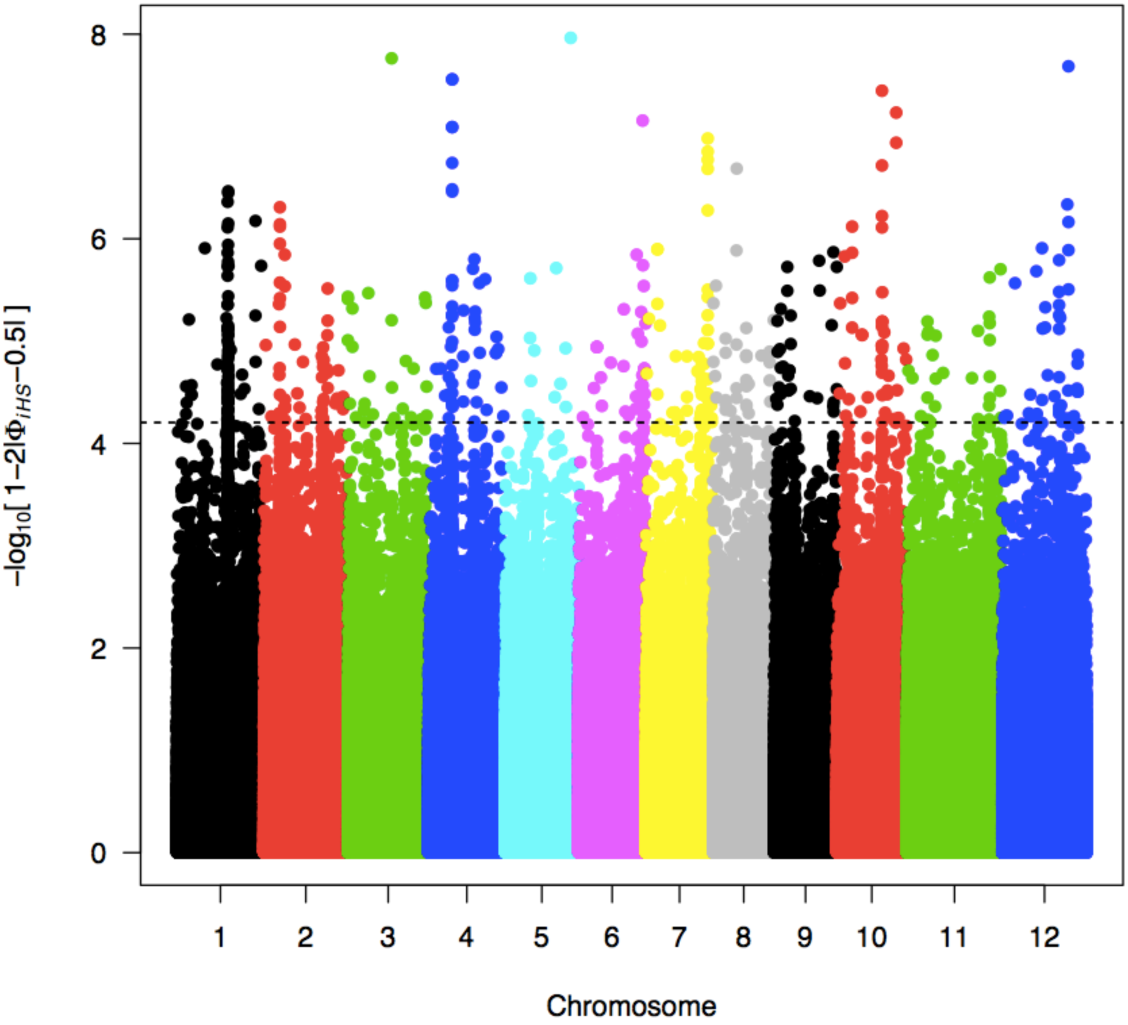
Supplementary Figure 12 Reduction of nucleotide diversity (π) between *O. glaberrima* and *O. barthii* group V (OB-V) using nonoverlapping 100 kb windows. The threshold of 2.5 percentile ($\pi_w/\pi_c > 5.8$) is indicated as a red dotted line. The regions within the 2.5 percentile are considered as candidate regions under artificial selection. One orthologue of the known domestication-related gene (*badh2*), identified in *O. sativa*, was located in a candidate region on chromosome 4.



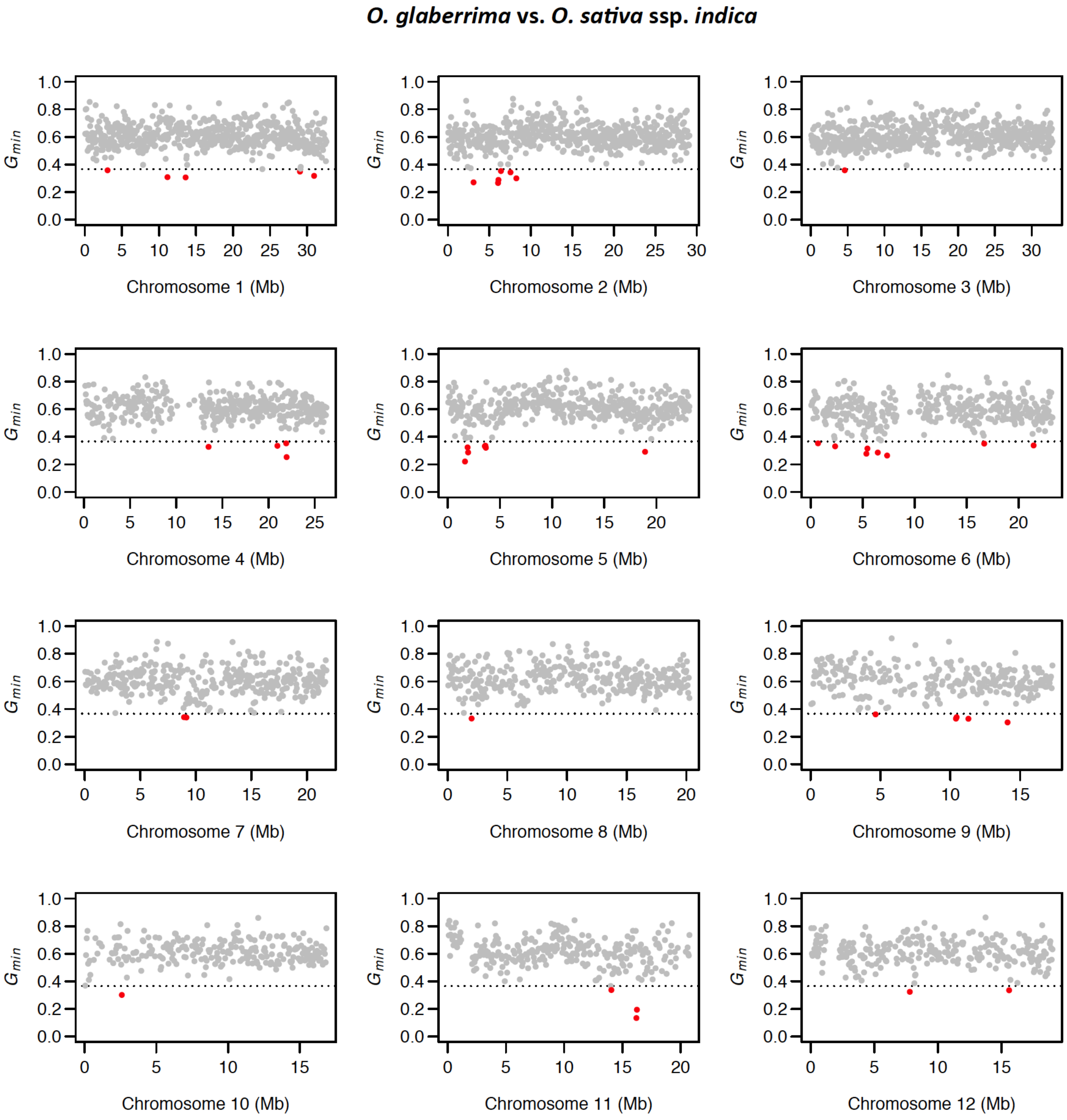
Supplementary Figure 13 Identification of regions that have experienced recent selective sweeps in *O. barthii* subgroup OB-V. Manhattan plot of the composite likelihood ratio (CLR) across the *O. barthii* genome. The red dotted line indicates the cutoff value for the 0.5% outlier regions with significant deviations from neutrality (CLR = 66.32), indicating evidence of recent selective sweeps. Only two of the significant peaks match regions showing evidence of selective sweeps in *O. glaberrima* (Chr 4: 26.1 Mb; Chr 12: 14.4 Mb).



Supplementary Figure 14 Genome-wide distribution of the *iHS* statistic in *O. glaberrima*. Large negative values of *iHS* indicate high haplotype homozygosity due to an increase in frequency of the derived allele, and positive values reflect high frequency of the ancestral allele. The dashed lines represent the cutoff for the 500 SNPs with the top |*iHS*| values (representing 0.028% of SNPs genome-wide).

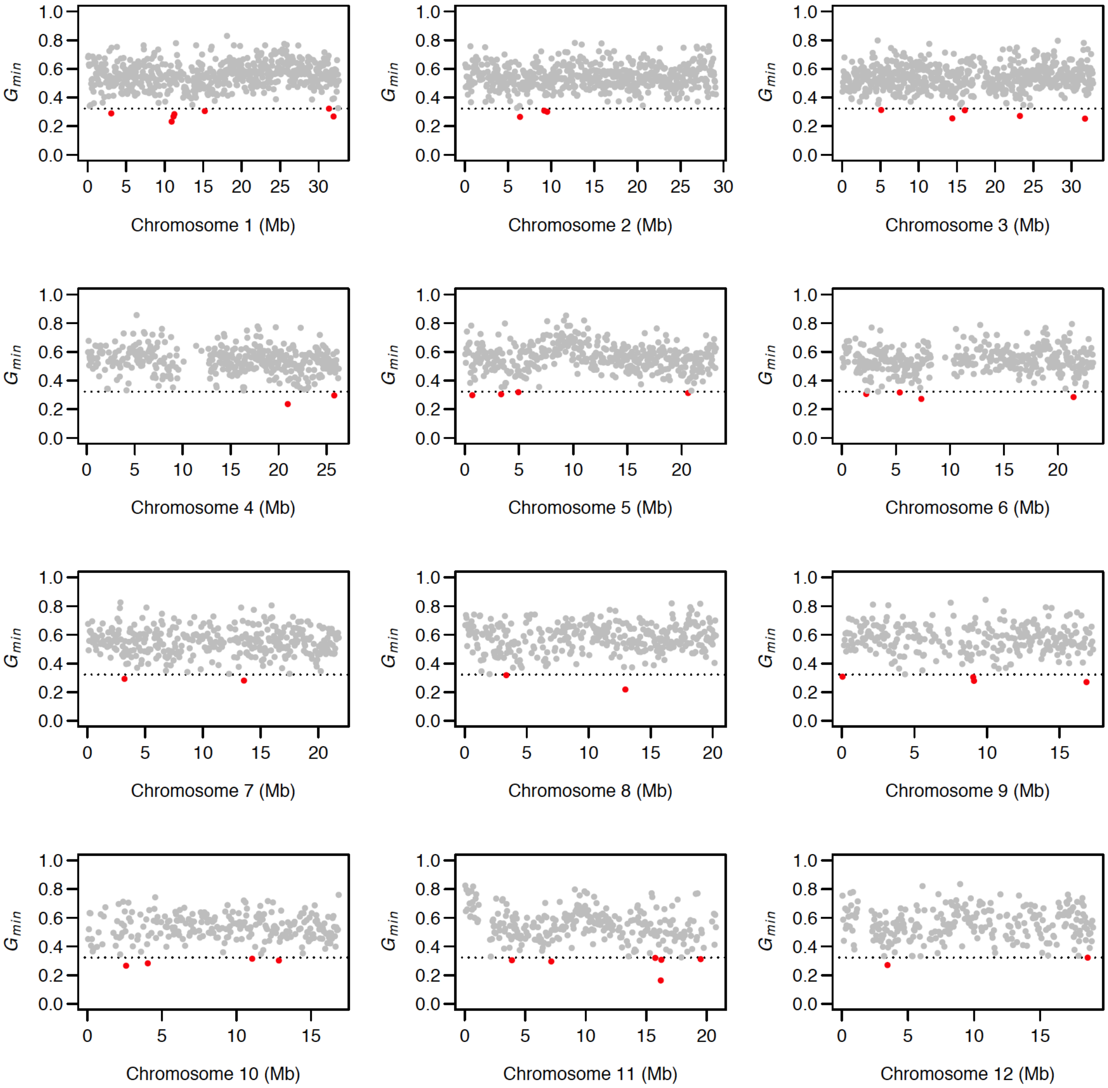


Supplementary Figure 15 Genome-wide distribution of the *iHS* P value in *O. glaberrima*. The P value corresponds to iHS scores that were transformed into *piHS* = -log[1-2|Φ(his)-0.5|] where Φ(x) represents the Gaussian cumulative distribution function. Assuming *iHS* values are normally distributed (under neutrality), *piHS* is interpreted as $\log_{10}(1/P)$ where P is the two-sided P-value associated to the neutral hypothesis of no selection. The dashed line represents the cutoff for the 500 SNPs with the top *piHS* values (representing 0.028% of SNPs genome-wide).

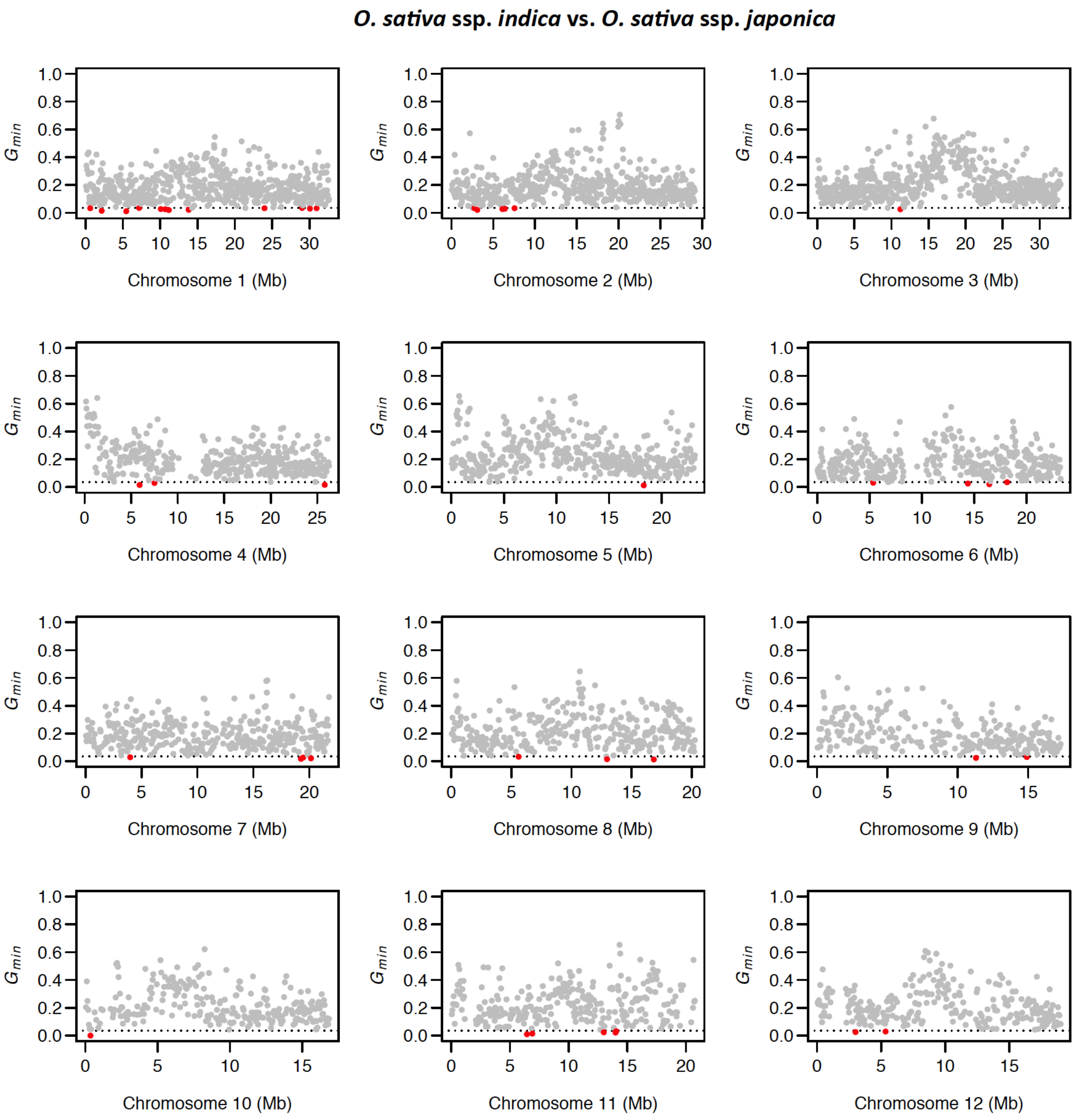


Supplementary Figure 16A. Plot of the G_{min} statistic in 50 kb windows across all chromosomes based on comparison of 20 *O. glaberrima* and 7 *O. sativa* ssp. *indica* resequenced accessions. Red points are in the lower 1% tail of the genome-wide distribution, also indicated by the dashed line ($G_{min} = 0.3661$)

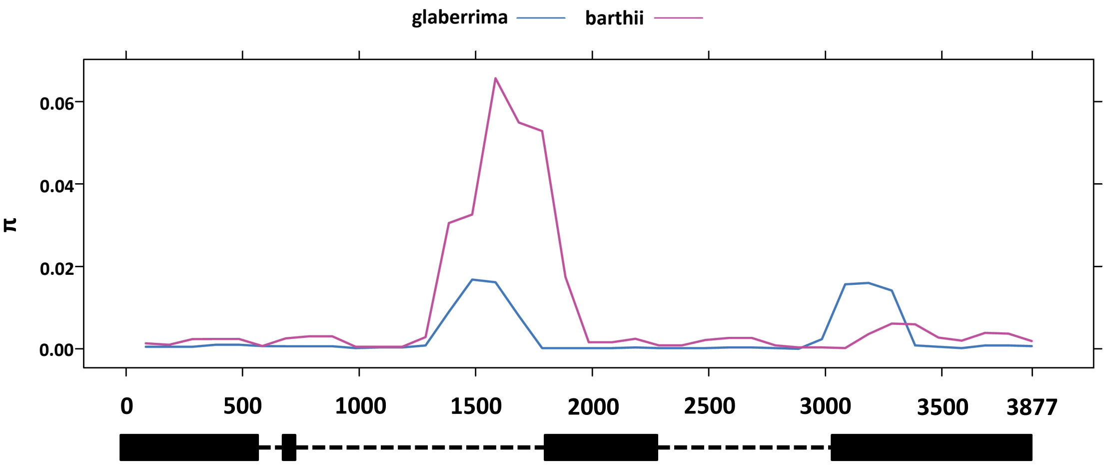
O. glaberrima* vs. *O. sativa* ssp. *japonica



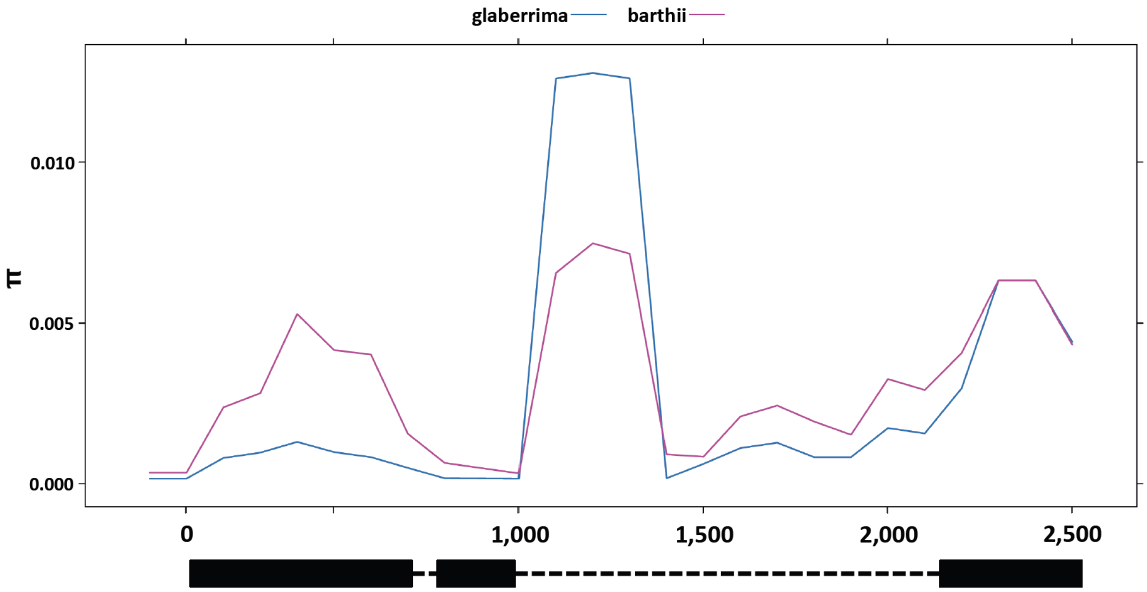
Supplementary Figure 16B. Plot of the G_{min} statistic in 50 kb windows across all chromosomes based on comparison of 20 *O. glaberrima* and 7 *O. sativa* ssp. *japonica* resequenced accessions. Red points are in the lower 1% tail of the genome-wide distribution, also indicated by the dashed line line ($G_{min} = 0.3225$).



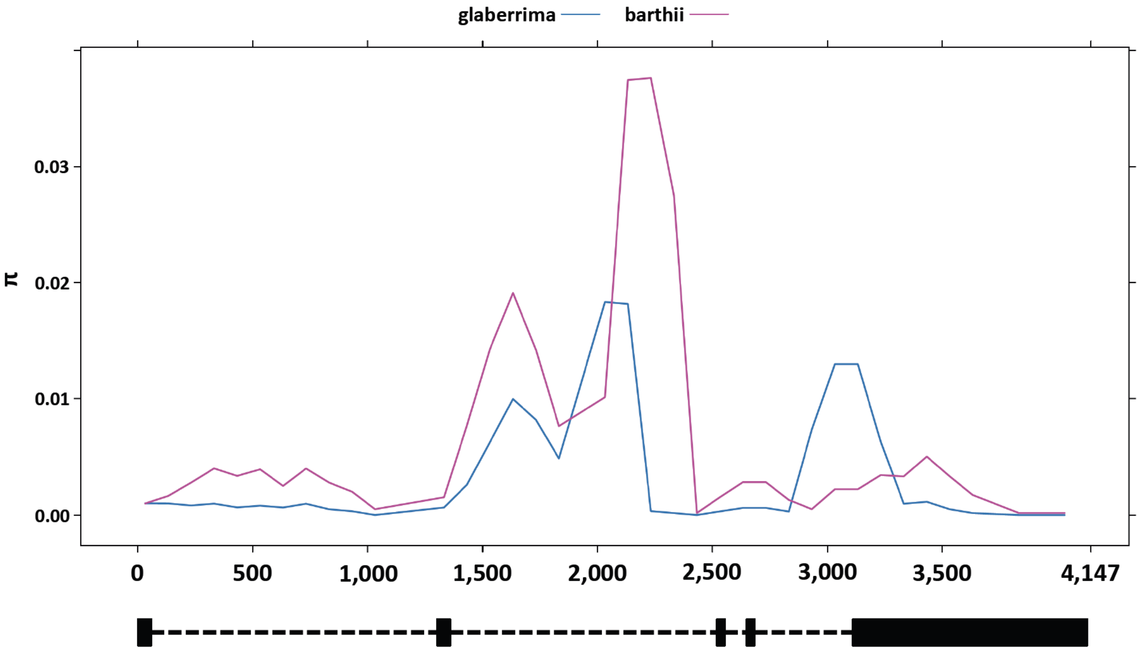
Supplementary Figure 16C. Plot of the G_{min} statistic in 50 kb windows across all chromosomes based on comparison of 7 *O. sativa* ssp. *indica* and 7 *O. sativa* ssp. *japonica* resequenced accessions. Red points are in the lower 1% tail of the genome-wide distribution, also indicated by the dashed line line ($G_{min} = 0.0353$). This figure is presented for comparison with Supplementary Figures 16A and 16B.



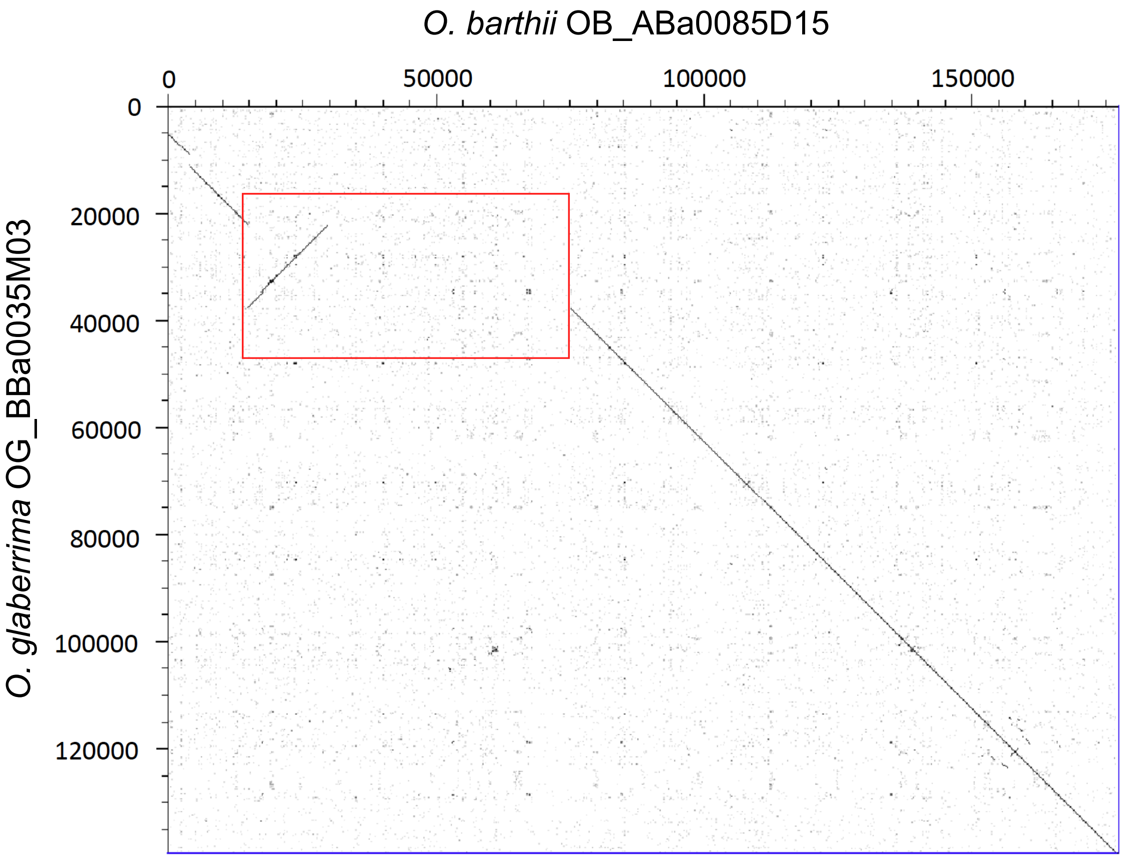
Supplementary Figure 17 Distribution of genetic diversity (π) of the *qSh1* gene between *O. glaberrima* and *O. barthii*. Genetic diversity was calculated in 300 bp sliding windows that overlap by 100 bp. The gene structure is shown below the plot.



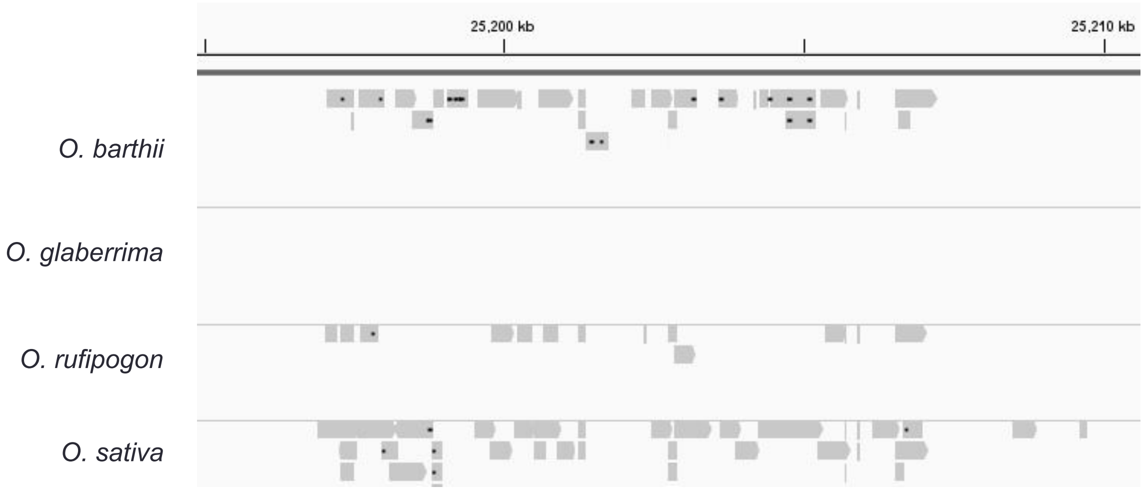
Supplementary Figure 18 Distribution of genetic diversity (π) of the *Sd1* gene between *O. glaberrima* and *O. barthii* of the *Sd1* gene. Genetic diversity was calculated in 300 bp sliding windows that are overlap by100 bp. The gene structure is shown below the plot.



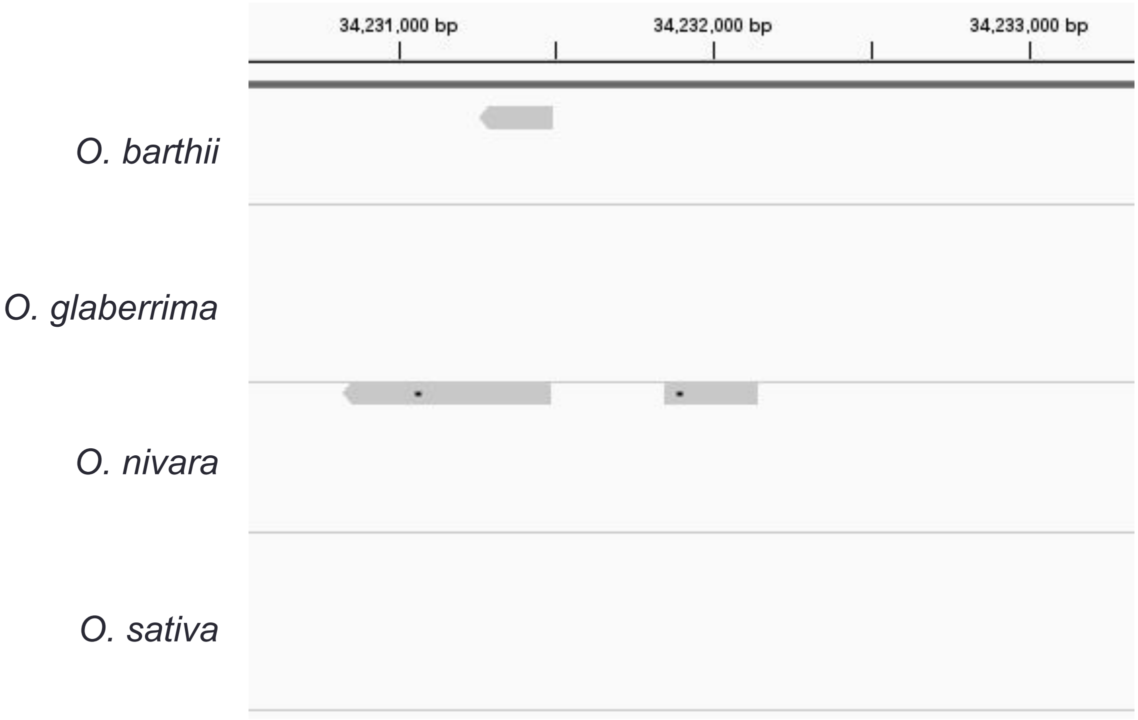
Supplementary Figure 19 Distribution of genetic diversity (π) of the *Dep1* gene between *O. glaberrima* and *O. barthii*. Genetic diversity was calculated in 300 bp sliding windows that overlap by 100 bp. The gene structure is shown below the plot.



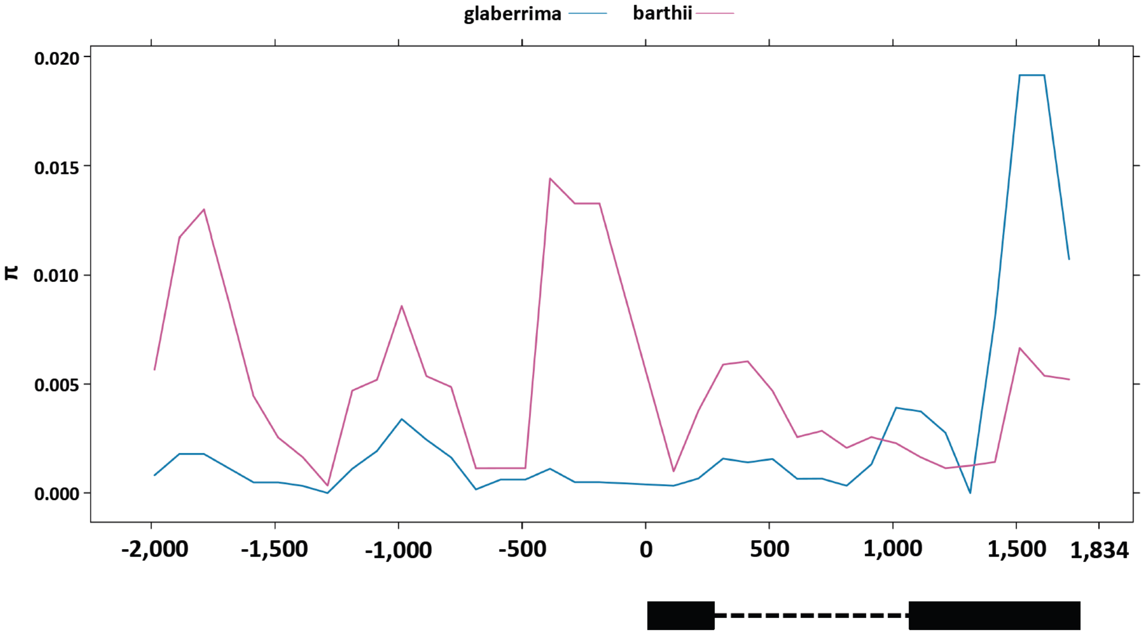
Supplementary Figure 20 Dot plot of the sequence alignment from two BACs, OG_BBa0035M03 (*O. glaberrima*) and OB_ABa0085D15 (*O. barthii*). The red rectangle shows a 15 kb inversion and a 45 kb deletion in *O. glaberrima*.



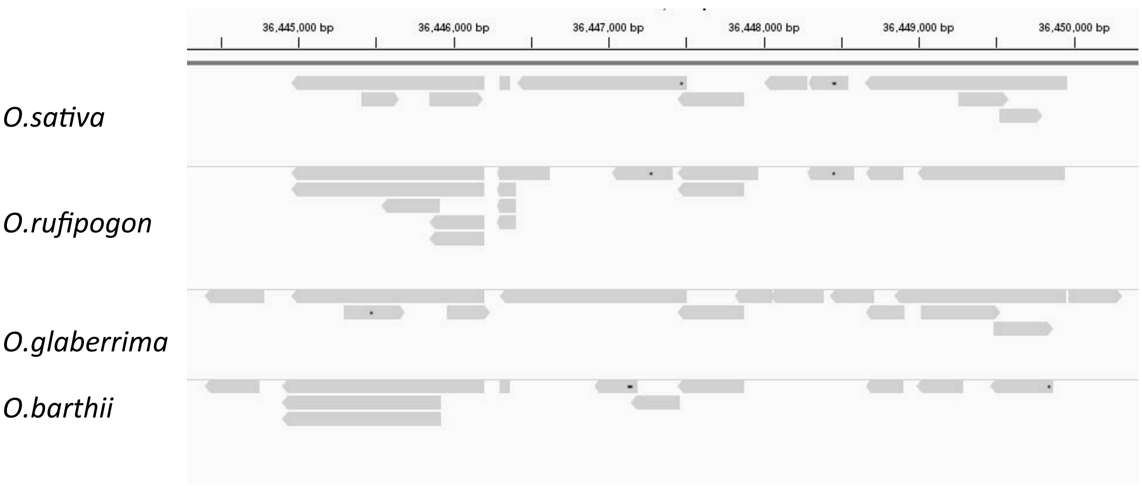
Supplementary Figure 21 Expression pattern of the *OsSh1* gene in mixed stage panicles in *O. barthii*, *O. glaberrima*, *O. rufipogon*, and *O. sativa* ssp. *japonica*. *OsSh1* was expressed in *O. barthii*, *O. rufipogon*, and *O. sativa* ssp. *japonica* but not in *O. glaberrima*. The raw RNA sequencing (RNA-Seq) reads from the four species were assembled independently and then mapped to the *O. sativa* ssp. *japonica* reference sequence. Results of these alignments were visualized using integrated genomics viewer (IGV). Each gray bar represents one assembled RNA-Seq fragment.



Supplementary Figure 22 Expression pattern of *Sh4* gene in mixed stage panicles in *O. barthii*, *O. glaberrima*, *O. nivara*, and *O. sativa* ssp. *japonica*. *Sh4* was expressed in *O. barthii*, *O. nivara*, but not in *O. sativa* ssp. *japonica* and *O. glaberrima.* The raw RNA sequencing (RNA-Seq) reads from the four species were assembled independently and then mapped to the *O. sativa* ssp. *japonica* reference sequence. Results of these alignments were visualized using integrated genomics viewer (IGV). Each gray bar represents one assembled RNA-Seq fragment.



Supplementary Figure 23 Distribution of genetic diversity (π) of the *Sh4* gene between *O. glaberrima* and *O. barthii*. Genetic diversity was calculated in 300 bp sliding windows that overlap by 100 bp. The gene structure is shown below the plot.



Supplementary Figure 24 Expression pattern of the *qSh1* gene in mixed stage panicles in *O. sativa* ssp. *japonica*, *O. rufipogon*, *O. glaberrima*, and *O. barthii*. *qSh1* was expressed in all four species. The raw RNA sequencing (RNA-Seq) reads from the four species were assembled independently and then mapped to the *O. sativa* ssp. *japonica* reference sequence. Results of these alignments were visualized using integrated genome viewer (IGV). Each gray bar represents one assembled RNA-Seq fragment.

Supplementary Table 1. Production sequencing summary of *O. glaberrima* genome

Chr	# BAC pools	BAC pool shotgun			Paired end (3kb)	
		# reads	Mb	SeqCov	# reads	PhyCov
1	13	3,128,847	962.94	22.3	991,322	34.4
2	11	2,790,111	901.06	25.1	511,558	21.3
3	6	2,771,920	982.82	54.3	433,497	36.1
4	11	2,153,556	751.04	21.2	529,294	22.4
5	9	2,011,749	753.08	25.3	387,732	19.6
6	10	3,265,626	1,099.70	35.8	361,325	17.6
7	10	4,258,849	1,122.59	37.9	834,476	33.8
8	10	3,833,583	986.57	34.7	793,834	33.5
9	8	1,996,521	581.41	25.6	741,616	39.2
10	8	1,958,453	732.57	32.3	439,512	29.1
11	10	2,119,388	792.21	27.9	380,534	20.1
12	9	1,460,208	570.69	20.7	382,199	20.8
WGS		5,623,228	2,203.4	5.5		
Total	**115**	**37,372,039**	**12,440.1**	**31.1**	**6,786,899**	**22.0**

Supplementary Table 2. Assembly statistics for the *O. glaberrima* genome sequence

Chromosome	Contig No.	Contig N50 (bp)	Scaffold no.		Scaffold N50 (bp)	Chromosome Size (bp)
			All	>500kb		
1	4661	19589	399	11	248470	34175486
2	5189	17013	835	4	183020	31644697
3	3661	20200	321	5	270287	35155656
4	3924	20502	451	7	180503	28085285
5	2634	27937	238	10	279325	23673821
6	4679	24486	686	5	186350	28779073
7	4073	19257	322	11	350076	25886956
8	3920	18484	348	7	305750	24024543
9	4025	13997	377	2	156364	19373526
10	2596	29139	240	5	266416	18658832
11	5021	15802	522	2	145504	24070662
12	6197	13316	570	2	152076	22555237

Supplementary Table 3. *O. glaberrima* assembly statistics for mapped and unmapped scaffolds

Chromosome	Mapped			UnMapped			Sequence ordered and oriented (%)
	Scaffold No.	Scaffold N50	Chromosome Length	Scaffold No.	Scaffold N50	Chromosome Length	
1	319	257774	32581612	80	86638	1593874	95.33
2	456	203532	29097369	379	8398	2547328	91.95
3	201	272213	33033699	120	59999	2121957	93.96
4	329	198126	26266211	122	35260	1819074	93.52
5	177	296234	23175214	61	14164	498601	97.89
6	392	200552	24135385	294	33500	4643688	83.86
7	158	368856	21783724	164	128781	4103232	84.14
8	194	307545	20273431	154	142929	3751112	84.38
9	272	168307	17580332	105	48487	1793184	90.74
10	142	275935	16896573	98	48131	1762259	90.55
11	357	163730	19804601	164	57063	3309911	85.68
12	372	159501	19117423	198	77405	3437818	85.75

Supplementary Table 4. Evaluation of *O. glaberrima* final assembly using four Sanger sequenced BACs

BAC ID	Query	Genbank accession	BAC length	Query start	Query end	Subject Chromosome	Subject start	Subject end	Query align length	Subject align length	%I
BAC1	OB_BBa0048H17	AC211606.1	120667	1	17280	Chr01	1514320	1531594	17280	17275	99.
				17331	59109	Chr01	1531977	1573747	41779	41771	99
				59141	59944	Chr01	1573991	1574794	804	804	99.
				60119	62811	Chr01	1575050	1577742	2693	2693	99.
				62793	73571	Chr01	1577993	1588771	10779	10779	99.
				73744	78039	Chr01	1589303	1593599	4296	4297	99.
				78884	80817	Chr01	1594231	1596160	1934	1930	99.
				80878	80949	Chr01	1597129	1597200	72	72	10
				81061	104434	Chr01	1597140	1620513	23374	23374	99.
				104921	120667	Chr01	1621022	1636760	15747	15739	99.
BAC2	OB_BBa0090P20	AC211289.1	144149	1	13897	Chr05	7153512	7139616	13897	13897	99.
		2 contigs		14194	22925	Chr05	7139447	7130717	8732	8731	99.
				22945	24055	Chr05	7130336	7129228	1111	1109	99.
				24179	46136	Chr05	7129097	7107142	21958	21956	99.
				46246	128969	Chr05	7106582	7023897	82724	82686	99.
				129002	144149	Chr05	7023767	7008620	15148	15148	10
BAC3	OG_BBa0066E18	FP340545.1	127790	1	24800	Chr06	1948614	1923819	24800	24796	99.
				25579	31510	Chr06	1922882	1916951	5932	5932	10
				32052	62356	Chr06	1916437	1886138	30305	30300	99.
				62962	82013	Chr06	1885748	1866697	19052	19052	10
				82081	127790	Chr06	1866152	1820456	45710	45697	99.
BAC4	OG_BBa0049I08	FP340543.1	164664	1	4434	Chr06	2130509	2126076	4434	4434	99.
				5603	65887	Chr06	2125001	2064727	60285	60275	99.

65832	73633	Chr06	2064138	2056338	7802	7801	99.
73770	147041	Chr06	2055931	1982684	73272	73248	99.
147068	158134	Chr06	1982465	1971398	11067	11068	99.
159930	164664	Chr06	1970272	1965539	4735	4734	99.

Supplementary Table 7. Number of WRKY family genes in *O. sativa* ssp. *japonica* and *O. glaberrima* genomes

Chromosome	*O. sativa* ssp. *japonica*		*O. glaberrima*	
	Number of genes	**Percentage**	**Number of genes**	**Percentage**
1	24	23.53%	20	21.74%
2	6	5.88%	6	6.52%
3	11	10.78%	10	10.87%
4	5	4.90%	5	5.43%
5	19	18.63%	15	16.30%
6	4	3.92%	5	5.43%
7	5	4.90%	5	5.43%
8	4	3.92%	3	3.26%
9	5	4.90%	5	5.43%
10	2	1.96%	2	2.17%
11	8	7.84%	9	9.78%
12	9	8.82%	7	7.61%
Unplaced	-	-	-	-
Total	102		92	

Supplementary Table 8. Distribution of WRKY gene groups in the *O. sativa* ssp. *japonica* and *O. glaberrima* genomes

WRKY	**Motif**	***O. sativa* ssp. *japonica***	***O. glaberrima***
Group I a	WRKY-C_2H_2/WRKY-C_2H_2	10	9
Group I b	WRKY-C_2HC/WRKY-C_2HC	0	0
Group II	WRKY-C_2H_2	56	44
Group III	WRKY-C_2HC	27	23
Group IV a	WRKY-C_2XX	5	1
Group IV b	WRKY	3	2
Truncated		0	9
Pseudogene		1	4
Total		102	92

Supplementary Table 10. Deletion status of 8 missing genes in the comparative analysis of flowering time, light response, and WRKY genes in the *O. glaberrima* genome assembly and RNA expression data

No.	**Gene name (*Arabidopsis thaliana*)**	**Gene name (*O. sativa*)**	**Locus ID (*O. sativa* ssp. *japonica*)**	**Genome assembly**	**RNA-Seq alignment**
1	*GI*	-	LOC_Os01g08700	No	Yes
2	*ARP6*	-	LOC_Os01g16414	No	Yes
3	*CO*	*Hd1*	LOC_Os06g16370	No	No
4	*RFT1*	-	LOC_Os07g30250	No	Yes
5	*SPL3*	*SPL13*	LOC_Os07g32170	No	Yes
6	*PIF3*	*OsPIL15*	Os01g0286200	No	Yes
7	-	*WRKY77*	LOC_Os01g40260	No	Yes
8		*WRKY*	LOC_Os08g17400	No	Yes

Supplementary Table 11. Summary of annotated repeats in *O. glaberrima* in comparison with *O. sativa*

	***O. glaberrima* Unplaced scaffolds**		***O. glaberrima* Placed scaffolds**		***O. glaberrima* Total**		***O. sativa***	
	bp	**%**	bp	**%**	bp	**%**	bp	**%**
Total length (excluding Ns)	29683365		273629535		303312900		372793407	
bases masked	12184150	**41.05**	91715285	**33.52**	103899435	**34.25**	155950693	**41.83**
SINEs	87897	**0.3**	935520	**0.34**	1023417	**0.34**	1169553	**0.31**
LINEs	350929	**1.18**	3651968	**1.33**	4002897	**1.32**	4563747	**1.22**
Ty1-*copia*	920904	**3.1**	7270849	**2.66**	8191753	**2.7**	10620343	**2.85**
Ty3-*gypsy*	4861777	**16.38**	26701823	**9.76**	31563600	**10.41**	56323148	**15.11**
LTR-RT Unclassified	1373726	**4.63**	9379202	**3.43**	10752928	**3.55**	15599470	**4.18**
LTR-RT (Total)	7156407	**24.11**	43351874	**15.84**	50508281	**16.65**	82542961	**22.14**
MuDR	230792	**0.78**	1983187	**0.72**	2213979	**0.73**	4572034	**1.23**
EnSPM	282399	**0.95**	1821224	**0.67**	2103623	**0.69**	5623429	**1.51**
hAT	102536	**0.35**	910674	**0.33**	1013210	**0.33**	1295672	**0.35**
Other classes	1456044	**4.91**	13137445	**4.8**	14593489	**4.81**	19576924	**5.25**
Unclassified	317839	**1.07**	3477453	**1.27**	3795292	**1.25**	4616171	**1.24**
DNA TE (Total-excluding Helitrons and MITEs)	2389610	**8.05**	21329983	**7.8**	23719593	**7.82**	35684230	**9.57**
MITEs	1357389	**4.57**	15530483	**5.68**	16887872	**5.57**	19534724	**5.24**
HELITRONs	433961	**1.46**	4498010	**1.64**	4931971	**1.63**	6746709	**1.81**
Unclassified	407957	**1.37**	2417447	**0.27**	2825404	**0.93**	5708769	**1.53**

Supplementary Table 12. TE fragments inside predicted genes in *O. glaberrima*

Repeat Type	Count
MITE	9685
UNKNOWN	558
DNA_TE	2112
Helitron	149
SINE	508
LINE	421
LTR-RT	2703

Supplementary Table 13. Comparison of LTR-RT insertion times in *O. sativa* and *O. glaberrima*

MYA	*O. glaberrima* (%)	*O. sativa* (%)
<1	29.18	65.54
1-2	50.52	24.29
2-3	15.17	6.38
>3	5.13	3.8

Supplementary Table 14. ***O. barthii* accessions sequenced for population structure analysis***

No.	Accession No.	Country of Origin	No.	Accession No.	Country of Origin	No.	Accession No.	Country of Origin
1	WAB0029000	Botswana	**33**	WAB0028957	Guinea	**65**	WAB0028877	Niger
2	IRGC105608	Cameroon	**34**	WAB0028876	Guinea	**66**	IRGC101381	Niger
3	WAB0028942	Cameroon	**35**	WAB0028956	Guinea	**67**	IRGC104084	Nigeria
4	WAB0028983	Cameroon	**36**	WAB0028889	Guinea	**68**	WAB0028987	Nigeria
5	WAB0028947	Cameroon	**37**	IRGC100931	Mali	**69**	WAB0028996	Nigeria
6	WAB0009240	Cameroon	**38**	IRGC100934	Mali	**70**	WAB0028997	Nigeria
7	WAB0028993	Cameroon	**39**	WAB0028896	Mali	**71**	WAB0028994	Nigeria
8	WAB0028944	Cameroon	**40**	WAB0028875	Mali	**72**	WAB0024904	Nigeria
9	WAB0028946	Cameroon	**41**	WAB0028961	Mali	**73**	WAB0026769	Nigeria
10	WAB0028948	Cameroon	**42**	WAB0028975	Mali	**74**	WAB0026770	Nigeria
11	WAB0028882	Cameroon	**43**	WAB0028980	Mali	**75**	WAB0028937	Nigeria
12	WAB0028884	Cameroon	**44**	WAB0028897	Mali	**76**	WAB0009239	Nigeria
13	IRGC104119	Chad	**45**	WAB0028910	Mali	**77**	WAB0028938	Nigeria
14	WAB0028934	Chad	**46**	WAB0028958	Mali	**78**	WAB0028940	Nigeria
15	WAB0028917	Chad	**47**	WAB0028959	Mali	**79**	IRGC103895	Senegal
16	WAB0028930	Chad	**48**	WAB0028976	Mali	**80**	WAB0028905	Senegal
17	WAB0028931	Chad	**49**	WAB0028977	Mali	**81**	WAB0028907	Senegal
18	WAB0028989	Chad	**50**	WAB0028979	Mali	**82**	IRGC101959	Senegal
19	WAB0028991	Chad	**51**	WAB0028981	Mali	**83**	IRGC106234	Sierra Leone
20	WAB0028919	Chad	**52**	WAB0028916	Mali	**84**	IRGC100927	Sierra Leone
21	WAB0028925	Chad	**53**	WAB0028911	Mali	**85**	WAB0028887	Tanzania
22	WAB0028926	Chad	**54**	WAB0028912	Mali	**86**	IRGC103912	Tanzania
23	WAB0028927	Chad	**55**	WAB0028913	Mali	**87**	WAB0028952	Zambia
24	WAB0028929	Chad	**56**	WAB0028900	Mali	**88**	WAB0028903	Zambia
25	WAB0028985	Chad	**57**	WAB0028885	Mali	**89**	WAB0030173	Unknown
26	WAB0028992	Chad	**58**	WAB0028893	Mali	**90**	WAB0026768	Unknown
27	WAB0028998	Chad	**59**	WAB0028894	Mali	**91**	IRGC100921	Unknown
28	WAB0030151	Chad	**60**	WAB0028915	Mali	**92**	IRGC100922	Unknown
29	IRGC100122	Gambia	**61**	WAB0030186	Mali	**93**	IRGC100939	Unknown
30	WAB0028874	Gambia	**62**	WAB0012712	Mali	**94**	IRGC101252	Unknown
31	WAB0028967	Gambia	**63**	IRGC101240	Mali			
32	WAB0028972	Gambia	**64**	IRGC103534	Mali			

***Accession numbers starting with "WAB" were obtained from African Rice Center. The remaining accessions were obtained from the International Rice Research Institute.**

Supplementary Table 15. Partitioning of ancestry within *O. barthii* population (94 accessions) based on SNPs data

No.	Accession No.	Subpopulation	No.	Accession No.	Subpopulation	No.	Accession No.	Subpopulation
1	IRGC100931	OB-I	**33**	WAB0028944	OB-II	**65**	WAB0028896	OB-V
2	WAB0028875	OB-I	**34**	WAB0028958	OB-II	**66**	WAB0028876	OB-V
3	WAB0028877	OB-II	**35**	WAB0028985	OB-II	**67**	WAB0028874	OB-V
4	WAB0028975	OB-I	**36**	WAB0028994	OB-II	**68**	WAB0028947	OB-V
5	WAB0028946	OB-I	**37**	WAB0028882	OB-II	**69**	WAB0028956	OB-V
6	WAB0028976	OB-I	**38**	WAB0028884	OB-II	**70**	WAB0028957	OB-V
7	WAB0028977	OB-I	**39**	WAB0028952	OB-III	**71**	WAB0028967	OB-V
8	WAB0028979	OB-I	**40**	WAB0028927	OB-III	**72**	WAB0028961	OB-V
9	WAB0028980	OB-I	**41**	WAB0028992	OB-III	**73**	WAB0028972	OB-V
10	WAB0028981	OB-I	**42**	WAB0028937	OB-III	**74**	WAB0028905	OB-V
11	WAB0028897	OB-I	**43**	WAB0009239	OB-III	**75**	WAB0028907	OB-V
12	WAB0028916	OB-I	**44**	WAB0028938	OB-III	**76**	WAB0009240	OB-V
13	WAB0028931	OB-I	**45**	WAB0028940	OB-III	**77**	WAB0028996	OB-V
14	IRGC101240	OB-I	**46**	IRGC106234	OB-IV	**78**	WAB0028997	OB-V
15	IRGC103534	OB-I	**47**	WAB0028987	OB-IV	**79**	WAB0028919	OB-V
16	IRGC103912	OB-II	**48**	WAB0028910	OB-IV	**80**	WAB0028948	OB-V
17	IRGC104119	OB-II	**49**	WAB0028911	OB-IV	**81**	WAB0030173	OB-V
18	IRGC105608	OB-II	**50**	WAB0028912	OB-IV	**82**	WAB0028889	OB-V
19	WAB0029000	OB-II	**51**	WAB0028913	OB-IV	**83**	WAB0028959	OB-V
20	WAB0028934	OB-II	**52**	WAB0028900	OB-IV	**84**	WAB0028998	OB-V
21	WAB0028917	OB-II	**53**	WAB0028893	OB-IV	**85**	WAB0028885	OB-V
22	WAB0028942	OB-II	**54**	WAB0028894	OB-IV	**86**	WAB0024904	OB-V
23	WAB0028903	OB-II	**55**	WAB0028915	OB-IV	**87**	WAB0026769	OB-V
24	WAB0028983	OB-II	**56**	WAB0030186	OB-IV	**88**	WAB0026770	OB-V
25	WAB0028930	OB-II	**57**	WAB0030151	OB-IV	**89**	WAB0012712	OB-V
26	WAB0028887	OB-II	**58**	IRGC100922	OB-IV	**90**	WAB0026768	OB-V
27	WAB0028989	OB-II	**59**	IRGC100927	OB-IV	**91**	IRGC100939	OB-V
28	WAB0028991	OB-II	**60**	IRGC100122	OB-V	**92**	IRGC101252	OB-V
29	WAB0028993	OB-II	**61**	IRGC100921	OB-V	**93**	IRGC101381	OB-V
30	WAB0028925	OB-II	**62**	IRGC100934	OB-V	**94**	IRGC101959	OB-V
31	WAB0028926	OB-II	**63**	IRGC103895	OB-V			
32	WAB0028929	OB-II	**64**	IRGC104084	OB-V			

Supplementary Table 16. ***O. glaberrima* accessions used for analyzing the genetic relationship between *O. glaberrima* and *O. barthii****

No.	Species	Accession No.	Country of Origin
1	*O. glaberrima*	IRGC103469	Burkina Faso
2	*O. glaberrima*	TOG5923	Liberia
3	*O. glaberrima*	TOG5949	Liberia
4	*O. glaberrima*	TOG5467	Nigeria
5	*O. glaberrima*	TOG5457	Nigeria
6	*O. glaberrima*	TOG7025	Sierra Leone
7	*O. glaberrima*	TOG7102	Mali
8	*O. glaberrima*	IRGC67563	Ghana
9	*O. glaberrima*	IRGC68939	Madagascar
10	*O. glaberrima*	IRGC68976	Guyana
11	*O. glaberrima*	IRGC75500	Burkina Faso
12	*O. glaberrima*	IRGC96841	Zimbabwe
13	*O. glaberrima*	IRGC101049	Senegal
14	*O. glaberrima*	IRGC103472	Burkina Faso
15	*O. glaberrima*	IRGC103520	Mali
16	*O. glaberrima*	IRGC103632	Mali
17	*O. glaberrima*	IRGC103937	Liberia
18	*O. glaberrima*	IRGC104206	Ghana
19	*O. glaberrima*	IRGC104574	Mali
20	*O. glaberrima*	IRGC104955	Sierra Leone

***Accession numbers starting with "TOG" were obtained from African Rice Center. The remaining accessions were obtained from the International Rice Research Institute.**

Supplementary Table 17. F_{ST} values between *O. glaberrima* (OG) and each *O. barthii* subgroups (OB-I to OB-V)

	OG
OB_I	0.14438
OB_II	0.099778
OB_III	0.36987
OB_IV	0.055195
OB_V	0.0096748

Supplementary Table 18. Summary of geographic distribution of *O. barthii* accessions of different subgroups

Country of Origin	Groups					Number of Accessions
	OB-I	OB-II	OB-III	OB-IV	OB-V	
Gambia	0	0	0	0	1	4
Guinea	0	0	0	0	1	4
Senegal	0	0	0	0	1	4
Sierra Leone	0	0	0	1	0	2
Nigeria	0	0.08	0.34	0.08	0.5	12
Mali	0.43	0.04	0	0.32	0.21	28
Niger	0.5	0	0	0	0.5	2
Cameroon	0.09	0.64	0	0	0.27	11
Chad	0.06	0.6	0.125	0.06	0.125	16
Zambia	0	0.5	0.5	0	0	2
Botswana	0	1	0	0	0	1
Tanzania	0.5	0.5	0	0	0	2
Unknown	0	0	0	0.33	0.67	6

Supplementary Table 19. *O. barthii* accessions used for investigating artificial selection in *O. glaberrima*

No	Species	Accession No.	Country of Origin
1	*O. barthii*	IRGC100122	Gambia
2	*O. barthii*	IRGC100931	Mali
3	*O. barthii*	IRGC100934	Mali
4	*O. barthii*	IRGC103895	Senegal
5	*O. barthii*	IRGC104084	Nigeria
6	*O. barthii*	IRGC104119	Chad
7	*O. barthii*	IRGC105608	Cameroon
8	*O. barthii*	IRGC106234	Sierra Leone
9	*O. barthii*	IRGC103912	Tanzania
10	*O. barthii*	WAB0028952	Zambia
11	*O. barthii*	WAB0028903	Zambia
12	*O. barthii*	WAB0028980	Mali
13	*O. barthii*	WAB0028987	Nigeria
14	*O. barthii*	WAB0028958	Mali
15	*O. barthii*	WAB0028976	Mali
16	*O. barthii*	WAB0028979	Mali
17	*O. barthii*	WAB0028992	Chad
18	*O. barthii*	WAB0028938	Nigeria
19	*O. barthii*	WAB0030151	Chad

***Accession numbers starting with "WAB" were obtained from African Rice Center. The remaining accessions were obtained from the International Rice Research Institute.**

Supplementary Table 20. Genome wide detection of regions under selective sweeps during the domestication of *O. glaberrima* by comparing genetic diversity of *O. glaberrima* and *O. barthii* group V (OB-V)

Chromosome	Interval (Mb)	π_w/π_c	***O.glaberrima*** **orthologs of known gene or loci in *O. sativa* ssp. *japonica***
1	16.9-17	6.165017386	
1	17.7-17.8	5.833727381	
1	18.3-18.4	6.156131663	
1	30.3-30.4	5.822852867	
1	31.4-31.5	6.617159947	
1	31.5-31.6	8.135254577	
1	31.6-31.7	7.584467728	
1	31.7-31.8	6.053773847	
2	2.6-2.7	6.294149825	
2	2.7-2.8	6.08949706	
2	6.2-6.3	6.414645744	
2	13.9-14	6.524117999	
3	1-1.1	6.767033941	
3	12-12.1	6.140314363	
3	31.5-31.6	6.272661611	
4	5.5-5.6	6.025091367	
4	6.4-6.5	5.823483665	
4	8.3-8.4	6.742639939	
4	8.5-8.6	5.929918359	
4	8.6-8.7	6.685373561	
4	8.7-8.8	7.312704191	
4	9.1-9.2	6.901842129	
4	15.6-15.7	6.845406423	*badh2*
4	17.2-17.3	6.64734209	
4	18.7-18.8	10.27095385	
4	20.7-20.8	6.592118188	
4	21.1-21.2	7.145963698	
5	0.8-0.9	6.820534889	
5	19.7-19.8	6.153361673	
5	20.5-20.6	6.460170114	
6	4.4-4.5	5.946290135	
6	5.2-5.3	6.157487611	
6	15-15.1	7.782344867	
6	16.7-16.8	7.935773318	
6	19.1-19.2	6.956370115	

6	19.2-19.3	8.080605029
6	19.3-19.4	6.463177199
6	19.6-19.7	7.14640548
6	20.9-21	5.94268207
6	21.1-21.2	6.01502746
6	21.8-21.9	7.433686046
6	21.9-22	6.407995227
6	22-22.1	8.776432892
6	22.8-22.9	6.050073719
6	22.9-23	8.073736639
6	23-23.1	8.021476407
7	8.3-8.4	5.869798702
7	15.6-15.7	6.483087638
7	17.2-17.3	6.266461188
7	19.3-19.4	6.007596813
7	21-21.1	6.081882037
8	1.4-1.5	7.16874163
8	1.5-1.6	6.052208826
8	2.4-2.5	5.90584113
8	2.5-2.6	7.156352859
8	3.8-3.9	5.980708514
8	11-11.1	8.905743933
8	11.1-11.2	8.822678726
8	13.6-13.7	5.931757793
9	14.7-14.8	8.793988931
9	16.2-16.3	6.975116499
10	0.2-0.3	6.07593137
10	3.3-3.4	5.822625614
10	10.9-11	5.906312458
11	15-15.1	5.836107484
11	17.8-17.9	6.521209811
12	1.4-1.5	6.442102085
12	1.9-2	5.965681589
12	2-2.1	7.603855214
12	7.6-7.7	7.113233549
12	15.3-15.4	9.876422448
12	15.4-15.5	7.035851392

Supplementary Table 21. List of agronomically important genes whose homologues are located close, within the window of LD decay (<251 Kb), to genomic regions with strong evidence of recent selective sweeps in *O. glaberrima*

Chromosome	*Gene*	Gene Function	Gene Coordinates	Position of CLR peak	CLR
1	*Sd1*	*Semi-dwarfism*	28,541,899-28,544,581	28,529,624	159.39
1	*OsNAC6*	*Abiotic resistance, salt tolerance*	28,555,290-28,556,940	28,529,624	159.39
2	*EP3*	*Grain number*	8,382,639-8,383,865	8,490,987	174.19
4	*Sh4*	*Shattering*	25,150,788-25,152,622	25,087,584	135.53
7	*Ep2*	*Panicle*	18,073,478-18,080,578	18,264,800	152.23
9	*Sub1*	*Abiotic resistance, submerge*	4,605,563-4,606,261	4,436,071	150.22
11	*Xa26*	*Biotic resistance, blight*	20,470,000-20,485,675	20,380,430	146.19

Supplementary Table 23. *O. sativa* accessions used for investigating introgression

No	Species	Accessions No.	Country of Origin
1	*O. sativa* ssp. *japonica*	IRGC58398	Brunel
2	*O. sativa* ssp. *japonica*	IRGC58716	Bangladesh
3	*O. sativa* ssp. *japonica*	IRGC3860	Philippines
4	*O. sativa* ssp. *japonica*	IRGC28275	Pakistan
5	*O. sativa* ssp. *japonica*	IRGC32298	Iran
6	*O. sativa* ssp. *japonica*	IRGC36420	Sri Lanka
7	*O. sativa* ssp. *japonica*	IRGC56350	Senegal
8	*O. sativa* ssp. *indica*	IRGC9505	Iraq
9	*O. sativa* ssp. *indica*	IRGC12880	Iran
10	*O. sativa* ssp. *indica*	IRGC13793	Brunel
11	*O. sativa* ssp. *indica*	IRGC32441	Bhutan
12	*O. sativa* ssp. *indica*	IRGC58267	Afghanistan
13	*O. sativa* ssp. *indica*	IRGC50409	Brazil
14	*O. sativa* ssp. *indica*	IRGC69559	Egypt

Supplementary Table 24. Homologues of domestication genes in *O. glaberrima*

No.	Gene name	Function	Locus ID	
			Oryza sativa* ssp. *japonica	***Oryza glaberrima***
1	*Gn1a*[1]	Grain productivity	LOC_Os01g10110	ORGLA01g0059300
2	*qSH1*[2]	Shattering	LOC_Os01g62920	ORGLA01G0311100
3	*sd1*[3]	Semi-dwarfing	LOC_Os01g66100	ORGLA01g0334700
4	*GW2*[4]	Grain width	LOC_Os02g14720	ORGLA02g0093000
5	*GS3*[5]	Grain size/shape	LOC_Os03g29380	ORGLA03g0195600
6	*OsSh1*[6]	Shattering	LOC_Os03g44710	**Deleted in *Oryza glaberrima***
7	*GIF1*[7]	Grain incomplete filling	LOC_Os04g33740	ORGLA04g0089800
8	*badh2.1*[8]	Flavor or fragrance	LOC_Os04g39020	ORGLA04g0122100
9	*Phr1*[9]	Grain discoloration	LOC_Os04g53300	ORGLA04g0229000
10	*OsLG1*[10]	Closed panicle	LOC_Os04g56170	ORGLA04G0242500
11	*Sh4*[11]	Shattering	LOC_Os04g57530	ORGLA04G0254300
12	*Hd1*[12]	Flowering time	LOC_Os06g16370	**Deleted in *Oryza glaberrima***
13	*MOC1*[13]	Tillering	LOC_Os06g40780	ORGLA06g0172100
14	*PROG1*[14]	Plant Architecture	LOC_Os07g05900	Not found
15	*Rc*[15]	Red pericarp	LOC_Os07g11020	ORGLA07g0060000
16	*Sdr4*[16]	Seed dormancy	LOC_Os07g39700	ORGLA07g0155600
17	*Ep2*[17]	Erect panicle	LOC_Os07g42410	ORGLA07g0170500
18	*Ipa1*[18]	Ideal plant architecture	LOC_Os08g39890	ORGLA08g0176500
19	*Dep1*[19]	Panicle architecture	LOC_Os09g26999	ORGLA09g0089300

Supplementary Table 25. Tajima’s D of 20 domestication genes in *O. glaberrima and O. barthii*

Gene	Species	Sample size	Sequence length	Tajima's D
Gn1a	*O. glaberrima*	20	4472	0.35195
	O. barthii	19	4472	0.07865
qSH1	*O. glaberrima*	20	3888	**-2.12828***
	O. barthii	19	3888	0.08885
Sd1	*O. glaberrima*	20	3094	**-2.14531***
	O. barthii	19	3094	-1.27688
Gw2	*O. glaberrima*	20	6616	-1.64913
	O. barthii	19	6616	-1.84678
GS3	*O. glaberrima*	20	5718	0.17269
	O. barthii	19	5718	-1.69595
GIF1	*O. glaberrima*	20	14338	-1.95942
	O. barthii	19	14338	-0.93243
badh2	*O. glaberrima*	20	4441	-1.8539
	O. barthii	19	4441	-1.97349
Phr1	*O. glaberrima*	20	2326	-1.16689
	O. barthii	19	2326	-0.58515
OsLG1	*O. glaberrima*	20	3424	-1.4346
	O. barthii	19	3424	-1.03152
Sh4	*O. glaberrima*	20	1823	-1.6767
	O. barthii	19	1823	-0.77275
Moc1	*O. glaberrima*	20	14572	-1.78003
	O. barthii	19	14572	-0.12163
Rc	*O. glaberrima*	20	6890	-1.51284
	O. barthii	19	6890	-0.02904
Sdr4	*O. glaberrima*	20	1035	-0.39341
	O. barthii	19	1035	-1.63067
Ep2	*O. glaberrima*	20	7655	-0.98129
	O. barthii	19	7655	-1.98277
Ipa1	*O. glaberrima*	20	3782	-1.86788
	O. barthii	19	3782	-0.01227
Dep1	*O. glaberrima*	20	4613	**-2.12144***
	O. barthii	19	4613	-1.3807
***Sh4* (promoter)**	*O. glaberrima*	20	2000	**-2.1306***
	O. barthii	19	2000	-0.652774

***Genes in the top 5 percentile of Tajima’s D values of all the genes in the genome**

Supplementary Table 26. Comparison of polymorphism levels of the *qSh1* gene and genome-wide intergenic regions in *O. glaberrima* and *O. barthii*

	qSh1	Intergenic regions	Ratio (intergenic/*qSh1*)
π (*O. glaberrima*)	0.00072	0.004799	6.665278
π (*O. barthii*)	0.00536	0.008506	1.586941
Ratio (*O. barthii*/*O. glaberrima*)	7.4444	1.77245	

Supplementary Table 27. Comparison of polymorphism levels of the *Sd1* gene and genome-wide intergenic regions in *O. glaberrima* and *O. barthii*

	Sd1	Intergenic regions	Ratio (intergenic/*Sd1*)
π (*O. glaberrima*)	0.00093	0.004799	5.160215
π (*O. barthii*)	0.0038	0.008506	2.238421
Ratio (*O. barthii*/*O. glaberrima*)	4.08602	1.77245	

Supplementary Table 28. Comparison of polymorphism levels of the *Dep1* gene and genome-wide intergenic regions in *O. glaberrima* and *O. barthii*

	Dep1	Intergenic regions	Ratio (intergenic/*Dep1*)
π (*O. glaberrima*)	0.00048	0.004799	9.997917
π (*O. barthii*)	0.00307	0.008506	2.770684
Ratio (*O. barthii*/*O. glaberrima*)	6.39583	1.77245	

Supplementary Table 29. Comparison of polymorphism levels of the promoter region of the *Sh4* gene in *O. glaberrima* and *O. barthii*

	Sh4 (Promoter)	Intergenic regions	Ratio (intergenic/Sh4 Promoter)
π (*O. glaberrima*)	0.00095	0.004799	5.051578947
π (*O. barthii*)	0.005	0.008506	1.7012
Ratio (*O. barthii*/*O. glaberrima*)	5.263157895	1.77245	

Reference:

1. Ashikari, M. Cytokinin Oxidase Regulates Rice Grain Production. *Science* **309**, 741–745 (2005).
2. Konishi, S. *et al.* An SNP caused loss of seed shattering during rice domestication. *Science* **312**, 1392–1396 (2006).
3. Spielmeyer, W., Ellis, M. H. & Chandler, P. M. Semidwarf (*sd-1*), 'green revolution' rice, contains a defective gibberellin 20-oxidase gene. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 9043–9048 (2002).
4. Song, X.-J., Huang, W., Shi, M., Zhu, M.-Z. & Lin, H.-X. A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat. Genet.* **39**, 623–630 (2007).
5. Fan, C. *et al. GS3*, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. *Theor. Appl. Genet.* **112**, 1164–1171 (2006).
6. Lin, Z. *et al.* Parallel domestication of the *Shattering1* genes in cereals. *Nat. Genet.* **44**, 720–724 (2012).
7. Wang, E. *et al.* Control of rice grain-filling and yield by a gene with a potential signature of domestication. *Nat. Genet.* **40**, 1370–1374 (2008).
8. Kovach, M. J., Calingacion, M. N., Fitzgerald, M. A. & McCouch, S. R. The origin and evolution of fragrance in rice (*Oryza sativa* L.). *Proc. Natl. Acad. Sci. U.S.A.* **106**, 14444–14449 (2009).
9. Yu, Y. *et al.* Independent losses of function in a polyphenol oxidase in rice: differentiation in grain discoloration between subspecies and the role of positive selection under domestication. *Plant Cell* **20**, 2946–2959 (2008).
10. Ishii, T. *et al. OsLG1* regulates a closed panicle trait in domesticated rice. *Nat. Genet.* **45**, 462–465 (2013).
11. Li, C., Zhou, A. & Sang, T. Rice domestication by reducing shattering. *Science* **311**, 1936–1939 (2006).
12. Yano, M. *et al. Hd1*, a Major Photoperiod Sensitivity Quantitative Trait Locus in Rice, Is Closely Related to the Arabidopsis Flowering Time Gene *CONSTANS*. *Plant Cell* (2000).
13. Li, X. *et al.* Control of tillering in rice. *Nature* **422**, 618–621 (2003).
14. Tan, L. *et al.* Control of a key transition from prostrate to erect growth in rice domestication. *Nat. Genet.* **40**, 1360–1364 (2008).
15. Sweeney, M. T., Thomson, M. J., Pfeil, B. E. & McCouch, S. Caught red-handed: Rc encodes a basic helix-loop-helix protein conditioning red pericarp in rice. *Plant Cell* **18**, 283–294 (2006).
16. Sugimoto, K. *et al.* Molecular cloning of *Sdr4*, a regulator involved in seed dormancy and domestication of rice. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 5792–5797 (2010).
17. Zhu, K. *et al. ERECT PANICLE2* Encodes a Novel Protein That Regulates Panicle Erectness in Indica Rice. *Genetics* **184**, 343–350 (2010).
18. Jiao, Y. *et al.* Regulation of *OsSPL14* by OsmiR156 defines ideal plant architecture in rice. *Nat. Genet.* **42**, 541–544 (2010).
19. Huang, X. *et al.* Natural variation at the *DEP1* locus enhances grain yield in rice.

Nat. Genet. **41**, 494–497 (2009).