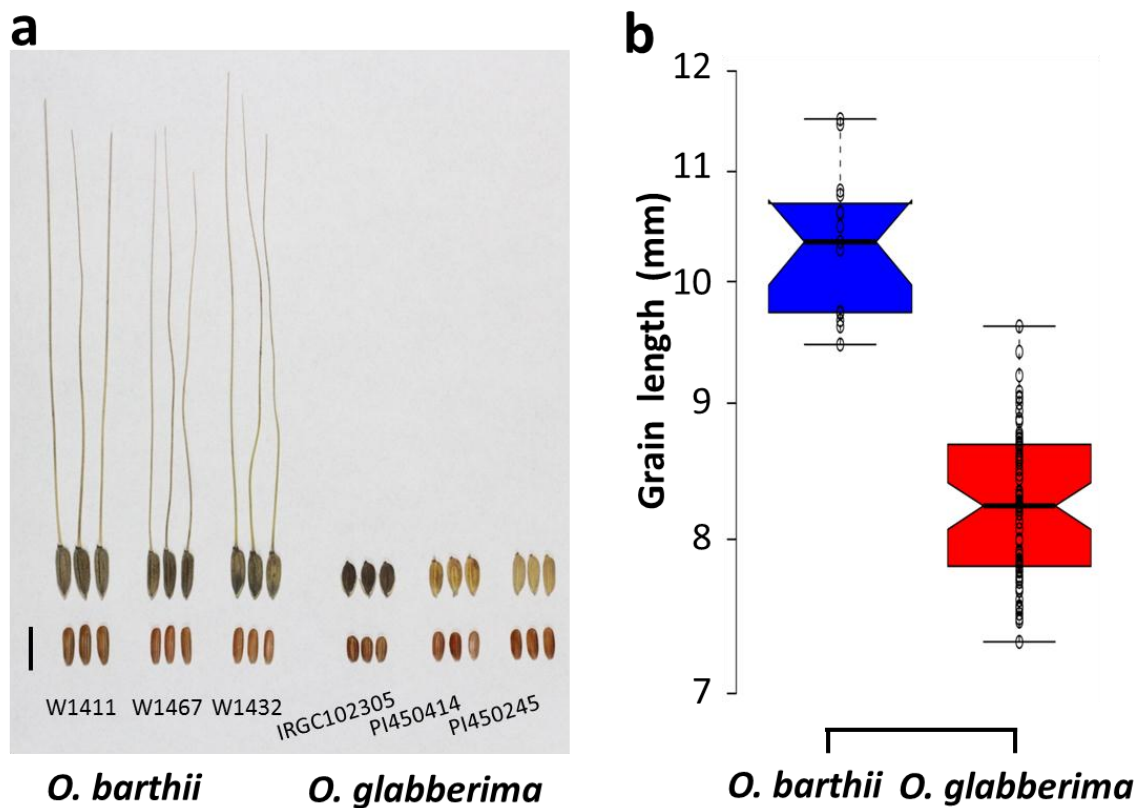


In the format provided by the authors and unedited.

A single-nucleotide polymorphism causes smaller grain size and loss of seed shattering during African rice domestication

Wenguang Wu^{1†}, Xiaoyun Liu^{1†}, Muhua Wang^{2†‡}, Rachel S. Meyer³, Xiaojin Luo⁴, Marie-Noelle Ndjondjop⁵, Lubin Tan¹, Jianwei Zhang², Jianzhong Wu⁶, Hongwei Cai¹, Chuanqing Sun^{1,7}, Xiangkun Wang¹, Rod A. Wing^{2,8*} and Zuofeng Zhu^{1*}

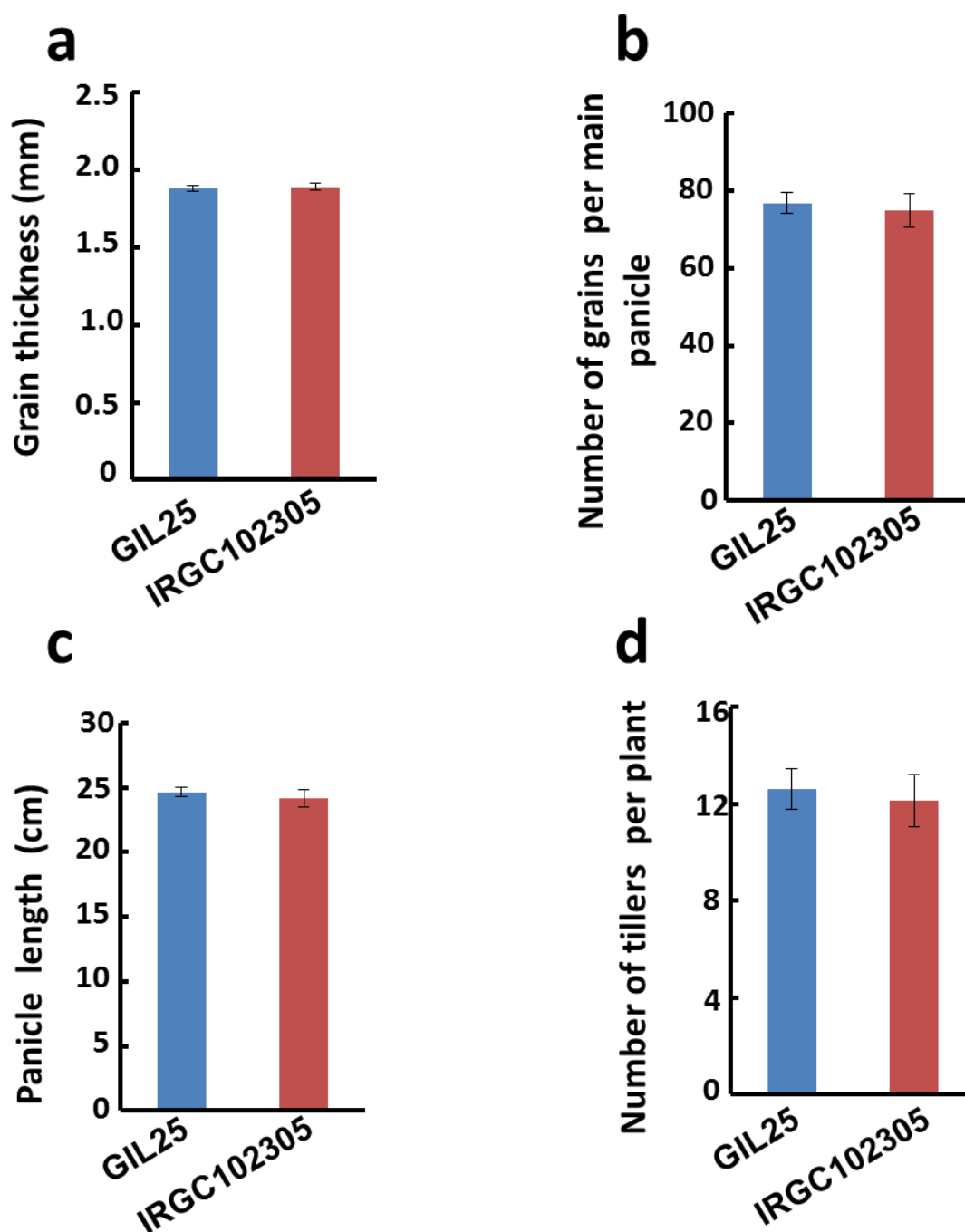
¹National Centre for Evaluation of Agricultural Wild Plants (Rice), MOE Key Laboratory of Crop Heterosis and Utilization, Department of Plant Genetics and Breeding, China Agricultural University, Beijing 100193, China. ²Arizona Genomics Institute, School of Plant Sciences, University of Arizona, Tucson, Arizona 85721, USA. ³Department of Ecology and Evolutionary Biology, University of California Los Angeles, Los Angeles, California 90095, USA. ⁴State Key Laboratory of Genetic Engineering, Institute of Genetics, School of Life Sciences, Fudan University, Shanghai 200433, China. ⁵Africa Rice Centre, Cotonou, Benin 2031. ⁶Advanced Genomics Breeding Section, Institute of Crop Science, National Agriculture and Food Research Organization, Tsukuba 305-8602, Japan. ⁷State Key Laboratory of Plant Physiology and Biochemistry, China Agricultural University, Beijing 100193, China. ⁸International Rice Research Institute, T.T. Chang Genetic Resources Centre, Los Baños, Laguna, Philippines. [†]Present address: Friedrich Miescher Laboratory of the Max Planck Society, Tübingen 72076, Germany. [‡]These authors contributed equally to this work. *e-mail: zhuzf@cau.edu.cn; rwing@mail.arizona.edu



Supplementary Figure 1. Comparison of grain length between *O. barthii* and *O. glabberima*.

(a) Grain of *O. barthii* and *O. glaberrima*. Bar=1 cm.

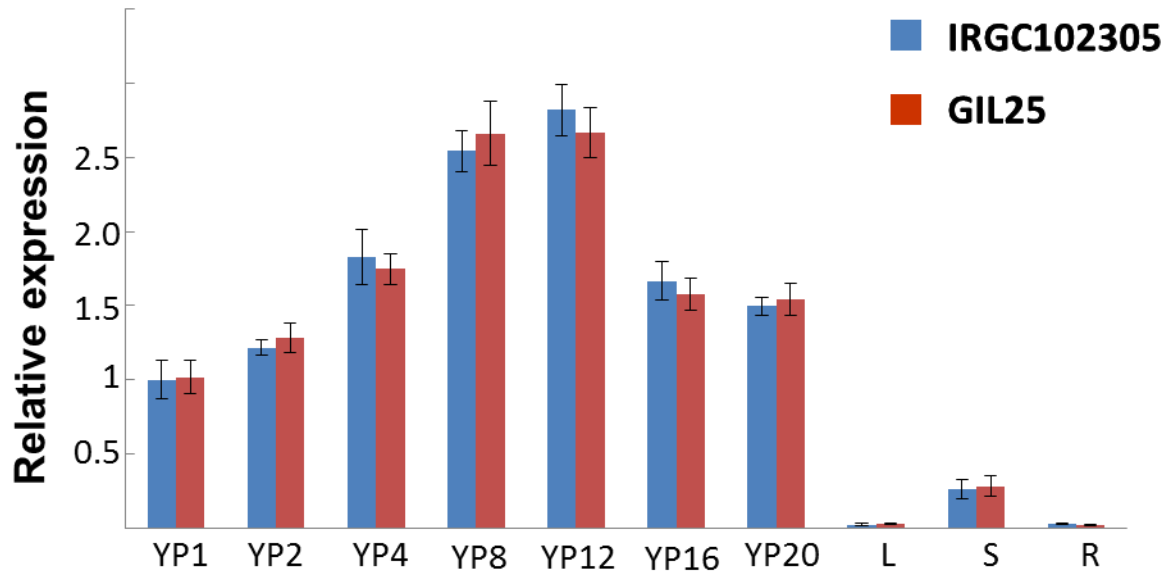
(b) Boxplot of Grain length in *O. barthii* (n=16) and *O. glabberima* (n=67).



Supplementary Figure 2. Agronomical traits Comparison of GIL25 and IRGC102305.

(a) Grain thickness. (b) Number of grains per main panicle. (c) Panicle length. (d) Number of tillers per plant. $n=10$ in a and $n=3$ in b-d.

All data are given as mean $\pm$ s.d. A Student's t -test was used.



Supplementary Figure 3. Expression patterns of the *GL4* gene.

Relative expression levels were detected by qPCR. Expression of *GL4* in different tissues and panicle development stages of GIL25 and IRGC102305. YP1, young panicle about 1 cm; YP2, young panicle about 2 cm; YP4, young panicle about 4 cm; YP8, young panicle about 8 cm; YP12, young panicle about 12 cm; YP16, young panicle about 16 cm; YP20, young panicle about 20 cm; L, leaf; S, stem; R, root.

Data are given as mean $\pm$ s.d.

GIL25	1	ATGTCGGGCTCCTCTGCTGCCGACCCCTCGCCATCCGCGTCGACCGGGGGCGGCGGCC	60
IRGC102305	1	ATGTCGGGCTCCTCTGCTGCCGACCCCTCGCCATCCGCGTCGACCGGGGGCGGCGGCC	60
GIL25	61	TCGCCGCTCGCGTGCTCCGCGCGCACGGGCACGGGCACCTACGCCGCCGTCGGCGGCG	120
IRGC102305	61	TCGCCGCTCGCGTGCTCCGCGCGCACGGGCACGGGCACCTACGCCGCCGTCGGCGGCG	120
GIL25	121	ACGGGGCCGGCGCCGCCGCCGCGTCGCCGGCGTCGGGGTCGGCGCCGCGGGACTACCGC	180
IRGC102305	121	ACGGGGCCGGCGCCGCCGCCGCGTCGCCGGCGTCGGGGTCGGCGCCGCGGGACTACCGC	180
GIL25	181	AAGGGGAAGTGGACGCTCCACGAGACGCTCATCCTCATCACCGCCAAGCGTCTCGACGAC	240
IRGC102305	181	AAGGGGAAGTGGACGCTCCACGAGACGCTCATCCTCATCACCGCCAAGCGTCTCGACGAC	240
GIL25	241	GACCGCCGCGCCGGCGTTGGGGGTGGCGCGGCGGCTGGTGGCGGTGGCGCCGGCTCGCCG	300
IRGC102305	241	GACCGCCGCGCCGGCGTTGGGGGTGGCGCGGCGGCTGGTGGCGGTGGCGCCGGCTCGCCG	300
		SNP1	
GIL25	301	CCGACGCCGAGGTCGGCGGAGCAGCGGTGGAAGTGGGTGGAGAACTACTGCTGGAAGAAC	360
IRGC102305	301	CCGACGCCGAGGTCGGCGGAGCAGCGGTGGAAGTGGGTGGAGAACTACTGCTGGAAGAAC	360
GIL25	361	GGCTGCCTCCGACGCCAGAACCAGTGCAATGACAAGTGGGACAACCTCCTCCGCGACTAC	420
IRGC102305	361	GGCTGCCTCCGACGCCAGAACCAGTGCAATGACAAGTGGGACAACCTCCTCCGCGACTAC	420
GIL25	421	AAGAAGGTCCGCGACTACGAGTCCCGCGTCGCCGCCGCCGCCGTCACCGGGCGCGCGGCC	480
IRGC102305	421	AAGAAGGTCCGCGACTACGAGTCCCGCGTCGCCGCCGCCGCCGTCACCGGGCGCGCGGCC	480
GIL25	481	GCCGCCAACTCCGCCCCCTCCCGTCGTAAGTGGACGATGGAGCGGCACGAGCGCAAGGAC	540
IRGC102305	481	GCCGCCAACTCCGCCCCCTCCCGTCGTAAGTGGACGATGGAGCGGCACGAGCGCAAGGAC	540
GIL25	541	TGCAACCTCCCCACCAACCTGGCGCCGGAGGTCTACGACGCGCTCTCCGAGGTGCTCTCC	600
IRGC102305	541	TGCAACCTCCCCACCAACCTGGCGCCGGAGGTCTACGACGCGCTCTCCGAGGTGCTCTCC	600
GIL25	601	CGCCGCGCGGCGCGACGCGGGCGGCGCCACGATCGCGCCACCCCGCGCCACCACCGCTC	660
IRGC102305	601	CGCCGCGCGGCGCGACGCGGGCGGCGCCACGATCGCGCCACCCCGCGCCACCACCGCTC	660
GIL25	661	GCGCTGCCGCCGCCGCCGCCGCCGCGCCCTCGCGCCGAAGCCTCTCGTCGCGCAG	720
IRGC102305	661	GCGCTGCCGCCGCCGCCGCCGCCGCGCCCTCGCGCCGAAGCCTCTCGTCGCGCAG	714
		Indel1	
GIL25	721	CAGCACCACCACCATCACGGCCATCACCACCACCCACCTCCTCCTCAGCCGCCCGCTCG	780
IRGC102305	715	CAGCACCACCACCATCACGGCCATCACCACCACCCACCTCCTCCTCAGCCGCCCGCTCG	774
		SNP2	
GIL25	781	TCGCTGCAGTCCCTCCGGCGGTCTGGCTCCGCCGCCGGCGTCCGTTTCCGGTAATGGT	840
IRGC102305	775	TCGCTGCAGTCCCTCCGGCGGTCTGGCTCCGCCGCCGGCGTCCGTTTCCGGTAATGGT	834
GIL25	841	CGGTGCGCGCACCGTACACTTAATACTCGTAGTAGCTGTTACATCCCTCCCTCCAAAC	900
IRGC102305	835	CGGTGCGCGCACCGTACACTTAATACTCGTAGTAGCTGTTACATCCCTCCCTCCAAAC	894
GIL25	901	CATTACTACTGTTCTCTCACAACGACATGTGGGGCCACCTCGCAGCGAGCTGAGCTCC	960
IRGC102305	895	CATTACTACTGTTCTCTCACAACGACATGTGGGGCCACCTCGCAGCGAGCTGAGCTCC	954

GIL25	961	GCCACTATACGTTATTAAAAAGCCCGCGTTATGATTGGGCTAGTTACGTTGTTGAGTTGAG	1020
IRGC102305	955	GCCACTATACGTTATTAAAAAGCCCGCGTTATGATTGGGCTAGTTACGTTGTTGAGTTGAG	1014
GIL25	1021	CTGGTCGTAATTATTTACTACCGCTACTTTTTTTTTTACCTTTTTTACCGTGGGGTTCGGGA	1080
IRGC102305	1015	CTGGTCGTAATTATTTACTACCGCTACTTTTTTTTTTACCTTTTTTACCGTGGGGTTCGGGA	1074
GIL25	1081	GAGGGTGGTCGCGGTAATAATAATGTCCTCAACTCAGGGGTGGGAGAATAAAGCTGCGT	1140
IRGC102305	1075	GAGGGTGGTCGCGGTAATAATAATGTCCTCAACTCAGGGGTGGGAGAATAAAGCTGCGT	1134
GIL25	1141	GCAGTGTGGTGCAGTTCATGCATGGGAAAGGTGATGCGAATCCGGATATTTTATGGGGGT	1200
IRGC102305	1135	GCAGTGTGGTGCAGTTCATGCATGGGAAAGGTGATGCGAATCCGGATATTTTATGGGGGT	1194
GIL25	1201	TTAATTGAAAGATTTACTCCACGACGATACTACCCTGTACTCCTGCCATGCTGCAAGCAT	1260
IRGC102305	1195	TTAATTGAAAGATTTACTCCACGACGATACTACCCTGTACTCCTGCCATGCTGCAAGCAT	1254
GIL25	1261	GCGTAATGCGTTACATTGCGAAATCACTCGCTTTGAAAGAAAAAAAAGCCTAAAACTTTG	1320
IRGC102305	1255	GCGTAATGCGTTACATTGCGAAATCACTCGCTTTGAAAGAAAAAAAAGCCTAAAACTTTG	1314
GIL25	1321	GAGCAAAAAAAGCACCTTTTGTCTCTCCTCGTGCATGCATGCCGCGCTGCCTATCTT	1380
IRGC102305	1315	GAGCAAAAAAAGCACCTTTTGTCTCTCCTCGTGCATGCATGCCGCGCTGCCTATCTT	1374
GIL25	1381	GAAGTACTTTGGACTTTTGTATCGATCAGCAAACTATACCTATATTAGCAGTAATTAAT	1440
IRGC102305	1375	GAAGTACTTTGGACTTTTGTATCGATCAGCAAACTATACCTATATTAGCAGTAATTAAT	1434
GIL25	1441	ACTACATTTGTAGATATCCTTTGACCGTTCTATCTTATTTTGTATAATAAAAAATTAG	1500
IRGC102305	1435	ACTACATTTGTAGATATCCTTTGACCGTTCTATCTTATTTTGTATAATAAAAAATTAG	1494
SNP3			
GIL25	1501	TTACATTTAAAAATACCTATTATATTTATCATCTAATAACAATAAAAGTATTAGTTAAA	1560
IRGC102305	1495	TTACATTTAAAAATGCTATTATATTTATCATCTAATAACAATAAAAGTATTAGTTAAA	1554
SNP4			
GIL25	1561	TATTAAACGTTGGATATGAATAGTTTAAAACTGTATTGTTTTGGGGCGGAGGGAGTAATT	1620
IRGC102305	1555	TATTAAACGTTGGATATGAATAGTTTAAAACTGTATTGTTTTGGGGCGGAGGGAGTAATT	1614
GIL25	1621	GATTGATGGATTAATTAAGTGGTTGACTAATGTGTGTGTGATTATTTGTGTAGCGGAGG	1680
IRGC102305	1615	GATTGATGGATTAATTAAGTGGTTGACTAATGTGTGTGTGATTATTTGTGTAGCGGAGG	1674
GIL25	1681	AGGAGATGTCGGGGTCGTCGGAGTCGGGGGAGGAGGAGGGGTGGGGCGGGGAGCCGG	1740
IRGC102305	1675	AGGAGATGTCGGGGTCGTCGGAGTCGGGGGAGGAGGAGGGGTGGGGCGGGGAGCCGG	1734
GIL25	1741	AGGCGAAGCGGCGGCGGCTGAGCCGGCTGGGGTCGAGCGTGGTGAGGAGCGCGACGGTGG	1800
IRGC102305	1735	AGGCGAAGCGGCGGCGGCTGAGCCGGCTGGGGTCGAGCGTGGTGAGGAGCGCGACGGTGG	1794
GIL25	1801	TGGCGAGGACGCTGGTGGCGTGCGAGGAGAAGCGGGAGCGCCGGCACCAGGAGCTGCTGC	1860
IRGC102305	1795	TGGCGAGGACGCTGGTGGCGTGCGAGGAGAAGCGGGAGCGCCGGCACCAGGAGCTGCTGC	1854

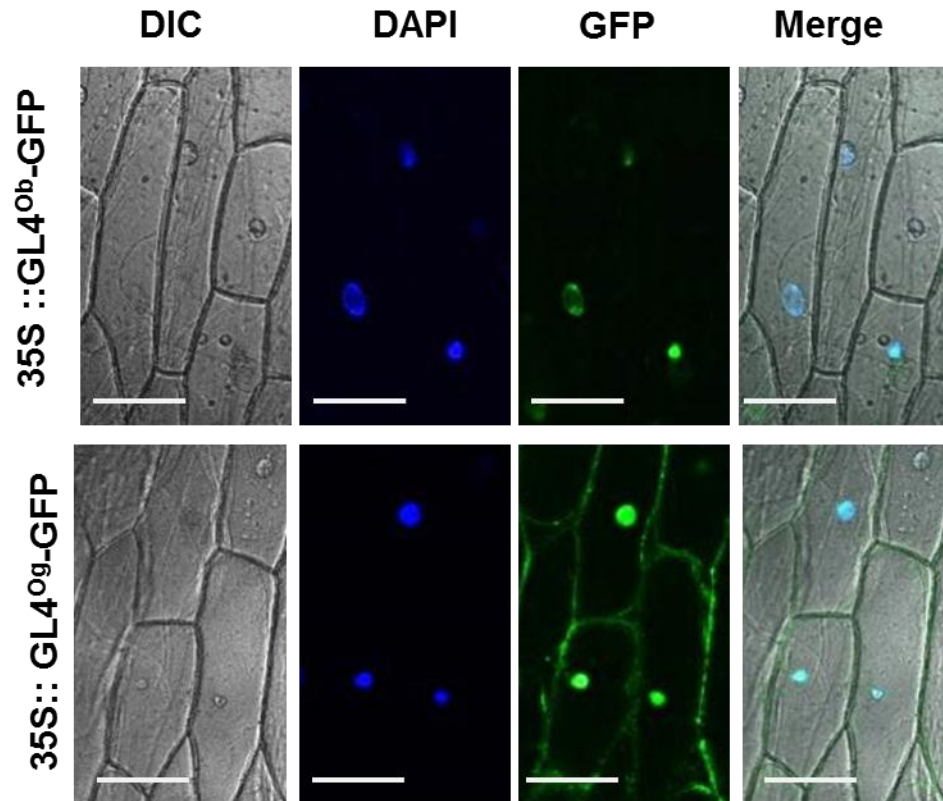
GIL25	1861	AGCTGGAGGAGCGGCGGCTGCGCCTCGAGGAGGAGCGCACCGAGGTCCGCCGCCAGGGCT	1920
IRGC102305	1855	AGCTGGAGGAGCGGCGGCTGCGCCTCGAGGAGGAGCGCACCGAGGTCCGCCGCCAGGGCT	1914
GIL25	1921	TCGCCGGCCTCATCGCCGCCGTCAACAGCCTCTCCTCCGCCATCCAGCCCCTCGTCTCCG	1980
IRGC102305	1915	TCGCCGGCCTCATCGCCGCCGTCAACAGCCTCTCCTCCGCCATCCAGCCCCTCGTCTCCG	1974
GIL25	1981	ACCACCGCAGCGGCGACTCCTCCGGCCGATGA	2012
IRGC102305	1975	ACCACCGCAGCGGCGACTCCTCCGGCCGATGA	2006

Supplementary Figure 4. The variations of nucleotide sequence in the *GL4* gene coding region between GIL25 and IRGC102305.

GIL25	MSGSSAADPSPSASTAGAAASPLALLRAHGHGHLTPPSAATGPAPPPSPASGSAPRDYRKG	62
IRGC102305	MSGSSAADPSPSASTAGAAASPLALLRAHGHGHLTPPSAATGPAPPPSPASGSAPRDYRKG	62
GIL25	NWTLHETLILITAKRLDDRRAGVGGGAAAGGGGAGSPPTPRSAEQRWKWVENYCWKNGCLR	124
IRGC102305	NWTLHETLILITAKRLDDRRAGVGGGAAAGGGGAGSPPTPRSAEQRWKWVENYCWKNGCLR	124
GIL25	SQNCNDKWDNLLRDYKKVRDYESRVAAAAVTGAAAAANSAPLPSYWTMERHERKDCNLPTN	186
IRGC102305	SQNCNDKWDNLLRDYKKVRDYESRVAAAAVTGAAAAANSAPLPSYWTMERHERKDCNLPTN	186
GIL25	LAPEVYDALSEVLSRRAARRGGATIAPTPPPPLALPPPPPPPPSPPKPLVAQQHHHHHGH	248
IRGC102305	LAPEVYDALSEVLSRRAARRGGATIAPTPPPPLALPPPPPP- -SPPKPLVAQQHHHHHGH	246
GIL25	HHHPPPPQPPSSLQLPPAVVAPPPASVSAEEEMSGSSESIEEEEGSGGEPEAKRRRLSRLG	310
IRGC102305	HHHPPPP*	253
GIL25	SSVVRSATVVARTLVACEEKRERRHRELLQLEERRLRLEEERTEVRRQGFAGLIAAVNSLSS	372
IRGC102305		
GIL25	AIHALVSDHRSGDSSGR*	389
IRGC102305		

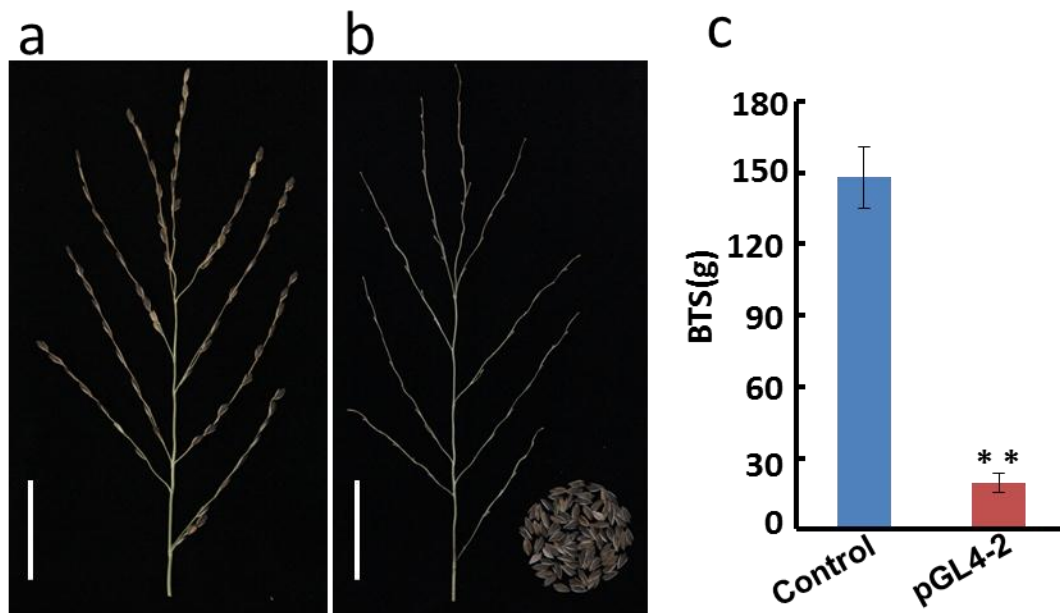
Supplementary Figure 5. Comparison of the deduced amino acid sequence of GL4 between IRGC102305 and GIL25.

The numbers on the right side indicate the position of the residues in the proteins. The predicted NIL was written in red color.



Supplementary Figure 6. Subcellular localization of ObGL4 and OgGL4.

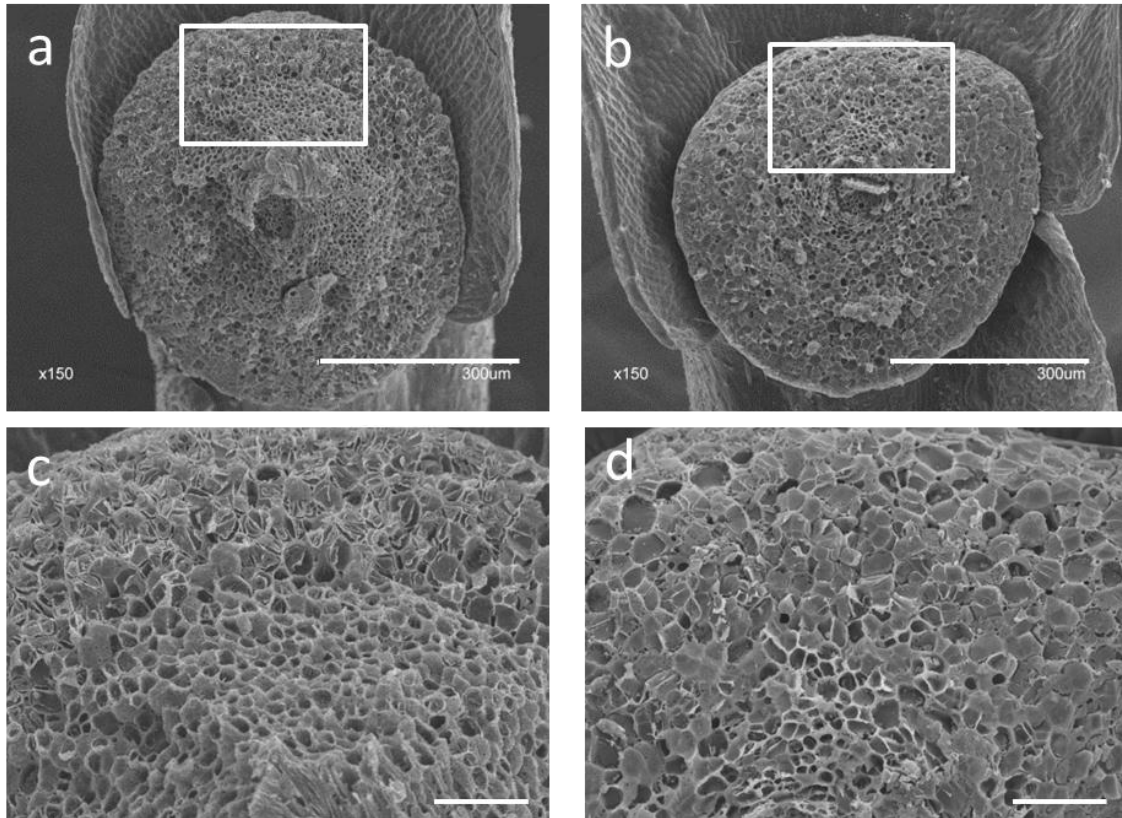
The *ObGL4* -GFP and *OgGL4* -GFP fusion gene under the control of the *CaMV* 35S promoter were expressed transiently in onion epidermal cells. Bars = 50 μ m.



Supplementary Figure 7. Comparison of the seed shattering between control and the transgenic plant.

(a) The main panicle of Control.(b) The main panicle of the transgenic plant with pGL4-2.Bars = 5 cm.

(c) Force required to pull grains away from pedicels on 35 day of flower. n=3. Data are given as mean $\pm$ s.d. **P < 0.01. A Student's *t*-test was used.

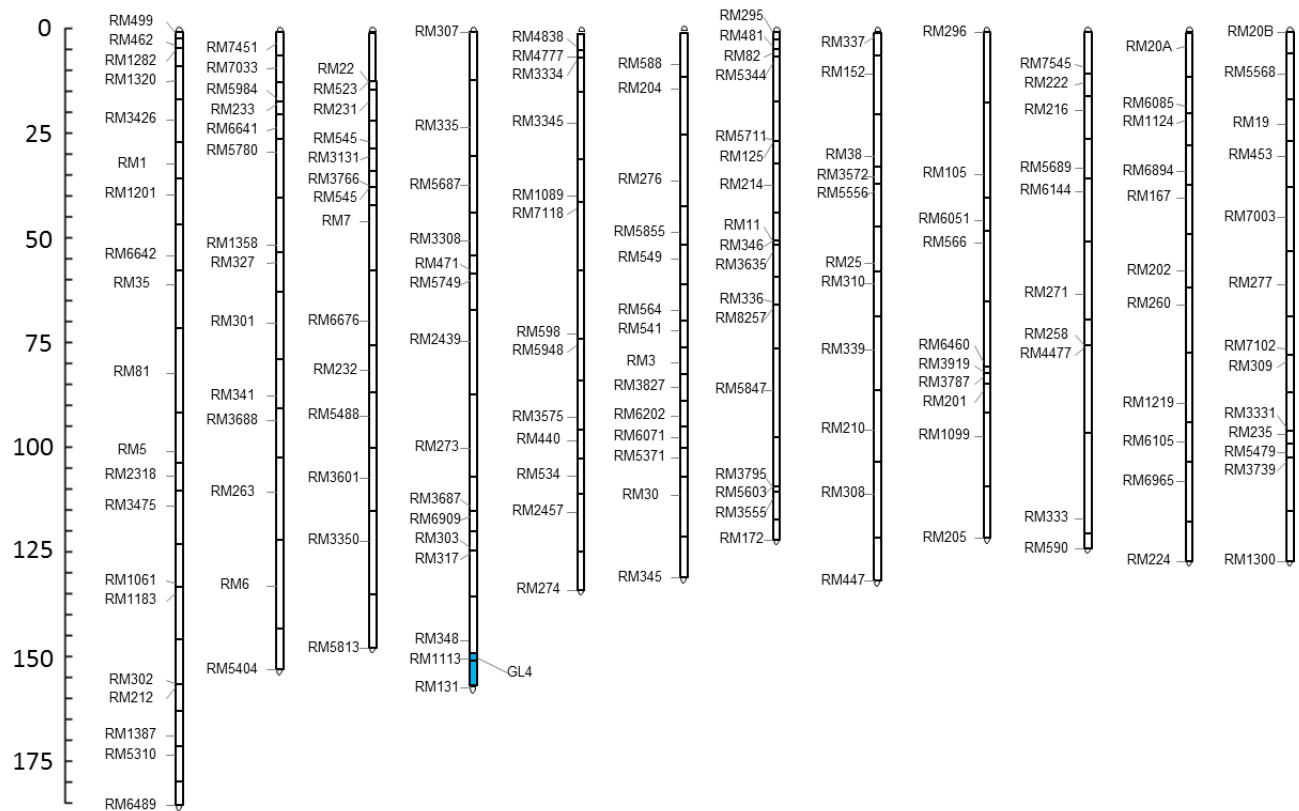


Supplementary Figure 8. Comparison of the abscission lays between control and the transgenic plant.

The scanning electron microscope (SEM) photo of pedicel junction after detachment of seeds.

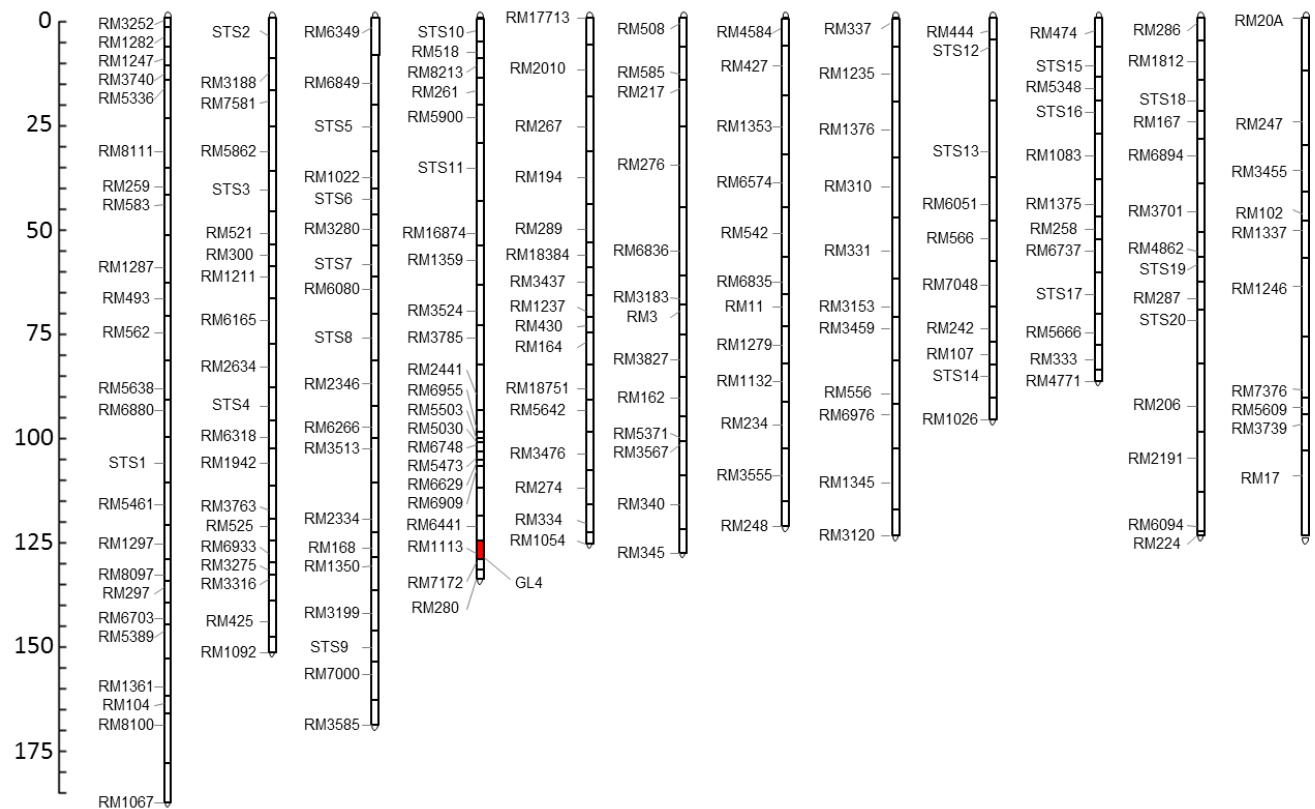
(a) Control; (b) pGL4. Bars=300 μm.

(c) and (d) magnified from the boxed region in (a) and (b). Bars = 50 μm.



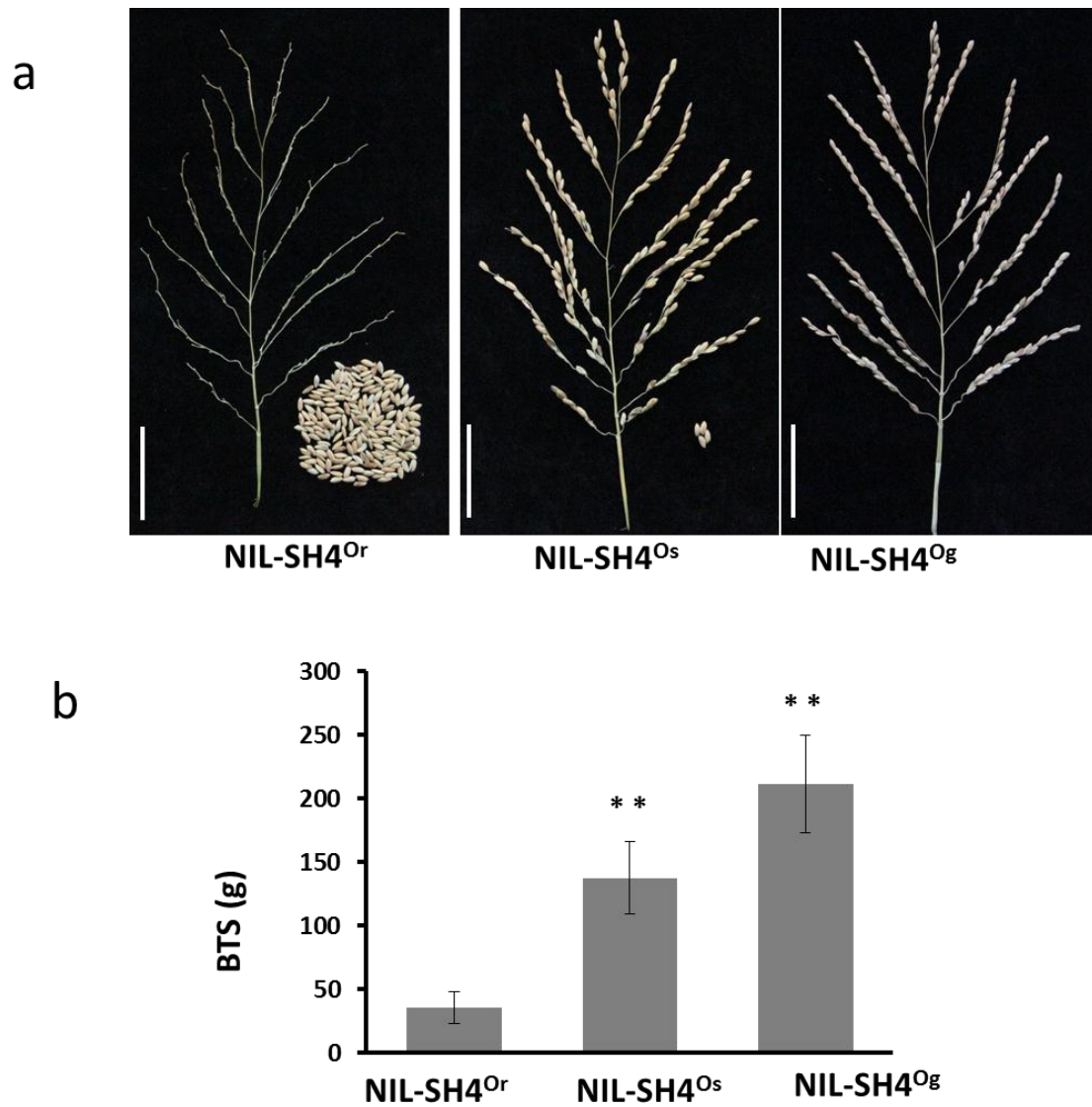
Supplementary Figure 9. Graphical genotypes of NIL-GL4^{Or}.

The blue regions indicate the homozygous regions for Yuanjiang common wild rice (*O. rufipogon*). The white regions indicate the homozygous regions for Teqing (*O. sativa* ssp. *indica*).



Supplementary Figure 10. Graphical genotypes of NIL-GL4^{Og}.

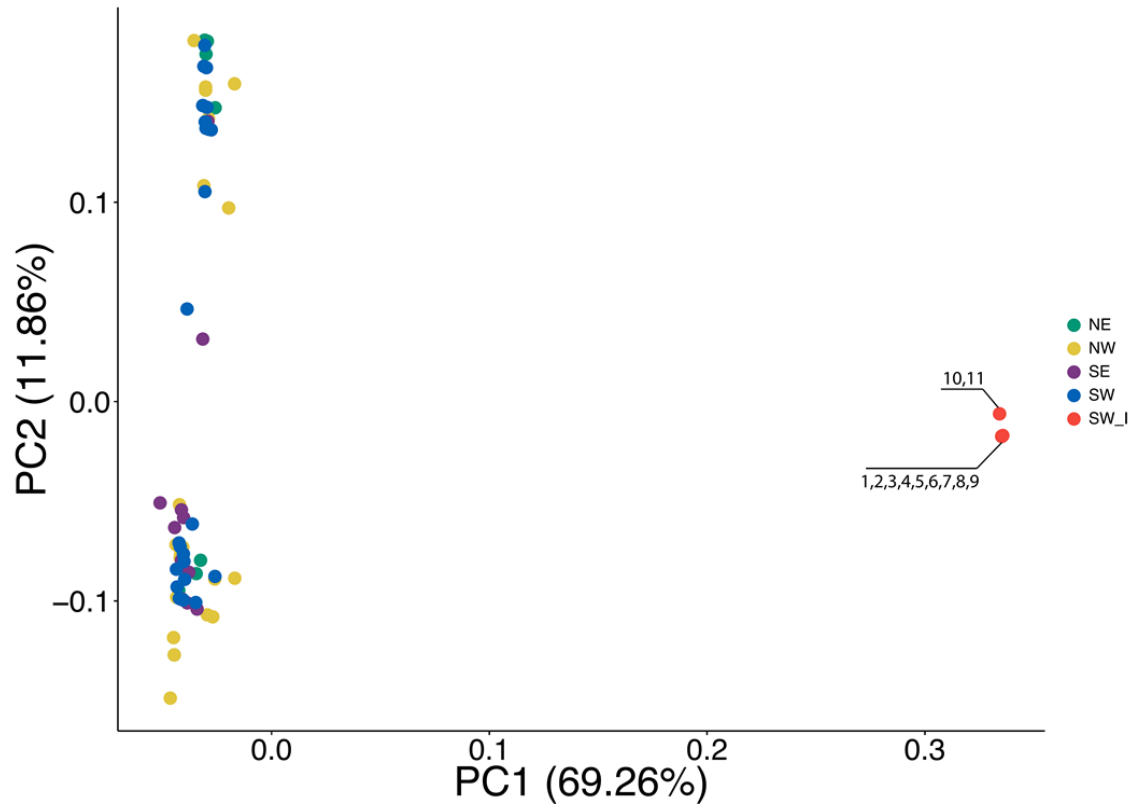
The red regions indicate the homozygous regions for (*O. glaberrima*). The white regions indicate the homozygous regions for Teqing (*O. sativa* ssp. *indica*).



Supplementary Figure 11. Comparison of the phenotype of seed shattering.

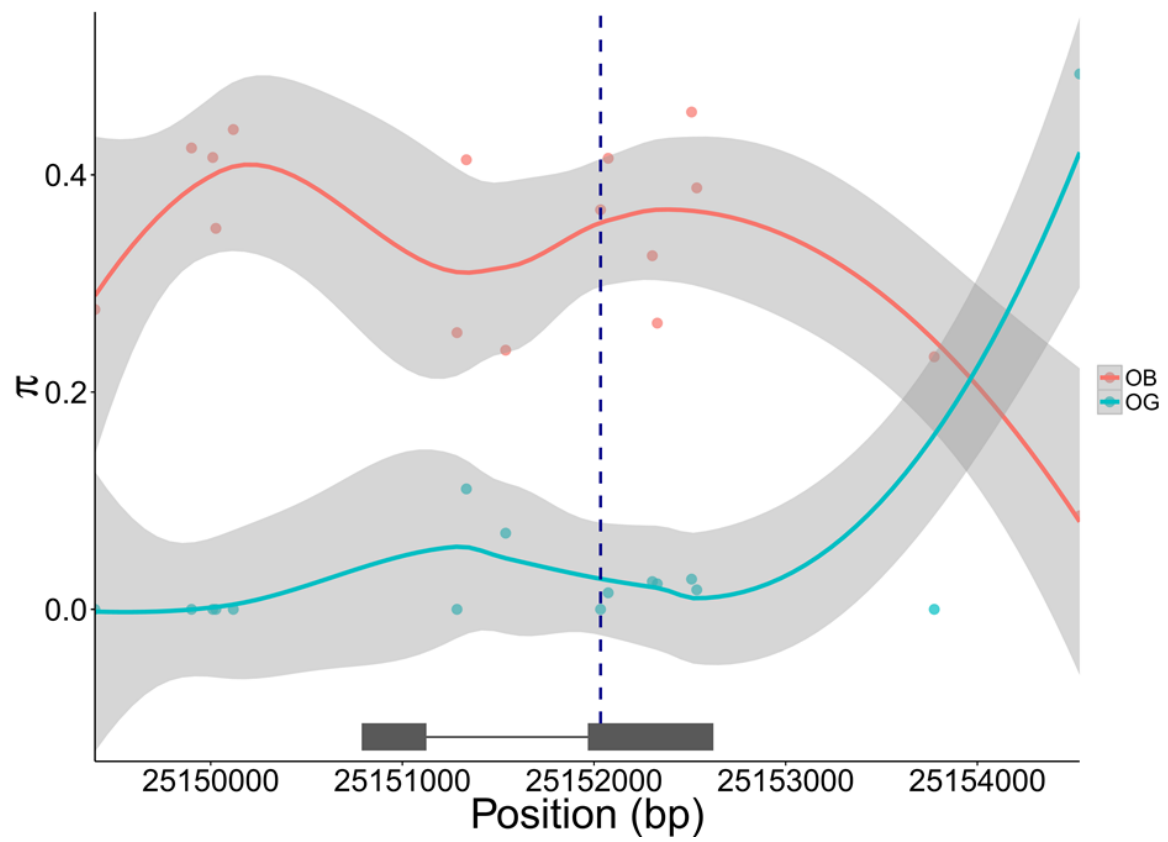
(a) Shattering Phenotype Comparison of three NIL-SH4 lines.

(b) Force required to pull grains away from pedicels on 35 day of flower. n=3. Data are given as mean $\pm$ s.d. **P < 0.01. A Student's *t*-test was used.

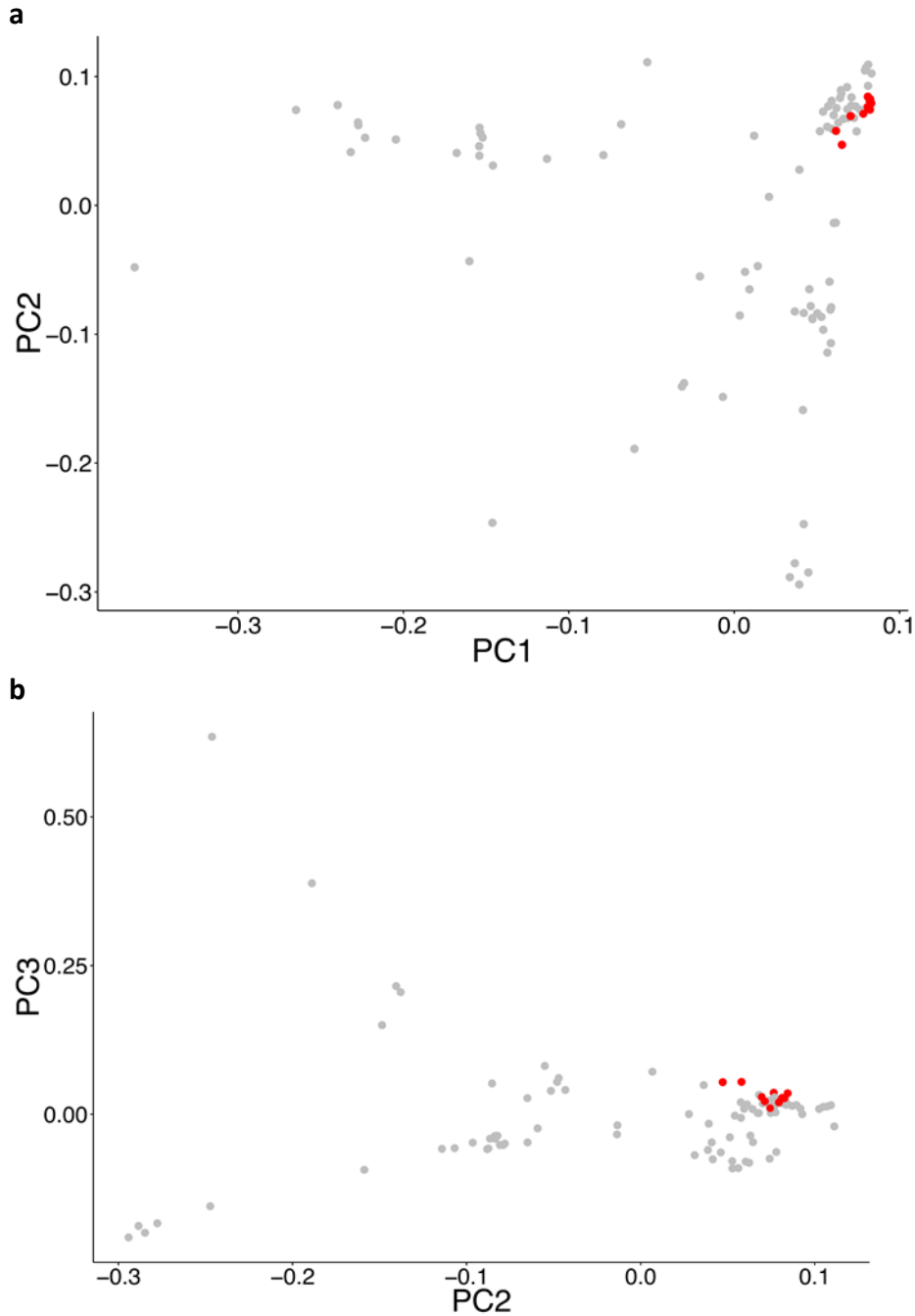


Supplementary Figure 12. Principal Component Analysis (PCA) of genetic variations 5kb upstream and downstream of *GL4*.

As *O. glaberrima* accessions carrying C at SNP2 are all from Southwestern subgroup and clustered together, they are named as SW_I subgroup.



Supplementary Figure 13. Comparison of nucleotide diversity (π) in *O. glaberrima* accessions carrying T at SNP2 and *O. barthii* accessions.
 SNP2 is indicated as dotted blue line.



Supplementary Figure 14. Principal Component Analysis (PCA) of genome-wide genetic variations of *O. glaberrima*.

(a) The first principal component (PC1) and the second principal component (PC2) are plotted. *O. glaberrima* accessions carrying C at SNP2 are colored in red.

(b) The second principal component (PC2) and the third principal component (PC3) are plotted. *O. glaberrima* accessions carrying C at SNP2 are colored in red.

Supplementary Table 1. Primers used in this study

Primer name	Forward primer sequence	Reverse primer sequence	Purpose
RM3740	ATCCCAACTCTAAGCCACCC	CTACCCGTCACCAACTCACC	Genotype analysis
RM283	GTCTACATGTACCCTTGTTGGG	CGGCATGAGAGTCTGTGATG	Genotype analysis
RM243	GATCTGCAGACTGCAGTTGC	AGCTGCAACGATGTTGTCC	Genotype analysis
RM140	TGCCTCTTCCCTGGCTCCCTG	GGCATGCCGAATGAAATGCATG	Genotype analysis
RM594	GCCACCAGTAAAAGCAATAC	TTGATCTGCTAGTGAGACCC	Genotype analysis
RM11113	GTCTTATCTCATTTAGTCCCGTTCC	TGGAGTAGGGTTAGGGTTTGTGG	Genotype analysis
RM9	GGTGCCATTGTCGTCTC	ACGGCCCTCATCACCTTC	Genotype analysis
RM246	GAGCTCCATCAGCCATTGAG	CTGAGTGCTGCTGCGACT	Genotype analysis
RM297	TCTTTGGAGGCGAGCTGAG	CGAAGGGTACATCTGCTTAG	Genotype analysis
RM212	CCACTTTCAGCTACTACCAG	CACCCATTTGTCTCTCATTATG	Genotype analysis
RM1339	ATCAAAGCATGTAAACCAGC	CGTAAGATCTCCCTACCACC	Genotype analysis
RM431	TCCTGCGAACTGAAGAGTTG	AGAGCAAAACCCTGGTTCAC	Genotype analysis
RM5310	TAGACAAAGCAACGGGTTCC	CGGAAGCAGGAGAATCGTAG	Genotype analysis
RM2770	TAGGCCCTGATTAGTTTCC	ATATATGTGTCCCTTCTCCATAC	Genotype analysis
RM279	GCGGGAGAGGGATCTCCT	GGCTAGGAGTTAACCTCGCG	Genotype analysis
RM555	TTGGATCAGCCAAAGGAGAC	CAGCATTGTGGCATGGATAC	Genotype analysis
RM5699	ATCGTTTCGCATATGTTT	ATCGGTAAAAGATGAGCC	Genotype analysis
RM5210	CTAAATGACAAAAGATAGCA	AAATTCTGACTTGTATGACA	Genotype analysis
RM6374	TGAGGACGCTGATTGTCAAC	GCTGCCCTATTATTTCACC	Genotype analysis
RM4499	AGCAACTTGCAAGCTTTAAT	GCTGAACCCTGAGAATATGT	Genotype analysis
RM1920	CAAACACAGTGTTGACAGAA	GCTATTGACTTATCCGTTCA	Genotype analysis
RM5631	CGTCCAAGAAATATTGCAGT	GTGAGACAGAATCCTTACGC	Genotype analysis
RM3316	TTCGACGATTCTGTACACGC	CATGATCCCAAATGCATGGG	Genotype analysis
RM3774	TTTGACCTCAACTCCCAAGC	GCTCCCTCACTCACAACCTCC	Genotype analysis
RM266	TAGTTTAACCAAGACTCTC	GGTGAACCCAAATCTGCA	Genotype analysis
RM60	AGTCCCATGTTCCAATTCCG	ATGGCTACTGCCTGTACTAC	Genotype analysis
RM81A	GAGTGCTTGTGCAAGATCCA	CTTCTTCACTCATGCAGTTC	Genotype analysis
RM433	TGCGCTGAACTAAACACAGC	AGACAAACCTGGCCATTAC	Genotype analysis
RM7	TTCGCCATGAAGTCTCTCG	CCTCCCATCATTTTCGTTGTT	Genotype analysis
RM5748	CAGTTGGCAATTGTCACGAG	TCGAACATATCCAAGCCTCC	Genotype analysis
RM6676	GTTACGGTCCAATAAGAAT	CTTTCAAGCTTACGAAAACA	Genotype analysis
RM411	ACACCAACTCTTGCCTGCAT	TGAAGCAAAAACATGGCTAGG	Genotype analysis
RM3525	AACTCTCAGCTCATCAAGACC	GGGCAAGTGGTCAAATCTTG	Genotype analysis
RM571	GGAGGTGAAAGCGAATCATG	CCTGCTGCTCTTTCATCAGC	Genotype analysis
RM85	CCAAAGATGAAACCTGGATTG	GCACAAGGTGAGCAGTCC	Genotype analysis
RM551	AGCCCAGACTAGCATGATTG	GAAGGCGAGAAGGATCACAG	Genotype analysis
RM261	CTACTTCTCCCCTTGTGTCG	TGTACCATCGCCAAATCTCC	Genotype analysis
RM7181	GTGAGAGTTCGTAGAAAGTGCC	GGATAAGCAAGATCCGTGAC	Genotype analysis

RM471	ACGCACAAGCAGATGATGAG	GGGAGAAGACGAATGTTTGC	Genotype analysis
RM1223	CAGCGTCTCCAAGAACTCC	GCTACCAGGTCAGAGTTGCC	Genotype analysis
RM3735	GCGACCGATCAGCTAGCTAG	ATAACTCCTCCCTTGCTGCC	Genotype analysis
RM3335	TAATCCACTGTGTCAATTTAA	ACCATCATCTTGTACCTAGT	Genotype analysis
RM3648	TACCCTTTCTTCCCCAAACC	ACCTCCTCCTCCACTTCTCC	Genotype analysis
RM5478	ATCCAATGCGATGCTACTCC	CATCACGAGACCACGACAAG	Genotype analysis
RM5506	AGGCGATGTTTGATCTCGAC	CTGGACGTACACACACGTACG	Genotype analysis
RM567	ATCAGGGAAATCCTGAAGGG	GGAAGGAGCAATCACCCTG	Genotype analysis
RM3332	ATTCTCCTCGCTTCCTCCTC	AAAGAGAGAGCCGAGCACTG	Genotype analysis
RM13	TCCAACATGGCAAGAGAGAG	GGTGGCATTCGATTCCAG	Genotype analysis
RM267	TGCAGACATAGAGAAGGAAGTG	AGCAACAGCACAACCTTGATG	Genotype analysis
RM169	TGGCTGGCTCCGTGGGTAGCTG	TCCCGTTGCCGTTTCATCCCTCC	Genotype analysis
RM18329	GCGTATGGATGGAAACATGTACG	CCACCCAGCCCTGTAGATCG	Genotype analysis
RM509	TAGTGAGGGAGTGGAACGG	ATCGTCCCCACAATCTCATC	Genotype analysis
RM163	ATCCATGTGCGCCTTTATGAGGA	CGCTACCTCCTTCACTTACTAGT	Genotype analysis
RM3663	CATCAACCTCCACGAACATG	CTCGGTGGTGATCCTCCTC	Genotype analysis
RM274	CCTCGCTTATGAGAGCTTCG	CTTCTCCATCACTCCCATGG	Genotype analysis
RM589	ATCATGGTCGGTGGCTTAAC	CAGGTTCCAACCAGACACTG	Genotype analysis
RM585	CAGTCTTGCTCCGTTTGTTG	CTGTGACTGACTTGGTCATAGG	Genotype analysis
RM5531	TTTGTGTTGGTAAGTTGCTTC	TTAAGGAGAGTGTTTTCTTTCTC	Genotype analysis
RM1130	AGATCGGATTGGGATGGC	ACCCAACCAATTAGTGCCAC	Genotype analysis
RM400	ACACCAGGCTACCCAAACTC	CGGAGAGATCTGACATGTGG	Genotype analysis
RM1210	TTTTTACCCAGTGGAGCAG	CTTATGTGCACTGCAGATGC	Genotype analysis
RM427	TCACTAGCTCTGCCCTGACC	TGATGAGAGTTGGTTGCGAG	Genotype analysis
RM125	ATCAGCAGCCATGGCAGCGACC	AGGGGATCATGTGCCGAAGGCC	Genotype analysis
RM500	GAGCTTGCCAGAGTGGAAG	GTTACACCGAGAGCCAGCTC	Genotype analysis
RM418	TCGCGTATCGTCATGCATAG	GAGCACATATGCCACGTACG	Genotype analysis
RM1279	GGGTATAAAATGCGTGGCAC	ATGGATGGTACGAGGACGAG	Genotype analysis
RM455	AACAACCCACCACCTGTCTC	AGAAGGAAAAGGGCTCGATC	Genotype analysis
RM234	ACAGTATCCAAGGCCCTGG	CACGTGAGACAAAGACGGAG	Genotype analysis
RM429	TCCCTCCAGCAATGTCTTTC	CCTTCATCTTGCTTTCCACC	Genotype analysis
RM420	GGACAGAATGTGAAGACAGTCG	ACTAATCCACCAACGCATCC	Genotype analysis
RM337	GTAGGAAAGGAAGGGCAGAG	CGATAGATAGCTAGATGTGGCC	Genotype analysis
RM2420	AACATAACTTTGACTTCTTG	GATTTGTCTTGAAAAATACT	Genotype analysis
RM32	AGTCTACGTGGTGTACACGTGG	TGCGGCCTGCCGTTTGTGAG	Genotype analysis
RM331	GAACCAGAGGACAAAAATGC	CATCATACATTTGCAGCCAG	Genotype analysis
RM404	CCAATCATTAACCCCTGAGC	GCCTTCATGCTTCAGAAGAC	Genotype analysis
RM4668	GGAGCACCAAACAAATTTAC	CAGTCGTCCCACTTCTTG	Genotype analysis
RM5485	CTTCCACAAGCTTGGCTAGG	AATGCCATCCCCTACTCATG	Genotype analysis
RM477	TCTCGCGGTATAGTTTGTGC	ACCACTACCAGCAGCCTCTG	Genotype analysis
RM23652	GCCTCTCCTCTCCTTCACTCC	CGTTAGGCTTCGGATCTTGTGC	Genotype analysis

RM5688	GCAGTGTCCAACCATCTGTG	ATCTGGTCACCCTTTGCTTG	Genotype analysis
RM105	GTCGTCGACCCATCGGAGCCAC	TGGTCGAGGTGGGGATCGGGTC	Genotype analysis
RM5526	TCAGCCTGGCCTCTCTTATC	ATGATCCTCCACCCACTAGC	Genotype analysis
RM460	TGATCGACAGCGTTCTTGAC	GCCTGGCCCACATAATTAAG	Genotype analysis
RM434	GCCTCATCCCTCTAACCCTC	CAAGAAAGATCAGTGCGTGG	Genotype analysis
RM201	CTCGTTTATTACCTACAGTACC	CTACCTCCTTTCTAGACCGATA	Genotype analysis
RM7217	TTTGTAGGATGACACGTGGC	CGGGATTTTCAGTACCTCACG	Genotype analysis
RM2125	TACCTCCTAGCTTTACTTAT	ACTGATCTCTATCTCATTGT	Genotype analysis
RM311	TGGTAGTATAGGTAATAACAT	TCCTATACACATACAAACATAC	Genotype analysis
RM5758	GAGGCCCTGATATTCAATGG	TATGGCTTAGCGTTAGACCG	Genotype analysis
RM467	GGTCTCTCTCTCTCTCTCTCTC	CTCCTGACAATTCAACTGCG	Genotype analysis
RM7300	TCCGTATCCTAGTCGCGATC	CGCCGTCATGACTCATACTC	Genotype analysis
RM7093	CCTCTCTTTGCCTAATTGCC	GTAGGACTGGAATTGGCACC	Genotype analysis
RM7203	ACGTGGTGCCTTCTTTCAAG	CTCTGGCCTTCTACTCATGG	Genotype analysis
RM3137	GTTAGGAATCCATGCTGCG	TGCCCGTGCTCGATAAGG	Genotype analysis
RM287	TTCCCTGTAAAGAGAGAAATC	GTGTATTTGGTGAAAGCAAC	Genotype analysis
RM229	CACTCACACGAACGACTGAC	CGCAGGTTCTTGTGAAATGT	Genotype analysis
RM206	CCCATGCGTTTAACTATTCT	CGTTCCATCGATCCGTATGG	Genotype analysis
RM7654	GTGTCGTGGTCGTAACCTG	TAAGAGACGGAAGAGTGAGC	Genotype analysis
RM7443	TGCTGCGTGTTACTTTGGTG	AACCCTTCATCAGGCTACGC	Genotype analysis
RM20A	ATCTTGTCCTGCAGGTCAT	GAAACAGAGGCACATTTTCATTG	Genotype analysis
RM247	TAGTGCCGATCGATGTAACG	CATATGGTTTTTGACAAAGCG	Genotype analysis
RM3455	TGAATCCACACTCGCAGATC	GCCAGTCCACGATTGGTC	Genotype analysis
RM7119	AGGCTGAGGCTTATAGGCAG	GGATGATACAACCTTGACCCC	Genotype analysis
RM2935	CAGCAAATTTGTTACTTATG	TGCTATGTTTTTTTATAACG	Genotype analysis
RM1261	GTCCATGCCCCAAGACACAAC	GTTACATCATGGGTGACCCC	Genotype analysis
RM1986	TAACGGAGGGAGTAGTTTTTC	GAACCTACATATCGAGAGCA	Genotype analysis
RM6306	CACCGGTCTAAGTCGACTCC	CCACTCGTTGTCGTCGTATG	Genotype analysis
RM3335	TAATCCACTGTGTCATTTAA	ACCATCATCTTGTACCTAGT	Genotype analysis
RM3648	TACCCTTTCTTCCCCAAACC	ACCTCCTCCTCCACTTCTCC	Genotype analysis
RM5478	ATCCAATGCGATGCTACTCC	CATCACGAGACCACGACAAG	Genotype analysis
RM5506	AGGCGATGTTTGATCTCGAC	CTGGACGTACACACACGTACG	Genotype analysis
RM6441	CGAAAGGTCGGCATCTTTC	ATGGCAATGATACGGAGGAG	Genotype analysis
RM567	ATCAGGGAAATCCTGAAGGG	GGAAGGAGCAATCACCCTG	Genotype analysis
RM3332	ATTCTCCTCGCTTCCTCCTC	AAAGAGAGAGCCGAGCACTG	Genotype analysis
RM470	CCCTCCCGTAGACCTTGTACCC	CCACAGCTAACCAATCCTTCTCC	Genotype analysis
RM255	GAGGAGGAGGAGGAGAGATCAGG	AACGAAACCGCTCAGTTCAACC	Genotype analysis
RM5608	AGAGAGCCCTGCTGCATTTGTCG	TGGGTCTTTTCGTTCTGTCCTACCC	Genotype analysis
STS1	CATCTCAAAGTCGTGAGTGCAAC	ATACGTGACCCCTCCGAGCT	Genotype analysis
STS2	ATGTGTGCTTCGGTCGTGAT	TTCTCACATCTGAACCTCTCC	Genotype analysis
STS3	TGCTGATGTTAATCTGTGGGTG	TGGCGGCTTGAGAGTGTTTGTAG	Genotype analysis

STS4	ATCCTGATCTCTCCTGTGCC	AGATGTAAACCCAATAGGACC	Genotype analysis
STS5	CAGCTGAGAAGCCATCC	CACATGACAGCCAGTCGG	Genotype analysis
STS6	TCCTGTGTTTCGCTGCTCGTC	TTCGTGGTCGCTGAAAAGTCC	Genotype analysis
STS7	GCGTCGTAATGTAATGGCTGG	TTGTTTGACCTCACACACTCTGC	Genotype analysis
STS8	CACCGTTTGGTTTGGCAG	AGCATGGTGTACCTCGTCG	Genotype analysis
STS9	GCAAGACGAGGCAAGGCTG	AGTTGTACTCTGGTTGCTTCCGC	Genotype analysis
STS10	AGAATAGAGTGCATCATCGTC	AACCTGATAGGTGGAAGATGTAC	Genotype analysis
STS11	TATGCGAAGGATGTGCGAC	ACGAATACATGTGCCTGCC	Genotype analysis
STS12	CCTGGTTAGCACTACAGCTC	TAGTTGGCTATGTCCACACC	Genotype analysis
STS13	AAGTCGAAGGAGGAGTTGTC	CACCGAACTAAAGACGAG	Genotype analysis
STS14	TGACTCAAGGCATATGCAGC	TTGTCCATGCTGTTAGCCTG	Genotype analysis
STS15	CTTTGTAGACCGTATCATTGTCC	GAAACGTCTGGCGAGTTCC	Genotype analysis
STS16	GCAGGAGTATATACGCAGGG	ACTCAGCGTGCTCAACCTG	Genotype analysis
STS17	TTAGGCAAACGTGATTGGC	TACCGTTTCCATCCTTCGC	Genotype analysis
STS18	CCAACTCCTAATTATGCTGAAGC	GAGTGGTAGTGGCCCAATGG	Genotype analysis
STS19	AGTTCCTGATTCTTCCCTCG	CTCAGTTCCTTATACCAAGCC	Genotype analysis
STS20	CGGCATGTCATGGACTACG	CAGGAAATCTGTAACCAGAGG	Genotype analysis
M1	TGTCTCAGCCGTTAGAAGCA	TTGCCTCAAGAAAAGCTCGC	Sequence
M2	GGCCAACTCGATGATCAACC	CTCTGTCCTAAATCGTGCGC	Sequence
M3	ACCATTTCCCAAACGTGCTT	CGAATGCAATCTCCTCACGG	Sequence
M4	CTGGAGAATCAGCACGTACT	ACAGTACTACGTTTCGTGTGG	Sequence
WWG227	ACCATTTCCCAAACGTGCTT	CGAATGCAATCTCCTCACGG	Sequence
WWG226-5	TGCAAGCAAAGGACAAGGAA	TGCTGCTTTACCTTAATCGCA	Sequence
WWG161	CGGAGAGCGCGTCGTAGA	CTCGGAGAGCGCGTCGTAGAC	Sequence
WWG163	TGGAAGTGGGTGGAGAACTAC	TATCCGGATTTCGCATCACCTT	Sequence
WWG165	ACCTCGCAGCGAGCTGAGCTC	CGGAGGCGTTGTGCATGCAAT	Sequence
WWG231	GCTGTGCTCGATCTCATGC	CCGTCAACAGCCTCTCCTC	Sequence
WWG232	CTGGAGAATCAGCACGTACT	ACAGTACTACGTTTCGTGTGG	Sequence
actin1	GACTCTGGTGATGGTGTCAGC	GGCTGGAAGAGGACCTCAGG	qRT-PCR
QRTGL4-1	ATGTCGGGCTCCTCTGCTGCC	GCAGTAGTTCTCCACCCACTT	qRT-PCR
QRTGL4-2	AGCATGTCGGGCTCCTCTGCTG	CCACTTGTCATTGCACTGGTTC	qRT-PCR
pGL4-1	TGCTCTAGAGCTGGCCGTAGAAGTC AAG	ACAAAGCTTGTATAGCTCAGCACG AGC	Vector construct
pGL4-2a	TACCCGGGGATCCTCTAGGCTGGCC GTAGAAGTCAAG	CGACGACGGCGGGCGGCTGAGGAGG AGGTGGGTGGTG	Vector construct
pGL4-2b	AGCCGCCGCCGTCGTCGCTGCA	TGTAACACGACGGCCAGTGCCACT TGTATAGCTCAGCACGAGC	Vector construct
ObGL4-GFP	ATAGAATTCATGTCGGGCTCCTCTG CTGC	ATAGTCGACTCGGCCGGAGGAGTC GCCGC	Vector construct
Oggl4-GFP	ATAATAGAATTCATGTCGGGCTCCT CTGCTGC	ATAGTCGACAGGAGGAGGTGGGTG GTGGT	Vector construct

Supplementary Table 2. Materials used in this study

NO.	Species	Accession	Origin	SNP2	Haplotype	Grain Length(mm)
1	<i>O. barthii</i>	W1643	Botswana	C	H2	10.61
2	<i>O. barthii</i>	W1642	Botswana	C	H2	10.35
3	<i>O. barthii</i>	W1467	Cameroun	C	H1	10.78
4	<i>O. barthii</i>	W1588	Cameroun	C	H1	9.67
5	<i>O. barthii</i>	W1050	Gambia	C	H4	9.62
6	<i>O. barthii</i>	W0698	Guinea	C	H4	9.73
7	<i>O. barthii</i>	W0747	Mali	C	H4	10.35
8	<i>O. barthii</i>	W1432	Mali	C	H4	10.82
9	<i>O. barthii</i>	W1443A	Mali	C	H1	9.74
10	<i>O. barthii</i>	W1063	Mary, Gambia	C	H1	9.47
11	<i>O. barthii</i>	W1410	Sierra Leone	C	H4	10.49
12	<i>O. barthii</i>	W1416	Sierra Leone	C	H4	10.28
13	<i>O. barthii</i>	W0652	Sierra Leone	C	H4	10.35
14	<i>O. barthii</i>	W1583	Tchad	C	H1	10.62
15	<i>O. barthii</i>	W0042	West Africa, de Koba	C	H3	11.51
16	<i>O. barthii</i>	W1411*	Sierra Leone	C	H3	11.46
17	<i>O. glaberrima</i>	PI450527	Cameroon	T	H5	8.94
18	<i>O. glaberrima</i>	PI450528	Cameroon	T	H5	8.50
19	<i>O. glaberrima</i>	PI450532	Cameroon	T	H5	8.86
20	<i>O. glaberrima</i>	GSOR311693	Cameroon	T	H5	8.73
21	<i>O. glaberrima</i>	PI450511	Chad	T	H5	9.00
22	<i>O. glaberrima</i>	PI450515	Chad	T	H5	9.04
23	<i>O. glaberrima</i>	PI450516	Chad	T	H5	8.57
24	<i>O. glaberrima</i>	GSOR311692	Chad	T	H5	8.77
25	<i>O. glaberrima</i>	PI450241	Cote D'Ivoire	T	H5	8.23
26	<i>O. glaberrima</i>	PI450448	Cote D'Ivoire	T	H5	8.21
27	<i>O. glaberrima</i>	PI450453	Cote D'Ivoire	T	H5	7.59
28	<i>O. glaberrima</i>	PI450454	Cote D'Ivoire	T	H5	8.55
29	<i>O. glaberrima</i>	PI636797	Cote D'Ivoire	C	H4	8.66
31	<i>O. glaberrima</i>	PI450408	Gambia	T	H5	7.75
32	<i>O. glaberrima</i>	PI450430	Gambia	T	H5	7.50
33	<i>O. glaberrima</i>	PI450203	Ghana	T	H5	9.09
34	<i>O. glaberrima</i>	IRGC104339	Guinea	C	H4	8.34
35	<i>O. glaberrima</i>	PI450253	Liberia	T	H5	7.84
36	<i>O. glaberrima</i>	PI450255	Liberia	T	H5	7.86
37	<i>O. glaberrima</i>	PI450307	Liberia	T	H5	8.07
38	<i>O. glaberrima</i>	PI450319	Liberia	T	H5	7.78
39	<i>O. glaberrima</i>	PI450320	Liberia	T	H5	7.65
40	<i>O. glaberrima</i>	PI450330	Liberia	T	H5	7.32
41	<i>O. glaberrima</i>	PI450333	Liberia	T	H5	7.55
42	<i>O. glaberrima</i>	IRGC102244	Liberia	T	H5	7.82
43	<i>O. glaberrima</i>	IRGC102266	Liberia	T	H5	9.41

44	<i>O. glaberrima</i>	IRGC102303	Liberia	T	H5	8.12
45	<i>O. glaberrima</i>	IRGC102305*	Liberia	T	H5	8.43
46	<i>O. glaberrima</i>	PI450322	Liberia	C	H4	8.69
47	<i>O. glaberrima</i>	PI450312	Liberia	C	H4	9.62
48	<i>O. glaberrima</i>	PI450243	Mali	T	H5	8.27
49	<i>O. glaberrima</i>	PI450365	Mali	T	H5	8.21
50	<i>O. glaberrima</i>	PI450367	Mali	T	H5	7.67
51	<i>O. glaberrima</i>	PI450368	Mali	T	H5	8.33
52	<i>O. glaberrima</i>	PI450471	Mali	T	H5	7.47
53	<i>O. glaberrima</i>	PI450472	Mali	T	H5	8.46
54	<i>O. glaberrima</i>	PI450474	Mali	T	H5	8.21
55	<i>O. glaberrima</i>	PI450477	Mali	T	H5	7.45
56	<i>O. glaberrima</i>	PI450488	Mali	T	H5	8.24
57	<i>O. glaberrima</i>	PI450489	Mali	T	H5	7.78
58	<i>O. glaberrima</i>	PI450490	Mali	T	H5	7.81
59	<i>O. glaberrima</i>	PI450507	Mali	T	H5	8.70
60	<i>O. glaberrima</i>	PI450509	Mali	T	H5	8.59
61	<i>O. glaberrima</i>	PI450510	Mali	T	H5	8.31
62	<i>O. glaberrima</i>	GSOR311689	Mali	T	H5	8.18
63	<i>O. glaberrima</i>	IRGC103604	Mali	T	H5	7.73
64	<i>O. glaberrima</i>	IRGC103612	Mali	T	H5	8.00
65	<i>O. glaberrima</i>	IRGC103632	Mali	T	H5	8.21
66	<i>O. glaberrima</i>	TOG7102	Mali	T	H5	9.05
67	<i>O. glaberrima</i>	PI450335	Nigeria	T	H5	8.68
68	<i>O. glaberrima</i>	IRGC103930	Nigeria	T	H5	8.69
69	<i>O. glaberrima</i>	IRGC104574	Nigeria	T	H5	9.22
70	<i>O. glaberrima</i>	IRGC104583	Nigeria	T	H5	8.72
71	<i>O. glaberrima</i>	PI450392	Senegal	T	H5	8.75
72	<i>O. glaberrima</i>	PI450396	Senegal	T	H5	7.89
73	<i>O. glaberrima</i>	PI450399	Senegal	T	H5	7.54
74	<i>O. glaberrima</i>	PI450405	Senegal	T	H5	7.85
75	<i>O. glaberrima</i>	PI450406	Senegal	T	H5	7.71
76	<i>O. glaberrima</i>	PI450414	Senegal	T	H5	8.26
77	<i>O. glaberrima</i>	PI450439	Senegal	T	H5	8.34
78	<i>O. glaberrima</i>	GSOR311690	Senegal	T	H5	8.30
79	<i>O. glaberrima</i>	GSOR311691	Senegal	T	H5	7.53
80	<i>O. glaberrima</i>	PI450346	Sierra Leone	T	H5	7.91
81	<i>O. glaberrima</i>	PI450353	Sierra Leone	T	H5	7.79
82	<i>O. glaberrima</i>	PI450354	Sierra Leone	T	H5	7.99
83	<i>O. glaberrima</i>	PI450356	Sierra Leone	T	H5	7.92
84	<i>O. glaberrima</i>	PI450352	Sierra Leone	C	H4	8.88

* These two accessions was used to development introgression line GIL25

**Supplementary Table 3. Principal Component Analysis of SNP
Variants 5kb upstream and downstream of GL4**

Accession	Quadrant	PC1	PC2	PC3	Individual No. of SW_I
IRRI_103517	NE	-0.0259	0.1475	-0.0202	
IRRI_103530	NE	-0.0307	0.1814	-0.049	
IRRI_103922	NE	-0.0447	-0.1271	-0.5546	
IRRI_103982	NE	-0.0415	-0.0745	0.0255	
IRRI_104533	NE	-0.0346	-0.0863	0.0472	
IRRI_104545	NE	-0.0326	-0.0796	0.0451	
IRRI_75546	NE	-0.0295	0.1809	-0.0486	
IRRI_75618	NE	-0.0426	-0.0947	0.0363	
IRRI_75729	NE	-0.0301	0.1744	-0.0367	
IRRI_103442	NW	-0.0418	-0.0765	0.017	
IRRI_103450	NW	-0.027	-0.108	0.0682	
IRRI_103452	NW	-0.0439	-0.0718	0.0219	
IRRI_103456	NW	-0.045	-0.1184	0.1092	
IRRI_103461	NW	-0.0423	-0.0517	0.0185	
IRRI_103463	NW	-0.0408	-0.0731	0.0164	
IRRI_103955	NW	-0.0356	0.1812	-0.0211	
IRRI_103956	NW	-0.0169	-0.0886	0.0583	
IRRI_103957	NW	-0.042	-0.0783	0.0566	
IRRI_103958	NW	-0.0295	-0.107	0.0719	
IRRI_103959	NW	-0.0303	0.1563	-0.0244	
IRRI_103960	NW	-0.0303	0.1563	-0.0244	
IRRI_103963	NW	-0.0465	-0.1488	-0.5119	
IRRI_103967	NW	-0.0424	-0.0987	0.0602	
IRRI_104022	NW	-0.042	-0.0727	0.0166	
IRRI_104023	NW	-0.042	-0.0727	0.0166	
IRRI_104024	NW	-0.0447	-0.1271	-0.5546	
IRRI_104025	NW	-0.0439	-0.0718	0.0219	
IRRI_104028	NW	-0.0261	-0.089	0.0416	
IRRI_104029	NW	-0.0408	-0.0787	0.057	
IRRI_104030	NW	-0.0289	0.1367	0.0108	
IRRI_104032	NW	-0.0436	-0.0982	0.0597	
IRRI_104177	NW	-0.017	0.1595	0.0284	
IRRI_104178	NW	-0.0297	0.1477	-0.0254	
IRRI_104180	NW	-0.0292	0.1407	-0.013	
IRRI_104566	NW	-0.0408	-0.073	0.0165	
IRRI_104571	NW	-0.0414	-0.0797	0.029	
IRRI_104595	NW	-0.0197	0.0972	0.0139	
IRRI_105005	NW	-0.045	-0.1184	0.1092	
IRRI_105011	NW	-0.041	-0.0823	0.0381	
IRRI_105043	NW	-0.0311	0.1084	0.0227	
IRRI_105044	NW	-0.029	0.1421	-0.0161	
TOg7135	NW	-0.0303	0.1579	-0.011	
IRRI_103592	SE	-0.0414	-0.0797	0.029	
IRRI_103599	SE	-0.0342	-0.1041	0.068	

IRRI_103981 SE	-0.0414	-0.0543	0.0497	
IRRI_104011 SE	-0.0445	-0.0632	0.0229	
IRRI_104044 SE	-0.0417	-0.0991	0.0544	
IRRI_104047 SE	-0.0411	-0.099	0.0601	
IRRI_104195 SE	-0.041	-0.0823	0.0381	
IRRI_104260 SE	-0.0512	-0.0508	0.0511	
IRRI_104294 SE	-0.0404	-0.0582	0.0532	
IRRI_104573 SE	-0.038	-0.0856	0.0574	
IRRI_104904 SE	-0.0387	-0.101	0.0631	
IRRI_104934 SE	-0.0292	0.1407	-0.013	
IRRI_67563 SE	-0.0316	0.0314	-0.0215	
IRRI_103948 SW	-0.0306	0.1054	-0.0183	
IRRI_103989 SW	-0.0402	-0.0802	0.0294	
IRRI_103991 SW	-0.0424	-0.0987	0.0602	
IRRI_103992 SW	-0.0316	0.1486	-0.02	
IRRI_103993 SW	-0.0277	0.1364	0.0107	
IRRI_103994 SW	-0.0297	0.1477	-0.0254	
IRRI_103995 SW	-0.0301	0.1372	0.0104	
IRRI_104173 SW	-0.0348	-0.1008	0.0546	
IRRI_104181 SW	-0.0434	-0.093	0.0661	
IRRI_104182 SW	-0.0399	-0.0892	0.0104	
IRRI_104187 SW	-0.0425	-0.0709	0.0077	
IRRI_104190 SW	-0.0388	0.0465	0.0076	
IRRI_104194 SW	-0.042	-0.0727	0.0166	
IRRI_104231 SW	-0.0437	-0.0841	0.0307	
IRRI_104561 SW	-0.0316	0.1486	-0.02	
IRRI_104562 SW	-0.0289	0.1367	0.0108	
IRRI_105021 SW	-0.0301	0.1372	0.0104	
IRRI_105026 SW	-0.0307	0.1788	-0.0195	
IRRI_105034 SW	-0.0297	0.1477	-0.0254	
IRRI_105036 SW	-0.0299	0.1676	0.0173	
IRRI_105038 SW	-0.0405	-0.0764	0.029	
IRRI_105050 SW	-0.026	-0.0877	0.0386	
IRRI_105052 SW	-0.0311	0.1683	0.0162	
IRRI_58622 SW	-0.0364	-0.0614	0.0351	
IRRI_61457 SW	-0.0405	-0.0996	0.0548	
TOg7197 SW	-0.0305	0.1404	-0.0016	
IRRI_104036 SW_I	0.3353	-0.0174	0.0047	1
IRRI_104165 SW_I	0.3355	-0.0172	0.0047	2
IRRI_105048 SW_I	0.3353	-0.0173	0.0046	3
IRRI_103946 SW_I	0.3358	-0.017	0.0047	4
IRRI_103949 SW_I	0.3357	-0.0171	0.0047	5
IRRI_103953 SW_I	0.3356	-0.0172	0.0047	6
IRRI_103988 SW_I	0.3355	-0.0173	0.0047	7
IRRI_104035 SW_I	0.3354	-0.0174	0.0047	8
TOg6203 SW_I	0.3353	-0.0175	0.0047	9
IRRI_103937 SW_I	0.3343	-0.0061	-0.0256	10
IRRI_105049 SW_I	0.3343	-0.0061	-0.0256	11

Supplementary Table 4. Geographical Analysis of SNP2

Postion in *O. glaberrima*

Accession	Quadrant	Country of Origin	Genotype
IRRI_75546	NE	Burkina Faso	T/T
IRRI_75618	NE	Burkina Faso	T/T
IRRI_75729	NE	Burkina Faso	T/T
IRRI_103517	NE	Mali	T/T
IRRI_103530	NE	Mali	T/T
IRRI_103922	NE	Nigeria	T/T
IRRI_103982	NE	Nigeria	T/T
IRRI_104533	NE	Nigeria	T/T
IRRI_104545	NE	Nigeria	T/T
IRRI_103450	NW	Gambia	T/T
IRRI_104177	NW	Guinea	T/T
IRRI_104178	NW	Guinea	T/T
IRRI_104180	NW	Guinea	T/T
IRRI_105005	NW	Guinea	T/T
IRRI_105011	NW	Guinea	T/T
IRRI_105043	NW	Guinea	T/T
IRRI_105044	NW	Guinea	T/T
IRRI_104022	NW	Guinea-Bissau	T/T
IRRI_104023	NW	Guinea-Bissau	T/T
IRRI_104024	NW	Guinea-Bissau	T/T
IRRI_104025	NW	Guinea-Bissau	T/T
IRRI_104028	NW	Guinea-Bissau	T/T
IRRI_104029	NW	Guinea-Bissau	T/T
IRRI_104030	NW	Guinea-Bissau	T/T
IRRI_104032	NW	Guinea-Bissau	T/T
IRRI_104595	NW	Mali	T/T
IRRI_103442	NW	Senegal	T/T
IRRI_103452	NW	Senegal	T/T
IRRI_103456	NW	Senegal	T/T
IRRI_103461	NW	Senegal	T/T
IRRI_103463	NW	Senegal	T/T
IRRI_103955	NW	Senegal	T/T
IRRI_103956	NW	Senegal	T/T
IRRI_103957	NW	Senegal	T/T
IRRI_103958	NW	Senegal	T/T
IRRI_103959	NW	Senegal	T/T
IRRI_103960	NW	Senegal	T/T
IRRI_103963	NW	Senegal	T/T
IRRI_103967	NW	Senegal	T/T
IRRI_104566	NW	Senegal	T/T
IRRI_104571	NW	Senegal	T/T
TOg7135	NW	Senegal	T/T
IRRI_104934	SE	Burkina Faso	T/T
IRRI_103592	SE	Cameroon	T/T
IRRI_103599	SE	Cameroon	T/T

IRRI_104047	SE	Cameroon	T/T
IRRI_104044	SE	Chad	T/T
IRRI_104294	SE	Chad	T/T
IRRI_104195	SE	Ghana	T/T
IRRI_104260	SE	Ghana	T/T
IRRI_67563	SE	Ghana	T/T
IRRI_103981	SE	Nigeria	T/T
IRRI_104011	SE	Nigeria	T/T
IRRI_104904	SE	Nigeria	T/T
IRRI_104034	SW	Cote d'Ivoire	T/T
TOg7197	SW	Cote d'Ivoire	T/T
IRRI_104173	SW	Guinea	T/T
IRRI_104181	SW	Guinea	T/T
IRRI_104182	SW	Guinea	T/T
IRRI_104187	SW	Guinea	T/T
IRRI_104190	SW	Guinea	T/T
IRRI_104194	SW	Guinea	T/T
IRRI_105021	SW	Guinea	T/T
IRRI_105026	SW	Guinea	T/T
IRRI_105034	SW	Guinea	T/T
IRRI_105036	SW	Guinea	T/T
IRRI_105038	SW	Guinea	T/T
IRRI_105052	SW	Guinea	T/T
IRRI_103948	SW	Liberia	T/T
IRRI_105050	SW	Liberia	T/T
IRRI_61457	SW	Liberia	T/T
IRRI_103989	SW	Sierra Leone	T/T
IRRI_103991	SW	Sierra Leone	T/T
IRRI_103992	SW	Sierra Leone	T/T
IRRI_103994	SW	Sierra Leone	T/T
IRRI_103995	SW	Sierra Leone	T/T
IRRI_104231	SW	Sierra Leone	T/T
IRRI_104561	SW	Sierra Leone	T/T
IRRI_104562	SW	Sierra Leone	T/T
IRRI_58622	SW	Sierra Leone	T/T
IRRI_104035	SW_I	Cote d'Ivoire	C/C
IRRI_104036	SW_I	Cote d'Ivoire	C/C
IRRI_104165	SW_I	Guinea	C/C
TOg6203	SW_I	Guinea	C/C
IRRI_103937	SW_I	Liberia	C/C
IRRI_103946	SW_I	Liberia	C/C
IRRI_103949	SW_I	Liberia	C/C
IRRI_105048	SW_I	Liberia	C/C
IRRI_105049	SW_I	Liberia	C/C
IRRI_103953	SW_I	Sierra Leone	C/C
IRRI_103988	SW_I	Sierra Leone	C/C