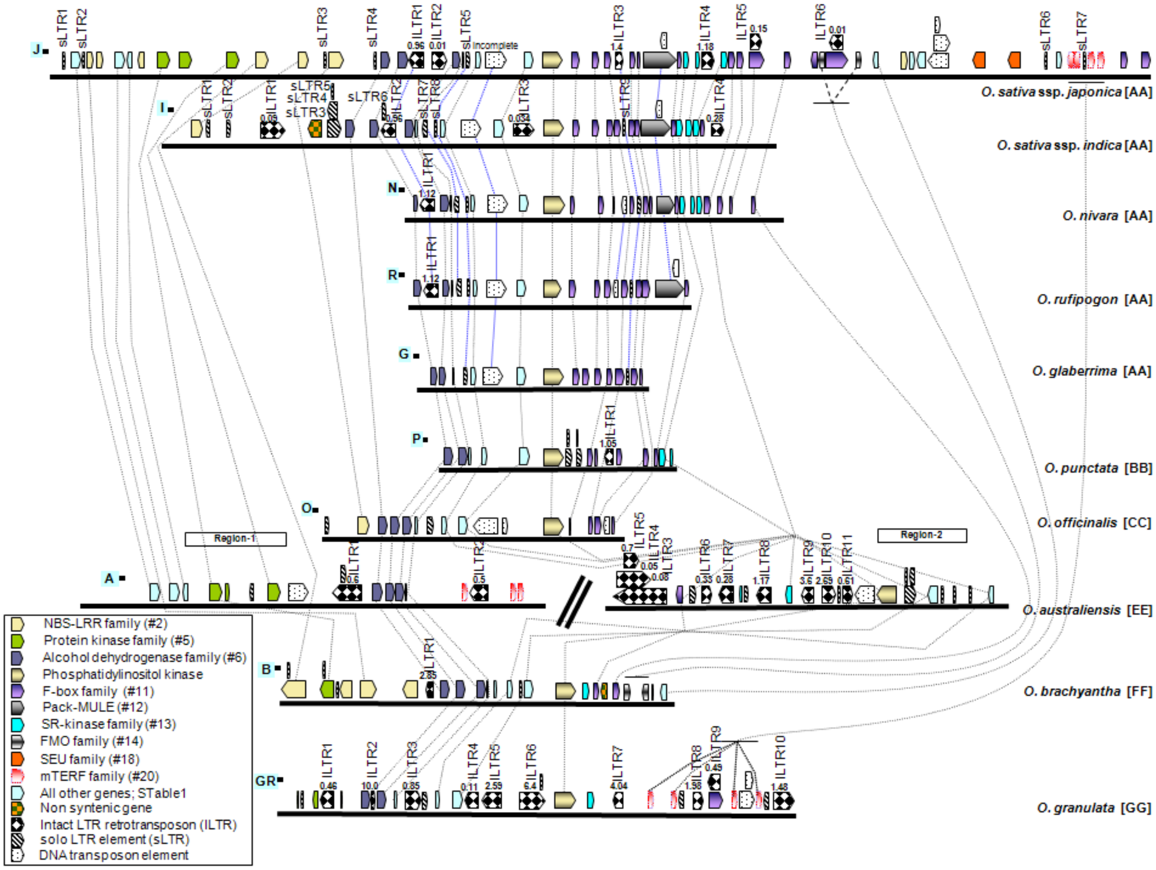
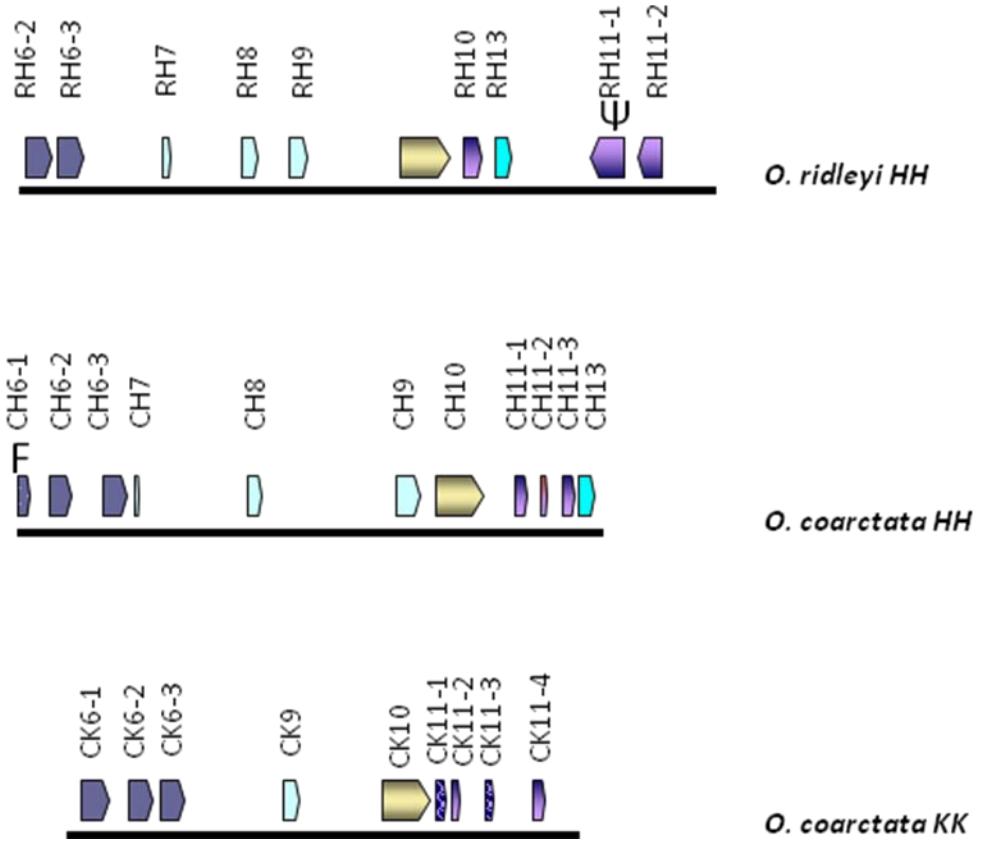
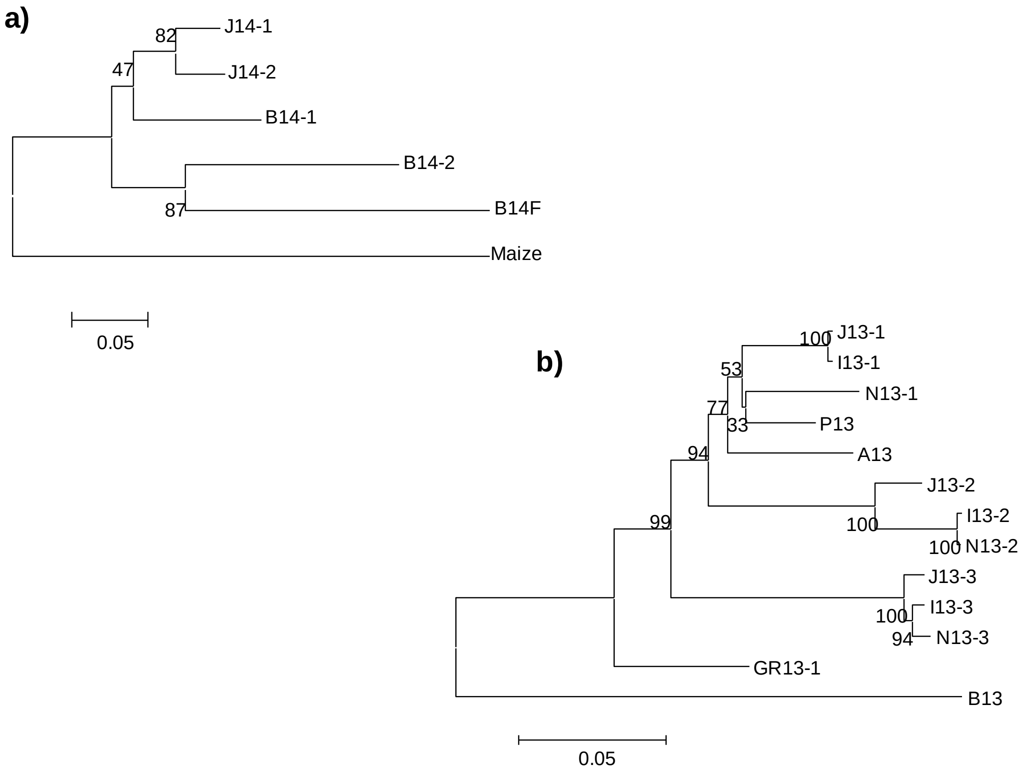
Supplemental Data. Ammiraju et al. (2008). Dynamic Evolution of *Oryza* Genomes Is Revealed by Comparative Genomic Analysis of a Genus-wide Vertical Dataset



Supplemental Figure 1. Phylogenomic View of the Genus-wide *Adh1* Vertical Sequence Dataset Depicting Both Gene and Transposon Dynamics. TEs larger than 3 Kb are shown with the exception of all intact (represented as ILTR) and solo LTR (sLTR) retrotransposons. Insertion time (in million years ago) was indicated above each ILTR. Orthologous TEs are highlighted in blue connectors.

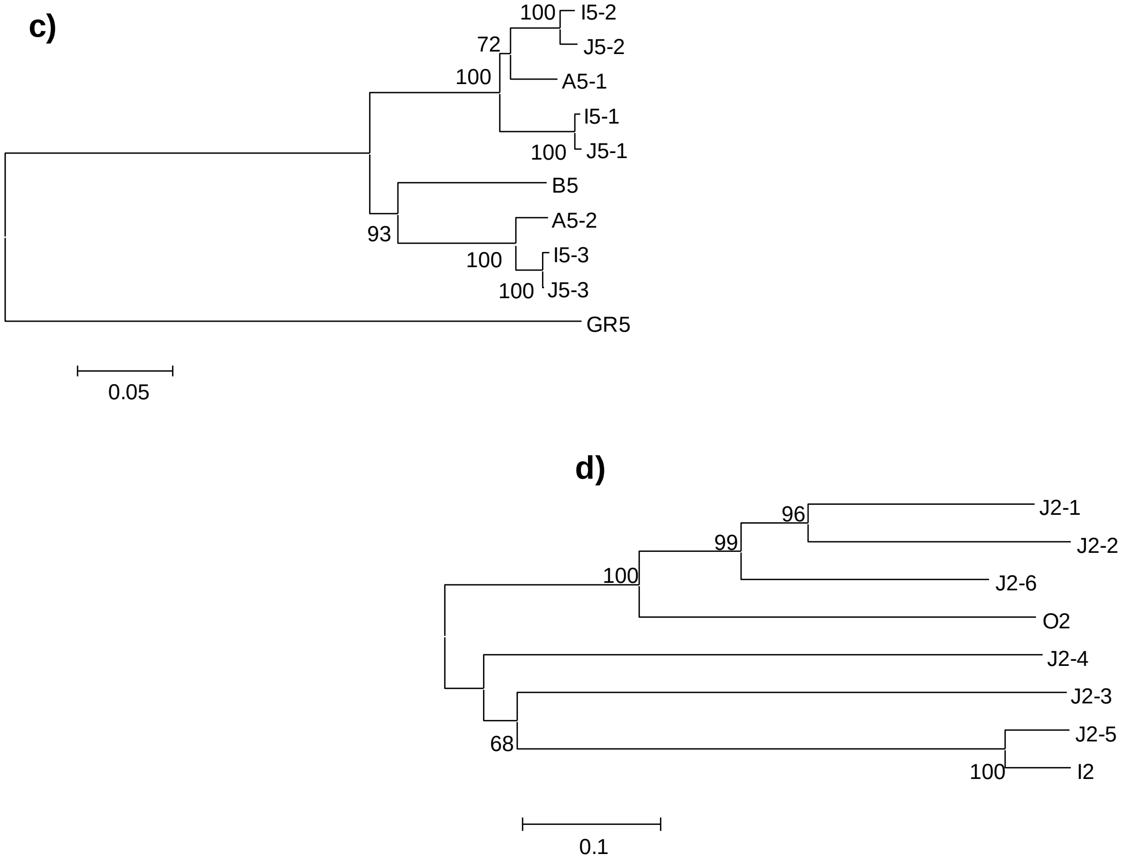


Supplemental Figure 2. Figure Showing The Presence of F-box Gene Family (gene family #11) In The Polyploid *Oryza* Species At The Orthologous *Adh1* Regions. Genes from each *O. coarctata* [HHKK] sub-genomes are denoted by CH and CK, and HH sub-genome from *O. ridleyi* are denoted by RI. F indicates fragment. Gene numbering is in accordance with Figure 1.



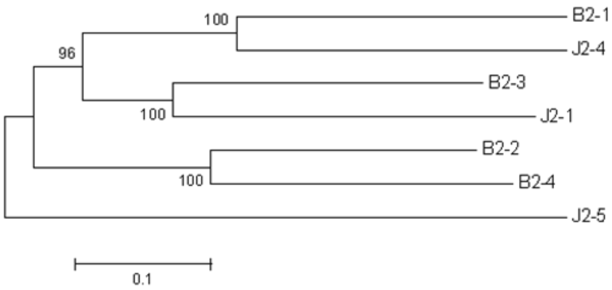
Supplemental Figure 3. Evolutionary Dynamics of Different Gene Families in Their Orthologous Regions. **a)** Phylogenetic relationships of gene family 14; **b)** Phylogenetic relationships of gene family 13; **c)** Phylogenetic relationships of gene family 5; **d)** Phylogenetic relationships of gene family 2; **e)** Phylogenetic relationships of gene family 2: *japonica* and *O. brachyantha* comparison; **f)** Phylogenetic relationships of gene family 20.

Supplemental Figure 3. Contd

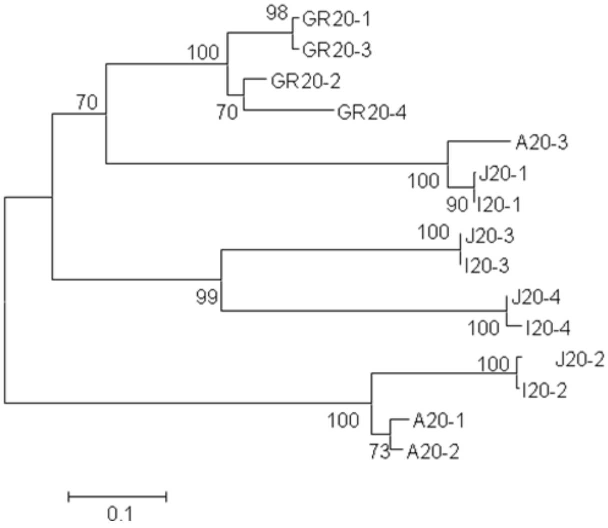


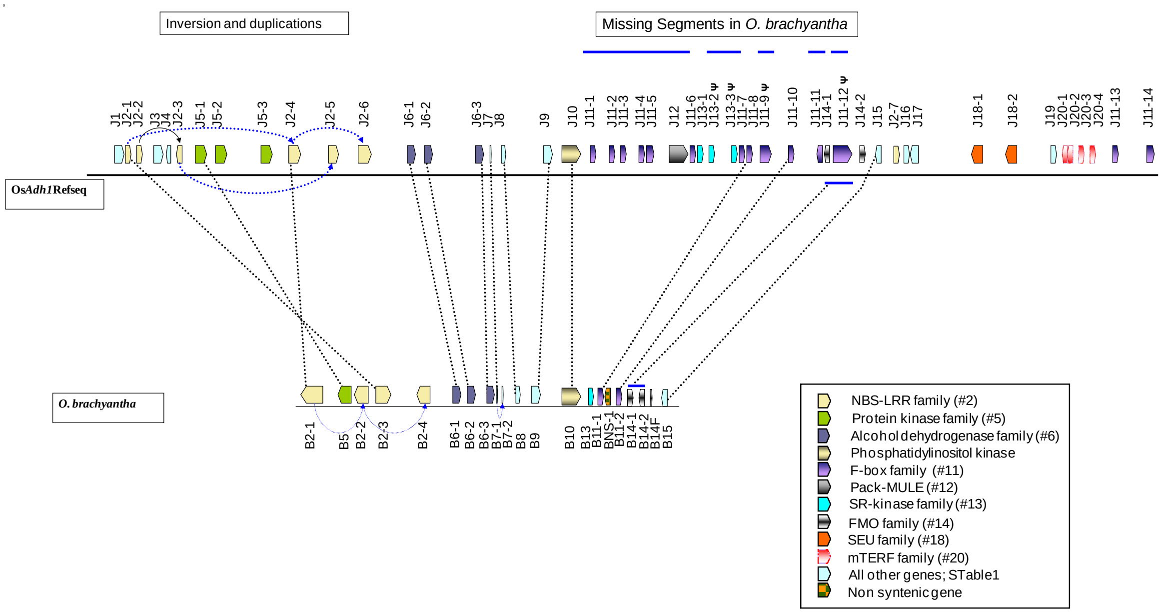
Supplemental Figure 3. Contd

e)

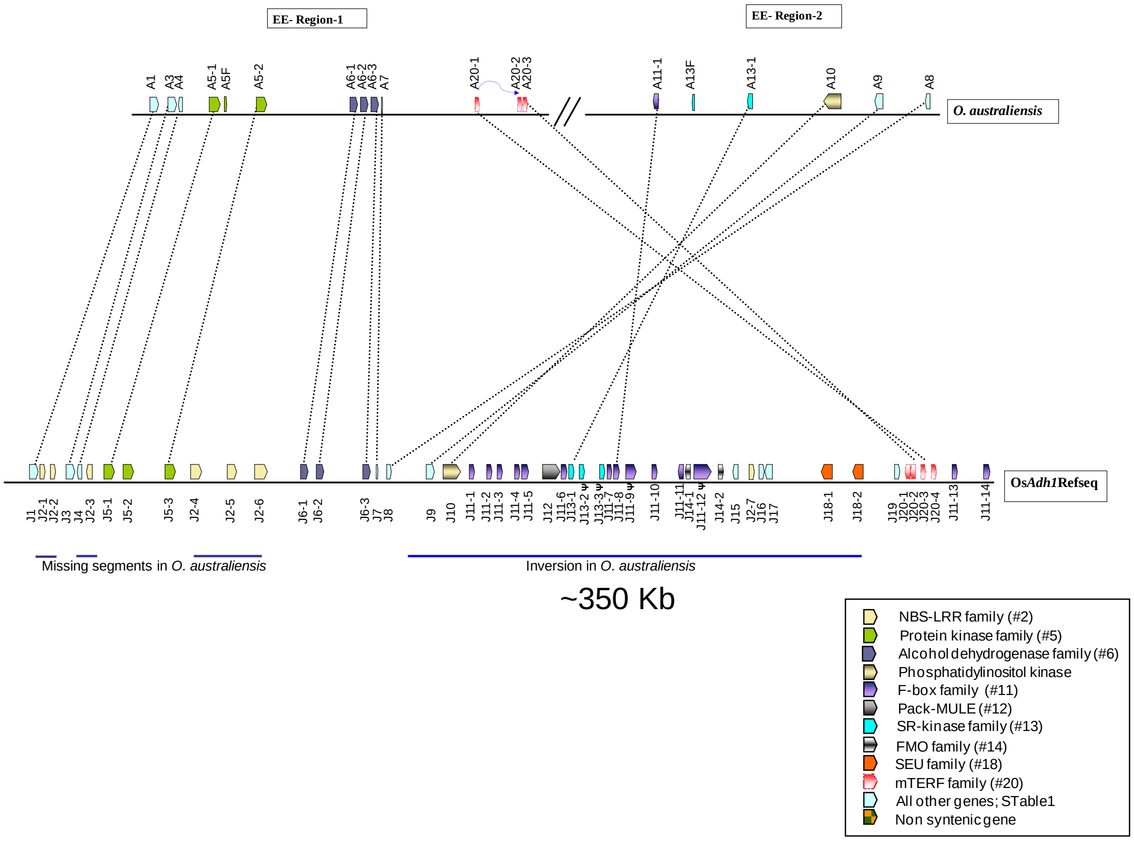


f)

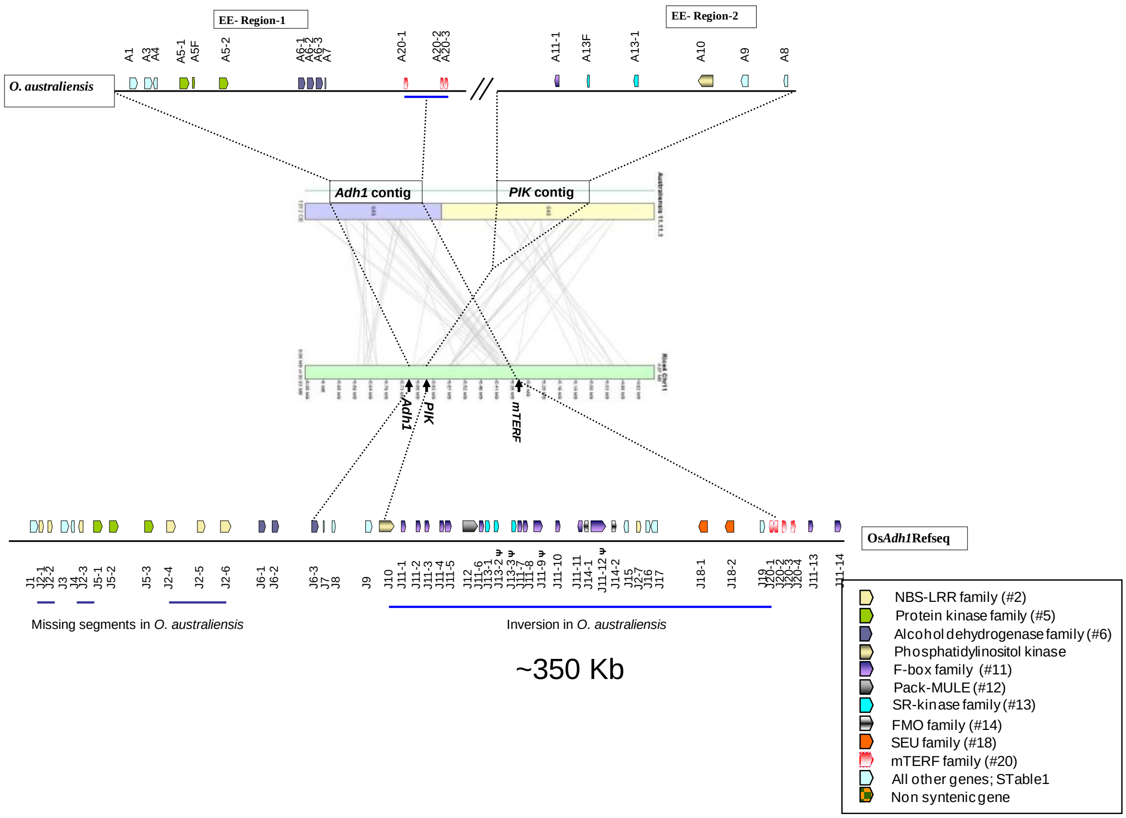




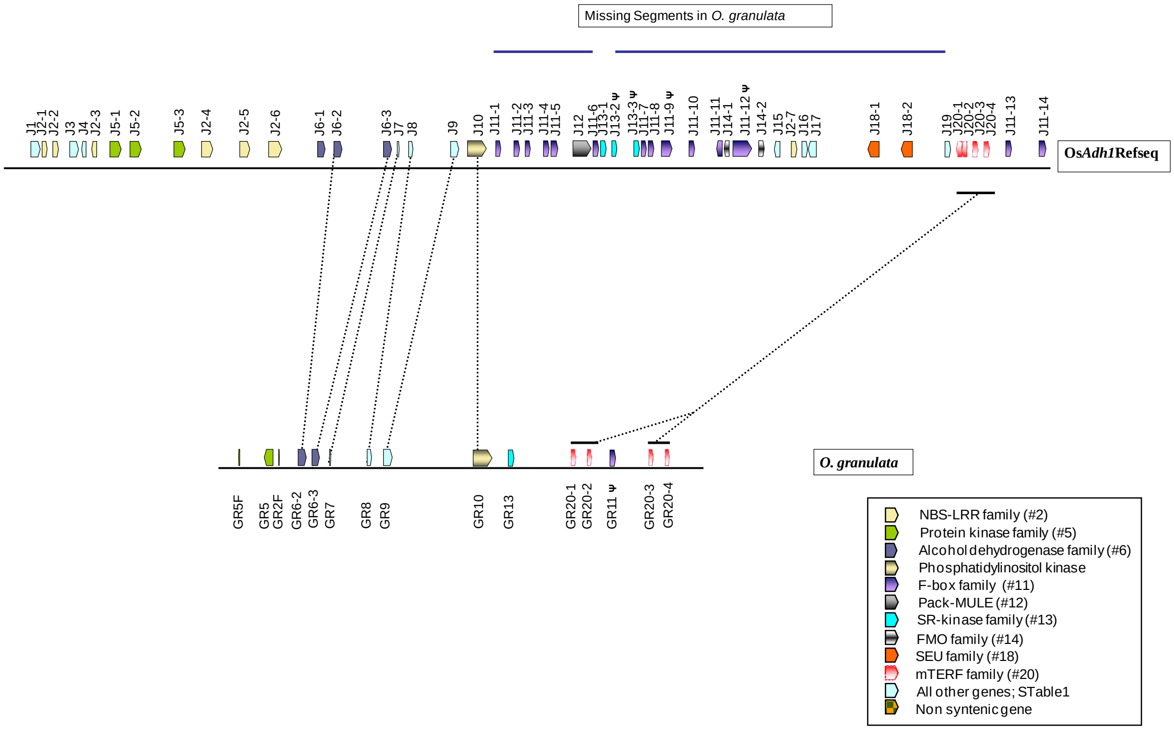
Supplemental Figure 4. Complex Rearrangements in *O. brachyantha* [FF] relative to the Os*Adh1*Refseq.



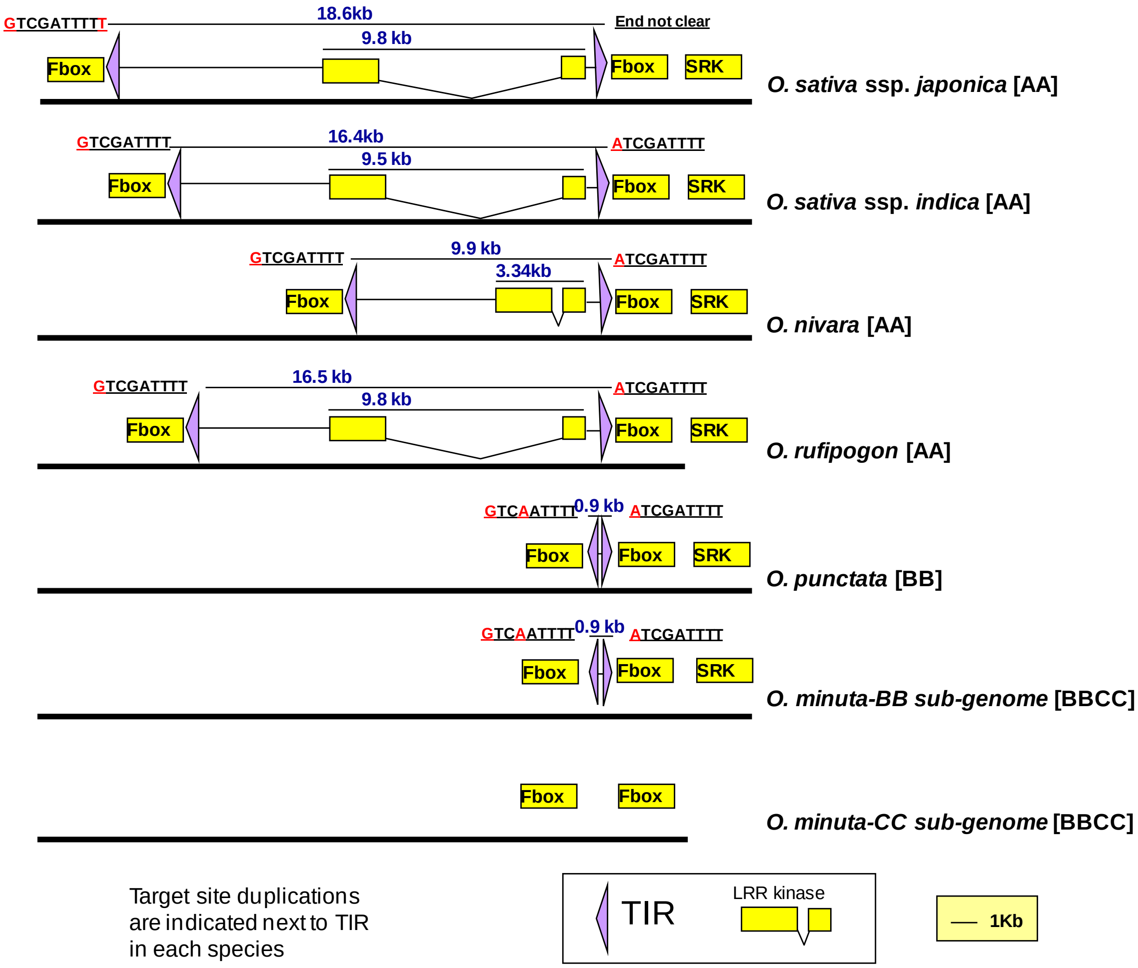
Supplemental Figure 5. Putative Inversion in *O. australiensis* [EE] Relative to the Os*Adh*1RefSeq.



Supplemental Figure 6. Putative Inversion in *O. australiensis* [EE] Relative to the Os*Adh1*Refseq, and Validation of the Inversion by Physical Mapping.



Supplemental Figure 7. Large Rearrangement in *O. granulata* [GG] Relative to the Os*Adh1*Refseq.



Supplemental Figure 8. Pack-MULE Mediated Gene Movement/Acquisition in the AA Genome Lineages.

Supplemental Table 1. List of Genes in the Os*Adh1*Refseq.

<table>
<tr><th rowspan="3">Gene#</th><th rowspan="3">Classification</th><th rowspan="3">Putative gene product</th><th colspan="4">Identification method</th><th rowspan="3">TIGR Rice model V5</th></tr>
<tr><th colspan="2">Transcript evidence</th><th colspan="2">Known functional domain</th></tr>
<tr><th>EST</th><th>Fl-cDNA</th><th>accession#</th><th>domain name</th></tr>
<tr><td>1</td><td>Expressed</td><td>UP1 protease family</td><td>TC316418</td><td>gi|32968360|dbj|AK058342.1|</td><td>COG5160</td><td>Protease, Ulp1 family</td><td>LOC_Os11g10780</td></tr>
<tr><td>2-1</td><td>Expressed</td><td>NBS LRR-1</td><td>TC314273</td><td>gi|32974139|dbj|AK064121.1|</td><td>PF00931</td><td>NB-ARC domain</td><td>LOC_Os11g10770</td></tr>
<tr><td>2-2</td><td>Expressed</td><td>NBS-LRR-2</td><td>TC309960</td><td>NONE</td><td>PF00560</td><td>Leucine Rich Repeat</td><td>LOC_Os11g10760</td></tr>
<tr><td>3</td><td>Expressed</td><td>serine carboxy peptidase</td><td>NONE</td><td>gi|32988670|dbj|AK103461.1|</td><td>PF00450</td><td>serine carboxypeptidase</td><td>LOC_Os11g10750</td></tr>
<tr><td>4</td><td>Hypothetical</td><td>tetratricopeptide</td><td>NONE</td><td>NONE</td><td>TIGR00756/PF01535</td><td>pentatricopeptide repeat domain</td><td>LOC_Os11g10740</td></tr>
<tr><td>2-3</td><td>Hypothetical</td><td>LRR -3</td><td>NONE</td><td>NONE</td><td>PF00560</td><td>Leucine Rich Repeat (2repeats)</td><td>LOC_Os11g10720</td></tr>
<tr><td>5-1</td><td>Expressed</td><td>protein kinase domain-1</td><td>TC331261</td><td>NONE</td><td>PF00069/PF00400</td><td>protein kinase domain/WD domain, G-beta repeat</td><td>LOC_Os11g10710</td></tr>
<tr><td>5-2</td><td>Expressed</td><td>protein kinase domain-2</td><td>TC316203</td><td>NONE</td><td>PF00069/PF07714</td><td>protein kinase domain/Protein tyrosine kinase /</td><td>LOC_Os11g10690
+
LOC_Os11g10680</td></tr>
<tr><td>5-3</td><td>Hypothetical</td><td>protein kinase-3</td><td>NONE</td><td>NONE</td><td>PF00069/PF00400</td><td>protein kinase domain/WD domain, G-beta repeat/</td><td>LOC_Os11g10640</td></tr>
<tr><td>2-4</td><td>Hypothetical</td><td>NBS-LRR-4</td><td>NONE</td><td>NONE</td><td>PF00931/PF00560</td><td>NB-ARC domain/Leucine Rich Repeat</td><td>LOC_Os11g10610
+
LOC_Os11g10620</td></tr>
<tr><td>2-5</td><td>Expressed</td><td>NBS-LRR-_5</td><td>CF196923</td><td>NONE</td><td>PF00560</td><td>Leucine Rich Repeat</td><td>LOC_Os11g10570</td></tr>
<tr><td>2-6</td><td>Expressed</td><td>NBS-LRR-6</td><td>NONE</td><td>gi|32982196|dbj|AK072173.1|
;
gi|32969144|dbj|AK059126.1|
gi|32969144|dbj|AK059126.1|</td><td>PF00560/PF00931</td><td>NB-ARC domain /Leucine Rich Repeat/</td><td>LOC_Os11g10550</td></tr>
<tr><td>6-1</td><td>Expressed</td><td>Adh3</td><td>TC311799</td><td>gi|32986497|dbj|AK101288.1|</td><td>PF00107</td><td>oxidoreductase, zinc-binding dehydrogenase family</td><td>LOC_Os11g10520</td></tr>
<tr><td>6-2</td><td>Expressed</td><td>Adh2</td><td>TC294581</td><td>gi|32988008|dbj|AK102799.1|</td><td>PF00107</td><td>oxidoreductase, zinc-binding dehydrogenase family</td><td>LOC_Os11g10510</td></tr>
<tr><td>6-3</td><td>Expressed</td><td>Adh1</td><td>TC318729</td><td>gi|32979354|dbj|AK069330.1|</td><td>PF00107</td><td>oxidoreductase, zinc-binding dehydrogenase family</td><td>LOC_Os11g10480</td></tr>
</table>

Supplemental Table 1. Contd

Gene#	Classification	Putative gene product	Identification method				TIGR Rice model V5
			Transcript evidence		Known functional domain		
			EST	Fl-cDNA	accession#	domain name	
7	Expressed	RZ53	TC310740	NONE	NONE	NONE	LOC_Os11g10470
8	Expressed	Peroxidase	TC313637	NONE	PF00141	Peroxidase	LOC_Os11g10460
9	Hypothetical	NifS	NONE	NONE	PF00266	aminotransferase, class V	LOC_Os11g10430
10	Expressed	1-phosphatidylinositol 4-kinase	TC316980	gi|32983758|dbj|AK073735.1|	PF00357	Integrin alpha cytoplasmic region	LOC_Os11g10420
11-1	Expressed	Fbox-1	TC342142	NONE	PF00646	F-box domain	LOC_Os11g10400
11-2	Expressed	Fbox-2	TC355497	NONE	PF00647	F-box domain	LOC_Os11g10384
11-3	Expressed	Fbox-3	NP176576	NONE	PF00646	F-box domain	LOC_Os11g10370
11-4	Expressed	Fbox-4	TC291291	gi|32992111|dbj|AK106902.1|	PF00646	F-box domain	LOC_Os11g10340
11-5	Hypothetical	Fbox-5	NONE	NONE	PF00647	F-box domain	LOC_Os11g10330
12	Expressed	LRRkinase/PACKMULE	NONE	gi|32985249|dbj|AK111797.1|	PF00560/PF00069	Leucine Rich Repeat/protein kinase domain	LOC_Os11g10310
11-6	Expressed	Fbox-6	NP176582	NONE	PF00647	F-box domain	LOC_Os11g10300
13-1	Expressed	SRK kinase-1	TC305271	NONE	PF01453/PF00954	D-mannose binding lectin/S-locus glycoprotein family	LOC_Os11g10290
13-2	Psuedo	SRKkinase-2	NONE	NONE	PF00069/ PF01453/PF00954	protein kinase domain /D-mannose binding lectin/S-locus glycoprotein family	LOC_Os11g10280
13-3	Psuedo	SRKkinase-3	NONE	NONE	PF00069/PF01453/ PF00954	protein kinase domain/D-mannose binding lectin/S-locus glycoprotein family	LOC_Os11g10260
11-7	Expressed	Fbox-7	TC321228	gi|32991667|dbj|AK106458.1|	NONE	NONE	LOC_Os11g10250
11-8	Expressed	Fbox-8	TC305667	gi|32977905|dbj|AK067887.1|	PF00646	F-box domain	LOC_Os11g10240
11-9	Psuedo	Fbox-9	NONE	NONE	NONE	NONE	LOC_Os11g10230 + LOC_Os11g10220
11-10	Expressed	Fbox-10	TC317227	NONE	PF00646	F-box domain	LOC_Os11g10200

Supplemental Table 1. Contd

Gene#	Classification	Putative gene product	Identification method				TIGR Rice model V5
			Transcript evidence		Known functional domain		
			EST	Fl-cDNA	accession#	domain name	
11-11	Expressed	Fbox-11	TC291677	gi|37988956|dbj|AK119333.1| ; gi|32995027|dbj|AK109818.1|	PF00646	F-box domain	LOC_Os11g10180
14-1	Expressed	FMO-1	NP176587	NONE	PF00744	FMO-like flavin-binding monooxygenase.	LOC_Os11g10170
11-12	Pseudo	Fbox-12-pseuso	TC310614 ; CK012051 ; CA765367	NONE	NONE	NONE	LOC_Os11g10160
14-2	Hypothetical	FMO-2	NONE	NONE	PF00743	FMO-like flavin-binding monooxygenase.	LOC_Os11g10140
15	Hypothetical	Myb	NONE	NONE	PF00249	Myb-like DNA-binding domain	LOC_Os11g10130
2-7	Expressed	NBS-LRR-7	TC300544	gi|32976100|dbj|AK066082.1|	PF00931	NB-ARC domain	LOC_Os11g10120
16	Expressed, unknown	unknown	TC317117	NONE	none	has EST and Arabidopsis protein homolog	LOC_Os11g10110
17	Expressed	MAPKinase	TC299118	gi|37989913|dbj|AK120290.1|	PF00069/PF07714	protein kinase domain/Protein tyrosine kinase	LOC_Os11g10100
18-1	Expressed	SEU3A	TC303603 ; TC358935	NONE	NONE	NONE	LOC_Os11g10070
18-2	Expressed	SEU3B	NONE	gi|32972092|dbj|AK062074.1|	NONE	NONE	LOC_Os11g10060
19	Hypothetical	alpha-G-protein	NONE	NONE	PF00503	G-protein alpha subunit	LOC_Os11g10050
20-1	Expressed	mTRF-1	TC288935	gi|32979988|dbj|AK069964.1|	PF02536	mTERF	LOC_Os11g10040
20-2	Expressed	mTRF-2	TC315166	gi|37991445|dbj|AK121822.1|	PF02536	mTERF	LOC_Os11g10020
20-3	Expressed	mTRF-3	TC291747	gi|32974188|dbj|AK064170.1|	PF02536	mTERF	LOC_Os11g10000
20-4	Expressed	mTRF-4	TC290677	gi|32986082|dbj|AK100873.1|	PF02536	mTERF	LOC_Os11g09990
11-13	Expressed	Fbox-13	NP1431302	NONE	PF00646	F-box domain	LOC_Os11g09970
11-14	Hypothetical	Fbox -14	NONE	NONE	NONE	NONE	LOC_Os11g09960

Supplemental Table 2. Previously Identified Putative Gene Models from the Os*Adh1*Refseq that were Omitted from the Comparative Analysis.

Number	Gene model	classification	Cause for omission
1	LOC_Os11g10670	Hypothetical	No homolog outside *Oryza*; No conservation across *Adh1* vertical dataset; No conserved functional domain
2	LOC_Os11g10650	Hypothetical	No homolog outside *Oryza*; No conservation across *Adh1* vertical dataset; No conserved functional domain
3	LOC_Os11g10600	Hypothetical	TE (SC-3) fragment
4	LOC_Os11g10560	Hypothetical	MULE like TE fragment: No conserved functional domain
5	LOC_Os11g10540	Expressed	No homolog outside *Oryza*; No conservation across *Adh1* vertical dataset; No conserved functional domain. Possible gene fragment
6	LOC_Os11g10530	Expressed	No homolog outside *Oryza*; No conservation across *Adh1* vertical dataset; No conserved functional domain. Possible gene fragment
7	LOC_Os11g10434	Expressed	Appears fragment; No intact ORFs. No conservation across *Adh1* vertical dataset; No conserved domain
8	LOC_Os11g10350	Conserved	Ty1-copia-like fragment
9	LOC_Os11g10320	Expressed	Embedded within a Pack-MULE (J12) No homolog outside *Oryza*; No conservation across *Adh1* vertical dataset; No conserved functional domain
10	LOC_Os11g10210	Hypothetical	TE: MULE (Os0177-RC1633)
11	LOC_Os11g10080	Hypothetical	No homolog outside *Oryza*; No conservation across *Adh1* vertical dataset; No conserved functional domain TE
12	LOC_Os11g10010	Hypothetical	No homolog outside *Oryza*; No conservation across *Adh1* vertical dataset; No conserved functional domain

Supplemental Table 3. Previously Identified Gene Models that were Corrected in this Study.

Number	Gene model	classification	Cause for correction
1	LOC_Os11g10384	TE	This predicted TE model also integrated an F-box family gene (11-2) from this region. This gene is phylogenetically similar to other genes in this region.
2	LOC_Os11g10260	TE	This predicted TE model also integrated an S-receptor kinase family member (13-3) from this region. This gene (pseudo) is phylogenetically similar to other genes in this region (LOC_Os11g10290 and LOC_Os11g10280; gene sequence AF172282.20; Tarchini et al, 2000).
3	LOC_Os11g10230	TE	The two TEs are actually a single LTR retrotransposon with intact target site duplications, and inserted within the merged gene model - LOC_Os11g10230 + LOC_Os11g10220. This resulted in a pseudo gene (F-box gene family (11-9ψ) The actual gene sequence is AF172282.23; Tarchini, 2000).
	LOC_Os11g10220	TE	
4	LOC_Os11g10690	Expressed	These two gene models were merged on Fl-cDNA evidence and the structure of other members of this family (i.e. LOC_Os11g10710 and LOC_Os11g10640) in *Adh1*RefSeq and the genus-wide vertical sequence dataset.
	LOC_Os11g10680	Expressed	
5	LOC_Os11g10620	Expressed	These two gene models were merged based on the structure of paralogous members of other genes in this region (eg: LOC_Os11g10550).
	LOC_Os11g10610	Expressed	

Supplemental Table 4. Total Number of Genes (Intact, Pseudo, TE Embedded, and Gene Fragments) Identified in the *Adh1* Genus-wide Vertical Sequence Dataset.

Species	**Genome type**	**Intact**	**Pseudo**[3***]	**TE embedded**	**Fragment**[****]
O. sativa ssp *japonica*	AA	45	4	12	
O. sativa ssp *indica*	AA	18	2	2	
O. nivara	AA	19	1	2	
O. rufipogon	AA	12	1	1	
O. glaberrima	AA	12	1	1	
O. punctata	BB	11	1		1
O. officinalis	CC	11	1	1	
O. australiensis[*]	EE	17			2
O. brachyantha	FF	20			1
O. granulata	GG	12	3	1	1
Total		178[**]	14	82	5

[*] From both *O. australiensis* BACs

[**] Pack-MULE carried a full-length cDNA supported gene similar to the *Xa21* gene, therefore it was included in the total gene count.

[***] Pseudo genes contain frameshift mutations resulting in premature stop codons and/or contain TE insertions.

[****] Fragments are sequences that have significant homology to known genes in the *Adh1* region, but only have partial coverage of the parental genes without a clear beginning or end.

Supplemental Table 5. Synonymous (Ks) and Nonsynonymous (Ka) Values for Six Core Genes Spanning the *Adh1* Genus-wide Vertical Sequence Dataset.

Gene	**Species and genome type**	**Ks**	**Ka**	**Ka/Ks**	**Divergence time (MYA)***
Adh1_japonica (J6-3)	*O. sativa_indica*_AA	0	0	0	0
	*O. nivara*_AA	0	0	0	0
	*O. rufipogon*_AA	0	0	0	0
	*O. glaberrima*_AA	0.0038	0	0	0.3
	O. punctata _BB	0.1049	0.0023	0.02	8.1
	*O. officinalis*_CC	0.1451	0.0023	0.02	11.2
	*O. australiensis*_EE	0.1409	0.0058	0.04	10.8
	*O. brachyantha*_FF	0.2018	0.0069	0.03	15.5
	*O. granulata*_GG	0.2216	0.0092	0.04	17
Adh2_japonica (J6-2)	*O. sativa_indica*_AA	0	0	0	0
	*O. nivara*_AA	0	0	0	0
	*O. rufipogon*_AA	0	0	0	0
	*O. glaberrima*_AA	0.0276	0.0013	0.05	2.1
	O. punctata _BB	0.1013	0.0122	0.12	7.8
	*O. officinalis*_CC	0.117	0.0109	0.09	9
	*O. australiensis*_EE	0.1276	0.0136	0.11	9.8
	*O. brachyantha*_FF	0.2284	0.0275	0.12	17.6
	*O. granulata*_GG	0.3624	0.0285	0.08	27.9
RZ53 (J7)	*O. sativa_indica*_AA	0	0	0	0
	*O. nivara*_AA	0	0	0	0
	*O. rufipogon*_AA	0	0	0	0
	*O. glaberrima*_AA	0.0484	0.0192	0.4	3.7
	O. punctata _BB	0.1521	0.0664	0.44	11.7
	*O. officinalis*_CC	0.1256	0.0732	0.58	9.7
	*O. australiensis*_EE	0.4961	0.0807	0.16	38.2
	*O. brachyantha*_FF	0.1777	0.1246	0.7	13.7
	*O. granulata*_GG	0.4556	0.1596	0.35	35
Peroxidase_japonica (J8)	*O. sativa_indica*_AA	0.0042	0.0029	0.69	0.3
	*O. nivara*_AA	0	0	0	0
	*O. rufipogon*_AA	0	0	0	0
	*O. glaberrima*_AA	0	0.0014	0	0
	O. punctata _BB	0.063	0.0169	0.27	4.8

	*O. officinalis*_CC	0.1037	0.0141	0.14	8
	*O. australiensis*_EE	0.0788	0.0205	0.26	6.1
	*O. brachyantha*_FF	0.18	0.0481	0.27	13.8
	*O. granulata*_GG	0.1403	0.0524	0.37	10.8
NifS_Exon2_japonica (J9)	*O. sativa_indica*_AA	0	0	0	0.3
	*O. nivara*_AA	0	0	0	0
	*O. rufipogon*_AA	0	0	0	0
	*O. glaberrima*_AA	0.022	0.0027	0.12	1.7
	O. punctata _BB	0.0567	0.0242	0.43	4.4
	*O. officinalis*_CC	0.0766	0.0284	0.37	5.9
	*O. australiensis*_EE	0.087	0.0401	0.46	6.7
	*O. brachyantha*_FF	0.1869	0.0999	0.53	14.4
	*O. granulata*_GG	0.1403	0.081	0.58	10.8
PIK_japonica (J10)	*O. sativa_indica*_AA	0.0023	0.0017	0.74	0.2
	*O. nivara*_AA	0.0014	0.0016	1.14	0.1
	*O. rufipogon*_AA	0.0014	0.0016	1.14	0.1
	*O. glaberrima*_AA	0.0039	0.0017	0.44	0.3
	O. punctata _BB	0.0753	0.007	0.09	5.8
	*O. officinalis*_CC	0.069	0.0066	0.1	5.3
	*O. australiensis*_EE	0.0823	0.0086	0.1	6.3
	*O. brachyantha*_FF	0.19190	0.02380	0.1241	14.8
	*O. granulata*_GG	0.1824	0.0306	0.17	14

* estimated using synonymous substitution rate of *Adh1* from grasses (Gaut et al, 1999)

Supplemental Table 6. Molecular Divergence Times of Individual *Oryza* Species using four Os*Adh1*Refseq Core Genes to Derive an Approximate Molecular Clock.

	Gene	***O. sativa* ssp. *indica* *(AA)***	***O. nivara (AA)***	***O. rufipogon (AA)***	***O. glaberrima (AA)***	***O. punctata (BB)***	***O. officinalis (CC)***	***O. australinesis (EE)***	***O. brachyantha (FF)***	***O. granulata (GG)***
O. sativa* ssp. *japonica (AA)	*Adh1*	0.0	0.0	0.0	0.3	8.1	11.2	10.8	15.5	17.0
	NifS (EX2)	0.3	0.0	0.0	1.7	4.4	5.9	6.7	14.4	10.8
	Peroxidase	0.3	0.0	0.0	0.0	4.8	8.0	6.1	13.8	10.8
	PIK	0.2	0.1	0.1	0.3	5.79	5.3	6.3	14.8	14.0
	Average divergence time (in MYA)	0.27	0.10*	0.10*	0.58	5.76	7.60	7.49	14.62	13.17

*molecular divergence time estimated for *O. nivara* and *O. rufipogon* is from a single gene.

Supplemental Table 7. Sequence Conservation in Coding and Intergenic Regions Derived from Pair-wise Comparisons Between a "Core" Segment (J*6-2*—J*10* interval) of the *Adh1* Genus-wide Vertical Sequence Dataset.

Species comparison	**Genome type**	**Length of the aligned region in pair wise comparison for *6-2--10* interval**		**Total length sequence in the alignment (L) * (bp)**	**Total length of conserved sequence in the alignment (H) * (bp)**	**Percentage of total sequence conserved %C-=H/L ***	**Percentage of total intergenic sequence conserved %I***	**Replaced portion in the intergenic region (100-I)***
		Aln1 (bp)	**Aln2 (bp)**					
O. sativa ssp. *indica (93-11)*	AA	99589	108341	207930	165524	82.7	74	26
O. nivara	AA	101697	93519	195216	169796	89.3	83.3	16.7
O. rufipogon	AA	99208	89759	188967	165682	89.9	84.2	15.8
O. glaberrima	AA	101697	82153	183850	140436	79.4	71.3	28.7
O. punctata	BB	96077	67859	163936	78500	43.7	28.3	71.7
O. officinalis	CC	147172	135027	282199	118216	24.3	14.7	85.3
O. australiensis	EE	168972	134149	303121	95732	24.4	12.4	87.6
O. brachyantha	FF	97309	65845	163154	46802	13.2	4.7	95.3
O. granulata	GG	96416	117622	214038	51978	12.7	5.3	94.7

*estimated as described by Dubcovsky and Dvorak (2007) and as described in methods.

Supplemental Table 8. List of Intact LTR Retrotransposons and Solo LTRs; Presence or Absence of TSDs, and Their coordinates.

Species	Type	No. of intact retro/similarity to known TE	Coordinate	Character	Size	Divergence	Insertion time (MYA)
O. sativa ssp *japonica* [AA]							
intact LTR		6					
	LTR/*copia*	Os3040LTR-COPIAA	5719376-5719647------5726676-5726932	TGT---ACA; TSD: TAAAC	7557	0.02	0.96
	LTR/*copia*	Os0076LTR-SZ18	5602680-5603629------5606488-5607196	TGT---ATA; TSD: ATATA	4423	0.04	1.40
	LTR/*copia*	Os0282LTR-Copia-D	5530144-5530646------5535517-5536019	TGT---ACA; TSD: TTTTG	5445	0.00	0.15
	LTR/*copia*	Os0196LTR-Gypsy-B	5555894-5556595------5562259-5562960	TGT---CCA; TSD: GTAAT---CTAAT	7066	0.03	1.18
	LTR/*copia*	Os0946LTR	5701860-5705226------5707443-5710809	TGT---ACA; TSD: TTATC	8949	0.00	0.01
	LTR/*copia*	Os0119LTR-RETROSAT2LTR	5482617-5485688------5491080-5494150	TGT---ACA; TSD: TTAGC	11533	0.00	0.01
soloLTR		7					
	LTR/*gypsy*	Os0019LTR-RC1067	5915014-5916510	TGT---ACA; TSD: AACTT---AATTT	1496		
	LTR/*gypsy*	Os0165LTR-SZ31	5771606-5772956	TGT---ACA; no TSD	1350		
	LTR/*gypsy*	Os0985LTR-Gypsy-B	5371152-5371806	TGT---ACA; TSD:GAAAG---GGAAG	654		
	LTR/*gypsy*	Os0019LTR-RC1067	5348620-5350129	TGT---ACA; TSD: GCCGC	1509		
	LTR/TRIM	Os0829LTR	5744466-5745247	TGT---ACA; TSD: GACAC	781		
	LTR/TRIM	Os0281LTR-SZ105	5689552-5689854	TGT---TGA; no TSD	303		
	LTR/unknown	Os0020LTR-BajieLTR	5903378-5904108	TGT---ACA; TSD: CGATG	730		
O. sativa ssp *indica* [AA]							
intact LTR		4					
	LTR/*gypsy*	Os0016LTR-SZ2	178132-179228---189924-191020	TGT---TCA; TSD: GTGGT	12888	0.00	0.03
	LTR/*gypsy*	Os0030LTR-RIRE2	35349-35789---49314-49754	TGT---ACA; TSD: CGTAC	14405	0.00	0.09
	LTR/*copia*	Os0083LTR-SZ9LTR	289601-290566---295035-296000	GT---ACA; no TSD	6397	0.01	0.28
	LTR/*copia*	Os3040LTR-COPIAA	103258-103511---113698-113969	TGT---ACA; no TSD	7318	0.02	0.96
soloLTR		9					
	LTR/*gypsy*	Os0200LTR-SZ107	126039-129790	TGT---ACA; TSD:AATAG	3751		
	LTR/*gypsy*	Os0119LTR-RETROSAT2	106534-109673	TGT---ACA; TSD: ATTCC	3139		
	LTR/*gypsy*	Os0139LTR-SZ106	76959-83963	TGT---ACA; TSD: ACAGC	3964		
	LTR/*gypsy*	Os0376LTR-SZ102	77992-81032	TAT---ACA; TSD: CAACC	3040		
	LTR/*gypsy*	Os0419LTR-SZ	12605-13793	TGT---TCA; TSD: GTGTG	1188		
	LTR/*copia*	Os0076LTR-SZ18	238915-239610	TGT---GCA; TSD: ATATA	695		
	LTR/*copia*	Os0245LTR-SC-25	76391-84181	TGT---ACA; TSD: GACAC	786		
	LTR/*copia*	Os0101LTR-SZ25	6078-6408	TGA---ACA; TSD: GCTTA-ACTTA	330		
	LTR/TRIM	Os0281LTR-SZ105	132964-133268	TCA---ACA; no TSD	305		
O. nivara [AA]							
intact LTR		1					
	LTR/*copia*	Os0988LTR-Copia	7377-7630---14672-14943	TGT---ACA; no TSD	7566	0.03	1.12
soloLTR		3					
	LTR/*gypsy*	Os0264LTR-SZ101	26996-30736	TGT---ACA; TSD: AATAG	3740		
	LTR/TRIM	Os0281LTR-SZ105	33902-34206	TGT---ACA; no TSD	305		
	LTR/*copia*	Os0076LTR-SZ18	129869-130564	TGT---GCA; TSD: ATATA	696		

O. rufipogon [AA]							
intact LTR		1					
	LTR/*copia*	Os0988LTR-Copia	3676-3929---10970-11241	TGT---ACA; no TSD	7565	0.03	1.12
soloLTR		3					
	LTR/*gypsy*	Os0264LTR-SZ101	23296-27037	TGT---ACA; TSD: AATAG	3741		
	LTR/TRIM	Os0281LTR-SZ105	30203-30507	TGT---TGA; no TSD	305		
	LTR/*copia*	Os0076LTR-SZ18	126181-126876	TGT---GCA; TSD: ATATA	696		
O. glaberrima [AA]							
soloLTR		2					
	LTR/*copia*	Os0076LTR-SZ18	113579-114291	TGT---ACA; TSD: ATATA	712		
	LTR/TRIM	Os0281LTR-SZ105	21864-22167	TCA---ACA; no TSD	304		
O. punctata [BB]							
intact LTR		1					
	LTR/unknown	Os0173LTR-Dasheng	87956-88392---94504-94729; 2nd LTR incomplete	TGT---ACA; TSD: CATTG	6773	0.03	1.05
soloLTR		9					
	LTR/*gypsy*	Os0119LTR/Os0329LTR-RETRO	73379-78257	TGG---ACA; no TSD	2682		
	LTR/*gypsy*	Os3344LTR-SZ7	117595- 118406	TGT---GCA; TSD: CATGG-CCTGG	811		
	LTR/*gypsy*	Os3312LTR-Osr31	104433-108315	TGA---TCA; TSD: AGCTC	3589		
	LTR/*copia*	Os0176LTR-SZ37	17061-18529	TGA---ACA; TSD: CTCTC	1468		
	LTR/*copia*	Os0176LTR-SZ37	18641-20186	TGT---TCA; no TSD	1545		
	LTR/*copia*	Os3340LTR-SZ61	75929-78125	TGA---TCA; TSD: AAGAC	2196		
	LTR/*copia*	Os0253LTR-SZ13L	79225-82255	TGA---CCA; TSD: ACCAC	1560		
	LTR/unknown	Os0310LTR-SZ29	80568-82037	TGT---ATA; TSD: CTAGT	1469		
	LTR/unknown	Os1002LTR	107127-107419	TGT---CCA; no TSD	292		
O. officinalis [CC]							
soloLTR		2					
	LTR/*gypsy*	Os0077LTR-SZ35	56805-59859	TGT---ACA; TSD: TGAGG	3054		
	LTR/Gypsy	Os0046LTR-SZ45	3848-5266	TGT---ACA; TSD: CCGTT	1418		
O. australiensis-OA_CBa0016E12 [EE]							
intact LTR		2					
	LTR/*copia*	Os0195LTR-RIRE1	130500-132054---146421-147992	TGA---ACA; TSD: CCAAC	13078	0.02	0.60
	LTR/unknown	Os0030TR-RIRE2	207744-208224---218988-219295	TGT---TCA; TSD: GTTTG	11551	0.01	0.50
soloLTR		5					
	LTR/*gypsy*	Os0019LTR-RC1067	185392-186946	TCT---ACA; no TSD	1582		
	LTR/*gypsy*	Os0240LTR-Gypsy-A	136672-139913	TGT---ACA; TSD: ACATC	3241		
	LTR/*copia*	Os0195LTR-RIRE1	150980-152618	TGT---TCA; TSD: GATGT	1638		
	LTR/*copia*	Os0195LTR-RIRE1	86985-88503	TGA---ACA; TSD: GGCTT	1578		
	LTR/unknown	Os0310LTR-SZ29	19751-21302	TGT---ACA; no TSD	727		
O. australiensis-OA_CBa0062H21 [EE]							
intact LTR		9					
	LTR/*gypsy*	Os0019LTR-RC1067	94-1529---35712-37154	TGT---ACA; TSD: CATAT---TATAT	13419	0.08	3.06
	LTR/*gypsy*	Os0019LTR-RC1067	2450-3943---24621-26091	TGT---ACA; TSD: GCGGC---ACGGC	13511	0.05	1.96
	LTR/*copia*	Os1031LTR-SZ5	67194-68287---71866-72960	TGT---ACA; TSD: CGGTA	5766	0.01	0.28
	LTR/*copia*	Os0195LTR-RIRE1	81917-83098---87957-89144	GGG---AAA; no TSD	7227	0.03	1.17
	LTR/*copia*	Os0008LTR-SZ5	109209-110200---114096-	TGT---ACA; TSD:	5765	0.09	3.60

			114974	CAGAT			
	LTR/*copia*	Os1031LTR-SZ5	118710-119918---124414-125629	TGT---AGA; TSD: GTTGT	6919	0.07	2.69
	LTR/*copia*	Os0156LTR-SC-3	130734-131052---136068-136414	TGA---ACA; no TSD	5680	0.02	0.61
	LTR/unknown	Os0173LTR-Dasheng	7599-8044---17285-17729	TGT---ACA; TSD: ATAAA	10130	0.02	0.70
	LTR/unknown	Os0173LTR-Dasheng	54656-55126---58265-58735	TGT---ACA; TSD: TATTT	4079	0.01	0.33
soloLTR		9					
	LTR/*gypsy*	Os0119LTR-RETROSAT2	41710-44559	TGT---ACA; TSD: GCCGA---GCCGG	2849		
	LTR/*gypsy*	Os1082LTR	165134-166603	TGT---ACA; no TSD	1469		
	LTR/*gypsy*	Os0264LTR-SZ101	167201-170714	TGT---ACA; TSD: TCAGA	3513		
	LTR/*gypsy*	Os0391LTR-RCS1	185155-186036	TGA---TCA; TSD: GTCTT	881		
	LTR/*gypsy*	Os0019LTR-RC1067	194135-195612	TGT---ACA; TSD: ACCAT	1477		
	LTR/*gypsy*	Os0019LTR-RC1067	200497-201966	TGT---ACA; TSD: CCAGG	853		
	LTR/*copia*	Os1031LTR-SZ5	77199-78384	TGT---ACA; TSD: TATCA	1185		
	LTR/*copia*	Os0356LTR-SZ27	127062-127329	TGT---ACA; no TSD	267		
	LTR/*copia*	Os0008LTR-SZ5	163983-170904	TGT---AC; no TSD	1173		
O. brachyantha [FF]							
intact LTR		1					
	LTR/*copia*	Os0356LTR-SZ27	80879-81164---85462-85773	TGT---ACA; TSD: TTCCC---TTCTC	4894	0.07	2.85
soloLTR		6					
	LTR/*copia*	Os0171LTR-SZ30	1345-2266	TGT---AGA; no TSD	921		
	LTR/*copia*	Os0356LTR-SZ27	9297-9619	TGT---ACA; TSD: GAGCC---GGGCC	322		
	LTR/*copia*	Os1031LTR-SZ5	26842-27967	TGT---ACA; TSD: GAAAG	990		
	LTR/TRIM	Os0281LTR-SZ105	32409-32683	TGT---ACA; TSD: AGATA	274		
	LTR/unknown	Os0173LTR-Dasheng	135326-135740	TGT---ATA; TSD: ACTTG	414		
	LTR/unknown	Os0173LTR-Dasheng	198005-198435	TGT---ACA; TSD: AATAT---AATAA	430		
O. granulata [GG]							
intact LTR		10					
	LTR/*gypsy*	Os0419	21361-22381---28431-29450	TGT---ACA; TSD: ACAGG	8089	0.01	0.46
	LTR/*gypsy*	Os0077	101014-103624---108677-111176	TGT---ACA; TSD: CATGG	10162	0.00	0.11
	LTR/*gypsy*	Os1082	269476-270813---279364-280701	TGT---ACA; TSD: GAGCC	11225	0.04	1.48
	LTR/*copia*	Os1036	50920-51314--- 52315-52712	TGT---ACA; TSD: GTATT---TTATT	1792	0.26	10.03
	LTR/*copia*	Os0466	67399-68612---74228-75085	TGA---TCA; TSD: ACTGC	5110	0.02	0.85
	LTR/*copia*	Os0176	120771-122317---131064-132601	TGT---TTA; TSD: GCCTC	11830	0.07	2.59
	LTR/*copia*	Os1214	235696-236091---240863-241258	TGT---ACA; TSD: ATCTT	5562	0.01	0.49
	LTR/unknown	Os0173	136217-136635---150503-150919	TGG---ACA; no TSD	13566	0.17	6.40
	LTR/unknown	Os0173	227698-228132---231883-232310	TGT---ACA; TSD: ATCGG	4612	0.05	1.98
	LTR/unknown	Os0972	184097-184526---188667-189097	TGT---ACA; TSD: CCACT	5000	0.11	4.04
soloLTR		5					
	LTR/*gypsy*	Os0141	264977-268099	TGT---ACA; TSD: GTACA---GTAGA	3122		
	LTR/*gypsy*	Os1082	75760-78470	TGG---ACA; no TSD	2710		
	LTR/*gypsy*	Os0119	218445-221617	TGT---ACA; no TSD	3172		
	LTR/*copia*	Os0171	147928-148683	TGT---ACA; TSD: TACCT	755		
	LTR/*copia*	Os0984	13767-15515	TGT---ATA; TSD: GCCGC	399		

Supplemental Table 9. Summary of Comparisons of the Number of Shared and Unshared Genes in the across the *Adh1* Genus-wide Vertical Sequence Dataset.

Species comparison	Total genes	Unshared genes	% of unshared genes
japonica-indica	37	3	8.1
japonica-nivara	39	1	2.6
japonica-rufipogon	25	1	4.0
japonica-glaberrima	24	2	8.3
japonica-punctata	25	7	28.0
japonica-officinalis	25	3	12.0
japonica-australiensis	51	19	37.3
japonica-brachyantha	56	28	50.0
japonica-granulata	46	34	73.9

Supplemental Table 10. List of Shared or Unshared Genes or Gene Families Across the *Adh1* Genus-wide Vertical Sequence Dataset.

	gene/ gene family#	**Gene family name**	***Species***									
			J	**I**	**N**	**R**	**G**	**P**	**O**	**A**	**B**	**GR**
	1	UP1 protease	√	ND	ND	ND	ND	ND	ND	√	ND	ND
	2	NBS-LRR	√	√	ND	ND	ND	ND	√	χ	√	χ*
	3	serine carboxy peptidase	√	ND	ND	ND	ND	ND	ND	√	χ	ND
	4	tetratricopeptide	√	ND	ND	ND	ND	ND	ND	√	χ	ND
	5	protein kinase	√	ND	ND	ND	ND	ND	ND	√	√	χ*
Core	6	Adh	√	√	√	√	√	√	√	√	√	√
	7	RZ53	√	√	√	√	√	√	√	√	√	√
	8	peroxidase	√	√	√	√	√	√	√	√	√	√
	9	NifS	√	√	√	√	√	√	√	√	√	√
	10	PIK	√	√	√	√	√	√	√	√	√	√
	11	F-box	√	√	√	√	√	√	√	√	√	χ
	12	LRRkinase/ PACKMULE	√	√	√	√	ND	χ	ND	χ	χ	χ
	13	SRKinase	√	√	√	ND	ND	√	ND	√	χ*	χ*
	14	FMO	√	ND	ND	ND	ND	ND	ND	ND	√	χ
	15	MYB	√	ND	ND	ND	ND	ND	ND	ND	√	χ
	16	unknown	√	ND	ND	ND	ND	ND	ND	ND	ND	χ
	17	MAP kinase	√	ND	ND	ND	ND	ND	ND	ND	ND	χ
	18	SEU3	√	ND	ND	ND	ND	ND	ND	ND	ND	χ
	19	alpha-G-protein	√	ND	ND	ND	ND	ND	ND	ND	ND	χ
	20	mTERF	√	ND	ND	ND	ND	ND	ND	√	ND	√

J = *japonica*, I = *indica*, N = *O. nivara*, R = *O. rufipogon*, G = *O. glaberrima*, P = *O. punctata*, O = *O. officinali*s, A = *O. australiensis*, B = *O. brachyantha* and GR = *O. granulata*.

Presence or absence of individual gene or whole gene families (*O. sativa* ssp japonica reference sequence) is indicated. A detailed description of the presence or absence of each individual family member(s) are indicated in the Figure 1.

ND = not determined because of lack of sequence coverage

√ = present

χ = absent

* Present but no phylogenetic relationship because of sequence divergence

Supplemental Table 11. Summary of Intact LTR Retrotransposons and Solo LTRs Based on the Presence or Absence of Target Site Duplications (TSD).

Status of Retrotransposon	Total No.	%
Intact Element with TSDs	28	80
Intact Element without TSDs	7	20
soloLTR with TSDs	43	71.7
soloLTR without TSDs	17	28.3

Supplemental Table 12. Detailed Breakdown of Intact LTR Retrotransposons and Solo LTRs Based on the Presence or Absence of Target Site Duplications.

Status of Retrotransposon	***Oryza* species**																					
	ssp. japonica		***ssp. indica***		***nivara***		***rufipogon***		***glaberrima***		***punctata***		***officinalis***		***australiensis*-1**		***australiensis*-2**		***brachyantha***		***granulata***	
	No.	**%**	**No.**	**%**	**No.**	**%**	**No.**	**%**	**No.**	**%**	**No.**	**%**	**No.**	**%**	**No.**	**%**	**No.**	**%**	**No.**	**%**	**No.**	**%**
Intact Element with TSDs	6	100	2	50	0	0	0	0	0	0	1	100	0	0	2	100	7	77.8	1	100	9	90
Intact Element without TSDs	0	0	2	50	1	100	1	100	0	0	0	0	0	0	0	0	2	22.2	0	0	1	10

Supplemental Table 13. Molecular Timing of Various Duplication Events.

a). Molecular Timing of Duplication for Gene Family #13 in ssp. *japonica*

Seq1	Seq2	SynDif	SynPos	Ks	NSynDif	NSynPos	Ka	Ka/Ks	Mol. clock
13-1_JAP	13-1_IND	1	163.33	0.0061	1	523.67	0.0019	0.3115	0.47
13-1_JAP	13-2_JAP	39.83	161.58	0.2989	35.17	525.42	0.0701	0.2345	23
13-1_JAP	13-3_JAP	53	164.42	0.4213	55	522.58	0.1134	0.2692	32.4
13-2_JAP	13-3_JAP	52.33	162.67	0.4202	59.67	524.33	0.1234	0.2937	32.3

b). Molecular Timing of Duplication for Gene Family #6 in ssp. *japonica*

Seq1	Seq2	SynDif	SynPos	Ks	NSynDif	NSynPos	Ka	ka/ks	Mol. clock
6-3-japonica	6-2-Japonica	120.42	268.08	0.6852	73.58	865.92	0.090	0.132	52.71
6-3-japonica	6-1-japonica	151.5	269.5	1.0383	79.5	864.5	0.098	0.094	79.87
6-2-Japonica	6-1-japonica	125.17	263.25	0.7538	69.83	870.75	0.085	0.112	57.98

c). Molecular Timing of Duplication in Gene Family #7 in *O. brachyantha* [BB].

Seq1	Seq2	SynDif	SynPos	Ks	NSynDif	NSynPos	Ka	ka/ks	Mol. clock
B7-1	B7-2	10.17	46.17	0.2607	14.83	160.83	0.0984	0.377445	20.1