

Supplementary information

Relationships of wild and domesticated rices (*Oryza* AA genome species) based upon whole chloroplast genome sequences

Peterson W. Wambugu¹, Marta Brozynska¹, Agnelo Furtado¹, Daniel L. Waters² and Robert J. Henry^{1*}

¹Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Brisbane, St Lucia, Qld, 4072, Australia

²Southern Cross Plant Science, Southern Cross University, Lismore, NSW, Australia

Supplementary Tables

Supplementary Table S1: Details of seed samples used for DNA extraction

Species	Accession number	Country of origin
<i>Oryza glaberrima</i>	104040	Chad
<i>Oryza longistaminata</i>	103905	Tanzania
<i>Oryza longistaminata</i>	81953	Zambia
<i>Oryza barthii</i>	103590	Cameroon
<i>Oryza barthii</i>	101381	Niger
<i>Oryza barthii</i>	104124	Chad
<i>Oryza barthii</i>	86481	Zambia
<i>Oryza glumaepatula</i>	AusTRCF 309281	Suriname
<i>Oryza officinalis</i>	AusTRCF 309302	Myanmar

Supplementary Table S2: Alignment statistics for whole chloroplast genome and with one inverted repeat (IR) excluded

Statistics/parameter	Whole cp genome	Cp genome with one IR excluded ^a
Alignment length	143,331bp	114,151bp
Conserved sites	134,353	113,591
Variable sites	484	439
Parsimony informative sites	221	214
Singleton sites	262	225
GC content	39.0	38.0

^a *Oryza glaberrima* 2 was excluded in this analysis as it was not possible to accurately identify the inverted repeats (IR) due to suspected assembly errors.

Supplementary Table S3: GenBank sequences used in this study

Species	Length (bp)	Accession number
<i>O. sativa</i> sp. <i>Japonica</i> 1	134,551	GU592207
<i>O. sativa</i> sp. <i>Japonica</i> 2	134,525	NC_001320
<i>O. sativa</i> sp. <i>Indica</i> 1	134,448	JN861109
<i>O. sativa</i> sp. <i>Indica</i> 2	134,496	NC_008155
<i>O. rufipogon</i> 1	134,537	KF562709
<i>O. rufipogon</i> 2	134,557	NC_022668
<i>O. rufipogon</i> 3	134,,557	JN005833
<i>O. meridionalis</i>	134,558	NC_016927
<i>O. nivara</i>	134,494	NC_005973
<i>O. glaberrima</i> 2	132,629	KJ513090

Supplementary results

Sequencing of 9 genomes produced a varying number of paired reads that ranged from 34562744 - 38275022. Reference guided assembly resulted in mapping rates that ranged from 5.1% to 12.6% which is equivalent to chloroplast genome fold coverage of between 1319X and 3443X (Table 1). The assembled genomes had a quadripartite structure comprising of a long single copy (80512-80684bp) and a short single copy (12336-12386bp) separated by a pair of inverted repeat regions (20794-20809bp) (Fig. 1). Each genome had a total of 124 unique genes comprising of 83 protein coding genes, 33 transfer RNA (tRNA) and 8 ribosomal RNA (rRNA) genes.