The Complete Chloroplast Genome of Wild Rice (*Oryza minuta*) and Its Comparison to Related Species

Sajjad Asaf†1, Muhammad Waqas1,2†, Abdul L. Khan3, Muhammad A. Khan†, Sang-Mo Kang†, Qari M. Imran†, Raheem Shahzad†, Saqib Bilal†, Byung-Wook Yun† and In-Jung Lee*1

† School of Applied Biosciences, Kyungpook National University, Daegu, South Korea, 1 Department of Agriculture, Abdul Wali Khan University Mardan, Mardan, Pakistan, 2 Chair of Oman’s Medicinal Plants and Marine Natural Products, University of Nizwa, Nizwa, Oman

*Correspondence:
In-Jung Lee
ijlee@knu.ac.kr

† These authors have contributed equally to this work.

**Specialty section:**
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Plant Science

**Received:** 02 December 2016
**Accepted:** 20 February 2017
**Published:** 07 March 2017

**INTRODUCTION**

*Oryza minuta*, a tetraploid wild relative of cultivated rice (family Poaceae), possesses a BBCC genome and contains genes that confer resistance to bacterial blight (BB) and white-backed (WBPH) and brown (BPH) plant hoppers. Based on the importance of this wild species, this study aimed to understand the phylogenetic relationships of *O. minuta* with other *Oryza* species through an in-depth analysis of the composition and diversity of the chloroplast (cp) genome. The analysis revealed a cp genome size of 135,094 bp with a typical quadripartite structure and consisting of a pair of inverted repeats separated by small and large single copies, 139 representative genes, and 419 randomly distributed microsatellites. The genomic organization, gene order, GC content and codon usage are similar to those of typical angiosperm cp genomes. Approximately 30 forward, 28 tandem and 20 palindromic repeats were detected in the *O. minuta* cp genome. Comparison of the complete *O. minuta* cp genome with another eleven *Oryza* species showed a high degree of sequence similarity and relatively high divergence of intergenic spacers. Phylogenetic analyses were conducted based on the complete genome sequence, 65 shared genes and *matK* gene showed same topologies and *O. minuta* forms a single clade with parental *O. punctata*. Thus, the complete *O. minuta* cp genome provides interesting insights and valuable information that can be used to identify related species and reconstruct its phylogeny.

**Keywords:** wild rice (*Oryza minuta*), cp genome, repeat analysis, codon usage, phylogeny, sequence divergence, SSRs
Previously, phylogenetic analyses have been based on sequencing one or a few loci from the chloroplast. Due to the availability of complete chloroplast sequences in public databases and advances in next-generation sequencing techniques, analyses based on the entire chloroplast genome are achievable and yield higher quality and more valuable information, which could reveal detailed insight into genomic organization (Martin et al., 2005). Indeed, examining the entire cp genome can resolve previously ambiguous phylogenetic relationships among species (Jansen et al., 2007; Moore et al., 2010). Due to availability of high-throughput sequencing technology as well as the comparatively small size and structural similarity of cp genomes, hundreds of sequencing projects in terrestrial plants have recently been reported (Wu, 2016b).

Rice is an important cereal crop that provides essential food and energy for more than half of the world’s population. In addition, rice is considered a model crop for studies on cereal genomics. Two species of the genus Oryza (O. sativa, and O. glaberrima) are cultivated, though there are more than 20 wild species (Evenson and Gollin, 1997; Sang and Ge, 2007). Different species are categorized into 10 genome types, six are diploid (AA, BB, CC, EE, FF, and GG) (2n = 2x = 24) and the other four are allotetraploid (BBCC, CCDD, HHJJ, and HHKK) (2n = 4x = 28) (Ge et al., 1999). About one half of the species in Oryza genus are allotetraploids that originated through interspecific hybridization and genome doubling (Vaughan, 1989; Bao and Ge, 2008; Jacquemin et al., 2013). Rice (O. sativa) with an AA genome type, is one of the most important species, and it is further divided into the subspecies japonica and indica, which are distributed globally (Chang, 1976; Wambugu et al., 2015).

Because of the importance of Oryza as a major food crop, great attention has been given to understanding the genetic makeup and phylogeny of this genus, both within the genus and species (Guo and Ge, 2005). In plants, sequencing functional genes in cpDNA (chloroplast DNA) is helpful for resolving issues related to molecular taxonomy and phylogenetic reconstruction (Jansen et al., 2007; Moore et al., 2010; Wu and Ge, 2012), and such approaches can yield vast benefits in plant breeding and conservation strategies. Currently, 10 cp genomes belonging to Oryzeae have been published (Waters et al., 2012; Brozynska et al., 2014). Some wild Oryza species are better able than cultivated Oryza species to resist biotic and abiotic stresses and attack from insect pests. Thus, cultivated species can be improved through introgression of resistance genes from wild species (Heinrichs et al., 1985). For example, resistance traits from wild O. minuta, a tetraploid wild relative of cultivated rice, have been reported. O. minuta has a BBCC genome type and exhibits significant potential to resist blast blight, bacterial blight (BB), and white-backed plant hopper (WBPH) and brown plant hopper (BPH) diseases (Vaughan, 1994). Such diseases are damaging to the growth and yield of cultivated rice. In addition, stress tolerance genes from O. minuta have been successfully transferred to cultivated rice through introgression (Amante-Bordeos et al., 1992; Rahman et al., 2009). Overall, wild species such as O. minuta possess valuable genetic diversity that can contribute greatly to improving the growth and yield of various crops (Amante-Bordeos et al., 1992). To identify desirable genes and ensure effective conservation, it is essential to analyze phylogenetic and evolutionary relationships among species (Guo et al., 2013). Previously, it was reported that O. minuta was originated from allotopolyploidization of O. officinalis (paternal) and O. punctata (maternal) (Ammiraju et al., 2010; Zou et al., 2015).

In this study, we assembled for the first time the complete chloroplast genome sequence of O. minuta, and performed detailed phylogenetic analyses on the basis of complete cp genome and 65 shared genes. The complete cp genome of O. minuta, in conjunction with previously reported cp genome sequences, will improve our understanding of O. minuta and the evolutionary history of genus Oryza. Hence, we analyzed the fully assembled cp genome of O. minuta and compared it to eleven closely related species: O. australiensis EE, O. nivara, O. rufipogon, O. sativa L. ssp. indica, O. sativa L. ssp. japonica, O. barthii, O. glumipatula, O. longistaminata, O. meridionalis, O. officinalis CC, and O. punctata BB.

**MATERIALS AND METHODS**

In this study, a standard protocol for DNA extraction was used as described in detail by Sierró et al. (2014). The extracted DNA was sequenced using an Illumina HiSeq-2000 (Illumina, San Diego, CA, USA) platform at Macrogen (Macrogen, Seoul, Korea), and the O. minuta cp genome was obtained by de novo assembly of the entire genome sequence via a bioinformatics pipeline (http://phyzen.com). A 400-bp paired-end library was produced according to the Illumina PE standard protocol, generating 28,110,596 bp of total reads with a 120-bp average read length. Raw reads with Phred scores of 20 or less were removed from the total PE reads using the CLC-quality trim tool, and de novo assembly was conducted on trimmed reads using CLC Genomics Workbench v7.0 (CLC Bio, Aarhus, Denmark) with parameters of minimum (200 to 600 bp) autonomously controlled overlap size. All contigs were then mapped and assembled against the reference cp genomes of O. officinalis and O. punctata by following a previously described method (Wu, 2016a,b). Primers were designed (Table S1) to test for correct sequence assembly. PCR amplification was performed in a total volume of 20 µl containing 1× reaction buffer, 0.4 µl dNTPs (10 mM), 0.1 µl Taq (SogTM h-Taq DNA Polymerase), 1 µl (10 pm/µl) primers, and 1 µl (10 ng/µl) DNA. The PCR program consisted of initial denaturation at 95°C for 5 min followed by 35 cycles of 95°C for 30 s, 65°C for 20 s and 72°C for 30 s, with a final extension step at 72°C for 5 min. After incorporation of the sequencing results, the finished cp genome was applied as a reference to map previously obtained short reads to refine the assembly based on maximum sequence coverage.

**Genome Annotation and Sequence Architecture**

The program DOGMA was used to annotate the O. minuta cp genome (Wyman et al., 2004). The annotation results were checked manually, and codon positions were adjusted by
TABLE 1 | Summary of complete chloroplast genomes for twelve Oryza species.

| Region | O. aust | O. min | O. niv | O. rufi | O. s. ind | O. s. jap | O. offi | O. Barth | O. punc | O. meri | O. long | O. glum |
|--------|---------|--------|--------|---------|----------|----------|--------|---------|---------|---------|---------|---------|
| LSC    | 81,074  | 80,974 | 80,544 | 80,594  | 80,512   | 80,594   | 80,952  | 80,684  | 80,621  | 80,604  | 80,596  | 80,612  |
| GC(%)  | 37.07   | 37.1   | 37.12  | 37.11   | 37.09    | 37.1     | 37.1   | 37.1    | 37.1    | 37.1    | 37.1    | 37.1    |
| Length (%) | 59.95  | 59.9   | 59.8   | 59.9    | 59.8     | 59.9     | 59     | 59.9    | 59.8    | 59.9    | 59.8    | 59.8    |

| SSC    | 12,470  | 12,446 | 12,346 | 12,347  | 12,345   | 12,345   | 12,330  | 12,381  | 12,347  | 12,347  | 12,357  | 12,356  |
| GC(%)  | 33.18   | 33.3   | 33.33  | 33.3    | 33.34    | 33.3     | 33.3   | 33.3    | 33.3    | 33.3    | 33.3    | 33.3    |
| Length (%) | 9.22   | 9.2    | 9.1    | 9.1     | 9.1      | 9.1      | 9.1    | 9.1     | 9.1     | 9.1     | 9.1     | 9.1     |

| IR     | 20,840  | 20,836 | 20,802 | 20,802  | 20,795   | 20,795   | 20,813  | 20,804  | 20,797  | 20,803  | 20,807  | 20,807  |
| GC(%)  | 44.33   | 44.3   | 44.35  | 44.35   | 44.3     | 44.3     | 44.3   | 44.4    | 44.4    | 44.4    | 44.33   | 44.33   |
| Length (%) | 15.4   | 15     | 15.4   | 15.4    | 15.4     | 15.4     | 15.4   | 15.4    | 15.4    | 15.4    | 15.4    | 15.4    |

| Total  | 135,224 | 135,094| 134,494| 134,544 | 134,448  | 134,525  | 134,911| 134,674 | 134,604 | 134,558 | 134,583 | 134,583 |
| GC(%)  | 38.95   | 39     | 39.1   | 39.1    | 39       | 39       | 39     | 39      | 39      | 39      | 39      | 39      |
| Length (%) | 135,224| 135,094| 134,494| 134,544 | 134,448  | 134,525  | 134,911| 134,674 | 134,604 | 134,558 | 134,583 | 134,583 |

O. aust, O. australiensis; O. min, O. minuta; O. niv, O. nivara; O. rufi, O. rufipogon; O. sat. ind, O. sativa indica; O. s. jap, O. sativa japonica; O. offi, O. officinalis; O. Barth, O. barthii; O. punc, O. punctata; O. meri, O. meridionalis; O. long, O. longistaminata; O. glum, O. glumipatula.

comparison to homologs from the cp genomes of O. australiensis and O. sativa ssp. indica in the database. All transfer RNA sequences were verified using tRNAscan-SE version 1.21 (Schattner et al., 2005) with the default settings. OGDRAW (Lohse et al., 2007) was applied to illustrate the structural features of the O. minuta cp genome. To examine deviations in synonymous codon usage by avoiding the influence of amino acid composition, the relative synonymous codon usage (RSCU) was determined using MEGA 6 software (Kumar et al., 2008). mVISTA software was used in the Shuffle-LAGAN mode to compare the complete variation in the O. minuta cp genome with eleven other cp genomes using the O. minuta annotation as a reference (Frazer et al., 2004).

Characterization of Repeat Sequences and SSRs

We employed REPuter to identify repeat sequences, including palindromic, reverse, and direct repeats, within the cp genome (Kurtz et al., 2001). The following settings for repeat identification were used: (1) Hamming distance of 3; (2) 90% or greater sequence identity; (3) a minimum repeat size of 30 bp. Phobos version 3.3.12 (Leese et al., 2008) was used to detect (SSRs) within the cp genome, with the search parameters set at ten repeat units ≥10 for mononucleotides, eight repeat units ≥8 for dinucleotides, four repeat units ≥4 for trinucleotides and tetrancleotides, and three repeat units ≥3 for pentanucleotide and hexanucleotide SSRs. Tandem repeats in the O. minuta cp genome were identified using Tandem Repeats Finder version 4.07 b (Benson, 1999) with the default settings.

Sequence Divergence and Phylogenetic Analysis

Complete cp genomes as well as a separate partition using only 65 shared genes were employed to analyze the average pairwise sequence divergence for 11 Oryza species: O. australiensis, O. nivara, O. rufipogon, O. sativa L. ssp. indica, O. sativa L. ssp. japonica, O. barthii, O. glumipatula, O. longistaminata, O. meridionalis, O. officinalis, and O. punctata. Missing and ambiguous gene annotations were confirmed by comparative sequence analysis after a multiple sequence alignment and gene order comparison. These regions were aligned using MAFFT (version 7.222) (Katoh and Standley, 2013) with the default parameters. Kimura’s two-parameter (K2P) model was selected to calculate pairwise sequence divergences (Kimura, 1980). To resolve the O. minuta phylogenetic position within the rice tribe (Oryzeae), 13 published cp genomes were downloaded from the NCBI database for analyses. First, multiple alignments were performed using the complete cp genomes based on the conserved structure and gene order of the chloroplast genomes (Wicke et al., 2011). Four methods were employed to construct phylogenetic trees, including Bayesian inference (BI) implemented with MrBayes 3.12 (Ronquist and Huelsenbeck, 2003), maximum parsimony (MP) with PAUP 4.0 (Swofford, 1993), and maximum likelihood (ML) and neighbor-joining (NJ) with MEGA 6 (Kumar et al., 2008) using described settings (Wu et al., 2015; Asaf et al., 2016a). In the second phylogenetic analysis, 65 shared genes from the cp genomes of 12 Oryza species and two Zizania outgroup species were aligned in ClustalX using the default settings, followed by manual adjustment to preserve reading frames. The above four phylogenetic-inference methods were used to infer trees from the 65 concatenated genes using the same settings (Wu et al., 2015; Asaf et al., 2016a).

RESULTS AND DISCUSSION

Chloroplast Genome Organization of O. minuta

The O. minuta cp genome was assembled by mapping all Illumina reads to the draft cp genome sequence using CLC Genomics Workbench v7.0. A total of 1,577,251 reads were obtained, with
an average length of 120 bp, for 504.211X coverage of the cp genome. The consensus sequence for a specific position was generated by assembling reads mapped with at least 875 reads per position and was used to construct the complete sequence of the *O. minuta* cp genome. The complete *O. minuta* cp genome is 135,094 bp in size (GenBank: KU179220), which is similar to the already reported cp genome sizes of related *Oryza* species and is within the range of other angiosperms (Yang et al., 2010). The cp genome possesses a typical quadripartite structure, which includes a pair of inverted repeats (IRa and IRb 20,836 bp) and separate SSC (12,446 bp) and LSC (80,974 bp) regions (Table 1, Figure 1). The GC content (39%) of the *O. minuta* cp genome is very similar to that of other *Oryza* species cp genomes (Table 1) (Wu et al., 2015). However, the GC content is unequally distributed in the *O. minuta* cp genome: it is highest in the IR regions (44.3%), moderate in the LSC regions (37.1%) and lowest in the SSC regions (33.3%). This high IR GC percentage is due to the presence of eight ribosomal RNA (rRNA) sequences in these regions. These results are similar to a previously reported high GC percentage in IR regions (Qian et al., 2013).

A total of 139 genes were found in the *O. minuta* cp genome, of which 110 are unique, including 91 protein-coding genes, 40 tRNA genes, and 8 rRNA genes (Figure 1, Table 2). Of these, 11 protein-coding, four rRNA, and eight tRNA genes are duplicated in the IR regions. The LSC region comprises 62 protein-coding
TABLE 2 | Genes in the sequenced *O. minuta* chloroplast genome.

| Category | Group of genes | Name of genes |
|----------|----------------|---------------|
| Self-replication | Large subunit of ribosomal proteins | rpl2, 14, 16, 20, 22, 23, 32, 33, 36 |
| | Small subunit of ribosomal proteins | rps2, 3, 4, 7, 8, 11, 12, 14, 15, 16, 18, 19 |
| | DNA dependent RNA polymerase | rpoA, B, C1, C2 |
| | rRNA genes | RNA |
| | tRNA genes | tRNA-UAG, tRNA-GCA, tRNA-D-GUC, tRNA-E-UUC, tRNA-F-GAA, tRNA-M-CAU, tRNA-G-AUC, tRNA-G-GUG, tRNA-CAU, tRNA-G-GAU, tRNA-UAA, tRNA-L-UAG, tRNA-M-CAU, tRNA-V-GU1, tRNA-P-GGG, tRNA-UUG, tRNA-Q-UUG, tRNA-A-CG, tRNA-U-UCU, tRNA-S-GCU, tRNA-S-GGA, tRNA-S-UGA, tRNA-T-GGU, tRNA-T-UGU, tRNA-V-GAC, tRNA-W-CCA, tRNA-Y-GUA |
| Photosynthesis | Photosystem I | psaA, B, C, I, J |
| | Photosystem II | psbA, C, D, E, F, H, I, K, L, M, N, T, lhbA |
| | NadH oxidoreductase | ndhA, B, C, D, E, F, G, H, I, J, K |
| | Cytochrome b6/f complex | petA, B, D, G, L, N |
| | ATP synthase | atpA, B, E, F, H, I |
| | Rubisco | rbcL |
| Other genes | Translational initiation factor | infA |
| | Maturase | matK |
| | Protease | clpP |
| | Envelop membrane protein | cemA |
| | Subunit Acetyl-CoA-Carboxylate | accD |
| | c-type cytochrome synthesis gene | ccsA |
| Unknown | Conserved Open reading frames | ycf2, 3, 4, 15, 68 |

TABLE 3 | Comparison of coding and non-coding region sizes among twelve *Oryza* species.

| Region | *O. aust* | *O. min* | *O. niv* | *O. rufi* | *O. s. ind* | *O. s. jap* | *O. offi* | *O. barth* | *O. punc* | *O. meri* | *O. long* | *O. glum* |
|--------|-----------|----------|---------|---------|---------|-----------|---------|----------|---------|---------|---------|---------|
| PROTEIN CODING | | | | | | | | | | | | |
| Length (bp) | 59,700 | 61,062 | 68,598 | 56,133 | 61,464 | 66,444 | 59,385 | 59,296 | 62,964 | 55,329 | 59,499 | 59,496 |
| GC(%) | 39.3 | 39.5 | 39.7 | 39.3 | 39.5 | 39.6 | 39.4 | 39.4 | 39.3 | 39.1 | 39.3 | 39.3 |
| Length (%) | 44.1 | 45.1 | 51 | 41.7 | 45.7 | 49.3 | 44 | 44 | 59.8 | 41.1 | 44.2 | 44.2 |
| tRNA | | | | | | | | | | | | |
| Length (bp) | 2,866 | 3,031 | 2,865 | 2,772 | 2,795 | 2,784 | 2,474 | 2,474 | 3,043 | 3,049 | 2,474 | 2,474 |
| GC(%) | 53.2 | 52.1 | 53.7 | 52.3 | 53 | 52.9 | 52.7 | 52.7 | 52.7 | 52.7 | 52.7 | 52.7 |
| Length (%) | 2.1 | 2.2 | 2.1 | 2 | 2 | 2 | 1.83 | 1.83 | 2.2 | 2.2 | 1.83 | 1.83 |
| rRNA | | | | | | | | | | | | |
| Length (bp) | 9,190 | 9,190 | 9,190 | 9,190 | 9,190 | 9,190 | 9,190 | 9,190 | 9,190 | 9,190 | 9,190 | 9,190 |
| GC(%) | 54.8 | 54.8 | 54.8 | 54.8 | 54.7 | 54.8 | 54.8 | 54.8 | 54.8 | 54.8 | 54.8 | 54.8 |
| Length (%) | 6.7 | 6.8 | 6.8 | 6.8 | 6.8 | 6.8 | 6.8 | 6.8 | 6.8 | 6.8 | 6.8 | 6.8 |
| Intergenic | 63,468 | 61,811 | 53,841 | 66,449 | 60,999 | 56,115 | 63,814 | 63,625 | 59,407 | 66,990 | 63,404 | 63,423 |
| GC(%) | 36 | 36 | 37 | 37 | 36 | 36 | 35 | 35 | 36 | 36 | 37 | 35 |
| Length (%) | 47 | 45.8 | 41 | 50 | 45.4 | 41.8 | 47.4 | 47.3 | 44.2 | 49.8 | 47.2 | 47.2 |

and 24 tRNA genes, whereas the SSC region comprises 11 protein-coding genes and one tRNA gene. The protein-coding genes present in the *O. minuta* cp genome include nine genes encoding large ribosomal proteins (*rpl2, 14, 16, 20, 22, 23, 32, 33, 36*), 12 genes encoding small ribosomal proteins (*rps2, 3, 4, 7, 8, 11, 12, 14, 15, 16, 18, 19*), five genes encoding photosystem I components (*psaA, B, C, I, J*), 10 genes related to photosystem II (*Table 2*), and six genes (*atpA, B, E, F, H, I*) encoding ATP synthase and electron transport chain components (*Table 2*). A similar pattern of protein-coding genes is also present in *O. sativa* (Zhang et al., 2012) and *O. glaberrima* (Wambugu et al., 2015). There are 11 intron-containing genes, 10 of which contain one intron, with only *ycf3* genes having two introns (*Table S2*). The *ndhA* gene has the longest intron (965 bp).
Asaf et al. Chloroplast Genome of Wild Rice

Protein, rRNAs, and tRNAs are encoded by 45.1, 6.83, and 2.2% of the entire cp genome, respectively, and the remaining 45.8% is composed of non-coding regions (Table 3). The total protein-coding sequences (CDSs) are 60,948 bp in length and consist of 91 genes encoding 20,354 codons (Tables 1, 4). The O. minuta cp genome codon usage frequency was determined based on tRNA and protein-coding gene sequences (Table 5). Leucine (10.7%) and cysteine (1.2%) are the maximum and minimum commonly encoded amino acids, and isoleucine, serine, glycine, arginine and alanine are encoded by 7.9, 7.5, 7.4, 6.5, and 6.1% of CDSs, respectively (Figure S1). Similar ratios for amino acids are present in previously reported cp genomes (Qian et al., 2013; Chen et al., 2015).

Among these, the maximum and minimum codons used are ATT (820), encoding isoleucine, and TTG and ATT (1, 1), encoding methionine. The AT content is 52.5, 60.0, and 68.7% at the 1st, 2nd, and 3rd codon positions, respectively (Table 4). The preference for a high AT content at the 3rd codon position is similar to the A and T concentrations reported in various terrestrial plant cp genomes (Morton, 1998; Nie et al., 2012; Qian et al., 2013). In total, 42.65 and 57% of all

### Table 4: Base compositions in the O. minuta cp genome.

|          | T/U | C   | A   | G   | Length (bp) |
|----------|-----|-----|-----|-----|-------------|
| Genome   | 30.4| 19.4| 30.7| 19.6| 135,094     |
| LSC      | 31.6| 18.3| 31.3| 18.8| 80,974      |
| SSC      | 30.8| 17.3| 35.9| 16.0| 12,446      |
| IR       | 27.7| 23.1| 28.1| 21.3| 20,836      |
| tRNA     | 23.5| 26.1| 24.3| 26.0| 3,031       |
| rRNA     | 22.6| 27.4| 22.6| 27.4| 9,190       |
| Protein-coding genes | 29.9| 19.5| 30.5| 20.0| 60,948      |
| 1st position | 23.27 | 19.0 | 29.3 | 28.2 | 20,354     |
| 2nd position | 32.72 | 21.1 | 27.3 | 18.82| 20,354     |
| 3rd position | 37.04 | 14.9 | 31.66| 16.5 | 20,354     |

### Table 5: The codon–anticodon recognition pattern and codon usage for the O. minuta chloroplast genome.

| Amino acid | Codon | No | RSCU | tRNA | Amino acid | Codon | No | RSCU | tRNA |
|------------|-------|----|------|------|------------|-------|----|------|------|
| Phe        | UUU   | 733| 1.28 | Ala  | UCA        | GCA   | 378| 1.18 | trnA-UGC |
| Phe        | UUC   | 407| 0.7  | Ala  | GCG        | 160  | 0.5|      |      |
| Leu        | UUA   | 710| 1.9  | Tyr  | UAU        | 567  | 1.5|      |      |
| Leu        | UUG   | 402| 1.1  | Tyr  | UAC        | 176  | 0.47| trnY-GUA |
| Leu        | CUU   | 473| 1.29 | Stop | UAG        | 22   | 0.74|      |      |
| Leu        | CUC   | 165| 0.4  | Stop | UGA        | 24   | 0.80|      |      |
| Leu        | CUA   | 319| 0.87 | trnL-UAG |       | Stop | UAA | 43   | 1.44 |
| Leu        | CGU   | 120| 0.32 | His  | CAU        | 351  | 1.49|      |      |
| Ile        | AUU   | 820| 1.51 | His  | CAC        | 119  | 0.50| trnH-GUG |
| Ile        | AUC   | 323| 0.5  | Gin  | CAA        | 521  | 1.53| trnQ-UUG |
| Ile        | AUA   | 485| 0.89 | Gin  | CAG        | 167  | 0.49|      |      |
| Met        | AUG   | 499| 1    | trnM-CAU |       | Asn  | AAU | 579  | 1.44 |
| Val        | GGU   | 450| 1.50 | Asn  | AAC        | 222  | 0.55| trnQ-UUG |
| Val        | GUC   | 140| 0.46 | trnV-GAC |       | Lys  | AAA | 752  | 1.44 |
| Val        | GUA   | 442| 1.47 | trnV-UAC |       | Lys  | AAG | 291  | 0.55 |
| Val        | GUG   | 163| 0.54 | Asp  | GUA        | 558  | 1.55|      |      |
| Ser        | UCU   | 383| 1.56 | Asp  | GAC        | 159  | 0.44| trnD-GUC |
| Ser        | UCC   | 304| 1.23 | Gln  | GAA        | 764  | 1.48| trnE-UUC |
| Ser        | UCA   | 254| 1.03 | Gln  | GAG        | 267  | 0.51|      |      |
| Ser        | UCG   | 120| 0.48 | Cys  | UGU        | 177  | 1.50|      |      |
| Ser        | AGU   | 306| 1.24 | Cys  | UGC        | 58   | 0.49|      |      |
| Ser        | AGC   | 105| 0.42 | trnS-GCU |       | Trp  | UGG | 356  | 1    | trnW-CCA |
| Pro        | CCU   | 351| 1.59 | Arg  | CGU        | 290  | 1.36| trnR-ACG |
| Pro        | CCC   | 190| 0.86 | Arg  | CGC        | 110  | 0.51|      |      |
| Pro        | CCA   | 236| 1.07 | Arg  | CGA        | 284  | 1.24|      |      |
| Pro        | CCG   | 105| 0.47 | Arg  | CGG        | 102  | 0.48|      |      |
| Thr        | ACU   | 455| 1.68 | Arg  | AGA        | 377  | 1.77| trnR-UCC |
| Thr        | ACC   | 208| 0.76 | Arg  | AGG        | 131  | 0.61|      |      |
| Thr        | ACA   | 294| 1.08 | Gly  | GGU        | 493  | 1.28|      |      |
| Thr        | ACG   | 124| 0.45 | Gly  | GGC        | 161  | 0.42|      |      |
| Ala        | GCU   | 553| 1.72 | Gly  | GGA        | 582  | 1.52| trnG-UCC |
| Ala        | GCC   | 189| 0.59 | Gly  | GGG        | 295  | 0.77|      |      |
types of preferred synonymous codons (RSCU > 1) ending with A and U and C and G, respectively, were found. Non-preferred synonymous codons (RSCU < 1) are 42.40 and 57.50% for C and G and A and U. Usage of the start codon AUG and UGG, the latter encoding tryptophan, has no bias (RSCU = 1) (Table 5).

Repeat Analysis
Repeat sequences, which play a role in genome rearrangements, are very helpful in phylogenetic studies (Cavalier-Smith, 2002; Nie et al., 2012). Furthermore, analyses of various cp genomes revealed that repeat sequences are essential to induce indels and substitutions (Yi et al., 2013). Repeat analysis of the O. minuta cp genome showed 20 palindromic repeats, 30 forward repeats, and 28 tandem repeats (Figure 2A). Among these, 17 forward repeats are 30–44 bp in length, with only three tandem repeats of the same length and 18 15–29 bp in length (Figures 2A–D). Similarly, 11 palindromic repeats are 30–44 bp, and 6 repeats are 45–59 bp in length (Figure 2B). Overall, 78 repeats were found in the O. minuta cp genome. Similarly, 73, 73, 76, 71 72, 78, 72, 71, 73, 77, and 74 repeat pairs were found in previously reported O. australiensis, O. nivara, O. rufipogon, O. sativa L. ssp. indica, O. sativa L. ssp. japonica, O. barthii, O. glumipatula, O. longistaminata, O. meridionalis, O. officinalis and O. punctata genomes, respectively (Figure 2A). This suggests that O. minuta is more similar to O. barthii and O. officinalis in terms of repeats. Approximately 29.4% of these repeats are distributed in protein-coding regions. Previous reports suggest that sequence variation and genome rearrangement occur due to the slipped-strand mispairing and improper recombination of these repeat sequences (Cavalier-Smith, 2002; Asano et al., 2004; Timme et al., 2007). Furthermore, the presence of these repeats indicates that the locus is a crucial hotspot for genome reconfiguration (Gao et al., 2009; Nie et al., 2012). Additionally, these repeats are an informative source for developing genetic markers for phylogenetic and population studies (Nie et al., 2012).

SSR Analysis
Simple sequence repeats (SSRs), or microsatellites, are repeating sequences of typically 1–6 bp that are distributed throughout the genome. In this study, we detected perfect SSRs in O. minuta together with 11 other Oryza species cp genomes (Figure 3A). Certain parameters were set because SSRs of 10 bp or longer are prone to slipped-strand mispairing, which is believed to be the main mechanism for SSR polymorphisms (Rose and Falush, 1998; Raubeson et al., 2007; Huotari and Korpelainen, 2012). A

![FIGURE 2](https://example.com/figure2.png)

**FIGURE 2** | Analysis of repeated sequences in twelve Oryza chloroplast genomes. (A) Total of three repeat types; (B) frequency of the palindromic repeat by length; (C) frequency of the tandem repeat by length; (D) frequency of forward repeat by length.
Asaf et al. Chloroplast Genome of Wild Rice

FIGURE 3 | Analysis of simple sequence repeats (SSRs) in twelve Oryza chloroplast genomes. (A) Number of different SSR types detected in twelve genomes; (B) frequency of identified SSR motifs in different repeat class types; (C) frequency of identified SSRs in coding regions; (D) frequency of identified SSRs in LSC, SSC and IR regions.

total of 419 perfect microsatellites were found in the O. minuta cp genome (Figure 3A). Similarly, 418, 413, 416, 416, 419, 420, 419, 419, 421, 429, and 422 SSRs were detected in O. australiensis, O. nivara, O. rufipogon, O. sativa L. ssp. indica, O. sativa L. ssp. japonica, O. barthii, O. glumipatula, O. longistaminata, O. meridionalis, O. officinalis and O. punctata, respectively (Figure 3A). The majority of SSRs in these cp genomes possess a dinucleotide repeat motif, varying in quantity from 269 in O. sativa ssp. indica to 276 in O. officinalis. Mononucleotide SSRs are the second most common, ranging from 92 in O. nivara to 100 in O. officinalis. Using our search criterion, only one pentanucleotide SSR was found in O. nivara, O. rufipogon, O. indica and O. officinalis (Figure 3A). In O. minuta, most mononucleotide SSRs are A (97%) and T (2.12.30%) motifs, with the majority of dinucleotide SSRs being A/G (47.05%) and A/T (38.60%) motifs (Figure 3B). Approximately 62% of SSRs are located in non-coding regions; approximately 4.3% are present in rRNA sequences and 2.3% in tRNA genes (Figure 3C).

Further analysis revealed that approximately 66.82% of SSRs occur in the LSC region, whereas 24.34 and 8.83% were found in IR and SSC regions, respectively (Figure 3D). These results are similar to previous reports that SSRs are unevenly distributed in cp genomes, and the findings might provide more information for selecting effective molecular markers for detecting intra- and interspecific polymorphisms (Powell et al., 1995a,b; Provan et al., 1997; Pauwels et al., 2012). Furthermore, most mononucleotides and dinucleotides are composed of A and T, which may contribute to bias in base composition, consistent with other cp genomes (Li et al., 2013). Our findings are comparable to previous reports that SSRs found in cp genome are generally composed of polythymine (polyT) or polyadenine (polyA) repeats and infrequently contain tandem cytosine (C) and guanine (G) repeats (Kuang et al., 2011). Therefore, these SSRs identified contribute to the AT richness of the O. minuta cp genome, as previously reported for various species (Kuang et al., 2011; Chen et al., 2015).
FIGURE 4 | Alignment of twelve chloroplast genome sequences. VISTA-based identity plot showing sequence identity among twelve Oryza species using O. minuta as a reference. The thick black line shows the inverted repeats (IRs) in the chloroplast genomes.
Asaf et al. Chloroplast Genome of Wild Rice

**FIGURE 5** | Pairwise sequence distances of *Oryza minuta* genes with *O. australiensis*, *O. nivara*, *O. rufipogon*, *O. sativa* L. ssp. *indica*, *O. sativa* L. ssp. *japonica*, *O. barthii*, *O. glumipatula*, *O. longistaminata*, *O. meridionalis*, *O. officinalis*, and *O. punctata*.

**Structural and Sequence Comparisons of cp Genomes in *Oryza***

Eleven complete cp genomes within the *Oryza* genus (*O. australiensis*, *O. nivara*, *O. rufipogon*, *O. sativa* L. ssp. *indica*, *O. sativa* L. ssp. *japonica*, *O. barthii*, *O. glumipatula*, *O. longistaminata*, *O. meridionalis*, *O. officinalis*, and *O. punctata*) were selected for comparison with that of *O. minuta* (135,094 bp). *O. australiensis* has the largest genome, and this difference is mostly attributed to variation in the length of the LSC region (*Table 1*). Analysis of genes with known functions showed that *O. minuta* shares 65 protein-coding genes with eleven other *Oryza* species. The number of unique genes found in *O. australiensis*, *O. nivara*, *O. rufipogon*, *O. sativa* L. ssp. *indica*, *O. sativa* L. ssp. *japonica*, *O. barthii*, *O. glumipatula*, *O. longistaminata*, *O. meridionalis*, *O. officinalis*, and *O. punctata* was 110, 100, 101, 108, 80, 104, 104, 100, 104 and 114, respectively (*Table S3*). Furthermore, the *O. minuta* cp genome has a gene content and organization that are similar to other *Oryza* species and members of Poaceae (Wicke et al., 2011); however, as for other grasses, it lacks a *ycf1* gene, and the accD gene is a truncated pseudogene. Because these genes are essential for the survival of photosynthetic plants (Drescher et al., 2000; Kode et al., 2005), they were most likely functionally transferred to the nucleus or functionally replaced by a eukaryotic gene, as observed for the accD plastid gene in other plant families (Babiychuk et al., 2011; Rousseau-Gueutin et al., 2011).

Pairwise cp genomic alignment between *O. minuta* and the 11 other genomes showed a high degree of synteny. The *O. minuta* cp genome annotation was used as a reference for plotting the overall sequence identity of the cp genomes of the 11 *Oryza* species in mVISTA (*Figure 4*), and the results revealed high sequence identity with all 11 *Oryza* species. However, except for *O. australiensis*, relatively lower identity was also observed with these species in various comparable genomic regions, particularly the *rps3*, *rpl22*, *rpl23*, *rpl2*, and *rps19* regions (*Figure 4*). In addition, the LSC and SSC regions show less similarity than the two IR regions in all *Oryza* species. In addition, non-coding regions exhibit greater divergence than coding regions. These highly divergent regions include *rbcL*, *rps16-trnQ, trnM-trnM, psbM-petN, rpoC2, atpF-atpH, ndhA rpl32, petA-psbJ, ccsA, ndhF-rpl32*, and *ycf3*. Similar results related to these genes were also reported by Qian et al. (2013). Our results also confirm similar differences among various coding regions in the analyzed species, as suggested by Kumar et al. (2009).

We compared the cp genomes and calculated the average pairwise sequence divergence among the 12 species (*Table S4*).
Of these, the *O. minuta* genome has 0.005 average sequence divergence, and high divergence was found for *O. australiensis* (0.00725); *O. officinalis* has the lowest average sequence divergence (0.0044). Furthermore, the twelve most divergent genes among these genomes are *petG*, *matK*, *infA*, *ccsA*, *rpoC2*, *clcP*, *psbE*, *rbcL*, *psbN*, *rps18*, *rpl36*, and *ndhF*. The highest average sequence distance was found for *rpoC2* (0.01983), followed by *petG* (0.0154) (*Figure 5*). Both these genes are located in LSC regions and display a trend toward more rapid evolution.

**IR Contraction and Expansion**

Expansion and contraction at the borders of IR regions are the main reason for size variations in the cp genome and play a vital role in its evolution (Raubeson et al., 2007; Wang et al., 2008; Yang et al., 2010, 2014). A detailed comparison on four junctions (*J*<sub>LA</sub>, *J*<sub>LB</sub>, *J*<sub>SA</sub>, and *J*<sub>SB</sub>) between the two IRs (IRa and IRb) and the two single-copy regions (LSC and SSC) was performed among *O. australiensis*, *O. nivara*, *O. rufipogon*, *O. sativa* L. ssp. *indica*, *O. sativa* L. ssp. *japonica*, *O. barthii*, *O. glumipatula*, *O. longistaminata*, *O. meridionalis*, *O. officinalis* and *O. punctata*. This analysis revealed that the IRs are more likely to contract than expand, which is consistent with previous studies (Raubeson et al., 2007; Wang et al., 2008; Yang et al., 2010, 2014). The results suggest that the IRs have undergone multiple contractions and expansions, leading to the observed variation in size among the chloroplast genomes of wild rice species.
Asaf et al. Chloroplast Genome of Wild Rice *O. punctata* with regard to *O. minuta* by carefully analyzing the exact IR border positions and adjacent genes (Figure 6). Despite the similar length of the *O. minuta* IR region with the other eleven *Oryza* species, from 20,836 bp to 20,840 bp, some IR expansion and contraction was observed. JLa is located between *rps19* and *psbA*, and variation in distances between *rps19* and JLa range from 40 to 49 bp across all species; the distance in *O. minuta* is 46 bp. The distance between *psbA* and JLa is 81 bp in *O. minuta*, which is similar to the other genomes (81 bp). The distance between *rpl22* and JLa varies from 23 bp to 29 bp. In *O. minuta*, 1-bp variations exist in the JSa border region compared to the other cp genomes. The *ndhH* gene traverses the SSC and IRa regions, with approximately 164 bp located in the IR region for *O. minuta*. Furthermore, there are 16-bp variations observed compared with *O. officinalis* for *ndhF*, *ndhH* and *rps15* in the SSC and IRb regions, located 41 bp, 164 bp and 302 bp from the JSB and JSA border regions, respectively.

**Phylogenetic Analysis**

The *Oryza* genus is composed of 23 species distributed in different regions of America, Africa, Asia, and Australia (Ge et al., 1999). Continued efforts have expanded our ability to differentiate among and to understand the genomic structure and phylogenetic relationships of rice species (Khus, 1997). Taxonomy and phylogeny of the rice genus have been extensively investigated at genus level (Ge et al., 1999; Zhu and Ge, 2005; Jacquemin et al., 2013). Previous evolutionary relationships among different rice genomes and species were estimated by nuclear and chloroplast DNA restriction fragment-length polymorphisms (Ge et al., 1999; Zou et al., 2015), but complete genome sequencing provides more detailed insight (Wambugu et al., 2015; Wu et al., 2015; Asaf et al., 2016b). In this regard, *O. minuta* has been poorly investigated. In this study, the phylogenetic position of *O. minuta* within *Oryza* was established by utilizing complete cp genomes and 65 shared genes among 12 *Oryza* members (Figures 7A,B). Two species, *Zizania aquatic* and *Zizania latifolia* were set as outgroups. Phylogenetic analysis using Bayesian inference (BI), maximum parsimony (MP), maximum likelihood (ML) and neighbor-joining (NJ) methods were performed. The results showed same phylogenetic signals for the complete cp genomes and 65 shared genes of *O. minuta*. The complete genome sequences (Tables S3, S6) and 65 shared genes (Tables S3, S6) from all species generated phylogenetic trees with same topologies (Figures 7A,B). In these phylogenetic trees based on the entire genome data set and 65 shared genes, *O. minuta* formed a single clade with *O. punctata*, with high BI and bootstrap support using four different methods (Figures 7A,B). Furthermore, the tree topology confirmed the
relationship inferred from the phylogenetic work conducted by Ge et al. (1999) and Zou et al. (2015). This position of *O. minuta* confirms the previously published phylogeny described by Ge et al. (1999). Ge et al. (1999) reported that *O. minuta* BBCC shares a clade with *O. punctata* BB with regard to Adh1, whereas it forms a clade with *O. officinalis* CC in the Adh2 phylogenetic analysis. Similar results was suggested by Zou et al. (2015), whereby phylogenetic analysis of the four nuclear loci and three maternally interted chloroplast fragments from different *Oryza* species grouped *O. minuta* in a clade with maternal parent *O. punctata* BB (Zou et al., 2015). As the phylogenetic tree based on the matK gene represents the maternal genealogy of rice species, which can offer an opportunity to identify maternal parents of allotetraploid species, we performed an additional phylogenetic analysis of *O. minuta* using the matK gene from related species (Figure S2). The results revealed a single clade for *O. minuta* with parental *O. punctata*. Similar results was also suggested by Ge et al. (1999), whereby phylogenetic analysis of the matK gene from different *Oryza* species grouped *O. minuta* in a clade with the maternal parent *O. punctata* BB instead of *O. officinalis* CC. Furthermore, the result suggests that there is no conflict between the entire genome data set and 65 shared genes of these cp genomes.

**CONCLUSION**

This study reports the first complete chloroplast genome sequence of *O. minuta* (135,094 bp). The structure and organization of this genome is very similar to previously reported cp genomes from the tribe Oryzeae. The location and distribution of repeat sequences was detected, and sequence divergences among cp genomes and 65 shared genes were identified with related species. No major structural rearrangement of *Oryza* species cp genomes was observed. Phylogenetic analyses showed that data sets based on the entire genome and 65 shared genes generate trees with same topologies regarding the placement of *O. minuta*. These findings provide a valuable analysis of the complete cp genome of *O. minuta*, which can be used to identify species and clarify taxonomic questions.

**AUTHOR CONTRIBUTIONS**

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

**FUNDING**

All the research work was financially supported by National Research Foundation of Korea (NRF), Ministry of Science, ICT and Future-Planning through Basic-Science Research Program (2014R1A1A1004918).

**SUPPLEMENTARY MATERIAL**

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpls.2017.00304/full#supplementary-material

---

**REFERENCES**

Amante-Bordeos, A., Sitch, L. A., Nelson, R., Dalmacio, R. D., Oliva, N. P., Aswiddinnoor, H., et al. (1992). Transfer of bacterial blight and blast resistance from the tetraploid wild rice *Oryza minuta* to cultivated rice, *Oryza sativa*. Theor. Appl. Genet. 84, 345–354. doi: 10.1007/bf00294939

Amirirauj, J. S. S., Fan, C. Z., Yu, Y. S., Song, X. A., Cranston, K. A., Pontoroli, A., et al. (2010). Spatio-temporal patterns of genome evolution in allotetraploid species of the genus *Oryza*. Plant J. 63, 430–442. doi: 10.1111/j.1365-3113.2010.04251.x

Asaf, S., Khan, A. L., Khan, A. R., Waqas, M., Kang, S.-M., Khan, M. A., et al. (2016a). Complete chloroplast genome of *Nicotiana sutorphora* and its comparison with related species. Front. Plant Sci. 7:843. doi: 10.3389/fpls.2016.00843

Asaf, S., Khan, A. L., Khan, A. R., Waqas, M., Kang, S.-M., Khan, M. A., et al. (2016b). Mitochondrial genome analysis of wild rice (*Oryza minuta*) and its comparison with other related species. PLoS ONE 11:e0152937. doi: 10.1371/journal.pone.0152937

Asano, T., Tsudzuki, T., Takahashi, S., Shimada, H., and Kadowaki, K. (2004). Complete nucleotide sequence of the sugarcane (*Saccharum officinarum*) chloroplast genome: a comparative analysis of four monocot chloroplast genomes. DNA Res. 11, 93–99. doi: 10.1093/dnares/11.2.93

Babiychuk, E., Vandesipe, K., Wisss, J., Garcia-Diaz, M., De Rycke, R., Akbari, H., et al. (2012). Plastid gene expression and plant development require a plastidic protein of the mitochondrial transcription termination factor family. Proc. Natl. Acad. Sci. U.S.A. 109, 6674–6679. doi: 10.1073/pnas.1103442108

Bao, Y., and Ge, S. (2008). Historical retrospect and the perplexity on the studies of the *Oryza* polyploids. J. Syst. Evol. 46, 3–12. doi: 10.3724/SP.J.1002.2008.07069

Brozynska, M., Furtado, A., and Henry, R. J. (2014). Direct chloroplast sequencing: comparison of sequencing platforms and analysis tools for whole chloroplast barcoding. PLoS ONE 9:e101387. doi: 10.1371/journal.pone.0101387

Cavalier-Smith, T. (2002). Chloroplast evolution: secondary symbiogenesis and multiple losses. *Curr. Biol.* 12, R62–R64. doi: 10.1016/S0960-9822(01)00675-3

Chang, C. C., Lin, H. C., Lin, I. P., Chow, T. Y., Chen, H. H., Chen, W. H., et al. (2006). The chloroplast genome of *Phalaenopsis aphrodite*
Chloroplast Genome of Wild Rice

Kumar, S., Hahn, F. M., McMahan, C. M., Cornish, K., and Whalen, M. C. (2009). Comparative analysis of the complete sequence of the plastid genome of *Parthenium argentatum* and identification of DNA barcodes to differentiate Parthenium species and lines. *BMC Plant Biol.* 9:131. doi: 10.1186/1471-2229-9-131

Leese, F., Mayer, C., and Held, C. (2008). Isolation of microsatellites from unknown genomes using known genomes as enrichment templates. *Linnmol. Oceanogr. Methods* 6, 412–426. doi: 10.4319/lom.2008.6.412

Li, X. W., Gao, H. H., Wang, Y. T., Song, J. Y., Henry, R., Wu, H. Z., et al. (2013). Complete chloroplast genome sequence of *Magnolia grandiflora* and comparative analysis with related species. *Sci. China Life Sci.* 56, 189–198. doi: 10.1007/s11427-012-4430-8

Lohse, M., Drechsel, O., and Bock, R. (2007). OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Genet.* 52, 267–274. doi: 10.1007/s00294-007-0161-y

Martin, W., Deusch, O., Stawski, N., Grünheit, N., and Grevenklin, V. (2005). Chloroplast genome phylogenetics: why we need independent approaches to plant molecular evolution. *Trends Plant Sci.* 10, 203–209. doi: 10.1016/j.tplants.2005.03.007

Moore, M. J., Soltis, P. S., Bell, C. D., Burleigh, J. G., and Soltis, D. E. (2010). Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Curr. Natl. Acad. Sci. U.S.A.* 107, 4623–4628. doi: 10.1073/pnas.0907801107

Morton, B. R. (1998). Selection on the codon bias of chloroplast and cyanelle genes in different plant and algal lineages. *J. Mol. Evol.* 46, 449–459. doi: 10.1007/PL00006325

Nadachowska-Brzyska, K., Li, C., Smeds, L., Zhang, G. J., and Ellegren, H. (2015). Temporal dynamics of avian populations during pleistocene revealed by whole-genome sequences. *Curr. Biol.* 25, 1375–1380. doi: 10.1016/j.cub.2015.03.047

Nie, X. J., Lv, S. Z., Zhang, Y. X., Du, X. H., Wang, L., Biradar, S. S., et al. (2012). Complete chloroplast genome sequence of a major invasive species, crofton weed (*Ageratina adenophora*). *PLoS ONE* 7:e36869. doi: 10.1371/journal.pone.0036869

Powels, V., Vekemans, X., Gode, C., Frerot, H., Castric, V., and Samitou-Provan, J., Corbett, G., McNicol, J. W., and Powell, W. (2012). Nuclear and chloroplast DNA phylogography reveals vicariance among European populations of the model species for the study of metal tolerance, *Arabidopsis halleri* (Brassicaceae). *New Phytol.* 193, 916–928. doi: 10.1111/j.1469-8137.2011.04003.x

Powell, W., Morgante, M., Andre, C., McNicol, J. W., Machray, C. G., Doyle, J. J., et al. (1995a). Hypervariable microsatellites provide a general source of polymorphic DNA markers for the chloroplast genome. *Curr. Biol.* 5, 1023–1029. doi: 10.1016/0960-9822(95)00206-5

Powell, W., Morgante, M., McDevitt, R., Vendramin, G. G., and Rafalski, J. A. (1995b). Polymorphic simple sequence repeat regions in chloroplast genomes - applications to the population-genetics of pines. *Proc. Natl. Acad. Sci. U.S.A.* 92, 7759–7763. doi: 10.1073/pnas.92.17.7759

Provan, J., Corbett, G., McNicol, J. W., and Powell, W. (1997). Chloroplast DNA variability in wild and cultivated rice (*Oryza spp.*) revealed by polymorphic chloroplast simple sequence repeats. *Genome* 40, 104–110. doi: 10.1139/g97-014

Qian, J., Song, J., Gao, H., Zhu, Y., Xu, J., and Pang, X. (2013). The complete chloroplast genome sequence of the medicinal plant *Salvia miltiorrhiza*. *PLoS ONE* 8:e57607. doi: 10.1371/journal.pone.0057607

Rahman, M. L., Jiang, W., Qiao, Y., Ham, T. H., Woo, M. O., et al. (2009). High-resolution mapping of two rice brown planthopper resistance genes, *Bph20(t)* and *Bph21(t)*, originating from *Oryza minuta*. *Theor. Appl. Genet.* 119, 1237–1246. doi: 10.1007/s00122-009-1125-z

Raubeson, L. A., Peery, R., Chumley, T. W., Dzubiak, C., Fourcade, H. M., Boone, J. L., et al. (2007). Comparative chloroplast genomics: analyses including new
sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. BMC Genomics 8:174. doi: 10.1186/1471-2164-8-174

Ronquist, F., and Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574. doi: 10.1093/bioinformatics/btg180

Rose, O., and Falush, D. (1998). A threshold size for microsatellite expansion. Mol. Biol. Evol. 15, 613–615. doi: 10.1093/oxfordjournals.molbev.a025964

Rousseau-Gueutin, M., Ayliffe, M. A., and Timmis, J. N. (2011). Conservation of plastid sequences in the plant nuclear genome for millions of years facilitates endosymbiotic evolution. Plant Physiol. 157, 2181–2193. doi: 10.1104/pp.111.185074

Sang, T., and Ge, S. (2007). Genetics and phylogenetics of rice domestication. Curr. Opin. Genet. Dev. 17, 533–538. doi: 10.1016/j.gde.2007.09.005

Schattner, P., Brooks, A. N., and Lowe, T. M. (2005). The tRNAscan-SE, snocGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res. 33, W686–W689. doi: 10.1093/nar/gki366

Sierro, N., Battey, J. N. D., Ouadi, S., Bakaher, N., Bovet, L., Willig, A., et al. (2014). The tobacco genome sequence and its comparison with those of tomato and potato. Nat. Commun. 5:3833. doi: 10.1038/ncomms4833

Swafford, D. L. (1993). Paup - a computer-program for phylogenetic inference using maximum parsimony. J. Gen. Physiol. 102, A9.

Timme, R. E., Kuehl, J. V., Boone, J. L., and Jansen, R. K. (2007). A comparative analysis of the Lactuca and Helianthus (Asteraceae) plastid genomes: identification of divergent regions and categorization of shared repeats. Am. J. Bot. 94, 302–312. doi: 10.3732/ajb.94.3.302

Timmis, J. N., Ayliffe, M. A., Huang, C. Y., and Martin, W. (2004). Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat. Rev. Genet. 5, 123–133. doi: 10.1038/nrg1271

Vaughan, D. A. (1989). The genus *Oryza L.:* current status of taxonomy. IRRI Res. Pap. Ser. 138:21.

Vaughan, D. A. (1994). *The Wild Relatives of Rice: A Genetic Resources Handbook.* Manila: International Rice Research Institute.

Wambu, P. A., Brozynska, M., Furtado, A., Waters, D. L., and Henry, R. J. (2015). Relationships of wild and domesticated rice (*Oryza A A genome species*) based upon whole chloroplast genome sequences. Sci. Rep. 5:13957. doi: 10.1038/srep13957

Wang, M. X., Cui, L. C., Meng, K. W., Deng, P. C., Du, X. H., Wan, F. H., et al. (2015). Comparative analysis of asteraceae chloroplast genomes: structural organization, RNA editing and evolution. Plant Mol. Biol. Rep. 33, 1526–1538. doi: 10.1007/s11105-015-0853-2

Wang, R. J., Cheng, C. L., Chang, C. C., Wu, C. L., Su, T. M., and Chaw, S. M. (2008). Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. BMC Evol. Biol. 8:36. doi: 10.1186/1471-2148-8-36

Waters, D. L., Nock, C. I., Ishikawa, R., Rice, N., and Henry, R. J. (2012). Chloroplast genome sequence confirms distinctness of Australian and Asian wild rice. Ecol. Evol. 2, 211–217. doi: 10.1002/ece3.66

Wicke, S., Schneeweiss, G. M., dePamphilis, C. W., Muller, K. F., and Quandt, D. (2011). The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol. Biol. 76, 273–297. doi: 10.1007/s11103-011-9762-4

Wu, Z. (2016a). The whole chloroplast genome of shrub willows (*Salix suchowensis*). Mitochondrial DNA A DNA Mapp Seq. Anal. 27, 2153–2154. doi: 10.3109/19401736.2014.982602

Wu, Z. (2016b). The completed eight chloroplast genomes of tomato from Solanum genus. Mitochondrial DNA Part A 27, 4155–4157. doi: 10.3109/19401736.2014.1003890

Wu, Z., and Ge, S. (2012). The phylogeny of the BEP clade in grasses revisited: evidence from the whole-genome sequences of chloroplasts. Mol. Phylogenet. Evol. 62, 573–578. doi: 10.1016/j.ympev.2011.10.019

Wu, Z., Tembrock, L. R., and Ge, S. (2015). Are differences in genomic data sets due to true biological variants or errors in genome assembly: an example from two chloroplast genomes. PLoS ONE 10:e0118019. doi: 10.1371/journal.pone.0118019

Wyman, S. K., Jansen, R. K., and Boore, J. L. (2004). Automatic annotation of organelar genomes with DOGMA. Bioinformatics 20, 3252–3255. doi: 10.1093/bioinformatics/bth352

Yang, M., Zhang, X., Liu, G., Yin, Y., Chen, K., and Yun, Q. (2010). The complete chloroplast genome sequence of date palm (*Phoenix dactylifera L.*). PLoS ONE 5:e12762. doi: 10.1371/journal.pone.0012762

Yang, Y., Dang, Y. Y., Li, Q., Lu, J. J., Li, X. W., and Wang, Y. T. (2014). Complete chloroplast genome sequence of poisonous and medicinal plant datura stramonium: organizations and implications for genetic engineering. PLoS ONE 9:e106586. doi: 10.1371/journal.pone.0106586

Yi, X., Gao, L., Wang, B., Su, Y. J., and Wang, T. (2013). The complete chloroplast genome sequence of *Cephalotaxus olivers* (*Cephalotaxaceae*): evolutionary comparison of *Cephalotaxus* chloroplast DNAs and insights into the loss of inverted repeat copies in gymnosperms. Genome Biol. Evol. 5, 688–698. doi: 10.1093/gbe/evt042

Yuria, N. P., and Odintsova, M. S. (1998). Comparative structural organization of plant chloroplast and mitochondrial genomes. Genetika 34, 5–22.

Zhang, T., Hu, S., Zhang, G., Pan, L., Zhang, X., Al-Mssallem, I. S., et al. (2012). The organelle genomes of Hassawi rice (*Oryza sativa L.*) and its hybrid in saudi arabic: genome variation, rearrangement, and origins. PLoS ONE 7:e42041. doi: 10.1371/journal.pone.0042041

Zhu, Q., and Ge, S. (2005). Phylogenetic relationships among A-genome species of the genus *Oryza revealed by intro* sequences of four nuclear genes. New Phytol. 167, 249–265. doi: 10.1111/j.1469-8137.2005.01406.x

Zou, X. H., Du, Y. S., Tang, L., Xu, X. W., Doyle, J. J., Sang, T., et al. (2015). Multiple origins of BBCC allopolyploid species in the rice genus (*Oryza*). Sci. Rep. 5:14876. doi: 10.1038/srep14876

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Asaf, Waqas, Khan, Khan, Kang, Imran, Shahzad, Bilal, Yun and Lee. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.