A systematic comparison of eight new plastome sequences from *Ipomoea* L

Jianying Sun¹,²,*, Xiaofeng Dong¹,²,*, Qinghe Cao³, Tao Xu¹,², Mingku Zhu¹,², Jian Sun¹,², Tingting Dong¹,², Daifu Ma³, Yonghua Han¹,² and Zongyun Li¹,²

¹ Institute of Integrative Plant Biology, School of Life Sciences, Jiangsu Normal University, Xuzhou, China
² Jiangsu Key Laboratory of Phylogenomics and Comparative Genomics, Jiangsu Normal University, Xuzhou, China
³ Jiangsu Xuhuai Regional Xuzhou Institute of Agricultural Sciences, Chinese Academy of Agricultural Sciences, Xuzhou, China

* These authors contributed equally to this work.

ABSTRACT

Background: *Ipomoea* is the largest genus in the family Convolvulaceae. The species in this genus have been widely used in many fields, such as agriculture, nutrition, and medicine. With the development of next-generation sequencing, more than 50 chloroplast genomes of *Ipomoea* species have been sequenced. However, the repeats and divergence regions in *Ipomoea* have not been well investigated. In the present study, we sequenced and assembled eight chloroplast genomes from sweet potato’s close wild relatives. By combining these with 32 published plastome genomes, we conducted a detailed comparative analysis of a broad range of *Ipomoea* species.

Methods: Eight chloroplast genomes were assembled using short DNA sequences generated by next-generation sequencing technology. By combining these chloroplast genomes with 32 other published *Ipomoea* chloroplast genomes downloaded from GenBank and the Oxford Research Archive, we conducted a comparative analysis of the repeat sequences and divergence regions across the *Ipomoea* genus. In addition, separate analyses of the Batata group and Quamoclit group were also performed.

Results: The eight newly sequenced chloroplast genomes ranged from 161,225 to 161,721 bp in length and displayed the typical circular quadripartite structure, consisting of a pair of inverted repeat (IR) regions (30,798–30,910 bp each) separated by a large single copy (LSC) region (87,575–88,004 bp) and a small single copy (SSC) region (12,018–12,051 bp). The average guanine-cytosine (GC) content was approximately 40.5% in the IR region, 36.1% in the LSC region, 32.2% in the SSC regions, and 37.5% in complete sequence for all the generated plastomes. The eight plastome genome sequences from this study included 80 protein-coding genes, four rRNAs (rrn23, rrn16, rrn5, and rrn4.5), and 37 tRNAs. The boundaries of single copy regions and IR regions were highly conserved in the eight chloroplast genomes. In *Ipomoea*, 57–89 pairs of repetitive sequences and 39–64 simple sequence repeats were found. By conducting a sliding window analysis, we found six relatively high variable regions (*ndhA* intron, *ndhH*-ndhF, *ndhF*-rpl32, rpl32-trnL, rps16-trnQ, and *ndhF*) in the *Ipomoea* genus, eight (trnG, rpl32-trnL, *ndhA* intron, *ndhF*-rpl32, *ndhH*-ndhF, ccsA-ndhD, trnG-trnR, and pasA-ycf3) in the Batatas group, and eight (*ndhA* intron, petN-psbM, rpl32-trnL, trnG-trnR, trnK-rps16, ndhC-trnV, rps16-trnQ, *ndhF*).
and trnG) in the Quamoclit group. Our maximum-likelihood tree based on whole chloroplast genomes confirmed the phylogenetic topology reported in previous studies. **Conclusions:** The chloroplast genome sequence and structure were highly conserved in the eight newly-sequenced Ipomoea species. Our comparative analysis included a broad range of Ipomoea chloroplast genomes, providing valuable information for Ipomoea species identification and enhancing the understanding of Ipomoea genetic resources.

**Subjects**  Bioinformatics, Genetics, Genomics, Plant Science  
**Keywords**  Ipomoea, Chloroplast genome, Divergence hotspot, Repeats, Genome structure

**INTRODUCTION**

*Ipomoea* belongs to the family Convolvulaceae comprises 600–700 species and is distributed throughout the tropical and subtropical regions (*Austin & Huáman, 1996*). *Ipomoea* mainly contains three subgenera: Eriospermum, Quamoclit, and Ipomoea. The species in this genus have been used in many fields, such as nutrition, medicine, rituals, and agriculture (*Meira et al., 2012*). Sweet potato, *Ipomoea batatas* (L.) Lam, is one of the most important food crops in the world and is grown in more than 100 countries (*Food and Agriculture Organization of the United Nations, 2017*).

Sweet potato is in *Ipomoea* series Batatas. After few corrections, more than 16 species were included in *I.* ser. Batatas (*Austin, 1987; McDonald & Austin, 1990; Muñoz-Rodríguez et al., 2018*). The phylogenetic relationships in *I.* ser. Batatas and the evolutionary origin of *I. batatas* have been a research focus for many years. Molecular markers, such as restriction fragment length polymorphisms (RFLPs) of genomic DNA, random amplified polymorphic DNAs (RAPDs), inter-simple sequence repeats (ISSRs), and some gene sequences have been used to investigate phylogenetic relationships of *I. batatas* and its wild relatives (*Gao et al., 2011; Rajapakse et al., 2004; Huang & Sun, 2000; Jarret & Austin, 1994; Jarret, Gawel & Whittemore, 1992*). In *Ipomoea*, various studies including morphology of pollen, flower, and seed; systemic characteristics, and floral anthocyanin regulators have been conducted (*Jayeola & Oladunjoye, 2012; Das, 2011; Streisfeld & Rausher, 2007*).

With the development of next-generation sequencing, large amounts of genomic data have been obtained to solve the taxonomic relationships in plants. Of these, chloroplast genome sequences are one of the most important sources. The chloroplast, the photosynthetic organelle in plants, has its own genome. The chloroplast genome of *Nicotiana tabacum* was the first chloroplast genome to be sequenced in 1986 (*Shinozaki et al., 1986*). The lengths of land-dwelling plant chloroplast genomes are generally 120–165 kb, the chloroplast DNA inheritance is mostly maternal and the genome is highly conserved in gene content and structure (*Harris & Ingram, 1991; Raubeson & Jansen, 2005*). Chloroplast genomes carry a large amount of valuable phylogenetic information and have been widely used to study the evolutionary relationships at almost any taxonomic level in plants (*Tong, Kim & Park, 2016; Zhang et al., 2016; Carbonell-Caballero et al., 2015*);
In *Ipomoea*, Eserman *et al.* (2014) utilized chloroplast genomes to properly resolve the phylogenetic relationships. Muñoz-Rodríguez *et al.* (2018) explained the origin of sweet potato by chloroplast genome sequence combined with nuclear regions. Although many chloroplast genomes of the *Ipomoea* genus have been obtained and have undergone phylogeny analysis, the divergence regions and variations of repeats have not been studied in the *Ipomoea* genus.

In this study, we sequenced and assembled eight chloroplast genomes from *I. ser. Batatas*. In order to have a better understanding of *Ipomoea* chloroplast genomes, we also included 32 published chloroplast genomes for comparison. Our two aims were as follows: first, to understand the conservation and diversity of the *Ipomoea* chloroplast genome through comparative genomics approaches and second, to analyze the repeat sequences in the *Ipomoea* genus.

**MATERIALS AND METHODS**

**Sampling and DNA extraction**

Total genomic DNA was extracted from eight species for sequencing (Table 1). The seeds of *I. trifida* (CIP 460377) were provided by the International Potato Center (Lima, Peru), those of *I. triloba* (NCNSP0323) were provided by North Carolina State University, and the rest were provided by the US National Plant Germplasm System (images of three species are shown in Supplementary Figures). Plants were grown in the greenhouse of the Xuzhou Sweet Potato Research Centre in China. Young leaves were collected from one plant of each species and were subsequently frozen in liquid nitrogen and stored at −80 °C until further use. Total genomic DNA was extracted using the Takara miniBEST plant genomic DNA extraction kit (Dalian, China). The integrity of genomic DNA was assessed by performing gel electrophoresis using a 1% agarose gel.

**Chloroplast genome assembling and annotation**

We constructed genomic libraries using the TruSeq DNA Nano kit with a DNA insert size of 350 bp. Sequencing was conducted on the Illumina X Ten platform, which
generated at least 21 Gb raw data from each species. Sequence data were quality trimmed using SOAPnuke (with the options: -n 0.1 -| 20 –q 0.1 -5 1) (Chen et al., 2018). Sequences were assembled into contigs according to the protocol described by Hahn, Bachmann & Chevreux (2013) using MIRA sequence assembler software (default settings) with the reference genome *I. trifida* (accession number: KF242476.1). The reads were blasted into contigs with perl scripts and contigs were extended to obtain the complete sequence. Finally, we blasted the reads to the complete sequence, detected the coverage of reads and manually corrected the sequence. The services of library construction, sequencing, and assembly were provided by Macrogen (http://www.macrogen.cn/sy, Shenzhen, China).

The eight chloroplast genome sequences were initially annotated using the online CpGAVAS (Liu et al., 2012) software with default settings, and then manually corrected using Genious 11.0.5. The circular chloroplast genome maps were constructed using the OrganellarGenome DRAW tool (Lohse, Drechsel & Bock, 2007).

**Repeat structure analysis**

Simple sequence repeats (SSRs) were detected using MISA-web (Beier et al., 2017). Thresholds of 10, six, and four repeat units for mono-, di-, and trinucleotides were used, respectively; while a threshold of three was used for tetra-, penta-, and hexanucleotides, respectively. REPuter (Kurtz et al., 2001) was used to visualize forward, palindrome, reverse, and complementary sequences, with a minimum repeat size of 30 bp and a sequence identity >90%.

**Divergence hotspot identification**

In order to study the differences among these genomes, the chloroplast genome sequences were aligned using MAFFT v7.307 (Katoh & Toh, 2010) including whole chloroplast genomes, then sliding window analysis was conducted. To calculate the proportion of mutational events, we used a modified formula previously described by Gielly & Taberlet (1994): proportion of mutational events = [(NS + ID)/L] × 100%, where NS, number of nucleotide substitutions; ID, number of indels; L, length of sequence. The step size was set to 400 bp, with an 800 bp window (Fu et al., 2017).

**Phylogenetic analysis**

We downloaded 32 published *Ipomoea* chloroplast genomes, and *Merremia quinquefolia* (KF242501) and *Operculina macrocarpa* (KF242502) were included in the analysis as the outgroup taxa to perform the phylogenetic analyses (Table S1). Before maximum-likelihood (ML) analyses, 42 chloroplast genomes were aligned using the MAFFT plugin in Geneious 11.0.5. The gaps in the alignment were stripped. ML analyses were performed using RAxML-HPC2 on XSEDE with 1,000 bootstrap replicates and the GTR+γ model on CPIRES Science Gateway (Miller, Pfeiffer & Schwartz, 2010) (https://www.phylo.org/).
RESULTS

Genome sequencing and assembly
At least 21 Gb of raw data from each species were generated using Illumina sequencing technology, then through filtering, we obtained 12.58–17.77 Gb of clean data. Using the *I. trifida* (accession number: KF242476.1) genome as a reference, we obtained the mapped reads. Finally, there were 0.59–1.28 Gb remaining bases and we assembled the chloroplast genomes of eight *Ipomoea* species using these data. The coverage of chloroplast genomes in each species ranged from 3,664 (in *I. splendor-sylvae*) to 7,941 (in *I. cordatotriloba*) (Table S2). All eight newly-sequenced chloroplast genomes were submitted to GenBank (accession numbers: MH173253; MH173254; MH173257; MH173259–MH173263) (Table 1).

Genome features
*Genome size and GC content*

The chloroplast genome size of the eight *Ipomoea* species ranged from 161,225 (*I. tabascana*) to 161,721 bp (*I. splendor-sylvae*) (Table 2; Fig. 1). These genomes displayed the typical circular quadripartite structure, consisting of a pair of inverted repeat (IR) regions (30,798–30,940 bp) separated by a large single copy (LSC) region (87,575–88,004 bp) and a small single copy (SSC) region (12,018–12,051 bp). The average guanine-cytosine content is approximately 40.5% in the IR regions, 36.1% in the LSC region, 32.2% in the SSC region, and 37.5% in the entire sequence of all plastomes (Table 2).

*Genes*

The chloroplast genomes of these wild species included 80 protein-coding genes, four rRNAs (*rrn23, rrn16, rrn5*, and *rrn4.5*), and 37 tRNAs. Based on their predicted

### Table 2 The chloroplast genome features of eight newly sequenced *Ipomoea* species.

| IPOMOEA SPECIES                      | TOTAL cpDNA SIZE | LENGTH OF LSC REGION | PROPORTION OF LSC (%) | LENGTH OF IR REGION | PROPORTION OF IR (%) | LENGTH OF SSC REGION | PROPORTION OF SSC (%) | TOTAL GC CONTENT (%) | PROPORTION OF GC IN LSC (%) | PROPORTION OF GC IN IR (%) | PROPORTION OF GC IN SSC (%) | TOTAL NUMBER OF GENES | PROTEIN-ENCODING GENES | tRNA | rRNA |
|-------------------------------------|------------------|----------------------|-----------------------|---------------------|----------------------|----------------------|-----------------------|-----------------------|-----------------------------|-----------------------------|-----------------------------|-------------------------|-------------------------|------|------|
| *Ipomoea trifida*                   | 161,531          | 87,606               | 54.23                 | 30,940              | 38.31                | 12,045               | 7.46                  | 37.52                 | 54.23                       | 38.31                       | 7.46                        | 121                     | 80          | 37   | 4    |
| *Ipomoea cynanchifolia*             | 161,386          | 87,591               | 54.27                 | 30,882              | 38.27                | 12,031               | 7.45                  | 37.53                 | 54.27                       | 38.27                       | 7.45                        | 121                     | 80          | 37   | 4    |
| *Ipomoea splendor-sylvae*           | 161,721          | 88,004               | 54.42                 | 30,833              | 38.13                | 12,051               | 7.45                  | 37.52                 | 54.42                       | 38.13                       | 7.45                        | 121                     | 80          | 37   | 4    |
| *Ipomoea cordatotriloba*            | 161,242          | 87,581               | 54.32                 | 30,906              | 38.21                | 12,049               | 7.46                  | 37.53                 | 54.32                       | 38.21                       | 7.46                        | 121                     | 80          | 37   | 4    |
| *Ipomoea lacunosa*                  | 161,446          | 87,575               | 54.24                 | 30,840              | 38.24                | 12,051               | 7.45                  | 37.55                 | 54.24                       | 38.24                       | 7.45                        | 121                     | 80          | 37   | 4    |
| *Ipomoea leucantha*                 | 161,296          | 87,598               | 54.31                 | 30,798              | 38.20                | 12,046               | 7.47                  | 37.55                 | 54.31                       | 38.20                       | 7.47                        | 121                     | 80          | 37   | 4    |
| *Ipomoea tabascana*                 | 161,225          | 87,583               | 54.32                 | 30,978              | 38.21                | 12,024               | 7.46                  | 37.54                 | 54.32                       | 38.21                       | 7.46                        | 121                     | 80          | 37   | 4    |
| *Ipomoea triloba*                   | 161,269          | 87,625               | 54.33                 | 30,810              | 38.21                | 12,018               | 7.46                  | 37.54                 | 54.33                       | 38.21                       | 7.46                        | 121                     | 80          | 37   | 4    |
functions, these genes can be divided into four categories, (1) genes related to photosynthesis; (2) genes related to self-replication; (3) genes related to the biosynthesis of cytochrome, protein, etc., and (4) functionally unknown ycf genes (Table S3). A total of 73 single-copy genes located in the LSC/SSC regions and seven two-copy genes in the IRs. In the chloroplast genomes analyzed, there were 16 genes harboring introns. In these genes, 14 genes had only one intron; ycf3 and clpP had two introns each (Table S4).

**Codon usage**

The chloroplast genomes of the eight *Ipomoea* species we studied contained 23,766–23,804 codons in total, possessing similar codon usage distribution. AUU (Ile) was the most...
abundant codon in all samples (Table S5). The relative synonymous codon usage of the third position showed more A/T than C/G.

**Boundary between LSC/SSC and IRs**
The junctions of LSC/IRa, SSC/IRa, and LSC/IRb regions are located in the IGS region between *rpl23* and *trnI*, *ndhH* and *ndhF*, and *trnI* and *trnH*, respectively, and the location of the SSC/IRb junction within the coding region of the *ndhA* gene made pseudogenes of the *ndhA* gene of 484 bp in length (Fig. 2). The distance from *ndhH* to the IRA/SSC boundary and the length of the *ndhA* gene located in the IRB is the same.

**Repetitive sequence analysis**
In total, we used 40 *Ipomoea* species for comparison (Table S1). The number of repetitive sequences in these chloroplast genomes ranged from 57 (*I. pedicellaris*) to 89 pairs (Table S6). Half of the chloroplast genomes had 88 or 89 pairs of repeats. The forward and palindromic repeats occupied 50.37% and 49.63% of the total repeats, respectively (Figs. 3A and 3C). The number of forward repeats ranged from 19 (*I. pes-caprae*) to 50 (*I. biflora*) and the number of palindromic repeats ranged from 28 (*I. pedicellaris*) to 61 (*I. pes-caprae*) (Table S6). The most common repetitive sequence lengths were 30–59 bp.
(58.55%), while those with >140 bp were least abundant (3.40%) (Figs. 3A and 3C). Most of the repeats were located in the ycf1 gene region (most abundant region), ycf2 gene region, and the intergenic regions between trnN-GUU and ycf1 or trnI-CAU and ycf2; all of which were in IR regions. Those repeats distributed in the LSC region were mostly in the accD gene, ycf3 gene, and the intergenic regions between rpl23 and trnI-CAU. We also identified few repeats located in ndhH-ndhF within the SSC regions (Table S6).

The separate analysis of the two large clades (Batatas clade and Quamoclit clade) showed similar results with genus Ipomoea (Fig. 3C). Interestingly, the I. nil belonging to the
Quamoclit clade was different from the other species. It had 43.2% (38 in 88) repeats distributed in the intergenic regions between ycf2 and trnL-CAU (Table S6).

Simple sequence repeats are tandemly-repeated nucleotides in DNA sequences. We found the Ipomoea chloroplast genomes we analyzed contained 39–64 SSRs. The most common SSRs were mononucleotides which accounted for 70.77%, followed by tetrancleotide repeats which accounted for 16.95% (Figs. 3B and 3D). Almost all of the mononucleotide repeat sequences were comprised of A/T repeats, while the tetranucleotides varied among different species. All the dinucleotide repeats were AT/TA repeats comprising 1.51% of the total SSRs. The proportion of pentanucleotides and hexanucleotides were 3.56% and 1.78%, respectively, while not all the species had them. Trinucleotides accounted for 0.40% in all the SSRs and were the least abundant, with only 10 trinucleotides detected across all the 40 Ipomoea species (Figs. 3B and 3D). Except for the mononucleotides and dinucleotides, five tetrancleotide repeats (AATA, CAAT, GAAA, TATC, and TTTC) were shared in the Quamoclit clade (Table S7). In the Batatas clade, more tetranucleotides were shared including AAAT, AATA, AGAT, ATAG, CAAT, GAAA, TATC, and TTTC. The tetranucleotide GAAA was only absent in I. splendor-sylvae (Table S7). In addition, the pentanucleotides (TTCTA) were present in 93.7% (15/16) and the hexanucleotides were present in 81.2% (13/16) of the Batatas clade. SSRs are different from the repetitive sequences identified by REPuter; they are almost all located in LSC regions (Table S7).

Divergence hotspot regions

We conducted divergence analysis of the Ipomoea genus, including 40 species. Additionally, separate analysis in the Batatas group (16 species) and Quamoclit group (10 species) were conducted. Across the Ipomoea genus, the percentage of identical sites was 78.5%. The mean value of the variation was 5.12% and six regions which had a variation rate >13% were thought to be highly variable regions. Five of them were intergenic regions (ndhA intron, ndhH-ndhF, ndhF-rpl32, rpl32-trnL, rps16-trnQ) and one was in a gene coding region (ndhF) (Fig. 4A). Except rps16-trnQ which was located in the LSC region, all other highly variable regions were located in the SSC region.

In the Batatas group, the percentage of identical sites was 94.8% and the mean value of the variation was 1.26%. The regions with a variation rate >3.2% were considered to be highly variable regions. They were trnG, rpl32-trnL, ndhA intron, ndhF-rpl32, ndhH-ndhF, ccsA-ndhD, trnG-trnR, and pasA-ycf3 (Fig. 4B). One of these was located in the IRA/SSC boundary (ndhH-ndhF), three of eight were located in the LSC region, and four of them were located in the SSC region (rpl32-trnL, ndhA intron, ndhF-rpl32, and ccsA-ndhD).

In the Quamoclit group, 10 species showed highly consistent sequences, with 92.0% identical sites. The mean value of variation was 2.66%. Eight relatively high variable regions with a variation rate >7.5% were detected, including ndhA intron, petN-psbM, rpl32-trnL, trnG-trnR, trnK-rps16, ndhG-trnV, rps16-trnQ, and trnG (Fig. 4C). Two of them were in the SSC region (rpl32-trnL, ndhA intron), the others were located in the LSC region, and all of them were from non-coding regions.
Phylogenetic analysis

The phylogenetic tree was constructed based on 40 Ipomoea whole chloroplast genomes, using M. quinquefolia (KF242501) and O. macrocarpa (KF242502) as outgroups (Fig. 5). These 40 Ipomoea species were divided into two major clades and were further divided into seven small clades with strong support values, including Batatas, Murucoides, Pes-caprae, Quamoclit, Cairica, Obscura, and Pes-tigridis (Eserman et al., 2014). The Cairica, Pes-tigridis, and Obscura groups formed one major clade and the other four groups formed another clade. The support value for the Cairica clade was lower than those of other small clades. Among these seven groups, two larger groups, Quamoclit and Batatas, included the most species. Ipomoea setosa and I. sepacuitensis were located as the sister group of I. ser. Batatas and constituted the Batatas group. In I. ser. Batatas, I. splendor-sylvae was located at the most basal position. In the Quamoclit group, I. dumetorum was clustered as the sister group with others.

Figure 4 Percentages of variable sites in homologous regions among the Ipomoea chloroplast genomes. (A) In 40 Ipomoea species; (B) Batatas group; (C) Quamoclit group. DOI: 10.7717/peerj.6563/fig-4
DISCUSSION

Variations among the Ipomoea species

In the present study, eight chloroplast genomes in I. ser. Batatas were assembled. They were highly conserved in terms of genome structure and size. They displayed a typical quadripartite structure and length of the chloroplast genome sequence in the eight-wild species, ranging from 161,225 to 161,721 bp. The expansion or contraction of IR regions was thought to be the main reason for the divergences in chloroplast genome size (Ravi et al., 2008). Two models were proposed to explain the expansion of the IR region. Gene conversion is used to explain the small IR expansion and movements, and double-strand DNA breaks and recombination were considered to be the reason for major IR expansion (Goulding et al., 1996). Among the eight species, the LSC/IRa/SSC/IRb boundary genes were highly conserved, with slight structural variations and there was no significant expansion/contraction of IRs among these species. According to the studies conducted by Eserman et al. (2014), the SC/IR boundary was highly conserved even across the whole tribe Ipomoeae with some exceptions only occurring in a few species. Our eight chloroplast genomes contained 121 genes and the infA gene, which encodes a translation initiation factor and is almost lost in all rosid species, has also been lost in our species (Millen et al., 2001; Gitzendanner et al., 2018).

Repeat motifs are thought to have a significant impact on genome phylogeny and rearrangement (Yue et al., 2008). Larger and more complex repeat sequences may play an important role in the rearrangement of chloroplast genomes and sequence divergence (Timme et al., 2007; Weng et al., 2014). Here, we conducted repeat analysis in 40 Ipomoea species and separate analyses in the Batatas and Quamoclit groups. In Ipomoea, only forward and palindromic repeats were found, and they were almost equal in measure in each species. The repeats distributed mainly in the ycf1 and ycf2 genes. Interestingly, I. pes-caprae had 19 forward repeats and 61 palindromic repeats; in I. nil, 43.2% of...
repeats were located in the ycf2-trnI-CAU region rather than in the ycf1 and ycf2 genes. Additionally, in contrast with the Quamoclit group, the Batatas group had a slightly higher proportion of 30–59 bp repeats and a lower proportion of 60–79 bp and >140 bp repeats.

We also investigated and compared the numbers and distributions of SSRs across the 40 *Ipomoea* species. Compared with the long repeats, the SSRs were distributed more widely throughout the chloroplast genomes and were usually located in LSC regions. Most of the SSRs were found in non-coding regions; only a few were located in the coding regions (e.g., *rpoC2*, *rpoB*, *atpB*, *ycf1*, *ycf2*, and *ndhF*). The predominant type of SSRs were mononucleotides and almost all of them were A or T repeats. This is consistent with the previous findings that suggested that chloroplast SSRs are generally comprised of short polyA or polyT repeats and rarely contain tandem G or C repeats (*Kuang et al., 2011*). By comparing the Quamoclit and Batatas groups, extremely rare trinucleotides were detected in the Batatas group (0.1%) and rare hexanucleotides were detected in the Quamoclit group (0.89%). Because of the high polymorphism of SSRs in the chloroplast genome, SSRs are potentially important molecular markers in the analysis of plant population genetics as well as evolutionary and ecological studies (*Kuang et al., 2011*). In sweet potato, chloroplast SSRs ((A)$_n$ and (T)$_n$) combined with nuclear SSRs have been used to investigate the genetic diversity of sweet potato and further provided strong support for the prehistoric transfer of sweet potato (*Roullier et al., 2011, 2013a*).

In addition, nucleotide substitution (SNVs, indels, and proportions of variability) may play a critical role in plant evolutionary processes. We conducted three independent analyses which included one in the *Ipomoea* genus and two in small clades (Batatas and Quamoclit groups) and they showed different divergence regions. Only two regions presented in all the three analyses—*rpl32-trnL* and *ndhA* intron. One of them (*rpl32-trnL*) has been used in previous studies to disentangle the origins of sweet potato (*Roullier et al., 2013b*). Like other angiosperms (*Liu et al., 2017*), the IR regions were more conserved than the SC regions in *Ipomoea* chloroplast genomes. This phenomenon possibly occurred because of copy correction between IR sequences by gene conversion (*Khakhlova & Bock, 2006*).

**Phylogenetic relationships**

The phylogenetic relationships in tribe Ipomoeae have been constructed based on whole chloroplast genomes of 30 species including *Ipomoea* and nine other genera (*Eserman et al., 2014*). Here, we constructed a phylogenetic tree of the *Ipomoea* with 21 chloroplast genomes downloaded from GenBank, eleven from Muñoz-Rodríguez’s group’s research (*Muñoz-Rodríguez et al., 2018*), and our eight chloroplast genomes. Our results also divided the 40 *Ipomoea* species into seven groups that confirmed previous studies (*Eserman et al., 2014*). Determination of taxonomy and species in *I. ser. Batatas* is particularly difficult because individuals often exhibit intermediate morphologies between descriptions of named species (*McDonald & Austin, 1990; Austin, 1978*), and several species may be of hybrid origin (*Diaz, Schmiediche & Austin, 1996*). Phylogenetic analysis in *I. ser. Batatas* has been performed using DNA markers, such as RFLP,
RAPD, ISSR, chloroplast restriction site variation, gene sequences, and morphological analyses (Rajapakse et al., 2004; Huang & Sun, 2000; Jarret & Austin, 1994; Jarret, Gawel & Whittemore, 1992). These studies have indicated the phylogenetic relationships between sweet potato and its wild relatives; however, the support values were low. The latest study conducted by Muñoz-Rodríguez et al. (2018) presented strong support for the relationships in this series based on the amounts of chloroplast genomes and nuclear data. Our studies showed a similar result as that shown in a study by Muñoz-Rodríguez et al. (2018), which suggested that the chloroplast genome is a very useful tool for resolving the phylogenetic relationships of *I. ser. Batatas*. *Ipomoea setosa* and *I. sepacuitensis* were clustered, and together with *I. ser. Batatas* formed the Batatas group. The taxonomy of the Quamoclit clade was mostly consistent with the ITS phylogeny (Miller, McDonald & Manos, 2004); for example, *I. indica* was more closely related to *I. nil* than *I. purpurea*. Conversely, *I. alba* grouped with *I. tricolor* rather than with *I. nil* or *I. purpurea*.

**CONCLUSIONS**

In this study, we sequenced, assembled, and annotated eight chloroplast genomes derived from close wild relatives of *I. batatas*. Based on these data, we conducted a broad range analysis in the genus *Ipomoea*. Our results showed that the chloroplast genome of these eight species is highly consistent in sequence and structure. Along with 32 published *Ipomoea* species, a detailed repeat analysis was conducted. We also identified six highly variable regions which could be useful in investigating the population genetics and biogeography of closely-related *Ipomoea* species. In addition, different divergence regions were also identified in the Batatas and Quamoclit groups.

**ADDITIONAL INFORMATION AND DECLARATIONS**

**Funding**

This work was supported by the National Natural Science Foundation of China (31771367), the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD), the China Agriculture Research System (Grant No. CARS-10-B03), and the Colleges and Universities in Jiangsu Province plans to graduate research and innovation (KYLX16_1318). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Grant Disclosures**

The following grant information was disclosed by the authors:
National Natural Science Foundation of China: 31771367.
The Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD), China Agriculture Research System: CARS-10-B03.
The Colleges and Universities in Jiangsu Province plans to graduate research and innovation: KYLX16_1318.

**Competing Interests**

The authors declare that they have no competing interests.
Author Contributions

- Jianying Sun conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Xiaofeng Dong performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Qinghe Cao contributed reagents/materials/analysis tools, approved the final draft.
- Tao Xu performed the experiments, approved the final draft.
- Mingku Zhu contributed reagents/materials/analysis tools, approved the final draft.
- Jian Sun contributed reagents/materials/analysis tools, approved the final draft.
- Tingting Dong contributed reagents/materials/analysis tools, approved the final draft.
- Daifu Ma contributed reagents/materials/analysis tools, approved the final draft.
- Yonghua Han conceived and designed the experiments, approved the final draft.
- Zongyun Li conceived and designed the experiments, authored or reviewed drafts of the paper, approved the final draft.

DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

The chloroplast genome sequences of ten species are accessible via GenBank accession numbers MH173253, MH173254, MH173257, and MH173259–MH173263.

Data Availability

The following information was supplied regarding data availability:

The data can be found at GenBank (accession numbers MH173253, MH173254, MH173257, and MH173259–MH173263) and the Oxford Research Archive (DOI 10.5287/bodleian:yrYKneXED).

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.6563#supplemental-information.

REFERENCES

Austin DF. 1978. The Ipomoea-Batatas complex-1. Taxonomy. Bulletin of the Torrey Botanical Club 105(2):114–129 DOI 10.2307/2484429.

Austin DF. 1987. The taxonomy evolution and genetic diversity of sweetpotatoes and related wild species. Lima: International Potato Center, 27–60.

Austin DF, Huáman Z. 1996. A synopsis of Ipomoea (Convolvulaceae) in the Americas. Taxon 45(1):3–38 DOI 10.2307/1222581.

Beier S, Thiel T, Munch T, Scholz U, Mascher M. 2017. MISA-web: a web server for microsatellite prediction. Bioinformatics 33(16):2583–2585 DOI 10.1093/bioinformatics/btx198.

Carbonell-Caballero J, Alonso R, Ibanez V, Terol J, Talon M, Dopazo J. 2015. A phylogenetic analysis of 34 chloroplast genomes elucidates the relationships between wild and domestic species within the genus citrus. Molecular Biology and Evolution 32(8):2015–2035 DOI 10.1093/molbev/msv082.
Chen Y, Chen Y, Shi C, Huang Z, Zhang Y, Li S, Li Y, Ye J, Yu C, Li Z, Zhang X, Wang J, Yang H, Fang L, Chen Q. 2018. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. GigaScience 7:1 DOI 10.1093/gigascience/gix120.

Das S. 2011. Congruence between morphological and molecular approach in understanding species relationship in Ipomoea spp.: a rare event in taxonomy. Asian Journal of Plant Sciences 10(4):263–268 DOI 10.3923/ajps.2011.263.268.

Diaz J, Schmiediche P, Austin DF. 1996. Polygon of crossability between eleven species of Ipomoea: Section Batatas (convolvulaceae). Euphytica 88(3):189–200 DOI 10.1007/BF00023890.

Eserman LA, Tiley GP, Jarret RL, Leebens-Mack JH, Miller RE. 2014. Phylogenetics and diversification of morning glories (tribe Ipomoeae, Convulvulaceae) based on whole plastome sequences. American Journal of Botany 101(1):92–103 DOI 10.3732/ajb.1300207.

Food and Agriculture Organization of the United Nations. 2017. FAOSTAT Statistics Database. Available at http://www.fao.org/faostat/ (accessed 1 August 2017).

Fu C, Li H, Milne R, Zhang T, Ma P, Yang J, Li D, Gao L. 2017. Comparative analyses of plastid genomes from fourteen Cornales species: inferences for phylogenetic relationships and genome evolution. BMC Genomics 18(1):956 DOI 10.1186/s12864-017-4319-9.

Gao M, Ashu GM, Stewart L, Akwe WA, Njiti V, Barnes S. 2011. Wx intron variations support an allohexaploid origin of the sweetpotato [Ipomoea batatas (L.) Lam]. Euphytica 177(1):111–133 DOI 10.1007/s10681-010-0275-z.

Gielly L, Taberlet P. 1994. The use of chloroplast DNA to resolve plant phylogenies: noncoding versus rbcL sequences. Molecular Biology and Evolution 11(5):769–777 DOI 10.1093/oxfordjournals.molbev.a040157.

Gitzendanner MA, Soltis PS, Wong GK, Ruhfel BR, Soltis DE. 2018. Plastid phylogenomic analysis of green plants: a billion years of evolutionary history. American Journal of Botany 105(3):291–301 DOI 10.1002/ajb2.1048.

Hahn C, Bachmann L, Chevreux B. 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. Nucleic Acids Research 41(13):e129 DOI 10.1093/nar/gkt371.

Harris SA, Ingram R. 1991. Chloroplast DNA and biosystematics: the effects of intraspecific diversity and plastid transmission. Taxon 40(3):393–412 DOI 10.2307/1223218.

Huang JC, Sun M. 2000. Genetic diversity and relationships of sweetpotato and its wild relatives in Ipomoea series Batatas (Convulvulaceae) as revealed by inter-simple sequence repeat (ISSR) and restriction analysis of chloroplast DNA. Theoretical and Applied Genetics 100(7):1050–1060 DOI 10.1007/s001220051386.

Jansen RK, Cai Z, Raubeson LA, Daniell H, Delpamphilis CW, Leebens-Mack J, Muller KF, Guisinger-Bellian M, Haberle RC, Hansen AK, Chumley TW, Lee SB, Peery R, McNeal JR, Kuehl JV, Boore JL. 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proceedings of the National Academy of Sciences of the United States of America 104(49):19369–19374 DOI 10.1073/pnas.0709121104.

Jarret RL, Austin DF. 1994. Genetic diversity and systematic relationships in sweetpotato (Ipomoea batatas (L.) Lam.) and related species as revealed by RAPD analysis. Genetic Resources and Crop Evolution 41(3):165–173 DOI 10.1007/BF00051633.
Jarret RL, Gawel N, Whittomere A. 1992. Phylogenetic Relationships of the Sweetpotato [Ipomoea batatas (L.) Lam. Journal of the American Society for Horticultural Science 117(4):633–637 DOI 10.21273/jashs.117.4.633.

Jayeola AA, Oladunjoye OR. 2012. Systematic studies in some ipomoea Linn, species using pollen and flower morphology. Annales of West University of Timisoara. Series of Biology 15(2):177–187.

Katoh K, Toh H. 2010. Parallelization of the MAFFT multiple sequence alignment program. Bioinformatics 26(15):1899–1900 DOI 10.1093/bioinformatics/btq224.

Khakhlova O, Bock R. 2006. Elimination of deleterious mutations in plastid genomes by gene conversion. Plant Journal 46(1):85–94 DOI 10.1111/j.1365-313X.2006.02673.x.

Kuang D-Y, Wu H, Wang Y-L, Gao L-M, Zhang S-Z, Lu L. 2011. Complete chloroplast genome sequence of Magnolia kwangsiensis (Magnoliaceae): implication for DNA barcoding and population genetics. Genome 54(8):663–673 DOI 10.1139/G11-026.

Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. 2001. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Research 29(22):4633–4642 DOI 10.1093/nar/29.22.4633.

Liu L, Li R, Worth JRP, Li X, Li P, Cameron KM, Fu C. 2017. The complete chloroplast genome of chinese bayberry (Morella rubra, Myricaceae): implications for understanding the evolution of Fagales. Frontiers in Plant Science 8:968 DOI 10.3389/fpls.2017.00968.

Liu C, Shi L, Zhu Y, Chen H, Zhang J, Lin X, Guan X. 2012. CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. BMC Genomics 13(1):715 DOI 10.1186/1471-2164-13-715.

Lohse M, Drechsel O, Bock R. 2007. OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. Current Genetics 52(5–6):267–274 DOI 10.1007/s00294-007-0161-y.

McDonald JA, Austin DF. 1990. Changes and additions in ipomoea section batatas (Convolvulaceae). Brittonia 42(2):116–120 DOI 10.2307/2807625.

Meira M, Da Silva EP, David JM, David JP. 2012. Review of the genus Ipomoea: traditional uses, chemistry and biological activities. Revista Brasileira de Farmacognosia 22(3):682–713 DOI 10.1590/s0102-695x2012005000025.

Millen RS, Olmstead RG, Adams KL, Palmer JD, Lao NT, Heggie L, Kavanagh TA, Hibberd JM, Gray JC, Morden CW, Calie PJ, Jermiin LS, Wolfe KH. 2001. Many parallel losses of infA from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. Plant Cell 13(3):645–658 DOI 10.2307/3871412.

Miller RE, McDonald JA, Manos PS. 2004. Systematics of Ipomoea subgenus Quamoclit (Convolvulaceae) based on ITS sequence data and a Bayesian phylogenetic analysis. American Journal of Botany 91(8):1208–1218 DOI 10.3732/ajb.91.8.1208.

Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: 2010 Gateway Computing Environments Workshop (GCE), New Orleans, LA, USA: IEEE Computer Society, 1–8.

Muñoz-Rodriguez P, Carruthers T, Wood JRI, Williams BRM, Weitemier K, Kronmiller B, Ellis D, Anglin NL, Longway L, Harris SA, Rausher MD, Kelly S, Liston A, Scotland RW. 2018. Reconciling conflicting phylogenies in the origin of sweet potato and dispersal to polynesia. Current Biology 28(8):1246–1256.e12 DOI 10.1016/j.cub.2018.03.020.

Rajapakse S, Nilmaloda SD, Molnar M, Ballard RE, Austin DF, Bohac JR. 2004. Phylogenetic relationships of the sweetpotato in Ipomoea series Batatas (Convolvulaceae) based on nuclear β-amylase gene sequences. Molecular Phylogenetics and Evolution 30(3):623–632 DOI 10.1016/S1055-7903(03)00249-5.
Raubeson LA, Jansen RK. 2005. Chloroplast genomes of plants. In: Henry RJ, ed. Plant Diversity and Evolution: Genotypic and Phenotypic Variation in Higher Plants, Wallingford: CABI, 44–68.

Ravi V, Khurana JP, Tyagi AK, Khurana P. 2008. An update on chloroplast genomes. Plant Systematics and Evolution 271(1–2):101–122 DOI 10.1007/s00606-007-0608-0.

Roullier C, Benoit L, McKey DB, Lebot V. 2013a. From the Cover: Cozzarelli Prize Winner: Historical collections reveal patterns of diffusion of sweet potato in Oceania obscured by modern plant movements and recombination. Proceedings of the National Academy of Sciences of the United States of America 110(6):2205–2210 DOI 10.1073/pnas.1211049110.

Roullier C, Duputie A, Wennekes P, Benoit L, Fernandez Bringas VM, Rossel G, Tay D, McKey D, Lebot V. 2013b. Disentangling the origins of cultivated sweet potato (Ipomoea batatas (L.) Lam.). PLOS ONE 8(5):e62707 DOI 10.1371/journal.pone.0062707.

Roullier C, Rossel G, Tay D, Mckey D, Lebot V. 2011. Combining chloroplast and nuclear microsatellites to investigate origin and dispersal of New World sweet potato landraces. Molecular Ecology 20(19):3963–3977 DOI 10.1111/j.1365-294X.2011.05229.x.

Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohito C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaia F, Kato A, Tohdoh N, Shimada H, Sugiura M. 1986. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO Journal 5(9):2043–2049 DOI 10.1002/j.1460-2075.1986.tb04464.x.

Streisfeld MA, Rausher MD. 2007. Relaxed constraint and evolutionary rate variation between basic helix-loop-helix floral anthocyanin regulators in ipomoea. Molecular Biology and Evolution 24(12):2816–2826 DOI 10.1093/molbev/msm216.

Timme RE, Kuehl JV, Boore JL, Jansen RK. 2007. A comparative analysis of the Lactuca and Helianthus (Asteraceae) plastid genomes: identification of divergent regions and categorization of shared repeats. American Journal of Botany 94(3):302–312 DOI 10.3732/ajb.94.3.302.

Tong W, Kim T-S, Park Y-J. 2016. Rice chloroplast genome variation architecture and phylogenetic dissection in diverse oryza species assessed by whole-genome resequencing. Rice 9(1):57 DOI 10.1186/s12284-016-0129-y.

Weng ML, Blazier JC, Govindu M, Jansen RK. 2014. Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. Molecular Biology and Evolution 31(3):645–659 DOI 10.1093/molbev/mst257.

Yue F, Cui L, DePamphilis CW, Moret BM, Tang J. 2008. Gene rearrangement analysis and ancestral order inference from chloroplast genomes with inverted repeat. BMC Genomics 9(Supp 1):S25 DOI 10.1186/1471-2164-9-S1-S25.

Zhang Y, Du L, Liu A, Chen J, Wu L, Hu W, Zhang W, Kim K, Lee SC, Yang TJ, Wang Y. 2016. The complete chloroplast genome sequences of five epimedium species: lights into phylogenetic and taxonomic analyses. Frontiers in Plant Science 7:306 DOI 10.3389/fpls.2016.00306.