Complete Chloroplast Genome Sequences of Mongolia Medicine Artemisia frigida and Phylogenetic Relationships with Other Plants

Yue Liu1, 2, 3, 4, Naxin Huo3, Lingli Dong3, Yi Wang3, Shuixian Zhang1, 5, Hugh A. Young3, Xiaoxiao Feng1, 5, Yong Qiang Gu3*

1 College of Life and Environmental Sciences, Minzu University of China, Beijing, China, 2 Institute of Chinese Materia Medica, China Academy of Traditional Chinese Medicine, Beijing, China, 3 Western Regional Research Center, Agricultural Research Service, United States Department of Agriculture (USDA-ARS), Albany, California, United States of America, 4 Department of Plant Science, University of California Davis, Davis, California, United States of America, 5 Graduate School, Minzu University of China, Beijing, China

Abstract

Background: Artemisia frigida Willd. is an important Mongolian traditional medicinal plant with pharmacological functions of stanch and detumescence. However, there is little sequence and genomic information available for Artemisia frigida, which makes phylogenetic identification, evolutionary studies, and genetic improvement of its value very difficult. We report the complete chloroplast genome sequence of Artemisia frigida based on 454 pyrosequencing.

Methodology/Principal Findings: The complete chloroplast genome of Artemisia frigida is 151,076 bp including a large single copy (LSC) region of 82,740 bp, a small single copy (SSC) region of 18,394 bp and a pair of inverted repeats (IRs) of 24,971 bp. The genome contains 114 unique genes and 18 duplicated genes. The chloroplast genome of Artemisia frigida contains a small 3.4 kb inversion within a large 23 kb inversion in the LSC region, a unique feature in Asteraceae. The gene order in the SSC region of Artemisia frigida is inverted compared with the other 6 Asteraceae species with the chloroplast genomes sequenced. This inversion is likely caused by an intramolecular recombination event only occurred in Artemisia frigida. The existence of rich SSR loci in the Artemisia frigida chloroplast genome provides a rare opportunity to study population genetics of this Mongolian medicinal plant. Phylogenetic analysis demonstrates a sister relationship between Artemisia frigida and four other species in Asteraceae, including Ageratina adenophora, Helianthus annuus, Guizotia abyssinica and Lactuca sativa, based on 61 protein-coding sequences. Furthermore, Artemisia frigida was placed in the tribe Anthemideae in the subfamily Asteroideae (Asteraceae) based on ndhF and trnL-F sequence comparisons.

Conclusion: The chloroplast genome sequence of Artemisia frigida was assembled and analyzed in this study, representing the first plastid genome sequenced in the Anthemideae tribe. This complete chloroplast genome sequence will be useful for molecular ecology and molecular phylogeny studies within Artemisia species and also within the Asteraceae family.

Introduction

Artemisia frigida Willd., named as “Agi” in the Mongolian language, is an important Mongolian traditional medicinal plant [1], distributed widely in the Inner Mongolia Autonomous Region and the northern part of China. This plant has medicinal application for stanch and detumescence, so it is often used to care for bleeding, arthroncus, rheumatism, menoxenia, and other ailments [1]. Besides its medicinal efficacy, it is also valued as an important food resource for livestock, and a remarkable component of the desert ecosystem [1].

Artemisia frigida belongs to the largest genus in the tribe Anthemideae of the family Asteraceae, which is the second largest family of plants in the world, consisting of over 20,000 species [2]. Artemisia frigida is a diploid species (2n = 2X = 18) and its haploid genome size is estimated to be 2,567 Mb [3]. However, polyploid A. frigida species with 2n = 4X = 36 have been identified in nature [4]. In recent years, there has been extensive research focused on the medicinal and pharmacological aspects and effects of the Artemisia frigida plant [5–9]. However, there has not been a comprehensive study of the genetic variability found in natural populations [1]. With the increasing demand for commercial use and the important ecological value of this traditional medicinal plant, large-scale breeding efforts need to be developed for Artemisia frigida. Selection of germplasm with high pharmaceutical efficacy at the molecular level is important and requires the availability of efficient genetic and molecular marker data. Access
to genetic information will not only improve the genetic breeding process, but also will aid in downstream analysis of sequence data and improvement of *Artemisia frigida*’s medicinal qualities. Currently, there are only 24 sequences available for *Artemisia frigida*, including 6 nrDNA sequences and 18 chloroplast DNA sequences listed in GenBank [10–16] (http://www.ncbi.nlm.nih.gov/nucore/?term = Artemisia%20frigida%20). Therefore, there is a clear need to develop genomic resources for *Artemisia frigida* in order to efficiently apply molecular and biotechnological approaches for the improvement of its value as an important medicinal plant.

Chloroplasts are plant organelles that contain the entire enzymatic machinery necessary for photosynthesis and other biochemical pathways. Most land plants have a highly conserved chloroplast genome organized into a single circular chromosome [17] that contains two copies of an inverted repeat (IR) separating a large single copy region (LSC) and a small single copy region (SSC). To date, over 200 chloroplast (cp) genome sequences are available in The Chloroplast Genome Database (http://chloroplast.ocean.washington.edu/cpbase/run). The vast majority of angiosperm cp genomes are highly conserved [18]. However, the gene order found in the LSC region of the Asteraceae, Fabaceae, and Poaceae families [19–21] is reversed when compared with *Nicotiana tabacum* [22], due to the presence of a large inversion in the Asteraceae, Fabaceae, and Poaceae family [19–21]. These structural differences in cp genomes can be exploited in the phylogenetic classification and molecular improvement of plants like *Artemisia frigida*. In addition, comparative analysis of cp genomes from distant and closely related species will not only allow for understanding the molecular evolution of cp genomes, but also facilitate the association of important traits controlled by plastid genomes.

One strategy for improving a plant species is through chloroplast genetic engineering to add high-value agronomic traits via transgenic expression [23], or to engineer multi-gene expression components in a single transformation event [23–25]. Plastid transformation, achieved via homologous recombination, is very advantageous compared to nuclear genome transformation mainly because it can generate high levels of expression and the recombinant DNA is more easily contained since chloroplasts are maternally inherited in most species of angiosperm [26]. Furthermore, chloroplast genetic engineering has also been widely used in basic research to understand plastid biogenesis and function [27–29].

Traditionally, sequencing of plastid genomes is done by isolation of chloroplasts followed by purification and amplification of plastid DNA for library construction and sequencing. Recently, a number of cp genome sequences are being reported using next-generation sequencing techniques due to the advantages of high-throughput, time-savings, and low-cost [30–33]. We report the complete cp genome sequence of *Artemisia frigida*, a kind of Mongolian traditional medicinal plant, using 454 pyrosequencing methods (Roche GS FLX+). We also describe details in the cp genome assembly, annotation, and comparative analysis with the sequences of cp genomes from other angiosperm species, including the six completed Asteraceae cp genomes. We identified and characterized a unique sequence rearrangement event in the *Artemisia frigida* cp genome, which resulted in the inversion of gene order in the SSC region as compared with other Asteraceae species. This work will lay a foundation for the molecular biology study and genetic improvement of *Artemisia frigida* in the future.

**Methods**

**DNA Sequencing**

A wild diploid *Artemisia frigida* (accession number NM1) from our germplasm collection from the Naimanqi area in Inner Mongolia Autonomous Region, China, was used for total DNA isolation from one gram of leaf tissue using the DNeasy Plant Mini Kit (Qiagen, CA, USA). The DNA (1 µg) was sheared by nebulization, subjected to 454 library preparation and shotgun sequencing using the Genome Sequencer (GS FLX+) platform [34] at the in-house facility (USDA-ARS, Western Regional Research Center, USA). The obtained nucleotide sequence reads were assembled using the GS De Novo Assembler version 2.6 and visualized by CONSED [35]. The assembled sequences and unassembled sequences were analyzed by BlastN and BlastX program against GenBank cp genome data to find *Artemisia frigida* cp genome sequence.

**Genome Analysis**

The genome was annotated using the program DOGMA (Dual Organellar GenoMe Annotator) [36]. The predicted annotations were verified using BLAST similarity search [37]. All genes, rRNAs and tRNAs were identified using the plastid/bacterial genetic code. The frequency of codon usage was calculated from exon sequences of all protein-coding genes in the *Artemisia frigida* genome. Inversions in the *Artemisia frigida* cp genome were identified by comparison to the sequence in the inverted region of *Lactuca sativa* (DQ993916) [30], *Helianthus annuus* (NC_007977) [38], and *Ageratina adenophora* (NC_015621) [41], and *Jabouina vulgaris* (NC_015543) [42], which are all in the Asteraceae family, was performed using the mVISTA program in Shufle-LAGAN mode [43], using the sequence annotation information of *Artemisia frigida*.

The gene orders in SSC regions of *Artemisia frigida* were compared with the above 6 Asteraceae species and 4 other species including *Nicotiana tabacum* (Solanaeaeae) (ncucore), *Piper crenatum* (magnolilids) (NC_000457) [44], *Dioscorea elephantipes* (Dioscoreaceae) (monocots) (NC_009601) [45], and *Chloranthus spicatus* (Chloranthaceae) (NC_009398) [45]. REPuter [46] was used to identify and locate forward, palindromic, reverse, and complement sequences with n ≥ 20 bp and a sequence identity ≥90%. We ran the same REPuter analyses against the other 6 Asteraceae species chloroplast genomes that were used for mVISTA to assess the relative number of repeats in chloroplast genomes. Microsatellite markers were predicted using MISA [47]. In the search for SSR standards, we defined SSRs as mononucleotide repeats ≥10 bases, dinucleotide repeats ≥12 bases, trinucleotide repeats ≥13 bases, tetranucleotide repeats ≥20 bases, pentanucleotide repeats ≥20 bases, and hexanucleotide or greater repeats ≥24 bases.

**PCR Amplification**

To acquire a high quality complete chloroplast genome sequence, 129 primers (Table S1) were designed to increase the sequence accuracy by correcting 454 sequencing errors occurred in the homopolymer regions and to confirm the four junction regions between the IRs and SSC/LSC. PCR products were sequenced using BigDye V3.1 Terminator kit for ABI3730XL (Applied Biosystems, Foster City CA) and assembled into the complete chloroplast genome sequence using CONSED software.
Figure 1. The map of the chloroplast genome of *Artemisia frigida*. IR, inverted repeat; LCS, large single copy region; SSC, small single copy region; Inv1, inverted sequence region 1; Inv2, inverted sequence region 2. Genes containing introns are marked with *.
doi:10.1371/journal.pone.0057533.g001
### Table 1. Genes present in *Artemisia frigida* chloroplast genome.

| No. | Gene types | Gene products |
|-----|------------|---------------|
| 1   | photosystem I | psaA, B, C, I, J, ycf3*, ycf4 |
| 2   | photosystem II | psbA, B, C, D, E, F, H, I, J, K, L, M, N, T, Z |
| 3   | Cytochrome b6/f | petA, B*, D*, G, L, N |
| 4   | Atp synthase | atpA, B, E, F*, H, I |
| 5   | Rubisco | rbcL |
| 6   | NADH oxidoreductase | ndhA*, B*, C, D, E, F, G, H, I, J, K |
| 7   | Large subunit ribosomal proteins | rpl20c, 14, 16s, 20, 22, 23s, 32, 33, 36 |
| 8   | Small subunit ribosomal proteins | rps2, 3, 4, 7s, 8, 11, 12s, 14, 15, 16s, 18, 19 |
| 9   | RNAP | rpoA, B, C1*, C2 |
| 10  | Other proteins | accD, ccsA, cemA, cipP*, matK, infA |
| 11  | Proteins of unknown function | ycf1, 2, ycf15* |
| 12  | Ribosomal RNAs | rnl16, 23, 4.5, 5 |
| 13  | Transfer RNAs | tma(UGC)*, tggA, di(guc), i(uuc), f(gaa), gi(ucc)*, gi(uc), hi(gu), icau*, igaA*, i(ku)u*, i(ka)A, i(lua)A, i(lua)G, mi(cau), mi(cau), ni(gu)u*, pl(ugg), qi(uug), ra(cg)*, ri(ucu), si(gcu), si(gga), si(uga), t(ugg), t(ugu), v(gac)*, v(ua)c*, w(cca), y(gua) |

*Gene containing two introns.
*Gene containing a single intron.
Two gene copies in IRs.
Gene divided into two independent transcription units.

To confirm the assembly accuracy at the junction regions of IRb with SSC and SSC with IRa in the *Artemisia frigida* cp genome, four primers (Table S1) were designed for the junction of IRb/SSC and SSC/IRa in *Artemisia frigida*. These primers were also used to examine the junction regions in other accessions of *Artemisia frigida* originated from Mongolia (PI 639180) and United States (W6 30042 from Colorado and AG 258 from Alaska) (available at Germplasm Resources Information Network http://www.ars-grin.gov/). The same strategy was also used to examine the junctions in *Helianthus annuus* and *Lactuca sativa* based on the sequence NC_007977 and DQ383816, respectively. The accession of HA410 for *Helianthus annuus* and the accession of LS01 for *Lactuca sativa* were used as template for PCR analyses. For PCR, each 20 μl PCR reaction system included 1× GoTaq buffer, 0.25 mM dNTP, 4 μM primers, 1 unit of homemade Taq polymerase, 6% DMSO, and 20 ng of DNA. The PCR amplification reactions were performed with 35 cycles of 50 sec denaturation at 94°C, 50 sec annealing at 52°C, and 90 sec extension at 72°C. PCR products were separated by electrophoresis in 1.5% agarose gel.

### Phylogenetic Analysis

A set of 61 protein-coding genes which have been analyzed in other species [48–50] were used to infer the phylogenetic relationships among *Artemisia frigida*, 36 angiosperm lineages previously published in the GenBank database, and 2 gymnosperms, *Picea abies* and *Ginkgo biloba* (Table S2). Sequences were aligned using ClustalW in MEGA5 [51], the alignment was edited manually. Phylogenetic analyses using maximum parsimony (MP) and maximum likelihood (ML) were performed with MEGA5 and the parameters were the same as Young described [52]. The high sequence diversity region found in the ndhF gene and the trnL-trnF region [53] were utilized for phylogenetic analyses among Asteraceae species. Both ndhF and trnL-trnF sequences of 92 species were downloaded from GenBank (Table S3). The concatenated sequence of ndhF and trnL-trnF were aligned using MUSCLE version 3.8 [54]. Maximum parsimony (MP) and maximum likelihood (ML) trees were reconstructed using above parameters with MEGA5. The gaps in the sequence alignment were treated as missing data.

### Results and Discussion

#### Chloroplast Genome Assembly and Validation

One sequencing run of *Artemisia frigida* genomic DNA was carried out using Roche 454 sequencing technology on the GS FLX+ system. A total of 645,965 quality-filtered sequence reads were generated with the average read length of 398 bp, representing 387 Mb sequence data. Assembly of the nucleotide sequence reads was performed to obtain non-redundant contigs and singletons using the GS De Novo Assembler. In total, 28,129 contigs were assembled with a N50 contig size of 910 bp and a total accumulated length of 15,021,516 bp, representing only 0.15× coverage of the *Artemisia frigida* nuclear genome (2,567 Mb). The resulting contigs were searched against NCBI GenBank chloroplast genome database using BlastN and BlastX. Five contigs, with nucleotide length of 43,781 bp, 37,022 bp, 24,972 bp, 18,397 bp, and 1,937 bp were identified to be part of the cp genome. The number of sequence reads that were assembled into these five contigs were 4,465 (0.69% of the total 645,965 sequence reads) with an average read length of 598 bp. CONSED was used to reassemble these sequence reads extracted from the 454 sequence dataset. With the involvement of manual editing, a single sequence contig representing the entire *Artemisia frigida* cp genome was achieved. The average sequence depth of each nucleotide on the *Artemisia frigida* cp genome was 17.67×. The high sequence coverage from 454 reads allows for generation of consensus sequence with high accuracy.

Traditionally, sequencing of chloroplast genomes involved chloroplast isolation followed by purification of its DNA for library construction and sequencing [41]. Recently, several chloroplast genomes have been sequenced from nuclear genomic DNA with the use of high-throughput sequence systems such as...
SOLiD [55], Illumina [56], and 454 GS FLX platforms [30,32,57]. The chloroplast genomes are present in a high copy number in a single cell and often co-purified with nuclear genomic DNA as by-product or contamination. Because of their relative small genome sizes, the low percentage of chloroplast DNA sequence reads from the total nuclear genomic sequences generated by the next-generation high-throughput sequencing technologies can provided sufficient coverage for the assembly of chloroplast genomes [57]. Compared with the sequence read length generated by Illumina (~150 bp) and SOLiD (~50 bp) sequencing methods, 454 GS FLX can generate longer sequence reads (~400 bp). In general, longer reads will provide better sequence assembly at the same or similar sequence coverage, particularly for complex genomes with high repeat contents [57]. In our study, we used the 454 GS FLX+ platform which produced an average read length of 638 bp for the Artemisia frigida cp genome sequence reads. In the previous reports of cp genome sequencing by Roche 454, the average read length of mungbean, date palm and Boea hygrometrica are 217 bp [30], 347 bp [32], and 339 bp [57], respectively. Therefore, the sequence read length for the Artemisia frigida cp genome is more than 300 bp longer than that for these three cp genomes. However, the percentage of reads representing chloroplast DNA for Artemisia frigida (0.69%) is lower than mungbean (5.22%) [30], date palm (8.8%) [32], and Boea hygrometrica (0.91%) [57]. In our study, 387 Mb sequences representing 0.15× coverage of the Artemisia frigida genome (2,567 Mb) had enough chloroplast reads to assemble its entire cp genome, while 1× genome coverage (300 Mb) are required for the complete assembly of the cp genome in Boea hygrometrica. Our results indicated that sequence reads generated by 454 GS FLX+ platforms may be a better choice for de novo sequencing and assembly of organelle genomes since it can produce longer reads and make assembly easier and more robust.

Figure 2. Sequence alignment of seven sequenced cp genomes in the Asteraceae family. Sequences of chloroplast genomes were aligned and compared using mVISTA program. A cut-off of 70% identity was used for the plot and the Y-scale represents the percent identity ranging from 50–100%. Blue represents exons, green-blue represents tRNA and rRNA genes, and pink represents conserved non-coding sequences (CNS). Grey arrows the direction of transcription; horizontal blue lines indicate the position of IRa and IRb; horizontal red lines indicate the position of Inv1 and Inv2.

doi:10.1371/journal.pone.0057533.g002
The homopolymer issues in the 454 sequencing method usually cannot be overcome by increasing the coverage of the sequence data [30–32,44]. To provide an accurate sequence for the *Artemisia frigida* chloroplast genome, resequencing of homopolymer regions by Sanger sequencing method was performed to determine the exact homopolymer lengths. PCR primer pairs (Table S1) were designed to cover 125 homopolymer regions (n > 7 bp) based on the sequence of the initial *Artemisia frigida* cp genome assembly. Most of these homopolymer regions occurred in the non-coding regions. The results from resequencing of homopolymer regions showed that 29 base pairs were added or excluded in 125 homopolymers. This final *Artemisia frigida* cp genome sequence has been submitted to GenBank (GenBank ID: JX293720).

The complete cp genome size of *Artemisia frigida* is 151,076 bp, including the LSC of 82,740 bp, the SSC of 18,394 bp and a pair of IRs of 24,971 bp each (Figure 1). The IRs span from *rpl2* to a portion of *ycf1*. The average AT content of the *Artemisia frigida* cp genome is 62.52%, which is consistent with the AT content reported for other plant cp genomes [41]. The AT contents of the LSC and SSC regions are 64.42% and 69.17%, respectively, whereas that of the IR regions is 56.93%.

**Genome Organization and Gene Content**

The *Artemisia frigida* cp genome contains 114 unique genes, including 30 tRNA genes, 4 rRNA genes, and 80 predicted protein-coding genes (Table 1). In addition, there are 18 genes duplicated in the IR, making a total of 132 genes present in the *Artemisia frigida* cp genome (Figure 1). Protein-coding genes, tRNAs, and rRNAs make up 52.08%, 1.85%, and 5.99% of the genome, respectively, while the remaining 40.08% are non-coding introns, intergenic spacers, and pseudo genes. There are 18 intron-containing genes, including 6 tRNA genes and 12 protein-coding genes, almost all of which are single-intron genes except for *ycf3* and *clpP*, each having two introns. The *tmk*-UUU gene has the largest intron (2,564 bp) where another gene, *matK*, is located in it. We found that the two *rps12* genes, one in each IR region, are trans-spliced, with one of its exons located in the LSC (5') and the other exon in the IR regions. Among the three pseudo genes, *ycf68* in the IR become pseudogenization due to several premature stop codons present in its coding sequence (Figure 1). Another two pseudo genes, *ycf1* and *rps19*, are located in the boundary regions between IRb/SSC and IRa/LSC, respectively. Incomplete duplication of the normal copy of *ycf1* and *rps19* at these boundaries has resulted in a lack of protein-coding ability.

Instead of a common ATG start codon, we identified two instances where ACG is used as a start codon: in *ndhD* and *psbL*. In addition, one GUG start codon is found in *rps19*. The ACG start codon has been shown to convert to an AUG initiation site as reported in *Nicotiana tabacum* [58]. Such RNA editing in the translation process likely also occurs in the *Artemisia frigida* cp genome.
There are 30 unique tRNA genes (7 tRNA genes duplicated in the IR) including two trnG-UCC genes in LSC region because of one with intron. These tRNA genes represented 20 amino acids identified in the cp genome (Table S2). A total of 26,226 codons represent the coding capacity of 86 protein-coding genes in the Artemisia frigida cp genome (Table S2). Isoleucine (2,208, 8.42%) and cysteine (288, 1.10%) are the most and the least abundant amino acids, respectively.

The cp genome size of Artemisia frigida is the third smallest among the seven completed Asteraceae cp genomes (after including Artemisia frigida). It is larger than Jacobaea vulgaris (150,689 bp) and Ageratina adenophora (150,698 bp) (Table S3), but smaller than the cp genomes of Lactuca sativa, Helianthus annuus, Guizotia abyssinica, and Parthenium argentatum by 1.70 kb, 28 bp, 0.69 kb, and 1.73 kb, respectively. Artemisia frigida has the smallest LSC region among these sequenced Asteraceae cp genomes.

Although chloroplast genomes are considered highly conserved among land plants, regions with highly sequence polymorphisms were often observed even among closely related species [59]. Alignments of seven sequenced cp genome sequences available in the Asteraceae family were performed using mVISTA program family with the new annotation of Artemisia frigida to reveal their sequence variations. This analysis showed that the coding region is more conserved than the non-coding region, and that the most divergent coding regions in the seven genomes were ycf1, accD, ccsA, rps16 and rpoC1 (Figure 2).

In addition to the various nucleotide divergence in different regions, sequence arrangements also occurred in cp genomes. Comparing with the cp genome of Nicotiana tabacum, the cp genome of Artemisia frigida had two inversion events in the LSC region. The sizes of the two inversions were 22,837 bp (Inv1) and 3,421 bp (Inv2). The large inversion (Inv1) changed the order genes located in this inversion region as compared to that in Nicotiana tabacum (Figure 1 and Figure 2). The second small inversion (Inv2) is within the region of the large inversion. Both inversions started at the position of 8,837 bp while Inv2 ended at 12,257 bp and Inv1 at 31,674 bp. It appears that the two inversions occurred within the same evolutionary time period as...
what existed in most of the Asteraceae family, including
*Lactuca sativa, Helianthus annuus, Guizotia abyssinica, Parthenium argentatum, Ageratina adenophora,* and *Jacobaea vulgaris* [38–42,60] (Figure 1 and Figure 2).

We also analyzed the gene order in the SSC region. The tobacco cp genome is often regarded to be unaltered [22] and therefore used as reference here (Figure 3). The gene order in the SSC region in tobacco and *Artemisia frigida* begins with *ndhF,* and

**Table 2.** Repeat sequences in the *Artemisia frigida* chloroplast genome.

| Repeat No. | Repeat size (bp) | Repeat 1 start | Repeat 2 start | Repeat type | Repeat 1 location | Repeat 2 location |
|------------|------------------|----------------|----------------|-------------|------------------|------------------|
| 1          | 33               | 5019           | 5034           | F           | IGS (trnK-UUU-psa16) | IGS (trnK-UUU-psa16) |
| 2          | 32               | 8441           | 34788          | F           | *trnS-GCU*       | *trnS-UGA*       |
| 3          | 30               | 27926          | 29521          | F           | IGS (atpF-atpA)  | IGS (atpA-trnR-UCU) |
| 4          | 32               | 37989          | 40213          | F           | *psaB*           | *psaA*           |
| 5          | 30               | 38000          | 40224          | F           | *psaB*           | *psaA*           |
| 6          | 41               | 42994          | 96732          | F           | Intron (ycf3)    | IGS (ips7-ycf15) |
| 7          | 39               | 42996          | 118042         | F           | Intron (ycf3)    | Intron (ndhA)    |
| 8          | 35               | 42999          | 93683          | F           | Intron (ycf3)    | Intron (ndhB)    |
| 9          | 30               | 56190          | 56214          | F           | *rbcL*           | IGS (rbcL-accD)  |
| 10         | 30               | 66201          | 97916          | F           | IGS (*psa1*-rpl33) | IGS (ycf15-tnV-GAC) |
| 11         | 30               | 86397          | 147393         | F           | *ycf2*           | *ycf2*           |
| 12         | 60               | 89967          | 89985          | F           | *ycf2*           | *ycf2*           |
| 13         | 42               | 89967          | 90003          | F           | *ycf2*           | *ycf2*           |
| 14         | 30               | 89979          | 90005          | F           | *ycf2*           | *ycf2*           |
| 15         | 45               | 89982          | 90000          | F           | *ycf2*           | *ycf2*           |
| 16         | 39               | 96734          | 118042         | F           | IGS (*ips7*-ycf15) | Intron (ndhA)    |
| 17         | 30               | 105646         | 105678         | F           | IGS (rnm4.5-nds5) | IGS (rnm4.5-nds5) |
| 18         | 30               | 121294         | 121295         | F           | IGS (ips15-ycf1) | IGS (ips15-ycf1) |
| 19         | 36               | 122447         | 122777         | F           | *ycf1*           | *ycf1*           |
| 20         | 30               | 128108         | 128140         | F           | IGS (rnm4.5-nds5) | IGS (rnm4.5-nds5) |
| 21         | 60               | 143771         | 143789         | F           | *ycf2*           | *ycf2*           |
| 22         | 42               | 143771         | 143807         | F           | *ycf2*           | *ycf2*           |
| 23         | 52               | 143779         | 143797         | F           | *ycf2*           | *ycf2*           |
| 24         | 34               | 143779         | 143815         | F           | *ycf2*           | *ycf2*           |
| 25         | 30               | 8443           | 44679          | P           | *trnS-GCU*       | *trnS-UGA*       |
| 26         | 30               | 34790          | 44679          | P           | *trnS-GCU*       | *trnS-UGA*       |
| 27         | 41               | 42994          | 137043         | P           | Intron (ycf3)    | IGS (ycf15-ips7) |
| 28         | 35               | 42999          | 140098         | P           | Intron (ycf3)    | Intron (ndhB)    |
| 29         | 30               | 66201          | 135870         | P           | IGS (*psa1*-rpl33) | IGS (tnV-GAC-ycf15) |
| 30         | 48               | 72915          | 72915          | P           | IGS (*psbT-psbB*) | IGS (*psbT-psbB*) |
| 31         | 60               | 89967          | 143771         | P           | *ycf2*           | *ycf2*           |
| 32         | 42               | 89967          | 143771         | P           | *ycf2*           | *ycf2*           |
| 33         | 30               | 89979          | 143771         | P           | *ycf2*           | *ycf2*           |
| 34         | 45               | 89982          | 143771         | P           | *ycf2*           | *ycf2*           |
| 35         | 60               | 89985          | 143789         | P           | *ycf2*           | *ycf2*           |
| 36         | 45               | 90000          | 143789         | P           | *ycf2*           | *ycf2*           |
| 37         | 42               | 90003          | 143807         | P           | *ycf2*           | *ycf2*           |
| 38         | 30               | 90015          | 143807         | P           | *ycf2*           | *ycf2*           |
| 39         | 30               | 105646         | 128108         | P           | IGS (rnm4.5-nds5) | IGS (rnm4.5-nds5) |
| 40         | 30               | 105678         | 128140         | P           | IGS (rnm4.5-nds5) | IGS (rnm4.5-nds5) |
| 41         | 43               | 114927         | 114927         | P           | IGS (ndhD-psaC)  | IGS (ndhD-psaC)  |
| 42         | 39               | 118042         | 137043         | P           | Intron (ndhA)    | IGS (ycf15-ips7) |
| 43         | 31               | 121291         | 121291         | R           | IGS (*ips15-ycf1) | IGS (*ips15-ycf1) |

*partly in the IGS region; F-forward, P-palindrome, R-reverse, IGS-Intergenic spacer region.

doi:10.1371/journal.pone.0057533.t002
then is followed by the order of rpl32, trnL, ccsA, ndhD, psaC, ndhE, ndhG, ndhI, ndhA, ndhH and rps15, and ends with ycf1, which is extended into IRa regions. The gene orders of the other 6 species in the Asteraceae family are the completely same, but inverted compared to Artemisia frigida. Given the notion that most species in the Asteraceae family have the same gene order in the SSC region, it is likely that an inversion in the SSC region occurred before the divergence of species in the Asteraceae family. The fact that Artemisia frigida has the same gene order in the SSC region with Nicotiana tabacum suggests that re-inversion in the SSC region occurred in Artemisia frigida lineage.

To further confirm that the gene order in the SSC in Artemisia frigida is different from those in the Asteraceae family, four primers were designed for each species to amplify the junctions of IRb/SSC and SSC/IRa in Artemisia frigida, Helianthus annuus, Lactuca sativa and from different accessions of Artemisia frigida (Figure 4A). The primer pairs of P1F/P1R and P2F/P2R amplified PCR products in Artemisia frigida while the primer combinations of P1F/P2F and P1R/P2R had no PCR products (Figure 4B). In two other species in the Asteraceae family, HA410 (Helianthus annuus) and LS01 (Lactuca sativa) provided amplified PCR products using the primer pairs of P1F/P1R and P2F/P2R. No PCR products were amplified with the primer pairs of P1F/P2F and P1R/P2R (Figure 4B). These results indicated that the SSC region in Artemisia frigida is re-inverted comparing to Helianthus annuus and Lactuca sativa. We further examined this re-inversion event in other Artemisia frigida accessions collected from different geographical regions (PI 639180 from Mogol, W6 30042 from Colorado, and AG 258 from Alaska). The results showed that these three accessions provided PCR products with the primer pairs of P1F/P1R and P2F/P2R (Figure 4C), indicating that they have the same gene order with the sequenced Artemisia frigida accession. It is likely that these accessions in Artemisia frigida shared the same re-inversion event.

The identification and characterization of inversion and re-inversion events in Artemisia frigida suggests that the SSC might be an active region for sequence rearrangements in plant cp genomes. We therefore searched the SSC regions of sequenced cp genomes in plants. Most species share the same cpDNA organization in the SSC region with Nicotiana tabacum [22]. However, some angiosperm species such as Piper cernualum (magnoliids) [44], Dioscorea elephantipes (Dioscoreaceae) (monocots) [45], and Chloranthus spicatus (Chloranthaceae) (basal angiosperm) [45] have an inverted SSC region (data not shown). Although chloroplast genomes are generally conserved in gene order in land plants [17,61], several sequence rearrangements in cp genomes from different plant species have been reported, including a large inversion in LSC region [19–21,62], IR contraction or expansions into single copy region with inversions [30,63], and SSC region as shown in this study. It has been proposed that intramolecular recombination events are the causes of sequence rearrangements in the cp genomes [64,65]. These sequence rearrangements that alter cp genome structures in related species could provide the genetic diversity useful for molecular classification and evolution studies.

Repeat Sequence Analysis and Distribution of cp SSR

We used REPuter to analyze the repeat sequences in the Artemisia frigida cp genome and found 24 direct (forward) repeats, 18 inverted (palindrome) repeats, and 1 reverse repeat of at least 30 bp long and sequence identity ≥90% in the cp genomes. F, P, R, and C indicate that the repeats matching in forward, palindrome, reverse, and complement orientations, respectively. Different repeat unit lengths are indicted with different color.

doi:10.1371/journal.pone.0057533.g005

Figure 5. Repeat sequence analysis in seven sequenced Asteraceae chloroplast genomes. REPuter was used to identify repeat sequences with length ≥30 bp long and sequence identity ≥90% in the cp genomes. F, P, R, and C indicate that the repeats matching in forward, palindrome, reverse, and complement orientations, respectively. Different repeat unit lengths are indicted with different color.

The repeat structures of the other six species within Asteraceae were also analyzed by REPuter (Figure 5). Forward repeats and inverted repeats are common in these species. The repeat structure of Artemisia frigida, which is from the Anthemideae tribe, is similar to those of Lactuca sativa, Guizotia abyssinica, and Jacobea vulgaris, which are from the Cichorieae, Heliantheae alliance, and Senecioneae tribe.
Parthenium argentatum, and Ageratina adenophora contains the greatest number of repeats that are 40 bp or greater in length. The reason for this may be because of the different subtribe and genus to which Ageratina adenophora belongs.

Another type of repeat sequences frequently occurred in the cp genomes is the simple sequence repeats (SSRs). The distribution of SSRs was analyzed for the Artemisia frigida cp genome. Thirty-eight (38) mononucleotide SSRs (92.68%), also called homopolymers, 2 dinucleotide SSRs (5.26%), and 1 trinucleotide SSR (2.63%) were identified (Table 3). Thirty-three (33) of the 41 SSR loci were found in the intergenic regions, 3 were located in introns, and the other 5 SSRs were located in genes. Among the 38 mononucleotide SSRs, only one C/G type was found, while all others belonged to the A/T type. The repeat number of mononucleotide motifs ranged from 10 to 27, and 52.63% of the repeats were A/T type with repeat number 10 (Figure 6).

Chloroplast SSRs (cpSSRs) are generally short mononucleotide tandem repeats that, when located in the noncoding regions of the cp genome, commonly show intraspecific variation in repeat number [66,67]. In our study, 34 of 38 mononucleotide SSR loci (≥10 bases) occurred in non-coding regions, including 31 in the intergenic regions and 3 in introns (Table 2). Compared with other species of angiosperms, the number of mononucleotide cpSSR in Artemisia frigida found in non-coding regions of the cp genome was much greater. Several species contain less than 34 mononucleotide cpSSRs in non-coding regions, including Helianthus annuus (Asteraceae) (30), Panax ginseng (Araliaceae) (9), Daucus carota (Apiaceae) (23), 7 species from three genera in Solanaceae (28–31), 5 species from two genera in Convolvulaceae (12–33), as well as other species [68]. However, Artemisia frigida also contains less non-coding mononucleotide cpSSRs than Cucumis sativus (Cucurbitaceae) (47), Citrus sinensis (Rutaceae) (60), Vitis vinifera (Vitaceae) (46), and other species [68]. Like other chloroplast markers which are uniparental in inheritance, cpSSRs have been widely used in the analysis of plant population structure, diversity, differentiation and maternity analysis. Inter- and intra-specific chloroplast variation has also been studied within plant populations, including many species of Poaceae [69–71], Solanaceae [72], and Brassicaceae [73,74]. While the applicable use of cpSSR is still largely centered on economically important plants and their relatives, the potential for cpSSRs to offer unique insights into ecological and evolutionary processes in wild plant species is quite substantial and not yet fully realized [68]. Our results provide cpSSR markers for the

Table 3. Simple sequence repeat in Artemisia frigida chloroplast genome.

| No. | location     | SSR type | SSR start | SSR end |
|-----|--------------|----------|-----------|---------|
| 1   | psbA-trnK-UUU | A(T)10   | 1674      | 1683    |
| 2   | trnK-UUU-matK | A(T)15   | 1957      | 1971    |
| 3   | trnK-UUU-rps16 | A(T)10   | 4733      | 4742    |
| 4   | rps16 intron  | C/G14    | 5430      | 5443    |
| 5   | psbM-trnD-GUC | A(T)10   | 11205     | 11215   |
| 6   | trnT-UCU-trpB | A(T)15   | 12645     | 12659   |
| 7   | rpoB         | A(T)10   | 13335     | 13344   |
| 8   | rpoC1 intron  | A(T)10   | 16604     | 16613   |
| 9   | rpoC1        | A(T)10   | 17957     | 17966   |
| 10  | rps16        | A(T)10   | 18383     | 18392   |
| 11  | rpoC2-rps2   | A(T)12   | 23084     | 23095   |
| 12  | atpF-atpH    | A(T)10   | 26017     | 26026   |
| 13  | atpF-atpA    | A(T)12   | 27936     | 27947   |
| 14  | atpA-trnR-UCU| A(T)11   | 29531     | 29541   |
| 15  | trnR-UCU-trnG-UCC | A(T)12 | 29749 | 29760 |
| 16  | psbC-trnS-UAA | A(T)19  | 34644     | 34662   |
| 17  | psbZ-trnG    | A(T)10   | 35513     | 35522   |
| 18  | psaA-trnC    | A(T)13   | 41570     | 41582   |
| 19  | psaA-trnC    | A(T)10   | 41754     | 41763   |
| 20  | ndhC-trnV-UAC| A(T)10   | 50173     | 50182   |
| 21  | atpB-rbcL    | A(T)14   | 54366     | 54379   |
| 22  | rbcL-accD    | A(T)10   | 56590     | 56599   |
| 23  | petA-psbJ    | A(T)11   | 62051     | 62061   |
| 24  | psbE-psbL    | A(T)10   | 64272     | 64281   |
| 25  | rpl18-rps20  | A(T)11   | 67100     | 67110   |
| 26  | cplP intron  | A(T)11   | 69197     | 69207   |
| 27  | rpoA         | A(T)10   | 76711     | 76720   |
| 28  | rps8-rpl14   | A(T)14   | 79128     | 79141   |
| 29  | rpl14-rpl16  | A(T)10   | 79631     | 79640   |
| 30  | rns-trnR-ACG | A(T)11   | 100607    | 100617  |
| 31  | trnR-ACG-trnN-GUU | A(T)10 | 106475 | 106484 |
| 32  | ndhF-rpl12   | A(T)10   | 110893    | 110902  |
| 33  | rpl32-trnL-UGA | A(T)10 | 111959 | 111968 |
| 34  | ndhD-psaC    | A(T)10   | 114977    | 114986  |
| 35  | rps15-ycf1   | A(T)27   | 121295    | 121321  |
| 36  | rps15-xcf1   | A(T)10   | 121534    | 121544  |
| 37  | trnN-GUU-trnR-ACG | A(T)10 | 127333 | 127342 |
| 38  | trnR-ACG-rrs5 | A(T)10   | 127800    | 127810  |
| 39  | petN-psbM    | A(T)6    | 10217     | 10228   |
| 40  | trnT-GGU-psbD | A(T)6   | 30963     | 30974   |
| 41  | ycf1         | A(T)6    | 125892    | 125906  |

doi:10.1371/journal.pone.0057533.t003

Figure 6. The distribution and frequency of simple sequence repeats in the Artemisia frigida cp genome. The SSR length is implied by the number of repeat for each SSR type in the Y-axes The percentage of a repeat with specific length from the total number of repeats is indicated above the bars. doi:10.1371/journal.pone.0057533.g006

tribe, respectively. Helianthus annuus, Parthenium argentatum, and Ageratina adenophora are all in the same Heliantheae alliance tribe, but the repeat structures of these species are different. The Helianthus annuus cp genome contains the greatest number and variety of repeats, while Parthenium argentatum shares the same repeat structure, but has fewer overall repeats. Of the 7 Asteraceae cp genomes studied, Ageratina adenophora contains the greatest total number of repeats that are 40 bp or greater in length. The reason for this may be because of the different subtribe and genus to which Ageratina adenophora belongs.
Phylogenetic Analysis

*Artemisia frigida* belongs to the tribe Anthemideae in the Asteraceae family. Several studies have been conducted to analyze the phylogenetic relationship in the Asteraceae family based on chloroplast coding or non-coding sequences [53,75,76]. The phylogenetic evolution of *Artemisia frigida* has only been studied by using *tnmSUGA-tnmMCAU, trnSGCU-trnCGCA* [13], *pokA-trnH, rpl32-trnL* [16], and nucleic DNA sequence 3-ETS, ITS [11] within the genus *Artemisia* L. The chloroplast gene *ndhF* has been used successfully to conduct phylogenies at the intergenetic and interfamilial levels within Asteraceae [59], Bromeliaceae [77], and Acanthaceae [70], among others [79]. The *rpl32-trnL* non-coding region has been widely used for reconstructing phylogenies between closely related species and for identifying plant species [80–82]. Many uncertainties are still remaining in the molecular phylogeny of the Asteraceae family and molecular evidence to support the phylogenetic position of *Artemisia frigida* is still lacking. The availability of completed *Artemisia frigida* cp genome provided us with the sequence information to study the molecular evolution and phylogeny of *Artemisia frigida* with closely related species. We first extracted 61 protein-coding genes from sequenced cp genomes from species belonging to 59 taxa, including 5 Asteraceae species (Table S4). After sequence alignment, all positions containing gaps and missing data were eliminated, leaving a total of 39,140 positions in the final dataset. ML analysis based on the Tamura-Nei model [83] resulted in a single tree with ln L = −415091.42 (Figure 7). Bootstrap analysis indicated that 44 of 55 nodes were supported by values ≥95% and 40 of these with bootstrap values of 100%. MP analysis resulted in a single tree with a length of 81, 210, a consistency index of 0.3447, and a retention index of 0.5978 (data not shown). The ML and MP trees had similar phylogenetic topologies. *Artemisia frigida* grouped together with *Helianthus annuus*, *Guizotia abyssinica*, and *Ageratina adenophora* in the supertribe Helianthodae, all within the subfamily Asteroideae. *Lactuca sativa* was grouped with the tribe Lactuceae of another subfamily, Cichorioideae, within the Asteraceae. The five species in the Asteraceae family were clustered into Asterales and placed within the euasterids II. In addition, the tribe Anthemideae demonstrates a closer relationship with the tribe Heliantheae than with Lactuceae. Through our analysis it was determined that *Cucumis*, whose phylogenetic position was not yet completely determined [84], was grouped within the eurosids I clade, which is comparable to the result of Nie et al. [41].

Further phylogenetic analysis was performed using *ndhF* and *trnL-F* sequences on 90 species in the Asteraceae family including *Artemisia frigida* (Table S5 and Figure S1). Both ML and MP trees were reconstructed for placement of phylogenic positions of these selected species. There were a total of 2,417 nucleotide alignment sites in the final dataset for the tree reconstructions. ML analysis based on the Tamura-Nei model [83] generated a single tree with ln L = −20049.26 (Figure S1). MP analysis resulted in a single tree with a length of 2,625, a consistency index of 0.4383, and a retention index of 0.6292 (data not shown). Both ML and MP trees provide strong support for *Artemisia frigida* being clustered into the Anthemideae tribe in the subfamily Asteroideae. As for the 6 species which have sequenced cp genomes, *Helianthus annuus, Parthenium argentatum, Ageratina adenophora, and Guizotia abyssinica* fall into the Heliantheae alliance tribe of Asteroideae, *Jacobsia vulgaris* is located in the Senercionae tribe of Asteroideae, and *Lactuca sativa* is grouped into the Cichorieae tribe in the Cichorioideae subfamily. Finally, *Artemisia frigida* grouped into the Anthemideae tribe has a closer relationship with the Heliantheae alliance and Seneconae tribes than with Cichorieae in the phylogenetic tree (Figure S1). The phylogeny obtained with the molecular data is consistent with the classification based on phenotypic observation [85].

Conclusions

Genomic DNA of *Artemisia frigida* was sequenced using 454 pyrosequencing technology and the complete chloroplast genome was identified and annotated. This is the first cp genome sequenced in the Anthemideae tribe within the Asteraceae family. We found that most Asteraceae species have an inverted SSC region in comparison with the unaltered tobacco cp genome. However, re-inversion event has occurred in the SSC region in *Artemisia frigida* lineage, suggesting that SSC might be an active region for inversion events. Repeat sequences were also analyzed in this study to explore the use of polymorphic microsatellites at the intra- and inter-specific level among *Artemisia* species. Sixty-one (61) protein-coding sequences from 59 species were employed to construct phylogenic trees, providing a strong support for a monophyletic group of the asteroids II clade. *Artemisia frigida* also demonstrated a close relationship to *Helianthus annuus, Guizotia abyssinica*, and *Ageratina adenophora*, which belong to the subfamily Asteroideae. In the Asteraceae family, *Artemisia frigida* clustered into the Anthemideae tribe in the subfamily Asteroideae based on *ndhF* and *trnL-F* gene sequence analysis. *Artemisia frigida* is the seventh cp genome of the Asteraceae family to be described. It will be useful for molecular ecology and molecular phylogeny studies within this species and also within the Asteraceae family.

Acknowledgments

The authors would like to thank Professor Chunlin Long, Minzu University of China, for critical review of the manuscript. The authors would like to thank Zhao Liu, North Dakota State University, for providing HA410 genomic DNA for *Helianthus annuus*.

Supporting Information

**Figure S1** Reconstruction of phylogenetic tree of Asteraceae and related families. The tree topology was constructed with the maximum likelihood method using the *ndhF* and *trnL-F* gene sequence regions. Bootstrap proportions shown above the branches. ln L = −20049.26. The position of the sequenced *Artemisia frigida* species is indicated with a red arrow.

(TIF)

**Table S1** List of primer pairs used in sequence verification and improvement of the *Artemisia frigida* chloroplast genome.

(DOC)

**Table S2** The codon-anticodon recognition pattern and codon usage for *Artemisia frigida* chloroplast genome.

(DOC)
Table S3  Size comparison of Artemisia frigida chloroplast genomic regions with those in other species of Asteraceae. (DOC)

Table S4  The GenBank accession numbers of all the 58 cp genomes used for phylogenetic analysis. (DOC)

Table S5  The GenBank accession numbers of ndhF gene and trnL-F gene region from all the 92 species used for phylogenetic analysis. (DOC)

References

1. Liu Y, Liu J, Tang L, Zhang W, Zhang L, et al. (2008) Current Status Research of Mongolian Medicine Artemisia Frigida Wild. Studies of Trace Elements and Health 25: 55–57.
2. Bremer K (1994) Asteraceae - Clades and Classification. Portland, Oregon: Timber Press.
3. Garcia S, Sana M, Garnatje T, Krevisch A, McArthur ED, et al. (2004) Variation of DNA amount in 47 populations of the subtribe Artemisinae and related taxa (Asteraceae, Anthemideae): karyological, ecological, and systematic implications. Genome 47: 1004–1014.
4. Wen T, Sun Q, Gai F, Meng X, Yi W, et al. (2011) Observation of chromosome karyotypes of Artemisia frigida in different ecological areas of Inner Mongolia. Acta Botanica Boreali-Occidentalia Sinica 31: 456–461.
5. Tian YH (2000) Chinese national medicine processing integration. Beijing: Chinese Ancient Books Press.
6. Zhang W, Tang L, Xie K, Cui J (2008) Determination of total flavonoids in raw and charred Mongolian drug Agr. Li Shizhen Medicine and Materia Medica Research 19: 2952–2953.
7. Cui J, Tang L, Lan R, Baoyinmadai (2006) Clinical analysis of 63 cases on curative effect of Artemisia frigida Wild in treatment of hemoptysis caused by bronchectasis. Journal of Minzu University of China (natural sciences edition) 1: 149–155.
8. Pao XL, Xie K, Cui J (2009) Fatty components of Artemisia frigida before and after carbonization by GC-MS. Li Shizhen Medicine and Materia Medica Research 20: 1848–1849.
9. Wang QH, Ao WL, Wang XL, Bao XH, Wang JH (2010) Two new flavonoid glycosides from Artemisia frigida Wild. J Asian Nat Prod Res 12: 950–954.
10. Taberlet P, Gielly L, Pautou G, Bouvet J (1994) Universal primers for amplification of three non-coding regions of chloroplast DNA Plant Mol Biol 17: 1105–1109.
11. Tkach NV, Hoffmann MH, Roser M, Korobkov AA, von Hagens KB (2008) Parallel evolutionary patterns in multiple lineages of arctic Artemisia L. (Asteraceae). Evolution 62: 184–190.
12. Soininen EM, Valentini A, Coisac E, Miquel C, Gielly L, et al. (2009) Analysis of diet of small herbivores: the efficiency of DNA barcoding coupled with high-throughput pyrosequencing for deciphering the composition of complex plant mezonts. Front Ecol 6: 16.
13. Garcia S, McArthur ED, Pellicer J, Sanderom SC, Valles J, et al. (2011) A molecular phylogenetic approach to western North America endemic Artemisia and allies (Asteraceae). Untangling the sagebrushes. Am J Bot 98: 638–653.
14. Híasola I, Opik M, Mes, M, Lijs L, Davinson J, et al. (2012) Plant species richness belowground: higher richness and new patterns revealed by next-generation sequencing. Mol Ecol Evol 21: 2004–2016.
15. Frank DA, Pontes AW, Marine EM, Carnana J, Raina R, et al. (2010) Grassland root communities: species distributions and how they are linked to aboveground abundance. Ecology 91: 3201–3209.
16. Riggins CW, Seigler DS (2012) The genus Artemisia (Asteraceae: Anthemideae) at a continental crossroads: Molecular insights into migrations, disjunctions, and recitals among Old and New World species from a Beringian perspective. Mol Phylogenet Evol 67: 471–490.
17. Palmer JD (1991) Plastid chromosomes: structure and evolution; Bogorad L, Chase CD, editors: Springer, Dordrecht, The Netherlands.
18. Chumley TW, Palmer JD, Mower JP, Forcade HM, Calie PJJ, et al. (2006) The complete chloroplast genome sequence of Pelargonium x hortorum: organisation and evolutionary implications of complete chloroplast genome sequence of a major invasive species, crofton weed (Ageratina adenonora). PLoS One 5: e12762.
19. Hagemann R (2004) The sexual inheritance of plant organelles; Daniell H, Chase CD, editors: Springer, Dordrecht, The Netherlands.
20. Doorduin L, Gravendeel B, Lammers Y, Ariyurek Y, Chin AWT, et al. (2011) Comparative analysis of organellar genomes with DOGMA. Bioinformaties 20: 3252–3255.
21. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410.
22. Timme RE, Kuehl JV, Boore JI, Rosen KJ, Cai Z, et al. (2006) A comparative analysis of the Lactuca and Helianthus (Asteraceae) plastid genomes: identification of divergent regions and categorization of shared repeats. J Mol Biol 36: 302–312.
23. De Costa B, Moar W, Lee SB, Miller M, Daniel H (2001) Overexpression of the Bc cry2Aa2 operon in chloroplasts leads to formation of insecticidal crystals. Nat Biotechnol 19: 71–74.
24. Ruiz ON, Hussein HS, Terry N, Daniel H (2003) Phytoremediation of organomercurials compounds via chloroplast genetic engineering. Plant Physiol 132: 1344–1352.
25. Quesada-Vargas T, Ruiz ON, Daniel H (2005) Characterization of heterologous multigene operons in transgenic chloroplasts: transcription, processing, and translation. Plant Physiol 138: 1746–1762.
26. Soininen EM, Valentini A, Coissac E, Miquel C, Gielly L, et al. (2009) Analysing of Montserrat platin genome and maternal inheritance of tranplastome. Plant Mol Biol 56: 203–216.
27. Chumley TW, Palmer JD, Mower JP, Forcade HM, Calie PJJ, et al. (2006) The complete plastid DNA inversion marks an ancient evolutionary split in the sunflower family (Asteraceae). Proc Natl Acad Sci USA 103: 5818–5822.
28. Joyce DJ, Joyce JL, Bellinger JA, Palmer JD (1996) The distribution and phylogenetic significance of a 500kbp chloroplast DNA inversion in the flowering plant family Leguminosae. Mol Phylogenet Evol 5: 429–438.
29. Joyce DJ, Davis J, Soergen RF, Garvin D, Anderson MJ (1992) Chloroplast DNA inversions and the origin of the grass family (Poaceae). Proc Natl Acad Sci USA 89: 7727–7732.
30. Shinozaki K, Obuse M, Tanaka M, Wakasugi T, Hayashida N, et al. (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J 5: 2043–2049.

Author Contributions

Conceived and designed the experiments: YL YQG. Performed the experiments: YL LD SZ XF. Contributed reagents/materials/analysis tools: YL QG HAY. Wrote the paper: YL YQG HAY.

Complete Chloroplast Genome of Artemisia frigida

PLOS ONE | www.plosone.org 13 February 2013 | Volume 8 | Issue 2 | e57533
sequences of four early-diverging angiosperms: Buxus (Buxaceae), Chlordanthus (Chloranthaceae), Dioscorea (Dioscoreaceae), and Illicium (Schisandraceae). Mol Phylogenet Evol 45: 547–563.

46. Kurtz S, Schleimercher C (1999) REPuter: fast computation of maximal repeats in complete genomes. Bioinformatics 15: 426–427.

47. Thiel T, Michalski V, Warshawsky RK, Grauer A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.). Theor Appl Genet 106: 411–422.

48. Sasaki C, Lee SB, Fjellheim S, Gueda C, Jansen RK, et al. (2007) Complete chloroplast genome sequences of Hordeum vulgare, Sorghum bicolor and Agrostis stolonifera, and comparative analyses with other grass genomes. Theor Appl Genet 115: 571–580.

49. Gormlykin VN, Hirsch-Elrst KL, Wolf S, Hellwig FH (2003) Analysis of the Amborella trichopoda chloroplast genome sequence suggests that amborella is not a basal angiosperm. Mol Biol Evol 20: 1499–1505.

50. Jansen RK, Kaittanis C, Sasaki C, Lee SB, Tomkins J, et al. (2006) Phylogenetic analyses of Vitis (Vitaceae) based on complete chloroplast genome sequences: effects of taxon sampling and phylogenetic methods on resolving relationships among lineages. BMC Evol Biol 6: 32.

51. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28: 2731–2739.

52. Young HA, Lanzatella CL, Sarath G, Tobias CM (2011) Chloroplast genome variation in upland and lowland switchgrass. PLoS One 6: e23980.

53. Panero JL, Funk VA (2008) The value of sampling anomalous taxa in phylogenetic studies: major clades of the Asteraceae revealed. Mol Phylogenet Evol 47: 757–782.

54. Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5: 113.

55. Wang W, Messing J (2011) High-throughput sequencing of three Lemnoideae chloroplast genomes from total DNA. PLoS One 6: e24670.

56. Noeck CJ, Waters DL, Edwards MA, Bowen SG, Rice N, et al. (2011) Chloroplast genome sequences from total DNA for plant identification. Plant Biotechnol J 9: 328–333.

57. Zhang T, Zhang X, Hu S, Yu J (2011) An efficient procedure for plant organellar genome assembly, based on whole-genome data from the 454 GS FLX sequencing platform. Plant Methods 7: 38.

58. Neckermann K, Zeha P, Igloi GL, Kossel H, Maier RM (1994) The role of RNA editing in conservation of start codons in chloroplast genomes. Gene 146: 177–182.

59. Kim KJ, Jansen RK (1995) ndhF sequence evolution and the major clades in the sunflower family (Asteraceae). Mol Biol Evol 12: 1783–1792.

60. Provan J, Allen EJ, Bancroft IA, Taheri S, Wetter CM, et al. (2000) Chloroplast DNA evidence on the ancient origin of the cereals. Mol Gen Genet 217: 185–194.

61. Provan J (2000) Novel chloroplast microsatellites reveal cytoplastic variation in Arabidopsis thaliana. Mol Biol Evol 9: 2183–2195.

62. Provan J, Allen EJ, Bancroft IA, Taheri S, Wetter CM, et al. (2000) Chloroplast DNA: A triple-sized inverted repeat, extensive gene duplications, and multiple inversions, and two repeat families. Proc Natl Acad Sci U S A 84: 769–773.

63. Provan J, Allen EJ, Bancroft IA, Taheri S, Wetter CM, et al. (2000) Chloroplast DNA variability in wild and cultivated rice (Oryza spp.) revealed by polymorphic chloroplast simple sequence repeats. Genome 40: 104–110.

64. Provan J, Corbet G, McNicol JW, Powell W (1997) Chloroplast DNA evidence on the ancient origin of the cereals. Mol Gen Genet 217: 185–194.

65. Provan J, Corbett G, Maier RM (2001) Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28: 2731–2739.

66. Provan J, Powell W, Hollingsworth PM (2001) Chloroplast microsatellites: new tools for studies in plant ecology and evolution. Trends Ecol Evol 16: 142–147.

67. Provan J, Powell W, Ruff DC, investigators (1999) Chloroplast microsatellite variation in upland and lowland switchgrass. PLoS One 6: e23980.

68. Provan J, Powell W, Ruff DC, investigators (1999) Chloroplast microsatellite variation in upland and lowland switchgrass. PLoS One 6: e23980.

69. Provan J, Powell W (1999) Polymorphic chloroplast simple sequence repeat primers for population and phylogenetic studies: major clades of the Asteraceae revealed. Mol Phylogenet Evol 8: 505–511.

70. Provan J, Powell W, Ruff DC, investigators (1999) Chloroplast microsatellite variation in upland and lowland switchgrass. PLoS One 6: e23980.

71. Provan J, Powell W, Ruff DC, investigators (1999) Chloroplast microsatellite variation in upland and lowland switchgrass. PLoS One 6: e23980.

72. Provan J, Powell W, Ruff DC, investigators (1999) Chloroplast microsatellite variation in upland and lowland switchgrass. PLoS One 6: e23980.

73. Provan J, Powell W, Ruff DC, investigators (1999) Chloroplast microsatellite variation in upland and lowland switchgrass. PLoS One 6: e23980.

74. Provan J, Powell W, Ruff DC, investigators (1999) Chloroplast microsatellite variation in upland and lowland switchgrass. PLoS One 6: e23980.

75. Provan J, Powell W, Ruff DC, investigators (1999) Chloroplast microsatellite variation in upland and lowland switchgrass. PLoS One 6: e23980.

76. Provan J, Powell W, Ruff DC, investigators (1999) Chloroplast microsatellite variation in upland and lowland switchgrass. PLoS One 6: e23980.

77. Provan J, Powell W, Ruff DC, investigators (1999) Chloroplast microsatellite variation in upland and lowland switchgrass. PLoS One 6: e23980.

78. Provan J, Powell W, Ruff DC, investigators (1999) Chloroplast microsatellite variation in upland and lowland switchgrass. PLoS One 6: e23980.

79. Provan J, Powell W, Ruff DC, investigators (1999) Chloroplast microsatellite variation in upland and lowland switchgrass. PLoS One 6: e23980.

80. Provan J, Powell W, Ruff DC, investigators (1999) Chloroplast microsatellite variation in upland and lowland switchgrass. PLoS One 6: e23980.

81. Provan J, Powell W, Ruff DC, investigators (1999) Chloroplast microsatellite variation in upland and lowland switchgrass. PLoS One 6: e23980.

82. Provan J, Powell W, Ruff DC, investigators (1999) Chloroplast microsatellite variation in upland and lowland switchgrass. PLoS One 6: e23980.

83. Provan J, Powell W, Ruff DC, investigators (1999) Chloroplast microsatellite variation in upland and lowland switchgrass. PLoS One 6: e23980.

84. Provan J, Powell W, Ruff DC, investigators (1999) Chloroplast microsatellite variation in upland and lowland switchgrass. PLoS One 6: e23980.

85. Provan J, Powell W, Ruff DC, investigators (1999) Chloroplast microsatellite variation in upland and lowland switchgrass. PLoS One 6: e23980.

86. Provan J, Powell W, Ruff DC, investigators (1999) Chloroplast microsatellite variation in upland and lowland switchgrass. PLoS One 6: e23980.

87. Provan J, Powell W, Ruff DC, investigators (1999) Chloroplast microsatellite variation in upland and lowland switchgrass. PLoS One 6: e23980.

88. Provan J, Powell W, Ruff DC, investigators (1999) Chloroplast microsatellite variation in upland and lowland switchgrass. PLoS One 6: e23980.

89. Provan J, Powell W, Ruff DC, investigators (1999) Chloroplast microsatellite variation in upland and lowland switchgrass. PLoS One 6: e23980.