

RESEARCH ARTICLE

Open Access



# Directed Repeats Co-occur with Few Short-Dispersed Repeats in Plastid Genome of a Spikemoss, *Selaginella vardei* (Selaginellaceae, Lycopodiopsida)

Hong-Rui Zhang<sup>1,2</sup>, Xian-Chun Zhang<sup>1</sup> and Qiao-Ping Xiang<sup>1\*</sup>

## Abstract

**Background:** It is hypothesized that the highly conserved inverted repeats (IR) structure of land plant plastid genomes (plastomes) is beneficial for stabilizing plastome organization, whereas the mechanism of the occurrence and stability maintenance of the recently reported direct repeats (DR) structure is yet awaiting further exploration. Here we describe the DR structure of the *Selaginella vardei* (Selaginellaceae) plastome, to elucidate the mechanism of DR occurrence and stability maintenance.

**Results:** The plastome of *S. vardei* is 121,254 bp in length and encodes 76 genes, of which 62 encode proteins, 10 encode tRNAs, and four encode rRNAs. Unexpectedly, the two identical rRNA gene regions (13,893 bp) are arranged in a direct orientation (DR), rather than inverted. Comparing to the IR organization in *Isoetes flaccida* (Isoetaceae, Lycopodiopsida) plastome, a ca. 50-kb *trnN-trnF* inversion that spans one DR copy was found in the plastome of *S. vardei*, which might cause the orientation change. In addition, we find extremely rare short dispersed repeats (SDRs) in the plastomes of *S. vardei* and its closely related species *S. indica*.

**Conclusions:** We suggest that the ca. 50-kb inversion resulted in the DR structure, and the reduction in SDRs plays a key role in maintaining the stability of plastomes with DR structure by avoiding potential secondary recombination. We further confirmed the presence of homologous recombination between DR regions, which are able to generate subgenomes and form diverse multimers. Our study deepens the understanding of *Selaginella* plastomes and provides new insights into the diverse plastome structures in land plants.

**Keywords:** direct repeat, inversion, inverted repeat, plastome evolution, short dispersed repeats

## Background

Chloroplasts in land plants have generally conserved genome structure, due to the constantly high selective pressures of photosynthesis [1]. Most plastomes are characterized by a quadripartite structure, which comprises two copies of an inverted repeat (IR) separating the large (LSC) and small (SSC) single copy regions [2]. The size of land plant plastomes usually ranges from 108 to 165 kb, and they generally contain 110–130 distinct genes including about 30 transfer RNA (tRNA)

genes, four ribosomal RNA (rRNA) genes, and approximately 80 protein-coding genes involved in photosynthesis or other metabolic processes [2, 3]. The typical IR is usually 20–30 kb and the genes that form the core of the IR encode the ribosomal RNAs (23S, 16S, 5S, and 4.5S) [4]. While plastomes of most land plants possess the typical IR structure, several lineages of land plants only retain one copy of the IR, such as *Carnegiea gigantea* (Cactaceae) [5], *Erodium* (Geraniaceae) [6], and an IR-lacking clade of Fabaceae [7], as well as conifers [8], in which one IR has been either extremely shortened or completely lost.

The conserved IR structure across land plants is hypothesized to function in stabilizing the plastomes

\* Correspondence: [qpxiang@ibcas.ac.cn](mailto:qpxiang@ibcas.ac.cn)

<sup>1</sup>State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, the Chinese Academy of Sciences, Beijing 100093, China  
Full list of author information is available at the end of the article



against major sequence rearrangements [9]. Palmer and Thompson [10] showed that rearrangement events were extremely rare in genomes with an IR, but increased remarkably in frequency when one IR copy was absent. Their hypothesis is supported by some recent studies using next generation sequencing. For example, *Trifolium* [11] and *Erodium texanum* [12] lack one copy of the IR and have highly rearranged plastomes. However, plastomes of *Pelargonium* [13] and *Trachelium* [14] are also highly rearranged, despite the presence of the IR structure, so that the existence of the IR appears to be insufficient to stabilize plastome structure [15]. A related hypothesis suggests that the incidence of short dispersed repeats (SDRs) is in fact more correlated with plastome instability [15]. Organellar DNA sequences of around 30 bp in length that may occur in direct or inverted forms are usually referred to as SDRs, and constitute, for example, more than 20% of the plastome of *Chlamydomonas reinhardtii* (green alga) [16]. Repeats larger than 1 kb can frequently recombine intra- or intermolecularly, and homologous recombination also occurs sporadically between short repeats larger than 50 bp [17]. Extensive studies have shown that genomes with massive rearrangement events tend to contain a high frequency of SDRs (e.g., *Trifolium* [11]; *Trachelium* [14]; *Pelargonium* [13]), whereas genomes containing virtually no SDRs have the conserved organization (*Erodium* [12]; algae [18, 19]).

Variation in number and orientation of rRNA-encoding repeats has been shown to be much more variable in algae than in land plants [20]. One to five copies of rRNA-encoding repeat are tandemly arranged in *Euglena* (green alga) [21], whereas in *Porphyra purpurea* (red alga), the two copies of the rRNA-encoding region are arranged into a direct repeat (DR) [22]. The mechanism of creating and maintaining this diversity remains unknown.

The lycophyte species *Selaginella tamariscina* and *S. kraussiana* have been recently reported to possess plastomes with two copies of rRNA-encoding repeat arranged into DR [23, 24]. The DR structure was proposed to have originated from the canonical IR via a large-scale inversion [24]. Here we confirmed plastomes with DR structure in *Selaginella* subg. *Rupestrae sensu* Weststrand and Korall [25]. We discovered that the two copies of DR were probably caused by a ca. 50-kb *trnF-trnN* inversion containing one DR copy. We also found SDRs to be rare in lycophyte plastomes with the DR organization in comparison to those with an IR. Considering the generation of sub-genomes by the recombination between DR regions, we predict that the reduction of SDRs, co-occurring with the DR structure, plays an important role in maintaining the stability of plastome and surviving over long evolutionary history.

## Results

### The Unconventional Structure of the *S. vardei* Plastome

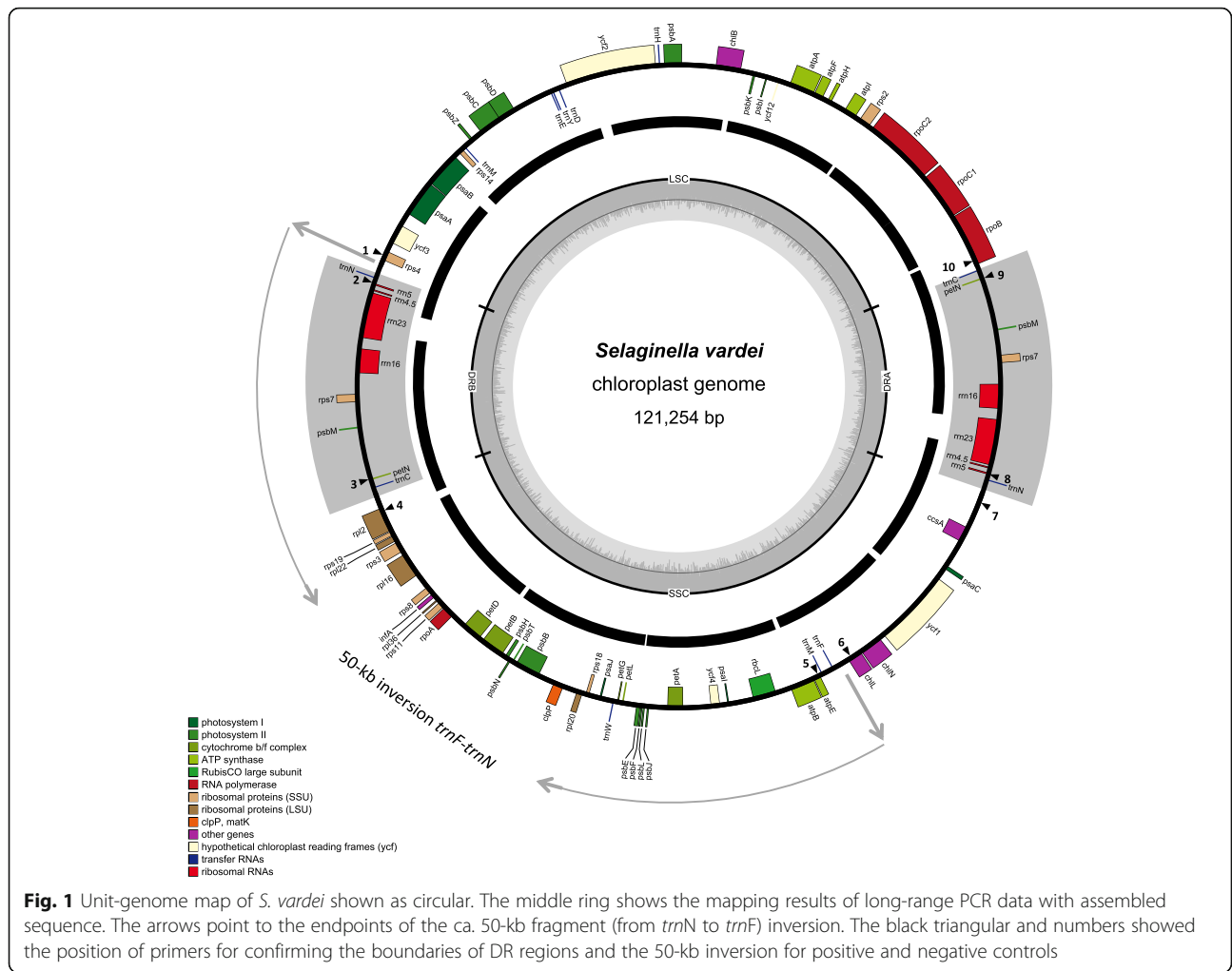
We sequenced the plastome of *S. vardei* (Fig. 1; MG272482) using Illumina HiSeq 2500 sequencing and assembled reads using de novo assembly methods. Coverage of the *S. vardei* plastome sequence is shown in Additional file 1: Figure S1c. The assembly result in Bandage 0.8.1 [26] was identical with the de novo methods, resulting in a circular plastome with the same organization (Additional file 1: Figure S1d). The plastome of *S. indica* (MK156801) was also sequenced and assembled; the organization and gene content are basically identical to that of *S. vardei*, therefore, we only described the detailed plastome characteristics of *S. vardei*.

The plastome of *S. vardei* is 121,254 bp long and contains 76 genes, including 62 protein-coding genes, four rRNA genes, and 10 tRNA genes (Additional file 6: Table S1). It is, therefore, much smaller than those of *S. uncinata* [27] and *S. moellendorffii* [28], apparently owing to gene losses of, for example, all the 11 *ndh* genes and two tRNA genes (*trnQ* and *trnR*). The plastome of *S. vardei* is similar to that of most land plants in having a quadripartite structure of two single-copy regions separated by two copies of a large rRNA-encoding repeat (13,893 bp). However, the two repeat copies were arranged into direct rather than inverted orientation, and the two single-copy regions are almost equal in size (47,676 bp and 45,792 bp, respectively).

### Confirmation of the Plastome Structure of *S. vardei*

The long PCR experiments across the whole plastome resulted in 12 products of the expected length (ca. 7–11 kb; Additional file 1: Figure S1a) using 12 pairs of primers (Additional file 7: Table S2) and the sequences were obtained using newly designed internal primers (Additional file 8: Table S3). We mapped the sequences of the long PCR products to the assembled plastome of *S. vardei* (Fig. 1). All the intergenic regions were covered by Sanger sequencing.

In addition, the amplification results of the DR structure and the 50-kb inversion, including six primer pairs for positive control and six for negative control, also supported the assembled plastome structure of *S. vardei* (Additional file 1: Figure S1b). The amplified fragments of positive control were consistent with the DR structure (1-2: *rps4-rrn5* and 3-4: *petN-rpl2*; 7-8: *ccsA-rrn5* and 9-10: *petN-rpoB*) and the ca. 50-kb inversion (1-2: *rps4-rrn5* and 5-6: *atpE-chlL*) in *S. vardei*. Accordingly, the negative control consistent with amplification of a canonical IR structure (1-3: *rps4-petN* and 2-4: *rrn5-rpl2*; 7-9: *ccsA-petN* and 8-10: *rrn5-rpoB*) and without the ca. 50-kb inversion structure (1-5: *rps4-atpE* and 2-6: *rrn5-chlL*) yielded no PCR product. We further checked the sequences at boundaries of the DR region and the ca.



50-kb inversion. The assembled sequences from NGS were identical with sequences from Sanger sequencing. The congruent results from Sanger sequencing with those from our assembly strongly support the DR structure and the ca. 50-kb inversion in the *S. vardei* plastome.

**PCR Confirmation of DR Structure in Representatives of subg. *Rupestreae***

The PCR experiments on one additional individual of *S. vardei*, two individuals of *S. indica*, and one individual of *S. dregei* (all from *Selaginella* subg. *Rupestreae*; Table 1) yielded the expected products with consistent length (Additional file 2: Figure S2a). PCR products were consistent with DR structure and the ca. 50-kb inversion. The aligned sequences of the four samples showed several indels (Additional file 2: Figure S2b). Sequences in *S. dregei* is much more divergent in the number of indels and nucleotide sites when compared with *S. vardei* and *S. indica*. For example, a 475 bp deletion occurred at the region of *ccsA-rrn5* in the *S. dregei* (collected from Kenya, Africa), which also exhibits

a smaller PCR product than the other species (Additional file 2: Figure S2b: 26698, 7-8).

**Gene Order of Plastomes in *S. vardei* and Other Lycophytes**

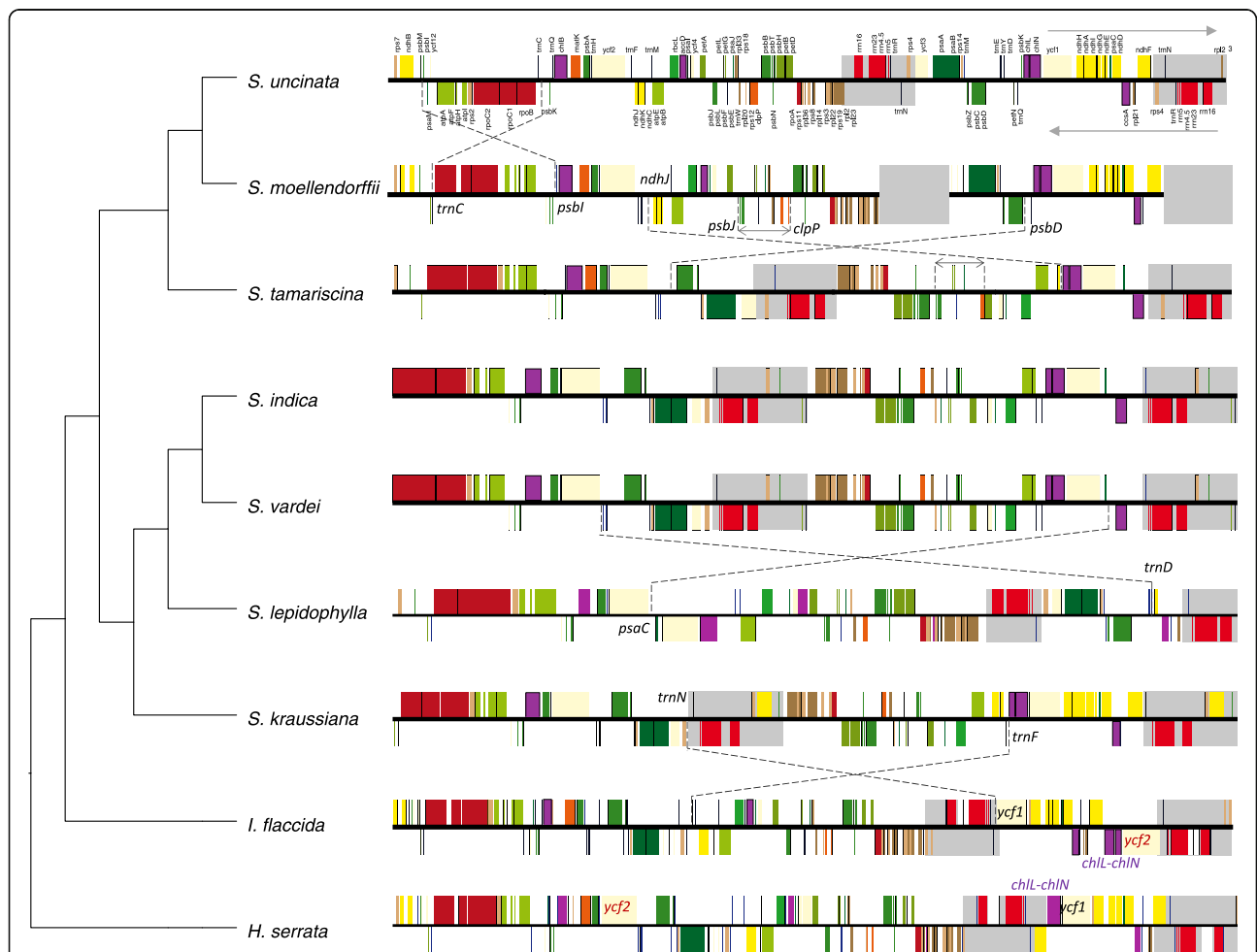
Dot plot analysis (Additional file 3: Figure S3) showed that plastomes of *Huperzia serrata* and *Isoetes flaccida* are basically syntenic, except the translocation of *ycf2* and inversion of *chlL-chlN* in *I. flaccida* (Additional file 3: Figure S3a), whereas, plastomes of *S. vardei* were quite divergent from both of them (Additional file 3: Figure S3b). A large

**Table 1** Representatives related for confirmation of plastome structure of *S. vardei*

Species	Voucher	Locality
<i>S. vardei</i>	Zhang X. C. 6948 (PE)	Sichuan, China
<i>S. vardei</i>	Zhang X. C. et al. 836 (PE)	Tibet, China
<i>S. indica</i>	Zhang X. C. 5868 (PE)	Sichuan Province, China
<i>S. indica</i>	Zhang X. C. 6255 (PE)	Yunnan Province, China
<i>S. dregei</i>	Liu B. 26698 (PE)	Kenya

inversion was present in *S. vardei* that encompasses a ca. 50-kb region from *trnN* to *trnF*, its endpoints lying between *rps4* and *trnN* at one end and between *trnF* and *chlL* at the other (Figs. 1 and 2). In *I. flaccida*, two tRNA genes, *trnL*-UAA and *trnT*-UGU were situated between *rps4* and *trnF* in LSC region, whereas *trnN* was at the border of the IRb/SSC adjacent to *ycf2*. In *S. vardei*, however, *trnL*-UAA and *trnT*-UGU were absent, *trnF* was located in the SSC, and *trnN* was located at the border of DRb/LSC next to *rps4*. Thus, we infer that the DR in *S. vardei* was caused by this ca. 50-kb inversion, which spans one copy of the DR (*trnN*-*trnC*). In addition, *S. vardei* has an expanded DR with genes originally located in LSC (*psbM* to *trnC*) being included, and *rpoB* was located at the start position of LSC region. Gene *ycf2* relocated to LSC, and *chlL*/*chlN* relocated to be adjacent to *ycf1* in *S. vardei*, which was consistent with the plastome of *H. serrata*.

Plastomes of *S. vardei*, *S. indica*, *S. kraussiana*, and *S. tamariscina* have DR structure and are basically syntenic in gene organization (Additional file 3: Figure S3c), whereas those of *S. lepidophylla*, *S. moellendorffii*, and *S. uncinata* possess canonical IR structure, but with rearrangement events among them (Fig. 2 and Additional file 3: Figure S3d, e, f). One remarkable difference between the plastomes of *S. uncinata* and *S. moellendorffii* is that *S. moellendorffii* lacks a ca. 20-kb inversion (from *trnC* to *psbI*), which exists in *S. uncinata* (Fig. 2 and Additional file 3: Figure S3f). This inversion is absent from all other published lycophyte plastomes, suggesting that the absence of this *trnC*-*psbI* inversion might be the ancestral state in lycophytes (Fig. 2). Two inversions (*ndh*-*psbD* region and *clpP*-*psbJ* region) exist in *S. moellendorffii* plastome in comparison with *S. tamariscina* (Additional file 3: Figure S3e). A 65-kb inversion (*trnD*-*psaC* region) exists in *S. lepidophylla* in comparison with *S. vardei* (Additional file 3: Figure S3d).



**Fig. 2** Linear maps of plastomes between *S. vardei* and other lycophytes showing the rearrangements. Upper arrow indicates genes transcribed in forward direction; lower arrow indicates genes transcribed in reverse direction. The red *ycf2* indicates the translocation from LSC to SSC and the purple *chlL*-*chlN* indicates the inversion in *I. flaccida*

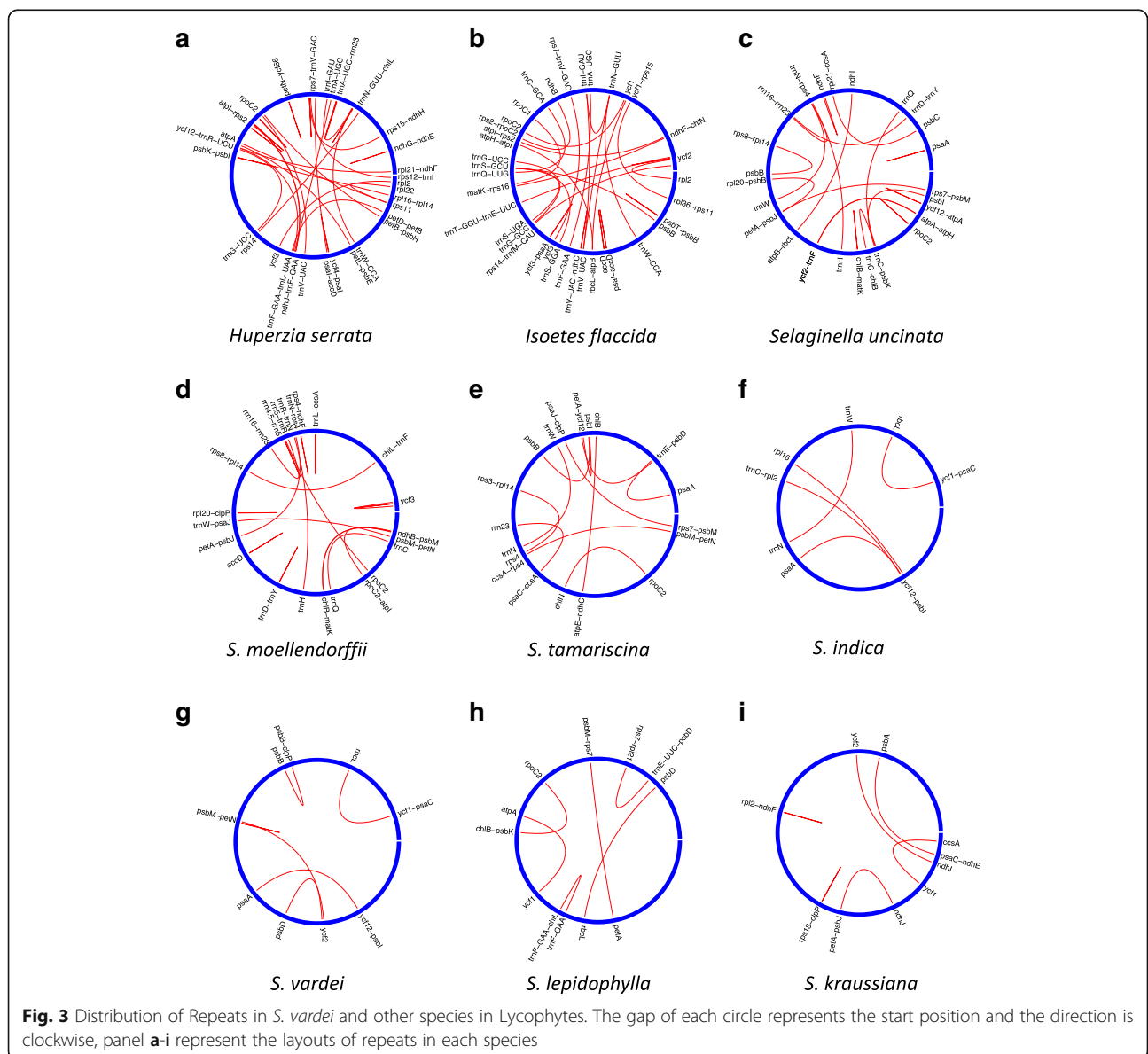
**Short Dispersed Repeats (SDRs) in Plastomes of *S. vardei* and Other Species**

Repeat analyses showed that fewer SDRs exist in the Selaginellaceae plastomes in comparison with those in other lycophytes (Isoetaceae and Lycopodiaceae; Fig. 3, Table 2 and Additional file 10: Table S5), and many fewer SDRs exist in Selaginellaceae with DR plastid structure in comparison with those with the IR structure. *Huperzia serrata* had the most SDRs (31; Fig. 3a), whereas *S. indica* contained the fewest (5; Fig. 3f). Furthermore, no SDRs were found at the endpoints of the inversions in plastomes with the DR structure, and all the SDRs was basically less than 50 bp, except one in *S. kraussiana*. In the plastome of *S. lepidophylla*, 16 copies of a 17 bp repeat unit dispersed in the intergenic region

of *trnF-chlL* (Additional file 10: Table S5-9). The 16 copies of repeat units can pair into a number of repeats, which are not displayed in the circular layouts (Fig. 3). In *S. uncinata*, a pair of SDR (46 bp) was located between the *psbI* and the *trnC-psbK* intergenic region, flanking the ca. 20-kb inversion from *psbI-trnC* (Fig. 3c and Additional file 10: Table S5-3). Another pair of 98 bp SDR located between *petA-psbJ/rpl20-psbB* in plastome of *S. uncinata* could also potentially cause inversions (Fig. 3c and Additional file 10: Tables S5-3).

**The Evolution of DR in Plastomes of Land Plants**

Phylogenetic relationships based on 32 plastid protein-coding genes (Fig. 4) showed that subg. *Rupestrae* containing *S. vardei* and *S. indica*, subg. *Lepidophyllae*



**Fig. 3** Distribution of Repeats in *S. vardei* and other species in Lycophytes. The gap of each circle represents the start position and the direction is clockwise, panel a-i represent the layouts of repeats in each species



**Table 2** Representatives related for confirmation of plastome structure of *S. vardei*

Species	No. SDRs	>=16 bp	>=30 bp	>=50 bp
<i>S. uncinata</i>	14	11	1	2
<i>S. moellendorffii</i>	18	14	3	1
<i>S. tamariscina</i>	11	11	0	0
<i>S. indica</i>	5	5	0	0
<i>S. vardei</i>	6	6	0	0
<i>S. lepidophylla</i>	6	6	0	0
<i>S. kraussiana</i>	6	4	1	1
<i>I. flaccida</i>	29	22	5	1
<i>H. serrata</i>	31	24	6	1

Note: species are listed following phylogenetic relationships

and subg. *Gymnogynum* formed into one clade and was sister to the remaining *Selaginella* species with sequenced plastomes (*S. tamariscina*, *S. moellendorffii* and *S. uncinata*), which is congruent with the recent published phylogenies of Selaginellaceae [29, 30] (Fig. 4). When we mapped the simplified structure of plastomes onto the phylogenetic tree, it is clear that most sequenced land plant plastomes possess the typical IR structure.

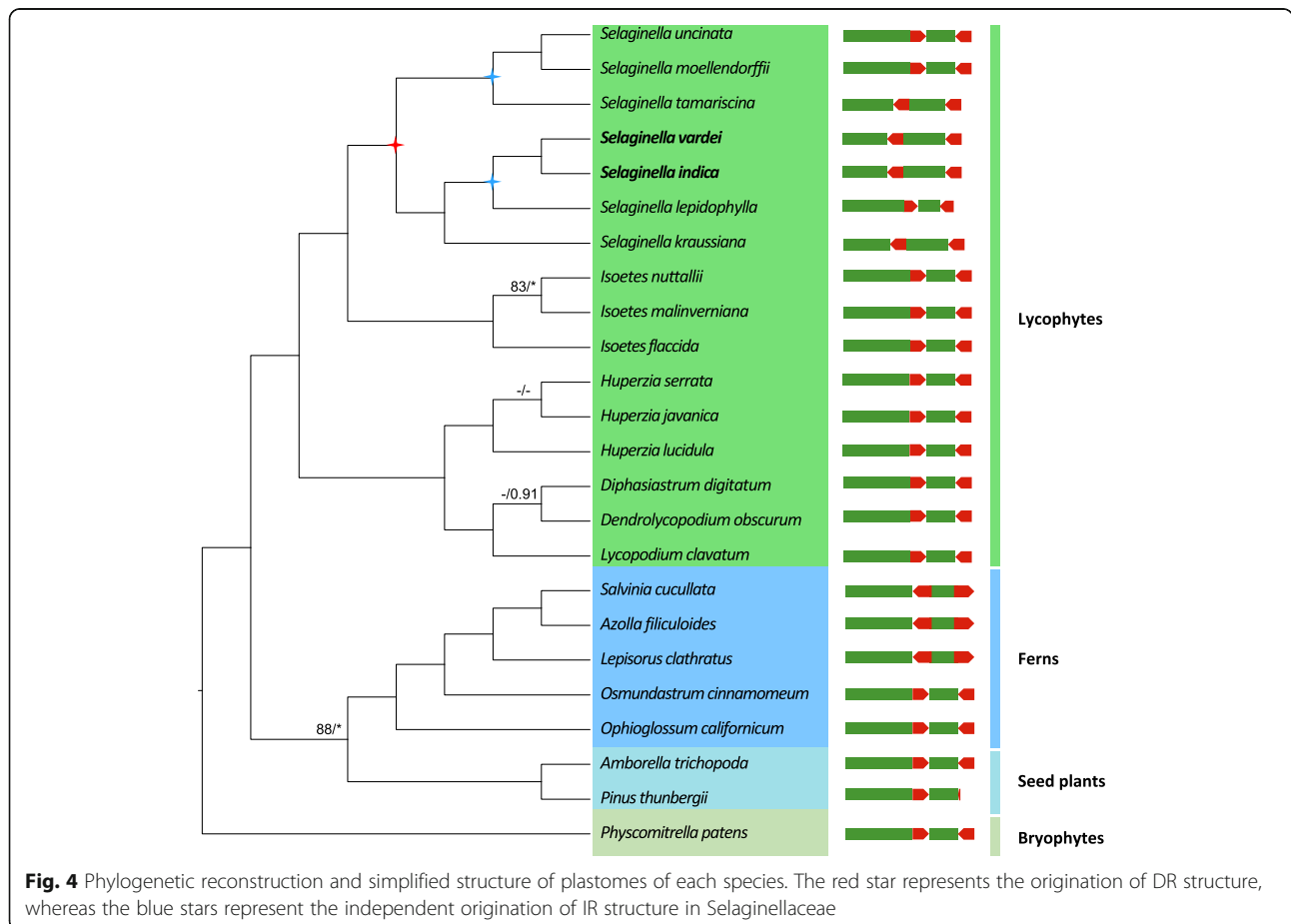
Compared to the plastomes of the early diverged ferns, such as Hymenophyllales, Marattiales, Psilotales, Ophioglossales, and Equisetales, the rRNA genes in the IR region were arranged in reverse order in Schizaeales and core leptosporangiates [31–33], with plastomes still maintaining the IR structure. To sum up, the plastomes with a DR structure only occurred in Selaginellaceae (Fig. 4).

**Discussion**

**The Characterization of the DR Structure in the Plastomes of Selaginellaceae**

The plastomes with DR structure, first reported in a red alga [22], originated independently in the phylogenetically distant Selaginellaceae. Considering the ancestral plastome in lycophytes (Lycopodiopsida; PPG 2016 [34]) likely has the canonical IR as shown in plastomes of Lycopodiaceae [35–37] and Isoetaceae [38], we propose that the innovative DR structure of plastomes in Selaginellaceae [23, 24] originated in the stem lineage of the family.

The DR structure in *S. vardei*, *S. indica*, and *S. dregei* (Table 1, Additional file 2: Figure S2), suggest that the DR structure is a shared characteristic in subg. *Rupestreae*. Following the infrageneric classification of Weststrand



and Korall [25], plastomes from subg. *Gymnogynum* (*S. kraussiana*) and subg. *Rupestrae* (*S. vardei* and *S. indica*) possess the DR structure, subg. *Lepidophyllae* (*S. lepidophylla*) possesses the IR structure, whereas subg. *Stachygynandrum* possesses both DR (*S. tamariscina*) and IR (*S. moellendorffii* and *S. uncinata*) structure. Therefore, two primary potential explanations exist for DR structure origin: 1) DR structure is the ancestral state of Selaginellaceae plastomes with at least two independent origins of IR structure (in *S. lepidophylla* and the clade including *S. moellendorffii* and *S. uncinata*); 2) IR structure is the ancestral state and at least three independent origin events are necessary for the three clades possessing DR structure (or two evolutions of DR and a subsequent re-evolution of IR). Since the gene organization in plastomes with the DR structure is basically congruent, and the organization of the plastomes with the IR structure is divergent (Fig. 2), we propose that the first hypothesis is more reasonable (Fig. 4).

The inversion of ca. 50-kb *trnN-trnF* region detected in *S. vardei* and other plastomes with a DR structure presumably resulted in the change to a DR structure. The ca. 50-kb inversion starts from the LSC and terminates at the border of DRb/SSC, spanning one DR copy, thus, results in the orientation change from inverted to direct (Fig. 1).

#### Conformation of the Plastome with DR Structure of Selaginellaceae

The standard depiction of the plastome is a genome-sized circular DNA molecule, and plastomes with IR structure are able to generate two potentially equimolar isomers [9, 39]. The two isomers differ only in the relative orientation of their single-copy regions, which was hypothesized to occur through interconversion within the two IR regions intramolecularly [9, 39]. However, it is now recognized that most plastomes display a great structural diversity existing as linear/concatemeric/highly branched complex molecules [40]. The orientation change of single-copy regions has been reinterpreted as the results of a BIR (break-induced replication)-like, recombination-dependent replication (RDR) mechanism among linear plastome templates [17, 41]. However, the recombinational activities of plastomes with DR structure have not been discussed. As proposed in mitogenomes, the existence of direct repeats could promote multipartite chromosomal architecture of master chromosomes and subgenomes [42]. Repeats larger than 1 kb appear to mediate high frequency, reciprocal recombination that can result in subgenomes of the genome in approximately equal stoichiometry (the quantities of DNA molecules in different form) [17], as shown in the mitogenome of *Ginkgo biloba* L. [43]. Thus, progress in understanding the structure of both mitogenomes and

plastomes suggest that both linear and circular forms of *Selaginella* plastomes exist in vivo, and recombination between the two copies of DR region could promote the generation of subgenomes following the RDR mechanism, with master chromosomes and subgenomes potentially occurring at similar stoichiometries. Two alternative read assemblies mapped at both ends of two copies of DR region reflect the existence of subgenomes in the plastome of *S. vardei* (Additional file 4: Figure S4). Furthermore, we screened the PacBio reads of the *S. tamariscina* plastome from GenBank and selected seven reads spanning the whole DR with both ends adjacent to genes from the LSC, and two reads spanning the DR with both ends adjacent to genes of the SSC (Additional file 5: Figure S5) confirming the recombination between DR regions and the existence of subgenomes in plastomes with DR structure. Either the master chromosome or subgenomic chromosomes could form head-to-tail multimers or branched complexes based on the RDR mechanism [40] (Additional file 4: Figure S4e).

#### The Co-occurrence of DR structure with reductions in SDRs in the Plastomes of Selaginellaceae

Remarkably, there are extremely few SDRs in Selaginellaceae plastomes with DR structure (Fig. 3). No SDRs are located at the ends of this ca. 50-kb inversion and the size of almost all the SDRs is less than 30 bp in plastomes with DR structure. Repeats less than 50 bp is only possibly able to invoke microhomology-mediated recombination, whereas as repeats larger than 50 bp is mostly able to mediate homologous recombination intermolecularly and intermolecularly in plastomes and mitogenomes [17]. A low frequency of illegitimate recombination can be presumably induced by the microhomologous repeats (less than 30 bp) only at the absence of plant-specific single-strand DNA (ssDNA)-binding protein the Whirlies, which is related to the recombination surveillance machinery [44]. In addition to the recombination of DR region, other repeats large enough to recombine efficiently often result in unintended secondary homologous recombination events [45]. The secondary recombination caused by direct or inverted short repeats could result in plastome fragmentation or gene loss, which may have the potential to destabilize the plastome of land plants [17]. Therefore, we hypothesize that plastomes with DR structure are possibly susceptible to SDRs scattered in single copy regions [43, 46], thus, the recombinational SDRs are selected against to maintain the integrity and stability of plastomes.

#### Conclusions

We documented the unconventional DR structure in plastomes of subg. *Rupestrae* and hypothesize that the

DR structure is the ancestral state of Selaginellaceae plastomes. We propose that the ca. 50-kb *trnN-trnF* inversion resulted in the DR structure, and the DR regions could mediate the recombination activity, generating approximately equimolar subgenomes. Subsequently, the extreme reduction in SDRs of lineages with the DR structure is presumably the result of selection to avoid the potential secondary recombination and plays a key role in maintaining the integrity and stability of the unconventional plastomes with DR structure.

## Materials and Methods

### Taxon Sampling, DNA Extraction, Sequencing and Assembly

*Selaginella vardei* is a member of the monophyletic subg. *Rupestrae sensu* Weststrand and Korall [25], characterized by having monomorphic and helically arranged vegetative leaves and tetrastichous strobili. We collected a sample of *S. vardei* from the wild in Sichuan Province and the closely related *S. indica* from Yunnan Province for this study and deposited the voucher specimens of this collection in the Herbarium of Institute of Botany, CAS (PE) (Table 1).

Total genomic DNA was isolated from silica gel-dried materials with a modified cetyltrimethylammonium bromide (CTAB) method [47]. Library construction was performed with the NEBNext DNA Library Prep Kit (New England Biolabs, Ipswich, Massachusetts, USA) and sequencing was performed on the Illumina HiSeq 2500 (Illumina, San Diego, California, USA). Illumina paired-end reads of 150 bp were mapped to *S. uncinata* (AB197035) [27] and *S. moellendorffii* (FJ755183) [28], with medium-low sensitivity in five to ten iterations in Geneious 9.1.4 (Biomatters, Inc., Auckland, New Zealand; <https://www.geneious.com>) [48]. The mapped reads were then assembled into contigs in Geneious. Additionally, the cleaned reads were assembled de novo with SPAdes v. 3.10.1 [49] using a range of kmer sizes from 21 to 99. Putative plastome contigs were identified using BLASTN 2.2.29 [50], with the previously published *S. uncinata* and *S. moellendorffii* plastomes as reference. We also used bandage v. 0.8.1 [26], a program for visualizing de novo assembly graphs, to help select plastome contigs and analyze de novo assembly results by importing the fastg file created by SPAdes. The contigs obtained above were then combined and imported into Geneious to extend and assemble into the complete plastomes.

### Gene Annotation

Gene annotation was performed using local BLAST with default parameter settings [51]. Putative start and stop codons were defined based on similarity with genes of published plastomes [27, 28]. The tRNAs were verified

using tRNAscan-SE version 1.21 [52] and ARAGORN [53]. Circular and linear genome maps were drawn with OGDRAW version 1.2 [54].

### PCR Confirmation of *S. vardei* Plastome

To confirm the accuracy of our *S. vardei* plastome assembly, we designed 12 primer pairs (Additional file 7: Table S2) using Primer v 3.0 [55] based on the assembled sequence of *S. vardei* for long-range PCR. We also designed five additional primer pairs (Additional file 9: Table S4) at the boundaries of the DR structure and the ca. 50-kb inversion (marked on Fig. 1) to confirm the accuracy of assembly. Both positive (50-kb inversion and DR structure) and negative (without 50-kb inversion and DR structure) controls were carried out using these five primer pairs (Additional file 1: Figure S1b). The PCR amplifications were performed in a total volume of 20  $\mu$ L containing 4  $\mu$ L of 5X PrimeSTAR GXL Buffer, 1.6  $\mu$ L of dNTP Mixture (2.5 mM each), 1.2  $\mu$ L of each primer (5 mM), 0.4  $\mu$ L of PrimeSTAR GXL DNA Polymerase and 20 ng of template DNA. Cycling conditions were 98 °C for 3 min, followed by 40 cycles of 98 °C for 10 s, 58 °C for 30 s and 72 °C for 5 min for long-range PCR and 1.5 min for normal PCR, and a final extension of 72 °C for 10 min. The PCR products were verified by electrophoresis in 0.8% agarose gels stained with ethidium bromide. Then, we designed internal primers (Additional file 8: Table S3) to get sequences of these long-range PCR products using Sanger sequencing. The PCR products of the DR and inversion confirmation were sequenced by Majorbio, Beijing, China.

### PCR Confirmation of DR Structure in Related Representatives of subg. *Rupestrae*

To further test whether the structure found in *S. vardei* existed in other species of subg. *Rupestrae*, PCR amplification using primers designed at the boundaries of the DR structure and the ca. 50-kb inversion in plastome of *S. vardei* (marked on Fig. 1, Additional file 9: Table S4) were carried out with another individual of *S. vardei*, two individuals of *S. indica*, and one individual of *S. dregei* (Table 1). Only positive control was carried out in these species. The PCR procedure follows the conditions of normal PCR mentioned above. The PCR products were verified by electrophoresis in 0.8% agarose gels stained with ethidium bromide. The PCR products were sequenced by Majorbio, Beijing, China.

### Comparison of the Plastomes of *S. vardei* and Other Lycophytes

Dot plot analyses of the plastomes of *S. vardei* and other lycophytes (*S. lepidophylla*, *S. kraussiana*, *S. tamariscina*, *S. uncinata*, *S. moellendorffii*, *Isoetes flaccida*, and *Huperzia serrata*) were performed using Gepard [56] in



order to identify the putative structural rearrangements in the *S. vardei* plastome. The syntenic analyses of linear plastome maps of *S. vardei* and other lycophytes were carried out based on the dot plot analyses. The first site of the LSC at the border of the LSC/IRa was considered as the starting point.

### Repeat Analyses

Short dispersed repeats (SDRs) were identified using RepeatsFinder [57] with default parameters. The circular layout of SDRs in plastomes was then visualized using the *circlize* package [58] in R v. 3.4.1 (R Development Core Team 2012). The locations of SDRs were marked outside the circle in order to find possible correlations with rearrangements. Nine plastomes (*S. lepidophylla*, *S. vardei*, *S. indica*, *S. kraussiana*, *S. tamariscina*, *S. uncinata*, *S. moellendorffii*, *I. flaccida* and *H. serrata*) were included. One copy of the IR/DR was removed from all plastomes used.

### Phylogenetic Analyses

Thirty-two conserved plastid protein-coding genes were used to reconstruct the phylogenetic framework using 19 species from previously published plastomes of land plants (from bryophytes to seed plants; Additional file 11: Table S6). For *S. tamariscina*, we downloaded the raw reads from GenBank (SRR6228814, SRR7135413) [23], assembled plastid contigs and extracted the 32 gene sequences since the complete plastome has not been released on GenBank. A total of 19,248 bp sequences were aligned at the protein level by MAFFT [59] using the translation-aligned function in Geneious v. 9.1.4. Poorly aligned regions were removed using Gblocks v. 0.91b [60]. Maximum-likelihood (ML) analysis was performed using RAxML v. 7.4.2 with 1000 bootstrap replicates and the GTR+G model [61] based on Akaike information criterion (AIC) in jModeltest 2.1.7 [62]. The simplified structure of the plastome of each species (Additional file 11: Table S6) was then mapped on the phylogenetic tree showing the direction of rRNA-encoding repeat.

### Additional files

**Additional file 1: Figure S1.** Confirmation of *S. vardei* plastome structure. **a:** Long-range PCR amplification results of *S. vardei*. Primer pairs used are indicated at the top of each lane. Size markers are in bp. Gene names of these 12 products are as follows: 1: *rpoB* – *rps2*; 2: *rps2* – *chlB*; 3: *chlB* – *ycf2*; 4: *ycf2* – *psaB*; 5: *psaB* – *rrn23*; 6: *rrn23* – *rpl2*; 7: *rpl2* – *petB*; 8: *petB* – *petE*; 9: *petE* – *atpB*; 10: *atpB* – *ycf1*; 11: *ycf1* – *rrn23*; 12: *rrn23* – *rpoB*. **b:** PCR confirmation results of DR structure and Inversion. Primer pairs used are indicated at the top of each lane. Size markers are in bp. Gene names of these 12 fragments are as follows: positive control: 1-2, *rps4* – *rrn5*; 3-4, *petN* – *rpl2*; 1-2, *rps4* – *rrn5*; 5-6, *atpE* – *chlL*; 7-8, *ccsA* – *rrn5*; 9-10, *petN* – *rpoB*; negative control: 1-3, *rps4* – *petN*; 2-4, *rrn5* – *rpl2*; 1-5, *rps4* – *atpE*; 2-6, *rrn5* – *chlL*; 7-9, *ccsA* – *petN*; 8-10, *rrn5* – *rpoB*; **c:** Reads coverage of *S. vardei* plastomes; **d:** The assembled graph in Bandage showing DR structure of *S. vardei*. (PDF 175 kb)

**Additional file 2: Figure S2.** Plastome structure confirmation and sequence alignments of related representatives. **a:** PCR confirmation results of DR structure in another individual of *S. vardei* and two representative species within subg. *Rupestrae*. Primer pairs used are indicated at the top of each lane. Size markers are in bp. Species name and gene names of these fragments are as follows: 836: *S. vardei*; 5868: *S. indica*; 6255: *S. indica*; 26698: *S. dregei*; 1-2: *rps4* – *rrn5*; 3-4: *petN* – *rpl2*; 5-6: *atpE* – *chlL*; 7-8: *ccsA* – *rrn5*; 9-10: *petN* – *rpoB*; **b:** Alignment results of sequences from each product in **a**, the light grey color represents regions with identical base pairs among individuals, whereas the dark color highlights regions with mismatched base pairs. (PDF 115 kb)

**Additional file 3: Figure S3.** Dot-plot analyses of plastomes between *S. vardei* and other lycophyte species. (PDF 489 kb)

**Additional file 4: Figure S4.** Reads coverages of putative subgenomes and alternative reads assemblies in *S. vardei*. **a, b:** Two alternative reads assemblies at LSC/DRa and SSC/DRb boundaries: the unmatched reads in black box in **a** is consistent with assembled sequence in black box in **b**, the unmatched reads in red box in **b** is consistent with assembled sequence in red box in **a**; **c, d:** Two alternative reads assemblies at DRa/SSC and DRb/LSC boundaries: the unmatched reads in black box in **c** is consistent with assembled sequence in black box in **d**, the unmatched reads in red box in **d** is consistent with assembled sequence in red box in **c**; **e:** the simplified structure for master chromosomes and subgenomes based on **a, b, c, d** in this figure. We define the arrow end as B, and the other end as A. End B of either LSC (**a**) or SSC (**b**) can be assembled with end B of DR. End A of either LSC (**c**) or SSC (**d**) can be assembled with end A of DR. (PDF 1549 kb)

**Additional file 5: Figure S5.** Screened PacBio reads of *S. tamariscina* showing evidence of the existence of subgenomes in plastomes with DR. **a:** the simplified plastome structure of *S. tamariscina* based on Xu et al. (2018); **b, c:** the simplified subgenome structure of plastomes with DR, supported by the screened PacBio reads of *S. tamariscina* plastome as listed. (PDF 29 kb)

**Additional file 6: Table S1.** Genes present in the plastome of *S. vardei*. (DOCX 15 kb)

**Additional file 7: Table S2.** Primers designed for long range PCR amplification of plastome of *S. vardei*. (DOCX 14 kb)

**Additional file 8: Table S3.** Internal primers designed for Sanger sequencing of long-range PCR products of *S. vardei*. (DOCX 18 kb)

**Additional file 9: Table S4.** Primers designed for Sanger sequencing of PCR confirmation of *Selaginella* subg. *Rupestrae*. (DOCX 14 kb)

**Additional file 10: Table S5.** Detailed dispersed short repeats in plastomes of each species. (DOCX 45 kb)

**Additional file 11: Table S6.** Species selected in the phylogenetic analyses. (DOCX 15 kb)

### Abbreviations

AIC: Akaike information criterion; BIR: Break-induced replication; CTAB: Cetyltrimethylammonium bromide; DR: Direct repeat; IR: Inverted repeat; LSC: Large single copy; ML: Maximum likelihood; NGS: Next-generation sequencing; RDR: Recombination-dependent replication; rRNA: Ribosomal RNA; SDRs: Short dispersed repeats; SSC: Small single copy; ssDNA: single-strand DNA; tRNA: Transfer RNA

### Acknowledgments

We are grateful to Shi-Liang Zhou, Ran Wei, Shan-Shan Dong, Wen-Pan Dong, and Jong-Soo Kang for helpful discussions. We thank Bing Liu for providing materials, Changhao Li and Lian-Sheng Xu for help with analyses, and Li Wang, Zhi-Qiang Wu, and AJ Harris for helpful revision and polishing of the whole manuscript. We also thank two anonymous reviewers for insightful comments.

### Authors' contributions

XCZ collected samples from the field, QPX and XCZ planned and designed the research, HRZ performed experiments and analyzed data, HRZ and QPX wrote the manuscript. All authors discussed the results and commented on the manuscript. All authors read and approved the final manuscript.

### Funding

This study was supported by the National Natural Science Foundation of China (NSFC, No. 31670205 and No. 31770237). The National Natural Science Foundation of China (NSFC, No. 31670205) supported us for materials collection from fieldwork; the National Natural Science Foundation of China (NSFC, No. 31770237) supported us with design of the research, experiments, data sequencing and data analyses, results interpretation.

### Availability of data and materials

The newly sequenced plastome sequences have been deposited in GenBank under the accession number MG272482 and MK156801.

### Ethics approval and consent to participate

Not applicable.

### Competing interests

The authors declare no competing financial interests.

### Author details

<sup>1</sup>State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, the Chinese Academy of Sciences, Beijing 100093, China. <sup>2</sup>University of Chinese Academy of Sciences, Beijing 100049, China.

Received: 16 November 2018 Accepted: 24 May 2019

Published online: 11 June 2019

### References

- Mower JP, Vickrey TL. Structural diversity among plastid genomes of land plants. In: Chaw S-M, Jansen RK, editors. *Plastid Genome Evolution*, *Advances in Botanical Research*, vol. 85. Cambridge: Academic Press; 2018. p. 263–92.
- Raubeson LA, Jansen RK. Chloroplast genomes of plants. In: Henry RJ, editor. *Plant diversity and evolution: genotypic and phenotypic variation in higher plants*. London: CAB International; 2005. p. 45–68.
- Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol Biol*. 2011;76(3):273–97.
- Simpson CL, Stern DB. The treasure trove of algal chloroplast genomes. Surprises in architecture and gene content, and their functional implications. *Plant Physiol*. 2002;129(3):957–66.
- Sanderson MJ, Copetti D, Búrquez A, Bustamante E, Charboneau JL, Eguiarte LE, Kumar S, Lee HO, Lee J, McMahon M. Exceptional reduction of the plastid genome of saguaro cactus (*Carnegiea gigantea*): Loss of the *ndh* gene suite and inverted repeat. *Am J Bot*. 2015;102(7):1115–27.
- Blazier JC, Guisinger MM, Jansen RK. Recent loss of plastid-encoded *ndh* genes within *Erodium* (Geraniaceae). *Plant Mol Biol*. 2011;76(3-5):263–72.
- Jansen RK, Wojciechowski MF, Sanniyasi E, Lee S-B, Daniell H. Complete plastid genome sequence of the chickpea (*Cicer arietinum*) and the phylogenetic distribution of *rps12* and *clpP* intron losses among legumes (Leguminosae). *Mol Phylogeny Evol*. 2008;48(3):1204–17.
- Wu C-S, Wang Y-N, Hsu C-Y, Lin C-P, Chaw S-M. Loss of different inverted repeat copies from the chloroplast genomes of Pinaceae and Cupressophytes and influence of heterotachy on the evaluation of gymnosperm phylogeny. *Gen Biol Evol*. 2011;3:1284–95.
- Palmer JD. Chloroplast DNA exists in two orientations. *Nature*. 1983;301(5895):92–3.
- Palmer JD, Thompson WF. Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. *Cell*. 1982;29(2):537–50.
- Cai Z, Guisinger M, Kim H-G, Ruck E, Blazier JC, McMurtry V, Kuehl JV, Boore J, Jansen RK. Extensive reorganization of the plastid genome of *Trifolium subterraneum* (Fabaceae) is associated with numerous repeated sequences and novel DNA insertions. *J Mol Evol*. 2008;67(6):696–704.
- Guisinger MM, Kuehl JV, Boore JL, Jansen RK. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. *Mol Biol Evol*. 2011;28(1):583–600.
- Chumley TW, Palmer JD, Mower JP, Fourcade HM, Calie PJ, Boore JL, Jansen RK. The complete chloroplast genome sequence of *Pelargonium x hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. *Mol Biol Evol*. 2006;23(11):2175–90.
- Haberle RC, Fourcade HM, Boore JL, Jansen RK. Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes. *J Mol Evol*. 2008;66(4):350–61.
- Blazier JC, Jansen RK, Mower JP, Govindu M, Zhang J, Weng M-L, Ruhlman TA. Variable presence of the inverted repeat and plastome stability in *Erodium*. *Ann Bot*. 2016;117(7):1209–20.
- Maul JE, Lilly JW, Cui L, Miller W, Harris EH, Stern DB. The *Chlamydomonas reinhardtii* plastid chromosome islands of genes in a sea of repeats. *Plant Cell*. 2002;14(11):2659–79.
- Maréchal A, Brisson N. Recombination and the maintenance of plant organelle genome stability. *New Phytol*. 2010;186(2):299–317.
- Pombert J-F, Otis C, Lemieux C, Turmel M. The chloroplast genome sequence of the green alga *Pseudoclonium akinetum* (Ulvophyceae) reveals unusual structural features and new insights into the branching order of chlorophyte lineages. *Mol Biol Evol*. 2005;22(9):1903–18.
- Pombert J-F, Lemieux C, Turmel M. The complete chloroplast DNA sequence of the green alga *Oltmannsiellopsis viridis* reveals a distinctive quadripartite architecture in the chloroplast genome of early diverging ulvophytes. *BMC Biol*. 2006;4(1):3.
- Reith M. Molecular biology of rhodophyte and chromophyte plastids. *Annu Rev Plant Biol*. 1995;46(1):549–75.
- Hallick RB, Hong L, Drager RG, Favreau MR, Monfort A, Orsat B, Spielmann A, Stutz E. Complete sequence of *Euglena gracilis* chloroplast DNA. *Nucleic Acids Res*. 1993;21(15):3537–44.
- Reith M, Munholland J. The ribosomal RNA repeats are non-identical and directly oriented in the chloroplast genome of the red alga *Porphyra purpurea*. *Curr Genet*. 1993;24(5):443–50.
- Xu Z, Xin T, Bartels D, Li Y, Gu W, Yao H, Liu S, Yu H, Pu X, Zhou J, et al. Genome analysis of the ancient tracheophyte *Selaginella tamariscina* reveals evolutionary features relevant to the acquisition of desiccation tolerance. *Mol Plant*. 2018;11:983–94.
- Mower JP, Ma P-F, Grewe F, Taylor A, Michael TP, VanBuren R, Qiu Y-L. Lycophyte plastid genomics: extreme variation in GC, gene and intron content and multiple inversions between a direct and inverted orientation of the rRNA repeat. *New Phytol*. 2019;222(2):1061–75.
- Weststrand S, Korall P. A subgeneric classification of *Selaginella* (Selaginellaceae). *Am J Bot*. 2016;103(12):2160–9.
- Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics*. 2015;31(20):3350–2.
- Tsuji S, Ueda K, Nishiyama T, Hasebe M, Yoshikawa S, Konagaya A, Nishiuchi T, Yamaguchi K. The chloroplast genome from a lycophyte (microphylophyte), *Selaginella uncinata*, has a unique inversion, transpositions and many gene losses. *J Plant Res*. 2007;120(2):281–90.
- Smith DR. Unparalleled GC content in the plastid DNA of *Selaginella*. *Plant Mol Biol*. 2009;71(6):627–39.
- Weststrand S, Korall P. Phylogeny of Selaginellaceae: There is value in morphology after all. *Am J Bot*. 2016;103(12):2136–59.
- Zhou X-M, Rothfels CJ, Zhang L, He Z-R, Le Péchon T, He H, Lu NT, Knapp R, Lorence D, He X-J, et al. A large-scale phylogeny of the lycophyte genus *Selaginella* (Selaginellaceae: Lycopodiopsida) based on plastid and nuclear loci. *Cladistics*. 2016;32(4):360–89.
- Grewe F, Guo W, Gubbels EA, Hansen AK, Mower JP. Complete plastid genomes from *Ophioglossum californicum*, *Ptilotum nudum*, and *Equisetum hyemale* reveal an ancestral land plant genome structure and resolve the position of Equisetales among monilophytes. *BMC Evol Biol*. 2013;13(1):1–16.
- Kuo L-Y, Qi X, Ma H, Li F-W. Order-level fern plastome phylogenomics: new insights from Hymenophyllales. *Am J Bot*. 2018;105(9):1–11.
- Li F-W, Kuo L-Y, Pryer KM, Rothfels CJ. Genes translocated into the plastid inverted repeat show decelerated substitution rates and elevated GC content. *Gen Biol Evol*. 2016;8(8):2452–8.
- Ppg I. A community-derived classification for extant lycophytes and ferns. *J Syst Evol*. 2016;54(6):563–603.
- Guo Z-Y, Zhang H-R, Shrestha N, Zhang X-C. Complete chloroplast genome of a valuable medicinal plant, *Huperzia serrata* (Lycopodiaceae), and comparison with its congener. *Appl Plant Sci*. 2016;4(11):1600071.
- Zhang H-R, Kang J-S, Viane RL, Zhang X-C. The complete chloroplast genome sequence of *Huperzia javanica* (Sw.) CY Yang in Lycopodiaceae. *Mitochondrial DNA Part B*. 2017;2(1):216–8.

37. Wolf PG, Karol KG, Mandoli DF, Kuehl J, Arumuganathan K, Ellis MW, Mishler BD, Kelch DG, Olmstead RG, Boore JL. The first complete chloroplast genome sequence of a lycophyte, *Huperzia lucidula* (Lycopodiaceae). *Gene*. 2005;350(2):117–28.
38. Karol KG, Arumuganathan K, Boore JL, Duffy AM, Everett KD, Hall JD, Hansen SK, Kuehl JV, Mandoli DF, Mishler BD, et al. Complete plastome sequences of *Equisetum arvense* and *Isoetes flaccida*: implications for phylogeny and plastid genome evolution of early land plant lineages. *BMC Evol Biol*. 2010;10(1):1–16.
39. Kolodner R, Tewari K. Inverted repeats in chloroplast DNA from higher plants. *Proc Natl Acad Sci*. 1979;76(1):41–5.
40. Oldenburg DJ, Bendich AJ. Most chloroplast DNA of maize seedlings in linear molecules with defined ends and branched forms. *J Mol Biol*. 2004;335(4):953–70.
41. Ruhlman TA, Zhang J, Blazier JC, Sabir JS, Jansen RK. Recombination-dependent replication and gene conversion homogenize repeat sequences and diversify plastid genome structure. *Am J Bot*. 2017;104(4):559–72.
42. Smith DR, Keeling PJ. Mitochondrial and plastid genome architecture: reoccurring themes, but significant differences at the extremes. *Proc Natl Acad Sci*. 2015;112(33):10177–84.
43. Guo W, Grewe F, Fan W, Young GJ, Knoop V, Palmer JD, Mower JP. *Ginkgo* and *Welwitschia* mitogenomes reveal extreme contrasts in gymnosperm mitochondrial evolution. *Mol Biol Evol*. 2016;33(6):1448–60.
44. Maréchal A, Parent J-S, Véronneau-Lafortune F, Joyeux A, Lang BF, Brisson N. Whirly proteins maintain plastid genome stability in *Arabidopsis*. *Proc Natl Acad Sci U S A*. 2009;106(34):14693–8.
45. Gray BN, Ahner B, Hanson M. Extensive homologous recombination between introduced and native regulatory plastid DNA elements in transplastomic plants. *Transgenic Res*. 2009;18(4):559–72.
46. Fauron C, Casper M, Gao Y, Moore B. The maize mitochondrial genome: dynamic, yet functional. *Trends Genet*. 1995;11(6):228–35.
47. Li J, Wang S, Yu J, Wang L, Zhou S. A modified CTAB protocol for plant DNA extraction. *Bulletin of Botany*. 2013;48(1):72–8.
48. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012;28(12):1647–9.
49. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012;19(5):455–77.
50. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009;10(1):421.
51. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215(3):403–10.
52. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res*. 1997;25(5):955–64.
53. Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res*. 2004;32(1):11–6.
54. Lohse M, Drechsel O, Bock R. OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr Genet*. 2007;52(5-6):267–74.
55. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. Primer3—new capabilities and interfaces. *Nucleic Acids Res*. 2012;40(15):e115.
56. Krumsiek J, Arnold R, Rattei T. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics*. 2007;23(8):1026–8.
57. Volfovsky N, Haas BJ, Salzberg SL. A clustering method for repeat analysis in DNA sequences. *Genome Biol*. 2001;2(8):1–11.
58. Gu Z, Gu L, Eils R, Schlesner M, Brors B. Circize implements and enhances circular visualization in R. *Bioinformatics*. 2014;30(19):2811–2.
59. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772–80.
60. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol*. 2000;17(4):540–52.
61. Stamatakis A. RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 2006;22(21):2688–90.
62. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods*. 2012;9(8):772.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://www.biomedcentral.com/submissions)

