


ORIGINAL ARTICLE

Open Access



Complete chloroplast genome sequences of two endangered *Phoebe* (Lauraceae) species

Yingang Li¹, Wuqin Xu², Wentao Zou³, Dongyue Jiang¹ and Xinhong Liu^{1*} 

Abstract

Background: *Phoebe* (Lauraceae) comprises of evergreen trees or shrubs with approximately 100 species, distributed in tropical and subtropical Asia and Neotropical America. A total of 34 species and three varieties occur in China. Despite of economic and ecological value, only limited genomic resources are available for this genus.

Results: We sequenced the two complete chloroplast (cp) genomes of *Phoebe chekiangensis* and *P. bournei* using Illumina sequencing technology via a combined strategy of de novo and reference-guided assembly. We also performed comparative analyses with the cp genomes of *P. sheareri* and *P. sheareri* var. *oineiensis* previously reported. The chloroplast genomes of *P. chekiangensis* and *P. bournei* identically contain 112 genes consisting of 78 protein coding genes, 30 tRNA genes, and 4 rRNA genes, with the size of 152,849 and 152,853 bp, respectively. From the two chloroplast genomes, 131 SSRs were identified and 12 different SSRs located in five protein coding genes. The analysis showed the extremely conserved structure of chloroplast genomes with surprisingly little variations at the LSC/IR and SSC/IR boundaries. Moreover, the mean nucleotide diversity was found to be 0.162% for 77 regions, suggesting an extraordinarily low level of sequence divergence. Four highest divergent regions (*trnH-psbA*, *rps14-trnT*, *petA-psbJ*, *ccsA-ndhD*) with the percentage of nucleotide diversity higher than 0.50% were identified, which had potential use for species identification and phylogenetic studies.

Conclusion: This study will facilitate our understanding of population genetics, phylogenetic relationship and plant evolution of *Phoebe* species.

Keywords: *Phoebe chekiangensis*, *Phoebe bournei*, Chloroplast genomes, Repeat analysis, SSRs, Divergent regions

Background

Phoebe is a genus of evergreen trees or shrubs belonging to family, Lauraceae. *Phoebe* comprises of approximately 100 species, distributed in tropical and subtropical Asia and neotropical America. A total of 34 species and three varieties are endemic to China (Wu et al. 2008). *Phoebe* species, with their high-quality wood, were widely used to make column during palace construction in the Ming and Qing dynasties and built high valuable furniture which stood for the power and status of the noble (Ding et al. 2015). The most famous and valuable *Phoebe* wood

are called ‘wood with golden wire’, which comes from several specific and sporadically distributed rare *Phoebe* species, endemics to China, including *Phoebe chekiangensis*, *Phoebe bournei*, *Phoebe sheareri*, *Phoebe zhennan*, and *Phoebe lichuanensis*. Our target species, *P. chekiangensis* is distributed in Zhejiang and its adjacent areas (including Fujian, Jiangxi, and Anhui Province), whereas *P. bournei* is distributed in Yangtze River Basin and its south region in China. The *Phoebe* species have a very high economic and ecological value but with extremely limited studies on biochemical compound and population diversity (Hegde et al. 1997; Zhang et al. 2012a, b).

The chloroplast (cp) is an important organelle that plays a key role in plant photosynthesis providing energy to green plants and carbon fixation (Douglas 1990). With

*Correspondence: lkyluxh@163.com

¹ Zhejiang Academy of Forestry, Hangzhou 310023, China

Full list of author information is available at the end of the article

the rapid development of next-generation sequencing, it is now cheaper and faster to obtain genomes than by traditional Sanger sequencing. Therefore, cp genome-scale data have been increasingly used to infer phylogenetic relationships at high taxonomical levels, and even in lower taxa (Moore et al. 2007; Parks et al. 2009; Huang et al. 2014; Yang et al. 2017; Zeng et al. 2017). Although the cp genome is more conserved than the nuclear genome in plants, many mutation events in the chloroplast DNA sequence have been identified, including indels, substitutions, and inversions (Ingvarsson et al. 2003). Most angiosperm cp genomes have a quadripartite circular structure ranging from 115 to 165 kbp in length, and are composed of two copies of inverted repeat (IR) regions that are separated by a large single copy (LSC) region and a small single copy (SSC) region (Wicke et al. 2011; Wang et al. 2015). Massive information for genetics, taxonomy, and phylogeny could be mined from cp genomes, because of their relatively conserved gene structure and sequence divergence between species and individuals (Parks et al. 2009; Huang et al. 2014). Furthermore, cp genomes can provide effective genetic markers for evolutionary studies from population level. In addition, analyzing the composition and structure of the cp genomes for such an important genus *Phoebe* can explore further genetic variations, which could improve quantitative and quality traits.

In this study, we reported the complete and annotated DNA sequences for the cp genomes of *P. chekiangensis* and *P. bournei* using next-generation sequencing platform, which is also the first comprehensive analysis on cp genomes for *Phoebe* combining the cp genomes of *P. sheareri* and *P. sheareri* var. *oineiensis* previously reported (Song et al. 2016). The specific aims of the present study were to: (1) present the complete chloroplast genome sequences and investigate global structural patterns of *P. chekiangensis* and *P. bournei*; (2) examine variations of repeat sequences and simple sequence repeats (SSRs) among the two *Phoebe* chloroplast genomes; (3) screen sequence divergence hotspot regions in the four *Phoebe* chloroplast genomes.

Methods

Plant material and DNA extraction

Young leaves of *P. chekiangensis* and *P. bournei* were sampled from single seedlings growing in the nursery located at 30.21N, 120.02E, of Zhejiang Academy of Forestry. The provenance was Hangzhou, Zhejiang Province and Mingxi, Fujian Province, PR China, which was the primary distribution area of *P. chekiangensis* and *P. bournei*, respectively. Total genomic DNA per species was extracted from 30 mg of the silica-dried leaf using the modified CTAB method (Porebski et al. 1997). The

quality and concentration of the genomic DNA were assessed using agarose gel electrophoresis and an Agilent BioAnalyzer 2100 (Agilent Technologies).

Chloroplast genome illumina sequencing, assembly and annotation

Genomic DNA was used to generate short-insert (500 bp) paired-end sequencing libraries according to the Illumina standard protocol. Genomic DNA from each species was sequenced using a HiSeq™ 2000 analyzer (Illumina, San Diego, California, USA) at Beijing Genomics Institute (BGI, Shenzhen, China). The raw reads (20,777,674 and 20,787,108 bp for *P. chekiangensis* and *P. bournei*) were generated with 125 bp length and assembled into whole chloroplast genomes in a multi-step approach employing a modified pipeline that involved a combination of both reference guided and de novo assembly approaches. First, paired-end sequence reads were trimmed to remove low-quality bases ($Q < 20$, 0.01 probability error) and adapter sequences using CLC-quality trim tool (quality_trim software included in CLC ASSEMBLY CELL package, <http://www.clcbio.com/products/clc-assembly-cell/>) before undertaking sequence assembly. Second, the contigs were assembled using CLC de novo assembler with the following optimized parameters: bubble size of 98, minimum contig length of 200, mismatch cost of two, deletion and insertion costs of three, length fraction of 0.9, and similarity fraction of 0.8. Third, all the contigs were aligned to the reference chloroplast genome of *Machilus yunnanensis* (NC028073) using BLAST (<http://blast.ncbi.nlm.nih.gov/>), and aligned contigs ($\geq 90\%$ similarity and query coverage) were ordered according to the reference chloroplast genome. Then, contigs were aligned with the reference genome to construct the draft chloroplast genome of each species in Geneious 9.0.5 software (<http://www.geneious.com>). Finally, clean reads were re-mapped to the draft cp genomes of two *Phoebe* species, and the mapping ratio were 3.41% for *P. chekiangensis* and 2.15% for *P. bournei*, and average coverage depth were 578.2 and 367.1 for *P. chekiangensis* and *P. bournei*, respectively. Using the Dual Organellar Genome Annotator (DOGMA) program (Wyman et al. 2004), the two *Phoebe* chloroplast genomes were annotated. Protein-coding genes were identified by using the plastid/bacterial genetic code. Intron/exon boundaries were further determined using MAFFT v7 with those of the *M. yunnanensis* chloroplast genomes as a reference (Kato and Standley 2013). Using the program tRNAscan-SE with default settings (Schattner et al. 2005), tRNA boundaries were verified. The circular chloroplast genome maps of the *Phoebe* were drawn using the Organellar Genome DRAW (OGDRAW) software, with subsequent manual editing (Lohse et al. 2007).

Characterization of repeat sequences and SSRs

REPuter was used to visualize both forward, palindrome, reverse and complement repeats, with a minimum repeat size of 30 bp and a sequence identity greater than 90% with hamming distance equal to 3 in *P. chekiangensis* and *P. bournei* (Kurtz and Schleiermacher 1999). The SSRs, which usually have a higher mutation rate, are easily genotyped using PCR, used as markers for phylogenetic analysis and in marker-assisted breeding when located on nuclear chromosomes. Microsatellites, or simple sequence repeats (SSRs) were detected using MISA perl script with thresholds of ten repeat units for mononucleotide SSRs, five repeat units for dinucleotide SSRs, four repeat units for trinucleotide SSRs, and three repeat units for tetra-, penta-, and hexanucleotide SSRs (Thiel et al. 2003).

Divergence hotspot identification

The four complete chloroplast genome sequences (*P. chekiangensis*, *P. bournei*, *P. shearereri*, and *P. shearereri* var. *oineiensis*) were aligned, using the chloroplast genome of *M. yunnanensis* as a reference by mVISTA program (Frazer et al. 2004). Default parameters were utilized to align the chloroplast genomes in Shuffle-LAGAN mode and a sequence conservation profile was visualized in an mVISTA plot. To discover the divergence hotspot regions in *Phoebe*, protein coding gene, intron, and intergenic spacer region were evaluated with DnaSP 5.10 (Librado and Rozas 2009). All the regions were sequentially extracted under the following criteria: (1) total number of mutation (Eta) >0; (2) an aligned length >200 bp. Any large structural events, such as gene order rearrangements and IR expansions/contractions, were recorded.

Results

Chloroplast genomes organization of *P. chekiangensis* and *P. bournei*

For these two *Phoebe* species, just three contigs which were found to be significantly homologous to the reference genome were combined to generate each chloroplast genome, with no gaps or missing nucleotides (Ns) found. The total genome size of *P. chekiangensis*, with a length of 152,849 bp (deposited in GenBank, Accession No. KY346511), was in close proximity to those of other *Phoebe* species and only 4, 6, and 27 bp smaller than that of *P. bournei* (152,853 bp, deposited in GenBank, Accession No. KY346512), *P. shearereri* var. *oineiensis* (152,855 bp, GenBank Accession No. KX437772), and *P. shearereri* (152,876 bp, GenBank Accession No. KX437773), respectively. The size of complete chloroplast genome of the four *Phoebe* species was within the range of angiosperms (Yang et al. 2010). Similar to the vast majority of angiosperms, showing a typical quadripartite

structure, the two *Phoebe* chloroplast genomes consist of a pair of inverted repeats (IRs) of 18,927 bp in *P. chekiangensis* and 18,928 bp in *P. bournei*, a large single copy (LSC) region of 93,772 bp in *P. chekiangensis* and 93,777 bp in *P. bournei* and a small single copy (SSC) region of 20,775 bp in *P. chekiangensis* and 20,774 bp in *P. bournei* (Fig. 1; Table 1). Protein-coding regions accounted for 46.23% of the whole genome, while tRNA and rRNA regions accounted for 1.79 and 5.91%, respectively, and the remaining 46.07% was non-coding regions (Table 1). The overall GC content was 39.1%, whereas the GC content in the LSC, SSC and IR regions were 38.0, 34.4, and 44.4%, respectively, indicating identical levels among the two *Phoebe* chloroplast genomes.

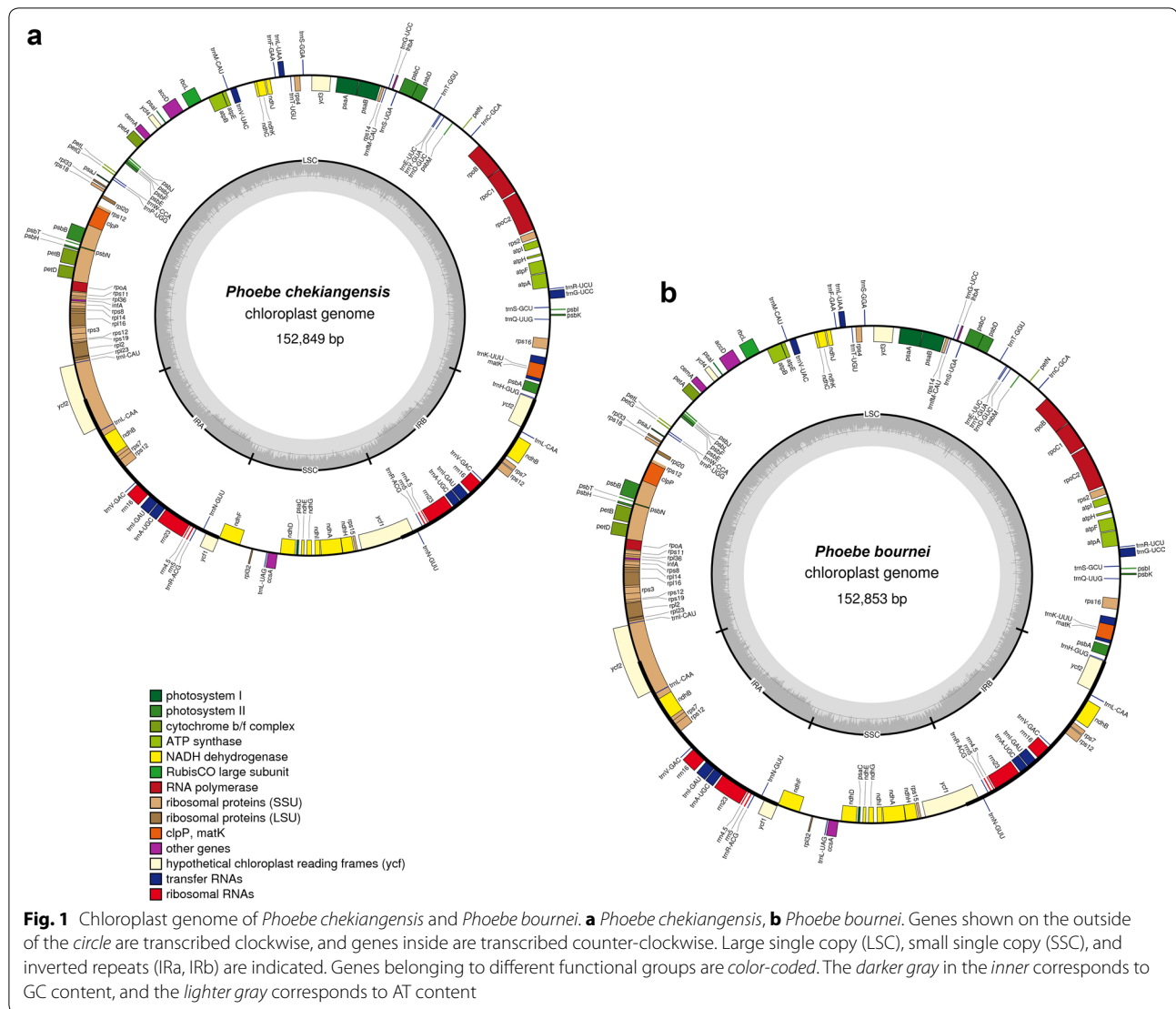
SSR and repeat sequences analysis

Repeat sequences played a vital role in phylogenetic analysis and genome rearrangement (Nie et al. 2012). We used REPuter to analyze the repeat sequence of *Phoebe* cp genomes and found forward repeats, palindrome repeats and reverse repeats of at least 30 bp per repeat unit with a sequence identity of $\geq 90\%$. *P. chekiangensis* contained 36 repeats comprising of 13 forward repeats, 15 palindromic repeats, 6 reverse repeats and 2 complement repeats. The only quantity difference in repeat type between *P. chekiangensis* and *P. bournei* was that the latter contained 14 palindromic repeats, one less than the former (Fig. 2a; Additional file 1). The quantity of repeats with 30–40 bp length was 33 and 32 for *P. chekiangensis* and *P. bournei*, respectively. In addition, *P. chekiangensis* and *P. bournei* both contained one repeat with 41, 42, and 48 bp length (Fig. 2b; Additional file 1).

With MISA analysis, *Phoebe* chloroplast genome was found to contain 66 (*P. chekiangensis*) and 65 (*P. bournei*) SSRs longer than 10 bp, of which 49 SSRs were the same for the two chloroplast genomes (similar repeat units located in similar genomic regions) (Fig. 3a; Additional file 2). Among the total 131 SSRs, most loci were located in intergenic spacer (IGS) regions (61.07%), followed by introns (23.37%) and CDS (17.56%) (Fig. 3b). We observed that 12 different SSRs were located in five protein-coding genes [*ycf1* ($\times 5$), *cemA* ($\times 2$), *rpoC2* ($\times 2$), *ycf2* ($\times 2$), and *matK*] of the two *Phoebe* chloroplast genomes.

Inverted repeats (IRs) contraction and expansion of four *Phoebe* species

The IR region expanded into the *ycf2* gene, creating a pseudogene fragment $\psi ycf2$ at the IRa/LSC border with length of 3,161 bp (*P. bournei* and *P. shearereri* var. *oineiensis*) and 3,162 bp (*P. chekiangensis* and *P. shearereri*). The *ycf1* gene crossed the SSC/IRa region and the pseudogene fragment $\psi ycf1$ was located at the IRb region with 1381 and 1399 bp. The *trnH* was the unique gene with



large difference among *Phoebe* species. The *trnH* genes of *P. sheareri* and *P. sheareri* var. *oineiensis* were separated by 21 bp at the IRa/LSC border, whereas those of *P. bournei* and *P. chekiangensis* were separated by 43 bp (Fig. 4).

Divergence sequence hotspots in *Phoebe* species

The overall sequence identity of the four *Phoebe* chloroplast genomes was compared and plotted using the mVISTA program (Frazer et al. 2004), with the annotation of *M. yunnanensis* as a reference to elucidate the level of sequence divergence (Fig. 5). Being largely consistent with recent studies (Yao et al. 2015; Zhang et al. 2016), most of the sequence variations were found to be located in the LSC and SSC regions, while the IR regions exhibited comparatively lower sequence diversity. The lower sequence divergence observed in the IRs rather

than LSC and SSC regions for *Phoebe* species is likely due to copy correction between IR sequences by gene conversion (Khakhlova and Bock 2006).

Seventy-seven regions (30 coding regions, 38 intergenic spacers, eight introns, and one rRNA) with more than 200 bp in length were eventually identified. Of these 77 regions, nucleotide diversity (P_i) ranged from 0.00018 (*rrn23*) to 0.01389 (*ccsA-ndhD*) among four *Phoebe* species (Fig. 6; Additional file 3). As found in most angiosperms (Choi et al. 2016), sequence divergence in intergenic regions was higher than that in genic regions of these four *Phoebe* chloroplast genomes. The mean value of P_i in non-coding regions was 0.221%, which was nearly more than twice as much as that (0.123% on average) in the coding regions. Intergenic regions with a percentage of P_i exceeding 0.5% were *trnH-psbA* (0.506%),

Table 1 The basic characteristics of the sequencing data for *P. chekiangensis* and *P. bournei*

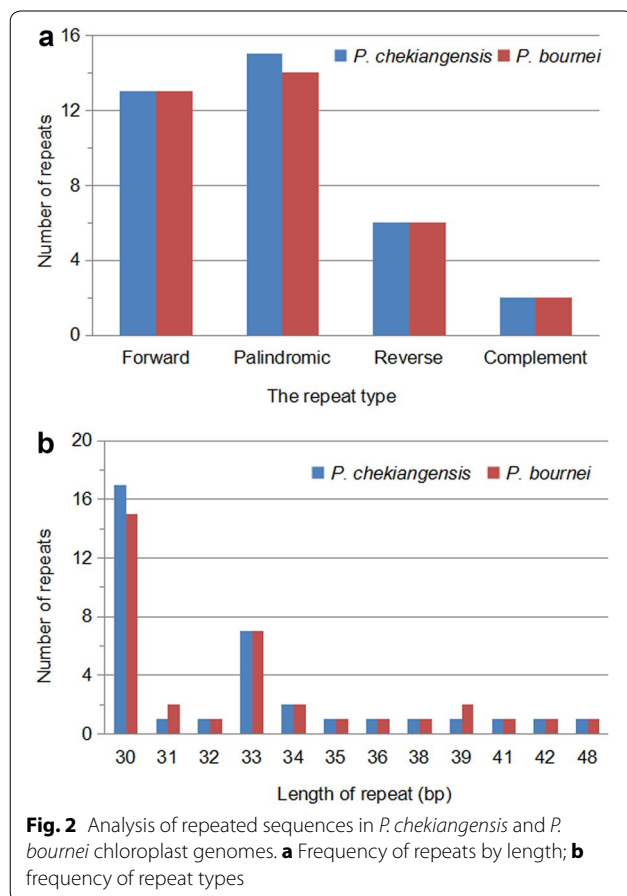
| Characteristics | <i>P. chekiangensis</i> | <i>P. bournei</i> |
|--|-------------------------|-------------------|
| Clean reads | 20,777,674 | 20,787,108 |
| Average read length | 125 | 125 |
| Total cpDNA size (bp) | 152,849 | 152,853 |
| Length of large single copy (LSC) region | 93,772 | 93,777 |
| Length of inverted repeat (IR) region | 18,927 | 18,928 |
| Length of small single copy (SSC) region | 20,075 | 20,074 |
| Total CDS length | 70,668 | 70,668 |
| Total tRNA length | 2740 | 2740 |
| Total rRNA length | 9038 | 9038 |
| Total GC content (%) | 39.1 | 39.1 |
| LSC | 38.0 | 38.0 |
| IR | 44.4 | 44.4 |
| SSC | 34.4 | 34.4 |
| Total number of genes | 113 | 113 |
| Protein-coding genes | 79 | 79 |
| rRNAs genes | 4 | 4 |
| tRNAs genes | 30 | 30 |

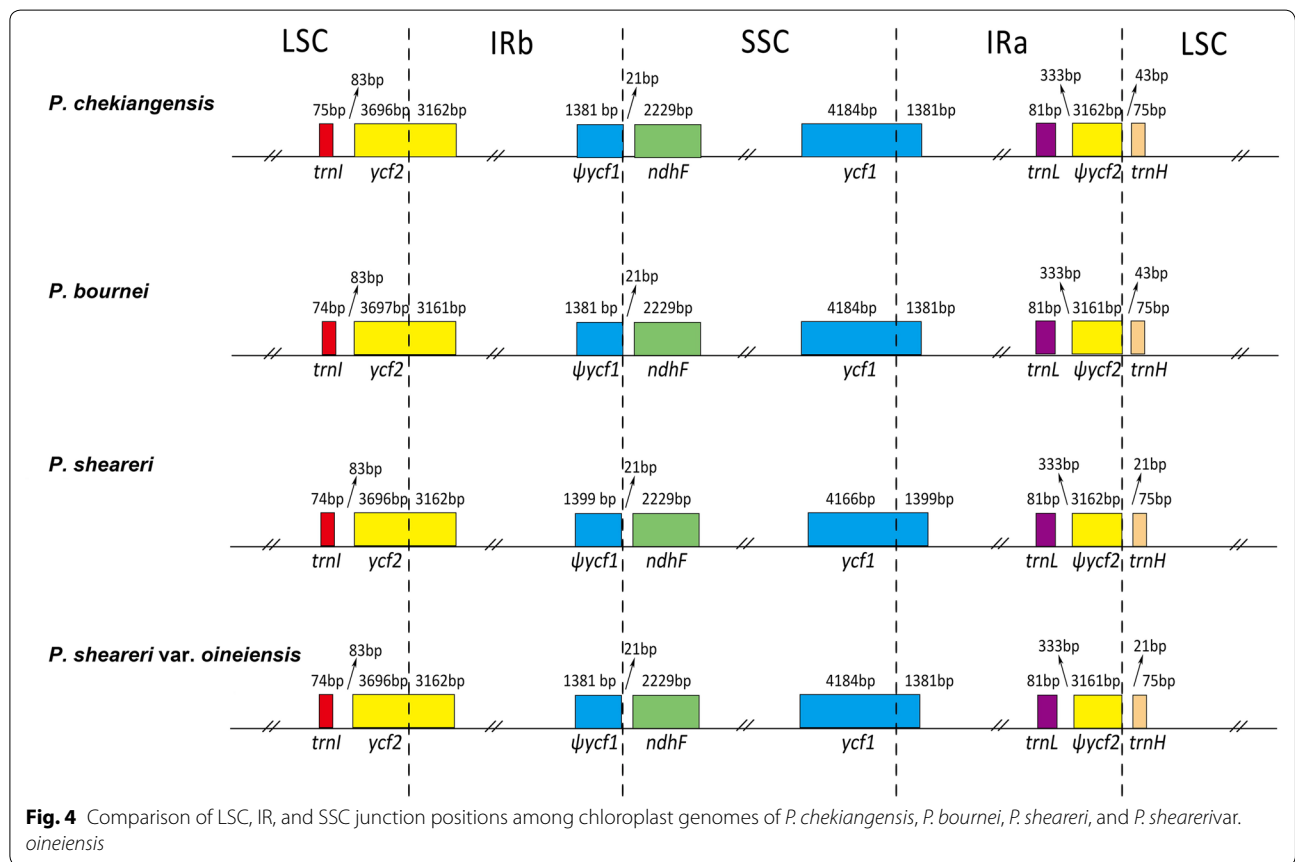
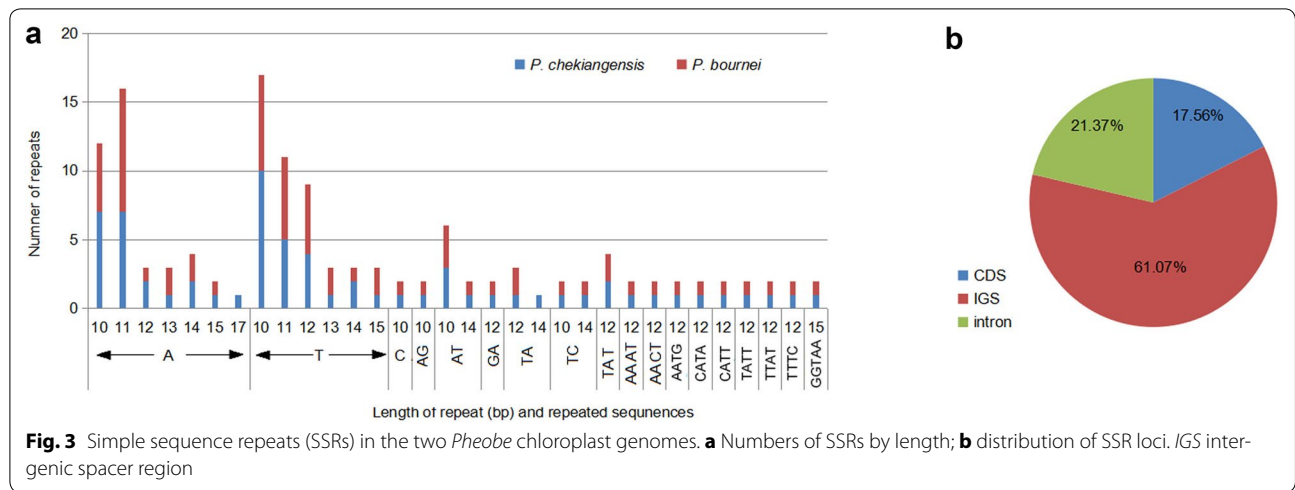
rps4-trnT (0.716%), *petA-psbJ* (0.887%), and *ccsA-ndhD* (1.389%). However, the highest proportion of variability in genic regions was 0.251% (*rps8*) (Fig. 6; Additional file 3).

Discussion

Results from this study showed the GC content of *P. chekiangensis* and *P. bournei* chloroplast genomes is close to that reported in other Lauraceae chloroplast genomes (Song et al. 2015, 2016). Although the GC percentage in the IR regions (44.4%) of the two *Pheobe* was higher than that of *Nicotiana otophora* (43%) (Asaf et al. 2016), the presence of rRNA in *Pheobe* (four) was lower than that of *N. otophora* (eight). The results in this study were incompatible to previous report which suggested that a high GC percentage in the IR regions could be due to the presence of rRNA (Qian et al. 2013). Similarly, 112 different genes, including 78 protein-coding genes (47 genes encoding photosynthesis-related proteins, four DNA dependent RNA polymerases, 20 ribosomal proteins, one translation initiation factor, four genes encoding other proteins, and two genes of unknown function), 30 tRNA genes, and four rRNA genes were annotated in *P. chekiangensis* and *P. bournei* chloroplast genomes (Fig. 1; Additional file 4). The pattern of protein coding genes was similar to that of *Persea americana* (Song et al. 2016). Among all the protein-coding genes, ten genes possessed a single intron, two genes (*ycf3* and *clpP*) contain two introns, whereas six tRNA genes contain a single intron (Additional file 4).

Overall, a total of 71 repeats, including 36.6% forward repeats (26), 40.8% palindromic repeats (29), 16.9% reverse repeats (12), and 5.7% complement repeats (4), were detected in *P. chekiangensis* and *P. bournei* chloroplast genomes. About 61.07% of these repeats were distributed in intergenic spacer. The result was comparable to chloroplast genomes of most angiosperm plant (Uthapaisanwong et al. 2012; Yao et al. 2015). Previous studies suggested that the presence of these repeats indicates that the region is a crucial hotspot for genome reconfiguration (Gao et al. 2009). One 30 bp forward repeat occurred in the *ndhC-trnV* intergenic spacer and one 30 bp palindromic repeat in *Ψycf1* were unique to *P. chekiangensis*. In contrast, one 31 bp forward repeat occurred in the *ndhC-trnV* intergenic spacer were only contained by *P. bournei* (Additional file 1). Apart from the above three repeats, the others were shared between two *Pheobe* species. SSRs in the chloroplast with high polymorphism in copy numbers have been recognized as one of the main sources of molecular markers, and extensively used for population genetics and phylogenetic investigation (Pauwels et al. 2012; Zhang et al. 2012a, b; Zhao et al. 2015).





Among 131 SSRs longer than 10 bp, almost all mono-nucleotide was composed of A/T (97.7%), and majority of dinucleotides was composed of AT (60.0%). The AT richness in SSRs of the two *Pheobe* genome was similar to previous reports suggesting that SSRs found in the chloroplast genome were generally composed of polythymine

(T) or polyadenine (A) repeats, and infrequently contained tandem cytosine (C) and guanine (G) repeats (Kuang et al. 2011; Qian et al. 2013; Chen et al. 2015). The SSRs identified in the two *Pheobe* species might be used in future population genetic studies as well as similar studies of other species, such as *Panax ginseng* (Kim and

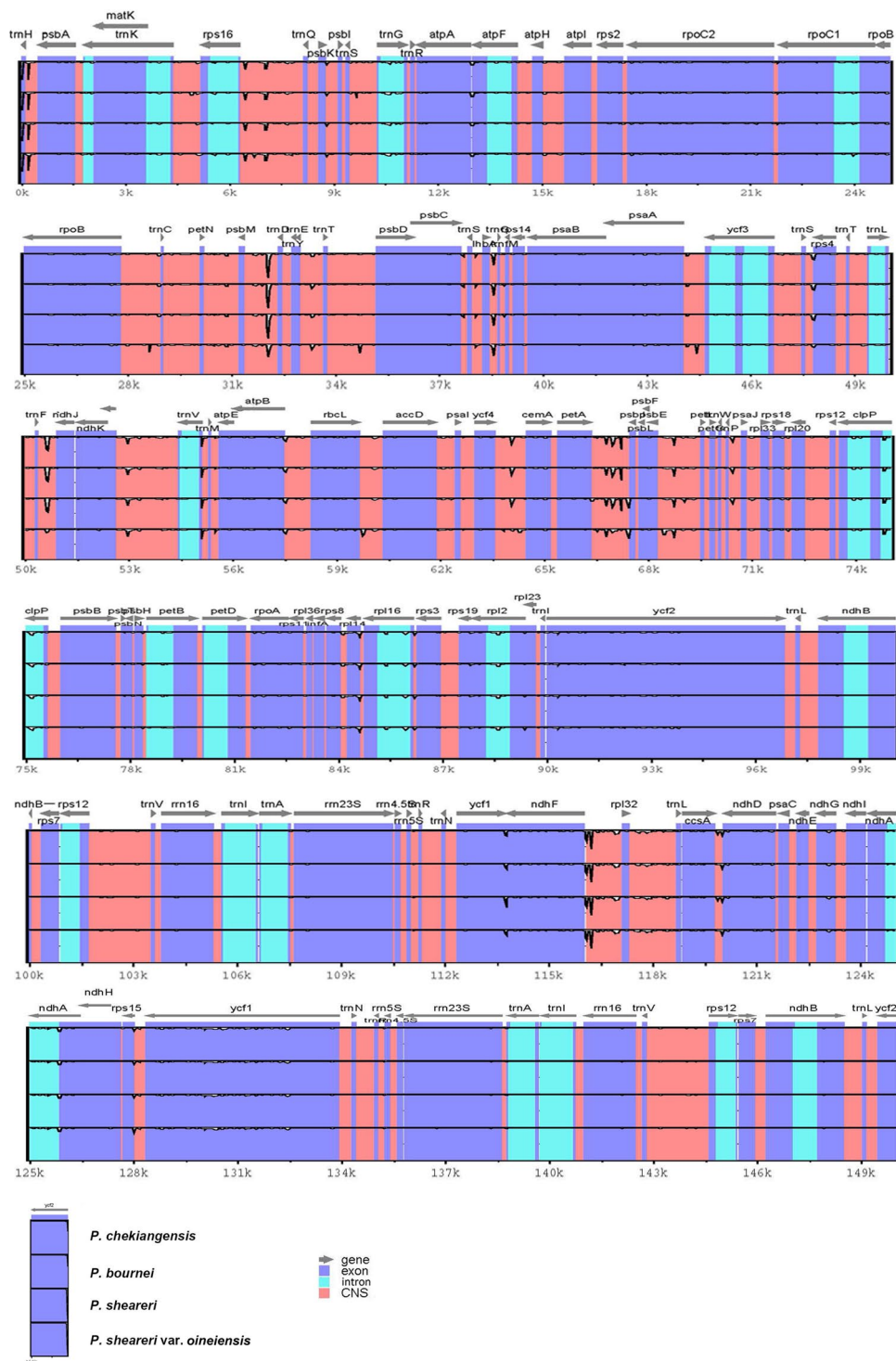


Fig. 5 Sequence identity plots among chloroplast genomes of *P. chekiangensis*, *P. bournei*, *P. shearerii*, and *P. shearerii var. oineiensis*, with *Machilus yunnanensis* as a reference. Annotated genes are displayed along the top. The vertical scale represents the percent identity between 50 and 100%. Genome regions are color coded as exon, intron, and conserved non-coding sequences (CNS)

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

All data are deposited in NCBI.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Funding

This work was supported by the Major Science and Technology Program (No. 2010C12009), Key Research & Development Program (No. 2017C02028) of Zhejiang Province, and National Key Research & Development Program (No. 2016YFD0600603) of China.

Publisher's Note

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

Received: 12 May 2017 Accepted: 20 August 2017

Published online: 13 September 2017

References

- Asaf S, Khan AL, Khan AR, Waqas M, Kang SM, Khan MA, Lee SM, Lee IJ (2016) Complete chloroplast genome of *Nicotiana glauca* and its comparison with related species. *Front Plant Sci* 7:843–854
- Chen JH, Hao ZD, Xu HB, Yang LM, Liu GX, Sheng Y, Zheng C, Zheng WW, Chen TL, Shi JS (2015) The complete chloroplast genome sequence of the relict woody plant *Metasequoia glyptostroboides* Hu et Cheng. *Front Plant Sci* 6:447
- Choi KS, Chung MG, Park SJ (2016) The complete chloroplast genome sequences of three Veroniceae species (Plantaginaceae): comparative analysis and highly divergent regions. *Front Plant Sci* 7:342–355
- Ding YJ, Zhang JH, Lu YF, Lin EP, Lou LH, Tong ZK (2015) Development of EST-SSR markers and analysis of genetic diversity in natural populations of endemic and endangered plant *Phoebe chekiangensis*. *Biochem Syst Ecol* 63:183–189
- Dong WP, Liu H, Xu C, Zuo YJ, Chen Z, Zhou SL (2014) A chloroplast genomic strategy for designing taxon specific DNA mini-barcodes: a case study on ginsengs. *BMC Genet* 15:138–145
- Douglas SE (1990) Plastid evolution: origins, diversity, trends. *Curr Opin Genet Dev* 8:655–661
- Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I (2004) VISTA: computational tools for comparative genomics. *Nucleic Acids Res* 32:W273–W279
- Gao L, Yi X, Yang YX, Su YJ, Wang T (2009) Complete chloroplast genome sequence of a tree fern *Alsophila spinulosa*: insights into evolutionary changes in fern chloroplast genomes. *BMC Evol Biol* 9:130–142
- Hegde VR, Dai P, Ladislav C, Patel MG, Puar MS, Pachter JA (1997) D₄ dopamine receptor-selective compounds from the Chinese plant *Phoebe chekiangensis*. *Bioorg Med Chem Lett* 9:1207–1212
- Huang H, Shi C, Liu Y, Mao SY, Gao LZ (2014) Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. *BMC Evol Biol* 14:151–167
- Ingvansson PK, Ribstein S, Taylor DR (2003) Molecular evolution of insertions and deletion in the chloroplast genome of *Silene*. *Mol Biol Evol* 20:1737–1740
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780
- Khakhlova O, Bock R (2006) Elimination of deleterious mutations in plastid genomes by gene conversion. *Plant J* 46:85–94
- Kim KJ, Lee HL (2004) Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Res* 11:247–261
- Kim JS, Jung JD, Lee JA, Park HW, Oh KH, Jeong WJ, Choi DW, Liu JR (2006) Complete sequence and organization of the cucumber (*Cucumis sativus* L. cv. Baekmibaekdadagi) chloroplast genome. *Plant Cell Rep* 25:334–340
- Korotkova N, Nauheimer L, Ter-Voskanyan H, Allgaier M, Borsch T (2014) Variability among the most rapidly evolving plastid genomic regions is lineage-specific: implications of pairwise genome comparisons in *Pyrus* (Rosaceae) and other angiosperms for marker choice. *PLoS ONE* 11:e112998
- Kuang DY, Wu H, Wang YL, Gao LM, Zhang SZ, Lu L, Bonen L (2011) Complete chloroplast genome sequence of *Magnolia kwangsiensis* (Magnoliaceae): implication for DNA barcoding and population genetics. *Genome* 54:663–673
- Kurtz S, Schleiermacher C (1999) REPuter: fast computation of maximal repeats in complete genomes. *Bioinformatics* 15:426–427
- Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452
- Lohse M, Drechsel O, Bock R (2007) OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr Genet* 52:267–274
- Moore MJ, Bell CD, Soltis PS, Soltis DE (2007) Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proc Natl Acad Sci USA* 104:19363–19368
- Nie XJ, Lv SZ, Zhang YX (2012) Complete chloroplast genome sequence of a major invasive species Crofton Weed (*Ageratina adenophora*). *PLoS ONE* 7:e36869
- Parks M, Cronn R, Liston A (2009) Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biol* 7:84–100
- Pauwels M, Vekemans X, Gode C, Frerot H, Castric V, Saumitou-Laprade P (2012) Nuclear and chloroplast DNA phylogeography reveals vicariance among European populations of the model species for the study of metal tolerance *Arabidopsis halleri* (Brassicaceae). *New Phytol* 193:916–928
- Porebski S, Bailey LG, Baum BR (1997) Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Mol Biol Rep* 15:8–15
- Qian J, Song JY, Gao HH (2013) The complete chloroplast genome sequence of the medicinal plant *Salvia miltiorrhiza*. *PLoS ONE* 8:e57607
- Rohwer JG (2000) Toward a phylogenetic classification of the Lauraceae: evidence from *matK* sequences. *Syst Bot* 25:60–71
- Särkinen T, George M (2013) Predicting plastid marker variation: can complete plastid genomes from closely related species help? *PLoS ONE* 11:e82266
- Schattner P, Brooks AN, Lowe TM (2005) The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res* 33:686–689
- Song Y, Dong WP, Liu B, Xu C, Yao X, Gao J, Corlett RT (2015) Comparative analysis of complete chloroplast genome sequences of two tropical trees *Machilus yunnanensis* and *Machilus balansae* in the family Lauraceae. *Front Plant Sci* 6:662–669
- Song Y, Yao X, Tan YH, Gan Y, Richard TC (2016) Complete chloroplast genome sequence of the avocado: gene organization, comparative analysis, and phylogenetic relationships with other Lauraceae. *Can J Forest Res* 46:1293–1301
- Tangphatsornruang S, Sangsrakru D, Chanprasert J, Uthapaisanwong P, Yoocha T, Jomchai N (2010) The chloroplast genome sequence of mungbean (*Vigna radiata*) determined by high-throughput pyrosequencing: structural organization and phylogenetic relationships. *DNA Res* 17:11–22
- Terakami S, Matsumura Y, Kurita K, Kanamori H, Katayose Y, Yamamoto T (2012) Complete sequence of the chloroplast genome from pear (*Pyrus pyrifolia*): genome structure and comparative analysis. *Tree Genet Genomes* 8:1–14
- Thiel T, Michalek W, Varshney RK, Graner 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
- Uthapaisanwong P, Chanprasert J, Shearman J (2012) Characterization of the chloroplast genome sequence of oil palm (*Elaeis guineensis* Jacq). *Gene* 500:172–180
- Wang MX, Cui LC, Feng KW, Deng PC, Du XH, Wan FH, Song WN, Nie XJ (2015) Comparative analysis of Asteraceae chloroplast genomes: structural organization, RNA editing and evolution. *Plant Mol Biol Rep* 33:1526–1538

- Wicke S, Schneeweiss GM, Müller KF, Quandt D (2011) The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol Biol* 76:273–297
- Wu ZY, Peter HR, Hong DY (2008) *Flora of China*. Science press, Beijing
- Wyman SK, Jansen RK, Boore JL (2004) Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20:3252–3255
- Xu Q, Xiong G, Li P, He F, Huang Y, Wang K, Li ZH, Hua JP (2012) Analysis of complete nucleotide sequences of 12 gossypium chloroplast genomes: origin and evolution of allotetraploids. *PLoS ONE* 8:e37128
- Yang M, Zhang XW, Liu GM, Yin YX, Chen KF, Yun QZ, Zhao DJ, Al-Mssallem IS, Yu J (2010) The complete chloroplast genome sequence of date palm (*Phoenix dactylifera* L.). *PLoS ONE* 5:e12762
- Yang J, Yue M, Niu C, Ma XF, Li ZH (2017) Comparative analysis of the complete chloroplast genome of four endangered herbals of *Notopterygium*. *Genes* 8:124
- Yao X, Tang P, Li Z, Li D, Liu Y, Huang H (2015) The first complete chloroplast genome sequences in Actinidiaceae: genome structure and comparative analysis. *PLoS ONE* 10:e0129347
- Zeng SY, Zhou T, Han K, Yang YC, Zhao JH, Liu ZL (2017) The complete chloroplast genome sequences of six *Rehmannia* species. *Genes* 8:103
- Zhang Q, Li J, Zhao Y, Korba SS, Han Y (2012a) Evaluation of genetic diversity in Chinese wild apple species along with apple cultivars using SSR markers. *Plant Mol Biol Rep* 30:539–546
- Zhang R, Zhou ZC, Jin GQ, Wang SH, Wang XH (2012b) Genetic diversity and differentiation within three species of the family Lauraceae in southeast China. *Biochem Syst Ecol* 44:317–324
- Zhang YJ, Du LW, Liu A, Chen JJ, Wu L, Hu WM, Zhang W, Kim KH, Lee SC, Yang TJ, Wang Y (2016) The complete chloroplast genome sequences of five *Epimedium* species: lights into phylogenetic and taxonomic analyses. *Front Plant Sci* 7:306–317
- Zhao Y, Yin J, Guo H (2015) The complete chloroplast genome provides insight into the evolution and polymorphism of *Panax ginseng*. *Front Plant Sci* 5:696–707

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
