

Research article

Open Access

## Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots

Rui-Jiang Wang<sup>†1,2</sup>, Chiao-Lei Cheng<sup>†2</sup>, Ching-Chun Chang<sup>3</sup>, Chun-Lin Wu<sup>2</sup>, Tian-Mu Su<sup>2</sup> and Shu-Miaw Chaw<sup>\*2</sup>

Address: <sup>1</sup>South China Botanical Garden, the Chinese Academy of Sciences, Guangzhou 510650, China, <sup>2</sup>Research Center for Biodiversity, Academia Sinica, Taipei 115, Taiwan and <sup>3</sup>Institute of Biotechnology, National Cheng Kung University, Tainan, Taiwan

Email: Rui-Jiang Wang - [rwang@graduate.hku.hk](mailto:rwang@graduate.hku.hk); Chiao-Lei Cheng - [chiaolei@yahoo.com.tw](mailto:chiaolei@yahoo.com.tw); Ching-Chun Chang - [chingcc@mail.ncku.edu.tw](mailto:chingcc@mail.ncku.edu.tw); Chun-Lin Wu - [chun\\_lin0201@yahoo.com.tw](mailto:chun_lin0201@yahoo.com.tw); Tian-Mu Su - [imidase@gmail.com](mailto:imidase@gmail.com); Shu-Miaw Chaw\* - [smchaw@sinica.edu.tw](mailto:smchaw@sinica.edu.tw)

\* Corresponding author †Equal contributors

Published: 31 January 2008

Received: 15 July 2007

*BMC Evolutionary Biology* 2008, **8**:36 doi:10.1186/1471-2148-8-36

Accepted: 31 January 2008

This article is available from: <http://www.biomedcentral.com/1471-2148/8/36>

© 2008 Wang et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Various expansions or contractions of inverted repeats (IRs) in chloroplast genomes led to fluxes in the IR-LSC (large single copy) junctions. Previous studies revealed that some monocot IRs contain a *trnH-rps19* gene cluster, and it has been speculated that this may be an evidence of a duplication event prior to the divergence of monocot lineages. Therefore, we compared the organizations of genes flanking two IR-LSC junctions in 123 angiosperm representatives to uncover the evolutionary dynamics of IR-LSC junctions in basal angiosperms and monocots.

**Results:** The organizations of genes flanking IR-LSC junctions in angiosperms can be classified into three types. Generally each IR of monocots contains a *trnH-rps19* gene cluster near the IR-LSC junctions, which differs from those in non-monocot angiosperms. Moreover, IRs expanded more progressively in monocots than in non-monocot angiosperms. IR-LSC junctions commonly occurred at polyA tract or A-rich regions in angiosperms. Our RT-PCR assays indicate that in monocot IR<sub>A</sub> the *trnH-rps19* gene cluster is regulated by two opposing promoters, *S10<sub>A</sub>* and *psbA*.

**Conclusion:** Two hypotheses are proposed to account for the evolution of IR expansions in monocots. Based on our observations, the inclusion of a *trnH-rps19* cluster in majority of monocot IRs could be reasonably explained by the hypothesis that a DSB event first occurred at IR<sub>B</sub> and led to the expansion of IRs to *trnH*, followed by a successive DSB event within IR<sub>A</sub> and led to the expansion of IRs to *rps19* or to *rpl22* so far. This implies that the duplication of *trnH-rps19* gene cluster was prior to the diversification of extant monocot lineages. The duplicated *trnH* genes in the IR<sub>B</sub> of most monocots and non-monocot angiosperms have distinct fates, which are likely regulated by different expression levels of *S10<sub>A</sub>* and *S10<sub>B</sub>* promoters. Further study is needed to unravel the evolutionary significance of IR expansion in more recently diverged monocots.

### Background

Typically the cpDNAs of land plants contain two identical

segments, the inverted repeats (IRs: IR<sub>A</sub> and IR<sub>B</sub>), separated by two single copy (SC) sequences, the large single

copy (LSC) region and the small single copy (SSC) region [1,2]. Thus four junctions, termed  $J_{LA}$ ,  $J_{SA}$ ,  $J_{SB}$ ,  $J_{LB}$ , are between the two IRs and the SC regions [3,4]. A major constraint on cpDNA is its organization into large clusters of polycistronically transcribed genes [5-7]. As a result, large structural changes in cpDNA, such as segmental duplication or deletion and mutation in gene order, are relatively rare and evolutionarily useful in making phylogenetic inferences [8].

In land plants, the sizes of rRNA gene-containing IRs are notably variable, ranging from 10 kb in liverworts to 20–25 kb in most angiosperms [2,9,10], and up to 76 kb in *Pelargonium* (a eudicot) [11]. Successive IR expansions, either within angiosperms or between non-vascular plants and angiosperms, have led to floating of  $J_{LA}$  and  $J_{LB}$  [12] and have evolutionary significance [13-15]. Several models concerning the expansion and contraction of IR regions have been proposed to explain the possible mechanisms that result in shift of the IR-LSC junctions. For example, the unusual triple-sized expansion of the *Geranium* IR was hypothesized as an outcome of inversion due to recombination between homologous dispersed repeats [16]. Similarly, the at least 4 kb expansion of the IR in buckwheat (*Fagopyrum esculentum*) cpDNA was also considered to be associated with an inversion [17].

Goulding et al. [15] found that in most *Nicotiana* species IR regions have both expanded and contracted with slight variations in length during the evolution of the genus. The exception is *N. acuminata*, which underwent a large IR expansion of over 12 kb. Goulding et al. [15] proposed two mechanisms of IR expansion: (i) gene conversion to account for the small IR expansion or movements in most species of the genus, and (ii) a DNA double-strand break (DSB) to explain the extensive incorporation of the LSC region into the IR of *N. acuminata*. Perry et al. [18] analyzed the endpoint sequence of a large 78 kb rearrangement in adzuki bean (*Vigna angularis*) and concluded that the unusual organization was caused by a two-step process of expansion and contraction of the IR, rather than a large inversion.

Recent phylogenetic studies using various molecular markers have yielded robust support for the hypothesis of either *Amborella* alone or *Amborella*-Nymphaeales together as the basal-most clade of angiosperms [13,19-26], and the genus *Acorus* has been identified as the earliest splitting lineage in monocots. However, the sister group of monocots is still unclear [26].

Monocots include about one-fourth of the world's flowering plants and represent one of the oldest angiosperm lineages [27]. However, no comparative study has been conducted to investigate the diversity and evolutionary

dynamics at the IR-LSC junctions of cpDNAs in basal angiosperms and monocots as a whole. Goulding et al. [15] found that each IR in rice and maize (Poaceae) contains a fully duplicated *trnH-rps19* gene cluster. Chang et al. [20] further discovered that the IRs of two other remote monocot taxa, *Acorus* and Orchidaceae, also include *trnH* and *rps19* (although the 3' region of *rps19* was truncated in *Acorus*), and speculated that the clustering of *rps19* and *trnH* was probably duplicated before the diversification of extant monocot lineages.

As a result of expansion and contraction, the IRs in the cpDNA of angiosperms have been suggested as an evolutionary marker for elucidating relationships among some taxa [14,28]. To improve understanding of the dynamics and evolution of IR-LSC junctions from basal angiosperms to the emergence and diversification of monocots (assuming that this evolutionary course is correct), we sampled 52 key species and determined the sequences of the two regions spanning  $J_{LA}$  (Fig. 1, between the 3' end of *rpl2* and the 5' end of *psbA*) and  $J_{LB}$  (Fig. 1, between the 3' end of *rpl2* and the 5' end of *rpl22*). A total of 123 representative angiosperms, including 12 basal angiosperms, 16 magnoliids, 62 eudicots, and 33 monocots (see the additional file 1), were analyzed. Three types of gene arrangements flanking the  $J_{LA}$  and  $J_{LB}$  regions were recognized and mapped onto the angiosperm phylogeny. In order to explain this arrangements we propose two alternative hypotheses concerning the evolutionary history of the flux of IR-LSC. Furthermore, to verify the transcriptional status of the duplicated *trnH-rps19* gene cluster near the  $IR_A$  junctions the activity of two operons in *Asparagus densiflorus*, *S10\_A* and *psbA*, was investigated.

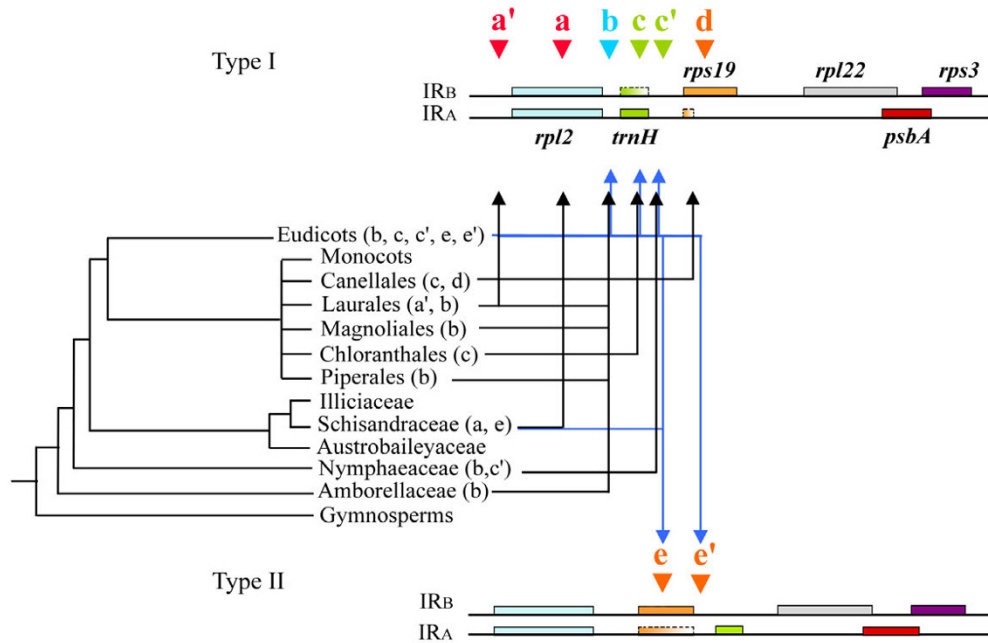
## Results

Several terms used in this section are briefly explained here. Types of IR-LSC junction are based on the organization of genes flanking  $J_{LB}$  and  $J_{LA}$  in angiosperms. Type I is found in most non-monocot angiosperm dicots. It refers to an intact *trnH* gene being located directly downstream of the *rpl2* sequence in  $IR_A$  and an intact *rps19* gene being located directly downstream of the *rpl2* sequence in  $IR_B$ . No full-length *rps19* or *trnH* sequence is present in  $IR_A$  or  $IR_B$  respectively. Type II refers to a partial sequence of *rps19* being located directly between *rpl2* and *trnH* in  $IR_A$ . Type II pattern is only found in some eudicots while type III characterizes the IRs of most monocots, in which each IR contains a *trnH-rps19* cluster. The letters a, a', c, ... and g used in the text and in Figure 1 refer to the IR-LSC junctions found in cpDNAs of sampled angiosperms.

### **In non-monocot angiosperms IR-LSC junctions of $IR_B$ are largely located between *rpl2* and *rps19***

Figure 1 shows that the IR-LSC junctions in 90 non-monocot angiosperms usually drift around position b

A Basal Angiosperms, Magnoliids, and Eudicots



B Monocots

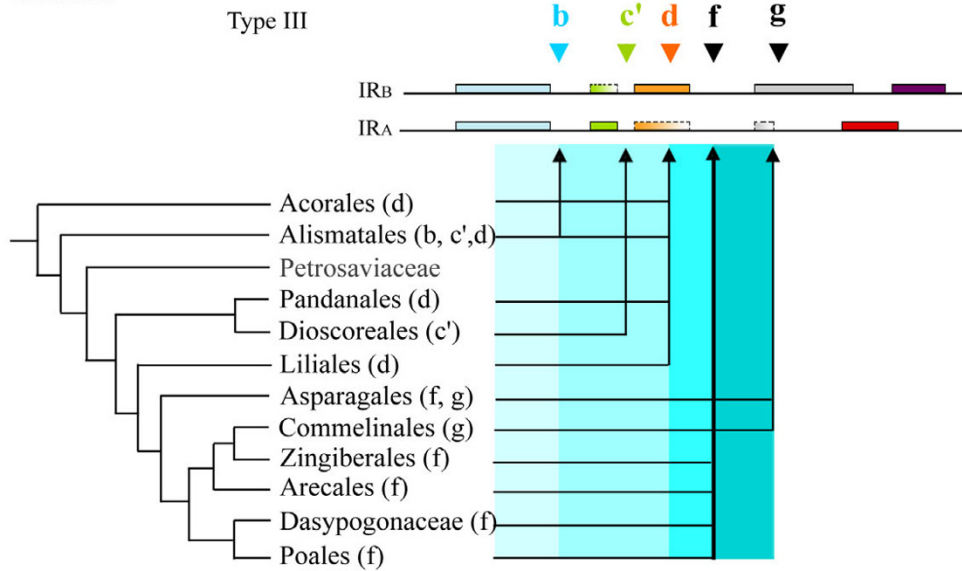


Figure 1

Classification of IR-LSC junction types based on the organization of genes flanking  $J_{LB}$  and  $J_{LA}$  in angiosperms. Triangles coded by different colors and letters indicate various locations of IR-LSC junctions in corresponding angiosperm lineages. Shaded boxes denote protein-coding genes, and boxes with broken margins and gradient color stand for genes that are variable in length. Relationships of major non-monocot (A) and monocot (B) lineages followed the phylogenetic trees of Soltis et al. (2005) [27]. (A) In type I the IR-LSC junction is located downstream of *rpl2* and upstream of *rps19*. In type II *rps19* is located downstream of *rpl2* in  $IRA$ . (B) In type III each IR has a copy of the *trnH-rps19* cluster, although in the  $IRA$  regions the *rps19* genes are variously truncated at the 3' regions in sampled taxa. The blue gradient on the right side of the monocot phylogenetic tree denotes the progressively expanded IRs.

(data shown in the additional file 1). In these cases, designated as type I, an intact *trnH* gene is always present near the  $J_{LA}$  but absent from the  $J_{LB}$ . In *Chloranthus oldhami*, *C. spicatus*, *Sarcandra glabra* (Chloranthales), *Canella winterana* (Canellales), *Ranunculus japonica* and *R. macranthus* (eudicot), a partial *trnH* sequence is found extending to position c in  $IR_B$  (Fig. 1A, additional file 1). The IR-LSC junctions were located upstream of position c' (i.e. upstream of *trnH*) in *Nuphar advena* (Nymphaeaceae) and *Elaeagnus formosana* (Elaeagnaceae, eudicot), at position a in *Kadsura japonica* (Schisandraceae, Austrobaileyales), and at position a' in *Calycanthus fertilis* and *C. floridus* (Calycanthaceae, Laurales, [29,30]) (Fig. 1A). However, *Vitis vinifera* (Vitaceae, eudicot) showed a complete loss of *rpl2* near  $J_{LA}$  [31].

The Winteraceae (Canellales), exemplified by *Zygogynum pauciflorum* and *Drimys granadensis* [29], were exceptional in that the organization of the genes flanking the IR-LSC junctions resembled the one found in most monocots, rather than the organization seen in other non-monocot angiosperms. Notably, each of their IRs contained a *trnH-rps19* cluster and their IR-LSC junctions were located within the 5' portion of *rps19* (position d, Fig. 1).

Type II IR-LSC junctions were found in *Schisandra arisanensis* (Schisandraceae; Austrobaileyales) and some 41 representative eudicots (Fig. 1A; additional file 1). Unlike type I, the  $J_{LA}$  of type II shifted to the 5' end of the truncated *rps19* in  $IR_A$  (position e and e', Fig. 1A, additional file 1).

#### **IRs of monocots generally contain *trnH-rps19* clusters**

In contrast to basal angiosperms and eudicots, most monocots (Fig. 1B) had *trnH-rps19* clusters present in each of the two IRs, and the IR-LSC junctions were generally at position f (Arecaceae, Dasygongonaceae, *Asparagus densiflorus* [Liliales], Poales and Zingiberales) or g (in Asparagales and Commelinales) (Fig. 1B). This type of gene organization was classified as type III. In addition, IR-LSC junctions of some monocots were located downstream of *rpl2* (position b; in Araceae, most Alismataceae, and Hydrocharitaceae), of *trnH* (position c' in Potamogetonaceae and Dioscoreaceae), or within *rps19* (position d, Fig. 1; in Acorales, *Lilium formosamum* [Liliales] and Pandanales). When the IR-LSC junction was at position d, the *rps19* sequence in  $IR_A$  was found to be partially truncated most of the times.

#### **Sequences flanking IR-LSC junctions are more variable in monocots than in non-monocot angiosperms**

Figure 2 illustrates alignment of the sequences flanking the  $J_{LA}$  regions in some representatives of basal angiosperms and eudicots (A) and monocots (B). Of particular interest is the observation that the IR-LSC junctions

of basal angiosperms, eudicots and monocots are commonly found at either polyA tract or A-rich regions (Fig. 2). We also found that the dicot IR sequences near the IR-LSC junctions varied little and could be aligned among orders having the same or different IR-LSC junction types, while in monocots the corresponding regions were very different and difficult to align across different orders (Fig. 2B). Moreover, within the sampled angiosperm families the sequences flanking the  $J_{LA}$  regions were very similar.

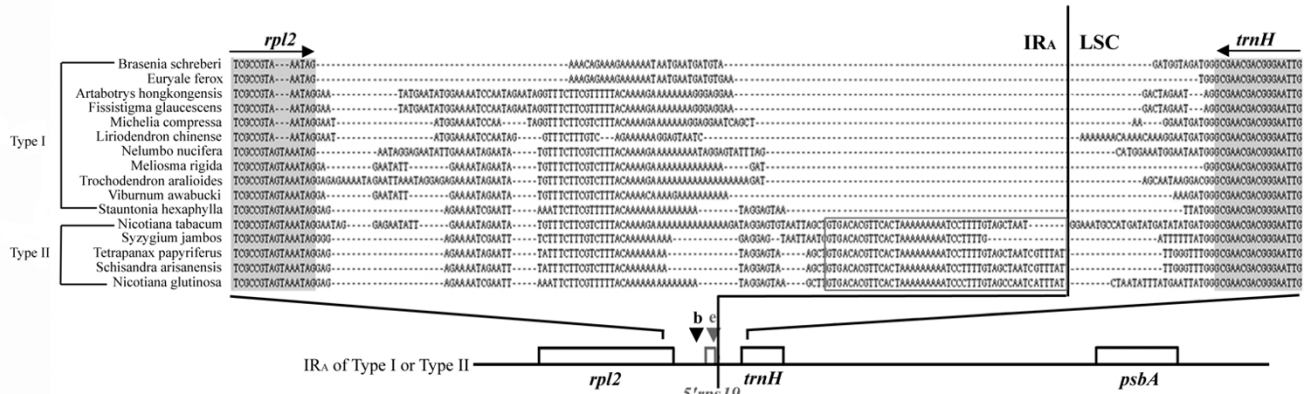
#### **Transcription of monocot *trnH-rps19* of $IR_A$ is regulated by both chloroplast $S10_A$ and *psbA* promoters**

Among the chloroplast operons, the *S10* ribosomal protein operon is the largest. It contains genes encoding both small (*rps*) and large (*rpl*) ribosomal protein subunits that are organized into a polycistronic transcription unit conserved in known cpDNAs [32]. In angiosperms, the 5' end of the *S10* operon is initiated within the IR, but only in  $IR_B$  does the operon extend into the LSC region, and the *S10* operon is only partially in  $IR_A$  (viz. the  $S10_A$  operon). However, a second operon in  $IR_A$ , the *psbA* operon, is transcribed from LSC towards  $IR_A$  [32] and opposite to the  $S10_A$  operon.

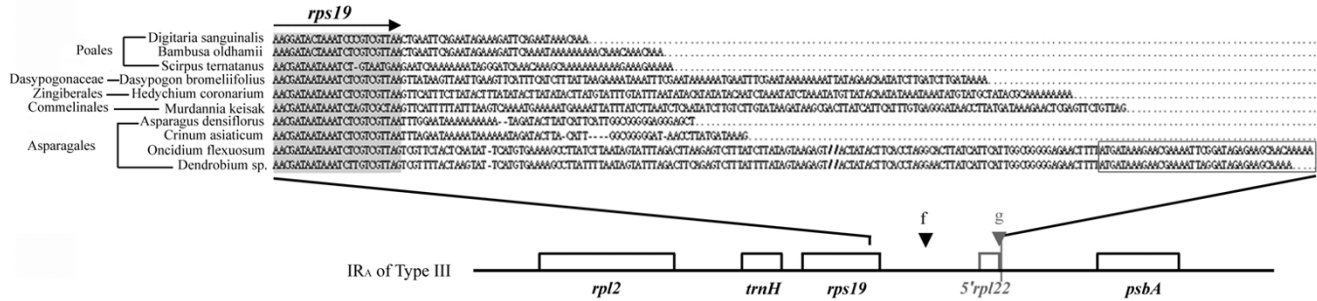
In the Winteraceae and a majority of monocots, the *trnH-rps19* cluster of  $IR_A$  is included in both the *S10* and *psbA* operons. Therefore, this gene cluster may be regulated by two opposing promoters, the  $S10_A$  and the *psbA* (Fig. 3A). In monocots, if the *trnH* in  $IR_A$  is indeed regulated by the above-mentioned two opposing promoters, the function of the *trnH* gene may be repressed because antisense-*trnH* RNAs would be generated by both the  $S10_A$  and  $S10_B$  promoters. To verify this possibility, we conducted RT-PCR assays using specific primers for a type III representative, *Asparagus densiflorus*, with the IR-LSC junction located at position f (Fig. 1B).

Our results indicate that expression of the *trnH* gene in  $IR_A$  is regulated by both the  $S10_A$  and *psbA* promoters. This suggests that the duplicated *trnH* gene located in the  $IR_B$  region of most monocots and in some non-monocots has different fates (i.e. functional or degenerate in different lineages; see Fig. 1). Figure 3B shows that two RT-PCR products, a 250 bp and a 700 bp fragment, respectively, were generated when specific primer pairs for each were used (Fig. 3A). The former fragment was amplified from the transcripts made by the *psbA* promoter, and the latter by the  $S10$  promoter. This result confirms that the *trnH-rps19* cluster of  $IR_A$  is regulated by two opposing promoters (Fig. 3B), indicating that the transcription machinery in IRs of monocots may differ from that of basal angiosperms and eudicots.

A Basal Angiosperm, Magnoliids, and Eudicots



B Monocots



**Figure 2**

Alignment of sequences flanking  $J_{LA}$  regions in some basal angiosperms, Magnoliids, and eudicots (type I at position b, and type II at position e), and the sequences within the  $J_{LA}$  in some monocots (type III at position f or g). Dashed lines denote gaps. Grey segments and the arrow lines above indicate coding regions and transcriptional directions of specified genes, respectively. (A) Grey box denotes degenerate *rps19* genes (5' segment) found in the  $IR_A$  of the type II (position e) pattern. (B) A degenerate *rpl22* gene (boxed sequences) found in the  $IR_A$  of type III (position g). "/" stands for abbreviated base pairs in the sequences of *Oncidium* and *Dendrobium*.

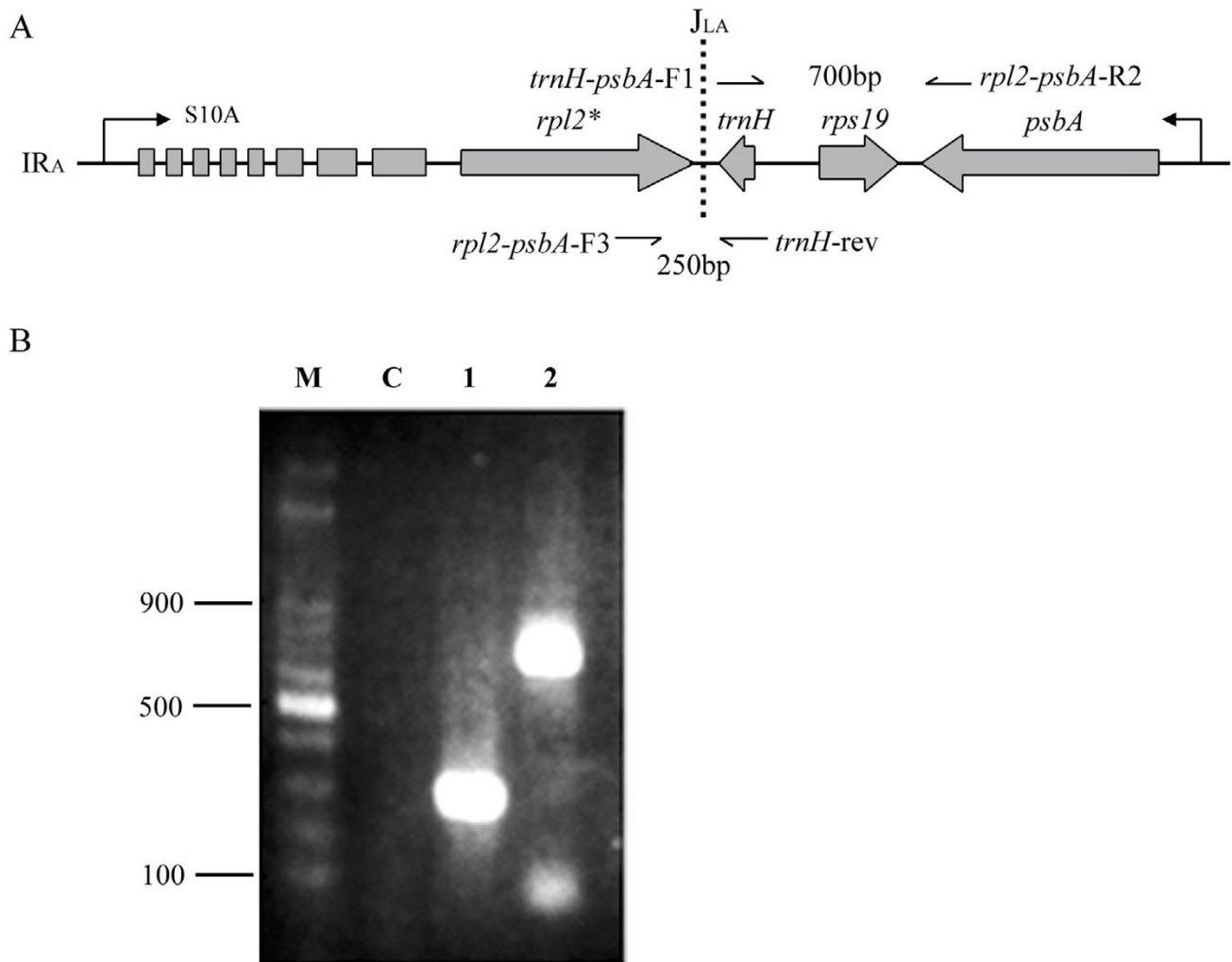
**Discussion**

**Two evolutionary hypotheses for the flux of IR-LSC junctions in monocots**

As shown in Figure 1A, IR-LSC junctions of the *Amborella* + Nymphaeales are mainly located at position b, but junctions of monocots are further expanded to encompass LSC genes and are located at positions f or g. Since the two IRs of monocots usually include the *trnH-rps19* cluster (position f or g, further downstream of *rpl2*; Fig. 1B), we hypothesize that at least two duplication events are required to explain the expansion of IRs in monocots during the course of IR evolution from an *Amborella*-like ancestor to present-day monocots. If this hypothesis is correct, it is expected that an intermediate junction type could be traceable in the cpDNAs of some early divergent monocot lineages between the two duplication events.

Narayanan et al. [33] have recently presented a model of gene amplification in eukaryotes that argues strongly for the involvement of hairpin-capped DSBs in the initiation.

Based on this model and our observations, we propose two hypotheses to account for the evolution of IR expansions in monocots (Fig. 4). In hypothesis A, a DSB event (Fig. 4, red arrowhead in step 1) occurs first within the  $IR_B$  of an *Amborella*-like ancestor, and then the free 3' end of the broken strand is repaired against the homologous sequence in  $IR_A$ . The repaired sequence extends over the original IR-LSC junction and reaches the area downstream of *trnH* (Fig. 4, step 1), so that duplication of a *trnH* gene in the newly repaired  $IR_B$  is achieved. Similarly, a second DSB event occurs in  $IR_A$  adjacent to the  $IR_A$ -LSC junction (Fig. 4, red arrowhead at step 2) so that duplication of *rps19* at  $IR_A$  can be initiated, and a *trnH-rps19* cluster nearby  $J_{LB}$  (Fig. 4, step 2) is created. The newly formed IRs might cover the *trnH-rps19* cluster and extend further into the intergenic spacer between *rps19* and *rpl22* (Fig. 4, step 1 to step 2). Furthermore, if one additional DSB event took place within the intergenic spacer located between *rps19* and *rpl22* in the LSC region, a partial *rpl22* gene would be duplicated at  $IR_A$  using the *rpl22* sequence of

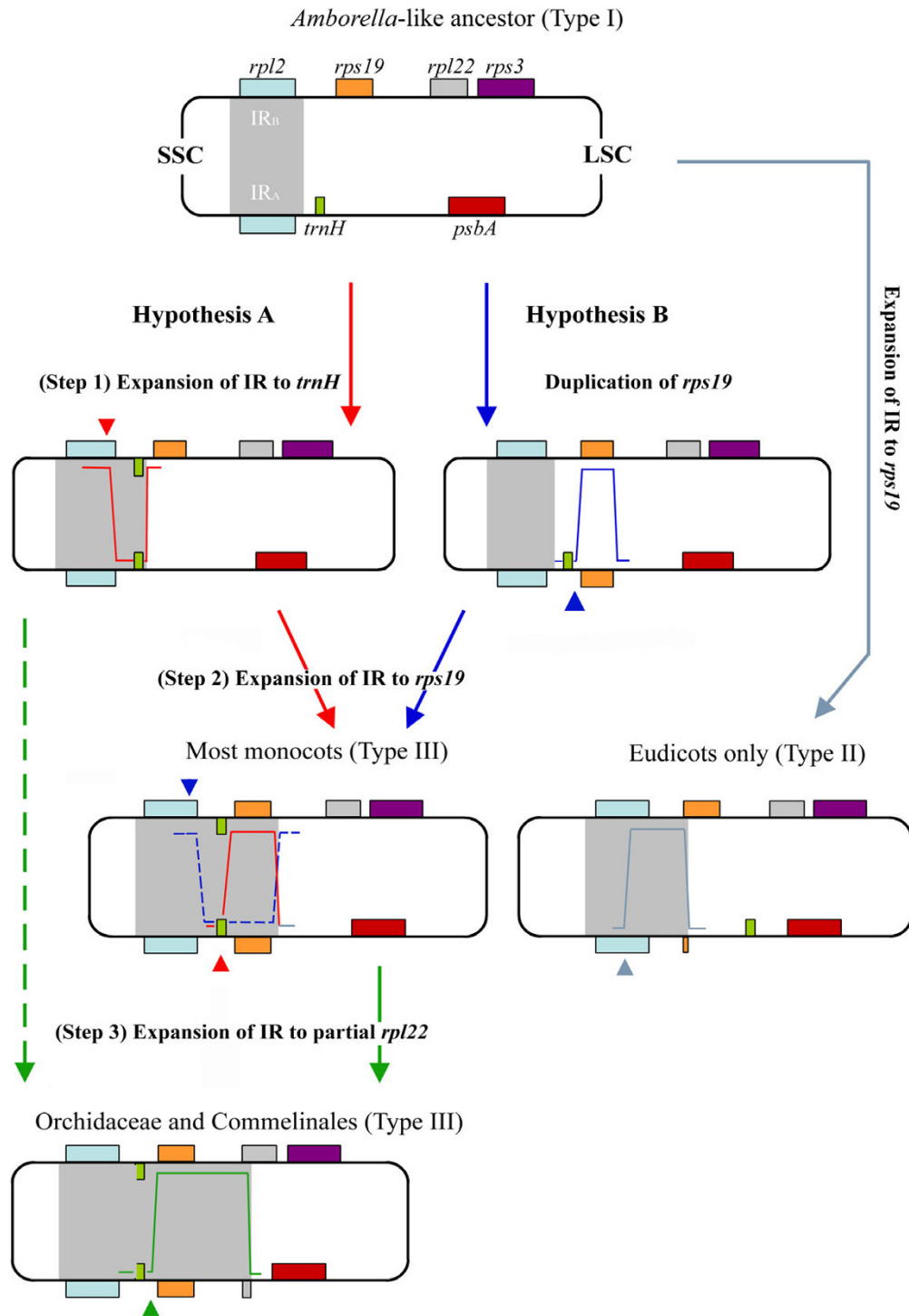


**Figure 3**

Transcription analysis of the *S10* and *psbA* operons in a monocot representative, *Asparagus densiflorus*. (A) The relative position of the *S10* and *psbA* operons at the flanking region of the  $IR_A$ -LSC junction. An arrow line denotes the transcriptional direction. One-side arrow lines indicate primers. (B) Transcripts obtained by reverse transcription PCR (RT-PCR). Lane M, 100 bp ladder; lane C, negative control using the same RNA as the template in lanes 1 and 2; lane 1, RT-PCR with the primer pair *trnH*-rev and *rpl2-psbA*-F3; lane 2, RT-PCR with the primer pair *trnH-psbA*-F1 and *rpl2-psbA*-R2.

LSC as a template, and from then on the repaired IRs might have expanded towards the 5' region of the *rpl22* (Fig. 4, step 2 to step 3). The exceptionally long IRs observed in the Orchidaceae and Commelinales are likely to have been generated by this process. The same outcomes could also result if the process proceeded directly from step 1 to step 3 without step 2 (Fig. 4, path indicated by green dashed arrow).

Hypothesis B, on the other hand, assumes that *rps19* would be duplicated or converted prior to the duplication of *trnH* through a DSB event that takes place at  $IR_A$  first (Fig. 4; blue arrowhead of step 1). A second DSB event (Fig. 4; blue arrowhead of step 2) then would take place within the  $IR_B$  region through a similar repair process to the one mentioned before, so that a duplicated *trnH* is generated at  $IR_B$ . Finally, the IRs expand downstream of



**Figure 4**

Two hypotheses for the evolutionary derivation of the *trnH-rps19* cluster in IRs of monocots from an *Amborella*-like ancestor. Arrow lines coded by different colors indicate distinct evolutionary pathways. Arrowheads denote possible breakpoints when DSB events occurred (different DSB colors are associated with different IR expansions). The light blue arrow line refers to a scenario in which a type II IR-LSC junction was established (see Fig. 1) in some eudicots (note that the *rps19* residue is situated between *rpl2* and *trnH* in IR<sub>A</sub>). The grey area in each cpDNA molecule highlights the IRs at all evolutionary stages.

rps19. In hypothesis B subsequent extension of IRs is assumed to resemble step 3 of hypothesis A.

Duplication of a partial or complete rps19 gene was also observed in some eudicots and Schisandraceae (type II) with their respective IR-LSC junctions located at position e or e' (additional file 1; Fig. 1). However, these duplicated rps19 genes (both partial and complete) are situated between the rpl2 and trnH genes of the IR\_A (refer to type II in Fig. 1A and Fig. 4 [see the light blue line at the right side leading to eudicots]) rather than downstream of trnH or upstream of psbA (refer to step (2) and (3) of hypothesis A in Figure 4). Therefore, the gene arrangement flanking the IR\_A-LSC of type II deviates from that of type I, suggesting that duplication of rps19 genes in type II must have a distinct evolutionary history.

Based on comparisons of aligned rpl2-trnH and trnH-rps19 intergenic spacer sequences from representatives of major monocot orders (Figure 5A, B), it is apparent that these two spacer sequences are separately highly similar across the sampled monocot orders. These data give strong support to hypothesis A that in monocots expansion and inclusion of trnH-rps19 gene cluster in IRs might require at

least two common DSBs (please refer to steps 1 to 3 of hypothesis A in Figure 4): one occurring within IR\_B (refer to Fig. 4, step 1), and the within IR\_A (refer to Fig. 4 step 2 or 3).

However, we did not discover any inverted repeats that might have led to the formation of hairpins in the monocot intergenic spacers of trnH and rps19. Therefore, we are inclined to conclude that the expansions of monocot IRs took the path depicted in hypothesis A.

IR expansion may be initiated by DSB and end in the nearby polyA region in angiosperms

Goulding et al. [15] proposed two models to account for two kinds of IR expansion: (1) small and random IR expansions, caused by gene conversion (viz. single strand break); and (2) large IR expansions, like those found in the Nicotiana species, rice and maize, generated via DSB events. Narayanan et al. [33] further demonstrated that DSBs can trigger gene amplification through a variety of mechanisms, and that breakage at the inverted repeats of chromosomes can cause gene amplification.

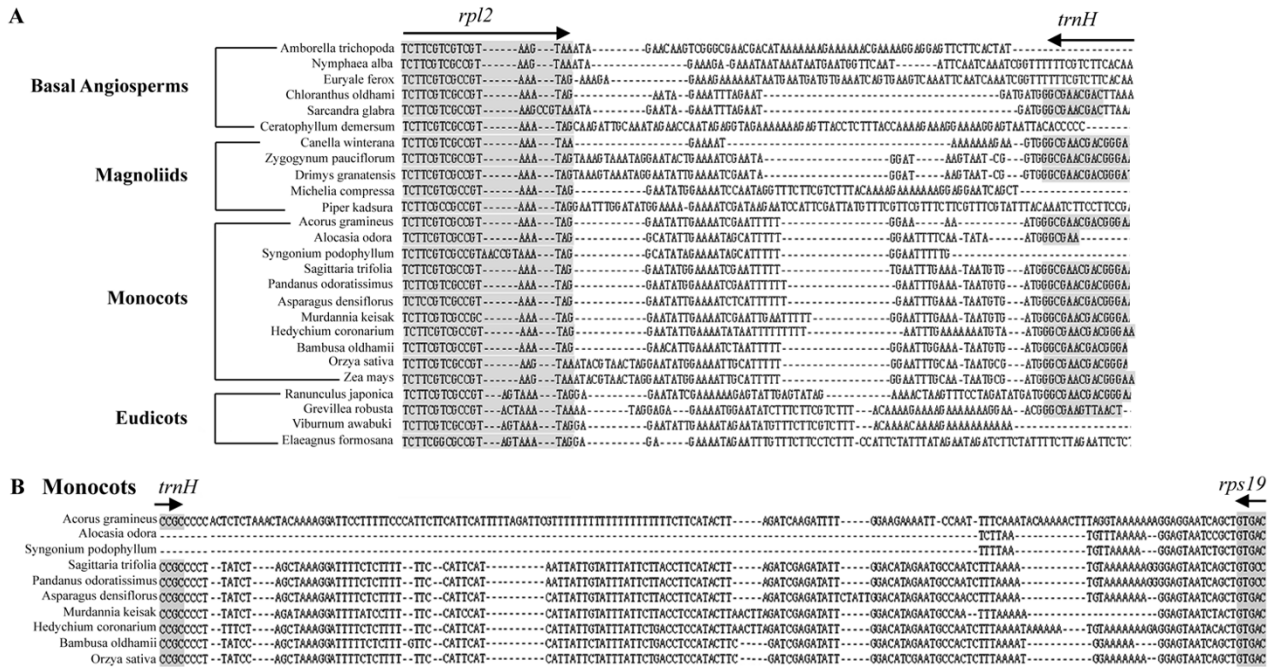


Figure 5

Comparisons of sequences that flank J\_LA regions in angiosperms. (A) Alignment of rpl2-trnH intergenic spacers in representative basal angiosperms, magnoliids, monocots, and eudicots. Grey regions and the arrow lines above indicate locations and transcriptional directions of rpl2 and trnH, respectively. (B) Alignment of the trnH-rps19 intergenic spacer sequences at IR\_A strand among representatives of major monocot orders. Grey regions with arrow lines indicate locations and transcriptional directions of trnH and rps19, respectively.



After a critical comparison of genes or sequences adjacent to the IR-LSC junctions in 33 major orders and 8 families of angiosperms (following the classification system proposed by Soltis et al. 2005 [27]), we hypothesize that IR expansions resulted principally from the DSB events that occurred during IR evolution from the *Amborella*-like ancestor to monocots. This hypothesis is founded on the following 5 observations: (1) the length of IR expansion from basal angiosperms to monocots is large (more than 100 bp); (2) *trnH* and *rps19* are situated downstream of IR<sub>A</sub> and IR<sub>B</sub>, respectively, in all sampled basal angiosperms (Fig. 1A). This type of gene arrangement might represent the ancestral gene pattern in basal angiosperms; (3) IRs of several basal angiosperms (e.g. Schisandraceae, Chloranthales and Magnoliales, Winteraceae) and eudicots (Fig. 1A) have partially or completely duplicated *trnH* genes located at IR<sub>B</sub>; (4) in comparison with other angiosperms, monocot IRs have expanded further to include a duplicated *rps19* in IR<sub>A</sub>, and this expansion should have occurred before the diversification of major monocot orders; and (5) the IRs of advanced monocots (from Asparagales to Poales) have expanded to encompass more LSC sequences or genes (Fig. 1B). Nevertheless, the latter expansions did not apparently result from another common DSB event but from independent ones, because among sampled monocot orders the downstream regions of *rps19* genes have low sequence similarity (Fig. 2). At the infra-order level of angiosperms, gene conversion might occur frequently at meiosis and cause small IR expansion or contraction during evolution, as found in Apiaceae [14] and *Nicotiana* [15].

Studies on the IR-LSC junctions of *Nicotiana* species [15] and Apiaceous plants [14] have indicated that short repeats or "polyA tract" sequences associated with tRNAs at the IR-LSC boundaries might be likely hotspots for recombination. We also observed that polyA tract sequences are commonly present near the IR-LSC junctions in all the basal angiosperms, eudicots and monocots examined (Fig. 2), indicating that such sequences are closely linked with the dynamics of IR-LSC junctions and expansion of IRs. In this regard, we further propose that IR expansion may initiate at the DSBs and finish at the polyA tract regions, where recombination may actively occur, and that the recombination mechanism in cpDNA may resemble that reported for nuclear genomes by Narayanan et al. [33].

According to our hypothesis, DSBs within IRs must have been frequent during angiosperm evolution. However, only those which led to successful IR expansions, and have subsequently been retained in the extant taxa, are detectable. Based on our observations, it is evident that the type of IR-LSC junction appears to be informative, at

least at the level of order, and is therefore useful for inferring phylogenetic relationships at this rank and above.

#### **Expansion of monocot IRs is correlated with the divergence pattern of monocot phylogeny**

As shown in Figure 1B, IR-LSC junctions of basal monocots including Acorales, Pandanales and Liliales are usually located at position d. This type might represent a primitive state. In contrast, IR-LSC junctions of the derived monocots, such as Asparagales and Poales, have generally expanded to position f or g. This trend in IR expansion seems to correlate well with the divergence pattern of monocot lineages in the multigene tree of Soltis et al. [27,34], which shows Acorales to be a sister group to other monocots. This correlation connotes the ancient status of the order and the continuous IR expansion experienced by the more terminal and derived lineages, viz. Asparagales, Commelinales, Zingiberales, Arecales, Dasygongonaceae and Poales.

It is worth mentioning that in some monocots (e.g. Pandanales and Liliales) the IR-LSC junctions are located at position d, with a truncated *rps19* gene at IR<sub>A</sub>. According to hypothesis A (Fig. 4), duplication of *rps19* at IR<sub>A</sub> was due to a second DSB event in IR<sub>A</sub> (Fig. 4, red arrowhead at step 2), followed by a sequence repair supposed to have been terminated within or downstream of the *rps19* gene. Duplication of the *rps19* gene will lead to a shift of the IR-LSC junction to position d or f (Fig. 1B). However, in Pandanales and Liliales, the *rps19* sequences of IR<sub>A</sub> are incomplete or degraded. We considered these common degradations likely to be secondary rather than primary, since the majority of monocot orders have the *trnH-rps19* clusters (Fig. 1B). Moreover, among the major monocot orders (except Alismatales) the intergenic spacer sequences within the *trnH-rps19* cluster (Fig. 5B) have a high degree of similarity, suggesting that among the sampled monocots a common DSB event might have taken place adjacent to the *trnH* gene. Therefore, the IRs in Acorales, Pandanales and Liliales are likely to have contracted, causing a shift of the IR-LSC junctions from around position f to position d.

A comparison of the downstream non-coding or spacer sequences of the *rps19* genes in monocots reveals that the sequences do not have a common origin (Fig. 2B), as they are highly variable and a reliable sequence alignment is impossible except between closely related con-ordinal taxa (e.g. Zingiberales and Asparagales). This indicates that these spacer sequences had diverse origins and are likely to have resulted from independent DSB events occurring at different points within the IRs.

In contrast, it appears that expansion of IR-LSC junctions did not parallel the evolutionary diversification of basal

angiosperms and eudicot lineages (Fig. 1A). In type I (Fig. 1), IR expansion downstream of *rps19* is extremely rare in eudicots, with the exception of Adzuki bean (Perry et al. [18]) and a *Pelargonium* species (Palmer et al. [16], Chumley et al. [11]). According to our hypothesis A (Fig. 4), the scenario of IR expansion in these two eudicots may have different origins from those of monocots and other eudicots (i.e. type II, Fig. 1), with IRs that have expanded downstream of *rps19* genes. Similarly, significant IR contractions in the basal angiosperm *Illicium oligandrum* (about 1 kb), coriander (4 kb) [13,14], and *Cuscuta reflexa* (about 700 bp to 8 kb) [35] seem to be separate events in their respective lineages.

#### Implications of sequences flanking IR-LSC junctions for angiosperm phylogeny

In extant angiosperms, the relationships among the remaining 5 lineages (magnoliids, monocots, eudicots, Chloranthaceae and *Ceratophyllum*) are unresolved [19,26,27]. To what extent the dicot lineage is a sister group of monocots remains uncertain, probably a reflection of the rapid radiation and extinction of early angiosperms soon after they originated [36,37].

Recent phylogenetic analyses based on plastid sequence data have suggested that monocots and eudicots are sister taxa (Graham et al. [38] and Cai et al. [39]), but with low bootstrap support (67% and 72%, respectively). In addition, several lines of evidence have indicated that Ceratophyllaceae could be the sister group of monocots [40-44].

Here we present an alternative view on this issue. As illustrated in Figure 1, an intact *trnH* is duplicated in IR<sub>B</sub> of all monocots, one basal angiosperm (*Nuphar advena*, position c'), and two winteraceae magnoliid species (*Zygogynum paucifolium* and *Drimys granadensis*, position d) [29]. Sequence comparison revealed that only Winteraceae and monocots have highly similar spacer sequences between the *rpl2* and *trnH* genes (Fig. 5B), suggesting that duplication of *trnH* gene in IR<sub>B</sub> of the two taxa might be common or similar (viz. convergent). On the other hand, Acorales (the most basal lineage in monocots, [27]) has its IR endpoint at position d, suggesting that those lineages with IR-LSC junctions at position b and c' (most Alismatales and Dioscoreales) might have resulted from separate, independent contractions. Our alternative view on the relationships among monocots and their relatives is preliminary, as it is only based on comparison of genic organizations at IR-LSC junctions. Additional molecular and morphological data are required to improve our understanding of monocot phylogeny.

#### The presence of two anti-sense strands of *trnH* in monocot IRs is mysterious

The presence of a *trnH-rps19* cluster in the IRs appears to be a common feature in monocots other than some Alismatids (additional file 1, Fig. 1), in which IR-LSC junctions are located at position b and strongly resemble those of most non-monocot angiosperms. However, alignment of the intergenic spacers between *rpl2* and *trnH* in some Alismatales (e.g. *Alocasia odora*) and other monocots, basal angiosperms and eudicots (Fig. 5) reveals that sequences of the Alismatids are more similar to other monocots than to non-monocot angiosperms. This implies that IR expansions in some Alismatids might share evolutionary scenarios similar to those proposed for other monocots, and that the short IRs (or IR contraction) in some other Alismatids are likely due to either an early termination of the repair-extension reaction after the first DSB in step 1 of hypothesis A (Fig. 4), or to a contraction after this step.

In monocots, each IR usually contains a *trnH* gene, while in most basal angiosperms and eudicots the gene is rarely present in IR<sub>B</sub> (see Fig. 1A: type I and type II). Why is the duplicated *trnH* gene able to survive in IR<sub>B</sub> of most monocots but is absent, degraded or truncated in most non-monocot angiosperms? In two studied eudicots, *Lotus japonicus* [18] and *Spinacea oleracea* [45], the transcriptional activity of *S10<sub>A</sub>* dropped significantly because of either the high transcription levels of the *psbA* and *trnH* genes or the termination of *S10<sub>A</sub>* proximal to J<sub>LA</sub> [32]. Therefore, in non-monocot angiosperms, *trnH*-encoded mRNA molecules constitute only one sense strand, transcribed solely by the *psbA* operon rather than by the *S10<sub>A</sub>* operon. Because anti-sense RNA molecules may interfere with the normal function of the sense RNA molecules [32], in monocots the mechanism by which anti-sense *trnH* is regulated by two *S10<sub>A</sub>* promoters is mysterious. Further study on the evolution and survival of the duplicated *trnH* gene in IR<sub>B</sub> of monocots is desirable.

#### Conclusion

Extensive comparisons of the genic organizations flanking the IR-LSC junctions in 123 diversified angiosperm lineages revealed that monocots and non-monocot angiosperms generally have different IR-LSC junction types. Notably, IRs expanded more progressively in monocots than in non-monocot angiosperms, with more LSC genes being converted into IRs. With the exceptions of Alismatales and a few Acorales, the monocot IR<sub>A</sub> regions either encompass a *trnH-rps19* cluster or extend as far as the 5' portion of the *rpl22* gene, which is typically situated at the LSC region in non-monocot angiosperms. Various expansions of IRs in monocots have resulted in corresponding fluxes of IR-LSC junctions. Our results further indicate that the IR expansions in angiosperms can be

explained by initiation of a DSB event and ending at a polyA tract region.

We proposed two hypotheses to explain the evolutionary derivation of the *trnH-rps19* cluster in the IRs of monocots from an *Amborella*-like ancestor (Fig. 4). Hypothesis A proposes that a DSB event occurs first within the IR<sub>B</sub> of an *Amborella*-like ancestor, and then the free 3' end of the broken strand is repaired against the homologous sequence in IR<sub>A</sub>. The repaired sequence extends and results in the duplication of a *trnH* gene in the newly repaired IR<sub>B</sub>. A subsequent DSB event may occur in IR<sub>A</sub> so that the *rps19* at IR<sub>A</sub> is duplicated, whereby a *trnH-rps19* cluster is created. Hypothesis B assumes that *rps19* is duplicated or converted before the duplication of *trnH* via a DSB event that occurs at IR<sub>A</sub>.

It is worth noting that IR expansions in monocots appear to correlate well with the divergence pattern of monocot phylogeny. The present study highlights the use of sequences flanking the IR-LSC junctions to address the evolutionary dynamics of IRs from basal angiosperms to monocots. Taken together with the evidence from the IR-LSC junctions, we conclude that (i) monocots may be closely related to the Winteraceae (magnoliids) than to other basal angiosperms or eudicots, (ii) the shorter IRs in Alismatids are probably due to either an early termination of repair-extension after the first DSB, or to a contraction after this step, and (iii) the duplicated *trnH* genes in the IR<sub>B</sub> of most monocots and non-monocot angiosperms have distinct fates, which are likely regulated by different expression levels of *S10<sub>A</sub>* and *S10<sub>B</sub>* promoters. Further study is needed to unravel the evolutionary significance or advantage of the presence of an additional *trnH* in monocot IRs, and of IR expansion in more recently diverged monocots.

## Methods

### Plant materials and DNA preparation

Species sampled in this study were listed in the additional file 1. Total cellular DNA was extracted using the method of Saghai-Marooof et al. [46]. The extracted DNAs were used directly for PCR amplification.

### PCR amplification

Primer design was based on published sequence data for conserved regions flanking the IR-LSC junctions. The J<sub>LA</sub> regions were amplified with the primer pair *rpl2-psbA-F3* and *rpl2-psbA-R2*, which correspond to the 3' end of *rpl2* and the 5' end of *psbA* respectively (Fig. 1). The J<sub>LB</sub> region was amplified using two forward primers, *rps3-F1* and *rps3-F2*, that respectively pair with a reverse primer *rps3-rpl2-R2*. The sequences of these primers are listed in Table 1. Amplicons were cleaned using the Gel Extraction System (Viogene, Taipei) and cloned into a pGEM T-Easy vector (Promega, Fitchsburg). Plasmid DNAs were purified using the Plasmid DNA Miniprep System (Viogene) and sequenced on an ABI 3730 automated sequencer (Applied Biosystems, Foster City). For each species two independent PCR clones were sequenced. Sequence alignments were made using GeneDoc (Ver. 2.6.02.)

### Reverse Transcriptase-Polymerase Chain Reaction (RT-PCR) Assay

To verify the transcription of *trnH-rps19* that flanks the IR<sub>A</sub> region, total RNAs were extracted and purified by RNeasy<sup>®</sup> Plant Mini Kit (Qiagen, Hilden). The resulting RNAs were reversely transcribed to synthesize cDNA with Superscript II reverse transcriptase (Invitrogen, Indianapolis) and a specific primer (either *trnH-psbA-F1* or *trnH-rev*), according to the manufacturer's protocol. The two synthesized cDNAs were then used with the primer pair *trnH-psbA-F1* and *rpl2-psbA-R2* to amplify a 674 bp fragment, and the primer pair *trnH-rev* and *rpl2-psbA-F3* to amplify a 298 bp fragment. Each of the two reactions was conducted under

**Table 1: Primers used for analyses of IR-LSC junctions and in RT-PCR**

Primer number	Name	Sequence	Application
1	<i>rpl2-psbA-F1</i>	5'-GACCCTAATCGAAATGCRTMCATTTG-3'	IR <sub>A</sub>
2	<i>rpl2-psbA-F2</i>	5'-TAATTGGAGATACATTKKTTCTGGTACA-3'	IR <sub>A</sub>
3	<i>rpl2-psbA-R1</i>	5'-ATGGCDTTC AAYYTR AAYGGMTTYAATTT-3'	IR <sub>A</sub>
4	<i>rpl2-psbA-R2</i>	5'-CTTGGTATGGARGTMATGCAYGARGCTAA-3'	IR <sub>A</sub>
5	<i>rps3-rpl2-F1</i>	5'-GYTAA YTCRATRRCTTTTTTCATTGC-3'	IR <sub>B</sub>
6	<i>rps3-rpl2-F2</i>	5'-AWABYYKTTGGTTKTMRAACCA AA-3'	IR <sub>B</sub>
7	<i>rps3-rpl2-R1</i>	5'-AATGGGAAATGCCCTACCTTTG-3'	IR <sub>B</sub>
8	<i>rps3-rpl2-R2</i>	5'-GTAGTAAGAGGRGTRGTTATGAACCC-3'	IR <sub>B</sub>
9	<i>rpl2-F1</i>	5'-TRRTTATTTCBGCAGCVGCAATGC-3'	IR <sub>B</sub>
10	<i>rps3-F1</i>	5'-ATAWATTTCYGAAGAATRTTAGG-3'	IR <sub>B</sub>
11	<i>rps3-F2</i>	5'-AGTCKGAAACCRAGTGGATTT-3'	IR <sub>B</sub>
12	<i>rpl2-psbA-F3</i>	5'-GGTAARCGYCCYGTAGTAAGAGG-3'	IR <sub>A</sub>
13	<i>trnH-psbA-F1</i>	5'-GGCGAACGACGGGAATTGAAC-3'	IR <sub>A</sub>
14	<i>trnH-rev</i>	5'-GGATGTAGCCAAGTGGATCAAGG-3'	IR <sub>A</sub>

the following conditions: 94°C for 5 min, followed by 30 cycles of 94°C for 30s, 55°C for 30s, and 72°C for 30s, and ending with an extension of 72°C for 10 min.

### Abbreviations

cpDNA, chloroplast genome; IR, inverted repeat; SSC, small single copy; LSC, large single copy; bp, base pair; J<sub>LA</sub>, junction between LSC and IR<sub>A</sub>; J<sub>LB</sub>, junction between LSC and IR<sub>B</sub>; DSB, double-strand break; RT-PCR: reverse transcriptase-polymerase chain reaction.

### Authors' contributions

SMC conceived the study. CLC, CLW, TMS and RJW carried out the sequence analysis, and CCC provided the unpublished orchid data. CLC and CLW prepared the sequence data and submitted it to GenBank. CLC prepared the figures. RJW, SMC, and CLC wrote the manuscript. All authors read and approved the final manuscript.

### Additional material

#### Additional file 1

Studied taxa and their GenBank accession numbers, references and IR-LSC junction positions. This table (Table S1) provides detailed information about the studied 123 taxa, including 12 basal angiosperms, 16 magnoliids, 62 eudicots, and 33 monocots, involved in the analysis.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-8-36-S1.PDF>]

### Acknowledgements

This work was supported by a research grant from the Research Center for Biodiversity, Academia Sinica, to SMC, and in part by a grant from Guangzhou Forestry Administration to RJW. We thank Yin-Long Qiu for providing DNA of some basal angiosperms, and the staff of the RBG Kew DNA Bank for some plant genomic DNA materials. We gratefully acknowledge the critical reading of the manuscript by Pablo Bolanos-Villegas and Yu-Ting Lai and the valuable comments by three anonymous reviewers.

### References

- Kolodner R, Tewari KK: **Inverted repeats in chloroplast DNA from higher plants.** *Proc Natl Acad Sci USA* 1979, **76**:41-45.
- Palmer JD: **Comparative organization of chloroplast genomes.** *Annu Rev Genet* 1985, **19**:325-354.
- Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M: **The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression.** *EMBO J* 1986, **5**:2043-2049.
- Sugiura M: **The chloroplast chromosomes in land plants.** *Annu Rev Cell Biol* 1989, **5**:51-70.
- Kanno A, Hirai A: **A transcription map of the chloroplast genome from rice (*Oryza sativa*).** *Curr Genet* 1993, **23**:166-174.
- Palmer JD, Osorio B, Thompson WF: **Evolutionary significance of inversions in legume chloroplast DNAs.** *Curr Genet* 1988, **14**:65-74.
- Woodbury NW, Roberts LL, Palmer JD, Thompson WF: **A transcription map of the pea chloroplast genome.** *Curr Genet* 1988, **14**:75-89.
- Raubeson LA, Jansen RK: **Chloroplast genomes of plants.** In *Plant diversity and evolution: genotypic and phenotypic variation in higher plants* Edited by: Henry RJ. Wallingford: CABI Publishing; 2005:45-68.
- Maier RM, Neckermann K, Igloi GL, Kössel H: **Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing.** *J Mol Biol* 1995, **251**:614-628.
- Sugiura M: **The chloroplast genome.** *Plt Mol Biol* 1992, **19**:149-168.
- 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**:2175-2190.
- Palmer JD, Stein DB: **Conservation of chloroplast genome structure among vascular plants.** *Curr Genet* 1986, **10**:823-833.
- Hansen DR, Dastidar SG, Cai Z, Penafior C, Kuehl JV, Boore JL, Jansen RK: **Phylogenetic and evolutionary implications of complete chloroplast genome sequences of four early-diverging angiosperms: *Buxus* (Buxaceae), *Chloranthus* (Chloranthaceae), *Dioscorea* (Dioscoreaceae), and *Illicium* (Schisandraceae).** *Mol Phylogenet Evol* 2007, **45**:547-563.
- Plunkett GM, Downie SR: **Expansion and contraction of the chloroplast inverted repeat in Apiaceae subfamily Apioideae.** *Syst Bot* 2000, **25**:648-667.
- Goulding SE, Olmstead RG, Morden CW, Wolfe KH: **Ebb and flow of the chloroplast inverted repeat.** *Mol Gen Genet* 1996, **252**:195-206.
- Palmer JD, Nugent JM, Herbon LA: **Unusual structure of geranium chloroplast DNA: a triple-sized inverted repeat, extensive gene duplications, multiple inversions, and two repeat families.** *Proc Natl Acad Sci USA* 1987, **84**:769-773.
- Aii J, Kishima Y, Mikami T, Adachi T: **Expansion of the IR in the chloroplast genomes of buckwheat species is due to incorporation of an SSC sequence that could be mediated by an inversion.** *Curr Genet* 1997, **31**:276-279.
- Perry AS, Brennan S, Murphy DJ, Kavanagh TA, Wolfe KH: **Evolutionary re-organisation of a large operon in Adzuki bean chloroplast DNA caused by inverted repeat movement.** *DNA Res* 2002, **9**:157-162.
- APGII: **An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG II.** *Bot J Linn Soc* 2003, **141**:399-436.
- Chang C-C, Lin H-C, Lin I-P, Chow T-Y, Chen H-H, Chen W-H, Cheng C-H, Lin C-Y, Liu S-M, Chang C-C, Chaw S-M: **The chloroplast genome of *Phalaenopsis aphrodite* (Orchidaceae): comparative analysis of evolutionary rate with that of grasses and its phylogenetic implications.** *Mol Biol Evol* 2006, **23**:279-291.
- Leebens-Mack J, Raubeson LA, Cui LY, Kuehl JV, Fourcade MH, Chumley TW, Boore JL, Jansen RK, dePamphilis CW: **Identifying the basal angiosperm node in chloroplast genome phylogenies: sampling one's way out of the Felsenstein zone.** *Mol Biol Evol* 2005, **22**:1948-1963.
- Mathews S, Donoghue MJ: **The root of angiosperm phylogeny inferred from duplicate phytochrome genes.** *Science* 1999, **286**:947-950.
- Qiu YL, Dombrowska O, Lee J, Li L, Whitlock BA, Bernasconi-Quadroni F, Rest JS, Davis CC, Borsch T, Hilu KW, Renner SS, Soltis DE, Soltis PS, Zanis MJ, Cannone JJ, Gutell RR, Powell M, Savolainen V, Chatrou LW, Chase MW: **Phylogenetic analyses of basal angiosperms based on nine plastid, mitochondrial, and nuclear genes.** *Int J Plt Sci* 2005, **166**:815-842.
- Savolainen V, Chase MW, Hoot SB, Morton CM, Soltis DE, Bayer C, Fay MF, de Bruijn AY, Sullivan S, Qiu YL: **Phylogenetics of flowering plants based on combined analysis of plastid *atpB* and *rbcl* gene sequences.** *Syst Biol* 2000, **49**:306-362.
- Soltis DE, Soltis PS: ***Amborella* not a "basal angiosperm"? not so fast.** *Amer J Bot* 2004, **91**:997-1001.
- Qiu YL, Li L, Hendry TA, Li R, Taylor DW, Issa MJ, Ronen AJ, Vekaria ML, White AM: **Reconstructing the basal angiosperm phylogeny: evaluation information content of mitochondrial genes.** *Taxon* 2006, **55**:837-856.

27. Soltis DS, Soltis PS, Chase MW: *Phylogeny and evolution of angiosperms* Sunderland, MA: Sinauer Associates, Inc; 2005.
28. Kim Y-D, Jansen RK: **Characterization and phylogenetic distribution of a chloroplast DNA rearrangement in the Berberidaceae.** *Plt Syst Evol* 1994, **193**:107-114.
29. Goremykin V, Hirsch-Ernst KI, Wölfl S, Hellwig FH: **Analysis of the *Amborella trichopoda* chloroplast genome sequence suggests that *Amborella* is not a basal angiosperm.** *Mol Biol Evol* 2003, **20**:1499-1505.
30. Goremykin V, Hirsch-Ernst KI, Wölfl S, Hellwig FH: **The chloroplast genome of the "basal" angiosperm *Calycanthus fertilis* – structural and phylogenetic analyses.** *Plant Syst Evol* 2003, **242**:119-135.
31. Jansen RK, Kaittinen C, Saski C, Lee S-B, Tomkins J, Alverson AJ, Daniell H: **Phylogenetic analyses of *Vitis* (Vitaceae) based on complete chloroplast genome sequences: effects of taxon sampling and phylogenetic methods on resolving relationships among rosids.** *BMC Evol Biol* 2006, **6**:32.
32. Tonkyn JC, Gruissem W: **Differential expression of the partially duplicated chloroplast S10 ribosomal protein operon.** *Mol Gen Genet* 1993, **241**:141-152.
33. Narayanan V, Mieczkowski PA, Kim H-M, Petes TD, Lobachev KS: **The pattern of gene amplification is determined by the chromosomal location of hairpin-capped breaks.** *Cell* 2006, **125**:1283-1296.
34. Soltis PS, Soltis DE, Chase MW: **Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology.** *Nature* 1999, **402**:402-404.
35. Bömmer D, Haberhausen G, Zetsche K: **A large deletion in the plastid DNA of the holoparasitic flowering plant *Cuscuta reflexa* concerning two ribosomal proteins (*rpl2*, *rpl23*), one transfer RNA (*trnI*) and an ORF 2280 homologue.** *Curr Genet* 1993, **24**(1-2):171-176.
36. Friis E, Pedersen K, Crane PR: **Reproductive structure and organization of basal angiosperms from the early Cretaceous (Barremian or Aptian) of western Portugal.** *Int J Plt Sci* 2000, **161**:S169-S182.
37. Friis EM, Pedersen KR, Crane PR: **Early angiosperm diversification: the diversity of pollen associated with angiosperm reproductive structures in early Cretaceous floras from Portugal.** *Ann Missouri Bot Gard* 1999, **86**:259-296.
38. Graham SW, Zgurski JM, McPherson MA, Cherniawsky DM, M. SJ, Horne ESC, Smith SY, Wong WA, O'Brien HE, Biron VL, Pires JC, Olmstead RG, Chase MW, Rai HS: **Robust inference of monocot deep phylogeny using an expanded multigene plastid data set.** In *Monocots: comparative biology and evolution Volume 1*. Edited by: Columbus JT, Friar EA, Hamilton CW, Porter JM, Prince LM, Simpson MG. Claremont: Rancho Santa Ana Botanic Garden; 2006:3-20.
39. Cai Z, Penafior C, Kuehl JV, Leebens-Mack J, Carlson JE, dePamphilis CW, Boore JL, Jansen RK: **Complete plastid genome sequences of *Drimys*, *Liriodendron*, and *Piper*: implications for the phylogenetic relationships of magnoliids.** *BMC Evol Biol* 2006, **6**:77.
40. Qiu YL, Lee J, Bernasconi-Quadroni F, Soltis DE, Soltis PS, Zanis M, Zimmer EA, Chen ZD, Savolainen V, Chase MW: **Phylogeny of basal angiosperms: analyses of five genes from three genomes.** *Int J Plt Sci* 2000, **161**:S3-S27.
41. Qiu YL, Lee J, Bernasconi-Quadroni F, Soltis DE, Soltis PS, Zanis M, Zimmer EA, Chen Z, Savolainen V, Chase MW: **The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes.** *Nature* 1999, **402**:404-407.
42. Soltis DE, Soltis PS, Chase MW, Mort ME, Albach DC, Zanis M, Savolainen V, Hahn WH, Hoot SB, Fay MF, Axtell M, Swensen SM, Prince LM, Kress WJ, Nixon KC, Farris JS: **Angiosperm phylogeny inferred from 18S rDNA, *rbcL*, and *atpB* sequences.** *Bot J Linn Soc* 2000, **133**:381-461.
43. Zanis M, Soltis DE, Soltis PS, Mathews S, Donoghue MJ: **The root of the angiosperms revisited.** *Proc Natl Acad Sci USA* 2002, **99**:6848-6853.
44. Zanis MJ, Soltis PS, Qiu YL, Zimmer E, Soltis DE: **Phylogenetic analyses and perianth evolution in basal angiosperms.** *Ann Missouri Bot Gard* 2003, **90**:129-150.
45. Zurawski G, Bottomley W, Whitfield PR: **Junctions of the large single copy region and the inverted repeats in *Spinacia oleracea* and *Nicotiana debneyi* chloroplast DNA: sequence of the genes for tRNA<sup>His</sup> and the ribosomal proteins S19 and L2.** *Nucl Acid Res* 1984, **12**(16):6547-6558.
46. Saghai-Marooof MA, Soliman KM, Jorgensen RA, Allard RW: **Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location and population dynamics.** *Proc Natl Acad Sci USA* 1984, **81**:8014-8018.
47. Goremykin VV, Hirsch-Ernst KI, Wölfl S, Hellwig FH: **The chloroplast genome of *Nymphaea alba*: whole-genome analyses and the problem of identifying the most basal angiosperm.** *Mol Biol Evol* 2004, **21**:1445-1454.
48. Raubeson LA, Peery R, Chumley TW, Dziubek C, Fourcade HM, Boore JL, Jansen RK: **Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*.** *BMC Genomics* 2007, **8**:174.
49. Moore MJ, Bell CD, Soltis PS, Soltis DE: **Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms.** *Proc Natl Acad Sci USA* 2007, **104**:19363-19368.
50. Moore MJ, Dhingra A, Soltis PS, Shaw R, Farmerie WG, Folta KM, Soltis DE: **Rapid and accurate pyrosequencing of angiosperm plastid genomes.** *BMC Plt Biol* 2006, **6**:17.
51. Schmitz-Linneweber C, Maier RM, Alcaraz J-P, Cottet A, Herrmann RG, Mache R: **The plastid chromosome of spinach (*Spinacia oleracea*): complete nucleotide sequence and gene organization.** *Plant Mol Biol* 2001, **45**(3):307-315.
52. Kim K-J, Lee HL: **Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants.** *DNA Res* 2004, **11**:247-261.
53. Ruhlman T, Lee S-B, Jansen RK, Hostetler JB, Tallon LJ, Town CD, Daniell H: **Complete plastid genome sequence of *Daucus carota*: implications for biotechnology and phylogeny of angiosperms.** *BMC Genomics* 2006, **7**:222.
54. Samson N, Bausher MG, Lee S-B, Jansen RK, Daniell H: **The complete nucleotide sequence of the coffee (*Coffea arabica* L.) chloroplast genome: organization and implications for biotechnology and phylogenetic relationships amongst angiosperms.** *Plant Biotechnology Journal* 2007, **5**:339-353.
55. Wolfe KH, Morden CW, Palmer JD: **Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant.** *Proc Natl Acad Sci USA* 1992, **89**:10648-10652.
56. Lee H-L, Jansen RK, Chumley TW, Kim K-J: **Gene relocations within chloroplast genomes of *Jasminum* and *Menodora* (Oleaceae) are due to multiple, overlapping inversions.** *Mol Biol Evol* 2007, **24**:1161-1180.
57. Schmitz-Linneweber C, Regel R, Du TG, Hupfer H, Herrmann RG, Maier RM: **The plastid chromosome of *Atrapa belladonna* and its comparison with that of *Nicotiana tabacum*: the role of RNA editing in generating divergence in the process of plant speciation.** *Mol Biol Evol* 2002, **19**:1602-1612.
58. Yukawa M, Tsudzuki T, Sugiura M: **The chloroplast genome of *Nicotiana sylvestris* and *Nicotiana tomentosiformis*: complete sequencing confirms that the *Nicotiana sylvestris* progenitor is the maternal genome donor of *Nicotiana tabacum*.** *Mol Genet Genomics* 2006, **275**:367-373.
59. Aldrich J, Cherney BV, Williams C, Merlin E: **Sequence analysis of the junction of the large single copy region and the large inverted repeat in the petunia chloroplast genome.** *Curr Genet* 1988, **14**:487-492.
60. Kahlau S, Aspinall S, Gray JC, Bock R: **Sequence of the tomato chloroplast DNA and evolutionary comparison of solanaceous plastid genomes.** *J Mol Evol* 2006, **63**:194-207.
61. Hupfer H, Swiatek M, Hornung S, Herrmann RG, Maier RM, Chiu WL, Sear B: **Complete nucleotide sequence of the *Oenothera elata* plastid chromosome, representing plastome I of the five distinguishable euoenothera plastomes.** *Mol Gen Genet* 2000, **263**:581-585.
62. Steane DA, Jones RC, Vaillancourt RE: **A set of chloroplast microsatellite primers for *Eucalyptus* (Myrtaceae).** *Mol Ecol Notes* 2005, **5**:538-541.
63. Sato S, Nakamura Y, Kaneko T, Asamizu E, Tabata S: **Complete structure of the chloroplast genome of *Arabidopsis thaliana*.** *DNA Res* 1999, **6**:283-290.
64. Nickelsen J, Link G: **Nucleotide sequence of the mustard chloroplast genes *trnH* and *rps19*'.** *Nucleic Acids Res* 1990, **18**:1051.
65. Bausher MG, Singh ND, Lee S-B, Jansen RK, Daniell H: **The complete chloroplast genome sequence of *Citrus sinensis* (L.) Osbeck var 'Ridge Pineapple': organization and phylogenetic relationships to other angiosperms.** *BMC Plt Biol* 2006, **6**:21.

66. Ibrahim RIH, Azuma J-I, Sakamoto M: **Complete nucleotide sequence of the cotton (*Gossypium barbadense* L.) chloroplast genome with a comparative analysis of sequences among 9 dicot plants.** *Genes Genet Syst* 2006, **81**:311-321.
67. Lee SB, Kaittanis C, Jansen RK, Hostetler JB, Tallon LJ, Town CD, Daniell H: **The complete chloroplast genome sequence of *Gossypium hirsutum*: organization and phylogenetic relationships to other angiosperms.** *BMC Genomics* 2006, **7**:61.
68. Spielmann A, Roux E, von Allmen J-M, Stutz E: **The soybean chloroplast genome: complete sequence of the *rps19* gene, including flanking parts containing exon 2 or *rpl2* (upstream), but lacking *rpl22* (downstream).** *Nucl Acids Res* 1988, **16**:1199.
69. Sasaki C, S-B L, Daniell H, Wood TC, Tomkins J, Kim HG, Jansen RK: **Complete chloroplast genome sequence of *Glycine max* and comparative analyses with other legume genomes.** *Plt Mol Biol* 2005, **59**(2):309-322.
70. Kato T, Kaneko T, Sato S, Nakamura Y, Tabata S: **Complete structure of the chloroplast genome of a legume, *Lotus japonicus*.** *DNA Res* 2000, **7**:323-330.
71. Ravi V, Khurana JP, Tyagi AK, Khurana P: **The chloroplast genome of mulberry: complete nucleotide sequence, gene organization and comparative analysis.** *Tree Genet Genomes* 2006, **3**:49-59.
72. Goremykin VV, Holland B, Hirsch-Ernst KI, Hellwig FH: **Analysis of *Acorus calamus* chloroplast genome and its phylogenetic implications.** *Mol Biol Evol* 2005, **22**:1813-1822.
73. Masooda MS, Nishikawaa T, Fukuokaa S-I, Njengaa PK, Tsudzukib T, Kadowakia K-I: **The complete nucleotide sequence of wild rice (*Oryza nivara*) chloroplast genome: first genome wide comparative sequence analysis of wild and cultivated rice.** *Gene* 2004, **340**:133-139.
74. Hiratsuka J, Shimada H, Whittier R, Ishibashi T, Sakamoto M, Mori M, Kondo C, Honji Y, Sun C-R, Meng B-Y, Li Y-Q, Kanno A, Nishizawa Y, Hirai A, Shinozaki K, Sugiura M: **The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals.** *Mol Gen Genet* 1989, **217**:185-194.
75. Asano T, Tsudzuki T, Takahashi S, Ssimada H, Kadowaki K: **Complete nucleotide sequence of the sugarcane (*Saccharum officinarum*) chloroplast genome: a comparative analysis of four monocot chloroplast genomes.** *DNA Res* 2004, **11**:93-99.
76. Ogiwara Y, Isono K, Kojima T, Endo A, Hanaoka M, Shiina T, Terachi T, Utsugi S, Murata M, Mori N, Takumi S, Ikeo K, Gojobori T, Murai R, Murai K, Matsuoka Y, Ohnishi Y, Tajiri H, Tsunewaki K: **Structural features of a wheat plastome as revealed by complete sequencing of chloroplast DNA.** *Mol Genet Genomics* 2002, **266**:740-746.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

