University of Nebraska - Lincoln
DigitalCommons@University of Nebraska - Lincoln

Faculty Publications from the Center for Plant Science Innovation

Plant Science Innovation, Center for

2016

Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates

Andan Zhu

Wenhu Guo

Weishu Fan

Jeffrey P. Mower

Follow this and additional works at: http://digitalcommons.unl.edu/plantscifacpub

Part of the Plant Biology Commons, Plant Breeding and Genetics Commons, and the Plant Pathology Commons

This Article is brought to you for free and open access by the Plant Science Innovation, Center for at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications from the Center for Plant Science Innovation by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates

Andan Zhu[1,2], Wenhu Guo[1,3], Sakshi Gupta[1], Weishu Fan[1,2] and Jeffrey P. Mower[1,2]

[1]Center for Plant Science Innovation, University of Nebraska, Lincoln, NE 68588, USA; [2]Department of Agronomy and Horticulture, University of Nebraska, Lincoln, NE 68583, USA; [3]School of Biological Sciences, University of Nebraska, Lincoln, NE 68588, USA

Author for correspondence:
Jeffrey P. Mower
Tel: +1 402 472 2130
Email: jpmower@unl.edu

Received: *19 June 2015*
Accepted: *1 October 2015*

New Phytologist (2016) **209:** 1747–1756
doi: 10.1111/nph.13743

Key words: inverted repeat (IR), localized hypermutation, plastid genome (plastome), recombination-mediated repair, synonymous substitution rate, vascular plants.

Summary

- Rates of nucleotide substitution were previously shown to be several times slower in the plastid inverted repeat (IR) compared with single-copy (SC) regions, suggesting that the IR provides enhanced copy-correction activity.
- To examine the generality of this synonymous rate dependence on the IR, we compared plastomes from 69 pairs of closely related species representing 52 families of angiosperms, gymnosperms, and ferns.
- We explored the breadth of IR boundary shifts in land plants and demonstrate that synonymous substitution rates are, on average, 3.7 times slower in IR genes than in SC genes. In addition, genes moved from the SC into the IR exhibit lower synonymous rates consistent with other IR genes, while genes moved from the IR into the SC exhibit higher rates consistent with other SC genes. Surprisingly, however, several plastid genes from *Pelargonium*, *Plantago*, and *Silene* have highly accelerated synonymous rates despite their IR localization.
- Together, these results provide strong evidence that the duplicative nature of the IR reduces the substitution rate within this region. The anomalously fast-evolving genes in *Pelargonium*, *Plantago*, and *Silene* indicate localized hypermutation, potentially induced by a higher level of error-prone double-strand break repair in these regions, which generates substitutional rate variation.

Introduction

The plastid genome (plastome) of nearly all land plants has a highly conserved quadripartite structure composed of two copies of an inverted repeat (IR) and two single-copy (SC) regions, termed the large single-copy (LSC) and small single-copy (SSC) regions. The land plant IR typically ranges in size from 15 to 30 kb and contains a core set of four rRNA genes (encoding 4.5S, 5S, 16S and 23S rRNA) and five tRNA genes (encoding *trnA*-UGC, *trnI*-GAU, *trnN*-GUU, *trnR*-ACG and *trnV*-GAC). In addition to this core rRNA/tRNA cluster, the IRs of many land plants, particularly vascular plants, also contain a variety of other genes as a result of lineage-specific expansions and contractions. Among more closely related species, these IR boundary shifts tend to be relatively minor, resulting in the gain or loss of a small number of genes (Goulding *et al.*, 1996; Wang *et al.*, 2008; Wicke *et al.*, 2014; Downie & Jansen, 2015; Wu & Chaw, 2015). However, recent large-scale expansions (exceeding several kb) were reported for a few lineages, such as *Pelargonium*, *Psilotum*, and Trochodendraceae (Chumley *et al.*, 2006; Grewe *et al.*, 2013; Sun *et al.*, 2013), which transferred numerous genes from the SC regions into the IR. At the opposite extreme, some plants have lost most, or even all, of the IR, as observed for conifers, many legumes, and some species of *Erodium* (Palmer *et al.*, 1987; Raubeson & Jansen, 1992; Tsudzuki *et al.*, 1992; Guisinger *et al.*, 2011; Guo *et al.*, 2014).

The presence of the IR has a major impact on the rate of plastome sequence evolution. The synonymous, nonsynonymous, and noncoding substitution rates have been shown to be several times lower for the IR relative to the SC regions among several angiosperms (Wolfe *et al.*, 1987; Maier *et al.*, 1995; Gaut, 1998; Perry & Wolfe, 2002; Yamane *et al.*, 2006; Kim *et al.*, 2009; Yi & Kim, 2012; Yi *et al.*, 2012). This pattern of lower IR substitution rates was recently extended to carnivorous plants (Wicke *et al.*, 2014) and outside of angiosperms to cycads (Wu & Chaw, 2015), suggesting that it is a hallmark feature of the IR in the plastome. Similarly, the frequencies of indels between maize and sugarcane and among carnivorous Lentibulariaceae are a few times lower in the IR than in the SC regions (Yamane *et al.*, 2006; Wicke *et al.*, 2014). When the IR becomes lost, however, as in the IR-lacking clade of legumes, the synonymous substitution rate of the former IR genes was shown to increase to a value similar to that of other SC genes, providing strong evidence that the reduced substitution rate is dependent on the duplicative nature of the IR (Perry & Wolfe, 2002). These findings suggest that the depressed substitution rate in the IR is a result of a copy-dependent repair mechanism (Wolfe *et al.*, 1987; Perry & Wolfe, 2002), such as gene conversion that is biased against new

© 2015 The Authors
New Phytologist © 2015 New Phytologist Trust

mutations (Birky & Walsh, 1992). While biased gene conversion can occur throughout the genome via intergenomic interactions, the duplicative nature of the IR provides a twofold higher copy number in the population of genome copies within each plastid, which enables a relatively higher rate of intergenomic gene conversion for the IR and also allows for intragenomic gene conversion between IR copies. Gene conversion activity was demonstrated in plastids using a transformation system (Khakhlova & Bock, 2006) and has been implicated as the mechanism generating small IR expansions and contractions (Goulding *et al.*, 1996).

In addition to regional effects of the IR on mutation rates, several studies have identified additional examples of intragenomic variation in substitution rates that appear to be independent of their IR or nonIR localization. Both synonymous and nonsynonymous rates are substantially higher in several ribosomal protein and RNA polymerase genes for species in Geraniaceae (Guisinger *et al.*, 2008). Similar rate accelerations were observed for ribosomal protein genes, *clpP*, *ycf1* and *ycf2* in *Silene* (Erixon & Oxelman, 2008; Sloan *et al.*, 2012b). In some legumes, a mutational hotspot was observed, affecting the *ycf4* and *psaI* genes (Magee *et al.*, 2010), which was attributed to a hotspot of double-strand breaks and their repair. Localized hypermutation has also been observed in several plant mitochondrial lineages (Mower *et al.*, 2007), including both *Silene* (Sloan *et al.*, 2012a) and *Ajuga* (Zhu *et al.*, 2014).

Although the reduction in IR substitution rates has been consistently demonstrated in several studies, comparisons have been made between relatively few taxa, and nearly all have been limited to angiosperms. With the proliferation of new plastome sequences available today, it is now possible to comprehensively examine the evolutionary effect of the IR on substitution rates. In addition, the abundance of IR boundary shifts (expansion, contraction, loss) in multiple lineages makes it possible to perform parallel, independent analyses to examine the generality of rate variation between IR and SC regions. Furthermore, large-scale shifts of IR boundaries have occurred at different evolutionary depths, allowing both short- and long-term impacts to be investigated. To assess the influence of the IR on plastome substitution rates, we first examined representative species to establish ancestral IR boundaries and subsequent boundary shifts during land plant evolution. Next, we performed parallel analysis of 69 species pairs to establish the evolutionary patterns of substitution rate variation between the SC and IR and to determine the effects of IR boundary shifts on substitution rates of genes that were relocated into or out of the IR. Finally, we looked at potential mechanistic causes for the patterns of rate variation observed among taxa.

Materials and Methods

Plastome sequencing, assembly and annotation

Total genomic DNAs from *Angiopteris angustifolia* C. Presl, *Gnetum gnemon* L., *Plantago maritima* L., and *Plantago media* L. were each isolated from fresh leaf tissue from a single plant using a simplified CTAB protocol (Doyle & Doyle, 1987). The *Angiopteris* and *Gnetum* DNAs were Illumina-sequenced at the Indiana University Center for Genomics and Bioinformatics, as described previously (Guo *et al.*, 2014), generating 6 Gb of 250 bp paired-end reads from an 800 bp library. The two *Plantago* DNAs were Illumina-sequenced at BGI Corp. (Shenzhen, China) from 5 kb mate-pair libraries, generating 7 Gb of 100 bp paired-end reads. Organelle-enriched DNAs from *Acorus gramineus* Sol. ex Aiton, *Ginkgo biloba* L., *Magnolia tripetala* L., and *Pinus strobus* L. were each isolated from leaf tissue from a single plant using differential centrifugation and CTAB extraction and then Illumina-sequenced at BGI Corp. as described previously (Grewe *et al.*, 2013; Zhu *et al.*, 2014), generating 4 Gb of 100 bp paired-end reads from an 800 bp library. All data were assembled with VELVET 1.2.03 (Zerbino & Birney, 2008), annotated with DOGMA (Wyman *et al.*, 2004), and checked for sequence and annotation accuracy using established procedures (Grewe *et al.*, 2013; Guo *et al.*, 2014; Zhu *et al.*, 2014). The annotated genome sequences were deposited in GenBank with accession numbers KJ408574, KP099646–KP099650, KR297244 and KR297245.

Estimation of sequence divergence and repeat content

In addition to the eight newly sequenced plastomes, another 130 plastomes were obtained from GenBank (Supporting Information Table S1). Plastomes were chosen to obtain pairs of closely related species from within the same genus. To increase taxon sampling, additional plastome pairs from species of the same family or from individuals of the same species were also included. This sampling strategy resulted in 69 pairs of closely related plastomes from 52 vascular plant families. For each plastome pair, pairwise synonymous substitution rates were compared between SC genes and IR genes. Individual protein-coding genes were aligned at the protein level using the CLUSTALW2 software (Larkin *et al.*, 2007) and then reverse-translated into codon-based alignments via PAL2NAL v.1.4 (Suyama *et al.*, 2006). A concatenated data set of all IR genes and a second data set of all SC genes were generated with FASconCAT (Kuck & Meusemann, 2010), except that genes located across IR-SC boundaries or genes whose IR or SC localization differed between the taxon pair were excluded. Synonymous rates were estimated for the concatenated SC and IR data sets via KAKS_CALCULATOR 2.0 (Wang *et al.*, 2010) under the GY-HKY substitution model.

To assess whether it was appropriate to combine the LSC and SSC genes into a single data set, synonymous sequence divergence was compared between LSC and SSC genes for 61 of the 69 pairs of species (excluding those pairs lacking an IR or with a highly reduced IR). LSC and SSC divergence values were strongly and significantly correlated ($R^2 = 0.96$; $P < 0.0001$) using a linear regression model ($y = 1.04x + 0.00$), providing justification for combining the LSC and SSC genes into a single SC data set. For those pairs of species that had an unusual pattern of synonymous rate variation in the IR relative to the SC, we calculated synonymous divergence for individual genes using the

© 2015 The Authors
New Phytologist © 2015 New Phytologist Trust

KAKS_CALCULATOR 2.0 and sequence divergence for individual introns using DNASP 5.10 (Librado & Rozas, 2009). These values were then plotted against their corresponding genomic positions.

Results

General features of new plastome sequences

We sequenced and assembled the complete sequences of eight plastomes from four angiosperms (*Acorus gramineus*, *Magnolia tripetala*, *Plantago maritima* and *Plantago media*), three gymnosperms (*Ginkgo biloba*, *Gnetum gnemon* and *Pinus strobus*), and one fern (*Angiopteris angustifolia*). These species were selected to represent distinct evolutionary lineages among vascular plants, to complement closely related genomes available in public sequence databases, and/or because of their distinct properties of IR expansion or contraction. Among the newly sequenced genomes, there is moderate variation in genome size, gene and intron content, guanosine-cytosine (GC) content, the size and frequency of nonIR repeats, and the number of duplicated genes (Table 1), in agreement with known degrees of diversity among euphyllophytes (Wicke *et al.*, 2011; Jansen & Ruhlman, 2012; Wolf & Karol, 2012). With the exception of the two *Plantago* genomes, the newly sequenced genomes are fully syntenic and show minimal sequence divergence in comparison to close relatives from the same genus (Fig. S1). These patterns are consistent with a generally slow rate of sequence and structural evolution of plant plastomes.

In contrast to the conserved evolution of most plant plastomes, the two *Plantago* plastomes exhibit increased levels of sequence and structural divergence. Although the two species diverged only 14 million yr ago (Cho *et al.*, 2004), their genomes have accumulated 5.0% sequence divergence as well as several rearranged segments (Fig. S2). Compared with the ancestral angiosperm genome structure (represented by *Nicotiana tabacum*), both genomes contain inverted repeats that have increased markedly in size to 33.7 kb in *P. maritima* and 38.4 kb in *P. media* (Fig. S3), resulting in the transfer of five former SSC genes into the IR of *P. maritima* and nine former SSC genes into the IR of *P. media* (Table 1). Both genomes have also experienced a large-scale inversion within the expanded IR, spanning 14 kb for *P. media* and 21 kb for *P. maritima* (Fig. S3). The breakpoints are inferred to be at *trnL*-*ndhB* and *trnN*-*trnR* for *P. media*, but at *trnL*-*ndhB* and *ycf1*-*rps15* for *P. maritima*. The *P. maritima* genome has another small-scale inversion associated with the *ycf1* gene (Fig. S3). By contrast, no inversions or gene relocations were found in the SC regions. In addition to IR expansion and genomic rearrangement, the *Plantago* plastomes have accumulated more repeats than most other angiosperms (Table 1). Most of the repeats are < 100 bp in length, although seven repeats in *P. maritima* and five repeats in *P. media* range from 100 to 450 bp. Finally, there is variation in intron content between the two *Plantago* genomes as a result of the loss of the *rpl2* intron and both *clpP* introns from *P. maritima* (Fig. S3).

Inverted repeat expansion, contraction, and loss among land plants

During land plant evolution, there have been multiple instances of IR expansion or contraction that have moved entire genes from the SC regions into the IR or vice versa (Fig. 1). Across land plants, the terminal IR gene adjacent to the SSC region is highly conserved. In most species, the last full-length IR gene at the IR/SSC boundary is *trnN*-GUU, providing strong evidence that this was the ancestral IR/SSC endpoint which has been retained in most lineages. Several minor IR extensions into the SSC have occurred in *Selaginella*, *Psilotum*, gnetophytes, and some angiosperms, but their sporadic distribution and general lack of homology indicate that they were independent events for each lineage. Within gnetophytes, the distinct IR boundaries were proposed to result from a multistep process involving several expansions, inversions, and gene losses (Wu *et al.*, 2009).

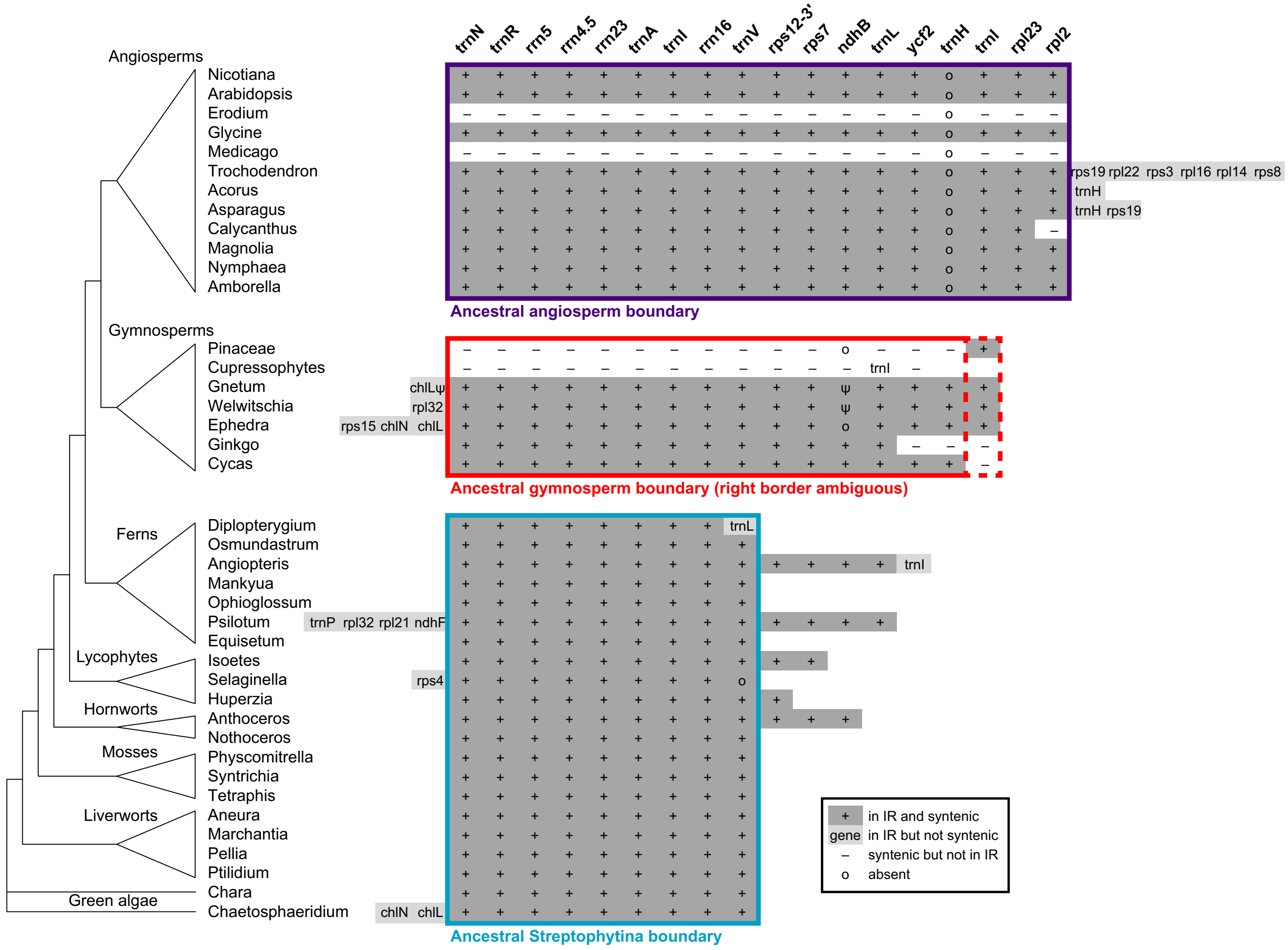
The IR/LSC boundary has shifted more dynamically during land plant evolution (Fig. 1). Excluding seed plants (i.e. angiosperms and gymnosperms), the IR generally terminates at the *trnV*-GAC gene at the IR/LSC boundary. The most

Table 1 General characteristics of vascular plant plastomes

	Ferns		Gymnosperms			Angiosperms			
	Ehye	Aang	Gbil	Ggne	Pstr	Agra	Mtri	Pmar	Pmed
Genome size (bp)	131 760	153 596	157 002	115 022	115 576	152 849	160 037	158 358	164 130
IR size (bp)	10 093	21 676	17 733	20 051	472	25 822	26 572	33 735	38 398
SC size (bp)	111 574	110 244	121 536	74 920	114 632	101 205	106 893	90 888	87 334
GC content (%)	33.7	35.5	39.6	38.2	38.8	38.7	39.3	38.6	38.0
Unique genes	120	122	118	99	108	112	112	113	113
Protein genes in IR	0	3	3	3	0	6	6	11	15
rRNAs in IR	4	4	4	4	0	4	4	4	4
tRNAs in IR	5	8	6	8	1	7	7	7	7
% nonIR repeats	1.8	1.6	1.2	1.4	4.3	0.47	0.93	2.5	2.7

Ehye, *Equisetum hyemale*; Aang, *Angiopteris angustifolia*; Gbil, *Ginkgo biloba*; Ggne, *Gnetum gnemon*; Pstr, *Pinus strobus*; Agra, *Acorus gramineus*; Mtri, *Magnolia tripetala*; Pmar, *Plantago maritima*; Pmed, *Plantago media*; IR, inverted repeat; SC, single-copy.

© 2015 The Authors
New Phytologist © 2015 New Phytologist Trust


Angiosperms
Gymnosperms
Ferns
Lycophytes
Hornworts
Mosses
Liverworts
Green algae
Ancestral angiosperm boundary
Ancestral gymnosperm boundary (right border ambiguous)
Ancestral Streptophytina boundary

		trnN	trnR	rrn5	rrn4.5	rrn23	trnA	trnI	rrn16	trnV	rps12-3'	rps7	ndhB	trnL	ycf2	trnH	trnI	rpl23	rpl2	
Nicotiana		+	+	+	+	+	+	+	+	+	+	+	+	+	+	o	+	+	+	
Arabidopsis		+	+	+	+	+	+	+	+	+	+	+	+	+	+	o	+	+	+	
Erodium		–	–	–	–	–	–	–	–	–	–	–	–	–	–	o	–	–	–	
Glycine		+	+	+	+	+	+	+	+	+	+	+	+	+	+	o	+	+	+	
Medicago		–	–	–	–	–	–	–	–	–	–	–	–	–	–	o	–	–	–	
Trochodendron		+	+	+	+	+	+	+	+	+	+	+	+	+	+	o	+	+	+	rps19 rpl22 rps3 rpl16 rpl14 rps8
Acorus		+	+	+	+	+	+	+	+	+	+	+	+	+	+	o	+	+	+	trnH
Asparagus		+	+	+	+	+	+	+	+	+	+	+	+	+	+	o	+	+	+	trnH rps19
Calycanthus		+	+	+	+	+	+	+	+	+	+	+	+	+	+	o	+	+	–	
Magnolia		+	+	+	+	+	+	+	+	+	+	+	+	+	+	o	+	+	+	
Nymphaea		+	+	+	+	+	+	+	+	+	+	+	+	+	+	o	+	+	+	
Amborella		+	+	+	+	+	+	+	+	+	+	+	+	+	+	o	+	+	+	
Pinaceae		–	–	–	–	–	–	–	–	–	–	–	o	–	–	–	+			
Cupressophytes		–	–	–	–	–	–	–	–	–	–	–	–	trnI	–					
Gnetum	chlLψ	+	+	+	+	+	+	+	+	+	+	+	ψ	+	+	+	+			
Welwitschia	rpl32	+	+	+	+	+	+	+	+	+	+	+	ψ	+	+	+	+			
Ephedra	rps15 chlN chlL	+	+	+	+	+	+	+	+	+	+	+	o	+	+	+	+			
Ginkgo		+	+	+	+	+	+	+	+	+	+	+	+	+	–	–	–			
Cycas		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	–			
Diplopterygium		+	+	+	+	+	+	+	+	trnL										
Osmundastrum		+	+	+	+	+	+	+	+	+										
Angiopteris		+	+	+	+	+	+	+	+	+	+	+	+	+	trnI					
Mankyua		+	+	+	+	+	+	+	+	+										
Ophioglossum		+	+	+	+	+	+	+	+	+										
Psilotum	trnP rpl32 rpl21 ndhF	+	+	+	+	+	+	+	+	+	+	+	+	+						
Equisetum		+	+	+	+	+	+	+	+	+										
Isoetes		+	+	+	+	+	+	+	+	+	+	+								
Selaginella	rps4	+	+	+	+	+	+	+	+	o										
Huperzia		+	+	+	+	+	+	+	+	+	+									
Anthoceros		+	+	+	+	+	+	+	+	+	+	+	+							
Nothoceros		+	+	+	+	+	+	+	+	+										
Physcomitrella		+	+	+	+	+	+	+	+	+										
Syntrichia		+	+	+	+	+	+	+	+	+										
Tetraphis		+	+	+	+	+	+	+	+	+										
Aneura		+	+	+	+	+	+	+	+	+										
Marchantia		+	+	+	+	+	+	+	+	+										
Pellia		+	+	+	+	+	+	+	+	+										
Ptilidium		+	+	+	+	+	+	+	+	+										
Chara		+	+	+	+	+	+	+	+	+										
Chaetosphaeridium	chlN chlL	+	+	+	+	+	+	+	+	+										

+	in IR and syntenic
gene	in IR but not syntenic
–	syntenic but not in IR
o	absent

Fig. 1 Inference of ancestral inverted repeat (IR) content during land plant evolution. Genes which are only partially duplicated in the IR are not shown. Genomes with highly rearranged IR content (e.g. from *Plantago*, *Pelargonium*, *Silene*, and most leptosporangiate ferns) were not included because these lineage-specific changes have no bearing on ancestral reconstruction.

parsimonious interpretation is that the *trnV*-GAC gene represents the ancestral IR/LSC endpoint among land plants, with several independent expansions in the hornwort *Anthoceros*, the lycophytes *Isoetes* and *Huperzia*, the ferns *Psilotum* and *Angiopteris*, and the common ancestor of angiosperms and gymnosperms. This scenario of independent expansions is further supported by the observation that the IR expanded to different endpoints among these land plant lineages. However, more complicated scenarios involving multiple expansions and contractions cannot be excluded. Within ferns, for example, it is only slightly less parsimonious to propose an ancestral expansion to *trnL*-CAA in the common ancestor of all ferns followed by independent contractions back to *trnV*-GAC in *Equisetum*, ophioglossoid ferns (*Ophioglossum* and *Mankyua*), and early diverging leptosporangiate ferns (*Diplopterygium* and *Osmundastrum*).

In addition to these IR boundary shifts, there are a few cases where the IR has been severely reduced or even eliminated (Fig. 1), as previously described for several legumes (Palmer *et al.*, 1987), some species of *Erodium* (Guisinger *et al.*, 2011), cupressophytes (Guo *et al.*, 2014), and Pinaceae (Tsudzuki *et al.*, 1992; Wu *et al.*, 2011).

Lower substitution rates in the IR are consistent with copy-dependent repair activity

To comprehensively examine the effect of the IR on plastome substitution rates, we used 69 pairs of closely related taxa (within the same family, genus, or species) from angiosperms, gymnosperms, and ferns to compare synonymous sequence divergence (d_S) of a concatenated set of genes in the IR and SC regions (Fig. 2). In nearly all species, d_S was markedly higher for SC genes than for IR genes. Linear regression showed a tight and significant correlation ($R^2 = 0.93$; $n = 54$; $P < 0.001$) between d_S values in the IR and the SC region, and the line of best fit indicated that d_S was 3.7-fold lower in the IR than in the SC region, consistent with results from previous studies. In contrast to most vascular plants, however, three plant lineages (*Pelargonium*,

© 2015 The Authors
New Phytologist © 2015 New Phytologist Trust

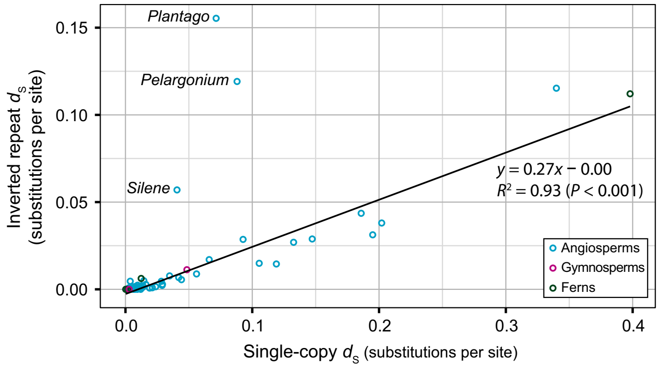


Fig. 2 Correlation of synonymous divergence between inverted repeat (IR) and single-copy (SC) regions among vascular plants. Pairwise synonymous rate analyses were based on genes located fully within the IR or SC region. Linear regression analyses were based on all values except for the outliers *Pelargonium*, *Plantago* and *Silene*. A significance test was calculated using the COR.TEST function available in the R package.

Plantago, and *Silene*) did not follow the trend of reduced IR rates. Instead, protein-coding genes in the IR of these three genera exhibited slightly higher d_S values (1.4- to 2.1-fold) when compared with their SC genes.

To examine the pattern of rate variation in these three genera in more detail, we plotted d_S for individual genes and sequence divergence for individual introns against their genomic positions (Fig. 3). The d_S plot for individual genes identified several extreme outliers in the IR and SC regions. In *Pelargonium*, the SC-localized genes *rpl33* and *psaC* have several-fold higher d_S values than other SC genes, while about half of the IR genes have values ranging from twofold (e.g. *rpl36*) to > 40-fold (*rpoA*) higher than the other half of the IR genes with lower values. Similarly, in *Plantago* there are two SC genes (*accD*, *clpP*) with substantially higher d_S values than other SC genes, while a single IR gene (*ycf1*) has a 10-fold higher d_S than the remaining IR genes. Likewise, d_S in the *Silene* plastome is much higher for the IR gene *ycf2* than for other IR genes and for the SC gene *clpP* than for other SC genes. Because there are so few protein-coding genes in the IR, these few outliers, especially the very large *ycf1* and *ycf2* genes, have a large effect on d_S calculations in the concatenated analysis, which explains the anomalously high d_S values for the concatenated IR genes for these three genera. Importantly, more than half of the IR genes in each species have a lower d_S than do the SC genes, as expected for a model of enhanced copy-correction activity in the IR. Intron divergence values are also consistent with this pattern; divergence is consistently lower for IR introns and higher for SC introns, as expected for their respective localization.

Substitution rate shifts of relocated genes provides further support for IR copy-dependent repair activity

With the numerous expansions and contractions of the IR region that have occurred during vascular plant evolution (Fig. 1), there are now many examples of genes that have

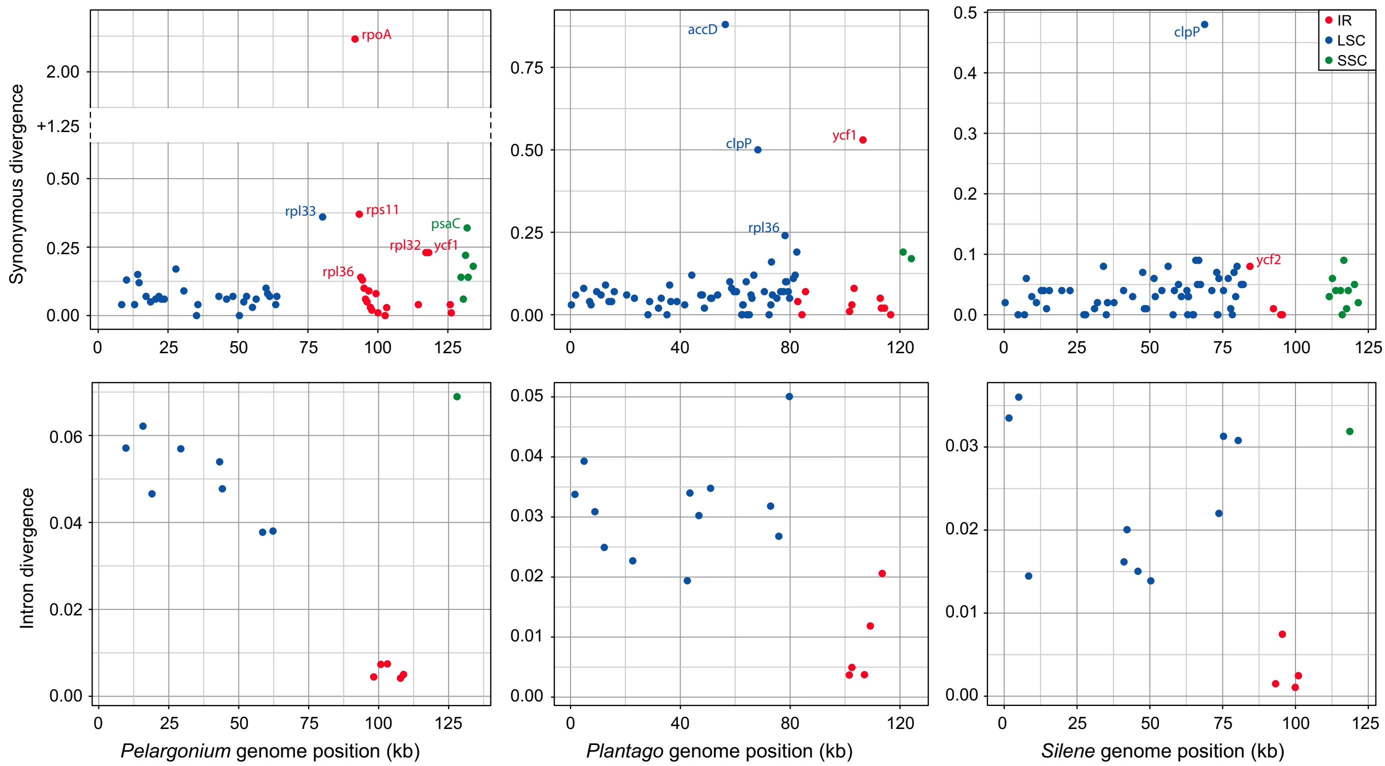


Fig. 3 Intronic and synonymous divergence for individual loci. Synonymous divergence levels for individual genes (upper) and sequence divergence levels for individual introns (lower) are plotted as a function of their genomic positions for *Pelargonium*, *Plantago* and *Silene*. For each species, only one copy of the inverted repeat (IR) is included. LSC, large single-copy region; SSC, small single-copy region.

© 2015 The Authors
New Phytologist © 2015 New Phytologist Trust

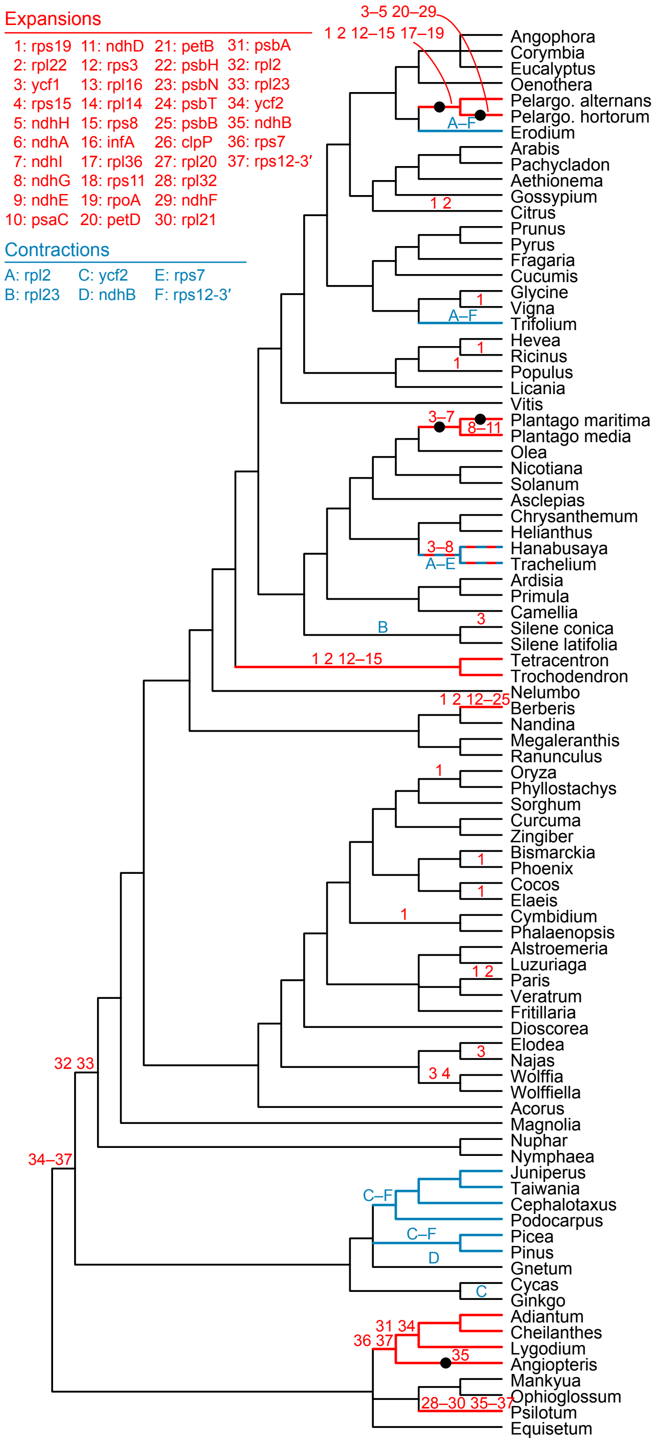


Fig. 4 Relocation of genes into or out of the inverted repeat (IR) in vascular plant plastomes. Only protein-coding genes were shown here, for simplicity, given the focus on synonymous divergence. Red, lineages with substantial IR expansion; blue, lineages with substantial IR contraction. Black dots on particular branches indicate inversion events affecting genes in the newly expanded IRs. Cladogram relationships follow the results of Ruhfel *et al.* (2014) and the Angiosperm Phylogeny Website (http://www.mobot.org/MOBOT/research/APweb/).

relocated into or out of the IR during vascular plant evolution (Fig. 4). Examination of the shift in substitution rates of these relocated genes provides strong support for the role of the IR in providing enhanced copy-correction activity among diverse plants (Fig. 5).

In angiosperms, the six protein-coding genes (*rps12-3′*, *rps7*, *ndhB*, *ycf2*, *rpl23* and *rpl2*) that were ancestrally present in the IR exhibit an approximately four- to sixfold reduction in d_S relative to SC genes, as exemplified by d_S ratios of 0.17–0.23 relative to SC genes for representative angiosperms, including the monocot *Acorus*, the magnoliid *Magnolia*, and the eudicot *Nicotiana* (Fig. 5). In Trochodendraceae, the IR has expanded into the LSC region, resulting in the movement of six genes (*rps19*, *rpl22*, *rps3*, *rpl16*, *rpl14* and *rps8*) into the IR (Fig. 4). These SC-to-IR genes show a sixfold reduction in d_S compared with SC genes, consistent with expectations for IR gene localization (Fig. 5). An opposite pattern is observed for genes that have moved out of the IR. The loss of the IR from *Erodium* and papillionoid legumes (e.g. *Trifolium*) shifted the six ancestral IR genes into the SC regions (Fig. 4). This transfer resulted in SC-like substitution rates for these IR-to-SC genes as demonstrated by a d_S ratio much closer to 1 (Fig. 5). The major shift of the IR in the Campanulaceae species *Hanabusaya* and *Trachelium* provides examples of SC-to-IR transitions for six genes (*ycf1*, *rps15*, *ndhH*, *ndhA*, *ndhI*, *ndhG*) and IR-to-SC transitions for five of the six ancestral IR genes (all except *rps12-3′*) (Fig. 4). Consistent with their current genomic locations in *Hanabusaya* and *Trachelium*, the SC-to-IR genes show a threefold reduction in d_S compared with SC genes, whereas d_S for the IR-to-SC genes have increased to values comparable with SC genes (Fig. 5).

Similar patterns are observed outside of angiosperms. The gymnosperm ancestor was inferred to have four protein-coding genes (*rps12-3′*, *rps7*, *ndhB* and *ycf2*) in the IR (Fig. 1). *Cycas* has retained all four of these genes in the IR, while *Ginkgo* and *Gnetum* have retained three out of the four (Fig. 4). For all three genera, the d_S values for their IR genes are substantially lower than for their SC genes (Fig. 4; note that the extremely low ratios for *Cycas* and *Ginkgo* are probably a result of the low overall plastid substitution rate for these species (Wu & Chaw, 2015) and the small number of genes available for estimation of the even lower IR rate). By contrast, the loss of the IR from cupressophytes (*Cephalotaxus*, *Juniperus*, *Podocarpus* and *Taiwania*) and the nearly complete loss of the IR from Pinaceae (*Picea* and *Pinus*) moved these ancestral IR genes into the SC (Fig. 4). This shift resulted in substantially increased d_S values for these IR-to-SC genes, with values close to (i.e. less than twofold higher or lower than) SC genes (Fig. 5).

Among ferns, multiple IR shifts (expansions, contractions, and/or rearrangements) are required to explain the IR diversity among species, making it difficult to unambiguously determine the ancestral IR content. Nevertheless, there is clear variation among species in terms of the presence and absence of protein-coding genes in the IR. Assuming an ancestral IR that lacked any protein-coding genes, SC-to-IR transitions must have occurred for six, three, and four protein-coding genes in *Psilotum*, *Angiopteris*, and Pteridaceae (*Adiantum* + *Cheilanthes*), respectively (Fig. 4). Consistent with their current location in the IR, these putative SC-to-IR genes have reduced d_S values relative to SC genes (Fig. 5). As mentioned, however, it is only slightly less

© 2015 The Authors
New Phytologist © 2015 New Phytologist Trust

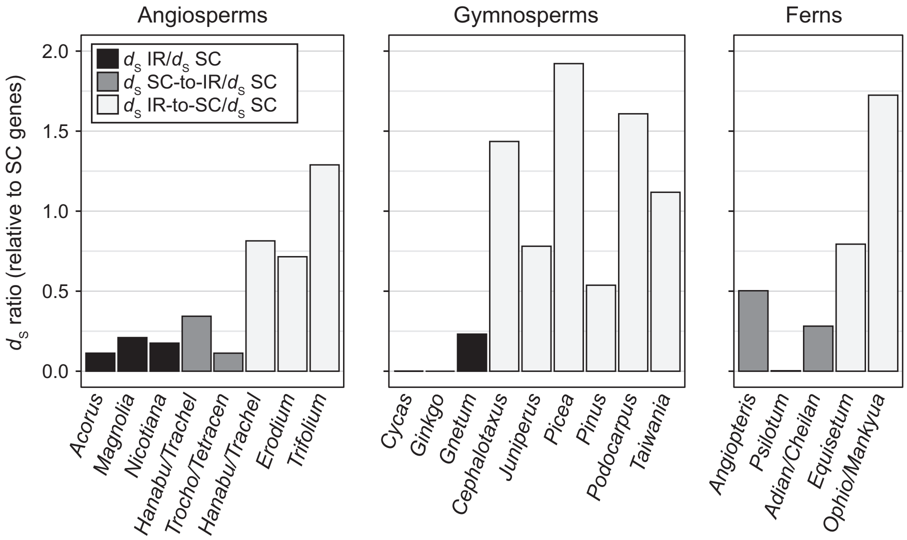


Fig. 5 Relative rates of synonymous divergence for genes relocated into or out of inverted repeat (IR). For each species pair, pairwise synonymous divergence (d_S) values for ancestral IR genes (black), single-copy (SC)-to-IR shifted genes (dark grey), and IR-to-SC shifted genes (light grey) were divided by the d_S value for ancestral SC genes, and this ratio is plotted. *Trocho/Tetracen*, *Trochodendron/Tetracentron*; *Hanabu/Trachel*, *Hanabusaya/Trachelium*; *Adian/Cheilan*, *Adiantum/Cheilanthes*.

parsimonious to assume that the ancestral IR contained three protein-coding genes (*rps12-3′*, *rps7* and *ndhB*). In this scenario, the absence of these genes from *Equisetum* and Ophioglossaceae (*Ophioglossum* + *Mankyua*) would be a result of IR-to-SC transitions. Consistent with their location in the SC, these three putative IR-to-SC genes have d_S values consistent with other SC genes (Fig. 5).

Discussion

Evolutionary models for IR boundary shifts

In this study, we first explored the conservation and evolutionary dynamics of IR boundaries among land plants. In particular, we identified the ancestral structure of the IR at several ancestral nodes and demonstrated that shifts in IR endpoints have occurred multiple times at different evolutionary depths (Fig. 1). While most shifts are small, involving up to several hundred bp, others have expanded or contracted the IR by several kb, which relocated multiple genes into or out of the IR (Fig. 4). By comparing the IR/SC junctions in closely related species, several elegant models have been proposed to explain the expansion and contraction of the IR. By examination of IR/LSC junctions in 13 *Nicotiana* species, Goulding *et al.* (1996) proposed a stepwise model involving a single-strand break, heteroduplex formation via a Holliday junction, and then small IR expansions via gene conversion. This same model may also apply to other small boundary shifts which have occasionally incorporated *rps19* and *rpl22* into the IR of several dicot lineages (Fig. 4). Goulding and colleagues also proposed a different model that starts with a double-strand break followed by strand invasion and recombination to explain the larger IR expansion in *N. acuminata*. A subsequent study by Wang *et al.* (2008) suggested that the double-strand break model could also apply to a small IR extension that incorporated the *trnH-rps19* cluster into the ancestral monocot IR.

In many ways, the IR-expanded plastomes of *Pelargonium* and *Plantago* have distinct features compared with other enlarged IR lineages, such as *N. acuminata*, *Trochodendron* and *Berberis*. These features include extensive genomic rearrangements, accelerated substitution rates, loss of genes and introns, and the presence of multiple large (> 100 bp), nonidentical repeats, which suggests that a different mechanism of IR expansion may be involved. For the *Pelargonium* IR expansion, Chumley *et al.* (2006) proposed a model involving multiple inversions promoted by these dispersed repeats, along with several rounds of ebb-and-flow expansions and contractions. Small dispersed repeats are also located at all inversion breakpoints in the *Plantago* genomes (Fig. S3), indicating that they may have promoted the inversion events. Given the many similarities between the *Plantago* and *Pelargonium* plastomes, this same model may also be applicable to the *Plantago* IR expansions.

Copy-dependent repair and reduced IR substitution rates

Previous studies have also shown that the substitution rate is slower in the IR than in the SC region of angiosperm plastomes. The seminal study by Wolfe *et al.* (1987) examined pairs of taxa at several different evolutionary depths (within Solanaceae, between rosids and asterids, or between monocots and eudicots), but their rate estimates were based on a small subset of genes. Subsequent studies have focused on pairwise comparisons of complete plastomes, yet only a few distinct lineages have been compared to date, including Poaceae (Maier *et al.*, 1995; Gaut, 1998; Yamane *et al.*, 2006), Fabaceae (Perry & Wolfe, 2002), Ranunculaceae (Kim *et al.*, 2009), Araliaceae (Yi *et al.*, 2012), Lamiales (Yi & Kim, 2012; Wicke *et al.*, 2014); and Cycadaceae (Wu & Chaw, 2015). In our study, we have vastly expanded sampling to include not only 39 angiosperm families but also seven gymnosperm and six fern families. Importantly, each of our 69 pairwise comparisons represents nonoverlapping segments of phylogenetic tree space, ensuring that they are independent data points suitable for statistical analysis. Furthermore, each comparison is between a pair of close relatives (intrafamilial, intrageneric, or intraspecific), providing higher confidence that the shared IR genes in each pair have been maintained in the IR since their

© 2015 The Authors
New Phytologist © 2015 New Phytologist Trust

divergence from a common ancestor. With this diverse and extensive sampling, we estimated that IR genes are evolving *c.* 4 times more slowly, on average, than SC genes in a wide variety of vascular plants (Fig. 2). The significant regression analysis provides strong evidence that reduced IR rates are a fundamental property of plant plastomes.

In an important follow-up study, Perry & Wolfe (2002) found that substitution rates increased to SC levels for former IR genes in the clade of IR-lacking legumes. A similar finding was reported by Gaut (1998), who used relative ratio tests for former IR genes in the pine plastome, although there were no statistical data presented to support this conclusion. Here, we examined IR boundary shifts in 13 pairs of vascular plants, which resulted in multiple examples of IR-to-SC gene transitions and SC-to-IR transitions (Figs 4, 5). Consistent with previous results, IR-to-SC genes exhibited substitution rates comparable to those of ancestral SC genes in several independent lineages including *Erodium*, *Trifolum*, cupressophytes (*Cephalotaxus, Juniperus*, *Podocarpus* and *Taiwania*), and Pinaceae (*Pinus* and *Picea*). Conversely, the major IR expansion in Trochodendrales and several ferns (*Angiopteris*, *Psilotum*, and Pteridaceae) resulted in IR-like substitution rates for genes moved from the SC into the IR. This reduction in substitution rates for SC-to-IR gene transitions has not been demonstrated previously. Perhaps the most illustrative single example of the effect of IR duplication on substitution rates comes from the *Hanabusaya* and *Trachelium* comparison, in which a major IR shift transferred some former SC genes into the IR and some former IR genes into the SC. Consistent with our other comparisons, the SC-to-IR genes in *Hanabusaya* and *Trachelium* show IR-like substitution rates, while their IR-to-SC genes show SC-like substitution rates (Fig. 5). Together, these results clearly demonstrate that IR localization, rather than gene identity or function, is the key factor in conferring reduced substitution rates in plant plastomes.

Localized hypermutation as another source of intragenomic rate heterogeneity

Surprisingly, however, we discovered that this pattern of reduced IR substitution rates does not apply universally to all vascular plants. We observed that IR genes from species in the genera *Pelargonium*, *Plantago* and *Silene* have comparable, and in fact slightly higher, synonymous rates, on average, relative to their SC genes. How did this unusual evolutionary pattern arise? Given that the enhanced copy-correction activity in the IR is probably a result of increased amounts of homologous recombination and gene conversion, one straightforward explanation might be the loss or reduction of homologous recombination activity in the IR of these three genera. However, closer inspection of sequence divergence of individual loci revealed that, instead of a general increase of the IR substitution rates for all genes, the increased substitution rates are confined to a few mutation hotspots: *rpoA-rps11-rpl36* and *ycf1-rpl32* in *Pelargonium*, *ycf1* in *Plantago*, and *ycf2* in *Silene* (Fig. 3). Overall, the observation of locus-specific increases in sequence divergence coupled with elevated levels of rearrangements, gene/intron loss, and repetitiveness in the plastomes of *Plantago*, *Pelargonium* and *Silene* suggests a common process driving the correlated evolution of all phenomena.

These striking locus-specific rate increases in *Pelargonium*, *Plantago*, and *Silene* are not unique, as examples have been observed in plastid or mitochondrial genomes of several plants (Mower *et al.*, 2007; Erixon & Oxelman, 2008; Guisinger *et al.*, 2008; Sloan *et al.*, 2009, 2012a,b; Magee *et al.*, 2010; Zhu *et al.*, 2014). In *Oenthera* and several *Sileneae* lineages (including *Silene*), the *clpP* gene was previously shown to have elevated synonymous and nonsynonymous substitution rates associated with the proliferation of repetitive amino acid sequence motifs and loss of the introns, although the evolutionary processes connecting these various phenomena were not determined (Erixon & Oxelman, 2008). Similar repetitive amino acid motifs were identified in *Medicago accD* and *ycf1* genes, and their active proliferation over a short evolutionary timescale was suggested to be recombinationally driven, but it was not determined if these genes also had accelerated substitution rates (Gurdon & Maliga, 2014). For the substantial intragenomic variation in synonymous substitution rates within some plant mitochondrial genomes, possible evolutionary processes were suggested to be recombination between maternal and paternal genome copies (Sloan *et al.*, 2009) or gene conversion via recombination with processed transcripts and their originating genes (Zhu *et al.*, 2014). A more direct link between recombination and mutation hotspots was postulated for the IR-lacking plastome from *Lathyrus*, where repeated DNA breakage and repair were suggested to cause a *c.* 1.5 kb localized hypermutation region around *ycf4* (Magee *et al.*, 2010). In *Plantago media*, the three fastest-evolving genes (*accD*, *clpP*, *ycf1*) have small repeats in their vicinity, suggesting a role for recombination in rate acceleration (Fig. S3B). Overall, the weight of evidence suggests that mutation hotspots are tied to increased recombinational activity, which itself may be driven by the proliferation of repeats within these genomes.

Acknowledgements

The authors gratefully acknowledge the efforts of Felix Grewe, Emily Gubbels, Nancy Hepburn, Volker Knoop, Monica Polsakiewicz, and Yizhong Zhang for preparation of DNAs sequenced in this study. We also thank Raghavendran Partha for assistance with *Plantago* plastome annotation and Amy Hilske and Samantha Link for care of plants in the Beadle Center glasshouse. This research was supported in part by the National Science Foundation (IOS-1027529 and MCB-1125386 to J.P.M.) and by startup funds from the University of Nebraska–Lincoln (to J.P.M.).

Author contributions

A.Z. and J.P.M. designed the research. A.Z., W.G., S.G., W.F. and J.P.M. performed experiments, analyzed data, and interpreted results. A.Z. and J.P.M. wrote the manuscript. All authors approved the final version of the manuscript.

© 2015 The Authors
New Phytologist © 2015 New Phytologist Trust

References

Birky CW, Walsh JB. 1992. Biased gene conversion, copy number, and apparent mutation-rate differences within chloroplast and bacterial genomes. *Genetics* **130**: 677–683.

Cho Y, Mower JP, Qiu YL, Palmer JD. 2004. Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. *Proceedings of the National Academy of Sciences, USA* **101**: 17741–17746.

Chumley TW, Palmer JD, Mower JP, Fourcade HM, Calie PJ, Boore JL, Jansen RK. 2006. The complete chloroplast genome sequence of *Pelargonium* × *hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. *Molecular Biology and Evolution* **23**: 2175–2190.

Downie SR, Jansen RK. 2015. A comparative analysis of whole plastid genomes from the Apiales: expansion and contraction of the inverted repeat, mitochondrial to plastid transfer of DNA, and identification of highly divergent noncoding regions. *Systematic Botany* **40**: 336–351.

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* **19**: 11–15.

Erixon P, Oxelman B. 2008. Whole-gene positive selection, elevated synonymous substitution rates, duplication, and indel evolution of the chloroplast *clpP1* gene. *PLoS ONE* **3**: e1386.

Gaut BS. 1998. Molecular clocks and nucleotide substitution rates in higher plants. In: Hecht MK, Macintyre RJ, Clegg MT, eds. *Evolutionary biology.* New York, NY, USA: Plenum Press, 93–120.

Goulding SE, Olmstead RG, Morden CW, Wolfe KH. 1996. Ebb and flow of the chloroplast inverted repeat. *Molecular and General Genetics* **252**: 195–206.

Grewe F, Guo W, Gubbels EA, Hansen AK, Mower JP. 2013. Complete plastid genomes from *Ophioglossum californicum*, *Psilotum nudum*, and *Equisetum hyemale* reveal an ancestral land plant genome structure and resolve the position of Equisetales among monilophytes. *BMC Evolutionary Biology* **13**: 8.

Guisinger MM, Kuehl JV, Boore JL, Jansen RK. 2008. Genome-wide analyses of Geraniaceae plastid DNA reveal unprecedented patterns of increased nucleotide substitutions. *Proceedings of the National Academy of Sciences, USA* **105**: 18424–18429.

Guisinger MM, Kuehl JV, Boore JL, Jansen RK. 2011. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. *Molecular Biology and Evolution* **28**: 583–600.

Guo W, Grewe F, Cobo-Clark A, Fan W, Duan Z, Adams RP, Schwarzbach AE, Mower JP. 2014. Predominant and substoichiometric isomers of the plastid genome coexist within *Juniperus* plants and have shifted multiple times during cupressophyte evolution. *Genome Biology and Evolution* **6**: 580–590.

Gurdon C, Maliga P. 2014. Two distinct plastid genome configurations and unprecedented intraspecies length variation in the accD coding region in *Medicago truncatula. DNA Research* **21**: 417–427.

Jansen RK, Ruhlman TA. 2012. Plastid genomes of seed plants. In: Bock R, Knoop V, eds. *Genomics of chloroplasts and mitochondria.* Dordrecht, the Netherlands: Springer, 103–126.

Khakhlova O, Bock R. 2006. Elimination of deleterious mutations in plastid genomes by gene conversion. *Plant Journal* **46**: 85–94.

Kim YK, Park CW, Kim KJ. 2009. Complete chloroplast DNA sequence from a Korean endemic genus, *Megaleranthis saniculifolia*, and its evolutionary implications. *Molecules and Cells* **27**: 365–381.

Kuck P, Meusemann K. 2010. FASconCAT: convenient handling of data matrices. *Molecular Phylogenetics and Evolution* **56**: 1115–1118.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R *et al.* 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947–2948.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**: 1451–1452.

Magee AM, Aspinall S, Rice DW, Cusack BP, Semon M, Perry AS, Stefanovic S, Milbourne D, Barth S, Palmer JD *et al.* 2010. Localized hypermutation and associated gene losses in legume chloroplast genomes. *Genome Research* **20**: 1700–1710.

Maier RM, Neckermann K, Igloi GL, Kossel H. 1995. Complete sequence of the maize chloroplast genome – gene content, hotspots of divergence and fine-tuning of genetic information by transcript editing. *Journal of Molecular Biology* **251**: 614–628.

Mower JP, Touzet P, Gummow JS, Delph LF, Palmer JD. 2007. Extensive variation in synonymous substitution rates in mitochondrial genes of seed plants. *BMC Evolutionary Biology* **7**: 135.

Palmer JD, Osorio B, Aldrich J, Thompson WF. 1987. Chloroplast DNA evolution among legumes – loss of a large inverted repeat occurred prior to other sequence rearrangements. *Current Genetics* **11**: 275–286.

Perry AS, Wolfe KH. 2002. Nucleotide substitution rates in legume chloroplast DNA depend on the presence of the inverted repeat. *Journal of Molecular Evolution* **55**: 501–508.

Raubeson LA, Jansen RK. 1992. A rare chloroplast-DNA structural mutation is shared by all conifers. *Biochemical Systematics and Ecology* **20**: 17–24.

Ruhfel BR, Gitzendanner MA, Soltis PS, Soltis DE, Burleigh JG. 2014. From algae to angiosperms-inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. *BMC Evolutionary Biology* **14**: 23.

Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer JD, Taylor DR. 2012a. Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. *PLoS Biology* **10**: e1001241.

Sloan DB, Alverson AJ, Wu M, Palmer JD, Taylor DR. 2012b. Recent acceleration of plastid sequence and structural evolution coincides with extreme mitochondrial divergence in the angiosperm genus *Silene. Genome Biology and Evolution* **4**: 294–306.

Sloan DB, Oxelman B, Rautenberg A, Taylor DR. 2009. Phylogenetic analysis of mitochondrial substitution rate variation in the angiosperm tribe Sileneae. *BMC Evolutionary Biology* **9**: 260.

Sun YX, Moore MJ, Meng AP, Soltis PS, Soltis DE, Li JQ, Wang HC. 2013. Complete plastid genome sequencing of Trochodendraceae reveals a significant expansion of the inverted repeat and suggests a paleogene divergence between the two extant species. *PLoS ONE* **8**: e60429.

Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research* **34**: W609–W612.

Tsudzuki J, Nakashima K, Tsudzuki T, Hiratsuka J, Shibata M, Wakasugi T, Sugiura M. 1992. Chloroplast DNA of black pine retains a residual inverted repeat lacking rRNA genes: nucleotide sequences of *trnQ*, *trnK*, *psbA*, *trnI* and *trnH* and the absence of *rps16*. *Molecular and General Genetics* **232**: 206–214.

Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. 2010. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics, Proteomics and Bioinformatics* **8**: 77–80.

Wang RJ, Cheng CL, Chang CC, Wu CL, Su TM, Chaw SM. 2008. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *BMC Evolutionary Biology* **8**: 36.

Wicke S, Schaferhoff B, dePamphilis CW, Muller KF. 2014. Disproportional plastome-wide increase of substitution rates and relaxed purifying selection in genes of carnivorous Lentibulariaceae. *Molecular Biology and Evolution* **31**: 529–545.

Wicke S, Schneeweiss GM, dePamphilis CW, Muller KF, Quandt D. 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Molecular Biology* **76**: 273–297.

Wolf PG, Karol KG. 2012. Plastomes of bryophytes, lycophytes and ferns. In: Bock R, Knoop V, eds. *Genomics of chloroplasts and mitochondria.* Dordrecht, the Netherlands: Springer, 89–102.

Wolfe KH, Li WH, Sharp PM. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences, USA* **84**: 9054–9058.

Wu CS, Chaw SM. 2015. Evolutionary stasis in cycad plastomes and the first case of plastome GC-biased gene conversion. *Genome Biology and Evolution* **7**: 2000–2009.

Wu CS, Lai YT, Lin CP, Wang YN, Chaw SM. 2009. Evolution of reduced and compact chloroplast genomes (cpDNAs) in gnetophytes: selection toward a lower-cost strategy. *Molecular Phylogenetics and Evolution* **52**: 115–124.

Wu CS, Lin CP, Hsu CY, Wang RJ, Chaw SM. 2011. Comparative chloroplast genomes of pinaceae: insights into the mechanism of diversified genomic organizations. *Genome Biology and Evolution* **3**: 309–319.

© 2015 The Authors
New Phytologist © 2015 New Phytologist Trust

Wyman SK, Jansen RK, Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* **20**: 3252–3255.

Yamane K, Yano K, Kawahara T. 2006. Pattern and rate of indel evolution inferred from whole chloroplast intergenic regions in sugarcane, maize and rice. *DNA Research* **13**: 197–204.

Yi DK, Kim KJ. 2012. Complete chloroplast genome sequences of important oilseed crop *Sesamum indicum* L. *PLoS ONE* **7**: e35872.

Yi DK, Lee HL, Sun BY, Chung MY, Kim KJ. 2012. The complete chloroplast DNA sequence of *Eleutherococcus senticosus* (Araliaceae); comparative evolutionary analyses with other three asterids. *Molecules and Cells* **33**: 497–508.

Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Research* **18**: 821–829.

Zhu A, Guo W, Jain K, Mower JP. 2014. Unprecedented heterogeneity in the synonymous substitution rate within a plant genome. *Molecular Biology and Evolution* **31**: 1228–1236.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 MAUVE alignments of conserved plastome sequences.

Fig. S2 MAUVE alignments of nonconserved plastome sequences.

Fig. S3 Plastome maps for *Plantago.*

Table S1 List of taxa used in this study

Please note: Wiley Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.

About *New Phytologist*

- *New Phytologist* is an electronic (online-only) journal owned by the New Phytologist Trust, a **not-for-profit organization** dedicated to the promotion of plant science, facilitating projects from symposia to free access for our Tansley reviews.
- Regular papers, Letters, Research reviews, Rapid reports and both Modelling/Theory and Methods papers are encouraged. We are committed to rapid processing, from online submission through to publication 'as ready' via *Early View* – our average time to decision is <27 days. There are **no page or colour charges** and a PDF version will be provided for each article.
- The journal is available online at Wiley Online Library. Visit **www.newphytologist.com** to search the articles and register for table of contents email alerts.
- If you have any questions, do get in touch with Central Office (np-centraloffice@lancaster.ac.uk) or, if it is more convenient, our USA Office (np-usaoffice@lancaster.ac.uk)
- For submission instructions, subscription and all the latest information visit **www.newphytologist.com**

© 2015 The Authors
New Phytologist © 2015 New Phytologist Trust