

# PUBLISHED VERSION

Dong Wang, Zhipeng Qu, David L. Adelson, Jian-Kang Zhu, and Jeremy N. Timmis  
Transcription of nuclear organellar DNA in a model plant system  
Genome Biology and Evolution, 2014; 6(6):1327-1334

© The Author(s) 2014. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

Originally published at:

<http://doi.org/10.1093/gbe/evu111>

## PERMISSIONS

<http://creativecommons.org/licenses/by-nc/3.0/>



Attribution-NonCommercial 3.0 Unported (CC BY-NC 3.0)

This is a human-readable summary of (and not a substitute for) the [license](#).

[Disclaimer](#)

### You are free to:

**Share** — copy and redistribute the material in any medium or format

**Adapt** — remix, transform, and build upon the material

The licensor cannot revoke these freedoms as long as you follow the license terms.

### Under the following terms:



**Attribution** — You must give **appropriate credit**, provide a link to the license, and **indicate if changes were made**. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.



**NonCommercial** — You may not use the material for **commercial purposes**.

**No additional restrictions** — You may not apply legal terms or **technological measures** that legally restrict others from doing anything the license permits.

<http://hdl.handle.net/2440/95292>

# Transcription of Nuclear Organellar DNA in a Model Plant System

Dong Wang<sup>1,\*†</sup>, Zhipeng Qu<sup>2,†</sup>, David L. Adelson<sup>2</sup>, Jian-Kang Zhu<sup>1</sup>, and Jeremy N. Timmis<sup>2</sup>

<sup>1</sup>Shanghai Center for Plant Stress Biology, Shanghai Institute for Biological Science, Chinese Academy of Sciences, Shanghai, China

<sup>2</sup>Discipline of Genetics, School of Molecular and Biomedical Science, The University of Adelaide, South Australia, Australia

\*Corresponding author: E-mail: dongwangbio@gmail.com.

†These authors contributed equally to this work.

Accepted: May 21, 2014

## Abstract

Endosymbiotic gene transfer from cytoplasmic organelles (chloroplasts and mitochondria) to the nucleus is an ongoing process in land plants. Although the frequency of organelle DNA migration is high, functional gene transfer is rare because a nuclear promoter is thought necessary for activity in the nucleus. Here we show that a chloroplast promoter, *16S mm*, drives nuclear transcription, suggesting that a transferred organellar gene may become active without obtaining a nuclear promoter. Examining the chromatin status of a known de novo chloroplast integrant indicates that plastid DNA inserts into open chromatin and that this relaxed condition is maintained after integration. Transcription of nuclear organelle DNA integrants was explored at the whole genome level by analyzing RNA-seq data of *Oryza sativa* subsp. *japonica*, and utilizing sequence polymorphisms to unequivocally discriminate nuclear organelle DNA transcripts from those of bona fide cytoplasmic organelle DNA. Nuclear copies of organelle DNA that are transcribed show a spectrum of transcriptional activity but at comparatively low levels compared with the majority of other nuclear genes.

**Key words:** endosymbiotic gene transfer, nuclear integrants of organellar DNA, open chromatin.

## Introduction

The transfer of prokaryotic DNA molecules into the nuclear genome that has occurred during two bacterial endosymbioses has played a major part in eukaryote evolution. Many endosymbiont genes were captured and activated by the nucleus and transferred DNA also contributed in more complex ways to the heterogeneity of the nuclear gene complement (Timmis et al. 2004). This endosymbiotic gene transfer (EGT) also resulted in massive functional relocation to the nucleus of genes that were formerly located in the endosymbionts (Gray et al. 1999; Timmis et al. 2004; Bock and Timmis 2008), explaining the much reduced size of extant mitochondrial and chloroplast genomes compared with their prokaryotic ancestors. Various steps in EGT have been recapitulated experimentally in yeast (Thorsness and Fox 1990), and in *Nicotiana tabacum* where the frequency of the first step—DNA transfer per se—was found to be surprisingly high (Huang et al. 2003; Stegemann et al. 2003; Wang, Lloyd, et al. 2012). However, for the successful relocation of functional organelle genes, mere insertion of DNA into the nuclear genome is not sufficient because of major differences in control of gene

expression between the nucleus and the prokaryotic endosymbionts. For this reason, functional activation of plastid-derived genes in the nucleus is much rarer. However, despite its rarity, the process has been demonstrated experimentally by two independent research teams (Stegemann and Bock 2006; Lloyd and Timmis 2011) to involve the acquisition of nuclear transcription and polyadenylation motifs. In a special case, the chloroplast *psbA* promoter has been reported to be weakly active in the nucleus without any modification (Cornelissen and Vandewiele 1989), and nuclear insertion of multiple copies of a spectinomycin resistance gene, *aadA*, driven by this promoter leads immediately to selectable spectinomycin resistance (Lloyd and Timmis 2011).

Cytoplasmic organellar DNAs integrate into the nuclear genome through nonhomologous end joining (NHEJ) (Ricchetti et al. 1999; Lloyd and Timmis 2011; Wang, Lloyd, et al. 2012), and they insert preferentially into open chromatin regions (Wang and Timmis 2013). Likewise, recent human nuclear integrants of mitochondrial DNAs (*numts*) are commonly located in, or closely adjacent to, regions of open chromatin (Tsuji et al. 2012). Open chromatin regions are often

depleted in nucleosome, a circumstance that permits greater access to interacting molecules (Hogan et al. 2006; Kim et al. 2007; Song et al. 2011), including the machinery of NHEJ and of transcription. Thus, although activity of the *psbA* in the nucleus may be explained by fortuitous nuclear transcription motifs (Lloyd and Timmis 2011), the likelihood exists of low level transcription of nuclear integrants of organellar DNA (*norgs*) simply because they tend to occupy open chromatin. Taking these observations together, we hypothesized that the majority of inserted organellar DNA may be transcribed directly after migration to the nucleus without the necessity to acquire nuclear transcription motifs, though the resulting mRNAs may lack the signals required for translation.

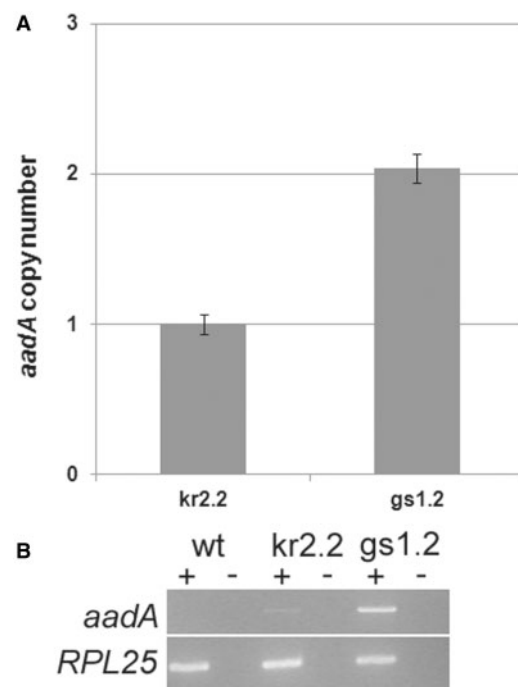
Here, we show that a *16S rrr* plastid promoter-driven reporter gene located in a de novo experimental chloroplast DNA integrant is transcribed after nuclear transfer, indicating that *16S rrr* promoter can be immediately active in nucleus, though it appears to contain none of the cryptic nuclear signals that appeared to explain the activity of the *psbA* promoter. We investigated the chromatin status of a fully sequenced de novo experimental chloroplast integrant (Lloyd and Timmis 2011) by DNase I-PCR (polymerase chain reaction). Plastid DNA was found to insert into open chromatin and the relaxed condition was maintained after *norg* insertion, suggesting that the chloroplast integrant might be transcribed immediately without acquiring a nuclear promoter. To further explore the transcription of *norgs* at the whole genome level, RNA-seq data of *Oryza sativa* subsp. *japonica* were analyzed by searching for polymorphic RNAs containing single nucleotide differences (SNPs) and indels which unequivocally distinguish *norg* transcripts from those of bona fide cytoplasmic organelle DNA. A set of *norg*-specific transcripts was identified in this way, and their transcriptional patterns showed a continuous distribution similar to that of other nuclear genes. However their average level of RNA abundance was much lower, suggesting that plastid promoters work weakly in the nucleus or that the *norgs* were nonspecifically transcribed because they were located in open chromatin. Some *norgs* with the highest transcriptional characteristics within the range of active nuclear genes were further investigated, and most were found to be inserted into a nuclear gene.

## Results

### Transcription of a Plastid Promoter-Driven Reporter Gene in the Nucleus

Because of the high sequence similarity between nuclear integrants of plastid DNAs (*nupts*) and their plastid counterparts, it is difficult to demonstrate their transcription unequivocally. However, the *gs1.2* tobacco line (Sheppard et al. 2008) allowed us to determine whether a plastid promoter other

than *psbA* is active in the nucleus. The *gs1.2* line contains a de novo experimental chloroplast integrant harboring two copies of a *16S rrr* promoter-driven *aadA* gene in figure 1A. Transcripts of *aadA* were demonstrated by reverse transcription (RT)-PCR demonstrating activity in the nucleus (fig. 1B) using *aadA* driven by the *psbA* promoter in the tobacco line *kr2.2* (Lloyd and Timmis 2011), as a positive control. The greater transcript accumulation of *aadA* in *gs1.2* (fig. 1B) is consistent with two copies of the reporter gene in *gs1.2* compared with a single copy in *kr2.2* (fig. 1A). These results suggest that the *16S rrr* and the *psbA* promoters are equally able to function directly in the nucleus. No cryptic nuclear transcription motifs such as TATA and CAAT are seen in *16S rrr* promoter (Sheppard et al. 2008). Therefore, it seems likely that transcriptional activity is facilitated by the *nupts* occupying open chromatin regions of the nucleus rather than the fortuitous presence of eukaryotic sequence motifs that were previously held responsible for promoter activity in the case of *aadA*.



**Fig. 1.**—Determination of *aadA* gene copy number and transcript accumulation. (A) The comparative copy number of *aadA* in *gs1.2* by real-time quantitative PCR. Both *kr2.2* and *gs1.2* are experimental gene-transfer lines of *Nicotiana tabacum*. The control (*kr2.2*) contains a single copy of *aadA* (Lloyd and Timmis 2011). (B) RT-PCR analysis of plastid promoter-driven *aadA* genes in the nucleus. Transcript accumulation of *aadA* genes driven by the *psbA* promoter (*kr2.2*) and the *16S rrr* (*gs1.2*) promoter is shown. Control RT-PCR using primers specific for *RPL25* is also shown. Lanes marked “+” and “-” indicate samples with and without reverse transcriptase.

### Chromatin Status within a De Novo Chloroplast DNA Integrant

Transcriptional activity is characteristic of open chromatin regions (Song et al. 2011), and recently formed human *numts* have been shown to favor open chromatin or regions flanking open chromatin (Tsuji et al. 2012). Therefore, the chromatin status of a fully characterized chloroplast DNA integrant in kr2.2 (Lloyd and Timmis 2011) and its preinsertion site in wild type (WT) seedlings was examined by DNase I-PCR. DNase I sensitivity is commonly used to interrogate DNA chromatin compaction.

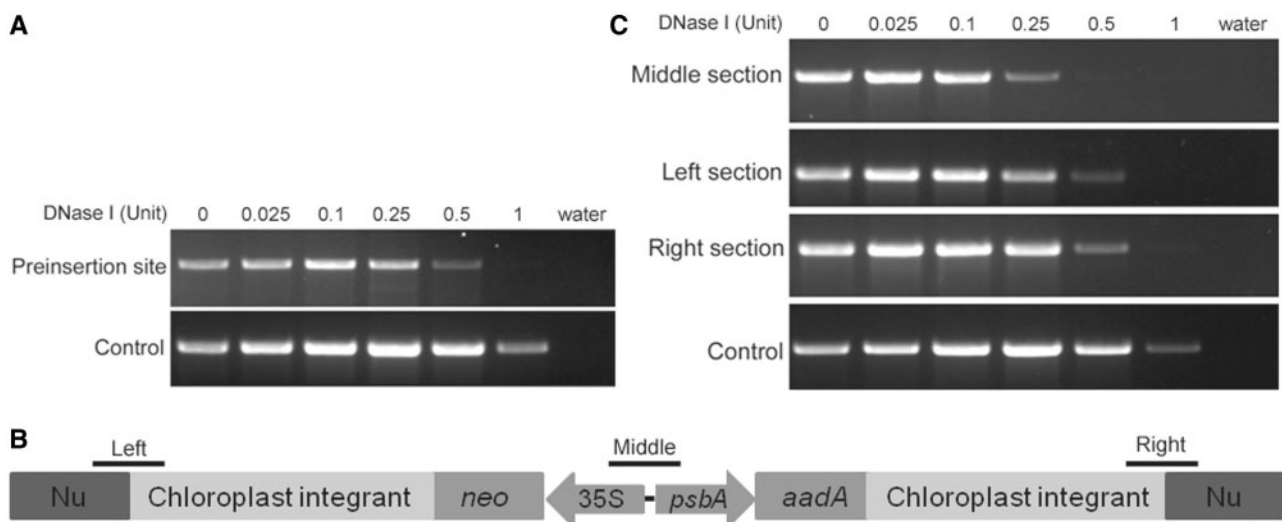
In WT tobacco, chromatin at the site of integration of the de novo *nupt* in kr2.2 was found to be less compacted compared with a control heterochromatic region (fig. 2A). This supports previous findings (Wang and Timmis 2013) that cytoplasmic organellar DNA inserts preferentially into open chromatin. After insertion of approximately 17 kb (Lloyd and Timmis 2011) of chloroplast DNA in kr2.2 (fig. 2B), this region of chromatin remained uncompacted over its entire length (fig. 2). The 17-kb integrant containing two reporter genes was examined with three different primer pairs. In particular, chromatin containing the DNA segment harboring the *neo* and *psbA* promoters of the reporter genes was more accessible (fig. 2C, middle section) than regions close to the insertion site (fig. 2C, left and right sections), suggesting that both *addA* and *neo* genes are transcriptionally active in the nucleus. Although the *neo* gene, driven by the 35S promoter, is known to be highly active as it was used to detect the

chloroplast DNA transfer event, *aadA* is driven by the chloroplast-specific *psbA* promoter which is expected to be much less active in the nucleus. It is possible that the undisturbed relaxed state of the chromatin is maintained because of the presence of the highly active 35S promoter driving *neo* or it may be that many *norgs* insert with minimal impact on neighboring genes.

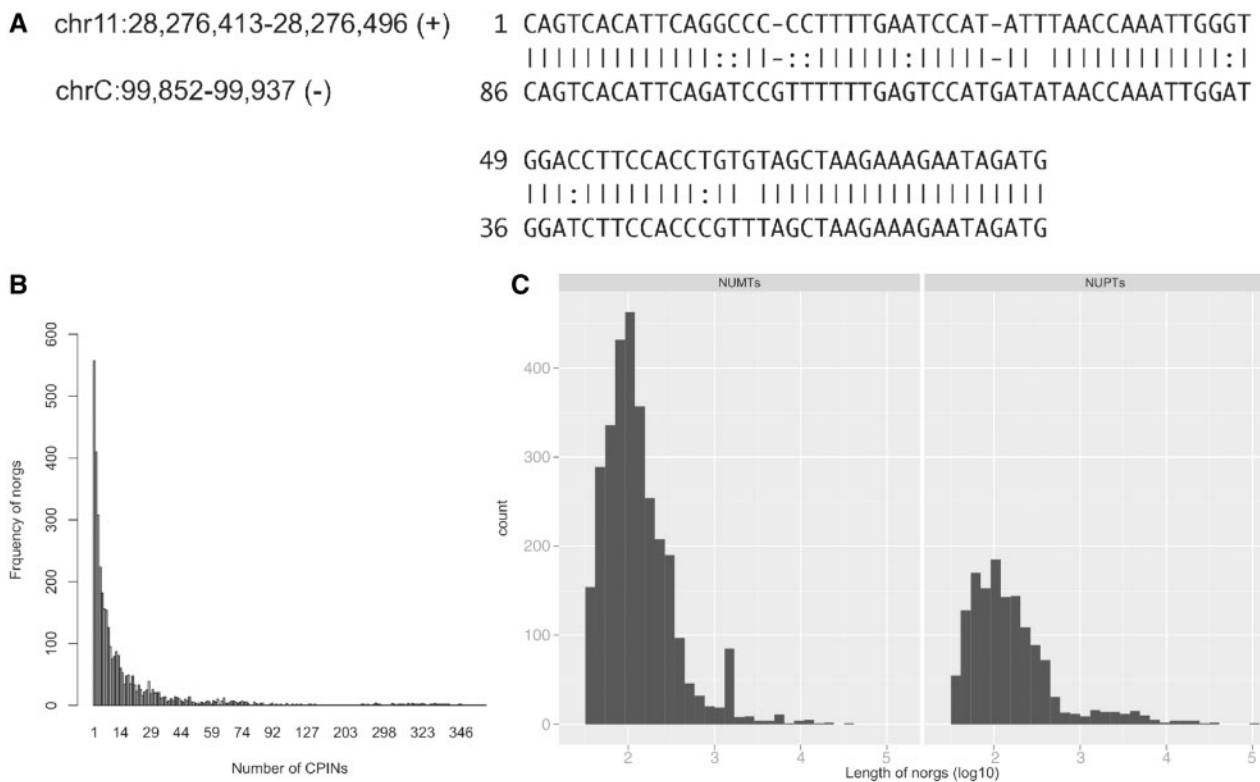
### Transcription of *norgs* in *Oryza sativa* subsp. *japonica*

The results in figure 1B confirm that some native chloroplast genes may be transcribed without modification after transfer to nucleus, in rare cases where they contain fortuitous eukaryotic promoters and, much more often, because they tend to integrate into active chromatin.

To investigate the generalized transcription of naturally occurring *norgs*, 3,032 *numts* and 1,417 *nupts* were identified in *O. sativa* subsp. *japonica*, and RNA-seq data (Zhang et al. 2012) searched for unambiguous matches. Base substitutions and indels located in *norgs* were utilized to distinguish unequivocal transcripts of *norgs* among total transcripts. These mutations were designated as checking points in *norgs* (CPINs) (fig. 3A), and only reads mapped with CPINs were retained for further analysis. A total of 90,413 CPINs were identified from 3,674 *norgs*. Most, often short, *norgs* were found to harbor one or two differences compared with their organellar counterparts, whereas longer ones regularly showed proportionately more CPINs (Spearman's correlation,  $r=0.581544$ ,  $P$  value  $< 2.2e-16$ ). This necessarily



**FIG. 2.**—Inspecting the chromatin status of a genomic region before and after chloroplast DNA insertion. (A) Testing the chromatin status at the preinsertion site of chloroplast insertion by DNase I-PCR in WT seedlings. (B) Structure of the chloroplast integrant of the kr2.2 line (Lloyd and Timmis 2011). Nu, nuclear DNA. Black lines indicate target sites of DNase I-PCR in (C). (C) Testing the chromatin status of chloroplast integrant and its flanking region by DNase I-PCR in kr2.2 homozygous plants. The final quantity of DNase I used in the experiment is listed at the top of the gel. Water, water added only in the PCR reaction. Control, a region containing transposons and repetitive sequences was used as positive heterochromatic control. For optimization of DNase I-PCR, all amplicons were approximately 1 kb according to previous research (Shu et al. 2013).



**Fig. 3.**—CPINs and *norgs* of *Oryza sativa* subsp. *japonica*. (A) An example of CPINs in a *nupt*. The sequence differences (CPINs) between nuclear genome (chr11, top line) and organellar genome (chrC, bottom line) were employed to identify unequivocal transcripts of a set of *norgs*. (B) The number of CPINs in *norgs*. (C) Length (log<sub>10</sub> bp) of *nupts* and *numts* within the *O. sativa* subsp. *japonica* genome.

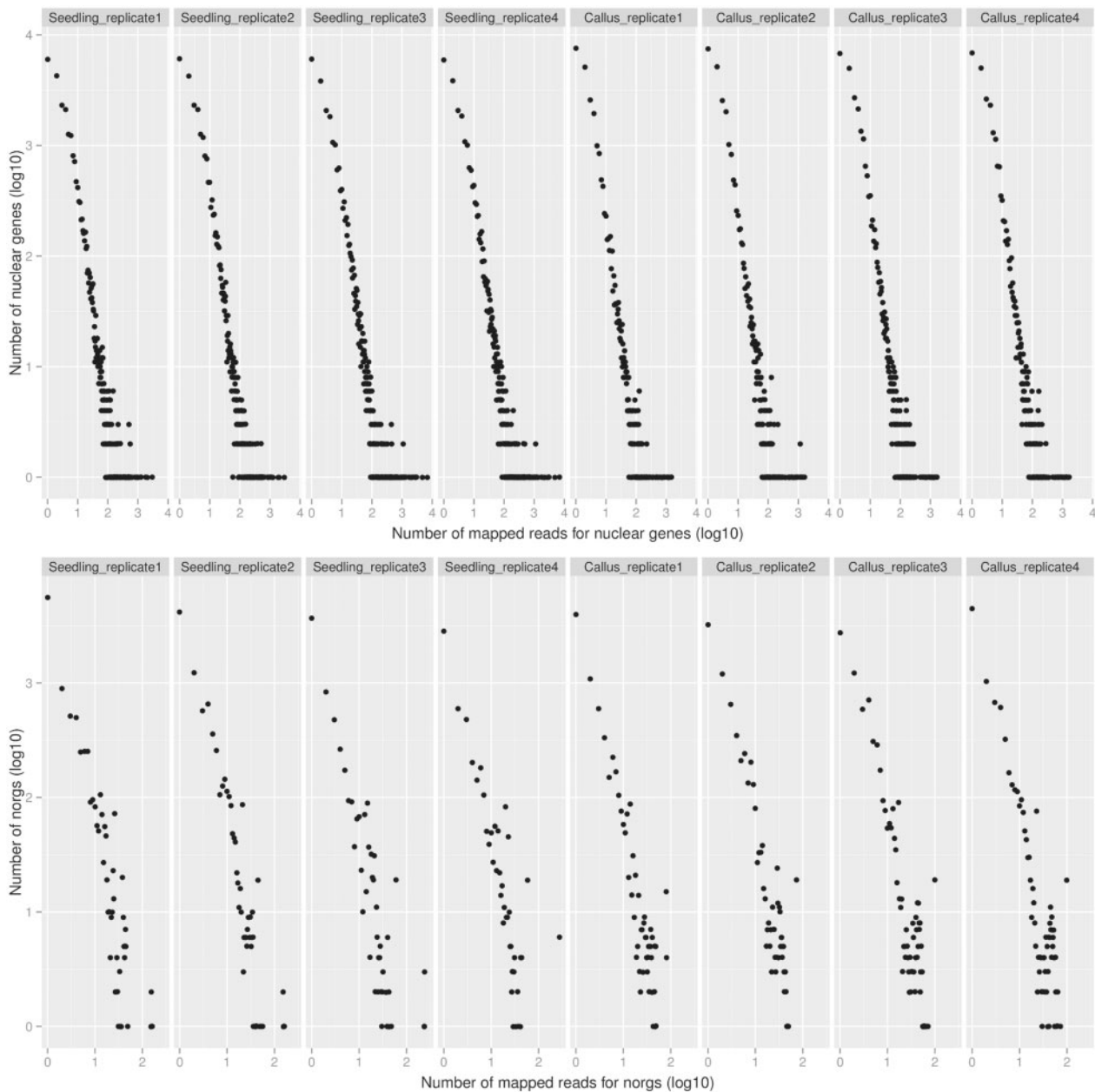
means that there are a large number of *norgs* with a minimal number of CPINs but sufficient polymorphisms were identified to permit a comprehensive and secure description of the transcription of *norgs* at the whole genome level.

After mapping the reads from eight RNA-seq samples to the genome of subsp. *japonica*, about 23% (21,156 out of 90,413) CPINs were found with matching reads in at least one sample. Of the 21,156 CPINs with matching reads, approximately 52% (11,125 out of 21,156) were found in only one RNA-seq sample, but around 8% (1,764 out of 21,156) were covered by reads in all eight available RNA-seq samples from seedling and callus (Zhang et al. 2012).

Next, we compared the transcription pattern between annotated genes and *norgs* in the eight RNA samples. The profiles of all the *norgs* presented a continuous pattern similar to the annotated genes, but their average transcript abundance was lower (~3.7 reads per *norg* compared with ~6.6 reads per gene) (fig. 4). This is consistent with the observation that plastid promoters work more weakly in the nucleus than the majority of eukaryotic promoters (Lloyd and Timmis 2011). Interestingly, approximately 0.4% of *norgs* were identified whose transcript abundance was equal to, or greater than, ten mapped read counts (fig. 4), which represents transcription well within the normal range of nuclear gene expression.

This subset of characterized loci with higher levels of transcriptional activity was extracted for further analysis. Using the method described in the legend to figure 4, 17 *norgs* were found to be preferentially transcribed in seedling tissues and 12 in callus, with six of these identified in both tissue types (table 1). Fifteen of these *norgs* overlapped with, or were embedded within, annotated nuclear genes. This finding explains why this subset of *norgs* shows transcript accumulation levels that are comparable with known active nuclear genes and suggests that they represent an integral part of the associated gene transcripts (supplementary fig. S1, Supplementary Material online). The identification of 35 transcriptionally active *norgs* provides an evolutionary context for the previous finding that *norgs* can create new functional exons in the nuclear genome (Noutsos et al. 2007), though none of the genes described by Noutsos et al. (2007) appeared in our analysis, probably because we examined only genes that showed at least ten reads per CPIN.

In order to determine whether organellar promoters function directly in the nucleus, the available 35 *norg* sequences were examined for organelle-type promoters. If promoter motifs were identified within 1 kb of upstream flanking DNA of the chloroplast or mitochondrial gene, the *norg* was classified as holding an organellar promoter. Overall, *norgs* with



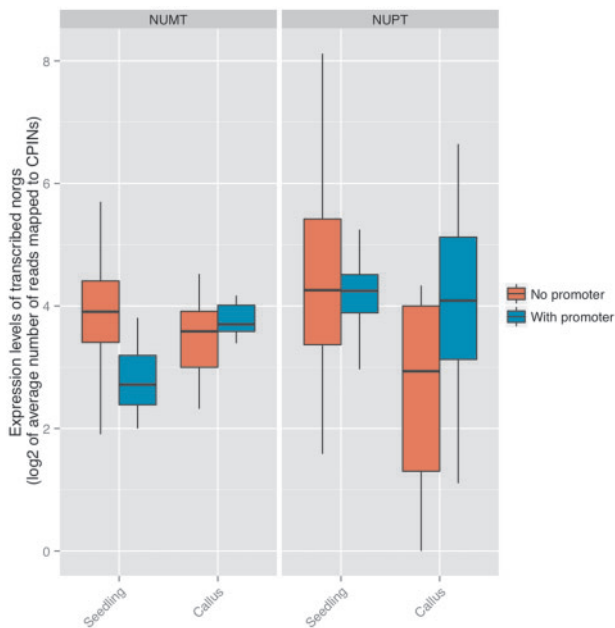
**Fig. 4.**—Plots of *norg* activity with respect to other nuclear genes. Transcription of annotated nuclear genes and *norgs* is shown in the top and bottom panels, respectively. The transcription for each annotated gene was normalized as follows: The midbase for each exon of subsp. *japonica* annotated gene was chosen, and then short reads uniquely mapped to the midbase without mismatches were counted to assess the transcription of each exon. Finally, transcription of the annotated gene was normalized using the mean transcription of all exons. Each dot shows the transcription level (number of mapped reads in the coordinate axes) of nuclear genes (top) or *norgs* (bottom).

**Table 1**

Information of Transcribed *norgs*

Tissue	No. of <i>norgs</i>	Overlapped or in Gene	Location in Gene	Flanked DH Sites
Seedling	17	3	1 is exon; 2 are 3'-UTRs	9
Callus	12	8	8 are exons	4
Both	6	4	3 are exons; 1 is 3'-UTR	2 (seedling) 4 (callus)

NOTE.—UTR, untranslated region.



**FIG. 5.**—Plots of activity of *norgs* containing or lacking organellar promoters. A total of 35 transcribed *norgs* (15 *nupts* and 20 *numts*) were classified into two groups: Those harboring and those lacking organellar promoters. Their transcriptional levels were compared separately in seedling and callus. The black line inside each box indicates the median transcription level in each RNA sample.

promoters identified in this way showed more transcripts than those that did not, but this difference was not statistically significant. More detailed analysis revealed a complex picture. Fifteen *norgs* (11 *nupts* and 4 *numts*) were found to harbor organellar promoters, whereas 20 (4 *nupts* and 16 *numts*) lacked promoters (fig. 5). Transcript levels of *numts* with promoters were significantly lower than the ones without promoters (Welch two sample *t*-test,  $P=5.485 \times 10^{-8}$ ) in seedling RNA, but there was an opposite trend in callus ( $P=0.009571$ ). The transcriptional level of *nupts* containing promoters was significantly higher than those without promoters in callus tissue ( $P=3.718 \times 10^{-6}$ ), but this difference was not significant in seedling RNA ( $P=0.07715$ ). This result indicates that organellar promoters may be active in the nucleus and that organellar promoters exhibit different transcriptional activity in different plant tissues. As the number of *norgs* we studied here is very limited, further work is required to elucidate these complex patterns.

Next, we investigated the connection between highly expressed *norgs* and open chromatin status by crosschecking *norgs* and 1 kb of flanking DNA with DNase I hypersensitive (DH) sites (Zhang et al. 2012) in *O. sativa* subsp. *japonica*. In seedling tissues, 9 of the 17 cases examined were located in open chromatin regions and 4 *norgs* of 12 tested were seen in callus tissues (table 1). For *norgs* transcribed in both tissues,

two of the six identified were located in open chromatin regions in seedling tissues and four in callus. Considering the possibility that a nuclear promoter was adjacent to the *norg*, 1-kb flanking DNA of these *norgs* in table 1 were compared with annotated rice promoter regions in PlantProm (Shahmuradov et al. 2003), and only one *norg* transcribed in seedling tissue was found to overlap with a promoter within 1 kb of its flanking DNA, but none was found in callus. Thus we conclude that certain rare *norgs* are transcribed because of fortuitous sequence motifs that occur in the mitochondrial or chloroplast genomes, but a larger number can achieve transcription by occupying a region of active open chromatin. However, the highest *norg* transcription levels are found after insertion into an existing nuclear gene.

## Discussion

Acquiring a eukaryotic promoter was thought to be essential for functional organellar gene transfer, but recent research (Lloyd and Timmis 2011) showed that transfer of multiple copies of the chloroplast *psbA* promoter-driven reporter gene leads directly to successful expression in the nucleus. Here we describe that the *16S rrr* plastid promoter is, like that of *psbA*, active in nucleus. Both *psbA* and *16S rrr* promoters are transcribed in the chloroplast by the plastid-encoded RNA polymerase (Sriraman et al. 1998; Hayashi et al. 2003) and they are most likely to be transcribed by nuclear DNA-dependent RNA polymerase II after they migrate to nucleus, suggesting that simple gene transfer from plastid-to-nucleus is immediately sufficient for transcriptional activity for some plastid genes. However, the power of these plastid promoters is much lower than that of evolved nucleus-specific promoters, a trend that we have now confirmed by whole genome analysis of *norg* transcription patterns in *O. sativa* subsp. *japonica* seedling and callus. Nonetheless, nuclear gene-comparable transcription levels were found for a set of 35 *norgs* and further analysis showed that a large proportion (15) of these were inserted into known nuclear genes. This finding demonstrates that plastid and mitochondrial DNAs not only diversify the nuclear genome but also contribute significantly to the transcriptome in rice. Interestingly, the number of actively transcribed *norgs* is different between seedling and callus tissues. This tissue-specific transcription of *norgs* is not surprising, because transcript accumulation levels of the large *numt* in *Arabidopsis thaliana* can be significantly increased by heat treatment (Pecinka et al. 2010; Tittel-Elmer et al. 2010), suggesting that cytoplasmic organellar promoters are environmentally responsive although they are weak under normal conditions. In addition, *norgs* may be transcribed by RNA polymerases IV or V, the homologs of DNA-dependent RNA polymerase II, and involved in RNA-directed DNA methylation. It will be interesting to study whether *norgs* are involved in nuclear DNA modification.

Plastid DNA is reported to contribute regulatory elements to a mitochondrial gene in its 3'-UTR region (Wang, Rousseau-Gueutin, et al. 2012). The finding of some *norgs* located in the 3'-UTR of annotated nuclear genes (table 1) suggests that cytoplasmic organellar DNA may likewise donate gene regulatory elements in the nucleus, and this thought is also supported by that *norgs* coincide with DH sites which can signify regulatory elements.

## Materials and Methods

### Plant Materials and PCR Assay

Tobacco plants were grown in sterile jars on half MS medium as previously described (Wang, Lloyd, et al. 2012). DNA and RNA were prepared using a DNeasy Plant Mini Kit (Qiagen) and an RNeasy Plant Mini Kit (Qiagen), respectively, according to the manufacturer's instructions.

Real-time quantitative PCR was performed as described previously (Yang et al. 2005). Primers used for determining copy number of *aadA* gene were 5'-GACTACCTTGGTGATC GCCTTTC-3' and 5'-GCCCCTCATACTTGAAGCTAGACAG-3'. For normalizing the total DNA contents used in quantitative PCR, *RPL25* amplified with primers 5'-CCCCTCACACAGAG TCTGC-3' and 5'-AAGGGTGTGTTGTCCTCAATCTT-3' was used as an internal standard (Schmidt and Delaney 2010).

Preparation of cDNA used in standard RT-PCR was carried out according to previous description (Wang, Lloyd, et al. 2012). Amplification of *aadA* cDNA used primers 5'-AGTAT CGACTCAACTATCAGAGG-3' and 5'-GACTACCTTGGTGATC TCGCCTTTC-3' (Lloyd and Timmis 2011).

### DNase I-PCR

DNase I-PCR was carried out as described (Shu et al. 2013). Primers used for determining chromatin status at the preinsertion site were 5'-GGGGTGGCCTGGTGGAAT-3' and 5'-TGG CCAGACGGGCCCCCTAAA-3'. A heterochromatic region used as negative control was amplified with primers 5'-GCTGCCTA TCGCGTTTCATCCAA-3' and 5'-CGGCCATATCGCTCTACCT CTTCG-3'. Primer pairs utilized for studying chromatin compaction after chloroplast DNA insertion in kr2.2 were 5'-CCCT CAGGGTGGCCTGGT-3' and 5'-TTCGGCAGCGGATCGCG AAA-3' for left section; 5'-TCCGACCCCTTTCCTTA GCGG-3' and 5'-ACCCACCTTCCCAGACCCT-3' for right section; 5'-GGCCACTCGAGGTCCTCTCCAAAT-3' and 5'-CG GAGAATCTCGCTCTCCAGGG-3' for middle section.

### Identification of *O. sativa* subsp. *japonica* *norgs*

The chloroplast, mitochondrial, and nuclear genome sequences and annotation data of *O. sativa* subsp. *japonica* were downloaded from TIGR database (Release 5). *Nupts* and *numts* present in the subsp. *indica* genome were identified by using local BLASTN (version 2.2.23) (Altschul et al.

1990) with the parameters previously described (Wang, Rousseau-Gueutin, et al. 2012).

### RNA-seq and Open Chromatin Regions

To investigate the transcription pattern of *norgs* from *O. sativa* subsp. *japonica* at whole-genome level, we downloaded RNA-seq and open chromatin data generated by DNase-seq (Zhang et al. 2012). The original FASTQ sequence file of RNA-seq data was downloaded from National Center for Biotechnology Information Gene Expression Omnibus (accession numbers: GSE26610 and GSE26734) and mapped back to *O. sativa* subsp. *japonica* genome (TIGR release 5) with no mismatch allowed using GSNAP (Wu and Nacu 2010). Then uniquely identified short reads were filtered for further analyses. CPINs were identified by pairwise sequence alignment between the nuclear DNA and chloroplast/mitochondrial DNA with LASTZ (Harris 2007). Reads mapped to CPINs were counted to represent the transcripts of *norgs*. The expression of genes was calculated as following description: Reads mapped to middle-base of each exon were counted as the expression of this exon, and then the mean value of all exons' expression was calculated as the expression of this gene. The read counting and overlap of genes and *norgs* were determined by R package "GenomicFeatures," and plots were generated with R package "ggplot2" (Wickham 2009; R Development Core Team 2010; Lawrence et al. 2013).

To compare the transcriptional levels of *norgs* with or without potential organelle promoters, we checked the genomic coordinates of 35 transcribed *norgs* in the chloroplast or mitochondrial genome. If a *norg* contained sequences that were within 1-kb upstream of an organellar gene, it was classified as *norg* with a potential organellar promoter. The average number of reads mapped to CPINs of each *norg* was calculated to reveal the transcriptional level of each *norg*. The transcription level difference between *norgs* (*nupts* and *numts*) with and without potential organelle promoters was calculated and tested using a Welch two sample *t*-test.

A Perl script was written for determining whether any highly transcribed *norg* insertion colocalized with nuclear gene. Open chromatin status of individual *norgs* was checked as previously described (Wang and Timmis 2013). Annotated rice promoter regions were retrieved from PlantProm (Shahmuradov et al. 2003). The overlap between promoter regions and 1-kb flanking regions of selected *norgs* was checked using R package "GenomicFeatures."

## Supplementary Material

Supplementary figure S1 is available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).



## Acknowledgments

The authors thank Huan Shu and Lars Henning for helpful advice on DNase I-PCR, and Jinbao Gu for technical assistance. This work was supported by the Chinese Academy of Sciences.

## Literature Cited

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* 215:403–410.
- Bock R, Timmis JN. 2008. Reconstructing evolution: gene transfer from plastids to the nucleus. *Bioessays* 30:556–566.
- Cornelissen M, Vandewiele M. 1989. Nuclear transcriptional activity of the tobacco plastid psbA promoter. *Nucleic Acids Res.* 17:19–29.
- Gray MW, Burger G, Lang BF. 1999. Mitochondrial evolution. *Science* 283:1476–1481.
- Harris RS. 2007. Improved pairwise alignment of genomic DNA [PhD thesis]. The Pennsylvania State University.
- Hayashi K, et al. 2003. A role of the -35 element in the initiation of transcription at psbA promoter in tobacco plastids. *Plant Cell Physiol.* 44:334–341.
- Hogan GJ, Lee CK, Lieb JD. 2006. Cell cycle-specified fluctuation of nucleosome occupancy at gene promoters. *PLoS Genet.* 2:e158.
- Huang CY, Ayliffe MA, Timmis JN. 2003. Direct measurement of the transfer rate of chloroplast DNA into the nucleus. *Nature* 422:72–76.
- Kim A, Song SH, Brand M, Dean A. 2007. Nucleosome and transcription activator antagonism at human beta-globin locus control region DNase I hypersensitive sites. *Nucleic Acids Res.* 35:5831–5838.
- Lawrence M, et al. 2013. Software for computing and annotating genomic ranges. *PLoS Comput Biol.* 9:e1003118.
- Lloyd AH, Timmis JN. 2011. The origin and characterization of new nuclear genes originating from a cytoplasmic organellar genome. *Mol Biol Evol.* 28:2019–2028.
- Noutsos C, Kleine T, Armbruster U, DalCorso G, Leister D. 2007. Nuclear insertions of organellar DNA can create novel patches of functional exon sequences. *Trends Genet.* 23:597–601.
- Pecinka A, et al. 2010. Epigenetic regulation of repetitive elements is attenuated by prolonged heat stress in *Arabidopsis*. *Plant Cell* 22:3118–3129.
- R Development Core Team. 2010. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing.
- Ricchetti M, Fairhead C, Dujon B. 1999. Mitochondrial DNA repairs double-strand breaks in yeast chromosomes. *Nature* 402:96–100.
- Schmidt GW, Delaney SK. 2010. Stable internal reference genes for normalization of real-time RT-PCR in tobacco (*Nicotiana tabacum*) during development and abiotic stress. *Mol Genet Genomics.* 283:233–241.
- Shahmuradov IA, Gammerman AJ, Hancock JM, Bramley PM, Solovyev VV. 2003. PlantProm: a database of plant promoter sequences. *Nucleic Acids Res.* 31:114–117.
- Sheppard AE, et al. 2008. Transfer of plastid DNA to the nucleus is elevated during male gametogenesis in tobacco. *Plant Physiol.* 148:328–336.
- Shu H, GUISSEM W, Hennig L. 2013. Measuring *Arabidopsis* chromatin accessibility using DNase I-polymerase chain reaction and DNase I-chip assays. *Plant Physiol.* 162:1794–1801.
- Song L, et al. 2011. Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Res.* 21:1757–1767.
- Sriraman P, Silhavy D, Maliga P. 1998. Transcription from heterologous rRNA operon promoters in chloroplasts reveals requirement for specific activating factors. *Plant Physiol.* 117:1495–1499.
- Stegemann S, Bock R. 2006. Experimental reconstruction of functional gene transfer from the tobacco plastid genome to the nucleus. *Plant Cell* 18:2869–2878.
- Stegemann S, Hartmann S, Ruf S, Bock R. 2003. High-frequency gene transfer from the chloroplast genome to the nucleus. *Proc Natl Acad Sci U S A.* 100:8828–8833.
- Thorsness PE, Fox TD. 1990. Escape of DNA from mitochondria to the nucleus in *Saccharomyces cerevisiae*. *Nature* 346:376–379.
- Timmis JN, Ayliffe MA, Huang CY, Martin W. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat Rev Genet.* 5:123–135.
- Tittel-Elmer M, et al. 2010. Stress-induced activation of heterochromatic transcription. *PLoS Genet.* 6:e1001175.
- Tsuji J, Frith MC, Tomii K, Horton P. 2012. Mammalian NUMT insertion is non-random. *Nucleic Acids Res.* 40:9073–9088.
- Wang D, Lloyd AH, Timmis JN. 2012. Environmental stress increases the entry of cytoplasmic organellar DNA into the nucleus in plants. *Proc Natl Acad Sci U S A.* 109:2444–2448.
- Wang D, Rousseau-Gueutin M, Timmis JN. 2012. Plastid sequences contribute to some plant mitochondrial genes. *Mol Biol Evol.* 29:1707–1711.
- Wang D, Timmis JN. 2013. Cytoplasmic organelle DNA preferentially inserts into open chromatin. *Genome Biol Evol.* 5:1060–1064.
- Wickham H. 2009. ggplot2: elegant graphics for data analysis. New York: Springer.
- Wu TD, Nacu S. 2010. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* 26:873–881.
- Yang L, et al. 2005. Estimating the copy number of transgenes in transformed rice by real-time quantitative PCR. *Plant Cell Rep.* 23:759–763.
- Zhang W, et al. 2012. High-resolution mapping of open chromatin in the rice genome. *Genome Res.* 22:151–162.

Associate editor: John Archibald