Correlation between Nuclear Plastid DNA Abundance and Plastid Number Supports the Limited Transfer Window Hypothesis

David Roy Smith*,†, Kate Crosby†, and Robert W. Lee†

†Department of Biology, Dalhousie University, Halifax, Nova Scotia, Canada
†Present address: Canadian Institute for Advanced Research, Department of Botany, University of British Columbia, Vancouver, British Columbia, Canada
*Corresponding author: E-mail: smithdr@dal.ca.

Abstract

The abundance of nuclear plastid DNA-like sequences (NUPTs) in nuclear genomes can vary immensely; however, the forces responsible for this variation are poorly understood. “The limited transfer window hypothesis” predicts that species with only one plastid per cell will have fewer NUPTs than those with many plastids per cell, but a lack of genome sequence data from monoplastidic species has made this hypothesis difficult to test. Here, by analyzing newly available genome sequences from diverse mono- and polyploid taxa, we show that the hypothesis holds. On average, the polyploid species we studied had 80 times more NUPTs than those that were monoplastic. Moreover, NUPT content was positively related to nuclear genome size, indicating that in addition to plastid number, NUPTs are influenced by the forces controlling the expansion and contraction of noncoding nuclear DNA. These findings are consistent with data on nuclear DNAs of mitochondrial origin (NUMTs), suggesting that similar processes govern the abundance of both NUPTs and NUMTs.

Key words: chloroplast, mitochondria, NUPT, NUMT, genome architecture, noncoding DNA.

The Limited Transfer Window Hypothesis

The movement of organelle DNA to the nucleus has been, and remains, a driving force in fashioning eukaryotic genomes (Timmis et al. 2004; Kleine et al. 2009). Early on in both mitochondrial and plastid evolution, there was a massive migration of organelle genes to the nuclear genome (Gray et al. 1999; Huang et al. 2003; Kleine et al. 2009); thus, present-day nuclear DNA is a mosaic of endosymbiont-derived organelle genes and “host” genes, and contemporary mitochondrial and plastid DNAs (mtDNAs and ptDNAs) are significantly more reduced than the endosymbiotic genomes from which they evolved (Gray et al. 1999; Archibald 2009). Aside from adding to the gene repertoire of nuclear genomes, organelle-to-nucleus DNA transfer events have generated, and continue to generate, forms of noncoding nuclear DNA (and occasionally exonic nuclear DNA, Noutsos et al. 2007) that share sequence identity with the coexisting organelle DNAs; these types of sequences are referred to as nuclear mitochondrial DNAs (NUMTs) and nuclear plastid DNAs (NUPTs) (Lopez et al. 1994; Richly and Leister 2004a, 2004b).

Although the nuclear genomes from at least 85 eukaryotic species have been analyzed for NUMTs (Hazkani-Covo et al. 2010 and references therein), there are relatively little data on NUPTs. This is because, until recently, there were only a small number of published nuclear genome sequences from plastid-harboring eukaryotes. Nonetheless, an intriguing observation has come from the NUPT data that are available: Species with one plastid per cell (monoplastic) have fewer NUPTs than those with many plastids per cell (polyplastic) (Lister et al. 2003; Martin 2003; Richly and Leister 2004b). For example, the monoplastic protists Chlamydomonas reinhardti and Plasmodium falciparum have <2.5 kb of NUPTs (Richly and Leister 2004b; Matsuo et al. 2005), whereas rice, which contains upwards of 100 plastids per cell, has around 900 kb of NUPTs (Guo et al. 2008). A possible explanation for these observations is that in monoplastic species, the transfer of ptDNA to the
nucleus is greatly reduced as compared with polyploidy
taxa because 1) there are fewer plastids to donate ptDNA
to the nucleus and 2) lysis of the plastid would almost cer-
tainly result in death to the cell, unlike the case for polyploidy-
species (Lister et al. 2003; Martin 2003; Richly and
Leister 2004b). This explanation has become known as the
“limited transfer window hypothesis” (Barbrook et al.
2006). When presenting this hypothesis, Barbrook et al.
(2006) predicted that “the sequencing of the nuclear ge-
neromes of organisms containing a single plastid should al-
ways reveal a low abundance of NUPTs.” But a lack of
nuclear DNA sequence data from monoploidy species
and from plastid-harboring taxa in general has made this
prediction difficult to test.

In this study, we take advantage of newly available geno-
ic sequence data from a series of diverse mono- and poly-
plastidic species to formally investigate the limited transfer
window hypothesis. Altogether, we calculate the number
and accumulative length of NUPTs in the nuclear DNAs of
11 polyploidy and 19 monoploidy (or effectively mono-
plastidic) eukaryotes. When possible, we also analyze these
same genomes for NUMTs and compare these data with the
corresponding NUPT statistics.

Testing the Limited Transfer Window
Hypothesis

To assess a genome for NUPTs, at least two things are re-
quired: complete nuclear DNA and ptDNA sequence data.
We found 30 species for which both these statistics are avail-
able, including 13 land plants, 7 green algae, 5 apicomplex-
ians, 3 stramenopiles, 1 haptophyte, and 1 red alga (table 1).
The sources for these genome sequence data are listed in sup-
plementary tables S1 and S2 (Supplementary Material
online). To the best of our knowledge, detailed NUPT
statistics for the majority of the above-mentioned taxa have
not been published elsewhere. For 20 of these species, com-
plete mtDNA sequence data are also available, allowing for
NUMT as well as NUPT analyses. Although most of these taxa
have already been explored for NUMTs (Hazkani-Covo et al.
2010, and references therein), we performed our own NUMT
investigations because in the past differences in search pa-
rameters among studies have led to discrepancies in NUPT/
NUMT tabulations. We did try, however, to use similar search
constraints as those employed in previous reports: BlastN
with an expectation value of 0.0001. Another source of dis-
crepancy in NUPT/NUMT assessments among earlier stud-
ies (Hazkani-Covo et al. 2010) were instances where one
segment of nuclear DNA matched to multiple sections of
organelle DNA. In our analyses, multiple organelle DNA hits
to the same nuclear DNA regions were counted only once.
Because many of the nuclear genomes that we scanned are
only in their draft assembly stage, the NUPT/NUMT data
presented here should be treated as approximations of
the true values. As these genome sequences become more
polished, the NUPT/NUMT estimates will change, but the
major trends that we observed among the different groups
should arguably remain the same.

Of the 30 species we investigated, 11 are polyploidy
and 19 are either monoploidy or effectively so. Thirteen
of the monoploidy species are also monomitochondrial
(i.e., they have one mitochondrion per cell). The number of
organelles per cell for each species and the references
used to determine these statistics are listed in table 1 and
supplementary table S3 (Supplementary Material online),
respectively. When possible, the decision to categorize a spe-
cies as having one or multiple organelles per cell was based
on published ultrastructural data. Two caveats should be
noted: The haptophyte Emiliania huxleyi and the stramenopile
Thalassiosira pseudonana can have up to two plastidst per cell but, for simplicity, were treated here as monoploidy
(Bagder et al. 1998; Dassow et al. 2008, and references
therein); and the lycophyte Selaginella moellendorfii and
the moss Physcomitrella patens both contain cells that are
polyploidy but, for the purpose of this study, they were
considered “effectively monoploidy” because mitosis and
meiosis only occur in cells that contain a single plastid
(Brown and Lemmon 1990).

Polyplastidy Means More NUPTs

Complete NUPT and, when attainable, NUMT statistics for
the various plastid-bearing species that we investigated
are shown in table 1. Overall, we found the difference in
NUPT content between mono- and polyploidy species to be
highly significant (fig. 1). Species with multiple plastidst per cell had on average 80 times more NUPTs than those
with one plastid per cell. The mean NUPT content for poly-
plastidic species was 460 kb as compared with only 6 kb
for monoploidy taxa. Moreover, the average number of
NUPTs (based on Blast hits, not accumulative length) for
polyploidy individuals was 20 times greater than that of
monoploidy species (1,540 vs. 79 hits). In species with only
a single plastid, the NUPT content ranged from undetectable
levels in the proists Aureococcus anophagaeferens, Babesia
bovis, Ostreococcus sp. RCC809, and Theileria parva to
~65 kb for the multicellular green alga Volvox carteri,
whereas in polyploidy species, it spanned from 50 kb
for Arabidopsis thaliana to >800 kb for Glycine max, Vitis
vinifera, and Oryza sativa. There is a clear separation in NUPT
content between mono- and polyploidy species, with the
members of the latter group having considerably more
NUPTs than the former (figs. 1 and 2). The only exceptions
were V. carteri and A. thaliana (fig. 2); the reasons for this
may be linked to their capacity (or deficiency) for purging
bulk nuclear DNA (discussed further below). It is also note-
worthy that of the 13 land plants that were explored the 2
that are effectively monoploidy (S. moellendorfii and

Downloaded from https://academic.oup.com/gbe/article-abstract/doi/10.1093/gbe/evr001/575194 by guest on 19 April 2019
NUPTs in Monoplastidic versus Polyplastidic Species

**Table 1**
Number and Total Amount (in Kilobases) of NUPTs and NUMTs in the Available Nuclear Genome Sequences from Plastid-Harboring Eukaryotes

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Number of Plastids per Cell</th>
<th>Number of Mitochondria per Cell</th>
<th>NUPTs</th>
<th>Number of Blast Hits</th>
<th>Accumulative Length (kb)</th>
<th>Average Length (kb)</th>
<th>NUMTs</th>
<th>Number of Blast Hits</th>
<th>Accumulative Length (kb)</th>
<th>Average Length (kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Land plants</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arabidopsis thaliana</td>
<td>Multiple</td>
<td>Multiple</td>
<td>332</td>
<td>50</td>
<td>0.15</td>
<td>1,173</td>
<td>549</td>
<td>0.46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brachypodium distachyon</td>
<td>Multiple</td>
<td>Multiple</td>
<td>310</td>
<td>114</td>
<td>0.37</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canica papaya</td>
<td>Multiple</td>
<td>Multiple</td>
<td>839</td>
<td>291</td>
<td>0.34</td>
<td>1,528</td>
<td>467</td>
<td>0.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cucumis sativus</td>
<td>Multiple</td>
<td>Multiple</td>
<td>751</td>
<td>265</td>
<td>0.35</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glycine max</td>
<td>Multiple</td>
<td>Multiple</td>
<td>3,414</td>
<td>822</td>
<td>0.24</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medicago truncatula</td>
<td>Multiple</td>
<td>Multiple</td>
<td>258</td>
<td>93.3</td>
<td>0.36</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Onzya sativa subsp. indica</td>
<td>Multiple</td>
<td>Multiple</td>
<td>1,541</td>
<td>782</td>
<td>0.50</td>
<td>2,544</td>
<td>818</td>
<td>0.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>O. sativa subsp. japonica</td>
<td>Multiple</td>
<td>Multiple</td>
<td>2,036</td>
<td>1,073</td>
<td>0.52</td>
<td>3,072</td>
<td>834</td>
<td>0.27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physcomitrella patens</td>
<td>Effectively monoplastidicb</td>
<td>Multiple</td>
<td>31</td>
<td>5</td>
<td>0.16</td>
<td>294</td>
<td>74</td>
<td>0.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Populus trichocarpa</em></td>
<td>Multiple</td>
<td>Multiple</td>
<td>2,036</td>
<td>428</td>
<td>0.30</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Selaginella moellendorfii</em></td>
<td>Effectively monoplastidicb</td>
<td>Multiple</td>
<td>114</td>
<td>11.4</td>
<td>0.10</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Sorghum bicolor</em></td>
<td>Multiple</td>
<td>Multiple</td>
<td>1,574</td>
<td>329</td>
<td>0.20</td>
<td>1,957</td>
<td>406</td>
<td>0.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Vitis vinifera</em></td>
<td>Multiple</td>
<td>Multiple</td>
<td>3,858</td>
<td>801</td>
<td>0.20</td>
<td>2,357</td>
<td>602</td>
<td>0.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Green algae</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chlamydomonas reinhardtii</td>
<td>Single</td>
<td>Multiple</td>
<td>35</td>
<td>1.9</td>
<td>0.05</td>
<td>35</td>
<td>3.3</td>
<td>0.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coccomyxa sp. C-169</td>
<td>Single</td>
<td>Single</td>
<td>73</td>
<td>7.5</td>
<td>0.10</td>
<td>107</td>
<td>12</td>
<td>0.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Ostreococcus</em> sp. RCC809</td>
<td>Single</td>
<td>Single</td>
<td>3</td>
<td>0.6</td>
<td>0.12</td>
<td>2</td>
<td>0.6</td>
<td>0.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Ostreococcus</em> tauri</td>
<td>Single</td>
<td>Single</td>
<td>3</td>
<td>0.5</td>
<td>0.16</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Micromonas pusilla</td>
<td>Single</td>
<td>Single</td>
<td>3</td>
<td>0.5</td>
<td>0.20</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Micromonas sp. RCC299</td>
<td>Single</td>
<td>Single</td>
<td>1,100</td>
<td>65</td>
<td>0.12</td>
<td>802</td>
<td>33</td>
<td>0.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Volvox carteri f. nagariensis</td>
<td>Single</td>
<td>Multiple</td>
<td>2</td>
<td>0.37</td>
<td>0.18</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Red algae</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyanidioschyzon merolae</td>
<td>Single</td>
<td>Single</td>
<td>2</td>
<td>0</td>
<td>0.18</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apicomplexans</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Babesia bovis</td>
<td>Single</td>
<td>Single</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Eimeria tenella</em></td>
<td>Single</td>
<td>Single</td>
<td>31</td>
<td>2.8</td>
<td>0.09</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Plasmodium falciparum</em></td>
<td>Single</td>
<td>Single</td>
<td>2</td>
<td>0.11</td>
<td>0.05</td>
<td>2</td>
<td>0.11</td>
<td>0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Theileria parva</em></td>
<td>Single</td>
<td>Single</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Toxoplasma gondii</em></td>
<td>Single</td>
<td>Single</td>
<td>77</td>
<td>10.3</td>
<td>0.03</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Haptophyte</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emiliania huxleyi</td>
<td>1–2</td>
<td>Single</td>
<td>2</td>
<td>0.15</td>
<td>0.07</td>
<td>2</td>
<td>0.1</td>
<td>0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Stramenopiles</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aureococcus anophagefferens</td>
<td>Single</td>
<td>Single</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phaeodactylum tricornutum</td>
<td>Single</td>
<td>Multiple</td>
<td>14</td>
<td>4</td>
<td>0.29</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thalassiosira pseudonana</td>
<td>1–2</td>
<td>Multiple</td>
<td>8</td>
<td>1.6</td>
<td>0.20</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*NOTE.—* NA—data not available (i.e., mitochondrial genome has not been sequenced; thus, we were unable to search the nuclear DNA for NUMTs).

*Bold parameters were as follows: BlastN (version 2.2.23) with an expectation value of 0.0001; a word size of 11; match and mismatch scores of 2 and −3, respectively; and gap cost values of 5 (existence) and 2 (extension). Multiple organelle DNA hits to the same nuclear DNA regions were counted only once. Regions of nuclear DNA that contained tight clusters of NUPTs/NUMTs (i.e., sections of organelle-like DNA interrupted by genomic sequence that did not show sequence identity to organelle DNA) were not counted as a single NUPT/NUMT but as separate hits. See supplementary table S3 (Supplementary Material online) for references and notes on the number of organelles per cell.

b *S. moellendorfii* and *P. patens* both contain cells that are polyplastidic, but for the purpose of this study, they are considered “effectively monoplastidic” because mitosis and meiosis only occurs in cells that contain a single plastid (Brown and Lemmon 1990).

*P. patens* had about 60 times fewer NUPTs than those that are polyplastidic (figs. 1 and 2).

Our NUPT analyses revealed similar trends and conclusions as those described above for NUPTs. Species with only a single mitochondrion per cell had significantly fewer NUPTs than those with many mitochondria per cell (figs. 1 and 2). The average NUPT content for monomitochondrial species (1.1 kb) was ~300 times less than that of poly-mitochondrial taxa (380 kb). These data are consistent with earlier observations on NUMTs (Hazzkani-Covo et al. 2010) and support the belief that organelle number influences both NUPT and NUMT content.

**Larger Genomes, Larger NUPT Content**

It was recently shown that NUMT content scales positively with nuclear genome size (Hazzkani-Covo et al. 2010). Here, we found that this is true for NUPTs as well: Bloated nuclear
genomes tend to have more NUPTs than those that are compact (fig. 3)—based on the 30 species investigated here, we found a reasonably strong relationship between nuclear genome size and NUPT content ($R^2 = 0.57$, $P = 8.6 \times 10^{-7}$). We expected to find this relationship because NUPTs are a type of excess DNA, and it is well established that all types of excess DNA mutually expand as the number of nucleotides in a genome increases (Lynch and Conery 2003). Ultimately, this suggests that the forces governing the expansion and contraction of noncoding nuclear DNA impact the accumulative length of NUPTs in a nuclear genome. Although the nature of these forces is hotly debated, there is evidence that the tendency for excess DNA to accumulate depends on the combined effects of the mutation rate ($\mu$) and the effective genetic population size ($N_e$; Lynch and Conery 2003). According to this hypothesis, one may expect species with a low $N_e \mu$ to have more NUPTs than those with a high $N_e \mu$. Although there are very few reliable data on this fundamental population genetic parameter, a recent study indicates that V. carteri has a very small $N_e \mu$, especially relative to other protists (Smith and Lee 2010). This could help explain why all the monoplastidic species that we studied, V. carteri had the largest NUPT content. Being monoplastidic, one would expect the transfer of ptDNA to the nuclear genome to be rare in V. carteri, but having a low $N_e \mu$ implies that it has a reduced ability to detect and eradicate excess DNA (Smith and Lee 2010) so that the few NUPTs that do arise avoid deletion and therefore can accumulate to reasonably high levels over time. Interestingly, the $N_e \mu$ estimates for the nuclear DNA of Arabidopsis spp. are about three times those of V. carteri (Wright et al. 2008); thus, one explanation for why A. thaliana, which is monoplastidic, had fewer NUPTs than the monoplastidic V. carteri could be that it is reasonably efficient at perceiving and purging non-coding nuclear DNA. It is worth mentioning that the V. carteri ptDNA, at ~525 kb, is the largest plastid genome sequenced to date (Smith and Lee 2010), being >300 kb larger than any other ptDNA employed in our data set. And although we did not find an association between plastid genome size and NUPT content (supplementary fig. S1, Supplementary Material online), there is still the possibility that the prodigious ptDNA of V. carteri is in some way contributing to its elevated NUPT content.

**The Evolution of NUPTs: It Is a Give and Take Relationship**

The data presented here provide support for the limited transfer window hypothesis and the notion that the number of plastids per cell in a eukaryotic species governs the amount of NUPTs found in its nuclear genome. We argue that the evolution of NUPTs is a “give and take” process where plastid number determines the potential for ptDNA to be donated (i.e., given) to the nuclear genome, and the

---

**Fig. 1.**—“Beanplot” depicting the difference in NUPT/NUMT content between mono- and polyplastidic (and polymitochondrial) species. Plot was generated using the beanplot package (Kampstra 2008) from R v. 2.1.1. The dashed line in the middle of each of the two plots is the overall average of the continuous variable (total NUPT or NUMT content in kilobases). The thick black line in the middle of each category (mono or poly) is the median of each continuous variable (NUPT or NUMT content) with respect to the categorical variable (mono- or polyplastidic/polymitochondrial). The colored curved beanpod surrounding the observations “beans” is the theoretical probability density distribution of these observations. If there were multiple observations with the same number (e.g., NUPTs content of 4 kb for two different taxa), then the line gets longer respective to the other measurements in the beanplot. A Wilcoxon signed rank test (nonparametric) was performed in R on all data because errors were not normally distributed. Note, the $P$ values shown are approximations—the exact values could not be computed due to ties in the data. For polyorganellar species, the lowest and highest data points are named. However, to avoid clutter, only the highest points are labeled for mono-organellar species. Note: At, Arabidopsis thaliana; Cc, Coccomyxa sp. C-169; Os, Oryza sativa; Ts, Thalassiosira pseudonana; and Vc, Volvox carteri.
probability that these ptDNA sequences will be accepted (i.e., taken) by the nuclear genome and persist as nuclear DNA is determined by a species’ ability (or lack thereof) to detect and eliminate excess DNA. In this study, being monoplastidic versus polyplastidic was used to define the number of plastids per cell in a species, and nuclear genome size was used as a proxy for defining a species’ ability to eradicate noncoding nuclear DNA. This same argument applies to NUMTs as well.

Again, it must be stressed that many of the nuclear DNA sequences that we used to calculate NUPT/NUMT abundance were in their draft assembly stage. As these genome assemblies improve, their NUPT/NUMT statistics may change, but we believe that the major trends reported here will be borne out by future investigations. Finally, given the wide diversity of eukaryotic species that we explored, there are certainly factors in addition to a limited transfer window and susceptibility to bulk DNA that are influencing NUPT and NUMT content; however, we argue these additional factors (whatever they may be) will turn out to be secondary to the forces outlined here.

Materials and Methods

The sources and references for the nuclear genome sequences employed in this study, as well the GenBank accession numbers, when available, are shown in supplementary table S1 (Supplementary Material online). All nuclear DNA data came from publicly available sources. The organelle DNA sequences (including their lengths, GenBank accession numbers, and noncoding DNA contents) used as queries for BlastN searches against nuclear genomes are listed in supplementary table S2 (Supplementary Material online). Some of the organelle DNA sequences that were used in this study are not deposited in GenBank but are available for download online from the given genome project Web site (Supplementary table S2, Supplementary Material online).
Nuclear genomes were scanned for NUPTs and NUMTs with BlastN (version 2.2.23) (Altschul et al. 1990) using the following parameters: an expectation value of 0.0001; a word size of 11; match and mismatch scores of 2 and \(-3\), respectively; and gap-cost values of 5 (existence) and 2 (extension). Hits under 30 nt and showing <70% sequence identity to the query were ignored. Spurious hits, such as those where organelle genes matched to homologous genes in the nuclear genome (e.g., organelle ribosomal DNA [rDNA] matching to nuclear rDNA) were ignored. Instances where one segment of nuclear DNA matched to multiple sections of organelle DNA (i.e., duplicate Blast hits) were reduced to a single Blast hit; in other words, NUPTs and NUMTs matching to multiple organelle genomic regions, such as repetitive elements, were counted only once. Regions of nuclear DNA that contained tight
clusters of NUPTs or NUMTs (i.e., sections of organelle-like DNA interrupted by genomic sequence that did not show sequence identity to organelle DNA) were not counted as a single NUPT or NUMT but as separate hits. Data and sources used to calculate the number of organelles per cell are shown in supplementary table S3 (Supplementary Material online).

Supplementary Material
Supplementary tables S1–S3 and figure S1 are available at Genome Biology and Evolution online (http://www.gbe.oxfordjournals.org/).

Acknowledgments
This work was supported by a grant to R.W.L. from the Natural Sciences and Engineering Research Council of Canada.

Literature Cited
Dassow PV, Petersen TW, Chepurnov VA, Armbrust EV. 2008. Inter- and intraspecific relationship between nuclear DNA content and cell size in selected members of the centric diatom genus Thalassiosira (Bacillariophyceae). J Phycol. 44:335–349.

Associated editor: Geoff McFadden