Genome Reports: Contracted Genes and Dwarfed Plastome in Mycoheterotrophic *Sciaphila thaidanica* (Triuridaceae, Pandanales)

Gitte Petersen*, Athanasios Zervas, Henrik Æ. Pedersen, and Ole Seberg

Natural History Museum of Denmark, University of Copenhagen, Denmark

*Corresponding author: E-mail: gittep@snm.ku.dk.

Accepted: March 15, 2018

Data deposition: The plastome sequence has been deposited in GenBank under the accession number MG757197.

Abstract

With a reduced need for photosynthesis, the plastome of parasitic and mycoheterotrophic plants degrades. In the tiny, fully mycoheterotrophic plant *Sciaphila thaidanica*, we find one of the smallest plastomes yet encountered. Its size is just 12,780 bp and it contains only 20 potentially functional housekeeping genes. Thus *S. thaidanica* fits the proposed model of gene loss in achlorophyllous plants. The most astonishing feature of the plastome is its extremely compact nature, with more than half of the genes having overlapping reading frames. Additionally, intergenic sequences have been reduced to a bare minimum, and the retained genes have been reduced in length both compared with the orthologous genes in another mycoheterotrophic species of *Sciaphila* and in the autotrophic relative *Carludovica*.

Key words: gene loss, genome reduction, mycoheterotrophy, plastid genome evolution.

Introduction

Sciaphila Blume is a genus of fully mycoheterotrophic plants that parasitize mycorrhizal fungi. The need for carbon fixation has been eliminated and photosynthesis has been lost completely. The loss of photosynthesis is shared between fully mycoheterotrophic and holoparasitic plants, and a number of studies have investigated how plastid genomes (plastomes), housing a large number of genes related to photosynthesis, evolve when photosynthesis is no longer needed (reviewed in Krause 2011; Wicke et al. 2011; Graham et al. 2017). Not surprisingly, plastomes gradually degrade as genes are under relaxed or no selective pressure, have increased substitution rates, are pseudogenized and eventually lost. Simultaneously, the length of noncoding sequences is reduced and structural rearrangements are more frequent. Most changes lead to compaction of the plastome and potentially it may be lost completely, as have been suggested in *Rafflesia* R. Br. (Molina et al. 2014). As the plastome includes both photosynthesis genes and housekeeping genes, gene loss is not random, but follows a pattern, which seems to be shared among the different clades of achlorophyllous plants (Barrett and Davis 2012; Barrett et al. 2014; Wicke et al. 2016; Graham et al. 2017).

Sciaphila belongs in the monocot family Triuridaceae (Pandanales). With more than 50 species, the Triuridaceae constitute one of the largest clades of fully mycoheterotrophic plants, but only one complete plastome from a species of *Sciaphila* has been described previously (Lam et al. 2015). In order to further explore plastome evolution in response to loss of photosynthesis further, we here describe a complete plastome sequence from another species of *Sciaphila*, *S. thaidanica* K. Larsen (fig. 1), and compare it to the previously sequenced plastome of *S. densiflora* Schltr. We also compare both *Sciaphila* plastomes with the recently sequenced plastome of an autotrophic representative of the Pandanales, *Carludovica palmata* Ruiz. & Pav. (Cyclanthaceae) (Lam et al. 2015), and assess whether *Sciaphila* fits the pattern of plastome modification previously described for achlorophyllous plants (Barrett and Davis 2012; Barrett et al. 2014; Wicke et al. 2016; Graham et al. 2017).

Materials and Methods

DNA Extraction, Library Preparation, and Sequencing

DNA was extracted from a silica gel sample of *Sciaphila thaidanica* (Collected: Thailand, Phang Nga Province, Sra Nang

Manora Forest Park, 17 Nov. 2014, specimen voucher Suddee et al. 4796; BKF) using the DNA Plant Minikit (Qiagen) according to the manufacturer's instructions, with the addition of a Proteinase K treatment step, for 2 hours at 65°C, immediately after the bead-beating step (Qiashredder).

The extracted DNA was sheared using a Bioruptor (Diagenode). A 100-µl dilution at a concentration of 10 ng/µl DNA was sheared using the following conditions: 5 cycles of 15 seconds ON, 90 seconds OFF. The sheared DNA was then run on a 2% agarose gel with a 100-bp DNA ladder (Thermo Scientific), to check the size variation of the DNA fragments, which were in the range of 200–600 bp. To prepare the fragments for inserting the adapter sequences and unique barcodes, the NEBNext DNA Sample Prep Master Mix Set 2 (New England Biolabs, Ipswich, Massachusetts) using blunt-end adapters specified by Meyer and Kircher (2010) was used. In order to determine the number of cycles needed for library construction, qPCR were performed on a 1:40 dilution of the DNA sample, using the SYBR Green Master Mix 1 (Agilent Technologies) on an Mx3500p qPCR machine (Agilent Technologies). Libraries were amplified 13 cycles in a total volume of 100 µl, containing 5 U AmpliTaq Gold polymerase (Applied Biosystems, Foster City, CA), 1× AmpliTaq Gold buffer, 2.5 mM $MgCl_2$, 0.4 mg/ml bovine serum albumin (BSA), 0.2 mM of each dNTP, 0.2 µM IS4 forward primer, 0.2 µM indexed reverse primer, and 20 µl DNA library template. Following amplification, libraries were purified using a QIAquick PCR Purification kit (Qiagen, Hilden, Germany), according to manufacturer's instructions. DNA was eluted in 32 µl EB buffer and the column was incubated for 10 minutes at 37°C prior to centrifugation. The libraries were first quantified on a Qubit 2.0 (Life Technologies, Carlsbad, CA) using a dsDNA high sensitivity assay, and then run on a TapeStation 2200 using the high sensitivity tapes (Agilent, Santa Clara, CA) to determine the average insert size and molarity of each library. About 100-bp paired-end sequencing was performed using an Illumina Hiseq 2500 at the National High-Throughput DNA Sequencing Centre.

Genome Assembly, Annotation, and Analysis

Raw reads were trimmed for quality, adapters, and unidentified nucleotides (Ns) using Adapter Removal (Lindgreen 2012). To roughly assess the gene content, reads were initially mapped to the complete plastome sequences of *Sciaphila densiflora* (NC027659) and *Carludovica palmata* (NC026786) using the Map to Reference option of Geneious version 11.0.2 (Biomatters Ltd) under default settings. As sequence reads only mapped to few genes, suggesting the plastome to be very small, we subsequently followed a procedure based on repeated reference mapping and de novo assembly for assembling a complete plastome: Consensus sequences calculated from the reads mapped to reference plastome genes were extracted and extended using several rounds of Map to Reference with custom settings using a higher sensitivity (maximum mismatches per read 5%; allow only 5% gaps per read) and up to 10 iterations where reads are automatically mapped to the consensus of the previous iteration. After each round of reference mapping, assemblies were inspected for errors and potential places where reads could split in two directions. Cases of the latter were not observed. Accepted new consensus sequences were used for de novo assembly using default options of Geneious version 11.0.2 (Biomatters Ltd). This procedure was continued until all contigs were assembled into a circular structure. The plastome genes were manually annotated in Geneious version 11.0.2 (Biomatters Ltd) following comparison to reference organelle genomes from Pandanales. Open reading frames of more than half the length of orthologous genes from autotrophic relatives were accepted as potentially functional genes.

Plastome sequences may be transferred to the mitochondrial genome, and this may confound plastome assembly. Thus, in order to roughly assess genomic location of sequences, we mapped the reads from *Sciaphila thaidanica* to five mitochondrial protein coding genes from *Carludovica palmata* available in GenBank (AF197734, AF197707, DQ406948, DQ508954, GU3511605) to determine mitogenome coverage.

Sequence similarity of the coding sequences retained in *Sciaphila* and those in *Carludovica palmata* were calculated from pairwise alignments using MUSCLE version 3.8.425 (Edgar 2004) as implemented in Geneious version 11.0.2 (Biomatters). For protein coding genes, we applied the translation alignment option.

Results and Discussion

From approximately 42 million paired-end reads, the plastome of *Sciaphila thaidanica* was assembled into a circular structure of 12,780 bp (fig. 2; GenBank accession number MG757197) with an average coverage of ~452×. Coverage is rather even across the plastome with differences being related to GC content. Coverage of the mitochondrial loci is ca. 60×, thus it is not likely that potentially transferred plastome sequences have been confounding for assembly of the plastome.

Like the previously sequenced plastome of *S. densiflora* (Lam et al. 2015), the *S. thaidanica* plastome lacks one of the inverted repeat regions. Although higher coverage of two regions of the *S. densiflora* plastome suggested possible duplications, this is not the case for *S. thaidanica*. The plastome is largely colinear with the *Carludovica palmata* plastome except for one inversion spanning a region from *rps12* exon1 to *rps2*. Thus, colinearity is better conserved in *S. thaidanica* than in *S. densiflora*, which differed from *Carludovica* by two inversions (Lam et al. 2015). The inversion in *S. thaidanica* is not similar to any of those in *S. densiflora* indicating that they represent three separate structural rearrangement events. Although plastome structure and gene order is usually

Fig. 1.—The tiny, achlorophyllous *Sciaphila thaidanica* in its natural habitat. Photo: H. Æ. Pedersen.

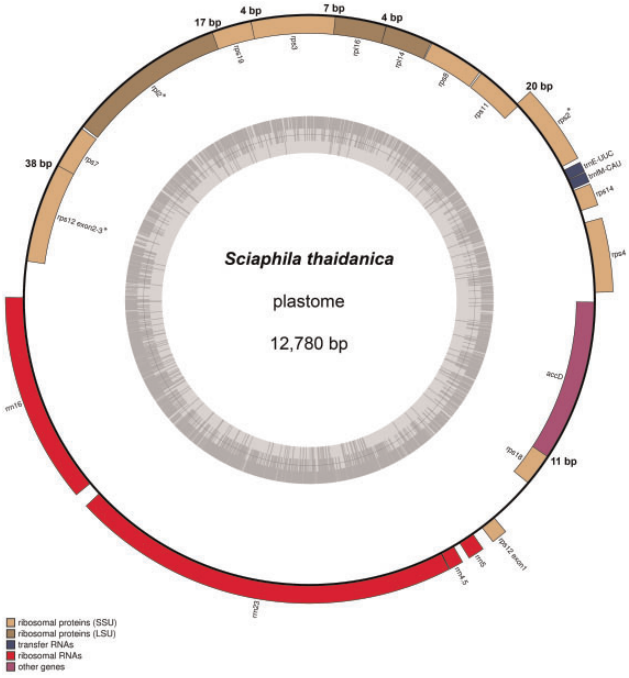

Fig. 2.—The plastome of *Sciaphila thaidanica*. Direction of transcription of genes is clockwise for those on the inside and counterclockwise for those on the outside. Asterisks (*) mark genes with introns. Overlapping reading frames are indicated with precise numbers of base pairs in the overlaps. Drawing made using OGDRAW v. 1.2 (Lohse et al. 2013).

highly conserved among angiosperms, achlorophyllous plants are particularly prone to structural changes (Wicke et al. 2016).

Although the plastome of *S. densiflora* is drastically reduced compared with *Carludovica palmata* (Lam et al. 2015), reduction has proceeded even further in *S. thaidanica* (table 1). The size has been reduced dramatically and the gene content is also substantially lower. All the 20 genes included in the *S. thaidanica* plastome are present in *S. densiflora*, but in addition, the latter includes putatively functional copies of *clpP*, *matK*, *rpl20*, *rpl36* and four more tRNA genes (*trnC*, *trnI*, *trnQ*, *trnW*) (Lam et al. 2015). The genes retained in *S. thaidanica* (*accD*, *rpl/s*, *rrn*, *trn*), which are all housekeeping genes, fit a model of gene loss from achlorophyllous plants perfectly (Barrett and Davis 2012; Barrett et al. 2014; Wicke et al. 2016), and the presence of *trnE* as one of only two tRNA genes supports the importance of this gene (Barbrook et al. 2006). The only nonribosomal protein coding gene, *accD*, is retained as in most other mycoheterotrophs yet sequenced (Lam et al. 2016; Graham et al. 2017). Only in some mycoheterotrophic Ericaceae has this gene possibly been functionally lost (Braukmann et al. 2017).

Two of the ribosomal protein genes (*rpl2*, *rps12*) retained in the *S. thaidanica* plastome contain a group IIA intron supposed to be spliced by the *matK* gene product (Zoschke et al. 2010). However, this gene has been lost from the plastome. Loss of *matK* coupled with retention of *rpl2* and *rps12* can be observed in a remarkable high number of parasitic and mycoheterotrophic plants suggesting that intron splicing may occur by alternative means (Graham et al. 2017). Alternatively, *matK* could be functionally transferred to the nuclear genome.

Reduction of the plastomes in *Sciaphila* is associated with compaction of the remaining genes. In *Carludovica palmata* genes occupy 70.2% of the plastome, in *Sciaphila densiflora*, the gene proportion is increased to 78.0% and in *S. thaidanica* to 91.9% leaving just 1.030 bp of intergenic sequence (table 1). This compaction of the *S. thaidanica* plastome has led to an unusual overlap of reading frames. Eleven genes have short overlapping reading frames in the 3′-end or 5′-end, and among them an array of five genes (*rpl14-rpl16-rps3-rps19-rpl2*) has no intergenic sequence at all (fig. 2). In addition, two ribosomal RNA genes (*rrn23*, *rrn4.5*) are just touching. Although reduction of the length of noncoding sequence, both through intron loss and deletions in intergenic regions, is well-known for plastomes from parasitic plants (Funk et al. 2007; McNeal et al. 2007; Petersen et al. 2015; Naumann et al. 2016), none of the previously described plastomes have attained a stage of gene compaction as observed here. To further understand how compaction has occurred a denser taxon sampling is crucial. Provided that the overlapping genes are functional (see below), they may be

Table 1
Comparison of Complete Plastomes of *Carludovica* and *Sciaphila*

	Carludovica palmata	*Sciaphila densiflora*	*Sciaphila thaidanica*
Complete plastome (bp)	158,545	21,485	12,780
Coding (% of total)	91,183 (57.5)	14,769 (68.7)	10,856 (84.9)
Introns (% of total)	20,174 (12.7)	1,998 (9.3)	894 (7.0)
Genes (% of total)	111,357 (70.2)	16,767 (78.0)	11,750 (91.9)
Intergenic (% of total)	47,189 (29.8)	4,718 (22.0)	1,030 (8.1)

Table 2
Comparison of Plastome Genes Shared by *Sciaphila* and *Carludovica*

	Carludovica	*Sciaphila densiflora*		*Sciaphila thaidanica*		
Locus	Length (bp)	Length (bp)	Similarity to *Carludovica* (%)	Length (bp)	Similarity to *Carludovica* (%)	Similarity to *S. densiflora* (%)
accD	1,476	1,401	67.2	1,179	52.4	53.5
rpl2	1,479	1,381		1,173		
rpl2 CDS	825	810	86.9	765	62.2	60.2
rpl14	369	369	89.2	345	64.8	65.0
rpl16	1,358	976		375		
rpl16 CDS	411	411	86.4	375	60.8	60.6
rps2	714	714	85.3	555	46.0	47.7
rps3	657	657	86.3	612	57.9	58.2
rps4	606	594	84.0	507	57.1	61.1
rps7	468	450	85.7	321	54.1	44.2
rps8	399	402	84.8	405	60.7	61.9
rps11	417	381	70.9	357	60.7	58.8
rps12	915	850		873		
rps12 CDS	369	369	86.2	387	71.8	68.3
rps14	303	306	87.9	156	37.3	40.6
rps18	306	252	66.3	240	66.3	58.6
rps19	279	279	87.1	291	58.4	57.8
rrn4.5	103	108	82.4	102	68.2	60.0
rrn5	121	121	89.3	116	55.4	51.5
rrn16	1,491	1,486	91.4	1,440	71.8	70.8
rrn23	2,811	2,816	90.5	2,656	70.4	69.6
trnE	73	78		75		
trnfM	74	74		73		
Sum coding[a]	12,272	12,078 (98.4%)		10,856[b] (88.5%)		
Sum introns[a]	2,147	1,617 (75.3%)		894 (41.6%)		
Sum genes[a]	14,419	13,695 (95.0%)		11,750 (81.5%)		

[a]For *Sciaphila*, the percentage relative to *Carludovica* is given in parentheses.
[b]The sum is 10,957, but due to overlap of several genes 101 bp are shared between two genes.

cotranscribed, but this possibility will need to be tested in live plant material.

The genes retained in the *Sciaphila* plastomes are also shorter in length; both with regard to coding and noncoding sequence. Intron coverage of the complete genomes drops from 12.7% in *Carludovica palmata* to 7.0% in *Sciaphila thaidanica* (table 1), but these differences are partly caused by loss of intron containing genes, thus a more precise picture is provided by the few remaining intron-containing genes in *Sciaphila* (table 2). Compared with *Carludovica palmata*, intron length of the three relevant genes (*rpl2, rpl16, rps12*) is reduced to 75.3% in *S. densiflora* and 41.6% in *S. thaidanica*. In the latter, the *rpl16* intron is lost completely. The length of the coding sequence of the genes present is reduced, too, although less drastically. The difference between *Carludovica palmata* and *S. densiflora* is hardly significant, but in *S. thaidanica*, the length of coding sequence has been reduced to 88.5% of the former, affecting most of the retained genes (table 2). Although intron loss is not uncommon, consistent reduction of coding sequence appears unusual. In *Thismia*

tentaculata K. Larsen & Aver. (Thismiaceae), having retained only 12 plastome genes, a 5% overall reduction in coding length can be observed, but the reduction applies to only five genes while another five have actually increased in length (Lim et al. 2016).

The reduction of coding sequence in the retained genes, coupled with extensive sequence divergence (see table 2), may indicate that some genes are no longer functional. However, if they are pseudogenized the reading frames would be expected to be even shorter. Most dubious in terms of functionality is *rps14*, which has been reduced to almost half the length compared with *Carludovica palmata*, and has a very low sequence similarity. With the exception of *rps18*, the genes of *Sciaphila thaidanica* are consistently more divergent than those of *S. densiflora* when compared with *Carludovica palmata* (table 2) indicating a further progressed state of degradation. It is, however, remarkable that the pairwise similarity between genes from the two species of *Sciaphila* is just as low as between genes from *S. thaidanica* and *Carludovica palmata* (table 2) suggesting a very high degree of randomness in the evolution of the retained genes. Sequence data from additional species of *Sciaphila* are strongly needed for understanding this extraordinary divergence in an evolutionary perspective.

In *S. densiflora* Lam et al. (2015) calculated that most genes were still under strong purifying selection. Only three genes were under significantly relaxed selection (*rps7*, *rpl14*, *clpP*), but of these only *clpP* has been deleted in *S. thaidanica* and none of the other genes missing from the *S. thaidanica* plastome (*matK*, *rpl20*, *rpl36*) showed any sign of relaxation in *S. densiflora*. This may indicate that gene loss follows different routes in different evolutionary lineages of *Sciaphila*. The pronounced sequence divergence of *S. thaidanica* is unfortunately a hindrance to meaningful analysis of selection pressure on individual genes, as such analyses are very sensitive to correct alignment. Neither is it possible to test transcription as we do not have access to live plants or suitably preserved tissue from *Sciaphila*.

With a maximum of 20 functional genes, the plastome of *S. thaidanica* is in the final stage of degradation. Only in the mycoheterotroph *Thismia tentaculata* and the holoparasitic genus *Pilostyles* Gull. (Apodanthaceae) have plastomes with even fewer genes been described (Bellot and Renner 2015; Lim et al. 2016). The plastome of *Sciaphila thaidanica* is also the second smallest yet described; only the 11,348-kb plastome of *Pilostyles aethiopica* Welw. is smaller. With more than 50 fully mycoheterotrophic species of Triuridaceae and a crown age of at least 50 Ma (Mennes et al. 2013), this clade represents one of the oldest examples of photosynthesis loss. Hence, the Triuridaceae offers an outstanding framework for evolutionary studies of the final stages of plastome degradation.

Acknowledgments

We thank Charlotte Hansen and the staff at the National High Throughput DNA Sequencing Centre, University of Copenhagen for skilful laboratory assistance, and we are indebted to Somran Suddee, herbarium BKF, for making the DNA sample available. We thank three anonymous reviewers for helpful comments. This work was supported by the Danish Council for Independent Research | Natural Sciences (grant number DFF-4002-00505).

Literature Cited

Barbrook AC, Howe CJ, Purton S. 2006. Why are plastid genomes retained in non-photosynthetic organisms? Trends Plant Sci. 11(2):101–108.

Barrett CF, Davis JI. 2012. The plastid genome of the mycoheterotrophic *Corallorhiza striata* (Orchidaceae) is in the relatively early stages of degradation. Am J Bot. 99(9):1513–1523.

Barrett CF, et al. 2014. Investigating the path of plastid genome degradation in an early transition clade of heterotrophic orchids, and implications for heterotrophic angiosperms. Mol Biol Evol. 31:3095–3112.

Bellot S, Renner SS. 2015. The plastomes of two species in the endoparasite genus *Pilostyles* (Apodanthaceae) each retain just five or six possibly functional genes. Genome Biol Evol. 8(1):189–201.

Braukmann TWA, Broe MB, Stefanović S, Freudenstein JV. 2017. On the brink: the highly reduced plastomes of nonphotosynthetic Ericaceae. New Phytol. 216(1):254–266.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32(5):1792–1797.

Funk HT, Berg S, Krupinska K, Maier UG, Krause K. 2007. Complete DNA sequences of the plastid genomes of two parasitic flowering plant species, *Cuscuta reflexa* and *Cuscuta gronovii*. BMC Plant Biol. 7:45.

Graham SW, Lam VKY, Merckx VSFT. 2017. Plastomes on the edge: the evolutionary breakdown of mycoheterotrophic plastid genomes. New Phytol. 214(1):48–55.

Krause K. 2011. Piecing together the puzzle of parasitic plant plastome evolution. Planta 234(4):647–656.

Lam VKY, Gomes MS, Graham SW. 2015. The highly reduced plastome of mycoheterotrophic *Sciaphila* (Triuridaceae) is colinear with its green relatives and is under strong purifying selection. Genome Biol Evol. 7(8):2220–2236.

Lam VKY, Merckx VSFT, Graham SW. 2016. A few-gene plastid phylogenetic framework for mycoheterotrophic monocots. Am J Bot. 103(4):692–708.

Lim GS, Barrett CF, Pang C-C, Davis JI. 2016. Drastic reduction of plastome size in the mycoheterotrophic *Thismia tentaculata* relative to that of its autotrophic relative *Tacca chantrieri*. Am J Bot. 103(6):1129–1137.

Lindgreen S. 2012. AdapterRemoval: easy cleaning of next-generation sequencing reads. BMC Res Notes 5:337.

Lohse M, Drechsel O, Kahlau S, Bock R. 2013. OrganellarGenomeDraw – a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucleic Acids Res. 41(W1):W575–W581.

McNeal JR, Kuehl JV, Boore JL, de Pamphilis CW. 2007. Complete plastid genome sequences suggest strong selection for retention of photosynthetic genes in the parasitic plant genus *Cuscuta*. BMC Plant Biol. 7:57.

Mennes CB, Smets EF, Moses SN, Merckx VSFT. 2013. New insights in the long-debated evolutionary history of Triuridaceae (Pandanales). Mol Phylogenet Evol. 69(3):994–1004.

Meyer M, Kircher M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb Protoc. 2010(6):pdb.prot5448.

Molina J, et al. 2014. Possible loss of the chloroplast genome in the parasitic flowering plant *Rafflesia lagascae* (Rafflesiaceae). Mol Biol Evol. 31(4):793–803.

Naumann J, et al. 2016. Detecting and characterizing the highly divergent plastid genome of the nonphotosynthetic parasitic plant *Hydnora visseri* (Hydnoraceae). Genome Biol Evol. 8(2):345–363.

Petersen G, Cuenca A, Seberg O. 2015. Plastome evolution in hemiparasitic mistletoes. Genome Biol Evol. 7(9):2520–2532.

Wicke S, et al. 2016. Mechanistic model of evolutionary rate variation en route to a nonphotosynthetic lifestyle in plants. Proc Natl Acad Sci U S A. 113:9045–9050.

Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D. 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol Biol. 76(3–5): 273–297.

Zoschke R, et al. 2010. An organellar maturase associates with multiple group II introns. Proc Natl Acad Sci U S A. 107:3245–3250.

Associate editor: John Archibald