

**Provided for non-commercial research and educational use only.  
Not for reproduction, distribution or commercial use.**

This chapter was originally published in the book *Advances in Botanical Research, Vol. 84* published by Elsevier, and the attached copy is provided by Elsevier for the author's benefit and for the benefit of the author's institution, for non-commercial research and educational use including without limitation use in instruction at your institution, sending it to specific colleagues who know you, and providing a copy to your institution's administrator.



All other uses, reproduction and distribution, including without limitation commercial reprints, selling or licensing copies or access, or posting on open internet sites, your personal or institution's website or repository, are prohibited. For exceptions, permission may be sought for such use through Elsevier's permissions site at:

<http://www.elsevier.com/locate/permissionusematerial>

From David R. Smith, *Let There Be Light: A Contemporary Primer on Primary Plastid Endosymbiosis*. In: Yoshihisa Hirakawa, editor, *Advances in Botanical Research, Vol. 84*, Oxford: Academic Press, 2017, pp. 31-56.

ISBN: 978-0-12-802651-9  
© Copyright 2017 Elsevier Ltd  
Academic Press



# Let There Be Light: A Contemporary Primer on Primary Plastid Endosymbiosis

David R. Smith<sup>1</sup>

University of Western Ontario, London, ON, Canada

<sup>1</sup>Corresponding author: e-mail address: dsmit242@uwo.ca

## Contents

1. Introduction	32
2. The ABC's of Primary Endosymbiosis	33
3. The Polychromatic Puzzle Pieces of Plastid Primary Endosymbiosis	36
4. <i>Paulinella chromatophora</i> : A Primary Plastid Revival Tour	39
5. Lost in the Light: Unexpected Insights From Nonphotosynthetic Primary Plastids	42
6. Three Decades of Primary Plastid Genomics	45
7. Concluding Thoughts: Too Much Sequencing, Not Enough Experimenting?	49
Acknowledgements	51
References	51

## Abstract

Endosymbiosis, more than any other process, perhaps, is the leading narrative upon which the history of eukaryotic evolution has been written. Primary endosymbiosis, which is the uptake of a prokaryote by another living cell, has arguably been *the* driving force for the origins and diversification of complex life on Earth. The genetic integration of, first, a nonphotosynthetic alphaproteobacterium and, later, a photosynthetic cyanobacterium into a eukaryotic cellular framework have shaped and altered the planet's biodiversity and biogeochemistry in countless ways, from the land, to the water, to the atmosphere. If you are alive today and reading these words, it is in no small part because of endosymbiosis. Like all eukaryotes, we are the product of an ancient endosymbiotic love affair, and for plants and algae the endosymbiotic romance was a complicated triangle. Here, I recount my own passions for the topic of endosymbiosis, highlighting past and present breakthroughs as well as some of the controversies and unanswered questions that have plagued the field. I focus on the evolution of primary plastids, their genomes, and the supergroup to which they are found (the Archaeplastida), including members that have lost photosynthetic capabilities but still retain a colourless plastid.



## 1. INTRODUCTION

*When you think about the complexity of our natural world—plants using quantum mechanics for photosynthesis, for example—a smartphone begins to look like a pretty dumb object.*

**Jeff VanderMeer**

Stroll through a flowered park, dip into a cool summer lake, or hike in a dense forest and you will find yourself in good company. I am talking about plants and algae, of course, but more specifically about chloroplasts—the factories of photosynthesis and hubs of countless crucial biochemical reactions. Yes, chloroplasts (and plastids as a whole) are one awesome eukaryotic organelle. Their pigments alone provide the world with much of its beautiful and mellifluous colours, not to mention their clever and life-sustaining ability to convert carbon dioxide and water into oxygen and sugar. Plastids are also the engines primarily responsible for generating the fruits, vegetables, and grains we eat, and they form the basis for the wine and beer many of us imbibe, some of us too often.

For me, the most intriguing thing about plastids is how they came to be. Their journey from a free-living, bacterial existence to integral and inalienable components of plants and algae involved countless winding, diverging, and colliding roads and is replete with whimsy, mystery, death, and disease. The story of plastids has many plots, many characters, and is still ongoing, as emphasized throughout the various chapters in this collection. And it is what first made me fall in love with biology.

Here, I recount some of what is currently known about the origins and early evolution of plastids, focusing first on the process by which they arose: primary endosymbiosis, which is when a prokaryote is taken up by another living cell. I discuss the immediate and downstream consequences of this momentous endosymbiotic event, and how it ultimately gave rise to the eukaryotic supergroup Archaeplastida, which is made up of red algae, glaucophyte algae, green algae, and land plants (Adl et al., 2012). I describe aspects of plastid genetics, genomics, and gene expression, topics which at first glance can appear straightforward and elementary as compared to nuclear systems but which have proven to be far from ordinary (Smith & Keeling, 2015). Although plastids are undoubtedly highly specialized organelles, in certain lineages they have drifted down peculiar evolutionary roads, including ones that have led to the restructuring of the photosynthetic apparatus (Morgan-Kiss, Prisco, Pocock, Gudynaite-Savitch, & Hüner, 2006)

or the outright loss of photosynthesis (Figuroa-Martinez, Nedelcu, Smith, & Reyes-Prieto, 2015). In other instances, they have adopted costly and bureaucracy-ridden infrastructures for gene editing (Knoop, 2011), acquired massive genomes made up almost entirely of noncoding DNA (Smith & Lee, 2010), or even forfeited their DNA and gene expression systems altogether (Molina et al., 2014; Smith & Asmail, 2014).

Despite the near-universal acceptance of the endosymbiotic theory, certain questions about plastids and plastid genome evolution remain unanswered and are mired in debate, confusion, and controversy. This is especially true for the steps leading up to and the nuances involved in the genetic merger between a cyanobacterium and a nonphotosynthetic eukaryote. I have witnessed many heated conversations and the occasional all-out shouting match surrounding the topic of whether, for example, the host was “primed” before the primary endosymbiotic event that produced plastids and the role that pathogenic bacteria might have played in this process. But I am getting ahead of myself. Before describing the subtleties of plastid evolution and primary endosymbiosis, one must first cover the basics.



## 2. THE ABC'S OF PRIMARY ENDOSYMBIOSIS

*You fit into me  
like a hook into an eye  
a fish hook  
an open eye*

**Margaret Atwood**

When I was an undergraduate student, I detested plant biology classes and anything to do with algae. I still cringe at having to memorize the life cycle of a fern, the encyclopaedic vascular system of a tree, and the multifarious flagellar apparatus of *Chlamydomonas*—and don't get me started on the electron transport chain of photosynthesis. But my interest in plants and algae grew immensely when I took advanced genetics and learned in detail about the endosymbiotic theory for the origins of mitochondria and plastids, which was made popular by Lynn Margulis in the late 1960s (Sagan, 1967), but articulated in various forms much earlier (Wallin, 1927), including by Russian biologist Mereschkowsky (1910). The endosymbiotic theory (or symbiogenesis) holds that the mitochondrion and the plastid were once free-living bacteria taken up via primary endosymbiosis by a host cell (Archibald, 2014). These two endosymbiotic events appear to have

occurred about 1.8 and 1.5 billion years ago, respectively (Gray, 2012; Yoon, Hackett, Ciniglia, Pinto, & Bhattacharya, 2004). And as every freshman biology student can tell you, mitochondria are closely related to present-day alphaproteobacteria (Gray, 2012), whereas chloroplasts show strong phylogenetic affinity to contemporary nitrogen-fixing filamentous cyanobacteria (Deusch et al., 2008)—but see de Vries and Archibald (2017).

I can still picture my advanced genetics professor writing on the blackboard the various steps involved in a primary endosymbiosis: feeding, entrapment, sharing, endosymbiotic gene transfer and gene loss, relinquishing control, and eventually complete genetic integration. “Most of you are fixated on procreation”, she said, staring down the boys in the class. “You are obsessed with the act of (and the act of avoiding) two becoming two plus one. But look closely, with symbiogenesis we’re talking about the opposite: two distinct lifeforms becoming one, and that is why it is so extraordinary”. Indeed, in a single 50-min lecture, this talented teacher ignited in me a deep passion for endosymbiosis, eukaryogenesis, and organellogenesis that persists to this day.

Up to that point in my undergraduate education, endosymbiotic theory had been an afterthought in my various biology courses—a subtext to an image of a eukaryotic cell, a single slide of a Pac-Man-esque organism eating a green dot, two simple circles labelled mitochondrial and plastid DNA (mtDNA and ptDNA), end of story. But in advanced genetics I learnt that primary endosymbiosis is not an afterthought or footnote to eukaryotic evolution, rather, it is a fundamental topic in biology, a driving force for the origins and diversification of complex life on Earth, and a leading narrative upon which the history of eukaryotes has been written (Archibald, 2014).

The endosymbiotic theory has been discussed, debated, and lauded in hundreds of research articles, review papers, textbooks, and commentaries over the past half century. There are now so much data in support of the theory that it has arguably completed the journey from hypothesis to scientific fact. Today’s scientists, with few exceptions (Kurland, Collins, & Penny, 2006), whole-heartedly accept as truth the endosymbiotic origins of mitochondria and plastids (Archibald, 2015). Consequently, contemporary evolutionary biologists are no longer concerned with proving or disproving symbiogenesis, but are instead focused on understanding the key events leading up to and following the primary endosymbioses that generated mitochondria and plastids, as well as the various eukaryote–eukaryote endosymbioses responsible for the horizontal spread of plastids across the eukaryotic domain (Zimorski, Ku, Martin, & Gould, 2014).

What was the nature of the host cells that acquired the mitochondrial and plastid endosymbionts? What precise alphaproteobacterial and cyanobacterial lineages were the progenitors of the mitochondrion and the plastid? What were the driving evolutionary forces that led to the success of these two ancient endosymbioses? Were they adaptive or nonadaptive? Driven by sharing or, more provocatively, parasitism? Did other earlier endosymbiotic events and/or waves of horizontal gene transfer play a role in the successful integration of mitochondria and plastids? Breakthroughs in molecular sequencing technologies and bioinformatics (Metzker, 2010), including single-cell sequencing, metagenomics, and phylogenomics (Kolisko, Boscaro, Burki, Lynn, & Keeling, 2014; Moran, 2009), are helping scientists address these and other important questions, are providing unexpected and occasionally paradigm-shifting answers, and are ultimately leading to a profound understanding of endosymbiosis at its deepest levels.

Take, for instance, the mitochondrial endosymbiotic event. In just the past 3 years our perception of the host cell and host lineage has improved considerably, and many of these advancements can be summarized in a single word: Loki. The Lokiarchaeota (Loki for short) is a newly discovered archaeal phylum that is more closely related to eukaryotes than any other prokaryotic lineage sampled to date (Spang et al., 2015). Metagenomic analyses of Loki uncovered genes that were previously thought to be unique to eukaryotes, including genes for endocytosis and/or phagocytosis (Embley & Williams, 2015). In other words, Loki likely resembles and descends directly from the archaeal ancestor of all eukaryotes, an ancestor that appears to have harboured the machinery needed to carry out primary endosymbiosis. The discovery of Loki further supports to the notion that the evolutionary transition from prokaryotic to eukaryotic life was predominantly driven by the primary engulfment of an alphaproteobacterium by an archaeal host possessing some eukaryotic-like features. But this does not tell the whole story.

Shortly after the discovery of Loki, Pittis and Gabaldón (2016) using phylogenomics traced the evolutionary histories of proteins believed to be present in the last eukaryotic common ancestor (LECA). In addition to identifying archaeal- and alphaproteobacterial-related proteins, the authors uncovered a third class of bacterial LECA proteins, predating those acquired from the mitochondrial endosymbiont, and some which function in the endoplasmic reticulum and Golgi apparatus. Pittis and Gabaldón (2016) believed that the genes for this third class of proteins came from an earlier bacterial endosymbiont, one that existed before that which gave

rise to the mitochondrion, or from waves of horizontal gene transfer. Whatever their origin, these genes might have endowed the host lineage with the complexity needed for the eventual genetic merger with an alphaproteobacterium.

The notion that genes obtained via endosymbiotic and/or horizontal gene transfer can provide the foundation for future endosymbiotic events is a reoccurring theme in evolutionary biology, including plastid evolution. But it is also a controversial and hotly debated idea. The suggestion that the mitochondrial endosymbiotic event occurred later in the evolution of complex cells than previously thought and preceded an earlier endosymbiosis was embraced by many in the scientific community but vehemently rejected by others. One research team went so far as to say that the paper of [Pittis and Gabaldón \(2016\)](#) had “multiple fatal flaws founded in inappropriate statistical methods and analysis, in addition to erroneous interpretations” ([Martin et al., 2016](#)). This debate echoes a continuous thread throughout the study of evolutionary endosymbioses: researchers can be heavily invested in specific views and hypotheses and not always accepting of alternative perspectives, often to the detriment of science. The legend goes that Lynn Margulis’ seminal paper on endosymbiotic theory ([Sagan, 1967](#)) was not only highly criticized but also rejected 15 times before finally being accepted by the *Journal of Theoretical Biology*. As discussed later, the soap opera surrounding the primary acquisition of plastids within the eukaryotic domain is no less controversial, multifaceted, and debated as its mitochondrial counterpart.



### 3. THE POLYCHROMATIC PUZZLE PIECES OF PLASTID PRIMARY ENDSYMBIOSIS

*Watermelons:  
Green Buddhas  
On the fruit stand.  
We eat the smile  
And spit out the teeth.*

*Charles Simic*

It makes intuitive sense that a heterotrophic organism would want to enslave a bacterium capable of converting sunlight into usable energy. Thus, it is not surprising that around 1.5 billion years ago a unicellular, nonphotosynthetic protist did just that, starting down the road towards eukaryotic phototrophy. The onramp to this road could have been quite simple: for example, a phagotrophic predatory lifestyle involving cyanobacteria as the engulfed

prey. However, the evolutionary leap from a predator–prey relationship to a symbiotic photosynthetic one is not so clear. Textbooks like to depict this transition as a single step—a jump from a free-living cyanobacterium to an entrapped and integrated endosymbiont. But, in fact, the segue to a full-fledged host–endosymbiont partnership was unquestionably more drawn out than most textbooks would have us believe, occurring at a population level and on an evolutionary timescale, and likely involving multiple contributing partners rather than being a singular event with only two players.

The early stages of this long-winded metamorphosis from food to endosymbiont to photosynthetic organelle are best depicted by the “shopping bag model” (Larkum, Lockhart, & Howe, 2007) of primary plastid evolution:

*It seems unlikely that the stable [cyanobacterial] symbiont ultimately acquired by the host cell would be the first one it had ever acquired. The acquisition would almost certainly have been preceded by the uptake of other photosynthetic organisms. ... early rounds of failed endosymbiosis, with some would-be endosymbionts eventually lysing and liberating DNA into the cytosol, would result in integration of endosymbiont DNA into the nuclear genome. This DNA would have persisted in the nucleus for a period of time, even if there were no longer functional symbionts in the host cytoplasm. If, finally, a symbiont [was] able to establish a balanced relationship with the host, the reservoir of sequences in the host nucleus that were derived from previous photosynthetic organisms would have provided a pool of sequences to encode proteins to be imported into the newly established plastid.*

*Howe, Barbrook, Nisbet, Lockhart, and Larkum (2008)*

In short, the shopping bag scenario implies that the protein machinery of plastids has a mixed origin with most proteins originating from the successful cyanobacterial endosymbiont but at least some coming from earlier unsuccessful endosymbionts.

It was long assumed that the preliminary events of primary plastid endosymbiosis, including those described in the shopping bag model, occurred in salt water—somewhere in the saline oceans of the mid-Proterozoic era (Deusch et al., 2008). Just recently, however, this assumption has been laid to rest, and it is now thought that eukaryotic phototrophy first evolved in freshwater and is directly linked to a newly uncovered clade of cyanobacteria called *Gloeomargarita* (Ponce-Toledo et al., 2017).

Like with Loki, it is thanks to the explorations of undersampled environments and the characterization of new organisms that we now know about *Gloeomargarita lithophora*. This recently discovered, deep-branching, and biofilm-forming cyanobacterium might hold countless untold secrets to the beginnings of primary plastids. *G. lithophora*, which is the only member



of the *Gloeomargarita* clade currently in culture, was first isolated in 2012 from a freshwater alkaline lake in the Oriental Basin of Mexico (Couradeau et al., 2012). But its claim to fame is its affiliation to plastids. Using a comprehensive phylogenomic data set, including plastid-encoded proteins, nucleus-encoded proteins of plastid origin, and extensive genomic data from cyanobacteria, Ponce-Toledo et al. (2017) showed that *G. lithophora* is the closest prokaryotic relative of primary plastids yet found. This implies that plastids evolved from an ancestor of the *Gloeomargarita*. What's more, the entire *Gloeomargarita* group appears to be restricted to freshwater environments, suggesting that eukaryotic photosynthesis first emerged in a terrestrial freshwater setting. This hypothesis is further supported by the fact that the earliest diverging archaeplastid lineage—the Glaucophyta—is exclusive to freshwater ecosystems (Delwiche & Cooper, 2015).

The notion that primary plastid endosymbiosis began as an assorted shopping bag with the central participant being an ancestral member of the *Gloeomargarita* is compelling. But a more provocative suggestion is that a *Chlamydia* pathogen had a helping hand in primary plastid evolution. A major hurdle en route to a successful endosymbiosis is the ability to overcome host defences. This presents a slight problem when pondering the genesis of primary plastids: present-day cyanobacteria, including *G. lithophora*, do not have the genetic capacity to evade host defences or interact with the host cellular machinery (Ball, Bhattacharya, & Weber, 2016a). So then how did the cyanobacterial progenitor of the plastid endure the initial stages of endosymbiosis? Well, as the shopping bag model posits, many early would-be cyanobacterial endosymbionts probably did not survive but over time gained ground with each subsequent failed attempt (Larkum, Lockhart, & Howe, 2007). An alternative view comes from the observation that the genomes of various archaeplastid plants and algae harbour several dozen genes of seemingly chlamydial origin (Huang, & Gogarten, 2007). These findings spurred the hypothesis “that a chlamydial bacterium entered the host cell together with a cyanobacterium ... [allowing] the cyanobacterium to escape host defenses and establish a tripartite symbiosis through the help of chlamydial-encoded effector proteins and transporters” (Ball, Bhattacharya, & Weber, 2016a). Simply put: eukaryotic photosynthesis might have resulted from a *ménage à trois* between a heterotrophic protist, a cyanobacterium, and a chlamydial bacterium, whereby key metabolic genes from the latter enabled the symbiotic capture of the cyanobacterium. It is argued that these key genes were ultimately horizontally transferred to

the host, thus permitting the eventual loss of the chlamydial partner from the relationship (Ball et al., 2013).

Unsurprisingly, not everyone supports such a salacious view of plastid evolution (Gould, 2016). A subsequent phylogenetic study found a mosaic origin for the putative chlamydial-derived enzymes and no strong evidence for *Chlamydia*-to-host gene transfer (Domman, Horn, Embley, & Williams, 2015). The authors of this study concluded that *Chlamydia* did not have any role in establishing the primary plastid endosymbiosis and noted the following:

*There is a deeper problem with inferring a special explanation for the presence of putative chlamydial genes on plant genomes ... [Recent] studies have demonstrated that in addition to organellar genes shared with Cyanobacteria and Alphaproteobacteria, the Archaeplastida share more genes with Gammaproteobacteria, Actinobacteria, Deltaproteobacteria, Bacilli, Bacteroidetes and Betaproteobacteria than with Chlamydiae. Given the extent of HGT, particularly of metabolic genes, among major cellular groups ... these patterns of gene sharing—including those involving Chlamydiae—are most simply explained as a mixture of genuine [horizontal gene transfer] and tree reconstruction artefacts.*

*Domman, Horn, Embley, and Williams (2015)*

Despite these data and that there are no known cases of *Chlamydia* infecting a species from the Archaeplastida, the authors of the “ménage à trois” (MAT) hypothesis—as it has become known—are, for now, sticking with their story: “We disagree that the phylogenetic analyses of Domman et al. (2015) reject our hypothesis ... [their trees] are very similar to ours and emphasize uncertainties that we have accounted for through straightforward interpretations of carbohydrate biochemistry” (Ball, Bhattacharya, & Weber, 2016b). The major challenge facing these hypotheses—and any hypothesis—about primary plastid endosymbiosis is that the events in question happened so long ago. If only there was a more recent primary endosymbiosis between a heterotrophic protist and a cyanobacterium that could be scrutinized. As anyone who works in this field will know, there is, and it is found within the rhizarian genus *Paulinella*.



#### **4. PAULINELLA CHROMATOPHORA: A PRIMARY PLASTID REVIVAL TOUR**

*Outside around the side  
Form a circle forward I'm an  
Outsider on the side*

Formerly a farm boy  
Inside I'm on the side  
I'm divided undecided  
Back then around again  
Second time's a charm, boy

**Enter the Haggis—Canadian Celtic rock band**

On the first day of my 4-year undergraduate genome evolution and endosymbiosis course, I always ask the class the same question: “Raise your hand if you’ve ever heard the word *Paulinella*?” Not once in my years of teaching this course have I seen a single hand go up. “Come on”, I say, “*Paulinella chromatophora*, anybody? I’ll even make it easy for you: it is unicellular, likes sunlight, lives in freshwater, belongs to the supergroup Rhizaria, the phylum Cercozoa ...” Still no hands. “It’s an alga, for Pete’s sake!” Even more blank faces.

It never ceases to amaze me how little undergraduate biology students are taught about eukaryotic microbes, and how excited they get when given the opportunity to learn about this broad assemblage of life called protists. When I finally tell the class that the primary endosymbiotic event that spawned the astonishing diversity of algae and land plants we see today is being played out again in *P. chromatophora*, their eyes go big and half of them are hooked. The other half are hooked when I go to the whiteboard and draw a long horizontal arrow from the words “distinct lifeform” to “organelle” and ask them to tell me at what precise point does the former become the latter. “Wherever that point falls on this arrow”, I exclaim, “it is where the photosynthetic endosymbiont of *P. chromatophora* is found”.

Yes, *P. chromatophora* deserves to be at the forefront of any course or review on endosymbiosis. Between 60 and 200 million years ago, the ancestor of this little-known amoeboid alga transitioned from a heterotrophic bacterivorous existence, sustained in part by feeding on cyanobacteria, to a phototrophic one, dependent on a cyanobacterial endosymbiont called a chromatophore (Nowack, 2014). *P. chromatophora* is truly one of a kind, being the only known example of primary acquisition of a photosynthetic organelle outside of that which generated the Archaeplastida, and about a billion years more recent, allowing researchers to study photosynthetic organellogenesis in real time. Unlike with primary plastid endosymbiosis, the shift to a phototrophic existence in *Paulinella* likely occurred in a marine environment through the uptake of an  $\alpha$ -cyanobacterium related to *Synechococcus* and *Cyanobium* species (Nowack, 2014) (note: *Gloeomargarita* species belong to the  $\beta$ -cyanobacterial clade). Two additional photosynthetic *Paulinella* algae

have been identified, including the marine *P. longichromatophora* (Kim & Park, 2016), but there is little doubt that the chromatophores of this genus have a single common origin (Yoon et al., 2009). Presently, there is no evidence of a chlamydial pathogen having any part in the chromatophore endosymbiotic event, but, as described later, endosymbiont gene loss might be driving the fixation of horizontally acquired bacterial genes in the host nuclear genome (Nowack et al., 2016).

What's been learned from examining the burgeoning *Paulinella* primary endosymbiosis? First, many of the concepts and conclusions formed from studying the procurement of primary plastids are echoed in the data from *P. chromatophora*, including, for example, the vital roles of intracellular gene loss, gene transfer, and protein targeting in endosymbiosis and organellogenesis (Gagat, Bodył, & Mackiewicz, 2016). Measuring around one million base pairs, the chromatophore genome is two-thirds smaller than that of its cyanobacterial ancestor (Reyes-Prieto et al., 2010). This severe genomic reduction is largely due to endosymbiotic gene loss, including the disposal of entire biosynthetic pathways for various amino acids and cofactors (Nowack, 2014), thus permanently securing the chromatophore to its host. There are also at least 30 chromatophore genes that have successfully relocated to the host nuclear genome, some of which were shown to be expressed on cytoplasmic ribosomes and posttranslationally targeted to the chromatophore where they assembled with endosymbiont-encoded proteins (Nowack & Grossman, 2012). The trafficking and import of host-encoded proteins into an endosymbiont are considered an early but obligatory step in organellogenesis, leading some to suggest that the chromatophore has officially matured into a bona fide photosynthetic organelle (Nakayama & Archibald, 2012). But new data have shown that the host–chromatophore integration is even further along than previously thought.

For *P. chromatophora*, endosymbiotic gene transfer alone cannot account for the missing genes in the chromatophore genome—or, more importantly, the resulting gaps in the corresponding biosynthetic pathways. Nowack et al. (2016) recently showed that these metabolic cracks have been filled in through horizontal gene transfer from various bacteria to the host nuclear genome, providing a new take on the rules governing primary endosymbiosis:

*[Chromatophore] genome reduction seems to drive the fixation of horizontally acquired 'compensatory' bacterial genes in the host genome. Thus, similar to endosymbiotic gene transfer, HGT-derived genes may facilitate integration of the*

*endosymbiont by providing the host with transcriptional/translational control over chromatophore metabolic functions, metabolite fluxes between the cytoplasm and chromatophore, and the processing of genetic information. Therefore, like endosymbiotic gene transfer, HGT establishes key connections that enable the host to coordinate host–chromatophore metabolism, growth, and proliferation.*

*Nowack et al. (2016)*

The authors of the study hypothesized that these horizontally acquired genes are the outcome of a phagotrophic lifestyle, which was maintained alongside phototrophy in the initial stages of chromatophore integration and eventually lost once the host–chromatophore relationship was firmly established. More broadly, these data underline the fundamental importance of horizontal gene transfer in the establishment of primary endosymbioses.

*Paulinella* is an excellent case study for the transition from heterotrophy to phototrophy. But how about the reverse direction—the switch from a phototrophic to a heterotrophic existence? As touched upon in the preceding section, there are many primary plastid-containing lineages that have forsaken photosynthesis, and some have even jettisoned their plastid genomes.



## 5. LOST IN THE LIGHT: UNEXPECTED INSIGHTS FROM NONPHOTOSYNTHETIC PRIMARY PLASTIDS

*When the dark comes down, oh, the wind is on the sea  
With lispng laugh and whimper to the red reef's threnody,  
The boats are sailing homeward now across the harbor bar  
With many a jest and many a shout from fishing grounds afar.  
So furl your sails and take your rest, ye fisher folk so brown,  
For task and quest are ended when the dark comes down.*

*Lucy Maud Montgomery*

When asked to present my research field to a general audience, I love to talk about the eclectic and unusual suspects that have ditched photosynthesis but still retain a colourless plastid (Figueroa-Martinez, Nedelcu, Smith, & Reyes-Prieto, 2015). “Let’s get high on some photosynthetic burnouts”, I say to the audience, showing them high-definition photos of chlorophyll-lacking flowers and algae, such as the ghostly white corpse plant *Monotropa uniflora* and the nonphotosynthetic green alga *Polytomella magna*, the  $\beta$ -carotene gleaming from its eroding eyespot (Asmail & Smith, 2016). These images usually illicit a few oohs and aahs from the audience, but by far the greatest reaction I get is when I show pictures of disease-causing non-photosynthetic algae: a close-up of a hand with lesions caused by the parasitic

trebouxiophyte *Prototheca wickerhamii*, the worm-like gut contents of an insect infected with the pathogenic green alga *Helicosporidium* sp. ... a red blood cell bursting with the malaria parasite *Plasmodium falciparum*, which has a secondary, red algal-derived plastid (Keeling, 2013). After such gruesome shots, I normally see surprised, inquisitive faces looking back at me. “What does a parasite do with a chloroplast anyway?” shouts someone from the back row. “Why hold on to a broken lightbulb?” mumbles another. “Seems like a stupid strategy to me”, says a third, referring to an obligate heterotrophic existence.

Indeed, after such a long, hard-fought battle to acquire a plastid, how could any sound-minded species give up photosynthesis? The short answer is that evolution is not always straightforward or adaptive. For instance, a nonlethal mutation that knocks out photosynthetic capabilities could be fixed in a population of algae or plants through random genetic drift, especially if the effective size of the population is small. Why would such a mutation not be lethal? Because many photosynthetic eukaryotes are mixotrophic, meaning that they can survive using inorganic (phototrophy) and organic (chemoheterotrophy) carbon sources—the latter involves prey consumption via phagocytosis or endocytosis, or the intake of small organic compounds through osmosis (Tittel et al., 2003).

It is no great surprise, therefore, that mixotrophs can sometimes find themselves going down the generally one-way road towards a mandatory heterotrophic existence. In spite of the obvious advantages of mixotrophy, it is metabolically expensive to preserve the molecular machineries needed for both trophic strategies. Mixotrophic algae spend about five times more energy on maintaining the photosynthetic apparatus than on heterotrophic abilities (Raven, 1997). Thus, under some conditions, like when the cost of maintaining photosynthesis outweighs its benefits, losing phototrophic capabilities should be advantageous (de Castro, Gaedke, & Boenigk, 2009). In fact, the large number of distinct nonphotosynthetic plant and algal lineages that exist today underscores just how dispensable photosynthesis can be (Stoecker, 1998).

Although it is possible to do away with photosynthesis, it is not so easy to dump a plastid—all known nonphotosynthetic members of the Archaeplastida retain one (Keeling, 2013). This is because as the primary plastid endosymbiosis took hold, the host became dependent upon its cyanobacterial partner for much more than photosynthesis. In plants and algae, many vital biochemical pathways unrelated to photosynthesis are outsourced entirely or partly to the plastid, such as the biosynthesis of aromatic

and hydrophobic side-chain amino acids, tetrapyrroles, and terpenoids (Gould, Waller, & McFadden, 2008). Although nearly all the enzymes involved in these pathways are nuclear encoded, most nonphotosynthetic plastids still retain a genome, albeit one that is highly reduced with a much smaller gene content than that in photosynthetic taxa (Figueroa-Martinez, Nedelcu, Smith, & Reyes-Prieto, 2015). The genes that remain in non-photosynthetic plastids are almost entirely dedicated to plastid gene expression—a complicated process involving both plastid- and nuclear-encoded machinery (Gould, Waller, & McFadden, 2008). In the absence of photosynthesis, this complex infrastructure, spanning two genetic compartments, exists just so that a few key metabolic genes from the ptDNA can be expressed. Consequently, it was long believed that nonphotosynthetic plastids were irreversibly tied to their genomes (Barbrook et al., 2006; Nair & Striepen, 2011). But in 2014 it was shown that genome-less plastids do exist. One of these examples came from my and my colleague's work on *Polytomella*—a genus of free-living, unicellular nonphotosynthetic green algae closely related to *Chlamydomonas reinhardtii* (Smith & Lee, 2014).

After a decade of working on the organelle genetics of *Polytomella*, we were finally able to provide sufficient data to support outright plastid genome loss in this colourless genus. On the day that the paper was accepted, we had champagne on ice ready to celebrate the first example of a primary plastid-bearing lineage with no ptDNA only to discover that another team had beaten us to the summit by only a few weeks. Like *Polytomella*, the non-photosynthetic and parasitic angiosperm *Rafflesia lagascae* (known for having the largest single flower of any plant) appears to have entirely shed its ptDNA (Molina et al., 2014). It is still not clear how these two lineages have managed to free themselves of a plastid genome and an associated gene expression system when so many other nonphotosynthetic species have not.

Detailed microscopy work and bioinformatic analyses of nuclear-encoded plastid-targeted proteins clearly show that *Polytomella* has a plastid (Moore, Cantor, Sheeler, & Kahn, 1970; Smith & Lee, 2014). But as pointed out by Krause (2015): “One concern with the interpretation of the data from *R. lagascae* may, for some, be the current lack of physical evidence for the existence of a plastid compartment at all”. Krauss goes on to suggest that something sneaky may be going on in *Rafflesia*:

*It is feasible that the intimate association between Rafflesia and its host has led to parasite cells being populated with host plastids. The sequestration of host plastids could have relieved the parasite of the selective pressure to keep its own plastid*

*genome. Thus, the phylogenetic loss of the plastid genome may be tolerable for the parasite because it can ontogenetically 'hijack' host organelles.*

*Krause (2015)*

The appropriation of plastids by nonphotosynthetic organisms (kleptoplasty) is a well-documented phenomenon, performed by some dinoflagellates (Gast, Moran, Dennett, & Caron, 2007) and even animals, such as the sea slug *Elysia chlorotica*, which steals plastids from the heterokont alga *Vaucheria litorea* (Pelletreau et al., 2011). However, there are currently no confirmed examples of kleptoplasty being performed by an archaeplastid species.

If forfeiting photosynthesis results in severe reduction or outright loss of the plastid genome, then the green alga *Polytoma uvella* is the exception to the rule. This nonphotosynthetic chlamydomonadalean has the largest plastid genome (>230 kb, 75% noncoding DNA) observed in a colourless plant or alga, eclipsing that of its close photosynthetic relative *Chlamydomonas leiostraca* by more than 60 kb (Figueroa-Martinez, Nedelcu, Smith, & Reyes-Prieto, 2017). *P. uvella* is also closely related to *Polytomella*, but each of these two lineages lost photosynthesis independently of one another and, at first glance, their plastid genomic architectures could not be more different: expansion and complete deletion, respectively. Although large, the *P. uvella* plastid genome has, like other nonphotosynthetic ptDNAs, undergone significant gene loss, shedding all coding regions for photosynthetic pathways, and most of its excess genomic baggage is in the form of noncoding DNA. Moreover, plastid genomic inflation and genome loss are not necessarily mutually exclusive—for all we know, the ancestor of *Polytomella* might have also had a bloated plastid genome before discarding its ptDNA (Figueroa-Martinez, Nedelcu, Reyes-Prieto, & Smith, 2017). As highlighted in the following section, the group to which both *Polytomella* spp. and *P. uvella* belong—the Chlamydomonadales—has an unparalleled propensity for plastid genomic inflation, with at least six members known to have ptDNAs exceeding 250 kb (Featherston, Arakaki, Nozaki, Durand, & Smith, 2016).



## 6. THREE DECADES OF PRIMARY PLASTID GENOMICS

*My soul is a hidden orchestra; I know not what instruments, what fiddlestrings and harps, drums and tamboura I sound and clash inside myself. All I hear is the symphony.*

*Fernando Pessoa*



One cannot emphasize enough how much the field of genomics has improved our understanding of primary endosymbiosis and primary plastid evolution (Daniell, Lin, Yu, Chang, 2016). Each advancement in molecular sequencing has brought with it a leap forward in plastid research—from the first ptDNA gene sequences (McIntosh, Poulsen, & Bogorad, 1980), to the first completely sequenced plastid genomes (tobacco and the common liverwort) (Ohyama et al., 1986; Shinozaki et al., 1986) to today's array of mind-boggling metagenomic techniques, which can generate dozens of ptDNAs in a few hours (Li et al., 2014; Stull et al., 2013; Twyford & Ness, 2017). I have benefitted professionally from these advancements: my PhD and postdoctoral work were largely built on plastid genomics, and so were those of many of my friends and colleagues. Personally, I think there is nothing more enjoyable than sitting down at a computer and piecing together an interesting genome, and plastids have more than their fair share of them.

Over the past 30 years, ptDNA has been—and will likely continue to be—among the most highly sequenced kinds of eukaryotic chromosome (Sanitá Lima, Woods, Cartwright, & Smith, 2016). As of February 2017, there are >1400 complete plastid genome sequences from land plants, and at least 125 for green algae, red algae, and glaucophytes, not to mention those from secondary plastids. These sequence data have provided the raw material for countless phylogenies, helping scientists trace the origin, diversification, and lateral spread of plastids (Keeling, 2013), and they have greatly enhanced our understanding of plastid biochemistry and physiology (Gould, Waller, & McFadden, 2008) as well as other sometimes puzzling cellular processes, such as intracellular gene transfer (Kleine, Maier, & Leister, 2009; Smith, Crosby, & Lee, 2011). Plastid DNA is also among the most widely used genetic markers in plant and algal research, both for population-level studies (Parks, Cronn, & Liston, 2009) and for broad-scale comparative analyses, like resolving the relationships among major groups on the eukaryotic tree of life (Ruhfel, Gitzendanner, Soltis, Soltis, & Burleigh, 2014). Some have even argued that plastid genes, such as *rbcl* and *matK*, should be used as universal genetic barcodes for characterizing plant and algal biodiversity (CBOL Plant Working Group, 2009), which seems appropriate given that such sequences have consistently helped researchers identify and/or characterize poorly studied lineages (Worden et al., 2012).

If all of that weren't enough, primary plastids have become a popular canvas for bioengineering: transgenic ptDNA is aiding in the mass

production of food, biofuels, vaccines, and other biopharmaceuticals, such as human blood proteins (Bock, 2001; Daniell, Lin, Yu, Chang, 2016; Su et al., 2015). Archaeologists and palaeobotanists, too, have long depended on plastid genomes for their research, extracting and sequencing ptDNA from fossils that are thousands—even millions—of years old (Hagelberg, Hofreiter, & Keyser, 2015). Using ptDNA isolated and sequenced from permafrost, for example, scientists pieced together the history of plant species populating the Arctic over the past 50,000 years (Willerslev et al., 2014). The results suggest that there was a surprisingly large amount of plant diversity in the ancient Arctic, more diversity, in fact, than the contemporary Arctic. What's more, the types of plants found in the Arctic have changed considerably over time, shifting from primarily wildflowers and other herbaceous vascular plants to more woody plants and grasses, which might have contributed to the decline of large mammals like the woolly mammoth (Willerslev et al., 2014). Similar kinds of studies have also been done on algae (Lejzerowicz et al., 2013; Stoof-Leichsenring et al., 2014).

More than anything, perhaps, the sequencing and characterization of ptDNAs have provided researchers with a treasure trove of interesting and unconventional genomes (Smith & Keeling, 2015), some of which have redefined well-established rules in genetics (Bendich, 2004; Knoop, 2011), helped generate leading hypotheses on molecular evolution (Lynch, 2007), and stimulated important discussions about the roles of adaptive vs non-adaptive processes in shaping genomic complexity (Smith, 2016).

Although not always as extreme as their mitochondrial counterparts (Smith & Keeling, 2015), plastid genomes are, by any measure, remarkably diverse. Within the Archaeplastida alone, ptDNAs can range in size by two orders of magnitude: from 11 kb in the parasitic land plant *Pilostyles aethiopica* (Bellot & Renner, 2016) to over 1000 kb in the beautiful and enormous unicellular green alga *Acetabularia acetabulum* (de Vries et al., 2013). This size variation is primarily a reflection of differences in noncoding DNA content, which has led to various hypotheses about why some plastids are so prone to genomic inflation while others are not (Smith, 2016). That said, plastid gene content can be highly variable: certain red algae have 250 unique genes encrypted in their ptDNAs (Janouškovec, Liu, et al., 2013), whereas heterotrophic archaeplastids can have as few as five (Bellot & Renner, 2016)—or none at all if you consider the ptDNA-lacking *Polytomella* and *Rafflesia*. In some species, particularly prasinophytes, the plastid genes are entirely devoid of introns (Lemieux, Otis, & Turmel, 2014), but in others, like the

chlamydomonadalean *Dunaliella salina*, introns have proliferated throughout nearly every coding region (Del Vasto et al., 2015). And the genes and introns themselves can be odd: in *C. reinhardtii*, and its close relatives, the exons and introns of *psaA* are fragmented and scrambled throughout the genome and sewed back together again at the RNA level through trans-splicing (Kück, Choquet, Schneider, Dron, & Bennoun, 1987); equally complex exon–intron arrangements exist in other primary plastids as well (Glanz & Kück, 2009).

Posttranscriptional processing is even more elaborate in land plant plastids where cytosine-to-uracil and uracil-to-cytosine substitutional RNA editing can be rampant (Knoop, 2011)—a trait not yet found in any other primary plastid-bearing lineages. For example, in the ptDNA of the seedless vascular plant *Selaginella*, 3415 C-to-U RNA editing sites were identified, including ones in noncoding regions (Oldenkott, Yamaguchi, Tsuji-Tsukinoki, Knie, & Knoop, 2014), paralleling the situation in the mitochondrial compartment (Hecht, Grewe, & Knoop, 2011). *Selaginella* is also one of the few known lineages (along with certain trebouxiophytes) with GC-rich ptDNA (Smith, 2012), contrasting with the typically high AT content of most other plastids (e.g., the green algae *Helicosporidium* sp. and *Pedinomonas minor*). Remarkably, all known species with GC-biased ptDNA also have GC-rich mtDNA, but, strangely, the reverse is not true (Smith, 2012).

The evolutionary roots of these ptDNA architectural features, including GC-richness, RNA editing, and genomic expansion, are poorly understood, but they are thought to have arisen through nonadaptive processes (Gray, Lukeš, Archibald, Keeling, & Doolittle, 2010), and might be connected to mutation rate, recombination, and/or aspects of the DNA maintenance machinery (Smith & Keeling, 2015). Unfortunately, we still have a relatively narrow understanding of plastid mutation rates, but recent studies, including mutation accumulation experiments and substitution rate data, suggest that they might be lower than once thought, especially in algae (Ness, Kraemer, Colegrave, & Keightley, 2016; Smith, 2015).

But it is the structure and conformation of plastid genomes that remain one of their most misunderstood and misrepresented features (Bendich, 2004). Pick up a recent review on plastid genetics or a ptDNA genome paper published in the past year and you are likely to find the words “circular chromosome”. This is because, with few exceptions (Janouškovec, Sobotka, et al., 2013), plastid genomes assemble and map as circular molecules. Their true conformation, however, is usually more complicated:

*Circular forms of [organelle] DNA from plants appear to have exerted a profound influence on 40 years of research, despite the weakness of the data in support of the notion that most or all functions of organellar chromosomes are served by circular DNA molecules (Williamson, 2002; Bendich, 2010). When in-gel methods are employed, chromosomal DNA molecules in the plastids and mitochondria of plants appear as linear and branched-linear forms of various sizes, are found in meristematic tissues, and are typically larger than the size of the genome. In maize, tobacco, and Medicago truncatula, restriction digest analysis showed that the linear molecules have ends at defined regions of the plastid genome and isomers with three to six distinct ends ... The circular forms account for a few percent or less of total [organelle] DNA ...*

*Oldenburg and Bendich (2015)*

Overall, the paucity of detailed data on plastid chromosome structure has contributed to the popular misconception that plastid genomes exist as intact, unit-sized circular molecules, when in fact some are clearly made up of complex multigenomic branched structures. Understanding and characterizing this level of chromosomal complexity requires much more than DNA sequencing and bioinformatics analyses: it demands a wide assortment of experiments, from pulsed-field gel electrophoresis to Southern blots—experiments that are often lacking from contemporary organelle genome studies (Sanitá Lima et al., 2016).



## **7. CONCLUDING THOUGHTS: TOO MUCH SEQUENCING, NOT ENOUGH EXPERIMENTING?**

*I was taught that the way of progress was neither swift nor easy.*

*Marie Currie*

One of my first projects as a graduate student was to sequence and assemble the plastid genome of *Volvox carteri*. Using PCR, cloning, and Sanger sequencing, I slowly pieced together the ptDNA of this model multicellular green alga. But it was slow going, let me tell you. This was partly because of the old-school PCR/cloning-based approach I was using, but also because the genome was big (>500kb) and full of repeats (Smith & Lee, 2010), which is a recurring theme among chlamydomonadalean ptDNAs (Featherston, Arakaki, Nozaki, Durand, & Smith, 2016). About halfway through the project, I stumbled upon a useful resource: the United States Department of Energy Joint Genome Institute (DOE JGI) was in the process of sequencing the nuclear genome of *Volvox* (Prochnik et al., 2010) and had deposited in GenBank all their raw Sanger-sequencing data, which were teeming with ptDNA-derived reads. By syphoning off these publically

available plastid reads, I quickly completed my *Volvox* plastid assembly. And, as I was told, I did not step on anybody's toes in doing so because the DOE JGI was not even interested in the plastid data, treating it more as a contamination rather than an asset—a mindset that persists today among many in the plant genomics community.

I've since gone on to make a modest research living out of mining sequencing databases, like GenBank's Sequence Read Archive (SRA), for organelle genomes and transcriptomes. With thousands of plant and algal high-throughput DNA and RNA sequencing data sets freely available in the SRA (Smith, 2013), and the recent development of open-source software for assembling organelle genomes from these kinds of data (Castandet, Hotto, Strickler, & Stern, 2016), there's never been a better time to be studying plastid genomics and transcriptomics. However, it is also hard not to get disheartened and disillusioned by this prevailing “sequence first, ask questions later” approach to contemporary organelle research.

Over 2600 organelle genome sequences were published in the last 5 years, and their rate of publication increases with each new year (Sanitá Lima et al., 2016). There are now journals entirely devoted to plastid and mitochondrial genome papers, many of which specialize in “genome announcements”—short 500-word articles that present a new chromosome sequence, its GenBank accession number, and little else (Smith, 2017). Often fast-tracked and sometimes not formally peer-reviewed, these pint-sized genome studies are being used as quick and easy routes to a publication. Approximately 75% of all organelle genome papers released in the past half-decade were genome announcements, and only 3% contained any “complementary” analyses to accompany the sequence data, such as gel electrophoresis, Southern or Northern blotting, or transcriptional profiling (Sanitá Lima et al., 2016). In other words, the plastid research community is generating more and more ptDNA sequences, but data on other crucial aspects of plastid genomes are going undescribed. This cannot be a good thing. What's the use of having thousands of plastid genomes if we don't understand how that information is stored, organized, and expressed, and how it is connected to deeper aspects of the genomic architecture, such as mutation rate, recombination, and DNA repair?

But I shouldn't complain too much. Compared to some other fields, we are lucky to have such large amounts of data to work with. And, again, there are hundreds, if not thousands, of plastid genomes and transcriptomes just waiting to be assembled from publically available data in the SRA, not to mention how new sequencing technologies, like single molecule real-time

sequencing, are making it possible to recover entire ptDNAs with a handful of reads. It is, without a doubt, exciting to think about how these and other advancements will change the field of plastid biology and endosymbiosis over the coming years. I, for one, will be watching closely, so keep an eye on your sequencing reads.

## ACKNOWLEDGEMENTS

D.R.S. is supported by a Discovery Grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada. He can be found online at [www.arogantgenome.com](http://www.arogantgenome.com).

## REFERENCES

- Adl, S. M., Simpson, A. G., Lane, C. E., Lukeš, J., Bass, D., Bowser, S. S., et al. (2012). The revised classification of eukaryotes. *Journal of Eukaryotic Microbiology*, *59*, 429–514.
- Archibald, J. (2014). *One plus one equals one: Symbiosis and the evolution of complex life*. USA: Oxford University Press.
- Archibald, J. M. (2015). Endosymbiosis and eukaryotic cell evolution. *Current Biology*, *25*, R911–R921.
- Asmail, S. R., & Smith, D. R. (2016). Retention, erosion, and loss of the carotenoid biosynthetic pathway in the nonphotosynthetic green algal genus *Polytomella*. *New Phytologist*, *209*, 899–903.
- Ball, S. G., Bhattacharya, D., & Weber, A. P. (2016a). Pathogen to powerhouse. *Science*, *351*, 659–660.
- Ball, S. G., Bhattacharya, D., & Weber, A. P. (2016b). Infection and the first eukaryotes—Response. *Science*, *352*, 1065–1066.
- Ball, S. G., Subtil, A., Bhattacharya, D., Moustafa, A., Weber, A. P., Gehre, L., et al. (2013). Metabolic effectors secreted by bacterial pathogens: Essential facilitators of plastid endosymbiosis? *Plant Cell*, *25*, 7–21.
- Barbrook, A. C., Howe, C. J., & Purton, S. (2006). Why are plastid genomes retained in non-photosynthetic organisms? *Trends in Plant Science*, *11*, 101–108.
- Bellot, S., & Renner, S. S. (2016). The plastomes of two species in the endoparasite genus *Pilostyles* (Apodanthaceae) each retain just five or six possibly functional genes. *Genome Biology and Evolution*, *8*, 189–201.
- Bendich, A. J. (2004). Circular chloroplast chromosomes: The grand illusion. *Plant Cell*, *16*, 1661–1666.
- Bendich, A. J. (2010). The end of the circle for yeast mitochondrial DNA. *Molecular Cell*, *39*, 831–832.
- Bock, R. (2001). Transgenic plastids in basic research and plant biotechnology. *Journal of Molecular Biology*, *312*, 425–438.
- Castandet, B., Hotto, A. M., Strickler, S. R., & Stern, D. B. (2016). ChloroSeq, an optimized chloroplast RNA-Seq bioinformatic pipeline, reveals remodeling of the organellar transcriptome under heat stress. *G3: Genes, Genomes, Genetics*, *6*, 2817–2827.
- CBOL Plant Working Group. (2009). A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 12794–12797.
- Couradeau, E., Benzerara, K., Gerard, E., Moreira, D., Bernard, S., Brown, G. E., et al. (2012). An early-branching microbialite cyanobacterium forms intracellular carbonates. *Science*, *336*, 459–462.

- Daniell, H., Lin, C. S., Yu, M., & Chang, W. J. (2016). Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. *Genome Biology*, *17*, 134.
- de Castro, F., Gaedke, U., & Boenigk, J. (2009). Reverse evolution: Driving forces behind the loss of acquired photosynthetic traits. *PLoS One*, *4*, e8465.
- Del Vasto, M., Figueroa-Martinez, F., Featherston, J., Gonzalez, M. A., Reyes-Prieto, A., Durand, P. M., et al. (2015). Massive and widespread organelle genomic expansion in the green algal genus *Dunaliella*. *Genome Biology and Evolution*, *7*, 656–663.
- Delwiche, C. F., & Cooper, E. D. (2015). The evolutionary origin of a terrestrial flora. *Current Biology*, *25*, R899–R910.
- Deusch, O., Landan, G., Roettger, M., Gruenheit, N., Kowallik, K. V., Allen, J. F., et al. (2008). Genes of cyanobacterial origin in plant nuclear genomes point to a heterocyst-forming plastid ancestor. *Molecular Biology and Evolution*, *25*, 748–761.
- de Vries, J., & Archibald, J. M. (2017). Endosymbiosis: Did plastids evolve from a freshwater cyanobacterium? *Current Biology*, *27*, R103–R105.
- de Vries, J., Habicht, J., Woehle, C., Huang, C., Christa, G., Wägele, H., et al. (2013). Is *ftsH* the key to plastid longevity in sacoglossan slugs? *Genome Biology and Evolution*, *5*, 2540–2548.
- Domman, D., Horn, M., Embley, T. M., & Williams, T. A. (2015). Plastid establishment did not require a chlamydial partner. *Nature Communications*, *6*, 6421.
- Embley, T. M., & Williams, T. A. (2015). Steps on the road to eukaryotes. *Nature*, *521*, 169–170.
- Featherston, J., Arakaki, Y., Nozaki, H., Durand, P. M., & Smith, D. R. (2016). Inflated organelle genomes and a circular-mapping mtDNA probably existed at the origin of colonality in volvocine green algae. *European Journal of Phycology*, *51*, 369–377.
- Figueroa-Martinez, F., Nedelcu, A. M., Smith, D. R., & Reyes-Prieto, A. (2015). When the lights go out: The evolutionary fate of free-living colorless green algae. *New Phytologist*, *206*, 972–982.
- Figueroa-Martinez, F., Nedelcu, A. M., Reyes-Prieto, A., & Smith, D. R. (2017). The plastid genomes of nonphotosynthetic algae are not so small after all. *Communicative & Integrative Biology*, *10*, e1283080.
- Figueroa-Martinez, F., Nedelcu, A. M., Smith, D. R., & Reyes-Prieto, A. (2017). The plastid genome of *Polytoma uwella* is the largest known among colorless algae and plants and reflects contrasting evolutionary paths to nonphotosynthetic lifestyles. *Plant Physiology*, *173*(2), 932–943.
- Gagat, P., Bodyl, A., & Mackiewicz, P. (2016). Breaking the endosymbiont-organelle border: The case of *Paulinella chromatophora*. *Endocytobiosis and Cell Research*, *27*, 45–51.
- Gast, R. J., Moran, D. M., Dennett, M. R., & Caron, D. A. (2007). Kleptoplasty in an Antarctic dinoflagellate: Caught in evolutionary transition? *Environmental Microbiology*, *9*, 39–45.
- Glanz, S., & Kück, U. (2009). Trans-splicing of organelle introns—A detour to continuous RNAs. *BioEssays*, *31*, 921–934.
- Gould, S. B. (2016). Infection and the first eukaryotes. *Science*, *352*, 1065.
- Gould, S. B., Waller, R. F., & McFadden, G. I. (2008). Plastid evolution. *Annual Review of Plant Biology*, *59*, 491–517.
- Gray, M. W. (2012). Mitochondrial evolution. *Cold Spring Harbor Perspectives in Biology*, *4*, a011403.
- Gray, M. W., Lukeš, J., Archibald, J. M., Keeling, P. J., & Doolittle, W. F. (2010). Irremediable complexity? *Science*, *330*, 920–921.
- Hagelberg, E., Hofreiter, M., & Keyser, C. (2015). Introduction. Ancient DNA: The first three decades. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*, 201303711.



- Hecht, J., Grewe, F., & Knoop, V. (2011). Extreme RNA editing in coding islands and abundant microsatellites in repeat sequences of *Selaginella moellendorffii* mitochondria: The root of frequent plant mtDNA recombination in early tracheophytes. *Genome Biology and Evolution*, *3*, 344–358.
- Howe, C. J., Barbrook, A. C., Nisbet, R. E. R., Lockhart, P. J., & Larkum, A. W. D. (2008). The origin of plastids. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*, 2675–2685.
- Huang, J., & Gogarten, J. P. (2007). Did an ancient chlamydial endosymbiosis facilitate the establishment of primary plastids? *Genome Biology*, *8*, R99.
- Janouškovec, J., Liu, S. L., Martone, P. T., Carré, W., Leblanc, C., Collén, J., et al. (2013). Evolution of red algal plastid genomes: Ancient architectures, introns, horizontal gene transfer, and taxonomic utility of plastid markers. *PLoS One*, *8*, e59001.
- Janouškovec, J., Sobotka, R., Lai, D. H., Flegontov, P., Koník, P., Komenda, J., et al. (2013). Split photosystem protein, linear-mapping topology, and growth of structural complexity in the plastid genome of *Chromera velia*. *Molecular Biology and Evolution*, *30*, 2447–2462.
- Keeling, P. J. (2013). The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. *Annual Review of Plant Biology*, *64*, 583–607.
- Kim, S., & Park, M. G. (2016). *Paulinella longichromatophora* sp. nov., a new marine photosynthetic testate amoeba containing a chromatophore. *Protist*, *167*, 1–12.
- Kleine, T., Maier, U. G., & Leister, D. (2009). DNA transfer from organelles to the nucleus: The idiosyncratic genetics of endosymbiosis. *Annual Review of Plant Biology*, *60*, 115–138.
- Knoop, V. (2011). When you can't trust the DNA: RNA editing changes transcript sequences. *Cellular and Molecular Life Sciences*, *68*, 567–586.
- Kolisko, M., Boscaro, V., Burki, F., Lynn, D. H., & Keeling, P. J. (2014). Single-cell transcriptomics for microbial eukaryotes. *Current Biology*, *24*, R1081–R1082.
- Krause, K. (2015). Grand-scale theft: Kleptoplasty in parasitic plants? *Trends in Plant Science*, *20*, 196–198.
- Kück, U., Choquet, Y., Schneider, M., Dron, M., & Bennoun, P. (1987). Structural and transcription analysis of two homologous genes for the P700 chlorophyll  $\alpha$ -apoproteins in *Chlamydomonas reinhardtii*: Evidence for in vivo trans-splicing. *The EMBO Journal*, *6*, 2185–2195.
- Kurland, C. G., Collins, L. J., & Penny, D. (2006). Genomics and the irreducible nature of eukaryote cells. *Science*, *312*, 1011–1014.
- Larkum, A. W. D., Lockhart, P. J., & Howe, C. J. (2007). Shopping for plastids. *Trends in Plant Science*, *12*, 189–195.
- Lejzerowicz, F., Esling, P., Majewski, W., Szczuciński, W., Decelle, J., Obadia, C., et al. (2013). Ancient DNA complements microfossil record in deep-sea subsurface sediments. *Biology Letters*, *9*, 20130283.
- Lemieux, C., Otis, C., & Turmel, M. (2014). Six newly sequenced chloroplast genomes from prasinophyte green algae provide insights into the relationships among prasinophyte lineages and the diversity of streamlined genome architecture in picoplanktonic species. *BMC Genomics*, *15*, 857.
- Li, Q., Li, Y., Song, J., Xu, H., Xu, J., Zhu, Y., et al. (2014). High-accuracy de novo assembly and SNP detection of chloroplast genomes using a SMRT circular consensus sequencing strategy. *New Phytologist*, *204*, 1041–1049.
- Lynch, M. (2007). *The origins of genome architecture*. Sunderland, MA: Sinauer Associates.
- Martin, W. F., Roettger, M., Ku, C., Garg, S. G., Nelson-Sathi, S., & Landan, G. (2016). Late mitochondrial origin is pure artefact. *bioRxiv*. <http://dx.doi.org/10.1101/055368>.
- McIntosh, L., Poulsen, C., & Bogorad, L. (1980). Chloroplast gene sequence for the large subunit of ribulose biphosphatocarboxylase of maize. *Nature*, *288*, 556–560.



- Mereshkowsky, C. (1910). Theorie der zwei Plasmaarten als Grundlage der Symbiogenesis, einer neuen Lehre von der Entstehung der Organismen. *Biologisches Centralblatt*, *30*, 278–288.
- Metzker, M. L. (2010). Sequencing technologies—The next generation. *Nature Reviews Genetics*, *11*, 31–46.
- Molina, J., Hazzouri, K. M., Nickrent, D., Geisler, M., Meyer, R. S., Pentony, M. M., et al. (2014). Possible loss of the chloroplast genome in the parasitic flowering plant *Rafflesia lagascae* (Rafflesiaceae). *Molecular Biology and Evolution*, *31*, 793–803.
- Moore, J., Cantor, M. H., Sheeler, P., & Kahn, W. (1970). The ultrastructure of *Polytomella agilis*. *The Journal of Protozoology*, *17*, 671–676.
- Moran, M. A. (2009). Metatranscriptomics: Eavesdropping on complex microbial communities: Large-scale sequencing of mRNAs retrieved from natural communities provides insights into microbial activities and how they are regulated. *Microbe*, *4*, 329–335.
- Morgan-Kiss, R. M., Priscu, J. C., Pockock, T., Gudynaite-Savitch, L., & Hüner, N. P. (2006). Adaptation and acclimation of photosynthetic microorganisms to permanently cold environments. *Microbiology and Molecular Biology Reviews*, *70*, 222–252.
- Nair, S. C., & Striepen, B. (2011). What do human parasites do with a chloroplast anyway? *PLoS Biology*, *9*, e1001137.
- Nakayama, T., & Archibald, J. M. (2012). Evolving a photosynthetic organelle. *BMC Biology*, *10*, 35.
- Ness, R. W., Kraemer, S. A., Colegrave, N., & Keightley, P. D. (2016). Direct estimate of the spontaneous mutation rate uncovers the effects of drift and recombination in the *Chlamydomonas reinhardtii* plastid genome. *Molecular Biology and Evolution*, *33*, 800–808.
- Nowack, E. C. (2014). *Paulinella chromatophora*—rethinking the transition from endosymbiont to organelle. *Acta Societatis Botanicorum Poloniae*, *83*, 387–397.
- Nowack, E. C., & Grossman, A. R. (2012). Trafficking of protein into the recently established photosynthetic organelles of *Paulinella chromatophora*. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 5340–5345.
- Nowack, E. C., Price, D. C., Bhattacharya, D., Singer, A., Melkonian, M., & Grossman, A. R. (2016). Gene transfers from diverse bacteria compensate for reductive genome evolution in the chromatophore of *Paulinella chromatophora*. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 12214–12219.
- Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., et al. (1986). Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. *Nature*, *322*, 572–574.
- Oldenburg, D. J., & Bendich, A. J. (2015). DNA maintenance in plastids and mitochondria of plants. *Frontiers in Plant Science*, *6*, 883.
- Oldenkott, B., Yamaguchi, K., Tsuji-Tsukinoki, S., Knie, N., & Knoop, V. (2014). Chloroplast RNA editing going extreme: More than 3400 events of C-to-U editing in the chloroplast transcriptome of the lycophyte *Selaginella uncinata*. *RNA*, *20*, 1499–1506.
- Parks, M., Cronn, R., & Liston, A. (2009). Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biology*, *7*, 84.
- Pelletreau, K. N., Bhattacharya, D., Price, D. C., Worful, J. M., Moustafa, A., & Rumpho, M. E. (2011). Sea slug kleptoplasty and plastid maintenance in a metazoan. *Plant Physiology*, *155*, 1561–1565.
- Pittis, A. A., & Gabaldón, T. (2016). Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry. *Nature*, *531*, 101–104.
- Ponce-Toledo, R. I., Deschamps, P., López-García, P., Zivanovic, Y., Benzerara, K., & Moreira, D. (2017). An early-branching freshwater cyanobacterium at the origin of plastids. *Current Biology*, *27*, 386–391.

- Prochnik, S. E., Umen, J., Nedelcu, A. M., Hallmann, A., Miller, S. M., Nishii, I., et al. (2010). Genomic analysis of organismal complexity in the multicellular green alga *Volvox carteri*. *Science*, *329*, 223–226.
- Raven, J. A. (1997). Phagotrophy in phototrophs. *Limnology and Oceanography*, *42*, 198–205.
- Reyes-Prieto, A., Yoon, H. S., Moustafa, A., Yang, E. C., Andersen, R. A., Boo, S. M., et al. (2010). Differential gene retention in plastids of common recent origin. *Molecular Biology and Evolution*, *27*, 1530–1537.
- Ruhfel, B. R., Gitzendanner, M. A., Soltis, P. S., Soltis, D. E., & Burleigh, J. G. (2014). From algae to angiosperms—inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. *BMC Evolutionary Biology*, *14*, 23.
- Sagan, L. (1967). On the origin of mitosing cells. *Journal of Theoretical Biology*, *14*, 225–274.
- Sanitá Lima, M., Woods, L. C., Cartwright, M. W., & Smith, D. R. (2016). The (in)complete organelle genome: Exploring the use and nonuse of available technologies for characterizing mitochondrial and plastid chromosomes. *Molecular Ecology Resources*, *16*, 1279–1286.
- Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., et al. (1986). The complete nucleotide sequence of the tobacco chloroplast genome: Its gene organization and expression. *The EMBO Journal*, *5*, 2043–2049.
- Smith, D. R. (2012). Updating our view of organelle genome nucleotide landscape. *Frontiers in Genetics*, *3*, 175.
- Smith, D. R. (2013). RNA-Seq data: A goldmine for organelle research. *Briefings in Functional Genomics*, *12*, 454–456.
- Smith, D. R. (2015). Mutation rates in plastid genomes: They are lower than you might think. *Genome Biology and Evolution*, *7*, 1227–1234.
- Smith, D. R. (2016). The mutational hazard hypothesis of organelle genome evolution: 10 years on. *Molecular Ecology*, *25*, 3769–3775.
- Smith, D. R. (2017). Goodbye genome paper, hello genome report: The increasing popularity of “genome announcements” and their impact on science. *Briefings in Functional Genomics*, *16*, 156–162.
- Smith, D. R., & Asmail, S. R. (2014). Next-generation sequencing data suggest that certain nonphotosynthetic green plants have lost their plastid genomes. *New Phytologist*, *204*, 7–11.
- Smith, D. R., Crosby, K., & Lee, R. W. (2011). Correlation between nuclear plastid DNA abundance and plastid number supports the limited transfer window hypothesis. *Genome Biology and Evolution*, *3*, 365–371.
- Smith, D. R., & Keeling, P. J. (2015). Mitochondrial and plastid genome architecture: Reoccurring themes, but significant differences at the extremes. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 10177–10184.
- Smith, D. R., & Lee, R. W. (2010). Low nucleotide diversity for the expanded organelle and nuclear genomes of *Volvox carteri* supports the mutational-hazard hypothesis. *Molecular Biology and Evolution*, *27*, 2244–2256.
- Smith, D. R., & Lee, R. W. (2014). A plastid without a genome: Evidence from the non-photosynthetic green algal genus *Polytomella*. *Plant Physiology*, *164*, 1812–1819.
- Spang, A., Saw, J. H., Jørgensen, S. L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A. E., et al. (2015). Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature*, *521*, 173–179.
- Stoecker, D. K. (1998). Conceptual models of mixotrophy in planktonic protists and some ecological and evolutionary implications. *European Journal of Protistology*, *34*, 281–290.
- Stoof-Leichsenring, K. R., Herzsich, U., Pestryakova, L. A., Klemm, J., Epp, L. S., & Tiedemann, R. (2014). Genetic data from algae sedimentary DNA reflect the influence of environment over geography. *Scientific Reports*, *5*, 12924.

- Stull, G. W., Moore, M. J., Mandala, V. S., Douglas, N. A., Kates, H. R., Qi, X., et al. (2013). A targeted enrichment strategy for massively parallel sequencing of angiosperm plastid genomes. *Applications in Plant Sciences*, *1*, 1200497.
- Su, J., Zhu, L., Sherman, A., Wang, X., Lin, S., Kamesh, A., et al. (2015). Low cost industrial production of coagulation factor IX bioencapsulated in lettuce cells for oral tolerance induction in hemophilia B. *Biomaterials*, *70*, 84–93.
- Tittel, J., Bissinger, V., Zippel, B., Gaedke, U., Bell, E., Lorke, A., et al. (2003). Mixotrophs combine resource use to outcompete specialists: Implications for aquatic food webs. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 12776–12781.
- Twyford, A. D., & Ness, R. W. (2017). Strategies for complete plastid genome sequencing. *Molecular Ecology Resources*. in press. <http://onlinelibrary.wiley.com/doi/10.1111/1755-0998.12626/abstract>.
- Wallin, I. E. (1927). *Symbiogenesis and the origin of species*. London: Bailliere, Tindall and Cox.
- Willerslev, E., Davison, J., Moora, M., Zobel, M., Coissac, E., Edwards, M. E., et al. (2014). Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature*, *506*, 47–51.
- Williamson, D. (2002). The curious history of yeast mitochondrial DNA. *Nature Reviews Genetics*, *3*, 475–481.
- Worden, A. Z., Janouskovec, J., McRose, D., Engman, A., Welsh, R. M., Malfatti, S., et al. (2012). Global distribution of a wild alga revealed by targeted metagenomics. *Current Biology*, *22*, R675–R677.
- Yoon, H. S., Hackett, J. D., Ciniglia, C., Pinto, G., & Bhattacharya, D. (2004). A molecular timeline for the origin of photosynthetic eukaryotes. *Molecular Biology and Evolution*, *21*, 809–818.
- Yoon, H. S., Nakayama, T., Reyes-Prieto, A., Andersen, R. A., Boo, S. M., Ishida, K. I., et al. (2009). A single origin of the photosynthetic organelle in different *Paulinella* lineages. *BMC Evolutionary Biology*, *9*, 98.
- Zimorski, V., Ku, C., Martin, W. F., & Gould, S. B. (2014). Endosymbiotic theory for organelle origins. *Current Opinion in Microbiology*, *22*, 38–48.