

# Unraveling chloroplast transcriptomes with ChloroSeq, an organelle RNA-Seq bioinformatics pipeline

David Roy Smith and Matheus Sanitá Lima

Corresponding author: David Roy Smith, Department of Biology, University of Western Ontario, London, Ontario N6A 5B7, Canada. Tel.: (519) 661 2111, ext; 86482; E-mail: dsmit242@uwo.ca

## Abstract

Online sequence repositories are teeming with RNA sequencing (RNA-Seq) data from a wide range of eukaryotes. Although most of these data sets contain large numbers of organelle-derived reads, researchers tend to ignore these data, focusing instead on the nuclear-derived transcripts. Consequently, GenBank contains massive amounts of organelle RNA-Seq data that are just waiting to be downloaded and analyzed. Recently, a team of scientists designed an open-source bioinformatics program called ChloroSeq, which systemically analyzes an organelle transcriptome using RNA-Seq. The ChloroSeq pipeline uses RNA-Seq alignment data to deliver detailed analyses of organelle transcriptomes, which can be fed into statistical software for further analysis and for generating graphical representations of the data. In addition to providing data on expression levels via coverage statistics, ChloroSeq can examine splicing efficiency and RNA editing profiles. Ultimately, ChloroSeq provides a well-needed avenue for researchers of all stripes to start exploring organelle transcription and could be a key step toward a more thorough understanding of organelle gene expression.

**Key words:** chloroplast transcription; organelle transcriptomics; plastid RNA; RNA editing; RNA-Seq

## Introduction

Massively parallel high-throughput sequencing of complementary DNA (cDNA) [RNA sequencing (RNA-Seq)] has become a preeminent technique in plant research, and life science investigations as a whole [1]. Consequently, open-access sequence repositories, such as GenBank, are expanding with RNA-Seq data from diverse land plants and algae (Figure 1). As of 17 June 2016, GenBank's Sequence Read Archive (SRA) [2] contained over 39 000 RNA-Seq data sets from streptophytes, and the Marine Microbial Eukaryotic Transcriptome Sequencing Project [3] recently sequenced and made publicly available the transcriptomes from hundreds of plastid-bearing protists.

RNA-Seq data sets from land plants and algae are obviously a great resource for investigating nuclear gene expression [1], but they are also an excellent but untapped means for exploring

plastid and mitochondrial transcription [4]. Given that organelle genomes are present in many copies per cell and are highly expressed, organelle transcripts can represent a significant proportion of plant cellular RNA [5]. Thus, eukaryotic RNA-Seq libraries typically contain large numbers (1–30%) of organelle-derived transcripts [6, 7], so much so that nearly complete organelle genome sequences can sometimes be assembled from RNA-Seq data alone [8, 9].

Unfortunately, researchers carrying out RNA-Seq on eukaryotes often ignore the organelle data, focusing instead on nuclear-derived transcripts [4]. In other words, GenBank contains a treasure trove of organelle RNA-Seq data that are just waiting to be examined (Figure 1). But there has not been a sophisticated bioinformatics pipeline designed for analyzing organelle reads from eukaryotic RNA-Seq studies. That is, until now.

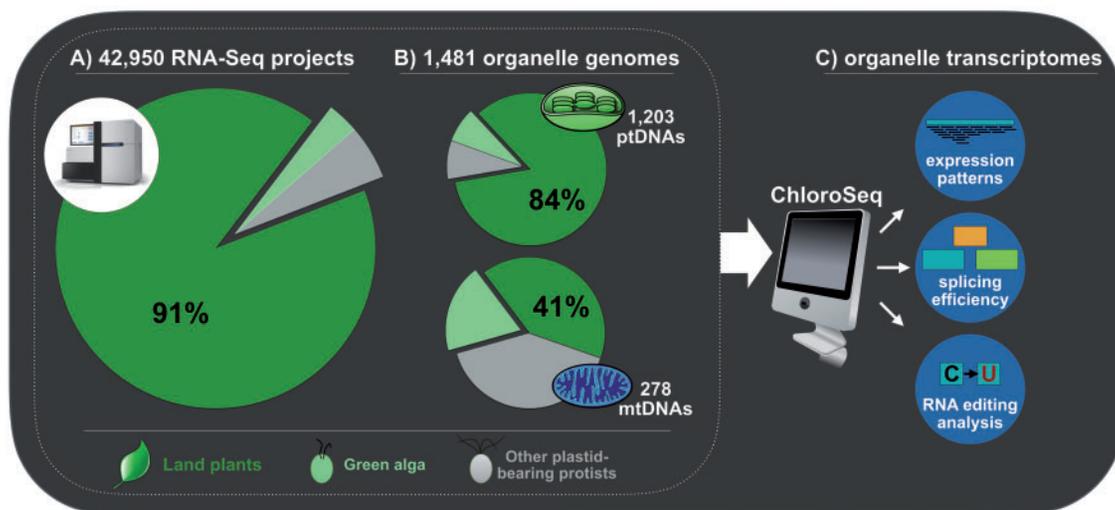
David Roy Smith is an assistant professor of biology at the University of Western Ontario. He studies genome evolution of eukaryotic microbes, and can be found online at [www.arrogantgenome.com](http://www.arrogantgenome.com) and <https://twitter.com/@arrogantgenome>.

Matheus Sanitá Lima is an MSc candidate in the biology department at the University of Western Ontario. He is currently studying organelle gene expression.

Submitted: 21 July 2016; Received (in revised form): 24 August 2016

© The Author 2016. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)



**Figure 1.** Available data in GenBank for exploring organelle transcription in plastid-bearing eukaryotes. (A) As of 17 June 2016, GenBank's SRA [<http://www.ncbi.nlm.nih.gov/sra>] contained 42950 publicly available RNA-Seq data sets from plastid-bearing species, 91% of which came from land plants. (B) Similarly, the most recent RefSeq release of mitochondrial and plastid organelle genome sequences (accessed 17 June 2016) [<http://www.ncbi.nlm.nih.gov/genome/organelle/>] included 1481 organelle genomes from land plants and algae, 1203 and 278 of which were ptDNAs and mtDNAs, respectively. This is an underestimate of the total number of available organelle genome sequences in GenBank because the RefSeq database often does not include genomes from different strains of the same species or nearly complete organelle DNAs. (C) These freely accessible RNA-Seq and organelle genome data can be used with the bioinformatics program ChloroSeq [6] to systematically analyze organelle transcriptomes.

## ChloroSeq: an organelle RNA-Seq bioinformatics pipeline

Recently, a team of scientists from the Boyce Thompson Institute at Cornell University designed a new bioinformatics program called ChloroSeq, which systematically analyzes a plastid transcriptome using RNA-Seq [6]. ChloroSeq is open-source and freely available from GitHub [<https://github.com/BenoitCastandet/chloroseq>; accessed 18 August 2016]. The program operates through command-line-driven Perl scripts, which can be easily implemented on most laptop computers, provided the user has some experience with Unix.

Once installed, ChloroSeq uses RNA-Seq alignment data (i.e. a BAM file) to deliver a detailed analysis of the plastid transcriptome. The program first indexes and then extracts the plastid reads from the alignment BAM file and uses these data for executing a variety of downstream analyses. The final output of ChloroSeq is in the form of text files (count tables), and it is important to emphasize that the program itself does not perform any statistical analyses on the transcriptional data; however, the count tables can be easily fed to other statistical software, such as R, for further investigations and for generating graphical representations of the data. Although most people associate transcriptomics with studies on differential gene expression, organelle genomes can undergo an assortment of other types of transcriptional modifications [10, 11]. Accordingly, in addition to providing data on expression levels via coverage statistics, ChloroSeq can examine splicing efficiency and RNA editing profiles.

To help carry out these different analyses, the ChloroSeq pipeline relies on other free, open-source bioinformatics programs, including the popular genomic software suites SAMtools [12] and BEDtools [13], which need to be installed on the host computer for the complete ChloroSeq workflow to run properly. And, again, users must provide an alignment BAM file, which

can be generated using most read mapping software, such as Bowtie2 and TopHat2 [14].

Not surprisingly, much of the RNA-Seq data within the SRA come from paired-end libraries that were enriched for polyadenylated transcripts and/or were depleted of ribosomal RNAs (rRNAs). These types of data sets can be used with ChloroSeq, but the software has been optimized for single-end, strand-specific RNA-Seq. Moreover, the creators of ChloroSeq advise against using data from poly(A)-enriched libraries. This is because plant organelle transcripts become unstable following polyadenylation [15] and are grossly underrepresented in these kinds of libraries. By comparing available RNA-Seq data from *Arabidopsis thaliana*, Castandet et al. [6] showed that around 1% of the reads from oligo(dT)-selected libraries mapped to the plastid genome, whereas when generated from poly(A)-depleted total RNA followed by rRNA subtraction, an astounding 30% of the reads came from the plastid. Nevertheless, if only 1% of the RNA-Seq data are plastid derived that still provides thousands and thousands of organelle reads for analysis, and means that researchers should be open to using ChloroSeq to explore any eukaryotic RNA-Seq dataset for organelle reads, no matter the protocol used to generate the library.

If you do decide to use poly(A)-enriched RNA for organelle studies, it is important to keep in mind that different types of organelle transcripts could be differentially represented in the data. Unlike the near ubiquity of polyadenylation of nuclear messenger RNAs, organelle transcripts are not necessarily polyadenylated [15, 16], and even when polyadenylation does occur, the transcripts for the various genes are often not polyadenylated at the same frequency. Moreover, polyadenylation is often a degradation signal in organelles [17], meaning that researchers using poly(A)-selected RNA-Seq for measuring differential expression in organelle systems may, in some instances, be measuring the opposite: differential degradation.

## Putting it to the test

To demonstrate the utility of ChloroSeq, Castandet *et al.* [6] applied the software to various *A. thaliana* RNA-Seq projects from the SRA for which the plastid transcript data had not been mined or studied. By comparing RNA-Seq information from plants grown under control and abiotic stress conditions, the authors showed that heat stress can result in a global reduction in plastid RNA splicing and editing efficiency as well as an increase in plastid transcript abundance, including transcripts from coding, noncoding and antisense regions of the genome. For instance, the authors used ChloroSeq to measure the ratio of spliced to un-spliced plastid RNAs and found that 12 hours of heat stress greatly inhibited the splicing efficiency of nearly all the plastid-encoded introns from *A. thaliana*, suggesting that organelle intron structure might be sensitive to temperature in a functionally significant manner [6].

By searching other available data in the SRA, one can easily identify a variety of interesting experiments to run with ChloroSeq. Members of the land plant genus *Selaginella*, for example, are known to undergo extremely high levels of organelle RNA editing [18, 19]. Indeed, transcriptome sequencing of *Selaginella uncinata* uncovered 3415 C-to-U RNA editing sites in the plastid genome, which is one of the highest levels of posttranscriptional editing ever observed for a plastid DNA (ptDNA). But detailed plastid RNA analyses have not yet been performed on any other members of the genus, even though the data needed to do so are available in GenBank. For *Selaginella moellendorffii*, there exists a complete plastid genome sequence (accession NC\_013086) and >15 different RNA-Seq data sets (e.g. SRA accessions SRX828740–5). Similarly, data from at least four RNA-Seq projects are available for *Selaginella kraussiana* (SRA accessions SRX1043962–5), and although the plastid genome of this species remains to be sequenced, one could easily generate a complete ptDNA from freely available whole genome shotgun sequencing data for *S. kraussiana* (SRA accession SRX1036537). Together, these data sets could be used in conjunction with ChloroSeq to generate complete RNA editing profiles for the ptDNAs of *S. moellendorffii* and *S. kraussiana* and provide insights into the evolution, conservation and diversity of plastid RNA editing in the *Selaginella* lineage.

If extreme RNA editing does not impress you, then widespread and bizarre intron splicing might. Expression of the *Euglena gracilis* plastid genome is a veritable circus act, requiring the removal of ~160 introns, including 15 twintrons (introns within introns), which need to be subtracted sequentially for accurate splicing [20]. Despite its record-breaking number of introns, RNA processing and intron splicing in the *E. gracilis* plastid remain poorly understood and poorly characterized. However, given that there are 22 freely available RNA-Seq data sets for this alga (e.g. SRA accessions ERX1051903–4) as well as a complete ptDNA sequence (accession NC\_001603), one could easily use ChloroSeq to investigate the plastid transcriptional architecture of *E. gracilis*.

Although designed with plastid transcriptomics in mind, ChloroSeq can also be used for studying plant and algal mitochondrial transcription [6]—or transcription from any organelle system for that matter (e.g. animal mitochondria). In fact, many of the same transcriptional modifications and peculiarities found in plastids can also occur in mitochondria, such as RNA editing [21] and trans-splicing [11]. Thus, the key features of ChloroSeq are equally as applicable to mitochondrial studies as they are to those on chloroplasts. Because of this, the software could help stimulate more thorough and extensive investigations of organelle gene expression.

Like with plants and algae, there is a plethora of publicly available RNA-Seq data from metazoans, which can be used for addressing interesting questions in organelle genetics. Medusozoans (jellyfish and hydras), for instance, can have linear or linear fragmented mitochondrial genomes [22] with elaborate telomere structures and homogenized gene sequences [23]. Although there exist dozens of completely sequenced mitochondrial DNAs (mtDNAs) and >200 RNA-Seq data sets for medusozoans, very few researchers have studied mitochondrial transcription in this lineage [24]. Using ChloroSeq to examine these mtDNA and RNA-Seq data (e.g. GenBank accessions JN593332 and SRX315373) could lead to an interesting synthesis.

## Bringing organelle transcriptomics to the forefront

Plastids and mitochondria harbor some of the most extreme and unconventional modes of gene expression identified from across the tree of life [11]. As noted above, posttranscriptional editing is rampant within the organelles of many plants and some algae. For instance, 11 of 12 possible types of substitution RNA editing (A-to-C, A-to-G, A-to-U, etc.) have been identified in the plastids of dinoflagellate algae [25], and both the plastid and mitochondrial transcripts of vascular plants can undergo moderate to severe C-to-U and/or U-to-C editing [21]. Similarly, various plastid-bearing protists use nonstandard genetic codes in their plastid and/or mitochondrion [26], and the organelle genomes of plants and algae often contain an abundance of introns, which in certain cases are trans-spliced or have unusual arrangements [27]. More recently, organelle noncoding RNAs have been shown to be possible regulators of gene expression, and certain cases might be integral components for nuclear gene regulation [28]. And organelle gene expression is integral to various aspects of cell signaling and cell physiology in plants, algae and eukaryotes as a whole, including animals [29].

Despite being so remarkable, organelle transcription remains a relatively poorly studied topic. In the past 5 years, >2500 organelle DNAs were sequenced, resulting in thousands of organelle genome papers [30]. But in the same period, only a few dozen high-quality organelle transcriptome analyses were published, most of which came from model species [31, 32]. Although the human mitochondrial genome was sequenced >35 years ago, it has only been in past half decade that a detailed human mitochondrial transcriptome was published [31]. But with over 300 000 RNA-Seq data sets from diverse eukaryotes currently sitting in the SRA and with new software like ChloroSeq arriving, the time is ripe for investigating organelle transcriptomes, and if the research community takes advantage of these freely available assets (Figure 1), we might soon uncover novel and critical facets of organelle gene expression.

One of the major limitations of ChloroSeq is that it requires the input of alignment data based on a reference organelle genome sequence on which RNA-Seq reads have been mapped. This means that RNA-Seq data for which there do not exist a corresponding organelle genome sequence (or one from a close relative) cannot be used with ChloroSeq. But with thousands of complete organelle DNAs available in GenBank, and hundreds more arriving each month, this should not be a hurdle for much longer. Moreover, there is always the strong possibility that researchers can reconstruct a near-complete organelle genome sequence from the RNA data itself and then use it as a ChloroSeq reference sequence [8, 9].

Although not mandatory, most of the key functions of ChloroSeq are dependent on the existence of a proper annotation file for the organelle genome of interest. One might assume that the organelle genome data in GenBank are completely and properly annotated, but there are a surprising number of mtDNA and ptDNA sequences that are poorly and/or incorrectly annotated, and some lack annotations altogether [33]. Thus, it would be smart to verify the organelle annotation files before using them with ChloroSeq.

RNA-Seq and ChloroSeq might be great starting points for investigating transcription, but a complete picture of organelle gene expression will likely require a broad range of techniques and experiments, in addition to sequencing and bioinformatics. If past work has proven anything, it is that a deep understanding of organelle transcription can entail years of painstaking experiments, and can involve everything from advanced polymerase chain reaction, gel electrophoresis and blotting methods to high-throughput transcriptomics and proteomics. For example, it has taken >20 years of detailed RNA work to resolve the large and small subunit rRNA genes from the *Plasmodium falciparum* mitochondrial genome, which are fragmented and scrambled into ~25 distinct coding modules [34]. ChloroSeq is not a panacea for organelle transcriptional studies, but it is certainly a well-needed tool in an environment where there are too few bioinformatics programs devoted to organelle research.

## The growth of bioinformatics software for organelle research

ChloroSeq is among a handful of free bioinformatics software packages dedicated to studying plastid and mitochondrial genetics. Other popular programs include RNAweasel and MFannot (<http://megasun.bch.umontreal.ca/RNAweasel/>), which predict and model complex organelle RNAs and annotate introns and exons, as well as the web servers MITOFY [35] and Organellar Genome Draw [36], which respectively annotate and graphically map organelle genomes. The ORGanelle ASSEMBler [<https://git.metabarcoding.org/org-asm/org-asm/wikis/home>; accessed 18 August 2016] is an open-source program designed to assemble complete organelle DNAs (and other small genomes) from whole genome shotgun sequencing data. Similar to ChloroSeq, the programs PREP-Mt [37] and PREPACT 2.0 [38] predict RNA editing sites in organelle genomes by searching against databases of known sequences, but unlike ChloroSeq, they cannot make use of raw RNA-Seq data and next-generation sequencing read mappers.

Together, these and other software suites [39] have helped streamline the study of organelle genomics, saving researchers time and energy. Yet, it is disappointing that there are not more bioinformatics programs specifically designed for analyzing organelle genomes. Organelle genetic data are used in a surprisingly wide variety of scientific disciplines, including medicine, forensics, genetic engineering and archeology, to name but a few, and they have yielded countless fundamental insights into our understanding of the origins, evolution and diversification of eukaryotic life, and continue to do so [40, 41].

As scientists, it is paramount that we use the data that are available to us now and that will become available in the near and distant future. For researchers that study organelles, ChloroSeq will help make this possible. As more bioinformatics programs devoted to plastid and mitochondrial genetics arise, we could soon find ourselves in a position where many (even most) aspects of organelle genomic and transcriptomic analyses

are automated—in fact, we have arguably nearly reached this point. Likewise, it will soon be possible to outsource nearly all of the laboratory and bioinformatics work required to generate, assemble, annotate and analyze an organelle genome. I recently received an e-mail from a company called Phyzen [<http://www.phyzen.com>; accessed 18 August 2016], advertising complete plastid genome assemblies, including annotations and GenBank submission files, for a few thousand US dollars. With ChloroSeq now freely available, I am betting that they will soon add plastid transcriptome analyses to their list of services.

### Key Points

- High-throughput sequencing of cDNA (RNA-Seq) has become a preeminent technique in life science research and, consequently, open-access sequence repositories are expanding with RNA-Seq data from diverse eukaryotes.
- Eukaryotic RNA-Seq data sets typically contain large numbers of organelle-derived reads, but researchers tend to ignore these data, focusing instead on the nuclear-derived transcripts. Moreover, there is a paucity of bioinformatics software for analyzing organelle transcriptomes.
- Recently, researchers designed a freely available bioinformatics program called ChloroSeq, which systematically analyzes an organelle transcriptome using RNA-Seq.
- The ChloroSeq pipeline uses RNA-Seq alignment data to deliver detailed analyses of organelle transcriptomes, including splicing efficiencies and RNA editing profiles.
- Our understanding of organelle transcription is surprisingly limited, despite the fact that mitochondria and chloroplasts harbor some of the most unusual modes of gene expression ever identified. ChloroSeq provides a well-needed avenue for researchers of all stripes to start exploring organelle transcription.

### Funding

The Natural Sciences and Engineering Research Council (NSERC) of Canada (Discovery Grant to D.R.S.)

### References

1. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 2009;10:57–63.
2. Kodama Y, Shumway M, Leinonen R. The sequence read archive: explosive growth of sequencing data. *Nucleic Acids Res* 2012;40:D54–6.
3. Keeling PJ, Burki F, Wilcox HM, et al. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol* 2014;12:e1001889.
4. Smith DR. RNA-Seq data: a goldmine for organelle research. *Brief Funct Genomics* 2013;12:454–6.
5. Loening UE, Ingle J. Diversity of RNA components in green plant tissues. *Nature* 1967;215:363–7.
6. Castandet B, Hotto AM, Strickler SR, et al. ChloroSeq, an optimized chloroplast RNA-Seq bioinformatic pipeline, reveals

- remodeling of the organellar transcriptome under heat stress. *G3* 2016, doi: 10.1534/g3.116.030783.
7. Raz T, Kapranov P, Lipson D, et al. Protocol dependence of sequencing-based gene expression measurements. *PLoS One* 2011;6:e19287.
  8. Tian Y, Smith DR. Recovering complete mitochondrial genome sequences from RNA-Seq: a case study of *Polytomella* non-photosynthetic green algae. *Mol Phylogenet Evol* 2016;98:57–62.
  9. Shi S, Wang S, Xia EH, et al. Full transcription of the chloroplast genome in photosynthetic eukaryotes. *Sci Rep* 2016;6:30135.
  10. Moreira S, Breton S, Burger G. Unscrambling genetic information at the RNA level. *Wiley Interdiscip Rev RNA* 2012;3:213–28.
  11. Smith DR, Keeling PJ. Protists and the wild, wild west of gene expression: new frontiers, lawlessness, and misfits. *Ann Rev Microbiol* 2016;70:161–78.
  12. Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAMtools. *Bioinformatics* 2009;25:2078–9.
  13. Quinlan AR. BEDTools: the Swiss-Army Tool for genome feature analysis. *Curr Protoc Bioinformatics* 2014;47:11.12.1–34.
  14. Kim D, Pertea G, Trapnell C, et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 2013;14:R36.
  15. Rorbach J, Bobrowicz A, Pearce S, et al. Polyadenylation in bacteria and organelles. *Methods Mol Biol* 2014;1125:211–27.
  16. Small ID, Rackham O, Filipovska A. Organelle transcriptomes: products of a deconstructed genome. *Curr Opin Microbiol* 2013;16:652–8.
  17. Hayes R, Kudla J, Gruissem W. Degrading chloroplast mRNA: the role of polyadenylation. *Trends Biochem Sci* 1999;24:199–202.
  18. Hecht J, Grewe F, Knoop V. 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 Biol Evol* 2011;3:344–58.
  19. Oldenkott B, Yamaguchi K, Tsuji-Tsukinoki S, et al. Chloroplast RNA editing going extreme: more than 3400 events of C-to-U editing in the chloroplast transcriptome of the lycophyte *Selaginella launinata*. *RNA* 2014;20:1499–506.
  20. Hallick RB, Hong L, Drager RG, et al. Complete sequence of *Euglena gracilis* chloroplast DNA. *Nucleic Acids Res* 1993;21:3537–44.
  21. Knoop V. When you can't trust the DNA: RNA editing changes transcript sequences. *Cell Mol Life Sci* 2011;68:567–86.
  22. Kayal E, Bentlage B, Collins AG, et al. Evolution of linear mitochondrial genomes in medusozoan cnidarians. *Genome Biol Evol* 2012;4:1–12.
  23. Smith DR, Kayal E, Yanagihara AA, et al. First complete mitochondrial genome sequence from a box jellyfish reveals a highly fragmented linear architecture and insights into telomere evolution. *Genome Biol Evol* 2012;4:52–8.
  24. Kayal E, Bentlage B, Cartwright P, et al. Phylogenetic analysis of higher-level relationships within Hydroidolina (Cnidaria: Hydrozoa) using mitochondrial genome data and insight into their mitochondrial transcription. *PeerJ* 2015; 3:e1403.
  25. Mungpakdee S, Shinzato C, Takeuchi T, et al. Massive gene transfer and extensive RNA editing of a symbiotic dinoflagellate plastid genome. *Genome Biol Evol* 2014;6:1408–22.
  26. Matsumoto T, Ishikawa SA, Hashimoto T, et al. A deviant genetic code in the green alga-derived plastid in the dinoflagellate *Lepidodinium chlorophorum*. *Mol Phylogenet Evol* 2011;60:68–72.
  27. Glanz S, Kück U. Trans-splicing of organelle introns—a detour to continuous RNAs. *Bioessays* 2009;31:921–34.
  28. Dietrich A, Wallet C, Iqbal RK, et al. Organellar non-coding RNAs: emerging regulation mechanisms. *Biochimie* 2015;117:48–62.
  29. Woodson JD, Chory J. Coordination of gene expression between organellar and nuclear genomes. *Nat Rev Genet* 2008;9:383–95.
  30. Sanitá Lima M, Woods LC, Cartwright MW, et al. The (in)complete organelle genome: exploring the use and non-use of available technologies for characterizing mitochondrial and plastid chromosomes. *Mol Ecol Res* 2016, doi: 10.1111/1755-0998.12585.
  31. Mercer TR, Neph S, Dinger ME, et al. The human mitochondrial transcriptome. *Cell* 2011;146:645–58.
  32. Zhelyazkova P, Sharma CM, Förstner KU, et al. The primary transcriptome of barley chloroplasts: numerous noncoding RNAs and the dominating role of the plastid-encoded RNA polymerase. *Plant Cell* 2012;24:123–36.
  33. Smith DR. Making your GenBank entry count. *Front Genet* 2012;3:123.
  34. Feagin JE, Harrell MI, Lee JC, et al. The fragmented mitochondrial ribosomal RNAs of *Plasmodium falciparum*. *PLoS One* 2012;7:e38320.
  35. Alverson AJ, Wei X, Rice DW, et al. Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). *Mol Biol Evol* 2010;27:1436–48.
  36. Lohse M, Drechsel O, Kahlau S, et al. OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res* 2013;41:W575–81.
  37. Mower JP. PREP-Mt: predictive RNA editor for plant mitochondrial genes. *BMC Bioinformatics* 2005;6:96.
  38. Lenz H, Knoop V. PREPACT 2.0: predicting C-to-U and U-to-C RNA editing in organelle genome sequences with multiple references and curated RNA editing annotation. *Bioinform Biol Insights* 2013;7:1–19.
  39. Picardi E, Regina TM, Verbitskiy D, et al. REDIdb: an upgraded bioinformatics resource for organellar RNA editing sites. *Mitochondrion* 2011;11:360–5.
  40. Gray MW. Mitochondrial evolution. *Cold Spring Harb Perspect Biol* 2012;4:a011403.
  41. Keeling PJ. The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. *Ann Rev Plant Biol* 2013;64: 583–607.