Correspondence

The strange mitochondrial genomes of *Metschnikowia* yeasts

Dong Kyung Lee[1], Tom Hsiang[2], Marc-André Lachance[1], and David Roy Smith[1,*]

While sequencing and characterizing the mitochondrial genomes of 71 strains from the yeast genus *Metschnikowia* [1] (close cousin to the model species *Candida albicans*), we uncovered one of the most extreme examples of mitochondrial genome architectural diversity observed to date. These *Metschnikowia* mitochondrial DNAs (mtDNAs) capture nearly the entire known gene-size and intron-content range for *cox1* and *cob* across all eukaryotic life and show remarkable differences in structure and noncoding content. This genomic variation can be seen both among species and between strains of the same species, raising the question: why are *Metschnikowia* mitogenomes so malleable?

Fungal mitochondrial genomes are no strangers to introns [2,3], but our investigation of *Metschnikowia* mtDNAs takes organelle intron accumulation to new levels (Figure 1). The number of introns within *cox1* (encoding subunit 1 of the cytochrome c oxidase) varies from 0–20 (avg. = 12), resulting in an astounding 30-fold size range for this gene across the genus: 1,599–49,088 nt (avg. = 22,699 nt). Even in strains of the same species, *cox1* differed in size by up to 5 kb. A survey of the >1.3 million *cox1* genes in GenBank shows that 12 of the 20 longest sequences now belong to *Metschnikowia* species. At 49 kb, *cox1* from *Metschnikowia cubensis* is the second largest organelle gene ever sequenced, and more than twice the size of the entire mtDNA of *Metschnikowia caudata* (24.1 kb). These trends are mirrored in the *Metschnikowia cob* genes (encoding cytochrome b), which contain 0–13 introns (avg. = 7), have lengths of 1,152–23,127 nt (avg. = 12,780 nt), and include 7 of the 10 largest *cob* sequences on record (Figure 1). Other genes within the *Metschnikowia* mtDNAs also vary extensively in size and intron content (Table S1).

	Genome		*cox1*		*cob*	
	Structure	Size (kb)	Introns	Size (kb)	Introns	Size (kb)
M. santaceciliae - a	C	62.2	14	19.3	5	11.0
M. santaceciliae - b	C	81.1	15	23.5	11	14.3
M. cerradonensis - a	L	97.2	18	32.2	11	14.6
M. cerradonensis - b	L	92.3	18	32.6	11	14.6
M. continentalis - a	C	79.7	17	28.5	10	12.0
M. continentalis - b	L	79.1	16	25.7	11	15.1
M. sp. *mer*	C	85.9	16	26.5	11	14.0
M. v. *maris*	C	75.4	16	24.9	10	12.2
M. matae - a	C	85.8	18	30.4	10	12.1
M. matae - b	C	82.1	17	27.1	10	12.7
M. lochheadii - a	C	89.7	14	32.4	9	15.1
M. lochheadii - b	C	76.3	16	24.9	10	12.5
M. borealis - a	C	96.2	18	44.4	6	12.0
M. borealis - b	C	96.6	18	44.4	6	12.0
M. sp. *flo*	C	105.2	20	46.7	9	20.1
M. sp. *sma*	C	81.2	16	26.0	11	14.1
M. cubensis - a	C	116.3	17	49.1	8	19.1
M. cubensis - b	C	116.3	17	49.1	8	19.1
M. ipomoeae - c	C	77.9	17	29.1	11	12.9
M. ipomoeae - a	C	76.9	16	24.9	8	10.1
M. ipomoeae - b	C	67.6	15	22.0	10	12.0
M. mauinuiana - a	C	55.5	6	9.6	4	7.8
M. mauinuiana - b	C	56.1	7	14.3	3	6.0
M. hamakuensis - a	C	43.5	5	9.2	3	4.0
M. hamakuensis - b	C	44.4	5	9.9	3	3.9
M. hawaiiensis - a	C	55.0	8	16.3	5	9.4
M. hawaiiensis - b	C	51.8	7	13.1	5	9.4
M. kamakouana - a	C	52.5	8	15.8	5	9.9
M. kamakouana - b	C	54.9	9	15.8	6	12.4
M. sp. *pal*	C	93.2	10	22.0	7	11.8
M. sp. *ipil*	C	60.3	5	13.5	3	6.4
M. bowlesiae - a	C	93.0	16	25.3	10	16.7
M. bowlesiae - b	L	93.1	15	22.2	10	16.2
M. bowlesiae - c	C	89.8	15	22.3	10	16.7
M. lacustris - c	C	99.0	15	25.6	10	19.7
M. lacustris - a	C	82.2	11	20.4	10	16.7
M. lacustris - b	C	83.9	14	22.9	10	16.4
M. dekortorum - a	C	76.4	14	18.8	10	16.1
M. dekortorum - b	C	79.8	14	18.9	10	16.4
M. dekortorum - c	C	81.0	14	18.4	10	16.4
M. similis - a	C	99.2	17	28.9	10	16.9
M. similis - b	L	104.9	16	31.0	11	20.3
M. sp. *cos*	C	63.9	11	19.2	6	11.0
M. arizonensis - a	C	187.5	15	35.7	6	15.0
M. arizonensis - b	C	187.0	15	35.5	6	14.9
M. colocasiae - a	C	62.6	11	17.2	0	1.2
M. colocasiae - b	L	64.7	11	15.1	6	6.9
M. amazonensis - a	C	132.6	18	36.1	13	22.4
M. amazonensis - b	C	132.2	18	36.0	13	22.5
M. drakensberg - a	C	109.1	17	36.0	9	18.7
M. drakensberg - b	C	113.4	17	36.0	11	23.1
M. proteae - a	C	120.3	18	37.3	11	23.1
M. proteae - b	C	121.5	19	38.5	11	23.1
M. shivogae - a	C	77.3	9	15.3	9	21.1
M. shivogae - b	C	81.2	11	23.2	8	18.9
M. aberdeeniae - a	C	69.3	7	13.9	8	19.7
M. aberdeeniae - b	C	70.5	8	18.5	7	17.3
M. hibisci - a	C	67.3	9	17.4	7	16.3
M. hibisci - b	C	66.6	9	17.6	6	14.6
M. sp. *cla*	C	52.1	8	16.5	3	7.2
M. kipukae	C	33.5	5	5.8	1	2.7
M. hawaiiana	C	35.6	0	1.6	0	1.2
M. orientalis - a	C	56.7	11	17.6	6	13.6
M. orientalis - b	C	47.0	7	11.6	4	8.6
M. caudata - a	C	24.1	1	3.4	0	1.2
M. caudata - b	C	25.0	1	4.2	0	1.2
M. torresii	C	48.2	4	8.3	1	2.9
M. drosophilae - a	C	33.6	3	5.9	2	4.8
M. drosophilae - b	C	33.5	3	5.9	2	4.8
M. agaves - a	C	61.4	7	12.3	2	4.4
M. agaves - b	C	61.6	7	12.2	2	4.4

Figure 1. Mitochondrial genome architectural diversity in *Metschnikowia* yeasts.
Genome structure: C, circular; L, linear. Notable features are shown in red font and boxed in grey. For further details on the various species and strains and their mtDNAs see Table S1. Information on branching order is found in the Supplemental Information.

Both group I and group II introns can be found in the *Metschnikowia* mitogenomes (with the former in slightly higher proportions) and many of them encode their own proteins, typically involved in intron mobility [4] (Table S1). In fact, the *Metschnikowia cubensis* mtDNA harbours twice as many intron-encoded proteins as standard mitochondrial ones (28 versus 14), including 18 maturases or reverse transcriptases, and 10 endonucleases. Equally astonishing is the number of distinct intron insertion sites across the *Metschnikowia* mitogenomes: at least 50, of which 25 are for *cox1* alone, making it the single biggest reservoir of introns among all available mtDNAs. With so many introns, a lot of RNA gets left on the cutting room floor. In *Metschnikowia borealis*, *cox1* and *cob* contain 30 introns between them, equalling ~59 kb of RNA (48% of the genome length) that needs to be excised from the transcripts of these genes to generate functional proteins.

The *Metschnikowia* mtDNAs also exhibit large differences in intergenic content (11–59%; avg. = 24.8%), contributing to a nearly 8-fold fluctuation in overall size among these genomes (24.1–187.5 kb; avg. = 79 kb), despite them having near identical gene contents (introns notwithstanding). The relative amounts of intergenic and intronic mtDNA within *Metschnikowia* scale positively with one another (Table S1), but the largest two genomes do not have the largest intron contents and vice versa.

Another unexpected observation concerns the structure of the mitogenomes. Fungal mtDNAs are typically circular-mapping molecules [5,6]; thus, it is no surprise that most of the sequenced *Metschnikowia* mtDNAs have circular maps. However, 6 of the 71 genomes assembled as linear chromosomes with defined telomeres arranged as inverted repeats (length 204–668 nt), a common theme among linear mitogenomes [7]. Support for the linear conformation of these mtDNAs comes from both sequencing-read coverage analyses and PCR work (see the Supplemental Information). The phylogenetic position of strains with linear genomes suggests that the transition from a circular to a linear mtDNA occurred multiple times independently within *Metschnikowia*, perhaps as many as five times (Figure 1). What's more, distinct isolates of the same species can have different mitogenome conformations, whereby one strain has a linear mtDNA (with telomeres) and the other a circular one (Figure 1).

How has this startling diversity in mtDNA architecture come to be? Organelle introns, which are widely distributed across eukaryotic life [7], are mobile elements, known to spread within a genome and horizontally between organelle DNAs of different species, sometimes disparate ones [4]. Once seeded within a lineage, mitochondrial introns can undergo waves of gain and loss [7]. Fungal mtDNAs, for whatever reason, are particularly susceptible to this process [2,3,5,6] and this is especially true in *Metschnikowia*. But keep in mind that our mitochondrial data for *Metschnikowia* are based on a wide sampling of the genus and that future broad sampling efforts of other fungal groups may uncover similarly extraordinary examples of intron gain or loss. Also, our sequencing efforts concentrated on heterothallic, haploid *Metschnikowia* species, which presumably reproduce sexually (not the case for many other yeast species), a feature that could facilitate the transmission of unconventional mitochondrial genomes across species.

It is noteworthy that the six *Metschnikowia* strains with putatively linear mtDNAs have high mitochondrial intron contents (>19; avg. = 32). Although speculative, it is possible that the genetic upheaval of having so many introns somehow contributed to the shift from circular- to linear-mapping genomes. Indeed, the mobility of group I introns involves site-specific double-strand-DNA cuts, mediated via homing endonucleases [6]. If an intron-encoded homing endonuclease targeted an intergenic region (one that randomly acquired a targeting sequence, for instance) and the cut went unrepaired, it could easily result in a genome linearization event. For the linear *Metschnikowia* mtDNAs, the linearization points all occur within large intergenic regions.

Studies of organelle DNA have provided a number of classic examples of unconventional genomic architectures [7]. The mtDNAs of land plants, for example, underscore the complexities of intron trans-splicing in organelle systems [8], and the chloroplast genome of *Euglena gracilis* has proven that even introns can have introns [9]. The data presented here reveal the consequences of unbridled intron gain and loss within a genus, and ultimately highlight the severe extent to which organelle-genome architectures can differ, even among the most closely related of species.

SUPPLEMENTAL INFORMATION

Supplemental Information includes one figure, one table, experimental procedures, author contributions, acknowledgements, and supplemental references, and can be found with this article online at https://doi.org/10.1016/j.cub.2020.05.075.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Lachance, M.A., Hurtado, E., and Hsiang, T. (2016). A stable phylogeny of the large-spored *Metschnikowia* clade. Yeast *33*, 261–275.
2. Férandon, C., Moukha, S., Callac, P., Benedetto, J.P., Castroviejo, M., and Barroso, G. (2010). The *Agaricus bisporus cox1* gene: the longest mitochondrial gene and the largest reservoir of mitochondrial group I introns. PLoS One *5*, e14048.
3. Zubaer, A., Wai, A., and Hausner, G. (2018). The mitochondrial genome of *Endoconidiophora resinifera* is intron rich. Sci. Rep. *8*, 17591.
4. Lang, B.F., Laforest, M.J., and Burger, G. (2007). Mitochondrial introns: a critical view. Trends Genet. *23*, 119–125.
5. Hausner, G. (2003). Fungal mitochondrial genomes, plasmids and introns. In Applied Mycology and Biotechnology, D.K. Arora and G.G. Khachatourians, eds. (Amsterdam: Elsevier), pp. 101–131.
6. Dujon, B. (2020). Mitochondrial genetics revisited. Yeast *37*, 191–205.
7. Smith, D.R. and Keeling, P.J. (2015). Mitochondrial and plastid genome architecture: reoccurring themes, but significant differences at the extremes. Proc. Natl. Acad. Sci. USA *112*, 10177–10184.
8. Guo, W., Zhu, A., Fan, W., Adams, R.P., and Mower, J.P. (2020). Extensive shifts from *cis*- to *trans*-splicing of gymnosperm mitochondrial introns. Mol. Biol. Evol. *37*, 1615–1620.
9. Hallick, R.B., Hong, L., Drager, R.G., Favreau, M.R., Monfort, A., Orsat, B., Spielmann, A., and Stutz, E. (1993). Complete sequence of *Euglena gracilis* chloroplast DNA. Nucleic Acids Res. *21*, 3537–3544.

[1]Department of Biology, University of Western Ontario, London, ON N6A 1W8, Canada. [2]School of Environmental Sciences, University of Guelph, Guelph, ON N1G 2W1, Canada. *E-mail: dsmit242@uwo.ca