



Evaluation of the potential of CO1 as DNA barcode for Basidiomycota

Agathe Viale¹, Nicolas Feu², Richard C. Hamelin^{1,2}

¹Laval University, Sainte Foy, QC, Canada, ²Natural Resources Canada, Canadian Forest Service, Québec, QC, Canada

BACKGROUND

The DNA barcoding system employs a short, standardized gene region to identify species. Studies on various groups of animals have shown that a 650-pb fragment of the mitochondrial gene cytochrome c oxidase 1 (CO1) is generally effective as a barcode sequence, delivering a strong species-level resolution (Herbert *et al.*, 2003).

Basidiomycota is one of the major Phyla in the fungi kingdom. They are morphologically, ecologically, and taxonomically very diverse, almost equivalent to the Ascomycota in terms of number of known species (ca. 30,000). The potential effectiveness of CO1 in species identification of Basidiomycota has not yet been evaluated. According to the literature, the mitochondrial fungal genome varies greatly in size between species and strains because of the presence of many introns (Yan and Xu, 2005). This characteristic reveals a potential complication to PCR amplification of CO1 sequences, a key step in the analytical chain from specimens to the barcodes (Hajibabaei *et al.*, 2005).

The objective of this work was to evaluate *in silico* the potential of the CO1 gene as a taxonomic tool for the fungal group Basidiomycota with publicly available sequences. We used bioinformatics approach to address two potential shortcomings of CO1 in fungi: the presence of introns and the presence of multiple copies of the CO1 gene in the mitochondrial and nuclear genome.

MATERIALS AND METHODS

Recovery of mitochondrial genomes: We designed strategies to recover the entire mitochondrial genome of 10 fungi including different strains of one species (*Cryptococcus neoformans*) depending on the available data. For some fungi, the entire annotated mitochondrial genome was available in public databases. For the others, however, the fully annotated genome was not available. In this case, we used two approaches to retrieve the mitochondrial genome: i) *in silico* probes using known mitochondrial genes employing EST databases were used to search for contigs comprising the mitochondrial genes, or ii) we reconstructed models of mitochondrial genomes with the assembly of potential mitochondrial sequences available in Tracefile on NCBI. These different approaches are described in Figure 1.

Identification and mapping of CO1 exons and introns: To better document the position and length of introns in the mitochondrial genome of fungi, we aligned complete mitochondrial genomes recuperated as described above (Table 1) with the Vista software (Loots *et al.*, 2002). Moreover, we obtained one additional CO1 complete sequence for the homobasidiomycete *Agrocybe aegerita*. Then we recovered and aligned by hand the CO1 gene sequences in order to identify and map the introns and exons in the different taxa.

Research of homologous sequence with CO1 gene: For each fungus under study, the CO1 gene sequence (exons + introns) was used as *in silico* probe to search through the entire nuclear and mitochondrial genome sequences using the *blastn* algorithm. Only the results with more than 20nt and e-value <0.01 have been considered. The number and length of intron and exon nuclear and mitochondrial copies were recorded.

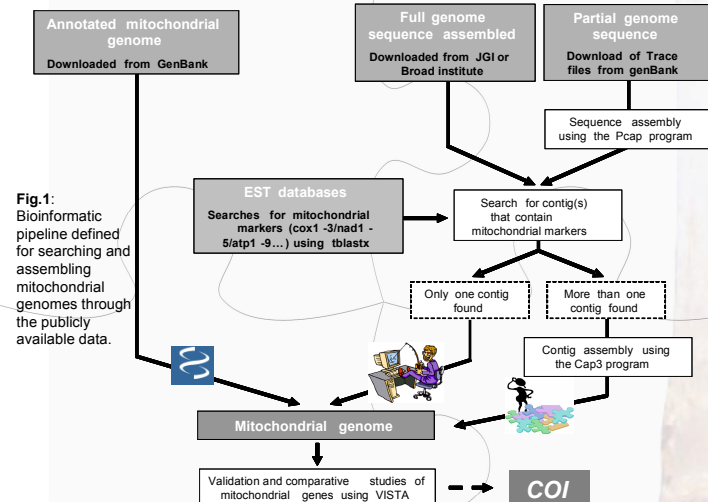


Fig. 1: Bioinformatic pipeline defined for searching and assembling mitochondrial genomes through the publicly available data.

Table 1: Mitochondrial genomes and sequences available for the Basidiomycota considered in this study.

Taxon	Description	Source
<i>Agrocybe aegerita</i>	cytochrome c oxidase subunit 1 gene	AF010257
<i>Coprinus cinereus</i>	Full genome sequence	Broad Institute
<i>Cryptococcus neoformans</i> H99	Full genome sequence	Broad Institute
<i>Cryptococcus neoformans</i> R265	Annotated mitochondrial genome	NC_004336
	Full genome sequence	Broad Institute
<i>Laccaria bicolor</i>	Full genome sequence (nuclear only)	Join Genome Institute (JGI)
	Annotated mitochondrial genome (F. Martin, personal communication)	
<i>Moniliophthora perniciosa</i>	Annotated mitochondrial genome	NC_005927
<i>Phanerochaete chrysosporium</i>	Full genome sequence (nuclear only)	
	Trace files	JGI
<i>Phakospora pachyrhizi</i>	Trace files	
<i>Puccinia graminis</i>	Full genome sequence (nuclear + mitochondrion)	Broad Institute
<i>Schizophyllum commune</i>	Annotated mitochondrial genome	NC_003049
	Full genome sequence	
<i>Ustilago maydis</i>	Annotated mitochondrial genome	NC_008368

RESULTS AND DISCUSSION

Characterization of CO1 gene in Basidiomycota fungi: We found and/or reconstituted 10 fungal mitochondrial genomes from public databases (Table 1). The comparative studies identified nine conserved protein-coding genes over these genomes, including CO1. **The vista alignment shows that the CO1 gene contains conserved regions that are suitable to design primers.**

The aligned CO1 for the 10 fungi studied revealed a mosaic of conserved exons interspersed with introns. Introns position and size varied across species and strains. **Exons 1&2 are promising for DNA barcoding as there were relatively few introns in that region.**

We retrieved 17 introns within the CO1 gene allowing us to limit the development of a barcode to a 232 bp region localised in the first exon using only one nucleotide primer pair (Figure 2). To enlarge this barcode to a 650 bp segment, we will design at least three different nucleotide primer pairs.

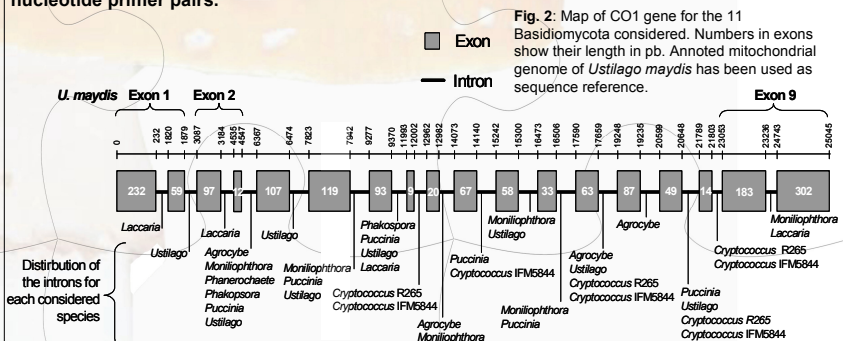


Fig. 2: Map of CO1 gene for the 11 Basidiomycota considered. Numbers in exons show their length in pb. Annotated mitochondrial genome of *Ustilago maydis* has been used as sequence reference.

CO1 copy numbers: No complete copy of the CO1 sequence (exons + introns) was found in the respective mitochondrial and nuclear genomes in these Basidiomycota. However, *blastn* searches against the mitochondrial and nuclear genome sequence available for these fungi identified 40 copies homologous to 20-200-bp CO1 segments among which two (in *Puccinia graminis* and *Cryptococcus neoformans*) were identical to the 232 bp mentioned previously (Table 2).

We observed a nuclear copy of a part of exon 1 in only one of the strains of *C. neoformans*. This intraspecific variability of CO1 copy number should be further investigated in other fungal species.

The presence of these homologous sequences must be considered during the design of the primers for the key PCR amplification step of the barcode protocol (Hajibabaei *et al.*, 2005).

Table 2: Copy of the CO1 gene complete sequence retrieved within nuclear and mitochondrial genome sequences. Numbers between brackets correspond to the length of the copy in bp.

	Nuclear		Mitochondrial	
	Exons	Introns	Exons	Introns
<i>C. cinereus</i>	1 (52)	-	-	-
<i>C. neoformans</i>	5 (38-200)	-	1 (20)	-
<i>M. perniciosa</i>	non déterminé	-	-	7 (21-30)
<i>P. pachyrhizi</i>	1 (157)	1 (1065)	-	1 (24)
<i>P. graminis</i>	3 (87-121)	5 (108-728)	-	3 (24)
<i>U. maydis</i>	-	4 (37-56)	-	8 (20-132)

ACKNOWLEDGEMENTS

This research was supported through funding to the Canadian Barcode of Life Research Network from Genome Canada through the Ontario Genomics Institute, NSERC, and other sponsors listed at www.BOLNET.org.

REFERENCES

- Hajibabaei M., deWaard J.R., Ivanova N.V., Ratnasingham S., Dooh R.T., Kirk S.L., Mackie P.M. and Herbert P.D.N., 2005. Critical factors for assembling a high volume of DNA barcodes. *Philos Trans R Soc B* 360:1959-1967.
- Herbert P.D.N., Cywinska A., Ball S.L. and deWaard J.R., 2003. Biological identification through DNA barcodes. *Proc. R. Soc. Lond. B* 270:313-321.
- Loots G.G., Ovcharenko I., Pachter L., Dubchak I. and Rubin E.M., 2002. rVista for comparative sequence-based discovery of functional transcription factor binding sites. *Gen. Res.* 12:832-839.
- Yan Z. and Xu J., 2005. Fungal mitochondrial inheritance and evolution. In *Evolutionary Genetics of Fungi*. Edited by Xu, J. Horizon Bioscience, Norfolk, UK, pp. 221-252.

