The largest reservoir of mitochondrial introns is a relic of an ancestral split gene

Cyril Férandon[1], Serge Moukha[2], Philippe Callac[3], Jean-Pierre Benedetto[1], Michel Castroviejo[1] & Gérard Barroso[1]†

1: UMR 5234 CNRS – Université Victor Segalen Bordeaux 2. Dept 1, 146 rue Léo Saignat, 33076 Bordeaux Cedex, France.

2: Laboratoire de Toxicologie et Hygiène Appliquée, UFR des Sciences Pharmaceutiques, Université Victor Segalen Bordeaux 2, 146 rue Léo Saignat, 33076 Bordeaux Cedex, France.

3: MycSA, Unité 1264 INRA Bordeaux-Aquitaine, BP 81, 33883 Villenave d'Ornon Cedex, France.

†To whom correspondence should be addressed. E-mail: gerard.barroso@u-bordeaux2.fr

1

In eukaryotes, introns are located in nuclear and organelle genes from several kingdoms (ref. 1-4) . Large introns (0.1 to 5 kbp) are frequent in mitochondrial genomes of plant and fungi (ref. 1,5) but scarce in Metazoa, despite these organisms are grouped with fungi among Opisthokonts. Introns are classified in two main groups (I and II) according to their RNA secondary structure involved in the intron self-splicing mechanism (ref. 5,6). Most of the group I introns carry a "Homing Endonuclease Gene" (ref. 7-9) encoding a DNA endonuclease acting in the transfer and site specific integration ("homing") and allowing the intron spreading and gain after lateral transfer even between species from different kingdoms (ref. 10,11). Opposite to this "late intron"paradigm, the "early intron" theory indicates that introns, which would have been abundant in the ancestral genes, would mainly evolve by loss (ref. 12,13).

Here we report the sequence of the *cox1* gene of the button mushroom *Agaricus bisporus*, the most worldwide cultivated mushroom. This gene is both the longest mitochondrial gene (29,902 nt) and the largest Group I intron reservoir reported to date. An analysis of the group I introns available in *cox1* genes shows that they are ancestral mobile genetic elements, whose frequent events of loss (according to the "late theory") and gain by lateral transfer ("early theory") must be combined to explain their wide and patchy distribution extending on several kingdoms. This allows the conciliation of the "early" and "late intron" paradigms, which are still matters of much debate (ref. 14,15). The overview of the intron distribution indicates that they evolve towards elimination. In such a landscape of eroded and lost intron sequences, the *A. bisporus* largest intron reservoir, by its singular dynamics of intron keeping and catching, constitutes the most fitted relic of an early split gene.

The complete sequence (29.902 nt, GenBank Accession number: EU314927) of the *cox1* gene of the button mushroom *Agaricus bisporus (Abi)*, was determined for the brown cultivar C9 (Supplementary Fig. 1). An alignment of this gene with the *cox1* exon sequence (CDS) of the related Agaricales *Agrocybe aegerita*[10] shows that the *Abi cox1* CDS is split by 19 large (>1kbp) introns. The borders of these introns, their group and sub-group were deduced from the design of each secondary structure (Supplementary information Fig. 2**)** and from a comparison of the gene with the *Abi* cDNA (1587 nt). The 19 introns are scattered on the whole gene sequence (Fig. 1). The size of the exons varies from 3 nt (exon 2) to 288 nt (exon 20), contrasting with the large size of the introns ranging from 1057 nt (*iAbi*13) to 2736 nt (*iAbi*2). The intron sequences represent 28,318 nt, i.e. 94.7 % of the gene. The *Abi cox1* gene is the longest mitochondrial gene reported to date in all kingdoms.

Only one (*iAbi2)* of the 19 introns belongs to the group II. This large intron (2736 nt) contains an eroded ORF (402 aa) whose the putative encoded protein possess low homologies with reverse transcriptases of other plant and fungal mitochondrial group II introns. In *Abi,* this intron appears as a remnant of an ancestral intron, present in both fungal and plant kingdoms, and in elimination process.

Sixteen out of the 18 group I introns harbour intact ORF encoding a putative functional HE, while only three (*iAbi6, iAbi11* and *iAbi14)* contain eroded *heg* leading to non functional HE (Fig. 1). The *heg* encoded by the *iAbi11* intron exhibits a singular organisation, resulting from the insertion in a *cox1 heg* of an other invading *heg* (Fig. 1) orthologous to a *heg* described in an intron of the mitochondrial *cob* gene of two Ascomycota species *Gibberella zeae* and *Podospora anserina*. In these two species, the invading and the recipient *heg* are located in introns of the *cob* and *cox1* genes, respectively. The invasion of the *iAbi11* intron *heg* by an *heg* issued from an other gene gives evidence of a recent transfer activity of

the enzyme encoded by the mobile invading element, as previously reported in the T4 phage or in the fungus *Podospora curvicolla*[16,17].

To understand the origin and dynamics of the *Abi* group I introns, a compilation of all the complete *cox1* genes in databases was carried out (Fig. 2a). Group I introns were reported in three kingdoms: Viridiplantae, Amoebozoa and Fungi/Metazoa. In the Fungi/Metazoa, mitochondrial group I introns have been described in a single species of the Choanoflagellida order (*Monosiga brevicollis*), in one species of the Porifera phylum (Metazoa kingdom) and in numerous species of the fungal kingdom. In contrast, all the *cox1* genes from the Eumetazoa lack group I intron, except two different intron types described in Cnidaria[18-20].

214 group I introns were characterized in 51 fungal *cox1* genes (including three species of the Zygomycota phylum, now fungi incertae sedis). They have been sorted in 34 classes first defined by the same location in the *cox1* CDS sequence. This distribution, based on the precise insertion site of each intron, is correlated to the highest sequence identities at the nucleotide level as well as amino acid identities of the putative encoded HE. Hence, each position class is also a sequence homology class. The 34 classes were named by increasing alphabetic characters from the 5'-end of the *cox 1* CDS. Since all the 34 classes are represented in at least one of the five fungal longest *cox1* genes, an alignment of the corresponding COXI proteins is used in Fig. 3 to represent their relative positions. The five longest *cox1* genes are found in the Basidiomycota *Abi* (18 group I introns), in the Ascomycota *Podospora anserina* (*Pan* : 14 introns) and *Gibberella zeae* (*Gze* : 12 introns), in the Chytridiomycota *Rhyzophidium sp.* (*Rsp* : 14 introns) and in the Blastocladiomycota *Allomyces macrogynus* (*Ama* : 11 introns).

The number of fungal *cox1* introns assigned to each class greatly varied from 23 in the class J to a unique representative in 11 different classes (Fig. 2a). Hence, these introns can be classified in rare and widely distributed introns. Rare introns are those detected either in a

single species (11 classes) or in two closely related species for the classes D and AE both containing introns reported in two Blastocladiales. The introns of the remaining 21 classes show a wide and patchy distribution. Indeed, these classes contain several representatives, from 2 in the class T to 23 in the class J, present in different phyla of the fungal kingdom when, at the same time, phylogenetically close species differ by the presence/absence of these introns. For instance, the two representatives of the class T have been described in the Chytridiomycota *Rsp* and in the distant Ascomycota *Gze* (Fig. 2a). The 23 representatives of the largest class J are present in one species (*Rsp*) among six Chytridiomycota, in one among two Blastocladiella, in one among three species of the basal fungal lineages, in six among nine Basidiomycota and in 14 among 31 Ascomycota. *Abi* contains 15 introns belonging to the 21 widely distributed classes, while *Schizophyllum commune* (*Sco*) belonging to the same order Agaricales is intronless.

Thus, fungal introns can be considered as mobile genetic element, enough frequently lost and gained to explain their wide and patchy distribution.

Among the 34 fungal intron classes, 17 contained orthologous introns defined by the same location in the *cox1* CDS and sequence homologies, in the Viridiplantae, Amoebozoa or Metazoa kingdoms. They represent 50 % (17/34) of the classes and up to 71.4 % (15/21) of the widely distributed ones. Excepted two group I introns specifically harboured by corals species (classes Co1 and Co2), all the *cox1* group I introns belong to classes previously defined by representatives from the fungal kingdom.

In Fig. 2a, species were ordered according to their systematic position indicated in the taxonomic and phylogenetic trees of Fig. 2b and 2c. The species phylogenetic relationships were deduced from the six-gene phylogeny established by James et al.[21]. The occurrence of the orthologous introns of the largestly distributed class J was reported in the trees (Fig. 2b

5

and 2c). This clearly shows that the patchy distribution of the species with and without orthologous introns requires a combination of multiple events of intron loss and gain.

A great sequence divergence appears between *heg* sequences of orthologous introns of the same class. The HE encoded by the introns of class J show percentages of aa identity (and similarity) ranging from 31% aa id. (54 % aa sim.) between the fungal Ascomycota *Gze* and the Viridiplantae *Mpo* to 56 % aa id (73 % aa sim.) between the Basidiomycota *Abi* and *Aae*. These high sequence divergences between orthologous introns of the same class, distributed over up to four kingdoms, but also observed between closely related species, strongly argue for an ancestral origin and high mobility of these introns.

Moreover, most of the classes of widely distributed introns (Fig. 2) contain introns with a potentially functional *heg* and also introns with an eroded *heg*. The classes X and AF are exceptions containing respectively 7 and 4 representatives with only intact *heg*. The erosion of the *heg* has been reported to be a preliminary step before the complete elimination of the intron[12]. For each class of intron, the presence of intact and eroded *heg,* even in closely related species, strengthens the hypothesis that numerous events of loss and gain occurred during evolution.

Most *cox1* genes of the Viridiplantae, Amoebozoa and Metazoa kingdoms lack intron and only a little number of species harbours 1 to 6 group I introns. But, the majority (43) of the 51 *cox1* sequences of the fungal kingdom possesses one to 18 introns. Only 8 genes do not contain intron.

In order to analyse the dynamics (loss and gain) of group I introns in eukaryots, a curve (Fig. 4) representing the number of species according to the number of introns of their *cox1* gene was established. This curve does not fit with a Gaussian distribution, as expected if the number of introns was reflecting random events of losses and gains. The plots can be represented by a theoretical curve showing a logarithmic decrease and thus, revealing a trend

of group I introns towards elimination from the *cox1* genes. However, the five long genes, distributed over four different fungal divisions, behave as Group I intron reservoirs, with 11 to 18 introns. Their compilation contains at least a representative of each class, except for the introns reported in corals. As the nine rare introns detected once were reported in four of these five long genes, the ability of bearing specific or rare introns appears to be a characteristic of these reservoir species.

Moreover, most of the introns described in the long *cox1* genes from the basal fungal lineages carry eroded *heg*: 10 introns among the 11 of *Ama* and 7 among the 14 of *Rsp.* This suggests that in both species these introns are in elimination process. Conceptually, the erosion of the *heg* should definitively prevent the spreading of the harboring intron. On the contrary, most of the introns of the long *cox1* genes of Dikarya (16/18, 12/14 and 12/12 for *Abi, Pan* and *Gze*, respectively) maintain putative functional *heg*.

In this context, the *Abi cox1* gene has to be considered as the most fitted relic of an ancestral *cox1* gene described to date. Indeed, this long gene possesses representatives of more than half (18/34 = 53 %) of the classes, and more than 69 % (16/23) of the widely distributed ones and all but two of its 18 group I introns possess intact and potentially functional *heg*.

In conclusion, it appears that the evolution of the mitochondrial genome of eukaryotes is still on the road to the ancestral group I intron elimination. Finally, the long genes described in the Dikarya fungi, whose the *Abi cox1* gene is the most amazing representative to date, ask an appealing and still unresolved question about the reasons and underlying mechanisms for which some fungal species maintain such an expensive organization of mitochondrial genes.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1   Haugen, P., Simon, D.M., & Bhattacharya, D., The natural history of group I introns. *Trends Genet* 21 (2), 111-119 (2005).

2   Lambowitz, A.M. & Belfort, M., Introns as mobile genetic elements. *Annu Rev Biochem* 62, 587-622 (1993).

3   Nesbo, C.L. & Doolittle, W.F., Active self-splicing group I introns in 23S rRNA genes of hyperthermophilic bacteria, derived from introns in eukaryotic organelles. *Proc Natl Acad Sci U S A* 100 (19), 10806-10811 (2003).

4   Sandegren, L. & Sjoberg, B.M., Distribution, sequence homology, and homing of group I introns among T-even-like bacteriophages: evidence for recent transfer of old introns. *J Biol Chem* 279 (21), 22218-22227 (2004).

5   Lang, B.F., Laforest, M.J., & Burger, G., Mitochondrial introns: a critical view. *Trends Genet* 23 (3), 119-125 (2007).

6   Saldanha, R., Mohr, G., Belfort, M., & Lambowitz, A.M., Group I and group II introns. *FASEB J* 7 (1), 15-24 (1993).

7   Lambowitz, A.M. & Perlman, P.S., Involvement of aminoacyl-tRNA synthetases and other proteins in group I and group II intron splicing. *Trends Biochem Sci* 15 (11), 440-444 (1990).

8   Lazowska, J., Jacq, C., & Slonimski, P.P., Sequence of introns and flanking exons in wild-type and box3 mutants of cytochrome b reveals an interlaced splicing protein coded by an intron. *Cell* 22 (2 Pt 2), 333-348 (1980).

9   Pellenz, S., Harington, A., Dujon, B., Wolf, K., & Schafer, B., Characterization of the I-Spom I endonuclease from fission yeast: insights into the evolution of a group I intron-encoded homing endonuclease. *J Mol Evol* 55 (3), 302-313 (2002).

10   Gonzalez, P., Barroso, G., & Labarere, J., Molecular analysis of the split cox1 gene from the Basidiomycota Agrocybe aegerita: relationship of its introns with homologous Ascomycota introns and divergence levels from common ancestral copies. *Gene* 220 (1-2), 45-53 (1998).

11   Vaughn, J.C., Mason, M.T., Sper-Whitis, G.L., Kuhlman, P., & Palmer, J.D., Fungal origin by horizontal transfer of a plant mitochondrial group I intron in the chimeric CoxI gene of Peperomia. *J Mol Evol* 41 (5), 563-572 (1995).

12   Goddard, M.R. & Burt, A., Recurrent invasion and extinction of a selfish gene. *Proc Natl Acad Sci U S A* 96 (24), 13880-13885 (1999).

13   Gonzalez, P., Barroso, G., & Labarere, J., Molecular gene organisation and secondary structure of the mitochondrial large subunit ribosomal RNA from the cultivated Basidiomycota Agrocybe aegerita: a 13 kb gene possessing six unusual nucleotide extensions and eight introns. *Nucleic Acids Res* 27 (7), 1754-1761 (1999).

14   Cusimano, N., Zhang, L.B., & Renner, S.S., Reevaluation of the cox1 group I intron in Araceae and angiosperms indicates a history dominated by loss rather than horizontal transfer. *Mol Biol Evol* 25 (2), 265-276 (2008).

15   Sanchez-Puerta, M.V., Cho, Y., Mower, J.P., Alverson, A.J., & Palmer, J.D., Frequent, phylogenetically local horizontal transfer of the cox1 group I Intron in flowering plant mitochondria. *Mol Biol Evol* 25 (8), 1762-1777 (2008).

16   Loizos, N., Tillier, E.R., & Belfort, M., Evolution of mobile group I introns: recognition of intron sequences by an intron-encoded endonuclease. *Proc Natl Acad Sci U S A* 91 (25), 11983-11987 (1994).

17   Saguez, C., Lecellier, G., & Koll, F., Intronic GIY-YIG endonuclease gene in the mitochondrial genome of Podospora curvicolla: evidence for mobility. *Nucleic Acids Res* 28 (6), 1299-1306 (2000).

18   Fukami, H., Chen, C.A., Chiou, C.Y., & Knowlton, N., Novel group I introns encoding a putative homing endonuclease in the mitochondrial cox1 gene of Scleractinian corals. *J Mol Evol* 64 (5), 591-600 (2007).

19   Goddard, M.R., Leigh, J., Roger, A.J., & Pemberton, A.J., Invasion and persistence of a selfish gene in the Cnidaria. *PLoS ONE* 1, e3 (2006).

20   Rot, C., Goldfarb, I., Ilan, M., & Huchon, D., Putative cross-kingdom horizontal gene transfer in sponge (Porifera) mitochondria. *BMC Evol Biol* 6, 71 (2006).

21   James, T.Y. *et al.*, Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature* 443 (7113), 818-822 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Authors contributions (optional)

Author information

Correspondence and requests should be addressed to G.B. (gerard.barroso@u-bordeaux2.fr)

Figure 1 Molecular organisation of the *Abi cox1* gene. Blue boxes represent exons, their nucleotide (nt) sizes are indicated below the exon E1 to E20. The name and nt size of the introns are indicated in the empty boxes. The *heg* of *cox1* group I introns are shown by green boxes, the eroded reverse transcriptase gene carried by the *iAbi 2* group II intron by an orange box. The invading *heg* with a *cob* origin harboured by the *iAbi11* intron is shown by a yellow box.

Figure 2 Schematic representation of the distribution of group I introns according to their reference class (line) and to the species harboring them (column). **a**, all introns of the same class are in the same line; all introns carried by a given species are in the same column. The 36 lines of the group I classes have been organized according to the decreasing number of representatives they contains, from the most distributed class J (23 members) to the 13 classes of "rare" introns containing either a single member or two (ID and IB2 classes) to four members (classes CO1 and CO2) shared by closely related species. Species were organized according to their taxonomic (superkingdoms and kingdoms, phyla, orders) and their phylogenetic positions, as indicated in **b** and **c**. The phylogenetic position of each fungal species was established according to the phylogenetic tree reported by James et al.[21] for the fungal kingdom using a six-gene phylogeny. When not present in the tree, species were located according to the nearest taxon (same genus or order) found in this tree. In **a**, Introns are coloured according to the kingdoms or phylla of the species that harbour them: green for Viridiplantae, purple for Amoebozoa, brown for Metazoa, pink for the fungi/metazoa incertae sedis, yellow for basal fungal lineages composed of fungi incertae sedis, Chytridiomycota and Blastocladiomycota, pale and dark orange for Basidiomycota and Ascomycota, respectively. The longest mitochondrial gene of *Abi* is in red. Presence of intact and eroded *heg* are shown by + and -, respectively. The representation of the ancestral gene shown on the right part of

10

the **a** was obtained by compiling an intron of each class which possesses an intact *heg* and belongs to one of the five longest *cox1* genes: from *Abi* (18 classes: 16 intact and 2 eroded *heg*), *Rsp* (6 classes: 4 eroded and 2 intact *heg*), *Gze* (5 classes with intact *heg*), *Ama* (4 classes with eroded *heg*) and *Pan* (1 class with intact *heg*). Aae : *Agrocybe aegerita* AF010257, Abi : *Agaricus bisporus* EU314927, Ago : *Ashbya gossypii* NC_005789, Ama : *Allomyces macrogynus* NC_001715, Apar : *Amoebidium parasiticum* AAN04062, Atha : *Arabidopsis thaliana* NP_085587, Atu : *Aspergillus tubingensis* NC_007597, Bba : *Beauveria bassiana* YP_001876504, Bem : *Blastocladiella emersonii* YP_002274319, Btau : *Bos taurus* ABV70623, Bvu : *Beta vulgaris* NP_064063, Cat : *Chlorokybus atmophyticus* YP_001315139, Cba : *Cordyceps bassiana* ABU50156, Cbr : *Cordyceps brongniartii* YP_002213602, Cgl : *Candida glabrata* NC_004691, Cglo : *Chaetosphaeridium globosum* NP_689386, Cnee : *Candida neerlandica* YP_002122387, Cnevne : *Cryptococcus neoformans var neoformans* AY560609, Cnevgr : *Cryptococcus neoformans var grubii* NC_004336, Cor : *Candida orthopsilosis* DQ026513, Cpa : *Candida parapsilosis* NC_005253, Cvu : *Chara vulgaris* NP_943703, Cze : *Candida zemplinina* NC_005972, Dci : *Dictyostelium citrinum* NC_007787, Dha : *Debaryomyces hansenii* NC_010166, Dmel : *Drosophila melanogaster* AAB59239, Drer : *Danio rerio* NP_059333, Efl : *Epidermophyton floccosum* NC_007394, Eni : *Emericella nidulans* X00790, Ggal : *Gallus gallus* BAC57577, Gze : *Gibberella zeae* NC_009493, Hcur : *Hyaloraphidium curvatum* NP_150103, Hje : *Hypocrea jecorina* NC_003388, Hsap : *Homo sapiens* ACA22152, Hsp94 : *Harpochytrium sp.* JEL94 NC_004760, Hsp105 : *Harpochytrium sp.* JEL105 NC_004623, Hur : *Hanseniaspora urvum* NC_007780, Kla : *Kluyveromyces lactis* NC_006077, Kth : *Kluyveromyces thermotholerans* NC_006626, Mbr : *Monosiga brevicollis* NP_696984, Mgr : *Mycosphaerella graminicola* NC_010222, Mpe : *Moniliophthora perniciosa* NC_005927, Mpo : *Marchantia polymorpha* NC_001660, Mse : *Metridium senile* NP_009253, Msp :

Monoblepharella sp. NC_004624, Mve : *Mortierella verticillata* NC_006838, Ncr : *Neurospora crassa* X14669, Oel : *Ochrosia elliptica* ABY83864, Osat : *Oryza sativa* BAD38494, Pan : *Podospora anserina* NC_001329, Pang : *Plakortis angulospiculatus* YP_001648679, Patr : *Plantago atrata* ABY83853, Pbr : *Paracoccioides brasiliensis* NC_007935, Pca : *Pichia canadensis* NC_001762, Pjvk : *Palythoa sp.* JVK-2006 ABF67639, Pma : *Penicillium marneffei* NC_005256, Pos : *Pleurotus ostreatus* NC_009905, Pwi :*Prototheca wickeramii* NP_042244, Rflo : *Ricordea florida* YP_654303, Ror : *Rhizopus oryzae* NC_006836, Rsp : *Rhizophydium sp.* 136 NC_003053, Rspca : *Rhodactis sp.* CASIZ 171755 YP_654290, Sca : *Saccharomyces castellii* NC_003920, Sce : *Saccharomyces cerevisiae* NC_001224, Sco : *Schizophyllum commune* NC_003049, Scu : *Smittium culisetae* NC_006837, Sja : *Schizosaccharomyces japonicum* NC_004332, Soc : *Schizosaccharomyces octosporus* NC_004312, Spo : *Schizosaccharomyces pombe* NC_001326, Spu : *Spizellomyces punctatus* NC_003052, Srad : *Siderastrea radians* YP_654418, Sse : *Saccharomyces servazii* NC_004918, Tin : *Tilletia indica* NC_009880, Uma : *Ustilago maydis* NC_008368, Yli : *Yarrowia lipolytica* NC_002659.
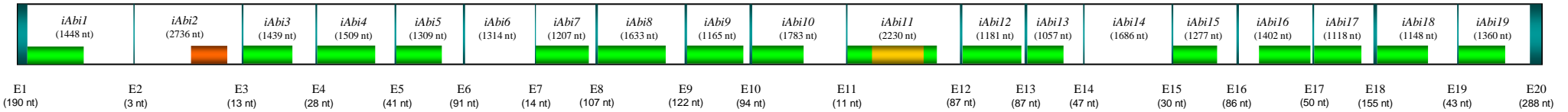
In Fig. 2a, a : number of introns per species; b: number of introns per class.

In Fig. 2b and 2c, the species harbouring a class J orthologous intron with an intact or an eroded *heg* are in red and green, respectively.
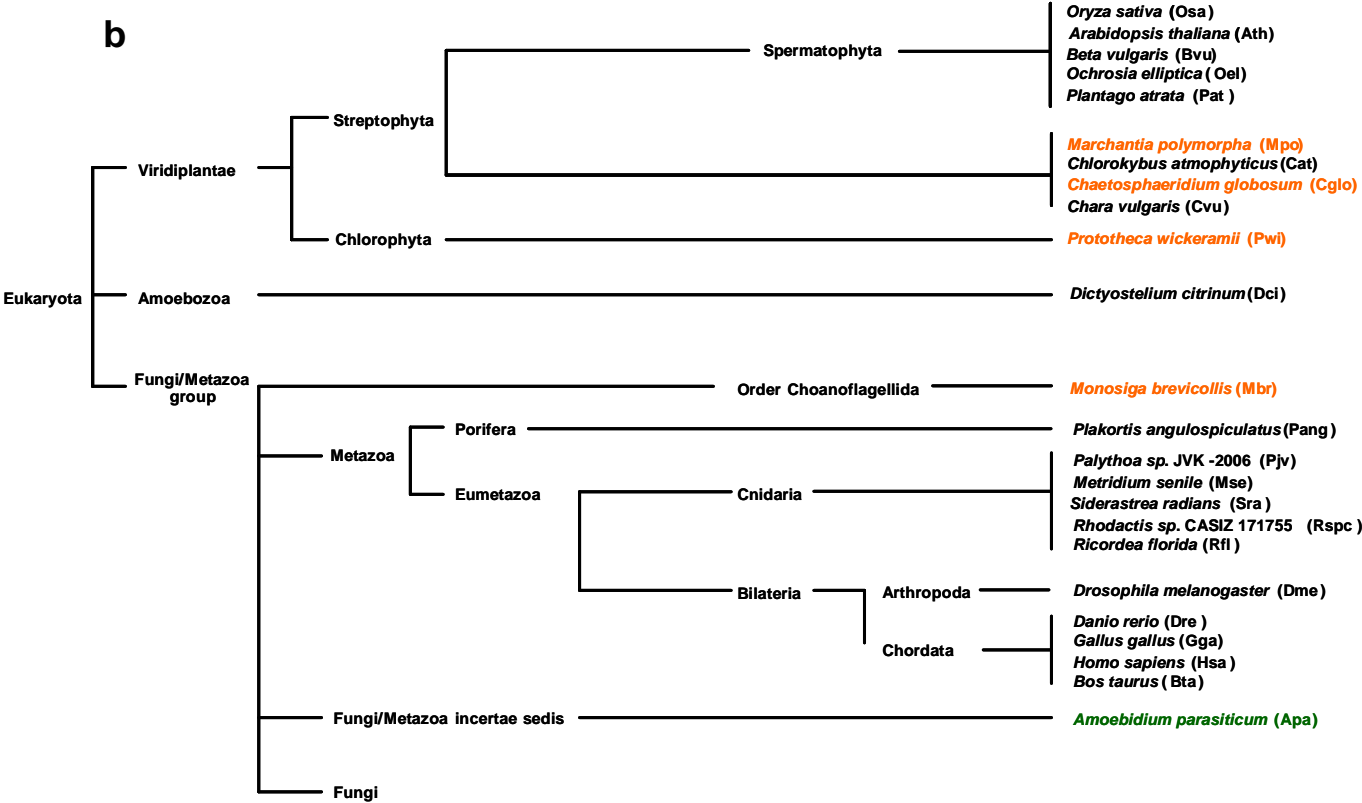
Figure 3 Comparison of the group I intron insertion sites in the amino acids alignment (Clustal W) of the COX1 protein encoded by the five longest *cox1* genes of fungi. Introns are designated by their class (From A to AH) and by their number in each fungal gene starting from the 5' end of each gene. Closed and open symbols (triangles) indicate introns with intact

12

and eroded *heg,* respectively. The insertion sites of the coral introns never reported to date in the fungal kingdom are indicated by a star.
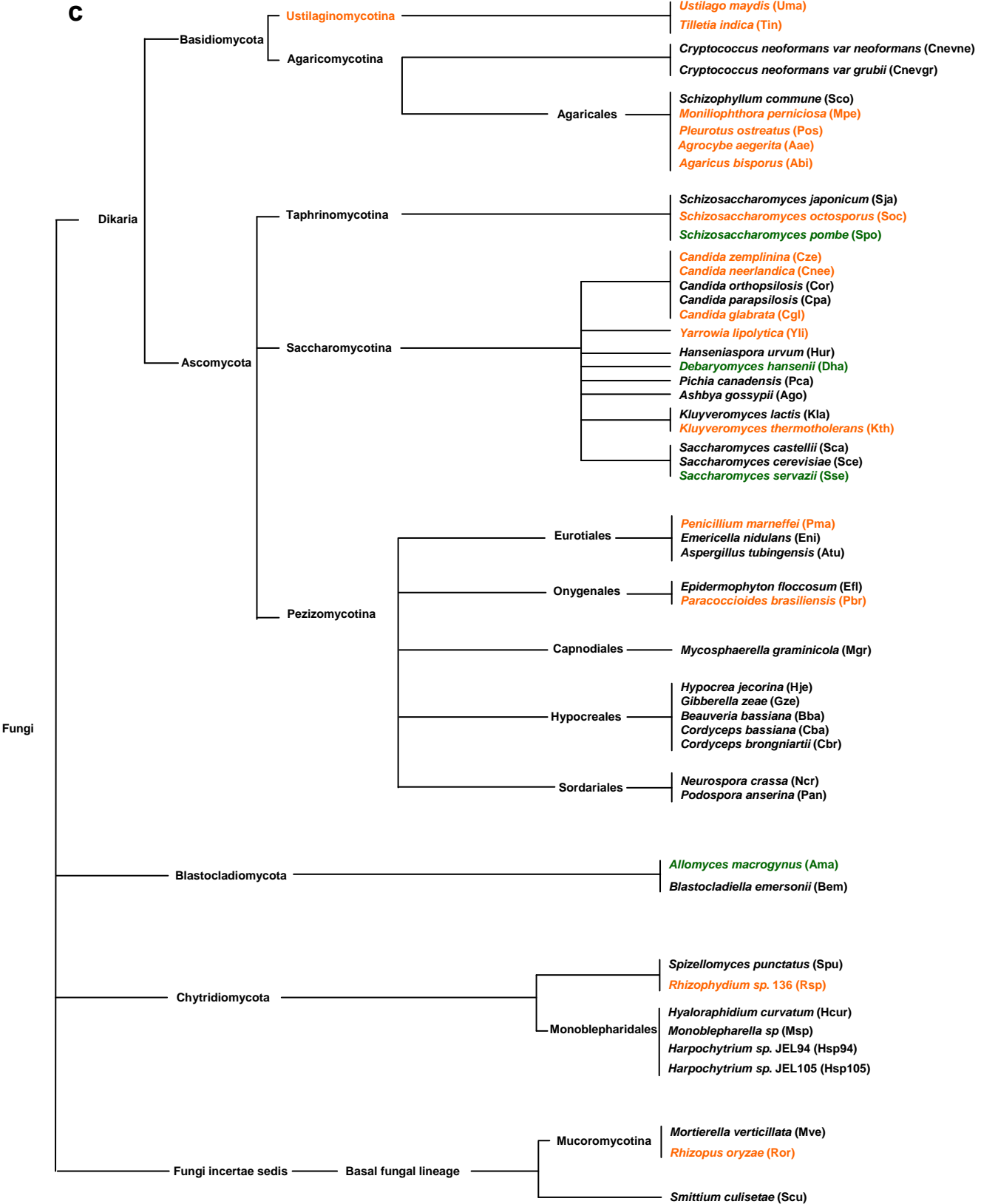
Figure 4 Graphical representation of the number of eukaryotic species (purple curve) and fungal species (blue curve) as a function of the number of introns carried by their *cox1* gene. The red curve shows the logarithmic regression model (R^2= 0.83) with all the analysed eukarotic species.

| iAbi1 (1448 nt) | iAbi2 (2736 nt) | iAbi3 (1439 nt) | iAbi4 (1509 nt) | iAbi5 (1309 nt) | iAbi6 (1314 nt) | iAbi7 (1207 nt) | iAbi8 (1633 nt) | iAbi9 (1165 nt) | iAbi10 (1783 nt) | iAbi11 (2230 nt) | iAbi12 (1181 nt) | iAbi13 (1057 nt) | iAbi14 (1686 nt) | iAbi15 (1277 nt) | iAbi16 (1402 nt) | iAbi17 (1118 nt) | iAbi18 (1148 nt) | iAbi19 (1360 nt) |

E1 (190 nt)　　E2 (3 nt)　　E3 (13 nt)　　E4 (28 nt)　　E5 (41 nt)　　E6 (91 nt)　　E7 (14 nt)　　E8 (107 nt)　　E9 (122 nt)　　E10 (94 nt)　　E11 (11 nt)　　E12 (87 nt)　　E13 (87 nt)　　E14 (47 nt)　　E15 (30 nt)　　E16 (86 nt)　　E17 (50 nt)　　E18 (155 nt)　　E19 (43 nt)　　E20 (288 nt)

b

Oryza sativa (Osa)
Arabidopsis thaliana (Ath)
Beta vulgaris (Bvu)
Ochrosia elliptica (Oel)
Plantago atrata (Pat)

Marchantia polymorpha (Mpo)
Chlorokybus atmophyticus (Cat)
Chaetosphaeridium globosum (Cglo)
Chara vulgaris (Cvu)

Prototheca wickeramii (Pwi)

Dictyostelium citrinum (Dci)

Monosiga brevicollis (Mbr)

Plakortis angulospiculatus (Pang)

Palythoa sp. JVK -2006 (Pjv)
Metridium senile (Mse)
Siderastrea radians (Sra)
Rhodactis sp. CASIZ 171755 (Rspc)
Ricordea florida (Rfl)

Drosophila melanogaster (Dme)

Danio rerio (Dre)
Gallus gallus (Gga)
Homo sapiens (Hsa)
Bos taurus (Bta)

Amoebidium parasiticum (Apa)

C

Fungi

- Dikaria
  - Basidiomycota
    - Ustilaginomycotina
      - *Ustilago maydis* (Uma)
      - *Tilletia indica* (Tin)
    - Agaricomycotina
      - *Cryptococcus neoformans var neoformans* (Cnevne)
      - *Cryptococcus neoformans var grubii* (Cnevgr)
      - Agaricales
        - *Schizophyllum commune* (Sco)
        - *Moniliophthora perniciosa* (Mpe)
        - *Pleurotus ostreatus* (Pos)
        - *Agrocybe aegerita* (Aae)
        - *Agaricus bisporus* (Abi)
  - Ascomycota
    - Taphrinomycotina
      - *Schizosaccharomyces japonicum* (Sja)
      - *Schizosaccharomyces octosporus* (Soc)
      - *Schizosaccharomyces pombe* (Spo)
    - Saccharomycotina
      - *Candida zemplinina* (Cze)
      - *Candida neerlandica* (Cnee)
      - *Candida orthopsilosis* (Cor)
      - *Candida parapsilosis* (Cpa)
      - *Candida glabrata* (Cgl)
      - *Yarrowia lipolytica* (Yli)
      - *Hanseniaspora urvum* (Hur)
      - *Debaryomyces hansenii* (Dha)
      - *Pichia canadensis* (Pca)
      - *Ashbya gossypii* (Ago)
      - *Kluyveromyces lactis* (Kla)
      - *Kluyveromyces thermotholerans* (Kth)
      - *Saccharomyces castellii* (Sca)
      - *Saccharomyces cerevisiae* (Sce)
      - *Saccharomyces servazii* (Sse)
    - Pezizomycotina
      - Eurotiales
        - *Penicillium marneffei* (Pma)
        - *Emericella nidulans* (Eni)
        - *Aspergillus tubingensis* (Atu)
      - Onygenales
        - *Epidermophyton floccosum* (Efl)
        - *Paracoccioides brasiliensis* (Pbr)
      - Capnodiales
        - *Mycosphaerella graminicola* (Mgr)
      - Hypocreales
        - *Hypocrea jecorina* (Hje)
        - *Gibberella zeae* (Gze)
        - *Beauveria bassiana* (Bba)
        - *Cordyceps bassiana* (Cba)
        - *Cordyceps brongniartii* (Cbr)
      - Sordariales
        - *Neurospora crassa* (Ncr)
        - *Podospora anserina* (Pan)
- Blastocladiomycota
  - *Allomyces macrogynus* (Ama)
  - *Blastocladiella emersonii* (Bem)
- Chytridiomycota
  - *Spizellomyces punctatus* (Spu)
  - *Rhizophydium sp.* 136 (Rsp)
  - Monoblepharidales
    - *Hyaloraphidium curvatum* (Hcur)
    - *Monoblepharella sp* (Msp)
    - *Harpochytrium sp.* JEL94 (Hsp94)
    - *Harpochytrium sp.* JEL105 (Hsp105)
- Fungi incertae sedis
  - Basal fungal lineage
    - Mucoromycotina
      - *Mortierella verticillata* (Mve)
      - *Rhizopus oryzae* (Ror)
    - *Smittium culisetae* (Scu)

Block 1 — markers (with insertion points): A (i1) C (i3) F (i4) H (i5) I (i6) J (i7) K (i8) M (i9)

```
Abi  MNWLNST NAKEIGTLYL IFSVFAGMVG TAFSVLIRLE LSSPGVQFLQ GDHQLFNVII SAHAFIMIFF MVMPGLLGGF GNYFLPIHCG SPDMAFPRLN NISFWLLPPS LTLLLMSSLV ESGAGTGWTV YPPLSGIQSH SGGSVDLAIF SLHLAGISSL LGAINFISTT LNMRTNGMSL HNLPLFVWAI FVTAVLLLLS LPVLAGAITM LLTDRNFNTS FFDPAGGGDP

Gze  MGMERWFNST NAKDIGTLYL IFALFSGLLG TAFSVLIRLE LSGPGVQYIS -NNQLYNSVI TAHAILMIFF MVMPALIGGF GNFLMPLMVG GPDMAFPRLN NISFWLLPPS LLLLIFSACI EGGVGTGWTL YPPLSGLQSH SGPSVDLAIF TLHLTGVSSL LGSINFITTI VNMRTPGIRL HKLALFGWAV VITAVLLLLS LPVLAGAITM VLTDRNFNTS FFEVAGGGDP

Pan  LWIERWMLST NAKDIGNLYL IFALFSGLLG TAFSVLIRME LSGPSVQYIA GDNQLYNSII TAHALLMIFF MVMPALIGGF GNFLMPLMVG GPDMAFPRLN NISFWLLPPS LILLVFSACI EGGAGTGWTI YPPLSGVQSH SGPSVDLAIF ALHLSGVSSL LGAMNFITTI MNMRTPSIRL HKLALFGWAV IITAVLLLLS LPVLAGAITM LLTDRNFNTS FFETAGGGDP

Rsp  YMINRWLLST NAKDIGVLYI IFGGFSGLVG SALSFIIRLE LSGGGQIYFL GNYHDYNVTI TGHGIVMIFF MVMPSLIGGF GNWLVPVMIG CPDMAFPRLN NVGFWLLPPS LILLITGLFS G-GAGTGWTV YPPLSDSAYH LGTAVDLSIL SLHIAGISSL LGALNLIVTI INIRSQGLTF ERLPLFVWSV KVTAWLLVLS LPVLAGAITL LLFDRNLNTS FYDPSGGGDP

Ama  NTVYRWLFST NAKDIGTLYL VFSIFAGMIG TAFSVLIRFE LAGPGVQYLY GDHQLYNVII TAHAFIMIFF LVMPAMLGGF GNYFVPIMIG APDMAFPRLN NISFWLLPPS LILLVGSAFV EQGAGTGWTV YPPLSSIGFH SGGSVDLAIF SLHLAGISSM LGSINFITTI LNMRAPGMTM HKLPLFVWSI LITAILLLLS LPVLAGAITM LLTDRNLNTT FYDPAGGGDP
```

Block 2 — markers: O (i10) P (i11, COII) S (i12) W (i13, COI) Y (i14) Z (i15) AB (i16) AC (i17) AF (i18) AH (i19)

```
ILYQHLFWFF GHPEVYIIII PGFGIVSQIV STFSGKPIFG YLGMVYAMFS IGILGFLVWS -HHMFSVGLD VDTRAYFTAA TMVIAVPTGI KIFSWLATLY GGSLRFN-TP LLFTIGFVAL FTIGGLTGVV LSNASIDIAF HDTYYVVAHF HYVLSMGAVF ALFAGFYYWA -PKILGR-TI NEFLGKIHFW TLFAGVNLTF FPQHFLGLAG MPRRIPDYPD AFSGWNAVSS

ILFQHLFWFF GHPEVYILIV PAFGIISTTI STNSNKPIFG YIGMVYAMMS IGILGFIVWS -HHMYTVGLD VDTRAYFTAA TLIIAVPTGI KIFSWLATCY GGSIKMT-PS MLFSLGFVFM FTIGGLSGVV LANASLDIAF HDTYYVVAHF HYVLSMGAVF AMFAGWYFWI -PKILGL-NY NLNLAKVQFW LLFIGVNLTF FPQHFLGLQG MPRRISDYPD AFAGWNLISS

ILFQHLFWFF GHPEVYILII PAFGIISTTI SAYSNKSVFG YIGMVYAMMS IGILGFIVWS -HHMYTVGLD VDTRAYFTAA TLIIAVPTGI KIFSWLATCY GGSIRLT-PS MLFALGFVFM FTIGGLSGVV LANASLDIAF HDTYYVVAHF HYVLSMGAVF AMFSGWYFWI -PKMLGL-NY NMTLSKVQFW ILFIGVNVTF FPQHFLGLQG MPRRISDYPD AFAGWNLISS

ILYQHLFWFF GHPEVYILIL PGFGIVSHIV SRFSQKTIFG QVGKVYAMIS IGVLGFIVWS -HHMYLVGLD IDSRAYFTAA TIIIALPTGV KVFSWIATIY GGKVHYT-VP MVFALGFILL FTFGGFTGVI LANASIDVAL HDTYYVVGHF HYVLSLGAVL SLFAGFYYWS -GKIFGY-QA NSKWAYVHYW VFLISINIVF FPMHFLGLQG KPRRIPDLAA GFEGWNNFMT

VLYQHLFWFF GHPEVYIIII PGFGIISQVI STFSRKPIFG YLGMVYAMAS IGILGFIVWS -HHMYTVGLD VDTRAYFTAA TMIIAVPTGI KIFSWLATLY GGNILYR-TP AYFALGFLFL FTIGGVTGVM LANASLDVAL HDTYYVVAHF HYVLSMGAVF ALFAGFYYWI -GKITGK-QY NEFWGQVHFW TMFIGVNVTF FPMHFLGLNG MPRRIPDYPD AFTQWNVISS
```

$y = -5,0316Ln(x) + 13,507$

$R^2 = 0,8335$