Insights into the Phylogeny or Arylamine *N*-Acetyltransferases in Fungi

**Marta Martins · Julien Dairou ·
Fernando Rodrigues-Lima · Jean-Marie Dupret ·
Philippe Silar**

Abstract Previous studies have shown that Eumycetes fungi can acylate arylamine thanks to arylamine *N*-acetyl-transferases, xenobiotic-metabolizing enzymes also found in animals and bacteria. In this article, we present the results of mining 96 available fungal genome sequences for arylamine *N*-acetyltransferase genes and propose their phylogeny. The filamentous Pezizomycotina are shown to possess many putative *N*-acetyltransferases, whilst these are often lacking in other fungal groups. The evolution of the *N*-acetyltransferases is best explained by the presence of at least one gene in the opisthokont ancestor of the fungi and animal kingdoms, followed by recurrent gene losses and gene duplications. A possible horizontal gene transfer event may have occurred from bacteria to the basidiomycetous yeast *Malassezia globosa*.

M. Martins · J. Dairou · F. Rodrigues-Lima · J.-M. Dupret
Univ Paris Diderot-Paris 7, Unité de Biologie Fonctionnelle et Adaptative (BFA), CNRS EAC 4413, Laboratoire des Réponses Moléculaires et Cellulaires aux Xénobiotiques, 75013 Paris, France

M. Martins · J. Dairou · F. Rodrigues-Lima · J.-M. Dupret · P. Silar
UFR des Sciences du Vivant, Univ Paris Diderot-Paris 7, 75013 Paris, France

P. Silar (✉)
Institut de Génétique et Microbiologie, CNRS UMR 8621, Univ Paris-Sud, 91405 Orsay Cedex, France
e-mail: philippe.silar@igmors.u-psud.fr

Present Address:
M. Martins
Cancer Research UK Centre for Cell and Molecular Biology, Chester Beatty Laboratories, The Institute of Cancer Research, Fulham Road, London SW3 6JB, UK

Introduction

Aromatic amines (AAs) are a group of chemicals used in industrial processes to produce many products, including polyurethane. These compounds are also by-products of chemical manufacturing (pesticides, dyes, rubbers or pharmaceuticals) and combustion. Many of them are cytotoxic or genotoxic (Kim and Guengerich 2005). It is, for example, estimated that 12% of the known carcinogenic substances in humans are AA (U.S. Department of Health and Human Services PHS, National Toxicology Program 2005). These chemicals may accumulate in the environment as xenobiotics, where they have noxious effects. This is especially true for 3,4-dichloroaniline (DCA). DCA is a by-product of aniline-derived herbicides such as diuron (Voelker et al. 2008). Indeed, in treated soil, diuron is transformed into DCA and is then acetylated by soil bacteria and fungi into the less toxic acetyl-DCA (Tixier et al. 2002; Westwood et al. 2005). Microorganisms are thus good candidates to be used for decontamination of AA-polluted soil, a process known as bioremediation.

Whilst attention has been first paid to bacteria in soil AA-remediation studies (Rodrigues-Lima et al. 2006), much less is known about the potential of fungi. This is surprising since fungi are the dominant life form in topsoil in term of biomass (Durrieu 1993), and they are known to carry out numerous metabolic reactions. Nevertheless, we

recently showed (Martins et al. 2009) that the fungus *Podospora anserina* is able to acetylate DCA using two arylamine *N*-acetyltransferases (NATs), PaNAT1 and PaNAT2. Although both enzymes acetylate AA in vitro, only the PaNAT2 isoform appears to confer to *P. anserina* resistance to a wide range of AA in vivo. Interestingly, the *P. anserina* enzymes are efficient enough to confer to this fungus the ability to efficiently acetylate DCA in contaminated soil and thereby to bioremediate contaminated soil (Martins et al. 2009). Another NAT enzyme has been characterized from *Fusarium verticillioides* (Glenn and Bacon 2009). This species infects maize and wheat in the field and is frequently found on stored seeds, probably because it is able to tolerate benzoxazolinones, which have antimicrobial activities and are produced by these plants. It does so by hydrolysing 2-benzoxazolinone into 2-aminophenol by enzyme(s) encoded at the FDB1 locus. 2-aminophenol is then acylated and inactivated into the nontoxic metabolite N-(2-hydroxyphenyl)malonamic acid by the FDB2 enzyme. The *FDB2* locus of *F. verticillioides* has been shown to encode a putative NAT enzyme (Glenn and Bacon 2009). To date, the full extent of NAT genes in fungal genomes and their role in fungal physiology are unknown.

NAT enzymes are found only in a subset of organisms (Vagena et al. 2008), and the reason for their presence or their absence in an organism is presently not clear. In our previous study (Martins et al. 2009), we reported that NATs are present in a wide range of fungi, but did not perform a phylogenetic analysis to understand how NAT genes have evolved in this important group of eukaryotes. The availability of 96 genome sequences (Table 1) and the recent deciphering of the fungal phylogeny (Hibbett et al. 2007; James et al. 2006) enable us to perform an analysis of NAT genes in Eumycetes. We present here the results of a phylogenetic survey of NAT genes in all the available fungal genomes. Data analysis reveals a complex evolutionary pattern for NAT genes in Eumycetes, including recurrent gene amplifications and losses, as well as one possible horizontal transfer from a bacterium.

Materials and Methods

Available fungal genomes and deduced protein sets from complete sequencing projects (Table 1) were searched for the presence of putative NAT genes by TBLASTN and BLASTP with e-value cutoffs of 10^{-4} and 10^{-6}, respectively (Altschul et al. 1990). Representative NAT genes from all the species were used as queries on genomic sequences in which no isoform was detected in the first round of queries. In many instances, because automatic annotation may not give accurate intron splicing boundaries, each putative NAT gene was manually

re-annotated. Manual annotation was based on conserved sequence with the closest species as well as the use of consensus for intron boundaries (Kupfer et al. 2004; Silar et al. 2003). Putative functional genes as well as putative pseudogenes were detected (Table 1). Pseudogenes were validated by checking the trace chromatograms to ensure that the mutations observed in the coding sequences were not sequencing errors. A few genes were incomplete due to gaps in the genomic sequences (Table 1). The deduced protein sequences for complete *bona fide* genes were aligned with the MUSCLE algorithm with the default parameters (Edgar 2004, Supplementary Fig. 1 online). The Neighbour Joining tree was made with MEGA4 (Tamura et al. 2007) with the default parameters (complete deletion, Poisson correction and uniform rates amongst sites) and 1,000 bootstraps. The PhyML tree (Guindon and Gascuel 2003) was constructed using the www.phylogeny.fr website with the default parameters (Dereeper et al. 2008) and 100 bootstrapped data sets. Trees were visualized with the iTOL server (Letunic and Bork 2007). Both methods yielded similar results (Fig. 1, Supplementary Fig. 2 online). In all the annotated proteins, the Cys-His-Asp catalytic triad was conserved, indicating that the enzymes are likely to be functional (Supplementary Fig. 1 online). Positions of introns were mapped onto the alignment (Supplementary Fig. 1 online).

Results and Discussion

Complex Set of Gain and Loss of NAT Genes in Eumycete Fungi

Metazoa is the sister group of Eumycetes (James et al. 2006). Most deuterostomian animals possess NAT genes, whilst other animals (Cnidarian, Acoelomata, Pseudocoelomata and Protostomia) do not (Vagena et al. 2008). Yet, the dog, a deuterostomian, lacks NAT genes (Trepanier et al. 1998). The presence of NAT in animals was once thought to be due to horizontal transfer, but this hypothesis was immediately challenged (see Genereux and Logsdon 2003 for a review) and is not favoured anymore (Rodrigues-Lima and Dupret 2002). A patchy phylogenetic distribution similar to that found in animals is also found in fungi. First, no NAT gene is detected in the genomes of unicellular protists related to animals and fungi, i.e. *Monosiga brevicollis* and *Proterospongia* sp. *ATCC50818*, two choanoflagellates, as well as *Capsaspora owczarzaki*, a mesomycetozoan (Table 1). Second, NAT genes are also absent from the genomes of all the investigated lower fungi (Spizellomycetales, Blastocladiomycota, Microsporidia and Mucoromycotina), except for the genome of *Batrachochytrium dendrobatidis* that contains a single

Table 1 NAT genes in fungal genomes

Taxonomic affiliation[a]	Species	Major lifestyle and habitat	Putative NAT genes (pseudogenes)[b]	NAT families	ID	Database[c]
Organisms related to fungi						
Choanoflagellate	*Monosiga brevicollis*	Bacteriovorous protist	0			JGI
	Proterospongia sp. *ATCC50818*	Bacteriovorous protist	0			OM
Mesomycetozoan	*Capsaspora owczarzaki*	Snail symbiont	0			OM
"Lower fungi"						
Chytridiomycota						
Chytriales	*Batrachochytrium dendrobatidis*	Frog pathogen	1		BDEG_00384.1	FGI
Spizellomycetales	*Spizellomyces punctatus*	Soil saprobe	0			OM
Blastocladiomycota	*Allomyces macrogynus*	Pond saprobe	0			OM
Mucoromycotina	*Rhizopus oryzae*	Soil saprobe	0			FGI
	Phycomyces blakesleeanus	Soil saprobe	0			JGI
	Mucor circinelloides	Soil saprobe	0			JGI
Microsporidia	*Encephalitozoon cuniculi*	Intracellular parasite	0			Genoscope
	Antonospora locustae	Intracellular parasite	0			Marine Lab.
	Nosema ceranae	Intracellular parasite	0			NCBI
"Higher fungi"						
Basidiomycota						
Ustilaginomycotina	*Ustilago maydis*	Maize pathogen	1		UM01576.1	FGI
	Malassezia globosa	Yeast, human scalp and skin commensal	1		MGL_4060	NCBI
Pucciniomycotina	*Puccinia graminis*	Wheat pathogen	1		PGTG_03435.2	FGI
	Sporobolomyces roseus	Yeast, phylloplan yeast saprobe	2		Sporo1l27929, Sporo1l33953	JGI
	Melampsora larici-populina	Yeast, poplar pathogen	1		Mellp1l108766	JGI
Agaricomycotina						
Tremellomycetes	*Cryptococcus neoformans*	Yeast, soil and dropping saprobe, human and bird opportunistic pathogen	0			FGI
	Tremella mesenterica	Mycoparasite	0			FGI
Agaricomycetes	*Phanerochaete chrysosporium*	Wood saprobe	0			JGI
	Laccaria bicolor	Mycorrhizal	0			JGI
	Pleurotus ostraceus	Wood saprobe	1		PleosPC15_1l30635	JGI
	Schizophyllum commune	Wood saprobe	1*		Schco1l102847	JGI
	Moniliophthora perniisiosa	Cocoa pathogen	1*		NA	NCBI
	Coprinopsis cinerea	Dung and soil saprobe	2		CC1G_07781.2, CC1G_07803.2	FGI
	Postia placenta	Wood saprobe	0			JGI
	Heterobasidion annosum	Wood saprobe	(1)		NA	JGI
	Serpula lacrymans	Wood saprobe	0			JGI

Table 1 continued

Taxonomic affiliation[a]	Species	Major lifestyle and habitat	Putative NAT genes (pseudogenes)[b]	NAT families	ID	Database[c]
Ascomycota						
Taphrinomycotina	*Schizosaccharomyces pombe*	Yeast saprobe	0			Sanger
	Schizosaccharomyces japonicus	Yeast saprobe	0			FGI
	Schizosaccharomyces octosporus	Yeast saprobe	0			FGI
	Pneumocystis carinii	Yeast-like lung commensal/pathogen	0			PGP
Saccharomycotina	*Yarrowia lipolytica*	Yeast soil saprobe	0			Genolevure
	Debaryomyces hanseni	Yeast saprobe	0			Genolevure
	Pichia stipitis	Yeast saprobe	0			JGI
	Candida albicans	Yeast, human commensal, opportunistic pathogen	0			FGI
	Candida guillermondi	Yeast, human commensal, opportunistic pathogen	0			FGI
	Candida lusitaniae	Yeast, human commensal, opportunistic pathogen	0			FGI
	Candida tropicalis	Yeast, human commensal, opportunistic pathogen	0			FGI
	Ashbya gossypii	Yeast-like, cotton pathogen	0			AGD
	Kluyveromyces lactis	Yeast saprobe	0			Genolevure
	Zygosaccharomyces rouxii	Yeast saprobe	0			Genolevure
	Kluyveromyces thermotolerans	Yeast saprobe	0			Genolevure
	Candida glabrata	Yeast, human commensal, opportunistic pathogen	0			Genolevure
	Saccharomyces cerevisae	Yeast saprobe	0			Genolevure
	Saccharomyces klyuveri	Yeast saprobe	0			Genolevure
Pezizomycotina						
Pezizomycetes	*Tuber melanosporum*	Mycorrhizal	0			Genoscope
Dothideomycetes	*Mycosphaerella fijiensis*	Banana pathogen	3	I III VI	Mycfi133629 Mycfi151513 Mycfi141649	JGI
	Mycosphaerella graminicolla	Wheat pathogen	2	V VI	Mycgr3l39307 Mycgr3l85442	JGI
	Alternaria brassicola	Brassicaceae pathogen	1	VI	AB05009.1	JGI
	Phaeosphaeria nodorum	Wheat pathogen	2	I VI	SNOG_06959.1 SNOG_06352.1	FGI
	Pyrenophora tritici	Wheat pathogen	2	III VI	PTRG_06864.1 PTRG_02465.1	FGI
	Cochliobolus heterostrophus	Maize pathogen	1	VI	CocheC5_1l107425	JGI

Table 1 continued

Taxonomic affiliation[a]	Species	Major lifestyle and habitat	Putative NAT genes (pseudogenes)[b]	NAT families	ID	Database[c]
Eurotiomycetes	Ascosphaera apis	Bee parasite	(1)	(II ?)	NA	HGSC
	Microsporum canis	Keratinophilic fungus	1	IV	MCYG_06389	FGI
	Trichophyton equinum	Keratinophilic fungus	1	IV	TEQG_03758.1	FGI
	Trichophyton tonsurans	Keratinophilic fungus	1	IV	NA	FGI
	Trichophyton rubrum	Keratinophilic fungus	1	IV	NA	FGI
	Arthroderma gypseum	Keratinophilic fungus	2	I	MGYG_02694.1	FGI
				IV	MGYG_06842.1	
	Uncinocarpus reesii	Keratinophilic fungus	1	IV	UREG_05260.1	FGI
	Coccidioides immitis	Soil saprobe, human opportunistic pathogen	1	IV	CISG_01266.1	FGI
	Coccidioides posadasii	Soil saprobe, human opportunistic pathogen	1	IV	CPAG_01321.1	FGI
	Paracoccidioides brasiliensis	Soil and dropping saprobe, human and bird opportunistic pathogen	1	IV	PADG_01689.1	FGI
	Blastomyces dematidis	Keratinophilic fungus	1	IV	BDBG_03471.1	FGI
	Histoplasma capsulatum	Soil saprobe, human opportunistic pathogen	1	IV	HCAG_01594.1	FGI
	Aspergillus nidulans	Soil saprobe	1	IV	ANID_10723.1	FGI
	Aspergillus carbonarius	Soil saprobe	2	I	Aspcal10395	JGI
				III	Aspcal14107	
				IV	Aspcal23693	
	Aspergillus niger	Soil saprobe	5	III	43936	JGI
				III	49151	
				IV	43842	
				IV	45379	
				VII	181999	
	Aspergillus oryzae	Soil saprobe	4	I	AO090010000051/52	DOGAN
				II	AO090012000406	
				III	AO090011000264	
				IV	AO090003001123	
	Aspergillus terreus	Soil saprobe	2	I	ATEG_05747.1	FGI
				IV	ATEG_03982.1	
	Aspergillus clavatus	Soil saprobe	1	IV	ACLA_050060	JVCI
	Aspergillus fumigatus	Soil saprobe, human opportunistic pathogen	1	IV	Afu4g10930	JVCI
	Neosartorya fischeri	Soil saprobe	2	IV	NFIA_104890	JVCI
				VII	NFIA_004340	
	Penicillium marneffei	Soil saprobe, opportunistic human pathogen	0			NCBI
	Talaromyces stipitatus	Soil saprobe	1	III	TSTA_053130	NCBI
	Penicillium chrysogenum	Soil saprobe	3	I	Pc12g02960	NCBI
				III	Pc20g04880	
				IV	Pc22g06840	

Table 1 continued

Taxonomic affiliation[a]	Species	Major lifestyle and habitat	Putative NAT genes (pseudogenes)[b]	NAT families	ID	Database[c]
Leotiomycetes	Blumeria graminis	Grass pathogen	0			NCBI
	Botrytis cinerea	Plant pathogen with wide host range	1	I	BC1G_14967.1	FGI
	Sclerotinia sclerotiorum	Plant pathogen with wide host range	1*	I	SS1G_08714.1	FGI
Sordariomycetes	Trichoderma virens	Soil saprobe and mycoparasite	2	I	Trive1|46725	JGI
				II	Trive1|69909	
	Trichoderma reiseii	Soil saprobe	1	II	Trire2|4737	JGI
	Trichoderma atroviride	Soil saprobe and mycoparasite	1	II	Triat1|91035	JGI
	Epichloë festucae	Grass mutualist	0			EFGP
	Fusarium verticillioides	Soil saprobe and corn pathogen	3 + (1)	I	FVEG_03961.3	FGI
				III	FDB2 = FVEG_12636.3	
				V	FVEG_12062.3	
				(I)	(FVEG_07425.3)	
	Fusarium graminearum	Wheat pathogen	3	I	FGSG_09400.3	FGI
				III	FGSG_00080.3	
				V	FGSG_07888.3	
	Fusarium oxysporum	Soil saprobe and broad range plant pathogen	4	I	FOXG_04301.2	FGI
				I	FOXG_06095.2	
				III	FOXG_15318.2	
				V	FOXG_03795.2	
	Nectria haematococca	Soil saprobe and broad range pathogen	5 + (2)	I	Necha2|51017	JGI
				I	Necha2|75515	
				II	Necha2|87499	
				II	Necha2|33380	
				V	Necha2|34602	
				(I)	(Necha2|44961)	
				(I)	NA	
	Verticillium dahliae	Soil saprobe and broad range plant pathogen	1	II	VDAG_09857.1	FGI
	Verticillium albo-atrum	Soil saprobe and broad range plant pathogen	(1)	(II)	(VDBG_07756.1)	FGI
	Ophiostoma (Grosmania) clavigerum	Pine pathogen	0			NCBI
	Cryphonectria parasitica	Chesnut pathogen	2	I	NA1	JGI
				I	NA2	
	Magnaporthe grisea	Rice pathogen	2 + (1)	II	MGG_10560	FGI
				V	MGG_04278	
				(V)	NA	
	Neurospora tetrasperma	Soil saprobe	1	II	Neute1|162225	JGI
	Neurospora discreta	Soil saprobe	1	II	Neudi1|79748	JGI
	Neurospora crassa	Soil saprobe	1	II	NCU09784.3	FGI

Table 1 continued

Taxonomic affiliation[a]	Species	Major lifestyle and habitat	Putative NAT genes (pseudogenes)[b]	NAT families	ID	Database[c]
	Chaetomium globosum	Soil and dung saprobe	2	I II	CHG06975.1 CHG09458.1	FGI
	Sporotrichum thermophile	Soil saprobe	1	II	Spoth189317	JGI
	Thielavia terrestris	Soil saprobe	0			JGI
	Podospora anserina	Dung saprobe	2	I	PaNAT1 = Pa_2_13150 PaNAT2 = Pa_4_4860	Genoscope/IGM
II						

[a] Hibbett et al. (2007) was followed for the taxonomic affiliation. The table is arranged according to the known phylogeny of Eumycetes

[b] Possible pseudogenes are in parenthesis; * incomplete genes due to gaps in genome sequence; *NA* genes not annotated; For the Pezizomycotina, the NAT families are given in roman uppercase

[c] Databases URL

AGD: http://agd.vital-it.ch/index.html

DOGAN: http://www.bio.nite.go.jp/dogan/MicroTop?GENOME_ID=ao

EFGP: http://www.endophyte.uky.edu/

FGI: http://www.broadinstitute.org/science/projects/fungal-genome-initiative/fungal-genome-initiative

Genolevure: http://www.genolevures.org/

Genoscope: http://www.genoscope.cns.fr/spip/Les-champignons-sequences-au.html

HGSC: http://www.hgsc.bcm.tmc.edu/microbial-index.xsp

JGI: http://genome.jgi-psf.org/

JVCI: http://www.jcvi.org/

Marine Lab: http://gmod.mbl.edu/perl/site/antonospora01?page=intro

NCBI: http://www.ncbi.nlm.nih.gov/genomes/leuks.cgi

OM: http://www.broadinstitute.org/annotation/genome/multicellularity_project/MultiHome.html

Sanger: http://old.genedb.org/genedb/pombe/

Fig. 1 Maximum Likelihood phylogenetic tree of fungal NAT genes Tree generated from the alignment of Supplementary Fig. 1 online by PhyML (Guindon and Gascuel 2003) with 100 bootstrap replicas. NAT families I–VII are from Pezizomycotina. *Arrow* indicates the *M. globosa* NAT gene
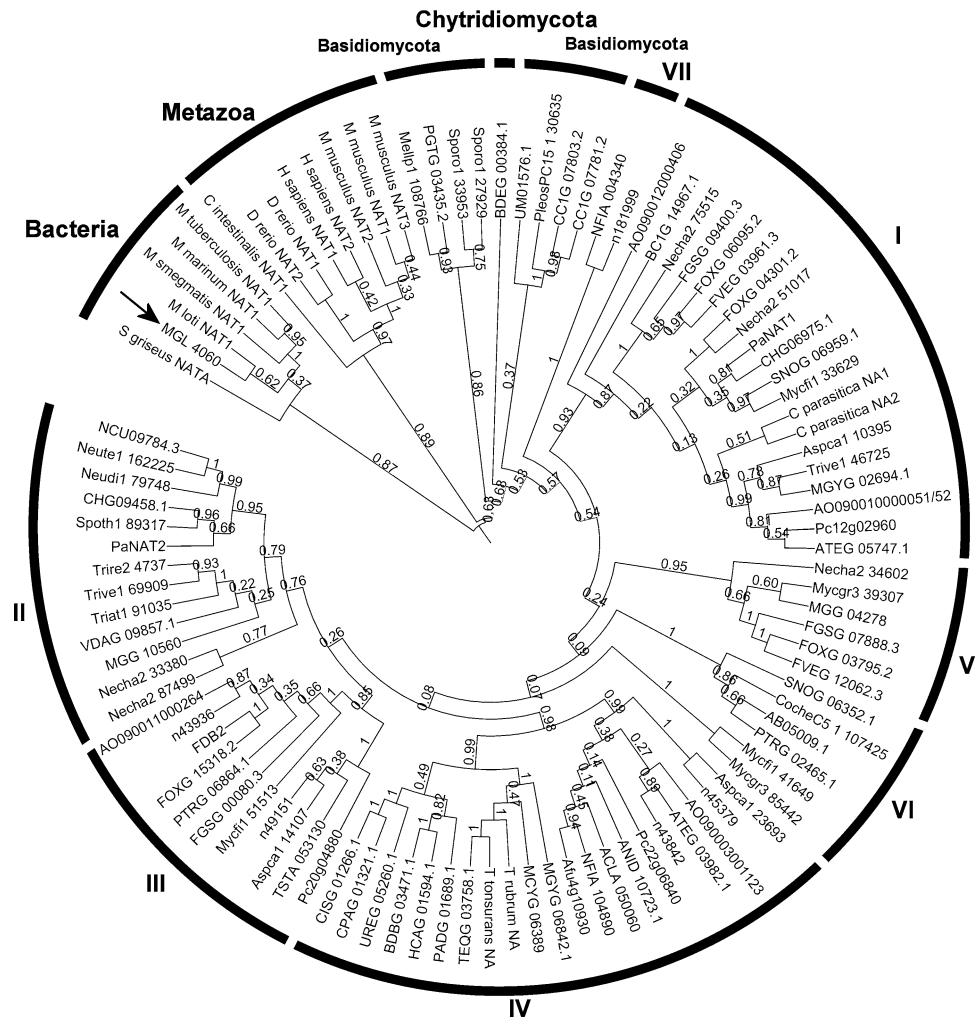
NAT gene (Table 1). Lack of NAT in the Mucoromycotina is in line with their sensitivity to AA (Martins et al. 2009). Third, NAT genes are found in both Basidiomycota and Ascomycota, albeit with a patchy phylogenetic distribution (Table 1).

The presence of NAT genes in both animals and fungi suggests that they were present in the common ancestor of both groups and lost independently in some lineages, as evidenced by the detection of pseudogenes in fungal genomes (Table 1). Trees for all the fungal full-length proteins and selected NAT from other organisms were constructed to reconstitute their evolution (Fig. 1, Supplementary Fig. 2 online). Unfortunately, due to the small size of the NAT proteins and their rapid evolution, the branching order for the most divergent NAT is not defined with good statistical support, especially when the whole diversity of fungi is considered. This precludes the decision as to whether the present patchy repartition is due to artefacts in the methods used for phylogenetic analysis, to recurrent gene losses or to multiple horizontal transfers early during evolution. For example, in both analyses, the *B. dendrobatidis* protein

branches within the NAT from Basidiomycetes, which could be indicative of horizontal transfer or which may result from a long-branch attraction. However, apart from this discrepancy and that of a probable horizontal transfer from a bacterium to *Malassezia globosa*, (see below), the NAT gene phylogeny follows that known for species, an argument in favour of an early presence of NAT genes followed by recurrent gene losses. Figure 2 proposes a plausible scenario for NAT gene evolution based on the data of Table 1 and the known phylogeny of fungi. Whatever the scenario for the early evolution of NAT, additional recent gene losses are clearly needed to explain the lack of NAT gene in particular organisms, such as the dog and some species of higher fungi (Table 1, Fig. 2).

In the higher fungi, NAT genes also present a patchy phylogenetic distribution and are found mostly in filamentous species (Table 1). Indeed, whilst they are present in the Pucciniomycotina yeasts and *M. globosa*, they are absent in all the other yeasts, i.e. *Cryptococcus neoformans* and the Taphrinomycotina and Saccharomycotina species. This pattern is in line with numerous gene losses that go

Fig. 2 Possible scenario of NAT genes evolution. The figure gives the most parsimonious scenario of NAT genes gains and losses based on the data of Table 1 and the currently known evolution of fungi (Hibbett et al. 2007; James et al. 2006). Alternative scenarios involving additional gene losses, gene duplication or lateral gene transfer are possible. **a** NAT gene evolution in fungi, but Pezizomycotina. Note the presumptive losses of NAT genes in many phyla. *Triangle* gene deletion, *2×* gene duplication, *HGT* horizontal gene transfer. **b** NAT gene evolution in Pezizomycotina. The NAT isoforms putatively present at the origins of the different classes are indicated on the *left*. This scenario suggests a highly dynamic pattern of gene gains and losses. Deletion (triangle) or duplication (2×) of specific isoforms is indicated above the corresponding branches. The lengths of the branches do not reflect the evolutionary distances between the species

together with the yeast life style (Arvas et al. 2007; Zeng et al. 2001). Intriguingly, a NAT activity has been reported in *Candida albicans* (Fang et al. 1997). It is likely that other enzymes supply the NAT activity in this species, whose genome does not contain canonical NAT gene. Therefore, if a species is missing canonical NAT genes, then it may still retain some NAT activity carried out by another acetyl transferase, although no such enzyme has yet been characterized enzymatically. In the filamentous Basidiomycota, NAT genes have been lost in many species of the higher Agaricomycotina (Fig. 2), whilst still present in Ustilaginomycotina and Pucciniomycotina. Interestingly, Basidiomycota genomes and especially Agaricomycotina ones are often poor in polyketide synthases (PKS) and non ribosomal peptide synthases (NRPS), which are enzymes involved in secondary metabolisms. For example, the genomes of the mycorrhizal fungus *Laccaria bicolor* and the saprobic *Coprinopsis cinerea* contain a single PKS as well as one and five NRPS, respectively, whilst the genome of the white rot *Phanaerochaete chrysosporium* has no PKS and three NRPS. Similarly, yeasts that do not contain NAT are devoid of PKS and contain few NRPS. In line with these observations, the most diverse array of NAT genes is found in the higher Ascomycota, the filamentous Pezizomycotina that are endowed with numerous PKS and NRPS genes involved in the production of secondary metabolites (Bushley and Turgeon 2010; Kroken et al. 2003).

The Genomes of Higher Ascomycetes Contain Numerous NAT Genes

Most species of Pezizomycotina contain at least one NAT gene, often several with up to five functional paralogs and two pseudogenes in *Haematonectria* (*Nectria*) *haematoccocca*, yet, a few species appear devoid of NAT genes (Table 1), which can only be accounted for by gene loss in these fungi. The history of NAT genes in Pezizomycotina appears quite complex since sequence comparison permits the definition of at least seven groups (I–VII) supported by strong statistical significance (Fig. 1, Supplementary Fig. 2 online). However, the branching order between these families is poorly resolved, and differs between the Neighbour Joining and PhyML methods. Intron presence and conserved position (see Supplementary Fig. 1 online) support a common ancestor for these genes and is in accordance with their grouping into seven families. Genes belonging to the family IV usually have two introns (I1 and I2, Supplementary Fig. 1 online) with conserved positions, except in *Coccidioides imitis*, *C. posadasii* and *Uncinocarpus reesii,* which only have I1. These three species are closely related, a single loss is, therefore, needed to explain lack of I2. I1 is present in many genes of families I, II, III

and VII, whilst a few genes of the same families lack I1. The intron pattern of Pezizomycotina NAT genes is thus best explained by recurrent intron loss, starting with I2 and then I1. However, there appears to be two exceptions, *Mycosphaerella fijiensis Mycfi1|51513* and *Cryphonectria parasitica NA2* genes. *M. fijiensis Mycfi1|51513* possesses two introns located at unique positions and may have gained its introns recently as seen for other fungal genes (Nielsen et al. 2004). *C. parasitica NA2* intron is slightly moved 7 nucleotides towards the 3′ end of the gene, a phenomenon known as intron sliding (Rogozin et al. 2005). It is noteworthy that none of the intron positions in Pezizomycotina genes is common with the ones from genes of Basidiomycota (Supplementary Fig. 1 online). The selection pressure and the mechanisms for intron gain and loss in fungi are still controversial (Nielsen et al. 2004).

Some NAT families are restricted to a specific clade of Pezizomycotina, whilst others are present in the whole diversity of Pezizomycotina. Family II is found only in Sordariomycetes, families IV and VII in Eurotiomycetes and family VI in Dothideomycetes. On the contrary, family I is found in the four investigated classes of Pezizomycotina. However, whilst it is frequent in Sordariomycetes and possibly Leotiomycetes, it is rare in Eurotiomycetes and Dothideomycetes. Retracing the evolution of the NAT families (Fig. 2) is difficult since they are too divergent to obtain good statistical support for branching order (Fig. 1, Supplementary Fig. 2 online). Interestingly, closely related species may harbour *NAT* genes belonging to different families. For example, *Aspergilli* may contain genes of families III, IV and VII (*A. niger*) or I, III and IV families (*A. oryzae*). Conversely, paralogs resulting from duplication events may be present in the same species, as seen in *H. haematoccocca*, which contains two functional paralogs of family I, two corresponding pseudogenes inactivated by mutations and two functional paralogs of family II. Overall, this complex pattern of the absence/presence of *NAT* paralogs suggests intense gene duplications and losses during the evolution of Pezizomycotina (Fig. 2), as observed for other protein families (Espagne et al. 2008; Wapinski et al. 2007). Amongst all these NAT, only the activities of PaNAT1/PaNAT2 of *P. anserina* (families I and II, respectively) and *F. verticillioides* FDB2, which belongs to the family III (Table 1), have been explored (Glenn and Bacon 2009; Martins et al. 2009). In both species, the enzymes are involved in the resistance to toxic compounds. It is known from studies in Streptomycetes that NAT-related enzymes may function as amide synthase that catalyses the ring closure of macro-cyclic chemicals produced by PKS, such as rifampicin (see Sim et al. 2008 for a review). There is thus a possibility that some enzymes detected in Pezizomycotina may act in secondary metabolite production. Intriguingly, there seems to be no

correlation between the known lifestyle of the fungi and the presence of one or several NAT genes (Table 1), suggesting that NAT genes respond to particular demands, as exemplified by the FDB2 enzyme of *F. verticillioides*, which may facilitate the growth of this species on some cereals. It is likely that a finer knowledge of the biotope and lifestyle of the various fungi will be necessary to understand the selection pressure that may lead to the retention or loss of NAT genes during evolution.

A Horizontal Transfer from a Bacterium to *Malassezia globosa*

Amongst the NAT genes present in Eumycota, the one present in *M. globosa* (MGL_4060) stands out by the fact that it branches with high statistical support amongst bacterial genes, as there is 81 and 87% boostrap support in the Neighbour Joining and PhyML analyses, respectively, for a bacterial clade of NAT + MGL_4060 (Fig. 1, Supplementary Fig. 2 online,). It is, therefore, likely that this gene is a recent transfer from a bacterium to *M. globosa*. GenBANK search by BLASTP shows that the NAT similar to the *M. globosa* NAT is encoded by the *NAT1* gene of *Mesorhizobium loti* (Rodrigues-Lima et al. 2006). *M. loti* is likely not the donor as it is involved in symbiosis with plant roots, whilst *M. globosa* is associated with several seborreic conditions in humans, such as dandruff. The sequenced strain (Xu et al. 2007) derives from a patient affected with pityriasis versicolor. Therefore, in addition to gain by duplication, fungi may gain NAT genes by horizontal transfer, a feat also achieved for PKS genes (Schmitt and Lumbsch 2009).

Lack of NAT Genes in Oomycetes and Presence in Related Algal Groups

The fungal lifestyle has evolved in another group of eukaryotes, the Oomycetes. Although similar in appearance and ecology, these organisms possess quite different biological properties from Eumycetes and are related to diatoms and other algae of the Chromista eukaryotic supergroup (Lamour et al. 2007). It has been proposed that genes were horizontally transferred from Eumycetes to Oomycetes during their long evolutionary cohabitation (Richards et al. 2006). To test, if NAT genes were transferred, then we mined the sequenced genomes of Oomycetes (Lamour et al. 2007). BLAST analysis did not reveal any NAT gene in the genomes of four oomycetes (*Hyaloperonospora parasitica* at http://genomeold.wustl.edu/index.cgi, *Phytophthora sojae*, *P. ramorum* and *P. infestans* at http://genome.jgi-psf.org/) nor in the genome of *Thalassiosira pseudonana*, a diatom related to oomycetes (at http://genome.jgi-psf.org/). However, one NAT gene is present in the diatom *Phaeodactylum tricornutum* (this

gene has not been detected by the automatic annotation), two in the genome of *Fragilariopsis cylindrus*, another diatom (Fracy1|264642 and Fracy1|243479 at http://genome.jgi-psf.org/) and at least two in the more distantly related haptophyte, *Emiliana huxleyi* (Emihu1|222756 and Emihu1|451400 at http://genome.jgi-psf.org/). Therefore, NAT genes present the same patchy distribution in Chromista as observed in Opisthokonts and have likely not been transferred from Eumycetes to Oomycetes

Conclusion

The genomes of filamentous Ascomycetes contain numerous NAT genes, unlike those of the other fungal clades. The fungal NAT genes are most likely derived from an ancestral gene that was present in the ancestor of all the opisthokonts, except for one gene potentially resulting from the horizontal transfer from a bacterium to *M. globosa*. During evolution, the fungal NAT gene family has undergone a complex history of duplications and gene losses (Fig. 2), as observed in other fungal gene families. The role of NAT in fungi is not clear. In the investigated cases, NAT enables fungi to resist toxic metabolites, be they xenobiotics or plant secondary metabolites. However, the presence of NAT genes follows a similar pattern of taxonomic distribution and amplification as seen for PKS and NRPS genes in fungal genomes, suggesting that some NAT could be involved in the production/modification of fungal secondary metabolites.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Arvas M, Kivioja T, Mitchell A, Saloheimo M, Ussery D, Penttila M, Oliver S (2007) Comparison of protein coding gene contents of the fungal phyla Pezizomycotina and Saccharomycotina. BMC Genomics 8:325

Bushley KE, Turgeon BG (2010) Phylogenomics reveals subfamilies of fungal nonribosomal peptide synthetases and their evolutionary relationships. BMC Evol Biol 10:26

Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie JM, Gascuel O (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res 36:W465–W469

Durrieu G (1993) Ecologie des champignons. Masson, Paris

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797

Espagne E, Lespinet O, Malagnac F, Da Silva C, Jaillon O, Porcel BM, Couloux A, Aury JM, Segurens B, Poulain J, Anthouard V, Grossetete S, Khalili H, Coppin E, Dequard-Chablat M, Picard M, Contamine V, Arnaise S, Bourdais A, Berteaux-Lecellier V,

Gautheret D, de Vries RP, Battaglia E, Coutinho PM, Danchin EG, Henrissat B, Khoury RE, Sainsard-Chanet A, Boivin A, Pinan-Lucarre B, Sellem CH, Debuchy R, Wincker P, Weissenbach J, Silar P (2008) The genome sequence of the model ascomycete fungus *Podospora anserina*. Genome Biol 9:R77

Fang SH, Chung JG, Chang WC, Chang SS (1997) Evidence for arylamine *N*-acetyltransferase activity in the fungi *Candida albicans*. Toxicol Lett 92:109–116

Genereux DP, Logsdon JM Jr (2003) Much ado about bacteria-to-vertebrate lateral gene transfer. Trends Genet 19:191–195

Glenn AE, Bacon CW (2009) FDB2 encodes a member of the arylamine *N*-acetyltransferase family and is necessary for biotransformation of benzoxazolinones by *Fusarium verticillioides*. J Appl Microbiol 107(2):657–671

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52:696–704

Hibbett DS, Binder M, Bischoff JF, Blackwell M, Cannon PF, Eriksson OE, Huhndorf S, James T, Kirk PM, Lucking R, Thorsten Lumbsch H, Lutzoni F, Matheny PB, McLaughlin DJ, Powell MJ, Redhead S, Schoch CL, Spatafora JW, Stalpers JA, Vilgalys R, Aime MC, Aptroot A, Bauer R, Begerow D, Benny GL, Castlebury LA, Crous PW, Dai YC, Gams W, Geiser DM, Griffith GW, Gueidan C, Hawksworth DL, Hestmark G, Hosaka K, Humber RA, Hyde KD, Ironside JE, Koljalg U, Kurtzman CP, Larsson KH, Lichtwardt R, Longcore J, Miadlikowska J, Miller A, Moncalvo JM, Mozley-Standridge S, Oberwinkler F, Parmasto E, Reeb V, Rogers JD, Roux C, Ryvarden L, Sampaio JP, Schussler A, Sugiyama J, Thorn RG, Tibell L, Untereiner WA, Walker C, Wang Z, Weir A, Weiss M, White MM, Winka K, Yao YJ, Zhang N (2007) A higher-level phylogenetic classification of the Fungi. Mycol Res 111:509–547

James TY, Kauff F, Schoch CL, Matheny PB, Hofstetter V, Cox CJ, Celio G, Gueidan C, Fraker E, Miadlikowska J, Lumbsch HT, Rauhut A, Reeb V, Arnold AE, Amtoft A, Stajich JE, Hosaka K, Sung GH, Johnson D, O'Rourke B, Crockett M, Binder M, Curtis JM, Slot JC, Wang Z, Wilson AW, Schussler A, Longcore JE, O'Donnell K, Mozley-Standridge S, Porter D, Letcher PM, Powell MJ, Taylor JW, White MM, Griffith GW, Davies DR, Humber RA, Morton JB, Sugiyama J, Rossman AY, Rogers JD, Pfister DH, Hewitt D, Hansen K, Hambleton S, Shoemaker RA, Kohlmeyer J, Volkmann-Kohlmeyer B, Spotts RA, Serdani M, Crous PW, Hughes KW, Matsuura K, Langer E, Langer G, Untereiner WA, Lucking R, Budel B, Geiser DM, Aptroot A, Diederich P, Schmitt I, Schultz M, Yahr R, Hibbett DS, Lutzoni F, McLaughlin DJ, Spatafora JW, Vilgalys R (2006) Reconstructing the early evolution of Fungi using a six-gene phylogeny. Nature 443:818–822

Kim D, Guengerich FP (2005) Cytochrome P450 activation of arylamines and heterocyclic amines. Annu Rev Pharmacol Toxicol 45:27–49

Kroken S, Glass NL, Taylor JW, Yoder OC, Turgeon BG (2003) Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes. Proc Natl Acad Sci USA 100:15670–15675

Kupfer DM, Drabenstot SD, Buchanan KL, Lai H, Zhu H, Dyer DW, Roe BA, Murphy JW (2004) Introns and splicing elements of five diverse fungi. Eukaryot Cell 3:1088–1100

Lamour KH, Win J, Kamoun S (2007) Oomycete genomics: new insights and future directions. FEMS Microbiol Lett 274:1–8

Letunic I, Bork P (2007) Interactive Tree OF Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics 23:127–128

Martins M, Rodrigues-Lima F, Dairou J, Lamouri A, Malagnac F, Silar P, Dupret JM (2009) An acetyltransferase conferring tolerance to toxic aromatic amine chemicals: molecular and functional studies. J Biol Chem 284:18726–18733

Nielsen CB, Friedman B, Birren B, Burge CB, Galagan JE (2004) Patterns of intron gain and loss in fungi. PLoS Biol 2:e422

Richards TA, Dacks JB, Jenkinson JM, Thornton CR, Talbot NJ (2006) Evolution of filamentous plant pathogens: gene exchange across eukaryotic kingdoms. Curr Biol 16:1857–1864

Rodrigues-Lima F, Dupret JM (2002) In silico sequence analysis of arylamine *N*-acetyltransferases: evidence for an absence of lateral gene transfer from bacteria to vertebrates and first description of paralogs in bacteria. Biochem Biophys Res Commun 293:783–792

Rodrigues-Lima F, Dairou J, Diaz CL, Rubio MC, Sim E, Spaink HP, Dupret JM (2006) Cloning, functional expression and characterization of *Mesorhizobium loti* arylamine *N*-acetyltransferases: rhizobial symbiosis supplies leguminous plants with the xenobiotic *N*-acetylation pathway. Mol Microbiol 60:505–512

Rogozin IB, Sverdlov AV, Babenko VN, Koonin EV (2005) Analysis of evolution of exon-intron structure of eukaryotic genes. Brief Bioinform 6:118–134

Schmitt I, Lumbsch HT (2009) Ancient horizontal gene transfer from bacteria enhances biosynthetic capabilities of fungi. PLoS One 4:e4437

Silar P, Barreau C, Debuchy R, Kicka S, Turcq B, Sainsard-Chanet A, Sellem CH, Billault A, Cattolico L, Duprat S, Weissenbach J (2003) Characterization of the genomic organization of the region bordering the centromere of chromosome V of *Podospora anserina* by direct sequencing. Fungal Genet Biol 39:250–263

Sim E, Sandy J, Evangelopoulos D, Fullam E, Bhakta S, Westwood I, Krylova A, Lack N, Noble M (2008) Arylamine N-acetyltransferases in mycobacteria. Curr Drug Metab 9:510–519

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

Tixier C, Sancelme M, Ait-Aissa S, Widehem P, Bonnemoy F, Cuer A, Truffaut N, Veschambre H (2002) Biotransformation of phenylurea herbicides by a soil bacterial strain, Arthrobacter sp. N2: structure, ecotoxicity and fate of diuron metabolite with soil fungi. Chemosphere 46:519–526

Trepanier LA, Cribb AE, Spielberg SP, Ray K (1998) Deficiency of cytosolic arylamine *N*-acetylation in the domestic cat and wild felids caused by the presence of a single NAT1-like gene. Pharmacogenetics 8:169–179

U.S. Department of Health and Human Services PHS, National Toxicology Program (2005) 11th Report on Carcinogens, Research Triangle Park, NC

Vagena E, Fakis G, Boukouvala S (2008) Arylamine *N*-acetyltransferases in prokaryotic and eukaryotic genomes: a survey of public databases. Curr Drug Metab 9:628–660

Voelker D, Stetefeld N, Schirmer K, Scholz S (2008) The role of cyp1a and heme oxygenase 1 gene expression for the toxicity of 3,4-dichloroaniline in zebrafish (Danio rerio) embryos. Aquat Toxicol 86:112–120

Wapinski I, Pfeffer A, Friedman N, Regev A (2007) Natural history and evolutionary principles of gene duplication in fungi. Nature 449:54–61

Westwood IM, Holton SJ, Rodrigues-Lima F, Dupret JM, Bhakta S, Noble ME, Sim E (2005) Expression, purification, characterization and structure of *Pseudomonas aeruginosa* arylamine *N*-acetyltransferase. Biochem J 385:605–612

Xu J, Saunders CW, Hu P, Grant RA, Boekhout T, Kuramae EE, Kronstad JW, Deangelis YM, Reeder NL, Johnstone KR, Leland M, Fieno AM, Begley WM, Sun Y, Lacey MP, Chaudhary T, Keough T, Chu L, Sears R, Yuan B, Dawson TL Jr (2007) Dandruff-associated *Malassezia* genomes reveal convergent and divergent virulence traits shared with plant and human fungal pathogens. Proc Natl Acad Sci USA 104:18730–18735

Zeng Q, Morales AJ, Cottarel G (2001) Fungi and humans: closer than you think. Trends Genet 17:682–684