FEBS Letters 584 (2010) 3158–3164

FEBS Letters

journal homepage: www.FEBSLetters.org

Comparative genomic and phylogenetic investigation of the xenobiotic metabolizing arylamine *N*-acetyltransferase enzyme family

Anthony E. Glenn [a], Eleni P. Karagianni [b], Antigona Ulndreaj [b], Sotiria Boukouvala [b,*]

[a] *Toxicology and Mycotoxin Research Unit, Russell Research Center, Agricultural Research Service, United States Department of Agriculture, 950 College Station Road, Athens, GA 30605, USA*

[b] *Department of Molecular Biology and Genetics, Democritus University of Thrace, Alexandroupolis 68100, Greece*

ARTICLE INFO

Article history:
Received 12 April 2010
Revised 24 May 2010
Accepted 31 May 2010
Available online 2 June 2010

Edited by Takashi Gojobori

Keywords:
Arylamine *N*-acetyltransferases
Xenobiotic metabolizing enzymes
Archaea
Fungi
Protista
Phylogeny

ABSTRACT

Arylamine *N*-acetyltransferases (NATs) are xenobiotic metabolizing enzymes characterized in several bacteria and eukaryotic organisms. We report a comprehensive phylogenetic analysis employing an exhaustive dataset of *NAT*-homologous sequences recovered through inspection of 2445 genomes. We describe the first *NAT* homologues in viruses, archaea, protists, many fungi and invertebrates, providing complete annotations in line with the consensus nomenclature. Contrary to the *NAT* genes of vertebrates, introns are commonly found within the homologous coding regions of lower eukaryotes. The NATs of fungi and higher animals are distinctly monophyletic, but evidence supports a mixed phylogeny of NATs among bacteria, protists and possibly some invertebrates.

1. Introduction

Arylamine *N*-acetyltransferases (NATs, E.C.2.3.1.5) are xenobiotic metabolizing enzymes extensively studied in a pharmacogenetic and toxicogenetic context, as their polymorphisms can modulate the biological effects of various drugs and carcinogens [1]. Substantial work has also been carried out with the NATs of various animal models and bacteria [2].

Population genetic studies have recently investigated the distribution of polymorphic human *NAT* haplotypes in relation with major demographic and dietary shifts in the history of our species [3,4]. The time is also ripe for investigations into the phylogenetic history of NATs across all major clades of life, exploiting the data generated by thousands of genome sequencing projects. The need for systematic annotation of *NAT* homologues in different species has been recognized, and detailed nomenclature guidelines have been discussed and published recently [5,6].

Here, we have expanded our previous genomic surveys [7,8], identifying and annotating newly described *NAT*-homologous genes in divergent taxonomic groups, such as the archaea, protists, fungi and invertebrates. We have used this extensive set of annotated genomic data to perform the first comprehensive phylogenetic analysis in the NAT field, encompassing all major taxa from bacteria, to animals. The study aims to provide a new perspective on the postulated roles and functions of NATs, as well as on the evolution of xenobiotic metabolism as a whole.

2. Methods

2.1. Retrieval and annotation of NAT genomic sequences

Genomic databases were accessed via Entrez of the US National Center for Biotechnology Information (NCBI).[1,2] Sequences were also retrieved from the Fungal Genome Initiative databases[3] of the Broad Institute. Initial searches were performed by tBLASTn, using the protein sequences of: (a) *Salmonella typhimurium* NAT1

Abbreviations: EMBL, European Molecular Biology Laboratory; EST, expressed sequence tag; HGT, horizontal gene transfer; MP, maximum parsimony; NAT, arylamine *N*-acetyltransferase; NCBI, National Center for Biotechnology Information; NJ, neighbor-joining; ORF, open reading frame; TPA, third-party annotation; UTR, untranslated region

* Corresponding author. Fax: +30 25510 30613.
E-mail address: sboukouv@mbg.duth.gr (S. Boukouvala).

[1] http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi.
[2] http://www.ncbi.nlm.nih.gov/genomes/leuks.cgi.
[3] http://www.broadinstitute.org/science/projects/fungal-genome-initiative/current-fgi-sequence-projects.

0014-5793/$36.00
doi:10.1016/j.febslet.2010.05.063

(GenBank ID: D90301) for genomic queries of prokaryotes, protists and plants, (b) *Gibberella monilliformis* NAT1 (alias FDB2; GenBank ID: EU552489) for fungi, protists and plants, and (c) human NAT1 (GenBank ID: X17059) for protists, plants and animals. We limited our searches for bacterial and animal *NAT* sequences to combined interrogations of all available genomes,[4,5] in order to update our previous catalogues of annotated *NAT* genes for these taxonomic groups [8]. The genomes of archaea,[1,4] protists[2,5] and fungi[2,3,5] were each one interrogated separately. Conventional (nBLAST) and tBLASTn queries of the GenBank complete nucleotide collection (nr/nt), as well as of the whole-genome shotgun and expressed sequence tag (EST) databases were additionally performed.

Annotation of the retrieved sequences involved reconstruction of the complete open reading frame (ORF) of each *NAT* gene, either by exploiting the locus information provided by the NCBI's Reference Sequence projects and the Broad Institute's Fungal Genome Initiative, or by running the GeneMark.hmm ES-3.0, FGenesh 2.6 and Geneid v.1.2 programs [9–11] with default settings for the appropriate organism category. Interrogation of the NCBI's Conserved Domain Database was helpful in many instances, while manual reconstruction of *NAT* ORFs was essential to overcome the limitations of the computational methods and verify their results. The annotation of the new *NAT* genes was validated by the presence, in the deduced amino acid sequences, of the characteristic semi-conserved VPFENL residues and the C-H-D catalytic triad, supported by successful alignment to pfam domain PF00797 ("acetyltransferase 2"). The *NAT* gene symbols were assigned according to current consensus recommendations[6] and the NAT Gene Nomenclature Committee was notified as required. Third-party annotation (TPA) numbers were provided by the European Molecular Biology Laboratory (EMBL) database.

2.2. Phylogenetic analysis of annotated NAT sequences

NAT amino acid sequences were determined based on standard code except for the unique translation codes of *Paramecium tetraurelia* and *Tetrahymena thermophila*. Highly divergent sequences (e.g. *Dictyostelium discoideum* NAT6), suspected pseudogenes, and ESTs were excluded from the analyses due to their various uncertainties. The protein sequences were aligned with ClustalX 2.0.10 [12], using default parameters and iteration at each alignment step. Alignments were only slightly modified by eye, if necessary.

Neighbor-Joining (NJ) and Maximum Parsimony (MP) analyses were performed using MEGA v4.1 [13]. For NJ, evolutionary distances were computed using the JTT matrix-based model and pairwise deletion of alignment gaps. MP analyses were conducted using the close-neighbor-interchange algorithm with search level 3 and initial trees obtained with random addition of sequences (10 replicates). Depending on the dataset, alignment gaps were either included or completely deleted from MP analyses. For both NJ and MP, 1000 bootstrap replications were performed to determine statistical support for clades.

To thoroughly evaluate the phylogenetic history of NAT sequences, we also performed Bayesian-Inference analyses of complete sequence alignments with MrBayes v3.1.2 [14], using mixed amino acid models with fixed rate matrices and at least 500 000 generations starting from an initial random tree. Two independent runs, with four simultaneous Markov chains per run (three heated, one cold), were performed per analysis. A tree was saved every 100th generation, and a burn-in sample of the first 1250 trees was discarded. The remaining trees were used to estimate branch lengths and posterior probabilities. All phylograms were edited using the MEGA tree editor.

3. Results and discussion

A total of 2445 (89 archaeal, 2019 bacterial, 45 protist, 146 fungal, 108 animal, 38 plant) genomes, representing 1532 (70 archaeal, 1210 bacterial, 43 protist, 97 fungal, 76 animal, 36 plant) species, were accessible via genomic BLAST at the time of our investigation and were searched for *NAT*-homologous sequences. A detailed account of the recovered genomic data concerning the distribution and genomic organization of *NAT* genes in those organisms is provided in Supplementary Tables S1–S5. EMBL database TPA accession numbers are also provided to enable access to the sequences of newly annotated *NAT* genes.

3.1. NATs in bacteria and archaea

In bacteria (Table S1), all newly retrieved *NAT* genes belonged to taxa previously characterized to contain such homologues [8]. During our phylogenetic analysis, the clustering of bacterial NATs did not deviate substantially from the consensus taxonomy (Fig. S1). In organisms with more than one *NAT* genes, there were examples of both paralogous (e.g. the two *NAT* genes of *Bradyrhizobium* sp., *Rhizobium etli* and *Novosphingobium aromaticivorans*, and the three *NAT* genes of the *Bacilli*) and independent (e.g. the *NAT* genes of *Mesorhizobium loti* and *Streptomyces griseus*) origin of the sister loci.

We demonstrate here for the first time the presence of *NAT* genes in archaea, their first representative found in the euryarchaeote *Halogeometricum borinquense* (Table S1 and Fig. S1). The archaeal NAT protein exhibited 30–40% match with the NATs of taxonomically diverse bacteria and certain fishes. This level of identity is expected for NATs belonging to distinct taxonomic groups, even within the same domain/kingdom of life [8, and this study].

3.2. NATs in protists and associated large DNA viruses

The existence of *NAT* genes is also demonstrated for the first time in certain taxa of the polyphyletic group of protists (Table S2). The presence of more than one *NAT* genes per genome was evident in most of the identified species, and these almost always represented recently diverged paralogues. Of remarkable exception were the NAT1, NAT2 and NAT3 paralogues of *D. discoideum*, which clustered with the two NAT paralogues of *Polysphondylium pallidum* (also a cellular slime mold), but also with the NAT paralogues of the *Bacillus* bacteria (Fig. 1). In contrast, two additional sequences (NAT4, NAT5) of *D. discoideum* appear to be of independent origin and paralogous to each other. These findings support the occurrence of horizontal gene transfer (HGT) between *Bacillus* and the cellular slime molds. In general, a large clade with limited internal resolution indicated a mixed phylogeny of NATs among bacteria and protists (Figs. 1 and S2).

Our pfam searches retrieved two paralogous *NAT*-like sequences belonging to the mimivirus of the amoeba *Acanthamoeba polyphaga* (Table S3), a giant virus carrying 1.2 Mb of genomic DNA with predicted 90.5% coding density. Phylogenetic analysis has placed this virus at the end of an independent branch, deeply rooted close to the origin of eukaryotes and most likely representing a distinct domain of life [15]. Our analysis suggested affinity of the mimivirus NATs to those of protists (Figs. 1 and S2), but our search for *NAT* genes in amoebozoa retrieved only one short EST from *Acanthamoeba castellanii* (Table S2), bearing limited similarity to the mimivirus NATs. Interestingly, the mimivirus genome has

[4] http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi.
[5] http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi?organism=euk.
[6] http://www.mbg.duth.gr/non-humannatnomenclature/background.html.

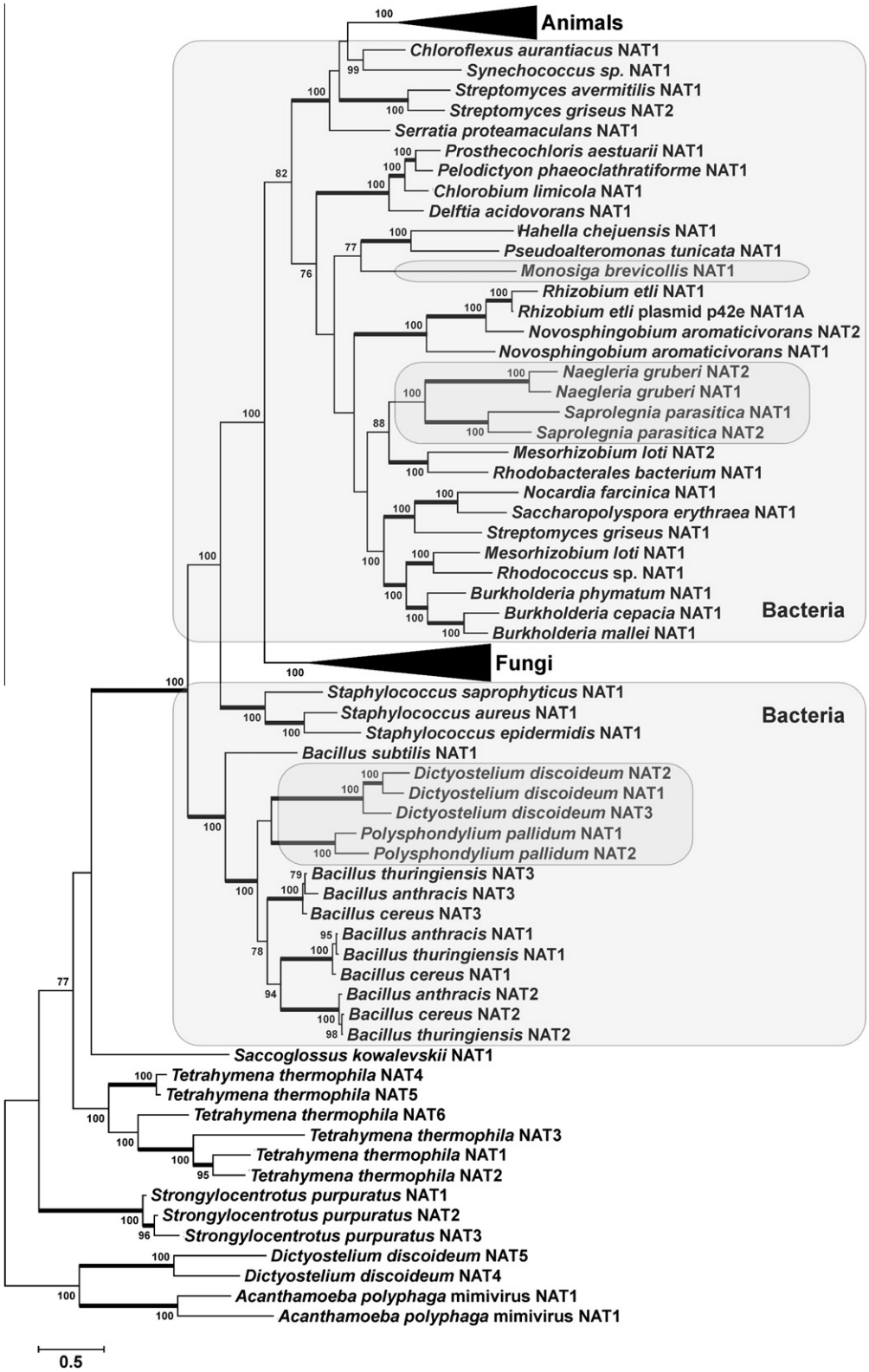

Fig. 1. Phylogenetic analysis of 115 prokaryotic and eukaryotic NAT amino acid sequences. The rooted phylogram resulting from Bayesian Inference is presented along with associated posterior probabilities (⩾75%). Bolded branches indicate bootstrap support (⩾90%) from Neighbor-Joining and Maximum Parsimony analyses. The fungal and animal clades were condensed to conserve space. Protist NATs with phylogenetic affinity to the bacterial NATs are shown circled within the two boxed areas of bacteria.

been reported to harbor two putative cytochrome P450 genes [16] which, together with the putative *NAT* genes identified here, could represent ancient forms of xenobiotic metabolizing enzymes.

Another interesting finding was the presence of a *NAT*-homologous gene in the choanoflagellates *M. brevicollis* and *M. ovata* (Table S2). These unicellular eukaryotes have the ability to form colonies and are considered as the sister group of metazoans [17]. The *M. brevicollis* NAT1 emerged in our analyses within the large collection of bacteria and protists phylogenetically situated between the fungal and animal clades (Fig. 1). However, the

M. brevicollis NAT1 amino acid sequence was not able to be resolved with strong statistical support, perhaps because of long-branch attraction (i.e. numerous unique amino acid changes to its sequence).

3.3. NATs in green algae and plants

A partial and highly divergent *NAT*-like sequence was found by pfam in the genome of the green alga *Chlamydomonas reinhardtii*

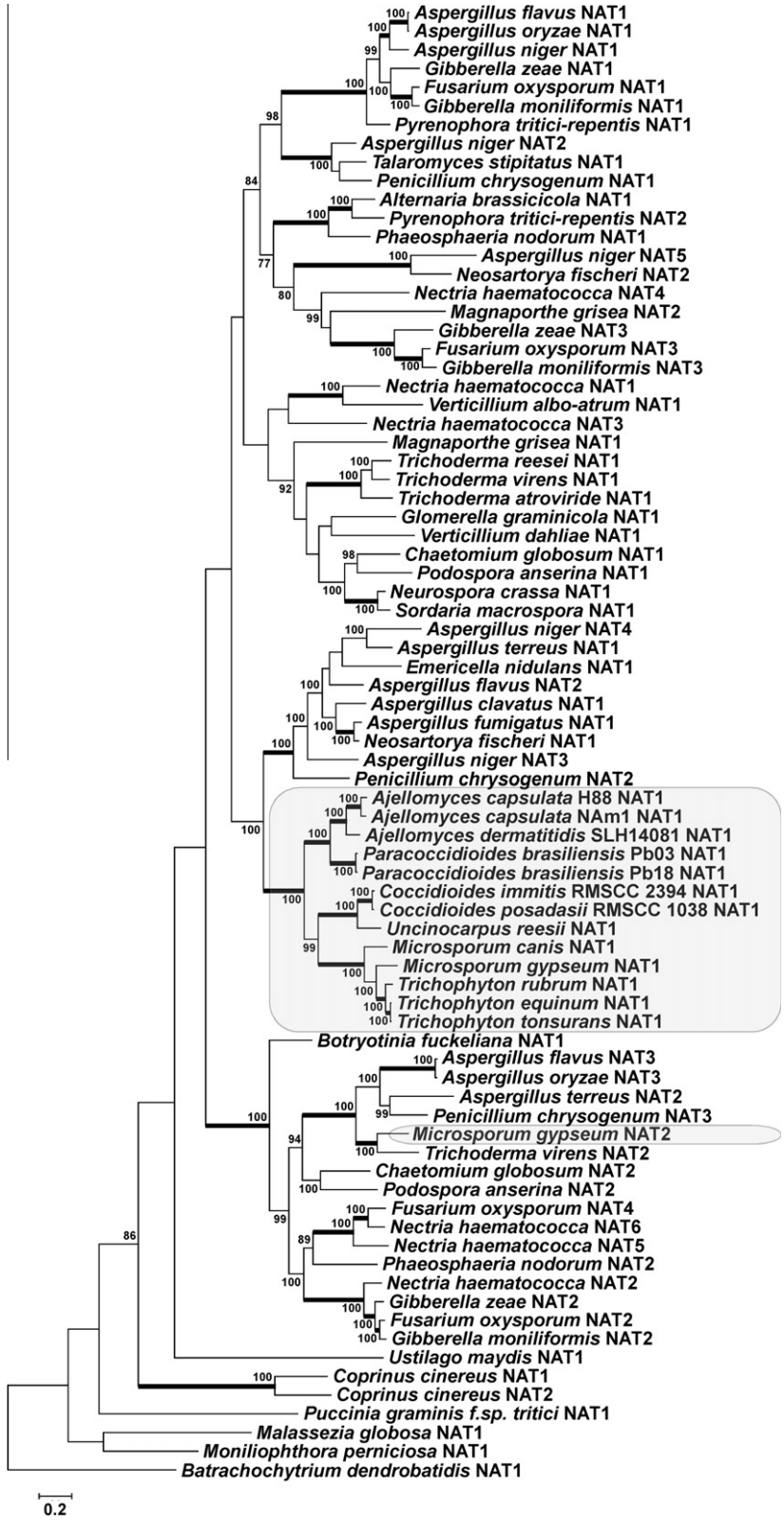

Fig. 2. Phylogenetic analysis of 79 NAT amino acid sequences from fungi. The phylogram is presented as in Fig. 1. The NATs of species typically pathogenic to humans or animals are circled.

(Table S2), an unexpected finding in view of the apparent lack of *NAT* genes in the 38 plant genomes interrogated. Our additional search of plant ESTs representative of 88 species retrieved only one partial sequence (dbEST ID: EY720533) in *Citrus sinensis*, but this could represent a bacterial contaminant of the library, as it bears very high similarity to the NATs of green sulfur bacteria. Overall, the available genomic data so far indicate complete absence of *NAT* genes in the kingdom of plants, suggesting that NATs were either never acquired or they were lost very early during plant evolution. Based on these observations, and the divergence of the *C. reinhardtii* sequence, the putative algal and citrus homologues were not included in our phylogenetic analyses.

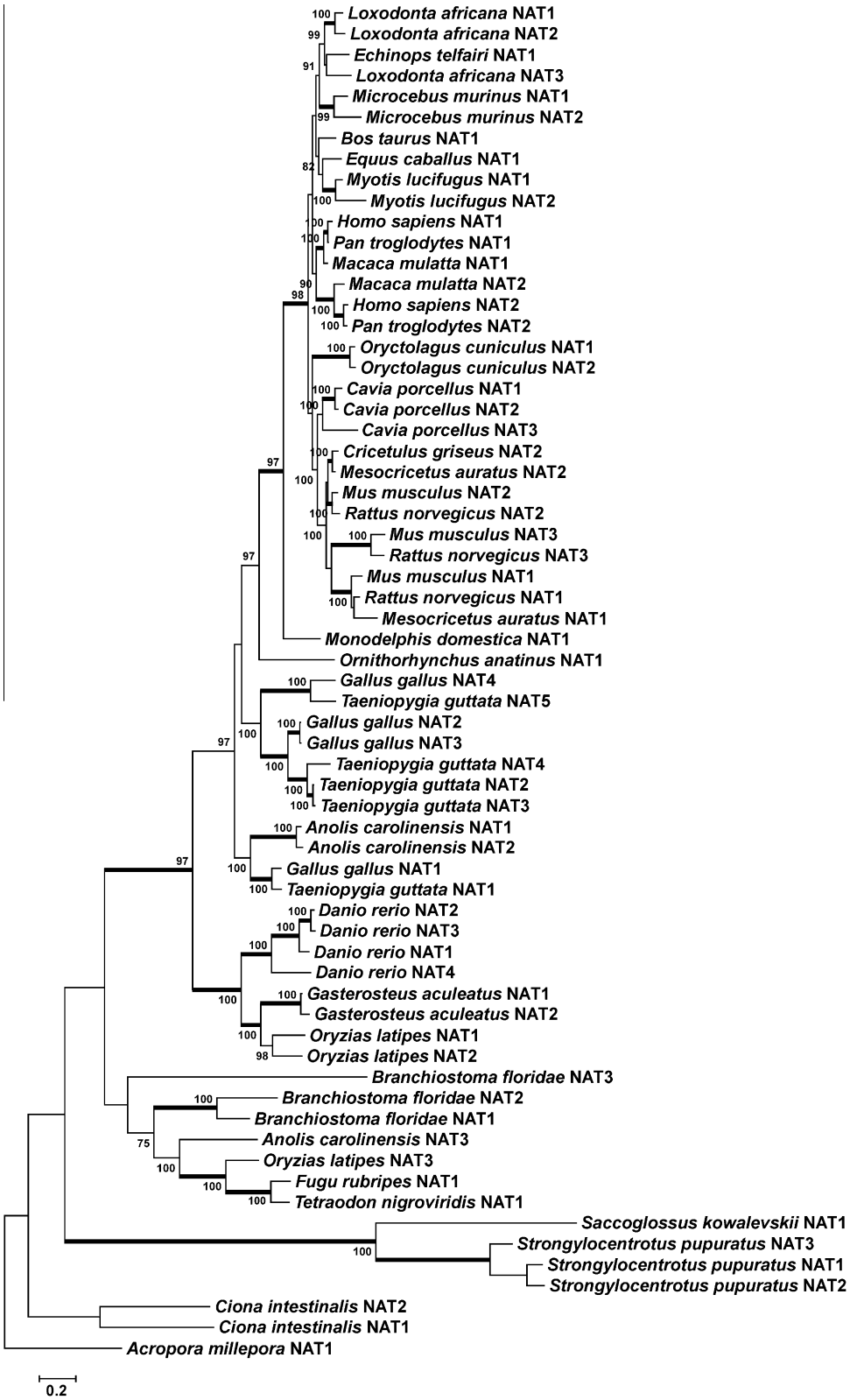

Fig. 3. Phylogenetic analysis of 65 NAT amino acid sequences from animals. The phylogram is presented as in Fig. 1.

3.4. NATs in fungi

Fungi have been demonstrated to possess *NAT* genes [8], and their enzyme products in the ascomycetes *G. monilliformis* and *Podospora anserina* were recently shown to transform xenobiotics [18,19]. Our search of 146 fungal genomes led to the annotation of 94 *NAT* genes from 62 genomes belonging to 1 chytridiomycete, 5 basidiomycetes and 39 ascomycetes (56 genomes). The latter belonged exclusively to the subphylum of *Pezizomycotina*. *NAT* genes were found neither in the remaining subtaxa of ascomycetes, nor in the phyla of microsporidians, blastocladiomycetes and zygomycetes. These findings were further supported by EST searches (Table S4).

The number of *NAT* genes per fungal genome varies considerably, with the ascomycetes described as typically pathogenic to animals or humans possessing only one NAT, while those infecting plants generally possessing multiple NATs (Table S4). The phylogenetic analysis clustered the NAT1 sequences of animal pathogens (species of *Ajellomyces*, *Paracoccidioides*, *Coccidioides, Microsporum* and *Trichophyton*) as a monophyletic sister clade to the monophyletic Eurotiales orthologues of *Aspergillus* and *Penicillium* (Fig. 2). Paralogy between the *NAT* genes of fungi was relatively limited, clearly evident between NAT1/NAT2 of *Coprinus cinereus*, as well as NAT1/NAT3 and NAT5/NAT6 of *Nectria haematococca*.

The overall pattern of the fungal NAT phylogeny supports distinct NAT lineages of orthologues, many of which are within plant pathogens (e.g. species of *Fusarium*, *Gibberella*, *Nectria*, *Magnaporthe* etc.). The NAT1, NAT2, and NAT3 lineages within *G. moniliformis* and related taxa are clearly evident. Also worth noting is the NAT2 of the animal pathogen *Microsporum gypseum*, which shows phylogenetic affinity to the NAT2 of the mycoparasite *Trichoderma virens*. The phylogenetic analyses did not support any substantial overlap between fungal and bacterial *NAT* genes, indicating that an exchange of *NAT* genes between these two groups was unlikely. A possible exception might be *Malassezia globosa* NAT1, which clustered among β-proteobacteria but with little statistical support, potentially representing an artifact of divergent sequences (data not shown).

3.5. NATs in the genomes of animals

Our exhaustive genomic analysis has also provided a picture of the distribution of *NAT* genes in the kingdom of animals (Table S5). Our previous genomic search [8] identified *NAT* genes only in deuterostomes, but we report here the presence of NATs in the lower phyla too, namely in cnidarians and potentially rotifers (various ESTs, Table S5). In contrast, the presence of NATs in protostomes remains uncertain; the search of 44 protostome genomes failed to identify any *NAT*-homologous sequences, but we were able to retrieve partial *NAT*-like ESTs of one mollusk (*Lottia gigantea*) and two arthropods (the crustacean *Litopenaeus vannamei* and the insect *Adelphocoris lineolatus*) (Table S5). Although the apparent sporadic incidence of *NAT* genes in these taxonomic groups could potentially be attributed to HGT (the genomes of rotifers, mollusks and insects have been demonstrated to host horizontally transferred genes [20,21]), the current genomic data is insufficient to support such speculations. Alternatively, the protostomes might employ bacterial endosymbionts to carry out detoxification of xenobiotics via *N*-acetylation.

In chordates, it has been possible to fully annotate *NAT* orthologues in the urochordate (tunicate) *Ciona intestinalis* and the cephalochordate (amphioxus) *Branchiostoma floridae*. NATs have been described previously in various ray-finned fishes (actinopterygii) and a range of mammals, as well as in the lizard *Anolis carolinensis* and the bird *Gallus gallus* [8]. Our genomic search has further identified five genes of the bird *Taeniopygia guttata*, while EST mining retrieved partial *NAT* sequences of additional actinopterygii, mainly of commercial significance. We also report the recovery of *NAT*-like ESTs of the sea lamprey *Petromyzon marinus* and the amphibia *Xenopus* (frog) and *Cynops* (salamander) (Table S5).

The phylogenetic relationships of chordate NATs (Fig. 3) followed the consensus taxonomy, and paralogous *NAT* genes appeared to be the outcome of recent genomic duplications. An exception was the clustering of *A. carolinensis* NAT1 and NAT2 with avian NATs, on the one hand, and the affinity of its NAT3 with certain NATs of fishes, on the other. Overall, our analysis indicates a monophyletic and fully supported clade of animal NATs, with the possible exception of the homologues of the hemichordate *Saccoglossus kowalevskii* and the echinoderm *Strongylocentrotus purpuratus*. When assessing just animal NAT sequences, these homologues represent a potential basal lineage to the vertebrates (Fig. 3), but when they are assessed more broadly with prokaryotic and eukaryotic sequences, they cluster among a polytomy of protists (Fig. 1). As with other examples noted above, this latter affinity may be an artifact due to numerous unique changes within their sequences, or it could represent an unresolved example of HGT.

3.6. Exon–intron structure of eukaryotic NAT genes

A major outcome of our genomic analysis and arduous reconstruction of hundreds of *NAT* ORFs was the discovery that many *NAT* genes possess coding regions that are segmented by introns (Tables S1–S5). Since the isolation from genomic libraries of their first mammalian representatives, the eukaryotic *NAT* genes have been considered to consist of intronless ORFs, often adjacent to segmented 5′ untranslated regions (UTRs) [7]. We now establish that 28% of the annotated *NAT* genes in protists and 78% of those in fungi harbor at least one intron, with some ORFs having as many as five exons. The highly divergent *NAT* gene of the green alga *C. reinhardtii* is exceptional in that it contains 12 or more introns. Intron-containing ORFs are also described for both *NAT* genes of the tunicate *C. intestinalis* and all three *NAT* genes of the amphioxus *B. floridae* (Table S5), suggesting that the loss and re-appearance of introns must have occurred independently more than once during the evolution of eukaryotic *NAT* genes. In many instances, we also identified putative upstream non-coding exons, presumably forming the 5′ UTR of *NAT* transcripts through intron splicing (data not shown).

4. Conclusions

Despite their considerable divergence and variability, xenobiotic metabolizing enzymes have persisted over evolutionary time, developing into elaborate biochemical systems that protect organisms from the toxic elements of their environment. Today, the sequencing of the genomes of hundreds of organisms presents scientists with unique opportunities for dissecting the evolutionary history of xenobiotic metabolism over a range of taxa inhabiting diverse environments. NATs are known to detoxify many harmful chemicals in humans and other mammals, while the tuberculous mycobacteria employ these enzymes to resist the lethal effect of isoniazid. More astoundingly, the NAT1/FDB2 enzyme of the corn pathogen *G. monilliformis* is involved in the detoxification of 2-benzoxazolinone, a natural toxin produced by the host plant as a defense against fungal infection [18]. This finding may have important implications in the management of agricultural environments. The microbial NATs could also prove useful in bioremediation processes. Our comparative genomic analysis has identified *NAT* genes in new taxonomic groups that include organisms of agronomical, biotechnological or ecological significance, and has looked into

the phylogenetic history of the NAT enzymes, an important component of xenobiotic metabolism.

Acknowledgements

We wish to acknowledge the NAT Gene Nomenclature Committee. We also thank Irene Vagena for supporting the training of E.P.K. and A.U. in data mining and analysis, and the Bodossaki Foundation (Greece) for funding.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.febslet.2010.05.063.

References

[1] Hein, D.W. (2009) *N*-Acetyltransferase SNPs: emerging concepts serve as a paradigm for understanding complexities of personalized medicine. Expert Opin. Drug Metab. Toxicol. 5, 353–366.
[2] Sim, E., Walters, K. and Boukouvala, S. (2008) Arylamine *N*-acetyltransferases: from structure to function. Drug Metab. Rev. 40, 479–510.
[3] Sabbagh, A., Langaney, A., Darlu, P., Gérard, N., Krishnamoorthy, R. and Poloni, E.S. (2008) Worldwide distribution of NAT2 diversity: implications for NAT2 evolutionary history. BMC Genet. 9, 21.
[4] Luca, F., Bubba, G., Basile, M., Brdicka, R., Michalodimitrakis, E., Rickards, O., Vershubsky, G., Quintana-Murci, L., Kozlov, A.I. and Novelletto, A. (2008) Multiple advantageous amino acid variants in the NAT2 gene in human populations. PLoS One 3, e3136.
[5] Boukouvala, S., Westwood, I.M., Butcher, N.J. and Fakis, G. (2008) Current trends in *N*-acetyltransferase research arising from the 2007 International NAT Workshop. Pharmacogenomics 9, 765–771.
[6] Hein, D.W., Boukouvala, S., Grant, D.M., Minchin, R.F. and Sim, E. (2008) Changes in consensus arylamine *N*-acetyltransferase gene nomenclature. Pharmacogenet. Genomics 18, 367–368.
[7] Boukouvala, S. and Fakis, G. (2005) Arylamine *N*-acetyltransferases: what we learn from genes and genomes. Drug Metab. Rev. 37, 511–564.
[8] Vagena, E., Fakis, G. and Boukouvala, S. (2008) Arylamine *N*-acetyltransferases in prokaryotic and eukaryotic genomes: a survey of public databases. Curr. Drug Metab. 9, 628–660.
[9] Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y.O. and Borodovsky, M. (2005) Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Res. 33, 6494–6506.
[10] Solovyev, V., Kosarev, P., Seledsov, I. and Vorobyev, D. (2006) Automatic annotation of eukaryotic genes, pseudogenes and promoters. Genome Biol. 7 (Suppl. 1), S10.1–S10.12.
[11] Blanco, E. and Abril, J.F. (2009) Computational gene annotation in new genome assemblies using GeneID. Methods Mol. Biol. 537, 243–261.
[12] Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.
[13] Tamura, K., Dudley, J., Nei, M. and Kumar, S. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24, 1596–1599.
[14] Ronquist, F. and Huelsenbeck, J.P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.
[15] Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Suzan, M. and Claverie, J.M. (2004) The 1.2-megabase genome sequence of mimivirus. Science 306, 1344–1350.
[16] Lamb, D.C., Lei, L., Warrilow, A.G., Lepesheva, G.I., Mullins, J.G., Waterman, M.R. and Kelly, S.L. (2009) The first virally encoded cytochrome p450. J. Virol. 83, 8266–8269.
[17] King, N., Westbrook, M.J., Young, S.L., Kuo, A., Abedin, M., Chapman, J., Fairclough, S., Hellsten, U., Isogai, Y., Letunic, I., Marr, M., Pincus, D., Putnam, N., Rokas, A., Wright, K.J., Zuzow, R., Dirks, W., Good, M., Goodstein, D., Lemons, D., Li, W., Lyons, J.B., Morris, A., Nichols, S., Richter, D.J., Salamov, A., Sequencing, J.G., Bork, P., Lim, W.A., Manning, G., Miller, W.T., McGinnis, W., Shapiro, H., Tjian, R., Grigoriev, I.V. and Rokhsar, D. (2008) The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. Nature 451, 783–788.
[18] Glenn, A.E. and Bacon, C.W. (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, 657–671.
[19] Martins, M., Rodrigues-Lima, F., Dairou, J., Lamouri, A., Malagnac, F., Silar, P. and Dupret, J.M. (2009) An acetyltransferase conferring tolerance to toxic aromatic amine chemicals: molecular and functional studies. J. Biol. Chem. 284, 18726–18733.
[20] Hotopp, J.C., Clark, M.E., Oliveira, D.C., Foster, J.M., Fischer, P., Torres, M.C., Giebel, J.D., Kumar, N., Ishmael, N., Wang, S., Ingram, J., Nene, R.V., Shepard, J., Tomkins, J., Richards, S., Spiro, D.J., Ghedin, E., Slatko, B.E., Tettelin, H. and Werren, J.H. (2007) Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. Science 317, 1753–1756.
[21] Bock, R. (2010) The give-and-take of DNA: horizontal gene transfer in plants. Trends Plant Sci. 15, 11–22.