Genomics 104 (2014) 45–57

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Phylogenetic and syntenic analyses of the 12-spanner drug:H+ antiporter family 1 (DHA1) in pathogenic *Candida* species: evolution of *MDR1* and *FLU1* genes

Paulo Jorge Dias, Isabel Sá-Correia *

IBB — Institute for Biotechnology and Bioengineering, Centre for Biological and Chemical Engineering, Department of Bioengineering, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais, 1049–001 Lisboa, Portugal

ARTICLE INFO

Article history:
Received 22 October 2013
Accepted 28 May 2014
Available online 5 June 2014

Keywords:
Multidrug resistance (MDR)
MFS transporters
Mdr1 and Flu1
Pathogenic *Candida* species
Phylogenetic and gene neighbourhood analyses

ABSTRACT

Candida albicans and other pathogenic *Candida* species can develop resistance to clinical fungicides through active drug export mediated by multidrug efflux pumps, in particular by members of the drug:H^+ antiporter family 1 (DHA1). The DHA1 proteins encoded in the genomes of 31 hemiascomycetous strains from 25 species were identified and homology relationships between these proteins and the functionally characterised DHA1 in the model yeast *Saccharomyces cerevisiae* were established. Gene neighbourhood analysis allowed the reconstruction of sixteen DHA1 lineages conserved during the CTG complex species evolution. The evolutionary history of *C. albicans MDR1* and *FLU1* genes and *Candida dubliniensis*, *Candida tropicalis* and *Candida parapsilosis MDR1* genes was detailed. *Candida* genomes show an abundant number of *MDR1* and *FLU1* homologues but the chromosome environment where *MDR1* homologues reside was poorly conserved during evolution. Gene duplication and loss are major mechanisms underlying the evolution of the DHA1 genes in *Candida* species.

© 2014 Elsevier Inc. All rights reserved.

1. Background

Multidrug resistance (MDR) is the simultaneous acquisition of resistance to a wide range of structurally and functionally unrelated cytotoxic chemicals [1]. Frequently, this phenomenon results from the activity of drug-efflux pumps [2,3]. Two types of efflux pumps have been associated with the development of MDR in the Hemiascomycetous yeasts: i) the ATP-Binding Cassette (ABC) proteins and ii) the Major Facilitator Superfamily (MFS) transporters [2,4–6]. According to the number of transmembrane spans (TMS), the MFS-MDR transporters belong to the drug: H^+ antiporter family 1 (DHA1) (12 TMS) or to the drug:H^+ antiporter family 2 (DHA2) (14 TMS) [2,6]. *S. cerevisiae* genome comprises twelve DHA1 genes: *QDR1*, *QDR2*, *AQR1*, *DTR1*, *QDR3*, *HOL1*, *TPO1*, *TPO2*, *TPO3*, *TPO4*, *YHK8* and *FLR1*, and the functional analysis of most of these transporters was already performed [2]. Although the use of *S. cerevisiae* as experimental benchmark has been important for understanding the genetic and biochemical basis of MDR, this yeast species is rarely pathogenic [7]. For this reason, *C. albicans* has been used as the main model to elucidate the mechanisms underlying the development of MDR in pathogenic yeasts.

* Corresponding author at: IBB — Institute for Biotechnology and Bioengineering, Centre for Biological and Chemical Engineering, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais, 1049–001 Lisboa, Portugal.

E-mail addresses: pjdias@tecnico.ulisboa.pt (P.J. Dias), isacorreia@tecnico.ulisboa.pt (I. Sá-Correia).

Candida species are commensal microorganisms that colonise different body sites in healthy hosts, especially the gastrointestinal tract, genital tract and the skin [8]. These yeasts are also opportunistic pathogens when an alteration in human microbial defences occur [8]. *Candida* infections account for around 10% of all nosocomial bloodstream infections [9]. The high mortality rate of *Candida* bloodstream infections (candidemia) is related to their ability to develop resistance against multiple antifungal agents [10]. Two DHA1 efflux pumps, Mdr1p and Flu1p, underlie the azole-resistance phenotype observed in *C. albicans* clinical isolates [4]. The *MDR1* gene was first described in a chemical screening for *C. albicans* genes involved in MDR [11]. Later, using the *MDR1* amino acid sequence as query sequence, *S. cerevisiae FLR1* gene was identified and shown to be the functional homologue of *MDR1* involved in fluconazole resistance [12]. Subsequently, other *MDR1* homologue genes were identified in the genomes of pathogenic fungi, including *C. dubliniensis* (*CdMDR1*), *C. tropicalis* (*CtMDR1*), *C. parapsilosis* (*CpMDR1*), *C. lusitaniae* (*ClMDR1*), *Aspergillus fumigatus* (*AfuMDR3*) and *Candida glabrata* (*CgFLR1*) [13–19]. The *FLU1* (fluconazole resistance 1) gene was identified in a screen for *C. albicans* genes that complemented the fluconazole hypersusceptibility of a *S. cerevisiae* Δ*pdr5* mutant strain [20]. *FLU1* amino acid sequence has strong sequence homology with *S. cerevisiae* Tpo1 transporter [21]. The *C. albicans TMP1* and *TMP2* (transmembrane protein) genes encode DHA1 transporters that confer resistance to cycloheximide, 4-nitroquinoline and 1,10-phenanthroline chemical compounds [22] and are required for virulence in the mouse model [23]. These two proteins are involved in the N-acetylglucosamine (NAG)

http://dx.doi.org/10.1016/j.ygeno.2014.05.005
0888-7543/© 2014 Elsevier Inc. All rights reserved.

catabolic pathway [24], being responsible for the uptake of this amino sugar [25]. In the Hemiascomycetes, the ability to use NAG as sole carbon source is an attribute of pathogenic *Candida* species [26]. Since the mucous membranes at the site of infection are rich in amino sugars and NAG is one of the known *in vitro* inducers of germ tube formation in *C. albicans* [8,27], thus promoting the yeast-to-hypha transition and yeast propagation into the host tissue, the NAG catabolic pathway is thought to be an important virulence factor [28]. Besides the four above referred efflux pumps, an *in silico* study focusing on the *C. albicans* MFS transportome identified eighteen additional ORFs showing amino acid sequence similarity to DHA1 proteins [29]. Recently, the *C. glabrata*, homologue of *S. cerevisiae QDR2* gene, *CgQDR2*, was found to confer resistance to the antifungal drugs clotrimazole, thioconazole, miconazole and ketoconazole [30].

The evolutionary history of the DHA1 proteins encoded in the genomes of thirteen hemiascomycete species was reconstructed in a previous study [31]. However, only the best annotated hemiascomycetous genomes available when this study was performed were used. This means that, with the exception of *Debaryomyces hansenii* and *Yarrowia lipolytica*, only the genomes of species taxonomically classified in the *Saccharomyces* complex [32] were examined. This impaired the inclusion of species translating CTG as serine instead of leucine, belonging to the phylogenetic complex (clade) where the majority of the pathogenic *Candida* species reside [33]. The goal of the present study was to extend the reconstruction of the evolutionary history of the hemiascomycete DHA1 genes to the species comprising the CTG phylogenetic complex. The objective was to get new insights into how this family of genes has evolved in *Candida* species, in particular the medically relevant *C. albicans MDR1* and *FLU1* genes and their homologues.

2. Materials and methods

2.1. Hemiascomycete yeast genomes

The translated ORFs of 31 sequenced hemiascomycetous strains, corresponding to 25 different species (14 of the *Saccharomyces* complex, 9 of the CTG complex and two early-divergent hemiascomycete species, *Pichia pastoris* and *Y. lipolytica*) were retrieved from genome databases (Table 1). Henceforth, the four letter code shown in Table 1 for species abbreviation will be used to designate both yeast genes and species. The letter displayed after the first four letters is used to abbreviate the strain name when the genome of more than one strain from a given species is available or when the genome of the same strain was sequenced by different research centres. Supplementary File 1 indicates the correspondence between the gene annotation used in this study and the original name of each ORF/gene analysed. To uniformise the annotation used, translated ORFs are always represented by lower case letters.

2.2. Sequence clustering of the translated ORFs

The comparative genomic approach used in this study is based on the sequence clustering of all translated ORFs of the 31 sequenced yeast strains examined. This required the compilation and organisation of a total of 172,422 translated ORFs. These translated ORFs were organised into a blast database and compared all-against-all using blastp algorithm made available in blast2 package [34]. The blastp algorithm used gapped alignment with the following parameters: expectation value (10^{-30}), open gap (-1), extend gap (-1), threshold for extending hits (11) and word size (3). This approach generated a total of 31 million pairwise alignments. In order to handle this amount of data, sequence clustering was formulated as a graph traversal problem, where the nodes are the translated ORFs and the edges indicate the existence of pairwise sequence similarity between amino acid sequences. Classification of the translated ORFs into clusters was achieved by breadth-first traversing this network at different e-value thresholds, ranging from E-30 to E-10.

2.3. Gene neighbourhood analysis

During this work, a MySQL genome database was built compiling genomic information regarding the previously mentioned translated ORFs, including gene name, chromosome/contig, sequence clustering classification, gene start position, gene end position, amino acid sequence length and encoded amino acid sequences. The package "sqldf" [35] and complementing scripting in R language was used to retrieve fifteen neighbour genes on each side of the query genes as well as the corresponding sequence clustering classification from this hemiascomycetous genome database. The rationale of synteny analysis resides on the assumption that two genes of different yeast species whose translation products belong to the same sequence cluster (homologues by similarity) will be members of the same gene lineage if they share at least one pair of neighbours that are also homologous to each other by similarity [31]. The process is reiterated for all possible heterospecific pairwise comparisons of homologues deduced from the sequence clusters. The sequence clustering classification of the thirty genes neighbouring each query gene was done using a conservative blastp e-value of E-30 to limit the number of false positive sequences incorporated together with true cluster members. When further evidences were needed to corroborate dubious synteny connections between genes, sequence clustering was performed at a less restrictive e-value threshold (E-15).

The chromosome neighbourhood of the query genes was converted into a format adequate for import into the Cytoscape environment [36]. In the resulting networks, nodes represent query genes and edges represent pairs of neighbouring genes classified in the same sequence cluster. Useful biological information indicated below was imported into the Cytoscape network as edges attributes. The existence of synteny between query genes was verified through the analysis of network topology (number of shared neighbour pairs) and the biological information associated with the corresponding edges. The advantage of this framework is that it allows scrutinising the synteny relationships established between genes in a simple mathematical context of network topology exploration. Three sources of biological information were used to assess the strength of each neighbour pair connection [31]: i) closeness of the connecting neighbours in relation to the query genes, ii) sequence similarity between connecting neighbours and iii) dimension of the sequence cluster to which the homologous neighbours belong; small dimension of the sequence cluster indicates that it is unlikely that two homologous neighbours are in the vicinity of two query genes by chance.

2.4. Topology prediction, sequence alignment and phylogenetic tree building

The topology of the amino acid sequences was analysed using HMMTOP and TMHMM 2.0 bioinformatic tools [37,38]. For each amino acid sequence not showing 12 TMS and with less than 450 residues in length, the TMS range was predicted by visual analysis of the topology probability and protein hydrophobicity plots generated by TMHMM 2.0 and TOPPRED 2 [38–40], respectively.

Multiple alignments of the amino acid sequences were calculated by MUSCLE [41] and processed using the PHYLIP package [42]. PROTDIST/NEIGHBOR and PROML packages were used to generate the phylogenetic trees based on distance and maximum likelihood methods, respectively. The Dendroscope application was used for tree visualisation [43].

Sequence identity and similarity shared between protein pairs was assessed using an all-against-all Needleman–Wunsch alignment approach. This algorithm was run using the needle package available in the EMBOSS suite [44]. All needle pairwise alignments made in this work used default values for the gap open and gap extension parameters, 10.0 and 0.5, respectively. After the construction of the DHA1 phylogenetic tree, all-against-all Needleman–Wunsch alignments were also constructed for the members of each phylogenetic cluster.

Download English Version:

https://daneshyari.com/en/article/2820603

Download Persian Version:

https://daneshyari.com/article/2820603

Daneshyari.com