The ASP3 locus in Saccharomyces cerevisiae originated by horizontal gene transfer from Wickerhamomyces

Garrett P. League, Jason C. Slot & Antonis Rokas
Department of Biological Sciences, Vanderbilt University, Nashville, TN, USA

Correspondence: Antonis Rokas, Department of Biological Sciences, Vanderbilt University, VU Station B #35-1634, Nashville, TN 37235, USA. Tel.: +1 615 936 3892; fax: +1 615 343 6707; e-mail: antonis.rokas@vanderbilt.edu

Received 26 April 2012; revised 11 June 2012; accepted 3 July 2012. Final version published online 31 July 2012.

DOI: 10.1111/j.1567-1364.2012.00828.x

Editor: Cletus Kurtzman

Keywords asparagine metabolism; horizontal gene transfer; gene duplication; metabolic diversity; phylogenetics.

The asparagine degradation pathway in the S288c laboratory strain of Saccharomyces cerevisiae is comprised of genes located at two separate loci. ASP1 is located on chromosome IV and encodes for cytosolic  \( \text{ASP} \) I-asparaginase, whereas ASP3 contains a gene cluster located on chromosome XII comprised of four identical genes, ASP3-1, ASP3-2, ASP3-3, and ASP3-4, which encode for cell wall-associated  \( \text{ASP} \) II-asparaginase II. Interestingly, the ASP3 locus appears to be only present, in variable copy number, in S. cerevisiae strains isolated from laboratory or industrial environments and is completely absent from the genomes of 128 diverse fungal species. Investigation of the evolutionary history of ASP3 across these 128 genomes as well as across the genomes of 43 S. cerevisiae strains shows that ASP3 likely arose in a S. cerevisiae strain via horizontal gene transfer (HGT) from, or a close relative of, the wine yeast Wickerhamomyces anomalus, which co-occurs with S. cerevisiae in several biotechnological processes. Thus, because the ASP3 present in the S288c laboratory strain of S. cerevisiae is induced in response to nitrogen starvation, its acquisition may have aided yeast adaptation to artificial environments. Our finding that the ASP3 locus in S. cerevisiae originated via HGT further highlights the importance of gene sharing between yeasts in the evolution of their remarkable metabolic diversity.

The asparagine degradation pathway in the S288c laboratory strain of the budding yeast Saccharomyces cerevisiae is comprised of five asparaginase (ASP) genes located at two separate loci. Specifically, a single copy of ASP1 is located on chromosome IV and four tandemly duplicated copies of ASP3 are located on chromosome XII (Jones, 1977a; b; Dunlop et al., 1978; Sinclair et al., 1994). ASP1 and ASP3 encode for the enzymes cytosolic  \( \text{ASP} \) I-asparaginase and cell wall  \( \text{ASP} \) II-asparaginase II, respectively, which hydrolyze L-asparagine as well as, in the case of ASP3, D-asparagine to aspartate and ammonia (Dunlop et al., 1978). Whereas ASP1 is constitutively expressed in the cell, ASP3 is upregulated during nitrogen starvation to facilitate the utilization of extracellular asparagine as a source of nitrogen (Jones, 1973; Jones & Mortimer, 1973; Dunlop & Roon, 1975).

The two yeast ASP loci are evolutionarily related to bacterial ASPs, which in Escherichia coli can also be distinguished into the cytosolic ASP I and the cell wall ASP II (Sinclair et al., 1994). However, both the yeast and the bacterial ASPs are more similar to each other than to their counterparts in the other organism, suggesting that eukaryotic and bacterial ASPs likely evolved via independent duplication events (Sinclair et al., 1994). To better understand the origin and evolution of the ASP3 locus, we first examined the synteny conservation of the two loci using the Yeast Gene Order Browser (Byrne & Wolfe, 2005). Whereas synteny is highly conserved at the ASP1 locus across yeast species, the ASP3 locus is uniquely present in S. cerevisiae and lacks any conservation. Further TBLASTN analysis (Altschul et al., 1997) of the genomes of 128 fungal species spanning the diversity of the kingdom using the ASP3-1 protein sequence as a query confirmed that the ASP3 gene cluster is unique to S. cerevisiae. Finally, examination of the genomes of 43 different S. cerevisiae strains showed that the ASP3 locus appears to be only present, in variable copy number, in 11 strains isolated from laboratory or industrial environments (Table 1).

We next constructed phylogenies of the homologs of the ASP1 and ASP3 loci, using a previously described approach (Slot & Rokas, 2010, 2011; Campbell et al.,...
Candida lineages (Fig. 1). In contrast, the theologs of the fied two robustly supported sister clades comprised of or-

ASP3 and mely long branches, as well as of distantly related homo-

poorly aligned and partial sequences that generated extre-

and 100 rapid bootstrap replicates. Following removal of

of amino acid substitution (Whelan & Goldman, 2001)

the non-

ported clade largely comprised of protein sequences from

the non-

S. cerevisiae strains (S288c, W303, 505, 1187, and 680) contain four copies of the ASP3 gene, six strains (2340, JAY270, CBS7960, CLIB324, FL100, and T73) contain a single copy, and a single strain (1056) contains two copies. Yeast strains containing the

locus largely comprised of protein sequences from

UM218 No Vineyard

UM237 No Vineyard

BB1235 No Vineyard

BB2453 No Vineyard

Lalvin EC-1118 No France

Lalvin ICV D254 No France

IOC 18-2007 No France

AEB Fermol Rougue No France

Davis Lalvin 522 No California

Y55 No Vineyard

YIM789 No Clinical

YAT7 No Clinical

RM3 No Vineyard

RM8 No Vineyard

S288c Yes (4) Laboratory

J940047 No Clinical

J940557 No Clinical

J940915 No Clinical

O6L3FF02 No Wine cellar

O6L1FF11 No Wine cellar

O6L3FF15 No Wine cellar

O6L6FF20 No Wine cellar

UM218 No Vineyard

UM237 No Vineyard

BB1235 No Vineyard

BB2453 No Vineyard

Lalvin EC-1118 No France

Lalvin ICV D254 No France

IOC 18-2007 No France

AEB Fermol Rougue No France

Davis Lalvin 522 No California

Y55 No Vineyard

YIM789 No Clinical

YAT7 No Clinical

RM3 No Vineyard

RM8 No Vineyard

ASP3 (copy #) Source

W303 Yes (4) Laboratory

505 Yes (4) Brewing

1187 Yes (4) Brewing

680 Yes (4) Brewing

2340 Yes (1) Brewing

1332 No No

1056 Yes (2) Brewing

530 No No

EC1118 No Wine

JAY270 Yes (1) Bioethanol

CBS7960 Yes (1) Bioethanol

CLIB215 No No

CLIB324 Yes (1) Baker's Yeast

CLIB382 No No

FL100 Yes (1) Laboratory

PW5 No Wine

T73 Yes (1) Wine

T7 No Oak tree

UC5 No No

Y10 No No

YJM269 No No

Grapes

S. cerevisiae strain project genomes; **Argueso et al. (2009); §§http://genome.wustl.edu/genomes/saccharomyces_ cerevisiae_strain_project_genomes; ††Lashkari et al. (1997); ††Wizneler et al. (1999)).

2012). Briefly, we used S. cerevisiae strain S288c ASP1 and ASP3 protein sequences to retrieve homologs from the nr database in GenBank using BLASTP (Altschul et al., 1997), we then grouped into gene families using OrthoMCL (Li et al., 2003). The resulting, nearly identi-

groups of 85 ASP1 and ASP3 homologous proteins were merged and aligned using MAFFT (Katoh & Toh, 2008). The sequence alignment was then used to con-

struct a maximum likelihood (ML) phylogeny in RAxML (Stamatakis, 2006), using a WAG + GAMMA + F model of amino acid substitution (Whelan & Goldman, 2001) and 100 rapid bootstrap replicates. Following removal of poorly aligned and partial sequences that generated extremely long branches, as well as of distantly related homo-

logs, we constructed a final ML phylogeny of the 30 ASP1 and ASP3 homologs found in Saccharomyces yeasts (Fig. 1).

Reconstruction of the phylogeny of the ASP loci identi-

two robustly supported sister clades comprised of or-

thologs of the ASP1 locus across the Saccharomyces and Candida lineages (Fig. 1). In contrast, the ASP3 sequences were nested within a different and also robustly sup-

ported clade largely comprised of protein sequences from the non-Saccharomyces wine yeast Wickerhamomyces anomalous (formerly known as Pichia anomala and Hansenula anomala) (Kurtzman, 2011; Schneider et al., 2012) (Fig. 1). Examination of ASP homologs with SignalP (Petersen et al., 2011) showed that none of the ASP1 sequences from the Saccharomyces and Candida lineages were predicted to contain signal peptides, consistent with their function in the cytosol. In contrast, the S. cerevisiae ASP3 sequences as well as two of the three W. anomalus sequences that grouped most closely to ASP3 were predicted to contain signal peptides, as expected from their function on the cell wall, whereas the third W. anomalus sequence was marginally nonsignificant for the presence of a signal peptide (Fig. 1). The unique presence of the ASP3 gene cluster in S. cerevisiae, the strong grouping of ASP3 with sequences containing putative signal peptides from the distantly related W. anomalus, and the co-occurrence of these two species in a variety of industrial niches (Daniel et al., 2011; Walker, 2011; Arroyo-Lopez et al., 2012), strongly suggest that S. cerevisiae acquired ASP3 via horizontal gene transfer (HGT) from W. anomalus or a close relative.

To test further the inference of HGT involvement in the making of the S. cerevisiae ASP3 locus, we forced ASP3 sequences to group with Saccharomyces and
Candida ASP1 sequences (Fig. 1) and used the Shimodaira–Hasegawa (SH) test (Shimodaira & Hasegawa, 1999; Stamatakis, 2006) to determine whether the ML estimate of the ASP phylogeny under this constraint differed significantly from the ML estimate of the unconstrained ASP phylogeny. The SH test showed a significant difference in the likelihood scores between the two phylogenies ($D_L = 127.61$; $P$-value < 0.0001), thus providing further support for the finding that ASP3 originated via HGT from Wickerhamomyces anomalus.

ASP3 is not only capable of acting on extracellular substrates that are inaccessible to ASP1 (Dunlop & Roon, 1975; Dunlop et al., 1978) but its expression and secretion is promoted during nitrogen starvation (Bon et al., 1997).

**Fig. 1.** Phylogeny of the yeast ASPs shows that Saccharomyces cerevisiae ASP3 originated via HGT from Wickerhamomyces anomalus. The ML topology recovered from the analysis of the 30 closest homologs of the S. cerevisiae ASP1 and ASP3 sequences is shown. Values near internodes correspond to bootstrap support values (only the values for key branches are shown) and branches lengths are in substitution/site units. Gene names for species with two or more sequences are shown below the species name. Note that the S. cerevisiae ASP1 sequence is nested within the Saccharomyces lineage, whose sister lineage is comprised of Candida species. In contrast, the four identical S. cerevisiae ASP3 gene copies are most closely related to ASP homologs from the wine yeast W. anomalus. The gene names of the six sequences predicted to contain signal peptides, and thus likely to function on the cell wall, are marked with two asterisks (**), whereas the gene name of the W. anomalus sequence whose signal peptide prediction was marginally nonsignificant is marked with a single asterisk (*). All other sequences show no evidence of containing a signal peptide and are likely functioning in the cytosol.
Furthermore, unlike ASP1, ASP3 is capable of hydrolyzing D-asparagine (Dunlop et al., 1978), and ASP3 has also been linked to the ability of *S. cerevisiae* strains S288c and W303 to utilize dipeptides with C-terminal asparagine residues, further expanding the list of potential substrates for ASP3 (Homann et al., 2005). Our finding that each of the *S. cerevisiae* strains where ASP3 is found was isolated from laboratory or industrial environments (Table 1) coupled with the fact that some of ASP3s known substrates are prevalent in industrial processes utilizing yeasts (Zagon et al., 1994; Pope et al., 2007), suggests that cell wall ASP3 function might be particularly favorable in human-constructed ecological niches.

The finding that the *S. cerevisiae* ASP3 was likely acquired through HGT from *W. anomalus* or a close relative adds to a growing body of evidence on the adaptive advantages of HGT in yeasts and fungi in general (Slot & Hibbett, 2007; Khaldi et al., 2008; Novo et al., 2009; Slot & Rokas, 2010, 2011; Andersen et al., 2011; Khaldi & Wolfe, 2011; Campbell et al., 2012), as well as sheds light to the types of environments that might favor HGT. For example, Novo et al. (2009) recently identified three separate events that introduced exogenous genes into the genome of the wine yeast *S. cerevisiae* EC1118, including one in which *Zygosaccharomyces bailii*, a major contaminant of wine fermentations, was identified as the donor species. Similarly, the wide use of *W. anomalus* in a variety of biotechnological applications, including ones that involve *S. cerevisiae* such as wine making, baking, and brewing (Walker, 2011; Schneider et al., 2012), suggests that human-constructed niches for yeasts, such as fermentation vats, might offer ample ecological opportunity for HGT and metabolic gene sharing between distantly related yeast species and strains.

**Acknowledgements**

This work was conducted in part using the resources of the Advanced Computing Center for Research and Education at Vanderbilt University. This work was supported by funds provided by the Searle Scholars Program (A.R.) and the National Science Foundation (DBI-0805625 to J.C.S. and DEB-0844968 to A.R.).

**References**


Horizontal transfer of the ASP3 gene cluster 863


