Genomic data mining reveals the transaminase repertoire of *Komagataella phaffii* (*Pichia pastoris*) strain GS115 and supports a systematic nomenclature

Maria José Henríquez, Rosalía Paula Cardós-Elena, Darren Nicholas Nesbeth*

*Corresponding author: Department of Biochemical Engineering, University College London, Bernard Katz Building, London WC1E 6BT, United Kingdom.

Tel: +44 (0) 20 7679 9582, Fax: +44 (0) 207 916 3943

Email: d.nesbeth@ucl.ac.uk

Keywords: Transaminase; Aminotransferase; bioinformatics; genome; phylogeny; nomenclature

ABSTRACT

Transaminases are an industrially important class of enzyme due to their ability to catalyse amination reactions for production of chiral amines, key building blocks of small molecule pharmaceuticals. We analysed the genome of strain GS115 of the methylotrophic yeast *Komagataella phaffii* (*K. phaffii*), formerly known as *Pichia pastoris* (*P. pastoris*), to identify transaminase genes and propose a systematic nomenclature based on both phylogeny and structuro-functional features. *K. phaffii* is an increasingly attractive industrial host cell due to its ability to grow to high biomass, up to 60% wet cell weight by volume, using methanol as carbon source and inducer of transgene expression. 39 UniProt database hits were reduced to 19 on the basis of sequence similarity and Hidden Markov Modelling. Of the 19 genes, the open reading frames of three (KpTam I-II.1b, KpTam I-II.7 and KpTam V.2) had strong homology with no characterised protein and four (KpTam III.1a, KpTam III.1b, KpTam III.2a and KpTam III.2b) had relatively high sequence similarity to ω-type transaminases, a subtype that typically accepts the broadest range of substrates. Comparison with *Saccharomyces cerevisiae* S288C suggested functions for KpTam I-II.1b and KpTam I-II.7. *K. phaffii* GS115 was originally generated by mutagenesis of *K. phaffii* CBS7435 and comparison revealed that one transaminase gene may have been deleted during this mutagenesis. These insights can advance fundamental understanding of yeast biology and can inform industrial screening and engineering of yeast transaminases.

Keywords: *Komagataella phaffii*, *Pichia pastoris*, transaminase, Hidden Markov; standard nomenclature

INTRODUCTION

Transaminases, also referred to as aminotransferases, are homodimeric enzymes that use pyridoxal-5´-phosphate (PLP) as a cofactor in catalysing the transfer of an amino group from a donor molecule, typically an amino acid, to the oxo- or carbonyl group of a keto-acid. In cellular metabolism transaminases play an important role in nitrogen utilisation and in synthesis of amino acids, vitamins, bacterial cell walls and antibiotics. As such, transaminases are ubiquitous in nature and the genomes of most microorganisms encode several of them (Ward and Wohlgemuth 2010). The Enzyme Commission (Bairoch 2000) has defined 112 transaminase types, assigning Enzyme Commission (EC) numbers 2.6.1.X to each type based on the reaction catalysed.

Mehta *et al*. (1993) analysed the homology and hydropathy patterns of 51 transaminase amino acid sequences to map out their structural relatedness. From this analysis they suggested four overall classes of transaminase existed, based on substrates preference, and concluded all transaminases had originated from the same ancestor. The growth in sequence data in the post-genomic era greatly increased the number of putative transaminase sequences available and Hwang *et al*. (2005) used protein sequences from the Pfam database (Finn et al. 2016) to suggest classification of transaminases into six classes (I-VI), within five subgroups (Table 1).

Transaminases can also be classified according to the position of an acceptor keto- or aldehyde group relative to the carboxylic acid or other major group of the amino donor molecule. Class I, II, IV and V transaminases are α-transaminases, as they transfer the amino group from the α position. Class III contains β-transaminases, γ-transaminases, and ω-transaminases. Class III transaminases are especially sought after as they have the widest substrate spectrum. Class III, ω-type transaminases are the only enzyme known to perform stereoselective amination of

ketones. As a result Class III, ω-type transaminases, have been investigated extensively for industrial production of amino acids, chiral amines and amino alcohols, all of which are valuable key intermediates for chemical synthesis of chiral, small-molecule therapeutics (Malik et al. 2012).

Class III, ω-type transaminases are highly active for a given substrate and also stable to a broad range of pH, substrate concentration, temperature and product concentration. Class III, ω-type transaminases with industrial promise have been identified by Shin *et al*. (2003) in *Vibrio fluvialis* JS17 (*V. fluvialis*), by Yonaha *et al*. (1992) in *Pseudomonas putida (P. putida)* and by Kaulmann *et al*. (2007) in *Chromobacterium violaceum (C. violaceum).*

These transaminases have also been co-expressed with other exogenous enzymes in a common host cells, such as *Escherichia coli* (*E. coli*), to provide *de novo* multi-step, whole cell biosynthetic pathways (Kaulmann et al. 2007). Cho *et al*. (2003) coupled two transaminase reactions (α and ω) for whole cell kinetic resolution of chiral amines. Ingram *et al*. (2007) achieved whole cell asymmetric synthesis of a chiral amino alcohol using co-overexpression of transketolase and transaminase in *E. coli*.

In addition to experimental screening, bioinformatic data-mining has emerged as a successful route to discovery of novel transaminases and prediction of their activity and substrate specificities (Hohne et al. 2010; Valli et al. 2016). Statistical techniques such as hidden Markov modelling (HMM), can be used to predict whether a given transaminase belongs to a particular classification using a temporal pattern recognition which enables the creation of protein structure profiles that may discriminate whether an input sequence belongs to a protein family or subtype, despite apparent non-significant sequence homology (Krogh et al. 1994). Using HMMER software (Finn et al. 2015) it is possible to search profile databases for sequence homologs employing Hidden Markov models.

The methylotrophic yeast *P. pastoris* was reassigned to the genus *Komagataella* following phylogenetic analysis (Kurtzman 2005), and the major strains split into three species: *Komagataella pastoris*, *K. phaffii GS115* (formerly *P. pastoris* strain GS115), and *Komagataella pseudopastoris*. *K. phaffii GS115* is a reliable and robust expression system (Invitrogen 2010) that has become widely used for production of recombinant protein in research and industrial settings (Templar et al. 2016) with well-established tools (Bollok et al. 2009) for genetic manipulation, strong native promoters to direct overexpression of transgenes and the ability to grow rapidly to high cell densities, up to 60% wet cell weight (wcw) by volume, on chemically defined culture media (Wei et al. 2018). Furthermore, *K. phaffii* GS115 is thermotolerant, able to grow at 47°C (Van der Klei et al. 2006), and it is tolerant to pH3-7 (Macauley-Patrick et al. 2005).

The availability of the complete genome sequence (De Schutter et al. 2009; Mattanovich et al. 2009) has made bioinformatic data-mining for transaminases possible in *K. phaffii* GS115. Here we attempt to predict the function of transaminase using sequence analyses and to assign each identified transaminases to a Hwang subclasses (Hwang et al. 2005) using HMM analysis. We also propose a rational nomenclature for *K. phaffii* GS115 transaminases. This information will assist future investigators who wish to exploit or rationally design transaminases for enhanced stability, substrate specificity, PLP binding and other properties.

MATERIAL AND METHODS

Identifying transaminase genes in yeast genome sequences

The UniProt database (The Uniprot Consortium 2015) was searched using first, the terms ‘pichia pastoris’, ‘gs115’ and ‘GO:0008483’ (gene ontology term for transaminase/aminotransferase activity); and then, replacing the GO terms by the

keywords ‘transaminase’ and ‘aminotransferase’ to generate a wider pool of initial sequences. JalView v.2.8.2 (Waterhouse et al. 2009) was then used to remove duplicate results by examining identity percentages and pairwise alignments. The first round of UniProt hits was further refined using the terms “pyridoxal binding site”, “aminotransferase” and “transaminase”. The nucleotide sequence of each transaminase identified in UniProt was obtained from the corresponding NCBI entry. The methods above were also used to identify transaminase genes in the published genome sequences of *K. phaffii* CBS7435 (Küberl et al. 2011) and *S. cerevisiae* S228C (Goffeau et al. 1996).

Assembly of transaminase dendrograms and phylograms

Protein sequences were aligned using ClustalX v.2.1 (Larkin et al. 2007) and a phylogenetic tree (Qian and Goldstein 2003) assembled using the neighbour-joining clustering algorithm and 1000 bootstrap replicates. The alignments were used to generate a tree diagram using TreeView v.1.6.6 (Page 1996) and dendrograms plotted using Dendroscope v.3.3.2 (Huson and Scornavacca 2012). Dendrogram images were edited for graphical brevity, to indicate bootstrap values and Hwang subgroup.

Assigning transaminases to a Hwang subgroup

Hits from the above UniProt search were then used to query the HMM database HMMER (Finn et al. 2015), provided by the European Molecular Biology Organisation - Biology European Bioinformatics Institute (EMBO-EBI) website (www.ebi.ac.uk/Tools/hmmer/search/hmmscan), by accession number. Predictions were deemed acceptable if the score was positive and expectation values (E-values) lower than 0.001. All protein families available on the search tool (Pfam, TIGRFAM, Gene3D, Superfamily and PIRSF) were selected for the search and those transaminases scoring highest for a HMM model pertaining to a particular subgroup (Hwang et al. 2005) were assigned to that subgroup.

Alignment of ω-transaminases

V. fluvialis JS17 and *C. violaceum* DSM30191 ω-transaminase sequences were aligned with the four putative *K. phaffii* GS115 Class III transaminases using ClustalW. The alignment file was inserted in JalView v.2.8.2 and residues colour-coded according to their BLOSUM62 score (Henikoff and Henikoff 1992). Secondary structure elements were generated in ESPript 3.0 (Robert and Gouet 2014). Pairwise alignments were performed in JalView v.2.8.2 to determine protein-to-protein identities. Analyses of *K. phaffii* CBS7435, *S. cerevisiae* S288C and *K. phaffii* GS115 Class III transaminases with respect to similarity were performing using protein BLAST.

RESULTS AND DISCUSSION

Analysis and classification of *K. phaffii* GS115 transaminases

A total of 39 potential transaminases/aminotransferase genes were identified in an initial search of the *K. phaffii* GS115 genome using the UniProt webtool (Table 2). A second UniProt search was performed within the 39 genes to rule out duplicates arising from the tautological terms ‘aminotransferase’ and ‘transaminase’. This identified 17 duplicate entries, which were then removed.

HMM screening with the HMMER tool was used to identify sequences for which a HMM model exists that is associated with a given transaminase subgroup, as set out by Hwang *et al.* (2005). This procedure revealed three proteins, C4R864, C4R277 and C4R194, which did not match any transaminase class by HMM profiling. UniProt entries for these three proteins also did not contain the search term ‘pyridoxal-5´-phosphate’. Future investigation of the structure of these proteins may resolve their status as transaminases.

A standard nomenclature for *K. phaffii* GS115 transaminases

The search methodology described above extracted 19 putative transaminases from the *K. phaffii* GS115 genome (Table 2). We used ClustalX v.2.1 to align the sequences and TreeView v.1.6.6 to make a tree diagram to illustrate the relatedness of the sequences and plotted a dendrogram (Figure 1). We proposed a systematic nomenclature for the 19 transaminases based on abbreviating ***K****omagataella* ***p****haffii* to 'Kp' and ***t***rans***am***inase to 'Tam' in 'KpTam', followed by modifiers that encompass structural and functional predictions of the Hwang *et al.* (2005) subgroupings (Table 2) and phylogenetic relationships (Figure 1). The modifier features firstly the Hwang subgroup, based on our HMMER analysis (Table 2), secondly phylogenetic branching within a Hwang subgroup and thirdly pairings, where they exist, of proteins of high similarity within a branch (Figure 1). For example, in 'KpTam III.2a', the 'III' refers to the Hwang subclass III, '2' indicates that the protein sequence falls within the second of at least two phylogenetic branches of proteins within Hwang subgroup III and the 'a' indicates the protein is one of a pair (a and b) of closely related proteins within type 'III.2'. In most cases closely related sequences appear also to be related with respect to function, where characterisation data has been reported (Table 2).

Putative *K. phaffii* GS115 Class III ω-transaminases

Currently HMMER (Finn et al. 2015) does not encompass sufficient mechanistic data to meaningfully predict the topology and chemistry of the active site of a given transaminase. As such, although HMMER analysis can be used to predict the Hwang subgroup of a given transaminase it cannot predict mechanistic information such as whether a given transaminase is of the β-, γ- or ω- type. Transaminases of the ω- type mechanism from *V. fluvialis* JS17 (Yonaha et al. 1992) and *C. violaceum* DSM30191 (Kaulmann et al. 2007) have been investigated mechanistically and shown to have industrially favourable substrate ranges. We

identified four putative Class III transaminases in *K. phaffii* GS115: KpTam III.1a, KpTam III.1b, KpTam III.2a and KpTam III.2b. We then determined the sequence similarity between these four genes and the ω-transaminases of *V. fluvialis* JS17 and *C. violaceum* DSM30191 by a multiple amino acid sequence alignment as a rudimentary *in silico* measure of the likelihood they act on substrates via an ω-type mechanism. Table 3 shows a level of similarity between the four putative *K. phaffii* GS115 Class III transaminases and the two known ω-transaminases of 21-28%. Although this is inconclusive, sequence similarity between the two proven ω-transaminases is only 38%, so the analysis in Table 3 at least recommends the four *K. phaffii* GS115 Class III transaminases for further investigation to establish if they are in fact ω-transaminases.

We next performed a substitution matrix to align the most highly conserved residues between the four *K. phaffii* Class III transaminases, KpTam III.1a, KpTam III.1b, KpTam III.2a and KpTam III.2b, the *C. violaceum* DSM30191transaminase, F2XBU9, and the *V. fluvialis* JS17 transaminase, Q7NWG4 (Figure 2). Previous alignment studies suggested a small number of residues are common to many transaminases, including a glutamic acid of unknown function which is only conserved in ω-transaminases (Mehta *et al.* 1993). Shen *et al.* (1998) also identified a conserved threonine residue, understood to form part of the cofactor phosphate binding site. Three conserved residues are most commonly reported across transaminases alignments studies: aspartic acid, lysine and arginine (Yonaha et al. 1992, Mehta et al. 1993, Shen et al. 1998, Hwang et al. 2005, Kaulmann et al. 2007). The invariant lysine is understood to participate in Schiff base formation with the 4'-aldehyde group of PLP. The invariant arginine participates in a hydrogen bond/salt bridge with the α-carboxylate group within substrates. The invariant aspartic acid generates a hydrogen bond/salt bridge to protonate the pyridine of the PLP cofactor.

Our alignment (Figure 2) identified 24 residues as being highly conserved among the six transaminases, including lysine, aspartic acid, arginine, threonine and the glutamic acid characteristic of ω-transaminases. Table 4 lists the topographically equivalent active site residues in the four putative *K. phaffii* GS115 Class III transaminases and the *C. violaceum* and *V. fluvialis* ω-transaminases.

Comparison with *K. phaffii* CBS7435 and *S. cerevisiae* S288C

We next compared the transaminase repertoire of *K. phaffi* GS115 to those of two other budding yeasts, *K. phaffi* CBS7435 and *S. cerevisiae* S288C (GenBank assembly accession: GCA_000146045.2), by performing the same data-mining procedures described previously for transaminase identification.

Strain GS115 of *K. phaffi* was originally developed by mutagenesis (Valli et al. 2016) of its parental strain *K. phaffii* CBS7435 (Küberl et al. 2011). Data-mining revealed a repertoire of 20 putative transaminases for *K. phaffii* CBS7435 (Table 5). All 19 putative transaminases of *K. phaffii* GS115 had orthologues in the *K. phaffii* CBS7435 genome, with 93.2-100% similarity. One *K. phaffii* CBS7435 transaminase (accession number F2QVZ3) had a zero similarity with any putative *K. phaffii* GS115 transaminase, suggesting this gene may have been lost during the mutagenesis procedure used to generate *K. phaffii* GS115.

S. cerevisiae S288C is a highly-characterised and widely-utilised strain whose genome was used as the reference sequence (Engel et al. 2014) for the Saccharomyces Genome Database (SGD). The genome of *S. cerevisiae* S288C is also the basis of efforts to synthetically refactor the entire genome of *S. cerevisiae* to improve its industrial utility (Richardson et al. 2017). Our data mining procedure identified 20 putative transaminases within the *S. cerevisiae* S288C genome (Table 5), all presenting some similarity to those of *K. phaffii* GS115 (27-73% identity). Comparing *S. cerevisiae* S288C and *K. phaffii* GS115 transaminases was illuminating as it revealed KpTam I-II.1b to have 57% identity with an aspartate

aminotransferase (accession number P23542) and KpTam I-II.7 to have 40% identity with a 2-aminoadipate transaminase (accession number P10356).

Overview

The information presented here will be useful for those seeking to engineer *K. phaffii* GS115 transaminases to build *de novo* biocatalytic pathways, modification of cellular metabolism and efforts to improve biological understanding of the organism. The data mining results provide a first overview of the transaminases of a methylotrophic yeast species. Our analysis suggested that in the order of 20 transaminases are encoded by the genomes of both *K. phaffii* GS115 and *S. cerevisiae* S288C. Phylogenetically divergent *K. phaffii* GS115 transaminase genes were often predicted to share the functional properties of Hwang *et al*. (2005) subgroup and reactive mechanism.

ACKNOWLEDGEMENTS

This work was supported by the Government of Chile and its Program CONICYT-Becas Chile (72120390), who supported the doctoral studies of M-J.H., and the ERA-IB IPCRES Consortium, funded in the UK by the BBSRC (BB/M004880/1), who supported all other authors.

REFERENCES

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of Molecular Biology 215:403-410 doi:10.1016/S0022-2836(05)80360-2

Bairoch A (2000) The ENZYME database in 2000. Nucleic Acids Research 28:304-305

Bollok M, Resina D, Valero F, Ferrer P (2009) Recent patents on the *Pichia pastoris* expression system: expanding the toolbox for recombinant protein production. Recent Patents on Biotechnology 3:192-201

Cho BK, Cho HJ, Park SH, Yun H, Kim BG (2003) Simultaneous synthesis of enantiomerically pure (S)-amino acids and (R)-amines using coupled transaminase reactions. Biotechnology and Bioengineering 81:783-789 doi:10.1002/bit.10526

De Schutter K, Lin YC, Tiels P, Van Hecke A, Glinka S, Weber-Lehmann J, Rouze P, Van de Peer Y, Callewaert N (2009) Genome sequence of the recombinant protein production host *Pichia pastoris*. Nature Biotechnology 27:561-566 doi:10.1038/nbt.1544

Engel SR, Dietrich FS, Fisk DG, Binkley G, Balakrishnan R, Costanzo MC, Dwight SS, Hitz BC, Karra K, Nash RS, Weng S, Wong ED, Lloyd P, Skrzypek MS, Miyasato SR, Simison M, Cherry JM (2014) The reference genome sequence of *Saccharomyces cerevisiae*: then and now. G3 (Bethesda) 4:389-398 doi:10.1534/g3.113.008995

Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F, Bateman A, Eddy SR (2015) HMMER web server: 2015 update. Nucleic Acids Research 43:W30-W38 doi:10.1093/nar/gkv397

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A (2016) The Pfam protein families database: towards a more sustainable future. Nucleic Acids Research 44:D279-D285 doi:10.1093/nar/gkv1344

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG (1996) Life with 6000 genes. Science 274:546, 563-547

Henikoff S, Henikoff JG (1992) Amino acid substitution matrices from protein blocks. Proceeding of the National Academy of Sciences of USA 89:10915-10919

Hohne M, Schatzle S, Jochens H, Robins K, Bornscheuer UT (2010) Rational assignment of key motifs for function guides in silico enzyme identification. Nature Chemical Biology 6:807-813 doi:10.1038/nchembio.447

Huson DH, Scornavacca C (2012) Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. Systematic Biology 61:1061-1067 doi:10.1093/sysbio/sys062

Hwang B-Y, Cho B-K, Yun H, Koteshwar K, Kim B-G (2005) Revisit of aminotransferase in the genomic era and its application to biocatalysis. Journal of Molecular Catalysis B: Enzymatic 37:47-55 doi:10.1016/j.molcatb.2005.09.004

Ingram CU, Bommer M, Smith ME, Dalby PA, Ward JM, Hailes HC, Lye GJ (2007) One-pot synthesis of amino-alcohols using a de-novo transketolase and beta-alanine: pyruvate transaminase pathway in *Escherichia coli*. Biotechnology and Bioengineering 96:559-569 doi:10.1002/bit.21125

Invitrogen (2010) Multi-Copy *Pichia* Expression Kit. For the Isolation and Expression of Recombinant Proteins from *Pichia pastoris* Strains Containing Multiple Copies of a Particular Gene. User Manual part no. 25-0170,

Kaulmann U, Smithies K, Smith MEB, Hailes HC, Ward JM (2007) Substrate spectrum of ω-transaminase from *Chromobacterium violaceum* DSM30191 and its

potential for biocatalysis. Enzyme and Microbial Technology 41:628-637 doi:10.1016/j.enzmictec.2007.05.011

Krogh A, Brown M, Mian IS, Sjolander K, Haussler D (1994) Hidden Markov models in computational biology. Applications to protein modeling. Journal of Molecular Biology 235:1501-1531 doi:10.1006/jmbi.1994.1104

Küberl A, Schneider J, Thallinger GG, Anderl I, Wibberg D, Hajek T, Jaenicke S, Brinkrolf K, Goesmann A, Szczepanowski R, Pühler A, Schwab H, Glieder A, Pichler H (2011) High-quality genome sequence of *Pichia pastoris* CBS7435. Journal of Biotechnology 154:312-320 doi:10.1016/j.jbiotec.2011.04.014

Kurtzman CP (2005) Description of *Komagataella phaffii* sp. nov. and the transfer of Pichia pseudopastoris to the methylotrophic yeast genus Komagataella. International Journal of Systematic and Evolutionary Microbiology 55:973-976 doi:10.1099/ijs.0.63491-0

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23:2947-2948 doi:10.1093/bioinformatics/btm404

Macauley-Patrick S, Fazenda ML, McNeil B, Harvey LM (2005) Heterologous protein production using the *Pichia pastoris* expression system. Yeast 22:249-270 doi:10.1002/yea.1208

Malik MS, Park ES, Shin JS (2012) Features and technical applications of omega-transaminases. Applied Microbiology and Biotechnology 94:1163-1171 doi:10.1007/s00253-012-4103-3

Mattanovich D, Callewaert N, Rouze P, Lin YC, Graf A, Redl A, Tiels P, Gasser B, De Schutter K (2009) Open access to sequence: browsing the *Pichia pastoris* genome. Microbial Cell Factories 8:53 doi:10.1186/1475-2859-8-53

Mehta PK, Hale TI, Christen P (1993) Aminotransferases: demonstration of homology and division into evolutionary subgroups. European Journal of Biochemistry 214:549-561

Payne WE, Kaiser CA, Bevis BJ, Soderholm J, Fu D, Sears IB, Glick BS (2000) Isolation of *Pichia pastoris* genes involved in ER-to-Golgi transport. Yeast 16:979-993 doi:10.1002/1097-0061(200008)16:11<979::AID-YEA594>3.0.CO;2-C

Qian B, Goldstein RA (2003) Detecting distant homologs using phylogenetic tree-based HMMs. Proteins 52:446-453 doi:10.1002/prot.10373

Richardson SM, Mitchell LA, Stracquadanio G, Yang K, Dymond JS, DiCarlo JE, Lee D, Huang CL, Chandrasegaran S, Cai Y, Boeke JD, Bader JS (2017) Design of a synthetic yeast genome. Science 355:1040-1044 doi:10.1126/science.aaf4557

Robert X, Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Research 42:W320-324 doi:10.1093/nar/gku316

Shen BW, Hennig M, Hohenester E, Jansonius JN, Schirmer T (1998) Crystal structure of human recombinant ornithine aminotransferase. Journal of Molecular Biology 277:81-102 doi:10.1006/jmbi.1997.1583

Shin JS, Yun H, Jang JW, Park I, Kim BG (2003) Purification, characterization, and molecular cloning of a novel amine:pyruvate transaminase from *Vibrio fluvialis*

JS17. Applied Microbiology and Biotechnology 61:463-471 doi:10.1007/s00253-003-1250-6

Templar A, Woodhouse S, Keshavarz-Moore E, Nesbeth DN (2016) Influence of *Pichia pastoris* cellular material on polymerase chain reaction performance as a synthetic biology standard for genome monitoring. Journal of Microbiological Methods 127, :111-122 doi:10.1016/j.mimet.2016.05.013

Valli M, Tatto NE, Peymann A, Gruber C, Landes N, Ekker H, Thallinger GG, Mattanovich D, Gasser B, Graf AB (2016) Curation of the genome annotation of *Pichia pastoris* (*Komagataella phaffii*) CBS7435 from gene level to protein function. FEMS Yeast Research 16:fow051 doi:10.1093/femsyr/fow051

Van der Klei IJ, Yurimoto H, Sakai Y, Veenhuis M (2006) The significance of peroxisomes in methanol metabolism in methylotrophic yeast. Biochimica Biophysica Acta (BBA)-Molecular Cell Research 1763:1453-1462 doi:10.1016/j.bbamcr.2006.07.016

Ward J, Wohlgemuth R (2010) High-Yield Biocatalytic Amination Reactions in Organic Synthesis. Current Organic Chemistry 14:1914-1927 doi:10.2174/138527210792927546

Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics 25:1189-1191 doi:10.1093/bioinformatics/btp033

Wei YC, Braun-Galleani S, Henriquez MJ, Bandara S, Nesbeth DN (2018) Biotransformation of beta-Hydroxypyruvate and Glycolaldehyde to L-Erythrulose by *Pichia pastoris* strain GS115 overexpressing native Transketolase. Biotechnology Progress 34:99-106 doi:10.1002/btpr.2577

Yonaha K, Nishie M, Aibara S (1992) The primary structure of omega-amino acid:pyruvate aminotransferase. Journal of Biological Chemistry 267:12506-12510

Figure 1 Dendrogram of the 19 *K. phaffii* (*P. pastoris*) GS115 transaminases identified *in silico*. The numbers in nodes are the bootstrap values. Transaminase class assigned by HMMER analysis. Branches in blue are indicative of Class I-II; green of Class III; red of Class IV; and orange of Class V. There are no transaminases belonging to Class VI. Accession numbers are indicated at the end of each branch followed by systematic nomenclature group in bold.

Figure 2 Amino acid sequence alignment of four putatively Class-III transminases from *K. phaffii* (*P. pastoris*) GS115, transaminase F2XBU9 from *V. fluvialis* JS17 and transaminase Q7NWG4 from *C. violaceum* DSM30192. Red colour indicates residues of highest homology by BLOSUM62 scores. Residues conserved in ω-TAms are marked by blue circles. Secondary structure elements in the spatial structure of CV2025 TAm (Sayer et al., 2007) are indicated at the top of each block: α-helices are displayed as squiggles, β-strands as arrows, strict β-turns as TT letters and 310-helix as η symbol

Figure 1

Click here to access/download;Figure;Figure 1 JGEN resub 200120.jpg

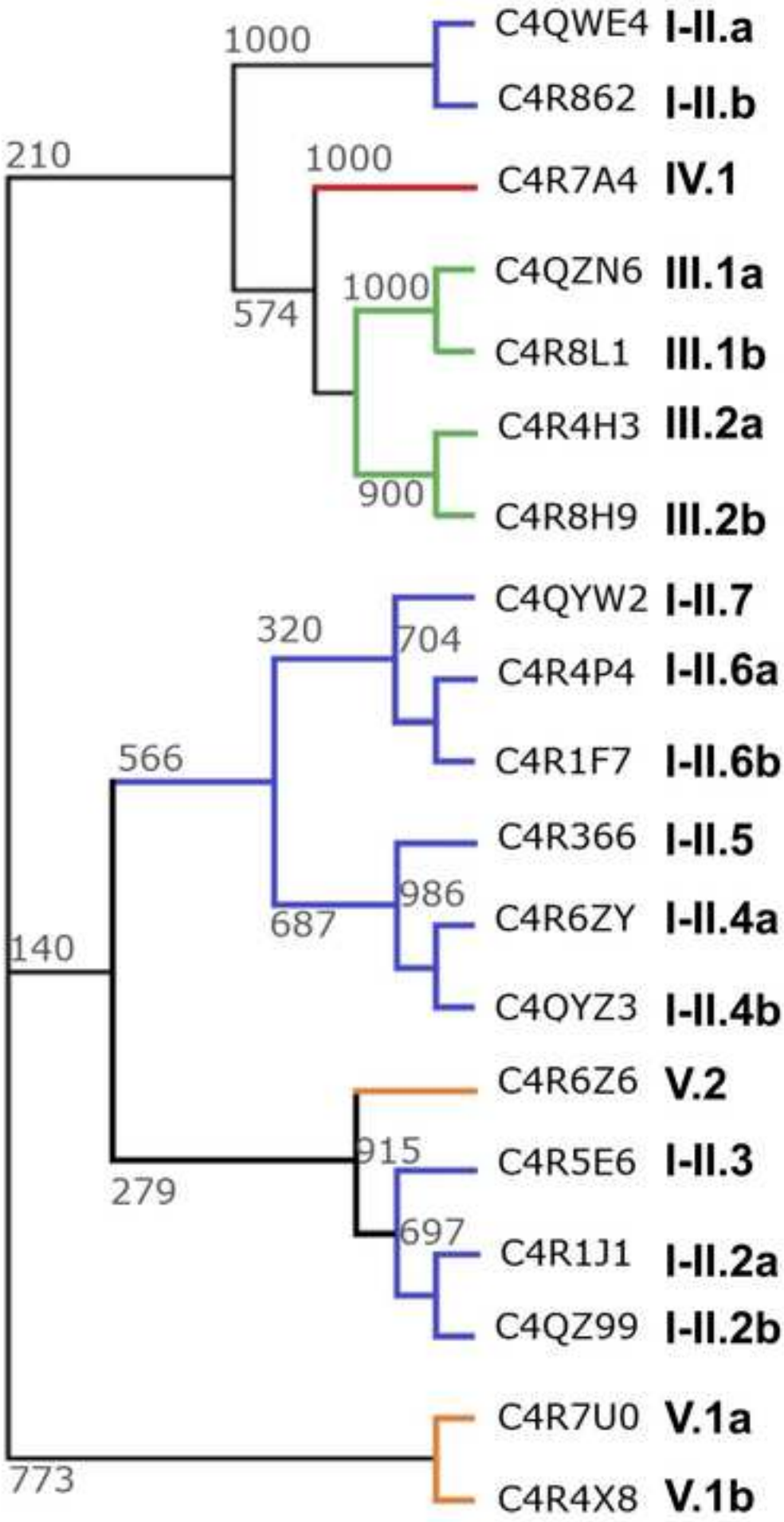

Figure 2

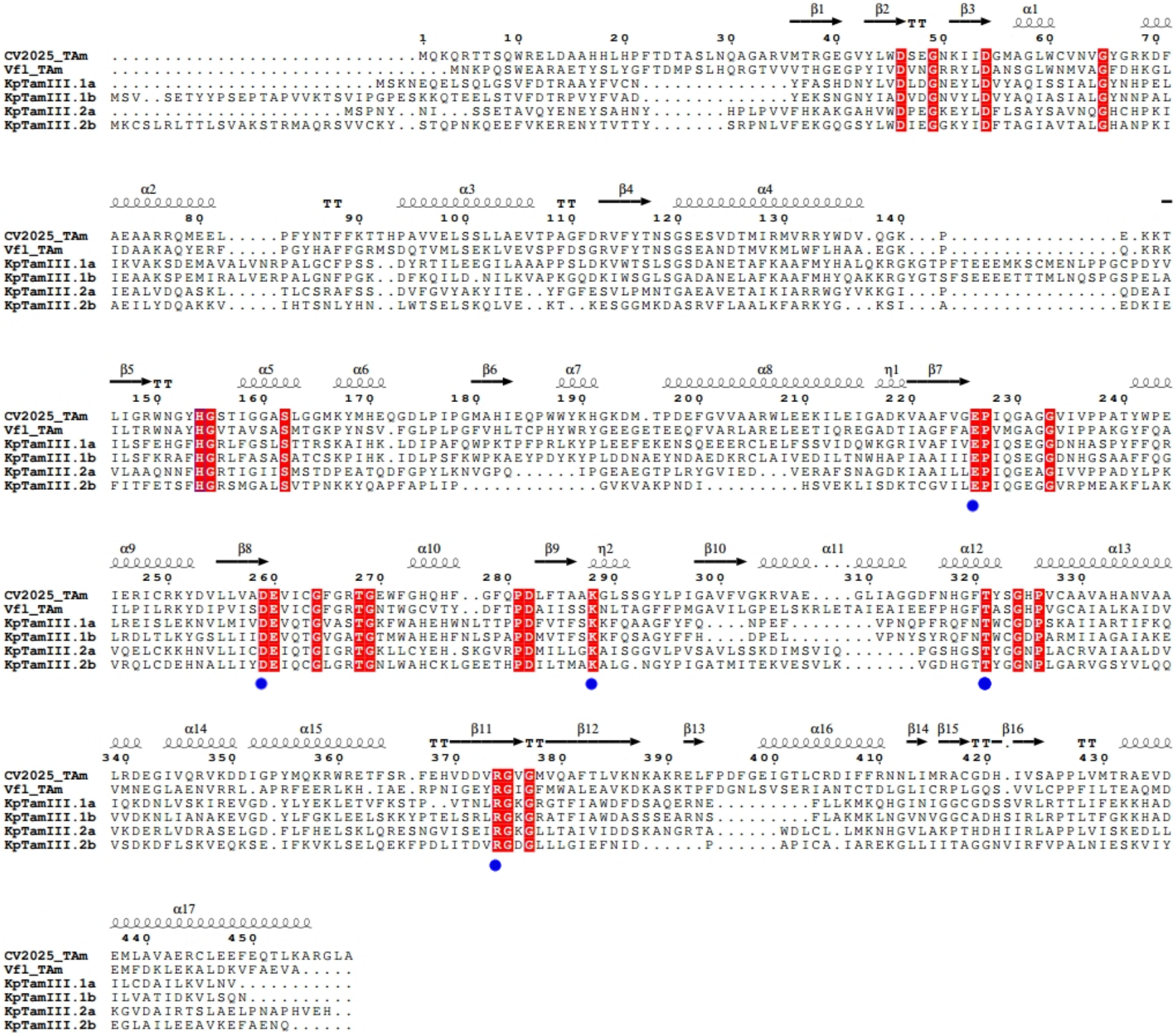
CV2025_TAm
Vfl_TAm
KpTamIII.1a
KpTamIII.1b
KpTamIII.2a
KpTamIII.2b

Table 1. The subgroups of transaminases set out by Hwang *et al*. (2005).

Subgroup	Notes
Class I+II.	Classes I and II were first proposed by Mehta *et al*. (1993) and comprise, respectively, the aspartate and aromatic transaminases. These are the most studied, and in general they use L-aspartate and L-tyrosine as an amino donor, respectively, and α-ketoglutarate as an amino acceptor. The difference between an aspartate and aromatic transaminase is the hydrophobicity of the active site binding pocket (Hwang *et al* 2005).
Class III.	Have a wide variety of substrate acceptors and can transfer an amino group to aldehydes and ketones of different types. Include β-transaminases, Ɣ-aminobutyrate transaminases, and ω-transaminases. Have no requirement for 2-ketocarboxylate (Ward and Wohlgemuth 2010).
Class IV.	Differ structurally from the other types in that the positions of the large and small binding pockets are reversed.
Class V.	Act on structurally and biosynthetically related substrates (Mehta *et al*. 1993).
Class VI.	Comprise sugar aminotransferases, the majority of which use L-glutamate as the amino donor. Are derived from antibiotic operons (Ward & Wohlgemuth 2010).

Table 2

Table 2. *K. phaffii* (*P. pastoris*) transaminases identified by genome query. Of the 39 genes, entries numbered 1-19 were assigned as unique, non-numbered entries were either duplicates or, indicated by asterisk, those that did not match any transaminase class by HMM profiling.

this report	Accession Number	Entry name	Gene Names	Predicted Function	ORF size (bp)	Entry description
1	C4R4P4	C4R4P4_PICPG	PAS_chr3_0482	Putative alanine transaminase (Glutamic pyruvic transaminase)	510	Transaminases
2	C4QWE4	C4QWE4_PICPG	PAS_chr1-1_0200	Aspartate aminotransferase (EC 2.6.1.1)	426	Transaminases
	C4QWE4	C4QWE4_PICPG	PAS_chr1-1_0200	Aspartate aminotransferase (EC 2.6.1.1)	426	Aminotranferases
3	C4QYW2	C4QYW2_PICPG	PAS_chr1-4_0579	Putative uncharacterized protein	439	Aminotranferases
4	C4QYZ3	C4QYZ3_PICPG	PAS_chr1-4_0608	Aromatic aminotransferase I, expression regulated by general control of amino acid biosynthesis	482	Transaminases
	C4QYZ3	C4QYZ3_PICPG	PAS_chr1-4_0608	Aromatic aminotransferase I, expression regulated by general control of amino acid biosynthesis	482	Aminotranferases
5	C4QZ99	C4QZ99_PICPG	PAS_FragB_0040	Serine palmitoyltransferase 1	550	Aminotranferases
6	C4QZN6	C4QZN6_PICPG	PAS_chr2-1_0107	Gamma-aminobutyrate (GABA) transaminase	446	Transaminases
	C4QZN6	C4QZN6_PICPG	PAS_chr2-1_0107	Gamma-aminobutyrate (GABA) transaminase	446	Aminotranferases
	C4R194*	C4R194_PICPG	PAS_chr2-1_0626	Glutamine-fructose-6-phosphate amidotransferase	696	Transaminases
	C4R194*	C4R194_PICPG	PAS_chr2-1_0626	Glutamine-fructose-6-phosphate amidotransferase	696	Aminotranferases
7	C4R1F7	C4R1F7_PICPG	PAS_chr2-1_0684	Histidinol-phosphate aminotransferase, catalyzes the seventh step in histidine biosynthesis	390	Transaminases
	C4R1F7	C4R1F7_PICPG	PAS_chr2-1_0684	Histidinol-phosphate aminotransferase, catalyzes the seventh step in histidine biosynthesis	390	Aminotranferases
8	C4R1J1	C4R1J1_PICPG	PAS_chr2-1_0716	5-aminolevulinate synthase (EC 2.3.1.37)	560	Aminotranferases
	C4R277*	C4R277_PICPG	PAS_chr2-2_0492	Aminomethyltransferase (EC 2.1.2.10)	392	Transaminases
	C4R277*	C4R277_PICPG	PAS_chr2-2_0492	Aminomethyltransferase (EC 2.1.2.10)	392	Aminotranferases
9	C4R366	C4R366_PICPG	PAS_chr3_1132	Kynurenine aminotransferase, catalyzes formation of kynurenic acid from kynurenine	434	Transaminases
	C4R366	C4R366_PICPG	PAS_chr3_1132	Kynurenine aminotransferase, catalyzes formation of kynurenic acid from kynurenine	434	Aminotranferases
10	C4R4H3	C4R4H3_PICPG	PAS_chr3_0410	L-ornithine transaminase (OTAse)	434	Transaminases
	C4R4H3	C4R4H3_PICPG	PAS_chr3_0410	L-ornithine transaminase (OTAse)	434	Aminotranferases
	C4R4P4	C4R4P4_PICPG	PAS_chr3_0482	Putative alanine transaminase (Glutamic pyruvic transaminase)	510	Aminotranferases
11	C4R4X8	C4R4X8_PICPG	PAS_chr3_0566	3-phosphoserine aminotransferase	390	Transaminases
	C4R4X8	C4R4X8_PICPG	PAS_chr3_0566	3-phosphoserine aminotransferase	390	Aminotranferases
12	C4R5E6	C4R5E6_PICPG	PAS_chr3_0733	Component of serine palmitoyltransferase	561	Aminotranferases
13	C4R6Z6	C4R6Z6_PICPG	PAS_chr4_0146	Uncharacterized protein	417	Aminotranferases
14	C4R6Z7	C4R6Z7_PICPG	PAS_chr4_0147	Aromatic aminotransferase II	462	Transaminases
	C4R6Z7	C4R6Z7_PICPG	PAS_chr4_0147	Aromatic aminotransferase II	462	Aminotranferases
15	C4R7A4	C4R7A4_PICPG	PAS_chr4_0248	Branched-chain-amino-acid aminotransferase (EC 2.6.1.42)	405	Transaminases
	C4R7A4	C4R7A4_PICPG	PAS_chr4_0248	Branched-chain-amino-acid aminotransferase (EC 2.6.1.42)	405	Aminotranferases
16	C4R7U0	C4R7U0_PICPG	PAS_chr4_0416	Alanine:glyoxylate aminotransferase (AGT), catalyzes synthesis of glycine from glyoxylate	413	Transaminases
	C4R7U0	C4R7U0_PICPG	PAS_chr4_0416	Alanine:glyoxylate aminotransferase (AGT), catalyzes synthesis of glycine from glyoxylate	413	Aminotranferases
17	C4R862	C4R862_PICPG	PAS_chr4_0974	Uncharacterized protein	385	Transaminases
	C4R862	C4R862_PICPG	PAS_chr4_0974	Uncharacterized protein	385	Aminotranferases
	C4R864*	C4R864_PICPG	PAS_chr4_0530	Cytosolic aspartate aminotransferase, involved in nitrogen metabolism	351	Transaminases
	C4R864*	C4R864_PICPG	PAS_chr4_0530	Cytosolic aspartate aminotransferase, involved in nitrogen metabolism	351	Aminotranferases
18	C4R8H9	C4R8H9_PICPG	PAS_chr4_0645	Acetylornithine aminotransferase	431	Transaminases
	C4R8H9	C4R8H9_PICPG	PAS_chr4_0645	Acetylornithine aminotransferase	431	Aminotranferases
19	C4R8L1	C4R8L1_PICPG	PAS_chr4_0677	Gamma-aminobutyrate (GABA) transaminase (4-aminobutyrate aminotransferase)	471	Transaminases
	C4R8L1	C4R8L1_PICPG	PAS_chr4_0677	Gamma-aminobutyrate (GABA) transaminase (4-aminobutyrate aminotransferase)	471	Aminotranferases

Table 3

Table 3. Identity percentages obtained in pairwise alignments between the Class III transaminases from *K. phaffii* (*P. pastoris*) GS115, and the reference ω-transaminases from *V. fluvialis* JS17 and *C. violaceum* DSM30191.

Systematic Nomenclature	***V. fluvialis* JS17**	***C. violaceum* DSM30191**
KpTam III.1a	22.46%	20.82%
KpTam III.1b	22.68%	22.45%
KpTam III.2a	25.76%	28.14%
KpTam III.2b	22.78%	26.91%

Table 4

Table 4. Topographical alignment of active site residues between the Class III transaminases from *K. phaffii* (*P. pastoris*) GS115, and the reference ω-transaminases from *C. violaceum* DSM30191 and *V. fluvialis* JS17.

C. violaceum **DSM30191**	***V. fluvialis*** **JS17**	**KpTam III.1a**	**KpTam III.1b**	**KpTam III.2a**	**KpTam III.2b**
E226	E223	E240	E262	E211	E218
D259	D256	D273	D295	D244	D251
K288	K285	K304	K326	K274	K282
T321	T322	T329	T351	T304	T311
R374	R373	R380	R404	R357	R364

Table 5

Table 5. Comparison of transaminases identified *in silico* in *K. phaffii* strains GS115 and CBS7435 and *S. cerevisiae* strain S288C. Hwang class assigned by HMMER HMMscan. Percentage identity shows sequence homology between strain pair.

Systematic Nomenclature GS115 Protein	Accession Number CBS7435	Accession Number S288C	Class (HMMER)	% Identity GS115 with CBS7435	% Identity GS115 with S288C
KpTam I-II.1a	F2QML6	Q01802	I-II	100%	40%
KpTam I-II.1b	F2QYY5	P23542	I-II	93.20%	57%
KpTamI-II.2a	F2QS71	P09950	I-II	100%	68%
KpTamI-II.2b	F2QQF8	P25045	I-II	100%	42%
KpTam I-II.3	F2QVJ5	P40970	I-II	100%	62%
KpTam I-II.4a	F2R043	P38840	I-II	100%	34%
KpTam I-II.4b	F2QQ55	P53090	I-II	100%	52%
KpTam I-II.5	F2QUC6	P47039	I-II	100%	56%
KpTam I-II.6a	F2QWA4	P52892 P52893	I-II	100%	58% 59%
KpTam I-II.6b	F2QSA5	P07172	I-II	100% 100%	53%
KpTam I-II.7	F2QQ23	P10356	I-II	100%	40%
KpTam III.1a	F2QTZ9	-	III	100%	-
KpTam III.1b	F2QYI5	P17649	III	100%	68%
KpTam III.2a	F2QWH8	P50277 P07991	III	100%	27% 70%
KpTam III.2b	F2QYL6	P18544	III	98%	51%
KpTam IV.1	F2QZT3	P38891 P47176	IV	100%	73% 69%
KpTam V.1a	F2QZA3	P43567	V	100%	48%
KpTam V.1b	F2QW19	P33330	V	100%	59%
KpTam V.2	F2R044	-	V	100%	-
-	F2QVZ3	-	V	-	-