

[Click here to view linked References](#)

1  
2  
3       **Genomic data mining reveals the transaminase repertoire of *Komagataella***  
4       ***phaffii* (*Pichia pastoris*) strain GS115 and supports a systematic**  
5       **nomenclature**  
6  
7  
8  
9

10  
11       Maria José Henríquez, Rosalía Paula Cardós-Elena, Darren Nicholas Nesbeth\*

12  
13  
14  
15       \*Corresponding author: Department of Biochemical Engineering, University College

16  
17       London, Bernard Katz Building, London WC1E 6BT, United Kingdom.

18  
19       Tel: +44 (0) 20 7679 9582, Fax: +44 (0) 207 916 3943

20  
21       Email: d.nesbeth@ucl.ac.uk  
22  
23  
24

25       **Keywords:** Transaminase; Aminotransferase; bioinformatics; genome; phylogeny;  
26       nomenclature  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3 **ABSTRACT**  
4

5 Transaminases are an industrially important class of enzyme due to their ability to  
6 catalyse amination reactions for production of chiral amines, key building blocks of  
7 small molecule pharmaceuticals. We analysed the genome of strain GS115 of the  
8 methylotrophic yeast *Komagataella phaffii* (*K. phaffii*), formerly known as *Pichia*  
9 *pastoris* (*P. pastoris*), to identify transaminase genes and propose a systematic  
10 nomenclature based on both phylogeny and structuro-functional features. *K. phaffii*  
11 is an increasingly attractive industrial host cell due to its ability to grow to high  
12 biomass, up to 60% wet cell weight by volume, using methanol as carbon source  
13 and inducer of transgene expression. 39 UniProt database hits were reduced to 19  
14 on the basis of sequence similarity and Hidden Markov Modelling. Of the 19 genes,  
15 the open reading frames of three (KpTam I-II.1b, KpTam I-II.7 and KpTam V.2) had  
16 strong homology with no characterised protein and four (KpTam III.1a, KpTam  
17 III.1b, KpTam III.2a and KpTam III.2b) had relatively high sequence similarity to  $\omega$ -  
18 type transaminases, a subtype that typically accepts the broadest range of  
19 substrates. Comparison with *Saccharomyces cerevisiae* S288C suggested  
20 functions for KpTam I-II.1b and KpTam I-II.7. *K. phaffii* GS115 was originally  
21 generated by mutagenesis of *K. phaffii* CBS7435 and comparison revealed that  
22 one transaminase gene may have been deleted during this mutagenesis. These  
23 insights can advance fundamental understanding of yeast biology and can inform  
24 industrial screening and engineering of yeast transaminases.  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47

48 **Keywords:** *Komagataella phaffii*, *Pichia pastoris*, transaminase, Hidden Markov;  
49 standard nomenclature  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

## 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  $\alpha$ -transaminases, as they transfer the amino group from the  $\alpha$  position. Class III contains  $\beta$ -transaminases,  $\gamma$ -transaminases, and  $\omega$ -transaminases. Class III transaminases are especially sought after as they have the widest substrate spectrum. Class III,  $\omega$ -type transaminases are the only enzyme known to perform stereoselective amination of

1  
2  
3 ketones. As a result Class III,  $\omega$ -type transaminases, have been investigated  
4 extensively for industrial production of amino acids, chiral amines and amino  
5 alcohols, all of which are valuable key intermediates for chemical synthesis of  
6 chiral, small-molecule therapeutics ([Malik et al. 2012](#)).

7  
8  
9  
10  
11 Class III,  $\omega$ -type transaminases are highly active for a given substrate and also  
12 stable to a broad range of pH, substrate concentration, temperature and product  
13 concentration. Class III,  $\omega$ -type transaminases with industrial promise have been  
14 identified by Shin *et al.* ([2003](#)) in *Vibrio fluvialis* JS17 (*V. fluvialis*), by Yonaha *et al.*  
15 ([1992](#)) in *Pseudomonas putida* (*P. putida*) and by Kaulmann *et al.* ([2007](#)) in  
16 *Chromobacterium violaceum* (*C. violaceum*).

17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
860  
861  
862  
863  
864  
865  
866  
867  
868  
869  
870  
871  
872  
873  
874  
875  
876  
877  
878  
879  
880  
881  
882  
883  
884  
885  
886  
887  
888  
889  
890  
891  
892  
893  
894  
895  
896  
897  
898  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913  
914  
915  
916  
917  
918  
919  
920  
921  
922  
923  
924  
925  
926  
927  
928  
929  
930  
931  
932  
933  
934  
935  
936  
937  
938  
939  
940  
941  
942  
943  
944  
945  
946  
947  
948  
949  
950  
951  
952  
953  
954  
955  
956  
957  
958  
959  
960  
961  
962  
963  
964  
965  
966  
967  
968  
969  
970  
971  
972  
973  
974  
975  
976  
977  
978  
979  
980  
981  
982  
983  
984  
985  
986  
987  
988  
989  
990  
991  
992  
993  
994  
995  
996  
997  
998  
999  
1000

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 ( $\alpha$  and  $\omega$ ) 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.

1  
2  
3 The methylotrophic yeast *P. pastoris* was reassigned to the genus *Komagataella*  
4 following phylogenetic analysis ([Kurtzman 2005](#)), and the major strains split into  
5 three species: *Komagataella pastoris*, *K. phaffii* GS115 (formerly *P. pastoris* strain  
6 GS115), and *Komagataella pseudopastoris*. *K. phaffii* GS115 is a reliable and  
7 robust expression system ([Invitrogen 2010](#)) that has become widely used for  
8 production of recombinant protein in research and industrial settings ([Templar et al.](#)  
9 [2016](#)) with well-established tools ([Bollok et al. 2009](#)) for genetic manipulation,  
10 strong native promoters to direct overexpression of transgenes and the ability to  
11 grow rapidly to high cell densities, up to 60% wet cell weight (wcv) by volume, on  
12 chemically defined culture media ([Wei et al. 2018](#)). Furthermore, *K. phaffii* GS115  
13 is thermotolerant, able to grow at 47°C ([Van der Klei et al. 2006](#)), and it is tolerant  
14 to pH3-7 ([Macauley-Patrick et al. 2005](#)).

15  
16 The availability of the complete genome sequence ([De Schutter et al. 2009](#);  
17 [Mattanovich et al. 2009](#)) has made bioinformatic data-mining for transaminases  
18 possible in *K. phaffii* GS115. Here we attempt to predict the function of  
19 transaminase using sequence analyses and to assign each identified  
20 transaminases to a Hwang subclasses ([Hwang et al. 2005](#)) using HMM analysis.  
21 We also propose a rational nomenclature for *K. phaffii* GS115 transaminases. This  
22 information will assist future investigators who wish to exploit or rationally design  
23 transaminases for enhanced stability, substrate specificity, PLP binding and other  
24 properties.

## 25 26 27 28 **MATERIAL AND METHODS**

### 29 30 **Identifying transaminase genes in yeast genome sequences**

31  
32 The UniProt database (The Uniprot Consortium 2015) was searched using first, the  
33 terms 'pichia pastoris', 'gs115' and 'GO:0008483' (gene ontology term for  
34 transaminase/aminotransferase activity); and then, replacing the GO terms by the  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3 keywords 'transaminase' and 'aminotransferase' to generate a wider pool of initial  
4 sequences. JalView v.2.8.2 ([Waterhouse et al. 2009](#)) was then used to remove  
5 duplicate results by examining identity percentages and pairwise alignments. The  
6 first round of UniProt hits was further refined using the terms "pyridoxal binding  
7 site", "aminotransferase" and "transaminase". The nucleotide sequence of each  
8 transaminase identified in UniProt was obtained from the corresponding NCBI  
9 entry. The methods above were also used to identify transaminase genes in the  
10 published genome sequences of *K. phaffii* CBS7435 ([Küberl et al. 2011](#)) and *S.*  
11 *cerevisiae* S228C ([Goffeau et al. 1996](#)).

### 21 **Assembly of transaminase dendrograms and phylograms**

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

### 37 **Assigning transaminases to a Hwang subgroup**

38 Hits from the above UniProt search were then used to query the HMM database  
39 HMMER ([Finn et al. 2015](#)), provided by the European Molecular Biology  
40 Organisation - Biology European Bioinformatics Institute (EMBO-EBI) website  
41 ([www.ebi.ac.uk/Tools/hmmer/search/hmmscan](http://www.ebi.ac.uk/Tools/hmmer/search/hmmscan)), by accession number. Predictions  
42 were deemed acceptable if the score was positive and expectation values (E-  
43 values) lower than 0.001. All protein families available on the search tool (Pfam,  
44 TIGRFAM, Gene3D, Superfamily and PIRSF) were selected for the search and  
45 those transaminases scoring highest for a HMM model pertaining to a particular  
46 subgroup ([Hwang et al. 2005](#)) were assigned to that subgroup.  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

### Alignment of $\omega$ -transaminases

*V. fluvialis* JS17 and *C. violaceum* DSM30191  $\omega$ -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 *Komagataella phaffii* to 'Kp' and *transaminase* 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 $\omega$ -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  $\beta$ -,  $\gamma$ - or  $\omega$ - type. Transaminases of the  $\omega$ - 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



1  
2  
3 identified four putative Class III transaminases in *K. phaffii* GS115: KpTam III.1a,  
4 KpTam III.1b, KpTam III.2a and KpTam III.2b. We then determined the sequence  
5 similarity between these four genes and the  $\omega$ -transaminases of *V. fluvialis* JS17  
6 and *C. violaceum* DSM30191 by a multiple amino acid sequence alignment as a  
7 rudimentary *in silico* measure of the likelihood they act on substrates via an  $\omega$ -type  
8 mechanism. Table 3 shows a level of similarity between the four putative *K. phaffii*  
9 GS115 Class III transaminases and the two known  $\omega$ -transaminases of 21-28%.  
10 Although this is inconclusive, sequence similarity between the two proven  $\omega$ -  
11 transaminases is only 38%, so the analysis in Table 3 at least recommends the  
12 four *K. phaffii* GS115 Class III transaminases for further investigation to establish if  
13 they are in fact  $\omega$ -transaminases.  
14

15 We next performed a substitution matrix to align the most highly conserved  
16 residues between the four *K. phaffii* Class III transaminases, KpTam III.1a, KpTam  
17 III.1b, KpTam III.2a and KpTam III.2b, the *C. violaceum* DSM30191 transaminase,  
18 F2XBU9, and the *V. fluvialis* JS17 transaminase, Q7NWX4 (Figure 2). Previous  
19 alignment studies suggested a small number of residues are common to many  
20 transaminases, including a glutamic acid of unknown function which is only  
21 conserved in  $\omega$ -transaminases (Mehta *et al.* [1993](#)). Shen *et al.* ([1998](#)) also  
22 identified a conserved threonine residue, understood to form part of the cofactor  
23 phosphate binding site. Three conserved residues are most commonly reported  
24 across transaminases alignments studies: aspartic acid, lysine and arginine  
25 (Yonaha *et al.* 1992, Mehta *et al.* 1993, Shen *et al.* 1998, Hwang *et al.* 2005,  
26 Kaulmann *et al.* 2007). The invariant lysine is understood to participate in Schiff  
27 base formation with the 4'-aldehyde group of PLP. The invariant arginine  
28 participates in a hydrogen bond/salt bridge with the  $\alpha$ -carboxylate group within  
29 substrates. The invariant aspartic acid generates a hydrogen bond/salt bridge to  
30 protonate the pyridine of the PLP cofactor.  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3 Our alignment (Figure 2) identified 24 residues as being highly conserved among  
4 the six transaminases, including lysine, aspartic acid, arginine, threonine and the  
5 glutamic acid characteristic of  $\omega$ -transaminases. Table 4 lists the topographically  
6 equivalent active site residues in the four putative *K. phaffii* GS115 Class III  
7 transaminases and the *C. violaceum* and *V. fluvialis*  $\omega$ -transaminases.  
8  
9

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

11 We next compared the transaminase repertoire of *K. phaffii* GS115 to those of two  
12 other budding yeasts, *K. phaffii* CBS7435 and *S. cerevisiae* S288C (GenBank  
13 assembly accession: GCA\_000146045.2), by performing the same data-mining  
14 procedures described previously for transaminase identification.  
15

16 Strain GS115 of *K. phaffii* was originally developed by mutagenesis ([Valli et al.](#)  
17 [2016](#)) of its parental strain *K. phaffii* CBS7435 ([Küberl et al. 2011](#)). Data-mining  
18 revealed a repertoire of 20 putative transaminases for *K. phaffii* CBS7435 (Table  
19 5). All 19 putative transaminases of *K. phaffii* GS115 had orthologues in the *K.*  
20 *phaffii* CBS7435 genome, with 93.2-100% similarity. One *K. phaffii* CBS7435  
21 transaminase (accession number F2QVZ3) had a zero similarity with any putative  
22 *K. phaffii* GS115 transaminase, suggesting this gene may have been lost during  
23 the mutagenesis procedure used to generate *K. phaffii* GS115.  
24

25 *S. cerevisiae* S288C is a highly-characterised and widely-utilised strain whose  
26 genome was used as the reference sequence ([Engel et al. 2014](#)) for the  
27 Saccharomyces Genome Database (SGD). The genome of *S. cerevisiae* S288C is  
28 also the basis of efforts to synthetically refactor the entire genome of *S. cerevisiae*  
29 to improve its industrial utility ([Richardson et al. 2017](#)). Our data mining procedure  
30 identified 20 putative transaminases within the *S. cerevisiae* S288C genome (Table  
31 5), all presenting some similarity to those of *K. phaffii* GS115 (27-73% identity).  
32 Comparing *S. cerevisiae* S288C and *K. phaffii* GS115 transaminases was  
33 illuminating as it revealed KpTam I-II.1b to have 57% identity with an aspartate  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3 aminotransferase (accession number P23542) and KpTam I-II.7 to have 40%  
4 identity with a 2-aminoadipate transaminase (accession number P10356).  
5  
6

### 7 **Overview**

8  
9 The information presented here will be useful for those seeking to engineer *K.*  
10 *phaffii* GS115 transaminases to build *de novo* biocatalytic pathways, modification  
11 of cellular metabolism and efforts to improve biological understanding of the  
12 organism. The data mining results provide a first overview of the transaminases of  
13 a methylotrophic yeast species. Our analysis suggested that in the order of 20  
14 transaminases are encoded by the genomes of both *K. phaffii* GS115 and *S.*  
15 *cerevisiae* S288C. Phylogenetically divergent *K. phaffii* GS115 transaminase genes  
16 were often predicted to share the functional properties of Hwang *et al.* ([2005](#))  
17 subgroup and reactive mechanism.  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28

### 29 **ACKNOWLEDGEMENTS**

30 This work was supported by the Government of Chile and its Program CONICYT-  
31 Becas Chile (72120390), who supported the doctoral studies of M-J.H., and the  
32 ERA-IB IPCRES Consortium, funded in the UK by the BBSRC (BB/M004880/1),  
33 who supported all other authors.  
34  
35  
36  
37  
38  
39  
40  
41

### 42 **REFERENCES**

43 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment  
44 search tool. *Journal of Molecular Biology* 215:403-410 doi:10.1016/S0022-  
45 2836(05)80360-2  
46  
47  
48

49  
50  
51 Bairoch A (2000) The ENZYME database in 2000. *Nucleic Acids Research* 28:304-  
52 305  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3 Bollok M, Resina D, Valero F, Ferrer P (2009) Recent patents on the *Pichia*  
4 *pastoris* expression system: expanding the toolbox for recombinant protein  
5 production. *Recent Patents on Biotechnology* 3:192-201  
6  
7

8  
9  
10 Cho BK, Cho HJ, Park SH, Yun H, Kim BG (2003) Simultaneous synthesis of  
11 enantiomerically pure ( S )-amino acids and ( R )-amines using coupled  
12 transaminase reactions. *Biotechnology and Bioengineering* 81:783-789  
13 doi:10.1002/bit.10526  
14  
15  
16

17  
18  
19 De Schutter K, Lin YC, Tiels P, Van Hecke A, Glinka S, Weber-Lehmann J, Rouze  
20 P, Van de Peer Y, Callewaert N (2009) Genome sequence of the recombinant  
21 protein production host *Pichia pastoris*. *Nature Biotechnology* 27:561-566  
22 doi:10.1038/nbt.1544  
23  
24  
25  
26

27  
28  
29 Engel SR, Dietrich FS, Fisk DG, Binkley G, Balakrishnan R, Costanzo MC, Dwight  
30 SS, Hitz BC, Karra K, Nash RS, Weng S, Wong ED, Lloyd P, Skrzypek MS,  
31 Miyasato SR, Simison M, Cherry JM (2014) The reference genome sequence of  
32 *Saccharomyces cerevisiae*: then and now. *G3 (Bethesda)* 4:389-398  
33 doi:10.1534/g3.113.008995  
34  
35  
36  
37

38  
39  
40 Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F, Bateman A,  
41 Eddy SR (2015) HMMER web server: 2015 update. *Nucleic Acids Research*  
42 43:W30-W38 doi:10.1093/nar/gkv397  
43  
44  
45  
46

47 Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta  
48 M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A (2016) The  
49 Pfam protein families database: towards a more sustainable future. *Nucleic Acids*  
50 *Research* 44:D279-D285 doi:10.1093/nar/gkv1344  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3 Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F,  
4 Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen  
5 P, Tettelin H, Oliver SG (1996) Life with 6000 genes. *Science* 274:546, 563-547  
6  
7

8  
9  
10 Henikoff S, Henikoff JG (1992) Amino acid substitution matrices from protein  
11 blocks. *Proceeding of the National Academy of Sciences of USA* 89:10915-10919  
12  
13

14  
15 Hohne M, Schatzle S, Jochens H, Robins K, Bornscheuer UT (2010) Rational  
16 assignment of key motifs for function guides in silico enzyme identification. *Nature*  
17 *Chemical Biology* 6:807-813 doi:10.1038/nchembio.447  
18  
19  
20

21  
22  
23 Huson DH, Scornavacca C (2012) Dendroscope 3: an interactive tool for rooted  
24 phylogenetic trees and networks. *Systematic Biology* 61:1061-1067  
25 doi:10.1093/sysbio/sys062  
26  
27

28  
29  
30 Hwang B-Y, Cho B-K, Yun H, Koteshwar K, Kim B-G (2005) Revisit of  
31 aminotransferase in the genomic era and its application to biocatalysis. *Journal of*  
32 *Molecular Catalysis B: Enzymatic* 37:47-55 doi:10.1016/j.molcatb.2005.09.004  
33  
34  
35

36  
37 Ingram CU, Bommer M, Smith ME, Dalby PA, Ward JM, Hailes HC, Lye GJ (2007)  
38 One-pot synthesis of amino-alcohols using a de-novo transketolase and beta-  
39 alanine: pyruvate transaminase pathway in *Escherichia coli*. *Biotechnology and*  
40 *Bioengineering* 96:559-569 doi:10.1002/bit.21125  
41  
42  
43  
44

45  
46 Invitrogen (2010) Multi-Copy *Pichia* Expression Kit. For the Isolation and  
47 Expression of Recombinant Proteins from *Pichia pastoris* Strains Containing  
48 Multiple Copies of a Particular Gene. User Manual part no. 25-0170,  
49  
50  
51

52  
53 Kaulmann U, Smithies K, Smith MEB, Hailes HC, Ward JM (2007) Substrate  
54 spectrum of  $\omega$ -transaminase from *Chromobacterium violaceum* DSM30191 and its  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3 potential for biocatalysis. *Enzyme and Microbial Technology* 41:628-637  
4  
5 doi:10.1016/j.enzmictec.2007.05.011  
6

7  
8 Krogh A, Brown M, Mian IS, Sjolander K, Haussler D (1994) Hidden Markov  
9 models in computational biology. Applications to protein modeling. *Journal of*  
10  
11 *Molecular Biology* 235:1501-1531 doi:10.1006/jmbi.1994.1104  
12  
13

14  
15 Küberl A, Schneider J, Thallinger GG, Anderl I, Wibberg D, Hajek T, Jaenicke S,  
16  
17 Brinkrolf K, Goesmann A, Szczepanowski R, Pühler A, Schwab H, Glieder A,  
18  
19 Pichler H (2011) High-quality genome sequence of *Pichia pastoris* CBS7435.  
20  
21 *Journal of Biotechnology* 154:312-320 doi:10.1016/j.jbiotec.2011.04.014  
22  
23

24  
25 Kurtzman CP (2005) Description of *Komagataella phaffii* sp. nov. and the transfer  
26  
27 of *Pichia pseudopastoris* to the methylotrophic yeast genus *Komagataella*.  
28  
29 *International Journal of Systematic and Evolutionary Microbiology* 55:973-976  
30  
31 doi:10.1099/ijs.0.63491-0  
32

33  
34 Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H,  
35  
36 Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG  
37  
38 (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947-2948  
39  
40 doi:10.1093/bioinformatics/btm404  
41

42  
43 Macauley-Patrick S, Fazenda ML, McNeil B, Harvey LM (2005) Heterologous  
44  
45 protein production using the *Pichia pastoris* expression system. *Yeast* 22:249-270  
46  
47 doi:10.1002/yea.1208  
48

49  
50 Malik MS, Park ES, Shin JS (2012) Features and technical applications of omega-  
51  
52 transaminases. *Applied Microbiology and Biotechnology* 94:1163-1171  
53  
54 doi:10.1007/s00253-012-4103-3  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3 Mattanovich D, Callewaert N, Rouze P, Lin YC, Graf A, Redl A, Tiels P, Gasser B,  
4 De Schutter K (2009) Open access to sequence: browsing the *Pichia pastoris*  
5 genome. *Microbial Cell Factories* 8:53 doi:10.1186/1475-2859-8-53  
6  
7

8  
9  
10 Mehta PK, Hale TI, Christen P (1993) Aminotransferases: demonstration of  
11 homology and division into evolutionary subgroups. *European Journal of*  
12 *Biochemistry* 214:549-561  
13  
14  
15

16  
17 Payne WE, Kaiser CA, Bevis BJ, Soderholm J, Fu D, Sears IB, Glick BS (2000)  
18 Isolation of *Pichia pastoris* genes involved in ER-to-Golgi transport. *Yeast* 16:979-  
19 993 doi:10.1002/1097-0061(200008)16:11<979::AID-YEA594>3.0.CO;2-C  
20  
21  
22

23  
24 Qian B, Goldstein RA (2003) Detecting distant homologs using phylogenetic tree-  
25 based HMMs. *Proteins* 52:446-453 doi:10.1002/prot.10373  
26  
27

28  
29 Richardson SM, Mitchell LA, Stracquadanio G, Yang K, Dymond JS, DiCarlo JE,  
30 Lee D, Huang CL, Chandrasegaran S, Cai Y, Boeke JD, Bader JS (2017) Design  
31 of a synthetic yeast genome. *Science* 355:1040-1044 doi:10.1126/science.aaf4557  
32  
33  
34  
35

36  
37 Robert X, Gouet P (2014) Deciphering key features in protein structures with the  
38 new ENDscript server. *Nucleic Acids Research* 42:W320-324  
39 doi:10.1093/nar/gku316  
40  
41  
42

43  
44 Shen BW, Hennig M, Hohenester E, Jansonius JN, Schirmer T (1998) Crystal  
45 structure of human recombinant ornithine aminotransferase. *Journal of Molecular*  
46 *Biology* 277:81-102 doi:10.1006/jmbi.1997.1583  
47  
48  
49

50  
51 Shin JS, Yun H, Jang JW, Park I, Kim BG (2003) Purification, characterization, and  
52 molecular cloning of a novel amine:pyruvate transaminase from *Vibrio fluvialis*  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3 JS17. Applied Microbiology and Biotechnology 61:463-471 doi:10.1007/s00253-  
4 003-1250-6  
5  
6

7  
8 Templar A, Woodhouse S, Keshavarz-Moore E, Nesbeth DN (2016) Influence of  
9 *Pichia pastoris* cellular material on polymerase chain reaction performance as a  
10 synthetic biology standard for genome monitoring. Journal of Microbiological  
11 Methods 127, :111-122 doi:10.1016/j.mimet.2016.05.013  
12  
13  
14  
15

16  
17 Valli M, Tatto NE, Peymann A, Gruber C, Landes N, Ekker H, Thallinger GG,  
18 Mattanovich D, Gasser B, Graf AB (2016) Curation of the genome annotation of  
19 *Pichia pastoris* (*Komagataella phaffii*) CBS7435 from gene level to protein function.  
20 FEMS Yeast Research 16:fow051 doi:10.1093/femsyr/fow051  
21  
22  
23  
24

25  
26 Van der Klei IJ, Yurimoto H, Sakai Y, Veenhuis M (2006) The significance of  
27 peroxisomes in methanol metabolism in methylotrophic yeast. Biochimica  
28 Biophysica Acta (BBA)-Molecular Cell Research 1763:1453-1462  
29 doi:10.1016/j.bbamcr.2006.07.016  
30  
31  
32  
33

34  
35 Ward J, Wohlgemuth R (2010) High-Yield Biocatalytic Amination Reactions in  
36 Organic Synthesis. Current Organic Chemistry 14:1914-1927  
37 doi:10.2174/138527210792927546  
38  
39  
40  
41

42  
43 Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ (2009) Jalview  
44 Version 2--a multiple sequence alignment editor and analysis workbench.  
45 Bioinformatics 25:1189-1191 doi:10.1093/bioinformatics/btp033  
46  
47  
48

49  
50 Wei YC, Braun-Galleani S, Henriquez MJ, Bandara S, Nesbeth DN (2018)  
51 Biotransformation of beta-Hydroxypyruvate and Glycolaldehyde to L-Erythulose by  
52 *Pichia pastoris* strain GS115 overexpressing native Transketolase. Biotechnology  
53 Progress 34:99-106 doi:10.1002/btpr.2577  
54  
55  
56  
57



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

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

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

**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 transaminases from *K. phaffii* (*P. pastoris*) GS115, transaminase F2XBU9 from *V. fluvialis* JS17 and transaminase Q7NWX4 from *C. violaceum* DSM30192. Red colour indicates residues of highest homology by BLOSUM62 scores. Residues conserved in  $\omega$ -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:  $\alpha$ -helices are displayed as squiggles,  $\beta$ -strands as arrows, strict  $\beta$ -turns as TT letters and 310-helix as  $\eta$  symbol

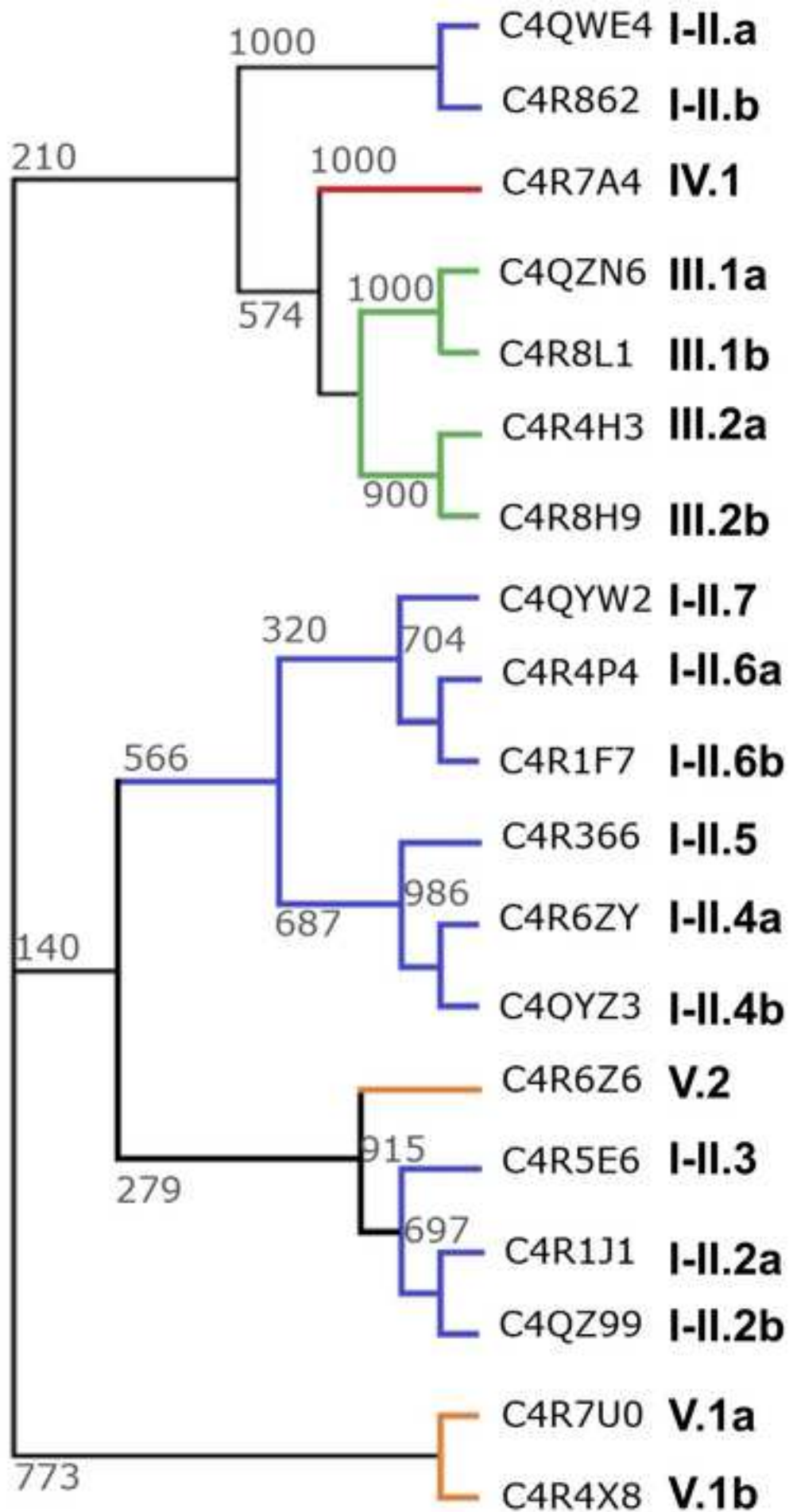


Figure 2

CV2025\_TAm .....MQKQRTTSQWRELDAAHHLHPFTDTASLNQAGARVMTRGEGVYLW**D**SE**G**NKI**I**DGMAGLWCNVN**G**YGRKDF  
Vf1\_TAm .....MNKPQSWEARAETYSLYGFTDMPSLHQRTVVVTHGEGPIY**D**VN**G**RRYL**D**DANSLWNMVA**G**FDHKGL  
KpTamIII.1a .....MSKNEQELSQLGSVFDTRAAYFVCN.....YFASHDNYL**D**LD**G**NEYL**D**VYQAQISSIAL**G**YNHP**E**L  
KpTamIII.1b MSV..SEYYPSEPTAPVVKTSVIPGPEKQSTVFDTRPVYFVAD.....YEKSNNGYI**D**V**D**GNVYL**D**VYQAQIASIAL**G**YNNPAL  
KpTamIII.2a .....MSPNY..NI...SSETAVQYENEYSAHNY.....HPLPVVFHAKAKGAHV**D**PE**G**KEYL**D**FLSAYSAVN**G**CHHPKI  
KpTamIII.2b MKCSLRLTTLVAKSTRMAQRSVVCKY..STQPNKQEEFVKERENYTVTTY.....SRPNLVFEKQGGSYLW**D**IE**G**GKY**I**DFTAGIAVTAL**G**HANPKI

CV2025\_TAm AEAARRQMEEL.....PFYNTFFKTTTHPAVVELSSLLAEVTPAGFDRVFYTN**S**GS**S**VD**T**MIRMVRRYWDV**.**Q**G**K...P.....E.KKT  
Vf1\_TAm IDAAKAQYERF.....PGYHAF**F**GRMSDQ**T**VMLSEKLV**E**VS**P**FD**S**GRV**F**YTN**S**GS**E**AN**D**TMVKMLWFLHAA**.**EG**K**...P.....Q.KRK  
KpTamIII.1a IKVAKSDEMAVALVNRPALGCFPS...DYRTILEEGILAAAPSLDKVWTSLSGSDAN**E**TAFKA**A**FMYHALQKR**G**K**T**PFTE**E**EMK**S**CMENLPPG**C**PDYV  
KpTamIII.1b IEAAKSP**E**MIRALVERPALGN**F**PGK..DFKQ**I**LD.NILKVAPK**Q**DKI**W**SGLSGADAN**E**LAFKA**A**FMHY**Q**AKKR**G**Y**G**TS**F**SE**E**ET**T**MLN**Q**SP**G**SP**E**LA  
KpTamIII.2a IEALVDQ**A**SKL.....TLCSRA**F**SS..D**V**FGVY**A**KY**I**TE..Y**F**GF**E**SVL**P**MNT**G**A**E**AV**E**TA**I**K**I**ARR**W**GV**K**K**G**I...P.....Q**D**E**A**I  
KpTamIII.2b AEILYDQ**A**KKV.....I**H**TSNLY**H**N..L**W**TSELSK**Q**L**V**E..K**T**..K**E**SG**G**M**K**DASRVFLAAL**K**F**A**R**K**Y**G**...K**S**I...A.....E**D**K**I**E

CV2025\_TAm LIGRWNGY**H**G**S**TIGGA**S**LGGMKYMHE**Q**GD**L**P**I**PGMAHIE**Q**P**W**WY**K**H**G**K**D**M.TPDEFGVVAARWLEEKILEIGADKVA**F**V**G****E**P**I**Q**G**AG**G**V**I**VPPATY**W**PE  
Vf1\_TAm ILTRWNAY**H**G**V**TA**V**SA**S**MTG**K**PN**S**V.FGLPLPGFVHL**T**CPHYWR**Y**G**E**E**G**E**T**E**E**Q**F**VARLARELE**E**T**I**Q**R**E**G**AD**T**I**A**G**F**FA**E**P**V**M**G**AG**G**V**I**PPAKGY**F**Q**A**  
KpTamIII.1a ILS**F**E**H**GF**H**G**R**LF**G**SL**S**TT**R**SKAI**H**K.LD**I**PA**F**Q**W**PK**T**PF**P**RL**K**Y**P**LE**F**E**K**EN**S**Q**E**E**R**C**L**E**L**F**S**S**V**I**D**Q**W**K**R**I**V**A**F**I**V****E**P**I**Q**S**E**G****D**N**H**AS**P**Y**F**Q**R**  
KpTamIII.1b ILS**F**K**R**A**F****H**G**R**LF**A**SA**S**AT**C**SK**P**I**H**K.ID**L**PS**F**K**W**PK**A**E**Y**PD**Y**K**P**LD**D**NA**E**Y**N**D**A**E**D**K**R**CL**A**I**V**E**D**IL**T**N**W**H**A**P**I**AA**I**I**I****E**P**I**Q**S**E**G****D**N**H**GS**A**A**F**F**Q**  
KpTamIII.2a VLAAQ**N**NE**H**G**R**T**I**G**I**I**S**MS**T**D**E**AT**Q**D**F**GP**Y**L**K**N**V**GP**Q**.....I**P**GE**A**E**G**T**P**LR**Y**GV**I**E**D**...V**E**RA**F**SN**A**GD**K**I**A**A**I**LL**E**P**I**Q**G**E**A**G**I**V**V**PP**A**D**Y**LP**K**  
KpTamIII.2b F**I**T**F**ET**S**F**H**G**R**SM**G**AL**S**V**T**PN**K**K**Y**Q**A**P**F**AP**L**IP.....G**V**K**V**AK**P**ND**I**.....H**S**VE**K**L**I**S**D**K**T**C**G**V**I**L**E**P**I**Q**G**E**G****G**V**R**P**M**E**A**K**F**L**A**K

CV2025\_TAm IERICRKYD**V**LLVA**D**E**V**IC**G**F**G**R**T**GE**W**FG**H**Q**H**F..GF**Q****P**D**L**FTAA**K**GLSS**G**Y**L**P**I**G**A**V**F**V**G**K**R**VA**E**...GL**I**AG**G**D**F**N**H**GF**T**Y**S****G**H**P**V**C**AA**V**A**H**A**N**V**A**A  
Vf1\_TAm IL**P**IL**R**K**Y**D**I**P**V**IS**D**E**V**IC**G**F**G**R**T**GN**T**W**G**CV**T**Y..D**F**T**P**DA**I**ISS**K**N**L**T**A**G**F**F**P**M**G**AV**I**L**G**PE**L**SK**R**LE**T**A**I**E**A**IE**F**F**H**GF**T**AS**G**H**P**V**G**CA**I**AL**K**A**I**D**V**  
KpTamIII.1a L**R**E**I**S**L**E**K**N**V**L**M**I**V****D**E**V**Q**T****G**V**A**S**T**G**K**F**W**A**H**E**H**W**N**L**T**T**P****D**F**V**T**F**S**K**K**F**Q**A**A**G**F**Y**F**Q**...N**P**E**F**...V**P**N**Q**PF**R**Q**F**N**T**W**C****G**D**P**SK**A**I**I**A**R**T**I**F**K**  
KpTamIII.1b L**R**D**L**T**L**K**Y**S**L**L**I**I**D**E**V**Q**T****G**V**G**A**T**G**T**M**W**A**H**E**H**F**N**L**S**P**A****P**D**M**V**T**F**S****K**K**F**Q**S**A**G**Y**F**F**H**...D**P**E**L**...V**P**N**Y**S**Y**R**Q**F**N**T**W**C**G**D**P**PA**R**M**I**I**A**G**A**I**A**K**E**  
KpTamIII.2a V**Q**E**L**C**K**K**H**N**V**L**L**I**C****D**E**I**Q**T****G**I**G**R**T**G**K**L**L**C**Y**E**H**.S**K**G**V**R**P**D**M**I**L**L**G****K**A**I**S**G**G**V**L**P**V**S**A**V**L**S**S**K**D**I**M**S**V**I**Q...P**G**S**H**G**S****T**Y**G****G**N**P**L**A**C**R**V**A**I**A**L**D**V  
KpTamIII.2b V**R**Q**L**C**D**E**H**N**A**L**L**I**Y****D**E**I**Q**C****G**L**G**R**T**GN**L**W**A**H**A**C**K**L**G**E**E**T**H****P**D**I**L**T**MA**K**AL**G**.N**G**Y**P**I**G**A**T**M**I**T**E**K**V**E**S**V**L**K...V**G**D**H**G**T****T**Y**G****G**N**P**L**G**A**R**V**G**S**Y**V**L**Q**Q**

CV2025\_TAm LRDEG**I**V**Q**R**V**K**D**D**I**G**P**Y**M**Q**K**R**W**R**E**T**F**S**R**.F**E**H**V**D**D**V**R**G**V****G**M**V**Q**A**F**T**L**V**K**N**K**A**K**R**E**L**F**P**D**F**G**E**I**G**T**L**C**R**D**I**F**F**R**N**N**L**I**M**R**A**C**G**D**H**.I**V**S**A**P**P**L**V**M**T**R**A**E**V**D  
Vf1\_TAm V**M**N**E**G**L**A**E**N**V**R**R**L.A**P**R**F**E**E**R**L**K**H**.I**A**E.R**P**N**I**G**E**Y**R**G**I****G**F**M**W**A**L**E**A**V**K**D**K**A**S**K**T**P**D**G**N**L**S**V**S**E**R**I**A**N**T**C**D**L**G**L**I**C**R**P**L**G**Q**S**.V**V**L**C**P**P**F**I**L**T**E**A**Q**M**D  
KpTamIII.1a I**Q**K**D**N**L**V**S**K**I**R**E**V**G**D.Y**L**Y**E**K**L**E**T**V**F**K**S**T**P**..V**T**N**L**R**G**K**G**R**T**F**I**A**W**D**F**D**S**A**Q**E**R**N**E**...F**L**L**K**M**K**Q**H**G**I**N**I**G**G**C**D**S**S**V**R**L**R**T**T**L**I**F**E**K**K**H**A**D  
KpTamIII.1b V**V**D**K**N**L**I**A**N**A**K**E**V**G**D.Y**L**F**G**K**L**E**E**L**S**K**K**Y**P**T**E**L**S**R**L**R**G**K**G**R**A**T**F**I**A**W**D**A**S**S**E**A**R**N**S**...F**L**A**K**M**K**L**N**G**V**N**V**G**G**C**A**D**H**S**I**R**L**R**P**T**L**T**F**G**K**K**H**A**D**  
KpTamIII.2a V**K**D**E**R**L**V**D**R**A**S**E**L**G**D.F**L**F**H**E**L**S**K**L**Q**R**E**S**N**G**V**I**S**E**I****R**G**K****G**L**L**T**A**I**V**I**D**D**S**K**A**N**G**R**T**A...W**D**L**C**L.L**M**K**N**H**G**V**L**A**K**P**T**H**D**H**I**R**L**A**P**P**L**V**I**S**K**E**D**L**L**  
KpTamIII.2b V**S**D**K**D**F**L**S**K**V**E**Q**K**S**E.I**F**K**V**K**L**S**E**L**Q**E**K**F**P**D**L**I**T**D**V****R**G**K****G**L**L**L**G**I**E**F**N**I**D**...P...A**P**I**C**A.I**A**R**E**K**G**L**L**I**I**T**A**G**G**N**V**I**R**F**V**P**A**L**N**I**E**S**K**V**I**Y

CV2025\_TAm EML**V**A**E**R**C**L**E**E**F**E**Q**T**L**K**A**R**G**L**A**  
Vf1\_TAm E**M**F**D**K**L**E**K**A**L**D**K**V**F**A**E**V**A**.....  
KpTamIII.1a I**L**C**D**A**I**L**K**V**L**N**V**.....  
KpTamIII.1b I**L**V**A**T**I**D**K**V**L**S**Q**N.....  
KpTamIII.2a K**G**V**D**A**I**R**T**S**L**A**E**L**P**N**A**P**H**V**E**H..  
KpTamIII.2b E**G**L**A**I**L**E**E**A**V**K**E**F**A**E**N**Q.....

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 <i>et al.</i> (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 $\alpha$ -ketoglutarate as an amino acceptor. The difference between an aspartate and aromatic transaminase is the hydrophobicity of the active site binding pocket (Hwang <i>et al.</i> 2005). |
| Class III.  | Have a wide variety of substrate acceptors and can transfer an amino group to aldehydes and ketones of different types. Include $\beta$ -transaminases, $\gamma$ -aminobutyrate transaminases, and $\omega$ -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 <i>et al.</i> 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. *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           | 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           | C4R277_PICPG | PAS_chr2-2_0492 | Aminomethyltransferase (EC 2.1.2.10)  | 392   | Aminotranferases  |
|             | C4R366           | C4R366_PICPG | PAS_chr3_1132   | Kynurenine aminotransferase, catalyzes formation of kynurenic acid from kynurenine              | 434   | Transaminases     |
| C4R366      | 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               |
| C4R4H3      | 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  |
|             | C4R6Z7           | C4R6Z7_PICPG | PAS_chr4_0147   | Aromatic aminotransferase II  | 462   | Transaminases     |
| C4R6Z7      | 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               |
| C4R7A4      | 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               |
| 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           | 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. Identity percentages obtained in pairwise alignments between the Class III transaminases from *K. phaffii* (*P. pastoris*) GS115, and the reference  $\omega$ -transaminases from *V. fluvialis* JS17 and *C. violaceum* DSM30191.

| <b>Systematic<br/>Nomenclature</b> | <b><i>V. fluvialis</i><br/>JS17</b> | <b><i>C. violaceum</i><br/>DSM30191</b> |
|------------------------------------|-------------------------------------|---|
| 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. Topographical alignment of active site residues between the Class III transaminases from *K. phaffii* (*P. pastoris*) GS115, and the reference  $\omega$ -transaminases from *C. violaceum* DSM30191 and *V. fluvialis* JS17.

| <i>C. violaceum</i><br>DSM30191 | <i>V. fluvialis</i><br>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. 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<br>GS115 Protein | Accession Number<br>CBS7435 | Accession Number<br>S288C | Class<br>(HMMER) | % Identity GS115<br>with CBS7435 | % Identity GS115<br>with S288C |
|--|-----------------------------|---------------------------|------------------|----------------------------------|--------------------------------|
| KpTam I-II.1a                            | F2QML6                      | Q01802                    | I-II             | 100%                             | 40%                            |
| KpTam I-II.1b                            | F2QYY5                      | P23542                    | I-II             | 93.20%                           | 57%                            |
| KpTam I-II.2a                            | F2QS71                      | P09950                    | I-II             | 100%                             | 68%                            |
| KpTam I-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<br>P52893          | I-II             | 100%                             | 58%<br>59%                     |
| KpTam I-II.6b                            | F2QSA5                      | P07172                    | I-II             | 100%<br>100%                     | 53%                            |
| KpTam I-II.7                             | F2QQ23                      | P10356                    | I-II             | 100%                             | 40%                            |
| KpTam III.1a                             | F2Q TZ9                     | -                         | III              | 100%                             | -                              |
| KpTam III.1b                             | F2QYI5                      | P17649                    | III              | 100%                             | 68%                            |
| KpTam III.2a                             | F2QWH8                      | P50277<br>P07991          | III              | 100%                             | 27%<br>70%                     |
| KpTam III.2b                             | F2QYL6                      | P18544                    | III              | 98%                              | 51%                            |
| KpTam IV.1                               | F2QZT3                      | P38891<br>P47176          | IV               | 100%                             | 73%<br>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                | -                                | -                              |