

File S4: Sequences and alignment of beta-xylosidase (GH3)

```
>1_Blastobotrys.raffinosisfermentans_ARAD1D50644
MIFSLLPFAGLALAAQAQTGYVNYSQEAQPNLTPQTVETITIKYPDCSSGPLANTVVCNASADYRERAWALVSMMTFDELVNNTQNTAPGVPRLG
LPYPQWWNEALHGLDRANFSAEFVSEEFGWATSFQPPILTAAALNRSLIHTIGDIIISTQARAFSNNKGYLDAYSNPNINGFRSPLWGRGQETPGE
DANFLSSVYAYEYITGLQGPDPANQIKIAATAKHFAGYDLENWDGHSRLGNDLFISSQDLAEYYTPQFLTAIRDAKSVSMLMCSYNAVNGVPSCA
SPYFLQTLILRESWGFDYDGYVSSDCDAVYNVFNPHNYSSTDALAAADSMRAGTDLDCGQEQYEPWHLTEAFANGDITRSEVELAVYRLYATLVRL
GYFDTNHYPYRKLGWKDVVDNTNNLNVTYEAAVQGITLLKNDGTLPKKEGKSVALIGPYANATEQMKGNYFGPAPYVITPLQAAKSMGLKATYAFG
NLISSTSGKEFFDDALKVAKESDEIIFFGGIDNTIEAEGMDRMNLTWPGNQLELISELSKCGKPLTVVQFGGGQVDSSSLKNNHNVNLSLVWAGYP
GQSAGAAIMDILTGTKRAPAGRLVTTQYPAEYAKQLPATDMSLRPNKSTNNPGQTYMWYTGKPVYEFHGHLFYTTFTKDTLTSEAKNTFNLTKVLT
AAHPGYDSPDLQTLNLTVTVTNTGDAVDYTNMLFANTKNAGPAPYPNKWLVGFDRLSSIQPGSSSKTLTIPVPVGALSRVDEDGNRVLYPGDY
ELHLGTDKEDKVKASVKLTGKAVTLTHWPADNF
```

```
>2_Blastobotrys.raffinosisfermentans_ARAD1C05676
MKWTTSTALTAVAVSGVQAANWFGPPYPDQNEFPDCANGPLKDNDVCNTELDPIITRAKGLVSAMTREEKIVNVQHNATGVPRGLGLAPYNWSEAL
HGVASSPAVVFEDESGDFSSATSFQPPILMGAAFDDKLI F DVATVVGTEARAFGNHQKAGLDFWTPNINPYRDPRWGRGQETPGEDPYHLQNYVY
NLSSTSGKEFFDDALKVAKESDEIIFFGGIDNTIEAEGMDRMNLTWPGNQLELISELSKCGKPLTVVQFGGGQVDSSSLKDNHNVNLSLVWAGYP
HWNWRGIGNWVTGDCGAVEDIYKTHNYTKDAASAAALALNAGTDLDCGETFPEGLPAAFDQGGVKESTLDQALVRLYSSLVHLGYFDPEGSQKY
RRLGWSVDVNQPHAQDLAYKAAVEGIVLLKNDGILPIKPDGQKVAVVGPPLSQATQMQGNYFGVAPHLHSPQYGAQQQNQWQVSQANGTFIESTNR
TGFDDAISAAKDADVIFYVGGIDNIEIEGLDRMTISWPQNQLDLIDELSKLGGKLVVQMGGGQVDDTPLKNNDKINAI VWAGYPGQDGGVAI
FDIITGKVSAPGRLPVTQYPAEYVDQVPMTDMDLRPSDSNPGRTYRWYDFGYGLHYTDFEVSFKGKNDKKYKVDQVIHSGKKVDLNDQK
VFDTFKVEVENKGDVTSYIALAFIKTDNAGPGPYPLKTLAGYTRAFSIAPGEKRVKVIDVPVQGVARVDEKGNRVLYPGDYTTLEIDVDGANGG
ASISFSISGDPYTIIEFPQPDF
```

```
>1_Blastobotrys.adeninivorans_XYL1
MIFSLLPFAGLALAAQAQTGYVDYDREAQPNLTPQSAGTITVKYPCDSNGPLANTIVCNASADYRERAWALVSMLSFDELVNNTQNTSPGVPRLG
LPYPQWWNEALHGLDRANFSADVGEEFGWATSFQPPILTAAALNRTL IHTIADIISTQARAFSNNKGYLDAYSNPNINGFRSPMWGRGQETPGE
DANFLSSVYAYEYITGLQGPDPANNIKIAATAKHFAGYDLENWDGHSRLGNDLTISQDLAEYYTPQFLTAIRDAKSVSMLMCSYNAVNGVPSCA
NPYFLQTLILRESWGFDYDGYVSSDCDAVYNVFNPHNYSSTDALAAADSMRAGTDLDCGQEQYEPWHLTEAFANGDITRSEVELAVYRLFATLVRL
GYFDTNHYPYRQLGWKDVVDNTNNLNVTYEAAVQGITLLKNDGTLPKQKGKSVALIGPYANATEQMKGNYFGPAPYIITPLQAAKSMGLKATYAFG
TNISSTSGKEFFDDALKVAKESDEIIFFGGIDNTVEAEGVDRMNVTWPGNQLELISELSKCGKPLTVVQFGGGQVDSSSLKDNHNVNLSLVWAGYP
GQSAGAAIMDILTGTKRAPAGRLVTTQYPAEYANQFPATDMSLRPNKSTNNPGQTYMWYTGKPVYEFHGHLFYTTFTKIDLTSEAKNTFNLTEVLS
TPHPGYDSPDLQTLNLTVTVTNTGDAVDYTNMLFANTKNAGPAPYPNKWLVGFDRLSSIQPGSSQTLTIPVPVGALSRVDEDGNRVLYPGDY
ELNLGTDNEDKVKASVKLTGKAVTLTHWPADDNF
```

```
>2_Blastobotrys.adeninivorans_XYL1
MKWTTSTALTAIAVSGVQAAEFWGHYPNQNEFPDCGNGPLKDNDICNTDLDPITRAKGLVSAMTRDEKIVNVQHNATGVPRGLGLAPYNWSEAL
HGVAESPAVVFEESGDFSSATSFQPPILMGAAFDDKLI FDIATVVGTEARAFGNQKAGLDFWTPNINPYRDPRWGRGQETPGEDPYHLQNYVY
SLLSGLEGPKDEPYKRI VATCKHYAGYDLENWNGITRHEFDAQISTQDLSEFYTPSFRTCARDAKVGAFMCTYNNMVNGIPTCGNPYFLDTLLRD
HWNWRGIGNWVTGDCGAVEDIYKYHNYTKDAASAAALALNAGTDLDCGETYPEGLPAAFDQDQVKESTLDQALVRLYSSLVHLGYFDPEGSQKY
RRLGWSVDVNQPHAQELAYKAAVEGIVLLKNDGFLPVKPEGQKVAVVGPWSQATGNGNVSRLGLNABEITQELSEYTPQFLAASRYAKTYSLMCSYNSVN
TGFDDAISAAKDADVIFYVGGIDNTIEIEGLDRMNISWPQNQLDLIDELTKLGGKLVVQMGGGQLDDTPLKNNDKINAILWG GYPGQDGGVAI
FDIITGKAAPAGRLPVTQYPAEYADQVPMTDMSLRPSDSNPGRTYRWYDKAVYQFGYGLHYTDFEVSFKGKNDKKYKVDQVIHSGKKVENLNDQK
VFDTFKVEVENKGDVTSYVALAFIKTDNAGPAPYPLKTLAGYTRAFSIGPGEKRVKVIDVPVIGQVARVDEKGNRVLYPGDYTTLEIDVDGANGG
ASVSFVSVDGDPFTIIEFPQPDF
```

```
>Blastobotrys.mokoenaii_XYL1
MVSSGTLTAVATALLATPALAQNNQTYANYSEKAQPDLYPLTLANIKRGFPDCENGPLKDNLVCDTSANYFERAQUALISLFTTEELVNNTNRNS
GPGVPRLGLPYPQVWNEALHGLARANETKEGEWSWSTSFQPPILSMAEFNRTLINQIGDIIATQARAFSNEERFGLNVYSPNININGFRSPLWGRG
QETPGEDPFPHISGYVYNLIQGLQGGLNPDQLKL VATCKHFAGYDLESWEEHSRLGFDAIISERDLVEYYLPAFQSCVRDAKAASVMCSYNAV N
GVPSCANSFLLQTLRLDSWDFPEHGYVSSDCGAVYNVFNPHGYAHNTTSAADALRAGTDIDCGETFPDFLDDSFDEGYVSRGEIEKGLTRLYA
TLVRLGYFDGKNAKYRDLSDWNVVQTDSDWNISYEAAVEGITLLKNDGTLP LSKSVKKVALIGPWANATEQLLGNYFGTTPPYMV SPLQGAKDAGY
EVNFATGTGISDGSTEGYSDAIAAAKSDV IIFAGGIDNTIEAEGQDRDIAWPGKQLDLIDKLSKLKPLVLVLMQGGGQVDSSSLKKNKNVNS
LVWGGYPQSGGHAIFDILRGKRAPAGRLVTTQYPAEYAHLPAYDMLNRPNGSNPGQTYI IWTGEPVYEFHGHLFYTFKFKETRSKGIKEHQTF
NLQDLLEKHPHGYKYPELAPFLNFTVDVKNTGHTESP YTGLVYARTTAGPKPLPNKWLVGFDRLPEIKPGHSSTMTIPIPIGAIVRADENGSKT
LYPGDYELALNNSERSVITFTLTGEPVTMENWPLYHQAVDGYVQTTIAGPVH
```

```
>Sugiyamaella.lignohabitan_XP_018735137.1_hypothetical_protein_AWJ20_922
MKSLTQLSAAIGLVGSALAQNGNNYVNFSSSEANPQLTSYTLLEHLNYSFPDCTREPLKSNLVCDTSAHYLDRARALVNEFTLQELINNTDNTAP
GVPRLGIPNYQWWEALHGIASSPGMQYAESGEYSHSTSFQPPILMGAAFDDDLIKRVATVSTEGRAFNNKGLYGLDVWSPNINPFKDPRWGR
GQETPGEDPFPHISGYVYNLIQGLQGGLNPDQLKL VATCKHFAGYDLESWEEHSRLGFDAIISERDLVEYYLPAFQSCVRDAKAASVMCSYNAV N
GVPSCANDFFLLSILRKEWGFDEYNGYVTSDCDAVYNVFNPHNYSDETFGAAAASINAGCDLNCGQTYSDTL PQAFAQNLVTRDTIEDSVVRLY
ANLVRLGFFDPANTQPYRQLGWTDVSTNSSNLAYQAAVEGIALLKNDGTLP LNSQIKKLALIGPWANATTQMKGNYEGVAPFLISPLQAAQNA
GFQVQYNI GTLINSTSTANFSQALQIAKDADAVIYVGGIDNTIEAEAQDRVNITWPGNQDLISQLSGLSKPLIVLQMGGGQIDSSSLKNNSNV
NGLLWGGYPGQSGGQAILDILTGVVAPAGRLPTTQYPADYINQVPMTNMSLRAGDNNVGRTYMWYTGEPVYEFHGHLFYTFKFKETRSKGIKEHQTF
NIQDLIQNAHPGVDYVDQINFATFTAGITNTGKVASDYVALLFASGDAGPAPQPKRRLVSYSRAHDIA PQQTAAQVKLPLTLGAISRVDENGNRV
LYPGTYNLALDDTTGLITTTVTLTGNAQIIITAWPTLGSDSFVKGQLLTQL
```

```
>Trichomonascus.ciferrii_KAA8906613.1
MRVSTLSYWTIASSLASLVSTAYPDCSKEPLKSNVDVCNTELDAKQRAALIKETWLEEKINNTQNASPGVSRLGIPPYDWWNEALHGVADRHGV
FFAEPGEEFSYASSFPSITMGSAFDDDLIFEVASTIGKEARAFGNHSGFDYWTNPNINPFKDPRWGRGQETPGEDPIHLQKYVYKLLSGLEG
DDDEYLQIAATCKHYAAYDLEDWQGMDFHFDAKVSMDLNEYLLPTFRTCGRDFKVGAFMCSYNSINGEPGCANSYLLQDVLDRDHGWDRDDQ
WVTSDCGAIENIWHDHNTTKDYQASAKALLAGTDLDCGTTFNNSLQEA YDEKLIKESDLDKALVRLYTLVLRVTGYFDPEDKVSYRQLGWKDVN
TEEAQKLT YEAAVEGIVLLDNKKDTP LPSADKCGKVAVVGPYANATDMLNGYHGIIPPYHISPLDAVKKMGLDVSFYDQIPMNSTNSTGIDEAL
EAKNADTILFYIEQEGHDRDTIVWPNQLK LIDKLAELKKTLLVVIQFGGGQVDDTPLLNDNDVNVAIVWVTGPGQEGGHAILDILT GK
KAPAARLPITQYPGDYVDQVPMTDMSLRPSDSNPGRTYRWYNEEVLPGYGLHYTTFDVSWADEPADSYDITDVVQSAKQNAEYVDKAPFDITYK
```

VSVKNTGDVASDYVALLFLKTQNAGPAPYPRKTLIGYARAFDVQPGESQTVSIEVEVGAIARYDEQGNVLYNGDYALQVDLEDNKGPTAKFSL
SGKSERIEEFPQPK

>1_Spathaspora.passalidarum_NRRL_Y-27907_XP_007373774.1

MLPIKLLATLALAI FANGETPPSSPDYNT EANPQLPPITFEAVHYTFPDCQNGPLKHNAV CNPHLPTEQRAKAVVDLFTVDEL IANMGNTSPGV
ERLGLPPYQWWSEALHGIARSNFTASGEYSHATSFPQPILMGGAFNNDLYKQVGNVIGTEARAFNNVGRAGLDFYSPNINPFRDARWGRGQEV
SESPVLVGNYALNYVQGLQGGGLDSNQDDTLQVAATCKHFVGYDMESWNQHSRLGYNAIISDQDLADFYLPFTFQSCVRDAKAAGAMCSYNAVNG
VPACASEFFLNTVLRDGFDFQNGVIHSDCAIYNVWNPHLYAQDLGGAAADAIKAGVDVNCGDYQNNLGYALGNKTINENQIRTSVTRQYSNL
IRLGYFDSFPQTNKYRKYDWNVDVSTPQANQLAYQA AVEGIALLLKNDGTLFPNKQKVRKVAVIGPWANATTQMLGDYAGTPPYMISPLQGAQSEGF
QVEYALGTQINTTDTSGYTAALNAAKGADAI VYFGGIDNSVENEALDRESLAWPGNQLDLVSKLSGLKKPLVVLQFGGGQIDDTEIKNNKNVNA
IVYAGYPGQSGGTAIWDILSGKYAPAGRLTTTQYPASYADQVPMTDMTLRPRQGYPGRTFMWYNGEPVYEFYGLHYTTFSASLANAPRGGHQS
FNIEQVVAAAKRSQYVD TGLITTFDVNIKNTGKTTSDYAALLYSKTTAGPGPHPNKILVSFDKLHQIHAGQTQTAKLPVTIGSLLQTD TNGNKW
LYPGTYTFFVDNDKKAQWEITLTGQAELIQKYPSQK

>2_Spathaspora.passalidarum_NRRL_Y-27907_XP_007375862.1

MLPIKLLATLALSIFAIGETPPSFDPDYNTESNPQLPPITFQAVHFSFPDCQNGPLKDNDVCNPYLPNNQRAKAVVDLFTVDEL IANMGNTSPGV
ERLGLPPYQWWSEGLHGIARSNFTASGEYSHATSFPQPILMGGAFNSDLYKQVGNVIGTEARAFNNVGRAGLDYYSNINPFKDPRWGRGQEV
SESPVLVGNYALNYVQGLQGGIDSNPNDDTLQVAATCKHFAGYDMESWKQHSRLGYNAIISDQDLADYYPFTFQSCVRDAKAAGAMCSYNAING
IPVCASEFFLGTVIREGDFQNGVIHSDCDSLYSIWNPHLYVQDLGAAAADGIKAGVDVNCGDYQNNLGYALGNKTINEDQIRASVTRQYSNL
IRLGYFDSFPQTNKYRTYNWSDVSTSQANQLAYQA AVEGITLLKNDGTLFPNKDKVKNVAVIGPWANATTDM LGDYAGTPPYLISPLQGAQDSGF
KVQYAYGTQINTTLTTNYTAALNAAKGADAI VYFGGIDNSIENEALDRESLAWPGNQLDLVSKLSGLNKPLVVVQFAGQVDDTEIKNNNNVNS
IVYAGYPGQSGGTAIWDVLNGIYAPAGRLSTTQYPASYADQVPMTDMTLRPRDGYPGRTFMWYNGEPVYEFYGLHYTTFSVSLANAPPKGAPQ
SFNIDQFIAAKSSQYVD TSLITTFDVNIKNTGKVTSDYAALLYSNNTTSGPGPHPNKILVSFDKLHQIHPGQIQTASLPVTIGSLLQTD TNGNKW
LYPGAYTFFVDNDMKAQWEITLTGQAALVQNYPSQS

1_Blastobotrys.raffii
1_Blastobotrys.adeni
Blastobotrys.nokoena
Sugiyamaella.lignoha
1_Spathaspora.passal
2_Spathaspora.passal
2_Blastobotrys.raffii
2_Blastobotrys.adeni
Trichomonascus.cifer

131 140 150 160 170 180 190 200 210 220 230 240 250 260

1_Blastobotrys.raffii
1_Blastobotrys.adeni
Blastobotrys.nokoena
Sugiyamaella.lignoha
1_Spathaspora.passal
2_Spathaspora.passal
2_Blastobotrys.raffii
2_Blastobotrys.adeni
Trichomonascus.cifer

261 270 280 290 300 310 320 330 340 350 360 370 380 390

1_Blastobotrys.raffii
1_Blastobotrys.adeni
Blastobotrys.nokoena
Sugiyamaella.lignoha
1_Spathaspora.passal
2_Spathaspora.passal
2_Blastobotrys.raffii
2_Blastobotrys.adeni
Trichomonascus.cifer

391 400 410 420 430 440 450 460 470 480 490 500 510 520

1_Blastobotrys.raffii
1_Blastobotrys.adeni
Blastobotrys.nokoena
Sugiyamaella.lignoha
1_Spathaspora.passal
2_Spathaspora.passal
2_Blastobotrys.raffii
2_Blastobotrys.adeni
Trichomonascus.cifer

521 530 540 550 560 570 580 590 600 610 620 630 640 650

1_Blastobotrys.raffii
1_Blastobotrys.adeni
Blastobotrys.nokoena
Sugiyamaella.lignoha
1_Spathaspora.passal
2_Spathaspora.passal
2_Blastobotrys.raffii
2_Blastobotrys.adeni
Trichomonascus.cifer

651 660 670 680 690 700 710 720 730 740 750 760 770 780

1_Blastobotrys.raffii
1_Blastobotrys.adeni
Blastobotrys.nokoena
Sugiyamaella.lignoha
1_Spathaspora.passal
2_Spathaspora.passal
2_Blastobotrys.raffii
2_Blastobotrys.adeni
Trichomonascus.cifer

781 790 800 810 820 827

1_Blastobotrys.raffii
1_Blastobotrys.adeni
Blastobotrys.nokoena
Sugiyamaella.lignoha
1_Spathaspora.passal
2_Spathaspora.passal
2_Blastobotrys.raffii
2_Blastobotrys.adeni
Trichomonascus.cifer

Alignments were performed with the Multalin server (<http://multalin.toulouse.inra.fr/multalin/>). The two *B. raffinosifermentans* ARAD1D50644 and ARAD1C05676 predicted proteins were aligned with two *B. adeninivorans* Xyll1 (personal data), *Blastobotrys nokoena* Xyll1, *Sugiyamaella lignohabitans* XP_018735137.1 *Trichomonascus ciferrii* KAA8906613.1 and *Spathaspora passalidarum* XP_007373774.1 and XP_007375862.1 conserved proteins

Amino acids 57 to 752 of the *B. raffinosifermentans* ARAD1D50644 protein and 46 to 740 of the *B. raffinosifermentans* ARAD1C05676 protein correspond to the conserved PLN03080 beta-xylosidase domain