SUPPLEMENTARY INFORMATION

METAGENOMIC AND FUNCTIONAL ANALYSIS OF HINDGUT MICROBIOTA OF A WOOD FEEDING HIGHER TERMITE

TABLE OF CONTENTS

GLYCOSIDE HYDROLASE CATALYTIC DOMAINS AND CARBOHYDRATE BINDING MODULES NOT REPRESENTED BY PFAMS. For several CAZy GH and CBM families, no Pfam HMM is available. In these cases, representative sequences were selected from the CAZy website, and the sequence region corresponding to the family of interest was determined. These sequence regions were then used in BLAST searches against the termite gut community metagenome to identify potential sequences belonging to these CAZy GH and CBM families. An e-value cutoff of 10^{-6} was used in these searches. The following are the partial sequences representing such GH and CBM families:

>GH110==29610934 [Streptomyces avermitilis MA-4680]
MAHGCSGGAMSRFVFLGVALALLGGATSPAAAAPRVTPVVVDVDDYGADPTGRTDSTPAVAA
ALRHAKSVDRPVRIVFSKGTYQLYPERAETRELYMSNTVGADQRYRDKKIGLLVEDMHDVTVD
GGGAKLVHHGLQTAFASIRSTDVTFQNFSFDYAAPEVIDATVATTGVTDGHAYRVLKIPAGSPYR
VNGTHITWLGETSPATGQPYWSGVDGLQYTQIHDPEAQRTWRGDNPLFNDVAAVTDLGGRRIRI
DYTTAARPADAGLVYQMRLIERTEPGAFIWESKNVTMRSMNAYYLQSFGVVGQFSENISIDKVN
FAPDPRSGRSTASFADFVQMSGVKGKVSITRSLFDGPHDDPINIHGTYLEVVGKPGPSTLTLAYKH
PQTAGFPQFAPGDEVEFATKRTMTPLADAHAQVTAVDGPSGMDHTKPLTTMTVTFDRPVPAGV
ETGGTVVENITATPSVVISGNVFRNVPTRGILVTTRKPVLITGNRFDGMSMASIYVSADAYQWYE
SGPVADLTIRGNSFTRPSGPVIFVEPTNQVIDPATPVHHNISVEHNSFDIGDVTVVNAKSVGGFAFT
GNTVRRLDGADHPPYTSPLFVFHGSSGIRIARNHYDKGLNTSVVTD

>GH109==6137031 [Streptomyces coelicolor A3(2)]
VRVAVIGLGNRGGGMITGWAAVPGCTVTAVCDIRADRAERAADRLESKGNPRPAEYGGSADSY
ARMLRRDDIDLVYIATPWEFHYEHGRAALLSGRHAVVELPVATELRQLWDLVDTSERTRRHLLL
SENCNYGRNELAMLKAAHDGLFGDLTNGHGGYLHDLRELLFSDTYYTDSWRRLWHTRSTASFY
PMHGLAPIAAAMDVNRGDRMTTLRATTTAPKGLADYRARFVPRDHPSWKETYINGDLVTCMIE
TAKGRTVRAEHDVSSPRPYSRINTLAGSRGIVEDYAGSAPTGARIYVEPDHGGHTWRDFETYRKE
YDHWLWQKVGDDAANNGGHGGMDYVLQWRTVQLMRAGLVPDIDVYDSAAWCSPVPLSVTSL
ARGGRPVEIPDFTRGAWRERRPGLDSAPTD

>GH109==124053159 [Elizabethkingia meningoseptica]
VRIAFIAVGLRGQTHVENMARRDDVEVVAFADPDPYMVGRAQEILKKNGRKPAKVFGNGNYDY
KNMLKDKNIDAVFVSSPWEWHHEHGVAAMKAGKIVGMEVCGALTMDECWDYVKVSEQTGVP
LMALENVCYRRDIMAGLNMVRKDMFGELIHGTGGYQHDLRPVLFNSGINGKNGDGVEFGDKAF
SEAKWRTNHYKNRNGELYPTHGLGPLHTMMDINRGNRLLRLSSFASKARGLHKYVVDKGGESH
PNAKVEWKQGDIVTTQIQCNNGETIVLTHDTSLQRPYNLGFKVQGTEGLWEDFGWGDAAQGFIY
FEKIMNHSHRWDGSEKWMKEYDHPMWKKHEQKAVGAGHGGMDYFLDNTFIECIKRNEAFPLD
VYDLATWYSITPLSEKSIAENGAVQEIPDFTNGKWKTVKNTFAINDDY

>GH107==115530293 [Flavobacteriaceae bacterium SW5]
MKKTLPTKKSNLWFLMACLIITFHKVEAQVPDPNQGLRAEWMRGALGMLWLPERTFNGNIEGIR
IDDFLTQIKDIRTVDYVQLPLTSPNIFSPTHVAPHPIIESLWQGDTDANGDPINLVAPRESVDDPLLS
WLKALRAAGLRTEIYVNSYNLLARIPEDTQADYPDVSARWMEWCDTNTEAQAFINSQTYHEGN
GRRKYMFCYAEFILKEYAQRYGDLIDAWCFDSADNVMEDECGDDPASEDVNDQRIYQAFADAC
HAGNPNAAIAFNNSVGDREGNPFTSATLFDDYTFGHPFGGAGNMVVPEALYTYNHDLVVFMQT
NNGYAFRDDTRTWNDNVVAHFFPKQSTTSWNAGNTPCLTDEQFVEWTSTGIVNGGGITWGTPL
VRTNLENAPVLTLQPYALNQFELTDTYLKEFQSPGKPNWSRQYTILPAIYPGQPYSHNLVEGVDF
WDPEGVGITGLTASGTLPAWLTISQTATGTWTLSGTPPVSEASNYTFELMAQDSDGVTNREVKLE
VISHPAGFTNPGDGTPVWFSNPMVLAKATALKDYGSLLKLGVDFYDFEGDVLTITKTSGPDWLV
LTQNSDDTWRLSGMPTAADAGENSFTFNVSDGILSSDTEIKITVDHVAGFTNLGNGAPVWSSPIL
NLTDGKGSFAYNYTLQLGTDYYDFEGDALTITKTSGPDWLTIQQTDANSWKLSGTPINSDAGENS
FTFNLSDDTNSTTAEILINVIATIISDGNVEIKATANTTYGINTVASMYSAVQTAPDGLATYRISIDV
TPPTDKGIYSGSSGGITTTTSWGIGDGTDAIQNTIFRGSDNEWTESINNIKMVDFNANGGSLTTDN
VTMFFKSISIGNSQSVNDFVSLKVGGVISNPGRSANQYETIDLTLATSVSNIANFAIGTGNDSDTNK
WSVEGITVFVDFAGTLSVTNPIQDDVDSFKLYPNPAKDRIFINKQPVTVQIFDVTGKLVKVDFKGK
NELDISTLKQGLYILKIQTAEGNMLFEKFIKK

>GH106==116228189 [Solibacter usitatus Ellin6076]
ALLAGLALGADPANITELQRGFQHPPDEARIMVRWWWFGPAVTKPELEREMRLMKQGGIGGFE
VQPTYALELDNAEKGIKNLPYMSPEFLEALRFTGEKAKELGLRMDLTLGSGWPFGGPHIPVELAS
PRLRVARAAVTAGAAAINVPELQPSEKLIAAFLGNQELEAAGSQSFRLPPSAAGQVTFFISSHTRQ
TVKRPAVGAEGWVLNHYDRAAIEQHFKSVGEPMLNALAKTPPAAIFSDSLEVYNSDWTGDFLA
QFAKRRGYDLKPYLPALAGDIGEKTGAIRNDWGQTLSELAEENYLKPLTAWAHQHNTKFRSQTY
GIPPVTLSSYSLVDLPEGEGTQWRHFSSTRWATSASHLYNRPVTSSETWTWLHSPVFRATPLDMK
AEADLHFLQGVNQLVAHGWPYSPPSVGEPGWRFYAAAVFNEHNPWWIVMPDIATYMQRVSWM
LRQGKPVNDVAVYLPTHDAFAGFTLGRDSVNQAMEGILGPKLVPTILDAGYGFDFIDDGAIAAG
GVPHKILIVPGAERIPLATYQKIEEYRRDGGKVFFTKRLPSLAPGLREEGDTAKIRELSAKHVVTG
DDGLAGALHDALPADFTAPPEIGVVHRKLAYADVYFLANTSNHAVKGAAKFRASGMPAAWWD
PVTGKSSKADLNLELAPYESRVLVFSRAGGAVMGQAMSLPEVDISSDWTITFPGEQPVKMATLRP
WTDEENRKFYSGTATYERKVTATPAMAAAHVFLNFGEGTVADPASERRAPSGMRAWLESPVRE
AAQVFVNGKPAGAVWKAPYEVDVTGLLQAGENSIRVVVANLALNVLAKGPLPDYKELTAKYG
EKFQAQDMQSVRALPGGMMGPVKLVS

>GH106==45219920 [Sphingomonas paucimobilis]
ALLACAASFAVPSWADELADGFHNPPQSARPRVWWHWMNGNVTKDGISKDLDWMQRVGIGG
VQNFDANLSTPQIVPERLIYMTEQWKDAFRHAVREADARGLEFAIAVSPGWSETGGPWVKPEDA
MKKLVWAEADLRGGQRLKGALPAPPSITGPFQSAQFHEALPGGAEMQTLPAFYRDARVLAYSLV
PKALPRPAVSEGDGSPVDAAPLLDGDLETVARIGKGDAARPGTMVLDYGKPVTVRSASLFVPHA
RPPFGDPDYRPALEVEQDGGWRRIGAFPLTEVATTISFAPVTGRRFRLVLAPNDAPRAPGLGEGAP
GAITMDVFARPQSDTLAIGDFRLLAEARIDRFEAKAGFALVSDYNALAEKSAADLPAIDPAKVVD
LTDRLRPDGTLDWTAPAGSDWRIVRMGYSLTGKTNHPATPEATGLEVDKYDPAAVRRYLETYL
GLYRDAVGAEWIGKKGIRALLTDSIEVGASNWTPRMVEEFRARRGYDPVPFLPALTGAVVGSAA
RSDAFLHDFRQTLADLLADAHYGTIAKVAHEQGLIVYGEALENGRPVLGDDLAMRAHADVPMA
ALWTYNRGSAPRPTLIGDMKGAASVAHIYGQNIVSAESMTAAFSPWAFAPADLKRVIDLEFVSG
VNRPIVHTSVHQPVDDKLPGLSLAIFGQYFNRHDSWAEMARPWVDYMARTGFLLQQGRDHADI
AWFHGEDMPVTALFEHGEPAGLPRRHGYDFVNADILAKVLKVEGGELVAPGGARYRALYLGGS
SRVMTLPTLRRIAQLAEQGATVIGTAPERSPALQDDPTAFRALVARLWNGAPVTPVGQGRVIAET
DVEKALAGIGIGPDFSFAGAGPDADLRFLHRKLADGDLYFIRNGLFRPEKIEARFRVTGRQPELWR
ATDGAVQPLS

>GH104==142853476 [Aeromonas salmonicida subsp. salmonicida A449]
NCHPQMAAFLDLLAFAEGTKGLGDDGYNKLVNPAGLFTDYRTHPNVKVQVNPHLVSTAAGRY
QFLSKHWSHYRDQLGLLDFGPESQDTWAIQLIRERKALADVVDGRIAQAVPKCANIWASLPGAG
YGQREHKLADLLAKFIEFGGVLA

>GH104==12514804 [Escherichia coli O157:H7 EDL933]
NLNPQRKAFLDMLAWSEGTDNGRQPTRNHGYDVIVGGELFTDYSDHPRKLVTLHPKLKSTAAG
RYQLLSRWWDAYRKQLGLKDFSPKSQDAVALQQIKERGALPMIDRGDIRQAIDRCSNIWASLPG
AGYGQFEHKADSLIAKFKEAGGTVR

>GH103==89952579 putative lytic murein transglycosylase [Saccharophagus degradans 2-40]
MFSTLKKVFSTTALAAVITASPLALADYSEHPEAKAFIAKMVNEYGLSEKEVVRILKDANKQQSI
LDAIARPAEKTKPWFEYRNIFLGESRITQGVEFWAENADTLAKASKEFGVPEQIIVAIIGVETRYGR
HAGSYRVIDALTTLGFDYPPRSTFFAKELENFLLLTTEQKQNPLALKGSYAGAMGYGQFMPSSYR
AYAVDFDGDKKADIWNNPKDAIGSVANYFRAHKWQTGEPVMVRARIAEGYDESILDSRSRPSLT
ATEVAAKGFTPVDVEIPGDTKVMPIAYDGEKGKEYWLGYDNFYVITRYNRSHMYARAVWELSE
EILYRHNS

>GH103==MLTB_ECOLI - Escherichia coli
MFKRRYVTLLPLFVLLAACSSKPKPTETDTTTGTPSGGFLLEPQHNVMQMGGDFANNPNAQQFI
DKMVNKHGFDRQQLQEILSQAKRLDSVLRLMDNQAPTTSVKPPSGPNGAWLRYRKKFITPDNV
QNGVVFWNQYEDALNRAWQVYGVPPEIIVGIIGVETRWGRVMGKTRILDALATLSFNYPRRAEY
FSGELETFLLMARDEQDDPLNLKGSFAGAMGYGQFMPSSYKQYAVDFSGDGHINLWDPVDAIGS
VANYFKAHGWVKGDQVAVMANGQAPGLPNGFKTKYSISQLAAAGLTPQQPLGNHQQASLLRL
DVGTGYQYWYGLPNFYTITRYNHSTHYAMAVWQLGQAVALARVQ

>GH101==O86617_STRCO - Streptomyces coelicolor
GRARVRRHGAVVAALGLTAGLLSAAALPAGAAPPRPAAAAAPAGAPTPVELSRGGLTVTVAKE
FPQVISYRLGRRGLDGRATALDGFTVNGESHRATTTVKAKGSRAVYTSTFEDLPGLTITSSITVTK
ETTVVFAVEKISGEAAPGVRTLAIPGQSLVSVDSAEPGANLARTKISTDSTTTADRFVPVTGDTAP
DKGPVGTPYAFVGNAQLSAGIITNATEDSPQDDNTDWNTRLQSRIVDEGEGRRRAELSAGTYTY
HPEGATDPRVDTYELPRATVVLAADANRDGTVDWQDGAIAHREHMRRPLGADRVPERVVQRIP
FNFASQATNPFLKTLDNTKRISMATDDLGQWVLEKGYASEGHDSAHPDYGGNENVRAGGWKD
LNRLTRTGAGYNADFAVHVNATEAYAQARTFTEDMVAGQADGWDWLNQAYHIDQRKDLGTG
AVLDRFKQLRKEAPGIRTVYIDAYYSSGWLADGLAAGLREMGFEVATEWAYKFEGTSVWSHW
AADKNYGGATNKGINSDIVRFIANADRDVWNVDPLLGGASVVEFEGWTGQDDWNAFYRNIWT
DNLPTKFLQHFQVLDWDRGRSARLTGGVDVKSVDGERRISMDGTEVLKGDTYLLPWQNAGKD
DGTSSPRDADKMYFYSASGGEHTFELTGQFAGTEDFTLYELTDQGRAEKARVTAHEGRVTLTAE
KGQPYVLVPNGGRAPHRDAHYGEFTGLSDPGFNGGDLDAWNASGGAEIVRAGNGDNVVRLGE
DASGIAQRVRGLTPGERYTLGADVGIGPGERRETTLRVRGGKDSEARTFDITPARNRMASDEKRD
TYSQRASVSFTAPRDGSVTVELGAVAGGAPVVLDDVRVMVDTTAPLPRSQDGTVVAHDDFEGN
RPGWGPFVKGDAGGVTDPRTSISDLHAPYSQKEWKNTYSPYDTGALKGRAVDDVLAGRHSLKS
HAENTGLVHRTTPATVPFEEGHRYRVSFSYQTNVEGQWAWVTGADRVADGTTTSRDITRDVLA
PALDTAAYSREFVAGCGDTWVGLRRLGSARGTDLVLDDFTVTDLGEADTGAACAAVTAPSGAE
LSPGVPGEYVTAFTNHESAGAENVGIALQGLPEGWKAEVKEKDGNLFERVQPGATVRTTWLLTP
PAGTAGTSATWQVTAAYAHEGATRTVSTGARAAVTDEPVLAPAST

>GH101==Q8XMJ5_CLOPE - Clostridium perfringens
GQVAVSNLPDVRGKVGLTGWFGNKNVTLDNLVVEELGGIMAPEVGPLEEQSIESDSMKVVLDN
RFPTVIRYEWKGTEDVLSGASVDDLEAQYMVEINGEKRIPKVTSEFANNEGIYTLNFEDIGMTITL
KMTVNENKLRMEVTDIQEGDVKLQTLNFPNHSLASVSSLNNGKTASVLTTGDWNNINEEFTDVA
KAKPGVKGKTYAFINDDKFAVTINNNTIEGGNRVVLTTENDTLPDNTNYKKVGISNGTWTYKEIL
QDTTDQGSKLYQGEKPWSEVIIARDENEDGQVDWQDGAIQYRKNMKIPVGGEEIKNQMSYIDFN
IGYTQNPFLRSLDTIKKLSNYTDGFGQLVLHKGYQGEGHDDSHPDYGGHIGMRQGGKEDFNTLIE
QGKEYNAKIGVHINATEYTMDAFEYPTELVNENAPGWGWLDQAYYVNQRGDITSGELFRRLDM
LMEDAPELGWIYVDVYTGNGWNAHQLGEKINDYGIMIATEMNGPLEQHVPWTHWGGDPAYPN
KGNASKIMRFMKNDTQDSFLADPLVKGNKHLLSGGWGTRHDIEGAYGTEVFYNQVLPTKYLQH
FQITKMSENEVLFENGVKAVRENSNINYYRNDRLVATTPENSIGNTGIGDTQLFLPWNPVDEANS
EKIYHWNPLGTTSEWTTGNDWNSDKVYLYELSDLGRTLVKEVPVVDGKVNLEVKQDTPYIVTK
EKVEEKRIEDWGYGSEIADPGFDSQTFDKWNKESTAENTDHITIENESVQKRLGNDVLKISGNEG
ADAKISQSISGLEEGVTYSVSAWVKNDNNREVTLGVNVGGKDFTNVITSGGKVRQGEGVKYIDD
TFVRMEVEFTVPKGVNSADVYLKASEGDADSVVLVDDFRIWDHPGHTNRDGYVFYEDFENVDE
GISPFYLSPGRGHSNRSHLAEKDISIDANQRMNWVLDGRFSLKSNQQPKEIGEMLTTDVSSFKLEP
NKTYEFGFLYSLENAAPGYSVNIKNRDGEKIVSIPLEATGSNYAQDIFTKTKSVTHEFTTGDFAGD
YYITLEKGDGFKEVILDNIYVKEIDKSIESPELAHVNLNTVEHDLEVGQSVPFAINALMNNGANVN
LEEAEVEYKVSKPEVLTIENGMMTGASEGFTDVQVNITVNGNKVSSNTVRVKVGNPEVEEEEVI
VNPVRNFKVTDKTKKNVTVSWEEPEKTYGLEGYVLYKDGKKVKEIGADKTEFTFKGLNRHTIY
NFKIAAKYSNGELSTKESITVRTAR

>GH99==119765938 [Shewanella amazonensis SB2B]
FYYGWYGNPQQDGQWQHWNHRVLPYGDIPLEGRLDFPGADDIGANFYPSLGSYSSHDPEIIEQH
LEMMRQAGIGVVSVSWLGADDFAARSIDFFMDKAAEKGLQINFHIEPNYRSAEEFHAIIAELMRK
FGTHPALYRYRGKPLYYVYDSYKMPVSEWQKLLLPGGELSLRTPELDGQFIGLWVNQGEEAFFL
DTGFDGFYTYFASEGFVWGSTSTNWPYLAGWASRHGKLFIPSVGPGYADDRIRPWNGANFKARE
QGRYYDRMFSQALNTKPDIVTITSFNEWHEGTQIEPAVVKQLPDYRYLDYGDLPEDYYLQRTLD
WSRKLSAIPVNS

>GH99==29341178 [Bacteroides thetaiotaomicron VPI-5482]
FYYNWYGNPSVDGEMKHWMHPIALAPGHSGDVGAISGLNDDIACNFYPELGTYSSNDPEIIRKHI
RMHIKANVGVLSVTWWGESDYGNQSVSLLLDEAAKVGAKVCFHIEPFNGRSPQTVRENIQYIVD
TYGDHPAFYRTHGKPLFFIYDSYLIKPAEWAKLFAAGGEISVRNTKYDGLFIGLTLKESELPDIETA
CMDGFYTYFAATGFTNASTPANWKSMQQWAKAHNKLFIPSVGPGYIDTRIRPWNGSTTRDREN

GKYYDDMYKAAIESGASYISITSFNEWHEGTQIEPAVSKKCDAFEYLDYKPLADDYYLIRTAYW
VDEFRKARSAS

>GH97==89951605 a-glucosidase-like protein [Saccharophagus degradans 2-40]
MHNNSQQIPQRTAHLPPVTPKWRSALRLPVIALALCSLSTISANALAKTFELTSPNEQLKIEVHTG
KQTTFEVHLNNKQVLAPSSIDLMLMSGEKLASNITVKQEKRYAVSNTLTPAVSHKSSKIAENYNA
LNLAFSNNFGIDFRAYNDGFAYRFTGERAGEIIIKNEHLALNFTQNATTLFPEEETFISHFERLYLPE
TVANISPKRFASLPTYIKSNDVNVVVTEADLYDYPGLFLFGTNKNGFTAGFPNAVAKTTPKAGSE
DRNQEFEYENYIAKTSGTRTFPWRVAIITKEDTDLVASQLVYLLSSENKIGDTSWIKPGRIAWDW
YNANNLFDVDFKAGLNTQTYKYYIDFAEKYGLEYVILDEGWTKTTTNTKESNPDIDVKELIAYG
KKKNVAIILWTLWEPLDKDYENILALYAQWGAAGIKVDFMQRADQYMVNYYEKIAKAAAKNK
LLVDYHGGFKPAGLRRAYPNVMTYEGVKGNENNKWSADITPEHNVTLPYIRMVAGPMDYTPGA
LRNKQLANHNVNHFRPEAIGTRAHQLAMYAVFESALQMLCESPSTYLREPLITEYIAKFPSVWDE
TRVLKGEVANYILLARRKGDTWYIGAMTDMTARTLTLNLSFLDGGKYSLHGVADGLNVEQYAE
DYQFIHQTVTAKDSLQLNLAAGGGWSAVITPAK

>GH97==P71094_BACTN - Bacteroides thetaiotaomicron
MKKRKILSLIAFLCISFIANAQQKLTSPDNNLVMTFQVDSKGAPTYELTYKNKVVIKPSTLGLELK
KEDNTRTDFDWVDRRDLTKLDSKTNLYDGFEVKDTQTATFDETWQPVWGEEKEIRNHYNELAV
TLYQPMNDRSIVIRFRLFNDGLGFRYEFPQQKSLNYFVIKEEHSQFGMNGDHIAFWIPGDYDTQE
YDYTISRLSEIRGLMKEAITPNSSQTPFSQTGVQTALMMKTDDGLYINLHEAALVDYSCMHLNLD
DKNMVFESWLTPDAKGDKGYMQTPCNTPWRTIIVSDDARNILASRITLNLNEPCKIADAASWVK
PVKYIGVWWDMITGKGSWAYTDELTSVKLGETDYSKTKPNGKHSANTANVKRYIDFAAAHGFD
AVLVEGWNEGWEDWFGNSKDYVFDFVTPYPDFDVKEIHRYAARKGIKMMMHHETSASVRNYE
RHMDKAYQFMADNGYNSVKSGYVGNIIPRGEHHYGQWMNNHYLYAVKKAADYKIMVNAHEA
TRPTGICRTYPNLIGNESARGTEYESFGGNKVYHTTILPFTRLVGGPMDYTPGIFETHCNKMNPAN
NSQVRSTIARQLALYVTMYSPLQMAADIPENYERFMDAFQFIKDVALDWDETNYLEAEPGEYITI
ARKAKDTDDWYVGCTAGENGHTSKLVFDFLTPGKQYIATVYADAKDADWKENPQAYTIKKGIL
TNKSKLNLHAANGGGYAISIKEVK

>GH96==121309479 [marine bacterium JAMB-A33]
AIPSPIYNPDDHFVAEIQGPQTDVTYLKKPVEIPANKKVLKSDVWYTYPQNRELEGYDNFGATGA
FWGHPPEHDFYDDTVIMDWAVDAVYAFQAEGEYYTARGEFDWGYGWFTEYTTNPQPHYVRTL
DDRNVRMTFMGYLSHDGYNNNWLSNHSPAFVPFMKSQVDQILKANPDKLMFDTQTNSTRSTD
MRDFGGDFSPYAMENFRVWLSKKYSTGELAALGINDINSFDYGDFLRAQGVTHTSWSNAGDTLS
GNIPLQEDYIYFNRDVWNQKFAEVLDYIRQQQPDIEIGASTHLFESRGYVFNENLTFLSGELNLGA
RTTISELPTNILVHLKGAQAVDKTLVYFPYPWEFDELRLQDAPRFGRGWVAQAYAYGGLFSIPAN
VWVGGEVWTWSPGADNYRDIYLFVRAQADLLDDYTSYSKVGLVHAMYSSMKAGFIDGGNQIQ
SSTKLLTEGNINFDLLVFGDEGYPVVPRPEDFDKFDHIFFDGDEQYLTAEQQALLDQQGDKVRHI
GQRGTVSGIEITVSISGTESNETVSAVSRIHETDAAAPYVVHLVNRPFAGGVTPTLNNVEVAIPQSY
FPEVVTGATLHLPDGTSTSLTLSTNADGDVVLPVNNLEVWGILELAH

>GH96==Q9LAP7_9ALTE - Alteromonas agarilytica
AVKQPRVYNPNEHIVAEIQGPATGLQYLKTPVEIPLANKVLKSDVWYTYPQNRNLVVDGDTPYA
DFGATGAFWGHPPEHDFYDDTVIMDWAVNVVDDFQSEGFEYTARGEFDWGYGWFTEFTTNPQP
HYVQTLDGRNVRMTFMGYLSHDGYNNNWLSNHSPAFVPFMKSQVDQILKANPDKLMFDTQTN
STRSTDMRTFGGDFSPYAMENFRVWLLKKYSNAQLVSMGINDITSFDYGAYLRAQGITHTDWSN
AGDTISGNIPMMEDFIYFNRDVWNQKFAEVLEYIRQQRPNIEIGASTHLFESRGYIFNENITFLSGE
LNLGARTSISELPTNILVHLKGAQAVDKTLAYFPYPWEFDELRIQNAPRFGRGWVAQAYAYGGL
FSIPANVWVGGEVFTWSPGADNYRDIYQFVRAQANLLDGYTSYAKAGYVHAMFSSMKAGFIDG
GNQVQSSVKILTEDNINFDMLVFGDAGYPVVPRQADFDKFEYIFYDGDLNYLTAEQQAVLDAQG
SKVKHIGQRGTIAGLQINVSINGSVSNETVSAVSRIHETDSTAPYVVHLINRPFAGGVTPILNNVEV
AIPASYFPQGVTSAKLHLPDGSSSTVAVSTNANGDTVVSVSNLEVWGILELAH

>GH95==71915747 [Thermobifida fusca YX]
MTYSTPTQGTATRLPAASWEEALITGNGRQGVLAHSMQTQIRLTLSHERLFWPREEPLPAPHTAP
ALDELRSLIRAGRPRAAAEHVTELARTEHPGYAHTRWIDPLTGAGTLTFIPDAPAWGPIHRTCDY
TTGVVREELPNGIRHDVFASRADNVIVIRLSADHELAGTLQLAPLPGEPPTPADTRLTSTPHTLTLH
TAFARTWPTTLTGYTLTCRIIAPGGHLRPLEPGRLHLSGARHVLLLVRTTVTSETTPAPGDPDLAR
LTPDFDTLLRRHTAHHTPLITRSRIRLSPRQPHPDPLPGEALLAQGPTPHLMERLVEAGRYATVCAI
GELPPTLQGVWSGTYSPPWSSGYTFDGNLASAVAALHPLGTPELMLPVFDLVDTLLDDFTHNAQ
RLYGCRGILVPPHASTHGRHNHFGPEWCLTLWTAGAAWLARLYWDHYSYTRDPGFLRERALPF
LHHAAAFYEDFLTDDGFVPSYSPENTPADSDSQACANATMDLAAVRDLVRNLLRAHRILDLPGA
RRWAELGARLPRYRIAPTGELAEWATPAGPLDQTDRHTHRHASHLFPFWYEGDPAFADPTLRAA
AVRAIRARLAWWRSAASDEMAFGLVQLGLAAANLGLATEAATTLELMATRYWRPTLVSTHNR
RALFNTDICGGFPAVVAAMLLRSREGRCDLLPALPPAWPEGEARGLCGRGGVLVDHLEWTPTR
MRARLRARHTTRVRIGLPDGATRVVTLDPDQAVDVTAPLTLHQ

>GH95==34451973 - [Bifidobacterium bifidum]
VKATSTADGTVKTLTEGNGEDNYTIDTSAFDSASIGVYPVTVKYNKDPEIAASFNAYVIASVEDG
GDGDTSKDDWLWYKQPASQTDATATAGGNYGNPDNNRWQQTTLPFGNGKIGGTVWGEVSRER
VTFNEETLWTGGPGSSTSYNGGNNETKGQNGATLRALNKQLANGAETVNPGNLTGGENAAEQG
NYLNWGDIYLDYGFNDTTVTEYRRDLNLSKGKADVTFKHDGVTYTREYFASNPDNVMVARLTA
SKAGKLNFNVSMPTNTNYSKTGETTTVKGDTLTVKGALGNNGLLYNSQIKVVLDNGEGTLSEGS
DGASLKVSDAKAVTLYIAAATDYKQKYPSYRTGETAAEVNTRVAKVVQDAANKGYTAVKKAH
IDDHSAIYDRVKIDLGQSGHSSDGAVATDALLKAYQRGSATTAQKRELETLVYKYGRYLTIGSSR
ENSQLPSNLQGIWSVTAGDNAHGNTPWGSDFHMNVNLQMNYWPTYSANMGELAEPLIEYVEGL
VKPGRVTAKVYAGAETTNPETTPIGEGEGYMAHTENTAYGWTAPGQSFSWGWSPAAVPWILQN
VYEAYEYSGDPALLDRVYALLKEESHFYVNYMLHKAGSSSGDRLTTGVAYSPEQGPLGTDGNT
YESSLVWQMLNDAIEAAKAKGDPDGLVGNTTDCSADNWAKNDSGNFTDANANRSWSCAKSLL
KPIEVGDSGQIKEWYFEGALGKKKDGSTISGYQADNQHRHMSHLLGLFPGDLITIDNSEYMDAA
KTSLRYRCFKGNVLQSNTGWAIGQRINSWARTGDGNTTYQLVELQLKNAMYANLFDYHAPFQI
DGNFGNTSGVDEMLLQSNSTFTDTAGKKYVNYTNILPALPDAWAGGSVSGLVARGNFTVGTTW
KNGKATEVRLTSNKGKQAAVKITAGGAQNYEVKNGDTAVNAKVVTNADG

>GH94==1336047-N [Clostridium stercorarium]
MKFGYFDDVNREYVITTPATPYPWINYLGCQDFFSLISNTSGGYCFYRDARLRRITRYRYNNVPID
SGGRYFYIYDSGDYWTPGWMPVKRELDRYECRHGLGYTRTGERNGVEVSQLAFVPLNYNGEV
NQVVITNKSGSEKEIALFSFVEFCLWNAMDDMTNFQRNFSTGEVEVEGSAIYHKTEYRERRNHY
AFFWVNSPIDGFDTDRESFLGLYNGFDSPKNVAAGKPTNSIASGWSPIASHYIKMSLKPGEKRSYI
FVLGYVE

>GH94==Q8VP44_CLOTM-N - Clostridium thermocellum
MKFGFFDDANKEYVITVPRTPYPWINYLGTENFFSLISNTAGGYCFYRDARLRRITRYRYNNVPID
MGGRYFYIYDNGDFWSPGWSPVKRELESYECRHGLGYTKIAGKRNGIKAEVTFFVPLNYNGEVQ
KLILKNEGQDKKITLFSFIEFCLWNAYDDMTNFQRNFSTGEVEIEGSVIYHKTEYRERRNHYAFY
SVNAKISGFDSDRDSFIGLYNGFDAPQAVVNGKSNNSVADGWAPIASHSIEIELNPGEQKEYVFIIG
YVE

>GH94==Q9X2G3_THEMA-N - Thermotoga maritima
MRFGYFDDVNREYVITTPQTPYPWINYLGTEDFFSIISHMAGGYCFYKDARLRRITRFRYNNVPTD
AGGRYFYIREENGDFWTPTWMPVRKDLSFFEARHGLGYTKITGERNGLRATITYFVPRHFTGEVH
YLVLENKAEKPRKIKLFSFIEFCLWNALDDMTNFQRNYSTGEVEIEGSVIYHKTEYRERRNHYAF
YSVNQPIDGFDTDRESFIGLYSGFEAPQAVVEGKPRNSVASGWAPIASHYLEIELAPSEKKELIFIL
GYVE

>GH94==89950565- [Saccharophagus degradans 2-40]
MKFGHFDDKAREYVITDPKTPYPWINYLGNEDFFSLVSNTGGGYSFYKDAKFRRLTRYRYNNVP
VDNGGKYFYINDSGDVWSPGWKPVKAELDAYSCAHGLSYTRITGERNGIQAEVLSFIPLGTWAEI
QKVSLKNTSGATKKFKLFSFAEWCLWNAEDDMTNFQRNFSTGEVEVEDSVIYHKTEFKERRNHY
AFYSVNAPIQGFDTDRDKWKGLYNDFDKPDAVFEGEPRNSEAHGWSPIASHYLEVELAPGESKD
LIFVLGYIE

>GH94==89950153-N [Saccharophagus degradans 2-40]
MLKAINNGERYQLTSPTAMPQSASFLWNKKMMIQVNCRGYAVAQFMQPEPAKYAYAPNLEAK
TFMQPEQPYYAHHPGRFFYIKDEETGEIFSAPYEPVRSQLNNFSFNAGKSDISWHIAALGIEVELCL
SLPVDDVVELWELKIKNGGAQPRKLSIYPYFPVGYMSWMNQSGDYSQTAGGIIASCVTPYQKVA
DYFKNKDFKDKTFFLHETAPAAWEVNQKNFEGEGGLHNPNAIQQETLGCGNALYETPTAVLQY
RRELAAQEQQTFRFIFGPA

>GH93==Q9K438_STRCO - Streptomyces coelicolor
MRNPSPTLRRSPALLGVLALLTALLSLWSQPASAAPVGQVLASPDLAAHPQGDNSYPRAVRLDH
DGSAGQTMLATYAKREQGATNTLPVHRSTDGGRTWSAAPISTITSHTPGWDIEAPVLYEVPRTAN
GLNQGDLLAAGTAWQAGDYTTQRVEVFKSTDHGQSWQYLSDCTRTSGLPDTIGHGIWEPWFLL
APDNTLACFISDERPANSPTNNQVIGHYTSTDGGRTWSGTLTQDVAFPADNLARPGMSIVVPLPD
GRFMMAYEMCRDATDADHACEVYTKTSPDGLNWAPADDPGTLVRTADGRELLHTPYLAWVPG
GGPDGTLLMSGQRVVSGPTGNKTVLSESGTVVFANTHLGTGEWTEIEAPVRTDPTGGYNPGEPS
CPGYSSPIVPRADGTSFLYLTATWLGTGNQCQVRFGTGGLGGP

>GH91==72132980 [Bacillus sp. snu-7]
YDVTTWRIKAHPEVTAQSDIGAVINDIIADIKQRQTSPDARPGAAIIIPPGDYDLHTQVVVDVSYLT
IAGFGHGFFSRSILDNSNPTGWQNLQPGASHIRVLTSPSAPQAFLVKRAGDPRLSGIVFRDFCLDG
VGFTPGKNSYHNGKTGIEVASDNDSFHITGMGFVYLEHALIVRGADALRVNDNMIAECGNCVEL
TGAGQATIVSGNHMGAGPDGVTLLAENHEGLLVTGNNLFPRGRSLIEFTGCNRCSVTSNRLQGFY
PGMLRLLNGCKENLITANHIRRTNEGYPPFIGRGNGLDDLYGVVHIAGDNNLISDNLFAYNVPPA
NIAPAGAQPTQILIAGGDANVVALNHVVSDVASQHVVLDASTTHSKVLDSGTASQITSYSSDTAIR
PTP

>GH91==1110443 [Arthrobacter globiformis]
YDVTTWSGATISPYVDIGAVINQIIADIKANQTSQAARPGAVIYIPPGHYDLLTRVVVDVSFLQIKG
SGHGFLSEAIRDESSTGSWVETQPGASHIRVKNTDGNREAFLVSRSGDPNVVGRLNSIEFKGFCLD
GVTDSKPYSPGNSKIGISVQSDNDSFHVEGMGFVYLEHAIIVKGADAPNITNNFIAECGSCIELTGA
SQVAKITNNFLISAWAGYSIYAENAEGPLITGNSLLWAANITLSDCNRVSISSNKLLSNFPSMVALL
GNCSENLIAANHFRRVSGDGTSTRFDDLFGLVHIEGNNNTVTGNMFSFNVPASSISPSGATPTIILV
KSGDSNYLATNNIVSNVSAMVVLDGSTTATRIIYSAKNSQLNAYTTSYTLVPTP

>GH90==TSPE_BPP22 - Bacteriophage P22
MTDITANVVVSNPRPIFTESRSFKAVANGKIYIGQIDTDPVNPANQIPVYIENEDGSHVQITQPLIIN
AAGKIVYNGQLVKIVTVQGHSMAIYDANGSQVDYIANVLKYDPDQYSIEADKKFKYSVKLSDYP
TLQDAASAAVDGLLIDRDYNFYGGETVDFGGKVLTIECKAKFIGDGNLIFTKLGKGSRIAGVFME
STTTPWVIKPWTDDNQWLTDAAAVVATLKQSKTDGYQPTVSDYVKFPGIETLLPPNAKGQNITS
TLEIRECIGVEVHRASGLMAGFLFRGCHFCKMVDANNPSGGKDGIITFENLSGDWGKGNYVIGGR
TSYGSVSSAQFLRNNGGFERDGGVIGFTSYRAGESGVKTWQGTVGSTTSRNYNLQFRDSVVIYPV
WDGFDLGADTDMNPELDRPGDYPITQYPLHQLPLNHLIDNLLVRGALGVGFGMDGKGMYVSNI
TVEDCAGSGAYLLTHESVFTNIAIIDTNTKDFQANQIYISGACRVNGLRLIGIRSTDGQGLTIDAPN
STVSGITGMVDPSRINVANLAEEGLGNIRANSFGYDSAAIKLRIHKLSKTLDSGALYSHINGGAGS
GSAYTQLTAISGSTPDAVSLKVNHKDCRGAEIPFVPDIASDDFIKDSSCFLPYWENNSTSLKALVK
KPNGELVRLTLATL

>GH90==56128763 [Salmonella enterica subsp. enterica serovar Paratyphi A str. ATCC 9150]
MTDITANVVVSNPRPIFTESRSFKAVANGKIYIGQIDTDPVNPANQIPVYIENEDGSHVQIAQPLIIN
AAGKIVYNGQLVKIVTVQGHSMAIYDANGSQVDYIANVLKYDPDQYSIEADKKFKYSVKLSDYP
TLQDAASAAVDGLLIDVDYHFYNGEKVDFGGKVLTIECKAKFIGDGNLIFTKLGKGSRIAGVFME
STTTPWVIKPWTDDNQWLTDAAAVVATLKQSKTDGYQPTVSDYVKFPGIETLLPPNAKGQNITS
TLEIRECIGVEVHRASGLMAGFLFRGCHFCKMVDANNPSGGKDGIITFENLSGDWGKGNYVIGGR
TSYGSVSSAQFLRNNGGFERDGGVIGFTSYRAGESGVKTWQGTVGSTTSRNYNLQFRDSVVIYPV
WDGFDLGADTDMNPELDRPGDYPITQYPLHQLPLNHLIDNLLVRGALGVGFGMDGKGMYVSNI
TVEDCAGSGAYLLTHESVFTNIAIIDTNTKDFQANQIYISGACRVNGLRLIGIRSTDGQGLTIDAPN
STVSGITGMVDPSRINVANLAEEGLGNIRANSFGYDSAAIKLRIHKLSKTLDSGALYSHINGGAGS
GSAWTQLTAISGNTPDAVSLKVNHKDCRGAEIPFVPDIASDDFIKDSSCFLPYWENNSTSLKALVK
KPNGELVRLTLATL

>GH87==Q93R89_9ACTO - Streptomyces sp. J-13-3
DTAASYTLDLIEAEETGSAATLPSGYVSAADLGVTPGGQDVTAALNNALNTAKNQGKGLFLPAG
TYKISDHINLSGVKLRGAGVWHTVLRGTGLKGGLFGQGGTSTIENLMIDGENTVRDDAGGQAAI
EGDFGTGSVIKNVWIQHTKVGLWISGPTTGLKAGDLRIRNTYADGVNLHGAVRDTVISNSSIRGT
GDDALAMWSDGAAVTNSAFRNNTVQLPSLANGAAVYGGNGNSVENNLISDTVVAAAGITVSTR
FGEPFSGTTTVSGNLLRRTGSMEVNWNTKLGALWVYADQHDITQPVVLRNNTVEDSTYSGLLIS
WEKQVSDLRVDGLTIDRTGANGIEGNAGGKGTFSNTKITGTAEAPLNMTGGFTIQRGSGNSGW

>GH87==110743487 [Bacillus circulans]
ADSLEYGVDFLEIEAVPAAIARPANSVSVTDFGAVANDGQDDLAAFEAAVNAAVTSGKILYIPAG
TFHLGNMWKIGSVANKINNITIMGAGIWHTNIQFTNPNQASGGISFRVTGQLDFSHIYMNSNLRSR
YGEQAVYKGFMDNFGTNSKVHNVWVEHFECGFWVGDYAHTPAIIANGLVIENSRIRNNLADGV
NFAQGTSNSTVRNSSIRNNGDDGLAVWTSNVNGAPAGVNNTFSYNTIENNWRAAGIAFFGGSGH

KATHNLIVDTVGGSAIRMNTVFPGYHFQNNTGIVFSDTTIINSGTSRDLYNGERGAIDLEASNDPIK
NVTFTNIDIINTQRSAIQFGYGGGFENIVFNNININGAGKDGVLTSRFSSPHPGAAIYTYTGNGSATF
NNLTTNDIAHPNLYFIQNGFNLTIQ

>GH86==89951900 [Saccharophagus degradans 2-40]
TCANTPALEFANEQGCSSSQVANTHVVNVSVNANFKRSVNGVFDFGRRRHMTAHTAIHEPDWV
GHTDKLNYLFNTLDVYMGRDNGSATWKFNDTTEDPNKPNWPNMDYMVERGKGLREAHDQNP
LFKRFSAEKQLLIAGTNPHALYPTLSWFPNAFTWSGWQPKNIETSAAWVGQYMEHYFANASNG
YVGEQLPEYWEVVNEPDMKMKTGQFMVTNQEAIWEYHNLVAQEIRDHLGAEAPPIGGMTWGQ
HDFYRRDGISRFADDSYDQWITNDDQVLQAEARAFYRNAMATTVDDTRDQDWYQWDVMWKG
FMDAAGDNMDFYSVHIYDWPGENVGDTTVVRRGGHTSAMLEMMEWYDVKRNGFNNRKPIVL
SEYGSVNGAWDNRAHEERYDIASIKAFNGMLMQFLERPDYVIKSLPFTPAKPLWGYLPGGCGYD
DAVACTTRYHYAMLIEDELNSGNWEWSSYIKFYELWADIDGTRVDSKSSDVDVQVDSYVKGNE
LFVILNNLEAADTTVNLDVSGIASVQNVELRNMHFDIQETHLDRHHMSAAPKTVTLAADATVVL
RYTLASSVAVNNTVVEKKYF

>GH86==AGAR_ALTAT - Alteromonas atlantica
MLKVIPWLLVTSSLVAIPTYIHATTEVVVNLNVKHSVEGKSEFERKNHIKLHSTLNDNDWQGEED
KLKYMMEELDVYFGRDNGGTVWNFNQAIEDPANIGYADPQNIIARGQAQRETNWGQNKSALHQ
YDGRGDLMIGGQPRAHYLGNTSPCCGGSAWQAKGGDAVGDFLGQYVNEFFRSAGDPVTKGHL
APVYFEVLNEPLYQVTDAPHELGLEQPIPPIDIFTFHNDVADAFRQHNTHIKIGGFTVAFPIFEQREF
ARWEERMKLFIDTSGSHMDVYSTHFYDLEDDNRFKGSRLEATLDMIDQYSLLALGETKPHVISEY
GGRNRPMENAPWSALRDWWFLKTASPMLMQFLSRPDSVLTSIPFVPIKALWGTAADGTPYNWR
LLRQQKEAPNETGENWVFTEMVKFYQLWSDVKGTRVDTFSTNSDFLIDSYVQNDKAYVLISNLT
EQAEKIVVHKYGAPASSQPTTRIKHLYLKGAAPRLMKQVMRQISKKSRLLLKRLW

>GH82==Q9F5I8_9ALTE - Alteromonas sp. ATCC 43554
MRLYFRKLWLTNLFLGGALASSAAIGAVSPKTYKDADFYVAPTQQDVNYDLVDDFGANGNDTS
DDSNALQRAINAISRKPNGGTLLIPNGTYHFLGIQMKSNVHIRVESDVIIKPTWNGDGKNHRLFEV
GVNNIVRNFSFQGLGNGFLVDFKDSRDKNLAVFKLGDVRNYKISNFTIDDNKTIFASILVDVTERN
GRLHWSRNGIIERIKQNNALFGYGLIQTYGADNILFRNLHSEGGIALRMETDNLLMKNYKQGGIR
NIFADNIRCSKGLAAVMFGPHFMKNGDVQVTNVSSVSCGSAVRSDSGFVELFSPTDEVHTRQSW
KQAVESKLGRGCAQTPYARGNGGTRWAARVTQKDACLDKAKLEYGIEPGSFGTVKVFDVTARF
GYNADLKQDQLDYFSTSNPMCKRVCLPTKEQWSKQGQIYIGPSLAAVIDTTPETSKYDYDVKTF
NVKRINFPVNSHKTIDTNTESSRVCNYYGMSECSSSRWER

>GH80==Q9S589_9SPHI - Sphingobacterium multivorum
MHSRSPSVRRIGVQAALTVLALVCGASAAVAAGKPKAAAQTNGQPSVYVQDGWVYTNTFTAT
GQPLVTATKAAAAAGVIPVGDSRVYGNVFDKGRKLTVNQWQAVLSMDAYPENGTTNYQDPEP
WRYCEVDYEASEGISDYRGNTFGPVGVTTVGDFPDYFKNAYAPYVLGKTGATNTDMKNWGVQ
VTGIAAADMKADDTRLDPYPNLARSNSKKRAALTKICQALQSDFDNRQAQYVMSHYAHIDSDK
LLPVLDALKKLGFTSFGQYNLVGLAFQVQVNTGSIGSISAFSSVKSAGNCGSMSNETCFATYLTD
QYIRWLKSSSLGDDAGNCWRASMALDIYKQDPTMGNVSVVTSIINSKYPNNSGKCPTSGVKWSK
NMAWN

>GH74==Q9WYE1_THEMA - Thermotoga maritima
WKSVEINGGGFVPGIIFHPASPGLLYARTDVGGLYRWDEETKRWKQLFDFLRRDQSDYMGVLSV
ALDPSDPKRIYAMTGKYTQDWAGYGAILISEDYGETWTIVNLDKYGIKVGGNEDGRNAGERLQV
DPNFSSVLFMGTTKYGLWKSEDFGKNWKKVDSFPSTSVTFVLFDEKSGEKGSPTPRIFVGCSEPK
GIFVTEDGGTTWNVLPNLPNDLIPLRGKIHDGILYVTLSNALGPNGATRGAVMKYVIADQKWYD
VTPMKGDFGYCGIDVQENVVIVSTLDRWYPHDEIFISLNGGETWRPLLEKANFDINKAPWIKDLN
PHWISDVKIDPFDMNRAIFTTGYGVWVTYELKKSFEGMGKPVKWIFENRGLEETVVLQLVPPIGE
RPLLSAIADWGGFRHESLDTPPSSMYKPLKWTSLGIAFAYQNSKFVARVHTYTYPFLSYSEDGGI
NWREIETVPEGITDGGRLSLAVSNDGKTLVWSPANHEVIVSSDKGKSWKKAISVPVPEFNYFPAS
DPVNPSKFYIFDWKNGDFLISKDGGKSFMKGAKLPSFDNWWVSLYSFPVLAPDREGDIWLALQW
NGLYRSKDGGITFERLGNVDIAYVIGFGAPKPGTDYPAIYLNGMVNGVYGIFMSTDEGKTWMRI
NNDKHQFGWIHYMIGDMNEFGRIFLGTEGRGIIVG

>GH74==71915745 [Thermobifida fusca YX]
WRNVEIVGGGFVPGIVFNQSEPDLIYARTDIGGAYRWDPATERWIPLLDHVGWDDWGHSGVVSI
ATDPVDPDRVYAAVGTYTNDWDPNNGAIKRSTDRGETWETTELPFKLGGNMPGRGMGERLAID
PNDNSVLYLGAPSGHGLWKSTDYGKTWQKVTSFPNPGNYVAGDPSDVGGYLGDNQGVVWVVFD
PTSSSPGHVTKDIYVGVADKQNTVYRSTDGGQTWERIPGQPTGFLAQKGVFDHVNGLLYIATSDT
GGPYDGSDGEVWRYDTTTGTWTDITPADPDGFEYGFSGLTIDRQNPDTIMVVSQILWWPDIQIWR
STDRGETWSRIWEFSGYPDRTLRYNHDISAAPWLDFNRQDNPPEVSPKLGWMTQAFEIDPFNSDR
MLYGTGATIYGSDNLTNWDEGKKIDIKVRAQGIEETAVQDLIAPPGDTELVSALGDIGGFVHDDI
TVVPDAMFDSPFHGNTRSIDFAELNPSVMARVGEAVDGEVDSHIGISTSGGSHWWAGQEPSGVT
GAGTVAVNADGSRIVWSPDGTGVHYSTTLGSSWTPSQGVPAGARVEADRVNPDKFYAFANGTF
YTSTDGGATFTKSSAAGLPTKGNIRFAAVPGHEGDIWLAGGETNSTYGMWRSTDSGATFTRITAV
DEGDVVGFGKPAPGRSYPAVYTSSKINGVRGIFRSDDAGTTWVRINDDQHQWAWTGAAITGDP
DVYGRVYIGTNGRGVIVG

>GH74==37651953 [Clostridium thermocellum]
WDNVVIGGGGGFMPGIVFNETEKDLIYARADIGGAYRWDPSTETWIPLLDHFQMDEYSYYGVES
IATDPVDPNRVYIVAGMYTNDWLPNMGAILRSTDRGETWEKTILPFKMGGNMPGRSMGERLAID
PNDNRILYLGTRCGNGLWRSTDYGVTWSKVESFPNPGTYIYDPNFDYTKDIIGVVWVVFDKSSST
PGNPTKTIYVGVADKKNESIYRSTDGGVTWKAVPGQPKGLLPHHGVLASNGMLYITYGDTCGPYD
GNGKGQVWKFNTRTGEWIDITPIPYSSSDNRFCFAGLAVDRQNPDIIMVTSMNAWWPDEYIFRST
DGGATWKNIWEWGMYPERILHYEIDISAAPWLDWGTEKQLPEINPKLGWMIGDIEIDPFNSDRM
MYVTGATIYGCDNLTDWDRGGKVKIEVKATGIEECAVLDLVSPPEGAPLVSAVGDLVGFVHDDL
KVGPKKMHVPSYSSGTGIDYAELVPNFMALVAKADLYDVKKISFSYDGGRNWFQPPNEAPNSV
GGGSVAVAADAKSVIWTPENASPAVTTDNGNSWKVCTNLGMGAVVASDRVNGKKFYAFYNGK
FYISTDGGLTFTDTKAPQLPKSVNKIKAVPGKEGHVWLAAREGGLWRSTDGGYTFEKLSNVDTA
HVVGFGKAAPGQDYMAIYITGKIDNVLGFFRSDDAGKTWVRINDDEHGYGAVDTAITGDPRVY
GRVYIATNGRGIVYG

>GH66==CTA1_BACCI - Bacillus circulans
MVRFMYALRKRRLSLLLAMSLLVMCVASVVSPPPQALASGSGGIERVFTDKARYNPGDAVSIRV
QAKNGTGSSWSGAARLEIFHLENSVYTSSQSLSLTNGQSTTLTFTWTAPSTDFRGYFVRIDAGTLG
QGATAIDVSSDFTKYPRYGYISEFESGETALESKAKVDQLAQDYHINAWQFYDWMWRHDKMIK
RTGGSIDSTWLDLFNREISWSTLQNQIDAVHDVNGKAMAYAMIYASRENYSPLGISPTWGIYEDS
SHTNQFDVDFGDGSTYLYMSDPQNPNWQNYIHAEYIDSINTAGFDGIHVDQMGQRSNVYDYNG
NSIDLSTRFSPFLDQAKSVLSANNPARDNLTYNIVDGTVNGWAVNDVSKNADLDFLYSEIWYLSD
SYNQLKNYIEQLRANGGNKAVVLAAYMNYADNAGTRY

>GH64==E13B_CELCE - Cellulosimicrobium cellulans (Arthrobacter luteus)
MPHDRKNSSRRAWAALCAAVLAVSGALVGVAAPASAVPATIPLTITNDSGRGPIYLYVLGERDG
VAGWADAGGTFHPWPGGVGPVPVPAPDASIAGPGPGQSVTIRLPKLSGRVYYSYGQKMTFQIVL
DGRLVQPAVQNDSDPNRNILFNWTEYTLNDGGLWINSTQVDHWSAPYQVGVQRADGQVLSTG
MLKPNGYEAFYTALEGAGWGGLVQRAPDGSRLRALNPSHGIDVGKISSASIDSYVTEVWNSYRT
RDMVVTPFSHEPGTQFRGRVDGDWFRFRSGSGQEVAAFKKPDASSVYGCHKDLQAPNDHVVGP
IARTLCAALVRTTALTNPNQPDANSAGFYQDARTNVYAKLAHAQQMANGKAYAFAFDDVGAHE
SLVHDGNPQAAYIKLDPFTGTATPLGNGGSTEQPGTPGGLP

>GH58==115513819 [Escherichia coli APECO1]
MTVSTEVDHNEYTGNGVTTSFPYTFRIFRKSDLVVQVSDLNGNVTELVLDTGYTVTGAGTYSGG
SVVLPSPLATGWRITIDRVLDVVQETDLRNQGKFPEVHEDAFDYLTMLIQQCFGWFRRALMKPS
LLAKYYDAKQNKISNLADPSFEQDAVNNRSMRNYVDAAIAGVVGGFGWFIQYGSGAVYRTFQD
KMRDAISPKDFGAVGDGINDDSTAISACLEASSPGYKIDGLGLTFKVSTLPDVSRFKNARFLFERIP
GQPLFYASEDFIQGELFKITDTPWYNAWTQDKTFVYDNVIYAPFMAGDRHGVNNLHVAWVRSG
DDGKTWTTPEWLTDLHENYPTVNYHCMSMGVVRNRLFAVIETRTVSGNKLQVAELWDRPMSR
SLRVYGGITKAANQQVAYIRITDHGLFAGDFVNFSNSGVTGVTGNMTVTTVIDKNTFTVTTQNTQ
DVDQNNEGRYWSFGTSFHSSPWRKTSLGTIPSFVDGSTPVTEIHSFATISDNSFAVGYHNGDIGPR
ELGILYFSDAFGSPGSFVRRRIPAEYEANASEPCVKYYDGILYLTTRGTLSTQPGSSLHRSSDLGTS
WNSLRFPNNVHHSNLPFAKVGDELIIFGSERAFGEWEGGEPDNRYAGNYPRTFMTRVNVNEWSL
DNVEWVNVTDQIYQGGIVNSAVGVGSVCIKDNWLYYFGGEDFLNPWSIGDNNRKYPYVHDGH
PADLYCFRVKIKQEEFVSRDFVYGATPNRTLPTFMSTSGVRTVPVPVDFTDDVAVQSLTVHAGTS
GQVRAEVKLEGNYAIIAKKVPSDDVTAQRLIVSGGETTSSADGAMITLHGSGSSTPRRAVYNALE
HLFENGDVKPYLDNVNALGGPGNRFSTVYLGSNPVVTSDGTLKTEPVSPDEALLDAWGDVRYIA
YKWLNAVAIKGEEGARIHHGVIAQQLRDVLISHGLMEEESTTCRYAFLCYDDYPAVYDDVITGQ
REMPLTDNDGSIIVDEDDNPVMVMEDIIERVEITPAGSRWGVRPDLLFYIEAAWQRREMDKIKERI
QSLEER

>GH55==Q8NIP1_ASPOR - Aspergillus oryzae
MLFLAHVLLLLGLPAGMVGAVPLGQETDITTNLAARAASEYWVGTIKRQGAVAFGNGTDYQVY
RNVKDFGAKGDGSTDDTAAINQAISSGNRCGKGCDSSTVTPALVYFPPGTYVVSKPIVQYYYTQI
VGDAVNLPVIKAAAGFAGMAVIDADPYEDDGSNWYTNQNNFFRAIRNLVIDLTAMPQGSGAGIH
WQVGQATSLQNIRFEMIKGGGDANKQQGIFMDNGSGGFMSDLTFNGGNYGMFLGNQQFTTRNL
TFNDCNTAIFMNWNWAWTFKSLSINNCQVGLNMSNAPQNQTVGSVLILDSQLTNTPTGVVSAFT
ENSIPIGGGVLILDNVDFSGSKVAVAGITGNTILAGGSVVTNWVQGNGYLPGSAKQKREASVKVT
TQTVTETVEVCTADYTDSPSAPTALPSSLGESRTAGLLPTLPLPNIPLLSGLLSGSQSSATQPAGVL
SSEVPEPTATPSTPEEAEPSTEVQSTPQPSAPAQSQPETPVESTVAAPLIPSQPSPTVQGSSSVVTGP
ASSSVAHATNQCSVKTVTKTRLQTALPTHAKPSSLLNGGKVYERSKPLYTSYDASSFVSVKSAGA
KGDGSTDDTAAIQKILNSAKEDQIVYFDHGAYIITDTIKVPKNVKITGEVWPVLMAYGQKFGDEK
NPIPMLQVGEVGETGSVEITDIALQTKGPAPGAILMQWNLAESSQGAAGMWDTHFRIGGSAGTEL
QSDKCAKTPKQTTTPNKECIAAFMLMHITEKASAYIENSWFWVADHELDLPDHNQINVYNGRGV
YIESQGPVWLYGTASEHNQLYNQLYNYQVTNAKNVFMGLIQTETPYYQANPNALTPFTPQTNW
NDPDFSYCKTDGCRKAWGLRVQNTSDMYVYGAGLYSFFENYGQTCLATESCQENMVEVDCSD
VHIYGLSTKASTNMITSNSGAGLVPQDENRSNFCSTIALFQQS

>GH55==116227509 [Solibacter usitatus Ellin6076]
MHPLTRCLLLATLLAASAHSTTRALYSSFSGALYQLQRGSDDTTTTISPLTAGGIADASAQDTFCANTT
CLITIIYDQSGNGNHLTQAPPGGFDGPDTDGYDNLASAIGAPVTLNGQKAYGVFMSPGTGYRNNE
ATGTATGDEAEGMYAVLDGTHYNDACCFDYGNAETSSTDTGAGHMEAIYLGNSTTWGYGAGD
GPWIMVDMENNLFSGADEGYNSGDPSISYRFVTAAVKGGADKWAIRGANAASGSLSTYYSGAR
PDYSGYNPMSKEGAIILGIGGDNSNGAQGTFYEGVMTSGYPSDDTENSVQENIVAAKYVVGSLVS
GPSFTSGEV

NSIDLSTRFSPFLDQAKSVLSANNPARDNLTYNIVDGTVNGWAVNDVSKNADLDFLYSEIWYLSD
SYNQLKNYIEQLRANGGNKAVVLAAYMNYADNAGTRY

>GH64==E13B_CELCE - Cellulosimicrobium cellulans (Arthrobacter luteus)
MPHDRKNSSRRAWAALCAAVLAVSGALVGVAAPASAVPATIPLTITNDSGRGPIYLYVLGERDG
VAGWADAGGTFHPWPGGVGPVPVPAPDASIAGPGPGQSVTIRLPKLSGRVYYSYGQKMTFQIVL
DGRLVQPAVQNDSDPNRNILFNWTEYTLNDGGLWINSTQVDHWSAPYQVGVQRADGQVLSTG
MLKPNGYEAFYTALEGAGWGGLVQRAPDGSRLRALNPSHGIDVGKISSASIDSYVTEVWNSYRT
RDMVVTPFSHEPGTQFRGRVDGDWFRFRSGSGQEVAAFKKPDASSVYGCHKDLQAPNDHVVGP
IARTLCAALVRTTALTNPNQPDANSAGFYQDARTNVYAKLAHAQQMANGKAYAFAFDDVGAHE
SLVHDGNPQAAYIKLDPFTGTATPLGNGGSTEQPGTPGGLP

>GH58==115513819 [Escherichia coli APECO1]
MTVSTEVDHNEYTGNGVTTSFPYTFRIFRKSDLVVQVSDLNGNVTELVLDTGYTVTGAGTYSGG
SVVLPSPLATGWRITIDRVLDVVQETDLRNQGKFPEVHEDAFDYLTMLIQQCFGWFRRALMKPS
LLAKYYDAKQNKISNLADPSFEQDAVNNRSMRNYVDAAIAGVVGGFGWFIQYGSGAVYRTFQD
KMRDAISPKDFGAVGDGINDDSTAISACLEASSPGYKIDGLGLTFKVSTLPDVSRFKNARFLFERIP
GQPLFYASEDFIQGELFKITDTPWYNAWTQDKTFVYDNVIYAPFMAGDRHGVNNLHVAWVRSG
DDGKTWTTPEWLTDLHENYPTVNYHCMSMGVVRNRLFAVIETRTVSGNKLQVAELWDRPMSR
SLRVYGGITKAANQQVAYIRITDHGLFAGDFVNFSNSGVTGVTGNMTVTTVIDKNTFTVTTQNTQ
DVDQNNEGRYWSFGTSFHSSPWRKTSLGTIPSFVDGSTPVTEIHSFATISDNSFAVGYHNGDIGPR
ELGILYFSDAFGSPGSFVRRRIPAEYEANASEPCVKYYDGILYLTTRGTLSTQPGSSLHRSSDLGTS
WNSLRFPNNVHHSNLPFAKVGDELIIFGSERAFGEWEGGEPDNRYAGNYPRTFMTRVNVNEWSL
DNVEWVNVTDQIYQGGIVNSAVGVGSVCIKDNWLYYFGGEDFLNPWSIGDNNRKYPYVHDGH
PADLYCFRVKIKQEEFVSRDFVYGATPNRTLPTFMSTSGVRTVPVPVDFTDDVAVQSLTVHAGTS
GQVRAEVKLEGNYAIIAKKVPSDDVTAQRLIVSGGETTSSADGAMITLHGSGSSTPRRAVYNALE
HLFENGDVKPYLDNVNALGGPGNRFSTVYLGSNPVVTSDGTLKTEPVSPDEALLDAWGDVRYIA
YKWLNAVAIKGEEGARIHHGVIAQQLRDVLISHGLMEEESTTCRYAFLCYDDYPAVYDDVITGQ
REMPLTDNDGSIIVDEDDNPVMVMEDIIERVEITPAGSRWGVRPDLLFYIEAAWQRREMDKIKERI
QSLEER

>GH55==Q8NIP1_ASPOR - Aspergillus oryzae
MLFLAHVLLLLGLPAGMVGAVPLGQETDITTNLAARAASEYWVGTIKRQGAVAFGNGTDYQVY
RNVKDFGAKGDGSTDDTAAINQAISSGNRCGKGCDSSTVTPALVYFPPGTYVVSKPIVQYYYTQI
VGDAVNLPVIKAAAGFAGMAVIDADPYEDDGSNWYTNQNNFFRAIRNLVIDLTAMPQGSGAGIH
WQVGQATSLQNIRFEMIKGGGDANKQQGIFMDNGSGGFMSDLTFNGGNYGMFLGNQQFTTRNL
TFNDCNTAIFMNWNWAWTFKSLSINNCQVGLNMSNAPQNQTVGSVLILDSQLTNTPTGVVSAFT
ENSIPIGGGVLILDNVDFSGSKVAVAGITGNTILAGGSVVTNWVQGNGYLPGSAKQKREASVKVT
TQTVTETVEVCTADYTDSPSAPTALPSSLGESRTAGLLPTLPLPNIPLLSGLLSGSQSSATQPAGVL
SSEVPEPTATPSTPEEAEPSTEVQSTPQPSAPAQSQPETPVESTVAAPLIPSQPSPTVQGSSSVVTGP
ASSSVAHATNQCSVKTVTKTRLQTALPTHAKPSSLLNGGKVYERSKPLYTSYDASSFVSVKSAGA
KGDGSTDDTAAIQKILNSAKEDQIVYFDHGAYIITDTIKVPKNVKITGEVWPVLMAYGQKFGDEK
NPIPMLQVGEVGETGSVEITDIALQTKGPAPGAILMQWNLAESSQGAAGMWDTHFRIGGSAGTEL
QSDKCAKTPKQTTTPNKECIAAFMLMHITEKASAYIENSWFWVADHELDLPDHNQINVYNGRGV
YIESQGPVWLYGTASEHNQLYNQLYNYQVTNAKNVFMGLIQTETPYYQANPNALTPFTPQTNW
NDPDFSYCKTDGCRKAWGLRVQNTSDMYVYGAGLYSFFENYGQTCLATESCQENMVEVDCSD
VHIYGLSTKASTNMITSNSGAGLVPQDENRSNFCSTIALFQQS

>GH55==116227509 [Solibacter usitatus Ellin6076]
MHPLTRCLLLATLLAASAHSTTRALYSSFSGALYQLQRGSDDTTTTISPLTAGGIADASAQDTFCANTT
CLITIIYDQSGNGNHLTQAPPGGFDGPDTDGYDNLASAIGAPVTLNGQKAYGVFMSPGTGYRNNE
ATGTATGDEAEGMYAVLDGTHYNDACCFDYGNAETSSTDTGAGHMEAIYLGNSTTWGYGAGD
GPWIMVDMENNLFSGADEGYNSGDPSISYRFVTAAVKGGADKWAIRGANAASGSLSTYYSGAR
PDYSGYNPMSKEGAIILGIGGDNSNGAQGTFYEGVMTSGYPSDDTENSVQENIVAAKYVVGSLVS
GPSFTSGEV

>GH54==116224401 [Solibacter usitatus Ellin6076]
PCDIYAAAGDPCVAAHSTTRALYAGYNGPLYQVLRQSDGKTLDIGVVQPAASPLQDAGGYANA
AAQDAFCANTYCWVSIIFDQSPKHNDVIHAPRGGFSGPSMGGFNNLPVADMAPITIMGHKAYGV
FITPGMGLRQNDAKGTAVDDQAEGQYWVINGHHYNSGCCFDYGNAEIDSRDDDNGTMETTYY
GNATSWYHGNPAGPWIMTDQENNLVGCVNPDGTKLCASLPNITWRFVTAMAKGEPHHWTSMG
GDAQHGALKVMFDGPRVDVTYDPMRKQGAILLGNGGDNSVGSQGTFYEGAMTAADTFPSDAT
DQLVQANVVAARYDLPRLSLAPASATAAPPGLQTFSPGSSQNTTLTFTNTTGAPVVGVQLSLSVP
RQWTSVVTGASTTSATIAGPVAPGASVSATFKVTSGPVAFNGDLVGNARFTNQATGAKQFETTS
EKVRNVSPVKINEFRITDGSPTNSTNSFIELYNAGARDVDISNWTLTQHPAQQAIFSAVKIPARTKL
TPGGFYLLGLANSGLAVPARAGDKTIHVRSTAGMNVGDTISIESGSGAETHKIVNVGTAAGAGTT
LWQPLPDGPVIKLPAGSTNVPVTSVAGFAAGEKIAIGYGAVYPAVARTIEQYEVATVTAVGKPGT
QAFLGVDAAAGATNIKVTSVANISAGDRIRLDIESVGHGIETVAVTKVGTQSSRTALSADAHAGA
TNIKVRNANGFTVGDKLTVGTPANHETVTITAVADGVDFTPALAKAHLNREFVVAQGTGLDLAA
P

>GH54==91690082 [Burkholderia xenovorans LB400]
PCDAAATANTPCSAAHSVTRLLTRNYRGPLFQVQRASDNAKQDIYPYTSSTLPHGADRSLIGSAN
VKSANTFCNNTTCSVTYIYDQIDLVAPLKGGAFGNVPATLTLNPDGTESLTVPNSGSRPQKTTTVP
VLNGFATITLPGTSGALAVQLSQPPTALTAQSPTLQLSIGNDLPALSGTPAKLGFVPLQGAQIPALG
TLAGQAYRNRFGTVNQSIGDSDIAEYMVAGAHYSQSSTCCGTYGNMENSANPGTYAHGEIEGE
MFALAFSNGNAAVFGYCDGNSNQPPNVKDPVCNARDTNWPGVDAEAGVYLYGPPQPAREKFV
TVLSKYSPVTASNIFAVKGGSASQSVLTALYDQAPPEAYNFGNDAGHAFNGQWQGGLSLGEGG
DGSAAPIQFFEGAVITKATSDTTDNAIQASIASFYGPPADKAVAACYANNLINSPLSLATPDIWSQQ
NGGTAVPATDISGVNRGAATVTSTNGSSVSNVREIIRVQGGQSYSFTDYVLATSNATVFPGGSIQT
DDPSGTEFAWVLNTNTGAVVRGTWGAGQPTALSAVRSGNWWKVTMTLTAPAGSNNASVFIDPP
TSNAAGARSQQASYLSATHYCPSLAIVASSGT

>GH51==ABFA_BACST - Geobacillus stearothermophilus
MATKKATMIIEKDFKIAEIDKRIYGSFIEHLGRAVYGGIYEPGHPQADENGFRQDVIELVKELQVPI
IRYPGGNFVSGYNWEDGVGPKEQRPRRLDLAWKSVETNEIGLNEFMDWAKMVGAEVNMAVNL
GTRGIDAARNLVEYCNHPSGSYYSDLRIAHGYKEPHKIKTWCLGNEMDGPWQIGHKTAVEYGRI
ACEAAKVMKWVDPTIELVVCGSSNRNMPTFAEWEATVLDHTYDHVDYISLHQYYGNRDNDTA
NYLALSLEMDDFIRSVVAIADYVKAKKRSKKTIHLSF

>GH51==125715257 [Clostridium thermocellum ATCC 27405]
MKKARMTVDKDYKIAEIDKRIYGSFVEHLGRAVYDGLYQPGNSKSDEDGFRKDVIELVKELNVP
IIRYPGGNFVSNYFWEDGVGPVEDRPRRLDLAWKSIEPNQVGINEFAKWCKKVNAEIMMAVNLG
TRGISDACNLLEYCNHPGGSKYSDMRIKHGVKEPHNIKVWCLGNEMDGPWQVGHKTMDEYGRI
AEETARAMKMIDPSIELVACGSSSKDMPTFPQWEATVLDYAYDYVDYISLHQYYGNKENDTADF
LAKSDDLDDFIRSVIATCDYIKAKKRSKKDIYLSF

>GH50==AGAA_VIBS7 - Vibrio sp. (strain JT0107)
SEVSTEQATDGNQSIKASFDAAFKPMVVWNWASWNWGAEDVMSVDVVNPNDTDVTFAIKLIDS
DILPDWVDESQTSLDYFTVSANTTQTFSFNLNGGNEFQTHGENFSKDKVIGVQFMLSENDPQVLY
FDNIMVDGETVTPPPSDGAVNTQTAPVATLAQIEDFETIPDYLRPDGGVNVSTTTEIVTKGAAAM
AAEFTAGWNGLVFAGTWNWAELGEHTAVAVDVSNNSDSNIWLYSRIEDVNSQGETATRGVLV
KAGESKTIYTSLNDNPSLLTQDERVSALGLRDIPADPMSAQNGWGDFVALDKSQITAIRYFIGELA
SGETSQTLVFDNMRVIKDLNHESAYAEMTDAMGQNNLVTYAGKVASKEELAKLSDPEMAALGE
LTNRNMYGGNPDSSPATDCVLATPASFNACKDADGNWQLVDPAGNAFFSTGVDNIRLQDTYTM
TGVSSDAESESALRQSMFTEIPSDYVNENYGPVHSGPVSQGQAVSFYANNLITRHASEDVWRDIT
VKRMKDWGFNTLGNWTDPALYANGDVPYVANGWSTSGADRLPVKQIGSGYWGPLPDPWDAN
FATNAATMAAEIKAQVEGNEEYLVGIFVDNEMSWGNVTDVEGSRYAQTLAVFNTDGTDATTSP
AKNSFIWFLENQRYTGGIADLNAAWGTDYASWDATSPAQELAYVAGMEADMQFLAWQFAFQY
FNTVNTALKAELPNHLYLGSRFADWGRTPDVVSAAAAVVDVMSYNIYKDSIAAADWDADALSQ
IEAIDKPVIIGEFHFGALDSGSFAEGVVNATSQQDRADKMVSFYESVNAHKNFVGAHWFQYIDSP
LTGRAWDGENYNVGFVSNTDTPYTLMTDAAREFNCGMYGTDCSSLSNATEAASRAGELY

>GH50==Q9RKF0_STRCO - Streptomyces coelicolor
MTVHKRACTTPPPRASRSFRVRWPVLIAAACAGLVLATTSPPAVAAGAHDLGDETMLYDFQDGL
VPAEVGPYQAKTTIVGRGDKKLRVDFQARKNYYSSFSVRPEPVWNWSAEESESLGIAMELTNPS
DRSVQLTIDLESSTGVATRSVNVPAGGGGTYYFDVDSPALHRDTGLRADPSWLADKDVTSAVW
MWGSKETDTSRISQLNFYVAGLLHDRSVIVDDIRVVRDAPADPDYLKGLVDAFGQNNKVDYKG
KVSRTSEILRQRAAEAKDLRRHPVPEDRSRYGGWLNGPRLEATGNFRVEKYQGRWTLVDPDGY
LFFSTGIDNARMFDSPTTTGYDFDHDAIQELPPPSLTAGGPEDLNRVQKSALPTRTKMSETRADLF
SKLPKYRTRAGEGFGYAPDTLAGPVAQGETYSFYKANVARKYPGSNYMERWRDNTVDRMLSW
GFTSFGNWTDPEMYDNDRIPYFAHGWIKGDFKTVSTGQDYWGPMPDPFDPAFSDAAARTARAV
ADEVADSPLAIGVFMDNELSWGNAGSFSTRYGVVIDTMSRDAAESPTKSAFSDELEEKYGTIDAL
NAAWQTTVPSWEALRSGSADLGSDETAKESDYSALMTLYATQYFKTVDAELDKVMPDHLYAG
SRFASWGRTPEVVEAASKYVDIMSYNEYREGLHPSEWAFLEELDKPSLIGEFHMGTTTTGQPHPG
LVSAGTQAERARMYAEYMEQLIDNPYMVGGHWFQYADSPVTGRALDGENYNIGFVSVTDRPYP
EIVAAARDVNQRLYDRRYGNLATAEGHYTGRRSAE

>GH44==MANB_CALSA - Caldocellum saccharolyticum
PSVVEITINTNAGRTQISPYIYGANQDIEGVVHSARRLGGNRLTGYNWENNFSNAGNDWYHSSDD
YLCWSMGISGEDAKVPAAVVSKFHEYSLKNNAYSAVTLQMAGYVSKDNYGTVSENETAPSNR
WAEVKFKKDAPLSLNPDLNDNFVYMDEFINYLINKYGMASSPTGIKGYILDNEPDLWASTHPRIH
PNKVTCKELIEKSVELAKVIKTLDPSAEVFGYASYGFMGYYSLQDAPDWNQVKGEHRWFISWYL
EQMKKASDSFGKRLLDVLDLHWYPEARGGNIRVCFDGENDTSKEVVIARMQAPRTLWDPTYKT
SVKGQITAGENSWINQWFSDYLPIIPNVKADIEKYYPGTKLAISEFDYGGRNHISGGIALADVLGIF
GKYGVNFAARWGDSGSYAAAAYNIYLNYDGKGSKYGNTNVSANTSDVENMPVYASINGQDDS
ELHIILINRNYDQKLQVKINITSTPKYTKAEIYGFDSNSPEYKKMGNIDNIESNVFTLEVPKFNGVS
HSITLDFNVSIKIIQNEVIKFIRNLVFMRALV

>GH44==P71140_CLOTM - Clostridium thermocellum
AKVVDIRIDTSAERKPISPYIYGSNQELDATVTAKRFGGNRTTGYNWENNFSNAGSDWLHYSDTY
LLEDGGVPKGEWSTPASVVTTFHDKALSKNVPYTLITLQAAGYVSADGNGPVSQEETAPSSRWK
EVKFEKGAPFSLTPDTEDDYVYMDEFVNYLVNKYGNASTPTGIKGYSIDNEPALWSHTHPRIHPD
NVTAKELIEKSVALSKAVKKVDPYAEIFGPALYGFAAYETLQSAPDWGTEGEGYRWFIDYYLDK
MKKASDEEGKRLLDVLDVHWYPEARGGGERICFGADPRNIETNKARLQAPRTLWDPTYIEDSWI
GQWKKDFLPILPNLLDSIEKYYPGTKLAITEYDYGGGNHITGGIAQADVLGIFGKYGVYLATFWG
DASNNYTEAGINLYTNYDGKGGKFGDTSVKCETSDIEVSSAYASIVGEDDSKLHIILLNKNYDQPT
TFNFSIDSSKNYTIGNVWAFDRGSSNITQRTPIVNIKDNTFTYTVPALTACHIVLEAAEPVVYGDLN
NDSKVNAVDIMMLKRYILGIIDNINLTAADI

>GH36==RAFA_ECOLI - Escherichia coli
LGLFSSPGLEGHRNGLDASPVFYTVDVEHTENTLRLTSEDSVAGLRLVSELVMTPSGILKVRHAL
TNLREGDWQINRFAITLPVAERAEEVMAFHGRWTREFQPHRVRLTHDAFVLENRRGRTSHEHFP
ALIVGTPGFSEQQGEVWAVHLGWSGNHRMRCEAKTDGRRYVQAEALWMPGEKALRKNETLYT
PWLYACHSADGLNGMSQQYHRFLRDEIIRFPEQKLRPVHLNTWEGIYFNHNPDYIMQMAERAAA
LGVERFIIDDGWFKGRNDDRAALGDWYTDEQKYPNGLMPVINHVKSLGMEFGIWVEPEMINPDS
DLFRLHPDWILSMPGYSQPTGRYQYVLNLNIPEAFDYIYKRFLWLLGEHPVDYVKWDMNRELVQ
AGHEGRAAADAQTRQFYRLLDLLRERFPHVEFESCASGGGRIDFEVLKRTHRFWASDNNDALER
CTIQRGMSYFFPPEVMGAHIGHRRCHATFRQHSIAFRGLTALFGHMGLELDPVAADAKESDGYR
RYALLYKEWRQLIHTGVLWRVDMPDSSIQVQGVVSPDQSQALFMISQLAMPDYTLPGILRFPGL
AAEVRYRLRVIDHPEIQLVGEGGHT

>GH36==O33835_THEMA - Thermotoga maritima
MEIFGKTFREGRFVLKEKNFTVEFAVEKIHLGWKISGRVKGSPGRLEVLRTKAPEKVLVNNWQS
WGPCRVVDAFSFKPPEIDPNWRYTASVVPDVLERNLQSDYFVAEEGKVYGFLSSKIAHPFFAVED
GELVAYLEYFDVEFDDFVPLEPLVVLEDPNTPLLLEKYAELVGMENNARVPKHTPTGWCSWYH
YFLDLTWEETLKNLKLAKNFPFEVFQIDDAYEKDIGDWLVTRGDFPSVEEMAKVIAENGFIPGIW
TAPFSVSETSDVFNEHPDWVVKENGEPKMAYRNWNKKIYALDLSKDEVLNWLFDLFSSLRKMG
YRYFKIDFLFAGAVPGERKKNITPIQAFRKGIETIRKAVGEDSFILGCGSPLLPAVGCVDGMRIGPD
TAPFWGEHIEDNGAPAARWALRNAITRYFMHDRFWLNDPDCLILREEKTDLTQKEKELYSYTCG
VLDNMIIESDDLSLVRDHGKKVLKETLELLGGRPRVQNIMSEDLRYEIVSSGTLSGNVKIVVDLNS
REYHLEKEGKSSLKKRVVKREDGRNFYFYEEGERE

>GH23==Q9HXN1_PSEAE - Pseudomonas aeruginosa
VITRNSPATYFQDRNGETGFEYELAKRFAERLGVELKIETADNLDDLYAQLSREGGPALAAAGLT
PGREDDASVRYSHTYLDVTPQIIYRNGQQRPTRPEDLVGKRIMVLKGSSHAEQLAELKKQYPELK
YEESDAVEVVDLLRMVDVGDIDLTLVDSNELAMNQVYFPNVRVAFDFGEARGLAWALPGGDD
DSLMNEVNAFLDQAKKEGLLQRLKDRYYGHVDVLGYVGAYTFAQHLQQRLPRYESHFKQSGK
QLDTDWRLLAAIGYQESLWQPGATSKTGVRGLMML TNRTAQAMGVSNRLDPKQSIQGGSKYFV
QIRSELPESIKEPDRSWFALAAYNIGGAHLEDARKMAEKEGLNPNKWLDVKKMLPRLAQKQWY
AKTRYGYARGGETVHFVQNVRRYYDILTWVTQPQMEGSQIAESGLHLPGVNKTRPEEDSGDEKL

>GH23==SLT_ECOLI - Escherichia coli
VVEQMMPGLKDYPLYPYLEYRQITDDLMNQPAVTVTNFVRANPTLPPARTLQSRFVNELARRED
WRGLLAFSPEKPGTTEAQCNYYYAKWNTGQSEEAWQGAKELWLTGKSQPNACDKLFSVWRAS
GKQDPLAYLERIRLAMKAGNTGLVTVLAGQMPADYQTIASAIISLANNPNTVLTFARTTGATDFT
RQMAAVAFASVARQDAENARLMIPSLAQAQQLNEDQIQELRDIVAWRLMGNDVTDEQAKWRD
DAIMRSQSTSLIERRVRMALGTGDRRGLNTWLARLPMEAKEKDEWRYWQADLLLERGREAEAK
EILHQLMQQRGFYPMVAAQRIGEEYELKIDKAPQNVDSALTQGPEMARVRELMYWNLDNTARS
EWANLVKSKSKTEQAQLARYAFNNQWWDLSVQATIAGKLWDHLEERFPLAYNDLFKRYTSGK
EIPQSYAMAIARQESAWNPKVKSPVGASGLMQIMPGTATHTVKMFSIPGYSSPGQLLDPETNINIG
TSYLQYVYQQFGNNRIFSSAAYNAGPGRVRTWLGNSAGRIDAVAFVESIPFSETRGYVKNVLAY
DAYYRYFMGDKPTLMSATEWG

>CBM8==GUN6_DICDI - Dictyostelium discoideum
PSSGESLSIYKSGLKNDFQDWSWGEHSLTDTTNVESGETNSISFTPKAYGAVFLGCFECIDTDTYN
NIEFDINGGSSGAQLLRITVVKNSKSVGSKLITDLNGGTPIEANSWTKIKASFIDDFKVSGKVDGIW
IQDIKGDTQSTVYISNIIATA

>CBM8==108462586 polysaccharide deacetylase domain protein [Myxococcus xanthus DK 1622]
PDLTAGLFVYTDALQNGFANWSWADHSLDETGVVHAGSAAIRFEPDTWNGLLFHHHGLDLAQY
QSIRVWVHGGTAGGQAVRLVLHDGTQMLGTTRLDTALGAPIAPGVWQQVTIPLSSIGATTGTLR
DIYFQDDSGTDQPALYLDDVVLLPQ

>CBM23==Q9XCV5_CELFI - Cellulomonas fimi
VAAAPAQPVVHIASPADGARVASAPTTVRVRVGGTDVQSVTVEVAQGGTVVDTLDLAYDGAL
WWTAPWSPTSAQLDNSTYTVTATATTAAGTLDVTNEVVLGPRPTFPAGVVDDFEGYGDDTALR
AEYVTYGANTISLETGSVGGGAKALRLDYDFATQTYTGVGKQLSGDWSDFNELAIWVDPDGSN
NRMVLQLNAGGVAYEAYPSLAGDEPQLVTIPFVDWRPAPWDTAHADRRMSDDDLRALTSFNVY
VNSAEGGAASGSLVVDDIAAH

>CBM24==Q9P563_NEUCR- - Neurospora crassa
CVEGTAAAGFDELCAFTCKYGYCPVGACLCTKMGPGNKLPGPETPGFGTVGFPAKGRTASYGGL
CSFACNYGYGCPNAYC

>CBM24==Q9P563_NEUCR-2 - Neurospora crassa
CTSGVGEGGLGGLCSFSCNFGFCPIHSCTCTSTGALHVPPGASDTITGKADPSVDERTYGPLCEFA
CSRGYCPEGAC

>CBM24==Q8WZM7_TRIHA Alpha-1,3-glucanase - Trichoderma harzianum
CVAGTVADGESGNYIGLCQFSCNYGYCPPGPCKCTAFGAPISPPASNGRNGCPLPGEGDGYLGLC
SFSCNHNYCPPTACQYC

>CBM26==Q9KFR4_BACHD - Bacillus halodurans
YYKTGWTHPHIHYSLNQGAWTTLPGVPLTKSEYEGYVKVTIEAEEGSQLRAAFNNGSGQWDNN
QGRDYDFSSGVHTLADGRILSGT

>CBM26==Q97GW3_CLOAB Hypothetical protein CAC2252 - Clostridium acetobutylicum
FKDPNGWSAPNIYYYDPAGKLTGPGWPGVKMNSDGNGWYSYTIQNWTSAKVLFDDGTNQIPGV
NQPGIDVTG

>CBM26==AMY_BACSU Alpha-amylase precursor - Bacillus subtilis
YQNPNHWSQVNAYIYKHDGSRVIELTGSWPGKPMTKNADGIYTLTLPADTDTTNAKVIFNNGSA
QVPGQNQPGFDYVLNGLYNDSGLSGS

>CBM27==P77847_CALSA-1 - Caldocellum saccharolyticum
VESSVNPVVLDFEDGTVMSFGEAWGDSLKCIKKVSVSQDLQRPGNKYALRLDVEFNPNNGWDQ
GDLGTWFGGVVEGQFDFTGYKSVEFEMFIPYDEFSKSQGGFPYKVVINDGWKELGSEFNITANA
GKKVKINGKDYTVIHKAFAIPEDFRTKKRAQLVFQFAGQNSNYKGPIYLDNVRIRPEDASN

>CBM27==P77847_CALSA-2 - Caldocellum saccharolyticum
YPEQGKSFVYNFEIDTMGFYKYSGDGFNKKTKSLEYSQDLKRSGNNGALKVNASLAGTAFDEM
NIAIKLTDKDDKKFDFSKYSTLEYTLYIPNPDKISGKLMVASAIDNPWQIIKDFTAINYKDKNAIQK
INGKDYAVIKCSDNLYNVNTKANVLVLRIAGSYVKYTGPIYIDNVKLVAGKKVA

>CBM27==Q9X0V4_THEMA - Thermotoga maritima
VNEARYVLAEEVDFSSPEEVKNWWNSGTWQAEFGSPDIEWNGEVGNGALQLNVKLPGKSDWE
EVRVARKFERLSECEILEYDIYIPNVEGLKGRLRPYAVLNPGWVKIGLDMNNANVESAEIITFGGK
EYRRFHVRIEFDRTAGVKELHIGVVGDHLRYDGPIFIDNVRLYKRTGGM

>CBM29==Q9C171_PIREQ-1 - Piromyces equi
RKVSATYSVVYETGKKLNSGFDNWGWDSKMSFKDNSLVLTADPDEYGAISLKNLNSNYYGKGG
CIYLQVKTETEGLVKVQGVRGYDETEAFNVGSFRSSSDFTEYKFEVDDEYQFDRIIVQDGPASNIP
IYMRYIIYSTGSCDDFNPPVDTTKVPVTT

>CBM29==Q9C171_PIREQ-2 - Piromyces equi
SNVRATYTVIFKNASGLPNGYDNWGWGCTLSYYGGAMIINPQEGKYGAVSLKRNSGSFRGGSLR
FDMKNEGKVKILVENSEADEKFEVETISPSDEYVTYILDVDFDLPFDRIDFQDAPGNGDRIWIKNL
VHSTGSADDFVDPIN

>CBM30==P71140_CLOTM - Clostridium thermocellum
KLLDVQIFKDSPVVGWSGSGMGELETIGDTLPVDTTVTYNGLPTLRLNVQTTVQSGWWISLLTLR
GWNTHDLSQYVENGYLEFDIKGKEGGEDFVIGFRDKVYERVYGLEIDVTTVISNYVTVTTDWQH
VKIPLRDLMKINNGFDPSSVTCLVFSKRYADPFTVWFSDIKITSEDNEKSAPAIKVNQLGFIPEAEK
YAL

>CBM30==P77865_FIBSU-1 - Fibrobacter succinogenes
KRPHLAVYKFFDEQYRPGGYDYSYGGTSKGVTITKSGGYKSKAALNIKLDPKEYSGASICLYNEF
FDLNKYMLDSKVEFMIKGKNGGESVKVGLLDEEVSDGKKTQVVLPMNKYIEGGAVTTDWKKV
SIPLVDFPDRGLYWDNTRKSEFPSRIDWDKIAEIRFSIDKSAASEFEVWVDNIEIVKGNKKAAPKK
QMVYWDENNDIID

>CBM30==P77865_FIBSU-2 - Fibrobacter succinogenes
KAKTLATFYDNQVKGFSYSYGGLTAQREAQSKTPGNKNVLAMYIDNNDWSGVTYSLGEGKFID
LSKVRDKGGLYFWIKGKLGGEKLYVGILDNQGNDIKSQTKVGLNDWIKVSKDWQLAKIPLKRFT
DKGKAWDANKSAEVAKDIKWDKIQEIRFSVGKGENAGEPGKPAPVTVFVDQITFTSNIDWVDPD
LKWDSFKSNAPDYVI

>CBM30==Q9AQF4_CLOCL - Clostridium cellulovorans
KLMDLEVFKSASITGWSGSAGGELEVASDSNLPIDTSATYNGLPSLRLNVTKASAQWWSSLLTLR
GWCTQDLTQYLANGYLEFNVKGKVGGEDFQIGLQDQTHERAAGDSVTSVKSIKNYVNISTNWQ
HVKIPLKDIMGPSTGFDPTTARCINIVKGSSEIFTAWINDLKITSTDNEKSYAPIKVNQDGYLPSSEK
YAL

>CBM31==Q8RS40_9BURK - Alcaligenes sp. XY-234
GTEPPENCQDDFNFNYVSDQEIEVYHVDKGWSAGWNYVCLNDYCLPGNKSNGAFRKTFNAVLG
QDYKLTFKVEDRYGQGQQILDRNITFTTQVCN

>CBM35==Q840C0_9GAMM - Cellvibrio japonicus
MAVPEGNSWTYTAASASITAPAQLVGNVGELQGAGSAVIWNVDVPVTGEYRINLTWSSPYSSKV
NTLVMDGTALSYAFAEATVPVTYVQTKTLSAGNHSFGVRVGSSDWGYMNVHSLKLELLGGLTI
RSPAN

>CBM35==Q9F1T9_CLOTM - Clostridium thermocellum
MENKTSIKVFLALMLAVLMPVVSCINVSNAVLSDGDKYEFEDGIHKGAQIYTDYVGQNEYGEVF
DLTGSTCSFIAQKGTSTSVNVEVDKEGLYEIFICYVQPYDKNKKVQYLNVNGVNQGEISFPFTLK
WREISAGIVKLNAGINNIELESYWGYTYFDYLIVKPAD

>CBM36==Q9RC94_9BACI - Bacillus sp. 41M-1
EAESMTKGGPYTSNITSPFNGVALYANGDNVSFNHSFTKANSSFSLRGASNNSNMARVDLRIGGQ
NRGTFYFGDQYPAVYTINNINHGIGNQLVELIVTADDGTWDAYLDYLE

>CBM36==P77853_DICTH - Dictyoglomus thermophilum
ECENMSLSGPYVSRITNPFNGIALYANGDTARATVNFPASRNYNFRLRGCGNNNNLARVDLRIDG
RTVGTFYYQGTYPWEAPIDNVYVSAGSHTVEITVTADNGTWDVYADYLV

>CBM36==XYND_PAEPO-1 - Paenibacillus polymyxa
EAETIAWQAGVTTEPTQASGGPISNLNVTNIHNGDWIAVGKADFGSAGAKTFKANVATNVGGNI
EVRLDSETGPLVGSLKVPSTGGMQTWREVETTINNATGVHNIYLVFTGSGSGNLLNLDAWQ

>CBM36==XYND_PAEPO-2 - Paenibacillus polymyxa
EAENMKIGGTYAGKISAPFDGVALYANADYVSYSQYFANSTHNISVRGASSNAGTAKVDLVIGG
VTVGSFNFTGKTPTVQTLSNITHATGDQEIKLALTSDDGTWDAYVDFIE

>CBM37==Q9AJB4_RUMAL - Ruminococcus albus
PEESEIEGTLYSATFENGTNYWSGRGSASVASSSSKAYEGSKSLYVSGRTDNWNGGEVEVNTSTF
KPGNSYSFSAMVNPAESTTVQLSMQYDQNGTTKYTQIAEGSCTAGKWTKLENTSFTIPSGASNV
KMYVEAPDSLCDIYVDRAIIADKGYVA

>CBM38==O52973_BACCI-1 - Bacillus circulans
TGWSVAEGGAFGPDSVSDETVWWAERIPYDQEGAYHLNGWKYPESETGVLRSSTFELGGSGWIS
FKLGGAKDPDKAFINIVEADTGQVVARYGNSAFTDVGFPDPAQGMRLANMEQYKADLSGHIGK
KLYIEIVDHATSDWGLIFADAF

>CBM38==O52973_BACCI-2 - Bacillus circulans
TGWTVIEGDAFGPNSVSDETVWWAERIPYDQEGAYHLNGWKYPESETGVLRSSTFELGGSGWIT
FKLGGGKHTDQVYVSVIEAETGNLIARYGNSAFTDVGFPDPAQGMRLANMEQYKADLSKHIGK
KLYLEIVDHGVSDWGLVFADAF

>CBM38==O52973_BACCI-3 - Bacillus circulans
NIIPAEIANRGFETGNLDGWTVEGDAFHVTDEAHAAKEGNFYALSSTEGQGSITSNTFTLQGAGII
NFTVLDILNPEGAYVALYDASSNTVIKITGNIGANEQISWKVQEHYNKKLYVKVVDQSGDAGIAV
DGF

>CBM38==O52973_BACCI-4 - Bacillus circulans
NEQAGGDSLTGTLRSQNFVLGGNGRIDFLMSGGRDIDRLYVALVRASDGKELFKETATNYEEYQ
RKIWDASDYIGEELYIKVVDQSTGGFGHLNVDDF

>CBM39==BGBP_PLOIN - Plodia interpunctella
YVVPSAKLEAIYPKGLRVSIPDDGFSLFAFHGKLNEEMDGLEAGHWARDITKPKEGRWTFRDRN
VKLKLGDKIYFWTYVIKDGLGYRQDNGEWTVTEFVNEDGTPADT

>CBM39==BGBP_BOMMO - Bombyx mori
YEAPPATLEAIHPKGLRVSVPDEGFSLFAFHGKLNEEMEGLEAGHWSRDITKPKNGRWIFRDRNA
ALKIGDKIYFWTFVIKDGLGYRQDNGEWTVEGFVDEAGNPVNT

>CBM41==PULA_KLEAE - Klebsiella aerogenes
LVDIAGITSSTPADYATKNLYLWNNETCDALSAPVADWNDVSTTPTGSDKYGPYWVIPLTKESGS
INVIVRDGTNKLIDSGRVSFSDFTDRTVSVIAGNSAVYDSRADAFR

>CBM41==145409304 [Caldicellulosiruptor saccharolyticus DSM 8903]
KTTLIIHYYRYNEDYQGWNLWIWPVEPVGTEGKAYEFTSKDDFGVKAVVELPGKVTKVGIIVRK
GNWEAKDVAVDRFISGINGTKEVWLIEGEEQIYTSQPQKTP

>CBM41==89949809 [Saccharophagus degradans 2-40]
PNQAILFFNKADGDYEGYRLHNWNSETCNAYAEGSLAASWSNGLVHSGVDPVYGAYWLLDLV
EGHDDCGHFIIHVGTDDAGKEMGGGDWIMPLAQDDPKFTRMNFTFSGVGSVFEFPVPSLGPQKV
AIAGAAAHWLDIDTLVWNLDTTDVADVKLHYSATAGITVDDDSNVSGSTADLTFATLSEELQQL
VPHMASWPAFGNMLTASD

>CBM41==PULA_THEMA - Thermotoga maritima
ETTIVVHYHRYDGKYDGWNLWIWPVEPVSQEGKAYQFTGEDDFGKVAVVKLPMDLTKVGIIVR
LNEWQAKDVAKDRFIEIKDGKAEVWILQGVEEIFYEKPDTSP

>CBM42==Q8NK89_ASPKA - Aspergillus kawachi
LVSGPSFTSGEVVSLRVTTPGYTTRYIAHTDTTVNTQVVDDDSSTTLKEEASWTVVTGLANSQCF
SFESVDTPGSYIRHYNFELLLNANDGTKQFHEDATFCPQAALNGEGTSLRSWSYPTRYFRHYENV
LYAASNGGVQTFDSKTSFNNDVSFEIETAFAS

>CBM42==125712765 [Clostridium thermocellum ATCC 27405]
FATSVCYSTNPITKAKFQSYNYPNMYIRHANFDARIDENVTPEMDSQWELVPGLANSGDGYVSIQ
SVNYPGYYLRHSNYDLSLEKNDGTSLFAESATFKIVPGLADPSYISFQSYNFPTRYIRHYNYLLRL
DEIVTELDRQDATFKIIS

>CBM43==Q94G86_OLEEU Beta-1,3-glucanase-like protein - Olea europaea
SWCVPKPGVSDDQLTGNINYACGQGIDCGPIQPGGACFEPNTVKAHAAYVMNLYYQSAGRNSW
NCDFSQTATLTNTNPSYGACNF

>CBM44==P71140_CLOTM - Clostridium thermocellum
TEKAFKGERGLKWTVTSEGEGTAELKLDGGTIVVPGTTMTFRIWIPSGAPIAAIQPYIMPHTPDWS
EVLWNSTWKGYTMVKTDDWNEITLTLPEDVDPTWPQQMGIQVQTIDEGEFTIYVDAIDW

>CBM45==Q9CAR6_ARATH - Arabidopsis thaliana
LHHSYLRHNSKVNRGNRSFIPISLNLRSHFTSNKLLHSIGKSVGVSSMNKSPVAIRATSSDTAVVET
AQS

>CBM45==3287270 R1 [Solanum tuberosum]
LTSTVLEHKSRISPPCVGGNSLFQQQVISKSPLSTEFRGNRLKVQKKKIPMEKKRAFSSSPHAVLTT
DTSSELAEKFSL

>CBM46==Q9KF82_BACHD-1 - Bacillus halodurans
RSSVAESNFIYLKQGDRIADATVSLQLHGNELTGLRANGQRLTPGQDYELNGERLTVKAHVLSAI
ASSGTLGTNGMVTAEFNRGADWHFRVNT

>CBM46==GUNB_PAELA-1 - Paenibacillus lautus
RSSTASSDLIHVKQGTAVKDTSVQLNLNGNTLTSLSVNGTTLKSGTDYTLNSSRLTFKASQLTKLT
SLGKLGVNATIVTKFNRGADWKFNVVL

>CBM46==Q9KF82_BACHD-2 - Bacillus halodurans
PVLQSTQGHVSNFSIPASFNGNSLATMEAVYVDGGNAGPQDWTSFKEFGYAFSPSYDANEMKLT
EAFFREVRDGEVRLTFHFWSGETVNYTIIKNGNQVTGIAA

>CBM46==GUNB_PAELA-2 - Paenibacillus lautus
PKLSSTTGTTSSFAIPTAFNGDQLATMEAVYVNGGNAGPHNWTSFKEFETTFSPAYSEGKIKLQQ
AFFNEVNDTTVTLKFQFWSGEIVNYTIKKSGSTVTGTAS

>CBM48==Q9RX51_DEIRA - Deinococcus radiodurans
KKEFGGFSGFSGEDVPDPQAEQTFLNSKLNWAEREGGEHARTLRLYRDLLRLRREDPVLHNRQR
ENLTTGHDGDVLWVRTVTGAGERVLLWNLGQDTRAVAEVKLPFTVPRRLLLHTEGREDLTLGA
GEAVLVG

>CBM48==GLGB_ECOLI - Escherichia coli
MSDRIDRDVINALIAGHFADPFSVLGMHKTTAGLEVRALLPDATDVWVIEPKTGRKLAKLECLDS
RGFFSGVIPRRKNFFRYQLAVVWHGQQNLIDDPYRFGPLIQEMDAWLLSEGTHLRPYE

>CBM48==PULA_KLEAE - Klebsiella aerogenes
KSSPLFTLGDGATVMKRVDFRNTGADQQTGLLVMTIDDGMQAGRQSGQPCRRHRGGDQRRAG
KPDAAGLRRHIAPAERYSAGGGRPVAGERVQVAADGSVTLPAWSVAVLELPQASRRALACR

>CBM49==GUN2_ORYSJ - Oryza sativa subsp. japonica (Rice)
AAGHGARARGRLGQSLQHGIAANHTSLPHGANHQHASPVEIEQKATASWEKDGRTYHRYAVTV
SNRSPAGGKTVEELHIGIGKLYGPVWGLEKAARYGYVLPSWTPSLPAGESAAFVYVHAAPPADV
WVTGYKLV

>CBM49==Q9ZSP9_SOLLC - Solanum lycopersicum (Tomato) (Lycopersicon esculentum)
HAGHSGYNQLLPVVPDPKPTPKPAPRTKVTPAPRPRVLPVPANAHVTIQQRATSSWALNGKTYY
RYSAVVTNKSGKTVKNLKLSIVKLYGPLWGLTKYGNSFIFPAWLNSLPAGKSLEFVYIHTASPAIV
SVSSYTLV

Table S1. **Non-parametric diversity estimators.** Diversity estimates were calculated using four different methods involving parametric (exponential and two parameter hyperbola) and non-parametric estimators (bootstrap and jackknife). 16S sequences were assigned to clusters (Operational Taxonomical Units or OTUs) using four different thresholds to differentiate OTUs: 3%, 2%, 1% and 0% difference in sequence. The highlighted 99% sequence identity threshold was chosen for the phylotype representation in the phylogenetic tree (Fig. 1b).

OTU definition	OTUs observed	Jackknife Estimate	% observed (Jackknife)	Bootstrap estimate	% observed (Bootstrap)
100%	463	739 +/- 48	63%	574 +/- 64	89%
99%	216	303 +/- 6	71%	254 +/- 24	84%
98%	122	165 +/- 1	74%	140 +/- 12	85%
97%	84	115 +/- 1	73%	97 +/- 9	87%

Table S2. Estimates of gross community structure based on sequence composition binning, and conserved single copy gene phylogenies (weighted for contig size). For the PhyloPythia binning only fragments classified at class or phylum level were considered.

Method	***% Spirochaetes***	***% Fibrobacteres***	**% other**
Sequence composition-based classifier:			
PhyloPythia	59	6	35
Single copy gene phylogenies:			
Ribosomal protein S3 (rpsC)	73	19	8
Ribosome recycling factor	79	4	17
Ribosome-binding factor A	59	0	41
Translation initiation factor 2 (IF-2; GTPase)	100	0	0
Polyribonucleotide nucleotidyl transferase	84	16	0
Imidazoleglycerol-phosphate dehydratase	0	20	80
Elongation factor EFG (fusA)	82	5	14
Membrane GTPase (lepA)	67	33	0
Ribosomal protein L2 (rplB)	63	26	11
Elongation factor EF-Tu (tufA)	82	18	0
Average +/- standard deviation	**68 +/- 26**	**13 +/- 11**	**19 +/- 25**

Table S3. Summary of glycoside hydrolases (GHs) and carbohydrate-binding modules (CBMs) discovered in the P3 luminal microbiota.

CAZy Family	Pfam HMM Name	Pfam Accession	Pfam Description	Known Activities (CAZy)	Termite Gut Community	
					metagenome ¶	fosmids ¶
JGI ID					**2004080001**	**2004175000**
Nucleotide sequence length					**61533145**	**459633**
Protein coding genes					**82789**	**436**
Glycoside hydrolase catalytic domains						
GH1	Glyco_hydro_1	PF00232.8	Glycosyl hydrolase family 1	β-glucosidase, β-galactosidase, β-mannosidase, others	22 (256 ppm)	0
GH2	Glyco_hydro_2_C	PF02836.6	Glycosyl hydrolases family 2, TIM barrel domain	β-galactosidase, β-mannosidase, others	23 (281 ppm)	2 (3818 ppm)
GH3	Glyco_hydro_3	PF00933.11	Glycosyl hydrolase family 3 N terminal domain	β-1,4-glucosidase, β-1,4-xylosidase, β-1,3-glucosidase, α-L-arabinofuranosidase, others	69 (670 ppm)	1 (1547 ppm)
GH4	Glyco_hydro_4	PF02056.5	Family 4 glycosyl hydrolase	α-glucosidase, α-galactosidase, α-glucuronidase, others	14 (192 ppm)	0
GH5	Cellulase	PF00150.7	Cellulase (glycosyl hydrolase family 5)	cellulase, β-1,4-endoglucanase, β-1,3-glucosidase, β-1,4-endoxylanase, β-1,4-endomannanase, others	56 (662 ppm)	1 (1971 ppm)
GH6	Glyco_hydro_6	PF01341.7	Glycosyl hydrolases family 6	endoglucanase, cellobiohydrolase	0	0
GH7	Glyco_hydro_7	PF00840.10	Glycosyl hydrolase family 7	endoglucanase, cellobiohydrolase (fungal only)	0	0
GH8	Glyco_hydro_8	PF01270.7	Glycosyl hydrolases family 8	cellulase, β-1,3-glucosidase, β-1,4-endoxylanase, β-1,4-endomannanase, others	5 (70 ppm)	0
GH9	Glyco_hydro_9	PF00759.8	Glycosyl hydrolase family 9	endoglucanase, cellobiohydrolase, β-glucosidase	9 (131 ppm)	1 (2428 ppm)
GH10	Glyco_hydro_10	PF00331.10	Glycosyl hydrolase family 10	xylanase, β-1,3-endoxylanase	46 (572 ppm)	1 (2187 ppm)
GH11	Glyco_hydro_11	PF00457.7	Glycosyl hydrolases family 11	xylanase	14 (110 ppm)	0
GH12	Glyco_hydro_12	PF01670.6	Glycosyl hydrolase family 12	endoglucanase, β-1,3-1,4-glucanase, xyloglucan hydrolase	0	0
GH13	Alpha-amylase	PF00128.11	Alpha amylase, catalytic domain	α-amylase, catalytic domain, and related enzymes	48 (661 ppm)	0
GH14	Glyco_hydro_14	PF01373.7	Glycosyl hydrolase family 14	β-amylase	0	0
GH15	Glyco_hydro_15	PF00723.10	Glycosyl hydrolases family 15	glucoamylase, glucodextranase	0	0
GH16	Glyco_hydro_16	PF00722.9	Glycosyl hydrolases family 16	β-1,3(4)-endoglucanase, others	1 (8 ppm)	0
GH17	Glyco_hydro_17	PF00332.8	Glycosyl hydrolases family 17	glucan endo-1,3-β-D-glucosidase glucan 1,3-β-glucosidase, others	0	0
GH18	Glyco_hydro_18	PF00704.16	Glycosyl hydrolases family 18	chitinase, endo-β-N-acetylglucosaminidase, non-catalytic proteins	17 (223 ppm)	0
GH19	Glyco_hydro_19	PF00182.8	Chitinase class I	chitinase	0	0
GH20	Glyco_hydro_20	PF00728.11	Glycosyl hydrolase family 20, catalytic domain	β-hexosaminidase, lacto-N-biosidase	15 (152 ppm)	0
GH21	deleted family					
GH22	Lys	PF00062.10	C-type lysozyme/alpha-lactalbumin family	C-type lysozyme, α-lactalbumin (eukaryotic only)	0	0
GH23	SLT #	PF01464.10	Transglycosylase SLT domain	G-type lysozyme, peptidoglycan lytic transglycosylase	52 (294 ppm)	1 (751 ppm)
GH24	Phage_lysozyme $	PF00959.9	Phage lysozyme	lysozyme	0	0
GH25	Glyco_hydro_25	PF01183.10	Glycosyl hydrolases family 25	lysozyme	1 (5 ppm)	0
GH26	Glyco_hydro_26	PF02156.5	Glycosyl hydrolase family 26	β-1,3-xylanase, mannanase	15 (197 ppm)	0
GH27	Melibiase	PF02065.8	Melibiase	α-galactosidase, α-N-acetylgalactosaminidase, isomalto-dextranase	4 (51 ppm)	0
GH28	Glyco_hydro_28	PF00295.7	Glycosyl hydrolases family 28	polygalacturonase, rhamnogalacturonase, others	6 (92 ppm)	0
GH29	Alpha_L_fucos	PF01120.7	Alpha-L-fucosidase	α-L-fucosidase	0	0
GH30	Glyco_hydro_30	PF02055.6	O-Glycosyl hydrolase family 30	β-1,6-glucanase, β-xylosidase	0	0
GH31	Glyco_hydro_31	PF01055.15	Glycosyl hydrolases family 31	α-glucosidase, α-xylosidase, others	26 (349 ppm)	1 (2807 ppm)
GH32	Glyco_hydro_32N	PF00251.9	Glycosyl hydrolases family 32 N terminal	invertase, others	0	0
GH33	BNR	PF02012.9	BNR/Asp-box repeat	sialidase, neuraminidase	0	0
GH34	Neur	PF00064.8	Neuraminidase	sialidase, neuraminidase (viral only)	0	0
GH35	Glyco_hydro_35	PF01301.9	Glycosyl hydrolases family 35	β-galactosidase	3 (38 ppm)	0
GH36	BLAST search §			α- galactosidase, α- N-acetylgalactosaminidase	5 to 7	0
GH37	Trehalase	PF01204.8	Trehalase	trehalase	0	0
GH38	Glyco_hydro_38	PF01074.11	Glycosyl hydrolases family 38 N-terminal domain	α-mannosidase	11 (106 ppm)	0
GH39	Glyco_hydro_39	PF01229.7	Glycosyl hydrolases family 39	β-xylosidase, α-L-iduronidase	3 (62 ppm)	0
GH40	deleted family					
GH41	deleted family					
GH42	Glyco_hydro_42	PF02449.5	Beta-galactosidase	β-galactosidase	24 (332 ppm)	1 (2754 ppm)
GH43	Glyco_hydro_43 #	PF04616.4	Glycosyl hydrolases family 43	xylanase, β-xylosidase, α-L-arabinofuranosidase, arabinanase, others	16 (226 ppm)	1 (1906 ppm)
GH44	BLAST search §			endoglucanase, xyloglucanase	6	0
GH45	Glyco_hydro_45	PF02015.6	Glycosyl hydrolase family 45	endoglucanase (mainly eukaryotic, 2 bacterial)	4 (38 ppm)	1 (1939 ppm)
GH46	Glyco_hydro_46	PF01374.8	Glycosyl hydrolase family 46	chitosanase	0	0
GH47	Glyco_hydro_47	PF01532.10	Glycosyl hydrolase family 47	α-mannosidase (mainly eukaryotic, 1 bacterial)	0	0
GH48	Glyco_hydro_48	PF02011.5	Glycosyl hydrolase family 48	endoglucanase, cellobiohydrolase	0	0
GH49	Glyco_hydro_49	PF03718.3	Glycosyl hydrolase family 49	dextranase, isopullulanase, others	0	0
GH50	BLAST search §			β-agarase	0	0
GH51	BLAST search §			endoglucanase, α-L-arabinofuranosidase	18 to 19	0
GH52	Glyco_hydro_52	PF03512.3	Glycosyl hydrolase family 52	β-xylosidase	3 (47 ppm)	0
GH53	Glyco_hydro_53 #	PF07745.2	Glycosyl hydrolase family 53	β-1,4-endogalactanase	12 (154 ppm)	0
GH54	ArabFuran-catal #	PF09206.1	Alpha-L-arabinofuranosidase B, catalytic	β-xylosidase, α-L-arabinofuranosidase (mainly eukaryotic, 2 bacterial)	0	0
GH55	BLAST search §			endo-1,3-glucanase, exo-1,3-glucanase (mainly eukaryotic, 2 bacterial)	0	0

CAZy Family	Pfam HMM Name	Pfam Accession	Pfam Description	Known Activities (CAZy)	Termite Gut Community	
					metagenome ¶	fosmids ¶
JGI ID					**2004080001**	**2004175000**
Nucleotide sequence length					**61533145**	**459633**
Protein coding genes					**82789**	**436**
Glycoside hydrolase catalytic domains						
GH56	Glyco_hydro_56	PF01630.8	Hyaluronidase	hyaluronidase (eukaryotic only)	0	0
GH57	Glyco_hydro_57	PF03065.4	Glycosyl hydrolase family 57	α-amylase, 4-α-glucanotransferase, α-galactosidase, others	17 (274 ppm)	0
GH58	BLAST search §			endo-N-acetylneuraminidase or endo-sialidase	1	0
GH59	Glyco_hydro_59	PF02057.5	Glycosyl hydrolase family 59	galactocerebrosidase (mainly eukaryotic, 2 bacterial)	0	0
GH60	deleted family					
GH61	Glyco_hydro_61	PF03443.4	Glycosyl hydrolase family 61	endoglucanase (fungal only)	0	0
GH62	Glyco_hydro_62	PF03664.3	Glycosyl hydrolase family 62	α-L-arabinofuranosidase	0	0
GH63	Glyco_hydro_63	PF03200.5	Mannosyl oligosaccharide glucosidase	processing α-glucosidase	0	0
GH64	BLAST search §			β-1,3-glucanase	0	0
GH65	Glyco_hydro_65m	PF03632.4	Glycosyl hydrolase family 65 central catalytic domain	trehalase, maltose phosphorylase, trehalose phosphorylase	6 (77 ppm)	0
GH66	BLAST search §			cycloisomaltooligosaccharide glucanotransferase, dextranase	0	0
GH67	Glyco_hydro_67M	PF07488.1	Glycosyl hydrolase family 67 middle domain	α-glucuronidase, others	10 (96 ppm)	0
GH68	Glyco_hydro_68	PF02435.6	Levansucrase/Invertase	levansucrase, others	0	0
GH69	renamed PL16					
GH70	Glyco_hydro_70	PF02324.6	Glycosyl hydrolase family 70	dextransucrase, others	0	0
GH71	Glyco_hydro_71	PF03659.4	Glycosyl hydrolase family 71	α-1,3-glucanase (mainly fungal, 1 bacterial)	0	0
GH72	Glyco_hydro_72	PF03198.3	Glycolipid anchored surface protein (GAS1)	β-1,3-glucanosyltransglycosylase (fungal only)	0	0
GH73	Glucosaminidase	PF01832.9	Mannosyl-glycoprotein endo-beta-N-acetylglucosaminidase	endo-β-N-acetylglucosaminidase, others	0	0
GH74	BLAST search §			endoglucanase, cellobiohydrolase, xyloglucanase	7	0
GH75	Chitosanase	PF07335.1	Fungal chitosanase	chitosanase (mainly fungal)	0	0
GH76	Glyco_hydro_76	PF03663.4	Glycosyl hydrolase family 76	α-1,6-mannanase	0	0
GH77	Glyco_hydro_77	PF02446.7	4-alpha-glucanotransferase	4-α-glucanotransferase, amylomaltase	14 (242 ppm)	0
GH78	Bac_rhamnosid	PF05592.1	Bacterial alpha-L-rhamnosidase	α-L-rhamnosidase	0	0
GH79	Glyco_hydro_79n	PF03662.3	Glycosyl hydrolase family 79, N-terminal domain	endo-β-glucuronidase	0	0
GH80	BLAST search §			chitosanase	0	0
GH81	Glyco_hydro_81	PF03639.3	Glycosyl hydrolase family 81	β-1,3-glucanase (mainly eukaryotic, 6 bacterial)	0	0
GH82	BLAST search §			ι-carrageenase (2 bacterial)	0	0
GH83	HN	PF00423.9	Hemagglutinin-neuraminidase	hemaglutinin-neuraminidase (viral only)	0	0
GH84	Hyaluronidase_2 #	PF07555.2	Hyaluronidase	hyaluronidase, N-acetyl b-glucosaminidase	0	0
GH85	Glyco_hydro_85	PF03644.3	Glycosyl hydrolase family 85	endo-β-N-acetylglucosaminidase	0	0
GH86	BLAST search §			β-agarase	0	0
GH87	BLAST search §			α-1,3-glucanase, mycodextranase	0	0
GH88	Glyco_hydro_88	PF07470.2	Glycosyl Hydrolase Family 88	d-4,5 unsaturated β-glucuronyl hydrolase	9 (133 ppm)	0
GH89	NAGLU	PF05089.2	Alpha-N-acetylglucosaminidase (NAGLU)	α-N-acetylglucosaminidase	0	0
GH90	BLAST search §			endorhamnosidase (mainly viral, 3 bacterial)	0	0
GH91	BLAST search §			inulin fructotransferase	1	0
GH92	Glyco_hydro_92 #	PF07971.2	Glycosyl hydrolase family 92	α-1,2-mannosidase	2 (29 ppm)	0
GH93	BLAST search §			exo-arabinanase (mainly fungal, 4 bacterial)	0	0
GH94	BLAST search §			cellobiose phosphorylase, chitobiose phosphorylase, cellodextrin phosphorylase	68 to 132	1 to 2
GH95	BLAST search §			α-L-fucosidase	12 to 31	0
GH96	BLAST search §			α-agarase (2 bacterial, 2 unclassified)	0	0
GH97	BLAST search §			α-glucosidase	0	0
GH98	Glyco_hydro_98M #	PF08306.1	Glycosyl hydrolase family 98	endo-β-galactosidase	1 (12 ppm)	0
GH99	BLAST search §			endo-α-1,2-mannosidase (mainly eukaryotic, 2 bacterial)	0	0
GH100	Invertase_neut	PF04853.2	Plant neutral invertase	alkaline and neutral invertase	0	0
GH101	BLAST search §			endo-α-N-acetylgalactosaminidase	0	0
GH102	MltA	PF03562.4	MltA N-terminal domain	peptidoglycan lytic transglycosylase	0	0
GH103	TIGR: MltB #	TIGR02282	MltB: lytic murein transglycosylase B	peptidoglycan lytic transglycosylase	3 (31 ppm)	0
GH104	Phage_lysozyme $	PF00959	Phage lysozyme	peptidoglycan lytic transglycosylase	see GH24	
GH105	Glyco_hydro_88 #	PF07470.2	Glycosyl Hydrolase Family 88	unsaturated rhamnogalacturonyl hydrolase	see GH88	
GH106	BLAST search §			α-L-rhamnosidase	2	0
GH107	BLAST search §			sulfated fucan endo-1,4-fucanase (3 bacterial)	0	0
GH108	DUF847 #	PF05838.2	Predicted lysozyme (DUF847)	N-acetylmuramidase	0	0
GH109	BLAST search §			α-N-acetylgalactosaminidase	3 to 5	0
GH110	BLAST search §			α-galactosidase	0	0

CAZy Family	Pfam HMM Name	Pfam Accession	Pfam Description	Known Activities (CAZy)	Termite Gut Community	
					metagenome ¶	fosmids ¶
JGI ID					**2004080001**	**2004175000**
Nucleotide sequence length					**61533145**	**459633**
Protein coding genes					**82789**	**436**
Carbohydrate-binding domains						
CBM1	CBM_1	PF00734.8	Fungal cellulose binding domain	cellulose-binding domain (mainly fungal, 1 viral)	0	0
CBM2	CBM_2	PF00553.9	Cellulose binding domain	cellulose-binding domain	0	0
CBM3	CBM_3	PF00942.8	Cellulose binding domain	cellulose-binding domain	0	0
CBM4	CBM_4_9 &	PF02018.7	Carbohydrate binding domain	amorphous cellulose-, xylan- and glucan-binding domain	5 (35 ppm)	0
CBM5	CBM_5_12 ¥	PF02839.4	Carbohydrate binding domain	cellulose-binding domain	0	0
CBM6	CBM_6	PF03422.5	Carbohydrate binding module (family 6)	amorphous cellulose- and xylan-binding domain	13 (79 ppm)	1 (862 ppm)
CBM7	deleted entry					
CBM8	BLAST search §			cellulose-binding domain (2 bacterial, 2 eukaryotic)	0	0
CBM9	CBM_4_9 &	PF02018.7	Carbohydrate binding domain	cellulose-binding domain	see CBM4	
CBM10	CBM_10	PF02013.6	Cellulose or protein binding domain	cellulose-binding domain (aerobic bacteria) and dockerin (anaerobic fungi)	0	0
CBM11	CBM_11 #	PF03425.3	Carbohydrate binding domain (family 11)	glucan-binding domain	3 (26 ppm)	0
CBM12	CBM_5_12 ¥	PF02839.4	Carbohydrate binding domain	chitin-binding domain	see CBM5	
CBM13	Ricin_B_lectin	PF00652.11	Ricin-type beta-trefoil lectin domain	mannose- and xylan-binding domain	0	0
CBM14	CBM_14	PF01607.12	Chitin binding Peritrophin-A domain	chitin-binding domain	0	0
CBM15	CBM_15	PF03426.4	Carbohydrate binding domain (family 15)	xylan-binding domain (2 bacterial)	0	0
CBM16	CBM_4_9 &	PF02018.7	Carbohydrate binding domain	carbohydrate-binding module 16	see CBM4	
CBM17	CBM_17_28 €	PF03424.3	Carbohydrate binding domain (family 17/28)	amorphous cellulose- and cellooligosaccharide-binding domain	0	0
CBM18	Chitin_bind_1	PF00187.8	Chitin recognition protein	chitin-binding domain (eukaryotic only)	0	0
CBM19	CBM_19	PF03427.3	Carbohydrate binding domain (family 19)	chitin-binding domain (eukaryotic only)	0	0
CBM20	CBM_20	PF00686.9	Starch binding domain	starch-binding domain	0	0
CBM21	CBM_21	PF03370.3	Putative phosphatase regulatory subunit	starch-binding domain (mainly eukaryotic, 1 bacterial)	0	0
CBM22	CBM_4_9 &	PF02018.7	Carbohydrate binding domain	xylan-binding domain	see CBM4	
CBM23	BLAST search §			mannan-binding domain (3 bacterial)	0	0
CBM24	BLAST search §			α-1,3-glucan (mutan)-binding domain (fungal only)	0	0
CBM25	CBM_25	PF03423.3	Carbohydrate binding domain (family 25)	starch-binding domain	0	0
CBM26	BLAST search §			starch-binding domain	0	0
CBM27	CBM27 #	PF09212.1	Carbohydrate binding module 27	mannan-binding domain	0	0
CBM28	CBM_17_28 €	PF03424.3	Carbohydrate binding domain (family 17/28)	amorphous cellulose- and cellooligosaccharide-binding domain	see CBM17	
CBM29	BLAST search §			mannan/glucomannan-binding domain (1 eukaryotic)	0	0
CBM30	BLAST search §			cellulose-binding domain	1 to 8	0
CBM31	BLAST search §			β-1,3-xylan binding domain (4 bacterial)	0	0
CBM32	F5_F8_type_C	PF00754.14	F5/8 type C domain	galactose- and lactose-binding domain	4 (22 ppm)	0
CBM33	Chitin_bind_3	PF03067.5	Chitin binding domain	chitin-binding domain	0	0
CBM34	Alpha_amylase_N	PF02903.4	Alpha amylase, N-terminal ig-like domain	starch-binding domain	0	0
CBM35	BLAST search §			mannan- xylan- and β-galactan-binding domain	0 to 1	0
CBM36	BLAST search §			xylan-binding domain	2 to 13	0 to 1
CBM37	BLAST search §			broad binding specificity	1	0
CBM38	BLAST search §			inulin-binding domain	0	0
CBM39	BLAST search §			β-1,3-glucan binding domain (eukaryotic only)	0	0
CBM40	Sialidase	PF02973.6	Sialidase, N-terminal domain	sialic acid-binding domain	0	0
CBM41	PUD#	PF03714.4	Bacterial pullanase-associated domain	α-glucan-, amylose-, amylopectin-, and pullulan-binding domain	0	0
CBM42	AbfB#	PF05270.3	Alpha-L-arabinofuranosidase B (ABFB)	arabinose-binding domain	0	0
CBM43	X8 #	PF07983.2	X8 domain	β-1,3-glucan binding domain (eukaryotic only)	0	0
CBM44	BLAST search §			cellulose- and xyloglucan-binding domain (2 bacterial)	0	0
CBM45	BLAST search §			starch-binding domain (eukaryotic only)	0	0
CBM46	BLAST search §			cellulose-binding domain	0	0
CBM47	F5_F8_type_C #	PF00754.14	F5/8 type C domain	fucose-binding domain	see CBM32	
CBM48	BLAST search §			glycogen-binding domain	0 to 1	0
CBM49	BLAST search §			crystalline cellulose-binding domain (eukaryotic only)	0	0

CAZy Family	Pfam HMM Name	Pfam Accession	Pfam Description	Known Activities (CAZy)	Termite Gut Community	
					metagenome ¶	fosmids ¶
JGI ID					**2004080001**	**2004175000**
Nucleotide sequence length					**61533145**	**459633**
Protein coding genes					**82789**	**436**
Other domains oftentimes associated with GH catalytic domains						
	Alpha-amylase_C	PF02806.6	Alpha amylase, C-terminal all-beta domain	α-amylase, C-terminal all-β domain; associated with GH13	0	0
	Alpha-L-AF_C	PF06964.1	Alpha-L-arabinofuranosidase C-terminus	αL-arabinofuranosidase, C-terminal domain; associated with GH51	10 (67 ppm)	0
	Alpha-mann_mid	PF09261.1	Alpha mannosidase, middle domain	middle domain; associated with GH38	15 (56 ppm)	0
	Big_1	PF02369.6	Bacterial Ig-like domain (group 1)	Bacterial Ig-like domain (group 1)	0	0
	Big_2	PF02368.7	Bacterial Ig-like domain (group 2)	Bacterial Ig-like domain (group 2)	140 (540 ppm)	0
	Big_3	PF07523.2	Bacterial Ig-like domain (group 3)	Bacterial Ig-like domain (group 3)	22 (72 ppm)	0
	Big_4	PF07532.1	Bacterial Ig-like domain (group 4)	Bacterial Ig-like domain (group 4)	0	0
	Bgal_small_N	PF02929.6	Beta galactosidase small chain	β-galactosidase, small chain; associated with GH2	5 (53 ppm)	1 (1932 ppm)
	CBM_X	PF06204.1	Putative carbohydrate binding domain	associated with GH94	44 (136 ppm)	1 (418 ppm)
	CelD_N	PF02927.4	N-terminal ig-like domain of cellulase	N-terminal Ig-like domain of cellulase; associated with GH9	5 (22 ppm)	1 (607 ppm)
	CHB_HEX	PF03173.3	Putative carbohydrate binding domain	Putative carbohydrate binding domain; associated with GH20	0	0
	CHB_HEX_C	PF03174.3	Chitobiase/beta-hexosaminidase C-terminal domain	Chitobiase/beta-hexosaminidase C-terminal domain; associated with GH20	0	0
	ChiC	PF06483.2	Chitinase C	Chitinase C; associated with GH18	0	0
	ChitinaseA_N	PF08329.1	Chitinase A, N-terminal domain	Chitinase A, N-terminal domain; associated with GH18	0	0
	Cohesin	PF00963.8	Cohesin domain	Cohesin domain	0	0
	Dockerin_1	PF00404.8	Dockerin type I repeat	Dockerin type I repeat	0	0
	fn3	PF00041.10	Fibronectin type III domain	Fibronectin type III domain	45 (181 ppm)	0
	GDE_C	PF06202.3	Amylo-alpha-1,6-glucosidase	Amylo-alpha-1,6-glucosidase	1 (12 ppm)	0
	Glucodextran_B	PF09136.1	Glucodextranase, domain B	Glucodextranase, domain B; associated with GH15	0	0
	Glucodextran_N	PF09137.1	Glucodextranase, domain N	Glucodextranase, domain N; associated with GH15	0	0
	Glyco_hydro_2_N	PF02837.7	Glycosyl hydrolases family 2, sugar binding domain	sugar-binding domain	26 (178 ppm)	2 (2206 ppm)
	Glyco_hydro_2	PF00703.10	Glycosyl hydrolases family 2, immunoglobulin-like beta-sandwich domain	immunoglobulin-like β-sandwich domain	4 (21 ppm)	0
	Glyco_hydro_3_C	PF01915.11	Glycosyl hydrolase family 3 C terminal domain	C-terminal domain (glycan-binding?)	42 (402 ppm)	1 (2317 ppm)
	Glyco_hydro_20b	PF02838.5	Glycosyl hydrolase family 20, domain 2	domain 2	0	0
	Glyco_hydro_32C	PF08244.1	Glycosyl hydrolases family 32 C terminal	C-terminal domain	0	0
	Glyco_hydro_38C	PF07748.2	Glycosyl hydrolases family 38 C-terminal domain	C-terminal domain	12 (170 ppm)	0
	Glyco_hydro_42M	PF08532.1	Beta-galactosidase trimerisation domain	trimerisation domain	20 (163 ppm)	1 (1371 ppm)
	Glyco_hydro_42C	PF08533.1	Beta-galactosidase C-terminal domain	C-terminal domain	8 (23 ppm)	1 (392 ppm)
	Glyco_hydro_65N	PF03636.4	Glycosyl hydrolase family 65, N-terminal domain	N-terminal domain	1 (10 ppm)	0
	Glyco_hydro_65C	PF03633.4	Glycosyl hydrolase family 65, C-terminal domain	C-terminal domain	3 (8 ppm)	0
	Glyco_hydro_67N	PF03648.4	Glycosyl hydrolase family 67 N-terminus	N-terminal domain	2 (11 ppm)	0
	Glyco_hydro_67C	PF07477.1	Glycosyl hydrolase family 67 C-terminus	C-terminal domain	11 (112 ppm)	0
	Glyco_hydro_98C	PF08307.1	Glycosyl hydrolase family 98 C-terminal domain	C-terminal domain	0	0
	Glyco_transf_36	PF06165.1	Glycosyltransferase family 36	associated with GH94	47 (248 ppm)	1 (738 ppm)
	GT36_AF	PF06205.2	Glycosyltransferase 36 associated family	associated with GH94	45 (189 ppm)	1 (568 ppm)
	He_PIG	PF05345.2	Putative Ig domain	Putative Ig-like domain	11 (24 ppm)	2 (627 ppm)
	Isoamylase_N	PF02922.7	Isoamylase N-terminal domain	Isoamylase N-terminal domain; associated with GH13	15 (67 ppm)	0
	TIG	PF01833.13	IPT/TIG domain	Ig-like domain	0	0
Footnotes						
¶	Pfam HMM hits are counted if their e-value is smaller than 1e-4. BLAST hits are counted if their e-value is smaller than 1e-6. The value in parenthesis is the ratio between the total nucleotide length of the hits for this family and the total nucleotide l					
#	CAZy does not provide weblinks to Pfam for this family, but this Pfam HMM recognizes the sequences of this family.					
§	No Pfam HMM is available for this family. BLAST searches were conducted with segments of selected sequences listed at CAZy corresponding to this family.					
$	The CAZy website links to Pfam PF00959.9 (Phage_lysozyme) for both, GH24 and GH104.					
&	The CAZy website links to Pfam PF02018.7 (CBM_4_9) for all, CBM4, CBM9, CBM16, and CBM22.					
¥	The CAZy website links to Pfam PF02839.4 (CBM_5_12) for both, CBM5 and CBM12.					
€	The CAZy website links to Pfam PF03424.3 (CBM_17_28) for both, CBM17 and CBM28.					

Table S4. Summary of the P3 luminal microbiota GH and CBM domains, their PhyloPhytia binning information, proteomics, and activity results.

CAZy Family	Pfam HMM Name	Known Activities (CAZy)	PhyloPhytia		Proteomics		Activity	
			treponema metagenome	fibrobacter (fosmids)		data mining	hybridization active (inactive)	activity screen
JGI ID								
Nucleotide sequence length								
Protein coding genes								
Glycoside hydrolase catalytic domains								
GH1	Glyco_hydro_1	β-glucosidase, β-galactosidase, β-mannosidase, others	5	0	1			
GH2	Glyco_hydro_2_C	β-galactosidase, β-mannosidase, others	4 (2)	0				
GH3	Glyco_hydro_3	β-1,4-glucosidase, β-1,4-xylosidase, β-1,3-glucosidase, α-L-arabinofuranosidase, others	14 (1)	0	1			
GH4	Glyco_hydro_4	α-glucosidase, α-galactosidase, α-glucuronidase, others	3	0	2			
GH5	Cellulase	cellulase, β-1,4-endoglucanase, β-1,3-glucosidase, β-1,4-endoxylanase, β-1,4-endomannanase, others	6 (1)	0	2	11 (0)	8 (2)	7 (2 partial)
GH6	Glyco_hydro_6	endoglucanase, cellobiohydrolase	0	0				
GH7	Glyco_hydro_7	endoglucanase, cellobiohydrolase (fungal only)	0	0				
GH8	Glyco_hydro_8	cellulase, β-1,3-glucosidase, β-1,4-endoxylanase, β-1,4-endomannanase, others	0	0				
GH9	Glyco_hydro_9	endoglucanase, cellobiohydrolase, β-glucosidase	0	2 (1)			3 (4)	4 (1)
GH10	Glyco_hydro_10	xylanase, β-1,3-endoxylanase	11 (1)	0	1			
GH11	Glyco_hydro_11	xylanase	4	0				
GH12	Glyco_hydro_12	endoglucanase, β-1,3-1,4-glucanase, xyloglucan hydrolase	0	0				
GH13	Alpha-amylase	αamylase, catalytic domain, and related enzymes	6	0	2			
GH14	Glyco_hydro_14	βamylase	0	0				
GH15	Glyco_hydro_15	glucoamylase, glucodextranase	0	0				
GH16	Glyco_hydro_16	β-1,3(4)-endoglucanase, others	0	0				
GH17	Glyco_hydro_17	glucan endo-1,3-β-D-glucosidase glucan 1,3-β-glucosidase, others	0	0				
GH18	Glyco_hydro_18	chitinase, endo-β-N-acetylglucosaminidase, non-catalytic proteins	4	0				
GH19	Glyco_hydro_19	chitinase	0	0				
GH20	Glyco_hydro_20	β-hexosaminidase, lacto-N-biosidase	1	0				
GH21	deleted family							
GH22	Lys	C-type lysozyme, α-lactalbumin (eukaryotic only)	0	0				
GH23	SLT #	G-type lysozyme, peptidoglycan lytic transglycosylase	11 (1)	0				
GH24	Phage_lysozyme $	lysozyme	0	0				
GH25	Glyco_hydro_25	lysozyme	1	0				
GH26	Glyco_hydro_26	β-1,3-xylanase, mannanase	3	0				
GH27	Melibiase	αgalactosidase, α-N-acetylgalactosaminidase, isomalto-dextranase	0	0				
GH28	Glyco_hydro_28	polygalacturonase, rhamnogalacturonase, others	1	0				
GH29	Alpha_L_fucos	αL-fucosidase	0	0				
GH30	Glyco_hydro_30	β-1,6-glucanase, β-xylosidase	0	0				
GH31	Glyco_hydro_31	α-glucosidase, α-xylosidase, others	7 (1)	0				
GH32	Glyco_hydro_32N	invertase, others	0	0				
GH33	BNR	sialidase, neuraminidase	0	0				
GH34	Neur	sialidase, neuraminidase (viral only)	0	0				
GH35	Glyco_hydro_35	βgalactosidase	1	0				
GH36	BLAST search §	α- galactosidase, α- N-acetylgalactosaminidase						
GH37	Trehalase	trehalase	0	0				
GH38	Glyco_hydro_38	α-mannosidase	3	0				
GH39	Glyco_hydro_39	βxylosidase, αL-iduronidase	2	0				
GH40	deleted family							
GH41	deleted family							
GH42	Glyco_hydro_42	βgalactosidase	4 (1)	0	1			
GH43	Glyco_hydro_43 #	xylanase, βxylosidase, αL-arabinofuranosidase, arabinanase, others	6 (1)	0				
GH44	BLAST search §	endoglucanase, xyloglucanase						
GH45	Glyco_hydro_45	endoglucanase (mainly eukaryotic, 2 bacterial)	0	(1)		1 (0)		1 (0)
GH46	Glyco_hydro_46	chitosanase	0	0				
GH47	Glyco_hydro_47	αmannosidase (mainly eukaryotic, 1 bacterial)	0	0				
GH48	Glyco_hydro_48	endoglucanase, cellobiohydrolase	0	0				
GH49	Glyco_hydro_49	dextranase, isopullulanase, others	0	0				
GH50	BLAST search §	β-agarase						
GH51	BLAST search §	endoglucanase, αL-arabinofuranosidase						
GH52	Glyco_hydro_52	βxylosidase	0	0				
GH53	Glyco_hydro_53 #	β-1,4-endogalactanase	2	0				
GH54	ArabFuran-catal #	βxylosidase, αL-arabinofuranosidase (mainly eukaryotic, 2 bacterial)	0	0				
GH55	BLAST search §	endo-1,3-glucanase, exo-1,3-glucanase (mainly eukaryotic, 2 bacterial)						

CAZy Family	Pfam HMM Name	Known Activities (CAZy)	PhyloPhytia		Proteomics		Activity	
			treponema metagenome	fibrobacter (fosmids)		data mining	hybridization active (inactive)	activity screen
JGI ID								
Nucleotide sequence length								
Protein coding genes								
Glycoside hydrolase catalytic domains								
GH56	Glyco_hydro_56	hyaluronidase (eukaryotic only)	0	0				
GH57	Glyco_hydro_57	α-amylase, 4- α-glucanotransferase, α-galactosidase, others	6	0				
GH58	BLAST search §	endo-N-acetylneuraminidase or endo-sialidase						
GH59	Glyco_hydro_59	galactocerebrosidase (mainly eukaryotic, 2 bacterial)	0	0				
GH60	deleted family							
GH61	Glyco_hydro_61	endoglucanase (fungal only)	0	0				
GH62	Glyco_hydro_62	α-L-arabinofuranosidase	0	0				
GH63	Glyco_hydro_63	processing α-glucosidase	0	0				
GH64	BLAST search §	β-1,3-glucanase						
GH65	Glyco_hydro_65m	trehalase, maltose phosphorylase, trehalose phosphorylase	0	0				
GH66	BLAST search §	cycloisomaltooligosaccharide glucanotransferase, dextranase						
GH67	Glyco_hydro_67M	α-glucuronidase, others	4	0				
GH68	Glyco_hydro_68	levansucrase, others	0	0				
GH69	renamed PL16							
GH70	Glyco_hydro_70	dextransucrase, others	0	0				
GH71	Glyco_hydro_71	α-1,3-glucanase (mainly fungal, 1 bacterial)	0	0				
GH72	Glyco_hydro_72	β-1,3-glucanosyltransglycosylase (fungal only)	0	0				
GH73	Glucosaminidase	endo- β-N-acetylglucosaminidase, others	0	0				
GH74	BLAST search §	endoglucanase, cellobiohydrolase, xyloglucanase						
GH75	Chitosanase	chitosanase (mainly fungal)	0	0				
GH76	Glyco_hydro_76	α-1,6-mannanase	0	0				
GH77	Glyco_hydro_77	4-α-glucanotransferase, amylomaltase	4	0				
GH78	Bac_rhamnosid	α-L-rhamnosidase	0	0				
GH79	Glyco_hydro_79n	endo- β-glucuronidase	0	0				
GH80	BLAST search §	chitosanase						
GH81	Glyco_hydro_81	β-1,3-glucanase (mainly eukaryotic, 6 bacterial)	0	0				
GH82	BLAST search §	ι−carrageenase (2 bacterial)						
GH83	HN	hemaglutinin-neuraminidase (viral only)	0	0				
GH84	Hyaluronidase_2 #	hyaluronidase, N-acetyl b-glucosaminidase	0	0				
GH85	Glyco_hydro_85	endo- β-N-acetylglucosaminidase	0	0				
GH86	BLAST search §	β-agarase						
GH87	BLAST search §	α-1,3-glucanase, mycodextranase						
GH88	Glyco_hydro_88	d-4,5 unsaturated β-glucuronyl hydrolase	3	0	1			
GH89	NAGLU	α-N-acetylglucosaminidase	0	0				
GH90	BLAST search §	endorhamnosidase (mainly viral, 3 bacterial)						
GH91	BLAST search §	inulin fructotransferase						
GH92	Glyco_hydro_92 #	α-1,2-mannosidase	0	0				
GH93	BLAST search §	exo-arabinanase (mainly fungal, 4 bacterial)						
GH94	BLAST search §	cellobiose phosphorylase, chitobiose phosphorylase, cellodextrin phosphorylase						
GH95	BLAST search §	α-L-fucosidase						
GH96	BLAST search §	α-agarase (2 bacterial, 2 unclassified)						
GH97	BLAST search §	α-glucosidase						
GH98	Glyco_hydro_98M #	endo- β-galactosidase	0	0				
GH99	BLAST search §	endo- α-1,2-mannosidase (mainly eukaryotic, 2 bacterial)						
GH100	Invertase_neut	alkaline and neutral invertase	0	0				
GH101	BLAST search §	endo- α-N-acetylgalactosaminidase						
GH102	MltA	peptidoglycan lytic transglycosylase	0	0				
GH103	TIGR: MltB #	peptidoglycan lytic transglycosylase	0	0				
GH104	Phage_lysozyme $	peptidoglycan lytic transglycosylase						
GH105	Glyco_hydro_88 #	unsaturated rhamnogalacturonyl hydrolase						
GH106	BLAST search §	α-L-rhamnosidase						
GH107	BLAST search §	sulfated fucan endo-1,4-fucanase (3 bacterial)						
GH108	DUF847 #	N-acetylmuramidase	0	0				
GH109	BLAST search §	α-N-acetylgalactosaminidase						
GH110	BLAST search §	α-galactosidase						

CAZy Family	Pfam HMM Name	Known Activities (CAZy)	PhyloPhytia		Proteomics		Activity	
			treponema	fibrobacter		data mining	hybridization	activity screen
			metagenome (fosmids)				active (inactive)	
JGI ID								
Nucleotide sequence length								
Protein coding genes								
Carbohydrate-binding domains								
CBM1	CBM_1	cellulose-binding domain (mainly fungal, 1 viral)	0	0				
CBM2	CBM_2	cellulose-binding domain	0	0				
CBM3	CBM_3	cellulose-binding domain	0	0				
CBM4	CBM_4_9 [&]	amorphous cellulose-, xylan- and glucan-binding domain	0	0	1			
CBM5	CBM_5_12 [¥]	cellulose-binding domain	0	0				
CBM6	CBM_6	amorphous cellulose- and xylan-binding domain	2 (1)	0				
CBM7	deleted entry							
CBM8	BLAST search [§]	cellulose-binding domain (2 bacterial, 2 eukaryotic)						
CBM9	CBM_4_9 [&]	cellulose-binding domain						
CBM10	CBM_10	cellulose-binding domain (aerobic bacteria) and dockerin (anaerobic fungi)	0	0				
CBM11	CBM_11 [#]	glucan-binding domain	0	0				
CBM12	CBM_5_12 [¥]	chitin-binding domain						
CBM13	Ricin_B_lectin	mannose- and xylan-binding domain	0	0				
CBM14	CBM_14	chitin-binding domain	0	0				
CBM15	CBM_15	xylan-binding domain (2 bacterial)	0	0				
CBM16	CBM_4_9 [&]	carbohydrate-binding module 16						
CBM17	CBM_17_28 [€]	amorphous cellulose- and cellooligosaccharide-binding domain	0	0				
CBM18	Chitin_bind_1	chitin-binding domain (eukaryotic only)	0	0				
CBM19	CBM_19	chitin-binding domain (eukaryotic only)	0	0				
CBM20	CBM_20	starch-binding domain	0	0				
CBM21	CBM_21	starch-binding domain (mainly eukaryotic, 1 bacterial)	0	0				
CBM22	CBM_4_9 [&]	xylan-binding domain						
CBM23	BLAST search [§]	mannan-binding domain (3 bacterial)						
CBM24	BLAST search [§]	α-1,3-glucan (mutan)-binding domain (fungal only)						
CBM25	CBM_25	starch-binding domain	0	0				
CBM26	BLAST search [§]	starch-binding domain						
CBM27	CBM27 [#]	mannan-binding domain	0	0				
CBM28	CBM_17_28 [€]	amorphous cellulose- and cellooligosaccharide-binding domain						
CBM29	BLAST search [§]	mannan/glucomannan-binding domain (1 eukaryotic)						
CBM30	BLAST search [§]	cellulose-binding domain						
CBM31	BLAST search [§]	β-1,3-xylan binding domain (4 bacterial)						
CBM32	F5_F8_type_C	galactose- and lactose-binding domain	0	1				
CBM33	Chitin_bind_3	chitin-binding domain	0	0				
CBM34	Alpha_amylase_N	starch-binding domain	0	0				
CBM35	BLAST search [§]	mannan- xylan- and β-galactan-binding domain						
CBM36	BLAST search [§]	xylan-binding domain						
CBM37	BLAST search [§]	broad binding specificity						
CBM38	BLAST search [§]	inulin-binding domain						
CBM39	BLAST search [§]	β-1,3-glucan binding domain (eukaryotic only)						
CBM40	Sialidase	sialic acid-binding domain	0	0				
CBM41	PUD [#]	α-glucan-, amylose-, amylopectin-, and pullulan-binding domain	0	0				
CBM42	AbfB [#]	arabinose-binding domain	0	0				
CBM43	X8 [#]	β-1,3-glucan binding domain (eukaryotic only)	0	0				
CBM44	BLAST search [§]	cellulose- and xyloglucan-binding domain (2 bacterial)						
CBM45	BLAST search [§]	starch-binding domain (eukaryotic only)						
CBM46	BLAST search [§]	cellulose-binding domain						
CBM47	F5_F8_type_C [#]	fucose-binding domain						
CBM48	BLAST search [§]	glycogen-binding domain						
CBM49	BLAST search [§]	crystalline cellulose-binding domain (eukaryotic only)						

CAZy Family	Pfam HMM Name	Known Activities (CAZy)	PhyloPhytia		Proteomics		Activity	
			treponema	fibrobacter		data mining	hybridization	activity screen
			metagenome	(fosmids)			active (inactive)	
JGI ID								
Nucleotide sequence length								
Protein coding genes								
Other domains oftentimes associated with GH catalytic domains								
	Alpha-amylase_C	α-amylase, C-terminal all-β domain; associated with GH13	0	0				
	Alpha-L-AF_C	α-L-arabinofuranosidase, C-terminal domain; associated with GH51	1	0				
	Alpha-mann_mid	middle domain; associated with GH38	3	0				
	Big_1	Bacterial Ig-like domain (group 1)	0	0				
	Big_2	Bacterial Ig-like domain (group 2)	7	0				
	Big_3	Bacterial Ig-like domain (group 3)	0	0				
	Big_4	Bacterial Ig-like domain (group 4)	0	0				
	Bgal_small_N	β-galactosidase, small chain; associated with GH2	(1)	0				
	CBM_X	associated with GH94	8 (1)	0	8			
	CelD_N	N-terminal Ig-like domain of cellulase; associated with GH9	0	1 (1)				
	CHB_HEX	Putative carbohydrate binding domain; associated with GH20	0	0				
	CHB_HEX_C	Chitobiase/beta-hexosaminidase C-terminal domain; associated with GH20	0	0				
	ChiC	Chitinase C; associated with GH18	0	0				
	ChitinaseA_N	Chitinase A, N-terminal domain; associated with GH18	0	0				
	Cohesin	Cohesin domain	0	0				
	Dockerin_1	Dockerin type I repeat	0	0				
	fn3	Fibronectin type III domain	4	0				
	GDE_C	Amylo-alpha-1,6-glucosidase	0	0				
	Glucodextran_B	Glucodextranase, domain B; associated with GH15	0	0				
	Glucodextran_N	Glucodextranase, domain N; associated with GH15	0	0				
	Glyco_hydro_2_N	sugar-binding domain	5 (2)	2				
	Glyco_hydro_2	immunoglobulin-like β-sandwich domain	0	0				
	Glyco_hydro_3_C	C-terminal domain (glycan-binding?)	8 (1)	1	3			
	Glyco_hydro_20b	domain 2	0	0				
	Glyco_hydro_32C	C-terminal domain	0	0				
	Glyco_hydro_38C	C-terminal domain	2	0				
	Glyco_hydro_42M	trimerisation domain	4 (1)	0				
	Glyco_hydro_42C	C-terminal domain	1 (1)	0				
	Glyco_hydro_65N	N-terminal domain	0	0				
	Glyco_hydro_65C	C-terminal domain	0	0				
	Glyco_hydro_67N	N-terminal domain	0	0				
	Glyco_hydro_67C	C-terminal domain	2	0				
	Glyco_hydro_98C	C-terminal domain	0	0				
	Glyco_transf_36	associated with GH94	12 (1)	0	8			
	GT36_AF	associated with GH94	13 (1)	0	8			
	He_PIG	Putative Ig-like domain	1	0 (1)				
	Isoamylase_N	Isoamylase N-terminal domain; associated with GH13	2	0				
	TIG	Ig-like domain	0	0				
Footnotes								
¶	Pfam HMM hits are counted if their e-value is smaller than 1e-4. BLAST hits are counted if their e-value is smaller than 1e-6. The value in parenthesis is the ratio between the total nucleotide length of the hits for this family and the total nucleotide l							
#	CAZy does not provide weblinks to Pfam for this family, but this Pfam HMM recognizes the sequences of this family.							
§	No Pfam HMM is available for this family. BLAST searches were conducted with segments of selected sequences listed at CAZy corresponding to this family.							
$	The CAZy website links to Pfam PF00959.9 (Phage_lysozyme) for both, GH24 and GH104.							
&	The CAZy website links to Pfam PF02018.7 (CBM_4_9) for all, CBM4, CBM9, CBM16, and CBM22.							
¥	The CAZy website links to Pfam PF02839.4 (CBM_5_12) for both, CBM5 and CBM12.							
€	The CAZy website links to Pfam PF03424.3 (CBM_17_28) for both, CBM17 and CBM28.							

Table S5. Comparison of abundance of selected GH and CBM domains in different single organism genomes and metagenome datasets.

CAZy Family	Pfam HMM Name	Known Activities (CAZy)	Human Gut Community		Soil (Diversa Silage)¶	Fibrobacter succinogenes¶	Saccharophagus degradans¶
			subject 7¶	subject 8¶			
JGI ID			**2004002000**	**2004002001**	**2001200001**	**2001800000**	**635640000**
Nucleotide sequence length			**15817685**	**20486813**	**152406385**	**4024184**	**5057531**
Protein coding genes			**20523**	**25983**	**184374**	**3341**	**4017**
Glycoside hydrolase catalytic domains							
GH1	Glyco_hydro_1	β-glucosidase, β-galactosidase, β-mannosidase, others	20 (1209 ppm)	23 (1158 ppm)	9 (46 ppm)	0	2 (536 ppm)
GH2	Glyco_hydro_2_C	β-galactosidase, β-mannosidase, others	13 (616 ppm)	17 (668 ppm)	7 (34 ppm)	1 (207 ppm)	6 (1102 ppm)
GH3	Glyco_hydro_3	β-1,4-glucosidase, β-1,4-xylosidase, β-1,3-glucosidase, α-L-arabinofuranosidase, others	20 (780 ppm)	23 (682 ppm)	44 (160 ppm)	2 (351 ppm)	6 (807 ppm)
GH4	Glyco_hydro_4	α-glucosidase, α-galactosidase, α-glucuronidase, others	4 (250 ppm)	5 (206 ppm)	7 (38 ppm)	0	0
GH5	Cellulase	cellulase, β-1,4-endoglucanase, β-1,3-glucosidase, β-1,4-endoxylanase, β-1,4-endomannanase, others	1 (49 ppm)	1 (44 ppm)	2 (10 ppm)	11 (2483 ppm)	17 (2834 ppm)
GH6	Glyco_hydro_6	endoglucanase, cellobiohydrolase	0	0	3 (12 ppm)	0	1 (209 ppm)
GH7	Glyco_hydro_7	endoglucanase, cellobiohydrolase (fungal only)	0	0	0	0	0
GH8	Glyco_hydro_8	cellulase, β-1,3-glucosidase, β-1,4-endoxylanase, β-1,4-endomannanase, others	0	1 (55 ppm)	1 (6 ppm)	6 (1602 ppm)	0
GH9	Glyco_hydro_9	endoglucanase, cellobiohydrolase, β-glucosidase	0	0	2 (12 ppm)	8 (2901 ppm)	3 (823 ppm)
GH10	Glyco_hydro_10	xylanase, β-1,3-endoxylanase	0	2 (97 ppm)	11 (47 ppm)	7 (1689 ppm)	5 (928 ppm)
GH11	Glyco_hydro_11	xylanase	0	0	0	4 (595 ppm)	2 (243 ppm)
GH12	Glyco_hydro_12	endoglucanase, β-1,3-1,4-glucanase, xyloglucan hydrolase	0	0	1 (3 ppm)	0	0
GH13	Alpha-amylase	α-amylase, catalytic domain, and related enzymes	27 (1636 ppm)	45 (2213 ppm)	42 (228 ppm)	3 (768 ppm)	10 (2272 ppm)
GH14	Glyco_hydro_14	β-amylase	0	0	0	0	0
GH15	Glyco_hydro_15	glucoamylase, glucodextranase	0	0	17 (90 ppm)	0	1 (267 ppm)
GH16	Glyco_hydro_16	β-1,3(4)-endoglucanase, others	0	1 (33 ppm)	8 (29 ppm)	4 (620 ppm)	8 (1059 ppm)
GH17	Glyco_hydro_17	glucan endo-1,3- β-D-glucosidase glucan 1,3- β-glucosidase, others	0	0	0	0	0
GH18	Glyco_hydro_18	chitinase, endo- β-N-acetylglucosaminidase, non-catalytic proteins	1 (58 ppm)	3 (156 ppm)	11 (49 ppm)	2 (458 ppm)	5 (1123 ppm)
GH19	Glyco_hydro_19	chitinase	0	0	0	0	0
GH20	Glyco_hydro_20	β-hexosaminidase, lacto-N-biosidase	2 (100 ppm)	6 (259 ppm)	15 (69 ppm)	0	2 (447 ppm)
GH22	Lys	C-type lysozyme, α-lactalbumin (eukaryotic only)	0	0	0	0	0
GH23	SLT #	G-type lysozyme, peptidoglycan lytic transglycosylase	4 (93 ppm)	1 (17 ppm)	37 (85 ppm)	2 (175 ppm)	4 (273 ppm)
GH24	Phage_lysozyme $	lysozyme	0	0	0	0	0
GH25	Glyco_hydro_25	lysozyme	8 (255 ppm)	4 (102 ppm)	6 (18 ppm)	0	0
GH26	Glyco_hydro_26	β-1,3-xylanase, mannanase	0	1 (33 ppm)	0	4 (1011 ppm)	1 (191 ppm)
GH27	Melibiase	α-galactosidase, α-N-acetylgalactosaminidase, isomalto-dextranase	3 (163 ppm)	8 (408 ppm)	1 (8 ppm)	1 (248 ppm)	1 (196 ppm)
GH28	Glyco_hydro_28	polygalacturonase, rhamnogalacturonase, others	0	1 (47 ppm)	1 (6 ppm)	0	1 (225 ppm)
GH29	Alpha_L_fucos	α-L-fucosidase	0	4 (230 ppm)	5 (25 ppm)	0	0
GH30	Glyco_hydro_30	β-1,6-glucanase, β-xylosidase	1 (67 ppm)	1 (65 ppm)	0	0	2 (578 ppm)
GH31	Glyco_hydro_31	α-glucosidase, α-xylosidase, others	9 (623 ppm)	14 (688 ppm)	11 (54 ppm)	0	1 (278 ppm)
GH32	Glyco_hydro_32N	invertase, others	11 (536 ppm)	17 (635 ppm)	1 (4 ppm)	0	0
GH33	BNR	sialidase, neuraminidase	0	0	0	0	0
GH34	Neur	sialidase, neuraminidase (viral only)	0	0	0	0	0
GH35	Glyco_hydro_35	β-galactosidase	2 (118 ppm)	1 (49 ppm)	3 (13 ppm)	0	0
GH37	Trehalase	trehalase	0	0	2 (9 ppm)	0	0
GH38	Glyco_hydro_38	α-mannosidase	2 (77 ppm)	2 (60 ppm)	6 (26 ppm)	0	0
GH39	Glyco_hydro_39	β-xylosidase, α-L-iduronidase	0	0	1 (7 ppm)	0	0
GH42	Glyco_hydro_42	β-galactosidase	6 (421 ppm)	6 (282 ppm)	3 (16 ppm)	0	0
GH43	Glyco_hydro_43 #	xylanase, β-xylosidase, α-L-arabinofuranosidase, arabinanase, others	4 (229 ppm)	4 (165 ppm)	5 (27 ppm)	13 (2795 ppm)	13 (2340 ppm)
GH45	Glyco_hydro_45	endoglucanase (mainly eukaryotic, 2 bacterial)	0	0	0	4 (741 ppm)	0
GH46	Glyco_hydro_46	chitosanase	0	0	0	0	0
GH47	Glyco_hydro_47	α-mannosidase (mainly eukaryotic, 1 bacterial)	0	0	1 (5 ppm)	0	0
GH48	Glyco_hydro_48	endoglucanase, cellobiohydrolase	0	0	0	0	0
GH49	Glyco_hydro_49	dextranase, isopullulanase, others	0	0	0	0	0
GH52	Glyco_hydro_52	β-xylosidase	0	0	0	0	0
GH53	Glyco_hydro_53 #	β-1,4-endogalactanase	1 (43 ppm)	7 (247 ppm)	4 (16 ppm)	2 (470 ppm)	3 (596 ppm)
GH54	ArabFuran-catal #	β-xylosidase, α-L-arabinofuranosidase (mainly eukaryotic, 2 bacterial)	0	0	2 (6 ppm)	1 (242 ppm)	0

CAZy Family	Pfam HMM Name	Known Activities (CAZy)	Human Gut Community		Soil (Diversa Silage) ¶	Fibrobacter succinogenes ¶	Saccharophagus degradans ¶
			subject 7 ¶	subject 8 ¶			
JGI ID			2004002000	2004002001	2001200001	2001800000	635640000
Nucleotide sequence length			15817685	20486813	152406385	4024184	5057531
Protein coding genes			20523	25983	184374	3341	4017
Glycoside hydrolase catalytic domains							
GH56	Glyco_hydro_56	hyaluronidase (eukaryotic only)	0	0	0	0	0
GH57	Glyco_hydro_57	α-amylase, 4-α-glucanotransferase, α-galactosidase, others	0	0	3 (18 ppm)	3 (753 ppm)	0
GH59	Glyco_hydro_59	galactocerebrosidase (mainly eukaryotic, 2 bacterial)	0	0	0	0	0
GH61	Glyco_hydro_61	endoglucanase (fungal only)	0	0	0	0	0
GH62	Glyco_hydro_62	α-L-arabinofuranosidase	0	0	1 (5 ppm)	0	0
GH63	Glyco_hydro_63	processing α-glucosidase	0	0	0	0	0
GH65	Glyco_hydro_65m	trehalase, maltose phosphorylase, trehalose phosphorylase	0	1 (32 ppm)	5 (24 ppm)	0	0
GH67	Glyco_hydro_67M	α-glucuronidase, others	0	1 (34 ppm)	2 (10 ppm)	0	1 (192 ppm)
GH68	Glyco_hydro_68	levansucrase, others	0	0	0	1 (324 ppm)	0
GH70	Glyco_hydro_70	dextransucrase, others	0	0	0	0	0
GH71	Glyco_hydro_71	α-1,3-glucanase (mainly fungal, 1 bacterial)	0	0	0	0	0
GH72	Glyco_hydro_72	β-1,3-glucanosyltransglycosylase (fungal only)	0	0	0	0	0
GH73	Glucosaminidase	endo- β-N-acetylglucosaminidase, others	0	2 (41 ppm)	3 (8 ppm)	0	2 (166 ppm)
GH75	Chitosanase	chitosanase (mainly fungal)	0	0	0	0	0
GH76	Glyco_hydro_76	α-1,6-mannanase	0	0	0	0	0
GH77	Glyco_hydro_77	4-α-glucanotransferase, amylomaltase	10 (664 ppm)	11 (608 ppm)	11 (68 ppm)	1 (471 ppm)	1 (317 ppm)
GH78	Bac_rhamnosid	αL-rhamnosidase	0	4 (276 ppm)	0	0	0
GH79	Glyco_hydro_79n	endo- β-glucuronidase	0	0	0	0	1 (190 ppm)
GH81	Glyco_hydro_81	β-1,3-glucanase (mainly eukaryotic, 6 bacterial)	0	0	0	0	1 (406 ppm)
GH83	HN	hemaglutinin-neuraminidase (viral only)	0	0	0	0	0
GH84	Hyaluronidase_2 #	hyaluronidase, N-acetyl b-glucosaminidase	0	3 (107 ppm)	1 (4 ppm)	0	0
GH85	Glyco_hydro_85	endo- β-N-acetylglucosaminidase	1 (56 ppm)	0	0	0	0
GH88	Glyco_hydro_88	d-4,5 unsaturated β-glucuronyl hydrolase	4 (234 ppm)	0	5 (25 ppm)	0	4 (836 ppm)
GH89	NAGLU	αN-acetylglucosaminidase	0	1 (65 ppm)	0	0	0
GH92	Glyco_hydro_92 #	α-1,2-mannosidase	0	0	9 (41 ppm)	0	0
GH98	Glyco_hydro_98M #	endo- β-galactosidase	0	0	0	0	0
GH100	Invertase_neut	alkaline and neutral invertase	0	0	0	0	0
GH102	MltA	peptidoglycan lytic transglycosylase	0	0	3 (9 ppm)	0	0
GH103	TIGR: MltB #	peptidoglycan lytic transglycosylase	0	0	19 (80 ppm)	0	1 (173 ppm)
GH108	DUF847 #	N-acetylmuramidase	0	0	2 (6 ppm)	0	1 (98 ppm)
Carbohydrate-binding domains							
CBM1	CBM_1	cellulose-binding domain (mainly fungal, 1 viral)	0	0	0	0	0
CBM2	CBM_2	cellulose-binding domain	0	2 (25 ppm)	3 (6 ppm)	0	19 (1070 ppm)
CBM3	CBM_3	cellulose-binding domain	0	0	1 (2 ppm)	0	0
CBM4	CBM_4_9 &	amorphous cellulose-, xylan- and glucan-binding domain	0	4 (86 ppm)	0	4 (434 ppm)	5 (412 ppm)
CBM5	CBM_5_12 ¥	cellulose-binding domain	0	0	0	0	2 (52 ppm)
CBM6	CBM_6	amorphous cellulose- and xylan-binding domain	0	1 (16 ppm)	2 (5 ppm)	32 (2895 ppm)	45 (3238 ppm)
CBM10	CBM_10	cellulose-binding domain (aerobic bacteria) and dockerin (anaerobic fungi)	0	0	0	0	4 (72 ppm)
CBM11	CBM_11 #	glucan-binding domain	0	0	0	2 (277 ppm)	0
CBM13	Ricin_B_lectin	mannose- and xylan-binding domain	1 (26 ppm)	5 (98 ppm)	4 (10 ppm)	0	9 (720 ppm)
CBM14	CBM_14	chitin-binding domain	0	0	0	0	0
CBM15	CBM_15	xylan-binding domain (2 bacterial)	0	0	0	0	0
CBM16	CBM_4_9 &	carbohydrate-binding module 16					
CBM17	CBM_17_28 €	amorphous cellulose- and cellooligosaccharide-binding domain	0	0	0	0	0
CBM18	Chitin_bind_1	chitin-binding domain (eukaryotic only)	0	0	0	0	0
CBM19	CBM_19	chitin-binding domain (eukaryotic only)	0	0	0	0	0
CBM20	CBM_20	starch-binding domain	0	0	1 (2 ppm)	0	2 (113 ppm)
CBM21	CBM_21	starch-binding domain (mainly eukaryotic, 1 bacterial)	0	0	0	0	0
CBM25	CBM_25	starch-binding domain	0	0	0	0	0
CBM27	CBM27 #	mannan-binding domain	0	0	0	0	0
CBM32	F5_F8_type_C	galactose- and lactose-binding domain	10 (243 ppm)	23 (447 ppm)	12 (30 ppm)	1 (99 ppm)	26 (1835 ppm)
CBM33	Chitin_bind_3	chitin-binding domain	0	0	0	0	1 (126 ppm)
CBM34	Alpha_amylase_N	starch-binding domain	3 (61 ppm)	4 (67 ppm)	0	0	0
CBM35	BLAST search §	mannan- xylan- and β-galactan-binding domain					
CBM40	Sialidase	sialic acid-binding domain	0	1 (27 ppm)	0	0	0
CBM41	PUD#	α-glucan-, amylose-, amylopectin-, and pullulan-binding domain	0	0	1 (2 ppm)	0	1 (63 ppm)
CBM42	AbfB#	arabinose-binding domain	0	0	1 (3 ppm)	0	0
CBM43	x8 #	β-1,3-glucan binding domain (eukaryotic only)	0	0	0	0	0

CAZy Family	Pfam HMM Name	Known Activities (CAZy)	Human Gut Community		Soil (Diversa Silage)¶	Fibrobacter succinogenes¶	Saccharophagus degradans¶
			subject 7¶	subject 8¶			
JGI ID			**2004002000**	**2004002001**	**2001200001**	**2001800000**	**635640000**
Nucleotide sequence length			**15817685**	**20486813**	**152406385**	**4024184**	**5057531**
Protein coding genes			**20523**	**25983**	**184374**	**3341**	**4017**
Other domains oftentimes associated with GH catalytic domains							
	Alpha-amylase_C	α-amylase, C-terminal all- β domain; associated with GH13	1 (15 ppm)	1 (10 ppm)	0	0	1 (52 ppm)
	Alpha-L-AF_C	α-L-arabinofuranosidase, C-terminal domain; associated with GH51	5 (162 ppm)	7 (207 ppm)	7 (22 ppm)	0	1 (116 ppm)
	Alpha-mann_mid	middle domain; associated with GH38	2 (29 ppm)	2 (18 ppm)	5 (8 ppm)	0	0
	Big_1	Bacterial Ig-like domain (group 1)	0	0	1 (2 ppm)	0	0
	Big_2	Bacterial Ig-like domain (group 2)	22 (330 ppm)	41 (472 ppm)	29 (47 ppm)	0	1 (49 ppm)
	Big_3	Bacterial Ig-like domain (group 3)	4 (54 ppm)	6 (65 ppm)	0	0	0
	Big_4	Bacterial Ig-like domain (group 4)	2 (24 ppm)	3 (27 ppm)	0	0	0
	Bgal_small_N	β-galactosidase, small chain; associated with GH2	9 (374 ppm)	7 (270 ppm)	0	1 (203 ppm)	0
	CBM_X	associated with GH94	3 (35 ppm)	5 (46 ppm)	18 (23 ppm)	1 (47 ppm)	1 (38 ppm)
	CelD_N	N-terminal Ig-like domain of cellulase; associated with GH9	0	0	2 (4 ppm)	4 (264 ppm)	3 (160 ppm)
	CHB_HEX	Putative carbohydrate binding domain; associated with GH20	0	0	1 (3 ppm)	0	1 (107 ppm)
	CHB_HEX_C	Chitobiase/beta-hexosaminidase C-terminal domain; associated with GH20	0	0	0	0	1 (38 ppm)
	ChiC	Chitinase C; associated with GH18	0	0	0	0	1 (107 ppm)
	ChitinaseA_N	Chitinase A, N-terminal domain; associated with GH18	0	0	0	0	2 (129 ppm)
	Cohesin	Cohesin domain	0	0	2 (5 ppm)	0	0
	Dockerin_1	Dockerin type I repeat	2 (8 ppm)	5 (15 ppm)	0	0	0
	fn3	Fibronectin type III domain	9 (136 ppm)	13 (153 ppm)	27 (44 ppm)	0	5 (235 ppm)
	GDE_C	Amylo-alpha-1,6-glucosidase	0	0	8 (44 ppm)	0	0
	Glucodextran_B	Glucodextranase, domain B; associated with GH15	0	0	0	0	0
	Glucodextran_N	Glucodextranase, domain N; associated with GH15	0	0	2 (9 ppm)	0	1 (170 ppm)
	Glyco_hydro_2_N	sugar-binding domain	12 (338 ppm)	16 (320 ppm)	18 (53 ppm)	2 (239 ppm)	8 (784 ppm)
	Glyco_hydro_2	immunoglobulin-like β-sandwich domain	8 (154 ppm)	8 (120 ppm)	5 (10 ppm)	1 (86 ppm)	5 (303 ppm)
	Glyco_hydro_3_C	C-terminal domain (glycan-binding?)	14 (671 ppm)	20 (655 ppm)	28 (104 ppm)	1 (157 ppm)	5 (817 ppm)
	Glyco_hydro_20b	domain 2	0	0	0	0	1 (79 ppm)
	Glyco_hydro_32C	C-terminal domain	2 (33 ppm)	1 (12 ppm)	1 (2 ppm)	0	0
	Glyco_hydro_38C	C-terminal domain	1 (46 ppm)	0	3 (21 ppm)	0	0
	Glyco_hydro_42M	trimerisation domain	5 (198 ppm)	5 (134 ppm)	3 (12 ppm)	0	0
	Glyco_hydro_42C	C-terminal domain	2 (22 ppm)	0	3 (3 ppm)	0	0
	Glyco_hydro_65N	N-terminal domain	0	0	2 (9 ppm)	0	0
	Glyco_hydro_65C	C-terminal domain	0	1 (8 ppm)	2 (2 ppm)	0	0
	Glyco_hydro_67N	N-terminal domain	0	1 (14 ppm)	0	0	1 (76 ppm)
	Glyco_hydro_67C	C-terminal domain	0	1 (19 ppm)	7 (22 ppm)	0	1 (139 ppm)
	Glyco_hydro_98C	C-terminal domain	0	0	0	0	0
	Glyco_transf_36	associated with GH94	3 (61 ppm)	5 (80 ppm)	21 (44 ppm)	0	2 (129 ppm)
	GT36_AF	associated with GH94	2 (34 ppm)	5 (66 ppm)	18 (30 ppm)	0	1 (53 ppm)
	He_PIG	Putative Ig-like domain	1 (8 ppm)	0	16 (14 ppm)	0	31 (874 ppm)
	Isoamylase_N	Isoamylase N-terminal domain; associated with GH13	6 (97 ppm)	10 (119 ppm)	25 (41 ppm)	1 (63 ppm)	3 (154 ppm)
	TIG	Ig-like domain	0	0	8 (13 ppm)	0	4 (209 ppm)
Footnotes							
¶	Pfam HMM hits are counted if their e-value is smaller than 1e-4. BLAST hits are counted if their e-value is smaller than 1e-6. The value in parenthesis is the ratio between the total nucleotide length of the hits for this family and the total nucleotide l						
#	CAZy does not provide weblinks to Pfam for this family, but this Pfam HMM recognizes the sequences of this family.						
§	No Pfam HMM is available for this family. BLAST searches were conducted with segments of selected sequences listed at CAZy corresponding to this family.						
$	The CAZy website links to Pfam PF00959.9 (Phage_lysozyme) for both, GH24 and GH104.						
&	The CAZy website links to Pfam PF02018.7 (CBM_4_9) for all, CBM4, CBM9, CBM16, and CBM22.						
¥	The CAZy website links to Pfam PF02839.4 (CBM_5_12) for both, CBM5 and CBM12.						
€	The CAZy website links to Pfam PF03424.3 (CBM_17_28) for both, CBM17 and CBM28.						

Table S6. Comparison of abundance of selected GH and CBM domains in different single organism genomes (continued).

CAZy Family	Pfam HMM Name	Known Activities (CAZy)	Cytophaga hutchinsonii ¶	Thermobifida fusca ¶	Caldicellulosiruptor saccharolyticus ¶	Clostridium thermocellum ¶
JGI ID			**400030000**	**620040000**	**625650000**	**400320000**
Nucleotide sequence length			**4433218**	**3642249**	**2788317**	**3783150**
Protein coding genes			**3592**	**3117**	**2602**	**3163**
Glycoside hydrolase catalytic domains						
GH1	Glyco_hydro_1	β-glucosidase, β-galactosidase, β-mannosidase, others	1 (300 ppm)	2 (756 ppm)	1 (486 ppm)	2 (688 ppm)
GH2	Glyco_hydro_2_C	β-galactosidase, β-mannosidase, others	0	0	3 (938 ppm)	1 (243 ppm)
GH3	Glyco_hydro_3	β-1,4-glucosidase, β-1,4-xylosidase, β-1,3-glucosidase, α-L-arabinofuranosidase, others	6 (928 ppm)	2 (384 ppm)	2 (482 ppm)	2 (363 ppm)
GH4	Glyco_hydro_4	α-glucosidase, α-galactosidase, α-glucuronidase, others	0	1 (356 ppm)	2 (893 ppm)	0
GH5	Cellulase	cellulase, β-1,4-endoglucanase, β-1,3-glucosidase, β-1,4-endoxylanase, β-1,4-endomannanase, others	4 (893 ppm)	3 (707 ppm)	5 (1484 ppm)	10 (2515 ppm)
GH6	Glyco_hydro_6	endoglucanase, cellobiohydrolase	0	2 (525 ppm)	0	0
GH7	Glyco_hydro_7	endoglucanase, cellobiohydrolase (fungal only)	0	0	0	0
GH8	Glyco_hydro_8	cellulase, β-1,3-glucosidase, β-1,4-endoxylanase, β-1,4-endomannanase, others	5 (1228 ppm)	0	0	1 (281 ppm)
GH9	Glyco_hydro_9	endoglucanase, cellobiohydrolase, β-glucosidase	6 (1870 ppm)	2 (749 ppm)	2 (919 ppm)	16 (5729 ppm)
GH10	Glyco_hydro_10	xylanase, β-1,3-endoxylanase	3 (646 ppm)	2 (528 ppm)	5 (1703 ppm)	5 (1262 ppm)
GH11	Glyco_hydro_11	xylanase	1 (127 ppm)	1 (147 ppm)	0	1 (148 ppm)
GH12	Glyco_hydro_12	endoglucanase, β-1,3-1,4-glucanase, xyloglucan hydrolase	0	0	0	0
GH13	Alpha-amylase	α-amylase, catalytic domain, and related enzymes	4 (1114 ppm)	6 (1824 ppm)	4 (1545 ppm)	2 (562 ppm)
GH14	Glyco_hydro_14	β-amylase	0	0	0	0
GH15	Glyco_hydro_15	glucoamylase, glucodextranase	1 (246 ppm)	2 (612 ppm)	1 (391 ppm)	1 (299 ppm)
GH16	Glyco_hydro_16	β-1,3(4)-endoglucanase, others	2 (267 ppm)	0	1 (152 ppm)	2 (324 ppm)
GH17	Glyco_hydro_17	glucan endo-1,3- β-D-glucosidase glucan 1,3- β-glucosidase, others	0	0	0	0
GH18	Glyco_hydro_18	chitinase, endo- β-N-acetylglucosaminidase, non-catalytic proteins	0	2 (638 ppm)	0	3 (746 ppm)
GH19	Glyco_hydro_19	chitinase	0	0	0	0
GH20	Glyco_hydro_20	β-hexosaminidase, lacto-N-biosidase	0	0	1 (340 ppm)	0
GH22	Lys	C-type lysozyme, α-lactalbumin (eukaryotic only)	0	0	0	0
GH23	SLT #	G-type lysozyme, peptidoglycan lytic transglycosylase	4 (317 ppm)	0	3 (393 ppm)	2 (182 ppm)
GH24	Phage_lysozyme $	lysozyme	0	0	0	0
GH25	Glyco_hydro_25	lysozyme	0	0	0	0
GH26	Glyco_hydro_26	β-1,3-xylanase, mannanase	1 (208 ppm)	0	1 (334 ppm)	2 (509 ppm)
GH27	Melibiase	α-galactosidase, α-N-acetylgalactosaminidase, isomalto-dextranase	0	0	1 (429 ppm)	0
GH28	Glyco_hydro_28	polygalacturonase, rhamnogalacturonase, others	0	0	2 (752 ppm)	0
GH29	Alpha_L_fucos	α-L-fucosidase	0	0	0	0
GH30	Glyco_hydro_30	β-1,6-glucanase, β-xylosidase	0	0	1 (476 ppm)	0
GH31	Glyco_hydro_31	α-glucosidase, α-xylosidase, others	1 (298 ppm)	1 (353 ppm)	1 (463 ppm)	0
GH32	Glyco_hydro_32N	invertase, others	0	0	0	0
GH33	BNR	sialidase, neuraminidase	0	0	0	0
GH34	Neur	sialidase, neuraminidase (viral only)	0	0	0	0
GH35	Glyco_hydro_35	β-galactosidase	0	0	0	0
GH37	Trehalase	trehalase	0	0	0	0
GH38	Glyco_hydro_38	α-mannosidase	0	0	0	0
GH39	Glyco_hydro_39	β-xylosidase, α-L-iduronidase	0	0	3 (1360 ppm)	0
GH42	Glyco_hydro_42	β-galactosidase	0	1 (338 ppm)	1 (459 ppm)	0
GH43	Glyco_hydro_43 #	xylanase, β-xylosidase, α-L-arabinofuranosidase, arabinanase, others	2 (369 ppm)	1 (241 ppm)	4 (1232 ppm)	4 (919 ppm)
GH45	Glyco_hydro_45	endoglucanase (mainly eukaryotic, 2 bacterial)	0	0	0	0
GH46	Glyco_hydro_46	chitosanase	0	0	0	0
GH47	Glyco_hydro_47	α-mannosidase (mainly eukaryotic, 1 bacterial)	0	0	0	0
GH48	Glyco_hydro_48	endoglucanase, cellobiohydrolase	0	1 (524 ppm)	1 (670 ppm)	2 (988 ppm)
GH49	Glyco_hydro_49	dextranase, isopullulanase, others	0	0	0	0
GH52	Glyco_hydro_52	β-xylosidase	0	0	0	0
GH53	Glyco_hydro_53 #	β-1,4-endogalactanase	0	0	0	1 (240 ppm)
GH54	ArabFuran-catal #	β-xylosidase, α-L-arabinofuranosidase (mainly eukaryotic, 2 bacterial)	0	0	0	0

CAZy Family	Pfam HMM Name	Known Activities (CAZy)	Cytophaga hutchinsonii ¶	Thermobifida fusca ¶	Caldicellulosiruptor saccharolyticus ¶	Clostridium thermocellum ¶
JGI ID			**400030000**	**620040000**	**625650000**	**400320000**
Nucleotide sequence length			**4433218**	**3642249**	**2788317**	**3783150**
Protein coding genes			**3592**	**3117**	**2602**	**3163**
Glycoside hydrolase catalytic domains						
GH56	Glyco_hydro_56	hyaluronidase (eukaryotic only)	0	0	0	0
GH57	Glyco_hydro_57	α-amylase, 4-α-glucanotransferase, α-galactosidase, others	1 (213 ppm)	0	0	0
GH59	Glyco_hydro_59	galactocerebrosidase (mainly eukaryotic, 2 bacterial)	0	0	0	0
GH61	Glyco_hydro_61	endoglucanase (fungal only)	0	0	0	0
GH62	Glyco_hydro_62	α-L-arabinofuranosidase	0	0	0	0
GH63	Glyco_hydro_63	processing α-glucosidase	0	0	0	0
GH65	Glyco_hydro_65m	trehalase, maltose phosphorylase, trehalose phosphorylase	0	1 (340 ppm)	2 (833 ppm)	0
GH67	Glyco_hydro_67M	α-glucuronidase, others	0	0	1 (356 ppm)	0
GH68	Glyco_hydro_68	levansucrase, others	0	0	0	0
GH70	Glyco_hydro_70	dextransucrase, others	0	0	0	0
GH71	Glyco_hydro_71	α-1,3-glucanase (mainly fungal, 1 bacterial)	0	0	0	0
GH72	Glyco_hydro_72	β-1,3-glucanosyltransglycosylase (fungal only)	0	0	0	0
GH73	Glucosaminidase	endo- β-N-acetylglucosaminidase, others	1 (97 ppm)	0	0	0
GH75	Chitosanase	chitosanase (mainly fungal)	0	0	0	0
GH76	Glyco_hydro_76	α-1,6-mannanase	0	0	0	0
GH77	Glyco_hydro_77	4-α-glucanotransferase, amylomaltase	1 (328 ppm)	1 (442 ppm)	0	0
GH78	Bac_rhamnosid	α-L-rhamnosidase	0	0	2 (867 ppm)	0
GH79	Glyco_hydro_79n	endo- β-glucuronidase	0	0	0	0
GH81	Glyco_hydro_81	β-1,3-glucanase (mainly eukaryotic, 6 bacterial)	0	1 (531 ppm)	0	1 (504 ppm)
GH83	HN	hemaglutinin-neuraminidase (viral only)	0	0	0	0
GH84	Hyaluronidase_2 #	hyaluronidase, N-acetyl b-glucosaminidase	0	0	0	0
GH85	Glyco_hydro_85	endo- β-N-acetylglucosaminidase	0	0	0	0
GH88	Glyco_hydro_88	d-4,5 unsaturated β-glucuronyl hydrolase	0	0	2 (755 ppm)	0
GH89	NAGLU	α-N-acetylglucosaminidase	0	0	0	0
GH92	Glyco_hydro_92 #	α-1,2-mannosidase	0	0	0	0
GH98	Glyco_hydro_98M #	endo- β-galactosidase	0	0	0	0
GH100	Invertase_neut	alkaline and neutral invertase	0	0	0	0
GH102	MltA	peptidoglycan lytic transglycosylase	0	0	0	0
GH103	TIGR: MltB #	peptidoglycan lytic transglycosylase	0	0	0	0
GH108	DUF847 #	N-acetylmuramidase	0	0	0	0
Carbohydrate-binding domains						
CBM1	CBM_1	cellulose-binding domain (mainly fungal, 1 viral)	0	0	0	0
CBM2	CBM_2	cellulose-binding domain	0	14 (1134 ppm)	0	0
CBM3	CBM_3	cellulose-binding domain	0	2 (133 ppm)	10 (898 ppm)	19 (1278 ppm)
CBM4	CBM_4_9 &	amorphous cellulose-, xylan- and glucan-binding domain	2 (208 ppm)	1 (115 ppm)	9 (1504 ppm)	12 (1458 ppm)
CBM5	CBM_5_12 ¥	cellulose-binding domain	0	0	0	0
CBM6	CBM_6	amorphous cellulose- and xylan-binding domain	2 (162 ppm)	1 (96 ppm)	1 (136 ppm)	17 (1653 ppm)
CBM10	CBM_10	cellulose-binding domain (aerobic bacteria) and dockerin (anaerobic fungi)	0	0	0	0
CBM11	CBM_11 #	glucan-binding domain	1 (129 ppm)	0	0	1 (139 ppm)
CBM13	Ricin_B_lectin	mannose- and xylan-binding domain	0	1 (112 ppm)	0	1 (108 ppm)
CBM14	CBM_14	chitin-binding domain	0	0	0	0
CBM15	CBM_15	xylan-binding domain (2 bacterial)	0	0	0	0
CBM17	CBM_17_28 €	amorphous cellulose- and cellooligosaccharide-binding domain	0	0	1 (238 ppm)	0
CBM18	Chitin_bind_1	chitin-binding domain (eukaryotic only)	0	0	0	0
CBM19	CBM_19	chitin-binding domain (eukaryotic only)	0	0	0	0
CBM20	CBM_20	starch-binding domain	0	1 (78 ppm)	1 (110 ppm)	0
CBM21	CBM_21	starch-binding domain (mainly eukaryotic, 1 bacterial)	0	0	0	0
CBM25	CBM_25	starch-binding domain	0	0	0	0
CBM27	CBM27 #	mannan-binding domain	0	0	0	0
CBM32	F5_F8_type_C	galactose- and lactose-binding domain	1 (83 ppm)	1 (108 ppm)	1 (125 ppm)	1 (92 ppm)
CBM33	Chitin_bind_3	chitin-binding domain	0	2 (334 ppm)	0	0
CBM34	Alpha_amylase_N	starch-binding domain	0	0	0	1 (90 ppm)
CBM40	Sialidase	sialic acid-binding domain	0	0	0	0
CBM41	PUD #	α-glucan-, amylose-, amylopectin-, and pullulan-binding domain	0	0	1 (113 ppm)	0
CBM42	AbfB #	arabinose-binding domain	0	0	0	4 (429 ppm)
CBM43	X8 #	β-1,3-glucan binding domain (eukaryotic only)	0	0	0	0

CAZy Family	Pfam HMM Name	Known Activities (CAZy)	Cytophaga hutchinsonii ¶	Thermobifida fusca ¶	Caldicellulosiruptor saccharolyticus ¶	Clostridium thermocellum ¶
JGI ID			**400030000**	**620040000**	**625650000**	**400320000**
Nucleotide sequence length			**4433218**	**3642249**	**2788317**	**3783150**
Protein coding genes			**3592**	**3117**	**2602**	**3163**
Other domains oftentimes associated with GH catalytic domains						
	Alpha-amylase_C	α-amylase, C-terminal all- β domain; associated with GH13	0	1 (67 ppm)	0	0
	Alpha-L-AF_C	α-L-arabinofuranosidase, C-terminal domain; associated with GH51	0	0	1 (219 ppm)	1 (163 ppm)
	Alpha-mann_mid	middle domain; associated with GH38	0	0	0	0
	Big_1	Bacterial Ig-like domain (group 1)	0	0	0	1 (84 ppm)
	Big_2	Bacterial Ig-like domain (group 2)	0	0	2 (169 ppm)	1 (63 ppm)
	Big_3	Bacterial Ig-like domain (group 3)	0	0	0	0
	Big_4	Bacterial Ig-like domain (group 4)	0	0	2 (130 ppm)	1 (44 ppm)
	Bgal_small_N	β-galactosidase, small chain; associated with GH2	0	0	2 (581 ppm)	0
	CBM_X	associated with GH94	0	0	2 (138 ppm)	3 (152 ppm)
	CelD_N	N-terminal Ig-like domain of cellulase; associated with GH9	4 (252 ppm)	1 (75 ppm)	0	4 (292 ppm)
	CHB_HEX	Putative carbohydrate binding domain; associated with GH20	0	0	0	0
	CHB_HEX_C	Chitobiase/beta-hexosaminidase C-terminal domain; associated with GH20	0	0	0	0
	ChiC	Chitinase C; associated with GH18	0	0	0	0
	ChitinaseA_N	Chitinase A, N-terminal domain; associated with GH18	0	0	0	0
	Cohesin	Cohesin domain	0	0	0	29 (3558 ppm)
	Dockerin_1	Dockerin type I repeat	0	0	0	138 (2297 ppm)
	fn3	Fibronectin type III domain	1 (53 ppm)	4 (275 ppm)	0	7 (440 ppm)
	GDE_C	Amylo-alpha-1,6-glucosidase	1 (248 ppm)	0	0	1 (290 ppm)
	Glucodextran_B	Glucodextranase, domain B; associated with GH15	0	0	0	0
	Glucodextran_N	Glucodextranase, domain N; associated with GH15	0	0	0	0
	Glyco_hydro_2_N	sugar-binding domain	0	1 (125 ppm)	5 (902 ppm)	1 (79 ppm)
	Glyco_hydro_2	immunoglobulin-like β-sandwich domain	0	0	3 (359 ppm)	1 (90 ppm)
	Glyco_hydro_3_C	C-terminal domain (glycan-binding?)	4 (658 ppm)	2 (397 ppm)	2 (510 ppm)	1 (170 ppm)
	Glyco_hydro_20b	domain 2	0	0	0	0
	Glyco_hydro_32C	C-terminal domain	0	0	0	0
	Glyco_hydro_38C	C-terminal domain	0	0	0	0
	Glyco_hydro_42M	trimerisation domain	0	1 (172 ppm)	1 (228 ppm)	0
	Glyco_hydro_42C	C-terminal domain	0	0	1 (62 ppm)	0
	Glyco_hydro_65N	N-terminal domain	0	1 (213 ppm)	2 (551 ppm)	0
	Glyco_hydro_65C	C-terminal domain	0	0	2 (111 ppm)	0
	Glyco_hydro_67N	N-terminal domain	0	0	1 (132 ppm)	0
	Glyco_hydro_67C	C-terminal domain	0	0	1 (242 ppm)	0
	Glyco_hydro_98C	C-terminal domain	0	0	0	0
	Glyco_transf_36	associated with GH94	0	0	2 (237 ppm)	4 (343 ppm)
	GT36_AF	associated with GH94	0	0	2 (186 ppm)	3 (213 ppm)
	He_PIG	Putative Ig-like domain	0	0	0	0
	Isoamylase_N	Isoamylase N-terminal domain; associated with GH13	1 (58 ppm)	2 (144 ppm)	2 (185 ppm)	1 (67 ppm)
	TIG	Ig-like domain	0	0	0	0
Footnotes						

¶	Pfam HMM hits are counted if their e-value is smaller than 1e-4. BLAST hits are counted if their e-value is smaller than 1e-6. The value in parenthesis is the ratio between the total nucleotide length of the hits for this family and the total nucleotide l
#	CAZy does not provide weblinks to Pfam for this family, but this Pfam HMM recognizes the sequences of this family.
§	No Pfam HMM is available for this family. BLAST searches were conducted with segments of selected sequences listed at CAZy corresponding to this family.
$	The CAZy website links to Pfam PF00959.9 (Phage_lysozyme) for both, GH24 and GH104.
&	The CAZy website links to Pfam PF02018.7 (CBM_4_9) for all, CBM4, CBM9, CBM16, and CBM22.
¥	The CAZy website links to Pfam PF02839.4 (CBM_5_12) for both, CBM5 and CBM12.
€	The CAZy website links to Pfam PF03424.3 (CBM_17_28) for both, CBM17 and CBM28.

Table S7. Phylogenetic characterization of the termite gut metagenome sequence dataset, based on compositional phylogenetic analysis [43]. Shown are clades containing ≥50 kb of sequence for fragments ≥1 kb.

Taxonomic group	# fragments	%[a]	# bp	%[a]
Bacteria	18,869	96	28,048,863	96
Bacteroidetes	189	1	287,830	1
Fibrobacteres	434	2	828,179	3
Firmicutes	640	3	963,754	3
Clostridia	39	<1	98,404	<1
Proteobacteria	1,614	8	2,579,469	9
Betaproteobacteria	197	1	287,933	1
Gammaproteobacteria	47	<1	92,769	<1
Deltaproteobacteria	43	<1	85,112	<1
Epsilonproteobacteria	111	1	125,930	<1
Spirochaetes	4,648	24	9,487,929	33
Spirochaetales				
Treponema	4,500	23	9,288,738	32
Archaea	584	3	762,065	3
Euryarchaeota	196	1	269,315	1
Other	251	1	311,181	1
Total	**19,720**		**29,164,892**	

[a] percentages are given at different taxonomic levels, hence add up to more than 100%; data in shaded rows was assigned to the sample-derived classes for the *Fibrobacteres* and *Treponema* clades of the model, all other assignments are to classes trained from publicly available data (generic model component).

Table S8. Counts of genes classified to COGs corresponding to different hydrogenase families.

Hydrogenase class	COG	Gene count in termite hindgut metagenome			
		total	*Fibrobacter*	*Treponema*	other
Ni-Fe hydrogen-uptake hydrogenase	COG0374 (Ni,Fe-hydrogenase I large subunit)	0	0	0	0
	COG1740 (Ni,Fe-hydrogenase I small subunit	0	0	0	0
Ni-Fe hydrogen-evolving hydrogenase	COG3261 (Ni,Fe-hydrogenase III large subunit)	1	0	0	1
	COG3260 (Ni,Fe-hydrogenase III small subunit)	1	0	0	1
Metal-free uptake hydrogenase	COG4074 (H2-forming N5,N10-methylenetetrahydromethanopterin dehydrogenase)	0	0	0	0
Iron-only hydrogenase (reversible)	COG4624 (Iron only hydrogenase large subunit, C-terminal domain)	124	0	24	100

Table S9. Fe-only hydrogenases (COG4624, large subunit, C-terminal domain) identified in the P3 luminal microbiota. Expressed genes identified in the metaproteomic analysis are highlighted in blue.

Family	Gene_OID	IMG/M annotation	Contig/read #
family 1	2004147267	Methyl-accepting chemotaxis protein	Contig41712
family 1	2004145065	Iron only hydrogenase large subunit, C-terminal domain	Contig41140
family 1	2004144833	Iron only hydrogenase large subunit, C-terminal domain	Contig41062
family 1	2004144069	Methyl-accepting chemotaxis protein	Contig40793
family 1	2004139220	Methyl-accepting chemotaxis protein	Contig38638
family 1	2004138928	Methyl-accepting chemotaxis protein	Contig38500
family 1	2004138343	Sigma-54 dependent transcriptional regulator, Fis family	Contig38190
family 1	2004135624	Iron only hydrogenase large subunit, C-terminal domain	Contig36773
family 1	2004135138	Iron only hydrogenase large subunit, C-terminal domain	Contig36510
family 1	2004133729	Methyl-accepting chemotaxis protein	Contig35731
family 1	2004131875	Methyl-accepting chemotaxis protein	Contig34710
family 1	2004130511	Methyl-accepting chemotaxis protein	Contig33880
family 1	2004130436	Ferredoxin 2	Contig33825
family 1	2004125995	Iron only hydrogenase large subunit, C-terminal domain	Contig31061
family 1	2004123920	Ferredoxin 2	Contig29740
family 1	2004123628	Methyl-accepting chemotaxis protein	Contig29555
family 1	2004120652	Iron only hydrogenase large subunit, C-terminal domain	Contig27679
family 1	2004119843	Iron only hydrogenase large subunit, C-terminal domain	Contig27164
family 1	2004119364	Iron only hydrogenase large subunit, C-terminal domain	Contig26877
family 1	2004116165	Iron only hydrogenase large subunit, C-terminal domain	Contig24838
family 1	2004115852	Methyl-accepting chemotaxis protein	Contig24630
family 1	2004115530	Iron only hydrogenase large subunit, C-terminal domain	Contig24431
family 1	2004115224	Iron only hydrogenase large subunit, C-terminal domain	Contig24239
family 1	2004113529	Sigma-54 dependent transcriptional regulator, Fis family	Contig23149
family 1	2004110612	Iron only hydrogenase large subunit, C-terminal domain	Contig21035
family 1	2004108337	Ferredoxin 2	Contig19382
family 1	2004107815	Iron only hydrogenase large subunit, C-terminal domain	Contig18989
family 1	2004106533	Sigma-54 dependent transcriptional regulator, Fis family	Contig18043
family 1	2004100524	Iron only hydrogenase large subunit, C-terminal domain	Contig13615
family 1	2004094811	Iron only hydrogenase large subunit, C-terminal domain	Contig9317
family 1	2004090341	Iron only hydrogenase large subunit, C-terminal domain	Contig5988
family 1	2004088633	Iron only hydrogenase large subunit, C-terminal domain	Contig4679
family 1	2004088033		Contig4225
family 1	2004086764	Iron only hydrogenase large subunit, C-terminal domain	Contig3281
family 1	2004086733	Iron only hydrogenase large subunit, C-terminal domain	Contig3259
family 1	2004084573	Iron only hydrogenase large subunit, C-terminal domain	Contig1675
family 1	2004083504	Iron only hydrogenase large subunit, C-terminal domain	Contig870
family 1	2004083482	Iron only hydrogenase large subunit, C-terminal domain	Contig856
family 2	2004083911	Uncharacterized anaerobic dehydrogenase	Contig1178
family 2	2004095107	Iron only hydrogenase large subunit, C-terminal domain	Contig9545
family 2	2004131651	Uncharacterized anaerobic dehydrogenase	Contig34592
family 2	2004145969	NADH dehydrogenase/NADH	Contig41400
family 3	2004160714	Iron only hydrogenase large subunit, C-terminal domain	BHZN39579_g2
family 3	2004148312	Iron only hydrogenase large subunit, C-terminal domain	BHZN3955_g1
family 3	2004146072	Iron only hydrogenase large subunit, C-terminal domain	Contig41428
family 3	2004145706	Iron only hydrogenase large subunit, C-terminal domain	Contig41331
family 3	2004144780	Iron only hydrogenase large subunit, C-terminal domain	Contig41045

Family	Gene_OID	IMG/M annotation	Contig/read #
family 3	2004142657	Iron only hydrogenase large subunit, C-terminal domain	Contig40246
family 3	2004142656	Iron only hydrogenase large subunit, C-terminal domain	Contig40246
family 3	2004141757	Iron only hydrogenase large subunit, C-terminal domain	Contig39862
family 3	2004139922	Iron only hydrogenase large subunit, C-terminal domain	Contig38965
family 3	2004138311	NADH dehydrogenase/NADH	Contig38171
family 3	2004136169	Iron only hydrogenase large subunit, C-terminal domain	Contig37066
family 3	2004133409	Iron only hydrogenase large subunit, C-terminal domain	Contig35567
family 3	2004133408	Iron only hydrogenase large subunit, C-terminal domain	Contig35567
family 3	2004132119	Iron only hydrogenase large subunit, C-terminal domain	Contig34839
family 3	2004128968	Iron only hydrogenase large subunit, C-terminal domain	Contig32917
family 3	2004128967	Iron only hydrogenase large subunit, C-terminal domain	Contig32917
family 3	2004126522	Uncharacterized anaerobic dehydrogenase	Contig31391
family 3	2004123602	Iron only hydrogenase large subunit, C-terminal domain	Contig29540
family 3	2004119568	Iron only hydrogenase large subunit, C-terminal domain	Contig27003
family 3	2004119452	Iron only hydrogenase large subunit, C-terminal domain	Contig26935
family 3	2004118909	Iron only hydrogenase large subunit, C-terminal domain	Contig26581
family 3	2004117939	Iron only hydrogenase large subunit, C-terminal domain	Contig25966
family 3	2004111139	Uncharacterized anaerobic dehydrogenase	Contig21412
family 3	2004110961	Iron only hydrogenase large subunit, C-terminal domain	Contig21282
family 3	2004107649	Iron only hydrogenase large subunit, C-terminal domain	Contig18868
family 3	2004106470	Iron only hydrogenase large subunit, C-terminal domain	Contig17999
family 3	2004105320	Iron only hydrogenase large subunit, C-terminal domain	Contig17144
family 3	2004101268	Iron only hydrogenase large subunit, C-terminal domain	Contig14158
family 3	2004100209	Iron only hydrogenase large subunit, C-terminal domain	Contig13393
family 3	2004098523	Iron only hydrogenase large subunit, C-terminal domain	Contig12149
family 3	2004097668	NADH dehydrogenase/NADH	Contig11484
family 3	2004096748	Iron only hydrogenase large subunit, C-terminal domain	Contig10773
family 3	2004093306	Iron only hydrogenase large subunit, C-terminal domain	Contig8200
family 3	2004091623	Iron only hydrogenase large subunit, C-terminal domain	Contig6946
family 3	2004091513	Iron only hydrogenase large subunit, C-terminal domain	Contig6870
family 3	2004090116	Iron only hydrogenase large subunit, C-terminal domain	Contig5808
family 3	2004087563	Iron only hydrogenase large subunit, C-terminal domain	Contig3857
family 3	2004086063	Iron only hydrogenase large subunit, C-terminal domain	Contig2785
family 3	2004085995	Iron only hydrogenase large subunit, C-terminal domain	Contig2738
family 3	2004083009	Iron only hydrogenase large subunit, C-terminal domain	Contig520
family 3	2004087291	Iron only hydrogenase large subunit, C-terminal domain	Contig3659
family 3	2004113334	Iron only hydrogenase large subunit, C-terminal domain	Contig23005
family 3	2004084376	Iron only hydrogenase large subunit, C-terminal domain	Contig1522
family 4	2004145520	Iron only hydrogenase large subunit, C-terminal domain	Contig41281
family 4	2004143447	Predicted Fe-S protein	Contig40571
family 4	2004143119	Iron only hydrogenase large subunit, C-terminal domain	Contig40438
family 4	2004142364	Predicted Fe-S protein	Contig40123
family 4	2004142003	Predicted Fe-S protein	Contig39974
family 4	2004141813	Iron only hydrogenase large subunit, C-terminal domain	Contig39888
family 4	2004140344	Iron only hydrogenase large subunit, C-terminal domain	Contig39175
family 4	2004137623	Iron only hydrogenase large subunit, C-terminal domain	Contig37820
family 4	2004136305	Iron only hydrogenase large subunit, C-terminal domain	Contig37142
family 4	2004132739		Contig35189
family 4	2004130999	Predicted Fe-S protein	Contig34175

Family	Gene_OID	IMG/M annotation	Contig/read #
family 4	2004122985	Predicted Fe-S protein	Contig29147
family 4	2004118032	Iron only hydrogenase large subunit, C-terminal domain	Contig26030
family 4	2004117531	Iron only hydrogenase large subunit, C-terminal domain	Contig25710
family 4	2004113736	Predicted Fe-S protein	Contig23289
family 4	2004112441	Iron only hydrogenase large subunit, C-terminal domain	Contig22346
family 4	2004110033	Iron only hydrogenase large subunit, C-terminal domain	Contig20612
family 4	2004107974	Iron only hydrogenase large subunit, C-terminal domain	Contig19110
family 4	2004106935	Predicted Fe-S protein	Contig18338
family 4	2004104986	Predicted Fe-S protein	Contig16892
family 4	2004094651	Iron only hydrogenase large subunit, C-terminal domain	Contig9194
family 4	2004092344	Iron only hydrogenase large subunit, C-terminal domain	Contig7486
family 4	2004085753	Predicted Fe-S protein	Contig2559
family 4	2004085275	Predicted Fe-S protein	Contig2200
family 4	2004084806	Iron only hydrogenase large subunit, C-terminal domain	Contig1848
family 4	2004086462	Predicted Fe-S protein	Contig3065
family 5	2004084868	Ferredoxin 2	Contig1895
family 5	2004103550		Contig15845
family 5	2004117823	Methyl-accepting chemotaxis protein	Contig25893
family 5	2004133476	Iron only hydrogenase large subunit, C-terminal domain	Contig35600
family 5	2004137292	Methyl-accepting chemotaxis protein	Contig37649
family 5	2004145010	Methyl-accepting chemotaxis protein	Contig41124
family 5	2004123386		Contig29398
family 6	2004088165	Iron only hydrogenase large subunit, C-terminal domain	Contig4324
family 6	2004088501	Fe-hydrogenase large subunit family protein	Contig4580
family 6	2004091626	Iron only hydrogenase large subunit, C-terminal domain	Contig6948
family 6	2004096774	Iron only hydrogenase large subunit, C-terminal domain	Contig10793
family 6	2004097531	Iron only hydrogenase large subunit, C-terminal domain	Contig11376
family 6	2004104141	Iron only hydrogenase large subunit, C-terminal domain	Contig16272
family 6	2004104158	Iron only hydrogenase large subunit, C-terminal domain	Contig16283
family 6	2004116450	Iron only hydrogenase large subunit, C-terminal domain	Contig25030
family 6	2004118259	Iron only hydrogenase large subunit, C-terminal domain	Contig26175
family 6	2004121757	Fe-hydrogenase large subunit family protein	Contig28364
family 6	2004123063	Iron only hydrogenase large subunit, C-terminal domain	Contig29195
family 6	2004131441	Fe-hydrogenase large subunit family protein	Contig34452
family 6	2004131442	Iron only hydrogenase large subunit, C-terminal domain	Contig34452
family 6	2004132679	Iron only hydrogenase large subunit, C-terminal domain	Contig35152
family 6	2004137321	Fe-hydrogenase large subunit family protein	Contig37664
family 6	2004145474	Iron only hydrogenase large subunit, C-terminal domain	Contig41267
family 6	2004141812		Contig39888
family 7	2004096510	Iron only hydrogenase large subunit, C-terminal domain	Contig10596
family 7	2004102798	Iron only hydrogenase large subunit, C-terminal domain	Contig15296
family 7	2004130014	Iron only hydrogenase large subunit, C-terminal domain	Contig33577
family 7	2004139317	NADH dehydrogenase/NADH	Contig38678
family 7	2004147498	Iron only hydrogenase large subunit, C-terminal domain	Contig41753
family 7	2004149219	Iron only hydrogenase large subunit, C-terminal domain	BHZN6343_g1
family 8	2004086463	Iron only hydrogenase large subunit, C-terminal domain	Contig3066
family 8	2004097310	Iron only hydrogenase large subunit, C-terminal domain	Contig11202
family 8	2004116251	Predicted Fe-S protein	Contig24896
family 8	2004125308	Iron only hydrogenase large subunit, C-terminal domain	Contig30621

Family	Gene_OID	IMG/M annotation	Contig/read #
family 8	2004127099	Predicted Fe-S protein	Contig31758
family 8	2004147499	Iron only hydrogenase large subunit, C-terminal domain	Contig41753
family 10	2004088865	Ferredoxin 2	Contig4854
family 10	2004124058	Iron only hydrogenase large subunit, C-terminal domain	Contig29832
family 10	2004128298	Iron only hydrogenase large subunit, C-terminal domain	Contig32507
family 10	2004145200	Methyl-accepting chemotaxis protein	Contig41182

Table S10. Gene clusters overrepresented in termite P3 luminal microbiota versus soil, ocean and human gut metagenome datasets.

COG/NOG/TOG	Description	Gene_OIDs of representative	FC[a]	p-value	Termite P3 lumen[b]	Termite metaproteome analysis[c]	Sargasso Sea[b]	Soil[b]	Human Colon[b]
NOG14630	Hypothetical repeat containing protein	2004131930		1.3E-04	82 (1.00)		0 (0.00)	0 (0.00)	0 (0.00)
COG2373	Large extracellular alpha-helical protein	2004146652	R	2.2E-04	201 (0.83)		28 (0.01)	83 (0.16)	0 (0.00)
NOG18930	Fibronectin type III domain protein	2004133748		4.0E-04	79 (0.87)		44 (0.04)	9 (0.05)	2 (0.04)
COG1262	Uncharacterized conserved protein	2004132642	S	4.4E-04	205 (0.75)		247 (0.08)	106 (0.18)	0 (0.00)
COG2604	Uncharacterized protein conserved in bacteria	2004142827	S	4.5E-04	51 (0.98)		5 (0.01)	1 (0.01)	0 (0.00)
NOG06353	Integrase domain protein	2004141667		4.9E-04	58 (0.95)		1 (0.00)	6 (0.04)	0 (0.00)
COG4624	Iron only hydrogenase large subunit, C-terminal domain	2004145706	R	5.3E-04	145 (0.69)		211 (0.08)	11 (0.02)	24 (0.20)
NOG45161	Alpha amylase family protein	2004146628		5.5E-04	60 (0.95)		0 (0.00)	7 (0.05)	0 (0.00)
COG4886	Leucine-rich repeat (LRR) protein	2004146824	S	7.2E-04	133 (0.65)		502 (0.20)	51 (0.11)	4 (0.04)
COG1543	Uncharacterized conserved protein	2004121747	S	8.3E-04	51 (0.89)		24 (0.04)	9 (0.07)	0 (0.00)
NOG18207	Hypothetical protein	2004145197		9.4E-04	53 (0.95)		0 (0.00)	6 (0.05)	0 (0.00)
COG3378	Predicted ATPase	2004143224	R	1.0E-03	98 (0.80)		48 (0.03)	11 (0.04)	9 (0.13)
COG1453	Predicted oxidoreductases of the aldo/keto reductase family	2004136416	R	1.0E-03	110 (0.78)		47 (0.03)	3 (0.01)	14 (0.18)
COG4804	Uncharacterized conserved protein	2004137258	S	1.3E-03	80 (0.76)		7 (0.01)	8 (0.03)	12 (0.20)
NOG44579	Conserved protein	2004146659		2.2E-03	100 (0.67)		1 (0.00)	0 (0.00)	28 (0.33)
COG1079	Uncharacterized ABC-type transport system, permease component	2004139432	R	2.4E-03	59 (0.65)		252 (0.21)	20 (0.12)	1 (0.02)
COG4603	ABC-type uncharacterized transport system, permease component	2004139431	R	2.4E-03	64 (0.65)		252 (0.23)	26 (0.10)	1 (0.02)
TOG1 (TIGR02145)	Hypothetical protein	2004082485			482 (0.98)	1 (2)	18 (0.00)	7 (0.01)	3 (0.01)
TOG2	Hypothetical protein	2004147571			145 (0.93)	13 (1833)	1 (0.00)	1 (0.00)	6 (0.07)
TOG3	Hypothetical protein	2004084370			121 (0.97)		40 (0.03)	0 (0.00)	0 (0.00)
TOG4 (TIGR0178)	Hypothetical protein	2004083157			103 (0.91)		0 (0.00)	0 (0.00)	6 (0.09)
TOG5 (TIGR0224)	Hypothetical protein	2004083056			97 (0.84)		11 (0.01)	6 (0.02)	8 (0.13)
TOG6	Hypothetical protein	2004083890			83 (1.00)	2 (2)	3 (0.00)	0 (0.00)	0 (0.00)
TOG7	Hypothetical protein	2004082393			78 (0.89)	1 (2)	21 (0.02)	5 (0.03)	3 (0.06)
TOG8 (PF01590)	Hypothetical protein	2004082587			72 (0.96)		2 (0.00)	6 (0.04)	0 (0.00)
TOG9	Hypothetical protein	2004082635			67 (0.97)	3 (24)	1 (0.00)	1 (0.01)	1 (0.03)
TOG10 (PF01627)	Hypothetical protein	2004082388			60 (0.99)		1 (0.00)	1 (0.01)	0 (0.00)
TOG11	Hypothetical protein	2004083498			60 (1.00)		1 (0.00)	0 (0.00)	0 (0.00)
TOG12	Hypothetical protein	2004084111			60 (0.81)		15 (0.02)	13 (0.08)	4 (0.10)
TOG13 (TIGR02722)	Hypothetical protein	2004088261			58 (0.95)		29 (0.04)	2 (0.01)	0 (0.00)
TOG14	Hypothetical protein	2004083452			56 (1.00)		0 (0.00)	0 (0.00)	0 (0.00)
TOG15	Hypothetical protein	2004084219			52 (0.96)		19 (0.03)	1 (0.01)	0 (0.00)
TOG16	Hypothetical protein	2004082452			51 (1.00)		0 (0.00)	0 (0.00)	0 (0.00)
TOG17	Hypothetical protein	2004085120			50 (0.97)		0 (0.00)	0 (0.00)	1 (0.03)

[a] FC – COG functional category, R= "general function prediction only", S="function unknown"; [b] Number of proteins in the dataset, relative representation in the specific metagenome is in parenthesis; [c] appearances of Termite metagenome orthologous group (TOG) peptides in the proteomic dataset: number of peptides identified (number of spectral counts)

Table S11. Operational taxonomic unit (OTU) representatives of 16S rRNA sequences obtained from the P3 luminal fluid of *Nasutitermes* spp.

	PCR library			Metagenome				Closest cultivated species / isolate		
OTU[1]	OTU rep.[2]	Acc. #[3]	%[4]		%[5]	Phylum	Phylog. group[6]	%[7]	Name	Acc. #[8]
1	290cost002-P3L-1850	EF454252	19			*Fibrobacteres*	TG3-1	79.6	*Thermanaeromonas toyohensis*	AB062280.1
2	290cost002-P3L-1281	EF453894	6	Contig41540	2	*Spirochaetes*	TTG-2	90.4	*Spirochaeta (Treponema) caldaria*	M71240.1
3	290cost002-P3L-1173	EF453818	3	Contig41291	2	*Spirochaetes*	TTG-1	89.9	*Spirochaeta (Treponema) caldaria*	M71240.1
4	290cost002-P3L-1616	EF454087	3	Contig38075	2	*Fibrobacteres*	TG3-1	80.2	*Thermanaeromonas toyohensis*	AB062280.1
5	290cost002-P3L-1831	EF454236	3			*Fibrobacteres*	TG3-1	79.7	*Thermanaeromonas toyohensis*	AB062280.1
6	290cost002-P3L-1330	EF453929	3	Contig38078	2	*Fibrobacteres*	TG3-1	79.8	*Thermanaeromonas toyohensis*	AB062280.1
7	290cost002-P3L-1580	EF454061	2			*Fibrobacteres*	TG3-1	79.8	*Thermanaeromonas toyohensis*	AB062280.1
8	290cost002-P3L-1788	EF454199	2			*Fibrobacteres*	TG3-1	80.0	*Thermanaeromonas toyohensis*	AB062280.1
9	290cost002-P3L-1581	EF454062	2			*Spirochaetes*	TTG-2	90.2	*Spirochaeta (Treponema) caldaria*	M71240.1
10	290cost002-P3L-2273	EF454524	2			*Fibrobacteres*	TG3-1	79.8	*Thermanaeromonas toyohensis*	AB062280.1
11	290cost002-P3L-1714	EF454151	2			*Fibrobacteres*	TG3-1	79.4	*Thermanaeromonas toyohensis*	AB062280.1
12	290cost002-P3L-1936	EF454311	2			*Fibrobacteres*	TG3-1	79.8	*Thermanaeromonas toyohensis*	AB062280.1
13	290cost002-P3L-1710	EF454148	2			*Fibrobacteres*	TG3-1	79.7	*Thermanaeromonas toyohensis*	AB062280.1
14	290cost002-P3L-1174	EF453819	1			*Fibrobacteres*	TG3-1	79.5	*Fibrobacter* isolate	L35547.1
15	290cost002-P3L-1474	EF454022	1	Contig41557	2	*Spirochaetes*	TTG-10	93.1	*Treponema* isolate strain SPIT5	AM182455.1
16	290cost002-P3L-1290	EF453902	1			*Spirochaetes*	TTG-1	90.2	*Spirochaeta (Treponema) caldaria*	M71240.1
17	290cost002-P3L-1362	EF453951	1			*Fibrobacteres*	TG3-1	79.4	*Thermanaeromonas toyohensis*	AB062280.1
18	290cost002-P3L-1792	EF454203	1			*Fibrobacteres*	TG3-1	79.6	*Fibrobacter* isolate	L35547.1
19	290cost002-P3L-1292	EF453904	1	Contig7288	2	*Spirochaetes*	TTG-10	91.7	*Treponema* isolate strain SPIT5	AM182455.1
20	290cost002-P3L-2618	EF454739	1			*Fibrobacteres*	TG3-1	79.4	*Thermanaeromonas toyohensis*	AB062280.1
21	290cost002-P3L-1195	EF453837	1			*Spirochaetes*	TTG-3	89.8	*Spirochaeta (Treponema) caldaria*	M71240.1
22	290cost002-P3L-2418	EF454620	1			*Fibrobacteres*	TG3-1	79.6	*Thermanaeromonas toyohensis*	AB062280.1
23	290cost002-P3L-1232	EF453863	1			*Spirochaetes*	TTG-10	92.0	*Treponema azotonutricium* ZAS-9	AF320287.1
24	290cost002-P3L-1484	EF454025	1			*Fibrobacteres*	TG3-1	79.9	*Thermanaeromonas toyohensis*	AB062280.1
25	290cost002-P3L-1542	EF454030	1			*Bacteroidetes*	*Bacteroidales*	84.9	*Alistipes finegoldii*	AJ518874.1
26	290cost002-P3L-2378	EF454595	1			*Proteobacteria*	*Deltaproteobacteria*	84.6	*Pelobacter acidigallici*	X77216.1
27	290cost002-P3L-1211	EF453849	1			*Fibrobacteres*	TG3-1	79.8	*Thermanaeromonas toyohensis*	AB062280.1
28	290cost002-P3L-1190	EF453832	1			*Spirochaetes*	TTG-2	90.2	*Spirochaeta (Treponema) caldaria*	M71240.1
29	290cost002-P3L-1297	EF453909	1	Contig41389	2	*Spirochaetes*	TTG-1	89.8	*Spirochaeta (Treponema) caldaria*	M71240.1
30	290cost002-P3L-1558	EF454042	1			*Fibrobacteres*	TG3-1	79.6	*Thermanaeromonas toyohensis*	AB062280.1
31	290cost002-P3L-1197	EF453839	1			*Acidobacteria*	*Holophagales*	86.9	*Holophaga foetida*	X77215.1
32	290cost002-P3L-1243	EF453870	1			*Spirochaetes*	TTG-1	90.1	*Spirochaeta (Treponema) caldaria*	M71240.1
33	290cost002-P3L-1249	EF453874	1			*Spirochaetes*	TTG-10	93.0	*Treponema* isolate strain SPIT5	AM182455.1

OTU[1]	PCR library OTU rep.[2]	Acc. #[3]	%[4]	Metagenome	%[5]	Phylum	Phylog. group[6]	Closest cultivated species / isolate %[7]	Name	Acc. #[8]
34	290cost002-P3L-1291	EF453903	1			*Fibrobacteres*	TG3-1	79.7	*Thermanaeromonas toyohensis*	AB062280.1
35	290cost002-P3L-1253	EF453876	1			*Bacteroidetes*	*Bacteroidales, Marinilabiliaceae*	85.2	*Ruminofilibacter xylanolyticum* str. S1	DQ141183.1
36	290cost002-P3L-2041	EF454386	1			*Proteobacteria*	*Deltaproteobacteria*	84.5	Deltaproteobacterium str. S2551	AF177428.1
37	290cost002-P3L-1750	EF454175	1			*Spirochaetes*	TSG-3	80.5	*"Spirochaeta bajacaliforniensis"*	M71239.1
38	290cost002-P3L-1777	EF454191	1	Contig40418	2	*Spirochaetes*	TTG-4	89.2	*Spirochaeta (Treponema) caldaria*	M71240.1
39	290cost002-P3L-2121	EF454429	1			*Fibrobacteres*	TG3-1	79.7	*Thermanaeromonas toyohensis*	AB062280.1
40	290cost002-P3L-1847	EF454249	1			*Fibrobacteres*	TG3-1	79.5	*Thermanaeromonas toyohensis*	AB062280.1
41	290cost002-P3L-1810	EF454220	1			*Spirochaetes*	TTG-1	90.2	*Spirochaeta (Treponema) caldaria*	M71240.1
42	290cost002-P3L-1194	EF453836	1	Contig39460	2	*Spirochaetes*	TTG-1	90.4	*Spirochaeta (Treponema) caldaria*	M71240.1
43	290cost002-P3L-1382	EF453966	1			*Fibrobacteres*	TG3-1	79.4	*Thermanaeromonas toyohensis*	AB062280.1
44	290cost002-P3L-1770	EF454186	1			*Fibrobacteres*	TG3-1	79.6	*Thermanaeromonas toyohensis*	AB062280.1
45	290cost002-P3L-1269	EF453887	<1	Contig13944	2	*Acidobacteria*	*Holophagales*	87.3	*Holophaga foetida*	X77215.1
46	290cost002-P3L-1573	EF454054	<1	Contig37655	2	*Proteobacteria*	*Deltaproteobacteria*	84.6	*Pelobacter acidigallici*	X77216.1
47	290cost002-P3L-1283	EF453896	<1			*Spirochaetes*	TTG-1	90.1	*Spirochaeta (Treponema) caldaria*	M71240.1
48	290cost002-P3L-1230	EF453861	<1			*Fibrobacteres*	TG3-2	76.7	*Brevibacillus* isolate WF146	AY950801.1
49	290cost002-P3L-1384	EF453967	<1			*Spirochaetes*	TTG-4	89.5	*Spirochaeta (Treponema) caldaria*	M71240.1
50	290cost002-P3L-1582	EF454063	<1	Contig24754	2	*Spirochaetes*	TTG-4	89.1	*Spirochaeta (Treponema) caldaria*	M71240.1
51	290cost002-P3L-1761	EF454182	<1			*Spirochaetes*	TTG-1	90.0	*Spirochaeta (Treponema) caldaria*	M71240.1
52	290cost002-P3L-1868	EF454266	<1			*Fibrobacteres*	TG3-1	79.7	*Thermanaeromonas toyohensis*	AB062280.1
53	290cost002-P3L-2001	EF454358	<1			*Fibrobacteres*	TG3-2	76.0	Butyrate-producing isolate L2-50	AJ270491.2
54	290cost002-P3L-1314	EF453919	<1			*ZB3*		79.5	*Clostridium* sp. str. PPf35E10	AY548785.1
55	290cost002-P3L-1628	EF454096	<1			*Bacteroidetes*	*Bacteroidales*	85.0	*Paludibacter propionicigenes*	AB078842.2
56	290cost002-P3L-1189	EF453831	<1	**Contig41631** [9]	2	*Fibrobacteres*	TFG-1	84.1	*Fibrobacter succinogenes*	M62685.1
57	290cost002-P3L-2170	EF454461	<1			*Fibrobacteres*	TFG-1	85.0	*Fibrobacter intestinalis*	M62686.1
58	290cost002-P3L-1613	EF454084	<1			*Spirochaetes*	TTG-10	92.1	*Treponema* isolate strain SPIT5	AM182455.1
59	290cost002-P3L-1300	EF453912	<1			*Spirochaetes*	TTG-10	92.1	*Treponema* isolate strain SPIT5	AM182455.1
60	290cost002-P3L-2057	EF454395	<1	Contig4242	5	*Fibrobacteres*	TG3-1	79.7	*Thermanaeromonas toyohensis*	AB062280.1
60				Contig3094		*Fibrobacteres*	TG3-1			
61	290cost002-P3L-2234	EF454503	<1			*Firmicutes*	*Lachnospiraceae*	86.9	*Clostridium* sp. str. XB90	AJ229234.1
62	290cost002-P3L-1869	EF454267	<1			*Spirochaetes*	TSG-2	81.6	*Treponema* isolate strain I:W:T040	AF182832.1
63	290cost002-P3L-1386	EF453968	<1			*Spirochaetes*	TTG-7	88.4	*Treponema primitia* ZAS-2	AF093252.1
64	290cost002-P3L-1393	EF453973	<1			*Spirochaetes*	TTG-1	90.0	*Spirochaeta (Treponema) caldaria*	M71240.1
65	290cost002-P3L-1376	EF453961	<1			*Fibrobacteres*	TG3-1	79.9	*Thermanaeromonas toyohensis*	AB062280.1
66	290cost002-P3L-1669	EF454124	<1			*Fibrobacteres*	TG3-1	79.4	*Thermanaeromonas toyohensis*	AB062280.1

OTU[1]	PCR library OTU rep.[2]	Acc. #[3]	%[4]	Metagenome	%[5]	Phylum	Phylog. group[6]	Closest cultivated species / isolate %[7]	Name	Acc. #[8]
67	290cost002-P3L-1178	EF453822	<1	Contig39436	2	*Fibrobacteres*	TFG-1	84.1	*Fibrobacter intestinalis*	M62687.1
68	290cost002-P3L-1618	EF454088	<1			*Spirochaetes*	TSG-2	81.7	*"Spirochaeta bajacaliforniensis"*	M71239.1
69	290cost002-P3L-1445	EF454015	<1			*Spirochaetes*	TTG-1	90.2	*Spirochaeta (Treponema) caldaria*	M71240.1
70	290cost002-P3L-556	EF454868	<1			*Spirochaetes*	TTG-1	90.3	*Spirochaeta (Treponema) caldaria*	M71240.1
71	290cost002-P3L-1377	EF453962	<1			*Spirochaetes*	TTG-10	91.6	*Treponema* isolate strain SPIT5	AM182455.1
72	290cost002-P3L-2083	EF454411	<1			*Spirochaetes*	TTG-8	91.5	*Treponema primitia* ZAS-1	AF093251.1
73	290cost002-P3L-1904	EF454293	<1			*Spirochaetes*	TTG-10	92.1	*Treponema* isolate strain SPIT5	AM182455.1
74	290cost002-P3L-1623	EF454093	<1	Contig18672	2	*Spirochaetes*	TTG-10	90.9	*Treponema primitia* ZAS-2	AF093252.1
75	290cost002-P3L-1405	EF453983	<1			*Spirochaetes*	TTG-10	92.5	*Treponema* isolate strain SPIT5	AM182455.1
76	290cost002-P3L-750	EF454953	<1			*ZB3*		79.6	*Clostridium* sp. str. PPf35E10	AY548785.1
77	290cost002-P3L-1947	EF454317	<1			*Bacteroidetes*	*Bacteroidales*	84.0	*Paludibacter propionicigenes*	AB078842.2
78	290cost002-P3L-2684	EF454785	<1			*Bacteroidetes*	*Bacteroidales, Marinilabiliaceae*	85.1	*Ruminofilibacter xylanolyticum* str. S1	DQ141183.1
79	290cost002-P3L-2433	EF454628	<1	Contig40689	2	*Fibrobacteres*	TFG-1	84.6	*Fibrobacter intestinalis*	M62686.1
80	290cost002-P3L-1576	EF454057	<1	Contig41384	2	*Fibrobacteres*	TFG-1	84.2	*Fibrobacter intestinalis*	M62687.1
81	290cost002-P3L-1210	EF453848	<1			*Spirochaetes*	TTG-4	89.4	*Spirochaeta (Treponema) caldaria*	M71240.1
82	290cost002-P3L-1325	EF453927	<1			*Spirochaetes*	TTG-1	89.8	*Spirochaeta (Treponema) caldaria*	M71240.1
83	290cost002-P3L-1395	EF453975	<1			*Spirochaetes*	TTG-1	89.9	*Spirochaeta (Treponema) caldaria*	M71240.1
84	290cost002-P3L-1342	EF453938	<1			*Spirochaetes*	TTG-5	90.5	*Treponema azotonutricium* ZAS-9	AF320287.1
85	290cost002-P3L-2660	EF454768	<1			*Spirochaetes*	TTG-1	90.3	*Spirochaeta (Treponema) caldaria*	M71240.1
86	290cost002-P3L-1607	EF454080	<1			*Spirochaetes*	TTG-1	90.0	*Spirochaeta (Treponema) caldaria*	M71240.1
87	290cost002-P3L-1204	EF453844	<1			*Spirochaetes*	TTG-10	92.5	*Treponema azotonutricium* ZAS-9	AF320287.1
88	290cost002-P3L-1653	EF454112	<1			*Spirochaetes*	TTG-10	91.8	*Treponema* isolate strain SPIT5	AM182455.1
89	290cost002-P3L-1331	EF453930	<1			*Spirochaetes*	TTG-10	91.7	*Treponema* isolate strain SPIT5	AM182455.1
90	290cost002-P3L-1340	EF453937	<1			*Spirochaetes*	TTG-10	92.7	*Treponema* isolate strain SPIT5	AM182455.1
91	290cost002-P3L-1895	EF454285	<1			*Spirochaetes*	TTG-10	92.6	*Treponema* isolate strain SPIT5	AM182455.1
92	290cost002-P3L-1360	EF453949	<1			*Fibrobacteres*	TG3-1	79.8	*Thermanaeromonas toyohensis*	AB062280.1
93	290cost002-P3L-1606	EF454079	<1			*Fibrobacteres*	TG3-1	79.5	*Thermanaeromonas toyohensis*	AB062280.1
94	290cost002-P3L-2652	EF454762	<1			*Fibrobacteres*	TG3-2	76.4	*Brevibacillus* isolate WF146	AY950801.1
95	290cost002-P3L-1153	EF453801	<1			*Acidobacteria*	*Holophagales*	86.6	*Holophaga foetida*	X77215.1
96	290cost002-P3L-2521	EF454677	<1			*Acidobacteria*		82.3	*Acidobacteria* str. Ellin345	CP000360.1
97	290cost002-P3L-2379	EF454596	<1			*Acidobacteria*		82.5	*Acidobacteriaceae* str. Gsoil 149	AB245339.1
98	290cost002-P3L-858	EF454982	<1			*Acidobacteria*		82.3	*Acidobacteriaceae* str. Gsoil 149	AB245339.1
99	290cost002-P3L-603	EF454889	<1			*Bacteroidetes*	*Bacteroidales*	83.3	*Alistipes finegoldii*	AJ518874.1
100	290cost002-P3L-2320	EF454557	<1			*Bacteroidetes*	*Bacteroidales*	85.4	*Paludibacter propionicigenes*	AB078842.2

OTU[1]	PCR library OTU rep.[2]	Acc. #[3]	%[4]	Metagenome	%[5]	Phylum	Phylog. group[6]	Closest cultivated species / isolate %[7]	Name	Acc. #[8]
101	290cost002-P3L-1182	EF453826	<1			*Fibrobacteres*	TFG-1	84.4	*Fibrobacter intestinalis*	M62686.1
102	290cost002-P3L-2166	EF454459	<1			*Fibrobacteres*	TFG-1	84.5	*Fibrobacter intestinalis*	M62686.1
103	290cost002-P3L-2390	EF454604	<1			*Fibrobacteres*	TFG-1	84.1	*Fibrobacter intestinalis*	M62687.1
104	290cost002-P3L-969	EF455006	<1	**Contig39720**[10]	10	*Fibrobacteres*	TFG-1	84.5	*Fibrobacter intestinalis*	M62687.1
104				Contig40980		*Fibrobacteres*	TFG-1			
104				Contig16110		*Fibrobacteres*	TFG-1			
104				Contig20300		*Fibrobacteres*	TFG-1			
105	290cost002-P3L-1224	EF453857	<1			*Fibrobacteres*	TFG-1	84.0	*Fibrobacter succinogenes*	M62685.1
106	290cost002-P3L-1410	EF453987	<1			*Proteobacteria*	*Deltaproteobacteria, Desulfovibrionales, Desulfovibrionaceae*	92.1	*Desulfovibrio* sp. str. LNB1	AY554145.1
107	290cost002-P3L-2069	EF454402	<1			*Proteobacteria*	*Gammaproteobacteria, Pseudomonadaceae*	100.0	*Pseudomonas* sp. str. DVS13A	AY864637.1
108	290cost002-P3L-1223	EF453856	<1			*Spirochaetes*	TSG-2	81.3	*Treponema* isolate strain I:W:T040	AF182832.1
109	290cost002-P3L-1614	EF454085	<1			*Spirochaetes*	TTG-2	89.8	*Spirochaeta (Treponema) caldaria*	M71240.1
110	290cost002-P3L-1603	EF454077	<1			*Spirochaetes*	TTG-1	90.2	*Spirochaeta (Treponema) caldaria*	M71240.1
111	290cost002-P3L-1681	EF454129	<1			*Spirochaetes*	TTG-1	89.6	*Spirochaeta (Treponema) caldaria*	M71240.1
112	290cost002-P3L-1733	EF454163	<1			*Spirochaetes*	TTG-4	89.2	*Treponema* isolate strain SPIT5	AM182455.1
113	290cost002-P3L-1753	EF454177	<1			*Spirochaetes*	TTG-1	90.2	*Spirochaeta (Treponema) caldaria*	M71240.1
114	290cost002-P3L-766	EF454959	<1			*Spirochaetes*	TTG-7	88.8	*Spirochaeta (Treponema) caldaria*	M71240.1
115	290cost002-P3L-1791	EF454202	<1			*Spirochaetes*	TTG-5	90.7	*Treponema azotonutricium* ZAS-9	AF320287.1
116	290cost002-P3L-443	EF454814	<1			*Spirochaetes*	TTG-6	90.5	*Treponema* isolate strain SPIT5	AM182455.1
117	290cost002-P3L-1207	EF453845	<1			*Spirochaetes*	TTG-1	90.4	*Spirochaeta (Treponema) caldaria*	M71240.1
118	290cost002-P3L-1272	EF453888	<1			*Spirochaetes*	TTG-10	91.9	*Treponema* isolate strain SPIT5	AM182455.1
119	290cost002-P3L-1458	EF454019	<1			*Spirochaetes*	TTG-10	92.1	*Treponema* isolate strain SPIT5	AM182455.1
120	290cost002-P3L-1586	EF454067	<1	Contig40676	2	*Spirochaetes*	TTG-10	92.0	*Treponema* isolate strain SPIT5	AM182455.1
121	290cost002-P3L-558	EF454869	<1			*Spirochaetes*	TTG-10	92.4	*Treponema* isolate strain SPIT5	AM182455.1
122	290cost002-P3L-1546	EF454034	<1			*Spirochaetes*	TTG-10	92.0	*Treponema* isolate strain SPIT5	AM182455.1
123	290cost002-P3L-1483	EF454024	<1			*Spirochaetes*	TTG-8	92.0	*Treponema* isolate strain SPIT5	AM182455.1
124	290cost002-P3L-636	EF454904	<1			*Spirochaetes*	TTG-10	92.2	*Treponema primitia* ZAS-2	AF093252.1
125	290cost002-P3L-2145	EF454444	<1			*Spirochaetes*	TTG-10	93.1	*Treponema* isolate strain SPIT5	AM182455.1
126	290cost002-P3L-1803	EF454213	<1			*Spirochaetes*	TTG-10	91.6	*Treponema azotonutricium* ZAS-9	AF320287.1
127	290cost002-P3L-1875	EF454273	<1			*Spirochaetes*	TTG-10	91.3	*Treponema azotonutricium* ZAS-9	AF320287.1
128	290cost002-P3L-2228	EF454497	<1			*Spirochaetes*	TTG-10	91.9	*Treponema* isolate strain SPIT5	AM182455.1
129	290cost002-P3L-2341	EF454571	<1			*Fibrobacteres*	TG3-1	79.4	*Thermanaeromonas toyohensis*	AB062280.1

OTU[1]	PCR library OTU rep.[2]	Acc. #[3]	%[4]	Metagenome	%[5]	Phylum	Phylog. group[6]	Closest cultivated species / isolate %[7]	Name	Acc. #[8]
130	290cost002-P3L-1175	EF453820	<1			*Acidobacteria*		82.1	*Acidobacteriaceae* str. Gsoil 149	AB245339.1
131	290cost002-P3L-1241	EF453868	<1	Contig30608	2	*Acidobacteria*	*Holophagales*	85.9	*Holophaga foetida*	X77215.1
132	290cost002-P3L-2407	EF454612	<1			*Acidobacteria*	*Holophagales*	85.6	*Holophaga foetida*	X77215.1
133	290cost002-P3L-693	EF454928	<1			*Acidobacteria*		82.6	*Acidobacteriaceae* str. Gsoil 149	AB245339.1
134	290cost002-P3L-685	EF454925	<1			*Aminanaerobia*	TG5	84.3	*Dethiosulfovibrio russensis*	AF234544.1
135	290cost002-P3L-1187	EF453830	<1			*Bacteroidetes*	*Bacteroidales*	84.4	*Paludibacter propionicigenes*	AB078842.2
136	290cost002-P3L-1305	EF453914	<1			*Bacteroidetes*	*Bacteroidales*	84.1	*Paludibacter propionicigenes*	AB078842.2
137	290cost002-P3L-1802	EF454212	<1			*Bacteroidetes*	*Bacteroidales*	84.7	*Alistipes finegoldii*	AJ518874.1
138	290cost002-P3L-1961	EF454327	<1			*Bacteroidetes*	*Bacteroidales*	83.4	*Ruminobacillus xylanolyticum* str. G1	DQ178248.1
139	290cost002-P3L-2080	EF454409	<1			*Bacteroidetes*	*Bacteroidales*	84.9	*Alistipes finegoldii*	AJ518874.1
140	290cost002-P3L-2184	EF454472	<1			*Bacteroidetes*	*Bacteroidales*	84.7	*Alistipes finegoldii*	AJ518874.1
141	290cost002-P3L-2216	EF454492	<1			*Bacteroidetes*	*Bacteroidales*	82.8	*Bacteroides* str. 22C	AY554420.1
142	290cost002-P3L-2517	EF454675	<1			*Chlorobi*		78.3	*Desulfobacteraceae* str. MSL86	AB110542.2
143	290cost002-P3L-1309	EF453917	<1			*Cyanobacteria*		91.7	Rumen isolate str. YS2	AF544207.1
144	290cost002-P3L-1371	EF453956	<1			*Cyanobacteria*		89.9	Rumen isolate str. YS2	AF544207.1
145	290cost002-P3L-421	EF454802	<1			*Deferribacteres*		86.4	*Geovibrio* str. Lincoln Park 3	AF157057.1
146	290cost002-P3L-2571	EF454709	<1			*Endomicrobia*		79.4	*Acidobacteriaceae* str. Gsoil 149	AB245339.1
147	290cost002-P3L-1177	EF453821	<1			*Fibrobacteres*	TFG-1	84.4	*Fibrobacter intestinalis*	M62687.1
148	290cost002-P3L-1894	EF454284	<1			*Fibrobacteres*	TFG-1	84.7	*Fibrobacter intestinalis*	M62686.1
149	290cost002-P3L-2095	EF454418	<1	Contig41368	2	*Fibrobacteres*	TFG-1	84.4	*Fibrobacter* isolate	L35547.1
150	290cost002-P3L-2239	EF454506	<1			*Fibrobacteres*	TFG-1	84.3	*Fibrobacter intestinalis*	M62687.1
151	290cost002-P3L-1289	EF453901	<1			*Firmicutes*	*Acidaminococcaceae*	87.9	*Dehalobacter* sp. E2	AY673991.1
152	290cost002-P3L-1430	EF454003	<1			*Firmicutes*	*Lachnospiraceae*	89.2	*Eubacterium siraeum*	L34625.1
153	290cost002-P3L-1599	EF454073	<1			*Firmicutes*	*Lachnospiraceae*	90.9	*Sporobacter termitidis*	Z49863.1
154	290cost002-P3L-1640	EF454103	<1			*Firmicutes*	*Catabacteraceae*	82.8	*Clostridium* sp. str. FCB90-3	AJ229251.1
155	290cost002-P3L-1650	EF454110	<1			*Firmicutes*	*Catabacteraceae*	82.6	*Sedimentibacter* sp. D7	AY766467.1
156	290cost002-P3L-1924	EF454305	<1			*Firmicutes*	*Peptostreptococcaceae*	92.1	*Anaerovorax odorimutans* str. NorPut	AJ251215.1
157	290cost002-P3L-2244	EF454508	<1			*Firmicutes*	*Lachnospiraceae*	91.8	*Sporobacter termitidis*	Z49863.1
158	290cost002-P3L-2299	EF454544	<1			*Firmicutes*	*Lachnospiraceae*	83.8	*Ethanologenbacterium harbinense* str. X-29	AY833421.1
159	290cost002-P3L-2304	EF454547	<1			*Firmicutes*	*Acidaminococcaceae*	87.6	*Sporomusa rhizae* str. RS	AM158322.1
160	290cost002-P3L-2413	EF454616	<1			*Firmicutes*	*Lachnospiraceae*	90.0	*Eubacterium siraeum*	L34625.1
161	290cost002-P3L-2450	EF454637	<1			*Firmicutes*	*Catabacteraceae*	82.9	*Sedimentibacter* sp. D7	AY766467.1
162	290cost002-P3L-2602	EF454727	<1			*Firmicutes*	*Lactobacillales, Enterococcaceae*	92.9	*Enterococcus malodoratus* str. ATCC43197	AF061012.1

OTU[1]	PCR library OTU rep.[2]	Acc. #[3]	%[4]	Metagenome	%[5]	Phylum	Phylog. group[6]	Closest cultivated species / isolate %[7]	Name	Acc. #[8]
163	290cost002-P3L-2633	EF454749	<1			*Firmicutes*	*Lactobacillales, Enterococcaceae*	92.8	*Enterococcus avium*	AF061008.1
164	290cost002-P3L-446	EF454815	<1			*Firmicutes*	*Lachnospiraceae*	87.6	*Clostridium scindens* str. JCM 10420 M-18	AB020729.1
165	290cost002-P3L-568	EF454875	<1			*Firmicutes*	*Lachnospiraceae*	83.9	*Acetanaerobacterium elongatum* str. Z1	AY518589.1
166	290cost002-P3L-2033	EF454380	<1			*Proteobacteria*	*Deltaproteobacteria*	83.4	Sulfate-reducing bacterium str. Hxd3	Y17698.1
167	290cost002-P3L-2608	EF454731	<1			*Proteobacteria*	*Deltaproteobacteria, Desulfovibrionales*	95.1	*Oxalobacter formigenes* str. OXC	U49755.1
168	290cost002-P3L-532	EF454852	<1			*Proteobacteria*	*Deltaproteobacteria, Desulfovibrionales, Desulfovibrionaceae*	90.7	*Desulfovibrio* sp. str. LNB2	AY554146.1
169	290cost002-P3L-618	EF454896	<1			*Proteobacteria*	*Deltaproteobacteria, Desulfovibrionales, Desulfovibrionaceae*	92.1	*Bilophila wadsworthia* str. NB-12	AB117562.1
170	290cost002-P3L-430	EF454808	<1			*Spirochaetes*	*T. primitia* group	91.2	*Treponema primitia* ZAS-2	AF093252.1
171	290cost002-P3L-657	EF454914	<1			*Spirochaetes*	TLG-1	83.2	*Leptospira inadai*	Z21634.1
172	290cost002-P3L-632	EF454901	<1			*Spirochaetes*	TSG-1	85.5	*"Spirochaeta bajacaliforniensis"*	M71239.1
173	290cost002-P3L-1129	EF453791	<1			*Spirochaetes*	TTG-1	89.8	*Spirochaeta (Treponema) caldaria*	M71240.1
174	290cost002-P3L-1231	EF453862	<1	Contig41516	7	*Spirochaetes*	TTG-7	88.8	*Treponema primitia* ZAS-2	AF093252.1
174				Contig38802		*Spirochaetes*	TTG-7			
174				Contig11194		*Spirochaetes*	TTG-7			
175	290cost002-P3L-1239	EF453867	<1			*Spirochaetes*	TTG-5	90.9	*Treponema azotonutricium* ZAS-9	AF320287.1
176	290cost002-P3L-1301	EF453913	<1			*Spirochaetes*	TTG-4	89.4	*Treponema* isolate strain SPIT5	AM182455.1
177	290cost002-P3L-1392	EF453972	<1			*Spirochaetes*	TTG-5	90.7	*Treponema azotonutricium* ZAS-9	AF320287.1
178	290cost002-P3L-1422	EF453997	<1			*Spirochaetes*	TTG-6	90.7	*Spirochaeta (Treponema) caldaria*	M71240.1
179	290cost002-P3L-1667	EF454122	<1			*Spirochaetes*	TTG-1	89.4	*Spirochaeta (Treponema) caldaria*	M71240.1
180	290cost002-P3L-1703	EF454143	<1			*Spirochaetes*	TTG-4	88.5	*Spirochaeta (Treponema) caldaria*	M71240.1
181	290cost002-P3L-1728	EF454159	<1			*Spirochaetes*	TTG-2	89.6	*Treponema primitia* ZAS-1	AF093251.1
182	290cost002-P3L-1731	EF454161	<1			*Spirochaetes*	TTG-1	89.0	*Spirochaeta (Treponema) caldaria*	M71240.1
183	290cost002-P3L-1794	EF454205	<1			*Spirochaetes*	TTG-1	89.8	*Spirochaeta (Treponema) caldaria*	M71240.1
184	290cost002-P3L-1823	EF454228	<1	Contig25203	2	*Spirochaetes*	TTG-7	88.3	*Spirochaeta (Treponema) caldaria*	M71240.1
185	290cost002-P3L-1859	EF454259	<1	Contig41088	2	*Spirochaetes*	TTG-7	88.1	*Treponema azotonutricium* ZAS-9	AF320287.1
186	290cost002-P3L-1879	EF454277	<1			*Spirochaetes*	TTG-2	89.0	*Spirochaeta (Treponema) caldaria*	M71240.1
187	290cost002-P3L-2106	EF454423	<1			*Spirochaetes*	TTG-5	89.6	*Spirochaeta (Treponema) caldaria*	M71240.1
188	290cost002-P3L-2386	EF454600	<1	Contig41269	10	*Spirochaetes*	TTG-7	88.4	*Spirochaeta (Treponema) caldaria*	M71240.1
188				Contig35356		*Spirochaetes*	TTG-7			
188				Contig8763		*Spirochaetes*	TTG-7			

OTU[1]	PCR library OTU rep.[2]	Acc. #[3]	%[4]	Metagenome	%[5]	Phylum	Phylog. group[6]	Closest cultivated species / isolate %[7]	Name	Acc. #[8]
188				Contig132		*Spirochaetes*	TTG-7			
189	290cost002-P3L-2594	EF454723	<1			*Spirochaetes*	TTG-2	89.0	*Spirochaeta (Treponema) caldaria*	M71240.1
190	290cost002-P3L-418	EF454799	<1			*Spirochaetes*	TTG-7	88.0	*Spirochaeta (Treponema) caldaria*	M71240.1
191	290cost002-P3L-599	EF454887	<1			*Spirochaetes*	TTG-7	88.5	*Treponema primitia* ZAS-2	AF093252.1
192	290cost002-P3L-742	EF454947	<1			*Spirochaetes*	TTG-2	89.8	*Spirochaeta (Treponema) caldaria*	M71240.1
193	290cost002-P3L-746	EF454951	<1			*Spirochaetes*	TTG-7	88.6	*Treponema azotonutricium* ZAS-9	AF320287.1
194	290cost002-P3L-899	EF454989	<1			*Spirochaetes*	TTG-5	90.1	*Treponema azotonutricium* ZAS-9	AF320287.1
195	290cost002-P3L-1157	EF453804	<1			*Spirochaetes*	TTG-9	92.9	*Treponema primitia* ZAS-2	AF093252.1
196	290cost002-P3L-1258	EF453880	<1			*Spirochaetes*	TTG-9	91.9	*Treponema primitia* ZAS-2	AF093252.1
197	290cost002-P3L-1293	EF453905	<1			*Spirochaetes*	TTG-9	91.3	*Treponema primitia* ZAS-2	AF093252.1
198	290cost002-P3L-1645	EF454106	<1			*Spirochaetes*	TTG-9	92.6	*Treponema primitia* ZAS-2	AF093252.1
199	290cost002-P3L-535	EF454854	<1			*Spirochaetes*	TTG-9	92.6	*Treponema primitia* ZAS-2	AF093252.1
200	290cost002-P3L-1160	EF453807	<1			*Spirochaetes*	TTG-10	91.8	*Treponema* isolate strain SPIT5	AM182455.1
201	290cost002-P3L-1171	EF453817	<1			*Spirochaetes*	TTG-10	92.3	*Treponema primitia* ZAS-2	AF093252.1
202	290cost002-P3L-1214	EF453850	<1			*Spirochaetes*	TTG-10	91.9	*Treponema* isolate strain SPIT5	AM182455.1
203	290cost002-P3L-1591	EF454071	<1			*Spirochaetes*	TTG-10	92.5	*Treponema primitia* ZAS-1	AF093251.1
204	290cost002-P3L-1598	EF454072	<1			*Spirochaetes*	TTG-10	91.4	*Treponema* isolate strain SPIT5	AM182455.1
205	290cost002-P3L-1656	EF454114	<1			*Spirochaetes*	TTG-10	92.1	*Treponema* isolate strain SPIT5	AM182455.1
206	290cost002-P3L-1688	EF454133	<1			*Spirochaetes*	TTG-10	91.8	*Spirochaeta (Treponema) caldaria*	M71240.1
207	290cost002-P3L-1708	EF454146	<1			*Spirochaetes*	TTG-10	92.4	*Treponema* isolate strain SPIT5	AM182455.1
208	290cost002-P3L-1874	EF454272	<1			*Spirochaetes*	TTG-10	91.8	*Treponema primitia* ZAS-2	AF093252.1
209	290cost002-P3L-505	EF454845	<1			*Spirochaetes*	TTG-10	92.5	*Treponema* isolate strain SPIT5	AM182455.1
210	290cost002-P3L-559	EF454870	<1			*Spirochaetes*	TTG-10	91.5	*Treponema* isolate strain SPIT5	AM182455.1
211	290cost002-P3L-660	EF454916	<1			*Spirochaetes*	TTG-10	92.0	*Treponema* isolate strain SPIT5	AM182455.1
212	290cost002-P3L-738	EF454945	<1			*Spirochaetes*	TTG-10	91.6	*Treponema* isolate strain SPIT5	AM182455.1
213	290cost002-P3L-762	EF454956	<1			*Spirochaetes*	TTG-10	91.8	*Treponema primitia* ZAS-2	AF093252.1
214	290cost002-P3L-2049	EF454390	<1			*Fibrobacteres*	TG3-1	79.3	*Thermanaeromonas toyohensis*	AB062280.1
215	290cost002-P3L-2381	EF454598	<1			*Fibrobacteres*	TG3-1	79.7	*Thermanaeromonas toyohensis*	AB062280.1
216	290cost002-P3L-2679	EF454781	<1			*ZB3*		78.7	*Clostridium* sp. str. JC3	AB093546.1
				Contig5625	2	*Proteobacteria*	*Alphaproteobacteria, Oleomonas*	77.3	Petrobacter sp. DM-3	DQ539621.1
				BHZN43990	2	*Bacteroidetes*	*Bacteroidales*	93.1	Bacillus strain NIH	L11886.1
				Contig20711	2	*Proteobacteria*	*Deltaproteobacteria, Desulfovibrionales, Desulfovibrionaceae*	90.8	Desulfovibrio vulgaris str. Hildenborough	AE017285.1

	PCR library			Metagenome				Closest cultivated species / isolate		
OTU[1]	OTU rep.[2]	Acc. #[3]	%[4]		%[5]	Phylum	Phylog. group[6]	%[7]	Name	Acc. #[8]
				Contig18107	2	*Acidobacteria*	*Holophagales*	83.9	Holophaga foetida	X77215.1
				Contig40113	2	*Spirochaetes*	TTG-7	91.0	*Treponema azotonutricium* ZAS-9	AF320287.1
				Contig2434	2	*Spirochaetes*	TTG-7	83.1	Spirochaeta xylanolyticus	AY735097.1

[1] OTUs were defined at 99% sequence identity; [2] OTU representative; [3] Genbank accession number for OTU representative; [4] Percentage in PCR library; [5] Percentage in metagenome; [6] phylogenetic group as depicted in Fig. 1b; [7] percent sequence identity with closest cultivated species; [8] Genbank accession number; [9] fosmid Contig88; [10] fosmid contig110.

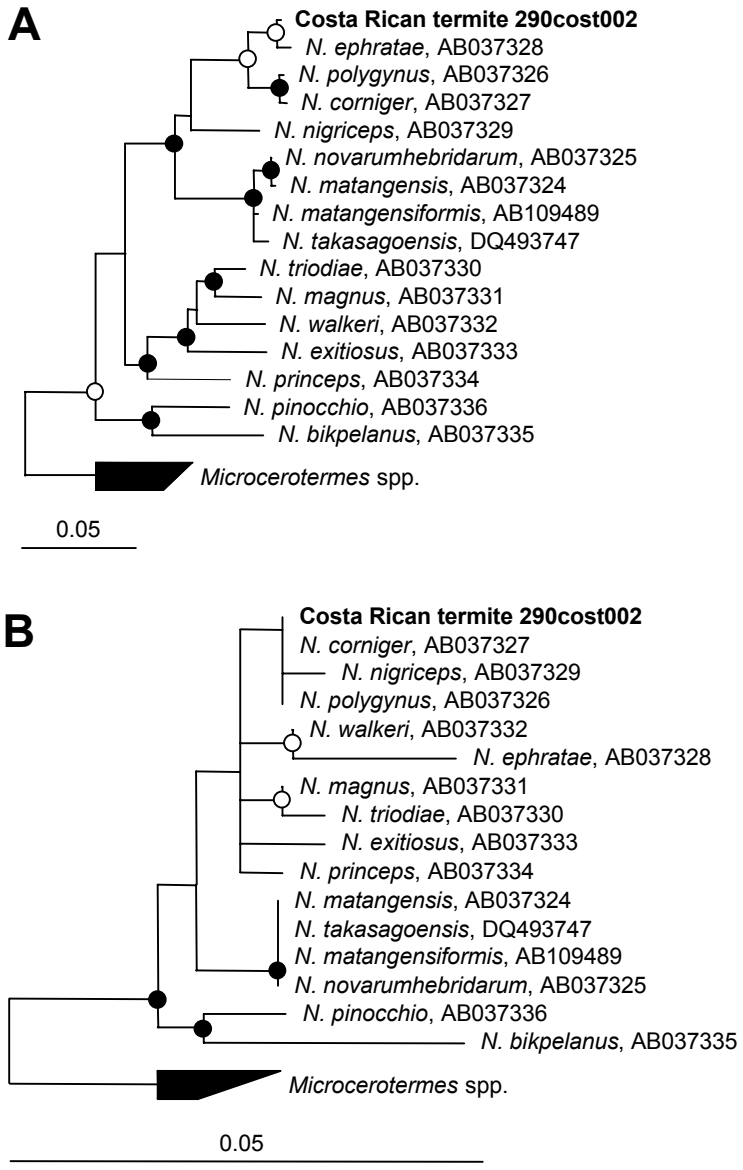


Fig. S1. Phylogenetic identification of termite host species 290cost002. The sequence of the mitochondrial cytochrome oxidase II (COII) gene was determined and aligned to all available *Isoptera* sequences, only species of the higher termite (*Termitidae*) genera *Nasutitermes* and *Microcerotermes* are shown. **A**, DNA nucleotide-based dendrogram using 661 unambigously aligned and filtered positions. **B,** DNA sequence was translated into amino acid sequence and a dendrogram was constructed using 228 unambiguously aligned and filtered positions. The trees were calculated using TREE-PUZZLE (Schmidt et al. 2002). Branching pattern confidence is visualised as follows: no circle, >50%; open circle, >70%; filled circle, >90%. The slightly contradictory tree topologies based on nucleotide and amino acid sequences is due to a large number of silent mutations in this gene. The collected specimens were most closely related to *N. ephratae* (by nt comparison, **A**) and *N. corniger* (by aa comparison, **B**) and we therefore only assign our specimens to *Nasutitermes* sp. in lieu of a comprehensive analysis of multiple collections of this and other closely related *Nasutitermes* throughout the region.

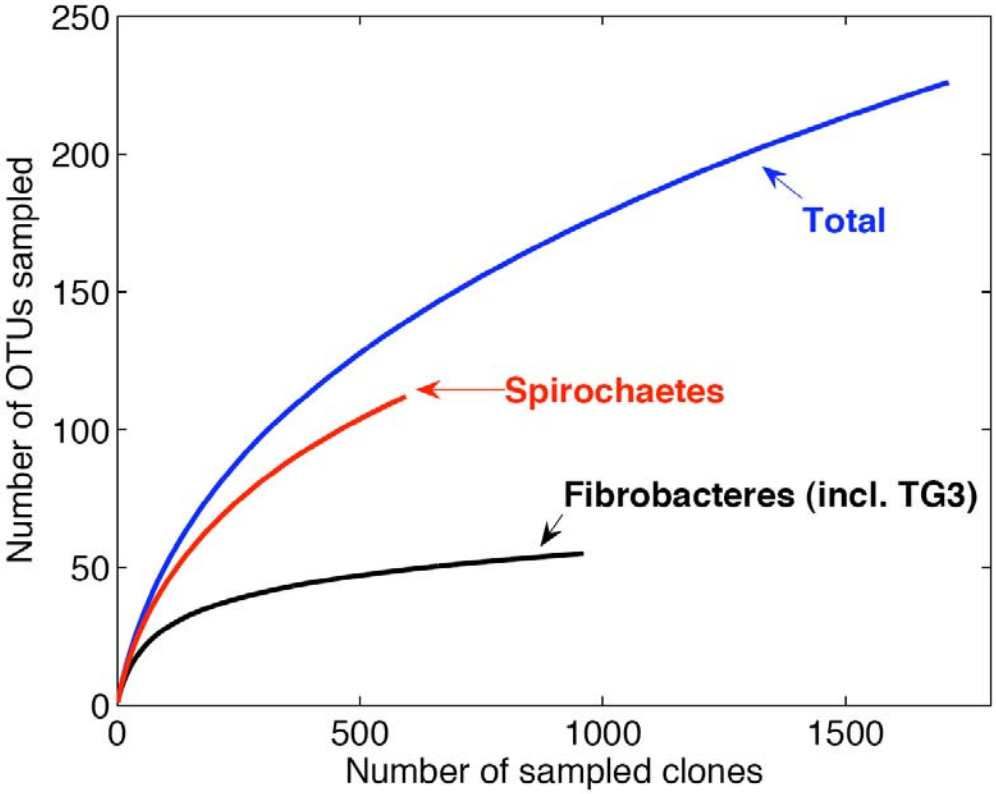


Fig. S2. Accumulation curves of 16S rRNA genes obtained from the P3 luminal microbiota . Accumulation curves were constructed by plotting the number of unique Operational taxonomic units (OTUs) at 99% sequence identity found in a given number of clones versus the number of clones (averaged over a 1000 random permutations of the initial clone order). For comparison, besides an accumulation curve for the total number of OTUs, accumulation curves for the *Spirochaetes* related OTUs and the *Fibrobacteres* including TG3-related OTUs are shown, too. For the the total OTU accumulation curve, neither the exponential nor the two parameter hyperbola models offered good fits, so they were dropped. The non-parametric estimates for this data set produced more meaningful estimates of the total diversity (Table S1). Based on this estimates, 71 to 84% of the total OTU diversity has been sequenced.

Fig. S3. Phylogenetic diversity of P3 luminal microbiota within the phylum *Spirochaetes*. From a PCR-based inventory and from the metagenome libraries 1703 and 14 near-full length 16S rRNA gene sequences, respectively, were used in a Maximum Likelihood analysis (RAxML). The Phylogram was constructed using 1289 unambiguously aligned and filtered nucleotide positions. Genbank accession numbers are given. Bar, 10% estimated sequence divergence.

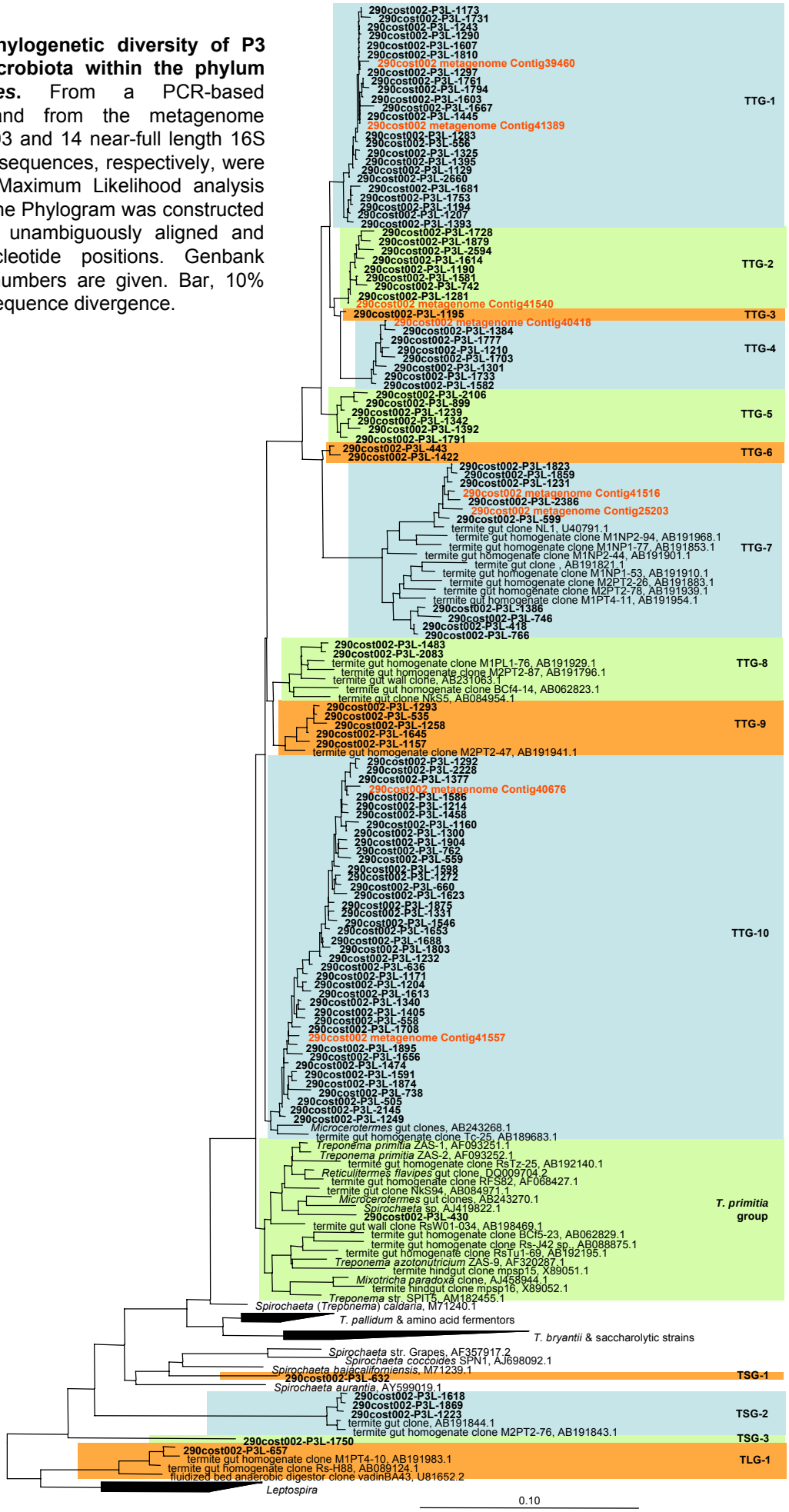


Fig. S4. Phylogenetic diversity of P3 luminal microbiota within the phylum *Fibrobacteres*. From a PCR-based inventory and from the metagenome libraries 1703 and 14 near-full length 16S rRNA gene sequences, respectively, were used in a Maximum Likelihood analysis (RaxML). The Phylogram was constructed using 1289 unambiguously aligned and filtered nucleotide positions. Genbank accession numbers are given. Bar, 10% estimated sequence divergence.

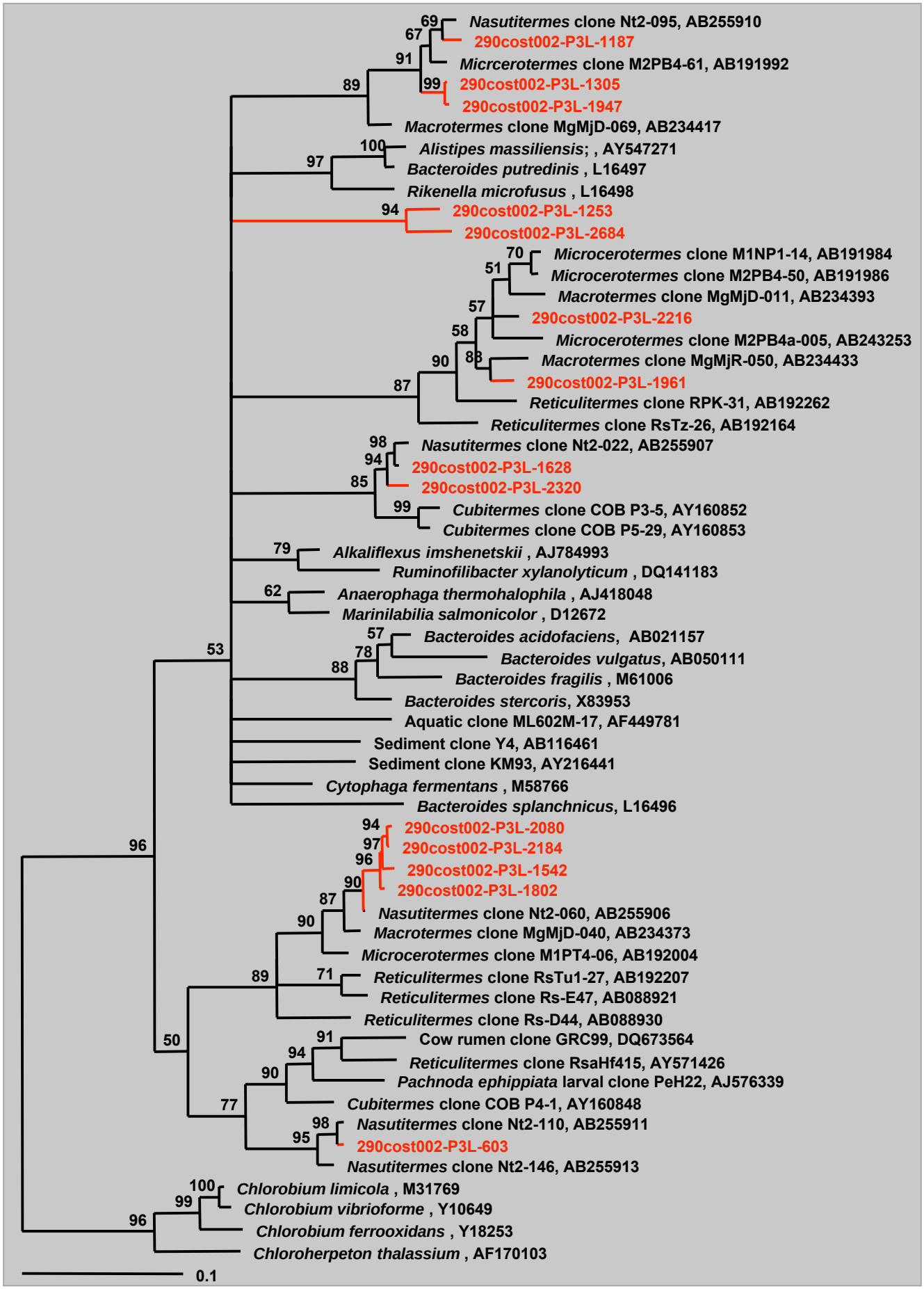


Fig. S5. 16S rRNA Phylogeny of Bacteroidetes-associated *Nasutitermes* gut clones. 99% OTU representatives of 16S rRNA sequencces amplified from *Nasutitermes* P3 lumen fluid are marked in red. Tree-Puzzle Maximum likelihood analysis was based on 844 unambiguously aligned nucleotides, HKY method for substitution across sites with uniform rate change. The Tree-Puzzle support values shown above nodes represent 5,000 puzzling steps.

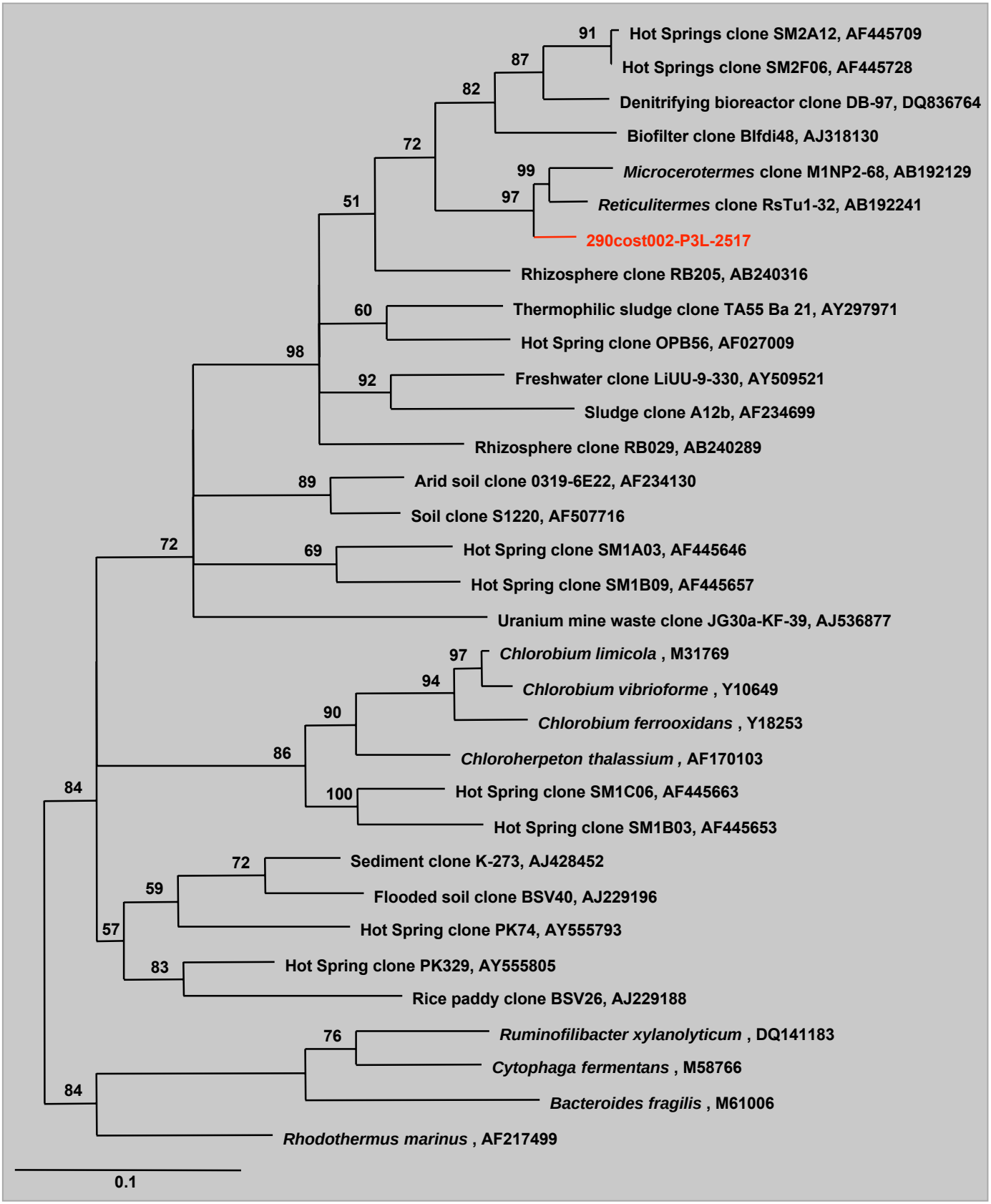


Fig. S6. Phylogeny of a Chlorobi-associated *Nasutitermes* gut clone. 16S rRNA sequences amplified from *Nasutitermes* P3 lumen fluid are marked in red. Tree-Puzzle Maximum likelihood analysis was based on 1028 unambiguously aligned nucleotides, HKY method for substitution across sites with uniform rate change. The Tree-Puzzle support values shown above nodes represent 10,000 puzzling steps.

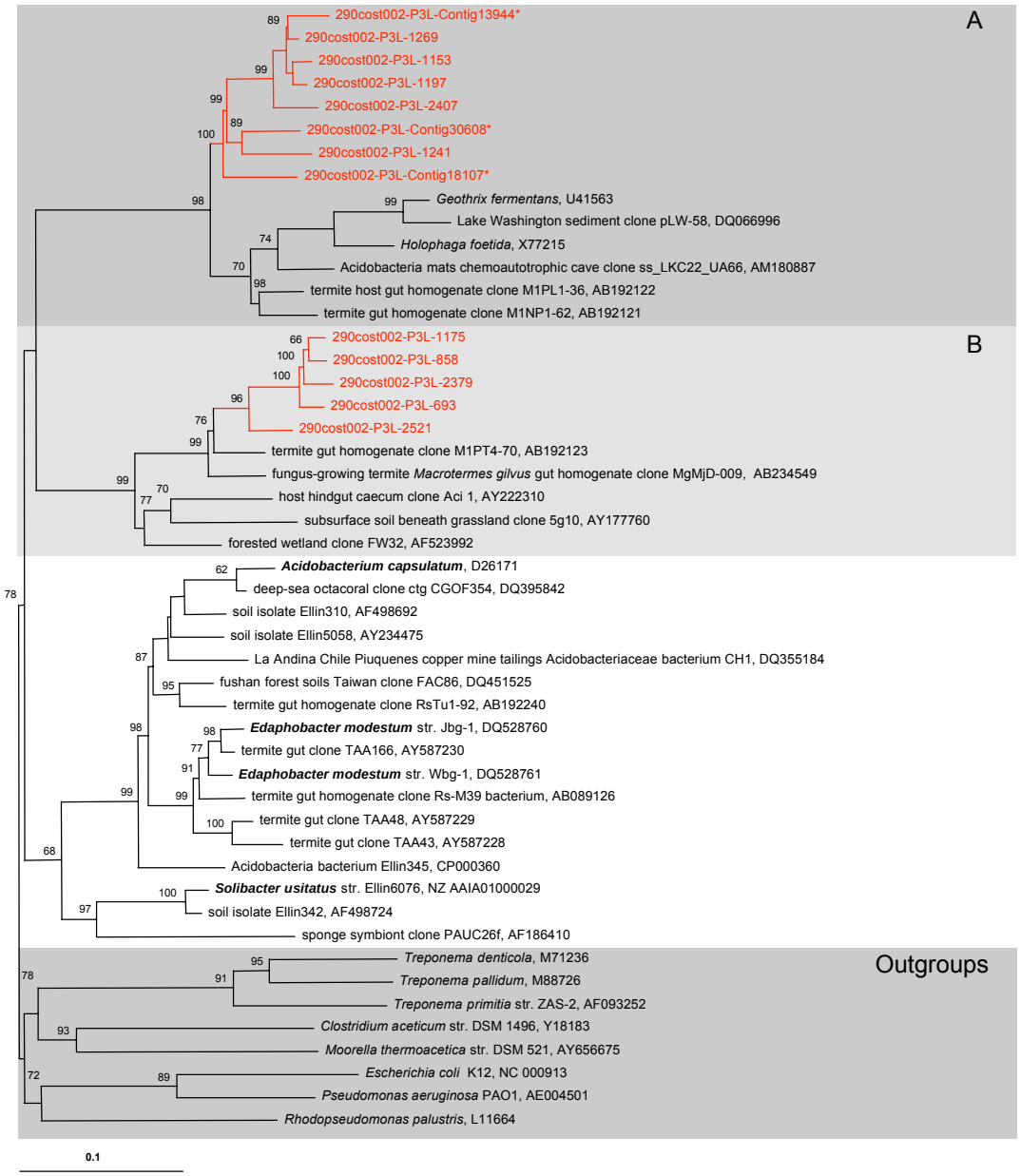


Fig. S7. Phylogeny of 16s rRNA sequences affiliating with the phylum *Acidobacteria* recovered from *Nasutitermes* P3 lumen contents. Tree-Puzzle Maximum likelihood analysis of metagenome sequences identified in the present study (highlighted in red) was based on 1044 unambiguously aligned nucleotides. Sequences marked with (*) are partial gene fragments and were added later to the tree by parsimony methods. 16s rRNA sequences from phyla *Spirochaetes*, *Firmicutes*, and *Proteobacteria* were used to outgroup the unrooted tree. The Tree-Puzzle support values shown above nodes represent confidence values from a 10,000 puzzling step analysis of the dataset. Bar represents evolutionary distance given as 0.1 changes per nucleotide position.

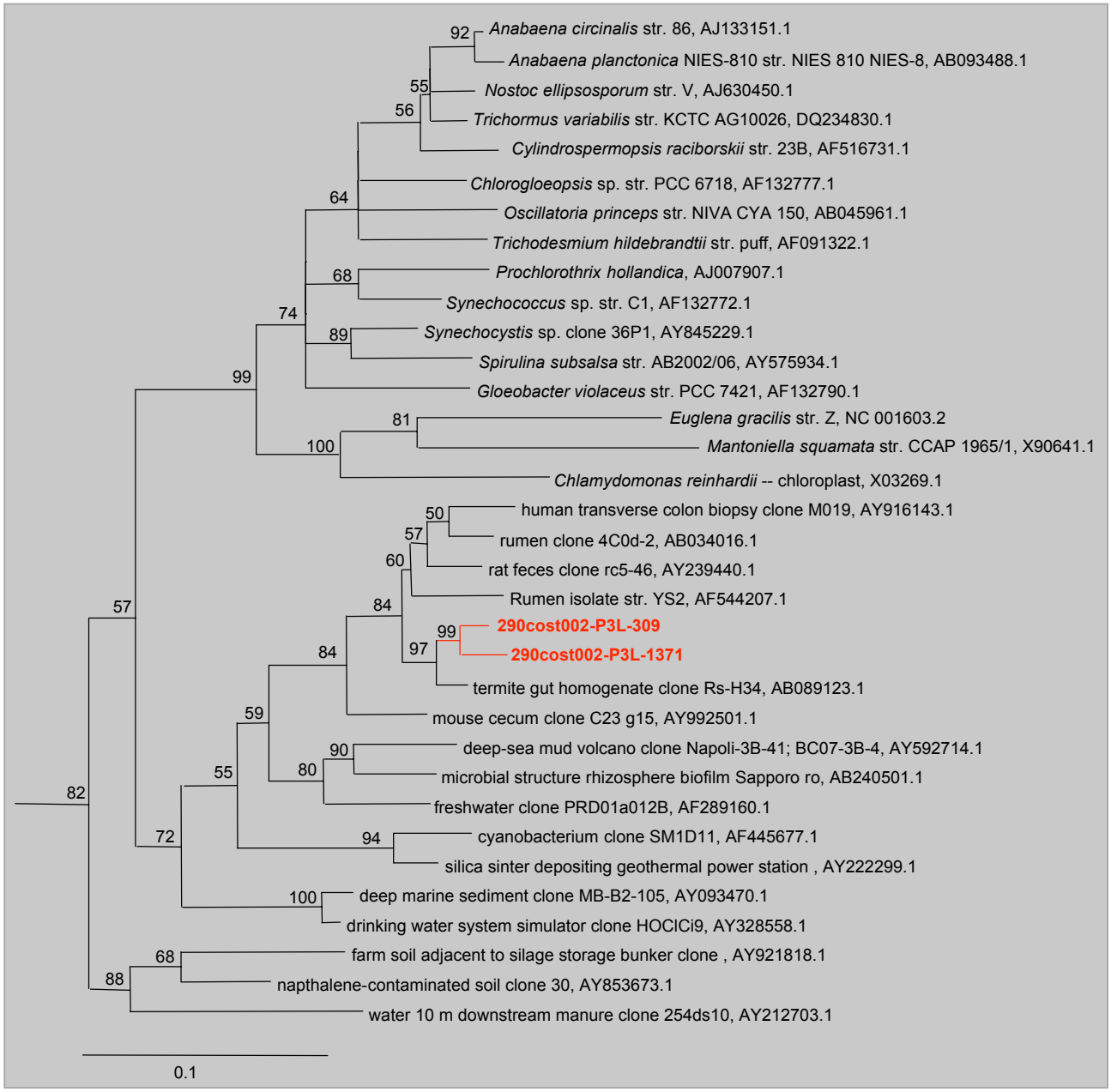


Fig. S8. 16S rRNA phylogeny of termite gut clones recovered from *Nasutitermes* P3 lumen fluid affiliating with the phylum *Cyanobacteria*. Termite metagenome sequences are highlighted in red. Tree-Puzzle Maximum likelihood analysis was based on 958 unambiguously aligned nucleotides, HKY method for substitution across sites with uniform rate change. *Firmicutes* sequences (not shown) were used to outgroup this tree. The Tree-Puzzle support values shown above nodes represent 10,000 puzzling steps.

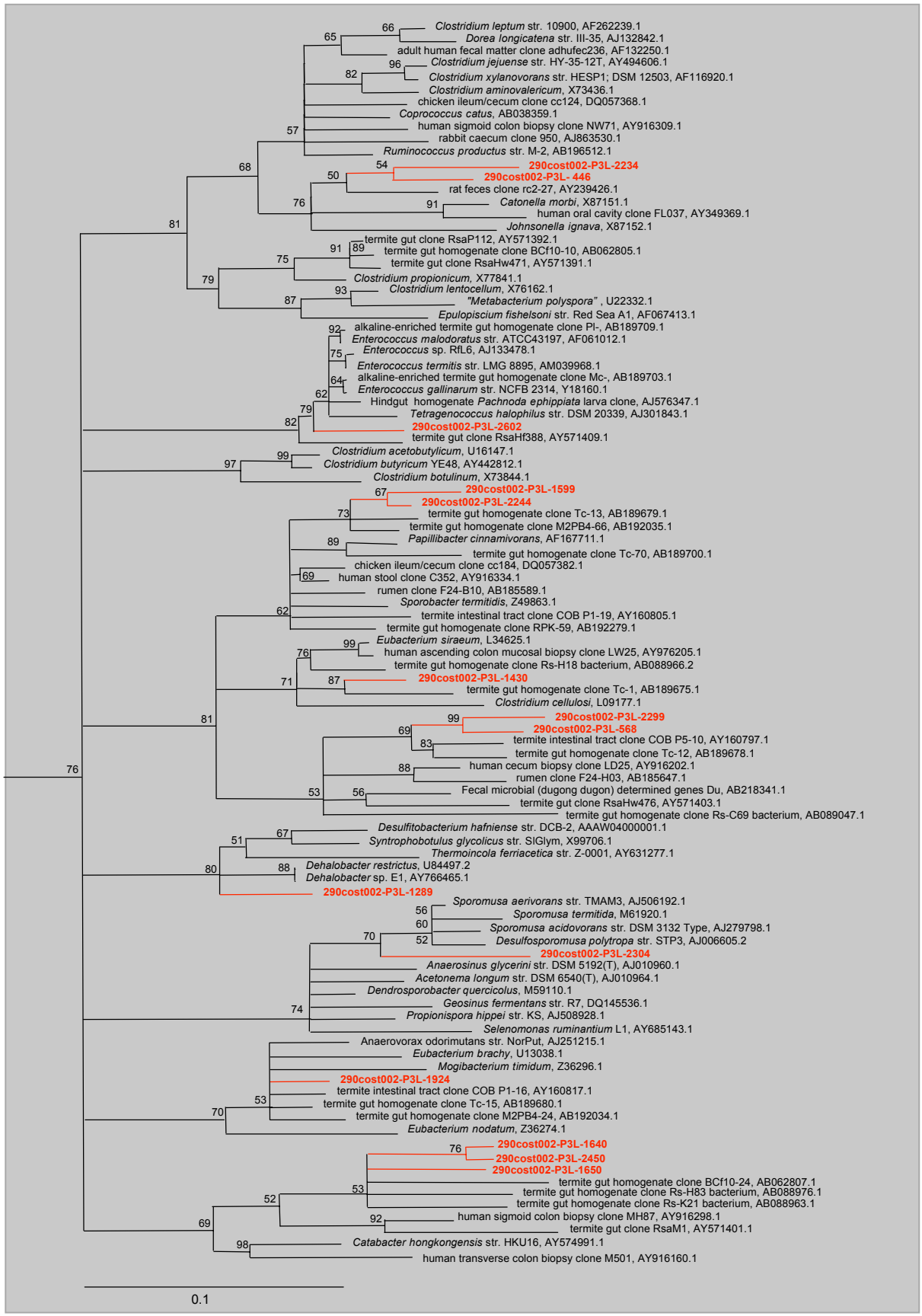


Fig. S9. 16S rRNA phylogeny of termite gut clones recovered from *Nasutitermes* P3 lumen fluid affiliating with the phylum *Firmicutes*. Termite metagenome sequences are highlighted in red. Tree-Puzzle Maximum likelihood analysis was based on 875 unambiguously aligned nucleotides, HKY method for substitution across sites with uniform rate change. *Cyanobacteria* sequences (not shown) were used to outgroup *Firmicutes*. The Tree-Puzzle support values shown above nodes represent 1,000 puzzling steps.

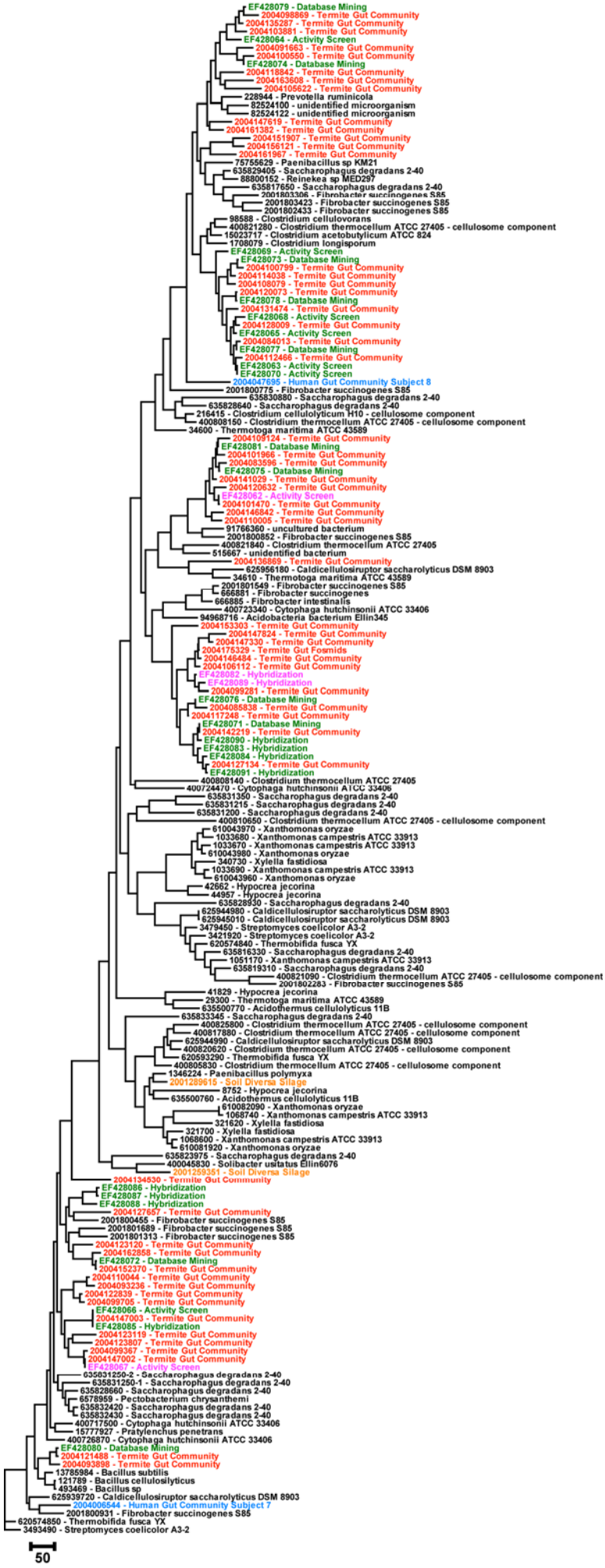


Fig. S10. Phylogenetic analysis of the glycoside hydrolase family 5 (GH5) diversity encoded by the termite gut P3 microbiota. GH5 sequences from the termite gut metagenome are colored in red, the soil metagenome in orange, human biome in cyan, and from various other sources in black. (Meta)genomic and additional public sequences are identified by their JGI Gene Object Identifier and GenBank GI number, respectively. Full-length subclone sequences recovered by database mining, activity screening, or hybridization (as indicated) are identified by their GenBank accession code. Subclones colored in green are active on phosphoric acid swollen cellulose (PASC), and subclones colored in magenta are inactive on PASC. The Pfam PF00150 (Cellulase – glycosyl hydrolase family 5) has a length of 380 residues, comprising domain sequences with an average length of 245 amino acids and 17% sequence identity. Analysis is based on a masked alignment of residues corresponding to the residues 57 to 346 of the endoglucanase from *Clostridium cellulovorans* (Genbank accession 98588).

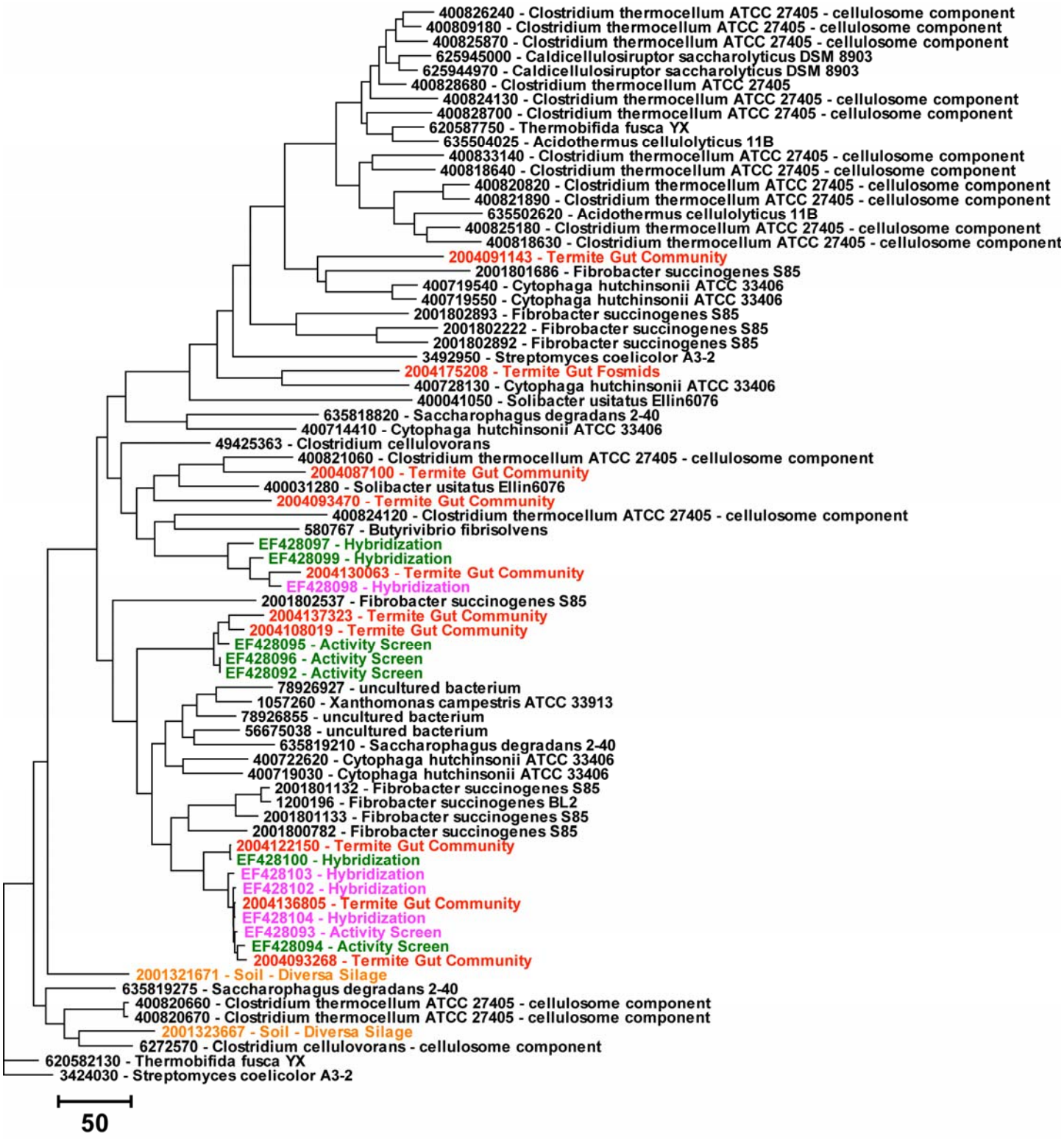
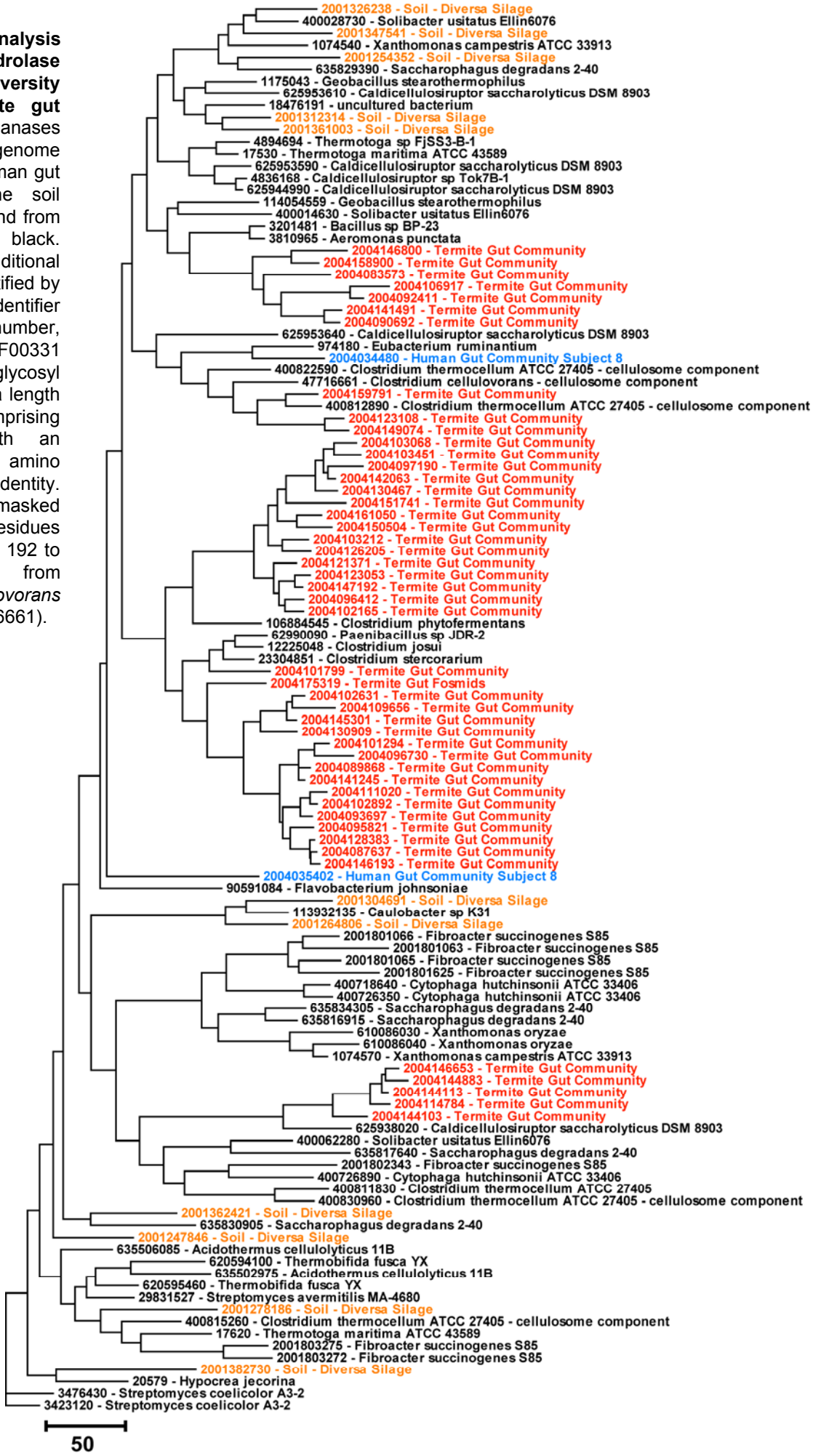


Fig. S11. Phylogenetic analysis of the glycoside hydrolase family 9 (GH9) diversity encoded by the termite gut microbiota. GH9 sequences from the termite gut metagenome are colored in red, the soil metagenome in orange, and from various other sources in black. (Meta)genomic and additional public sequences are identified by their JGI Gene Object Identifier and GenBank GI number, respectively. Full-length subclone sequences recovered by database mining, activity screening, or hybridization (as indicated) are identified by their GenBank accession code. Subclones colored in green are active on phosphoric acid swollen cellulose (PASC), and subclones colored in magenta are inactive on PASC. The Pfam PF00759 (Glyco_hydro_9 – glycosyl hydrolase family 9) has a length of 510 residues, comprising domain sequences with an average length of 373 amino acids and 34% sequence identity. Analysis is based on a masked alignment of residues corresponding to positions 286 to 721 of the non-cellulosomal enzyme EngO from *Clostridium cellulovorans* (Genbank accession 49425363).

Fig. S12. Phylogenetic analysis of the glycoside hydrolase family 10 (GH10) diversity encoded by the termite gut microbiota. GH10 xylanases from the termite gut metagenome are colored in red, the human gut microbiome in blue, the soil metagenome in orange, and from various other sources in black. (Meta)genomic and additional public sequences are identified by their JGI Gene Object Identifier and GenBank GI number, respectively. The Pfam PF00331 (Glyco_hydro_10 – glycosyl hydrolase family 10) has a length of 356 residues, comprising domain sequences with an average length of 232 amino acids and 32% sequence identity. Analysis is based on a masked alignment of 356 residues corresponding to positions 192 to 534 of xylanase B from *Clostridium cellulovorans* (Genbank accession 47716661).



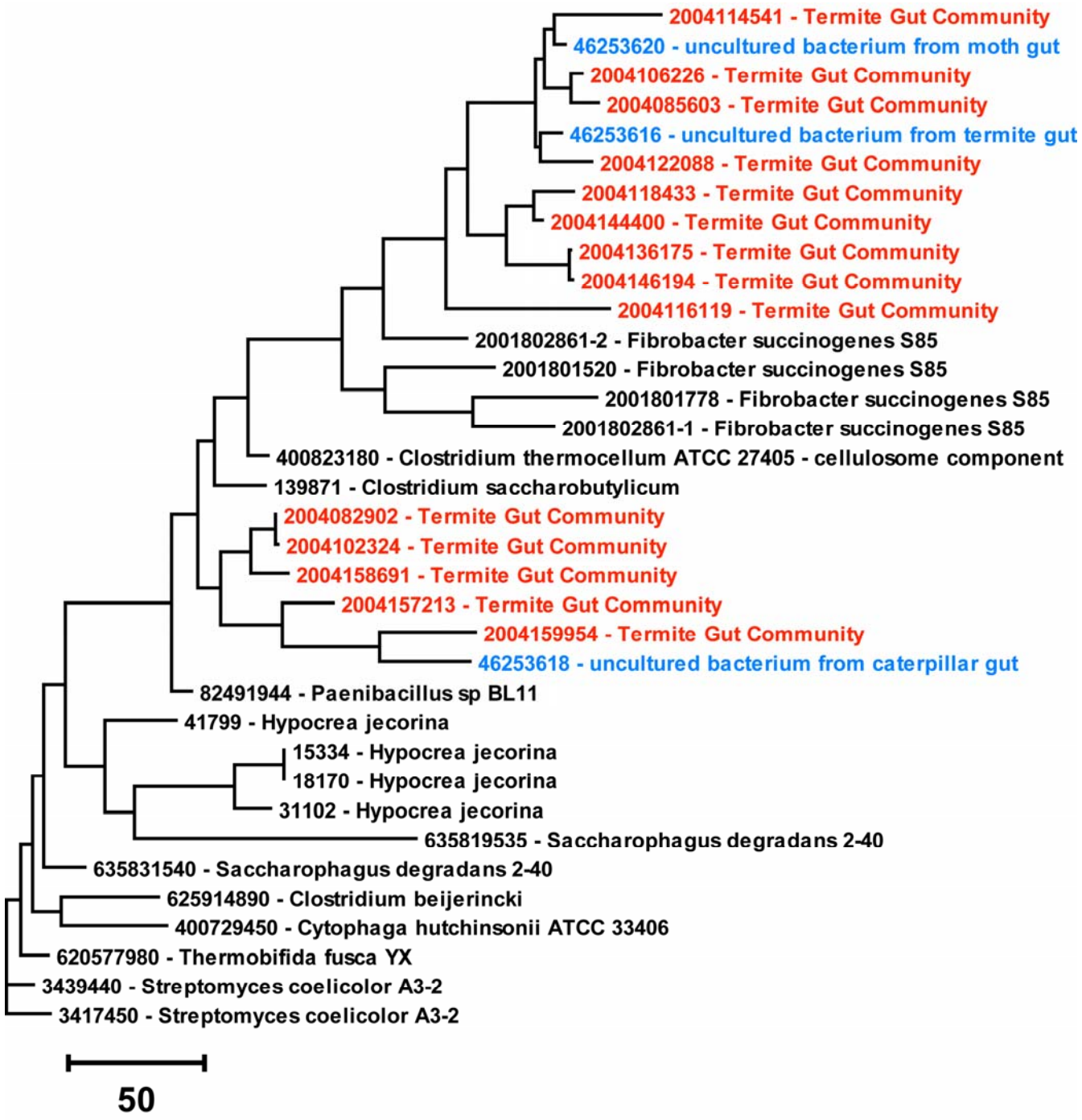


Fig. S13. Phylogenetic analysis of the glycoside hydrolase family 11 (GH11) diversity encoded by the termite gut microbiota. GH11 xylanases from the termite gut metagenome are colored in red, three previously discovered and characterized GH11 xylanases from insect guts (Brennan et al., 2004) are colored in blue, and GH11 xylanases from various other sources are colored in black. (Meta)genomic and additional public sequences are identified by their JGI Gene Object Identifier and GenBank GI number, respectively. The Pfam PF00457 (Glyco_hydro_11 – glycosyl hydrolase family 11) has a length of 194 residues, comprising domain sequences with an average length of 173 amino acids and 48% sequence identity. Analysis is based on a masked alignment of 194 residues corresponding to the positions 68 to 254 of the endo-1,4-β-xylanase from *Clostridium saccharobutylicum* (Genbank accession 139871).

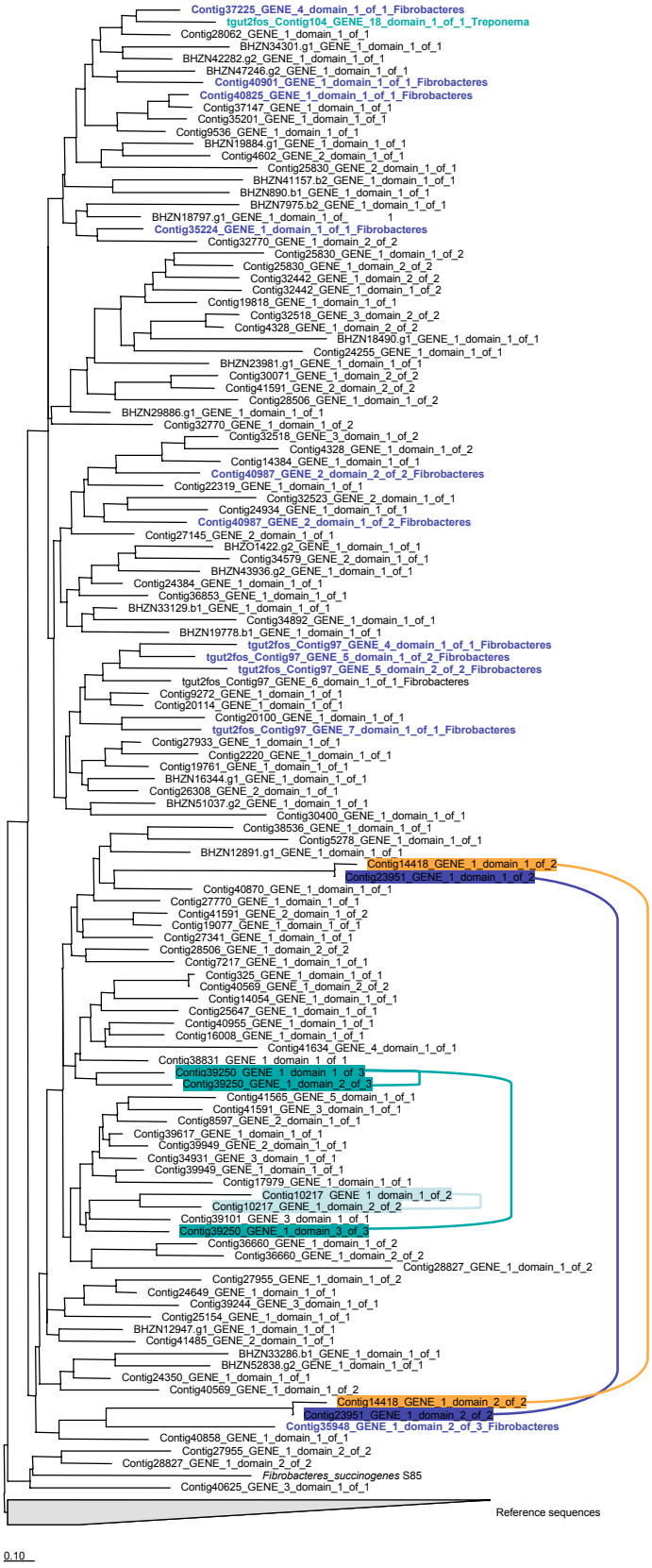
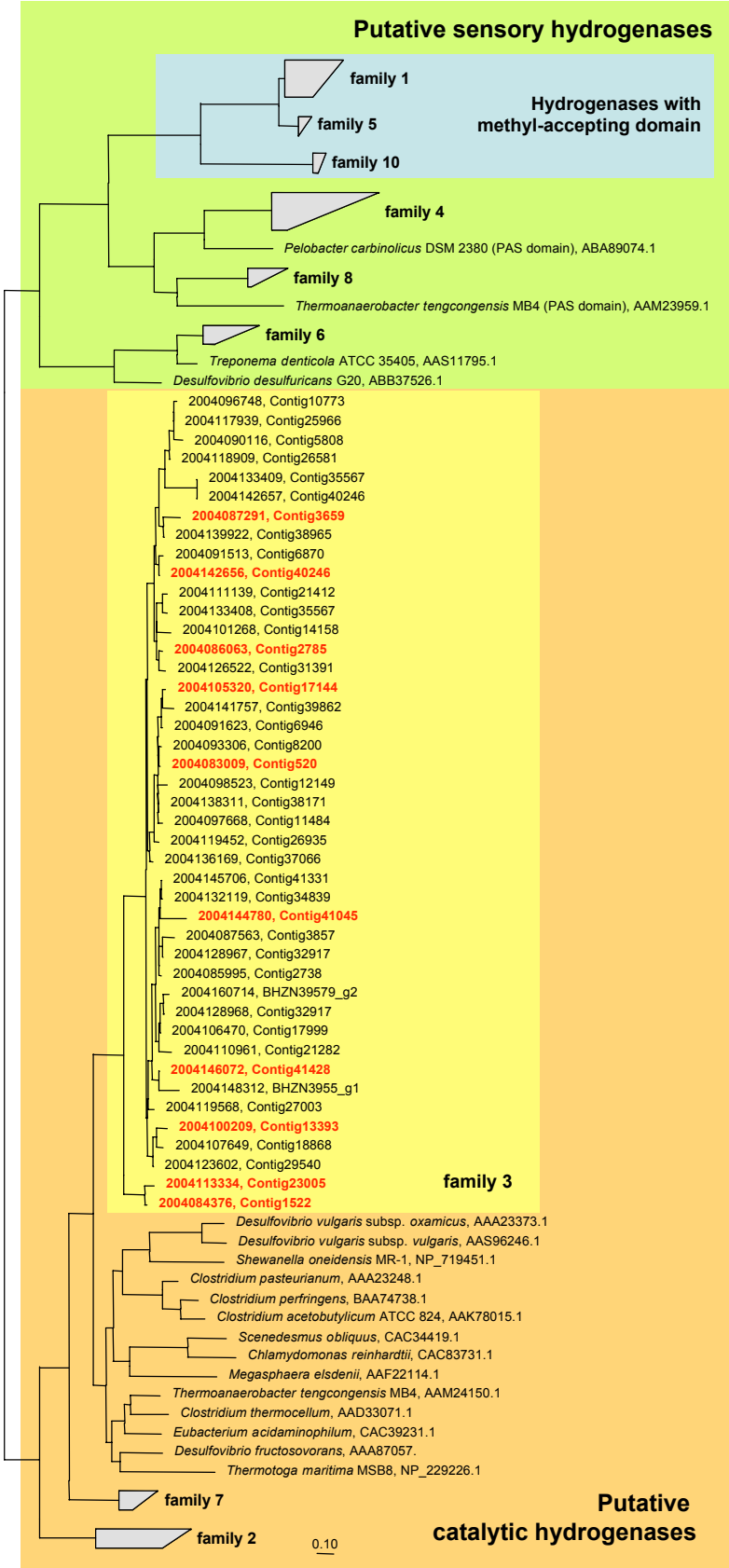


Fig. S14. Phylogenetic analysis of overrepresented conserved hypo-thetical gene family TOG1 (TIGR02145) encoded by the P3 luminal microbiota. For the analysis the amino acid sequences of the domain were aligned and a Maximum Likelihood tree was calculated (Phylip PROML, 124 positions considered). Reference sequences comprised: *Fibrobacteres succinogenes*, *Bacteroidetes fragilis* and *B. thetaiotaomicron*, *Chlorobium tepidum* and *Pelodictyon luteolum*. Names are printed bold and in blue color for contigs assigned to *Fibrobacteres* by PhyloPythia and green for *Treponema*. Color-coded boxes visualize examples of the phylogenetic position of domains belonging to the same gene. Bar, 10% estimated sequence divergence.

Fig. S15. Phylogenetic diversity of iron-only hydrogenases. In the metagenomic dataset >150 genes coding for hydrogenase catalytic domains (pfam02906, large subunit, C-terminal domain) were discovered. Based on the overall protein sequence similarity, Fe-only hydrogenase proteins were grouped into 11 families; since many sequences represented gene fragments rather than complete genes, proteins within each family were aligned and a consensus sequence for each family was generated, which was further used for analysis of domain composition (Fig S17). The tree shows that Fe-only hydrogenases in the termite hindgut metagenome are separated into 2 large groups, with one of them containing domains typical for hydrogenases catalyzing hydrogen evolution or uptake, while in the other group catalytic the domain was fused to various other conserved domains (MA – methyl-accepting chemotaxis protein signaling domain, REC – response regulator receiver domain, PAS and PAS_4 – PAS/PAC sensor domain). Remarkably, most of these domains fused with Fe-only hydrogenase catalytic core in the termite hindgut metagenome are known to be involved in signal transduction. Based on domain architecture of Fe-only hydrogenase proteins we tentatively assigned the first group as "catalytic" hydrogenases and the second group as "sensory" hydrogenases. One of the largest families of "sensory" hydrogenases in termite hindgut metagenome was family 1, which represented a fusion of the catalytic domain with a methyl-accepting chemotaxis protein signaling domain and response regulator receiver domain. We hypothesize that the representatives of this family function as hydrogen chemoreceptors with hydrogenase domain serving as hydrogen sensor rather than a catalyst of hydrogen evolution or uptake. Search for proteins with similar domain architecture using CDART (Conserved Domain Architecture Retrieval Tool) retrieved no hits in any other organism, indicating that this family is unique for *Treponema* species inhabiting termite hindguts. It has been demonstrated that strains of *Treponema* isolated from the hindgut of lower termites alternatively either produce or utilize molecular hydrogen. For the phylogenetic analysis 288 alignment positions were considered. Bar, 10% estimated sequence divergence.



Geer L. Y., Domrachev M., Lipman D. J., Bryant S. H. CDART: protein homology by domain architecture. Genome Res. 2002. 12:1619-1623.

Graber, J.R., Leadbetter, J.R. & Breznak, J.A. Description of *Treponema azotonutricium* sp. nov. and *Treponema primitia* sp. nov., the first spirochetes isolated from termite guts. *Appl Environ Microbiol* **70**, 1315-1320 (2004).

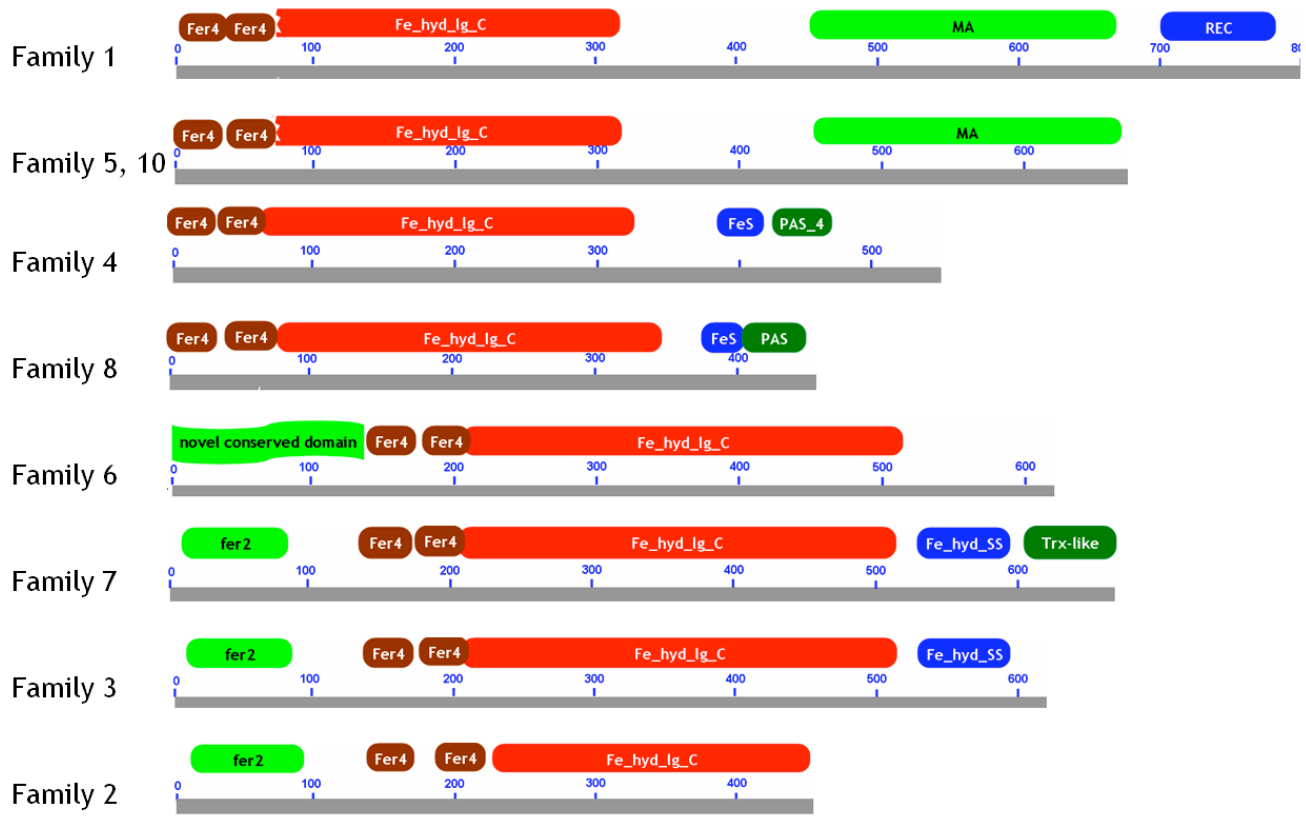


Fig. S16. Domain composition of iron-only hydrogenases. Conserved domains and motifs were detected in the consensus sequences of each family using InterProScan. The following domains were discovered: fer2 – pfam00111, 2Fe-2S iron-sulfur cluster binding domain; Fer4 – pfam00037, 4Fe-4S binding domain; Fe_hyd_lg_C – pfam02906, iron only hydrogenase large subunit, C-terminal domain; FeS – pfam04060, putative Fe-S cluster; MA – pfam00015, methyl-accepting chemotaxis protein (MCP) signaling domain; REC – pfam00072, response regulator receiver domain; Fe_hyd_SS – pfam02256, iron hydrogenase small subunit; PAS – pfam00989, PAS domain; PAS_4 – pfam08448, PAS_4 domain, part of the PAS domain clan; Trx-like – pfam00085, thioredoxin. Families 9 and 11 were not included in the phylogenetic analysis due to the very short sequences contained therein In the metagenomic dataset >150 genes coding for hydrogenase catalytic domains were discovered.

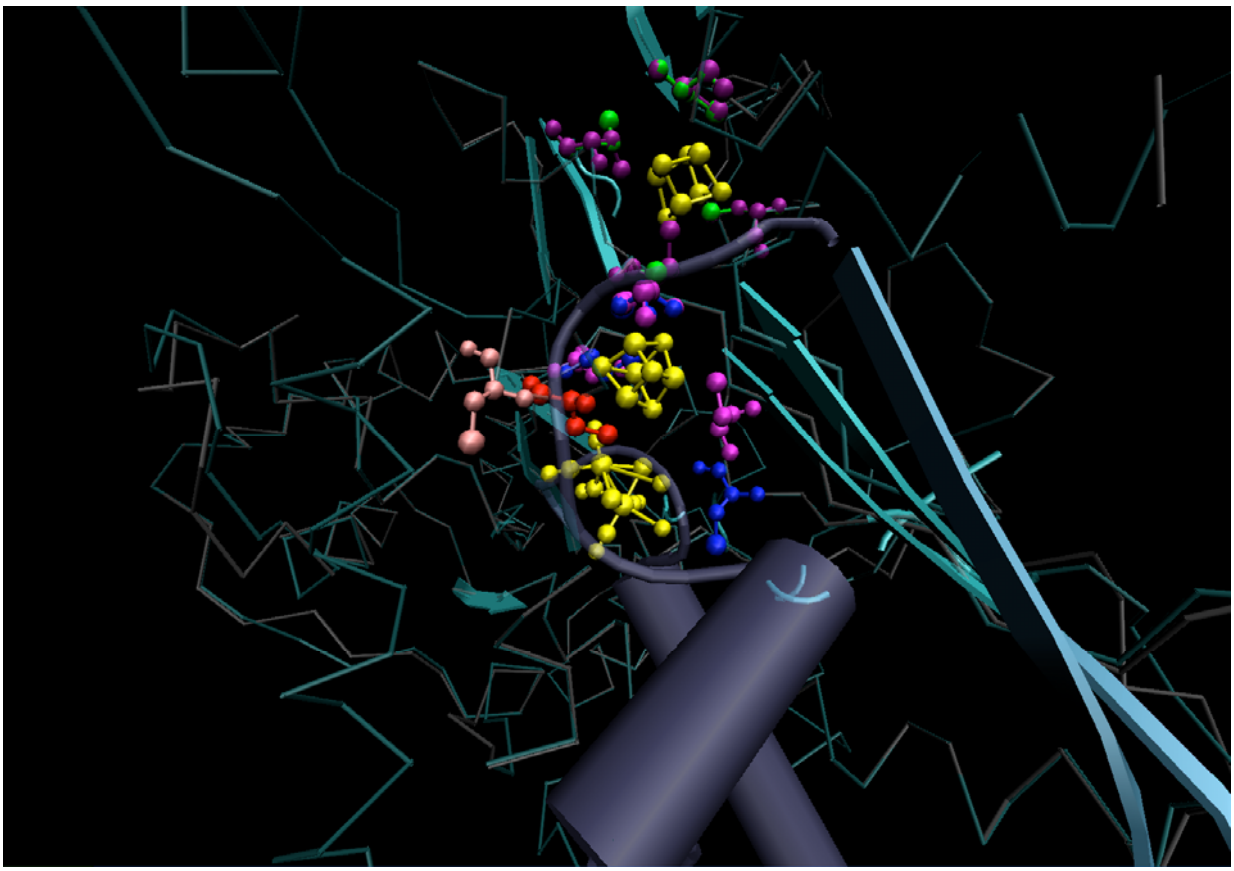


Fig. S17. Three-dimensional structure of a family 2 Fe-only hydrogenase representative obtained by comparative modeling using the periplasmatic *Clostridium pasteurianum* enzyme CpI as template and superimposed on its structure. The *C. pasteurianum* enzyme is colored cyan, family 2 protein is silver. The region in the *C. pasteurianum* protein that was deleted in family 2 is ice-blue. Iron-sulfur clusters in *C. pasteurianum* protein are presented as yellow CPK; cysteines coordinating proximal FS4A cluster are purple CPK in *C. pasteurianum* and green in family 2 protein. Cysteines coordinating H cluster are purple in *C. pasteurianum* and blue in family 2 protein, except for the subcluster-bridging cysteine, which is red in *C. pasteurianum* protein and pink in family 2 protein. Among the families of "catalytic" hydrogenases family 2 was the most remarkable one. Representatives of this family are very distant from any other Fe-only hydrogenase proteins and lack the C-terminal domain found in most Fe-only hydrogenases characterized so far; in addition, alignment of the catalytic domain showed that some of the cysteine residues coordinating the active-site iron-sulfur cluster may occupy different positions than in typical enzymes. The shown structure reveals that while the cysteine residues coordinating the proximal FS4A cluster align very well, cysteines coordinating the active-site iron-sulfur cluster (including the subcluster-bridging Cys) occupy different positions.

Peters J. W., Lanzilotta W. N., Lemon B. J., Seefeldt L. C. X-ray crystal structure of the Fe-only hydrogenase (CpI) from *Clostridium pasteurianum* to 1.8 angstrom resolution. Science. 1998. 282:1853-1858

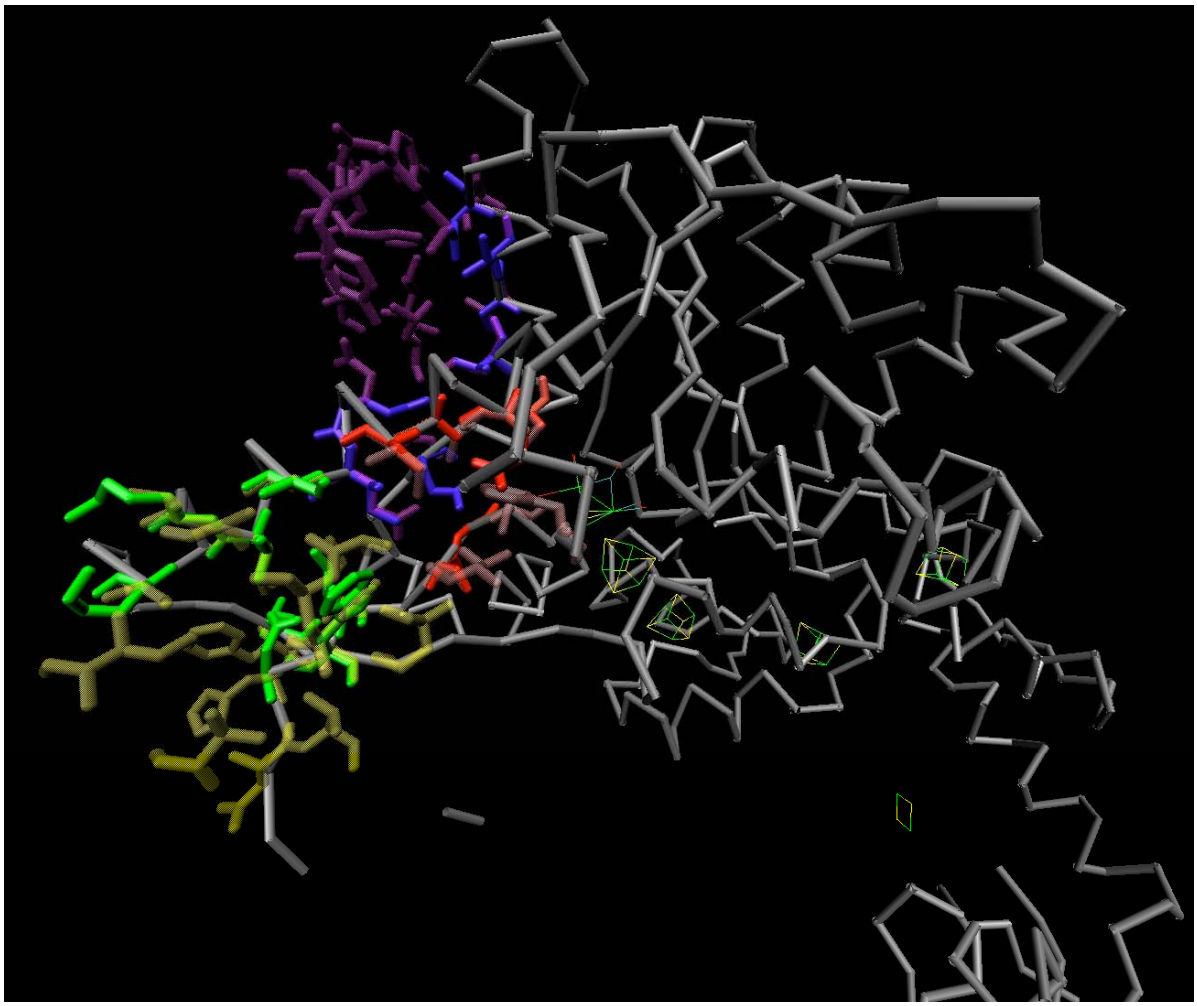


Fig. S18. Predicted hydrogen diffusion channels in *C. pasteurianum* Fe-only hydrogenase superimposed on the predicted 3D structure of a family 2 representative. Hydrogen diffusion channel 1 in *C. pasteurianum* protein is ochre, residues conserved in family 2 protein are green; hydrogen diffusion channel 2 in *C. pasteurianum* is dark magenta, residues conserved in family 2 protein are blue. Residues conserved in family 2 protein and corresponding to the central chamber in *C. pasteurianum* protein are red. In conclusion, residues forming predicted hydrogen diffusion channels in *C. pasteurianum* enzyme are partially conserved in family 2 proteins.

Cohen J., Kim K., Posewitz M., Ghirardi M. L., Schulten K., Seibert M., King P. Molecular dynamics and experimental investigation of H(2) and O(2) diffusion in [Fe]-hydrogenase. Biochem. Soc. Trans. 2005. 33:80-82

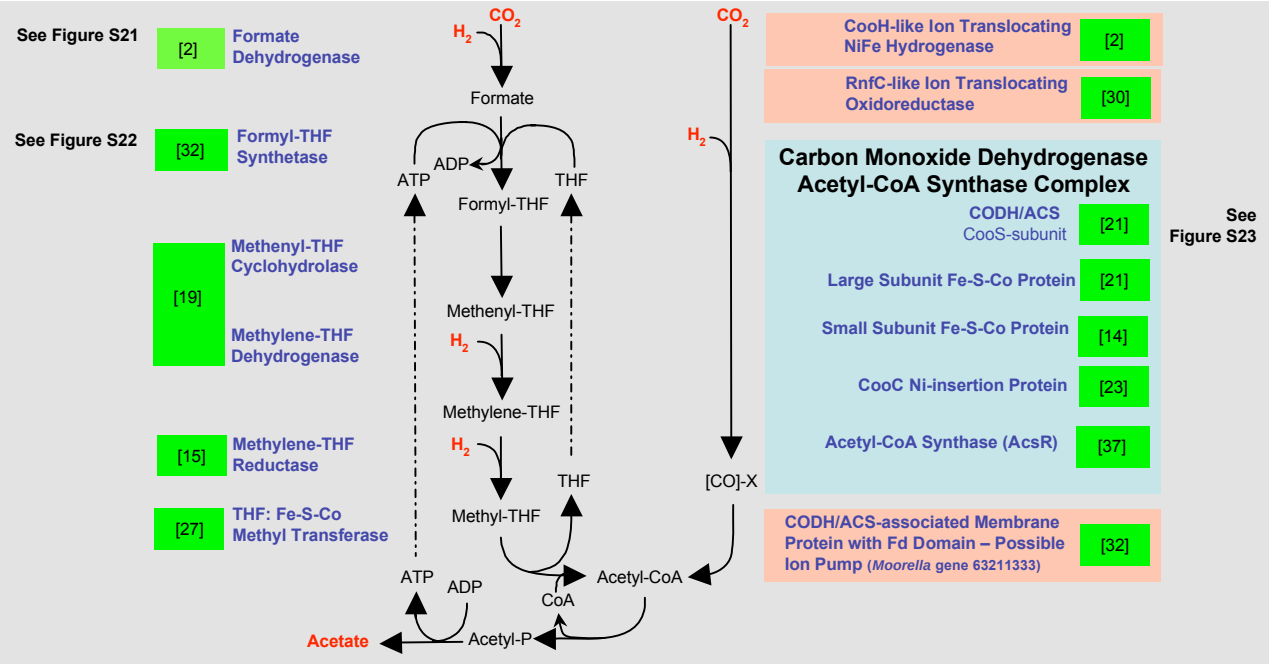


Fig. S19. Recovery of genes or partial genes encoding proteins relevant to the Wood-Ljungdahl pathway of CO_2-reductive homoacetogenesis. Green boxes denote the number of variants for each function recovered in the metagenome dataset. Note that for all of the functions, formate dehydrogenases appears to be particularly underrepresented, suggesting that either formate is generated from a precursor other than carbon dioxide (e.g. such as pyruvate), from within another gut compartment, or by an enzyme that belongs to a not yet recognizable FDH class of proteins. The blue box encompasses 5 proteins commonly encountered as being a part of the multi-enzyme acetyl synthase complex. The red boxes denote two classes of possible ion pumps that hypothetically might be used to drive either energy consuming steps in the pathway (e.g. CO_2 reduction to CO) or energy conserving steps (such as the reduction of Methylene-THF or the methyl group transfer from THF to the CODH/ACS complex).

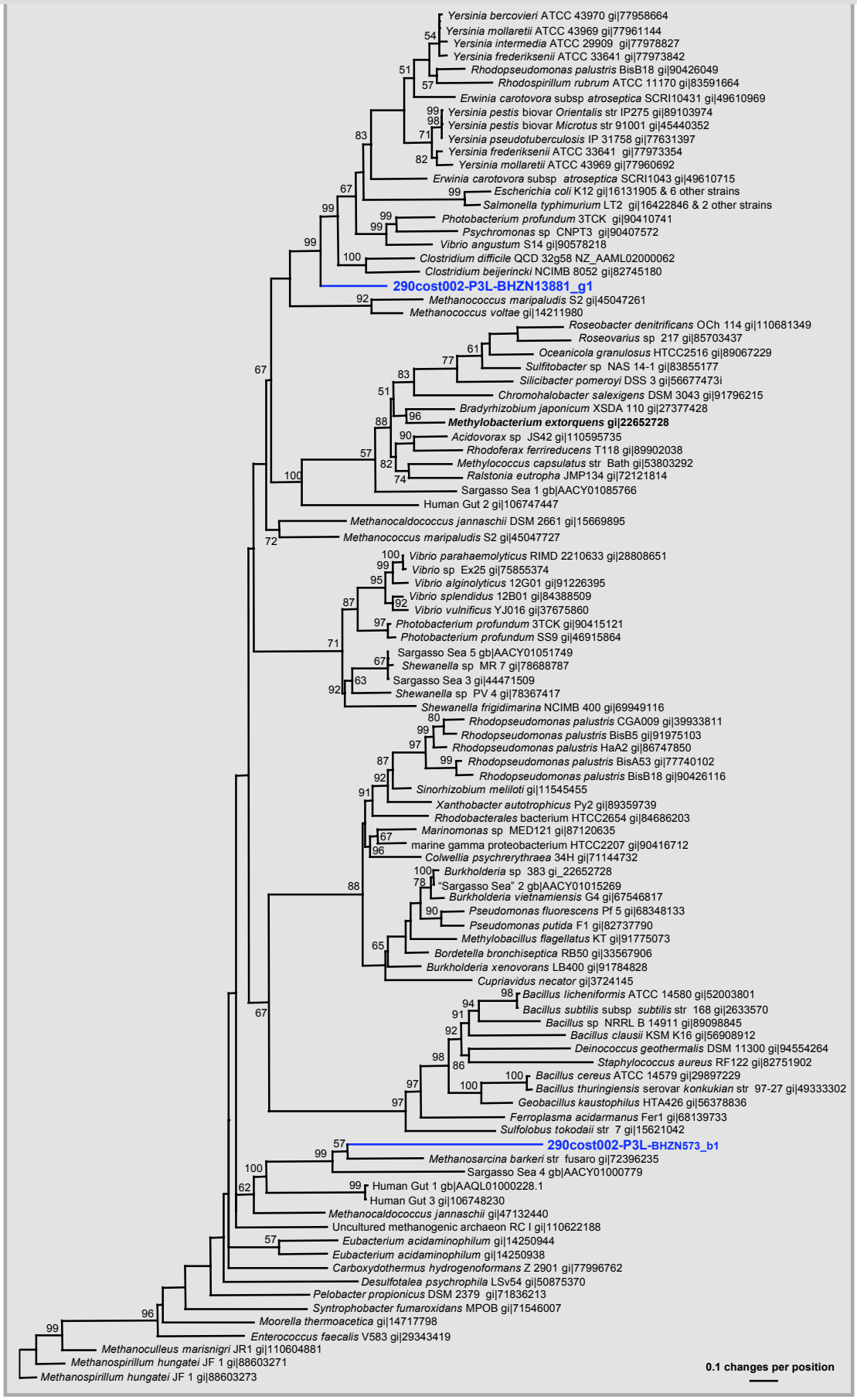


Fig. S20. Phylogeny of the only two putative formate dehydrogenases recovered during *Nasutitermes* P3 Metagenomic analysis. Tree-Puzzle Maximum likelihood analysis was based on 568 unambiguously aligned amino acid positions. One formate dehydrogenase affiliates with bona fide FdhF (FDH H) type homologs from *E. coli* and other enteric *Gammaproteobacteria,* whereas the other affiliates with less well studied variants.

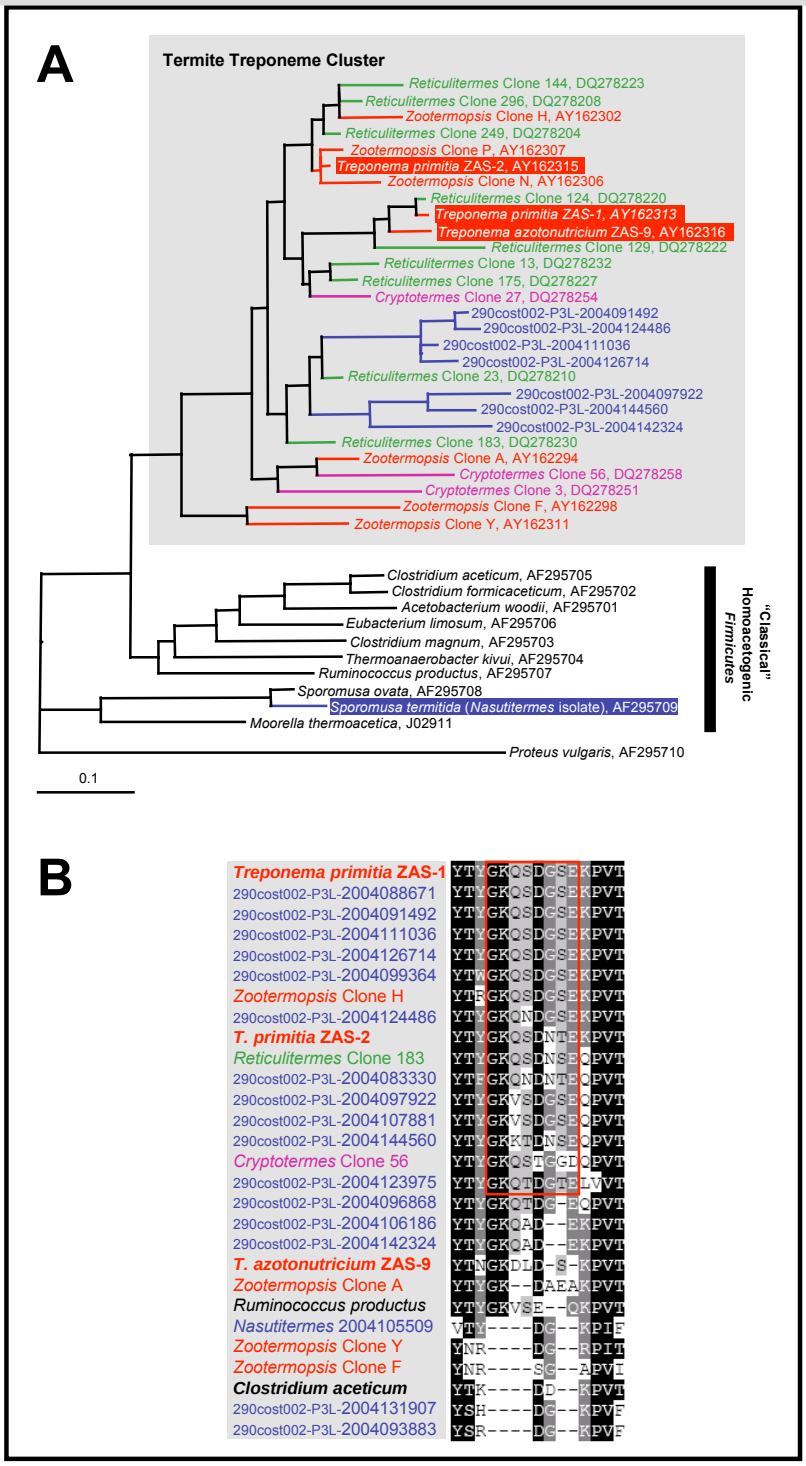


Fig. S21. Metagenomic formyl-tetrahydrofolate synthetases, markers of CO_2-reductive homoacetogenic metabolism, affiliate with known homoacetogenic *Spirochetaetes*, not *Firmicutes*. A) Protein Maximum Likelihood phylogram of homoacetogen-like FTHFS homologs. Analysis based on 235 unambigouosly aligned amino acid positions. Short FTHFS fragments identified in *Nasutitermes* metagenomic libraries were not included in the phylogenetic analysis. Shaded box highlights FTHFS clade of homologs cloned from isolated termite gut homoacetogenic spirochetes and from the gut contents of species corresponding to four of the 6 major termite families. **B)** Variable region of FTHFS**.** Box highlights a short peptide region of the FTHFS protein that is unique to termite gut associated homoacetogenic spirochetes (e.g. isolated *T. primitia* strains ZAS-1 and ZAS-2) and gut derived PCR and metagenomic clones from *Nasutitermes* and termites representing an additional three termite families. A non-acetogenic gut spirochete (e.g. isolated *T. azotonutricium* strain ZAS-9) and "classical" firmicutes homoacetogens do not contain this character. This peptide stretch was not included in the phylogenetic analysis, thus stands in independent support of it. Note that not all of the metagenomic clones shown in B) were used in the phylogenetic analysis for A) as the gene fragments were too short to analyze fully.

Salmassi, T. M., and J. R. Leadbetter. 2003. Analysis of genes of tetrahydrofolate-dependent metabolism from cultivated spirochaetes and the gut community of the termite *Zootermopsis angusticollis*. Microbiology 149:2529-2537.

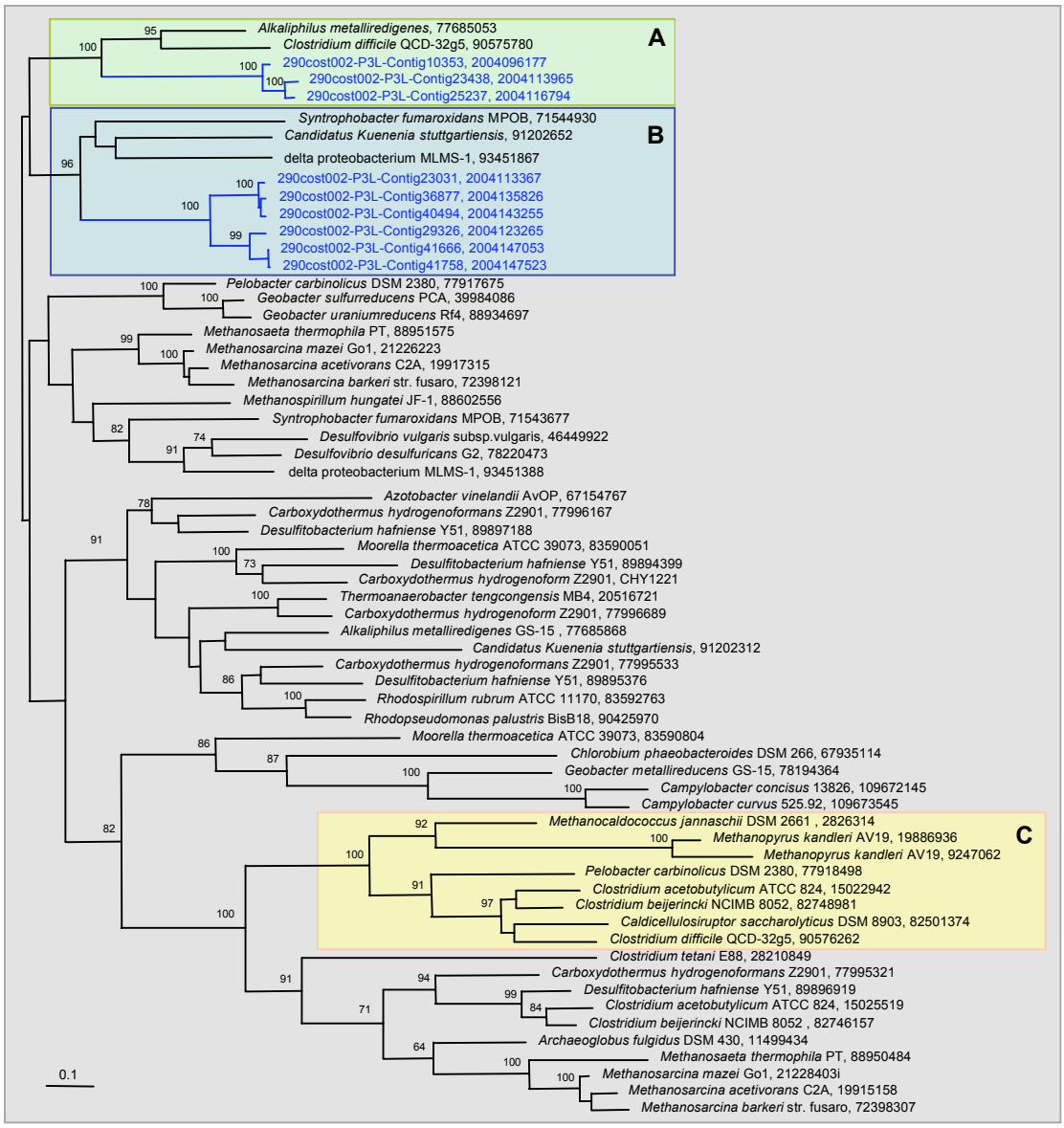


Fig. S22. Phylogeny of Carbon Monoxide Dehydrogenase Catalytic Subunits (CooS) recovered during *Nasutitermes* P3 Metagenomic analysis. The enzymatic activity catalyzed by CooS homologs can be used in either the oxidative or reducitive direction, and in either anabolism or energy metabolism, depending on the *Bacteria* or *Archaea* in question, or their mode of growth. In CO_2-reducing homoacetogens, the enzyme operates in the reductive direction (see Fig. S20). Maximum likelihood analysis was based on 245 unambigouosly aligned amino acid positions. Full length or near full length CooS are highlighted in blue and fell within two clusters, colored boxes A and B. Partial CooS fragments from the metagenomic dataset (not included in this phylogram) affiliated with either cluster A (1 sequence), B (9 sequences) or grouped within colored box C (4 sequences). One short CooS fragment affiliated with the two CooS homologs present in *Campylobacter* genomes (not shown). Values are given for a 1000 bootstrap analysis of the dataset. Bar represents evolutionary distance given as 0.1 changes per amino acid position.

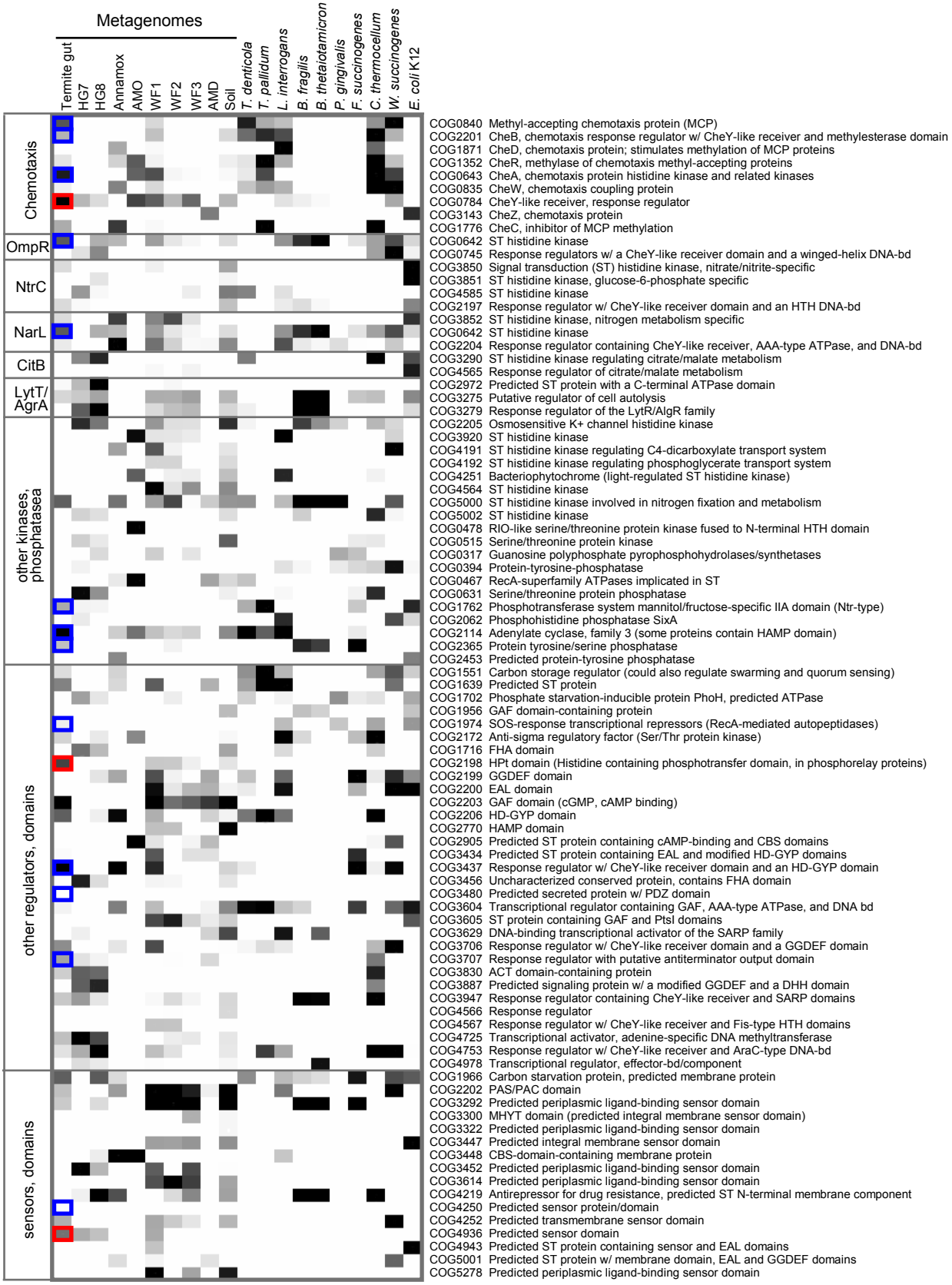


Fig. S23. Abundance distribution of "Signal Transduction Mechanism COG" genes and functional domains among metagenomic datasets and representative single bacterial genomes. To account for the differences in size between the different datasets and genomes, the frequencies were normalized according to a Z-score method in IMG/M. The resulting normalized abundance frequencies were represented as gray scale intensities using Genesis software. Blue rectangles indicate COG categories having a high abundance within the termite gut metagenomic dataset, red rectangles indicate the highest abundance among the analyzed metabiomes and individual genomes.