(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 December 2013 (19.12.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/188305 A2

(51) International Patent Classification:
C12N 9/24 (2006.01) *C12P 19/14* (2006.01)

(21) International Application Number:
PCT/US2013/045007

(22) International Filing Date:
10 June 2013 (10.06.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

61/658,166	11 June 2012 (11.06.2012)	US
61/673,358	19 July 2012 (19.07.2012)	US
61/774,695	8 March 2013 (08.03.2013)	US
61/774,706	8 March 2013 (08.03.2013)	US

(71) Applicant: **CODEXIS, INC.** [US/US]; 200 Penobscot Drive, Redwood City, California 94063 (US).

(72) Inventors: **FONG, Ryan**; 200 Penobscot Drive, Redwood City, California 94063 (US). **ZHANG, Xiyun**; 200 Penobscot Drive, Redwood City, California 94063 (US). **NORIEGA, Chris**; 200 Penobscot Drive, Redwood City, California 94063 (US). **AGARD, Nicholas**; 200 Penobscot Drive, Redwood City, California 94063 (US). **GOHEL, Anupam**; 200 Penobscot Drive, Redwood City, California 94063 (US). **SMITH, Derek**; 200 Penobscot Drive, Redwood City, California 94063 (US).

(74) Agents: **LOCKYER, Jean M.** et al.; Kilpatrick Townsend & Stockton LLP, Two Embarcadero Center, 8th Floor, San Francisco, California 94111 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC,

[Continued on next page]

(54) Title: FUNGAL BETA-XYLOSIDASE VARIANTS

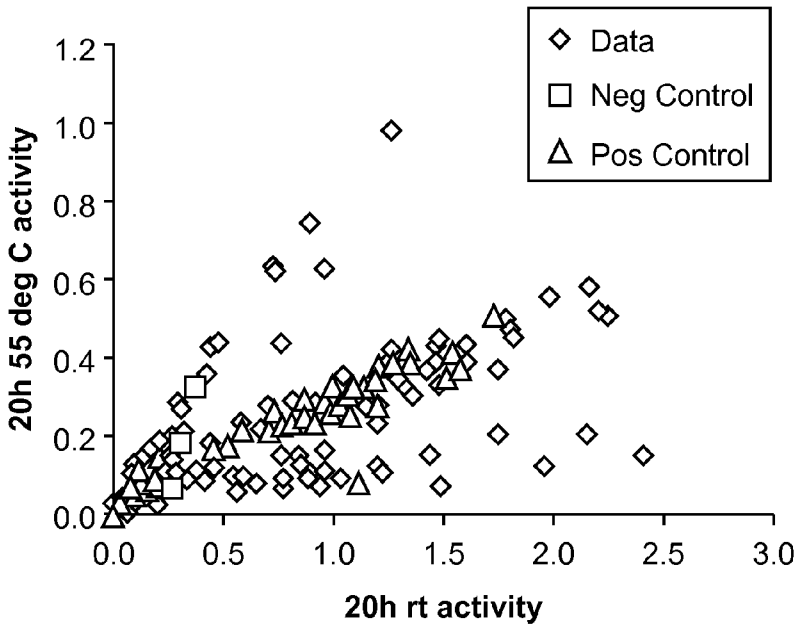

FIG. 3

[Continued on next page]

WO 2013/188305 A2

SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

— *with sequence listing part of description (Rule 5.2(a))*

(57) Abstract: The present invention provides fungal xylanase and/or beta-xylosidase enzymes suitable for use in saccharification reactions. The present application further provides genetically modified fungal organisms that produce xylanase and/or beta-xylosidases, as well as enzyme mixtures exhibiting enhanced hydrolysis of cellulosic material to fermentable sugars, enzyme mixtures produced by the genetically modified fungal organisms, and methods for producing fermentable sugars from cellulose using such enzyme mixtures.

FUNGAL BETA-XYLOSIDASE VARIANTS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims benefit of priority to U.S. provisional application no. 61/774,695, filed March 8, 2013; U.S. provisional application no. 61/774,706 filed March 8, 2103; U.S. provisional application no. 61/673,358, filed July 19, 2012; and U.S. provisional application no. 61/658,166, filed June 11, 2012, the entire contents of each of which are incorporated by reference for all purposes

REFERENCE TO A SEQUENCE LISTING APPENDIX SUBMITTED AS AN ASCII TEXT FILE

[0002] The Sequence Listing written in file 90834-877691_ST25.TXT, created on June 5, 2013, 496,945 bytes, machine format IBM-PC, MS-Windows operating system, is hereby incorporated by reference.

FIELD OF THE INVENTION

[0003] The present invention provides beta-xylosidase variant enzymes suitable for use in saccharification reactions. The present application further provides genetically modified fungal organisms that produce beta-xylosidase variants, as well as enzyme mixtures exhibiting enhanced hydrolysis of cellulosic material to fermentable sugars, enzyme mixtures produced by the genetically modified fungal organisms, and methods for producing fermentable sugars from cellulose using such enzyme mixtures.

BACKGROUND

[0004] Interest has arisen in fermentation of carbohydrate-rich biomass to provide alternatives to petrochemical sources for fuels and organic chemical precursors. There is great interest in using lignocellulosic feedstocks where the plant cellulose is broken down to sugars and subsequently converted to desired end products, such as organic chemical precursors. Lignocellulosic biomass is primarily composed of cellulose, hemicelluloses, and lignin. Cellulose and hemicellulose can be hydrolyzed in a saccharification process to sugars that can be subsequently converted to various products via fermentation. The major fermentable sugars obtained from lignocelluloses are glucose and xylose. For economical

product yields, a process that can effectively convert all the major sugars present in cellulosic feedstock would be highly desirable.

SUMMARY OF THE INVENTION

[0005] The present invention provides beta-xylosidase variant enzymes suitable for use in saccharification reactions. The present application further provides genetically modified fungal organisms that produce beta-xylosidase variants, as well as enzyme mixtures exhibiting enhanced hydrolysis of cellulosic material to fermentable sugars, enzyme mixtures produced by the genetically modified fungal organisms, and methods for producing fermentable sugars from cellulose using such enzyme mixtures.

[0006] The present application further provides genetically modified fungal organisms that produce beta-xylosidase variants, as well as enzyme mixtures exhibiting enhanced hydrolysis of cellulosic material to fermentable sugars, enzyme mixtures produced by the genetically modified fungal organisms, and methods for producing fermentable sugars from cellulose using such enzyme mixtures. In some embodiments, the beta-xylosidase variants are obtained from *Myceliophthora thermophila*.

[0007] In some embodiments, the present application provides recombinant beta-xylosidase variants and/or biologically active fragments of recombinant beta-xylosidase variants comprising at least one amino acid sequence comprising at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% sequence identity to SEQ ID NO:2 and comprising at least one mutation at position 31, 108, 115, 209, 211, 219, 235, 280, 320, 322, 345, 347, 379, 449, 499, 571, 572, 761, 763, and/or 798 , wherein the positions are numbered with reference to SEQ ID NO:2. In some embodiments, the recombinant beta-xylosidase variants and/or biologically active fragments of recombinant beta-xylosidase variants comprise at least one amino acid sequence comprising at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% sequence identity to SEQ ID NO:2 and comprising at least one mutation at position P31, S108, L115, V209, S211, N219, V235, M280, G320, G322, S345, G347, H379, G449, A499, N571, W572, L761,

G763, and/or I798 , wherein the positions are numbered with reference to SEQ ID NO:2. In some additional embodiments, the recombinant beta-xylosidase variants and/or biologically active fragments of recombinant beta-xylosidase variants comprise at least one amino acid sequence comprising at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% sequence identity to SEQ ID NO:2 and comprising at least one mutation at position P31G, S108A, L115I, V209I, S211A, N219Y, V235I, V235L, M280L, G320A, G322A, S345L, G347Q, H379Y, G449N, A499S, A499K, N571G, W572Y, L761I, G763P, and/or I798V, wherein the positions are numbered with reference to SEQ ID NO:2. In some further embodiments, the recombinant beta-xylosidase variants comprise at least one amino acid sequence comprising at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% sequence identity to SEQ ID NO:2, wherein said amino acid sequence comprises SEQ ID NO:5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, and/or 55. In some additional embodiments, the recombinant beta-xylosidase variant or biologically active beta-xylosidase variant fragment of Claim 1, wherein said beta-xylosidase variant is a *Myceliophthora thermophila* beta-xylosidase variant.

[0008] The present invention also provides enzyme compositions comprising at least one beta-xylosidase variant and/or at least one biologically active beta-xylosidase fragment as provided herein. In some embodiments, the enzyme composition further comprises at least one additional enzyme. In some further embodiments, the enzyme composition further comprises one or more enzymes selected from cellulases, hemicellulases, xylanases, amylases, glucoamylases, proteases, esterases, and lipases. In some additional embodiments, the enzyme composition further comprises one or more enzyme(s) selected from endoglucanases (EG), β-glucosidases (BGL), Type 1 cellobiohydrolases (CBH1), Type 2 cellobiohydrolases (CBH2), GH61 enzymes, and/or xylanases.

[0009] The present invention also provides recombinant organisms comprising at least one beta-xylosidase variant and/or at least one biologically active beta-xylosidase variant fragment as provided herein. In some embodiments, the present invention provides recombinant fungal organisms comprising at least one polynucleotide comprising at least one

nucleic acid sequence encoding at least one beta-xylosidase variant and/or biologically active beta-xylosidase fragment, and/or at least one polynucleotide that hybridizes under stringent hybridization conditions to the polynucleotide and/or a complement of a polynucleotide that encodes a polypeptide comprising the amino acid sequence provided herein. In some embodiments, the polynucleotide comprises a sequence that has least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 99%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identity to SEQ ID NOS:1, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, and/or 54.

[0010] The present invention also provides recombinant nucleic acid constructs comprising at least one polynucleotide sequence, wherein the polynucleotide is selected from: a polynucleotide that encodes a polypeptide comprising an amino acid sequence comprising at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identity to SEQ ID NO: 2, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, and/or 55; a polynucleotide that hybridizes under stringent hybridization conditions to at least a fragment of a polynucleotide that encodes a polypeptide comprising the amino acid sequence of SEQ ID NO: 2, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, and/or 55; and/or a polynucleotide that hybridizes under stringent hybridization conditions to the complement of at least a fragment of a polynucleotide that encodes a polypeptide comprising the amino acid sequence of SEQ ID NO: 2, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, and/or 55. In some embodiments, the polynucleotide sequence is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO:1,4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, and/or 54. In some further embodiments, the polynucleotide sequence is operably linked to a promoter. In some additional embodiments, the promoter is a heterologous promoter. In some additional embodiments, the nucleic acid sequence is operably linked to at least one additional regulatory sequence.

[0011] The present invention also provides recombinant host cells that express at least one polynucleotide sequence encoding at least one beta-xylosidase variant and/or biologically

active beta-xylosidase fragment, as provided herein. In some embodiments, host cell comprises at least one nucleic acid construct as provided herein. In some additional embodiments, the host cell comprises at least one polypeptide sequence set forth in SEQ ID NO:5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, and/or 55. In some further embodiments, the host cell comprises at least one polynucleotide sequence set forth in SEQ ID NO:1, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, and/or 54. In some still additional embodiments, at least beta-xylosidase variant and/or at least one biologically active beta-xylosidase fragment is produced by said cell. In some embodiments, the beta-xylosidase variant and/or biologically active beta-xylosidase fragment is secreted from the host cell. In some further embodiments, the host cell further produces at least one enzyme selected from endoglucanases (EG), β-glucosidases (BGL), Type 1 cellobiohydrolases (CBH1), Type 2 cellobiohydrolases (CBH2), GH61 enzymes, and xylanases. In some additional embodiments, the host cell produces at least two recombinant cellulases. In some further embodiments, the recombinant host cell produces at least three, at least four, or at least five recombinant cellulases. In some further embodiments, the recombinant cell is a prokaryotic or eukaryotic cell. In some embodiments, the recombinant host cell is a yeast cell or filamentous fungal cell. In some additional embodiments, the recombinant host cell is a filamentous fungal cell that is a *Myceliophthora*, a *Thielavia*, a *Trichoderma*, or an *Aspergillus* cell. In some alternative embodiments, the recombinant host cell is selected from *Saccharomyces* and *Myceliophthora*. In some further embodiments, the recombinant host is a *Myceliophthora thermophila*, while in some alternative embodiments, the recombinant host cell is *Saccharomyces cerevisiae*.

[0012] In some embodiments, the present invention provides methods for producing at least one fermentable sugar from a feedstock, comprising contacting the feedstock with at least one enzyme composition provided herein, under culture conditions whereby fermentable sugars are produced. In some embodiments, the enzyme composition further comprises at least one enzyme selected from endoglucanases (EG), β-glucosidases (BGL), Type 1 cellobiohydrolases (CBH1), Type 2 cellobiohydrolases (CBH2), GH61 enzymes, and xylanases. In some embodiments, at least one of the further enzymes is a recombinant enzyme. In some additional embodiments, the methods further comprise pretreating the feedstock prior to the contacting step. In some embodiments, the feedstock comprises wheat grass, wheat straw, barley straw, sorghum, rice grass, sugarcane, sugar beet, bagasse, switchgrass, corn stover, corn fiber, grains, or a combination thereof. In some additional

embodiments, the fermentable sugar comprises glucose and/or xylose. In some further embodiments, the methods further comprise recovering at least one fermentable sugar. In some still further embodiments the methods further comprise contacting the at least one fermentable sugar with a microorganism under conditions such that said microorganism produces at least one fermentation end product. In some additional embodiments, the fermentation end product is selected from alcohols, fatty acids, lactic acid, acetic acid, 3-hydroxypropionic acid, acrylic acid, succinic acid, citric acid, malic acid, fumaric acid, succinic acid, amino acids, 1,3-propanediol, ethylene, glycerol, and β-lactams. In some further embodiments, the fermentation product is an alcohol selected from ethanol and butanol. In some embodiments, alcohol is ethanol. In some additional embodiments, the feedstock is a cellulosic and/or lignocellulosic feedstock.

[0013] The present invention also provides methods of producing an end product from a feedstock, comprising: contacting the feedstock with at least one enzyme composition provided herein, under conditions whereby at least one fermentable sugar is produced from the substrate; and contacting the fermentable sugar with a microorganism under conditions such that the microorganism uses the fermentable sugar to produce an end-product. In some embodiments, the methods comprise simultaneous saccharification and fermentation reactions (SSF), while in some alternative embodiments, the methods comprise separate saccharification and fermentation reactions (SHF). In some embodiments, the feedstock is a cellulosic and/or lignocellulosic feedstock.

[0014] The present invention also provides methods of producing a fermentation end product from a feedstock, comprising: obtaining at least one fermentable sugar produced according to any method provided herein; and contacting the fermentable sugar with a microorganism in a fermentation to produce at least one fermentation end product. In some embodiments, the fermentation end product is selected from alcohols, fatty acids, lactic acid, acetic acid, 3-hydroxypropionic acid, acrylic acid, succinic acid, citric acid, malic acid, fumaric acid, succinic acid, amino acids, 1,3-propanediol, ethylene, glycerol, and β-lactams. In some additional embodiments, the fermentation end product is at least one alcohol selected from ethanol and butanol. In some further embodiments, the microorganism is a yeast. In some further embodiments, the methods further comprise recovering the fermentation end product.

[0015] The present invention also provides methods for producing at least one fermentable sugar from a feedstock, comprising contacting the feedstock with at least one recombinant

beta xylosidase and/or at least one biologically active beta-xylosidase fragment provided herein, and/or at least one enzyme composition provided herein, and/or at least one recombinant host cell as provided herein, under culture conditions whereby fermentable sugars are produced. In some embodiments, the enzyme composition and/or recombinant host cell further comprises at least one enzyme selected from endoglucanases (EG), β-glucosidases (BGL), Type 1 cellobiohydrolases (CBH1), Type 2 cellobiohydrolases (CBH2), GH61s, and xylanases. In some embodiments, at least one of the further enzymes is a recombinant enzyme. In some embodiments, at least one further enzyme is a heterologous enzyme. In some embodiments, the methods further comprise pretreating the feedstock prior to the contacting step. In some additional embodiments, the feedstock comprises wheat grass, wheat straw, barley straw, sorghum, rice grass, sugarcane, sugar beet, bagasse, switchgrass, corn stover, corn fiber, grains, or a combination thereof. In some further embodiments, the fermentable sugar comprises glucose and/or xylose. In some further embodiments, the methods further comprise recovering at least one fermentable sugar. In still some additional embodiments, the methods further comprise contacting at least one fermentable sugar with a microorganism under conditions such that the microorganism produces at least one fermentation end product. In some embodiments, the fermentation end product is selected from alcohols, fatty acids, lactic acid, acetic acid, 3-hydroxypropionic acid, acrylic acid, succinic acid, citric acid, malic acid, fumaric acid, succinic acid, amino acids, 1,3-propanediol, ethylene, glycerol, and β-lactams. In some further embodiments, the fermentation product is an alcohol selected from ethanol and butanol. In some further additional embodiments, the alcohol is ethanol. In some additional embodiments, the feedstock is a cellulosic and/or lignocellulosic feedstock.

[0016] The present invention also provides methods of producing an end product from a feedstock, comprising: contacting the feedstock with at least one recombinant beta xylosidase and/or at least one biologically active beta-xylosidase fragment provided herein, and/or an at least one enzyme composition provided herein, and/or at least one recombinant host cell provided herein, under conditions whereby at least one fermentable sugar is produced from the substrate; and contacting the fermentable sugar with a microorganism under conditions such that the microorganism uses the fermentable sugar to produce an end-product. In some embodiments, the enzyme composition and/or recombinant host cell further comprises at least one enzyme selected from endoglucanases (EG), β-glucosidases (BGL), Type 1 cellobiohydrolases (CBH1), Type 2 cellobiohydrolases (CBH2), GH61 enzymes, and

xylanases. In some embodiments, at least one of the further enzymes is a recombinant enzyme. In some additional embodiments, at least one of the further enzymes is a heterologous enzyme. In some embodiments, the methods comprise a simultaneous saccharification and fermentation reactions (SSF), while in some alternative embodiments, the methods comprise separate saccharification and fermentation reactions (SHF). In some embodiments, the feedstock is a cellulosic and/or lignocellulosic feedstock.

[0017] The present invention also provides methods of producing a fermentation end product from a feedstock, comprising: obtaining at least one fermentable sugar produced according to at least one method provided herein; and contacting the fermentable sugar with a microorganism in a fermentation to produce at least one fermentation end product. In some embodiments, the fermentation end product is selected from alcohols, fatty acids, lactic acid, acetic acid, 3-hydroxypropionic acid, acrylic acid, succinic acid, citric acid, malic acid, fumaric acid, succinic acid, amino acids, 1,3-propanediol, ethylene, glycerol, and β-lactams. In some further embodiments, the fermentation end product is at least one alcohol selected from ethanol and butanol. In some additional embodiments, the microorganism is a yeast. In some further embodiments, the methods further comprise the step of recovering the fermentation end product.

[0018] The present invention provides recombinant beta-xylosidase variants and/or biologically active recombinant beta-xylosidase variant fragments comprising at least one amino acid sequence comprising at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or at least about 100% sequence identity to SEQ ID NO:2 and comprising at least one mutation at position 31, 108, 115, 174, 177, 203, 209, 211, 219, 235, 264, 280, 309, 320, 322, 345, 347, 375, 379, 389, 394, 398, 431, 438, 449, 475, 482, 484, 499, 525, 539, 560, 565, 571, 572, 589, 662, 727, 761, 763, 798, and/or 842, wherein the positions are numbered with reference to SEQ ID NO:2. In some embodiments, the recombinant beta-xylosidase variants and/or biologically active recombinant beta-xylosidase variant fragments comprise at least one amino acid sequence comprising at least about 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID NO:2 and comprising at least one mutation at position 31, 108, 115, 174, 177, 203, 209, 211, 219, 235, 264, 280, 309, 320, 322, 345, 347, 375, 379, 389,

394, 398, 431, 438, 449, 475, 482, 484, 499, 525, 539, 560, 565, 571, 572, 589, 662, 727, 761, 763, 798, and/or 842, wherein the positions are numbered with reference to SEQ ID NO:2. In some embodiments, the recombinant beta-xylosidase variants and/or biologically active recombinant beta-xylosidase variant fragments comprise at least one mutation at position 31, 108, 115, 174, 177, 203, 209, 211, 219, 235, 264, 280, 309, 320, 322, 345, 347, 375, 379, 389, 394, 398, 431, 438, 449, 475, 482, 484, 499, 525, 539, 560, 565, 571, 572, 589, 662, 727, 761, 763, 798, and/or 842, wherein the positions are numbered with reference to SEQ ID NO:2.

[0019] In some additional embodiments, the recombinant beta-xylosidase variant and/or biologically active recombinant beta-xylosidase variant fragments comprise: (i) at least one amino acid sequence comprising at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or at least about 100% sequence identity to SEQ ID NO:2 and comprising at least one mutation at position P31, S108, L115, V174, L177, G203, V209, S211, N219, V235, A264, M280, A309, G320, G322, S345, G347, A354, P375, H379, R389, E394, R398, R431, F438, G449, G475, D482, D484, A499, G525, R539, E560, G565, N571, N572, W572, E589, D662, E727, L761, G763, I798 , and/or G842, wherein the positions are numbered with reference to SEQ ID NO:2; (ii) at least one amino acid sequence comprising at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity to SEQ ID NO:2 and comprising at least one mutation at position P31G, S108A, L115I, V174P, L177G, G203C, V209I, S211A, N219Y, V235I, V235L, V235R, A264S, M280L, A309L, G320A, G322A, S345L, G347Q, A354V, P375E, P375S, H379Y, R389T, E394L, R398N, R431W, F438P, G449N, G475T, D482G, D484P, A499S, A499K, G525R, R539G, R539H, R539Q, R539S, R560D, G565N, N571G, W572Y, R589K, D662N, E727D, E727T, L761I, G763P, I798V, and/or G842A, wherein the positions are numbered with reference to SEQ ID NO:2; or (iii) at least one amino acid sequence comprising at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity to SEQ ID NO:2, wherein said amino

acid sequence comprises SEQ ID NO:5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, and/or 59.

[0020] In some further embodiments, the recombinant beta-xylosidase variant and/or biologically active recombinant beta-xylosidase variant fragments comprise: (i) at least one amino acid sequence comprising at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID NO:2 and comprising at least one mutation at position P31, S108, L115, V174, L177, G203, V209, S211, N219, V235, A264, M280, A309, G320, G322, S345, G347, A354, P375, H379, R389, E394, R398, R431, F438, G449, G475, D482, D484, A499, G525, R539, E560, G565, N571, N572, W572, E589, D662, E727, L761, G763, I798 , and/or G842, wherein the positions are numbered with reference to SEQ ID NO:2; (ii) at least one amino acid sequence comprising at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID NO:2 and comprising at least one mutation at position P31G, S108A, L115I, V174P, L177G, G203C, V209I, S211A, N219Y, V235I, V235L, V235R, A264S, M280L, A309L, G320A, G322A, S345L, G347Q, A354V, P375E, P375S, H379Y, R389T, E394L, R398N, R431W, F438P, G449N, G475T, D482G, D484P, A499S, A499K, G525R, R539G, R539H, R539Q, R539S, R560D, G565N, N571G, W572Y, R589K, D662N, E727D, E727T, L761I, G763P, I798V, and/or G842A, wherein the positions are numbered with reference to SEQ ID NO:2; or (iii) at least one amino acid sequence comprising at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 100% sequence identity to SEQ ID NO:2, wherein said amino acid sequence comprises SEQ ID NO:5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, and/or 59. In some embodiments, the recombinant beta-xylosidase variant or biologically active beta-xylosidase variant are *Myceliophthora thermophila* beta-xylosidase variant or biologically active beta-xylosidase variant fragment.

[0021] The present invention also provides enzyme compositions comprising at least one beta-xylosidase variant and/or at least one biologically active beta-xylosidase variant fragment provided herein, and optionally further comprising: (i) at least one additional enzyme; (ii) one or more enzymes selected from cellulases, hemicellulases, xylanases,

amylases, glucoamylases, proteases, esterases, and lipases; and/or (iii) one or more enzyme(s) selected from endoglucanases (EG), β-glucosidases (BGL), Type 1 cellobiohydrolases (CBH1), Type 2 cellobiohydrolases (CBH2), GH61 enzymes, and/or xylanases. In some embodiments, the enzyme compositions comprise at least one polypeptide sequence selected from SEQ ID NOS:2, 3, 57, 61, 62, 64, 65, 67, 68, 70, 71, 73, 74, 76, 77, 79, 80, 82, 83, 85, 86, 88, 89, 91, 92, 94, 95, 97, 98, 100, 101, 103, 104, 106, 107, 109, 111, 113, 115, and/or 117; and/or at least one polypeptide sequence encoded by at least one polynucleotide sequence selected from SEQ ID NO:1, 56, 60, 63, 66, 69, 72, 75, 78, 81, 84, 87, 90, 93, 96, 99, 102, 105, 108, 110, 112, 114, and/or 116.

[0022] The present invention further provides recombinant organisms comprising at least one beta-xylosidase variant and/or at least one biologically active beta-xylosidase variant fragment as provided herein.

[0023] The present invention also provides recombinant fungal organisms comprising at least one polynucleotide comprising at least one nucleic acid sequence encoding at least one beta-xylosidase variant and/or at least one biologically active fragment as provided herein, and/or at least one polynucleotide that hybridizes under stringent hybridization conditions to the polynucleotide and/or a complement of a polynucleotide that encodes a polypeptide comprising the amino acid sequence provided herein, optionally wherein said polynucleotide comprises a sequence that has least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identity to SEQ ID NOS:1, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, and/or 58. In some embodiments, the recombinant fungal organisms comprise at least one polynucleotide comprising at least one nucleic acid sequence encoding at least one beta-xylosidase variant and/or at least one biologically active fragment as provided herein, and/or at least one polynucleotide that hybridizes under stringent hybridization conditions to the polynucleotide and/or a complement of a polynucleotide that encodes a polypeptide comprising the amino acid sequence provided herein, optionally wherein said polynucleotide comprises a sequence that has least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at at least 100% identity to SEQ ID NOS:1, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, and/or 58. In some additional

embodiments, the recombinant fungal organisms comprise at least one polynucleotide comprising at least one nucleic acid sequence encoding at least one beta-xylosidase variant and/or at least one biologically active fragment as provided herein, and/or at least one polynucleotide that hybridizes under stringent hybridization conditions to the polynucleotide and/or a complement of a polynucleotide that encodes a polypeptide comprising the amino acid sequence provided herein, optionally wherein said polynucleotide comprises at least one sequence selected from SEQ ID NOS:1, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, and/or 58.

[0024] The present invention also provides recombinant nucleic acid constructs comprising at least one polynucleotide sequence, wherein the polynucleotide is selected from: (a) a polynucleotide that encodes a polypeptide comprising an amino acid sequence comprising at least about 70%, at least about 75%, at least about 80%, at least about 81%, at least about 82%, at least about 83%, at least about 84%, at least about 85%, at least about 86%, at least about 87%, at least about 88%, at least about 89%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or at least about 100% identity to SEQ ID NO: 2, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, and/or 59; (b) a polynucleotide that hybridizes under stringent hybridization conditions to at least a fragment of a polynucleotide that encodes a polypeptide comprising the amino acid sequence of SEQ ID NO: 2, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, and/or 59; and/or (c) a polynucleotide that hybridizes under stringent hybridization conditions to the complement of at least a fragment of a polynucleotide that encodes a polypeptide comprising the amino acid sequence of SEQ ID NO: 2, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55 and/or 59. In some embodiments, the present invention also provides recombinant nucleic acid constructs comprising at least one polynucleotide sequence, wherein the polynucleotide is selected from: (a) a polynucleotide that encodes a polypeptide comprising an amino acid sequence comprising at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 2, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, and/or 59; (b) a polynucleotide that hybridizes under stringent

hybridization conditions to at least a fragment of a polynucleotide that encodes a polypeptide comprising the amino acid sequence of SEQ ID NO: 2, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, and/or 59; and/or (c) a polynucleotide that hybridizes under stringent hybridization conditions to the complement of at least a fragment of a polynucleotide that encodes a polypeptide comprising the amino acid sequence of SEQ ID NO: 2, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55 and/or 59.

[0025] The present invention also provides recombinant nucleic acid constructs comprising at least one polynucleotide sequence, wherein the polynucleotide is selected from: (a) a polynucleotide that encodes a polypeptide comprising an amino acid sequence comprising at least about 70%, at least about 75%, at least about 80%, at least about 81%, at least about 82%, at least about 83%, at least about 84%, at least about 85%, at least about 86%, at least about 87%, at least about 88%, at least about 89%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or at least about 100% identity to SEQ ID NO: 2, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, and/or 59; (b) a polynucleotide that hybridizes under stringent hybridization conditions to at least a fragment of a polynucleotide that encodes a polypeptide comprising the amino acid sequence of SEQ ID NO: 2, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, and/or 59; and/or (c) a polynucleotide that hybridizes under stringent hybridization conditions to the complement of at least a fragment of a polynucleotide that encodes a polypeptide comprising the amino acid sequence of SEQ ID NO: 2, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55 and/or 59. In some additional embodiments, the present invention also provides recombinant nucleic acid constructs comprising at least one polynucleotide sequence, wherein the polynucleotide is selected from: (a) a polynucleotide that encodes a polypeptide comprising an amino acid sequence comprising at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 2, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, and/or 59; (b) a polynucleotide that hybridizes under stringent hybridization conditions to at least a fragment of a polynucleotide that encodes a

polypeptide comprising the amino acid sequence of SEQ ID NO: 2, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, and/or 59; and/or (c) a polynucleotide that hybridizes under stringent hybridization conditions to the complement of at least a fragment of a polynucleotide that encodes a polypeptide comprising the amino acid sequence of SEQ ID NO: 2, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55 and/or 59. In some additional embodiments, the present invention also provides recombinant nucleic acid constructs comprising at least one polynucleotide sequence, wherein the polynucleotide is selected from: (a) a polynucleotide that encodes a polypeptide comprising an amino acid sequence comprising at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at 100% identity to SEQ ID NO: 2, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, and/or 59; (b) a polynucleotide that hybridizes under stringent hybridization conditions to at least a fragment of a polynucleotide that encodes a polypeptide comprising the amino acid sequence of SEQ ID NO: 2, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, and/or 59; and/or (c) a polynucleotide that hybridizes under stringent hybridization conditions to the complement of at least a fragment of a polynucleotide that encodes a polypeptide comprising the amino acid sequence of SEQ ID NO: 2, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55 and/or 59.

[0026] The present invention also provides recombinant nucleic acid constructs comprising at least one polynucleotide sequence, wherein the polynucleotide is selected from: (a) a polynucleotide that encodes a polypeptide comprising an amino acid sequence comprising at least about 70%, at least about 75%, at least about 80%, at least about 81%, at least about 82%, at least about 83%, at least about 84%, at least about 85%, at least about 86%, at least about 87%, at least about 88%, at least about 89%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identity to SEQ ID NO: 2, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, and/or 59; (b) a polynucleotide that hybridizes under stringent hybridization conditions to at least a fragment of a polynucleotide that encodes a polypeptide comprising the amino acid sequence of SEQ ID NO: 2, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23,

25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, and/or 59; and/or (c) a polynucleotide that hybridizes under stringent hybridization conditions to the complement of at least a fragment of a polynucleotide that encodes a polypeptide comprising the amino acid sequence of SEQ ID NO: 2, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55 and/or 59. In some embodiments, the recombinant nucleic acid constructs comprise at least one polynucleotide sequence, wherein the polynucleotide is selected from: (a) a polynucleotide that encodes a polypeptide comprising an amino acid sequence comprising at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 2, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, and/or 59; (b) a polynucleotide that hybridizes under stringent hybridization conditions to at least a fragment of a polynucleotide that encodes a polypeptide comprising the amino acid sequence of SEQ ID NO: 2, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, and/or 59; and/or (c) a polynucleotide that hybridizes under stringent hybridization conditions to the complement of at least a fragment of a polynucleotide that encodes a polypeptide comprising the amino acid sequence of SEQ ID NO: 2, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55 and/or 59.

[0027] In some further embodiments of the recombinant nucleic acid constructs provided herein, (i) the polynucleotide sequence is at least about 70%, at least about 75%, at least about 80%, at least about 81%, at least about 82%, at least about 83%, at least about 84%, at least about 85%, at least about 86%, at least about 87%, at least about 88%, at least about 89%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or at least about 100% identical to SEQ ID NO:1,4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, and/or 58; (ii) the polynucleotide sequence is operably linked to a promoter, optionally wherein said promoter is a heterologous promoter; and/or (iii) the nucleic acid sequence is operably linked to at least one additional regulatory sequence. In some still further embodiments of the recombinant nucleic acid constructs provided herein, (i) the polynucleotide sequence is at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at

least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO:1,4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, and/or 58; (ii) the polynucleotide sequence is operably linked to a promoter, optionally wherein said promoter is a heterologous promoter; and/or (iii) the nucleic acid sequence is operably linked to at least one additional regulatory sequence.

[0028] The present invention also provides recombinant host cells that expresses at least one polynucleotide sequence encoding at least one beta-xylosidase variant and/or at least one biologically active beta-xylosidase fragment as provided herein. In some embodiments, the (i) host cell comprises at least one nucleic acid construct as provided herein; (ii) the host cell comprises at least one polypeptide sequence set forth in SEQ ID NO: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, and/or 59; (iii) the host cell comprises at least one polynucleotide sequence set forth in SEQ ID NO: 1,4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, and/or 58; (iv) at least beta-xylosidase variant and/or at least one biologically active beta-xylosidase fragment is produced by said cell, optionally wherein at least one beta-xylosidase variant and/or at least one biologically active beta-xylosidase fragment is secreted from the host cell; (v) said host cell further produces at least one enzyme selected from endoglucanases (EG), β-glucosidases (BGL), Type 1 cellobiohydrolases (CBH1), Type 2 cellobiohydrolases (CBH2), GH61 enzymes, and xylanases; (vi) said cell produces at least two recombinant cellulases; (vii) said cell produces at least three, at least four, or at least five recombinant cellulases; (viii) said cell is a prokaryotic or eukaryotic cell, optionally wherein said cell is a yeast cell or filamentous fungal cell, such as wherein the filamentous fungal cell is a *Myceliophthora*, a *Thielavia*, a *Trichoderma*, and/or an *Aspergillus* cell; and/or (ix) said cell is selected from *Saccharomyces* and *Myceliophthora*, optionally wherein the filamentous fungal cell is a *Myceliophthora thermophila* or wherein the yeast cell is *Saccharomyces cerevisiae*.

[0029] The present invention also provides methods for producing at least one fermentable sugar from a feedstock, comprising contacting the feedstock with at least one recombinant beta xylosidase and/or biologically active beta xylosidase fragment, as provided herein, and/or a recombinant host cell as provided herein, and/or an enzyme composition as provided herein, under culture conditions whereby fermentable sugars are produced. In some embodiments of the methods the enzyme composition and/or recombinant host cell further

comprises at least one enzyme selected from endoglucanases (EG), β-glucosidases (BGL), Type 1 cellobiohydrolases (CBH1), Type 2 cellobiohydrolases (CBH2), GH61s, and xylanases, such as wherein said at least one enzyme is a recombinant enzyme and/or wherein said at least one enzyme is a heterologous enzyme; further comprise pretreating the feedstock prior to said contacting; wherein the feedstock comprises wheat grass, wheat straw, barley straw, sorghum, rice grass, sugarcane, sugar beet, bagasse, switchgrass, corn stover, corn fiber, grains, or a combination thereof; (iv) wherein the fermentable sugar comprises glucose and/or xylose; (v) further comprise recovering at least one fermentable sugar; (vi) further comprise contacting the at least one fermentable sugar with a microorganism under conditions such that said microorganism produces at least one fermentation end product, optionally wherein said fermentation end product is selected from alcohols, fatty acids, lactic acid, acetic acid, 3-hydroxypropionic acid, acrylic acid, succinic acid, citric acid, malic acid, fumaric acid, succinic acid, amino acids, 1,3-propanediol, ethylene, glycerol, and β-lactams, such as wherein said fermentation product is an alcohol selected from ethanol and butanol, preferably wherein said alcohol is ethanol; and/or wherein the feedstock is a cellulosic and/or lignocellulosic feedstock.

[0030] The present invention also provides methods of producing an end product from a feedstock, comprising: a) contacting the feedstock with at least one recombinant beta-xylosidase and/or at least one biologically active beta-xylosidase fragment as provided herein and/or a recombinant host cell as provided herein, and/or an enzyme composition as provided herein, under conditions whereby at least one fermentable sugar is produced from the substrate; and b) contacting the fermentable sugar with a microorganism under conditions such that the microorganism uses the fermentable sugar to produce an end-product. In some additional embodiments, the recombinant organism and/or recombinant host cell further comprises at least one enzyme selected from endoglucanases (EG), β-glucosidases (BGL), Type 1 cellobiohydrolases (CBH1),Type 2 cellobiohydrolases (CBH2), GH61 enzymes, and xylanases; at least one enzyme is a recombinant enzyme and/or said at least one enzyme is a heterologous enzyme; the method comprises a simultaneous saccharification and fermentation reactions (SSF) or wherein the method comprises separate saccharification and fermentation reactions (SHF); and/or the feedstock is a cellulosic and/or lignocellulosic feedstock.

[0031] The present invention further provides methods of producing a fermentation end product from a feedstock, comprising: a) obtaining at least one fermentable sugar produced

according to a method provided herein; and b) contacting the fermentable sugar with a microorganism in a fermentation to produce at least one fermentation end product, optionally wherein: (i) the fermentation end product is selected from alcohols, fatty acids, lactic acid, acetic acid, 3-hydroxypropionic acid, acrylic acid, succinic acid, citric acid, malic acid, fumaric acid, succinic acid, amino acids, 1,3-propanediol, ethylene, glycerol, and β-lactams, such as wherein said fermentation end product is at least one alcohol selected from ethanol and butanol; (ii) wherein the microorganism is a yeast; and/or (iii) further comprising recovering the fermentation end product.

[0032] The present invention also provides the following further embodiments:

1. A recombinant beta-xylosidase variant and/or biologically active recombinant beta-xylosidase variant fragment comprising at least one amino acid sequence comprising at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% sequence identity to SEQ ID NO:2 and comprising at least one mutation at position 31, 108, 115, 209, 211, 219, 235, 280, 320, 322, 345, 347, 379, 449, 499, 571, 572, 761, 763, and/or 798 , wherein the positions are numbered with reference to SEQ ID NO:2.

2. The recombinant beta-xylosidase variant and/or biologically active recombinant beta-xylosidase variant fragment of Embodiment 1, comprising at least one amino acid sequence comprising at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% sequence identity to SEQ ID NO:2 and comprising at least one mutation at position P31, S108, L115, V209, S211, N219, V235, M280, G320, G322, S345, G347, H379, G449, A499, N571, W572, L761, G763, and/or I798 , wherein the positions are numbered with reference to SEQ ID NO:2.

3. The recombinant beta-xylosidase variant and/or biologically active recombinant beta-xylosidase variant fragment of Embodiment 1, comprising at least one amino acid sequence comprising at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% sequence identity to SEQ ID NO:2 and comprising at least one mutation at position P31G, S108A, L115I, V209I, S211A, N219Y, V235I, V235L, M280L, G320A,

G322A, S345L, G347Q, H379Y, G449N, A499S, A499K, N571G, W572Y, L761I, G763P, and/or I798V, wherein the positions are numbered with reference to SEQ ID NO:2.

4. The recombinant beta-xylosidase variant and/or biologically active recombinant beta-xylosidase variant fragment of Embodiment 1, comprising at least one amino acid sequence comprising at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% sequence identity to SEQ ID NO:2, wherein said amino acid sequence comprises SEQ ID NO:5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, and/or 55.

5. The recombinant beta-xylosidase variant or biologically active beta-xylosidase variant fragment of Embodiment 1, wherein said beta-xylosidase variant is a *Myceliophthora thermophila* beta-xylosidase variant.

6. An enzyme composition comprising at least one beta-xylosidase variant and/or at least one biologically active beta-xylosidase variant fragment of any of Embodiments 1-4.

7. The enzyme composition of Embodiment 6, further comprising at least one additional enzyme.

8. The enzyme composition of Embodiment 6 and/or 7, further comprising one or more enzymes selected from cellulases, hemicellulases, xylanases, amylases, glucoamylases, proteases, esterases, and lipases.

9. The enzyme composition of any of Embodiments 6-8, further comprising one or more enzyme(s) selected from endoglucanases (EG), β-glucosidases (BGL), Type 1 cellobiohydrolases (CBH1), Type 2 cellobiohydrolases (CBH2), GH61 enzymes, and/or xylanases.

10. A recombinant organism comprising at least one beta-xylosidase variant and/or at least one biologically active beta-xylosidase variant fragment of any of Embodiments 1-4.

11. A recombinant fungal organism comprising at least one polynucleotide comprising at least one nucleic acid sequence encoding at least one beta-xylosidase variant and/or at least one biologically active fragment of any of Embodiments 1-4, and/or at least one polynucleotide that hybridizes under stringent hybridization conditions to the polynucleotide and/or a complement of a polynucleotide that encodes a polypeptide comprising the amino acid sequence provided in any of Embodiments 1-4.

12. The polynucleotide of Embodiment 11, wherein said polynucleotide comprises a sequence that has least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least

91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identity to SEQ ID NOS:1, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, and/or 54.

13. A recombinant nucleic acid construct comprising at least one polynucleotide sequence, wherein the polynucleotide is selected from:

(a) a polynucleotide that encodes a polypeptide comprising an amino acid sequence comprising at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identity to SEQ ID NO: 2, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, and/or 55;

(b) a polynucleotide that hybridizes under stringent hybridization conditions to at least a fragment of a polynucleotide that encodes a polypeptide comprising the amino acid sequence of SEQ ID NO: 2, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, and/or 55; and/or

(c) a polynucleotide that hybridizes under stringent hybridization conditions to the complement of at least a fragment of a polynucleotide that encodes a polypeptide comprising the amino acid sequence of SEQ ID NO: 2, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, and/or 55.

14. The recombinant nucleic acid construct of Embodiment 13, wherein the polynucleotide sequence is at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO:1,4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, and/or 54.

15. The nucleic acid construct of Embodiment 13 and/or 14, wherein the polynucleotide sequence is operably linked to a promoter.

16. The nucleic acid construct of Embodiment 15, wherein said promoter is a heterologous promoter.

17. The nucleic acid construct of any of Embodiments 13-16, wherein said nucleic acid sequence is operably linked to at least one additional regulatory sequence.

18. A recombinant host cell that expresses at least one polynucleotide sequence encoding at least one beta-xylosidase variant and/or at least one biologically active beta-xylosidase fragment of any of Embodiments 1-14.

19. The recombinant host cell of Embodiment 18, wherein said host cell comprises at least one nucleic acid construct as provided in any of Embodiments 13-16.

20. The recombinant host cell of Embodiment 18 or 19, wherein said host cell comprises at least one polypeptide sequence set forth in SEQ ID NO: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, and/or 55.

21. The recombinant host cell of any of Embodiments 18-20, wherein said host cell comprises at least one polynucleotide sequence set forth in SEQ ID NO: 1,4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, and/or 54.

22. The recombinant host cell of any of Embodiments 18-21, wherein at least beta-xylosidase variant and/or at least one biologically active beta-xylosidase fragment is produced by said cell.

23. The recombinant host cell of Embodiment 22, wherein at least one beta-xylosidase variant and/or at least one biologically active beta-xylosidase fragment is secreted from the host cell.

24. The recombinant host cell of any of Embodiments 18-23, wherein said host cell further produces at least one enzyme selected from endoglucanases (EG), β-glucosidases (BGL), Type 1 cellobiohydrolases (CBH1), Type 2 cellobiohydrolases (CBH2), GH61 enzymes, and xylanases.

25. The recombinant host cell of any of Embodiments 18-24, wherein said cell produces at least two recombinant cellulases.

26. The recombinant host cell of any of Embodiments 18-25, wherein said cell produces at least three, at least four, or at least five recombinant cellulases.

27. The recombinant cell of any of Embodiments 18-26, wherein said cell is a prokaryotic or eukaryotic cell.

28. The recombinant host cell of Embodiment 27, wherein said cell is a yeast cell or filamentous fungal cell.

29. The recombinant host cell of Embodiment 27 or 28, wherein the filamentous fungal cell is a *Myceliophthora*, a *Thielavia*, a *Trichoderma*, or an *Aspergillus* cell.

30. The recombinant host cell of any of Embodiments 18-28, wherein said cell is selected from *Saccharomyces* and *Myceliophthora*.

31. The recombinant host cell of Embodiment 30, wherein the filamentous fungal cell is a *Myceliophthora thermophila*.

32. The recombinant host cell of Embodiment 30, wherein the yeast cell is *Saccharomyces cerevisiae*.

33. A method for producing at least one fermentable sugar from a feedstock, comprising contacting the feedstock with at least one enzyme composition according to any of Embodiments 6 to 9, under culture conditions whereby fermentable sugars are produced.

34. The method of Embodiment 33, wherein the enzyme composition further comprises at least one enzyme selected from endoglucanases (EG), β-glucosidases (BGL), Type 1 cellobiohydrolases (CBH1), Type 2 cellobiohydrolases (CBH2), GH61 enzymes, and xylanases.

35. The method of Embodiment 33, wherein said at least one enzyme is a recombinant enzyme.

36. The method of any of Embodiments 33-35, further comprising pretreating the feedstock prior to said contacting.

37. The method of any of Embodiments 33 to 36, wherein the feedstock comprises wheat grass, wheat straw, barley straw, sorghum, rice grass, sugarcane, sugar beet, bagasse, switchgrass, corn stover, corn fiber, grains, or a combination thereof.

38. The method of any of Embodiments 33 to 37, wherein the fermentable sugar comprises glucose and/or xylose.

39. The method of any of Embodiments 33 to 37, further comprising recovering at least one fermentable sugar.

40. The method of any of Embodiments 33 to 39, further comprising contacting the at least one fermentable sugar with a microorganism under conditions such that said microorganism produces at least one fermentation end product.

41. The method of Embodiment 40, wherein said fermentation end product is selected from alcohols, fatty acids, lactic acid, acetic acid, 3-hydroxypropionic acid, acrylic acid, succinic acid, citric acid, malic acid, fumaric acid, succinic acid, amino acids, 1,3-propanediol, ethylene, glycerol, and β-lactams.

42. The method of Embodiment 41, wherein said fermentation product is an alcohol selected from ethanol and butanol.

43. The method of Embodiment 42, wherein said alcohol is ethanol.

44. The method of any of Embodiments 33-43, wherein the feedstock is a cellulosic and/or lignocellulosic feedstock.

45. A method of producing an end product from a feedstock, comprising:

a) contacting the feedstock with at least one enzyme composition according to any of Embodiments 6 to 9, under conditions whereby at least one fermentable sugar is produced from the substrate; and

b) contacting the fermentable sugar with a microorganism under conditions such that the microorganism uses the fermentable sugar to produce an end-product.

46. The method of Embodiment 45, wherein the method comprises a simultaneous saccharification and fermentation reactions (SSF).

47. The method of Embodiment 45, wherein the method comprises separate saccharification and fermentation reactions (SHF).

48. The method of any of Embodiments 45 to 47, wherein the feedstock is a cellulosic and/or lignocellulosic feedstock.

49. A method of producing a fermentation end product from a feedstock, comprising:

a) obtaining at least one fermentable sugar produced according to the method of any of Embodiments 33 to 48; and

b) contacting the fermentable sugar with a microorganism in a fermentation to produce at least one fermentation end product.

50. The method of Embodiment 49, wherein said fermentation end product is selected from alcohols, fatty acids, lactic acid, acetic acid, 3-hydroxypropionic acid, acrylic acid, succinic acid, citric acid, malic acid, fumaric acid, succinic acid, amino acids, 1,3-propanediol, ethylene, glycerol, and β-lactams.

51. The method of Embodiment 49 and/or 50, wherein said fermentation end product is at least one alcohol selected from ethanol and butanol.

52. The method of any of Embodiments 49 to 51, wherein the microorganism is a yeast.

53. The method of any of Embodiments 49 to 52, further comprising recovering the fermentation end product.

54. A method for producing at least one fermentable sugar from a feedstock, comprising contacting the feedstock with at least one recombinant beta xylosidase and/or biologically active beta xylosidase fragment of Embodiment 1 and/or the recombinant host cell set forth in any of Embodiments 14 to 28, and/or the enzyme composition provided in Embodiments 6-9, under culture conditions whereby fermentable sugars are produced.

55. The method of Embodiment 54, wherein the enzyme composition and/or recombinant host cell further comprises at least one enzyme selected from endoglucanases (EG), β-glucosidases (BGL), Type 1 cellobiohydrolases (CBH1), Type 2 cellobiohydrolases (CBH2), GH61s, and xylanases.

56. The method of Embodiment 55, wherein said at least one enzyme is a recombinant enzyme.

57. The method of Embodiment 55 and/or 56, wherein said at least one enzyme is a

heterologous enzyme.

58. The method of any of Embodiments 54-57, further comprising pretreating the feedstock prior to said contacting.

59. The method of any of Embodiments 54 to 58, wherein the feedstock comprises wheat grass, wheat straw, barley straw, sorghum, rice grass, sugarcane, sugar beet, bagasse, switchgrass, corn stover, corn fiber, grains, or a combination thereof.

60. The method of any of Embodiments 54 to 59, wherein the fermentable sugar comprises glucose and/or xylose.

61. The method of any of Embodiments 54 to 60, further comprising recovering at least one fermentable sugar.

62. The method of any of Embodiments 54 to 61, further comprising contacting the at least one fermentable sugar with a microorganism under conditions such that said microorganism produces at least one fermentation end product.

63. The method of Embodiment 62, wherein said fermentation end product is selected from alcohols, fatty acids, lactic acid, acetic acid, 3-hydroxypropionic acid, acrylic acid, succinic acid, citric acid, malic acid, fumaric acid, succinic acid, amino acids, 1,3-propanediol, ethylene, glycerol, and β-lactams.

64. The method of Embodiment 63, wherein said fermentation product is an alcohol selected from ethanol and butanol.

65. The method of Embodiment 64, wherein said alcohol is ethanol.

66. The method of any of Embodiments 54-65, wherein the feedstock is a cellulosic and/or lignocellulosic feedstock.

67. A method of producing an end product from a feedstock, comprising:

a) contacting the feedstock with at least one recombinant beta-xylosidase and/or at least one biologically active beta-xylosidase fragment of Embodiment 1 and/or the recombinant host cell set forth in any of Embodiments 18 to 32, under conditions whereby at least one fermentable sugar is produced from the substrate; and

b) contacting the fermentable sugar with a microorganism under conditions such that the microorganism uses the fermentable sugar to produce an end-product.

68. The method of Embodiment 67, wherein the recombinant organism and/or recombinant host cell further comprises at least one enzyme selected from endoglucanases (EG), β-glucosidases (BGL), Type 1 cellobiohydrolases (CBH1),Type 2 cellobiohydrolases (CBH2), GH61 enzymes, and xylanases.

69. The method of Embodiment 68, wherein said at least one enzyme is a recombinant

enzyme.

70. The method of Embodiment 68 and/or 69, wherein said at least one enzyme is a heterologous enzyme.

71. The method of any of Embodiment 67-70, wherein the method comprises a simultaneous saccharification and fermentation reactions (SSF).

72. The method of any of Embodiments 67-70, wherein the method comprises separate saccharification and fermentation reactions (SHF).

73. The method of any of Embodiments 67 to 72, wherein the feedstock is a cellulosic and/or lignocellulosic feedstock.

74. A method of producing a fermentation end product from a feedstock, comprising:

a) obtaining at least one fermentable sugar produced according to the method of any of Embodiments 67-73; and

b) contacting the fermentable sugar with a microorganism in a fermentation to produce at least one fermentation end product.

75. The method of Embodiment 74, wherein said fermentation end product is selected from alcohols, fatty acids, lactic acid, acetic acid, 3-hydroxypropionic acid, acrylic acid, succinic acid, citric acid, malic acid, fumaric acid, succinic acid, amino acids, 1,3-propanediol, ethylene, glycerol, and β-lactams.

76. The method of Embodiment 74 and/or 75, wherein said fermentation end product is at least one alcohol selected from ethanol and butanol.

77. The method of any of Embodiments 74 to 76, wherein the microorganism is a yeast.

78. The method of any of Embodiments 74-77, further comprising recovering the fermentation end product.

DESCRIPTION OF THE DRAWINGS

[0033] Figure 1 provides a map of pYTSEC72-trc.

[0034] Figure 2 provides a map of pC1DX10PhR.

[0035] Figure 3 provides a graph showing the relative thermostabilities of some beta-xylosidase variants.

[0036] Figure 4 provides a graph showing relative thermoactivities of some beta-xylosidase variants as compared to XOS.

[0037] Figure 5 provides a graph showing that some thermostable beta-xylosidase variants provide improve xylose yields in saccharification reactions.

DESCRIPTION OF THE INVENTION

[0038] The present invention provides beta-xylosidase variant enzymes suitable for use in saccharification reactions. The present application further provides genetically modified fungal organisms that produce beta-xylosidase variants, as well as enzyme mixtures exhibiting enhanced hydrolysis of cellulosic material to fermentable sugars, enzyme mixtures produced by the genetically modified fungal organisms, and methods for producing fermentable sugars from cellulose using such enzyme mixtures.

[0039] All patents and publications, including all sequences disclosed within such patents and publications, referred to herein are expressly incorporated by reference. Unless otherwise indicated, the practice of the present invention involves conventional techniques commonly used in molecular biology, fermentation, microbiology, and related fields, which are known to those of skill in the art. Unless defined otherwise herein, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the some methods and materials are described herein. Indeed, it is intended that the present invention not be limited to the particular methodology, protocols, and reagents described herein, as these may vary, depending upon the context in which they are used. The headings provided herein are not limitations of the various aspects or embodiments of the present invention. Nonetheless, in order to facilitate understanding of the present invention, a number of terms are defined below. Numeric ranges are inclusive of the numbers defining the range. Thus, every numerical range disclosed herein is intended to encompass every narrower numerical range that falls within such broader numerical range, as if such narrower numerical ranges were all expressly written herein. It is also intended that every maximum (or minimum) numerical limitation disclosed herein includes every lower (or higher) numerical limitation, as if such lower (or higher) numerical limitations were expressly written herein.

[0040] As used herein, the term "comprising" and its cognates are used in their inclusive sense (*i.e.*, equivalent to the term "including" and its corresponding cognates).

[0041] As used herein and in the appended claims, the singular "a", "an" and "the" include the plural reference unless the context clearly dictates otherwise. Thus, for example, reference to a "host cell" includes a plurality of such host cells.

[0042] Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation; amino acid sequences are written left to right in amino to carboxy orientation, respectively. The headings provided herein are not limitations of the various aspects or embodiments of the invention that can be had by reference to the specification as a whole. Accordingly, the terms defined below are more fully defined by reference to the specification as a whole.

[0043] As used herein, the terms "isolated" and "purified" are used to refer to a molecule (*e.g.,* an isolated nucleic acid, polypeptide, etc.) or other component that is removed from at least one other component with which it is naturally associated.

[0044] As used herein, "polynucleotide" refers to a polymer of deoxyribonucleotides or ribonucleotides in either single- or double-stranded form, and complements thereof.

[0045] As used herein, the term "cellulase" refers to any enzyme that is capable of degrading cellulose. Thus, the term encompasses enzymes capable of hydrolyzing cellulose (β-1,4-glucan or β-D-glucosidic linkages) to shorter cellulose chains, oligosaccharides, cellobiose and/or glucose. "Cellulases" are divided into three sub-categories of enzymes: 1,4-β-D-glucan glucanohydrolase ("endoglucanase" or "EG"); 1,4-β-D-glucan cellobiohydrolase ("exoglucanase," "cellobiohydrolase," or "CBH"); and β-D-glucoside-glucohydrolase ("β-glucosidase," "cellobiase," "BG," or "BGL"). These enzymes act in concert to catalyze the hydrolysis of cellulose-containing substrates. Endoglucanases break internal bonds and disrupt the crystalline structure of cellulose, exposing individual cellulose polysaccharide chains ("glucans"). Cellobiohydrolases incrementally shorten the glucan molecules, releasing mainly cellobiose units (a water-soluble β-1,4-linked dimer of glucose) as well as glucose, cellotriose, and cellotetrose. β-glucosidases split the cellobiose into glucose monomers.

[0046] As used herein, the term "wild-type" refers to naturally-occurring organisms, enzymes and/or other proteins (e.g., non-recombinant enzymes).

[0047] As used herein, the terms "enzyme variant" and "variant enzyme" are used in reference to enzymes that are similar to a reference enzyme, particularly in their function, but have mutations in their amino acid sequence that make them different in sequence from the wild-type or another reference enzyme. Enzyme variants can be made by a wide variety of

different mutagenesis techniques well known to those skilled in the art. In addition, mutagenesis kits are also available from many commercial molecular biology suppliers. Methods are available to make specific substitutions at defined amino acids (site-directed), specific or random mutations in a localized region of the gene (regio-specific) or random mutagenesis over the entire gene (*e.g.*, saturation mutagenesis). Numerous suitable methods are known to those in the art to generate enzyme variants, including but not limited to site-directed mutagenesis of single-stranded DNA or double-stranded DNA using PCR, cassette mutagenesis, gene synthesis, error-prone PCR, shuffling, and chemical saturation mutagenesis, or any other suitable method known in the art. After the variants are produced, they can be screened for the desired property (*e.g.*, high or increased; or low or reduced activity, increased thermal and/or alkaline stability, etc.). In some embodiments, "variant xylosidases" (also referred to as "variant xylosidase enzymes" and "xylosidase variants") find use. These variants are similar to a reference enzyme (e.g., wild-type xylosidase), particularly in their function, but have mutations in their amino acid sequence that make them different in sequence from the wild-type or another reference (e.g., another variant) xylosidase.

[0048] As used herein, "combinatorial variant" refers to any variant that has a combination of two or more mutations (*e.g.,* substitutions). In some embodiments, the combination of mutations results in changes in enzyme activity (*e.g.*, improved thermostability, improved thermoactivity, improved specific activity, etc.). In some embodiments, combinatorial variant xylosidases find use.

[0049] The terms "improved" or "improved properties," as used in the context of describing the properties of a beta-xylosidase (e.g., beta-xylosidase variants), refers to a beta-xylosidase polypeptide that exhibits an improvement in a property or properties as compared to another beta-xylosidase and/or a specified reference polypeptide. Improved properties include, but are not limited to such properties as increased protein expression, increased thermoactivity, increased thermostability, increased pH activity, increased stability (*e.g.,* increased pH stability), increased product specificity, increased specific activity, increased substrate specificity, increased resistance to substrate or end-product inhibition, increased chemical stability, reduced inhibition by glucose, increased resistance to inhibitors (*e.g.,* acetic acid, lectins, tannic acids, and phenolic compounds), and altered pH/temperature profile.

[0050] As used herein, the phrase "improved thermoactivity" or "increased thermoactivity" refers to an enzyme displaying an increase, relative to a reference enzyme, in the amount of beta-xylosidase enzymatic activity (e.g., substrate hydrolysis) in a specified time under specified reaction conditions, for example, elevated temperature. Exemplary methods for measuring beta-xylosidase activity are provided in the Examples herein. In addition, cells expressing and secreting the recombinant proteins can be cultured under the same conditions and the beta-xylosidase activity per volume culture medium can be compared.

[0051] As used herein, the term "improved thermostability" or "increased thermostability" refers to an enzyme displaying an increase in "residual activity" relative to a reference enzyme. Residual activity is determined by (1) exposing the test enzyme or reference enzyme to stress conditions of elevated temperature, optionally at lowered pH, for a period of time and then determining beta-xylosidase activity; (2) exposing the test enzyme or reference enzyme to unstressed conditions for the same period of time and then determining beta-xylosidase activity; and (3) calculating residual activity as the ratio of activity obtained under stress conditions (1) over the activity obtained under unstressed conditions (2). For example, the beta-xylosidase activity of the enzyme exposed to stress conditions ("a") is compared to that of a control in which the enzyme is not exposed to the stress conditions ("b"), and residual activity is equal to the ratio a/b. An enzyme with increased thermostability will have greater residual activity than the reference enzyme. In some embodiments, the enzymes are exposed to stress conditions of 55°C at pH 5.0 for 1 hr, but other cultivation conditions, such as conditions described herein, can be used. Exemplary methods for measuring residual beta-xylosidase activity are provided in the Examples herein.

[0052] As used herein, the terms "endoglucanase" and "EG" refer to a category of cellulases (EC 3.2.1.4) that catalyze the hydrolysis of internal β-1,4 glucosidic bonds of cellulose.

[0053] As used herein, "EG1" refers to a carbohydrate active enzyme expressed from a nucleic acid sequence coding for a glycohydrolase (GH) Family 7 catalytic domain classified under EC 3.2.1.4 or any protein, polypeptide or catalytically active fragment thereof. In some embodiments, the EG1 is functionally linked to a carbohydrate binding module (CBM), such as a Family 1 cellulose binding domain. As used herein, the term "EG1b polypeptide" refers to a polypeptide comprising EG1b activity.

[0054] As used herein, the term "EG2" refers to a carbohydrate active enzyme expressed from a nucleic acid sequence coding for a glycohydrolase (GH) Family 5 catalytic domain

classified under EC 3.2.1.4 or any protein, polypeptide or catalytically active fragment thereof. In some embodiments, the EG2 is functionally linked to a carbohydrate binding module (CBM), such as a Family 1 cellulose binding domain.

[0055] As used herein, the term "EG3" refers to a carbohydrate active enzyme expressed from a nucleic acid sequence coding for a glycohydrolase (GH) Family 12 catalytic domain classified under EC 3.2.1.4 or any protein, polypeptide or catalytically active fragment thereof. In some embodiments, the EG3 is functionally linked to a carbohydrate binding module (CBM), such as a Family 1 cellulose binding domain.

[0056] As used herein, the term "EG4" refers to a carbohydrate active enzyme expressed from a nucleic acid sequence coding for a glycohydrolase (GH) Family 61 catalytic domain classified under EC 3.2.1.4 or any protein, polypeptide or fragment thereof. In some embodiments, the EG4 is functionally linked to a carbohydrate binding module (CBM), such as a Family 1 cellulose binding domain.

[0057] As used herein, the term "EG5" refers to a carbohydrate active enzyme expressed from a nucleic acid sequence coding for a glycohydrolase (GH) Family 45 catalytic domain classified under EC 3.2.1.4 or any protein, polypeptide or fragment thereof. In some embodiments, the EG5 is functionally linked to a carbohydrate binding module (CBM), such as a Family 1 cellulose binding domain.

[0058] As used herein, the term "EG6" refers to a carbohydrate active enzyme expressed from a nucleic acid sequence coding for a glycohydrolase (GH) Family 6 catalytic domain classified under EC 3.2.1.4 or any protein, polypeptide or fragment thereof. In some embodiments, the EG6 is functionally linked to a carbohydrate binding module (CBM), such as a Family 1 cellulose binding domain.

[0059] As used herein, the terms "cellobiohydrolase" and "CBH" refer to a category of cellulases (EC 3.2.1.91) that hydrolyze glycosidic bonds in cellulose.

[0060] As used herein, the terms "CBH1" and "type 1 cellobiohydrolase" refer to a carbohydrate active enzyme expressed from a nucleic acid sequence coding for a glycohydrolase (GH) Family 7 catalytic domain classified under EC 3.2.1.91 or any protein, polypeptide or catalytically active fragment thereof. In some embodiments, the CBH1 is functionally linked to a carbohydrate binding module (CBM), such as a Family 1 cellulose binding domain.

[0061] As used herein, the terms "CBH2" and "type 2 cellobiohydrolase" refer to a carbohydrate active enzyme expressed from a nucleic sequence coding for a glycohydrolase (GH) Family 6 catalytic domain classified under EC 3.2.1.91 or any protein, polypeptide or catalytically active fragment thereof. Type 2 cellobiohydrolases are also commonly referred to as "the Cel6 family." The CBH2 may be functionally linked to a carbohydrate binding module (CBM), such as a Family 1 cellulose binding domain.

[0062] As used herein, the terms "β-glucosidase," "cellobiase," and "BGL" refers to a category of cellulases (EC 3.2.1.21) that catalyze the hydrolysis of cellobiose to glucose.

[0063] As used herein, the term "glycoside hydrolase 61" and "GH61" refers to a category of cellulases that enhance cellulose hydrolysis when used in conjunction with one or more additional cellulases. The GH61 family of cellulases is described, for example, in the Carbohydrate Active Enzymes (CAZY) database (*See e.g.*, Harris *et al.*, Biochem., 49(15):3305-16 [2010]).

[0064] A "hemicellulase" as used herein, refers to a polypeptide that can catalyze hydrolysis of hemicellulose into small polysaccharides such as oligosaccharides, or monomeric saccharides. Hemicelluloses include xylan, glucuonoxylan, arabinoxylan, glucomannan and xyloglucan. Hemicellulases include, for example, the following: endoxylanases, b-xylosidases, a-L-arabinofuranosidases, a-D-glucuronidases, feruloyl esterases, coumaroyl esterases, a-galactosidases, b-galactosidases, b-mannanases, and b-mannosidases. In some embodiments, the present invention provides enzyme mixtures that comprise at least one beta-xylosidase variant and one or more additional hemicellulases.

[0065] As used herein, "protease" includes enzymes that hydrolyze peptide bonds (peptidases), as well as enzymes that hydrolyze bonds between peptides and other moieties, such as sugars (glycopeptidases). Many proteases are characterized under EC 3.4, and are suitable for use in the present invention. Some specific types of proteases include but are not limited to, cysteine proteases including pepsin, papain and serine proteases including chymotrypsins, carboxypeptidases and metalloendopeptidases. In some embodiments, the present invention provides at least one variant beta-xylosidase and at least one protease.

[0066] As used herein, "lipase" includes enzymes that hydrolyze lipids, fatty acids, and acylglycerides, including phosphoglycerides, lipoproteins, diacylglycerols, and the like. In plants, lipids are used as structural components to limit water loss and pathogen infection. These lipids include waxes derived from fatty acids, as well as cutin and suberin. In some

additional embodiments, the present invention provides at least one variant beta-xylosidase and at least one lipase

[0067] As used herein, the term "xylanase" refers to enzymes within EC 3.2.1.8, that catalyze the hydrolysis of 1,4-beta-D-xylans, to cleave polymers or oligomers of xylose-containing xylans or hemicellulose into shorter chains. This enzyme may also be referred to as endo-1,4-beta-xylanase, 4-beta-D-xylan xylanohydrolase, endo-xylanase, or beta-xylanase.

[0068] As used herein, the term "xylanase polynucleotide" refers to a polynucleotide encoding a polypeptide comprising beta-xylanase activity.

[0069] As used herein, the term "xylanase activity" refers to the enzymatic activity of xylanase (*i.e.*, hydrolyzing a cellulose-containing substrate).

[0070] As used herein, the term "xylosidase" refers to a group of enzymes that catalyze the hydrolysis of alpha- or beta-xylosidic linkages. Enzymes in class EC 3.2.1.8 catalyze the endo-hydrolysis of 1,4-beta-D-xylosidic linkages; while those in class EC 3.2.1.32 catalyze the endo-hydrolysis of 1,3-beta-D-xylosidic linkages; those in class EC 3.2.1.37 catalyze the exo-hydrolysis of 1,4-beta-D-linkages from the non-reducing termini of xylans; and those in class EC 3.2.1.72 catalyze the exo-hydrolysis of 1,3-beta-D-linkages from the non-reducing termini of xylans. Additional xylosidases have been identified that catalyze the hydrolysis of alpha-xylosidic bonds. As used herein, the term encompasses alpha-xylosidases and beta-xylosidases, as well as any other enzymes that have xylosidase activity (e.g., gamma-xylosidases).

[0071] As used herein the term "xylosidase polynucleotide" refers to a polynucleotide encoding a polypeptide comprising xylosidase activity.

[0072] As used herein, the term "xylosidase activity" refers to the enzymatic activity of xylosidase (*i.e.*, hydrolyzing a cellulose-containing substrate).

[0073] As used herein, the term "alpha-xylosidase" refers to enzymes within EC 3.2.1 that remove the alpha-1,6-linked xylose residue from xyloglucan. In some embodiments, the removal of the alpha-1,6-linked xylose residue from xyloglucan facilitates the breakdown of xyloglucan to monomeric sugars (e.g., glucose and xylose).

[0074] As used herein the term "alpha-xylosidase polynucleotide" refers to a polynucleotide encoding a polypeptide comprising alpha-xylosidase activity.

[0075] As used herein, the term “alpha-xylosidase activity" refers to the enzymatic activity of alpha-xylosidase (*i.e.*, removing the alpha-1,6-linked xylose residues from xyloglucan).

[0076] As used herein, the term “beta-xylosidase” refers to enzymes within EC 3.2.1.37, that catalyze the hydrolysis of 1 ,4-beta-D-xylans, to remove successive D-xylose residues from the non-reducing termini. This enzyme may also be referred to as xylan 1, beta-β-xylosidase, 1 ,4-beta-D-xylan xylohydrolase, exo-1 ,4-beta-xylosidase or xylobiase.

[0077] As used herein, the term "beta-xylosidase polynucleotide" refers to a polynucleotide encoding a polypeptide comprising beta-xylosidase activity.

[0078] As used herein, the term “beta-xylosidase activity" refers to the enzymatic activity of beta-xylosidase (*i.e.*, hydrolyzing a cellulose-containing substrate).

[0079] As used herein, in some embodiments, the terms "wild-type beta-xylosidase polynucleotide," "wild-type beta-xylosidase DNA," and “wild-type beta-xylosidase nucleic acid” refer to SEQ ID NO:1, 56, and/or SEQ ID NO:58.

[0080] The terms "protein" and "polypeptide" are used interchangeably herein to refer to a polymer of amino acid residues. In addition, the terms "amino acid" “polypeptide,” and “peptide” encompass naturally-occurring and synthetic amino acids, as well as amino acid analogs. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified (*e.g.,* hydroxyproline, γ-carboxyglutamate, and O-phosphoserine). As used herein, the term “amino acid analogs” refers to compounds that have the same basic chemical structure as a naturally occurring amino acid (*i.e.*, an α-carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, including but not limited to homoserine, norleucine, methionine sulfoxide, and methionine methyl sulfonium). In some embodiments, these analogs have modified R groups (*e.g.,* norleucine) and/or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid.

[0081] Amino acids are referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes.

[0082] An amino acid or nucleotide base "position" is denoted by a number that sequentially identifies each amino acid (or nucleotide base) in the reference sequence based on its position

relative to the N-terminus (or 5'-end). Due to deletions, insertions, truncations, fusions, and the like that must be taken into account when determining an optimal alignment, the amino acid residue number in a test sequence determined by simply counting from the N-terminus will not necessarily be the same as the number of its corresponding position in the reference sequence. For example, in a case where a test sequence has a deletion relative to an aligned reference sequence, there will be no amino acid in the variant that corresponds to a position in the reference sequence at the site of deletion. Where there is an insertion in an aligned test sequence, that insertion will not correspond to a numbered amino acid position in the reference sequence. In the case of truncations or fusions there can be stretches of amino acids in either the reference or aligned sequence that do not correspond to any amino acid in the corresponding sequence.

[0083] As used herein, the terms "numbered with reference to" or "corresponding to," when used in the context of the numbering of a given amino acid or polynucleotide sequence, refers to the numbering of the residues of a specified reference sequence when the given amino acid or polynucleotide sequence is compared to the reference sequence. The following nomenclature may be used to describe substitutions in a test sequence relative to a reference sequence polypeptide or nucleic acid sequence: "R-#-V," where # refers to the position in the reference sequence, R refers to the amino acid (or base) at that position in the reference sequence, and V refers to the amino acid (or base) at that position in the test sequence. In some embodiments, an amino acid (or base) may be called "X," by which is meant any amino acid (or base).

[0084] As used herein, the term "reference enzyme" refers to an enzyme to which another enzyme of the present invention (*e.g.*, a "test" enzyme, such as wild-type beta-xylosidase) is compared in order to determine the presence of an improved property in the other enzyme being evaluated. In some embodiments, a reference enzyme is a wild-type enzyme (*e.g.*, a wild-type beta-xylosidase). In some embodiments, the reference enzyme is an enzyme to which a test enzyme of the present invention is compared in order to determine the presence of an improved property in the test enzyme being evaluated, including but not limited to improved thermoactivity, improved thermostability, and/or improved stability. In some embodiments, a reference enzyme is a wild-type enzyme (*e.g.*, a wild-type beta-xylosidase).

[0085] As used herein, the terms "biologically active fragment" and "fragment" refer to a polypeptide that has an amino-terminal and/or carboxy-terminal deletion(s) and/or internal

deletion(s), but where the remaining amino acid sequence is identical to the corresponding positions in the sequence to which it is being compared (*e.g.,* a full-length beta-xylosidase variant of the present invention) and that retains substantially all of the activity of the full-length polypeptide. In some embodiments, the biologically active fragment is a biologically active beta-xylosidase variant fragment. A biologically active fragment can comprise about 60%, about 65%, about 70%, about 75%, about 80%, about 85%, at about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% of a full-length beta-xylosidase variant polypeptide. In some embodiments, the biologically active fragments comprise about 20%, about 21%, about 22%, about 23%, about 24%, about 25%, about 26%, about 27%, about 28%, about 29%, about 30%, about 35%, about 40%, about 45%, about 50%, or about 55% of a full-length beta-xylosidase variant.

[0086] As used herein, the term "overexpress" is intended to encompass increasing the expression of a protein to a level greater than the cell normally produces. It is intended that the term encompass overexpression of endogenous, as well as heterologous proteins.

[0087] As used herein, the term "recombinant" refers to a polynucleotide or polypeptide that does not naturally occur in a host cell. In some embodiments, recombinant molecules contain two or more naturally-occurring sequences that are linked together in a way that does not occur naturally. In some embodiments, "recombinant cells" express genes that are not found in identical form within the native (*i.e.*, non-recombinant) form of the cell and/or express native genes that are otherwise abnormally over-expressed, under-expressed, and/or not expressed at all due to deliberate human intervention. As used herein, "recombinant cells," as well as recombinant host cells," "recombinant microorganisms," and "recombinant fungal cells," contain at least one recombinant polynucleotide or polypeptide.

[0088] As used herein, "recombinant" used in reference to a cell or vector, refers to a cell or vector that has been modified by the introduction of a heterologous nucleic acid sequence or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found in identical form within the native (non-recombinant) form of the cell or express native genes that are otherwise abnormally expressed, under expressed or not expressed at all as a result of deliberate human intervention. Thus, "recombinant" or "engineered" or "non-naturally occurring" when used with reference to a cell, nucleic acid, or polypeptide, refers to a material, or a material corresponding to the natural or native form of the material, that has been modified in a manner that would not otherwise exist in nature, or

is identical thereto but produced or derived from synthetic materials and/or by manipulation using recombinant techniques. Non-limiting examples include, among others, recombinant cells expressing genes that are not found within the native (non-recombinant) form of the cell or express native genes that are otherwise expressed at a different level. "Recombination," "recombining" and generating a "recombined" nucleic acid generally encompass the assembly of at least two nucleic acid fragments. In some embodiments, "Recombination," "recombining," and generating a "recombined" nucleic acid also encompass the assembly of two or more nucleic acid fragments wherein the assembly gives rise to a chimeric gene.

[0089] As used herein, when used with reference to a nucleic acid or polypeptide, the term "heterologous" refers to a sequence that is not normally expressed and secreted by an organism (*e.g.,* a wild-type organism). In some embodiments, the term encompasses a sequence that comprises two or more subsequences which are not found in the same relationship to each other as normally found in nature, or is recombinantly engineered so that its level of expression, or physical relationship to other nucleic acids or other molecules in a cell, or structure, is not normally found in nature. For instance, a heterologous nucleic acid is typically recombinantly produced, having two or more sequences from unrelated genes arranged in a manner not found in nature (*e.g.*, a nucleic acid open reading frame (ORF) of the invention operatively linked to a promoter sequence inserted into an expression cassette, such as a vector).

[0090] A nucleic acid construct, nucleic acid (*e.g.*, a polynucleotide), polypeptide, or host cell is referred to herein as "recombinant" when it is non-naturally occurring, artificial or engineered. The present invention also provides recombinant nucleic acid constructs comprising a beta-xylosidase variant polynucleotide sequence that hybridizes under stringent hybridization conditions to the complement of a polynucleotide which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, and/or 55.

[0091] Nucleic acids "hybridize" when they associate, typically in solution. Nucleic acids hybridize due to a variety of well-characterized physico-chemical forces, such as hydrogen bonding, solvent exclusion, base stacking and the like. As used herein, the term "stringent hybridization wash conditions" in the context of nucleic acid hybridization experiments, such as Southern and Northern hybridizations, are sequence dependent, and are different under different environmental parameters. An extensive guide to the hybridization of nucleic acids

is found in Tijssen, 1993, "Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes," Part I, Chapter 2 (Elsevier, New York), which is incorporated herein by reference. For polynucleotides of at least 100 nucleotides in length, low to very high stringency conditions are defined as follows: prehybridization and hybridization at 42°C in 5xSSPE, 0.3% SDS, 200 μg/ml sheared and denatured salmon sperm DNA, and either 25% formamide for low stringencies, 35% formamide for medium and medium-high stringencies, or 50% formamide for high and very high stringencies, following standard Southern blotting procedures. For polynucleotides of at least 100 nucleotides in length, the carrier material is finally washed three times each for 15 minutes using 2xSSC, 0.2% SDS 50°C (low stringency), at 55°C (medium stringency), at 60°C (medium-high stringency), at 65°C (high stringency), or at 70°C (very high stringency).

[0092] As used herein, "identity" or "percent identity," in the context of two or more polypeptide sequences, refers to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues that are the same (*e.g.,* share at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 88% identity, at least about 89%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identity) over a specified region to a reference sequence, when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using a sequence comparison algorithms or by manual alignment and visual inspection.

[0093] In some embodiments, the terms "percent identity," "% identity", "percent identical," and "% identical," are used interchangeably herein to refer to the percent amino acid or polynucleotide sequence identity that is obtained by ClustalW analysis (version W 1.8 available from European Bioinformatics Institute, Cambridge, UK), counting the number of identical matches in the alignment and dividing such number of identical matches by the length of the reference sequence, and using the following ClustalW parameters to achieve slow/more accurate pairwise optimal alignments – DNA/Protein Gap Open Penalty:15/10; DNA/Protein Gap Extension Penalty:6.66/0.1; Protein weight matrix: Gonnet series; DNA weight matrix: Identity..

[0094] As used herein the term "comparison window," includes reference to a segment of any one of a number of contiguous positions from about 20 to about 464 (*e.g.*, about 50 to about

300 contiguous positions, about 50 to 250 contiguous positions, or also about 100 to about 200 contiguous positions), in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. As noted, in some embodiments the comparison is between the entire length of the two sequences, or, if one sequence is a fragment of the other, the entire length of the shorter of the two sequences. Optimal alignment of sequences for comparison and determination of sequence identity can be determined by a sequence comparison algorithm or by visual inspection, as well-known in the art. When optimally aligning sequences and determining sequence identity by visual inspection, percent sequence identity is calculated as the number of residues of the test sequence that are identical to the reference sequence divided by the number of non-gap positions and multiplied by 100. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

[0095] Two sequences are "aligned" when they are aligned for similarity scoring using a defined amino acid substitution matrix (*e.g.*, BLOSUM62), gap existence penalty and gap extension penalty so as to arrive at the highest score possible for that pair of sequences. Amino acid substitution matrices and their use in quantifying the similarity between two sequences are well known in the art (*See, e.g.*, Dayhoff *et al.*, *in* Dayhoff [ed.], Atlas of Protein Sequence and Structure," Vol. 5, Suppl. 3, Natl. Biomed. Res. Round., Washington D.C. [1978]; pp. 345-352; and Henikoff *et al.*, Proc. Natl. Acad. Sci. USA, 89:10915-10919 [1992], both of which are incorporated herein by reference). The BLOSUM62 matrix is often used as a default scoring substitution matrix in sequence alignment protocols such as Gapped BLAST 2.0. The gap existence penalty is imposed for the introduction of a single amino acid gap in one of the aligned sequences, and the gap extension penalty is imposed for each additional empty amino acid position inserted into an already opened gap. The alignment is defined by the amino acid position of each sequence at which the alignment begins and ends, and optionally by the insertion of a gap or multiple gaps in one or both sequences so as to arrive at the highest possible score. While optimal alignment and scoring can be accomplished manually, the process is facilitated by the use of a computer-implemented alignment algorithm (*e.g.*, gapped BLAST 2.0; *See*, Altschul *et al.*, Nucleic Acids Res., 25:3389-3402 [1997], which is incorporated herein by reference), and made available to the

public at the National Center for Biotechnology Information Website). Optimal alignments, including multiple alignments can be prepared using readily available programs such as PSI-BLAST (*See e.g*,, Altschul *et al*., *supra*).

[0096] The present invention also provides a recombinant nucleic acid construct comprising at least one beta-xylosidase variant polynucleotide sequence that hybridizes under stringent hybridization conditions to the complement of a polynucleotide which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, and/or 55, wherein the polypeptide is capable of catalyzing the degradation of cellulose.

[0097] Two nucleic acid or polypeptide sequences that have 100% sequence identity are said to be "identical." A nucleic acid or polypeptide sequence are said to have "substantial sequence identity" to a reference sequence when the sequences have at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, or greater sequence identity as determined using the methods described herein, such as BLAST using standard parameters.

[0098] As used herein, the term "pre-protein" refers to a protein including an amino-terminal signal peptide (or leader sequence) region attached. The signal peptide is cleaved from the pre-protein by a signal peptidase prior to secretion to result in the "mature" or "secreted" protein.

[0099] As used herein, a "vector" is a DNA construct for introducing a DNA sequence into a cell. In some embodiments, the vector is an expression vector that is operably linked to a suitable control sequence capable of effecting the expression in a suitable host of the polypeptide encoded in the DNA sequence. An "expression vector" has a promoter sequence operably linked to the DNA sequence (*e.g.,* transgene) to drive expression in a host cell, and in some embodiments a transcription terminator sequence.

[0100] As used herein, the term "expression" includes any step involved in the production of the polypeptide including, but not limited to, transcription, post-transcriptional modification, translation, and post-translational modification. In some embodiments, the term also encompasses secretion of the polypeptide from a cell.

[0101] As used herein, the term "produces" refers to the production of proteins and/or other compounds by cells. It is intended that the term encompass any step involved in the production of polypeptides including, but not limited to, transcription, post-transcriptional modification, translation, and post-translational modification. In some embodiments, the term also encompasses secretion of the polypeptide from a cell.

[0102] As used herein, the term "operably linked" refers to a configuration in which a control sequence is appropriately placed at a position relative to the coding sequence of the DNA sequence such that the control sequence influences the expression of a polypeptide.

[0103] As used herein, an amino acid or nucleotide sequence (*e.g.*, a promoter sequence, signal peptide, terminator sequence, etc.) is "heterologous" to another sequence with which it is operably linked if the two sequences are not associated in nature.

[0104] As used herein, the terms "host cell" and "host strain" refer to suitable hosts for expression vectors comprising DNA provided herein. In some embodiments, the host cells are prokaryotic or eukaryotic cells that have been transformed or transfected with vectors constructed using recombinant DNA techniques as known in the art. Transformed hosts are capable of either replicating vectors encoding at least one protein of interest and/or expressing the desired protein of interest. In addition, reference to a cell of a particular strain refers to a parental cell of the strain as well as progeny and genetically modified derivatives. Genetically modified derivatives of a parental cell include progeny cells that contain a modified genome or episomal plasmids that confer for example, antibiotic resistance, improved fermentation, etc. In some embodiments, host cells are genetically modified to have characteristics that improve protein secretion, protein stability or other properties desirable for expression and/or secretion of a protein. For example, knockout of Alp1 function results in a cell that is protease deficient. Knockout of pyr5 function results in a cell with a pyrimidine deficient phenotype. In some embodiments, host cells are modified to delete endogenous cellulase protein-encoding sequences or otherwise eliminate expression of one or more endogenous cellulases. In some embodiments, expression of one or more endogenous cellulases is inhibited to increase production of cellulases of interest. Genetic modification can be achieved by any suitable genetic engineering techniques and/or classical microbiological techniques (*e.g.,* chemical or UV mutagenesis and subsequent selection). Using recombinant technology, nucleic acid molecules can be introduced, deleted, inhibited or modified, in a manner that results in increased yields of beta-xylosidase variant(s) within

the organism or in the culture. For example, knockout of Alp1 function results in a cell that is protease deficient. Knockout of pyr5 function results in a cell with a pyrimidine deficient phenotype. In some genetic engineering approaches, homologous recombination is used to induce targeted gene modifications by specifically targeting a gene *in vivo* to suppress expression of the encoded protein. In an alternative approach, siRNA, antisense, and/or ribozyme technology finds use in inhibiting gene expression.

[0105] As used herein, the term "introduced" used in the context of inserting a nucleic acid sequence into a cell, means transformation, transduction, conjugation, transfection, and/or any other suitable method(s) known in the art for inserting nucleic acid sequences into host cells. Any suitable means for the introduction of nucleic acid into host cells find use in the present invention.

[0106] As used herein, the terms "transformed" and "transformation" used in reference to a cell refer to a cell that has a non-native nucleic acid sequence integrated into its genome or has an episomal plasmid that is maintained through multiple generations.

[0107] As used herein, the term "C1" refers to strains of *Myceliophthora thermophila*, including the fungal strain described by Garg (*See*, Garg, Mycopathol., 30: 3-4 [1966]). As used herein, "*Chrysosporium lucknowense*" includes the strains described in U.S. Pat. Nos. 6,015,707, 5,811,381 and 6,573,086; US Pat. Pub. Nos. 2007/0238155, US 2008/0194005, US 2009/0099079; International Pat. Pub. Nos., WO 2008/073914 and WO 98/15633, all of which are incorporated herein by reference, and include, without limitation, *Chrysosporium lucknowense* Garg 27K, VKM-F 3500 D (Accession No. VKM F-3500-D), C1 strain UV13-6 (Accession No. VKM F-3632 D), C1 strain NG7C-19 (Accession No. VKM F-3633 D), and C1 strain UV18-25 (VKM F-3631 D), all of which have been deposited at the All-Russian Collection of Microorganisms of Russian Academy of Sciences (VKM), Bakhurhina St. 8, Moscow, Russia, 113184, and any derivatives thereof. Although initially described as *Chrysosporium lucknowense*, C1 may currently be considered a strain of *Myceliophthora thermophila*. Other C1 strains include cells deposited under accession numbers ATCC 44006, CBS (Centraalbureau voor Schimmelcultures) 122188, CBS 251.72, CBS 143.77, CBS 272.77, CBS122190, CBS122189, and VKM F-3500D. Exemplary C1 derivatives include modified organisms in which one or more endogenous genes or sequences have been deleted or modified and/or one or more heterologous genes or sequences have been introduced. Derivatives include, but are not limited to UV18#100f Δalpl, UV18#100f Δpyr5 Δalp1,

UV18#100.f Δalp1 Δpep4 Δalp2, UV18#100.f Δpyr5 Δalp1 Δpep4 Δalp2 and UV18#100.f Δpyr4 Δpyr5 ΔaIp1 Δpep4 Δalp2, as described in WO2008073914 and WO2010107303, each of which is incorporated herein by reference.

[0108] As used herein, the terms "improved thermoactivity" and "increased thermoactivity" refer to an enzyme (*e.g.,* a "test" enzyme of interest) displaying an increase, relative to a reference enzyme, in the amount of enzymatic activity (*e.g.*, substrate hydrolysis) in a specified time under specified reaction conditions, for example, elevated temperature.

[0109] As used herein, the terms "improved thermostability" and "increased thermostability" refer to an enzyme (*e.g.,* a "test" enzyme of interest) displaying an increase in "residual activity" relative to a reference enzyme. Residual activity is determined by (1) exposing the test enzyme or reference enzyme to stress conditions of elevated temperature, optionally at lowered pH, for a period of time and then determining beta-xylosidase activity; (2) exposing the test enzyme or reference enzyme to unstressed conditions for the same period of time and then determining beta-xylosidase activity; and (3) calculating residual activity as the ratio of activity obtained under stress conditions (1) over the activity obtained under unstressed conditions (2). For example, the beta-xylosidase activity of the enzyme exposed to stress conditions ("a") is compared to that of a control in which the enzyme is not exposed to the stress conditions ("b"), and residual activity is equal to the ratio a/b. A test enzyme with increased thermostability will have greater residual activity than the reference enzyme. In some embodiments, the enzymes are exposed to stress conditions of 55°C at pH 5.0 for 1 hr, but other cultivation conditions can be used.

[0110] As used herein, the term "culturing" refers to growing a population of microbial cells under suitable conditions in a liquid or solid medium.

[0111] As used herein, the term "saccharification" refers to the process in which substrates (*e.g.*, cellulosic biomass) are broken down via the action of cellulases to produce fermentable sugars (*e.g.* monosaccharides such as but not limited to glucose).

[0112] As used herein, the term "fermentable sugars" refers to simple sugars (*e.g.*, monosaccharides, disaccharides and short oligosaccharides), including but not limited to glucose, xylose, galactose, arabinose, mannose and sucrose. Indeed, a fermentable sugar is any sugar that a microorganism can utilize or ferment.

[0113] As used herein the term "soluble sugars" refers to water-soluble hexose monomers and oligomers of up to about six monomer units.

[0114] As used herein, the term "fermentation" is used broadly to refer to the cultivation of a microorganism or a culture of microorganisms that use simple sugars, such as fermentable sugars, as an energy source to obtain a desired product.

[0115] As used herein, the term "feedstock" refers to any material that is suitable for use in production of an end product. It is intended that the term encompass any material suitable for use in saccharification reactions. In some embodiments, the term encompasses material obtained from nature that is in an unprocessed or minimally processed state, although it is not intended that the term be limited to these embodiments. In some embodiments, the term encompasses biomass and biomass substrates comprising any suitable compositions for use in production of fermentable sugars. In some embodiments, the feedstock is "pre-treated" before and/or while it is being used as a substrate in a saccharification reaction.

[0116] The terms "biomass," and "biomass substrate," encompass any suitable materials for use in saccharification reactions. The terms encompass, but are not limited to materials that comprise cellulose (*i.e.,* "cellulosic biomass," "cellulosic feedstock," and "cellulosic substrate"). Biomass can be derived from plants, animals, or microorganisms, and may include, but is not limited to agricultural, industrial, and forestry residues, industrial and municipal wastes, and terrestrial and aquatic crops grown for energy purposes. Examples of biomass substrates include, but are not limited to, wood, wood pulp, paper pulp, corn fiber, corn grain, corn cobs, crop residues such as corn husks, corn stover, grasses, wheat, wheat straw, barley, barley straw, hay, rice, rice straw, switchgrass, waste paper, paper and pulp processing waste, woody or herbaceous plants, fruit or vegetable pulp, distillers grain, grasses, rice hulls, cotton, hemp, flax, sisal, sugar cane bagasse, sorghum, soy, switchgrass, components obtained from milling of grains, trees, branches, roots, leaves, tops, wood chips, sawdust, shrubs, bushes, seed pods, vegetables, fruits, and flowers and any suitable mixtures thereof. In some embodiments, the biomass comprises, but is not limited to cultivated crops (*e.g.*, grasses, including C4 grasses, such as switch grass, cord grass, rye grass, miscanthus, reed canary grass, or any combination thereof), sugar processing residues, for example, but not limited to, bagasse (*e.g.*, sugar cane bagasse, beet pulp [*e.g.*, sugar beet], or a combination thereof), agricultural residues (*e.g.*, soybean stover, corn stover, corn fiber, rice straw, sugar cane straw, rice, rice hulls, barley straw, corn cobs, wheat straw, canola straw, oat straw, oat

hulls, corn fiber, hemp, flax, sisal, cotton, tops, stems, leaves, seed pods, fruit pods, or any combination thereof), fruit pulp, vegetable pulp, distillers' grains, forestry biomass (*e.g.*, wood, wood pulp, paper pulp, recycled wood pulp fiber, sawdust, hardwood, such as aspen wood, softwood, or a combination thereof). Furthermore, in some embodiments, the biomass comprises cellulosic waste material and/or forestry waste materials, including but not limited to, paper and pulp processing waste, municipal paper waste, newsprint, cardboard and the like. In some embodiments, biomass comprises one species of fiber, while in some alternative embodiments, the biomass comprises a mixture of fibers that originate from different biomasses. In some embodiments, the biomass may also comprise transgenic plants that express ligninase and/or cellulase enzymes (*See e.g.,* US 2008/0104724 A1).

[0117] A biomass substrate is said to be "pretreated" when it has been processed by some physical and/or chemical means to facilitate saccharification. As described further herein, in some embodiments, the biomass substrate is "pretreated," or treated using methods known in the art, such as chemical pretreatment (*e.g.,* ammonia pretreatment, dilute acid pretreatment, dilute alkali pretreatment, or solvent exposure), physical pretreatment (*e.g.,* steam explosion or irradiation), mechanical pretreatment (*e.g.,* grinding or milling) and biological pretreatment (*e.g.,* application of lignin-solubilizing microorganisms) and combinations thereof, to increase the susceptibility of cellulose to hydrolysis. Thus, the term "biomass" encompasses any living or dead biological material that contains a polysaccharide substrate, including but not limited to cellulose, starch, other forms of long-chain carbohydrate polymers, and mixtures of such sources. It may or may not be assembled entirely or primarily from glucose or xylose, and may optionally also contain various other pentose or hexose monomers. Xylose is an aldopentose containing five carbon atoms and an aldehyde group. It is the precursor to hemicellulose, and is often a main constituent of biomass. In some embodiments, the substrate is slurried prior to pretreatment. In some embodiments, the consistency of the slurry is between about 2% and about 30% and more typically between about 4% and about 15%. In some embodiments, the slurry is subjected to a water and/or acid soaking operation prior to pretreatment. In some embodiments, the slurry is dewatered using any suitable method to reduce steam and chemical usage prior to pretreatment. Examples of dewatering devices include, but are not limited to pressurized screw presses (*See e.g.*, WO 2010/022511, incorporated herein by reference) pressurized filters and extruders.

[0118] In some embodiments, the pretreatment is carried out to hydrolyze hemicellulose, and/or a portion thereof present in the cellulosic substrate to monomeric pentose and hexose

sugars (*e.g.*, xylose, arabinose, mannose, galactose, and/or any combination thereof). In some embodiments, the pretreatment is carried out so that nearly complete hydrolysis of the hemicellulose and a small amount of conversion of cellulose to glucose occurs. In some embodiments, an acid concentration in the aqueous slurry from about 0.02% (w/w) to about 2% (w/w), or any amount therebetween, is typically used for the treatment of the cellulosic substrate. Any suitable acid finds use in these methods, including but not limited to, hydrochloric acid, nitric acid, and/or sulfuric acid. In some embodiments, the acid used during pretreatment is sulfuric acid. Steam explosion is one method of performing acid pretreatment of biomass substrates (*See e.g.*, U.S. Patent No. 4,461,648). Another method of pretreating the slurry involves continuous pretreatment (*i.e.*, the cellulosic biomass is pumped though a reactor continuously). This methods are well-known to those skilled in the art (*See e.g.,* U.S. Patent No. 7,754,457).

[0119] In some embodiments, alkali is used in the pretreatment. In contrast to acid pretreatment, pretreatment with alkali may not hydrolyze the hemicellulose component of the biomass. Rather, the alkali reacts with acidic groups present on the hemicellulose to open up the surface of the substrate. In some embodiments, the addition of alkali alters the crystal structure of the cellulose so that it is more amenable to hydrolysis. Examples of alkali that find use in the pretreatment include, but are not limited to ammonia, ammonium hydroxide, potassium hydroxide, and sodium hydroxide. One method of alkali pretreatment is Ammonia Freeze Explosion, Ammonia Fiber Explosion or Ammonia Fiber Expansion ("AFEX" process; *See e.g.*, U.S. Patent Nos. 5,171,592; 5,037,663; 4,600,590; 6,106,888; 4,356,196; 5,939,544; 6,176,176; 5,037,663 and 5,171,592). During this process, the cellulosic substrate is contacted with ammonia or ammonium hydroxide in a pressure vessel for a sufficient time to enable the ammonia or ammonium hydroxide to alter the crystal structure of the cellulose fibers. The pressure is then rapidly reduced, which allows the ammonia to flash or boil and explode the cellulose fiber structure. In some embodiments, the flashed ammonia is then recovered using methods known in the art. In some alternative methods, dilute ammonia pretreatment is utilized. The dilute ammonia pretreatment method utilizes more dilute solutions of ammonia or ammonium hydroxide than AFEX (*See e.g.*, WO2009/045651 and US 2007/0031953). This pretreatment process may or may not produce any monosaccharides.

[0120] An additional pretreatment process for use in the present invention includes chemical treatment of the cellulosic substrate with organic solvents, in methods such as those utilizing

organic liquids in pretreatment systems (*See e.g.,* U.S. Patent No. 4,556,430; incorporated herein by reference). These methods have the advantage that the low boiling point liquids easily can be recovered and reused. Other pretreatments, such as the Organosolv™ process, also use organic liquids (*See e.g.*, U.S. Patent No. 7,465,791, which is also incorporated herein by reference). Subjecting the substrate to pressurized water may also be a suitable pretreatment method (*See e.g.*, Weil *et al.* (1997) Appl. Biochem. Biotechnol., 68(1-2): 21-40 [1997], which is incorporated herein by reference). In some embodiments, the pretreated cellulosic biomass is processed after pretreatment by any of several steps, such as dilution with water, washing with water, buffering, filtration, or centrifugation, or any combination of these processes, prior to enzymatic hydrolysis, as is familiar to those skilled in the art. The pretreatment produces a pretreated feedstock composition (*e.g.*, a "pretreated feedstock slurry") that contains a soluble component including the sugars resulting from hydrolysis of the hemicellulose, optionally acetic acid and other inhibitors, and solids including unhydrolyzed feedstock and lignin. In some embodiments, the soluble components of the pretreated feedstock composition are separated from the solids to produce a soluble fraction. In some embodiments, the soluble fraction, including the sugars released during pretreatment and other soluble components (*e.g.*, inhibitors), is then sent to fermentation. However, in some embodiments in which the hemicellulose is not effectively hydrolyzed during the pretreatment one or more additional steps are included (*e.g.*, a further hydrolysis step(s) and/or enzymatic treatment step(s) and/or further alkali and/or acid treatment) to produce fermentable sugars. In some embodiments, the separation is carried out by washing the pretreated feedstock composition with an aqueous solution to produce a wash stream and a solids stream comprising the unhydrolyzed, pretreated feedstock. Alternatively, the soluble component is separated from the solids by subjecting the pretreated feedstock composition to a solids-liquid separation, using any suitable method (*e.g.*, centrifugation, microfiltration, plate and frame filtration, cross-flow filtration, pressure filtration, vacuum filtration, etc.). Optionally, in some embodiments, a washing step is incorporated into the solids-liquids separation. In some embodiments, the separated solids containing cellulose, then undergo enzymatic hydrolysis with cellulase enzymes in order to convert the cellulose to glucose. In some embodiments, the pretreated feedstock composition is fed into the fermentation process without separation of the solids contained therein. In some embodiments, the unhydrolyzed solids are subjected to enzymatic hydrolysis with cellulase enzymes to convert the cellulose to glucose after the fermentation process. In some embodiments, the pretreated cellulosic feedstock is subjected to enzymatic hydrolysis with cellulase enzymes.

[0121] As used herein, the term "lignocellulosic biomass" refers to any plant biomass comprising cellulose and hemicellulose, bound to lignin. In some embodiments, the biomass may optionally be pretreated to increase the susceptibility of cellulose to hydrolysis by chemical, physical and biological pretreatments (such as steam explosion, pulping, grinding, acid hydrolysis, solvent exposure, and the like, as well as combinations thereof). Various lignocellulosic feedstocks find use, including those that comprise fresh lignocellulosic feedstock, partially dried lignocellulosic feedstock, fully dried lignocellulosic feedstock, and/or any combination thereof. In some embodiments, lignocellulosic feedstocks comprise cellulose in an amount greater than about 20%, more preferably greater than about 30%, more preferably greater than about 40% (w/w). For example, in some embodiments, the lignocellulosic material comprises from about 20% to about 90% (w/w) cellulose, or any amount therebetween, although in some embodiments, the lignocellulosic material comprises less than about 19%, less than about 18%, less than about 17%, less than about 16%, less than about 15%, less than about 14%, less than about 13%, less than about 12%, less than about 11%, less than about 10%, less than about 9%, less than about 8%, less than about 7%, less than about 6%, or less than about 5% cellulose (w/w). Furthermore, in some embodiments, the lignocellulosic feedstock comprises lignin in an amount greater than about 10%, more typically in an amount greater than about 15% (w/w). In some embodiments, the lignocellulosic feedstock comprises small amounts of sucrose, fructose and/or starch. The lignocellulosic feedstock is generally first subjected to size reduction by methods including, but not limited to, milling, grinding, agitation, shredding, compression/expansion, or other types of mechanical action. Size reduction by mechanical action can be performed by any type of equipment adapted for the purpose, for example, but not limited to, hammer mills, tub-grinders, roll presses, refiners and hydrapulpers. In some embodiments, at least 90% by weight of the particles produced from the size reduction have lengths less than between about 1/16 and about 4 in (the measurement may be a volume or a weight average length). In some embodiments, the equipment used to reduce the particle size reduction is a hammer mill or shredder. Subsequent to size reduction, the feedstock is typically slurried in water, as this facilitates pumping of the feedstock. In some embodiments, lignocellulosic feedstocks of particle size less than about 6 inches do not require size reduction.

[0122] As used herein, the term "lignocellulosic feedstock" refers to any type of lignocellulosic biomass that is suitable for use as feedstock in saccharification reactions.

[0123] As used herein, the term "pretreated lignocellulosic feedstock," refers to lignocellulosic feedstocks that have been subjected to physical and/or chemical processes to make the fiber more accessible and/or receptive to the actions of cellulolytic enzymes, as described above.

[0124] As used herein, the term "recovered" refers to the harvesting, isolating, collecting, or recovering of protein from a cell and/or culture medium. In the context of saccharification, it is used in reference to the harvesting of fermentable sugars produced during the saccharification reaction from the culture medium and/or cells. In the context of fermentation, it is used in reference to harvesting the fermentation product from the culture medium and/or cells. Thus, a process can be said to comprise "recovering" a product of a reaction (such as a soluble sugar recovered from saccharification) if the process includes separating the product from other components of a reaction mixture subsequent to at least some of the product being generated in the reaction.

[0125] As used herein, the term "slurry" refers to an aqueous solution in which are dispersed one or more solid components, such as a cellulosic substrate.

[0126] As used herein, "increasing" the yield of a product (such as a fermentable sugar) from a reaction occurs when a particular component of interest is present during the reaction (*e.g.*, beta-xylosidase) causes more product to be produced, compared with a reaction conducted under the same conditions with the same substrate and other substituents, but in the absence of the component of interest (*e.g.*, without beta-xylosidase).

[0127] As used herein, a reaction is said to be "substantially free" of a particular enzyme if the amount of that enzyme compared with other enzymes that participate in catalyzing the reaction is less than about 2%, about 1%, or about 0.1% (wt/wt).

[0128] As used herein, "fractionating" a liquid (*e.g.*, a culture broth) means applying a separation process (*e.g.*, salt precipitation, column chromatography, size exclusion, and filtration) or a combination of such processes to provide a solution in which a desired protein (such as beta-xylosidase, a cellulase enzyme, and/or a combination thereof) comprises a greater percentage of total protein in the solution than in the initial liquid product.

[0129] As used herein, the term "enzymatic hydrolysis", refers to a process comprising at least one cellulases and at least one glycosidase enzyme and/or a mixture glycosidases that act on polysaccharides, (*e.g.*, cellulose), to convert all or a portion thereof to fermentable

sugars. "Hydrolyzing" cellulose or other polysaccharide occurs when at least some of the glycosidic bonds between two monosaccharides present in the substrate are hydrolyzed, thereby detaching from each other the two monomers that were previously bonded.

[0130] It is intended that the enzymatic hydrolysis be carried out with any suitable type of cellulase enzymes capable of hydrolyzing the cellulose to glucose, regardless of their source, including those obtained from fungi, such as *Trichoderma* spp., *Aspergillus* spp., *Hypocrea* spp., *Humicola* spp., *Neurospora* spp., *Orpinomyces* spp., *Gibberella* spp., *Emericella* spp., *Chaetomium* spp., *Chrysosporium* spp., *Fusarium* spp., *Penicillium* spp., *Magnaporthe* spp., *Phanerochaete* spp., *Trametes* spp., *Lentinula edodes, Gleophyllum trabeiu, Ophiostoma piliferum, Corpinus cinereus, Geomyces pannorum, Cryptococcus laurentii, Aureobasidium pullulans, Amorphotheca resinae, Leucosporidium scotti, Cunninghamella elegans, Thermomyces lanuginosus, Myceliopthora thermophila*, and *Sporotrichum thermophile*, as well as those obtained from bacteria of the genera *Bacillus, Thermomyces, Clostridium, Streptomyces* and *Thermobifida*. Cellulase compositions typically comprise one or more cellobiohydrolase, endoglucanase, and beta-glucosidase enzymes. In some cases, the cellulase compositions additionally contain hemicellulases, esterases, swollenins, cips, etc. Many of these enzymes are readily commercially available.

[0131] In some embodiments, the enzymatic hydrolysis is carried out at a pH and temperature that is at or near the optimum for the cellulase enzymes being used. For example, the enzymatic hydrolysis may be carried out at about 30°C to about 75°C, or any suitable temperature therebetween, for example a temperature of about 30°C, about 35°C, about 40°C, about 45°C, about 50°C, about 55°C, about 60°C, about 65°C, about 70°C, about 75°C, or any temperature therebetween, and a pH of about 3.5 to about 7.5, or any pH therebetween (*e.g.*, about 3.5, about 4.0, about 4.5, about 5.0, about 5.5, about 6.0, about 6.5, about 7.0, about 7.5, or any suitable pH therebetween). In some embodiments, the initial concentration of cellulose, prior to the start of enzymatic hydrolysis, is preferably about 0.1% (w/w) to about 20% (w/w), or any suitable amount therebetween (*e.g.*, about 0.1%, about 0.5%, about 1%, about 2%, about 4%, about 6%, about 8%, about 10%, about 12%, about 14%, about 15%, about 18%, about 20%, or any suitable amount therebetween.) In some embodiments, the combined dosage of all cellulase enzymes is about 0.001 to about 100 mg protein per gram cellulose, or any suitable amount therebetween (*e.g.*, about 0.001, about 0.01, about 0.1, about 1, about 5, about 10, about 15, about 20, about 25, about 30, about 40, about 50, about 60, about 70, about 80, about 90, about 100 mg protein per gram cellulose or

any amount therebetween. The enzymatic hydrolysis is carried out for any suitable time period. In some embodiments, the enzymatic hydrolysis is carried out for a time period of about 0.5 hours to about 200 hours, or any time therebetween (*e.g.*, about 2 hours to about 100 hours, or any suitable time therebetween). For example, in some embodiments, it is carried out for about 0.5, about 1, about 2, about 5, about 7, about 10, about 12, about 14, about 15, about 20, about 25, about 30, about 35, about 40, about 45, about 50, about 55, about 60, about 65, about 70, about 75, about 80, about 85, about 90, about 95, about 100, about 120, about 140, about 160, about 180, about 200, or any suitable time therebetween.)

[0132] In some embodiments, the enzymatic hydrolysis is batch hydrolysis, continuous hydrolysis, and/or a combination thereof. In some embodiments, the hydrolysis is agitated, unmixed, or a combination thereof. The enzymatic hydrolysis is typically carried out in a hydrolysis reactor. The cellulase enzyme composition is added to the pretreated lignocellulosic substrate prior to, during, or after the addition of the substrate to the hydrolysis reactor. Indeed it is not intended that reaction conditions be limited to those provided herein, as modifications are well-within the knowledge of those skilled in the art. In some embodiments, following cellulose hydrolysis, any insoluble solids present in the resulting lignocellulosic hydrolysate, including but not limited to lignin, are removed using conventional solid-liquid separation techniques prior to any further processing. In some embodiments, these solids are burned to provide energy for the entire process.

[0133] As used herein, the term "by-product" refers to an organic molecule that is an undesired product of a particular process (*e.g.*, saccharification).

DETAILED DESCRIPTION OF THE INVENTION

[0134] The present invention provides beta-xylosidase variant enzymes suitable for use in saccharification reactions. The present application further provides genetically modified fungal organisms that produce beta-xylosidase variants, as well as enzyme mixtures exhibiting enhanced hydrolysis of cellulosic material to fermentable sugars, enzyme mixtures produced by the genetically modified fungal organisms, and methods for producing fermentable sugars from cellulose using such enzyme mixtures.

[0135] In some embodiments, the present invention provides methods and compositions suitable for use in the degradation of cellulose. In some additional embodiments, the present invention provides beta-xylosidase variant enzymes suitable for use in saccharification

reactions to hydrolyze cellulose components in biomass feedstock. In some additional embodiments, the beta-xylosidase variant enzymes are used in combination with additional enzymes, including but not limited to EG1a, Eg1b, EG2, EG3, EG5, EG6, cellobiohydrolase(s), GH61 enzymes, etc., in saccharification reactions.

[0136] Fungi, bacteria, and other organisms produce a variety of cellulases and other enzymes that act in concert to catalyze decrystallization and hydrolysis of cellulose to yield fermentable sugars. One such fungus is *M. thermophila*, which is described herein. Cellulases of interest include the variant beta-xylosidase enzymes provided herein. The variant beta-xylosidase sequences provided herein are particularly useful for the production of fermentable sugars from cellulosic biomass. In some embodiments, the present invention provides methods of generating fermentable sugars from cellulosic biomass, by contacting the biomass with a cellulase composition comprising at least one variant beta-xylosidase described herein under conditions suitable for the production of fermentable sugars

[0137] For example, mutagenesis and directed evolution methods can be readily applied to polynucleotides to generate variant libraries that can be expressed, screened, and assayed. Mutagenesis and directed evolution methods are well known in the art (*See e.g.,* US Patent Nos. 5,605,793, 5,830,721, 6,132,970, 6,420,175, 6,277,638, 6,365,408, 6,602,986, 7,288,375, 6,287,861, 6,297,053, 6,576,467, 6,444,468, 5,811238, 6,117,679, 6,165,793, 6,180,406, 6,291,242, 6,995,017, 6,395,547, 6,506,602, 6,519,065, 6,506,603, 6,413,774, 6,573,098, 6,323,030, 6,344,356, 6,372,497, 7,868,138, 5,834,252, 5,928,905, 6,489,146, 6,096,548, 6,387,702, 6,391,552, 6,358,742, 6,482,647, 6,335,160, 6,653,072, 6,355,484, 6,03,344, 6,319,713, 6,613,514, 6,455,253, 6,579,678, 6,586,182, 6,406,855, 6,946,296, 7,534,564, 7,776,598, 5,837,458, 6,391,640, 6,309,883, 7,105,297, 7,795,030, 6,326,204, 6,251,674, 6,716,631, 6,528,311, 6,287,862, 6,335,198, 6,352,859, 6,379,964, 7,148,054, 7,629,170, 7,620,500, 6,365,377, 6,358,740, 6,406,910, 6,413,745, 6,436,675, 6,961,664, 7,430,477, 7,873,499, 7,702,464, 7,783,428, 7,747,391, 7,747,393, 7,751,986, 6,376,246, 6,426,224, 6,423,542, 6,479,652, 6,319,714, 6,521,453, 6,368,861, 7,421,347, 7,058,515, 7,024,312, 7,620,502, 7,853,410, 7,957,912, 7,904,249, and all related non-US counterparts; Ling *et al.*, Anal. Biochem., 254(2):157-78 [1997]; Dale *et al.*, Meth. Mol. Biol., 57:369-74 [1996]; Smith, Ann. Rev. Genet., 19:423-462 [1985]; Botstein *et al.*, Science, 229:1193-1201 [1985]; Carter, Biochem. J., 237:1-7 [1986]; Kramer *et al.*, Cell, 38:879-887 [1984]; Wells *et al.*, Gene, 34:315-323 [1985]; Minshull *et al.*, Curr. Op. Chem. Biol., 3:284-290 [1999]; Christians *et al.*, Nat. Biotechnol., 17:259-264 [1999]; Crameri *et al.*, Nature, 391:288-291

[1998]; Crameri, *et al.*, Nat. Biotechnol., 15:436-438 [1997]; Zhang *et al.*, Proc. Nat. Acad. Sci. U.S.A., 94:4504-4509 [1997]; Crameri *et al.*, Nat. Biotechnol., 14:315-319 [1996]; Stemmer, Nature, 370:389-391 [1994]; Stemmer, Proc. Nat. Acad. Sci. USA, 91:10747-10751 [1994]; WO 95/22625; WO 97/0078; WO 97/35966; WO 98/27230; WO 00/42651; WO 01/75767; and WO 2009/152336, all of which are incorporated herein by reference).

[0138] Beta-xylosidase activity and thermostability of the variants can be determined by methods described in the Examples, and/or using other suitable assay methods known in the art (e.g., the PAHBAH kit [Megazyme] and/or HPLC). Additional methods of xylose or xylobiose quantification include, but are not limited chromatographic methods (*e.g.*, HPLC; *See e.g.*, U.S. Patent Nos. 6,090,595 and 7,419,809, both of which are incorporated by reference in their entireties).

[0139] The present invention provides beta-xylosidase variants suitable for use in saccharification reactions. In some embodiments, the present invention provides methods and compositions suitable for use in the degradation of cellulose and/or hemicellulose. In some additional embodiments, the present invention provides variant beta-xylosidase enzymes suitable for use in saccharification reactions to hydrolyze cellulose components in biomass feedstock. In some additional embodiments, the variant beta-xylosidase(s) are used in combination with additional enzymes, including but not limited to at least one EG (*e.g.*, EG1b, EG1a, EG2, EG3, EG4, EG5, and/or EG6), cellobiohydrolase, GH61, and/or beta-glucosidases, etc., in saccharification reactions.

[0140] Fungi, bacteria, and other organisms produce a variety of cellulases and other enzymes that act in concert to catalyze decrystallization and hydrolysis of cellulose to yield fermentable sugars. One such fungus is *M. thermophila*, which is described herein. The variant beta-xylosidase sequences provided herein are particularly useful for the production of fermentable sugars from cellulosic biomass and other feedstocks. In some additional embodiments, the present invention provides methods for generating fermentable sugars from biomass, involving contacting the biomass with a cellulase composition comprising at least one beta-xylosidase variant as described herein, under conditions suitable for the production of fermentable sugars.

[0141] In some embodiments, the variant beta-xylosidases of the present invention further comprise additional sequences which do not alter the encoded activity of the enzyme. For

example, in some embodiments, the variant beta-xylosidases are linked to an epitope tag or to another sequence useful in purification.

[0142] In some embodiments, the variant beta-xylosidase polypeptides of the present invention are secreted from the host cell in which they are expressed (*e.g.*, a yeast or filamentous fungal host cell) and are expressed as a pre-protein including a signal peptide (*i.e.*, an amino acid sequence linked to the amino terminus of a polypeptide and which directs the encoded polypeptide into the cell secretory pathway). In some embodiments, the signal peptide is an endogenous *M. thermophila* beta-xylosidase signal peptide. In some other embodiments, signal peptides from other *M. thermophila* secreted proteins are used. In some embodiments, other signal peptides find use, depending on the host cell and other factors. Effective signal peptide coding regions for filamentous fungal host cells include, but are not limited to, the signal peptide coding regions obtained from *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* neutral amylase, *Aspergillus niger* glucoamylase, *Rhizomucor miehei* aspartic proteinase, *Humicola insolens* cellulase, *Humicola lanuginosa* lipase, and *T. reesei* cellobiohydrolase II. Signal peptide coding regions for bacterial host cells include, but are not limited to the signal peptide coding regions obtained from the genes for *Bacillus* NClB 11837 maltogenic amylase, *Bacillus stearothermophilus* alpha-amylase, *Bacillus licheniformis* subtilisin, *Bacillus licheniformis* β-lactamase, *Bacillus stearothermophilus* neutral proteases (nprT, nprS, nprM), and *Bacillus subtilis* prsA. In some additional embodiments, other signal peptides find use in the present invention (*See e.g.*, Simonen and Palva, Microbiol. Rev., 57: 109-137 [1993], incorporated herein by reference). Additional useful signal peptides for yeast host cells include those from the genes for *Saccharomyces cerevisiae* alpha-factor, *Saccharomyces cerevisiae* SUC2 invertase (*See e.g.,* Taussig and Carlson, Nucl. Acids Res., 11:1943-54 [1983]; SwissProt Accession No. P00724; and Romanos *et al.*, Yeast 8:423-488 [1992]). In some embodiments, variants of these signal peptides and other signal peptides find use. Indeed, it is not intended that the present invention be limited to any specific signal peptide, as any suitable signal peptide known in the art finds use in the present invention.

[0143] In some embodiments, the present invention provides polynucleotides encoding variant beta-xylosidase polypeptides, and/or biologically active fragments thereof, as described herein. In some embodiments, the polynucleotide is operably linked to one or more heterologous regulatory or control sequences that control gene expression to create a recombinant polynucleotide capable of expressing the polypeptide. In some embodiments,

expression constructs containing a heterologous polynucleotide encoding a variant beta-xylosidase is introduced into appropriate host cells to express the variant beta-xylosidase.

[0144] Those of ordinary skill in the art understand that due to the degeneracy of the genetic code, a multitude of nucleotide sequences encoding variant beta-xylosidase polypeptides of the present invention exist. For example, the codons AGA, AGG, CGA, CGC, CGG, and CGU all encode the amino acid arginine. Thus, at every position in the nucleic acids of the invention where an arginine is specified by a codon, the codon can be altered to any of the corresponding codons described above without altering the encoded polypeptide. It is understood that "U" in an RNA sequence corresponds to "T" in a DNA sequence. The invention contemplates and provides each and every possible variation of nucleic acid sequence encoding a polypeptide of the invention that could be made by selecting combinations based on possible codon choices.

[0145] A DNA sequence may also be designed for high codon usage bias codons (codons that are used at higher frequency in the protein coding regions than other codons that code for the same amino acid). The preferred codons may be determined in relation to codon usage in a single gene, a set of genes of common function or origin, highly expressed genes, the codon frequency in the aggregate protein coding regions of the whole organism, codon frequency in the aggregate protein coding regions of related organisms, or combinations thereof. A codon whose frequency increases with the level of gene expression is typically an optimal codon for expression. In particular, a DNA sequence can be optimized for expression in a particular host organism. A variety of methods are well-known in the art for determining the codon frequency (*e.g.*, codon usage, relative synonymous codon usage) and codon preference in specific organisms, including multivariate analysis (*e.g.*, using cluster analysis or correspondence analysis,) and the effective number of codons used in a gene. The data source for obtaining codon usage may rely on any available nucleotide sequence capable of coding for a protein. These data sets include nucleic acid sequences actually known to encode expressed proteins (*e.g.*, complete protein coding sequences-CDS), expressed sequence tags (ESTs), or predicted coding regions of genomic sequences, as is well-known in the art. Polynucleotides encoding variant beta-xylosidases can be prepared using any suitable methods known in the art. Typically, oligonucleotides are individually synthesized, then joined (*e.g.*, by enzymatic or chemical ligation methods, or polymerase-mediated methods) to form essentially any desired continuous sequence. In some embodiments, polynucleotides of the present invention are prepared by chemical synthesis using, any suitable methods known

in the art, including but not limited to automated synthetic methods. For example, in the phosphoramidite method, oligonucleotides are synthesized (*e.g.*, in an automatic DNA synthesizer), purified, annealed, ligated and cloned in appropriate vectors. In some embodiments, double stranded DNA fragments are then obtained either by synthesizing the complementary strand and annealing the strands together under appropriate conditions, or by adding the complementary strand using DNA polymerase with an appropriate primer sequence. There are numerous general and standard texts that provide methods useful in the present invention are well known to those skilled in the art.

[0146] The present invention also provides recombinant constructs comprising a sequence encoding at least one variant beta-xylosidase, as provided herein. In some embodiments, the present invention provides an expression vector comprising a variant beta-xylosidase polynucleotide operably linked to a heterologous promoter. In some embodiments, expression vectors of the present invention are used to transform appropriate host cells to permit the host cells to express the variant beta-xylosidase protein. Methods for recombinant expression of proteins in fungi and other organisms are well known in the art, and a number expression vectors are available or can be constructed using routine methods. In some embodiments, nucleic acid constructs of the present invention comprise a vector, such as, a plasmid, a cosmid, a phage, a virus, a bacterial artificial chromosome (BAC), a yeast artificial chromosome (YAC), and the like, into which a nucleic acid sequence of the invention has been inserted. In some embodiments, polynucleotides of the present invention are incorporated into any one of a variety of expression vectors suitable for expressing variant beta-xylosidase polypeptide(s). Suitable vectors include, but are not limited to chromosomal, nonchromosomal and synthetic DNA sequences (*e.g.*, derivatives of SV40), as well as bacterial plasmids, phage DNA, baculovirus, yeast plasmids, vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus, fowl pox virus, pseudorabies, adenovirus, adeno-associated virus, retroviruses, and many others. Any suitable vector that transduces genetic material into a cell, and, if replication is desired, which is replicable and viable in the relevant host finds use in the present invention. In some embodiments, the construct further comprises regulatory sequences, including but not limited to a promoter, operably linked to the protein encoding sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art. Indeed, in some embodiments, in order to obtain high levels of expression in a particular host it is often useful to express the variant beta-xylosidases of the present invention under the control of a heterologous

promoter. In some embodiments, a promoter sequence is operably linked to the 5' region of the variant beta-xylosidase coding sequence using any suitable method known in the art. Examples of useful promoters for expression of variant beta-xylosidases include, but are not limited to promoters from fungi. In some embodiments, a promoter sequence that drives expression of a gene other than a beta-xylosidase gene in a fungal strain finds use. As a non-limiting example, a fungal promoter from a gene encoding an endoglucanase may be used. In some embodiments, a promoter sequence that drives the expression of a beta-xylosidase gene in a fungal strain other than the fungal strain from which the beta-xylosidases were derived finds use. Examples of other suitable promoters useful for directing the transcription of the nucleotide constructs of the present invention in a filamentous fungal host cell include, but are not limited to promoters obtained from the genes for *Aspergillus* oryzae TAKA amylase, *Rhizomucor miehei* aspartic proteinase, *Aspergillus niger* neutral alpha-amylase, *Aspergillus niger* acid stable alpha-amylase, *Aspergillus niger* or *Aspergillus awamori* glucoamylase (glaA), *Rhizomucor miehei* lipase, *Aspergillus oryzae* alkaline protease, *Aspergillus oryzae* triose phosphate isomerase, *Aspergillus nidulans* acetamidase, and *Fusarium oxysporum* trypsin-like protease (*See e.g.,* WO 96/00787, incorporated herein by reference), as well as the NA2-tpi promoter (a hybrid of the promoters from the genes for *Aspergillus niger* neutral alpha-amylase and *Aspergillus oryzae* triose phosphate isomerase), promoters such as *cbh*1, *cbh*2, *egl*1, *egl*2, *pep*A, *hfb*1, *hfb*2, *xyn*1, *amy*, and *gla*A (*See e.g.,* Nunberg *et al.*, Mol. Cell Biol., 4:2306 -2315 [1984]; Boel *et al.*, EMBO J. 3:1581-85 [1984]; and European Patent Appln. 137280, all of which are incorporated herein by reference), and mutant, truncated, and hybrid promoters thereof. In a yeast host, useful promoters include, but are not limited to those from the genes for *Saccharomyces cerevisiae* enolase (*eno-1*), *Saccharomyces cerevisiae* galactokinase (*gal1*), *Saccharomyces cerevisiae* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase (*ADH2/GAP*), and *S. cerevisiae* 3-phosphoglycerate kinase. Additional useful promoters useful for yeast host cells are known in the art (*See e.g.*, Romanos *et al.*, Yeast 8:423-488 [1992], incorporated herein by reference). In addition, promoters associated with chitinase production in fungi find use in the present invention (*See e.g.*, Blaiseau and Lafay, Gene 120243-248 [1992]; and Limon *et al.*, Curr. Genet, 28:478-83 [1995], both of which are incorporated herein by reference).

[0147] In some embodiments, cloned variant beta-xylosidases of the present invention also have a suitable transcription terminator sequence, a sequence recognized by a host cell to terminate transcription. The terminator sequence is operably linked to the 3' terminus of the

nucleic acid sequence encoding the polypeptide. Any terminator that is functional in the host cell of choice finds use in the present invention. Exemplary transcription terminators for filamentous fungal host cells include, but are not limited to those obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* glucoamylase, *Aspergillus nidulans* anthranilate synthase, *Aspergillus niger* alpha-glucosidase, and *Fusarium oxysporum* trypsin-like protease (*See also*, US Patent No. 7,399,627, incorporated herein by reference). In some embodiments, exemplary terminators for yeast host cells include those obtained from the genes for *Saccharomyces cerevisiae* enolase, *Saccharomyces cerevisiae* cytochrome C (CYCl), and *Saccharomyces cerevisiae* glyceraldehyde-3-phosphate dehydrogenase. Other useful terminators for yeast host cells are well-known to those skilled in the art (*See e.g.*, Romanos *et al.*, Yeast 8:423-88 [1992]).

[0148] In some embodiments, a suitable leader sequence is part of a cloned variant beta-xylosidase sequence, which is a nontranslated region of an mRNA that is important for translation by the host cell. The leader sequence is operably linked to the 5' terminus of the nucleic acid sequence encoding the polypeptide. Any leader sequence that is functional in the host cell of choice finds use in the present invention. Exemplary leaders for filamentous fungal host cells include, but are not limited to those obtained from the genes for *Aspergillus oryzae* TAKA amylase and *Aspergillus nidulans* triose phosphate isomerase. Suitable leaders for yeast host cells include, but are not limited to those obtained from the genes for *Saccharomyces cerevisiae* enolase (ENO-1), *Saccharomyces cerevisiae* 3-phosphoglycerate kinase, *Saccharomyces cerevisiae* alpha-factor, and *Saccharomyces cerevisiae* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase (ADH2/GAP).

[0149] In some embodiments, the sequences of the present invention also comprise a polyadenylation sequence, which is a sequence operably linked to the 3' terminus of the nucleic acid sequence and which, when transcribed, is recognized by the host cell as a signal to add polyadenosine residues to transcribed mRNA. Any polyadenylation sequence which is functional in the host cell of choice finds use in the present invention. Exemplary polyadenylation sequences for filamentous fungal host cells include, but are not limited to those obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* glucoamylase, *Aspergillus nidulans* anthranilate synthase, *Fusarium oxysporum* trypsin-like protease, and *Aspergillus niger* alpha-glucosidase. Useful polyadenylation sequences for yeast host cells are known in the art (*See e.g.*, Guo and Sherman, Mol Cell Biol., 15:5983-5990 [1995]).

[0150] In some embodiments, the expression vector of the present invention contains one or more selectable markers, which permit easy selection of transformed cells. A "selectable marker" is a gene, the product of which provides for biocide or viral resistance, resistance to antimicrobials or heavy metals, prototrophy to auxotrophs, and the like. Any suitable selectable markers for use in a filamentous fungal host cell find use in the present invention, including, but are not limited to, amdS (acetamidase), argB (ornithine carbamoyltransferase), bar (phosphinothricin acetyltransferase), hph (hygromycin phosphotransferase), niaD (nitrate reductase), pyrG (orotidine-5'-phosphate decarboxylase), sC (sulfate adenyltransferase), and trpC (anthranilate synthase), as well as equivalents thereof. Additional markers useful in host cells such as *Aspergillus*, include but are not limited to the amdS and pyrG genes of *Aspergillus nidulans* or *Aspergillus oryzae* and the bar gene of *Streptomyces hygroscopicus*. Suitable markers for yeast host cells include, but are not limited to ADE2, HIS3, LEU2, LYS2, MET3, TRP1, and URA3.

[0151] In some embodiments, a vector comprising a sequence encoding at least one variant beta-xylosidase is transformed into a host cell in order to allow propagation of the vector and expression of the variant beta-xylosidase(s). In some embodiments, the variant beta-xylosidases are post-translationally modified to remove the signal peptide and in some cases may be cleaved after secretion. In some embodiments, the transformed host cell described above is cultured in a suitable nutrient medium under conditions permitting the expression of the variant beta-xylosidase(s). Any suitable medium useful for culturing the host cells finds use in the present invention, including, but not limited to minimal or complex media containing appropriate supplements. In some embodiments, host cells are grown in HTP media. Suitable media are available from various commercial suppliers or may be prepared according to published recipes (*e.g.,* in catalogues of the American Type Culture Collection).

[0152] In some embodiments, the host cell is a eukaryotic cell. Suitable eukaryotic host cells include, but are not limited to, fungal cells, algal cells, insect cells, and plant cells. Suitable fungal host cells include, but are not limited to, Ascomycota, Basidiomycota, Deuteromycota, Zygomycota, Fungi imperfecti. In some embodiments, the fungal host cells are yeast cells and filamentous fungal cells. The filamentous fungal host cells of the present invention include all filamentous forms of the subdivision Eumycotina and Oomycota. Filamentous fungi are characterized by a vegetative mycelium with a cell wall composed of chitin, cellulose and other complex polysaccharides. The filamentous fungal host cells of the present invention are morphologically distinct from yeast.

[0153] In some embodiments of the present invention, the filamentous fungal host cells are of any suitable genus and species, including, but not limited to *Achlya*, *Acremonium*, *Aspergillus*, *Aureobasidium*, *Bjerkandera*, *Ceriporiopsis, Cephalosporium*, *Chrysosporium*, *Cochliobolus*, *Corynascus*, *Cryphonectria*, *Cryptococcus*, *Coprinus*, *Coriolus*, *Diplodia*, *Endothis*, *Fusarium*, *Gibberella*, *Gliocladium*, *Humicola*, *Hypocrea*, *Myceliophthora*, *Mucor*, *Neurospora*, *Penicillium*, *Podospora*, *Phlebia*, *Piromyces*, *Pyricularia*, *Rhizomucor*, *Rhizopus*, *Schizophyllum*, *Scytalidium*, *Sporotrichum, Talaromyces*, *Thermoascus*, *Thielavia*, *Trametes*, *Tolypocladium*, *Trichoderma*, *Verticillium*, and/or *Volvariella*, and/or teleomorphs, or anamorphs, and synonyms, basionyms, or taxonomic equivalents thereof.

[0154] In some embodiments of the present invention, the filamentous fungal host cell is of the *Trichoderma* species (*e.g., T. longibrachiatum*, T. *viride* [*e.g.*, ATCC 32098 and 32086]), *Hypocrea jecorina* or *T. reesei* (NRRL 15709, ATTC 13631, 56764, 56765, 56466, 56767 and RL-P37 and derivatives thereof (*See e.g.*, Sheir-Neiss *et al.*, Appl. Microbiol. Biotechnol., 20:46 – 53 [1984]), *T. koningii*, and *T. harzianum*. In addition, the term "*Trichoderma*" refers to any fungal strain that was previously and/or currently classified as *Trichoderma*. In some embodiments of the present invention, the filamentous fungal host cell is of the *Aspergillus* species (*e.g., A. awamori*, *A. fumigatus*, *A. japonicus*, *A. nidulans*, *A. niger*, *A. aculeatus, A. foetidus, A. oryzae*, *A. sojae*, and *A. kawachi; See e.g.,* Kelly and Hynes, EMBO J., 4:475-479 [1985]; NRRL 3112, ATCC 11490, 22342, 44733, and 14331; Yelton *et al.*, Proc. Natl. Acad. Sci. USA, 81, 1470-1474 [1984]; Tilburn *et al.*, Gene 26:205-221 [1982]; and Johnston, *et al.,* EMBO J., 4:1307-1311 [1985]). In some embodiments of the present invention, the filamentous fungal host cell is a *Chrysosporium* species (*e.g.*, *C. lucknowense*, *C. keratinophilum*, *C. tropicum*, *C. merdarium*, *C. inops*, *C. pannicola*, and *C. zonatum*). In some embodiments of the present invention, the filamentous fungal host cell is a *Myceliophthora* species (*e.g.*, *M. thermophila*). In some embodiments of the present invention, the filamentous fungal host cell is a *Fusarium* species (*e.g.*, *F. bactridioides*, *F. cerealis*, *F. crookwellense*, *F. culmorum*, *F. graminearum*, *F. graminum*. *F. oxysporum*, *F. roseum*, and *F. venenatum*). In some embodiments of the present invention, the filamentous fungal host cell is a *Neurospora* species (*e.g.*, *N. crassa*; *See e.g.*, Case *et al.*, Proc. Natl. Acad. Sci. USA, 76:5259-5263 [1979]; US Pat. No. 4,486,553; and Kinsey and Rambosek (1984) Mol. Cell. Biol., 4:117–122 [1984], all of which are hereby incorporated by reference). In some embodiments of the present invention, the filamentous fungal host cell is a *Humicola* species (*e.g.*, *H. insolens*, *H. grisea*, and *H. lanuginosa*). In some embodiments

of the present invention, the filamentous fungal host cell is a *Mucor* species (*e.g.*, *M. miehei* and *M. circinelloides*). In some embodiments of the present invention, the filamentous fungal host cell is a *Rhizopus* species (*e.g.*, *R. oryzae* and *R.niveus*.). In some embodiments of the invention, the filamentous fungal host cell is a *Penicillium* species (*e.g.*, *P. purpurogenum*, *P. chrysogenum*, and *P. verruculosum*). In some embodiments of the invention, the filamentous fungal host cell is a *Talaromyces* species (*e.g., T. emersonii*, *T. flavus*, *T. helicus*, T. *rotundus*, and *T. stipitatus*). In some embodiments of the invention, the filamentous fungal host cell is a *Thielavia* species (*e.g.*, *T. terrestris* and *T. heterothallica*). In some embodiments of the present invention, the filamentous fungal host cell is a *Tolypocladium* species (*e.g.*, *T. inflatum* and *T. geodes*). In some embodiments of the present invention, the filamentous fungal host cell is a *Trametes* species (*e.g.*, *T. villosa* and *T. versicolor*). In some embodiments of the present invention, the filamentous fungal host cell is a *Sporotrichum* species. In some embodiments of the present invention, the filamentous fungal host cell is a *Corynascus* species.

[0155] In some embodiments of the present invention, the host cell is a yeast cell, including but not limited to cells of *Candida*, *Hansenula*, *Saccharomyces*, *Schizosaccharomyces*, *Pichia*, *Kluyveromyces*, or *Yarrowia* species. In some embodiments of the present invention, the yeast cell is *Hansenula polymorpha*, *Saccharomyces cerevisiae*, *Saccharomyces carlsbergensis*, *Saccharomyces diastaticus*, *Saccharomyces norbensis*, *Saccharomyces kluyveri*, *Schizosaccharomyces pombe*, *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia kodamae*, *Pichia membranaefaciens*, *Pichia opuntiae*, *Pichia thermotolerans*, *Pichia salictaria, Pichia quercuum, Pichia pijperi*, *Pichia stipitis*, *Pichia methanolica*, *Pichia angusta*, *Kluyveromyces lactis*, *Candida albicans*, or *Yarrowia lipolytica*.

[0156] In some embodiments of the invention, the host cell is an algal cell such as *Chlamydomonas* (*e.g.*, *C. reinhardtii*) and *Phormidium* (P. sp. ATCC29409).

[0157] In some other embodiments, the host cell is a prokaryotic cell. Suitable prokaryotic cells include, but are not limited to Gram-positive, Gram-negative and Gram-variable bacterial cells. Any suitable bacterial organism finds use in the present invention, including but not limited to *Agrobacterium*, *Alicyclobacillus*, *Anabaena*, *Anacystis, Acinetobacter, Acidothermus, Arthrobacter, Azobacter, Bacillus, Bifidobacterium, Brevibacterium, Butyrivibrio, Buchnera, Campestris, Camplyobacter, Clostridium, Corynebacterium,*

Chromatium, Coprococcus, Escherichia, Enterococcus, Enterobacter, Erwinia, Fusobacterium, Faecalibacterium, Francisella, Flavobacterium, Geobacillus, Haemophilus, Helicobacter, Klebsiella, Lactobacillus, Lactococcus, Ilyobacter, Micrococcus, Microbacterium, Mesorhizobium, Methylobacterium, Methylobacterium, Mycobacterium, Neisseria, Pantoea, Pseudomonas, Prochlorococcus, Rhodobacter, Rhodopseudomonas, Rhodopseudomonas, Roseburia, Rhodospirillum, Rhodococcus, Scenedesmus, Streptomyces, Streptococcus, Synecoccus, Saccharomonospora, Staphylococcus, Serratia, Salmonella, Shigella, Thermoanaerobacterium, Tropheryma, Tularensis, Temecula, Thermosynechococcus, Thermococcus, Ureaplasma, Xanthomonas, Xylella, Yersinia and *Zymomonas*. In some embodiments, the host cell is a species of *Agrobacterium*, *Acinetobacter*, *Azobacter*, *Bacillus*, *Bifidobacterium*, *Buchnera*, *Geobacillus*, *Campylobacter*, *Clostridium*, *Corynebacterium*, *Escherichia*, *Enterococcus*, *Erwinia*, *Flavobacterium*, *Lactobacillus*, *Lactococcus*, *Pantoea*, *Pseudomonas*, *Staphylococcus*, *Salmonella*, *Streptococcus*, *Streptomyces*, or *Zymomonas*. In some embodiments, the bacterial host strain is non-pathogenic to humans. In some embodiments the bacterial host strain is an industrial strain. Numerous bacterial industrial strains are known and suitable in the present invention. In some embodiments of the present invention, the bacterial host cell is an *Agrobacterium* species (*e.g., A. radiobacter*, *A. rhizogenes*, and *A. rubi*). In some embodiments of the present invention, the bacterial host cell is an *Arthrobacter* species (*e.g.*, *A. aurescens*, *A. citreus*, *A. globformis*, *A. hydrocarboglutamicus*, *A. mysorens*, *A. nicotianae*, *A. paraffineus*, *A. protophonniae*, *A. roseoparqffinus*, *A. sulfureus*, and *A. ureafaciens).* In some embodiments of the present invention, the bacterial host cell is a *Bacillus* species (*e.g.*, *B. thuringensis*, *B. anthracis*, *B. megaterium*, *B. subtilis*, *B. lentus*, *B. circulans*, *B. pumilus*, *B. lautus*, *B.coagulans, B. brevis, B. firmus, B. alkaophius*, *B. licheniformis*, *B. clausii*, *B. stearothermophilus*, *B. halodurans,* and *B. amyloliquefaciens*). In some embodiments, the host cell is an industrial *Bacillus* strain including but not limited to *B. subtilis*, *B. pumilus*, *B. licheniformis*, *B. megaterium, B. clausii*, *B. stearothermophilus,* or *B. amyloliquefaciens.* In some embodiments, the *Bacillus* host cells are *B. subtilis, B. licheniformis*, *B. megaterium, B. stearothermophilus,* and/or *B. amyloliquefaciens.* In some embodiments, the bacterial host cell is a *Clostridium* species *(e.g., C. acetobutylicum*, *C. tetani* E88, *C. lituseburense*, *C. saccharobutylicum*, *C. perfringens*, and *C. beijerinckii*). In some embodiments, the bacterial host cell is a *Corynebacterium* species (*e.g.*, *C. glutamicum* and *C. acetoacidophilum*). In some embodiments the bacterial host cell is a *Escherichia* species (*e.g.*, *E. coli*). In some embodiments, the bacterial host cell is an *Erwinia* species (*e.g.*, *E. uredovora*, *E. carotovora*,

E. ananas, *E. herbicola*, *E. punctata*, and *E. terreus*). In some embodiments, the bacterial host cell is a *Pantoea* species (*e.g.*, *P. citrea,* and *P. agglomerans).* In some embodiments the bacterial host cell is a *Pseudomonas* species (*e.g.*, *P. putida*, *P. aeruginosa*, *P. mevalonii*, and *P. sp.* D-0l 10). In some embodiments, the bacterial host cell is a *Streptococcus* species (*e.g.*, *S. equisimiles*, *S. pyogenes*, and *S. uberis*). In some embodiments, the bacterial host cell is a *Streptomyces* species (*e.g.*, *S. ambofaciens*, *S. achromogenes*, *S. avermitilis*, *S. coelicolor*, *S. aureofaciens, S. aureus*, *S. fungicidicus, S. griseus*, and *S. lividans*). In some embodiments, the bacterial host cell is a *Zymomonas* species (*e.g.*, *Z. mobilis*, and *Z. lipolytica*).

[0158] Many prokaryotic and eukaryotic strains that find use in the present invention are readily available to the public from a number of culture collections such as American Type Culture Collection (ATCC), Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSM), Centraalbureau Voor Schimmelcultures (CBS), and Agricultural Research Service Patent Culture Collection, Northern Regional Research Center (NRRL).

[0159] In some embodiments, host cells are genetically modified to have characteristics that improve protein secretion, protein stability and/or other properties desirable for expression and/or secretion of a protein. For example, knockout of Alp1 function results in a cell that is protease deficient. Knockout of pyr5 function results in a cell with a pyrimidine deficient phenotype. In some embodiments, the host cells are modified to delete endogenous cellulase protein-encoding sequences or otherwise eliminate expression of one or more endogenous cellulases. In some embodiments, expression of one or more endogenous cellulases is inhibited to increase production of cellulases of interest. Genetic modification can be achieved by genetic engineering techniques and/or classical microbiological techniques (*e.g.*, chemical or UV mutagenesis and subsequent selection). Indeed, in some embodiments, combinations of recombinant modification and classical selection techniques are used to produce the host cells. Using recombinant technology, nucleic acid molecules can be introduced, deleted, inhibited or modified, in a manner that results in increased yields of beta-xylosidase variant(s) within the host cell and/or in the culture medium. For example, knockout of Alp1 function results in a cell that is protease deficient, and knockout of pyr5 function results in a cell with a pyrimidine deficient phenotype. In one genetic engineering approach, homologous recombination is used to induce targeted gene modifications by specifically targeting a gene *in vivo* to suppress expression of the encoded protein. In alternative approaches, siRNA, antisense and/or ribozyme technology find use in inhibiting gene expression.

[0160] In some embodiments, host cells (*e.g., Myceliophthora thermophila*) used for expression of variant beta-xylosidases have been genetically modified to reduce the amount of endogenous cellobiose dehydrogenase (EC 1.1.3.4) and/or other enzymes activity that is secreted by the cell. A variety of methods are known in the art for reducing expression of protein in cells, including, but not limited to deletion of all or part of the gene encoding the protein and site-specific mutagenesis to disrupt expression or activity of the gene product. (*See e.g.*, Chaveroche *et al.*, Nucl. Acids Res., 28:22 e97 [2000]; Cho *et al.*, *MPMI* 19: 1:7-15 [2006]; Maruyama and Kitamoto, Biotechnol Lett., 30:1811-1817 [2008]; Takahashi *et al.*, *Mol. Gen. Genom.,* 272: 344-352 [2004]; and You *et al.* , Arch Micriobiol.,191:615-622 [2009], all of which are incorporated by reference herein). Random mutagenesis, followed by screening for desired mutations also finds use (*See e.g.*, Combier *et al.*, FEMS Microbiol Lett 220:141–8 [2003]; and Firon *et al.*, Eukary. Cell 2:247–55 [2003], both of which are incorporated by reference). In some embodiments, the host cell is modified to reduce production of endogenous cellobiose dehydrogenases. In some embodiments, the cell is modified to reduce production of cellobiose dehydrogenase (*e.g.,* CDH1 or CDH2). In some embodiments, the host cell has less than 75%, sometimes less than 50%, sometimes less than 30%, sometimes less than 25%, sometimes less than 20%, sometimes less than 15%, sometimes less than 10%, sometimes less than 5%, and sometimes less than 1% of the cellobiose dehydrogenase (*e.g.,* CDH1 and/or CDH2) activity of the corresponding cell in which the gene is not disrupted. Exemplary *Myceliophthora thermophila* cellobiose dehydrogenases include, but are not limited to CDH1 and CDH2. The genomic sequence for the *Cdh1* encoding CDH1 has accession number AF074951.1. In one approach, gene disruption is achieved using genomic flanking markers (*See e.g.*, Rothstein, Meth. Enzymol., 101:202-11 [1983]). In some embodiments, site-directed mutagenesis is used to target a particular domain of a protein, in some cases, to reduce enzymatic activity (*e.g.,* glucose-methanol-choline oxido-reductase N and C domains of a cellobiose dehydrogenase or heme binding domain of a cellobiose dehydrogenase; *See e.g.*, Rotsaert *et al.*, Arch. Biochem. Biophys., 390:206-14 [2001], which is incorporated by reference herein in its entirety).

[0161] Introduction of a vector or DNA construct into a host cell can be accomplished using any suitable method known in the art, including but not limited to calcium phosphate transfection, DEAE-Dextran mediated transfection, PEG-mediated transformation, electroporation, or other common techniques known in the art.

[0162] In some embodiments, the engineered host cells (*i.e.*, "recombinant host cells") of the present invention are cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants, or amplifying the cellobiohydrolase polynucleotide. Culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and are well-known to those skilled in the art. As noted, many standard references and texts are available for the culture and production of many cells, including cells of bacterial, plant, animal (especially mammalian) and archebacterial origin.

[0163] In some embodiments, cells expressing the variant beta-xylosidase polypeptides of the invention are grown under batch or continuous fermentations conditions. Classical "batch fermentation" is a closed system, wherein the compositions of the medium is set at the beginning of the fermentation and is not subject to artificial alternations during the fermentation. A variation of the batch system is a "fed-batch fermentation" which also finds use in the present invention. In this variation, the substrate is added in increments as the fermentation progresses. Fed-batch systems are useful when catabolite repression is likely to inhibit the metabolism of the cells and where it is desirable to have limited amounts of substrate in the medium. Batch and fed-batch fermentations are common and well known in the art. "Continuous fermentation" is an open system where a defined fermentation medium is added continuously to a bioreactor and an equal amount of conditioned medium is removed simultaneously for processing. Continuous fermentation generally maintains the cultures at a constant high density where cells are primarily in log phase growth. Continuous fermentation systems strive to maintain steady state growth conditions. Methods for modulating nutrients and growth factors for continuous fermentation processes as well as techniques for maximizing the rate of product formation are well known in the art of industrial microbiology.

[0164] In some embodiments of the present invention, cell-free transcription/translation systems find use in producing variant beta-xylosidase(s). Several systems are commercially available and the methods are well-known to those skilled in the art.

[0165] The present invention provides methods of making variant beta-xylosidase polypeptides or biologically active fragments thereof. In some embodiments, the method comprises: providing a host cell transformed with a polynucleotide encoding an amino acid sequence that comprises at least about 70% (or at least about 75%, at least about 80%, at least

about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%) sequence identity to SEQ ID NO:2 and comprising at least one mutation as provided herein; culturing the transformed host cell in a culture medium under conditions in which the host cell expresses the encoded variant beta-xylosidase polypeptide; and optionally recovering or isolating the expressed variant beta-xylosidase polypeptide, and/or recovering or isolating the culture medium containing the expressed variant beta-xylosidase polypeptide. In some embodiments, the methods further provide optionally lysing the transformed host cells after expressing the encoded beta-xylosidase polypeptide and optionally recovering and/or isolating the expressed variant beta-xylosidase polypeptide from the cell lysate. The present invention further provides methods of making a variant beta-xylosidase polypeptide comprising cultivating a host cell transformed with a variant beta-xylosidase polypeptide under conditions suitable for the production of the variant beta-xylosidase polypeptide and recovering the variant beta-xylosidase polypeptide. Typically, recovery or isolation of the beta-xylosidase polypeptide is from the host cell culture medium, the host cell or both, using protein recovery techniques that are well known in the art, including those described herein. In some embodiments, host cells are harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including, but not limited to freeze-thaw cycling, sonication, mechanical disruption, and/or use of cell lysing agents, as well as many other suitable methods well known to those skilled in the art.

[0166] In some embodiments, the resulting polypeptide is recovered/isolated and optionally purified by any of a number of methods known in the art. For example, in some embodiments, the polypeptide is isolated from the nutrient medium by conventional procedures including, but not limited to, centrifugation, filtration, extraction, spray-drying, evaporation, chromatography (*e.g.*, ion exchange, affinity, hydrophobic interaction, chromatofocusing, and size exclusion), or precipitation. In some embodiments, protein refolding steps are used, as desired, in completing the configuration of the mature protein. In addition, in some embodiments, high performance liquid chromatography (HPLC) is employed in the final purification steps. For example, in some embodiments, methods for purifying BGL known in the art, find use in the present invention (*See e.g.*, Parry *et al.*, Biochem. J., 353:117 [2001]; and Hong *et al.*, Appl. Microbiol. Biotechnol., 73:1331 [2007],

both incorporated herein by reference). Indeed, any suitable purification methods known in the art find use in the present invention.

[0167] In some embodiments, immunological methods are used to purify beta-xylosidase variants. In one approach, antibody raised against a variant beta-xylosidase polypeptide (*e.g.,* against a polypeptide comprising any of SEQ ID NOS:2, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, and/or 55, and/or an immunogenic fragment thereof) using conventional methods is immobilized on beads, mixed with cell culture media under conditions in which the variant beta-xylosidase is bound, and precipitated. In a related approach, immunochromatography finds use.

[0168] In some embodiments, the variant beta-xylosidases are expressed as a fusion protein including a non-enzyme portion. In some embodiments, the variant beta-xylosidase sequence is fused to a purification facilitating domain. As used herein, the term "purification facilitating domain" refers to a domain that mediates purification of the polypeptide to which it is fused. Suitable purification domains include, but are not limited to metal chelating peptides, histidine-tryptophan modules that allow purification on immobilized metals, a sequence which binds glutathione (*e.g.,* GST), a hemagglutinin (HA) tag (corresponding to an epitope derived from the influenza hemagglutinin protein; *See e.g.*, Wilson *et al.*, Cell 37:767 [1984]), maltose binding protein sequences, the FLAG epitope utilized in the FLAGS extension/affinity purification system (*e.g.*, the system available from Immunex Corp, Seattle, WA), and the like. One expression vector contemplated for use in the compositions and methods described herein provides for expression of a fusion protein comprising a polypeptide of the invention fused to a polyhistidine region separated by an enterokinase cleavage site. The histidine residues facilitate purification on IMIAC (immobilized metal ion affinity chromatography; *See e.g.*, Porath *et al.*, Prot. Exp. Purif., 3:263-281 [1992]) while the enterokinase cleavage site provides a means for separating the variant beta-xylosidase polypeptide from the fusion protein. pGEX vectors (Promega; Madison, Wis.) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to ligand-agarose beads (*e.g.,* glutathione-agarose in the case of GST-fusions) followed by elution in the presence of free ligand.

[0169] The variant beta-xylosidases and biologically active fragments thereof as described herein have multiple industrial applications, including but not limited to, sugar production

(*e.g.*, glucose syrups), biofuels production, textile treatment, pulp or paper treatment, bio-based chemical production, and applications in detergents and/or animal feed. A host cell containing at least one variant beta-xylosidase of the present invention finds use without recovery and purification of the recombinant variant beta-xylosidase(s) (*e.g.*, for use in a large scale biofermentor). Alternatively, recombinant variant beta-xylosidases are produced and purified from the host cell.

[0170] The variant beta-xylosidases provided herein are particularly useful in methods used to break down cellulose to smaller oligosaccharides, disaccharides and monosaccharides. In some embodiments, the variant beta-xylosidases are used in saccharification methods. In some embodiments, the variant beta-xylosidases are used in combination with other cellulase enzymes in conventional enzymatic saccharification methods to produce fermentable sugars. In some embodiments, the present invention provides methods for producing at least one end-product from a cellulosic substrate, the methods comprising contacting the cellulosic substrate with at least one variant beta-xylosidase as described herein (and optionally other cellulases) under conditions in which fermentable sugars are produced. The fermentable sugars are then used in a fermentation reaction comprising a microorganism (*e.g.*, a yeast) to produce at least one end-product. In some embodiments, the methods further comprise pretreating the cellulosic substrate to increase its susceptibility to hydrolysis prior to contacting the cellulosic substrate with at least one variant beta-xylosidase (and optionally other cellulases).

[0171] In some embodiments, enzyme compositions comprising at least one variant beta-xylosidase of the present invention are reacted with a biomass substrate in the range of about 25°C to about 100°C, about 30°C to about 90°C, about 30°C to about 80°C, or about 30°C to about 70°C. Also the biomass may be reacted with the enzyme compositions at about 25°C, at about 30°C, at about 35°C, at about 40°C, at about 45°C, at about 50°C, at about 55°C, at about 60°C, at about 65°C, at about 70°C, at about 75°C, at about 80°C, at about 85°C, at about 90°C, at about 95°C and at about 100°C. Generally the pH range will be from about pH 3.0 to about 8.5, about pH 3.5 to about 8.5, about pH 4.0 to about 7.5, about pH 4.0 to about 7.0 and about pH 4.0 to about 6.5. In some embodiments, the incubation time varies (*e.g.*, from about 1.0 to about 240 hours, from about 5.0 to about 180 hrs and from about 10.0 to about 150 hrs). In some embodiments, the incubation time is at least about 1 hr, at least about 5 hrs, at least about 10 hrs, at least about 15 hrs, at least about 25 hrs, at least about 50 hr, at least about 100 hrs, at least about 180 hrs, etc. In some embodiments, incubation of the

cellulase under these conditions and subsequent contact with the substrate results in the release of substantial amounts of fermentable sugars from the substrate (*e.g.*, glucose when the cellulase is combined with β-glucosidase). For example, in some embodiments, at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, or more fermentable sugar is available as compared to the release of sugar by a reference enzyme.

[0172] In some embodiments, an "end-product of fermentation" is any product produced by a process including a fermentation step using a fermenting organism. Examples of end-products of a fermentation include, but are not limited to, alcohols (*e.g.,* fuel alcohols such as ethanol and butanol), organic acids (*e.g.*, citric acid, acetic acid, acrylic acid, lactic acid, gluconic acid, and succinic acid), glycerol, ketones, diols, amino acids (*e.g.*, glutamic acid), antibiotics (*e.g.,* penicillin and tetracycline), vitamins (*e.g.*, beta-carotene and B12), hormones, and fuel molecules other than alcohols (*e.g.,* hydrocarbons).

[0173] In some embodiments, the fermentable sugars produced by the methods of the present invention are used to produce at least one alcohol (*e.g.*, ethanol, butanol, etc.). The variant beta-xylosidases of the present invention find use in any method suitable for the generation of alcohols or other biofuels from cellulose. It is not intended that the present invention be limited to the specific methods provided herein. Two methods commonly employed are separate saccharification and fermentation (SHF) methods (*See e.g.*, Wilke *et al.*, Biotechnol. Bioengin., 6:155-75 [1976]) and simultaneous saccharification and fermentation (SSF) methods (*See e.g.,* U.S. Pat. Nos. 3,990,944 and 3,990,945). In some embodiments, the SHF saccharification method comprises the steps of contacting a cellulase with a cellulose containing substrate to enzymatically break down cellulose into fermentable sugars (*e.g.*, monosaccharides such as glucose), contacting the fermentable sugars with an alcohol-producing microorganism to produce alcohol (*e.g.*, ethanol or butanol) and recovering the alcohol. In some embodiments, the method of consolidated bioprocessing (CBP) finds use, in which the cellulase production from the host is simultaneous with saccharification and fermentation either from one host or from a mixed cultivation. In addition, SSF methods find use in the present invention. In some embodiments, SSF methods provide a higher efficiency of alcohol production than that provided by SHF methods (*See e.g.*, Drissen *et al.,* Biocat. Biotrans., 27:27-35 [2009]).

[0174] In some embodiments, for cellulosic substances to be effectively used as substrates for the saccharification reaction in the presence of a cellulase of the present invention, it is desirable to pretreat the substrate. Means of pretreating a cellulosic substrate are well-known in the art, including but not limited to chemical pretreatment (*e.g.,* ammonia pretreatment, dilute acid pretreatment, dilute alkali pretreatment, or solvent exposure), physical pretreatment (*e.g.,* steam explosion or irradiation), mechanical pretreatment (*e.g.,* grinding or milling) and biological pretreatment (*e.g.,* application of lignin-solubilizing microorganisms), and the present invention is not limited by such methods.

[0175] In some embodiments, any suitable alcohol-producing microorganism known in the art (*e.g.*, *Saccharomyces cerevisiae*), finds use in the present invention for the fermentation of fermentable sugars to alcohols and other end-products. The fermentable sugars produced from the use of the variant beta-xylosidase(s) provided by the present invention find use in the production of other end-products besides alcohols, including, but not limited to biofuels and/or biofuels compounds, acetone, amino acids (*e.g.*, glycine, lysine, etc.), organic acids (*e.g.*, lactic acids, etc.), glycerol, ascorbic acid, diols (*e.g.,* 1,3-propanediol, butanediol, etc.), vitamins, hormones, antibiotics, other chemicals, and animal feeds. In addition, the variant beta-xylosidases provided herein further find use in the pulp and paper industry. Indeed, it is not intended that the present invention be limited to any particular end-products.

[0176] In some embodiments, the present invention provides an enzyme mixture that comprises at least one variant beta-xylosidase polypeptide as provided herein. The enzyme mixture may be cell-free, or in alternative embodiments, may not be separated from host cells that secrete an enzyme mixture component. A cell-free enzyme mixture typically comprises enzymes that have been separated from cells. Cell-free enzyme mixtures can be prepared by any of a variety of methodologies that are known in the art, such as filtration or centrifugation methodologies. In some embodiments, the enzyme mixtures are partially cell-free, substantially cell-free, or entirely cell-free.

[0177] In some embodiments, at least one variant beta-xylosidase and any additional enzymes present in the enzyme mixture are secreted from a single genetically modified fungal cell or by different microbes in combined or separate fermentations. Similarly, in additional embodiments, the variant beta-xylosidase(s) and any additional enzymes present in the enzyme mixture are expressed individually or in sub-groups from different strains of different organisms and the enzymes are combined *in vitro* to make the enzyme mixture. It is

also contemplated that the variant beta-xylosidase(s) and any additional enzymes in the enzyme mixture will be expressed individually or in sub-groups from different strains of a single organism, and the enzymes combined to make the enzyme mixture. In some embodiments, all of the enzymes are expressed from a single host organism, such as a genetically modified fungal cell.

[0178] In some embodiments, the enzyme mixture comprises at least one cellulase, selected from cellobiohydrolase (CBH), endoglucanase (EG), glycoside hydrolase 61 (GH61) and/or beta-glucosidase (BGL). In some embodiments, the cellobiohydrolase is *T. reesei* cellobiohydrolase II. In some embodiments, the endoglucanase comprises a catalytic domain derived from the catalytic domain of a *Streptomyces avermitilis* endoglucanase. In some embodiments, at least one cellulase is *Acidothermus cellulolyticus*, *Thermobifida fusca*, *Humicola grisea,* and/or a *Chrysosporium* sp. cellulase. Cellulase enzymes of the cellulase mixture work together in decrystallizing and hydrolyzing the cellulose from a biomass substrate to yield fermentable sugars, such as but not limited to glucose (*See e.g.*, Brigham *et al. in* Wyman ([ed.], Handbook on Bioethanol, Taylor and Francis, Washington DC [1995], pp 119–141, incorporated herein by reference). Indeed, it is not intended that the present invention be limited to any enzyme compositions comprising any particular cellulase component(s), as various combinations of cellulases find use in the enzyme compositions of the present invention.

[0179] Cellulase mixtures for efficient enzymatic hydrolysis of cellulose are known (*See e.g.*, Viikari *et al.*, Adv. Biochem. Eng. Biotechnol.,108:121-45 [2007]; and US Pat. Publns. 2009/0061484; US 2008/0057541; and US 2009/0209009, each of which is incorporated herein by reference). In some embodiments, mixtures of purified naturally occurring or recombinant enzymes are combined with cellulosic feedstock or a product of cellulose hydrolysis. In some embodiments, one or more cell populations, each producing one or more naturally occurring or recombinant cellulases, are combined with cellulosic feedstock or a product of cellulose hydrolysis.

[0180] In some embodiments, at least one variant beta-xylosidase polypeptide of the present invention is present in mixtures comprising enzymes other than cellulases that degrade cellulose, hemicellulose, pectin, and/or lignocellulose.

[0181] Cellulase mixtures for efficient enzymatic hydrolysis of cellulose are known (*See e.g.*, Viikari *et al.*, Adv. Biochem. Eng. Biotechnol.,108:121-45 [2007]; and US Pat. Publns.

2009/0061484; US 2008/0057541; and US 2009/0209009, each of which is incorporated herein by reference). In some embodiments, mixtures of purified naturally occurring or recombinant enzymes are combined with cellulosic feedstock or a product of cellulose hydrolysis. In some embodiments, one or more cell populations, each producing one or more naturally occurring or recombinant cellulases, are combined with cellulosic feedstock or a product of cellulose hydrolysis.

[0182] In some additional embodiments, the present invention provides at least one variant beta-xylosidase and at least one endoxylanase. Endoxylanases (EC 3.2.1.8) catalyze the endohydrolysis of 1,4-β-D-xylosidic linkages in xylans. This enzyme may also be referred to as endo-1,4-β-xylanase or 1,4-β-D-xylan xylanohydrolase. In some embodiments, an alternative is EC 3.2.1.136, a glucuronoarabinoxylan endoxylanase, an enzyme that is able to hydrolyze 1,4 xylosidic linkages in glucuronoarabinoxylans.

[0183] In some additional embodiments, the present invention provides at least one variant beta-xylosidase and at least one alpha-L-arabinofuranosidase. Alpha-L-arabinofuranosidases (EC 3.2.1.55) catalyze the hydrolysis of terminal non-reducing alpha-L-arabinofuranoside residues in alpha-L-arabinosides. The enzyme acts on alpha-L-arabinofuranosides, alpha-L-arabinans containing (1,3)- and/or (1,5)-linkages, arabinoxylans, and arabinogalactans. Alpha-L-arabinofuranosidase is also known as arabinosidase, alpha-arabinosidase, alpha-L-arabinosidase, alpha-arabinofuranosidase, arabinofuranosidase, polysaccharide alpha-L-arabinofuranosidase, alpha-L-arabinofuranoside hydrolase, L-arabinosidase and alpha-L-arabinanase.

[0184] In some additional embodiments, the present invention provides at least one variant beta-xylosidase and at least one alpha-glucuronidase. Alpha-glucuronidases (EC 3.2.1.139) catalyze the hydrolysis of an alpha-D-glucuronoside to D-glucuronate and an alcohol.

[0185] In some additional embodiments, the present invention provides at least one variant beta-xylosidase and at least one acetylxylanesterase. Acetylxylanesterases (EC 3.1.1.72) catalyze the hydrolysis of acetyl groups from polymeric xylan, acetylated xylose, acetylated glucose, alpha-napthyl acetate, and p-nitrophenyl acetate.

[0186] In some additional embodiments, the present invention provides at least one variant beta-xylosidase and at least one feruloyl esterase. Feruloyl esterases (EC 3.1.1.73) have 4-hydroxy-3-methoxycinnamoyl-sugar hydrolase activity (EC 3.1.1.73) that catalyzes the hydrolysis of the 4-hydroxy-3-methoxycinnamoyl (feruloyl) group from an esterified sugar,

which is usually arabinose in "natural" substrates, to produce ferulate (4-hydroxy-3-methoxycinnamate). Feruloyl esterase is also known as ferulic acid esterase, hydroxycinnamoyl esterase, FAE-III, cinnamoyl ester hydrolase, FAEA, cinnAE, FAE-I, or FAE-II.

[0187] In some additional embodiments, the present invention provides at least one variant beta-xylosidase and at least one coumaroyl esterase. Coumaroyl esterases (EC 3.1.1.73) catalyze a reaction of the form: coumaroyl-saccharide + H_2O = coumarate + saccharide. In some embodiments, the saccharide is an oligosaccharide or a polysaccharide. This enzyme may also be referred to as trans-4-coumaroyl esterase, trans-p-coumaroyl esterase, p-coumaroyl esterase or p-coumaric acid esterase. The enzyme also falls within EC 3.1.1.73 so may also be referred to as a feruloyl esterase.

[0188] In some additional embodiments, the present invention provides at least one variant beta-xylosidase and at least one alpha-galactosidase. Alpha-galactosidases (EC 3.2.1.22) catalyze the hydrolysis of terminal, non-reducing α-D-galactose residues in α-D-galactosides, including galactose oligosaccharides, galactomannans, galactans and arabinogalactans. This enzyme may also be referred to as melibiase.

[0189] In some additional embodiments, the present invention provides at least one variant beta-xylosidase and at least one beta-galactosidase. Beta-galactosidases (EC 3.2.1.23) catalyze the hydrolysis of terminal non-reducing β-D-galactose residues in beta-D-galactosides. In some embodiments, the polypeptide is also capable of hydrolyzing alpha-L-arabinosides. This enzyme may also be referred to as exo-(1->4)-β-D-galactanase or lactase.

[0190] In some additional embodiments, the present invention provides at least one variant beta-xylosidase and at least one beta-mannanase. Beta-mannanases (EC 3.2.1.78) catalyze the random hydrolysis of 1 ,4-beta-D-mannosidic linkages in mannans, galactomannans and glucomannans. This enzyme may also be referred to as mannan endo-1 ,4-beta-mannosidase or endo-1 ,4-mannanase.

[0191] In some additional embodiments, the present invention provides at least one variant beta-xylosidase and at least one beta-mannosidase. Beta-mannosidases (EC 3.2.1.25) catalyze the hydrolysis of terminal, non-reducing beta-D-mannose residues in beta-D-mannosides. This enzyme may also be referred to as mannanase or mannase.

[0192] In some additional embodiments, the present invention provides at least one variant beta-xylosidase and at least one glucoamylase. Glucoamylases (EC 3.2.1.3) catalyzes the release of D-glucose from non-reducing ends of oligo- and polysaccharide molecules. Glucoamylase is also generally considered a type of amylase known as amylo-glucosidase.

[0193] In some additional embodiments, the present invention provides at least one variant beta-xylosidase and at least one amylase. Amylases (EC 3.2.1.1) are starch cleaving enzymes that degrade starch and related compounds by hydrolyzing the alpha-1,4 and/or alpha-1,6 glucosidic linkages in an endo- or an exo-acting fashion. Amylases include alpha-amylases (EC 3.2.1.1); beta-amylases (3.2.1.2), amylo-amylases (EC 3.2.1.3), alpha-glucosidases (EC 3.2.1.20), pullulanases (EC 3.2.1.41), and isoamylases (EC 3.2.1.68). In some embodiments, the amylase is an alpha-amylase.

[0194] In some embodiments one or more enzymes that degrade pectin are included in enzyme mixtures that comprise at least one variant beta-xylosidase of the present invention. A pectinase catalyzes the hydrolysis of pectin into smaller units such as oligosaccharide or monomeric saccharides. In some embodiments, the enzyme mixtures comprise any pectinase, for example an endo- polygalacturonase, a pectin methyl esterase, an endo-galactanase, a pectin acetyl esterase, an endo-pectin lyase, pectate lyase, alpha rhamnosidase, an exo-galacturonase, an exo-polygalacturonate lyase, a rhamnogalacturonan hydrolase, a rhamnogalacturonan lyase, a rhamnogalacturonan acetyl esterase, a rhamnogalacturonan galacturonohydrolase and/or a xylogalacturonase.

[0195] In some additional embodiments, the present invention provides at least one variant beta-xylosidase and at least one endo-polygalacturonase. Endo-polygalacturonases (EC 3.2.1.15) catalyze the random hydrolysis of 1 ,4-alpha-D-galactosiduronic linkages in pectate and other galacturonans. This enzyme may also be referred to as polygalacturonase pectin depolymerase, pectinase, endopolygalacturonase, pectolase, pectin hydrolase, pectin polygalacturonase, poly-alpha-1 ,4-galacturonide glycanohydrolase, endogalacturonase; endo-D-galacturonase or poly(1 ,4-alpha-D-galacturonide) glycanohydrolase.

[0196] In some additional embodiments, the present invention provides at least one variant beta-xylosidase and at least one pectin methyl esterase. Pectin methyl esterases (EC 3.1.1.11) catalyze the reaction: pectin + n H_2O = n methanol + pectate. The enzyme may also been known as pectin esterase, pectin demethoxylase, pectin methoxylase, pectin methylesterase, pectase, pectinoesterase or pectin pectylhydrolase.

[0197] In some additional embodiments, the present invention provides at least one variant beta-xylosidase and at least one endo-galactanase. Endo-galactanases (EC 3.2.1.89) catalyze the endohydrolysis of 1 ,4-beta-D-galactosidic linkages in arabinogalactans. The enzyme may also be known as arabinogalactan endo-1 ,4-beta-galactosidase, endo-1 ,4-beta- galactanase, galactanase, arabinogalactanase or arabinogalactan 4-beta-D- galactanohydrolase.

[0198] In some additional embodiments, the present invention provides at least one variant beta-xylosidase and at least one pectin acetyl esterase. Pectin acetyl esterases catalyze the deacetylation of the acetyl groups at the hydroxyl groups of GaIUA residues of pectin.

[0199] In some additional embodiments, the present invention provides at least one variant beta-xylosidase and at least one endo-pectin lyase. Endo-pectin lyases (EC 4.2.2.10) catalyze the eliminative cleavage of (1 →4)-alpha-D-galacturonan methyl ester to give oligosaccharides with 4-deoxy-6-O-methyl-alpha-D-galact-4-enuronosyl groups at their non-reducing ends. The enzyme may also be known as pectin lyase, pectin trans-eliminase; endo-pectin lyase, polymethylgalacturonic transeliminase, pectin methyltranseliminase, pectolyase, PL, PNL or PMGL or (1 →4)-6-O-methyl-alpha-D-galacturonan lyase.

[0200] In some additional embodiments, the present invention provides at least one variant beta-xylosidase and at least one pectate lyase. Pectate lyases (EC 4.2.2.2) catalyze the eliminative cleavage of (1 →4)-alpha-D-galacturonan to give oligosaccharides with 4-deoxy-alpha-D-galact-4-enuronosyl groups at their non-reducing ends. The enzyme may also be known polygalacturonic transeliminase, pectic acid transeliminase, polygalacturonate lyase, endopectin methyltranseliminase, pectate transeliminase, endogalacturonate transeliminase, pectic acid lyase, pectic lyase, alpha-1 ,4-D-endopolygalacturonic acid lyase, PGA lyase, PPase-N, endo-alpha-1 ,4-polygalacturonic acid lyase, polygalacturonic acid lyase, pectin trans-eliminase, polygalacturonic acid trans-eliminase or (1 →4)-alpha-D- galacturonan lyase.

[0201] In some additional embodiments, the present invention provides at least one variant beta-xylosidase and at least one alpha-rhamnosidase. Alpha-rhamnosidases (EC 3.2.1.40) catalyze the hydrolysis of terminal non-reducing alpha-L-rhamnose residues in alpha-L-rhamnosides or alternatively in rhamnogalacturonan. This enzyme may also be known as alpha-L-rhamnosidase T, alpha-L-rhamnosidase N or alpha-L-rhamnoside rhamnohydrolase.

[0202] In some additional embodiments, the present invention provides at least one variant beta-xylosidase and at least one exo-galacturonase. Exo-galacturonases (EC 3.2.1.82)

hydrolyze pectic acid from the non-reducing end, releasing digalacturonate. The enzyme may also be known as exo-poly-α-galacturonosidase, exopolygalacturonosidase or exopolygalacturanosidase.

[0203] In some additional embodiments, the present invention provides at least one variant beta-xylosidase and at least one exo-galacturan 1,4-alpha galacturonidase. Exo-galacturonases (EC 3.2.1.67) catalyze a reaction of the following type: (1 ,4-alpha-D-galacturonide)n + H2O = (1 ,4-alpha-D-galacturonide)n-i + D- galacturonate. The enzyme may also be known as poly [1->4) alpha-D-galacturonide] galacturonohydrolase, , exopolygalacturonase, poly(galacturonate) hydrolase, exo-D-galacturonase, exo-D-galacturonanase, exopoly-D-galacturonase or poly(1 ,4-alpha-D-galacturonide) galacturonohydrolase.

[0204] In some additional embodiments, the present invention provides at least one variant beta-xylosidase and at least one exopolygalacturonate lyase. Exopolygalacturonate lyases (EC 4.2.2.9) catalyze eliminative cleavage of 4-(4-deoxy-alpha-D-galact-4-enuronosyl)-D-galacturonate from the reducing end of pectate (*i.e.*, de-esterified pectin). This enzyme may be known as pectate disaccharide-lyase, pectate exo-lyase, exopectic acid transeliminase, exopectate lyase, exopolygalacturonic acid-trans-eliminase, PATE, exo-PATE, exo-PGL or (1 →4)-alpha-D-galacturonan reducing-end-disaccharide-lyase.

[0205] In some additional embodiments, the present invention provides at least one xylanase variant beta-xylosidase and at least one rhamnogalacturonanase. Rhamnogalacturonanases hydrolyze the linkage between galactosyluronic acid and rhamnopyranosyl in an endo-fashion in strictly alternating rhamnogalacturonan structures, consisting of the disaccharide [(1,2-alpha-L-rhamnoyl-(1,4)-alpha-galactosyluronic acid].

[0206] In some additional embodiments, the present invention provides at least one variant beta-xylosidase and at least one rhamnogalacturonan lyase. Rhamnogalacturonan lyases cleave alpha-L-Rhap-(1 →4)-alpha-D-GalpA linkages in an endo-fashion in rhamnogalacturonan by beta-elimination.

[0207] In some additional embodiments, the present invention provides at least one variant beta-xylosidase and at least one rhamnogalacturonan acetyl esterase. Rhamnogalacturonan acetyl esterases catalyze the deacetylation of the backbone of alternating rhamnose and galacturonic acid residues in rhamnogalacturonan.

[0208] In some additional embodiments, the present invention provides at least one variant beta-xylosidase and at least one rhamnogalacturonan galacturonohydrolase. Rhamnogalacturonan galacturonohydrolases hydrolyze galacturonic acid from the non-reducing end of strictly alternating rhamnogalacturonan structures in an exo-fashion. This enzyme may also be known as xylogalacturonan hydrolase.

[0209] In some additional embodiments, the present invention provides at least one variant beta-xylosidase and at least one endo-arabinanase. Endo-arabinanases (EC 3.2.1.99) catalyze endohydrolysis of 1 ,5-alpha-arabinofuranosidic linkages in 1 ,5-arabinans. The enzyme may also be known as endo-arabinase, arabinan endo-1 ,5-alpha-L-arabinosidase, endo-1 ,5-alpha-L-arabinanase, endo-alpha-1 ,5-arabanase; endo-arabanase or 1 ,5-alpha-L-arabinan 1 ,5-alpha-L-arabinanohydrolase.

[0210] In some additional embodiments, the present invention provides at least one variant beta-xylosidase and at least one enzyme that participates in lignin degradation in an enzyme mixture. Enzymatic lignin depolymerization can be accomplished by lignin peroxidases, manganese peroxidases, laccases and cellobiose dehydrogenases (CDH), often working in synergy. These extracellular enzymes are often referred to as "lignin-modifying enzymes" or "LMEs." Three of these enzymes comprise two glycosylated heme-containing peroxidases: lignin peroxidase (LIP); Mn-dependent peroxidase (MNP); and, a copper-containing phenoloxidase laccase (LCC).

[0211] In some additional embodiments, the present invention provides at least one variant beta-xylosidase and at least one laccase. Laccases are copper containing oxidase enzymes that are found in many plants, fungi and microorganisms. Laccases are enzymatically active on phenols and similar molecules and perform a one electron oxidation. Laccases can be polymeric and the enzymatically active form can be a dimer or trimer.

[0212] In some additional embodiments, the present invention provides at least one variant beta-xylosidase and at least one Mn-dependent peroxidase. The enzymatic activity of Mn-dependent peroxidase (MnP) in is dependent on Mn2+. Without being bound by theory, it has been suggested that the main role of this enzyme is to oxidize Mn2+ to Mn3+ (*See e.g*,, Glenn *et al*., Arch. Biochem. Biophys., 251:688-696 [1986]). Subsequently, phenolic substrates are oxidized by the Mn3+ generated.

[0213] In some additional embodiments, the present invention provides at least one variant beta-xylosidase and at least one lignin peroxidase. Lignin peroxidase is an extracellular heme

that catalyses the oxidative depolymerization of dilute solutions of polymeric lignin in vitro. Some of the substrates of LiP, most notably 3,4-dimethoxybenzyl alcohol (veratryl alcohol, VA), are active redox compounds that have been shown to act as redox mediators. VA is a secondary metabolite produced at the same time as LiP by ligninolytic cultures of *P. chrysosporium* and without being bound by theory, has been proposed to function as a physiological redox mediator in the LiP-catalyzed oxidation of lignin *in vivo* (*See e.g.*, Harvey, *et al.*, FEBS Lett., 195:242–246 [1986]).

[0214] In some additional embodiments, the present invention provides at least one variant beta-xylosidase and at least one protease, amylase, glucoamylase, and/or a lipase that participates in cellulose degradation.

[0215] In some additional embodiments, the present invention provides at least one variant beta-xylosidase and at least one expansin or expansin-like protein, such as a swollenin (*See e.g.,* Salheimo *et al.,* Eur. J. Biochem., 269:4202-4211 [2002]) or a swollenin-like protein. Expansins are implicated in loosening of the cell wall structure during plant cell growth. Expansins have been proposed to disrupt hydrogen bonding between cellulose and other cell wall polysaccharides without comprising hydrolytic activity. In this way, they are thought to allow the sliding of cellulose fibers and enlargement of the cell wall. Swollenin, an expansin-like protein contains an N-terminal Carbohydrate Binding Module Family 1 domain (CBD) and a C-terminal expansin-like domain. In some embodiments, an expansin-like protein or swollenin-like protein comprises one or both of such domains and/or disrupts the structure of cell walls (such as disrupting cellulose structure), optionally without producing detectable amounts of reducing sugars.

[0216] In some additional embodiments, the present invention provides at least one variant beta-xylosidase and at least one polypeptide product of a cellulose integrating protein, scaffoldin or a scaffoldin-like protein, for example CipA or CipC from *Clostridium thermocellum* or *Clostridium cellulolyticum* respectively. Scaffoldins and cellulose integrating proteins are multi-functional integrating subunits which may organize cellulolytic subunits into a multi-enzyme complex. This is accomplished by the interaction of two complementary classes of domain (*i.e.* a cohesion domain on scaffoldin and a dockerin domain on each enzymatic unit). The scaffoldin subunit also bears a cellulose-binding module that mediates attachment of the cellulosome to its substrate. A scaffoldin or cellulose

integrating protein for the purposes of this invention may comprise one or both of such domains.

[0217] In some additional embodiments, the present invention provides at least one variant beta-xylosidase and at least one cellulose induced protein or modulating protein, for example as encoded by cip1 or cip2 gene or similar genes from *Trichoderma reesei* (*See e.g.*, Foreman *et al.*, J. Biol. Chem., 278:31988-31997 [2003]).

[0218] In some additional embodiments, the present invention provides at least one variant beta-xylosidase and at least one member of each of the classes of the polypeptides described above, several members of one polypeptide class, or any combination of these polypeptide classes to provide enzyme mixtures suitable for various uses.

[0219] In some embodiments, the enzyme mixture comprises other types of cellulases, selected from but not limited to cellobiohydrolase, endoglucanase, beta-glucosidase, and glycoside hydrolase 61 protein (GH61) cellulases. These enzymes may be wild-type or recombinant enzymes. In some embodiments, the cellobiohydrolase is a type 1 cellobiohydrolase (*e.g.*, a *T. reesei* cellobiohydrolase I). In some embodiments, the endoglucanase comprises a catalytic domain derived from the catalytic domain of a *Streptomyces avermitilis* endoglucanase (*See e.g.*, US Pat. Appln. Pub. No. 2010/0267089, incorporated herein by reference). In some embodiments, the at least one cellulase is derived from *Acidothermus cellulolyticus*, *Thermobifida fusca*, *Humicola grisea*, *Myceliophthora thermophila*, *Chaetomium thermophilum*, *Acremonium* sp., *Thielavia* sp, *Trichoderma reesei*, *Aspergillus* sp., or a *Chrysosporium* sp. Cellulase enzymes in the cellulase mixtures work together resulting in decrystallization and hydrolysis of the cellulose from a biomass substrate to yield fermentable sugars, such as but not limited to glucose.

[0220] Some cellulase mixtures for efficient enzymatic hydrolysis of cellulose are known (*See e.g.*, Viikari *et al.*, Adv. Biochem. Eng. Biotechnol., 108:121-45 [2007]; and US Pat. Appln. Publn. Nos. US 2009/0061484, US 2008/0057541, and US 2009/0209009, each of which is incorporated herein by reference in their entireties). In some embodiments, mixtures of purified naturally occurring or recombinant enzymes are combined with cellulosic feedstock or a product of cellulose hydrolysis. Alternatively or in addition, one or more cell populations, each producing one or more naturally occurring or recombinant cellulases, are combined with cellulosic feedstock or a product of cellulose hydrolysis.

[0221] In some embodiments, the enzyme mixture comprises commercially available purified cellulases. Commercial cellulases are known and available (*e.g.*, C2730 cellulase from *Trichoderma reesei* ATCC No. 25921 available from Sigma-Aldrich, Inc.).

[0222] In some embodiments, the enzyme mixture comprises at least one variant beta-xylosidase as provided herein and at least one or more cellobiohydrolase type 1a such as a CBH1a, CBH2b, endoglucanase (EG) such as a type 2 endoglucanase (EG2) or type 1 endoglucanse (EG1), β-glucosidase (Bgl), and/or a glycoside hydrolase 61 protein (GH61). In some embodiments, at least about 5%, at least about 6%, at least about 7%, at least about 8%, at least about 9%, at least about 10%, at least about 11%, at least about 12%, at least about 13%, at least about 14%, at least about 15%, at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, or at least about 50% of the enzyme mixture comprises at least one variant beta-xylosidase. In some embodiments, the enzyme mixture further comprises at least one cellobiohydrolase type 1 (*e.g.,* CBH1a), cellobiohydrolase type 2 (*e.g.,* CBH2b), and at least one variant beta-xylosidase, wherein the enzymes together comprise at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, or at least about 80% of the enzyme mixture. In some embodiments, the enzyme mixture further comprises at least one β-glucosidase (Bgl), at least one variant beta-xylosidase, CBH1a, and CBH2b, wherein the four enzymes together comprise at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, or at least about 85% of the enzyme mixture. In some embodiments, the enzyme mixture further comprises at least one additional endoglucanase (*e.g.,* EG2 and/or EG1), variant beta-xylosidase, xylananse, CBH2b, CBH1a, and/or Bgl, wherein the five enzymes together comprise at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, or at least about 90% of the enzyme mixture.

[0223] In some embodiments, the enzyme mixture comprises at least one or a combination of beta-xylosidase variants, CBH2b, CBH1a, Bgl, EG2, EG1, xyalanases, and/or glycoside hydrolase 61 protein (GH61), in any suitable proportion for the desired reaction. In some embodiments, the enzyme mixture composition comprises isolated cellulases in the following proportions by weight (wherein the total weight of the cellulases is 100%): about 20% to

about 0.5% of xylanase and/or beta-xylosidase (e.g., variant beta-xylosidase), about 20% to about 10% of Bgl, about 30% to about 15% of CBH1a, about 50% to about 0% of GH61, and about 10% to about 25% of CBH2b. In some embodiments, the enzyme mixture composition comprises isolated cellulases in the following proportions by weight: about 20% to about 1% of variant beta-xylosidase, about 25% to about 15% of Bgl, about 20% to about 30% of CBH1a, about 10% to about 15% of GH61, and about 25% to about 30% of CBH2b. In some embodiments, the enzyme mixture composition comprises isolated cellulases in the following proportions by weight: about 1% to about 15% of variant beta-xylosidase, about 20% to about 25% of Bgl, about 30% to about 20% of CBH1a, about 15% to about 5% of GH61, and about 25% to about 35% of CBH2b. In some embodiments, the enzyme mixture composition comprises isolated cellulases in the following proportions by weight: about 15% to about 5% of variant beta-xylosidase, about 15% to about 10% of Bgl, about 45% to about 30% of CBH1a, about 25% to about 5% of GH61, and about 40% to about 10% of CBH2b. In some embodiments, the enzyme mixture composition comprises isolated cellulases in the following proportions by weight: about 10% of variant beta-xylosidase, about 15% of Bgl, about 40% of CBH1a, about 25% of GH61, and about 10% of CBH2b. In some further embodiments, the enzyme mixture comprises cellulases in the following proportions: about 2% to about 100% xylanase and/or xylosidase (e.g., variant beta-xylosidase), about 0% to about 35% Bgl, about 0% to about 75% CBH1 (i.e., CBH1a and/or b), about 0% to about 75% CBH2 (i.e., CBH2a and/or CBH2b), about 0% to about 50% EG (i.e., EG2 and/or EG1, etc.), and/or about 0% to about 50% GH61 (i.e., GH61a, etc.). In some additional embodiments, the enzyme compositions comprise further enzymes.

[0224] In some embodiments, the enzyme mixture comprises isolated cellulases in the following proportions by weight: about 12% variant beta-xylosidase, about 33% GH61, about 10% Bgl, about 22% CBH1a, and about 23% CBH2b/EG2. In some other embodiments, the enzyme mixture comprises cellulases in the following proportions by weight: about 9% variant beta-xylosidase, about 9% EG2, about 28% GH61, about 10% about BGL1, about 30% CBH1a, and about 14% CBH2b. It is not intended that the present invention be limited to any specific proportions of enzymes, as the mixture compositions will vary, depending upon their intended use. Those of skill in the art know how to formulate the mixtures to provide optimal activity, performance and results. In some embodiments, additional enzymes, such as other cellulases, xyalanases, esterases, amylases, proteases, glucoamylases, etc., are included in the enzyme mixtures. Indeed, it is not intended that the

present invention be limited to any particular enzyme composition and/or any particular additional enzymes, as any suitable enzyme and/or composition find use in the present invention. It is also not intended that the present invention be limited to any particular combinations nor proportions of cellulases in the enzyme mixture, as any suitable combinations of cellulases and/or proportions of cellulases find use in various embodiments of the invention. In addition to the use of a single variant beta-xylosidase, any combination of variant beta-xylosidases provided herein find use in these embodiments.

[0225] In some embodiments, the enzyme component comprises more than one CBH2b, CBH1a, EG, Bgl, and/or GH61 enzyme (*e.g.*, 2, 3, 4 or more different variants of one or more of these enzymes) in addition to at least one variant beta-xylosidase, in any suitable combination. In some embodiments, an enzyme mixture composition of the invention further comprises at least one additional protein and/or enzyme. In some embodiments, enzyme mixture compositions of the present invention further comprise at least one additional enzyme other than Bgl, CBH1a, GH61, and/or CBH2b. In some embodiments, the enzyme mixture compositions of the invention further comprise at least one additional cellulase, other than the variant beta-xylosidase, EG2, EG1, Bgl, CBH1a, GH61, and/or CBH2b recited herein. In some embodiments, the variant beta-xylosidase polypeptide of the invention is also present in mixtures with non-cellulase enzymes that degrade cellulose, hemicellulose, pectin, and/or lignocellulose.

[0226] In some embodiments, a variant beta-xylosidase polypeptide of the present invention is used in combination with other optional ingredients such as at least one buffer, surfactant, and/or scouring agent. In some embodiments, at least one buffer is used with the variant beta-xylosidase polypeptide of the present invention (optionally combined with other enzymes) to maintain a desired pH within the solution in which the variant beta-xylosidase is employed. The exact concentration of buffer employed depends on several factors which the skilled artisan can determine. Suitable buffers are well known in the art. In some embodiments, at least one surfactant is used in with the variant beta-xylosidase(s) of the present invention. Suitable surfactants include any surfactant compatible with the variant beta-xylosidase(s) and, optionally, with any other enzymes being used in the mixture. Exemplary surfactants include anionic, non-ionic, and ampholytic surfactants. Indeed, it indeed that any suitable surfactant will find use in the present invention. Suitable anionic surfactants include, but are not limited to, linear or branched alkylbenzenesulfonates; alkyl or alkenyl ether sulfates comprising linear or branched alkyl groups or alkenyl groups; alkyl or alkenyl sulfates;

olefinsulfonates; alkanesulfonates, and the like. Suitable counter ions for anionic surfactants include, for example, alkali metal ions, such as sodium and potassium; alkaline earth metal ions, such as calcium and magnesium; ammonium ion; and alkanolamines comprising from 1 to 3 alkanol groups of carbon number 2 or 3. Ampholytic surfactants suitable for use in the practice of the present invention include, for example, quaternary ammonium salt sulfonates, betaine-type ampholytic surfactants, and the like. Suitable nonionic surfactants generally include polyoxalkylene ethers, as well as higher fatty acid alkanolamides or alkylene oxide adduct thereof, fatty acid glycerine monoesters, and the like. Mixtures of surfactants also find use in the present invention, as is known in the art.

[0227] The foregoing and other aspects of the invention may be better understood in connection with the following non-limiting examples.

EXPERIMENTAL

[0228] The present invention is described in further detail in the following Examples, which are not in any way intended to limit the scope of the invention as claimed.

[0229] In the experimental disclosure below, the following abbreviations apply: ppm (parts per million); M (molar); mM (millimolar), uM and μM (micromolar); nM (nanomolar); mol (moles); gm and g (gram); mg (milligrams); ug and μg (micrograms); L and l (liter); ml and mL (milliliter); ul, uL, μL, and μl (microliter); cm (centimeters); mm (millimeters); um and μm (micrometers); sec. and " " " (i.e., quote symbol) (seconds); min(s) and " ' " (i.e.,an apostrophe) (minute(s)); h(s) and hr(s) (hour(s)); U (units); MW (molecular weight); rpm (rotations per minute); rt (room temperature); °C (degrees Centigrade); DNA (deoxyribonucleic acid); RNA (ribonucleic acid); HPLC (high pressure liquid chromatography); MES (2-N-morpholino ethanesulfonic acid); LB (Luria Broth; commercially available from numerous sources, such as Sigma-Aldrich, Invitrogen, etc.); Cascade (Cascade Analytical Reagents and Biochemicals, Corvallis, OR); Calbiochem (Calbiochem, available from EMD Millipore Corp., Billerica, MA); Finnzymes (Finnzymes, part of Thermo Fisher Scientific, Lafayette, CO); NEB (New England Biolabs, Ipswich, MA); Megazyme (Megazyme International Ireland, Ltd., Wicklow, Ireland); Sigma-Aldrich (Sigma-Aldrich, St. Louis, MO); Infors (Infors AG, Bottminger/Basel, Switzerland); Difco (Difco Laboratories, BD Diagnostic Systems, Detroit, MI); KapaBiosystems (KapaBiosystems, Inc., Woburn, MA); Invitrogen (Invitrogen, Life Technologies, Grand Island, NY); Stratagene (Stratagene, now an Agilent Technologies company); Agilent

(Agilent Technologies, Inc., Santa Clara, CA); Molecular Devices (Molecular Devices, Sunnyvale, CA); Symbio (Symbio, Inc., Menlo Park, CA); USBio (US Biological, Swampscott, MA); Qiagen (Qiagen Inc., Germantown, MD); and Bio-Rad (Bio-Rad Laboratories, Hercules, CA).

[0230] Various culture media find use in the present invention. Indeed, any suitable media known in the art for growing filamentous fungi such as *M. thermophila* find use (See e.g., Berka et al., Nat. Biotechnol., 29:922-927 [2011).

[0231] Strain CF-410 is a derivative of a wild-type *M. thermophila* C1 strain with alp1 and pyr5 deleted (i.e., UV18#100fΔalplΔpyr5). Strain CF-415 is a *M. thermophila* strain developed from CF-410, having an overexpressed recombinant beta-glucosidase, an overexpressed *M. thermophila* wild-type GH61a, and deleted *cdh1* and *cdh2* genes. The beta-glucosidase is described in US Pat. No. 8,143,050; the wild-type GH61a is described in US Pat. Appln. Ser. No. 13/215,193, filed August 22, 2011, and the *cdh1* and *cdh2* deletions are described in US Pat. Appln. Ser. No. 13/286,972, filed November 1, 2011; all of which are incorporated by reference in their entireties.

[0232] The polypeptide and polynucleotide sequence of the wild-type M. thermophila C1 beta-xylosidase are provided below. Wild-type beta-xylosidase cDNA (SEQ ID NO:1) and amino acid (SEQ ID NO:2) sequences are provided below. SEQ ID NO:3 provides the sequence of the beta-xylosidase, without the signal sequence. SEQ ID NOS:84 and 85 are polynucleotide and polypeptide sequences (respectively) of a cloned cDNA *M. thermophila* beta-xylosidase ("bxyl8-233").

Beta-xylosidase WT1:

ATGAAGGCCTCTGTATCATGCCTCGTCGGCATGAGCGCCGTGGCCTACGGCCTCG
ATGGCCCTTTCCAGACCTACCCCGACTGCACCAAGCCCCCCCTGTCCGATATTAA
GGTGTGCGACCGGACACTGCCCGAGGCGGAGCGGGCGGCAGCCCTCGTGGCAGC
CCTGACCGACGAGGAGAAGCTGCAAAACCTGGTCAGCAAGGCGCCGGGGGCGCC
GCGGATCGGCCTGCCCGCGTACAACTGGTGGAGCGAGGCGCTGCACGGGGTGGC
CCACGCGCCCGGGACGCAGTTCCGCGACGGGCCGGGGGACTTCAACTCGTCCAC
GTCGTTCCCGATGCCGCTGCTGATGGCCGCCGCCTTCGACGACGAGCTGATCGAG
GCCGTCGGCGACGTCATCGGCACCGAGGCCCGCGCCTTTGGCAACGCCGGCTGG
TCCGGCCTCGACTACTGGACCCCCAACGTCAACCCCTTCCGGGACCCCCGCTGGG
GCCGCGGCTCCGAGACGCCGGGCGAGGACGTCGTGCGCCTCAAGCGCTACGCCG
CCTCCATGATCCGCGGGCTCGAGGGTCGTTCCTCCTCCTCCTCCTCCTGCTCCTTC
GGATCCGGAGGGGAGCCGCCGCGCGTCATCTCGACCTGCAAGCACTACGCCGGC
AACGACTTTGAGGACTGGAACGGCACGACGCGGCACGACTTCGACGCCGTCATC
TCGGCGCAGGACCTGGCCGAGTACTACCTGGCGCCGTTCCAGCAGTGCGCGCGC

GACTCGCGCGTCGGCTCCGTCATGTGCGCCTACAACGCCGTCAACGGGGTGCCGT
CGTGCGCCAACTCGTACCTCATGAACACGATCCTGCGCGGGCACTGGAACTGGA
CCGAGCACGACAACTACGTCACCAGCGACTGCGAGGCCGTCCTCGACGTCTCGG
CCCACCACCACTACGCCGACACCAACGCCGAGGGCACCGGCCTCTGCTTCGAGG
CCGGCATGGACACGAGCTGCGAGTACGAGGGCTCCTCCGACATCCCGGGCGCCT
CCGCCGGCGGCTTCCTGACCTGGCCCGCCGTCGACCGCGCCCTGACGCGGCTGTA
CCGGAGCCTGGTGCGGGTCGGCTACTTTGACGGCCCCGAGTCGCCGCACGCCTCG
CTGGGCTGGGCCGACGTCAACCGGCCCGAGGCGCAGGAGCTGGCCCTGCGCGCT
GCCGTCGAGGGCATCGTGCTGCTCAAGAACGACAACGACACGCTGCCGCTGCCG
CTGCCGGACGATGTCGTTGTCACCGCTGATGGTGGCCGCCGCCGCGTCGCCATGA
TCGGCTTCTGGGCCGACGCCCCGGACAAGCTGTTTGGCGGGTACAGCGGCGCGC
CCCCCTTCGCGCGCTCGCCCGCGAGCGCCGCCCGGCAGCTGGGCTGGAACGTCAC
GGTCGCCGGAGGGCCCGTCCTGGAGGGAGACTCGGACGAGGAGGAGGACACGT
GGACGGCGCCGGCCGTCGAGGCGGCCGCCGACGCCGACTACATCGTCTACTTTG
GCGGCCTGGACACGTCGGCGGCGGGCGAGACCAAGGACCGGATGACGATCGGGT
GGCCGGCGGCGCAGCTGGCGCTCATCTCGGAGCTGGCGCGGCTCGGCAAGCCCG
TCGTGGTGGTGCAGATGGGCGACCAGCTCGACGACACGCCCCTCTTCGAGCTGG
ACGGGGTGGGCGCCGTCCTGTGGGCCAACTGGCCGGGCCAGGACGGCGGCACGG
CCGTGGTCCGGCTGCTCAGCGGCGCCGAGAGCCCGGCCGGCCGCCTGCCCGTGA
CCCAGTACCCGGCCAACTACACCGACGCGGTGCCCCTGACCGACATGACCCTGC
GCCCGTCGGCGACCAACCCGGGCCGGACCTACCGCTGGTACCCGACTCCCGTCCG
GCCCTTCGGCTTCGGCCTCCACTATACCACCTTCCGGGCCGAGTTCGGCCCCCAC
CCCTTCTTCCCGGGGGCGGGCAAGGGCGATGGCGACGGCGAGGACAAGGGCGAG
AGCAAGAGCGAGATCAGGACGCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCG
CAGGGCGGCGGCGGCGGCCACCACGCCGATCCGGGACCTGCTCCGCGACTGCGA
CAAGACGTACCCGGACACGTGCCCGCTGCCGCCGCTGACGGTGCGCGTGACCAA
CGAGGGCGAGCGCGCGTCCGACTACGTGGTGCTGGCCTTCGTGTCGGGCGAGTA
CGGGCCGGCGCCGTACCCGATCAAGACGCTGGTCTCGTACGCGCGGGCGCGCGG
GCTAAAGGGGAAGGGCGGGACGGGCGCCGGCGACGGCGACGTCGCCACCACTA
CCGTCTCGCTCGACTGGACCGTCGGCAACCTGGCCCGCCACGACGAGCGCGGCA
ACACAATCCTGTACCCGGGAACTTACACCCTCACTCTCGACGAGCCGGCCCAGGC
GAGCGTGCAGTTCGCCCTCGAGGGCGAGCCCGTCGTGCTCGACGAGTGGCCTGC
GCCGCCGAGTGCCAACTCCACCGCCAGGGGGAGGCACAGG (SEQ ID NO:1)

MKASVSCLVGMSAVAYGLDGPFQTYPDCTKPPLSDIKVCDRTLPEAERAAALVAAL
TDEEKLQNLVSKAPGAPRIGLPAYNWWSEALHGVAHAPGTQFRDGPGDFNSSTSFP
MPLLMAAAFDDELIEAVGDVIGTEARAFGNAGWSGLDYWTPNVNPFRDPRWGRGS
ETPGEDVVRLKRYAASMIRGLEGRSSSSSSCSFGSGGEPPRVISTCKHYAGNDFEDWN
GTTRHDFDAVISAQDLAEYYLAPFQQCARDSRVGSVMCAYNAVNGVPSCANSYLM
NTILRGHWNWTEHDNYVTSDCEAVLDVSAHHHYADTNAEGTGLCFEAGMDTSCEY
EGSSDIPGASAGGFLTWPAVDRALTRLYRSLVRVGYFDGPESPHASLGWADVNRPE
AQELALRAAVEGIVLLKNDNDTLPLPLPDDVVVTADGGRRRVAMIGFWADAPDKLF
GGYSGAPPFARSPASAARQLGWNVTVAGGPVLEGDSDEEEDTWTAPAVEAAADAD
YIVYFGGLDTSAAGETKDRMTIGWPAAQLALISELARLGKPVVVVQMGDQLDDTPL
FELDGVGAVLWANWPGQDGGTAVVRLLSGAESPAGRLPVTQYPANYTDAVPLTDM
TLRPSATNPGRTYRWYPTPVRPFGFGLHYTTFRAEFGPHPFFPGAGKGDGDGEDKGE
SKSEIRTQQQQQQQQQQRRAAAAATTPIRDLLRDCDKTYPDTCPLPPLTVRVTNEGE
RASDYVVLAFVSGEYGPAPYPIKTLVSYARARGLKGKGGTGAGDGDVATTTVSLD
WTVGNLARHDERGNTILYPGTYTLTLDEPAQASVQFALEGEPVVLDEWPAPPSANST
ARGRHR (SEQ ID NO:2)

LDGPFQTYPDCTKPPLSDIKVCDRTLPEAERAAALVAALTDEEKLQNLVSKAPGAPRI
GLPAYNWWSEALHGVAHAPGTQFRDGPGDFNSSTSFPMPLLMAAAFDDELIEAVGD
VIGTEARAFGNAGWSGLDYWTPNVNPFRDPRWGRGSETPGEDVVRLKRYAASMIR
GLEGRSSSSSSCSFGSGGEPPRVISTCKHYAGNDFEDWNGTTRHDFDAVISAQDLAEY
YLAPFQQCARDSRVGSVMCAYNAVNGVPSCANSYLMNTILRGHWNWTEHDNYVT
SDCEAVLDVSAHHHYADTNAEGTGLCFEAGMDTSCEYEGSSDIPGASAGGFLTWPA
VDRALTRLYRSLVRVGYFDGPESPHASLGWADVNRPEAQELALRAAVEGIVLLKND
NDTLPLPLPDDVVVTADGGRRRVAMIGFWADAPDKLFGGYSGAPPFARSPASAARQ
LGWNVTVAGGPVLEGDSDEEEDTWTAPAVEAAADADYIVYFGGLDTSAAGETKDR
MTIGWPAAQLALISELARLGKPVVVVQMGDQLDDTPLFELDGVGAVLWANWPGQD
GGTAVVRLLSGAESPAGRLPVTQYPANYTDAVPLTDMTLRPSATNPGRTYRWYPTP
VRPFGFGLHYTTFRAEFGPHPFFPGAGKGDGDGEDKGESKSEIRTQQQQQQQQQRR
AAAAATTPIRDLLRDCDKTYPDTCPLPPLTVRVTNEGERASDYVVLAFVSGEYGPAP
YPIKTLVSYARARGLKGKGGTGAGDGDVATTTVSLDWTVGNLARHDERGNTILYPG
TYTLTLDEPAQASVQFALEGEPVVLDEWPAPPSANSTARGRHR (SEQ ID NO:3)

Beta-Xylosidase Variants:

[0233] The following sequences are the polynucleotide and polypeptide of some beta-xylosidase variants provided by the following invention. These polynucleotide sequences are genomic (i.e., introns are included), except for Variant V235L and Variant G347Q/G449N; these sequences were machine reverse translated from the polypeptide sequence.

Variant 985:

ATGAAGGCCTCTGTATCATGCCTCGTCGGCATGAGCGCCGTGGCCTACGGCCTCG
ATGGC
CCTTTCCAGACCTACCCCGACTGCACCAAGCCCCCCCTGTCCGATATTAAGGTGT
GCGAC
CGGACACTGCCCGAGGCGGAGCGGGCGGCAGCCCTCGTGGCAGCCCTGACCGAC
GAGGAG
AAGCTGCAAAACCTGGTCAGGTATGTATGCGGAGAGAGAGAAACACACACACAC
GCGCGC
GCGCACACACACACACACACACACTCTCTCTCTCTCTCTCTCGCGTACCATGGGT
GCCGT
CTGACGTTTTCCCTTTGTCTCTGTGTCCAGCAAGGCGCCGGGGGCGCCGCGGATC
GGCCT
GCCCGCGTACAACTGGTGGAGCGAGGCGCTGCACGGGGTGGCCCACGCGCCCGG
GACGCA
GTTCCGCGACGGGCCGGGGGACTTCAACTCGTCCACGTCGTTCCCGATGCCGCTG
CTGAT
GGCCGCCGCCTTCGACGACGAGCTGATCGAGGCCGTCGGCGACGTCATCGGCAC
CGAGGC
CCGCGCCTTTGGCAACGCCGGCTGGTCCGGCCTCGACTACTGGACCCCCAACGTC
AACCC
CTTCCGGGACCCCCGCTGGGGCCGCGGCTCCGAGACGCCGGGCGAGGACGTCGT
GCGCCT

CAAGCGCTACGCCGCCTCCATGATCCGCGGGCTCGAGGGTCGTTCCTCCTCCTCC
TCCTC
CTGCTCCTTCGGATCCGGAGGGGAGCCGCCGCGCGTCATCTCGACCTGCAAGCAC
TACGC
CGGCAACGACTTTGAGGACTGGAACGGCACGACGCGGCACGACTTCGACGCCGT
CATCTC
GGCGCAGGACCTGGCCGAGTACTACCTGGCGCCGTTCCAGCAGTGCGCGCGCGA
CTCGCG
CGTCGGCTCCGTCATGTGCGCCTACAACGCCGTCAACGGGGTGCCGTCGTGCGCC
AACTC
GTACCTCCTGAACACGATCCTGCGCGGGCACTGGAACTGGACCGAGCACGACAA
CTACGT
CACCAGCGACTGCGAGGCCGTCCTCGACGTCTCGGCCCACCACCACTACGCCGAC
ACCAA
CGCCGAGGGCACCGGCCTCTGCTTCGAGGCCGGCATGGACACGAGCTGCGAGTA
CGAGGG
CTCCTCCGACATCCCGGGCGCCTCCGCCGGCGGCTTCCTGACCTGGCCCGCCGTC
GACCG
CGCCCTGACGCGGCTGTACCGGAGCCTGGTGCGGGTCGGCTACTTTGACGGCCCC
GAGTC
GCCGCACGCCTCGCTGGGCTGGGCCGACGTCAACCGGCCCGAGGCGCAGGAGCT
GGCCCT
GCGCGCTGCCGTCGAGGGCATCGTGCTGCTCAAGAACGACAACGACACGCTGCC
GCTGCC
GCTGCCGGACGATGTCGTTGTCACCGCTGATGGTGGCCGCCGCCGCGTCGCCATG
ATCGG
CTTCTGGGCCGACGCCCCGGACAAGCTGTTTGGCGGGTACAGCGGCGCGCCCCCC
TTCGC
GCGCTCGCCCGCGAGCGCCGCCCGGCAGCTGGGCTGGAACGTCACGGTCGCCGG
AGGGCC
CGTCCTGGAGGGAGACTCGGACGAGGAGGAGGACACGTGGACGGCGCCGGCCG
TCGAGGC
GGCCGCCGACGCCGACTACATCGTCTACTTTGGCGGCCTGGACACGTCGGCGGCG
GGCGA
GACCAAGGACCGGATGACGATCGGGTGGCCGGCGGCGCAGCTGGCGCTCATCTC
GGAGCT
GGCGCGGCTCGGCAAGCCCGTCGTGGTGGTGCAGATGGGCGACCAGCTCGACGA
CACGCC
CCTCTTCGAGCTGGACGGGGTGGGCGCCGTCCTGTGGGCCAACTGGCCGGGCCA
GGACGG
CGGCACGGCCGTGGTCCGGCTGCTCAGCGGCGCCGAGAGCCCGGCCGGCCGCCT
GCCCGT
GACCCAGTACCCGGCCAACTACACCGACGCGGTGCCCCTGACCGACATGACCCT
GCGCCC
GTCGGCGACCAACCCGGGCCGGACCTACCGCTGGTACCCGACTCCCGTCCGGCCC
TTCGG
CTTCGGCCTCCACTATACCACCTTCCGGGCCGAGTTCGGCCCCCACCCCTTCTTCC
CGGG
GGCGGGCAAGGGCGATGGCGACGGCGAGGACAAGGGCGAGAGCAAGAGCGAGA
TCAGGAC

GCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCGCAGGGCGGCGGCGGCGGCCA
CCACGCC
GATCCGGGACCTGCTCCGCGACTGCGACAAGACGTACCCGGACACGTGCCCGCT
GCCGCC
GCTGACGGTGCGCGTGACCAACGAGGGCGAGCGCGCGTCCGACTACGTGGTGCT
GGCCTT
CGTGTCGGGCGAGTACGGGCCGGCGCCGTACCCGATCAAGACGCTGGTCTCGTA
CGCGCG
GGCGCGCGGGCTAAAGGGGAAGGGCGGCGACGGCGACGGCGACGGCGACGGCG
CCACCAC
TACCGTCTCGCTCGACTGGACCGTCGGCAACCTGGCCCGCCACGACGAGCGCGG
CAACAC
AATCCTGTACCCGGGAACTTACACCCTCACTCTCGACGAGCCGGCCCAGGCGAGC
GTGCA
GTTCGCCCTCGAGGGCGAGCCCGTCGTGCTCGACGAGTGGCCTGCGCCGCCGAGT
GCCAA
CTCCACCGCCAGGGGGAGGCACAGGTAA (SEQ ID NO:4)

MKASVSCLVGMSAVAYGLDGPFQTYPDCTKPPLSDIKVCDRTLPEAERAAALVAAL
TDEE
KLQNLVSKAPGAPRIGLPAYNWWSEALHGVAHAPGTQFRDGPGDFNSSTSFPMPLL
MAAA
FDDELIEAVGDVIGTEARAFGNAGWSGLDYWTPNVNPFRDPRWGRGSETPGEDVVR
LKRY
AASMIRGLEGRSSSSSSCSFGSGGEPPRVISTCKHYAGNDFEDWNGTTRHDFDAVISA
QD
LAEYYLAPFQQCARDSRVGSVMCAYNAVNGVPSCANSYLLNTILRGHWNWTEHDN
YVTSD
CEAVLDVSAHHHYADTNAEGTGLCFEAGMDTSCEYEGSSDIPGASAGGFLTWPAVD
RALT
RLYRSLVRVGYFDGPESPHASLGWADVNRPEAQELALRAAVEGIVLLKNDNDTLPL
PLPD
DVVVTADGGRRRVAMIGFWADAPDKLFGGYSGAPPFARSPASAARQLGWNVTVAG
GPVLE
GDSDEEEDTWTAPAVEAAADADYIVYFGGLDTSAAGETKDRMTIGWPAAQLALISE
LARL
GKPVVVVQMGDQLDDTPLFELDGVGAVLWANWPGQDGGTAVVRLLSGAESPAGR
LPVTQY
PANYTDAVPLTDMTLRPSATNPGRTYRWYPTPVRPFGFGLHYTTFRAEFGPHPFFPG
AGK
GDGDGEDKGESKSEIRTQQQQQQQQQRRAAAAATTPIRDLLRDCDKTYPDTCPLPP
LTV
RVTNEGERASDYVVLAFVSGEYGPAPYPIKTLVSYARARGLKGKGGDGDGDGDGA
TTTVS
LDWTVGNLARHDERGNTILYPGTYTLTLDEPAQASVQFALEGEPVVLDEWPAPPSA
NSTA
RGRHR (SEQ ID NO:5)

Variant 983:

ATGAAGGCCTCTGTATCATGCCTCGTCGGCATGAGCGCCGTGGCCTACGGCCTCG
ATGGC
CCTTTCCAGACCTACCCCGACTGCACCAAGCCCCCCCTGTCCGATATTAAGGTGT
GCGAC
CGGACACTGCCCGAGGCGGAGCGGGCGGCAGCCCTCGTGGCAGCCCTGACCGAC
GAGGAG
AAGCTGCAAAACCTGGTCAGGTATGTATGCGGAGAGAGAGAAACACACACACAC
GCGCGC
GCGCACACACACACACACACACACTCTCTCTCTCTCTCTCTCGCGTACCATGGGT
GCCGT
CTGACGTTTTCCCTTTGTCTCTGTGTCCAGCAAGGCGCCGGGGGCGCCGCGGATC
GGCCT
GCCCGCGTACAACTGGTGGAGCGAGGCGCTGCACGGGGTGGCCCACGCGCCCGG
GACGCA
GTTCCGCGACGGGCCGGGGGACTTCAACTCGTCCACGTCGTTCCCGATGCCGCTG
CTGAT
GGCCGCCGCCTTCGACGACGAGCTGATCGAGGCCGTCGGCGACGTCATCGGCAC
CGAGGC
CCGCGCCTTTGGCAACGCCGGCTGGTCCGGCCTCGACTACTGGACCCCCAACGTC
AACCC
CTTCCGGGACCCCCGCTGGGGCCGCGGCTCCGAGACGCCGGGCGAGGACGTCGT
GCGCCT
CAAGCGCTACGCCGCCTCCATGATCCGCGGGCTCGAGGGTCGTTCCTCCTCCTCC
TCCTC
CTGCTCCTTCGGATCCGGAGGGGAGCCGCCGCGCGTCATCGCGACCTGCAAGCA
CTACGC
CGGCTATGACTTTGAGGACTGGAACGGCACGACGCGGCACGACTTCGACGCCGT
CATCTC
GGCGCAGGACCTGGCCGAGTACTACCTGGCGCCGTTCCAGCAGTGCGCGCGCGA
CTCGCG
CGTCGGCTCCGTCATGTGCGCCTACAACGCCGTCAACGGGGTGCCGTCGTGCGCC
AACTC
GTACCTCCTGAACACGATCCTGCGCGGGCACTGGAACTGGACCGAGCACGACAA
CTACGT
CACCAGCGACTGCGAGGCCGTCCTCGACGTCTCGGCCCACCACCACTACGCCGAC
ACCAA
CGCCGAGGGCACCGGCCTCTGCTTCGAGGCCGGCATGGACACGAGCTGCGAGTA
CGAGGG
CTCCTCCGACATCCCGGGCGCCTCCGCCGGCGGCTTCCTGACCTGGCCCGCCGTC
GACCG
CGCCCTGACGCGGCTGTACCGGAGCCTGGTGCGGGTCGGCTACTTTGACGGCCCC
GAGTC
GCCGCACGCCTCGCTGGGCTGGGCCGACGTCAACCGGCCCGAGGCGCAGGAGCT
GGCCCT
GCGCGCTGCCGTCGAGGGCATCGTGCTGCTCAAGAACGACAACGACACGCTGCC
GCTGCC
GCTGCCGGACGATGTCGTTGTCACCGCTGATGGTGGCCGCCGCCGCGTCGCCATG
ATCGG

CTTCTGGGCCGACGCCCCGGACAAGCTGTTTGGCGGGTACAGCGGCGCGCCCCCC
TTCGC
GCGCTCGCCCGCGAGCGCCGCCCGGCAGCTGGGCTGGAACGTCACGGTCGCCGG
AGGGCC
CGTCCTGGAGGGAGACTCGGACGAGGAGGAGGACACGTGGACGGCGCCGGCCG
TCGAGGC
GGCCGCCGACGCCGACTACATCGTCTACTTTGGCGGCCTGGACACGTCGGCGGCG
GGCGA
GACCAAGGACCGGATGACGATCGGGTGGCCGGCGGCGCAGCTGGCGCTCATCTC
GGAGCT
GGCGCGGCTCGGCAAGCCCGTCGTGGTGGTGCAGATGGGCGACCAGCTCGACGA
CACGCC
CCTCTTCGAGCTGGACGGGGTGGGCGCCGTCCTGTGGGCCAACTATCCGGGCCAG
GACGG
CGGCACGGCCGTGGTCCGGCTGCTCAGCGGCGCCGAGAGCCCGGCCGGCCGCCT
GCCCGT
GACCCAGTACCCGGCCAACTACACCGACGCGGTGCCCCTGACCGACATGACCCT
GCGCCC
GTCGGCGACCAACCCGGGCCGGACCTACCGCTGGTACCCGACTCCCGTCCGGCCC
TTCGG
CTTCGGCCTCCACTATACCACCTTCCGGGCCGAGTTCGGCCCCCACCCCTTCTTCC
CGGG
GGCGGGCAAGGGCGATGGCGACGGCGAGGACAAGGGCGAGAGCAAGAGCGAGA
TCAGGAC
GCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCGCAGGGCGGCGGCGGCGGCCA
CCACGCC
GATCCGGGACCTGCTCCGCGACTGCGACAAGACGTACCCGGACACGTGCCCGCT
GCCGCC
GCTGACGGTGCGCGTGACCAACGAGGGCGAGCGCGCGTCCGACTACGTGGTGCT
GGCCTT
CGTGTCGGGCGAGTACGGGCCGGCGCCGTACCCGATCAAGACGCTGGTCTCGTA
CGCGCG
GGCGCGCGGGCTAAAGGGGAAGGGCGGCGACGGCGACGGCGACGGCGACGGCG
CCACCAC
TACCGTCTCGCTCGACTGGACCGTCGGCAACCTGGCCCGCCACGACGAGCGCGG
CAACAC
AATCCTGTACCCGGGAACTTACACCCTCACTCTCGACGAGCCGGCCCAGGCGAGC
GTGCA
GTTCGCCCTCGAGGGCGAGCCCGTCGTGCTCGACGAGTGGCCTGCGCCGCCGAGT
GCCAA
CTCCACCGCCAGGGGGAGGCACAGGTAA (SEQ ID NO:6)

MKASVSCLVGMSAVAYGLDGPFQTYPDCTKPPLSDIKVCDRTLPEAERAAALVAAL
TDEE
KLQNLVSKAPGAPRIGLPAYNWWSEALHGVAHAPGTQFRDGPGDFNSSTSFPMPLL
MAAA
FDDELIEAVGDVIGTEARAFGNAGWSGLDYWTPNVNPFRDPRWGRGSETPGEDVVR
LKRY
AASMIRGLEGRSSSSSSCSFGSGGEPPRVIATCKHYAGYDFEDWNGTTRHDFDAVISA
QD

LAEYYLAPFQQCARDSRVGSVMCAYNAVNGVPSCANSYLLNTILRGHWNWTEHDN
YVTSD
CEAVLDVSAHHHYADTNAEGTGLCFEAGMDTSCEYEGSSDIPGASAGGFLTWPAVD
RALT
RLYRSLVRVGYFDGPESPHASLGWADVNRPEAQELALRAAVEGIVLLKNDNDTLPL
PLPD
DVVVTADGGRRRVAMIGFWADAPDKLFGGYSGAPPFARSPASAARQLGWNVTVAG
GPVLE
GDSDEEEDTWTAPAVEAAADADYIVYFGGLDTSAAGETKDRMTIGWPAAQLALISE
LARL
GKPVVVVQMGDQLDDTPLFELDGVGAVLWANYPGQDGGTAVVRLLSGAESPAGRL
PVTQY
PANYTDAVPLTDMTLRPSATNPGRTYRWYPTPVRPFGFGLHYTTFRAEFGPHPFFPG
AGK
GDGDGEDKGESKSEIRTQQQQQQQQQRRAAAAATTPIRDLLRDCDKTYPDTCPLPP
LTV
RVTNEGERASDYVVLAFVSGEYGPAPYPIKTLVSYARARGLKGKGGDGDGDGDGA
TTTVS
LDWTVGNLARHDERGNTILYPGTYTLTLDEPAQASVQFALEGEPVVLDEWPAPPSA
NSTA
RGRHR (SEQ ID NO:7)

Variant 963:

ATGAAGGCCTCTGTATCATGCCTCGTCGGCATGAGCGCCGTGGCCTACGGCCTCG
ATGGC
CCTTTCCAGACCTACCCCGACTGCACCAAGCCCCCCCTGTCCGATATTAAGGTGT
GCGAC
CGGACACTGCCCGAGGCGGAGCGGGCGGCAGCCCTCGTGGCAGCCCTGACCGAC
GAGGAG
AAGCTGCAAAACCTGGTCAGGTATGTATGCGGAGAGAGAGAAACACACACACAC
GCGCGC
GCGCACACACACACACACACACACTCTCTCTCTCTCTCTCTCGCGTACCATGGGT
GCCGT
CTGACGTTTTCCCTTTGTCTCTGTGTCCAGCAAGGCGCCGGGGGCGCCGCGGATC
GGCCT
GCCCGCGTACAACTGGTGGAGCGAGGCGCTGCACGGGGTGGCCCACGCGCCCGG
GACGCA
GTTCCGCGACGGGCCGGGGGACTTCAACTCGTCCACGTCGTTCCCGATGCCGATT
CTGAT
GGCCGCCGCCTTCGACGACGAGCTGATCGAGGCCGTCGGCGACGTCATCGGCAC
CGAGGC
CCGCGCCTTTGGCAACGCCGGCTGGTCCGGCCTCGACTACTGGACCCCCAACGTC
AACCC
CTTCCGGGACCCCCGCTGGGGCCGCGGCTCCGAGACGCCGGGCGAGGACGTCGT
GCGCCT
CAAGCGCTACGCCGCCTCCATGATCCGCGGGCTCGAGGGTCGTTCCTCCTCCTCC
TCCTC

CTGCTCCTTCGGATCCGGAGGGGAGCCGCCGCGCGTCATCGCGACCTGCAAGCA
CTACGC
CGGCAACGACTTTGAGGACTGGAACGGCACGACGCGGCACGACTTCGACGCCGT
CATCTC
GGCGCAGGACCTGGCCGAGTACTACCTGGCGCCGTTCCAGCAGTGCGCGCGCGA
CTCGCG
CGTCGGCTCCGTCATGTGCGCCTACAACGCCGTCAACGGGGTGCCGTCGTGCGCC
AACTC
GTACCTCCTGAACACGATCCTGCGCGGGCACTGGAACTGGACCGAGCACGACAA
CTACGT
CACCAGCGACTGCGAGGCCGTCCTCGACGTCTCGGCCCACCACCACTACGCCGAC
ACCAA
CGCCGAGGGCACCGCGCTCTGCTTCGAGGCCGGCATGGACACGAGCTGCGAGTA
CGAGGG
CTCCTCCGACATCCCGGGCGCCTCCGCCGGCGGCTTCCTGACCTGGCCCGCCGTC
GACCG
CGCCCTGACGCGGCTGTACCGGAGCCTGGTGCGGGTCGGCTACTTTGACGGCCCC
GAGTC
GCCGCACGCCTCGCTGGGCTGGGCCGACGTCAACCGGCCCGAGGCGCAGGAGCT
GGCCCT
GCGCGCTGCCGTCGAGGGCATCGTGCTGCTCAAGAACGACAACGACACGCTGCC
GCTGCC
GCTGCCGGACGATGTCGTTGTCACCGCTGATGGTGGCCGCCGCCGCGTCGCCATG
ATCGG
CTTCTGGGCCGACGCCCCGGACAAGCTGTTTGGCGGGTACAGCGGCGCGCCCCCC
TTCGC
GCGCTCGCCCGCGAGCGCCGCCCGGCAGCTGGGCTGGAACGTCACGGTCGCCGG
AGGGCC
CGTCCTGGAGGGAGACTCGGACGAGGAGGAGGACACGTGGACGGCGCCGGCCG
TCGAGGC
GGCCGCCGACGCCGACTACATCGTCTACTTTGGCGGCCTGGACACGTCGGCGGCG
GGCGA
GACCAAGGACCGGATGACGATCGGGTGGCCGGCGGCGCAGCTGGCGCTCATCTC
GGAGCT
GGCGCGGCTCGGCAAGCCCGTCGTGGTGGTGCAGATGGGCGACCAGCTCGACGA
CACGCC
CCTCTTCGAGCTGGACGGGGTGGGCGCCGTCCTGTGGGCCAACTATCCGGGCCAG
GACGG
CGGCACGGCCGTGGTCCGGCTGCTCAGCGGCGCCGAGAGCCCGGCCGGCCGCCT
GCCCGT
GACCCAGTACCCGGCCAACTACACCGACGCGGTGCCCCTGACCGACATGACCCT
GCGCCC
GTCGGCGACCAACCCGGGCCGGACCTACCGCTGGTACCCGACTCCCGTCCGGCCC
TTCGG
CTTCGGCCTCCACTATACCACCTTCCGGGCCGAGTTCGGCCCCCACCCCTTCTTCC
CGGG
GGCGGGCAAGGGCGATGGCGACGGCGAGGACAAGGGCGAGAGCAAGAGCGAGA
TCAGGAC
GCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCGCAGGGCGGCGGCGGCGGCCA
CCACGCC

GATCCGGGACCTGCTCCGCGACTGCGACAAGACGTACCCGGACACGTGCCCGCT
GCCGCC
GCTGACGGTGCGCGTGACCAACGAGGGCGAGCGCGCGTCCGACTACGTGGTGCT
GGCCTT
CGTGTCGGGCGAGTACGGGCCGGCGCCGTACCCGATCAAGACGCTGGTCTCGTA
CGCGCG
GGCGCGCGGGCTAAAGGGGAAGGGCGGCGACGGCGACGGCGACGGCGACGGCG
CCACCAC
TACCGTCTCGCTCGACTGGACCGTCGGCAACCTGGCCCGCCACGACGAGCGCGG
CAACAC
AATCCTGTACCCGGGAACTTACACCCTCACTCTCGACGAGCCGGCCCAGGCGAGC
GTGCA
GTTCGCCCTCGAGGGCGAGCCCGTCGTGCTCGACGAGTGGCCTGCGCCGCCGAGT
GCCAA
CTCCACCGCCAGGGGGAGGCACAGGTAA (SEQ ID NO:8)

MKASVSCLVGMSAVAYGLDGPFQTYPDCTKPPLSDIKVCDRTLPEAERAAALVAAL
TDEE
KLQNLVSKAPGAPRIGLPAYNWWSEALHGVAHAPGTQFRDGPGDFNSSTSFPMPIL
MAAA
FDDELIEAVGDVIGTEARAFGNAGWSGLDYWTPNVNPFRDPRWGRGSETPGEDVVR
LKRY
AASMIRGLEGRSSSSSSCSFGSGGEPPRVIATCKHYAGNDFEDWNGTTRHDFDAVISA
QD
LAEYYLAPFQQCARDSRVGSVMCAYNAVNGVPSCANSYLLNTILRGHWNWTEHDN
YVTSD
CEAVLDVSAHHHYADTNAEGTALCFEAGMDTSCEYEGSSDIPGASAGGFLTWPAVD
RALT
RLYRSLVRVGYFDGPESPHASLGWADVNRPEAQELALRAAVEGIVLLKNDNDTLPL
PLPD
DVVVTADGGRRRVAMIGFWADAPDKLFGGYSGAPPFARSPASAARQLGWNVTVAG
GPVLE
GDSDEEEDTWTAPAVEAAADADYIVYFGGLDTSAAGETKDRMTIGWPAAQLALISE
LARL
GKPVVVVQMGDQLDDTPLFELDGVGAVLWANYPGQDGGTAVVRLLSGAESPAGRL
PVTQY
PANYTDAVPLTDMTLRPSATNPGRTYRWYPTPVRPFGFGLHYTTFRAEFGPHPFFPG
AGK
GDGDGEDKGESKSEIRTQQQQQQQQQRRAAAAATTPIRDLLRDCDKTYPDTCPLPP
LTV
RVTNEGERASDYVVLAFVSGEYGPAPYPIKTLVSYARARGLKGKGGDGDGDGDGA
TTTVS
LDWTVGNLARHDERGNTILYPGTYTLTLDEPAQASVQFALEGEPVVLDEWPAPPSA
NSTA
RGRHR (SEQ ID NO:9)

Variant 873:

ATGAAGGCCTCTGTATCATGCCTCGTCGGCATGAGCGCCGTGGCCTACGGCCTCG
ATGGC
CCTTTCCAGACCTACCCCGACTGCACCAAGCCCCCCCTGTCCGATATTAAGGTGT
GCGAC
CGGACACTGCCCGAGGCGGAGCGGGCGGCAGCCCTCGTGGCAGCCCTGACCGAC
GAGGAG
AAGCTGCAAAACCTGGTCAGGTATGTATGCGGAGAGAGAGAAACACACACACAC
GCGCGC
GCGCACACACACACACACACACACTCTCTCTCTCTCTCTCTCGCGTACCATGGGT
GCCGT
CTGACGTTTTCCCTTTGTCTCTGTGTCCAGCAAGGCGCCGGGGGCGCCGCGGATC
GGCCT
GCCCGCGTACAACTGGTGGAGCGAGGCGCTGCACGGGGTGGCCCACGCGCCCGG
GACGCA
GTTCCGCGACGGGCCGGGGGACTTCAACTCGTCCACGTCGTTCCCGATGCCGATT
CTGAT
GGCCGCCGCCTTCGACGACGAGCTGATCGAGGCCGTCGGCGACGTCATCGGCAC
CGAGGC
CCGCGCCTTTGGCAACGCCGGCTGGTCCGGCCTCGACTACTGGACCCCCAACGTC
AACCC
CTTCCGGGACCCCCGCTGGGGCCGCGGCTCCGAGACGCCGGGCGAGGACGTCGT
GCGCCT
CAAGCGCTACGCCGCCTCCATGATCCGCGGGCTCGAGGGTCGTTCCTCCTCCTCC
TCCTC
CTGCTCCTTCGGATCCGGAGGGGAGCCGCCGCGCGTCATCGCGACCTGCAAGCA
CTACGC
CGGCAACGACTTTGAGGACTGGAACGGCACGACGCGGCACGACTTCGACGCCGT
CATCTC
GGCGCAGGACCTGGCCGAGTACTACCTGGCGCCGTTCCAGCAGTGCGCGCGCGA
CTCGCG
CGTCGGCTCCGTCATGTGCGCCTACAACGCCGTCAACGGGGTGCCGTCGTGCGCC
AACTC
GTACCTCATGAACACGATCCTGCGCGGGCACTGGAACTGGACCGAGCACGACAA
CTACGT
CACCAGCGACTGCGAGGCCGTCCTCGACGTCTCGGCCCACCACCACTACGCCGAC
ACCAA
CGCCGAGGGCACCGGCCTCTGCTTCGAGGCCGGCATGGACACGAGCTGCGAGTA
CGAGGG
CTCCTCCGACATCCCGGGCGCCTCCGCCGGCGGCTTCCTGACCTGGCCCGCCGTC
GACCG
CGCCCTGACGCGGCTGTACCGGAGCCTGGTGCGGGTCGGCTACTTTGACGGCCCC
GAGTC
GCCGCACGCCTCGCTGGGCTGGGCCGACGTCAACCGGCCCGAGGCGCAGGAGCT
GGCCCT
GCGCGCTGCCGTCGAGGGCATCGTGCTGCTCAAGAACGACAACGACACGCTGCC
GCTGCC
GCTGCCGGACGATGTCGTTGTCACCGCTGATGGTGGCCGCCGCCGCGTCGCCATG
ATCGG

CTTCTGGGCCGACGCCCCGGACAAGCTGTTTGGCGGGTACAGCGGCGCGCCCCCC
TTCGC
GCGCTCGCCCGCGAGCGCCGCCCGGCAGCTGGGCTGGAACGTCACGGTCGCCGG
AGGGCC
CGTCCTGGAGGGAGACTCGGACGAGGAGGAGGACACGTGGACGGCGCCGGCCG
TCGAGGC
GGCCGCCGACGCCGACTACATCGTCTACTTTGGCGGCCTGGACACGTCGGCGGCG
GGCGA
GACCAAGGACCGGATGACGATCGGGTGGCCGGCGGCGCAGCTGGCGCTCATCTC
GGAGCT
GGCGCGGCTCGGCAAGCCCGTCGTGGTGGTGCAGATGGGCGACCAGCTCGACGA
CACGCC
CCTCTTCGAGCTGGACGGGGTGGGCGCCGTCCTGTGGGCCAACTGGCCGGGCCA
GGACGG
CGGCACGGCCGTGGTCCGGCTGCTCAGCGGCGCCGAGAGCCCGGCCGGCCGCCT
GCCCGT
GACCCAGTACCCGGCCAACTACACCGACGCGGTGCCCCTGACCGACATGACCCT
GCGCCC
GTCGGCGACCAACCCGGGCCGGACCTACCGCTGGTACCCGACTCCCGTCCGGCCC
TTCGG
CTTCGGCCTCCACTATACCACCTTCCGGGCCGAGTTCGGCCCCCACCCCTTCTTCC
CGGG
GGCGGGCAAGGGCGATGGCGACGGCGAGGACAAGGGCGAGAGCAAGAGCGAGA
TCAGGAC
GCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCGCAGGGCGGCGGCGGCGGCCA
CCACGCC
GATCCGGGACCTGCTCCGCGACTGCGACAAGACGTACCCGGACACGTGCCCGCT
GCCGCC
GCTGACGGTGCGCGTGACCAACGAGGGCGAGCGCGCGTCCGACTACGTGGTGCT
GGCCTT
CGTGTCGGGCGAGTACGGGCCGGCGCCGTACCCGATCAAGACGCTGGTCTCGTA
CGCGCG
GGCGCGCGGGCTAAAGGGGAAGGGCGGCGACGGCGACGGCGACGGCGACGGCG
CCACCAC
TACCGTCTCGCTCGACTGGACCGTCGGCAACCTGGCCCGCCACGACGAGCGCGG
CAACAC
AATCCTGTACCCGGGAACTTACACCCTCACTCTCGACGAGCCGGCCCAGGCGAGC
GTGCA
GTTCGCCCTCGAGGGCGAGCCCGTCGTGCTCGACGAGTGGCCTGCGCCGCCGAGT
GCCAA
CTCCACCGCCAGGGGGAGGCACAGGTAA (SEQ ID NO:10)

MKASVSCLVGMSAVAYGLDGPFQTYPDCTKPPLSDIKVCDRTLPEAERAAALVAAL
TDEE
KLQNLVSKAPGAPRIGLPAYNWWSEALHGVAHAPGTQFRDGPGDFNSSTSFPMPIL
MAAA
FDDELIEAVGDVIGTEARAFGNAGWSGLDYWTPNVNPFRDPRWGRGSETPGEDVVR
LKRY
AASMIRGLEGRSSSSSSCSFGSGGEPPRVIATCKHYAGNDFEDWNGTTRHDFDAVISA
QD

LAEYYLAPFQQCARDSRVGSVMCAYNAVNGVPSCANSYLMNTILRGHWNWTEHD
NYVTSD
CEAVLDVSAHHHYADTNAEGTGLCFEAGMDTSCEYEGSSDIPGASAGGFLTWPAVD
RALT
RLYRSLVRVGYFDGPESPHASLGWADVNRPEAQELALRAAVEGIVLLKNDNDTLPL
PLPD
DVVVTADGGRRRVAMIGFWADAPDKLFGGYSGAPPFARSPASAARQLGWNVTVAG
GPVLE
GDSDEEEDTWTAPAVEAAADADYIVYFGGLDTSAAGETKDRMTIGWPAAQLALISE
LARL
GKPVVVVQMGDQLDDTPLFELDGVGAVLWANWPGQDGGTAVVRLLSGAESPAGR
LPVTQY
PANYTDAVPLTDMTLRPSATNPGRTYRWYPTPVRPFGFGLHYTTFRAEFGPHPFFPG
AGK
GDGDGEDKGESKSEIRTQQQQQQQQQRRAAAAATTPIRDLLRDCDKTYPDTCPLPP
LTV
RVTNEGERASDYVVLAFVSGEYGPAPYPIKTLVSYARARGLKGKGGDGDGDGDGA
TTTVS
LDWTVGNLARHDERGNTILYPGTYTLTLDEPAQASVQFALEGEPVVLDEWPAPPSA
NSTA
RGRHR (SEQ ID NO:11)

Variant 989:

ATGAAGGCCTCTGTATCATGCCTCGTCGGCATGAGCGCCGTGGCCTACGGCCTCG
ATGGC
CCTTTCCAGACCTACCCCGACTGCACCAAGCCCCCCCTGTCCGATATTAAGGTGT
GCGAC
CGGACACTGCCCGAGGCGGAGCGGGCGGCAGCCCTCGTGGCAGCCCTGACCGAC
GAGGAG
AAGCTGCAAAACCTGGTCAGGTATGTATGCGGAGAGAGAGAAACACACACACAC
GCGCGC
GCGCACACACACACACACACACACTCTCTCTCTCTCTCTCTCGCGTACCATGGGT
GCCGT
CTGACGTTTTCCCTTTGTCTCTGTGTCCAGCAAGGCGCCGGGGGCGCCGCGGATC
GGCCT
GCCCGCGTACAACTGGTGGAGCGAGGCGCTGCACGGGGTGGCCCACGCGCCCGG
GACGCA
GTTCCGCGACGGGCCGGGGGACTTCAACTCGGCGACGTCGTTCCCGATGCCGCTG
CTGAT
GGCCGCCGCCTTCGACGACGAGCTGATCGAGGCCGTCGGCGACGTCATCGGCAC
CGAGGC
CCGCGCCTTTGGCAACGCCGGCTGGTCCGGCCTCGACTACTGGACCCCCAACGTC
AACCC
CTTCCGGGACCCCCGCTGGGGCCGCGGCTCCGAGACGCCGGGCGAGGACGTCGT
GCGCCT
CAAGCGCTACGCCGCCTCCATGATCCGCGGGCTCGAGGGTCGTTCCTCCTCCTCC
TCCTC

CTGCTCCTTCGGATCCGGAGGGGAGCCGCCGCGCGTCATCGCGACCTGCAAGCA
CTACGC
CGGCAACGACTTTGAGGACTGGAACGGCACGACGCGGCACGACTTCGACGCCGT
CATCTC
GGCGCAGGACCTGGCCGAGTACTACCTGGCGCCGTTCCAGCAGTGCGCGCGCGA
CTCGCG
CGTCGGCTCCGTCATGTGCGCCTACAACGCCGTCAACGGGGTGCCGTCGTGCGCC
AACTC
GTACCTCCTGAACACGATCCTGCGCGGGCACTGGAACTGGACCGAGCACGACAA
CTACGT
CACCAGCGACTGCGAGGCCGTCCTCGACGTCTCGGCCCACCACCACTACGCCGAC
ACCAA
CGCCGAGGGCACCGGCCTCTGCTTCGAGGCCGGCATGGACACGAGCTGCGAGTA
CGAGGG
CTCCTCCGACATCCCGGGCGCCTCCGCCGGCGGCTTCCTGACCTGGCCCGCCGTC
GACCG
CGCCCTGACGCGGCTGTACCGGAGCCTGGTGCGGGTCGGCTACTTTGACGGCCCC
GAGTC
GCCGCACGCCTCGCTGGGCTGGGCCGACGTCAACCGGCCCGAGGCGCAGGAGCT
GGCCCT
GCGCGCTGCCGTCGAGGGCATCGTGCTGCTCAAGAACGACAACGACACGCTGCC
GCTGCC
GCTGCCGGACGATGTCGTTGTCACCGCTGATGGTGGCCGCCGCCGCGTCGCCATG
ATCGG
CTTCTGGGCCGACGCCCCGGACAAGCTGTTTGGCGGGTACAGCGGCGCGCCCCCC
TTCGC
GCGCTCGCCCGCGAGCGCCGCCCGGCAGCTGGGCTGGAACGTCACGGTCGCCGG
AGGGCC
CGTCCTGGAGGGAGACTCGGACGAGGAGGAGGACACGTGGACGGCGCCGGCCG
TCGAGGC
GGCCGCCGACGCCGACTACATCGTCTACTTTGGCGGCCTGGACACGTCGGCGGCG
GGCGA
GACCAAGGACCGGATGACGATCGGGTGGCCGGCGGCGCAGCTGGCGCTCATCTC
GGAGCT
GGCGCGGCTCGGCAAGCCCGTCGTGGTGGTGCAGATGGGCGACCAGCTCGACGA
CACGCC
CCTCTTCGAGCTGGACGGGGTGGGCGCCGTCCTGTGGGCCAACTATCCGGGCCAG
GACGG
CGGCACGGCCGTGGTCCGGCTGCTCAGCGGCGCCGAGAGCCCGGCCGGCCGCCT
GCCCGT
GACCCAGTACCCGGCCAACTACACCGACGCGGTGCCCCTGACCGACATGACCCT
GCGCCC
GTCGGCGACCAACCCGGGCCGGACCTACCGCTGGTACCCGACTCCCGTCCGGCCC
TTCGG
CTTCGGCCTCCACTATACCACCTTCCGGGCCGAGTTCGGCCCCCACCCCTTCTTCC
CGGG
GGCGGGCAAGGGCGATGGCGACGGCGAGGACAAGGGCGAGAGCAAGAGCGAGA
TCAGGAC
GCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCGCAGGGCGGCGGCGGCGGCCA
CCACGCC

GATCCGGGACCTGCTCCGCGACTGCGACAAGACGTACCCGGACACGTGCCCGCT
GCCGCC
GCTGACGGTGCGCGTGACCAACGAGGGCGAGCGCGCGTCCGACTACGTGGTGCT
GGCCTT
CGTGTCGGGCGAGTACGGGCCGGCGCCGTACCCGATCAAGACGCTGGTCTCGTA
CGCGCG
GGCGCGCGGGCTAAAGGGGAAGGGCGGCGACGGCGACGGCGACGGCGACGGCG
CCACCAC
TACCGTCTCGCTCGACTGGACCGTCGGCAACCTGGCCCGCCACGACGAGCGCGG
CAACAC
AATCCTGTACCCGGGAACTTACACCCTCACTCTCGACGAGCCGGCCCAGGCGAGC
GTGCA
GTTCGCCCTCGAGGGCGAGCCCGTCGTGCTCGACGAGTGGCCTGCGCCGCCGAGT
GCCAA
CTCCACCGCCAGGGGGAGGCACAGGTAA (SEQ ID NO:12)

MKASVSCLVGMSAVAYGLDGPFQTYPDCTKPPLSDIKVCDRTLPEAERAAALVAAL
TDEE
KLQNLVSKAPGAPRIGLPAYNWWSEALHGVAHAPGTQFRDGPGDFNSATSFPMPLL
MAAA
FDDELIEAVGDVIGTEARAFGNAGWSGLDYWTPNVNPFRDPRWGRGSETPGEDVVR
LKRY
AASMIRGLEGRSSSSSSCSFGSGGEPPRVIATCKHYAGNDFEDWNGTTRHDFDAVISA
QD
LAEYYLAPFQQCARDSRVGSVMCAYNAVNGVPSCANSYLLNTILRGHWNWTEHDN
YVTSD
CEAVLDVSAHHHYADTNAEGTGLCFEAGMDTSCEYEGSSDIPGASAGGFLTWPAVD
RALT
RLYRSLVRVGYFDGPESPHASLGWADVNRPEAQELALRAAVEGIVLLKNDNDTLPL
PLPD
DVVVTADGGRRRVAMIGFWADAPDKLFGGYSGAPPFARSPASAARQLGWNVTVAG
GPVLE
GDSDEEEDTWTAPAVEAAADADYIVYFGGLDTSAAGETKDRMTIGWPAAQLALISE
LARL
GKPVVVVQMGDQLDDTPLFELDGVGAVLWANYPGQDGGTAVVRLLSGAESPAGRL
PVTQY
PANYTDAVPLTDMTLRPSATNPGRTYRWYPTPVRPFGFGLHYTTFRAEFGPHPFFPG
AGK
GDGDGEDKGESKSEIRTQQQQQQQQQRRAAAAATTPIRDLLRDCDKTYPDTCPLPP
LTV
RVTNEGERASDYVVLAFVSGEYGPAPYPIKTLVSYARARGLKGKGGDGDGDGDGA
TTTVS
LDWTVGNLARHDERGNTILYPGTYTLTLDEPAQASVQFALEGEPVVLDEWPAPPSA
NSTA
RGRHR (SEQ ID NO:13)

Variant 902:

ATGAAGGCCTCTGTATCATGCCTCGTCGGCATGAGCGCCGTGGCCTACGGCCTCG
ATGGC
CCTTTCCAGACCTACCCCGACTGCACCAAGCCCCCCCTGTCCGATATTAAGGTGT
GCGAC
CGGACACTGCCCGAGGCGGAGCGGGCGGCAGCCCTCGTGGCAGCCCTGACCGAC
GAGGAG
AAGCTGCAAAACCTGGTCAGGTATGTATGCGGAGAGAGAGAAACACACACACAC
GCGCGC
GCGCACACACACACACACACACACTCTCTCTCTCTCTCTCTCGCGTACCATGGGT
GCCGT
CTGACGTTTTCCCTTTGTCTCTGTGTCCAGCAAGGCGCCGGGGGCGCCGCGGATC
GGCCT
GCCCGCGTACAACTGGTGGAGCGAGGCGCTGCACGGGGTGGCCCACGCGCCCGG
GACGCA
GTTCCGCGACGGGCCGGGGGACTTCAACTCGTCCACGTCGTTCCCGATGCCGCTG
CTGAT
GGCCGCCGCCTTCGACGACGAGCTGATCGAGGCCGTCGGCGACGTCATCGGCAC
CGAGGC
CCGCGCCTTTGGCAACGCCGGCTGGTCCGGCCTCGACTACTGGACCCCCAACGTC
AACCC
CTTCCGGGACCCCCGCTGGGGCCGCGGCTCCGAGACGCCGGGCGAGGACGTCGT
GCGCCT
CAAGCGCTACGCCGCCTCCATGATCCGCGGGCTCGAGGGTCGTTCCTCCTCCTCC
TCCTC
CTGCTCCTTCGGATCCGGAGGGGAGCCGCCGCGCGTCATCGCGACCTGCAAGCA
CTACGC
CGGCTATGACTTTGAGGACTGGAACGGCACGACGCGGCACGACTTCGACGCCGT
CATCTC
GGCGCAGGACCTGGCCGAGTACTACCTGGCGCCGTTCCAGCAGTGCGCGCGCGA
CTCGCG
CGTCGGCTCCGTCATGTGCGCCTACAACGCCGTCAACGGGGTGCCGTCGTGCGCC
AACTC
GTACCTCATGAACACGATCCTGCGCGGGCACTGGAACTGGACCGAGCACGACAA
CTACGT
CACCAGCGACTGCGAGGCCGTCCTCGACGTCTCGGCCCACCACCACTACGCCGAC
ACCAA
CGCCGAGGGCACCGCGCTCTGCTTCGAGGCCGGCATGGACACGAGCTGCGAGTA
CGAGGG
CTCCTCCGACATCCCGGGCGCCTCCGCCGGCGGCTTCCTGACCTGGCCCGCCGTC
GACCG
CGCCCTGACGCGGCTGTACCGGAGCCTGGTGCGGGTCGGCTACTTTGACGGCCCC
GAGTC
GCCGCACGCCTCGCTGGGCTGGGCCGACGTCAACCGGCCCGAGGCGCAGGAGCT
GGCCCT
GCGCGCTGCCGTCGAGGGCATCGTGCTGCTCAAGAACGACAACGACACGCTGCC
GCTGCC
GCTGCCGGACGATGTCGTTGTCACCGCTGATGGTGGCCGCCGCCGCGTCGCCATG
ATCGG

CTTCTGGGCCGACGCCCCGGACAAGCTGTTTGGCGGGTACAGCGGCGCGCCCCCC
TTCGC
GCGCTCGCCCGCGAGCGCCGCCCGGCAGCTGGGCTGGAACGTCACGGTCGCCGG
AGGGCC
CGTCCTGGAGGGAGACTCGGACGAGGAGGAGGACACGTGGACGGCGCCGGCCG
TCGAGGC
GGCCGCCGACGCCGACTACATCGTCTACTTTGGCGGCCTGGACACGTCGGCGGCG
GGCGA
GACCAAGGACCGGATGACGATCGGGTGGCCGGCGGCGCAGCTGGCGCTCATCTC
GGAGCT
GGCGCGGCTCGGCAAGCCCGTCGTGGTGGTGCAGATGGGCGACCAGCTCGACGA
CACGCC
CCTCTTCGAGCTGGACGGGGTGGGCGCCGTCCTGTGGGCCAACTGGCCGGGCCA
GGACGG
CGGCACGGCCGTGGTCCGGCTGCTCAGCGGCGCCGAGAGCCCGGCCGGCCGCCT
GCCCGT
GACCCAGTACCCGGCCAACTACACCGACGCGGTGCCCCTGACCGACATGACCCT
GCGCCC
GTCGGCGACCAACCCGGGCCGGACCTACCGCTGGTACCCGACTCCCGTCCGGCCC
TTCGG
CTTCGGCCTCCACTATACCACCTTCCGGGCCGAGTTCGGCCCCCACCCCTTCTTCC
CGGG
GGCGGGCAAGGGCGATGGCGACGGCGAGGACAAGGGCGAGAGCAAGAGCGAGA
TCAGGAC
GCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCGCAGGGCGGCGGCGGCGGCCA
CCACGCC
GATCCGGGACCTGCTCCGCGACTGCGACAAGACGTACCCGGACACGTGCCCGCT
GCCGCC
GCTGACGGTGCGCGTGACCAACGAGGGCGAGCGCGCGTCCGACTACGTGGTGCT
GGCCTT
CGTGTCGGGCGAGTACGGGCCGGCGCCGTACCCGATCAAGACGCTGGTCTCGTA
CGCGCG
GGCGCGCGGGCTAAAGGGGAAGGGCGGCGACGGCGACGGCGACGGCGACGGCG
CCACCAC
TACCGTCTCGCTCGACTGGACCGTCGGCAACCTGGCCCGCCACGACGAGCGCGG
CAACAC
AATCCTGTACCCGGGAACTTACACCCTCACTCTCGACGAGCCGGCCCAGGCGAGC
GTGCA
GTTCGCCCTCGAGGGCGAGCCCGTCGTGCTCGACGAGTGGCCTGCGCCGCCGAGT
GCCAA
CTCCACCGCCAGGGGGAGGCACAGGTAA (SEQ ID NO:14)

MKASVSCLVGMSAVAYGLDGPFQTYPDCTKPPLSDIKVCDRTLPEAERAAALVAAL
TDEE
KLQNLVSKAPGAPRIGLPAYNWWSEALHGVAHAPGTQFRDGPGDFNSSTSFPMPLL
MAAA
FDDELIEAVGDVIGTEARAFGNAGWSGLDYWTPNVNPFRDPRWGRGSETPGEDVVR
LKRY
AASMIRGLEGRSSSSSSCSFGSGGEPPRVIATCKHYAGYDFEDWNGTTRHDFDAVISA
QD

LAEYYLAPFQQCARDSRVGSVMCAYNAVNGVPSCANSYLMNTILRGHWNWTEHD
NYVTSD
CEAVLDVSAHHHYADTNAEGTALCFEAGMDTSCEYEGSSDIPGASAGGFLTWPAVD
RALT
RLYRSLVRVGYFDGPESPHASLGWADVNRPEAQELALRAAVEGIVLLKNDNDTLPL
PLPD
DVVVTADGGRRRVAMIGFWADAPDKLFGGYSGAPPFARSPASAARQLGWNVTVAG
GPVLE
GDSDEEEDTWTAPAVEAAADADYIVYFGGLDTSAAGETKDRMTIGWPAAQLALISE
LARL
GKPVVVVQMGDQLDDTPLFELDGVGAVLWANWPGQDGGTAVVRLLSGAESPAGR
LPVTQY
PANYTDAVPLTDMTLRPSATNPGRTYRWYPTPVRPFGFGLHYTTFRAEFGPHPFFPG
AGK
GDGDGEDKGESKSEIRTQQQQQQQQQRRAAAAATTPIRDLLRDCDKTYPDTCPLPP
LTV
RVTNEGERASDYVVLAFVSGEYGPAPYPIKTLVSYARARGLKGKGGDGDGDGDGA
TTTVS
LDWTVGNLARHDERGNTILYPGTYTLTLDEPAQASVQFALEGEPVVLDEWPAPPSA
NSTA
RGRHR (SEQ ID NO:15)

Variant 914:

ATGAAGGCCTCTGTATCATGCCTCGTCGGCATGAGCGCCGTGGCCTACGGCCTCG
ATGGC
CCTTTCCAGACCTACCCCGACTGCACCAAGCCCCCCCTGTCCGATATTAAGGTGT
GCGAC
CGGACACTGCCCGAGGCGGAGCGGGCGGCAGCCCTCGTGGCAGCCCTGACCGAC
GAGGAG
AAGCTGCAAAACCTGGTCAGGTATGTATGCGGAGAGAGAGAAACACACACACAC
GCGCGC
GCGCACACACACACACACACACACTCTCTCTCTCTCTCTCTCGCGTACCATGGGT
GCCGT
CTGACGTTTTCCCTTTGTCTCTGTGTCCAGCAAGGCGCCGGGGGCGCCGCGGATC
GGCCT
GCCCGCGTACAACTGGTGGAGCGAGGCGCTGCACGGGGTGGCCCACGCGCCCGG
GACGCA
GTTCCGCGACGGGCCGGGGGACTTCAACTCGTCCACGTCGTTCCCGATGCCGATT
CTGAT
GGCCGCCGCCTTCGACGACGAGCTGATCGAGGCCGTCGGCGACGTCATCGGCAC
CGAGGC
CCGCGCCTTTGGCAACGCCGGCTGGTCCGGCCTCGACTACTGGACCCCCAACGTC
AACCC
CTTCCGGGACCCCCGCTGGGGCCGCGGCTCCGAGACGCCGGGCGAGGACGTCGT
GCGCCT
CAAGCGCTACGCCGCCTCCATGATCCGCGGGCTCGAGGGTCGTTCCTCCTCCTCC
TCCTC

CTGCTCCTTCGGATCCGGAGGGGAGCCGCCGCGCGTCATCTCGACCTGCAAGCAC
TACGC
CGGCTATGACTTTGAGGACTGGAACGGCACGACGCGGCACGACTTCGACGCCGT
CATCTC
GGCGCAGGACCTGGCCGAGTACTACCTGGCGCCGTTCCAGCAGTGCGCGCGCGA
CTCGCG
CGTCGGCTCCGTCATGTGCGCCTACAACGCCGTCAACGGGGTGCCGTCGTGCGCC
AACTC
GTACCTCATGAACACGATCCTGCGCGGGCACTGGAACTGGACCGAGCACGACAA
CTACGT
CACCAGCGACTGCGAGGCCGTCCTCGACGTCTCGGCCCACCACCACTACGCCGAC
ACCAA
CGCCGAGGGCACCGGCCTCTGCTTCGAGGCCGGCATGGACACGAGCTGCGAGTA
CGAGGG
CTCCTCCGACATCCCGGGCGCCTCCGCCGGCGGCTTCCTGACCTGGCCCGCCGTC
GACCG
CGCCCTGACGCGGCTGTACCGGAGCCTGGTGCGGGTCGGCTACTTTGACGGCCCC
GAGTC
GCCGCACGCCTCGCTGGGCTGGGCCGACGTCAACCGGCCCGAGGCGCAGGAGCT
GGCCCT
GCGCGCTGCCGTCGAGGGCATCGTGCTGCTCAAGAACGACAACGACACGCTGCC
GCTGCC
GCTGCCGGACGATGTCGTTGTCACCGCTGATGGTGGCCGCCGCCGCGTCGCCATG
ATCGG
CTTCTGGGCCGACGCCCCGGACAAGCTGTTTGGCGGGTACAGCGGCGCGCCCCCC
TTCGC
GCGCTCGCCCGCGAGCGCCGCCCGGCAGCTGGGCTGGAACGTCACGGTCGCCGG
AGGGCC
CGTCCTGGAGGGAGACTCGGACGAGGAGGAGGACACGTGGACGGCGCCGGCCG
TCGAGGC
GGCCGCCGACGCCGACTACATCGTCTACTTTGGCGGCCTGGACACGTCGGCGGCG
GGCGA
GACCAAGGACCGGATGACGATCGGGTGGCCGGCGGCGCAGCTGGCGCTCATCTC
GGAGCT
GGCGCGGCTCGGCAAGCCCGTCGTGGTGGTGCAGATGGGCGACCAGCTCGACGA
CACGCC
CCTCTTCGAGCTGGACGGGGTGGGCGCCGTCCTGTGGGCCAACTATCCGGGCCAG
GACGG
CGGCACGGCCGTGGTCCGGCTGCTCAGCGGCGCCGAGAGCCCGGCCGGCCGCCT
GCCCGT
GACCCAGTACCCGGCCAACTACACCGACGCGGTGCCCCTGACCGACATGACCCT
GCGCCC
GTCGGCGACCAACCCGGGCCGGACCTACCGCTGGTACCCGACTCCCGTCCGGCCC
TTCGG
CTTCGGCCTCCACTATACCACCTTCCGGGCCGAGTTCGGCCCCCACCCCTTCTTCC
CGGG
GGCGGGCAAGGGCGATGGCGACGGCGAGGACAAGGGCGAGAGCAAGAGCGAGA
TCAGGAC
GCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCGCAGGGCGGCGGCGGCGGCCA
CCACGCC

GATCCGGGACCTGCTCCGCGACTGCGACAAGACGTACCCGGACACGTGCCCGCT
GCCGCC
GCTGACGGTGCGCGTGACCAACGAGGGCGAGCGCGCGTCCGACTACGTGGTGCT
GGCCTT
CGTGTCGGGCGAGTACGGGCCGGCGCCGTACCCGATCAAGACGCTGGTCTCGTA
CGCGCG
GGCGCGCGGGCTAAAGGGGAAGGGCGGCGACGGCGACGGCGACGGCGACGGCG
CCACCAC
TACCGTCTCGCTCGACTGGACCGTCGGCAACCTGGCCCGCCACGACGAGCGCGG
CAACAC
AATCCTGTACCCGGGAACTTACACCCTCACTCTCGACGAGCCGGCCCAGGCGAGC
GTGCA
GTTCGCCCTCGAGGGCGAGCCCGTCGTGCTCGACGAGTGGCCTGCGCCGCCGAGT
GCCAA
CTCCACCGCCAGGGGGAGGCACAGGTAA (SEQ ID NO:16)

MKASVSCLVGMSAVAYGLDGPFQTYPDCTKPPLSDIKVCDRTLPEAERAAALVAAL
TDEE
KLQNLVSKAPGAPRIGLPAYNWWSEALHGVAHAPGTQFRDGPGDFNSSTSFPMPIL
MAAA
FDDELIEAVGDVIGTEARAFGNAGWSGLDYWTPNVNPFRDPRWGRGSETPGEDVVR
LKRY
AASMIRGLEGRSSSSSSCSFGSGGEPPRVISTCKHYAGYDFEDWNGTTRHDFDAVISA
QD
LAEYYLAPFQQCARDSRVGSVMCAYNAVNGVPSCANSYLMNTILRGHWNWTEHD
NYVTSD
CEAVLDVSAHHHYADTNAEGTGLCFEAGMDTSCEYEGSSDIPGASAGGFLTWPAVD
RALT
RLYRSLVRVGYFDGPESPHASLGWADVNRPEAQELALRAAVEGIVLLKNDNDTLPL
PLPD
DVVVTADGGRRRVAMIGFWADAPDKLFGGYSGAPPFARSPASAARQLGWNVTVAG
GPVLE
GDSDEEEDTWTAPAVEAAADADYIVYFGGLDTSAAGETKDRMTIGWPAAQLALISE
LARL
GKPVVVVQMGDQLDDTPLFELDGVGAVLWANYPGQDGGTAVVRLLSGAESPAGRL
PVTQY
PANYTDAVPLTDMTLRPSATNPGRTYRWYPTPVRPFGFGLHYTTFRAEFGPHPFFPG
AGK
GDGDGEDKGESKSEIRTQQQQQQQQQRRAAAAATTPIRDLLRDCDKTYPDTCPLPP
LTV
RVTNEGERASDYVVLAFVSGEYGPAPYPIKTLVSYARARGLKGKGGDGDGDGDGA
TTTVS
LDWTVGNLARHDERGNTILYPGTYTLTLDEPAQASVQFALEGEPVVLDEWPAPPSA
NSTA
RGRHR (SEQ ID NO:17)

Variant 016:

ATGAAGGCCTCTGTATCATGCCTCGTCGGCATGAGCGCCGTGGCCTACGGCCTCG
ATGGC
CCTTTCCAGACCTACCCCGACTGCACCAAGCCCCCCCTGTCCGATATTAAGGTGT
GCGAC
CGGACACTGCCCGAGGCGGAGCGGGCGGCAGCCCTCGTGGCAGCCCTGACCGAC
GAGGAG
AAGCTGCAAAACCTGGTCAGGTATGTATGCGGAGAGAGAGAAACACACACACAC
GCGCGC
GCGCACACACACACACACACACACTCTCTCTCTCTCTCTCTCGCGTACCATGGGT
GCCGT
CTGACGTTTTCCCTTTGTCTCTGTGTCCAGCAAGGCGCCGGGGGCGCCGCGGATC
GGCCT
GCCCGCGTACAACTGGTGGAGCGAGGCGCTGCACGGGGTGGCCCACGCGCCCGG
GACGCA
GTTCCGCGACGGGCCGGGGGACTTCAACTCGTCCACGTCGTTCCCGATGCCGATT
CTGAT
GGCCGCCGCCTTCGACGACGAGCTGATCGAGGCCGTCGGCGACGTCATCGGCAC
CGAGGC
CCGCGCCTTTGGCAACGCCGGCTGGTCCGGCCTCGACTACTGGACCCCCAACGTC
AACCC
CTTCCGGGACCCCCGCTGGGGCCGCGGCTCCGAGACGCCGGGCGAGGACGTCGT
GCGCCT
CAAGCGCTACGCCGCCTCCATGATCCGCGGGCTCGAGGGTCGTTCCTCCTCCTCC
TCCTC
CTGCTCCTTCGGATCCGGAGGGGAGCCGCCGCGCGTCATCTCGACCTGCAAGCAC
TACGC
CGGCTATGACTTTGAGGACTGGAACGGCACGACGCGGCACGACTTCGACGCCGT
CATCTC
GGCGCAGGACCTGGCCGAGTACTACCTGGCGCCGTTCCAGCAGTGCGCGCGCGA
CTCGCG
CGTCGGCTCCGTCATGTGCGCCTACAACGCCGTCAACGGGGTGCCGTCGTGCGCC
AACTC
GTACCTCATGAACACGATCCTGCGCGGGCACTGGAACTGGACCGAGCACGACAA
CTACGT
CACCAGCGACTGCGAGGCCGTCCTCGACGTCTCGGCCCACCACCACTACGCCGAC
ACCAA
CGCCGAGGGCACCGGCCTCTGCTTCGAGGCCGGCATGGACACGAGCTGCGAGTA
CGAGGG
CTCCTCCGACATCCCGGGCGCCTCCGCCGGCGGCTTCCTGACCTGGCCCGCCGTC
GACCG
CGCCCTGACGCGGCTGTACCGGAGCCTGGTGCGGGTCGGCTACTTTGACGGCCCC
GAGTC
GCCGCACGCCTCGCTGGGCTGGGCCGACGTCAACCGGCCCGAGGCGCAGGAGCT
GGCCCT
GCGCGCTGCCGTCGAGGGCATCGTGCTGCTCAAGAACGACAACGACACGCTGCC
GCTGCC
GCTGCCGGACGATGTCGTTGTCACCGCTGATGGTGGCCGCCGCCGCGTCGCCATG
ATCGG

CTTCTGGGCCGACGCCCCGGACAAGCTGTTTGGCGGGTACAGCGGCGCGCCCCCC
TTCGC
GCGCTCGCCCGCGAGCGCCGCCCGGCAGCTGGGCTGGAACGTCACGGTCGCCGG
AGGGCC
CGTCCTGGAGGGAGACTCGGACGAGGAGGAGGACACGTGGACGGCGCCGGCCG
TCGAGGC
GGCCGCCGACGCCGACTACATCGTCTACTTTGGCGGCCTGGACACGTCGGCGGCG
GGCGA
GACCAAGGACCGGATGACGATCGGGTGGCCGGCGGCGCAGCTGGCGCTCATCTC
GGAGCT
GGCGCGGCTCGGCAAGCCCGTCGTGGTGGTGCAGATGGGCGACCAGCTCGACGA
CACGCC
CCTCTTCGAGCTGGACGGGGTGGGCGCCGTCCTGTGGGCCAACTATCCGGGCCAG
GACGG
CGGCACGGCCGTGGTCCGGCTGCTCAGCGGCGCCGAGAGCCCGGCCGGCCGCCT
GCCCGT
GACCCAGTACCCGGCCAACTACACCGACGCGGTGCCCCTGACCGACATGACCCT
GCGCCC
GTCGGCGACCAACCCGGGCCGGACCTACCGCTGGTACCCGACTCCCGTCCGGCCC
TTCGG
CTTCGGCCTCCACTATACCACCTTCCGGGCCGAGTTCGGCCCCCACCCCTTCTTCC
CGGG
GGCGGGCAAGGGCGATGGCGACGGCGAGGACAAGGGCGAGAGCAAGAGCGAGA
TCAGGAC
GCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCGCAGGGCGGCGGCGGCGGCCA
CCACGCC
GATCCGGGACCTGCTCCGCGACTGCGACAAGACGTACCCGGACACGTGCCCGCT
GCCGCC
GCTGACGGTGCGCGTGACCAACGAGGGCGAGCGCGCGTCCGACTACGTGGTGCT
GGCCTT
CGTGTCGGGCGAGTACGGGCCGGCGCCGTACCCGATCAAGACGCTGGTCTCGTA
CGCGCG
GGCGCGCGGGCTAAAGGGGAAGGGCGGCGACGGCGACGGCGACGGCGACGGCG
CCACCAC
TACCGTCTCGCTCGACTGGACCGTCGGCAACCTGGCCCGCCACGACGAGCGCGG
CAACAC
AATCCTGTACCCGGGAACTTACACCCTCACTCTCGACGAGCCGGCCCAGGCGAGC
GTGCA
GTTCGCCCTCGAGGGCGAGCCCGTCGTGCTCGACGAGTGGCCTGCGCCGCCGAGT
GCCAA
CTCCACCGCCAGGGGGAGGCACAGGTAA (SEQ ID NO:18)

MKASVSCLVGMSAVAYGLDGPFQTYPDCTKPPLSDIKVCDRTLPEAERAAALVAAL
TDEE
KLQNLVSKAPGAPRIGLPAYNWWSEALHGVAHAPGTQFRDGPGDFNSSTSFPMPIL
MAAA
FDDELIEAVGDVIGTEARAFGNAGWSGLDYWTPNVNPFRDPRWGRGSETPGEDVVR
LKRY
AASMIRGLEGRSSSSSSCSFGSGGEPPRVISTCKHYAGYDFEDWNGTTRHDFDAVISA
QD

LAEYYLAPFQQCARDSRVGSVMCAYNAVNGVPSCANSYLMNTILRGHWNWTEHD
NYVTSD
CEAVLDVSAHHHYADTNAEGTGLCFEAGMDTSCEYEGSSDIPGASAGGFLTWPAVD
RALT
RLYRSLVRVGYFDGPESPHASLGWADVNRPEAQELALRAAVEGIVLLKNDNDTLPL
PLPD
DVVVTADGGRRRVAMIGFWADAPDKLFGGYSGAPPFARSPASAARQLGWNVTVAG
GPVLE
GDSDEEEDTWTAPAVEAAADADYIVYFGGLDTSAAGETKDRMTIGWPAAQLALISE
LARL
GKPVVVVQMGDQLDDTPLFELDGVGAVLWANYPGQDGGTAVVRLLSGAESPAGRL
PVTQY
PANYTDAVPLTDMTLRPSATNPGRTYRWYPTPVRPFGFGLHYTTFRAEFGPHPFFPG
AGK
GDGDGEDKGESKSEIRTQQQQQQQQQRRAAAAATTPIRDLLRDCDKTYPDTCPLPP
LTV
RVTNEGERASDYVVLAFVSGEYGPAPYPIKTLVSYARARGLKGKGGDGDGDGDGA
TTTVS
LDWTVGNLARHDERGNTILYPGTYTLTLDEPAQASVQFALEGEPVVLDEWPAPPSA
NSTA
RGRHR (SEQ ID NO:19)

Variant 920:

ATGAAGGCCTCTGTATCATGCCTCGTCGGCATGAGCGCCGTGGCCTACGGCCTCG
ATGGC
CCTTTCCAGACCTACCCCGACTGCACCAAGCCCCCCCTGTCCGATATTAAGGTGT
GCGAC
CGGACACTGCCCGAGGCGGAGCGGGCGGCAGCCCTCGTGGCAGCCCTGACCGAC
GAGGAG
AAGCTGCAAAACCTGGTCAGGTATGTATGCGGAGAGAGAGAAACACACACACAC
GCGCGC
GCGCACACACACACACACACACACTCTCTCTCTCTCTCTCTCGCGTACCATGGGT
GCCGT
CTGACGTTTTCCCTTTGTCTCTGTGTCCAGCAAGGCGCCGGGGGCGCCGCGGATC
GGCCT
GCCCGCGTACAACTGGTGGAGCGAGGCGCTGCACGGGGTGGCCCACGCGCCCGG
GACGCA
GTTCCGCGACGGGCCGGGGGACTTCAACTCGTCCACGTCGTTCCCGATGCCGCTG
CTGAT
GGCCGCCGCCTTCGACGACGAGCTGATCGAGGCCGTCGGCGACGTCATCGGCAC
CGAGGC
CCGCGCCTTTGGCAACGCCGGCTGGTCCGGCCTCGACTACTGGACCCCCAACGTC
AACCC
CTTCCGGGACCCCCGCTGGGGCCGCGGCTCCGAGACGCCGGGCGAGGACGTCGT
GCGCCT
CAAGCGCTACGCCGCCTCCATGATCCGCGGGCTCGAGGGTCGTTCCTCCTCCTCC
TCCTC

CTGCTCCTTCGGATCCGGAGGGGAGCCGCCGCGCGTCATCGCGACCTGCAAGCA
CTACGC
CGGCTATGACTTTGAGGACTGGAACGGCACGACGCGGCACGACTTCGACGCCGT
CATCTC
GGCGCAGGACCTGGCCGAGTACTACCTGGCGCCGTTCCAGCAGTGCGCGCGCGA
CTCGCG
CGTCGGCTCCGTCATGTGCGCCTACAACGCCGTCAACGGGGTGCCGTCGTGCGCC
AACTC
GTACCTCCTGAACACGATCCTGCGCGGGCACTGGAACTGGACCGAGCACGACAA
CTACGT
CACCAGCGACTGCGAGGCCGTCCTCGACGTCTCGGCCCACCACCACTACGCCGAC
ACCAA
CGCCGAGGGCACCGGCCTCTGCTTCGAGGCCGGCATGGACACGAGCTGCGAGTA
CGAGGG
CTCCTCCGACATCCCGGGCGCCTCCGCCGGCGGCTTCCTGACCTGGCCCGCCGTC
GACCG
CGCCCTGACGCGGCTGTACCGGAGCCTGGTGCGGGTCGGCTACTTTGACGGCCCC
GAGTC
GCCGCACGCCTCGCTGGGCTGGGCCGACGTCAACCGGCCCGAGGCGCAGGAGCT
GGCCCT
GCGCGCTGCCGTCGAGGGCATCGTGCTGCTCAAGAACGACAACGACACGCTGCC
GCTGCC
GCTGCCGGACGATGTCGTTGTCACCGCTGATGGTGGCCGCCGCCGCGTCGCCATG
ATCGG
CTTCTGGGCCGACGCCCCGGACAAGCTGTTTGGCGGGTACAGCGGCGCGCCCCCC
TTCGC
GCGCTCGCCCGCGAGCGCCGCCCGGCAGCTGGGCTGGAACGTCACGGTCGCCGG
AGGGCC
CGTCCTGGAGGGAGACTCGGACGAGGAGGAGGACACGTGGACGGCGCCGGCCG
TCGAGGC
GGCCGCCGACGCCGACTACATCGTCTACTTTGGCGGCCTGGACACGTCGGCGGCG
GGCGA
GACCAAGGACCGGATGACGATCGGGTGGCCGGCGGCGCAGCTGGCGCTCATCTC
GGAGCT
GGCGCGGCTCGGCAAGCCCGTCGTGGTGGTGCAGATGGGCGACCAGCTCGACGA
CACGCC
CCTCTTCGAGCTGGACGGGGTGGGCGCCGTCCTGTGGGCCAACTATCCGGGCCAG
GACGG
CGGCACGGCCGTGGTCCGGCTGCTCAGCGGCGCCGAGAGCCCGGCCGGCCGCCT
GCCCGT
GACCCAGTACCCGGCCAACTACACCGACGCGGTGCCCCTGACCGACATGACCCT
GCGCCC
GTCGGCGACCAACCCGGGCCGGACCTACCGCTGGTACCCGACTCCCGTCCGGCCC
TTCGG
CTTCGGCCTCCACTATACCACCTTCCGGGCCGAGTTCGGCCCCCACCCCTTCTTCC
CGGG
GGCGGGCAAGGGCGATGGCGACGGCGAGGACAAGGGCGAGAGCAAGAGCGAGA
TCAGGAC
GCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCGCAGGGCGGCGGCGGCGGCCA
CCACGCC

GATCCGGGACCTGCTCCGCGACTGCGACAAGACGTACCCGGACACGTGCCCGCT
GCCGCC
GCTGACGGTGCGCGTGACCAACGAGGGCGAGCGCGCGTCCGACTACGTGGTGCT
GGCCTT
CGTGTCGGGCGAGTACGGGCCGGCGCCGTACCCGATCAAGACGCTGGTCTCGTA
CGCGCG
GGCGCGCGGGCTAAAGGGGAAGGGCGGCGACGGCGACGGCGACGGCGACGGCG
CCACCAC
TACCGTCTCGCTCGACTGGACCGTCGGCAACCTGGCCCGCCACGACGAGCGCGG
CAACAC
AATCCTGTACCCGGGAACTTACACCCTCACTCTCGACGAGCCGGCCCAGGCGAGC
GTGCA
GTTCGCCCTCGAGGGCGAGCCCGTCGTGCTCGACGAGTGGCCTGCGCCGCCGAGT
GCCAA
CTCCACCGCCAGGGGGAGGCACAGGTAA (SEQ ID NO:20)

MKASVSCLVGMSAVAYGLDGPFQTYPDCTKPPLSDIKVCDRTLPEAERAAALVAAL
TDEE
KLQNLVSKAPGAPRIGLPAYNWWSEALHGVAHAPGTQFRDGPGDFNSSTSFPMPLL
MAAA
FDDELIEAVGDVIGTEARAFGNAGWSGLDYWTPNVNPFRDPRWGRGSETPGEDVVR
LKRY
AASMIRGLEGRSSSSSSCSFGSGGEPPRVIATCKHYAGYDFEDWNGTTRHDFDAVISA
QD
LAEYYLAPFQQCARDSRVGSVMCAYNAVNGVPSCANSYLLNTILRGHWNWTEHDN
YVTSD
CEAVLDVSAHHHYADTNAEGTGLCFEAGMDTSCEYEGSSDIPGASAGGFLTWPAVD
RALT
RLYRSLVRVGYFDGPESPHASLGWADVNRPEAQELALRAAVEGIVLLKNDNDTLPL
PLPD
DVVVTADGGRRRVAMIGFWADAPDKLFGGYSGAPPFARSPASAARQLGWNVTVAG
GPVLE
GDSDEEEDTWTAPAVEAAADADYIVYFGGLDTSAAGETKDRMTIGWPAAQLALISE
LARL
GKPVVVVQMGDQLDDTPLFELDGVGAVLWANYPGQDGGTAVVRLLSGAESPAGRL
PVTQY
PANYTDAVPLTDMTLRPSATNPGRTYRWYPTPVRPFGFGLHYTTFRAEFGPHPFFPG
AGK
GDGDGEDKGESKSEIRTQQQQQQQQQRRAAAAATTPIRDLLRDCDKTYPDTCPLPP
LTV
RVTNEGERASDYVVLAFVSGEYGPAPYPIKTLVSYARARGLKGKGGDGDGDGDGA
TTTVS
LDWTVGNLARHDERGNTILYPGTYTLTLDEPAQASVQFALEGEPVVLDEWPAPPSA
NSTA
RGRHR (SEQ ID NO:21)

Variant 949:

ATGAAGGCCTCTGTATCATGCCTCGTCGGCATGAGCGCCGTGGCCTACGGCCTCG
ATGGC
CCTTTCCAGACCTACCCCGACTGCACCAAGCCCCCCCTGTCCGATATTAAGGTGT
GCGAC
CGGACACTGCCCGAGGCGGAGCGGGCGGCAGCCCTCGTGGCAGCCCTGACCGAC
GAGGAG
AAGCTGCAAAACCTGGTCAGGTATGTATGCGGAGAGAGAGAAACACACACACAC
GCGCGC
GCGCACACACACACACACACACACTCTCTCTCTCTCTCTCTCGCGTACCATGGGT
GCCGT
CTGACGTTTTCCCTTTGTCTCTGTGTCCAGCAAGGCGCCGGGGGCGCCGCGGATC
GGCCT
GCCCGCGTACAACTGGTGGAGCGAGGCGCTGCACGGGGTGGCCCACGCGCCCGG
GACGCA
GTTCCGCGACGGGCCGGGGGACTTCAACTCGTCCACGTCGTTCCCGATGCCGCTG
CTGAT
GGCCGCCGCCTTCGACGACGAGCTGATCGAGGCCGTCGGCGACGTCATCGGCAC
CGAGGC
CCGCGCCTTTGGCAACGCCGGCTGGTCCGGCCTCGACTACTGGACCCCCAACGTC
AACCC
CTTCCGGGACCCCCGCTGGGGCCGCGGCTCCGAGACGCCGGGCGAGGACGTCGT
GCGCCT
CAAGCGCTACGCCGCCTCCATGATCCGCGGGCTCGAGGGTCGTTCCTCCTCCTCC
TCCTC
CTGCTCCTTCGGATCCGGAGGGGAGCCGCCGCGCGTCATCTCGACCTGCAAGCAC
TACGC
CGGCTATGACTTTGAGGACTGGAACGGCACGACGCGGCACGACTTCGACGCCGT
CATCTC
GGCGCAGGACCTGGCCGAGTACTACCTGGCGCCGTTCCAGCAGTGCGCGCGCGA
CTCGCG
CGTCGGCTCCGTCATGTGCGCCTACAACGCCGTCAACGGGGTGCCGTCGTGCGCC
AACTC
GTACCTCATGAACACGATCCTGCGCGGGCACTGGAACTGGACCGAGCACGACAA
CTACGT
CACCAGCGACTGCGAGGCCGTCCTCGACGTCTCGGCCCACCACCACTACGCCGAC
ACCAA
CGCCGAGGGCACCGCGCTCTGCTTCGAGGCCGGCATGGACACGAGCTGCGAGTA
CGAGGG
CTCCTCCGACATCCCGGGCGCCTCCGCCGGCGGCTTCCTGACCTGGCCCGCCGTC
GACCG
CGCCCTGACGCGGCTGTACCGGAGCCTGGTGCGGGTCGGCTACTTTGACGGCCCC
GAGTC
GCCGCACGCCTCGCTGGGCTGGGCCGACGTCAACCGGCCCGAGGCGCAGGAGCT
GGCCCT
GCGCGCTGCCGTCGAGGGCATCGTGCTGCTCAAGAACGACAACGACACGCTGCC
GCTGCC
GCTGCCGGACGATGTCGTTGTCACCGCTGATGGTGGCCGCCGCCGCGTCGCCATG
ATCGG

CTTCTGGGCCGACGCCCCGGACAAGCTGTTTGGCGGGTACAGCGGCGCGCCCCCC
TTCGC
GCGCTCGCCCGCGAGCGCCGCCCGGCAGCTGGGCTGGAACGTCACGGTCGCCGG
AGGGCC
CGTCCTGGAGGGAGACTCGGACGAGGAGGAGGACACGTGGACGGCGCCGGCCG
TCGAGGC
GGCCGCCGACGCCGACTACATCGTCTACTTTGGCGGCCTGGACACGTCGGCGGCG
GGCGA
GACCAAGGACCGGATGACGATCGGGTGGCCGGCGGCGCAGCTGGCGCTCATCTC
GGAGCT
GGCGCGGCTCGGCAAGCCCGTCGTGGTGGTGCAGATGGGCGACCAGCTCGACGA
CACGCC
CCTCTTCGAGCTGGACGGGGTGGGCGCCGTCCTGTGGGCCAACTATCCGGGCCAG
GACGG
CGGCACGGCCGTGGTCCGGCTGCTCAGCGGCGCCGAGAGCCCGGCCGGCCGCCT
GCCCGT
GACCCAGTACCCGGCCAACTACACCGACGCGGTGCCCCTGACCGACATGACCCT
GCGCCC
GTCGGCGACCAACCCGGGCCGGACCTACCGCTGGTACCCGACTCCCGTCCGGCCC
TTCGG
CTTCGGCCTCCACTATACCACCTTCCGGGCCGAGTTCGGCCCCCACCCCTTCTTCC
CGGG
GGCGGGCAAGGGCGATGGCGACGGCGAGGACAAGGGCGAGAGCAAGAGCGAGA
TCAGGAC
GCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCGCAGGGCGGCGGCGGCGGCCA
CCACGCC
GATCCGGGACCTGCTCCGCGACTGCGACAAGACGTACCCGGACACGTGCCCGCT
GCCGCC
GCTGACGGTGCGCGTGACCAACGAGGGCGAGCGCGCGTCCGACTACGTGGTGCT
GGCCTT
CGTGTCGGGCGAGTACGGGCCGGCGCCGTACCCGATCAAGACGCTGGTCTCGTA
CGCGCG
GGCGCGCGGGCTAAAGGGGAAGGGCGGCGACGGCGACGGCGACGGCGACGGCG
CCACCAC
TACCGTCTCGCTCGACTGGACCGTCGGCAACCTGGCCCGCCACGACGAGCGCGG
CAACAC
AATCCTGTACCCGGGAACTTACACCCTCACTCTCGACGAGCCGGCCCAGGCGAGC
GTGCA
GTTCGCCCTCGAGGGCGAGCCCGTCGTGCTCGACGAGTGGCCTGCGCCGCCGAGT
GCCAA
CTCCACCGCCAGGGGGAGGCACAGGTAA (SEQ ID NO:22)

MKASVSCLVGMSAVAYGLDGPFQTYPDCTKPPLSDIKVCDRTLPEAERAAALVAAL
TDEE
KLQNLVSKAPGAPRIGLPAYNWWSEALHGVAHAPGTQFRDGPGDFNSSTSFPMPLL
MAAA
FDDELIEAVGDVIGTEARAFGNAGWSGLDYWTPNVNPFRDPRWGRGSETPGEDVVR
LKRY
AASMIRGLEGRSSSSSSCSFGSGGEPPRVISTCKHYAGYDFEDWNGTTRHDFDAVISA
QD

LAEYYLAPFQQCARDSRVGSVMCAYNAVNGVPSCANSYLMNTILRGHWNWTEHD
NYVTSD
CEAVLDVSAHHHYADTNAEGTALCFEAGMDTSCEYEGSSDIPGASAGGFLTWPAVD
RALT
RLYRSLVRVGYFDGPESPHASLGWADVNRPEAQELALRAAVEGIVLLKNDNDTLPL
PLPD
DVVVTADGGRRRVAMIGFWADAPDKLFGGYSGAPPFARSPASAARQLGWNVTVAG
GPVLE
GDSDEEEDTWTAPAVEAAADADYIVYFGGLDTSAAGETKDRMTIGWPAAQLALISE
LARL
GKPVVVVQMGDQLDDTPLFELDGVGAVLWANYPGQDGGTAVVRLLSGAESPAGRL
PVTQY
PANYTDAVPLTDMTLRPSATNPGRTYRWYPTPVRPFGFGLHYTTFRAEFGPHPFFPG
AGK
GDGDGEDKGESKSEIRTQQQQQQQQQRRAAAAATTPIRDLLRDCDKTYPDTCPLPP
LTV
RVTNEGERASDYVVLAFVSGEYGPAPYPIKTLVSYARARGLKGKGGDGDGDGDGA
TTTVS
LDWTVGNLARHDERGNTILYPGTYTLTLDEPAQASVQFALEGEPVVLDEWPAPPSA
NSTA
RGRHR (SEQ ID NO:23)

Variant V209I:

ATGAAGGCCTCTGTATCATGCCTCGTCGGCATGAGCGCCGTGGCCTACGGCCTCG
ATGGCCCTTTCCAGACCTACCCCGACTGCACCAAGCCCCCCCTGTCCGATATTAA
GGTGTGCGACCGGACACTGCCCGAGGCGGAGCGGGCGGCAGCCCTCGTGGCAGC
CCTGACCGACGAGGAGAAGCTGCAAAACCTGGTCAGCAAGGCGCCGGGGGCGCC
GCGGATCGGCCTGCCCGCGTACAACTGGTGGAGCGAGGCGCTGCACGGGGTGGC
CCACGCGCCCGGGACGCAGTTCCGCGACGGGCCGGGGGACTTCAACTCGTCCAC
GTCGTTCCCGATGCCGCTGCTGATGGCCGCCGCCTTCGACGACGAGCTGATCGAG
GCCGTCGGCGACGTCATCGGCACCGAGGCCCGCGCCTTTGGCAACGCCGGCTGG
TCCGGCCTCGACTACTGGACCCCCAACGTCAACCCCTTCCGGGACCCCCGCTGGG
GCCGCGGCTCCGAGACGCCGGGCGAGGACGTCGTGCGCCTCAAGCGCTACGCCG
CCTCCATGATCCGCGGGCTCGAGGGTCGTTCCTCCTCCTCCTCCTCCTGCTCCTTC
GGATCCGGAGGGGAGCCGCCGCGCATCATCTCGACCTGCAAGCACTACGCCGGC
AACGACTTTGAGGACTGGAACGGCACGACGCGGCACGACTTCGACGCCGTCATC
TCGGCGCAGGACCTGGCCGAGTACTACCTGGCGCCGTTCCAGCAGTGCGCGCGC
GACTCGCGCGTCGGCTCCGTCATGTGCGCCTACAACGCCGTCAACGGGGTGCCGT
CGTGCGCCAACTCGTACCTCATGAACACGATCCTGCGCGGGCACTGGAACTGGA
CCGAGCACGACAACTACGTCACCAGCGACTGCGAGGCCGTCCTCGACGTCTCGG
CCCACCACCACTACGCCGACACCAACGCCGAGGGCACCGGCCTCTGCTTCGAGG
CCGGCATGGACACGAGCTGCGAGTACGAGGGCTCCTCCGACATCCCGGGCGCCT
CCGCCGGCGGCTTCCTGACCTGGCCCGCCGTCGACCGCGCCCTGACGCGGCTGTA
CCGGAGCCTGGTGCGGGTCGGCTACTTTGACGGCCCCGAGTCGCCGCACGCCTCG
CTGGGCTGGGCCGACGTCAACCGGCCCGAGGCGCAGGAGCTGGCCCTGCGCGCT
GCCGTCGAGGGCATCGTGCTGCTCAAGAACGACAACGACACGCTGCCGCTGCCG

CTGCCGGACGATGTCGTTGTCACCGCTGATGGTGGCCGCCGCCGCGTCGCCATGA
TCGGCTTCTGGGCCGACGCCCCGGACAAGCTGTTTGGCGGGTACAGCGGCGCGC
CCCCCTTCGCGCGCTCGCCCGCGAGCGCCGCCCGGCAGCTGGGCTGGAACGTCAC
GGTCGCCGGAGGGCCCGTCCTGGAGGGAGACTCGGACGAGGAGGAGGACACGT
GGACGGCGCCGGCCGTCGAGGCGGCCGCCGACGCCGACTACATCGTCTACTTTG
GCGGCCTGGACACGTCGGCGGCGGGCGAGACCAAGGACCGGATGACGATCGGGT
GGCCGGCGGCGCAGCTGGCGCTCATCTCGGAGCTGGCGCGGCTCGGCAAGCCCG
TCGTGGTGGTGCAGATGGGCGACCAGCTCGACGACACGCCCCTCTTCGAGCTGG
ACGGGGTGGGCGCCGTCCTGTGGGCCAACTGGCCGGGCCAGGACGGCGGCACGG
CCGTGGTCCGGCTGCTCAGCGGCGCCGAGAGCCCGGCCGGCCGCCTGCCCGTGA
CCCAGTACCCGGCCAACTACACCGACGCGGTGCCCCTGACCGACATGACCCTGC
GCCCGTCGGCGACCAACCCGGGCCGGACCTACCGCTGGTACCCGACTCCCGTCCG
GCCCTTCGGCTTCGGCCTCCACTATACCACCTTCCGGGCCGAGTTCGGCCCCCAC
CCCTTCTTCCCGGGGGCGGGCAAGGGCGATGGCGACGGCGAGGACAAGGGCGAG
AGCAAGAGCGAGATCAGGACGCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCG
CAGGGCGGCGGCGGCGGCCACCACGCCGATCCGGGACCTGCTCCGCGACTGCGA
CAAGACGTACCCGGACACGTGCCCGCTGCCGCCGCTGACGGTGCGCGTGACCAA
CGAGGGCGAGCGCGCGTCCGACTACGTGGTGCTGGCCTTCGTGTCGGGCGAGTA
CGGGCCGGCGCCGTACCCGATCAAGACGCTGGTCTCGTACGCGCGGGCGCGCGG
GCTAAAGGGGAAGGGCGGCGACGGCGACGGCGACGGCGACGGCGCCACCACTA
CCGTCTCGCTCGACTGGACCGTCGGCAACCTGGCCCGCCACGACGAGCGCGGCA
ACACAATCCTGTACCCGGGAACTTACACCCTCACTCTCGACGAGCCGGCCCAGGC
GAGCGTGCAGTTCGCCCTCGAGGGCGAGCCCGTCGTGCTCGACGAGTGGCCTGC
GCCGCCGAGTGCCAACTCCACCGCCAGGGGGAGGCACAGG (SEQ ID NO:24)

MKASVSCLVGMSAVAYGLDGPFQTYPDCTKPPLSDIKVCDRTLPEAERAAALVAAL
TDEEKLQNLVSKAPGAPRIGLPAYNWWSEALHGVAHAPGTQFRDGPGDFNSSTSFP
MPLLMAAAFDDELIEAVGDVIGTEARAFGNAGWSGLDYWTPNVNPFRDPRWGRGS
ETPGEDVVRLKRYAASMIRGLEGRSSSSSSCSFGSGGEPPRIISTCKHYAGNDFEDWN
GTTRHDFDAVISAQDLAEYYLAPFQQCARDSRVGSVMCAYNAVNGVPSCANSYLM
NTILRGHWNWTEHDNYVTSDCEAVLDVSAHHHYADTNAEGTGLCFEAGMDTSCEY
EGSSDIPGASAGGFLTWPAVDRALTRLYRSLVRVGYFDGPESPHASLGWADVNRPE
AQELALRAAVEGIVLLKNDNDTLPLPLPDDVVVTADGGRRRVAMIGFWADAPDKLF
GGYSGAPPFARSPASAARQLGWNVTVAGGPVLEGDSDEEEDTWTAPAVEAAADAD
YIVYFGGLDTSAAGETKDRMTIGWPAAQLALISELARLGKPVVVVQMGDQLDDTPL
FELDGVGAVLWANWPGQDGGTAVVRLLSGAESPAGRLPVTQYPANYTDAVPLTDM
TLRPSATNPGRTYRWYPTPVRPFGFGLHYTTFRAEFGPHPFFPGAGKGDGDGEDKGE
SKSEIRTQQQQQQQQQRRAAAAATTPIRDLLRDCDKTYPDTCPLPPLTVRVTNEGE
RASDYVVLAFVSGEYGPAPYPIKTLVSYARARGLKGKGGTGAGDGDVATTTVSLD
WTVGNLARHDERGNTILYPGTYTLTLDEPAQASVQFALEGEPVVLDEWPAPPSANST
ARGRHR (SEQ ID NO:25)

Variant S211A:

ATGAAGGCCTCTGTATCATGCCTCGTCGGCATGAGCGCCGTGGCCTACGGCCTCG
ATGGCCCTTTCCAGACCTACCCCGACTGCACCAAGCCCCCCCTGTCCGATATTAA
GGTGTGCGACCGGACACTGCCCGAGGCGGAGCGGGCGGCAGCCCTCGTGGCAGC
CCTGACCGACGAGGAGAAGCTGCAAAACCTGGTCAGCAAGGCGCCGGGGGCGCC

GCGGATCGGCCTGCCCGCGTACAACTGGTGGAGCGAGGCGCTGCACGGGGTGGC
CCACGCGCCCGGGACGCAGTTCCGCGACGGGCCGGGGGACTTCAACTCGTCCAC
GTCGTTCCCGATGCCGCTGCTGATGGCCGCCGCCTTCGACGACGAGCTGATCGAG
GCCGTCGGCGACGTCATCGGCACCGAGGCCCGCGCCTTTGGCAACGCCGGCTGG
TCCGGCCTCGACTACTGGACCCCCAACGTCAACCCCTTCCGGGACCCCCGCTGGG
GCCGCGGCTCCGAGACGCCGGGCGAGGACGTCGTGCGCCTCAAGCGCTACGCCG
CCTCCATGATCCGCGGGCTCGAGGGTCGTTCCTCCTCCTCCTCCTCCTGCTCCTTC
GGATCCGGAGGGGAGCCGCCGCGCGTCATCGCCACCTGCAAGCACTACGCCGGC
AACGACTTTGAGGACTGGAACGGCACGACGCGGCACGACTTCGACGCCGTCATC
TCGGCGCAGGACCTGGCCGAGTACTACCTGGCGCCGTTCCAGCAGTGCGCGCGC
GACTCGCGCGTCGGCTCCGTCATGTGCGCCTACAACGCCGTCAACGGGGTGCCGT
CGTGCGCCAACTCGTACCTCATGAACACGATCCTGCGCGGGCACTGGAACTGGA
CCGAGCACGACAACTACGTCACCAGCGACTGCGAGGCCGTCCTCGACGTCTCGG
CCCACCACCACTACGCCGACACCAACGCCGAGGGCACCGGCCTCTGCTTCGAGG
CCGGCATGGACACGAGCTGCGAGTACGAGGGCTCCTCCGACATCCCGGGCGCCT
CCGCCGGCGGCTTCCTGACCTGGCCCGCCGTCGACCGCGCCCTGACGCGGCTGTA
CCGGAGCCTGGTGCGGGTCGGCTACTTTGACGGCCCCGAGTCGCCGCACGCCTCG
CTGGGCTGGGCCGACGTCAACCGGCCCGAGGCGCAGGAGCTGGCCCTGCGCGCT
GCCGTCGAGGGCATCGTGCTGCTCAAGAACGACAACGACACGCTGCCGCTGCCG
CTGCCGGACGATGTCGTTGTCACCGCTGATGGTGGCCGCCGCCGCGTCGCCATGA
TCGGCTTCTGGGCCGACGCCCCGGACAAGCTGTTTGGCGGGTACAGCGGCGCGC
CCCCCTTCGCGCGCTCGCCCGCGAGCGCCGCCCGGCAGCTGGGCTGGAACGTCAC
GGTCGCCGGAGGGCCCGTCCTGGAGGGAGACTCGGACGAGGAGGAGGACACGT
GGACGGCGCCGGCCGTCGAGGCGGCCGCCGACGCCGACTACATCGTCTACTTTG
GCGGCCTGGACACGTCGGCGGCGGGCGAGACCAAGGACCGGATGACGATCGGGT
GGCCGGCGGCGCAGCTGGCGCTCATCTCGGAGCTGGCGCGGCTCGGCAAGCCCG
TCGTGGTGGTGCAGATGGGCGACCAGCTCGACGACACGCCCCTCTTCGAGCTGG
ACGGGGTGGGCGCCGTCCTGTGGGCCAACTGGCCGGGCCAGGACGGCGGCACGG
CCGTGGTCCGGCTGCTCAGCGGCGCCGAGAGCCCGGCCGGCCGCCTGCCCGTGA
CCCAGTACCCGGCCAACTACACCGACGCGGTGCCCCTGACCGACATGACCCTGC
GCCCGTCGGCGACCAACCCGGGCCGGACCTACCGCTGGTACCCGACTCCCGTCCG
GCCCTTCGGCTTCGGCCTCCACTATACCACCTTCCGGGCCGAGTTCGGCCCCCAC
CCCTTCTTCCCGGGGGCGGGCAAGGGCGATGGCGACGGCGAGGACAAGGGCGAG
AGCAAGAGCGAGATCAGGACGCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCG
CAGGGCGGCGGCGGCGGCCACCACGCCGATCCGGGACCTGCTCCGCGACTGCGA
CAAGACGTACCCGGACACGTGCCCGCTGCCGCCGCTGACGGTGCGCGTGACCAA
CGAGGGCGAGCGCGCGTCCGACTACGTGGTGCTGGCCTTCGTGTCGGGCGAGTA
CGGGCCGGCGCCGTACCCGATCAAGACGCTGGTCTCGTACGCGCGGGCGCGCGG
GCTAAAGGGGAAGGGCGGCGACGGCGACGGCGACGGCGACGGCGCCACCACTA
CCGTCTCGCTCGACTGGACCGTCGGCAACCTGGCCCGCCACGACGAGCGCGGCA
ACACAATCCTGTACCCGGGAACTTACACCCTCACTCTCGACGAGCCGGCCCAGGC
GAGCGTGCAGTTCGCCCTCGAGGGCGAGCCCGTCGTGCTCGACGAGTGGCCTGC
GCCGCCGAGTGCCAACTCCACCGCCAGGGGGAGGCACAGG (SEQ ID NO:26)

MKASVSCLVGMSAVAYGLDGPFQTYPDCTKPPLSDIKVCDRTLPEAERAAALVAAL
TDEEKLQNLVSKAPGAPRIGLPAYNWWSEALHGVAHAPGTQFRDGPGDFNSSTSFP
MPLLMAAAFDDELIEAVGDVIGTEARAFGNAGWSGLDYWTPNVNPFRDPRWGRGS
ETPGEDVVRLKRYAASMIRGLEGRSSSSSSCSFGSGGEPPRVIATCKHYAGNDFEDW
NGTTRHDFDAVISAQDLAEYYLAPFQQCARDSRVGSVMCAYNAVNGVPSCANSYL
MNTILRGHWNWTEHDNYVTSDCEAVLDVSAHHHYADTNAEGTGLCFEAGMDTSC

EYEGSSDIPGASAGGFLTWPAVDRALTRLYRSLVRVGYFDGPESPHASLGWADVNR
PEAQELALRAAVEGIVLLKNDNDTLPLPLPDDVVVTADGGRRRVAMIGFWADAPDK
LFGGYSGAPPFARSPASAARQLGWNVTVAGGPVLEGDSDEEEDTWTAPAVEAAAD
ADYIVYFGGLDTSAAGETKDRMTIGWPAAQLALISELARLGKPVVVVQMGDQLDDT
PLFELDGVGAVLWANWPGQDGGTAVVRLLSGAESPAGRLPVTQYPANYTDAVPLT
DMTLRPSATNPGRTYRWYPTPVRPFGFGLHYTTFRAEFGPHPFFPGAGKGDGDGED
KGESKSEIRTQQQQQQQQQQRRAAAAATTPIRDLLRDCDKTYPDTCPLPPLTVRVTN
EGERASDYVVLAFVSGEYGPAPYPIKTLVSYARARGLKGKGGTGAGDGDVATTTVS
LDWTVGNLARHDERGNTILYPGTYTLTLDEPAQASVQFALEGEPVVLDEWPAPPSA
NSTARGRHR (SEQ ID NO:27)

Variant S211A/ N219Y:

ATGAAGGCCTCTGTATCATGCCTCGTCGGCATGAGCGCCGTGGCCTACGGCCTCG
ATGGCCCTTTCCAGACCTACCCCGACTGCACCAAGCCCCCCCTGTCCGATATTAA
GGTGTGCGACCGGACACTGCCCGAGGCGGAGCGGGCGGCAGCCCTCGTGGCAGC
CCTGACCGACGAGGAGAAGCTGCAAAACCTGGTCAGCAAGGCGCCGGGGGCGCC
GCGGATCGGCCTGCCCGCGTACAACTGGTGGAGCGAGGCGCTGCACGGGGTGGC
CCACGCGCCCGGGACGCAGTTCCGCGACGGGCCGGGGGACTTCAACTCGTCCAC
GTCGTTCCCGATGCCGCTGCTGATGGCCGCCGCCTTCGACGACGAGCTGATCGAG
GCCGTCGGCGACGTCATCGGCACCGAGGCCCGCGCCTTTGGCAACGCCGGCTGG
TCCGGCCTCGACTACTGGACCCCCAACGTCAACCCCTTCCGGGACCCCCGCTGGG
GCCGCGGCTCCGAGACGCCGGGCGAGGACGTCGTGCGCCTCAAGCGCTACGCCG
CCTCCATGATCCGCGGGCTCGAGGGTCGTTCCTCCTCCTCCTCCTCCTGCTCCTTC
GGATCCGGAGGGGAGCCGCCGCGCGTCATCGCCACCTGCAAGCACTACGCCGGC
TACGACTTTGAGGACTGGAACGGCACGACGCGGCACGACTTCGACGCCGTCATC
TCGGCGCAGGACCTGGCCGAGTACTACCTGGCGCCGTTCCAGCAGTGCGCGCGC
GACTCGCGCGTCGGCTCCGTCATGTGCGCCTACAACGCCGTCAACGGGGTGCCGT
CGTGCGCCAACTCGTACCTCATGAACACGATCCTGCGCGGGCACTGGAACTGGA
CCGAGCACGACAACTACGTCACCAGCGACTGCGAGGCCGTCCTCGACGTCTCGG
CCCACCACCACTACGCCGACACCAACGCCGAGGGCACCGGCCTCTGCTTCGAGG
CCGGCATGGACACGAGCTGCGAGTACGAGGGCTCCTCCGACATCCCGGGCGCCT
CCGCCGGCGGCTTCCTGACCTGGCCCGCCGTCGACCGCGCCCTGACGCGGCTGTA
CCGGAGCCTGGTGCGGGTCGGCTACTTTGACGGCCCCGAGTCGCCGCACGCCTCG
CTGGGCTGGGCCGACGTCAACCGGCCCGAGGCGCAGGAGCTGGCCCTGCGCGCT
GCCGTCGAGGGCATCGTGCTGCTCAAGAACGACAACGACACGCTGCCGCTGCCG
CTGCCGGACGATGTCGTTGTCACCGCTGATGGTGGCCGCCGCCGCGTCGCCATGA
TCGGCTTCTGGGCCGACGCCCCGGACAAGCTGTTTGGCGGGTACAGCGGCGCGC
CCCCCTTCGCGCGCTCGCCCGCGAGCGCCGCCCGGCAGCTGGGCTGGAACGTCAC
GGTCGCCGGAGGGCCCGTCCTGGAGGGAGACTCGGACGAGGAGGAGGACACGT
GGACGGCGCCGGCCGTCGAGGCGGCCGCCGACGCCGACTACATCGTCTACTTTG
GCGGCCTGGACACGTCGGCGGCGGGCGAGACCAAGGACCGGATGACGATCGGGT
GGCCGGCGGCGCAGCTGGCGCTCATCTCGGAGCTGGCGCGGCTCGGCAAGCCCG
TCGTGGTGGTGCAGATGGGCGACCAGCTCGACGACACGCCCCTCTTCGAGCTGG
ACGGGGTGGGCGCCGTCCTGTGGGCCAACTGGCCGGGCCAGGACGGCGGCACGG
CCGTGGTCCGGCTGCTCAGCGGCGCCGAGAGCCCGGCCGGCCGCCTGCCCGTGA
CCCAGTACCCGGCCAACTACACCGACGCGGTGCCCCTGACCGACATGACCCTGC
GCCCGTCGGCGACCAACCCGGGCCGGACCTACCGCTGGTACCCGACTCCCGTCCG

GCCCTTCGGCTTCGGCCTCCACTATACCACCTTCCGGGCCGAGTTCGGCCCCCAC
CCCTTCTTCCCGGGGGCGGGCAAGGGCGATGGCGACGGCGAGGACAAGGGCGAG
AGCAAGAGCGAGATCAGGACGCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCG
CAGGGCGGCGGCGGCGGCCACCACGCCGATCCGGGACCTGCTCCGCGACTGCGA
CAAGACGTACCCGGACACGTGCCCGCTGCCGCCGCTGACGGTGCGCGTGACCAA
CGAGGGCGAGCGCGCGTCCGACTACGTGGTGCTGGCCTTCGTGTCGGGCGAGTA
CGGGCCGGCGCCGTACCCGATCAAGACGCTGGTCTCGTACGCGCGGGCGCGCGG
GCTAAAGGGGAAGGGCGGCGACGGCGACGGCGACGGCGACGGCGCCACCACTA
CCGTCTCGCTCGACTGGACCGTCGGCAACCTGGCCCGCCACGACGAGCGCGGCA
ACACAATCCTGTACCCGGGAACTTACACCCTCACTCTCGACGAGCCGGCCCAGGC
GAGCGTGCAGTTCGCCCTCGAGGGCGAGCCCGTCGTGCTCGACGAGTGGCCTGC
GCCGCCGAGTGCCAACTCCACCGCCAGGGGGAGGCACAGG (SEQ ID NO:28)

MKASVSCLVGMSAVAYGLDGPFQTYPDCTKPPLSDIKVCDRTLPEAERAAALVAAL
TDEEKLQNLVSKAPGAPRIGLPAYNWWSEALHGVAHAPGTQFRDGPGDFNSSTSFP
MPLLMAAAFDDELIEAVGDVIGTEARAFGNAGWSGLDYWTPNVNPFRDPRWGRGS
ETPGEDVVRLKRYAASMIRGLEGRSSSSSSCSFGSGGEPPRVIATCKHYAGYDFEDW
NGTTRHDFDAVISAQDLAEYYLAPFQQCARDSRVGSVMCAYNAVNGVPSCANSYL
MNTILRGHWNWTEHDNYVTSDCEAVLDVSAHHHYADTNAEGTGLCFEAGMDTSC
EYEGSSDIPGASAGGFLTWPAVDRALTRLYRSLVRVGYFDGPESPHASLGWADVNR
PEAQELALRAAVEGIVLLKNDNDTLPLPLPDDVVVTADGGRRRVAMIGFWADAPDK
LFGGYSGAPPFARSPASAARQLGWNVTVAGGPVLEGDSDEEEDTWTAPAVEAAAD
ADYIVYFGGLDTSAAGETKDRMTIGWPAAQLALISELARLGKPVVVVQMGDQLDDT
PLFELDGVGAVLWANWPGQDGGTAVVRLLSGAESPAGRLPVTQYPANYTDAVPLT
DMTLRPSATNPGRTYRWYPTPVRPFGFGLHYTTFRAEFGPHPFFPGAGKGDGDGED
KGESKSEIRTQQQQQQQQQQRRAAAAATTPIRDLLRDCDKTYPDTCPLPPLTVRVTN
EGERASDYVVLAFVSGEYGPAPYPIKTLVSYARARGLKGKGGTGAGDGDVATTTVS
LDWTVGNLARHDERGNTILYPGTYTLTLDEPAQASVQFALEGEPVVLDEWPAPPSA
NSTARGRHR (SEQ ID NO:29)

Variant S108A/S211A/M280L/L761I:

ATGAAGGCCTCTGTATCATGCCTCGTCGGCATGAGCGCCGTGGCCTACGGCCTCG
ATGGCCCTTTCCAGACCTACCCCGACTGCACCAAGCCCCCCCTGTCCGATATTAA
GGTGTGCGACCGGACACTGCCCGAGGCGGAGCGGGCGGCAGCCCTCGTGGCAGC
CCTGACCGACGAGGAGAAGCTGCAAAACCTGGTCAGCAAGGCGCCGGGGGCGCC
GCGGATCGGCCTGCCCGCGTACAACTGGTGGAGCGAGGCGCTGCACGGGGTGGC
CCACGCGCCCGGGACGCAGTTCCGCGACGGGCCGGGGGACTTCAACTCGGCCAC
GTCGTTCCCGATGCCGCTGCTGATGGCCGCCGCCTTCGACGACGAGCTGATCGAG
GCCGTCGGCGACGTCATCGGCACCGAGGCCCGCGCCTTTGGCAACGCCGGCTGG
TCCGGCCTCGACTACTGGACCCCCAACGTCAACCCCTTCCGGGACCCCCGCTGGG
GCCGCGGCTCCGAGACGCCGGGCGAGGACGTCGTGCGCCTCAAGCGCTACGCCG
CCTCCATGATCCGCGGGCTCGAGGGTCGTTCCTCCTCCTCCTCCTCCTGCTCCTTC
GGATCCGGAGGGGAGCCGCCGCGCGTCATCGCCACCTGCAAGCACTACGCCGGC
AACGACTTTGAGGACTGGAACGGCACGACGCGGCACGACTTCGACGCCGTCATC
TCGGCGCAGGACCTGGCCGAGTACTACCTGGCGCCGTTCCAGCAGTGCGCGCGC
GACTCGCGCGTCGGCTCCGTCATGTGCGCCTACAACGCCGTCAACGGGGTGCCGT
CGTGCGCCAACTCGTACCTCCTCAACACGATCCTGCGCGGGCACTGGAACTGGAC

CGAGCACGACAACTACGTCACCAGCGACTGCGAGGCCGTCCTCGACGTCTCGGC
CCACCACCACTACGCCGACACCAACGCCGAGGGCACCGGCCTCTGCTTCGAGGC
CGGCATGGACACGAGCTGCGAGTACGAGGGCTCCTCCGACATCCCGGGCGCCTC
CGCCGGCGGCTTCCTGACCTGGCCCGCCGTCGACCGCGCCCTGACGCGGCTGTAC
CGGAGCCTGGTGCGGGTCGGCTACTTTGACGGCCCCGAGTCGCCGCACGCCTCGC
TGGGCTGGGCCGACGTCAACCGGCCCGAGGCGCAGGAGCTGGCCCTGCGCGCTG
CCGTCGAGGGCATCGTGCTGCTCAAGAACGACAACGACACGCTGCCGCTGCCGC
TGCCGGACGATGTCGTTGTCACCGCTGATGGTGGCCGCCGCCGCGTCGCCATGAT
CGGCTTCTGGGCCGACGCCCCGGACAAGCTGTTTGGCGGGTACAGCGGCGCGCC
CCCCTTCGCGCGCTCGCCCGCGAGCGCCGCCCGGCAGCTGGGCTGGAACGTCAC
GGTCGCCGGAGGGCCCGTCCTGGAGGGAGACTCGGACGAGGAGGAGGACACGT
GGACGGCGCCGGCCGTCGAGGCGGCCGCCGACGCCGACTACATCGTCTACTTTG
GCGGCCTGGACACGTCGGCGGCGGGCGAGACCAAGGACCGGATGACGATCGGGT
GGCCGGCGGCGCAGCTGGCGCTCATCTCGGAGCTGGCGCGGCTCGGCAAGCCCG
TCGTGGTGGTGCAGATGGGCGACCAGCTCGACGACACGCCCCTCTTCGAGCTGG
ACGGGGTGGGCGCCGTCCTGTGGGCCAACTGGCCGGGCCAGGACGGCGGCACGG
CCGTGGTCCGGCTGCTCAGCGGCGCCGAGAGCCCGGCCGGCCGCCTGCCCGTGA
CCCAGTACCCGGCCAACTACACCGACGCGGTGCCCCTGACCGACATGACCCTGC
GCCCGTCGGCGACCAACCCGGGCCGGACCTACCGCTGGTACCCGACTCCCGTCCG
GCCCTTCGGCTTCGGCCTCCACTATACCACCTTCCGGGCCGAGTTCGGCCCCCAC
CCCTTCTTCCCGGGGGCGGGCAAGGGCGATGGCGACGGCGAGGACAAGGGCGAG
AGCAAGAGCGAGATCAGGACGCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCG
CAGGGCGGCGGCGGCGGCCACCACGCCGATCCGGGACCTGCTCCGCGACTGCGA
CAAGACGTACCCGGACACGTGCCCGCTGCCGCCGCTGACGGTGCGCGTGACCAA
CGAGGGCGAGCGCGCGTCCGACTACGTGGTGCTGGCCTTCGTGTCGGGCGAGTA
CGGGCCGGCGCCGTACCCGATCAAGACGCTGGTCTCGTACGCGCGGGCGCGCGG
GATAAAGGGGAAGGGCGGCGACGGCGACGGCGACGGCGACGGCGCCACCACTA
CCGTCTCGCTCGACTGGACCGTCGGCAACCTGGCCCGCCACGACGAGCGCGGCA
ACACAATCCTGTACCCGGGAACTTACACCCTCACTCTCGACGAGCCGGCCCAGGC
GAGCGTGCAGTTCGCCCTCGAGGGCGAGCCCGTCGTGCTCGACGAGTGGCCTGC
GCCGCCGAGTGCCAACTCCACCGCCAGGGGGAGGCACAGG (SEQ ID NO:30)

MKASVSCLVGMSAVAYGLDGPFQTYPDCTKPPLSDIKVCDRTLPEAERAAALVAAL
TDEEKLQNLVSKAPGAPRIGLPAYNWWSEALHGVAHAPGTQFRDGPGDFNSATSFP
MPLLMAAAFDDELIEAVGDVIGTEARAFGNAGWSGLDYWTPNVNPFRDPRWGRGS
ETPGEDVVRLKRYAASMIRGLEGRSSSSSSCSFGSGGEPPRVIATCKHYAGNDFEDW
NGTTRHDFDAVISAQDLAEYYLAPFQQCARDSRVGSVMCAYNAVNGVPSCANSYLL
NTILRGHWNWTEHDNYVTSDCEAVLDVSAHHHYADTNAEGTGLCFEAGMDTSCEY
EGSSDIPGASAGGFLTWPAVDRALTRLYRSLVRVGYFDGPESPHASLGWADVNRPE
AQELALRAAVEGIVLLKNDNDTLPLPLPDDVVVTADGGRRRVAMIGFWADAPDKLF
GGYSGAPPFARSPASAARQLGWNVTVAGGPVLEGDSDEEEDTWTAPAVEAAADAD
YIVYFGGLDTSAAGETKDRMTIGWPAAQLALISELARLGKPVVVVQMGDQLDDTPL
FELDGVGAVLWANWPGQDGGTAVVRLLSGAESPAGRLPVTQYPANYTDAVPLTDM
TLRPSATNPGRTYRWYPTPVRPFGFGLHYTTFRAEFGPHPFFPGAGKGDGDGEDKGE
SKSEIRTQQQQQQQQQRRAAAAATTPIRDLLRDCDKTYPDTCPLPPLTVRVTNEGE
RASDYVVLAFVSGEYGPAPYPIKTLVSYARARGIKGKGGTGAGDGDVATTTVSLDW
TVGNLARHDERGNTILYPGTYTLTLDEPAQASVQFALEGEPVVLDEWPAPPSANSTA
RGRHR (SEQ ID NO:31)

Variant N219Y:

ATGAAGGCCTCTGTATCATGCCTCGTCGGCATGAGCGCCGTGGCCTACGGCCTCG
ATGGCCCTTTCCAGACCTACCCCGACTGCACCAAGCCCCCCCTGTCCGATATTAA
GGTGTGCGACCGGACACTGCCCGAGGCGGAGCGGGCGGCAGCCCTCGTGGCAGC
CCTGACCGACGAGGAGAAGCTGCAAAACCTGGTCAGCAAGGCGCCGGGGGCGCC
GCGGATCGGCCTGCCCGCGTACAACTGGTGGAGCGAGGCGCTGCACGGGGTGGC
CCACGCGCCCGGGACGCAATTCCGCGACGGGCCGGGGGACTTCAACTCGTCCAC
GTCGTTCCCGATGCCGCTGCTGATGGCCGCCGCCTTCGACGACGAGCTGATCGAG
GCCGTCGGCGACGTCATCGGCACCGAGGCCCGCGCCTTTGGCAACGCCGGCTGG
TCCGGCCTCGACTACTGGACCCCCAACGTCAACCCCTTCCGGGACCCCCGCTGGG
GCCGCGGCTCCGAGACGCCGGGCGAGGACGTCGTGCGCCTCAAGCGCTACGCCG
CCTCCATGATCCGCGGGCTCGAGGGTCGTTCCTCCTCCTCCTCCTCCTGCTCCTTC
GGATCCGGAGGGGAGCCGCCGCGCGTCATCTCGACCTGCAAGCACTACGCCGGC
TACGACTTTGAGGACTGGAACGGCACGACGCGGCACGACTTCGACGCCGTCATC
TCGGCGCAGGACCTGGCCGAGTACTACCTGGCGCCGTTCCAGCAGTGCGCGCGC
GACTCGCGCGTCGGCTCCGTCATGTGCGCCTACAACGCCGTCAACGGGGTGCCGT
CGTGCGCCAACTCGTACCTCATGAACACGATCCTGCGCGGGCACTGGAACTGGA
CCGAGCACGACAACTACGTCACCAGCGACTGCGAGGCCGTCCTCGACGTCTCGG
CCCACCACCACTACGCCGACACCAACGCCGAGGGCACCGGCCTCTGCTTCGAGG
CCGGCATGGACACGAGCTGCGAGTACGAGGGCTCCTCCGACATCCCGGGCGCCT
CCGCCGGCGGCTTCCTGACCTGGCCCGCCGTCGACCGCGCCCTGACGCGGCTGTA
CCGGAGCCTGGTGCGGGTCGGCTACTTTGACGGCCCCGAGTCGCCGCACGCCTCG
CTGGGCTGGGCCGACGTCAACCGGCCCGAGGCGCAGGAGCTGGCCCTGCGCGCT
GCCGTCGAGGGCATCGTGCTGCTCAAGAACGACAACGACACGCTGCCGCTGCCG
CTGCCGGACGATGTCGTTGTCACCGCTGATGGTGGCCGCCGCCGCGTCGCCATGA
TCGGCTTCTGGGCCGACGCCCCGGACAAGCTGTTTGGCGGGTACAGCGGCGCGC
CCCCCTTCGCGCGCTCGCCCGCGAGCGCCGCCCGGCAGCTGGGCTGGAACGTCAC
GGTCGCCGGAGGGCCCGTCCTGGAGGGAGACTCGGACGAGGAGGAGGACACGT
GGACGGCGCCGGCCGTCGAGGCGGCCGCCGACGCCGACTACATCGTCTACTTTG
GCGGCCTGGACACGTCGGCGGCGGGCGAGACCAAGGACCGGATGACGATCGGGT
GGCCGGCGGCGCAGCTGGCGCTCATCTCGGAGCTGGCGCGGCTCGGCAAGCCCG
TCGTGGTGGTGCAGATGGGCGACCAGCTCGACGACACGCCCCTCTTCGAGCTGG
ACGGGGTGGGCGCCGTCCTGTGGGCCAACTGGCCGGGCCAGGACGGCGGCACGG
CCGTGGTCCGGCTGCTCAGCGGCGCCGAGAGCCCGGCCGGCCGCCTGCCCGTGA
CCCAGTACCCGGCCAACTACACCGACGCGGTGCCCCTGACCGACATGACCCTGC
GCCCGTCGGCGACCAACCCGGGCCGGACCTACCGCTGGTACCCGACTCCCGTCCG
GCCCTTCGGCTTCGGCCTCCACTATACCACCTTCCGGGCCGAGTTCGGCCCCCAC
CCCTTCTTCCCGGGGGCGGGCAAGGGCGATGGCGACGGCGAGGACAAGGGCGAG
AGCAAGAGCGAGATCAGGACGCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCG
CAGGGCGGCGGCGGCGGCCACCACGCCGATCCGGGACCTGCTCCGCGACTGCGA
CAAGACGTACCCGGACACGTGCCCGCTGCCGCCGCTGACGGTGCGCGTGACCAA
CGAGGGCGAGCGCGCGTCCGACTACGTGGTGCTGGCCTTCGTGTCGGGCGAGTA
CGGGCCGGCGCCGTACCCGATCAAGACGCTGGTCTCGTACGCGCGGGCGCGCGG
GCTAAAGGGGAAGGGCGGCGACGGCGACGGCGACGGCGACGGCGCCACCACTA
CCGTCTCGCTCGACTGGACCGTCGGCAACCTGGCCCGCCACGACGAGCGCGGCA
ACACAATCCTGTACCCGGGAACTTACACCCTCACTCTCGACGAGCCGGCCCAGGC
GAGCGTGCAGTTCGCCCTCGAGGGCGAGCCCGTCGTGCTCGACGAGTGGCCTGC
GCCGCCGAGTGCCAACTCCACCGCCAGGGGGAGGCACAGG (SEQ ID NO:32)

MKASVSCLVGMSAVAYGLDGPFQTYPDCTKPPLSDIKVCDRTLPEAERAAALVAAL
TDEEKLQNLVSKAPGAPRIGLPAYNWWSEALHGVAHAPGTQFRDGPGDFNSSTSFP
MPLLMAAAFDDELIEAVGDVIGTEARAFGNAGWSGLDYWTPNVNPFRDPRWGRGS
ETPGEDVVRLKRYAASMIRGLEGRSSSSSSCSFGSGGEPPRVISTCKHYAGYDFEDWN
GTTRHDFDAVISAQDLAEYYLAPFQQCARDSRVGSVMCAYNAVNGVPSCANSYLM
NTILRGHWNWTEHDNYVTSDCEAVLDVSAHHHYADTNAEGTGLCFEAGMDTSCEY
EGSSDIPGASAGGFLTWPAVDRALTRLYRSLVRVGYFDGPESPHASLGWADVNRPE
AQELALRAAVEGIVLLKNDNDTLPLPLPDDVVVTADGGRRRVAMIGFWADAPDKLF
GGYSGAPPFARSPASAARQLGWNVTVAGGPVLEGDSDEEEDTWTAPAVEAAADAD
YIVYFGGLDTSAAGETKDRMTIGWPAAQLALISELARLGKPVVVVQMGDQLDDTPL
FELDGVGAVLWANWPGQDGGTAVVRLLSGAESPAGRLPVTQYPANYTDAVPLTDM
TLRPSATNPGRTYRWYPTPVRPFGFGLHYTTFRAEFGPHPFFPGAGKGDGDGEDKGE
SKSEIRTQQQQQQQQQRRAAAAATTPIRDLLRDCDKTYPDTCPLPPLTVRVTNEGE
RASDYVVLAFVSGEYGPAPYPIKTLVSYARARGLKGKGGTGAGDGDVATTTVSLD
WTVGNLARHDERGNTILYPGTYTLTLDEPAQASVQFALEGEPVVLDEWPAPPSANST
ARGRHR (SEQ ID NO:33)

Variant N219Y/N571G:

ATGAAGGCCTCTGTATCATGCCTCGTCGGCATGAGCGCCGTGGCCTACGGCCTCG
ATGGCCCTTTCCAGACCTACCCCGACTGCACCAAGCCCCCCCTGTCCGATATTAA
GGTGTGCGACCGGACACTGCCCGAGGCGGAGCGGGCGGCAGCCCTCGTGGCAGC
CCTGACCGACGAGGAGAAGCTGCAAAACCTGGTCAGCAAGGCGCCGGGGGCGCC
GCGGATCGGCCTGCCCGCGTACAACTGGTGGAGCGAGGCGCTGCACGGGGTGGC
CCACGCGCCCGGGACGCAGTTCCGCGACGGGCCGGGGGACTTCAACTCGTCCAC
GTCGTTCCCGATGCCGCTGCTGATGGCCGCCGCCTTCGACGACGAGCTGATCGAG
GCCGTCGGCGACGTCATCGGCACCGAGGCCCGCGCCTTTGGCAACGCCGGCTGG
TCCGGCCTCGACTACTGGACCCCCAACGTCAACCCCTTCCGGGACCCCCGCTGGG
GCCGCGGCTCCGAGACGCCGGGCGAGGACGTCGTGCGCCTCAAGCGCTACGCCG
CCTCCATGATCCGCGGGCTCGAGGGTCGTTCCTCCTCCTCCTCCTCCTGCTCCTTC
GGATCCGGAGGGGAGCCGCCGCGCGTCATCTCGACCTGCAAGCACTACGCCGGC
TACGACTTTGAGGACTGGAACGGCACGACGCGGCACGACTTCGACGCCGTCATC
TCGGCGCAGGACCTGGCCGAGTACTACCTGGCGCCGTTCCAGCAGTGCGCGCGC
GACTCGCGCGTCGGCTCCGTCATGTGCGCCTACAACGCCGTCAACGGGGTGCCGT
CGTGCGCCAACTCGTACCTCATGAACACGATCCTGCGCGGGCACTGGAACTGGA
CCGAGCACGACAACTACGTCACCAGCGACTGCGAGGCCGTCCTCGACGTCTCGG
CCCACCACCACTACGCCGACACCAACGCCGAGGGCACCGGCCTCTGCTTCGAGG
CCGGCATGGACACGAGCTGCGAGTACGAGGGCTCCTCCGACATCCCGGGCGCCT
CCGCCGGCGGCTTCCTGACCTGGCCCGCCGTCGACCGCGCCCTGACGCGGCTGTA
CCGGAGCCTGGTGCGGGTCGGCTACTTTGACGGCCCCGAGTCGCCGCACGCCTCG
CTGGGCTGGGCCGACGTCAACCGGCCCGAGGCGCAGGAGCTGGCCCTGCGCGCT
GCCGTCGAGGGCATCGTGCTGCTCAAGAACGACAACGACACGCTGCCGCTGCCG
CTGCCGGACGATGTCGTTGTCACCGCTGATGGTGGCCGCCGCCGCGTCGCCATGA
TCGGCTTCTGGGCCGACGCCCCGGACAAGCTGTTTGGCGGGTACAGCGGCGCGC
CCCCCTTCGCGCGCTCGCCCGCGAGCGCCGCCCGGCAGCTGGGCTGGAACGTCAC
GGTCGCCGGAGGGCCCGTCCTGGAGGGAGACTCGGACGAGGAGGAGGACACGT
GGACGGCGCCGGCCGTCGAGGCGGCCGCCGACGCCGACTACATCGTCTACTTTG
GCGGCCTGGACACGTCGGCGGCGGGCGAGACCAAGGACCGGATGACGATCGGGT
GGCCGGCGGCGCAGCTGGCGCTCATCTCGGAGCTGGCGCGGCTCGGCAAGCCCG

TCGTGGTGGTGCAGATGGGCGACCAGCTCGACGACACGCCCCTCTTCGAGCTGG
ACGGGGTGGGCGCCGTCCTGTGGGCCGGCTGGCCGGGCCAGGACGGCGGCACGG
CCGTGGTCCGGCTGCTCAGCGGCGCCGAGAGCCCGGCCGGCCGCCTGCCCGTGA
CCCAGTACCCGGCCAACTACACCGACGCGGTGCCCCTGACCGACATGACCCTGC
GCCCGTCGGCGACCAACCCGGGCCGGACCTACCGCTGGTACCCGACTCCCGTCCG
GCCCTTCGGCTTCGGCCTCCACTATACCACCTTCCGGGCCGAGTTCGGCCCCCAC
CCCTTCTTCCCGGGGGCGGGCAAGGGCGATGGCGACGGCGAGGACAAGGGCGAG
AGCAAGAGCGAGATCAGGACGCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCG
CAGGGCGGCGGCGGCGGCCACCACGCCGATCCGGGACCTGCTCCGCGACTGCGA
CAAGACGTACCCGGACACGTGCCCGCTGCCGCCGCTGACGGTGCGCGTGACCAA
CGAGGGCGAGCGCGCGTCCGACTACGTGGTGCTGGCCTTCGTGTCGGGCGAGTA
CGGGCCGGCGCCGTACCCGATCAAGACGCTGGTCTCGTACGCGCGGGCGCGCGG
GCTAAAGGGGAAGGGCGGCGACGGCGACGGCGACGGCGACGGCGCCACCACTA
CCGTCTCGCTCGACTGGACCGTCGGCAACCTGGCCCGCCACGACGAGCGCGGCA
ACACAATCCTGTACCCGGGAACTTACACCCTCACTCTCGACGAGCCGGCCCAGGC
GAGCGTGCAGTTCGCCCTCGAGGGCGAGCCCGTCGTGCTCGACGAGTGGCCTGC
GCCGCCGAGTGCCAACTCCACCGCCAGGGGGAGGCACAGG (SEQ ID NO:34)

MKASVSCLVGMSAVAYGLDGPFQTYPDCTKPPLSDIKVCDRTLPEAERAAALVAAL
TDEEKLQNLVSKAPGAPRIGLPAYNWWSEALHGVAHAPGTQFRDGPGDFNSSTSFP
MPLLMAAAFDDELIEAVGDVIGTEARAFGNAGWSGLDYWTPNVNPFRDPRWGRGS
ETPGEDVVRLKRYAASMIRGLEGRSSSSSSCSFGSGGEPPRVISTCKHYAGYDFEDWN
GTTRHDFDAVISAQDLAEYYLAPFQQCARDSRVGSVMCAYNAVNGVPSCANSYLM
NTILRGHWNWTEHDNYVTSDCEAVLDVSAHHHYADTNAEGTGLCFEAGMDTSCEY
EGSSDIPGASAGGFLTWPAVDRALTRLYRSLVRVGYFDGPESPHASLGWADVNRPE
AQELALRAAVEGIVLLKNDNDTLPLPLPDDVVVTADGGRRRVAMIGFWADAPDKLF
GGYSGAPPFARSPASAARQLGWNVTVAGGPVLEGDSDEEEDTWTAPAVEAAADAD
YIVYFGGLDTSAAGETKDRMTIGWPAAQLALISELARLGKPVVVVQMGDQLDDTPL
FELDGVGAVLWAGWPGQDGGTAVVRLLSGAESPAGRLPVTQYPANYTDAVPLTDM
TLRPSATNPGRTYRWYPTPVRPFGFGLHYTTFRAEFGPHPFFPGAGKGDGDGEDKGE
SKSEIRTQQQQQQQQQRRAAAAATTPIRDLLRDCDKTYPDTCPLPPLTVRVTNEGE
RASDYVVLAFVSGEYGPAPYPIKTLVSYARARGLKGKGGTGAGDGDVATTTVSLD
WTVGNLARHDERGNTILYPGTYTLTLDEPAQASVQFALEGEPVVLDEWPAPPSANST
ARGRHR (SEQ ID NO:35)

Variant V235I:

ATGAAGGCCTCTGTATCATGCCTCGTCGGCATGAGCGCCGTGGCCTACGGCCTCG
ATGGCCCTTTCCAGACCTACCCCGACTGCACCAAGCCCCCCCTGTCCGATATTAA
GGTGTGCGACCGGACACTGCCCGAGGCGGAGCGGGCGGCAGCCCTCGTGGCAGC
CCTGACCGACGAGGAGAAGCTGCAAAACCTGGTCAGCAAGGCGCCGGGGGCGCC
GCGGATCGGCCTGCCCGCGTACAACTGGTGGAGCGAGGCGCTGCACGGGGTGGC
CCACGCGCCCGGGACGCAGTTCCGCGACGGGCCGGGGGACTTCAACTCGTCCAC
GTCGTTCCCGATGCCGCTGCTGATGGCCGCCGCCTTCGACGACGAGCTGATCGAG
GCCGTCGGCGACGTCATCGGCACCGAGGCCCGCGCCTTTGGCAACGCCGGCTGG
TCCGGCCTCGACTACTGGACCCCCAACGTCAACCCCTTCCGGGACCCCCGCTGGG
GCCGCGGCTCCGAGACGCCGGGCGAGGACGTCGTGCGCCTCAAGCGCTACGCCG
CCTCCATGATCCGCGGGCTCGAGGGTCGTTCCTCCTCCTCCTCCTCCTGCTCCTTC

GGATCCGGAGGGGAGCCGCCGCGCGTCATCTCGACCTGCAAGCACTACGCCGGC
AACGACTTTGAGGACTGGAACGGCACGACGCGGCACGACTTCGACGCCATTATC
TCGGCGCAGGACCTGGCCGAGTACTACCTGGCGCCGTTCCAGCAGTGCGCGCGC
GACTCGCGCGTCGGCTCCGTCATGTGCGCCTACAACGCCGTCAACGGGGTGCCGT
CGTGCGCCAACTCGTACCTCATGAACACGATCCTGCGCGGGCACTGGAACTGGA
CCGAGCACGACAACTACGTCACCAGCGACTGCGAGGCCGTCCTCGACGTCTCGG
CCCACCACCACTACGCCGACACCAACGCCGAGGGCACCGGCCTCTGCTTCGAGG
CCGGCATGGACACGAGCTGCGAGTACGAGGGCTCCTCCGACATCCCGGGCGCCT
CCGCCGGCGGCTTCCTGACCTGGCCCGCCGTCGACCGCGCCCTGACGCGGCTGTA
CCGGAGCCTGGTGCGGGTCGGCTACTTTGACGGCCCCGAGTCGCCGCACGCCTCG
CTGGGCTGGGCCGACGTCAACCGGCCCGAGGCGCAGGAGCTGGCCCTGCGCGCT
GCCGTCGAGGGCATCGTGCTGCTCAAGAACGACAACGACACGCTGCCGCTGCCG
CTGCCGGACGATGTCGTTGTCACCGCTGATGGTGGCCGCCGCCGCGTCGCCATGA
TCGGCTTCTGGGCCGACGCCCCGGACAAGCTGTTTGGCGGGTACAGCGGCGCGC
CCCCCTTCGCGCGCTCGCCCGCGAGCGCCGCCCGGCAGCTGGGCTGGAACGTCAC
GGTCGCCGGAGGGCCCGTCCTGGAGGGAGACTCGGACGAGGAGGAGGACACGT
GGACGGCGCCGGCCGTCGAGGCGGCCGCCGACGCCGACTACATCGTCTACTTTG
GCGGCCTGGACACGTCGGCGGCGGGCGAGACCAAGGACCGGATGACGATCGGGT
GGCCGGCGGCGCAGCTGGCGCTCATCTCGGAGCTGGCGCGGCTCGGCAAGCCCG
TCGTGGTGGTGCAGATGGGCGACCAGCTCGACGACACGCCCCTCTTCGAGCTGG
ACGGGGTGGGCGCCGTCCTGTGGGCCAACTGGCCGGGCCAGGACGGCGGCACGG
CCGTGGTCCGGCTGCTCAGCGGCGCCGAGAGCCCGGCCGGCCGCCTGCCCGTGA
CCCAGTACCCGGCCAACTACACCGACGCGGTGCCCCTGACCGACATGACCCTGC
GCCCGTCGGCGACCAACCCGGGCCGGACCTACCGCTGGTACCCGACTCCCGTCCG
GCCCTTCGGCTTCGGCCTCCACTATACCACCTTCCGGGCCGAGTTCGGCCCCCAC
CCCTTCTTCCCGGGGGCGGGCAAGGGCGATGGCGACGGCGAGGACAAGGGCGAG
AGCAAGAGCGAGATCAGGACGCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCG
CAGGGCGGCGGCGGCGGCCACCACGCCGATCCGGGACCTGCTCCGCGACTGCGA
CAAGACGTACCCGGACACGTGCCCGCTGCCGCCGCTGACGGTGCGCGTGACCAA
CGAGGGCGAGCGCGCGTCCGACTACGTGGTGCTGGCCTTCGTGTCGGGCGAGTA
CGGGCCGGCGCCGTACCCGATCAAGACGCTGGTCTCGTACGCGCGGGCGCGCGG
GCTAAAGGGGAAGGGCGGCGACGGCGACGGCGACGGCGACGGCGCCACCACTA
CCGTCTCGCTCGACTGGACCGTCGGCAACCTGGCCCGCCACGACGAGCGCGGCA
ACACAATCCTGTACCCGGGAACTTACACCCTCACTCTCGACGAGCCGGCCCAGGC
GAGCGTGCAGTTCGCCCTCGAGGGCGAGCCCGTCGTGCTCGACGAGTGGCCTGC
GCCGCCGAGTGCCAACTCCACCGCCAGGGGGAGGCACAGG (SEQ ID NO:36)

MKASVSCLVGMSAVAYGLDGPFQTYPDCTKPPLSDIKVCDRTLPEAERAAALVAAL
TDEEKLQNLVSKAPGAPRIGLPAYNWWSEALHGVAHAPGTQFRDGPGDFNSSTSFP
MPLLMAAAFDDELIEAVGDVIGTEARAFGNAGWSGLDYWTPNVNPFRDPRWGRGS
ETPGEDVVRLKRYAASMIRGLEGRSSSSSSCSFGSGGEPPRVISTCKHYAGNDFEDWN
GTTRHDFDAIISAQDLAEYYLAPFQQCARDSRVGSVMCAYNAVNGVPSCANSYLMN
TILRGHWNWTEHDNYVTSDCEAVLDVSAHHHYADTNAEGTGLCFEAGMDTSCEYE
GSSDIPGASAGGFLTWPAVDRALTRLYRSLVRVGYFDGPESPHASLGWADVNRPEA
QELALRAAVEGIVLLKNDNDTLPLPLPDDVVVTADGGRRRVAMIGFWADAPDKLFG
GYSGAPPFARSPASAARQLGWNVTVAGGPVLEGDSDEEEDTWTAPAVEAAADADYI
VYFGGLDTSAAGETKDRMTIGWPAAQLALISELARLGKPVVVVQMGDQLDDTPLFE
LDGVGAVLWANWPGQDGGTAVVRLLSGAESPAGRLPVTQYPANYTDAVPLTDMTL
RPSATNPGRTYRWYPTPVRPFGFGLHYTTFRAEFGPHPFFPGAGKGDGDGEDKGESK
SEIRTQQQQQQQQQRRAAAAATTPIRDLLRDCDKTYPDTCPLPPLTVRVTNEGERA

SDYVVLAFVSGEYGPAPYPIKTLVSYARARGLKGKGGTGAGDGDVATTTVSLDWTV
GNLARHDERGNTILYPGTYTLTLDEPAQASVQFALEGEPVVLDEWPAPPSANSTARG
RHR (SEQ ID NO:37)

Variant V235L:

ATGAAAGCGAGCGTGAGCTGCCTGGTGGGCATGAGCGCGGTGGCGTATGGCCTG
GATGGCCCGTTTCAGACCTATCCGGATTGCACCAAACCGCCGCTGAGCGATATTA
AAGTGTGCGATCGCACCCTGCCGGAAGCGGAACGCGCGGCGGCGCTGGTGGCGG
CGCTGACCGATGAAGAAAAACTGCAGAACCTGGTGAGCAAAGCGCCGGGCGCGC
CGCGCATTGGCCTGCCGGCGTATAACTGGTGGAGCGAAGCGCTGCATGGCGTGG
CGCATGCGCCGGGCACCCAGTTTCGCGATGGCCCGGGCGATTTTAACAGCAGCA
CCAGCTTTCCGATGCCGCTGCTGATGGCGGCGGCGTTTGATGATGAACTGATTGA
AGCGGTGGGCGATGTGATTGGCACCGAAGCGCGCGCGTTTGGCAACGCGGGCTG
GAGCGGCCTGGATTATTGGACCCCGAACGTGAACCCGTTTCGCGATCCGCGCTGG
GGCCGCGGCAGCGAAACCCCGGGCGAAGATGTGGTGCGCCTGAAACGCTATGCG
GCGAGCATGATTCGCGGCCTGGAAGGCCGCAGCAGCAGCAGCAGCAGCTGCAGC
TTTGGCAGCGGCGGCGAACCGCCGCGCGTGATTAGCACCTGCAAACATTATGCG
GGCAACGATTTTGAAGATTGGAACGGCACCACCCGCCATGATTTTGATGCGCTGA
TTAGCGCGCAGGATCTGGCGGAATATTATCTGGCGCCGTTTCAGCAGTGCGCGCG
CGATAGCCGCGTGGGCAGCGTGATGTGCGCGTATAACGCGGTGAACGGCGTGCC
GAGCTGCGCGAACAGCTATCTGATGAACACCATTCTGCGCGGCCATTGGAACTG
GACCGAACATGATAACTATGTGACCAGCGATTGCGAAGCGGTGCTGGATGTGAG
CGCGCATCATCATTATGCGGATACCAACGCGGAAGGCACCGGCCTGTGCTTTGAA
GCGGGCATGGATACCAGCTGCGAATATGAAGGCAGCAGCGATATTCCGGGCGCG
AGCGCGGGCGGCTTTCTGACCTGGCCGGCGGTGGATCGCGCGCTGACCCGCCTGT
ATCGCAGCCTGGTGCGCGTGGGCTATTTTGATGGCCCGGAAAGCCCGCATGCGA
GCCTGGGCTGGGCGGATGTGAACCGCCCGGAAGCGCAGGAACTGGCGCTGCGCG
CGGCGGTGGAAGGCATTGTGCTGCTGAAAAACGATAACGATACCCTGCCGCTGC
CGCTGCCGGATGATGTGGTGGTGACCGCGGATGGCGGCCGCCGCCGCGTGGCGA
TGATTGGCTTTTGGGCGGATGCGCCGGATAAACTGTTTGGCGGCTATAGCGGCGC
GCCGCCGTTTGCGCGCAGCCCGGCGAGCGCGGCGCGCCAGCTGGGCTGGAACGT
GACCGTGGCGGGCGGCCCGGTGCTGGAAGGCGATAGCGATGAAGAAGAAGATA
CCTGGACCGCGCCGGCGGTGGAAGCGGCGGCGGATGCGGATTATATTGTGTATTT
TGGCGGCCTGGATACCAGCGCGGCGGGCGAAACCAAAGATCGCATGACCATTGG
CTGGCCGGCGGCGCAGCTGGCGCTGATTAGCGAACTGGCGCGCCTGGGCAAACC
GGTGGTGGTGGTGCAGATGGGCGATCAGCTGGATGATACCCCGCTGTTTGAACTG
GATGGCGTGGGCGCGGTGCTGTGGGCGAACTGGCCGGGCCAGGATGGCGGCACC
GCGGTGGTGCGCCTGCTGAGCGGCGCGGAAAGCCCGGCGGGCCGCCTGCCGGTG
ACCCAGTATCCGGCGAACTATACCGATGCGGTGCCGCTGACCGATATGACCCTGC
GCCCGAGCGCGACCAACCCGGGCCGCACCTATCGCTGGTATCCGACCCCGGTGC
GCCCGTTTGGCTTTGGCCTGCATTATACCACCTTTCGCGCGGAATTTGGCCCGCAT
CCGTTTTTTCCGGGCGCGGGCAAAGGCGATGGCGATGGCGAAGATAAAGGCGAA
AGCAAAAGCGAAATTCGCACCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCG
CCGCGCGGCGGCGGCGGCGACCACCCCGATTCGCGATCTGCTGCGCGATTGCGA
TAAAACCTATCCGGATACCTGCCCGCTGCCGCCGCTGACCGTGCGCGTGACCAAC
GAAGGCGAACGCGCGAGCGATTATGTGGTGCTGGCGTTTGTGAGCGGCGAATAT
GGCCCGGCGCCGTATCCGATTAAAACCCTGGTGAGCTATGCGCGCGCGCGCGGC

CTGAAAGGCAAAGGCGGCACCGGCGCGGGCGATGGCGATGTGGCGACCACCACC
GTGAGCCTGGATTGGACCGTGGGCAACCTGGCGCGCCATGATGAACGCGGCAAC
ACCATTCTGTATCCGGGCACCTATACCCTGACCCTGGATGAACCGGCGCAGGCGA
GCGTGCAGTTTGCGCTGGAAGGCGAACCGGTGGTGCTGGATGAATGGCCGGCGC
CGCCGAGCGCGAACAGCACCGCGCGCGGCCGCCATCGC (SEQ ID NO:38)

MKASVSCLVGMSAVAYGLDGPFQTYPDCTKPPLSDIKVCDRTLPEAERAAALVAAL
TDEEKLQNLVSKAPGAPRIGLPAYNWWSEALHGVAHAPGTQFRDGPGDFNSSTSFP
MPLLMAAAFDDELIEAVGDVIGTEARAFGNAGWSGLDYWTPNVNPFRDPRWGRGS
ETPGEDVVRLKRYAASMIRGLEGRSSSSSSCSFGSGGEPPRVISTCKHYAGNDFEDWN
GTTRHDFDALISAQDLAEYYLAPFQQCARDSRVGSVMCAYNAVNGVPSCANSYLM
NTILRGHWNWTEHDNYVTSDCEAVLDVSAHHHYADTNAEGTGLCFEAGMDTSCEY
EGSSDIPGASAGGFLTWPAVDRALTRLYRSLVRVGYFDGPESPHASLGWADVNRPE
AQELALRAAVEGIVLLKNDNDTLPLPLPDDVVVTADGGRRRVAMIGFWADAPDKLF
GGYSGAPPFARSPASAARQLGWNVTVAGGPVLEGDSDEEEDTWTAPAVEAAADAD
YIVYFGGLDTSAAGETKDRMTIGWPAAQLALISELARLGKPVVVVQMGDQLDDTPL
FELDGVGAVLWANWPGQDGGTAVVRLLSGAESPAGRLPVTQYPANYTDAVPLTDM
TLRPSATNPGRTYRWYPTPVRPFGFGLHYTTFRAEFGPHPFFPGAGKGDGDGEDKGE
SKSEIRTQQQQQQQQQRRAAAAATTPIRDLLRDCDKTYPDTCPLPPLTVRVTNEGE
RASDYVVLAFVSGEYGPAPYPIKTLVSYARARGLKGKGGTGAGDGDVATTTVSLD
WTVGNLARHDERGNTILYPGTYTLTLDEPAQASVQFALEGEPVVLDEWPAPPSANST
ARGRHR (SEQ ID NO:39)

Variant S345L:

ATGAAGGCCTCTGTATCATGCCTCGTCGGCATGAGCGCCGTGGCCTACGGCCTCG
ATGGCCCTTTCCAGACCTACCCCGACTGCACCAAGCCCCCCCTGTCCGATATTAA
GGTGTGCGACCGGACACTGCCCGAGGCGGAGCGGGCGGCAGCCCTCGTGGCAGC
CCTGACCGACGAGGAGAAGCTGCAAAACCTGGTCAGCAAGGCGCCGGGGGCGCC
GCGGATCGGCCTGCCCGCGTACAACTGGTGGAGCGAGGCGCTGCACGGGGTGGC
CCACGCGCCCGGGACGCAGTTCCGCGACGGGCCGGGGGACTTCAACTCGTCCAC
GTCGTTCCCGATGCCGCTGCTGATGGCCGCCGCCTTCGACGACGAGCTGATCGAG
GCCGTCGGCGACGTCATCGGCACCGAGGCCCGCGCCTTTGGCAACGCCGGCTGG
TCCGGCCTCGACTACTGGACCCCCAACGTCAACCCCTTCCGGGACCCCCGCTGGG
GCCGCGGCTCCGAGACGCCGGGCGAGGACGTCGTGCGCCTCAAGCGCTACGCCG
CCTCCATGATCCGCGGGCTCGAGGGTCGTTCCTCCTCCTCCTCCTCCTGCTCCTTC
GGATCCGGAGGGGAGCCGCCGCGCGTCATCTCGACCTGCAAGCACTACGCCGGC
AACGACTTTGAGGACTGGAACGGCACGACGCGGCACGACTTCGACGCCGTCATC
TCGGCGCAGGACCTGGCCGAGTACTACCTGGCGCCGTTCCAGCAGTGCGCGCGC
GACTCGCGCGTCGGCTCCGTCATGTGCGCCTACAACGCCGTCAACGGGGTGCCGT
CGTGCGCCAACTCGTACCTCATGAACACGATCCTGCGCGGGCACTGGAACTGGA
CCGAGCACGACAACTACGTCACCAGCGACTGCGAGGCCGTCCTCGACGTCTCGG
CCCACCACCACTACGCCGACACCAACGCCGAGGGCACCGGCCTCTGCTTCGAGG
CCGGCATGGACACGAGCTGCGAGTACGAGGGCTCCTCCGACATCCCGGGCGCCT
TGGCCGGCGGCTTCCTGACCTGGCCCGCCGTCGACCGCGCCCTGACGCGGCTGTA
CCGGAGCCTGGTGCGGGTCGGCTACTTTGACGGCCCCGAGTCGCCGCACGCCTCG
CTGGGCTGGGCCGACGTCAACCGGCCCGAGGCACAGGAGCTGGCCCTGCGCGCT
GCCGTCGAGGGCATCGTGCTGCTCAAGAACGACAACGACACGCTGCCGCTGCCG

CTGCCGGACGATGTCGTTGTCACCGCTGATGGTGGCCGCCGCCGCGTCGCCATGA
TCGGCTTCTGGGCCGACGCCCCGGACAAGCTGTTTGGCGGGTACAGCGGCGCGC
CCCCCTTCGCGCGCTCGCCCGCGAGCGCCGCCCGGCAGCTGGGCTGGAACGTCAC
GGTCGCCGGAGGGCCCGTCCTGGAGGGAGACTCGGACGAGGAGGAGGACACGT
GGACGGCGCCGGCCGTCGAGGCGGCCGCCGACGCCGACTACATCGTCTACTTTG
GCGGCCTGGACACGTCGGCGGCGGGCGAGACCAAGGACCGGATGACGATCGGGT
GGCCGGCGGCGCAGCTGGCGCTCATCTCGGAGCTGGCGCGGCTCGGCAAGCCCG
TCGTGGTGGTGCAGATGGGCGACCAGCTCGACGACACGCCCCTCTTCGAGCTGG
ACGGGGTGGGCGCCGTCCTGTGGGCCAACTGGCCGGGCCAGGACGGCGGCACGG
CCGTGGTCCGGCTGCTCAGCGGCGCCGAGAGCCCGGCCGGCCGCCTGCCCGTGA
CCCAGTACCCGGCCAACTACACCGACGCGGTGCCCCTGACCGACATGACCCTGC
GCCCGTCGGCGACCAACCCGGGCCGGACCTACCGCTGGTACCCGACTCCCGTCCG
GCCCTTCGGCTTCGGCCTCCACTATACCACCTTCCGGGCCGAGTTCGGCCCCCAC
CCCTTCTTCCCGGGGGCGGGCAAGGGCGATGGCGACGGCGAGGACAAGGGCGAG
AGCAAGAGCGAGATCAGGACGCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCG
CAGGGCGGCGGCGGCGGCCACCACGCCGATCCGGGACCTGCTCCGCGACTGCGA
CAAGACGTACCCGGACACGTGCCCGCTGCCGCCGCTGACGGTGCGCGTGACCAA
CGAGGGCGAGCGCGCGTCCGACTACGTGGTGCTGGCCTTCGTGTCGGGCGAGTA
CGGGCCGGCGCCGTACCCGATCAAGACGCTGGTCTCGTACGCGCGGGCGCGCGG
GCTAAAGGGGAAGGGCGGCGACGGCGACGGCGACGGCGACGGCGCCACCACTA
CCGTCTCGCTCGACTGGACCGTCGGCAACCTGGCCCGCCACGACGAGCGCGGCA
ACACAATCCTGTACCCGGGAACTTACACCCTCACTCTCGACGAGCCGGCCCAGGC
GAGCGTGCAGTTCGCCCTCGAGGGCGAGCCCGTCGTGCTCGACGAGTGGCCTGC
GCCGCCGAGTGCCAACTCCACCGCCAGGGGGAGGCACAGG (SEQ ID NO:40)

MKASVSCLVGMSAVAYGLDGPFQTYPDCTKPPLSDIKVCDRTLPEAERAAALVAAL
TDEEKLQNLVSKAPGAPRIGLPAYNWWSEALHGVAHAPGTQFRDGPGDFNSSTSFP
MPLLMAAAFDDELIEAVGDVIGTEARAFGNAGWSGLDYWTPNVNPFRDPRWGRGS
ETPGEDVVRLKRYAASMIRGLEGRSSSSSSCSFGSGGEPPRVISTCKHYAGNDFEDWN
GTTRHDFDAVISAQDLAEYYLAPFQQCARDSRVGSVMCAYNAVNGVPSCANSYLM
NTILRGHWNWTEHDNYVTSDCEAVLDVSAHHHYADTNAEGTGLCFEAGMDTSCEY
EGSSDIPGALAGGFLTWPAVDRALTRLYRSLVRVGYFDGPESPHASLGWADVNRPE
AQELALRAAVEGIVLLKNDNDTLPLPLPDDVVVTADGGRRRVAMIGFWADAPDKLF
GGYSGAPPFARSPASAARQLGWNVTVAGGPVLEGDSDEEEDTWTAPAVEAAADAD
YIVYFGGLDTSAAGETKDRMTIGWPAAQLALISELARLGKPVVVVQMGDQLDDTPL
FELDGVGAVLWANWPGQDGGTAVVRLLSGAESPAGRLPVTQYPANYTDAVPLTDM
TLRPSATNPGRTYRWYPTPVRPFGFGLHYTTFRAEFGPHPFFPGAGKGDGDGEDKGE
SKSEIRTQQQQQQQQQRRAAAAATTPIRDLLRDCDKTYPDTCPLPPLTVRVTNEGE
RASDYVVLAFVSGEYGPAPYPIKTLVSYARARGLKGKGGTGAGDGDVATTTVSLD
WTVGNLARHDERGNTILYPGTYTLTLDEPAQASVQFALEGEPVVLDEWPAPPSANST
ARGRHR (SEQ ID NO:41)

Variant G347Q:

ATGAAGGCCTCTGTATCATGCCTCGTCGGCATGAGCGCCGTGGCCTACGGCCTCG
ATGGCCCTTTCCAGACCTACCCCGACTGCACCAAGCCCCCCCTGTCCGATATTAA
GGTGTGCGACCGGACACTGCCCGAGGCGGAGCGGGCGGCAGCCCTCGTGGCAGC
CCTGACCGACGAGGAGAAGCTGCAAAACCTGGTCAGCAAGGCGCCGGGGGCGCC

GCGGATCGGCCTGCCCGCGTACAACTGGTGGAGCGAGGCGCTGCACGGGGTGGC
CCACGCGCCCGGGACGCAGTTCCGCGACGGGCCGGGGGACTTCAACTCGTCCAC
GTCGTTCCCGATGCCGCTGCTGATGGCCGCCGCCTTCGACGACGAGCTGATCGAG
GCCGTCGGCGACGTCATCGGCACCGAGGCCCGCGCCTTTGGCAACGCCGGCTGG
TCCGGCCTCGACTACTGGACCCCCAACGTCAACCCCTTCCGGGACCCCCGCTGGG
GCCGCGGCTCCGAGACGCCGGGCGAGGACGTCGTGCGCCTCAAGCGCTACGCCG
CCTCCATGATCCGCGGGCTCGAGGGTCGTTCCTCCTCCTCCTCCTCCTGCTCCTTC
GGATCCGGAGGGGAGCCGCCGCGCGTCATCTCGACCTGCAAGCACTACGCCGGC
AACGACTTTGAGGACTGGAACGGCACGACGCGGCACGACTTCGACGCCGTCATC
TCGGCGCAGGACCTGGCCGAGTACTACCTGGCGCCGTTCCAGCAGTGCGCGCGC
GACTCGCGCGTCGGCTCCGTCATGTGCGCCTACAACGCCGTCAACGGGGTGCCGT
CGTGCGCCAACTCGTACCTCATGAACACGATCCTGCGCGGGCACTGGAACTGGA
CCGAGCACGACAACTACGTCACCAGCGACTGCGAGGCCGTCCTCGACGTCTCGG
CCCACCACCACTACGCCGACACCAACGCCGAGGGCACCGGCCTCTGCTTCGAGG
CCGGCATGGACACGAGCTGCGAGTACGAGGGCTCCTCCGACATCCCGGGCGCCT
CCGCCCAGGGCTTCCTGACCTGGCCCGCCGTCGACCGCGCCCTGACGCGGCTGTA
CCGGAGCCTGGTGCGGGTCGGCTACTTTGACGGCCCCGAGTCGCCGCACGCCTCG
CTGGGCTGGGCCGACGTCAACCGGCCCGAGGCGCAGGAGCTGGCCCTGCGCGCT
GCCGTCGAGGGCATCGTGCTGCTCAAGAACGACAACGACACGCTGCCGCTGCCG
CTGCCGGACGATGTCGTTGTCACCGCTGATGGTGGCCGCCGCCGCGTCGCCATGA
TCGGCTTCTGGGCCGACGCCCCGGACAAGCTGTTTGGCGGGTACAGCGGCGCGC
CCCCCTTCGCGCGCTCGCCCGCGAGCGCCGCCCGGCAGCTGGGCTGGAACGTCAC
GGTCGCCGGAGGGCCCGTCCTGGAGGGAGACTCGGACGAGGAGGAGGACACGT
GGACGGCGCCGGCCGTCGAGGCGGCCGCCGACGCCGACTACATCGTCTACTTTG
GCGGCCTGGACACGTCGGCGGCGGGCGAGACCAAGGACCGGATGACGATCGGGT
GGCCGGCGGCGCAGCTGGCGCTCATCTCGGAGCTGGCGCGGCTCGGCAAGCCCG
TCGTGGTGGTGCAGATGGGCGACCAGCTCGACGACACGCCCCTCTTCGAGCTGG
ACGGGGTGGGCGCCGTCCTGTGGGCCAACTGGCCGGGCCAGGACGGCGGCACGG
CCGTGGTCCGGCTGCTCAGCGGCGCCGAGAGCCCGGCCGGCCGCCTGCCCGTGA
CCCAGTACCCGGCCAACTACACCGACGCGGTGCCCCTGACCGACATGACCCTGC
GCCCGTCGGCGACCAACCCGGGCCGGACCTACCGCTGGTACCCGACTCCCGTCCG
GCCCTTCGGCTTCGGCCTCCACTATACCACCTTCCGGGCCGAGTTCGGCCCCCAC
CCCTTCTTCCCGGGGGCGGGCAAGGGCGATGGCGACGGCGAGGACAAGGGCGAG
AGCAAGAGCGAGATCAGGACGCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCG
CAGGGCGGCGGCGGCGGCCACCACGCCGATCCGGGACCTGCTCCGCGACTGCGA
CAAGACGTACCCGGACACGTGCCCGCTGCCGCCGCTGACGGTGCGCGTGACCAA
CGAGGGCGAGCGCGCGTCCGACTACGTGGTGCTGGCCTTCGTGTCGGGCGAGTA
CGGGCCGGCGCCGTACCCGATCAAGACGCTGGTCTCGTACGCGCGGGCGCGCGG
GCTAAAGGGGAAGGGCGGCGACGGCGACGGCGACGGCGACGGCGCCACCACTA
CCGTCTCGCTCGACTGGACCGTCGGCAACCTGGCCCGCCACGACGAGCGCGGCA
ACACAATCCTGTACCCGGGAACTTACACCCTCACTCTCGACGAGCCGGCCCAGGC
GAGCGTGCAGTTCGCCCTCGAGGGCGAGCCCGTCGTGCTCGACGAGTGGCCTGC
GCCGCCGAGTGCCAACTCCACCGCCAGGGGGAGGCACAGG (SEQ ID NO:42)

MKASVSCLVGMSAVAYGLDGPFQTYPDCTKPPLSDIKVCDRTLPEAERAAALVAAL
TDEEKLQNLVSKAPGAPRIGLPAYNWWSEALHGVAHAPGTQFRDGPGDFNSSTSFP
MPLLMAAAFDDELIEAVGDVIGTEARAFGNAGWSGLDYWTPNVNPFRDPRWGRGS
ETPGEDVVRLKRYAASMIRGLEGRSSSSSSCSFGSGGEPPRVISTCKHYAGNDFEDWN
GTTRHDFDAVISAQDLAEYYLAPFQQCARDSRVGSVMCAYNAVNGVPSCANSYLM
NTILRGHWNWTEHDNYVTSDCEAVLDVSAHHHYADTNAEGTGLCFEAGMDTSCEY

EGSSDIPGASAQGFLTWPAVDRALTRLYRSLVRVGYFDGPESPHASLGWADVNRPE
AQELALRAAVEGIVLLKNDNDTLPLPLPDDVVVTADGGRRRVAMIGFWADAPDKLF
GGYSGAPPFARSPASAARQLGWNVTVAGGPVLEGDSDEEEDTWTAPAVEAAADAD
YIVYFGGLDTSAAGETKDRMTIGWPAAQLALISELARLGKPVVVVQMGDQLDDTPL
FELDGVGAVLWANWPGQDGGTAVVRLLSGAESPAGRLPVTQYPANYTDAVPLTDM
TLRPSATNPGRTYRWYPTPVRPFGFGLHYTTFRAEFGPHPFFPGAGKGDGDGEDKGE
SKSEIRTQQQQQQQQQRRAAAAATTPIRDLLRDCDKTYPDTCPLPPLTVRVTNEGE
RASDYVVLAFVSGEYGPAPYPIKTLVSYARARGLKGKGGTGAGDGDVATTTVSLD
WTVGNLARHDERGNTILYPGTYTLTLDEPAQASVQFALEGEPVVLDEWPAPPSANST
ARGRHR (SEQ ID NO:43)

Variant G347Q/G449N:

ATGAAAGCGAGCGTGAGCTGCCTGGTGGGCATGAGCGCGGTGGCGTATGGCCTG
GATGGC
CCGTTTCAGACCTATCCGGATTGCACCAAACCGCCGCTGAGCGATATTAAAGTGT
GCGAT
CGCACCCTGCCGGAAGCGGAACGCGCGGCGGCGCTGGTGGCGGCGCTGACCGAT
GAAGAA
AAACTGCAGAACCTGGTGAGCAAAGCGCCGGGCGCGCCGCGCATTGGCCTGCCG
GCGTAT
AACTGGTGGAGCGAAGCGCTGCATGGCGTGGCGCATGCGCCGGGCACCCAGTTT
CGCGAT
GGCCCGGGCGATTTTAACAGCAGCACCAGCTTTCCGATGCCGCTGCTGATGGCGG
CGGCG
TTTGATGATGAACTGATTGAAGCGGTGGGCGATGTGATTGGCACCGAAGCGCGC
GCGTTT
GGCAACGCGGGCTGGAGCGGCCTGGATTATTGGACCCCGAACGTGAACCCGTTT
CGCGAT
CCGCGCTGGGGCCGCGGCAGCGAAACCCCGGGCGAAGATGTGGTGCGCCTGAAA
CGCTAT
GCGGCGAGCATGATTCGCGGCCTGGAAGGCCGCAGCAGCAGCAGCAGCAGCTGC
AGCTTT
GGCAGCGGCGGCGAACCGCCGCGCGTGATTAGCACCTGCAAACATTATGCGGGC
AACGAT
TTTGAAGATTGGAACGGCACCACCCGCCATGATTTTGATGCGGTGATTAGCGCGC
AGGAT
CTGGCGGAATATTATCTGGCGCCGTTTCAGCAGTGCGCGCGCGATAGCCGCGTGG
GCAGC
GTGATGTGCGCGTATAACGCGGTGAACGGCGTGCCGAGCTGCGCGAACAGCTAT
CTGATG
AACACCATTCTGCGCGGCCATTGGAACTGGACCGAACATGATAACTATGTGACC
AGCGAT
TGCGAAGCGGTGCTGGATGTGAGCGCGCATCATCATTATGCGGATACCAACGCG
GAAGGC
ACCGGCCTGTGCTTTGAAGCGGGCATGGATACCAGCTGCGAATATGAAGGCAGC
AGCGAT

ATTCCGGGCGCGAGCGCGCAGGGCTTTCTGACCTGGCCGGCGGTGGATCGCGCG
CTGACC
CGCCTGTATCGCAGCCTGGTGCGCGTGGGCTATTTTGATGGCCCGGAAAGCCCGC
ATGCG
AGCCTGGGCTGGGCGGATGTGAACCGCCCGGAAGCGCAGGAACTGGCGCTGCGC
GCGGCG
GTGGAAGGCATTGTGCTGCTGAAAAACGATAACGATACCCTGCCGCTGCCGCTG
CCGGAT
GATGTGGTGGTGACCGCGGATGGCGGCCGCCGCCGCGTGGCGATGATTGGCTTTT
GGGCG
GATGCGCCGGATAAACTGTTTGGCAACTATAGCGGCGCGCCGCCGTTTGCGCGCA
GCCCG
GCGAGCGCGGCGCGCCAGCTGGGCTGGAACGTGACCGTGGCGGGCGGCCCGGTG
CTGGAA
GGCGATAGCGATGAAGAAGAAGATACCTGGACCGCGCCGGCGGTGGAAGCGGC
GGCGGAT
GCGGATTATATTGTGTATTTTGGCGGCCTGGATACCAGCGCGGCGGGCGAAACCA
AAGAT
CGCATGACCATTGGCTGGCCGGCGGCGCAGCTGGCGCTGATTAGCGAACTGGCG
CGCCTG
GGCAAACCGGTGGTGGTGGTGCAGATGGGCGATCAGCTGGATGATACCCCGCTG
TTTGAA
CTGGATGGCGTGGGCGCGGTGCTGTGGGCGAACTGGCCGGGCCAGGATGGCGGC
ACCGCG
GTGGTGCGCCTGCTGAGCGGCGCGGAAAGCCCGGCGGGCCGCCTGCCGGTGACC
CAGTAT
CCGGCGAACTATACCGATGCGGTGCCGCTGACCGATATGACCCTGCGCCCGAGC
GCGACC
AACCCGGGCCGCACCTATCGCTGGTATCCGACCCCGGTGCGCCCGTTTGGCTTTG
GCCTG
CATTATACCACCTTTCGCGCGGAATTTGGCCCGCATCCGTTTTTTCCGGGCGCGG
GCAAA
GGCGATGGCGATGGCGAAGATAAAGGCGAAAGCAAAAGCGAAATTCGCACCCA
GCAGCAG
CAGCAGCAGCAGCAGCAGCGCCGCGCGGCGGCGGCGGCGACCACCCCGATT
CGCGAT
CTGCTGCGCGATTGCGATAAAACCTATCCGGATACCTGCCCGCTGCCGCCGCTGA
CCGTG
CGCGTGACCAACGAAGGCGAACGCGCGAGCGATTATGTGGTGCTGGCGTTTGTG
AGCGGC
GAATATGGCCCGGCGCCGTATCCGATTAAAACCCTGGTGAGCTATGCGCGCGCG
CGCGGC
CTGAAAGGCAAAGGCGGCACCGGCGCGGGCGATGGCGATGTGGCGACCACCACC
GTGAGC
CTGGATTGGACCGTGGGCAACCTGGCGCGCCATGATGAACGCGGCAACACCATT
CTGTAT
CCGGGCACCTATACCCTGACCCTGGATGAACCGGCGCAGGCGAGCGTGCAGTTT
GCGCTG
GAAGGCGAACCGGTGGTGCTGGATGAATGGCCGGCGCCGCCGAGCGCGAACAGC
ACCGCG

CGCGGCCGCCATCGC (SEQ ID NO:44)

MKASVSCLVGMSAVAYGLDGPFQTYPDCTKPPLSDIKVCDRTLPEAERAAALVAAL
TDEEKLQNLVSKAPGAPRIGLPAYNWWSEALHGVAHAPGTQFRDGPGDFNSSTSFP
MPLLMAAAFDDELIEAVGDVIGTEARAFGNAGWSGLDYWTPNVNPFRDPRWGRGS
ETPGEDVVRLKRYAASMIRGLEGRSSSSSSCSFGSGGEPPRVISTCKHYAGNDFEDWN
GTTRHDFDAVISAQDLAEYYLAPFQQCARDSRVGSVMCAYNAVNGVPSCANSYLM
NTILRGHWNWTEHDNYVTSDCEAVLDVSAHHHYADTNAEGTGLCFEAGMDTSCEY
EGSSDIPGASAQGFLTWPAVDRALTRLYRSLVRVGYFDGPESPHASLGWADVNRPE
AQELALRAAVEGIVLLKNDNDTLPLPLPDDVVVTADGGRRRVAMIGFWADAPDKLF
GNYSGAPPFARSPASAARQLGWNVTVAGGPVLEGDSDEEEDTWTAPAVEAAADAD
YIVYFGGLDTSAAGETKDRMTIGWPAAQLALISELARLGKPVVVVQMGDQLDDTPL
FELDGVGAVLWANWPGQDGGTAVVRLLSGAESPAGRLPVTQYPANYTDAVPLTDM
TLRPSATNPGRTYRWYPTPVRPFGFGLHYTTFRAEFGPHPFFPGAGKGDGDGEDKGE
SKSEIRTQQQQQQQQQRRAAAAATTPIRDLLRDCDKTYPDTCPLPPLTVRVTNEGE
RASDYVVLAFVSGEYGPAPYPIKTLVSYARARGLKGKGGTGAGDGDVATTTVSLD
WTVGNLARHDERGNTILYPGTYTLTLDEPAQASVQFALEGEPVVLDEWPAPPSANST
ARGRHR (SEQ ID NO:45)

Variant G347Q/G763P:

ATGAAGGCCTCTGTATCATGCCTCGTCGGCATGAGCGCCGTGGCCTACGGCCTCG
ATGGCCCTTTCCAGACCTACCCCGACTGCACCAAGCCCCCCCTGTCCGATATTAA
GGTGTGCGACCGGACACTGCCCGAGGCGGAGCGGGCGGCAGCCCTCGTGGCAGC
CCTGACCGACGAGGAGAAGCTGCAAAACCTGGTCAGCAAGGCGCCGGGGGCGCC
GCGGATCGGCCTGCCCGCGTACAACTGGTGGAGCGAGGCGCTGCACGGGGTGGC
CCACGCGCCCGGGACGCAGTTCCGCGACGGGCCGGGGGACTTCAACTCGTCCAC
GTCGTTCCCGATGCCGCTGCTGATGGCCGCCGCCTTCGACGACGAGCTGATCGAG
GCCGTCGGCGACGTCATCGGCACCGAGGCCCGCGCCTTTGGCAACGCCGGCTGG
TCCGGCCTCGACTACTGGACCCCCAACGTCAACCCCTTCCGGGACCCCCGCTGGG
GCCGCGGCTCCGAGACGCCGGGCGAGGACGTCGTGCGCCTCAAGCGCTACGCCG
CCTCCATGATCCGCGGGCTCGAGGGTCGTTCCTCCTCCTCCTCCTCCTGCTCCTTC
GGATCCGGAGGGGAGCCGCCGCGCGTCATCTCGACCTGCAAGCACTACGCCGGC
AACGACTTTGAGGACTGGAACGGCACGACGCGGCACGACTTCGACGCCGTCATC
TCGGCGCAGGACCTGGCCGAGTACTACCTGGCGCCGTTCCAGCAGTGCGCGCGC
GACTCGCGCGTCGGCTCCGTCATGTGCGCCTACAACGCCGTCAACGGGGTGCCGT
CGTGCGCCAACTCGTACCTCATGAACACGATCCTGCGCGGGCACTGGAACTGGA
CCGAGCACGACAACTACGTCACCAGCGACTGCGAGGCCGTCCTCGACGTCTCGG
CCCACCACCACTACGCCGACACCAACGCCGAGGGCACCGGCCTCTGCTTCGAGG
CCGGCATGGACACGAGCTGCGAGTACGAGGGCTCCTCCGACATCCCGGGCGCCT
CCGCCCAGGGCTTCCTGACCTGGCCCGCCGTCGACCGCGCCCTGACGCGGCTGTA
CCGGAGCCTGGTGCGGGTCGGCTACTTTGACGGCCCCGAGTCGCCGCACGCCTCG
CTGGGCTGGGCCGACGTCAACCGGCCCGAGGCGCAGGAGCTGGCCCTGCGCGCT
GCCGTCGAGGGCATCGTGCTGCTCAAGAACGACAACGACACGCTGCCGCTGCCG
CTGCCGGACGATGTCGTTGTCACCGCTGATGGTGGCCGCCGCCGCGTCGCCATGA
TCGGCTTCTGGGCCGACGCCCCGGACAAGCTGTTTGGCGGGTACAGCGGCGCGC
CCCCCTTCGCGCGCTCGCCCGCGAGCGCCGCCCGGCAGCTGGGCTGGAACGTCAC
GGTCGCCGGAGGGCCCGTCCTGGAGGGAGACTCGGACGAGGAGGAGGACACGT
GGACGGCGCCGGCCGTCGAGGCGGCCGCCGACGCCGACTACATCGTCTACTTTG

GCGGCCTGGACACGTCGGCGGCGGGCGAGACCAAGGACCGGATGACGATCGGGT
GGCCGGCGGCGCAGCTGGCGCTCATCTCGGAGCTGGCGCGGCTCGGCAAGCCCG
TCGTGGTGGTGCAGATGGGCGACCAGCTCGACGACACGCCCCTCTTCGAGCTGG
ACGGGGTGGGCGCCGTCCTGTGGGCCAACTGGCCGGGCCAGGACGGCGGCACGG
CCGTGGTCCGGCTGCTCAGCGGCGCCGAGAGCCCGGCCGGCCGCCTGCCCGTGA
CCCAGTACCCGGCCAACTACACCGACGCGGTGCCCCTGACCGACATGACCCTGC
GCCCGTCGGCGACCAACCCGGGCCGGACCTACCGCTGGTACCCGACTCCCGTCCG
GCCCTTCGGCTTCGGCCTCCACTATACCACCTTCCGGGCCGAGTTCGGCCCCCAC
CCCTTCTTCCCGGGGGCGGGCAAGGGCGATGGCGACGGCGAGGACAAGGGCGAG
AGCAAGAGCGAGATCAGGACGCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCG
CAGGGCGGCGGCGGCGGCCACCACGCCGATCCGGGACCTGCTCCGCGACTGCGA
CAAGACGTACCCGGACACGTGCCCGCTGCCGCCGCTGACGGTGCGCGTGACCAA
CGAGGGCGAGCGCGCGTCCGACTACGTGGTGCTGGCCTTCGTGTCGGGCGAGTA
CGGGCCGGCGCCGTACCCGATCAAGACGCTGGTCTCGTACGCGCGGGCGCGCGG
GCTAAAGCCGAAGGGCGGCGACGGCGACGGCGACGGCGACGGCGCCACCACTA
CCGTCTCGCTCGACTGGACCGTCGGCAACCTGGCCCGCCACGACGAGCGCGGCA
ACACAATCCTGTACCCGGGAACTTACACCCTCACTCTCGACGAGCCGGCCCAGGC
GAGCGTGCAGTTCGCCCTCGAGGGCGAGCCCGTCGTGCTCGACGAGTGGCCTGC
GCCGCCGAGTGCCAACTCCACCGCCAGGGGGAGGCACAGG (SEQ ID NO:46)

MKASVSCLVGMSAVAYGLDGPFQTYPDCTKPPLSDIKVCDRTLPEAERAAALVAAL
TDEEKLQNLVSKAPGAPRIGLPAYNWWSEALHGVAHAPGTQFRDGPGDFNSSTSFP
MPLLMAAAFDDELIEAVGDVIGTEARAFGNAGWSGLDYWTPNVNPFRDPRWGRGS
ETPGEDVVRLKRYAASMIRGLEGRSSSSSSCSFGSGGEPPRVISTCKHYAGNDFEDWN
GTTRHDFDAVISAQDLAEYYLAPFQQCARDSRVGSVMCAYNAVNGVPSCANSYLM
NTILRGHWNWTEHDNYVTSDCEAVLDVSAHHHYADTNAEGTGLCFEAGMDTSCEY
EGSSDIPGASAQGFLTWPAVDRALTRLYRSLVRVGYFDGPESPHASLGWADVNRPE
AQELALRAAVEGIVLLKNDNDTLPLPLPDDVVVTADGGRRRVAMIGFWADAPDKLF
GGYSGAPPFARSPASAARQLGWNVTVAGGPVLEGDSDEEEDTWTAPAVEAAADAD
YIVYFGGLDTSAAGETKDRMTIGWPAAQLALISELARLGKPVVVVQMGDQLDDTPL
FELDGVGAVLWANWPGQDGGTAVVRLLSGAESPAGRLPVTQYPANYTDAVPLTDM
TLRPSATNPGRTYRWYPTPVRPFGFGLHYTTFRAEFGPHPFFPGAGKGDGDGEDKGE
SKSEIRTQQQQQQQQQQRRAAAAATTPIRDLLRDCDKTYPDTCPLPPLTVRVTNEGE
RASDYVVLAFVSGEYGPAPYPIKTLVSYARARGLKPKGGTGAGDGDVATTTVSLDW
TVGNLARHDERGNTILYPGTYTLTLDEPAQASVQFALEGEPVVLDEWPAPPSANSTA
RGRHR (SEQ ID NO:47)

Variant A499K:

ATGAAGGCCTCTGTATCATGCCTCGTCGGCATGAGCGCCGTGGCCTACGGCCTCG
ATGGCCCTTTCCAGACCTACCCCGACTGCACCAAGCCCCCCCTGTCCGATATTAA
GGTGTGCGACCGGACACTGCCCGAGGCGGAGCGGGCGGCAGCCCTCGTGGCAGC
CCTGACCGACGAGGAGAAGCTGCAAAACCTGGTCAGCAAGGCGCCGGGGGCGCC
GCGGATCGGCCTGCCCGCGTACAACTGGTGGAGCGAGGCGCTGCACGGGGTGGC
CCACGCGCCCGGGACGCAGTTCCGCGACGGGCCGGGGGACTTCAACTCGTCCAC
GTCGTTCCCGATGCCGCTGCTGATGGCCGCCGCCTTCGACGACGAGCTGATCGAG
GCCGTCGGCGACGTCATCGGCACCGAGGCCCGCGCCTTTGGCAACGCCGGCTGG
TCCGGCCTCGACTACTGGACCCCCAACGTCAACCCCTTCCGGGACCCCCGCTGGG

GCCGCGGCTCCGAGACGCCGGGCGAGGACGTCGTGCGCCTCAAGCGCTACGCCG
CCTCCATGATCCGCGGGCTCGAGGGTCGTTCCTCCTCCTCCTCCTCCTGCTCCTTC
GGATCCGGAGGGGAGCCGCCGCGCGTCATCTCGACCTGCAAGCACTACGCCGGC
AACGACTTTGAGGACTGGAACGGCACGACGCGGCACGACTTCGACGCCGTCATC
TCGGCGCAGGACCTGGCCGAGTACTACCTGGCGCCGTTCCAGCAGTGCGCGCGC
GACTCGCGCGTCGGCTCCGTCATGTGCGCCTACAACGCCGTCAACGGGGTGCCGT
CGTGCGCCAACTCGTACCTCATGAACACGATCCTGCGCGGGCACTGGAACTGGA
CCGAGCACGACAACTACGTCACCAGCGACTGCGAGGCCGTCCTCGACGTCTCGG
CCCACCACCACTACGCCGACACCAACGCCGAGGGCACCGGCCTCTGCTTCGAGG
CCGGCATGGACACGAGCTGCGAGTACGAGGGCTCCTCCGACATCCCGGGCGCCT
CCGCCGGCGGCTTCCTGACCTGGCCCGCCGTCGACCGCGCCCTGACGCGGCTGTA
CCGGAGCCTGGTGCGGGTCGGCTACTTTGACGGCCCCGAGTCGCCGCACGCCTCG
CTGGGCTGGGCCGACGTCAACCGGCCCGAGGCGCAGGAGCTGGCCCTGCGCGCT
GCCGTCGAGGGCATCGTGCTGCTCAAGAACGACAACGACACGCTGCCGCTGCCG
CTGCCGGACGATGTCGTTGTCACCGCTGATGGTGGCCGCCGCCGCGTCGCCATGA
TCGGCTTCTGGGCCGACGCCCCGGACAAGCTGTTTGGCGGGTACAGCGGCGCGC
CCCCCTTCGCGCGCTCGCCCGCGAGCGCCGCCCGGCAGCTGGGCTGGAACGTCAC
GGTCGCCGGAGGGCCCGTCCTGGAGGGAGACTCGGACGAGGAGGAGGACACGT
GGACGGCGCCGGCCGTCGAGGCGGCCAAGGACGCCGACTACATCGTCTACTTTG
GCGGCCTGGACACGTCGGCGGCGGGCGAGACCAAGGACCGGATGACGATCGGGT
GGCCGGCGGCGCAGCTGGCGCTCATCTCGGAGCTGGCGCGGCTCGGCAAGCCCG
TCGTGGTGGTGCAGATGGGCGACCAGCTCGACGACACGCCCCTCTTCGAGCTGG
ACGGGGTGGGCGCCGTCCTGTGGGCCAACTGGCCGGGCCAGGACGGCGGCACGG
CCGTGGTCCGGCTGCTCAGCGGCGCCGAGAGCCCGGCCGGCCGCCTGCCCGTGA
CCCAGTACCCGGCCAACTACACCGACGCGGTGCCCCTGACCGACATGACCCTGC
GCCCGTCGGCGACCAACCCGGGCCGGACCTACCGCTGGTACCCGACTCCCGTCCG
GCCCTTCGGCTTCGGCCTCCACTATACCACCTTCCGGGCCGAGTTCGGCCCCCAC
CCCTTCTTCCCGGGGGCGGGCAAGGGCGATGGCGACGGCGAGGACAAGGGCGAG
AGCAAGAGCGAGATCAGGACGCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCG
CAGGGCGGCGGCGGCGGCCACCACGCCGATCCGGGACCTGCTCCGCGACTGCGA
CAAGACGTACCCGGACACGTGCCCGCTGCCGCCGCTGACGGTGCGCGTGACCAA
CGAGGGCGAGCGCGCGTCCGACTACGTGGTGCTGGCCTTCGTGTCGGGCGAGTA
CGGGCCGGCGCCGTACCCGATCAAGACGCTGGTCTCGTACGCGCGGGCGCGCGG
GCTAAAGGGGAAGGGCGGCGACGGCGACGGCGACGGCGACGGCGCCACCACTA
CCGTCTCGCTCGACTGGACCGTCGGCAACCTGGCCCGCCACGACGAGCGCGGCA
ACACAATCCTGTACCCGGGAACTTACACCCTCACTCTCGACGAGCCGGCCCAGGC
GAGCGTGCAGTTCGCCCTCGAGGGCGAGCCCGTCGTGCTCGACGAGTGGCCTGC
GCCGCCGAGTGCCAACTCCACCGCCAGGGGGAGGCACAGG (SEQ ID NO:48)

MKASVSCLVGMSAVAYGLDGPFQTYPDCTKPPLSDIKVCDRTLPEAERAAALVAAL
TDEEKLQNLVSKAPGAPRIGLPAYNWWSEALHGVAHAPGTQFRDGPGDFNSSTSFP
MPLLMAAAFDDELIEAVGDVIGTEARAFGNAGWSGLDYWTPNVNPFRDPRWGRGS
ETPGEDVVRLKRYAASMIRGLEGRSSSSSSCSFGSGGEPPRVISTCKHYAGNDFEDWN
GTTRHDFDAVISAQDLAEYYLAPFQQCARDSRVGSVMCAYNAVNGVPSCANSYLM
NTILRGHWNWTEHDNYVTSDCEAVLDVSAHHHYADTNAEGTGLCFEAGMDTSCEY
EGSSDIPGASAGGFLTWPAVDRALTRLYRSLVRVGYFDGPESPHASLGWADVNRPE
AQELALRAAVEGIVLLKNDNDTLPLPLPDDVVVTADGGRRRVAMIGFWADAPDKLF
GGYSGAPPFARSPASAARQLGWNVTVAGGPVLEGDSDEEEDTWTAPAVEAAKDAD
YIVYFGGLDTSAAGETKDRMTIGWPAAQLALISELARLGKPVVVVQMGDQLDDTPL
FELDGVGAVLWANWPGQDGGTAVVRLLSGAESPAGRLPVTQYPANYTDAVPLTDM

TLRPSATNPGRTYRWYPTPVRPFGFGLHYTTFRAEFGPHPFFPGAGKGDGDGEDKGE
SKSEIRTQQQQQQQQQQRRAAAAATTPIRDLLRDCDKTYPDTCPLPPLTVRVTNEGE
RASDYVVLAFVSGEYGPAPYPIKTLVSYARARGLKGKGGTGAGDGDVATTTVSLD
WTVGNLARHDERGNTILYPGTYTLTLDEPAQASVQFALEGEPVVLDEWPAPPSANST
ARGRHR (SEQ ID NO:49)

Variant A499S:

ATGAAGGCCTCTGTATCATGCCTCGTCGGCATGAGCGCCGTGGCCTACGGCCTCG
ATGGCCCTTTCCAGACCTACCCCGACTGCACCAAGCCCCCCCTGTCCGATATTAA
GGTGTGCGACCGGACACTGCCCGAGGCGGAGCGGGCGGCAGCCCTCGTGGCAGC
CCTGACCGACGAGGAGAAGCTGCAAAACCTGGTCAGCAAGGCGCCGGGGGCGCC
GCGGATCGGCCTGCCCGCGTACAACTGGTGGAGCGAGGCGCTGCACGGGGTGGC
CCACGCGCCCGGGACGCAGTTCCGCGACGGGCCGGGGGACTTCAACTCGTCCAC
GTCGTTCCCGATGCCGCTGCTGATGGCCGCCGCCTTCGACGACGAGCTGATCGAG
GCCGTCGGCGACGTCATCGGCACCGAGGCCCGCGCCTTTGGCAACGCCGGCTGG
TCCGGCCTCGACTACTGGACCCCCAACGTCAACCCCTTCCGGGACCCCCGCTGGG
GCCGCGGCTCCGAGACGCCGGGCGAGGACGTCGTGCGCCTCAAGCGCTACGCCG
CCTCCATGATCCGCGGGCTCGAGGGTCGTTCCTCCTCCTCCTCCTCCTGCTCCTTC
GGATCCGGAGGGGAGCCGCCGCGCGTCATCTCGACCTGCAAGCACTACGCCGGC
AACGACTTTGAGGACTGGAACGGCACGACGCGGCACGACTTCGACGCCGTCATC
TCGGCGCAGGACCTGGCCGAGTACTACCTGGCGCCGTTCCAGCAGTGCGCGCGC
GACTCGCGCGTCGGCTCCGTCATGTGCGCCTACAACGCCGTCAACGGGGTGCCGT
CGTGCGCCAACTCGTACCTCATGAACACGATCCTGCGCGGGCACTGGAACTGGA
CCGAGCACGACAACTACGTCACCAGCGACTGCGAGGCCGTCCTCGACGTCTCGG
CCCACCACCACTACGCCGACACCAACGCCGAGGGCACCGGCCTCTGCTTCGAGG
CCGGCATGGACACGAGCTGCGAGTACGAGGGCTCCTCCGACATCCCGGGCGCCT
CCGCCGGCGGCTTCCTGACCTGGCCCGCCGTCGACCGCGCCCTGACGCGGCTGTA
CCGGAGCCTGGTGCGGGTCGGCTACTTTGACGGCCCCGAGTCGCCGCACGCCTCG
CTGGGCTGGGCCGACGTCAACCGGCCCGAGGCGCAGGAGCTGGCCCTGCGCGCT
GCCGTCGAGGGCATCGTGCTGCTCAAGAACGACAACGACACGCTGCCGCTGCCG
CTGCCGGACGATGTCGTTGTCACCGCTGATGGTGGCCGCCGCCGCGTCGCCATGA
TCGGCTTCTGGGCCGACGCCCCGGACAAGCTGTTTGGCGGGTACAGCGGCGCGC
CCCCCTTCGCGCGCTCGCCCGCGAGCGCCGCCCGGCAGCTGGGCTGGAACGTCAC
GGTCGCCGGAGGGCCCGTCCTGGAGGGAGACTCGGACGAGGAGGAGGACACGT
GGACGGCGCCGGCCGTCGAGGCGGCCTCTGACGCCGACTACATCGTCTACTTTGG
CGGCCTGGACACGTCGGCGGCGGGCGAGACCAAGGACCGGATGACGATCGGGTG
GCCGGCGGCGCAGCTGGCGCTCATCTCGGAGCTGGCGCGGCTCGGCAAGCCCGT
CGTGGTGGTGCAGATGGGCGACCAGCTCGACGACACGCCCCTCTTCGAGCTGGA
CGGGGTGGGCGCCGTCCTGTGGGCCAACTGGCCGGGCCAGGACGGCGGCACGGC
CGTGGTCCGGCTGCTCAGCGGCGCCGAGAGCCCGGCCGGCCGCCTGCCCGTGAC
CCAGTACCCGGCCAACTACACCGACGCGGTGCCCCTGACCGACATGACCCTGCG
CCCGTCGGCGACCAACCCGGGCCGGACCTACCGCTGGTACCCGACTCCCGTCCGG
CCCTTCGGCTTCGGCCTCCACTATACCACCTTCCGGGCCGAGTTCGGCCCCCACC
CCTTCTTCCCGGGGGCGGGCAAGGGCGATGGCGACGGCGAGGACAAGGGCGAG
AGCAAGAGCGAGATCAGGACGCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCG
CAGGGCGGCGGCGGCGGCCACCACGCCGATCCGGGACCTGCTCCGCGACTGCGA
CAAGACGTACCCGGACACGTGCCCGCTGCCGCCGCTGACGGTGCGCGTGACCAA

CGAGGGCGAGCGCGCGTCCGACTACGTGGTGCTGGCCTTCGTGTCGGGCGAGTA
CGGGCCGGCGCCGTACCCGATCAAGACGCTGGTCTCGTACGCGCGGGCGCGCGG
GCTAAAGGGGAAGGGCGGCGACGGCGACGGCGACGGCGACGGCGCCACCACTA
CCGTCTCGCTCGACTGGACCGTCGGCAACCTGGCCCGCCACGACGAGCGCGGCA
ACACAATCCTGTACCCGGGAACTTACACCCTCACTCTCGACGAGCCGGCCCAGGC
GAGCGTGCAGTTCGCCCTCGAGGGCGAGCCCGTCGTGCTCGACGAGTGGCCTGC
GCCGCCGAGTGCCAACTCCACCGCCAGGGGGAGGCACAGG (SEQ ID NO:50)

MKASVSCLVGMSAVAYGLDGPFQTYPDCTKPPLSDIKVCDRTLPEAERAAALVAAL
TDEEKLQNLVSKAPGAPRIGLPAYNWWSEALHGVAHAPGTQFRDGPGDFNSSTSFP
MPLLMAAAFDDELIEAVGDVIGTEARAFGNAGWSGLDYWTPNVNPFRDPRWGRGS
ETPGEDVVRLKRYAASMIRGLEGRSSSSSSCSFGSGGEPPRVISTCKHYAGNDFEDWN
GTTRHDFDAVISAQDLAEYYLAPFQQCARDSRVGSVMCAYNAVNGVPSCANSYLM
NTILRGHWNWTEHDNYVTSDCEAVLDVSAHHHYADTNAEGTGLCFEAGMDTSCEY
EGSSDIPGASAGGFLTWPAVDRALTRLYRSLVRVGYFDGPESPHASLGWADVNRPE
AQELALRAAVEGIVLLKNDNDTLPLPLPDDVVVTADGGRRRVAMIGFWADAPDKLF
GGYSGAPPFARSPASAARQLGWNVTVAGGPVLEGDSDEEEDTWTAPAVEAASDAD
YIVYFGGLDTSAAGETKDRMTIGWPAAQLALISELARLGKPVVVVQMGDQLDDTPL
FELDGVGAVLWANWPGQDGGTAVVRLLSGAESPAGRLPVTQYPANYTDAVPLTDM
TLRPSATNPGRTYRWYPTPVRPFGFGLHYTTFRAEFGPHPFFPGAGKGDGDGEDKGE
SKSEIRTQQQQQQQQQRRAAAAATTPIRDLLRDCDKTYPDTCPLPPLTVRVTNEGE
RASDYVVLAFVSGEYGPAPYPIKTLVSYARARGLKGKGGTGAGDGDVATTTVSLD
WTVGNLARHDERGNTILYPGTYTLTLDEPAQASVQFALEGEPVVLDEWPAPPSANST
ARGRHR (SEQ ID NO:51)

Variant I798V:

ATGAAGGCCTCTGTATCATGCCTCGTCGGCATGAGCGCCGTGGCCTACGGCCTCG
ATGGCCCTTTCCAGACCTACCCCGACTGCACCAAGCCCCCCCTGTCCGATATTAA
GGTGTGCGACCGGACACTGCCCGAGGCGGAGCGGGCGGCAGCCCTCGTGGCAGC
CCTGACCGACGAGGAGAAGCTGCAAAACCTGGTCAGCAAGGCGCCGGGGGCGCC
GCGGATCGGCCTGCCCGCGTACAACTGGTGGAGCGAGGCGCTGCACGGGGTGGC
CCACGCGCCCGGGACGCAGTTCCGCGACGGGCCGGGGGACTTCAACTCGTCCAC
GTCGTTCCCGATGCCGCTGCTGATGGCCGCCGCCTTCGACGACGAGCTGATCGAG
GCCGTCGGCGACGTCATCGGCACCGAGGCCCGCGCCTTTGGCAACGCCGGCTGG
TCCGGCCTCGACTACTGGACCCCCAACGTCAACCCCTTCCGGGACCCCCGCTGGG
GCCGCGGCTCCGAGACGCCGGGCGAGGACGTCGTGCGCCTCAAGCGCTACGCCG
CCTCCATGATCCGCGGGCTCGAGGGTCGTTCCTCCTCCTCCTCCTCCTGCTCCTTC
GGATCCGGAGGGGAGCCGCCGCGCGTCATCTCGACCTGCAAGCACTACGCCGGC
AACGACTTTGAGGACTGGAACGGCACGACGCGGCACGACTTCGACGCCGTCATC
TCGGCGCAGGACCTGGCCGAGTACTACCTGGCGCCGTTCCAGCAGTGCGCGCGC
GACTCGCGCGTCGGCTCCGTCATGTGCGCCTACAACGCCGTCAACGGGGTGCCGT
CGTGCGCCAACTCGTACCTCATGAACACGATCCTGCGCGGGCACTGGAACTGGA
CCGAGCACGACAACTACGTCACCAGCGACTGCGAGGCCGTCCTCGACGTCTCGG
CCCACCACCACTACGCCGACACCAACGCCGAGGGCACCGGCCTCTGCTTCGAGG
CCGGCATGGACACGAGCTGCGAGTACGAGGGCTCCTCCGACATCCCGGGCGCCT
CCGCCGGCGGCTTCCTGACCTGGCCCGCCGTCGACCGCGCCCTGACGCGGCTGTA
CCGGAGCCTGGTGCGGGTCGGCTACTTTGACGGCCCCGAGTCGCCGCACGCCTCG

CTGGGCTGGGCCGACGTCAACCGGCCCGAGGCGCAGGAGCTGGCCCTGCGCGCT
GCCGTCGAGGGCATCGTGCTGCTCAAGAACGACAACGACACGCTGCCGCTGCCG
CTGCCGGACGATGTCGTTGTCACCGCTGATGGTGGCCGCCGCCGCGTCGCCATGA
TCGGCTTCTGGGCCGACGCCCCGGACAAGCTGTTTGGCGGGTACAGCGGCGCGC
CCCCCTTCGCGCGCTCGCCCGCGAGCGCCGCCCGGCAGCTGGGCTGGAACGTCAC
GGTCGCCGGAGGGCCCGTCCTGGAGGGAGACTCGGACGAGGAGGAGGACACGT
GGACGGCGCCGGCCGTCGAGGCGGCCGCCGACGCCGACTACATCGTCTACTTTG
GCGGCCTGGACACGTCGGCGGCGGGCGAGACCAAGGACCGGATGACGATCGGGT
GGCCGGCGGCGCAGCTGGCGCTCATCTCGGAGCTGGCGCGGCTCGGCAAGCCCG
TCGTGGTGGTGCAGATGGGCGACCAGCTCGACGACACGCCCCTCTTCGAGCTGG
ACGGGGTGGGCGCCGTCCTGTGGGCCAACTGGCCGGGCCAGGACGGCGGCACGG
CCGTGGTCCGGCTGCTCAGCGGCGCCGAGAGCCCGGCCGGCCGCCTGCCCGTGA
CCCAGTACCCGGCCAACTACACCGACGCGGTGCCCCTGACCGACATGACCCTGC
GCCCGTCGGCGACCAACCCGGGCCGGACCTACCGCTGGTACCCGACTCCCGTCCG
GCCCTTCGGCTTCGGCCTCCACTATACCACCTTCCGGGCCGAGTTCGGCCCCCAC
CCCTTCTTCCCGGGGGCGGGCAAGGGCGATGGCGACGGCGAGGACAAGGGCGAG
AGCAAGAGCGAGATCAGGACGCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCG
CAGGGCGGCGGCGGCGGCCACCACGCCGATCCGGGACCTGCTCCGCGACTGCGA
CAAGACGTACCCGGACACGTGCCCGCTGCCGCCGCTGACGGTGCGCGTGACCAA
CGAGGGCGAGCGCGCGTCCGACTACGTGGTGCTGGCCTTCGTGTCGGGCGAGTA
CGGGCCGGCGCCGTACCCGATCAAGACGCTGGTCTCGTACGCGCGGGCGCGCGG
GCTAAAGGGGAAGGGCGGCGACGGCGACGGCGACGGCGACGGCGCCACCACTA
CCGTCTCGCTCGACTGGACCGTCGGCAACCTGGCCCGCCACGACGAGCGCGGCA
ACACAGTCCTGTACCCGGGAACTTACACCCTCACTCTCGACGAGCCGGCCCAGGC
GAGCGTGCAGTTCGCCCTCGAGGGCGAGCCCGTCGTGCTCGACGAGTGGCCTGC
GCCGCCGAGTGCCAACTCCACCGCCAGGGGGAGGCACAGG (SEQ ID NO:52)

MKASVSCLVGMSAVAYGLDGPFQTYPDCTKPPLSDIKVCDRTLPEAERAAALVAAL
TDEEKLQNLVSKAPGAPRIGLPAYNWWSEALHGVAHAPGTQFRDGPGDFNSSTSFP
MPLLMAAAFDDELIEAVGDVIGTEARAFGNAGWSGLDYWTPNVNPFRDPRWGRGS
ETPGEDVVRLKRYAASMIRGLEGRSSSSSSCSFGSGGEPPRVISTCKHYAGNDFEDWN
GTTRHDFDAVISAQDLAEYYLAPFQQCARDSRVGSVMCAYNAVNGVPSCANSYLM
NTILRGHWNWTEHDNYVTSDCEAVLDVSAHHHYADTNAEGTGLCFEAGMDTSCEY
EGSSDIPGASAGGFLTWPAVDRALTRLYRSLVRVGYFDGPESPHASLGWADVNRPE
AQELALRAAVEGIVLLKNDNDTLPLPLPDDVVVTADGGRRRVAMIGFWADAPDKLF
GGYSGAPPFARSPASAARQLGWNVTVAGGPVLEGDSDEEEDTWTAPAVEAAADAD
YIVYFGGLDTSAAGETKDRMTIGWPAAQLALISELARLGKPVVVVQMGDQLDDTPL
FELDGVGAVLWANWPGQDGGTAVVRLLSGAESPAGRLPVTQYPANYTDAVPLTDM
TLRPSATNPGRTYRWYPTPVRPFGFGLHYTTFRAEFGPHPFFPGAGKGDGDGEDKGE
SKSEIRTQQQQQQQQQQRRAAAAATTPIRDLLRDCDKTYPDTCPLPPLTVRVTNEGE
RASDYVVLAFVSGEYGPAPYPIKTLVSYARARGLKGKGGTGAGDGDVATTTVSLD
WTVGNLARHDERGNTVLYPGTYTLTLDEPAQASVQFALEGEPVVLDEWPAPPSANS
TARGRHR (SEQ ID NO:53)

Variant P31G/H379Y:

ATGAAGGCCTCTGTATCATGCCTCGTCGGCATGAGCGCCGTGGCCTACGGCCTCG
ATGGCCCTTTCCAGACCTACCCCGACTGCACCAAGGGCCCCCTGTCCGATATTAA

GGTGTGCGACCGGACACTGCCCGAGGCGGAGCGGGCGGCAGCCCTCGTGGCAGC
CCTGACCGACGAGGAGAAGCTGCAAAACCTGGTCAGCAAGGCGCCGGGGGCGCC
GCGGATCGGCCTGCCCGCGTACAACTGGTGGAGCGAGGCGCTGCACGGGGTGGC
CCACGCGCCCGGGACGCAGTTCCGCGACGGGCCGGGGGACTTCAACTCGTCCAC
GTCGTTCCCGATGCCGCTGCTGATGGCCGCCGCCTTCGACGACGAGCTGATCGAG
GCCGTCGGCGACGTCATCGGCACCGAGGCCCGCGCCTTTGGCAACGCCGGCTGG
TCCGGCCTCGACTACTGGACCCCCAACGTCAACCCCTTCCGGGACCCCCGCTGGG
GCCGCGGCTCCGAGACGCCGGGCGAGGACGTCGTGCGCCTCAAGCGCTACGCCG
CCTCCATGATCCGCGGGCTCGAGGGTCGTTCCTCCTCCTCCTCCTCCTGCTCCTTC
GGATCCGGAGGGGAGCCGCCGCGCGTCATCTCGACCTGCAAGCACTACGCCGGC
AACGACTTTGAGGACTGGAACGGCACGACGCGGCACGACTTCGACGCCGTCATC
TCGGCGCAGGACCTGGCCGAGTACTACCTGGCGCCGTTCCAGCAGTGCGCGCGC
GACTCGCGCGTCGGCTCCGTCATGTGCGCCTACAACGCCGTCAACGGGGTGCCGT
CGTGCGCCAACTCGTACCTCATGAACACGATCCTGCGCGGGCACTGGAACTGGA
CCGAGCACGACAACTACGTCACCAGCGACTGCGAGGCCGTCCTCGACGTCTCGG
CCCACCACCACTACGCCGACACCAACGCCGAGGGCACCGGCCTCTGCTTCGAGG
CCGGCATGGACACGAGCTGCGAGTACGAGGGCTCCTCCGACATCCCGGGCGCCT
CCGCCGGCGGCTTCCTGACCTGGCCCGCCGTCGACCGCGCCCTGACGCGGCTGTA
CCGGAGCCTGGTGCGGGTCGGCTACTTTGACGGCCCCGAGTCGCCGTACGCCTCG
CTGGGCTGGGCCGACGTCAACCGGCCCGAGGCGCAGGAGCTGGCCCTGCGCGCT
GCCGTCGAGGGCATCGTGCTGCTCAAGAACGACAACGACACGCTGCCGCTGCCG
CTGCCGGACGATGTCGTTGTCACCGCTGATGGTGGCCGCCGCCGCGTCGCCATGA
TCGGCTTCTGGGCCGACGCCCCGGACAAGCTGTTTGGCGGGTACAGCGGCGCGC
CCCCCTTCGCGCGCTCGCCCGCGAGCGCCGCCCGGCAGCTGGGCTGGAACGTCAC
GGTCGCCGGAGGGCCCGTCCTGGAGGGAGACTCGGACGAGGAGGAGGACACGT
GGACGGCGCCGGCCGTCGAGGCGGCCGCCGACGCCGACTACATCGTCTACTTTG
GCGGCCTGGACACGTCGGCGGCGGGCGAGACCAAGGACCGGATGACGATCGGGT
GGCCGGCGGCGCAGCTGGCGCTCATCTCGGAGCTGGCGCGGCTCGGCAAGCCCG
TCGTGGTGGTGCAGATGGGCGACCAGCTCGACGACACGCCCCTCTTCGAGCTGG
ACGGGGTGGGCGCCGTCCTGTGGGCCAACTGGCCGGGCCAGGACGGCGGCACGG
CCGTGGTCCGGCTGCTCAGCGGCGCCGAGAGCCCGGCCGGCCGCCTGCCCGTGA
CCCAGTACCCGGCCAACTACACCGACGCGGTGCCCCTGACCGACATGACCCTGC
GCCCGTCGGCGACCAACCCGGGCCGGACCTACCGCTGGTACCCGACTCCCGTCCG
GCCCTTCGGCTTCGGCCTCCACTATACCACCTTCCGGGCCGAGTTCGGCCCCCAC
CCCTTCTTCCCGGGGGCGGGCAAGGGCGATGGCGACGGCGAGGACAAGGGCGAG
AGCAAGAGCGAGATCAGGACGCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCG
CAGGGCGGCGGCGGCGGCCACCACGCCGATCCGGGACCTGCTCCGCGACTGCGA
CAAGACGTACCCGGACACGTGCCCGCTGCCGCCGCTGACGGTGCGCGTGACCAA
CGAGGGCGAGCGCGCGTCCGACTACGTGGTGCTGGCCTTCGTGTCGGGCGAGTA
CGGGCCGGCGCCGTACCCGATCAAGACGCTGGTCTCGTACGCGCGGGCGCGCGG
GCTAAAGGGGAAGGGCGGCGACGGCGACGGCGACGGCGACGGCGCCACCACTA
CCGTCTCGCTCGACTGGACCGTCGGCAACCTGGCCCGCCACGACGAGCGCGGCA
ACACAATCCTGTACCCGGGAACTTACACCCTCACTCTCGACGAGCCGGCCCAGGC
GAGCGTGCAGTTCGCCCTCGAGGGCGAGCCCGTCGTGCTCGACGAGTGGCCTGC
GCCGCCGAGTGCCAACTCCACCGCCAGGGGGAGGCACAGG (SEQ ID NO:54)

MKASVSCLVGMSAVAYGLDGPFQTYPDCTKGPLSDIKVCDRTLPEAERAAALVAAL
TDEEKLQNLVSKAPGAPRIGLPAYNWWSEALHGVAHAPGTQFRDGPGDFNSSTSFP
MPLLMAAAFDDELIEAVGDVIGTEARAFGNAGWSGLDYWTPNVNPFRDPRWGRGS
ETPGEDVVRLKRYAASMIRGLEGRSSSSSSCSFGSGGEPPRVISTCKHYAGNDFEDWN

GTTRHDFDAVISAQDLAEYYLAPFQQCARDSRVGSVMCAYNAVNGVPSCANSYLM
NTILRGHWNWTEHDNYVTSDCEAVLDVSAHHHYADTNAEGTGLCFEAGMDTSCEY
EGSSDIPGASAGGFLTWPAVDRALTRLYRSLVRVGYFDGPESPYASLGWADVNRPE
AQELALRAAVEGIVLLKNDNDTLPLPLPDDVVVTADGGRRRVAMIGFWADAPDKLF
GGYSGAPPFARSPASAARQLGWNVTVAGGPVLEGDSDEEEDTWTAPAVEAAADAD
YIVYFGGLDTSAAGETKDRMTIGWPAAQLALISELARLGKPVVVVQMGDQLDDTPL
FELDGVGAVLWANWPGQDGGTAVVRLLSGAESPAGRLPVTQYPANYTDAVPLTDM
TLRPSATNPGRTYRWYPTPVRPFGFGLHYTTFRAEFGPHPFFPGAGKGDGDGEDKGE
SKSEIRTQQQQQQQQQRRAAAAATTPIRDLLRDCDKTYPDTCPLPPLTVRVTNEGE
RASDYVVLAFVSGEYGPAPYPIKTLVSYARARGLKGKGGTGAGDGDVATTTVSLD
WTVGNLARHDERGNTILYPGTYTLTLDEPAQASVQFALEGEPVVLDEWPAPPSANST
ARGRHR (SEQ ID NO:55)

Beta-xylosidase WT2:

ATGAAGGCCTCTGTATCATGCCTCGTCGGCATGAGCGCCGTGGCCTACGGCCTCG
ATGGCCCTTTCCAGACCTACCCCGACTGCACCAAGCCCCCCCTGTCCGATATTAA
GGTGTGCGACCGGACACTGCCCGAGGCGGAGCGGGCGGCAGCCCTCGTGGCAGC
CCTGACCGACGAGGAGAAGCTGCAAAACCTGGTCAGCAAGGCGCCGGGGGCGCC
GCGGATCGGCCTGCCCGCGTACAACTGGTGGAGCGAGGCGCTGCACGGGGTGGC
CCACGCGCCCGGGACGCAGTTCCGCGACGGGCCGGGGGACTTCAACTCGTCCAC
GTCGTTCCCGATGCCGCTGCTGATGGCCGCCGCCTTCGACGACGAGCTGATCGAG
GCCGTCGGCGACGTCATCGGCACCGAGGCCCGCGCCTTTGGCAACGCCGGCTGG
TCCGGCCTCGACTACTGGACCCCCAACGTCAACCCCTTCCGGGACCCCCGCTGGG
GCCGCGGCTCCGAGACGCCGGGCGAGGACGTCGTGCGCCTCAAGCGCTACGCCG
CCTCCATGATCCGCGGGCTCGAGGGTCGTTCCTCCTCCTCCTCCTCCTGCTCCTTC
GGATCCGGAGGGGAGCCGCCGCGCGTCATCTCGACCTGCAAGCACTACGCCGGC
AACGACTTTGAGGACTGGAACGGCACGACGCGGCACGACTTCGACGCCGTCATC
TCGGCGCAGGACCTGGCCGAGTACTACCTGGCGCCGTTCCAGCAGTGCGCGCGC
GACTCGCGCGTCGGCTCCGTCATGTGCGCCTACAACGCCGTCAACGGGGTGCCGT
CGTGCGCCAACTCGTACCTCATGAACACGATCCTGCGCGGGCACTGGAACTGGA
CCGAGCACGACAACTACGTCACCAGCGACTGCGAGGCCGTCCTCGACGTCTCGG
CCCACCACCACTACGCCGACACCAACGCCGAGGGCACCGGCCTCTGCTTCGAGG
CCGGCATGGACACGAGCTGCGAGTACGAGGGCTCCTCCGACATCCCGGGCGCCT
CCGCCGGCGGCTTCCTGACCTGGCCCGCCGTCGACCGCGCCCTGACGCGGCTGTA
CCGGAGCCTGGTGCGGGTCGGCTACTTTGACGGCCCCGAGTCGCCGCACGCCTCG
CTGGGCTGGGCCGACGTCAACCGGCCCGAGGCGCAGGAGCTGGCCCTGCGCGCT
GCCGTCGAGGGCATCGTGCTGCTCAAGAACGACAACGACACGCTGCCGCTGCCG
CTGCCGGACGATGTCGTTGTCACCGCTGATGGTGGCCGCCGCCGCGTCGCCATGA
TCGGCTTCTGGGCCGACGCCCCGGACAAGCTGTTTGGCGGGTACAGCGGCGCGC
CCCCCTTCGCGCGCTCGCCCGCGAGCGCCGCCCGGCAGCTGGGCTGGAACGTCAC
GGTCGCCGGAGGGCCCGTCCTGGAGGGAGACTCGGACGAGGAGGAGGACACGT
GGACGGCGCCGGCCGTCGAGGCGGCCGCCGACGCCGACTACATCGTCTACTTTG
GCGGCCTGGACACGTCGGCGGCGGGCGAGACCAAGGACCGGATGACGATCGGGT
GGCCGGCGGCGCAGCTGGCGCTCATCTCGGAGCTGGCGCGGCTCGGCAAGCCCG
TCGTGGTGGTGCAGATGGGCGACCAGCTCGACGACACGCCCCTCTTCGAGCTGG
ACGGGGTGGGCGCCGTCCTGTGGGCCAACTGGCCGGGCCAGGACGGCGGCACGG
CCGTGGTCCGGCTGCTCAGCGGCGCCGAGAGCCCGGCCGGCCGCCTGCCCGTGA

CCCAGTACCCGGCCAACTACACCGACGCGGTGCCCCTGACCGACATGACCCTGC
GCCCGTCGGCGACCAACCCGGGCCGGACCTACCGCTGGTACCCGACTCCCGTCCG
GCCCTTCGGCTTCGGCCTCCACTATACCACCTTCCGGGCCGAGTTCGGCCCCCAC
CCCTTCTTCCCGGGGGCGGGCAAGGGCGATGGCGACGGCGAGGACAAGGGCGAG
AGCAAGAGCGAGATCAGGACGCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCG
CAGGGCGGCGGCGGCGGCCACCACGCCGATCCGGGACCTGCTCCGCGACTGCGA
CAAGACGTACCCGGACACGTGCCCGCTGCCGCCGCTGACGGTGCGCGTGACCAA
CGAGGGCGAGCGCGCGTCCGACTACGTGGTGCTGGCCTTCGTGTCGGGCGAGTA
CGGGCCGGCGCCGTACCCGATCAAGACGCTGGTCTCGTACGCGCGGGCGCGCGG
GCTAAAGGGGAAGGGCGGCGACGGCGACGGCGACGGCGACGGCGCCACCACTA
CCGTCTCGCTCGACTGGACCGTCGGCAACCTGGCCCGCCACGACGAGCGCGGCA
ACACAATCCTGTACCCGGGAACTTACACCCTCACTCTCGACGAGCCGGCCCAGGC
GAGCGTGCAGTTCGCCCTCGAGGGCGAGCCCGTCGTGCTCGACGAGTGGCCTGC
GCCGCCGAGTGCCAACTCCACCGCCAGGGGGAGGCACAGG (SEQ ID NO:56)

MKASVSCLVGMSAVAYGLDGPFQTYPDCTKPPLSDIKVCDRTLPEAERAAALVAAL
TDEEKLQNLVSKAPGAPRIGLPAYNWWSEALHGVAHAPGTQFRDGPGDFNSSTSFP
MPLLMAAAFDDELIEAVGDVIGTEARAFGNAGWSGLDYWTPNVNPFRDPRWGRGS
ETPGEDVVRLKRYAASMIRGLEGRSSSSSSCSFGSGGEPPRVISTCKHYAGNDFEDWN
GTTRHDFDAVISAQDLAEYYLAPFQQCARDSRVGSVMCAYNAVNGVPSCANSYLM
NTILRGHWNWTEHDNYVTSDCEAVLDVSAHHHYADTNAEGTGLCFEAGMDTSCEY
EGSSDIPGASAGGFLTWPAVDRALTRLYRSLVRVGYFDGPESPHASLGWADVNRPE
AQELALRAAVEGIVLLKNDNDTLPLPLPDDVVVTADGGRRRVAMIGFWADAPDKLF
GGYSGAPPFARSPASAARQLGWNVTVAGGPVLEGDSDEEEDTWTAPAVEAAADAD
YIVYFGGLDTSAAGETKDRMTIGWPAAQLALISELARLGKPVVVVQMGDQLDDTPL
FELDGVGAVLWANWPGQDGGTAVVRLLSGAESPAGRLPVTQYPANYTDAVPLTDM
TLRPSATNPGRTYRWYPTPVRPFGFGLHYTTFRAEFGPHPFFPGAGKGDGDGEDKGE
SKSEIRTQQQQQQQQQRRAAAAATTPIRDLLRDCDKTYPDTCPLPPLTVRVTNEGE
RASDYVVLAFVSGEYGPAPYPIKTLVSYARARGLKGKGGDGDGDGDGATTTVSLD
WTVGNLARHDERGNTILYPGTYTLTLDEPAQASVQFALEGEPVVLDEWPAPPSANST
ARGRHR (SEQ ID NO:57)

BXyl Variant-233

ATGAAGGCCTCTGTATCATGCCTCGTCGGCATGAGCGCCGTGGCCTACGGCCTCG
ATGGCCCTTTCCAGACCTACCCCGACTGCACCAAGCCCCCCCTGTCCGATATTAA
GGTGTGCGACCGGACACTGCCCGAGGCGGAGCGGGCGGCAGCCCTCGTGGCAGC
CCTGACCGACGAGGAGAAGCTGCAAAACCTGGTCAGCAAGGCGCCGGGGGCGCC
GCGGATCGGCCTGCCCGCGTACAACTGGTGGAGCGAGGCGCTGCACGGGGTGGC
CCACGCGCCCGGGACGCAATTCCGCGACGGGCCGGGGGACTTCAACTCGTCCAC
GTCGTTCCCGATGCCGCTGCTGATGGCCGCCGCCTTCGACGACGAGCTGATCGAG
GCCGTCGGCGACGTCATCGGCACCGAGGCCCGCGCCTTTGGCAACGCCGGCTGG
TCCGGCCTCGACTACTGGACCCCCAACGTCAACCCCTTCCGGGACCCCCGCTGGG
GCCGCGGCTCCGAGACGCCGGGCGAGGACGTCGTGCGCCTCAAGCGCTACGCCG
CCTCCATGATCCGCGGGCTCGAGGGTCGTTCCTCCTCCTCCTCCTCCTGCTCCTTC
GGATCCGGAGGGGAGCCGCCGCGCGTCATCTCGACCTGCAAGCACTACGCCGGC
TACGACTTTGAGGACTGGAACGGCACGACGCGGCACGACTTCGACGCCGTCATC

TCGGCGCAGGACCTGGCCGAGTACTACCTGGCGCCGTTCCAGCAGTGCGCGCGC
GACTCGCGCGTCGGCTCCGTCATGTGCGCCTACAACGCCGTCAACGGGGTGCCGT
CGTGCGCCAACTCGTACCTCATGAACACGATCCTGCGCGGGCACTGGAACTGGA
CCGAGCACGACAACTACGTCACCAGCGACTGCGAGGCCGTCCTCGACGTCTCGG
CCCACCACCACTACGCCGACACCAACGCCGAGGGCACCGGCCTCTGCTTCGAGG
CCGGCATGGACACGAGCTGCGAGTACGAGGGCTCCTCCGACATCCCGGGCGCCT
CCGCCGGCGGCTTCCTGACCTGGCCCGCCGTCGACCGCGCCCTGACGCGGCTGTA
CCGGAGCCTGGTGCGGGTCGGCTACTTTGACGGCCCCGAGTCGCCGCACGCCTCG
CTGGGCTGGGCCGACGTCAACCGGCCCGAGGCGCAGGAGCTGGCCCTGCGCGCT
GCCGTCGAGGGCATCGTGCTGCTCAAGAACGACAACGACACGCTGCCGCTGCCG
CTGCCGGACGATGTCGTTGTCACCGCTGATGGTGGCCGCCGCCGCGTCGCCATGA
TCGGCTTCTGGGCCGACGCCCCGGACAAGCTGTTTGGCGGGTACAGCGGCGCGC
CCCCCTTCGCGCGCTCGCCCGCGAGCGCCGCCCGGCAGCTGGGCTGGAACGTCAC
GGTCGCCGGAGGGCCCGTCCTGGAGGGAGACTCGGACGAGGAGGAGGACACGT
GGACGGCGCCGGCCGTCGAGGCGGCCGCCGACGCCGACTACATCGTCTACTTTG
GCGGCCTGGACACGTCGGCGGCGGGCGAGACCAAGGACCGGATGACGATCGGGT
GGCCGGCGGCGCAGCTGGCGCTCATCTCGGAGCTGGCGCGGCTCGGCAAGCCCG
TCGTGGTGGTGCAGATGGGCGACCAGCTCGACGACACGCCCCTCTTCGAGCTGG
ACGGGGTGGGCGCCGTCCTGTGGGCCAACTGGCCGGGCCAGGACGGCGGCACGG
CCGTGGTCCGGCTGCTCAGCGGCGCCGAGAGCCCGGCCGGCCGCCTGCCCGTGA
CCCAGTACCCGGCCAACTACACCGACGCGGTGCCCCTGACCGACATGACCCTGC
GCCCGTCGGCGACCAACCCGGGCCGGACCTACCGCTGGTACCCGACTCCCGTCCG
GCCCTTCGGCTTCGGCCTCCACTATACCACCTTCCGGGCCGAGTTCGGCCCCCAC
CCCTTCTTCCCGGGGGCGGGCAAGGGCGATGGCGACGGCGAGGACAAGGGCGAG
AGCAAGAGCGAGATCAGGACGCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCG
CAGGGCGGCGGCGGCGGCCACCACGCCGATCCGGGACCTGCTCCGCGACTGCGA
CAAGACGTACCCGGACACGTGCCCGCTGCCGCCGCTGACGGTGCGCGTGACCAA
CGAGGGCGAGCGCGCGTCCGACTACGTGGTGCTGGCCTTCGTGTCGGGCGAGTA
CGGGCCGGCGCCGTACCCGATCAAGACGCTGGTCTCGTACGCGCGGGCGCGCGG
GCTAAAGGGGAAGGGCGGCGACGGCGACGGCGACGGCGACGGCGCCACCACTA
CCGTCTCGCTCGACTGGACCGTCGGCAACCTGGCCCGCCACGACGAGCGCGGCA
ACACAATCCTGTACCCGGGAACTTACACCCTCACTCTCGACGAGCCGGCCCAGGC
GAGCGTGCAGTTCGCCCTCGAGGGCGAGCCCGTCGTGCTCGACGAGTGGCCTGC
GCCGCCGAGTGCCAACTCCACCGCCAGGGGGAGGCACAGG (SEQ ID NO:58)

MKASVSCLVGMSAVAYGLDGPFQTYPDCTKPPLSDIKVCDRTLPEAERAAALVAAL
TDEEKLQNLVSKAPGAPRIGLPAYNWWSEALHGVAHAPGTQFRDGPGDFNSSTSFP
MPLLMAAAFDDELIEAVGDVIGTEARAFGNAGWSGLDYWTPNVNPFRDPRWGRGS
ETPGEDVVRLKRYAASMIRGLEGRSSSSSSCSFGSGGEPPRVISTCKHYAGYDFEDWN
GTTRHDFDAVISAQDLAEYYLAPFQQCARDSRVGSVMCAYNAVNGVPSCANSYLM
NTILRGHWNWTEHDNYVTSDCEAVLDVSAHHHYADTNAEGTGLCFEAGMDTSCEY
EGSSDIPGASAGGFLTWPAVDRALTRLYRSLVRVGYFDGPESPHASLGWADVNRPE
AQELALRAAVEGIVLLKNDNDTLPLPLPDDVVVTADGGRRRVAMIGFWADAPDKLF
GGYSGAPPFARSPASAARQLGWNVTVAGGPVLEGDSDEEEDTWTAPAVEAAADAD
YIVYFGGLDTSAAGETKDRMTIGWPAAQLALISELARLGKPVVVVQMGDQLDDTPL
FELDGVGAVLWANWPGQDGGTAVVRLLSGAESPAGRLPVTQYPANYTDAVPLTDM
TLRPSATNPGRTYRWYPTPVRPFGFGLHYTTFRAEFGPHPFFPGAGKGDGDGEDKGE
SKSEIRTQQQQQQQQQRRAAAAATTPIRDLLRDCDKTYPDTCPLPPLTVRVTNEGE
RASDYVVLAFVSGEYGPAPYPIKTLVSYARARGLKGKGGDGDGDGDGATTTVSLD

WTVGNLARHDERGNTILYPGTYTLTLDEPAQASVQFALEGEPVVLDEWPAPPSANST
ARGRHR (SEQ ID NO:59)

[0234] The following sequences comprise additional xylanase (Xyl), beta-xylosidase (Bxyl), and alpha-xylosidase (Axyl) sequences of interest. The first sequence provided in each set below comprises the cDNA sequence, the second sequence is the polypeptide sequence with the predicted signal sequence included and the third sequence is the polypeptide sequence without the signal sequence.

Xyl1974:

ATGGTTGCTCTCTCTTCTCTCCTCGTCGCTGCCTCTGCGGCGGCCGTGGCCGTGGC
TGCGCCGAGCGAGGCCCTCCAGAAGCGCCAGACGCTCACGAGCAGCCAGACGGG
CTTCCACGACGGCTTTTACTACTCCTTCTGGACCGACGGTGCCGGCAACGTCCGG
TACACGAACGAGGCCGGCGGCCGGTACAGTGTCACCTGGTCCGGCAACAACGGC
AACTGGGTTGGCGGCAAGGGCTGGAACCCGGGGGCTGCTCGCAACATCAGCTTC
ACGGGGCAGTATAACCCCAACGGCAACTCGTACCTGGCCGTGTACGGGTGGACG
CGCAACCCGCTGATCGAGTACTACATCGTCGAGAACTTCGGCACGTACGACCCGT
CGACGGGGGCGCAGCGGCTCGGCAGCATCACGGTGGACGGGTCGACGTACAACA
TCCTCAAGACGACGCGGGTCAACCAGCCGTCCATCGAGGGCACCAGCACCTTTG
ACCAGTTCTGGTCCGTCCGGACCAACAAGCGCAGCAGCGGCTCCGTCAACGTCA
AGGCTCACTTCGACGCTTGGGCCCAGGCCGGCCTCCGCCTGGGCACCCACGACTA
CCAGATCATGGCCACCGAGGGCTACTTCTCGAGCGGCTCCGCCACCATCACCGTC
GGCGAGGGCACCAGCAGCGGCGGCGGCGGCGACAATGGCGGCGGCAACAACGG
CGGCGGCGGCAACACCGGCACCTGCAGCGCCCTGTACGGCCAGTGCGGTGGCCA
GGGGTGGACGGGCCCGACTTGCTGCTCCCAGGGAACCTGCCGCGTCTCCAACCA
GTGGTACTCGCAGTGCTTGTAA (SEQ ID NO:60)

MVALSSLLVAASAAAVAVAAPSEALQKRQTLTSSQTGFHDGFYYSFWTDGAGNVR
YTNEAGGRYSVTWSGNNGNWVGGKGWNPGAARNISFTGQYNPNGNSYLAVYGWT
RNPLIEYYIVENFGTYDPSTGAQRLGSITVDGSTYNILKTTRVNQPSIEGTSTFDQFWS
VRTNKRSSGSVNVKAHFDAWAQAGLRLGTHDYQIMATEGYFSSGSATITVGEGTSS
GGGGDNGGGNNGGGGNTGTCSALYGQCGGQGWTGPTCCSQGTCRVSNQWYSQCL
(SEQ ID NO:61)

APSEALQKRQTLTSSQTGFHDGFYYSFWTDGAGNVRYTNEAGGRYSVTWSGNNGN
WVGGKGWNPGAARNISFTGQYNPNGNSYLAVYGWTRNPLIEYYIVENFGTYDPSTG
AQRLGSITVDGSTYNILKTTRVNQPSIEGTSTFDQFWSVRTNKRSSGSVNVKAHFDA
WAQAGLRLGTHDYQIMATEGYFSSGSATITVGEGTSSGGGGDNGGGNNGGGGNTG
TCSALYGQCGGQGWTGPTCCSQGTCRVSNQWYSQCL (SEQ ID NO:62)

Xyl40741:

ATGAAGGCCAATCTCCTGGTCCTCGCGCCGCTGGCCGTCTCGGCAGCGCCCGCGC
TCGAGCACCGCCAGGCAACTGAGAGCATCGACGCGCTCATTAAGGCCAAGGGCA
AGCTCTACTTTGGCACCTGTACCGACCAGGGCCGGCTGACGTCGGGCAAGAACG

CGGACATCATCAGGGCCAACTTCGGCCAGGTGACGCCCGAGAACAGCATGAAGT
GGCAGAGCATCGAGCCATCGCGGGGTCAGTTCACCTGGGGCCAGGCTGACTACC
TCGTCGACTGGGCCACTCAGAACAACAAGACCATCCGCGGCCACACGCTCGTCT
GGCACTCGCAGCTCGCCGGCTACGTTCAGCAGATCGGCGACCGGAACACCTTGA
CCCAGACCATCCAGGACCACATTGCCGCCGTCATGGGCCGCTACAAGGGCAAGA
TCTACGCCTGGGATGTCATCAACGAGATGTTCAACGAGGATGGCTCGCTTCGCAG
CAGCGTCTTCTCCAACGTCCTCGGAGAGGACTTTGTTGGGATCGCCTTCAAGGCG
GCGCGCGAGGCCGACCCCGACACCAAGTTGTACATCAACGACTACAACCTCGAC
AGCCCCAACTACGCCAAGCTGACCAACGGCATGGTCGCTCACGTCAAGAAGTGG
CTCGCGGCCGGCATCCCCATCGACGGCATCGGCACCCAGGGTCACCTGCAGTCTG
GCCAGGGTTCCGGTCTTGCGCAGGCCATCAAGGCTCTCGCCCAGGCTGGCGTCGA
GGAGGTTGCCGTCACCGAGCTCGATATCCAGAACCAGAACACCAACGACTACAC
TGCCGTTGTCCAGGGCTGCTTGGACGAGCCCAAGTGCGTCGGTATCACCGTCTGG
GGTGTCCGCGATCCCGACTCGTGGCGTCCCCAGGGCAACCCCTTGCTCTTCGACA
GCAACTTCAACCCCAAGGCGAACTACAATGCCATCGTCCAGCTCCTCAAGCAGTA
G (SEQ ID NO:63)

MKANLLVLAPLAVSAAPALEHRQATESIDALIKAKGKLYFGTCTDQGRLTSGKNADI
IRANFGQVTPENSMKWQSIEPSRGQFTWGQADYLVDWATQNNKTIRGHTLVWHSQ
LAGYVQQIGDRNTLTQTIQDHIAAVMGRYKGKIYAWDVINEMFNEDGSLRSSVFSN
VLGEDFVGIAFKAAREADPDTKLYINDYNLDSPNYAKLTNGMVAHVKKWLAAGIPI
DGIGTQGHLQSGQGSGLAQAIKALAQAGVEEVAVTELDIQNQNTNDYTAVVQGCLD
EPKCVGITVWGVRDPDSWRPQGNPLLFDSNFNPKANYNAIVQLLKQ (SEQ ID
NO:64)

APALEHRQATESIDALIKAKGKLYFGTCTDQGRLTSGKNADIIRANFGQVTPENSMK
WQSIEPSRGQFTWGQADYLVDWATQNNKTIRGHTLVWHSQLAGYVQQIGDRNTLT
QTIQDHIAAVMGRYKGKIYAWDVINEMFNEDGSLRSSVFSNVLGEDFVGIAFKAARE
ADPDTKLYINDYNLDSPNYAKLTNGMVAHVKKWLAAGIPIDGIGTQGHLQSGQGSG
LAQAIKALAQAGVEEVAVTELDIQNQNTNDYTAVVQGCLDEPKCVGITVWGVRDP
DSWRPQGNPLLFDSNFNPKANYNAIVQLLKQ (SEQ ID NO:65)

Xyl34208:

ATGGTCAAGCTCTCTCTCATCGCAGCGAGCCTTGTGGCACCTAGCGTGCTTGCGG
GTCCTCTCATCGGCCCCAAGACGCAAACCGAGAGCCAGCTGAACCCGCGTCAAG
GCGGCTACAACTACTTCCAGAATTGGTCCGAGGGAGGCAGCAATATCCGCTGCA
ACAACGGCCCTGGGGGTTCCTACACGGCCGACTGGAACAGCAGGGGCGGCTTCG
TCTGTGGCAAGGGCTGGAGCTATGGAGGCAATCGCGCCATCACGTACACCGGCG
AATACAACGCCAGCGGCCCCGGCTACCTCGCCGTCTACGGGTGGACCCGCAACC
CGCTGATTGAATACTACATCATCGAGGCCCATGCCGACCTCGCCCCCAACGAGCC
GTGGACATCCAAGGGTAATTTCAGCTTCGAGGAGGGCGAGTACGAGGTCTTCAC
CAGCACCCGCGTCAACAAGCCGTCCATCGAGGGCACCAGGACTTTTCAGCAGTA
CTGGTCGCTGCGCAAGGAGCAGCGGGTCGGCGGCACCGTCACCACCCAGAGGCA
CTTTGAAGAGTGGGCCAAGCTGGGCATGAAGCTGGGCAATCATGACTATGTCAT
CCTGGCGACCGAAGGATACACTGCCAACGGAGGATCCGGTAGCAGCGGGCACTC
GAGCATTACTCTGCAGTAG (SEQ ID NO:66)

MVKLSLIAASLVAPSVLAGPLIGPKTQTESQLNPRQGGYNYFQNWSEGGSNIRCNNG
PGGSYTADWNSRGGFVCGKGWSYGGNRAITYTGEYNASGPGYLAVYGWTRNPLIE
YYIIEAHADLAPNEPWTSKGNFSFEEGEYEVFTSTRVNKPSIEGTRTFQQYWSLRKEQ
RVGGTVTTQRHFEEWAKLGMKLGNHDYVILATEGYTANGGSGSSGHSSITLQ (SEQ
ID NO:67)

GPLIGPKTQTESQLNPRQGGYNYFQNWSEGGSNIRCNNGPGGSYTADWNSRGGFVC
GKGWSYGGNRAITYTGEYNASGPGYLAVYGWTRNPLIEYYIIEAHADLAPNEPWTS
KGNFSFEEGEYEVFTSTRVNKPSIEGTRTFQQYWSLRKEQRVGGTVTTQRHFEEWAK
LGMKLGNHDYVILATEGYTANGGSGSSGHSSITLQ (SEQ ID NO:68)

Xyl7143:

ATGGTCTCGTTCACTCTCCTCCTCACGGTCATCGCCGCTGCGGTGACGACGGCCA
GCCCTCTCGAGGTGGTCAAGCGCGGCATCCAGCCGGGCACGGGCACCCACGAGG
GGTACTTCTACTCGTTCTGGACCGACGGCCGTGGCTCGGTCGACTTCAACCCCGG
GCCCCGCGGCTCGTACAGCGTCACCTGGAACAACGTCAACAACTGGGTTGGCGG
CAAGGGCTGGAACCCGGGCCCGCCGCGCAAGATTGCGTACAACGGCACCTGGAA
CAACTACAACGTGAACAGCTACCTCGCCCTGTACGGCTGGACTCGCAACCCGCTG
GTCGAGTATTACATCGTGGAGGCATACGGCACGTACAACCCCTCGTCGGGCACG
GCGCGGCTGGGCACCATCGAGGACGACGGCGGCGTGTACGACATCTACAAGACG
ACGCGGTACAACCAGCCGTCCATCGAGGGGACCTCCACCTTCGACCAGTACTGGT
CCGTCCGCCGCCAGAAGCGCGTCGGCGGCACTATCGACACGGGCAAGCACTTTG
ACGAGTGGAAGCGCCAGGGCAACCTCCAGCTCGGCACCTGGAACTACATGATCA
TGGCCACCGAGGGCTACCAGAGCTCTGGTTCGGCCACTATCGAGGTCCGGGAGG
CCTAA (SEQ ID NO:69)

MVSFTLLLTVIAAAVTTASPLEVVKRGIQPGTGTHEGYFYSFWTDGRGSVDFNPGPR
GSYSVTWNNVNNWVGGKGWNPGPPRKIAYNGTWNNYNVNSYLALYGWTRNPLV
EYYIVEAYGTYNPSSGTARLGTIEDDGGVYDIYKTTRYNQPSIEGTSTFDQYWSVRR
QKRVGGTIDTGKHFDEWKRQGNLQLGTWNYMIMATEGYQSSGSATIEVREA (SEQ
ID NO:70)

SPLEVVKRGIQPGTGTHEGYFYSFWTDGRGSVDFNPGPRGSYSVTWNNVNNWVGG
KGWNPGPPRKIAYNGTWNNYNVNSYLALYGWTRNPLVEYYIVEAYGTYNPSSGTA
RLGTIEDDGGVYDIYKTTRYNQPSIEGTSTFDQYWSVRRQKRVGGTIDTGKHFDEWK
RQGNLQLGTWNYMIMATEGYQSSGSATIEVREA (SEQ ID NO:71)

Xyl42827:

ATGGTCTCGCTCAAGTCCCTCCTCCTCGCCGCGGCGGCGACGTTGACGGCGGTGA
CGGCGCGCCCGTTCGACTTTGACGACGGCAACTCGACCGAGGCGCTGGCCAAGC
GCCAGGTCACGCCCAACGCGCAGGGCTACCACTCGGGCTACTTCTACTCGTGGTG
GTCCGACGGCGGCGGCCAGGCCACCTTCACCCTGCTCGAGGGCAGCCACTACCA
GGTCAACTGGAGGAACACGGGCAACTTTGTCGGTGGCAAGGGCTGGAACCCGGG
TACCGGCCGGACCATCAACTACGGCGGCTCGTTCAACCCGAGCGGCAACGGCTA
CCTGGCCGTCTACGGCTGGACGCACAACCCGCTGATCGAGTACTACGTGGTCGAG

TCGTACGGGACCTACAACCCGGGCAGCCAGGCCCAGTACAAGGGCAGCTTCCAG
AGCGACGGCGGCACCTACAACATCTACGTCTCGACCCGCTACAACGCGCCCTCG
ATCGAGGGCACCCGCACCTTCCAGCAGTACTGGTCCATCCGCACCTCCAAGCGCG
TCGGCGGCTCCGTCACCATGCAGAACCACTTCAACGCCTGGGCCCAGCACGGCAT
GCCCCTCGGCTCCCACGACTACCAGATCGTCGCCACCGAGGGCTACCAGAGCAG
CGGCTCCTCCGACATCTACGTCCAGACTCACTAG (SEQ ID NO:72)

MVSLKSLLLAAAATLTAVTARPFDFDDGNSTEALAKRQVTPNAQGYHSGYFYSWW
SDGGGQATFTLLEGSHYQVNWRNTGNFVGGKGWNPGTGRTINYGGSFNPSGNGYL
AVYGWTHNPLIEYYVVESYGTYNPGSQAQYKGSFQSDGGTYNIYVSTRYNAPSIEGT
RTFQQYWSIRTSKRVGGSVTMQNHFNAWAQHGMPLGSHDYQIVATEGYQSSGSSDI
YVQTH (SEQ ID NO:73)

RPFDFDDGNSTEALAKRQVTPNAQGYHSGYFYSWWSDGGGQATFTLLEGSHYQVN
WRNTGNFVGGKGWNPGTGRTINYGGSFNPSGNGYLAVYGWTHNPLIEYYVVESYG
TYNPGSQAQYKGSFQSDGGTYNIYVSTRYNAPSIEGTRTFQQYWSIRTSKRVGGSVT
MQNHFNAWAQHGMPLGSHDYQIVATEGYQSSGSSDIYVQTH (SEQ ID NO:74)

BXyl1883:

ATGGCCTTCCTTTCCTCCTTTGCCCTTGCCGCCCTCGGGGCACTCGTCGTCCCGGC
GAGGGGCGGCGTGACGTACCCGGACTGCGCAAACGGACCGCTCAAGTCAAATAC
GGTGTGCGATACGTCGGCGTCCCCGGGAGCCCGAGCCGCTGCTCTTGTGAGTGTA
ATGAACAACAACGAAAAACTTGCAAATCTTGTCAACAATTCGCCCGGCGTCTCGC
GGCTCGGCCTGAGTGCGTACCAGTGGTGGAACGAAGCCCTCCACGGAGTAGCCC
ATAACCGCGGCATTACCTGGGGCGGCGAGTTCAGCGCGGCAACCCAGTTCCCGC
AGGCTATCACGACTTCCGCCACTTTCGATGACGCTTTGATCGAGCAAATCGGCAC
CATTATCAGCACCGAGGCCCGTGCCTTTGCCAACAATGGGCGCGCTCATCTCGAC
TTCTGGACGCCCAACGTCAACCCGTTTCGAGACCCGCGATGGGGTCGCGGACAC
GAGACGCCGGGAGAGGATGCATTCAAGAATAAGAAGTGGGCCGAGGCCTTCGTC
AAGGGCATGCAAGGACCCGGACCGACGCACCGAGTCATCGCCACATGTAAGCAC
TACGCCGCCTACGACCTCGAGAACTCCGGGAGCACGACCCGATTCAACTTCGATG
CGAAGGTGTCAACTCAAGATCTCGCCGAGTACTATCTCCCTCCGTTCCAACAGTG
CGCCCGGGACTCTAAGGTGGGCTCCATCATGTGCAGCTACAATGCGGTCAATGA
AATCCCGGCCTGCGCGAATCCTTACCTGATGGATACCATCCTGCGGAAACATTGG
AATTGGACCGACGAGCACCAGTATATTGTGAGCGACTGCGATGCCGTGTACTATC
TCGGCAATGCGAACGGCGGCCACCGATACAAGCCGAGCTATGCGGCGGCGATCG
GAGCATCTCTCGAGGCTGGTTGCGATAACATGTGCTGGGCGACCGGCGGCACCG
CCCCGGATCCCGCCTCAGCCTTCAATTCCGGCCAGTTCAGCCAGACGACACTGGA
CACGGCTATTTTGCGCCAGATGCAGGGCCTCGTCCTAGCGGGATACTTTGACGGT
CCGGGCGGTATGTACCGCAACCTGAGCGTGGCGGACGTGAACACGCAGACCGCC
CAGGACACTGCACTCAAGGCGGCGGAAGGAGGCATCGTGCTCCTCAAGAACGAT
GGGATCCTTCCGCTGTCGGTTAACGGTTCCAATTTCCAGGTCGCTATGATCGGGT
TCTGGGCGAACGCAGCCGACAAGATGCTCGGGGGTTACAGCGGGAGCCCGCCGT
TCAACCATGATCCCGTGACCGCTGCAAGATCGATGGGCATCACGGTCAACTACGT
CAACGGGCCATTGACGCAACCCAACGGGGATACGTCGGCAGCACTCAATGCGGC
CCAAAAGTCCAACGCGGTGGTATTCTTTGGTGGAATCGACAATACGGTGGAGAA
GGAGAGTCAGGACAGAACGTCCATCGAGTGGCCCTCAGGGCAACTGGCTCTGAT

TCGGAGGCTAGCCGAAACCGGCAAACCAGTCATCGTCGTCAGGCTCGGGACGCA
CGTCGACGACACCCCGCTCCTCAGCATTCCGAATGTGAGAGCCATTTTGTGGGCA
GGATACCCGGGTCAAGACGGCGGGACTGCTGTGGTGAAAATCATTACCGGCCTT
GCTAGTCCGGCGGGGAGGCTGCCCGCCACTGTGTATCCGTCTTCGTACACCAGCC
AAGCGCCCTTTACAAACATGGCCCTGAGGCCTTCTTCGTCCTATCCCGGGCGAAC
ATACCGCTGGTACAGTAACGCCGTCTTTCCATTTGGCCACGGCCTACATTATACC
AATTTCAGTGTCTCGGTGCGGGACTTTCCGGCCAGCTTCGCGATTGCCGATCTCCT
GGCTTCCTGCGGGGATTCCGTGGCGTATCTTGATCTTTGCCCCTTCCCGTCCGTGT
CGCTCAATGTGACCAATACAGGCACCCGCGTGTCCGATTACGTTGCGCTTGGGTT
CTTGTCGGGAGATTTTGGTCCCAGCCCACATCCCATCAAGACATTGGCGACGTAT
AAGCGCGTGTTTAACATCGAACCTGGGGAAACACAGGTGGCCGAGCTAGACTGG
AAGCTGGAGAGCCTGGTCCGGGTAGATGAGAAGGGCAACAGGGTACTCTACCCC
GGAACATATACGCTTCTTGTGGATCAGCCAACCTTGGCAAATATCACCTTTATTTT
GACAGGAGAAGAGGCAGTGTTGGATAGTTGGCCGCAGCCGTGA (SEQ ID NO:75)

MAFLSSFALAALGALVVPARGGVTYPDCANGPLKSNTVCDTSASPGARAAALVSVM
NNNEKLANLVNNSPGVSRLGLSAYQWWNEALHGVAHNRGITWGGEFSAATQFPQA
ITTSATFDDALIEQIGTIISTEARAFANNGRAHLDFWTPNVNPFRDPRWGRGHETPGE
DAFKNKKWAEAFVKGMQGPGPTHRVIATCKHYAAYDLENSGSTTRFNFDAKVSTQ
DLAEYYLPPFQQCARDSKVGSIMCSYNAVNEIPACANPYLMDTILRKHWNWTDEHQ
YIVSDCDAVYYLGNANGGHRYKPSYAAAIGASLEAGCDNMCWATGGTAPDPASAF
NSGQFSQTTLDTAILRQMQGLVLAGYFDGPGGMYRNLSVADVNTQTAQDTALKAA
EGGIVLLKNDGILPLSVNGSNFQVAMIGFWANAADKMLGGYSGSPPFNHDPVTAAR
SMGITVNYVNGPLTQPNGDTSAALNAAQKSNAVVFFGGIDNTVEKESQDRTSIEWPS
GQLALIRRLAETGKPVIVVRLGTHVDDTPLLSIPNVRAILWAGYPGQDGGTAVVKIIT
GLASPAGRLPATVYPSSYTSQAPFTNMALRPSSSYPGRTYRWYSNAVFPFGHGLHYT
NFSVSVRDFPASFAIADLLASCGDSVAYLDLCPFPSVSLNVTNTGTRVSDYVALGFLS
GDFGPSPHPIKTLATYKRVFNIEPGETQVAELDWKLESLVRVDEKGNRVLYPGTYTL
LVDQPTLANITFILTGEEAVLDSWPQP (SEQ ID NO:76)

GVTYPDCANGPLKSNTVCDTSASPGARAAALVSVMNNNEKLANLVNNSPGVSRLGL
SAYQWWNEALHGVAHNRGITWGGEFSAATQFPQAITTSATFDDALIEQIGTIISTEAR
AFANNGRAHLDFWTPNVNPFRDPRWGRGHETPGEDAFKNKKWAEAFVKGMQGPG
PTHRVIATCKHYAAYDLENSGSTTRFNFDAKVSTQDLAEYYLPPFQQCARDSKVGSI
MCSYNAVNEIPACANPYLMDTILRKHWNWTDEHQYIVSDCDAVYYLGNANGGHRY
KPSYAAAIGASLEAGCDNMCWATGGTAPDPASAFNSGQFSQTTLDTAILRQMQGLV
LAGYFDGPGGMYRNLSVADVNTQTAQDTALKAAEGGIVLLKNDGILPLSVNGSNFQ
VAMIGFWANAADKMLGGYSGSPPFNHDPVTAARSMGITVNYVNGPLTQPNGDTSA
ALNAAQKSNAVVFFGGIDNTVEKESQDRTSIEWPSGQLALIRRLAETGKPVIVVRLGT
HVDDTPLLSIPNVRAILWAGYPGQDGGTAVVKIITGLASPAGRLPATVYPSSYTSQAP
FTNMALRPSSSYPGRTYRWYSNAVFPFGHGLHYTNFSVSVRDFPASFAIADLLASCG
DSVAYLDLCPFPSVSLNVTNTGTRVSDYVALGFLSGDFGPSPHPIKTLATYKRVFNIE
PGETQVAELDWKLESLVRVDEKGNRVLYPGTYTLLVDQPTLANITFILTGEEAVLDS
WPQP (SEQ ID NO:77)

Xyl25453:

ATGCGTACTCTTACGTTCGTGCTGGCAGCCGCCCCGGTGGCTGTGCTTGCCCAAT
CTCCTCTGTGGGGCCAGTGCGGCGGTCAAGGCTGGACAGGTCCCACGACCTGCGT
TTCTGGCGCAGTATGCCAATTCGTCAATGACTGGTACTCCCAATGCGTGCCCGGA
TCGAGCAACCCTCCTACGGGCACCACCAGCAGCACCACTGGAAGCACCCCGGCT
CCTACTGGCGGCGGCGGCAGCGGAACCGGCCTCCACGACAAATTCAAGGCCAAG
GGCAAGCTCTACTTCGGAACCGAGATCGATCACTACCATCTCAACAACAATGCCT
TGACCAACATTGTCAAGAAAGACTTTGGTCAAGTCACTCACGAGAACAGCTTGA
AGTGGGATGCTACTGAGCCGAGCCGCAATCAATTCAACTTTGCCAACGCCGACG
CGGTTGTCAACTTTGCCCAGGCCAACGGCAAGCTCATCCGCGGCCACACCCTCCT
CTGGCACTCTCAGCTGCCGCAGTGGGTGCAGAACATCAACGACCGCAACACCTT
GACCCAGGTCATCGAGAACCACGTCACCACCCTTGTCACTCGCTACAAGGGCAA
GATCCTCCACTGGGACGTCGTTAACGAGATCTTTGCCGAGGACGGCTCGCTCCGC
GACAGCGTCTTCAGCCGCGTCCTCGGCGAGGACTTTGTCGGCATCGCCTTCCGCG
CCGCCCGCGCCGCCGATCCCAACGCCAAGCTCTACATCAACGACTACAACCTCGA
CATTGCCAACTACGCCAAGGTGACCCGGGGCATGGTCGAGAAGGTCAACAAGTG
GATCGCCCAGGGCATCCCGATCGACGGCATCGGCACCCAGTGCCACCTGGCCGG
GCCCGGCGGGTGGAACACGGCCGCCGGCGTCCCCGACGCCCTCAAGGCCCTCGC
CGCGGCCAACGTCAAGGAGATCGCCATCACCGAGCTCGACATCGCCGGCGCCTC
CGCCAACGACTACCTCACCGTCATGAACGCCTGCCTCCAGGTCTCCAAGTGCGTC
GGCATCACCGTCTGGGGCGTCTCTGACAAGGACAGCTGGAGGTCGAGCAGCAAC
CCGCTCCTCTTCGACAGCAACTACCAGCCAAAGGCGGCATACAATGCTCTGATTA
ATGCCTTGTAA (SEQ ID NO:78)

MRTLTFVLAAAPVAVLAQSPLWGQCGGQGWTGPTTCVSGAVCQFVNDWYSQCVP
GSSNPPTGTTSSTTGSTPAPTGGGGSGTGLHDKFKAKGKLYFGTEIDHYHLNNNALT
NIVKKDFGQVTHENSLKWDATEPSRNQFNFANADAVVNFAQANGKLIRGHTLLWH
SQLPQWVQNINDRNTLTQVIENHVTTLVTRYKGKILHWDVVNEIFAEDGSLRDSVFS
RVLGEDFVGIAFRAARAADPNAKLYINDYNLDIANYAKVTRGMVEKVNKWIAQGIP
IDGIGTQCHLAGPGGWNTAAGVPDALKALAAANVKEIAITELDIAGASANDYLTVM
NACLQVSKCVGITVWGVSDKDSWRSSSNPLLFDSNYQPKAAYNALINAL (SEQ ID
NO:79)

QSPLWGQCGGQGWTGPTTCVSGAVCQFVNDWYSQCVPGSSNPPTGTTSSTTGSTPA
PTGGGGSGTGLHDKFKAKGKLYFGTEIDHYHLNNNALTNIVKKDFGQVTHENSLKW
DATEPSRNQFNFANADAVVNFAQANGKLIRGHTLLWHSQLPQWVQNINDRNTLTQ
VIENHVTTLVTRYKGKILHWDVVNEIFAEDGSLRDSVFSRVLGEDFVGIAFRAARAA
DPNAKLYINDYNLDIANYAKVTRGMVEKVNKWIAQGIPIDGIGTQCHLAGPGGWNT
AAGVPDALKALAAANVKEIAITELDIAGASANDYLTVMNACLQVSKCVGITVWGVS
DKDSWRSSSNPLLFDSNYQPKAAYNALINAL (SEQ ID NO:80)

Xyl805:

ATGCATCTCTCCTCGTCTCTCCTCCTCCTCGCCGCCTTGCCCCTGGGCATCGCCGG
CAAGGGCAAGGGCCACGGCCACGGCCCCCATACCGGGCTCCACACCCTCGCCAA
GCAGGCCGGCCTCAAGTACTTCGGCTCTGCCACCGACTCTCCCGGCCAGCGTGAG
CGCGCCGGCTACGAGGACAAGTACGCCCAGTACGACCAGATCATGTGGAAGTCG

GGCGAGTTCGGCCTGACGACCCCGACCAACGGCCAAAAGTGGCTGTTTACTGAG
CCCGAGCGTGGCGTGTTCAACTTCACCGAGGGTGACATCGTGACGAACCTGGCCC
GGAAGCACGGTTTCATGCAGCGGTGCCACGCGCTCGTCTGGCACAGCCAGCTCG
CCCCTTGGGTCGAGTCGACCGAGTGGACGCCCGAGGAGCTGCGCCAGGTCATTG
TCAACCACATCACCCACGTGGCCGGCTACTACAAGGGCAAGTGCTATGCCTGGG
ACGTCGTCAACGAGGCCCTGAACGAGGACGGCACCTACCGCGAGTCCGTCTTCT
ACAAGGTGCTCGGCGAGGACTACATCAAGCTGGCCTTCGAGACGGCCGCCAAGG
TCGACCCCCACGCCAAGCTCTACTACAACGACTACAACCTCGAGTCCCCCAGCGC
CAAGACCGAGGGCGCCAAGCGCATCGTCAAGATGCTCAAGGACGCCGGCATCCG
CATCGACGGCGTCGGCCTGCAGGCCCACCTCGTCGCCGAGAGCCACCCGACCCTC
GACGAGCACATCGATGCCATCAAGGGCTTCACCGAGCTCGGCGTCGAGGTCGCC
CTGACCGAGCTCGACATCCGCCTCTCCATCCCGGCCAACGCCACCAACCTCGCCC
AGCAGAGGGAGGCGTACAAGAACGTCGTCGGCGCTTGCGTCCAGGTTCGCGGCT
GCATTGGCGTGGAGATCTGGGACTTCTATGACCCCTTCAGCTGGGTCCCTGCCAC
CTTCCCCGGCCAGGGCGCCCCCCTGCTCTGGTTCGAGGACTTTTCCAAGCACCCC
GCCTACGACGGCGTCGTCGAGGCCCTGACCAACAGGACCACGGGCGGGTGCAAG
GGCAAGGGCAAGGGCAAGGGCAAGGTTTGGAAGGCCTAA (SEQ ID NO:81)

MHLSSSLLLLAALPLGIAGKGKGHGHGPHTGLHTLAKQAGLKYFGSATDSPGQRER
AGYEDKYAQYDQIMWKSGEFGLTTPTNGQKWLFTEPERGVFNFTEGDIVTNLARKH
GFMQRCHALVWHSQLAPWVESTEWTPEELRQVIVNHITHVAGYYKGKCYAWDVV
NEALNEDGTYRESVFYKVLGEDYIKLAFETAAKVDPHAKLYYNDYNLESPSAKTEG
AKRIVKMLKDAGIRIDGVGLQAHLVAESHPTLDEHIDAIKGFTELGVEVALTELDIRL
SIPANATNLAQQREAYKNVVGACVQVRGCIGVEIWDFYDPFSWVPATFPGQGAPLL
WFEDFSKHPAYDGVVEALTNRTTGGCKGKGKGKGKVWKA (SEQ ID NO:82)

KGKGHGHGPHTGLHTLAKQAGLKYFGSATDSPGQRERAGYEDKYAQYDQIMWKS
GEFGLTTPTNGQKWLFTEPERGVFNFTEGDIVTNLARKHGFMQRCHALVWHSQLAP
WVESTEWTPEELRQVIVNHITHVAGYYKGKCYAWDVVNEALNEDGTYRESVFYKV
LGEDYIKLAFETAAKVDPHAKLYYNDYNLESPSAKTEGAKRIVKMLKDAGIRIDGVG
LQAHLVAESHPTLDEHIDAIKGFTELGVEVALTELDIRLSIPANATNLAQQREAYKNV
VGACVQVRGCIGVEIWDFYDPFSWVPATFPGQGAPLLWFEDFSKHPAYDGVVEALT
NRTTGGCKGKGKGKGKVWKA (SEQ ID NO:83)

Xyl36882:

ATGCACTCCAAAGCTTTCTTGGCAGCGCTTCTTGCGCCTGCCGTCTCAGGGCAAC
TGAACGACCTCGCCGTCAGGGCTGGACTCAAGTACTTTGGTACTGCTCTTAGCGA
GAGCGTCATCAACAGTGATACTCGGTATGCTGCCATCCTCAGCGACAAGAGCAT
GTTCGGCCAGCTCGTCCCCGAGAATGGCATGAAGTGGGATGCTACTGAGCCGTCC
CGTGGCCAGTTCAACTACGCCTCGGGCGACATCACGGCCAACACGGCCAAGAAG
AATGGCCAGGGCATGCGTTGCCACACCATGGTCTGGTACAGCCAGCTCCCGAGCT
GGGTCTCCTCGGGCTCGTGGACCAGGGACTCGCTCACCTCGGTCATCGAGACGCA
CATGAACAACGTCATGGGCCACTACAAGGGCCAATGCTACGCCTGGGATGTCAT
CAACGAGGCCATCAATGACGACGGCAACTCCTGGCGCGACAACGTCTTTCTCCG
GACCTTTGGGACCGACTACTTCGCCCTGTCCTTCAACCTAGCCAAGAAGGCCGAT
CCCGATACCAAGCTGTACTACAACGACTACAACCTCGAGTACAACCAGGCCAAG
ACGGACCGCGCTGTTGAGCTCGTCAAGATGGTCCAGGCCGCCGGCGCGCCCATC
GACGGTGTCGGCTTCCAGGGCCACCTCATTGTCGGCTCGACCCCGACGCGCTCGC
AGCTGGCCACCGCCCTCCAGCGCTTCACCGCGCTCGGCCTCGAGGTCGCCTACAC
CGAGCTCGACATCCGCCACTCGAGCCTGCCGGCCTCTTCGTCGGCGCTCGCGACC
CAGGGCAACGACTTCGCCAACGTGGTCGGCTCTTGCCTCGACACCGCCGGCTGCG
TCGGCGTCACCGTCTGGGGCTTCACCGATGCGCACTCGTGGATCCCGAACACGTT
CCCCGGCCAGGGCGACGCCCTGATCTACGACAGCAACTACAACAAGAAGCCCGC
GTGGACCTCGATCTCGTCCGTCCTGGCCGCCAAGGCCACCGGCGCCCCGCCCGCC
TCGTCCTCCACCACCCTCGTCACCATCACCACCCCTCCGCCGGCATCCACCACCG
CCTCCTCCTCCTCCAGTGCCACGCCCACGAGCGTCCCGACGCAGACGAGGTGGGG
ACAGTGCGGCGGCATCGGATGGACGGGGCCGACCCAGTGCGAGAGCCCATGGAC
CTGCCAGAAGCTGAACGACTGGTACTGGCAGTGCCTGTAA (SEQ ID NO:84)

MHSKAFLAALLAPAVSGQLNDLAVRAGLKYFGTALSESVINSDTRYAAILSDKSMFG
QLVPENGMKWDATEPSRGQFNYASGDITANTAKKNGQGMRCHTMVWYSQLPSWV
SSGSWTRDSLTSVIETHMNNVMGHYKGQCYAWDVINEAINDDGNSWRDNVFLRTF
GTDYFALSFNLAKKADPDTKLYYNDYNLEYNQAKTDRAVELVKMVQAAGAPIDGV
GFQGHLIVGSTPTRSQLATALQRFTALGLEVAYTELDIRHSSLPASSSALATQGNDFA
NVVGSCLDTAGCVGVTVWGFTDAHSWIPNTFPGQGDALIYDSNYNKKPAWTSISSV
LAAKATGAPPASSSTTLVTITTPPPASTTASSSSSATPTSVPTQTRWGQCGGIGWTGPT
QCESPWTCQKLNDWYWQCL (SEQ ID NO:85)

QLNDLAVRAGLKYFGTALSESVINSDTRYAAILSDKSMFGQLVPENGMKWDATEPS
RGQFNYASGDITANTAKKNGQGMRCHTMVWYSQLPSWVSSGSWTRDSLTSVIETH
MNNVMGHYKGQCYAWDVINEAINDDGNSWRDNVFLRTFGTDYFALSFNLAKKAD
PDTKLYYNDYNLEYNQAKTDRAVELVKMVQAAGAPIDGVGFQGHLIVGSTPTRSQL
ATALQRFTALGLEVAYTELDIRHSSLPASSSALATQGNDFANVVGSCLDTAGCVGVT
VWGFTDAHSWIPNTFPGQGDALIYDSNYNKKPAWTSISSVLAAKATGAPPASSSTTL
VTITTPPPASTTASSSSSATPTSVPTQTRWGQCGGIGWTGPTQCESPWTCQKLNDWY
WQCL (SEQ ID NO:86)

Xyl5123:

ATGGTCTCCTTCAAGGCCCTCGTTCTCGGCGCCGTTGGCGCCCTCTCCTTCCCTTT
CAACGTCACCGAGCTGTCCGAGGCGCACGCCCGGGGCGAGAATGTGACCGAGCT

CTTGATGTCTCGCGCCGGCACGCCGAGCCAGACCGGCTGGCACGGGGGCTACTA
CTTCTCCTTCTGGACCGACAACGGCGGCACCGTCAACTACTGGAACGGCGACAAT
GGCAGATACGGTGTCCAGTGGCAGAACTGCGGCAACTTTGTCGGCGGTAAGGGA
TGGAACCCCGGCGCGGCGCGGACCATCAACTTCAGCGGCTCCTTCAACCCGTCGG
GCAACGGGTACCTGGCCGTGTACGGGTGGACGCAGAACCCGCTGATCGAGTACT
ACATCGTCGAGTCGTTCGGCACGTACGACCCGTCGTCGCAGGCCCAGGTCCTCGG
CACCTTCTACCAGGACGGCAGCAACTACAAGATCGCCAAGACGACCCGCTACAA
CCAGCCCTCCATCGAGGGCACCAGCACCTTCGACCAGTTCTGGTCCGTCCGCGAG
AACCACCGCACCAGCGGCAGCGTCAACGTCGGCGCCCACTTCGCCCGCTGGCAG
CAGGCCGGCCTCCGCCTCGGCACCCACAACTACCAAATCATGGCCACCGAGGGC
TACCAGAGCAGCGGCTCCTCCGATATCACCGTCTGGTAA (SEQ ID NO:87)

MVSFKALVLGAVGALSFPFNVTELSEAHARGENVTELLMSRAGTPSQTGWHGGYYF
SFWTDNGGTVNYWNGDNGRYGVQWQNCGNFVGGKGWNPGAARTINFSGSFNPSG
NGYLAVYGWTQNPLIEYYIVESFGTYDPSSQAQVLGTFYQDGSNYKIAKTTRYNQPS
IEGTSTFDQFWSVRENHRTSGSVNVGAHFARWQQAGLRLGTHNYQIMATEGYQSSG
SSDITVW (SEQ ID NO:88)

FPFNVTELSEAHARGENVTELLMSRAGTPSQTGWHGGYYFSFWTDNGGTVNYWNG
DNGRYGVQWQNCGNFVGGKGWNPGAARTINFSGSFNPSGNGYLAVYGWTQNPLIE
YYIVESFGTYDPSSQAQVLGTFYQDGSNYKIAKTTRYNQPSIEGTSTFDQFWSVRENH
RTSGSVNVGAHFARWQQAGLRLGTHNYQIMATEGYQSSGSSDITVW (SEQ ID
NO:89)

Xyl2202:

ATGGTTTCTGTCAAGGCAGTCCTCCTCCTCGGCGCCGCCGGCACCACCCTGGCCT
TCCCGTTCAACGCTACCCAGTTCAGCGAGCTCGTTGCCCGGGCCGGCACCCCGAG
CGGCACCGGCACGCACGACGGCTTCTACTACTCCTTCTGGACCGACGGCGGCGGC
AACGTCAACTACGAGAACGGTCCTGGCGGCTCCTACACCGTCCAGTGGCAGAAC
TGCGGCAACTTTGTCGGCGGCAAGGGCTGGAACCCCGGCCAGGCCCGCACCATC
ACCTACTCGGGCACCGTCGACTTCCAGGGCGGCAACGGCTACCTGGCCATCTACG
GCTGGACGCAGAACCCGCTGATCGAGTACTACATCGTCGAGTCGTTCGGCTCGTA
CGACCCCTCGTCGCAGGCCCAGACTTTCGGCACCGTCGAGGTGGACGGCGGCAC
CTACACGCTGGCCAAGACGACGCGCGTCAACCAGCCCTCGATCGAGGGCACCAG
CACCTTCGACCAGTTCTGGTCCGTCCGCCAGCAGCACCGCACCTCCGGCTCCGTC
GACGTCGGCGCCCACTTCGACGCCTGGGCCAAGGCCGGCCTCCAGCTCGGCACC
CACAACTACAGATCGTCGCCACCGAGGGCTACCAGAGCAGCGGCTCCTCTTCCAT
CACCGTCCAGGCCTAAGAGGGCCCTCAGGCCTTTGCTCTACTGCCCTCTCCTCTCC
TCTGCGCTTTCCGTAAGGGAGATCTAA (SEQ ID NO:90)

MVSVKAVLLLGAAGTTLAFPFNATQFSELVARAGTPSGTGTHDGFYYSFWTDGGGN
VNYENGPGGSYTVQWQNCGNFVGGKGWNPGQARTITYSGTVDFQGGNGYLAIYG
WTQNPLIEYYIVESFGSYDPSSQAQTFGTVEVDGGTYTLAKTTRVNQPSIEGTSTFDQ
FWSVRQQHRTSGSVDVGAHFDAWAKAGLQLGTHNYRSSPPRATRAAAPLPSPSRPK
RALRPLLYCPLLSSALSVREI (SEQ ID NO:91)

FPFNATQFSELVARAGTPSGTGTHDGFYYSFWTDGGGNVNYENGPGGSYTVQWQN
CGNFVGGKGWNPGQARTITYSGTVDFQGGNGYLAIYGWTQNPLIEYYIVESFGSYDP
SSQAQTFGTVEVDGGTYTLAKTTRVNQPSIEGTSTFDQFWSVRQQHRTSGSVDVGA
HFDAWAKAGLQLGTHNYRSSPPRATRAAAPLPSPSRPKRALRPLLYCPLLSSALSVR
EI (SEQ ID NO:92)

BXyl17994:

ATGATAATGATGAGACTCAAGTCGGGACTGGCCGGGGCGCTGGCCTGGGGAACG
ACGGCGGCGGCGGCGGCGGCGGTGGCGAGAGTGGGAGCCGGCGCGGCCGCGAA
CTCGACCTACTACAACCCGATCCTCCCCGGGTGGCACTCGGACCCGTCGTGCGTG
CAGGTGGAGGGGATCTTCTACTGCGTGACGTCGACCTTCATCTCGTTCCCCGGCC
TGCCCATCTACGCGTCCCGGGACCTGATCAACTGGAAGCACGTCAGCCACGTGTG
GAACCGCGAGTCCCAGCTGCCCGGGTACAGCTGGGCGACGGAGGGCCAGCAGGA
GGGCATGTACGCGGCGACGATCCGGCACCGCGAGGGCGTCTTCTATGTCATCTGC
GAGTACCTGGGCGTCGGCGGCAGGGACGCCGGCGTGCTCTTCCGGGCGACGGAC
CCGTTCGACGACGCGGCCTGGAGCGACGCCCTGACCTTCGCCGCGCCCAAGATC
GACCCGGACCTGTTCTGGGACGACGACGGGACGGCCTACGTGGCGACGCAGGGC
GTGCAGGTGCAGCGCATGGACCTCGACACGGGCGCCATCGGCCCGCCCGTGCCG
CTGTGGAACGGGACGGGCGGGGTGTGGCCCGAGGGCCCGCACATCTACCGCCGC
GCCGACCACTTCTACCTCATGATCGCCGAGGGCGGCACGGCCGAGGACCACGCC
ATCACCATCGCCCGCAGCGACCGGCTGACGGGGCCCTACGTCTCCTGCCCGCACA
ACCCGATCCTGACCAACCGCGGCACGGACGAGTACTTCCAGACGGTCGGCCACG
GCGACCTCTTCCAGGACGCCGCCGGCAACTGGTGGGGCGTCGCCCTGGCCACGC
GCTCCGGCCCGGAGTACCGCGTCTACCCGATGGGGCGCGAGACCGTGCTGTTCCC
CGTCACCTGGCGCGAGGGCGACTGGCCGGTCCTGCAGCCCGTGCGCGGCGCCAT
GTCGGGCTGGCCGCTGCCGCCGCCGACGCGCGACCTGCCCGGCGACGGGCCCTT
CAACGCGGACCCGGACGTGAAGGCGATGCCGCGGAACCTGGTGCACTGGCGGGT
CCCGCGCGAGGGCGCCTTCGCGACCACGGCGCGCGGGCTCCGCGTCGCGCTGGG
GCGCAACCGGCTCGACGGCTGGCCCGGGGGCGCCGAGCCGGCCGCCAGGGCCGT
CTCCTTCGTGGGGCGCCGCCAGACCGACAGCCTCTTCACCTTCAGCGAGGCCGGC
GTGACCGCGTTCCTGACCCAGCTCGCCAACCTGCAGCTCGGCCTGGTCCTCCCTG
GACGGCGGGCCAGCTGCGGCTCCGCTTCATCGCGTCGGGCCACGTCACGCGATA
CCGCGGTGCCGGAGGACTGCACCGATGTCGGCAGCTGTGACGGCGGTGACGACG
GCGGTGACGGCGGGTACCGGTTCGCGGCCATGCTGGCGTCCGACCCGGACCCGG
ACCGGACCCGGATCGAGGTCGGCACCGCGCCGGCCGAGCTGCTCAGCGGCGGCT
CCGGCTCCTTCGTCGGCACCCTGCTCGGCGTCTACGCCACCTGCAACGGGGCCGG
GGAGGGCATCGACTGCCCCGCCGGCACGCCCGACGCTTACTTCACCCGGTGGAG
GTACACGGGCGAGGGCCAGTTCTACACCGAGACCGATCTCGTCCCGCCCGACGA
GGGCCAGGGCAAGGGTAAAGGTAAAGGGAACGGTAAAGGCAAGGGCAACGGCA
ACGGCAACGGCAAAGCCGCCAAGAGAAGCAGGTTTCCAAGGTGGACGCCGGGTC
TAAATGGCGTCGTTATCCCGCCCCTGTGGATCATGGAGGACGACCCGGAGACCC
GCTGGCCGGCCCAGAAGCGGGCTGGGGCGGGCGGGCAGAGCTACGTCTTCCGCC
ACGGCAACCTGCACACAGTTCGGGATGAGAATGATGCCTTCAAGGGCGCCTCTCT
CTGCGTACCTTACCATACCTACCTTGCCAAGGTGATCCAGGCACTTACTCTCAAC
TTTGCGCATCTTTTCGGGGCGTGGAGACTGACGGTGTAG (SEQ ID NO:93)

MIMMRLKSGLAGALAWGTTAAAAAAVARVGAGAAANSTYYNPILPGWHSDPSCV
QVEGIFYCVTSTFISFPGLPIYASRDLINWKHVSHVWNRESQLPGYSWATEGQQEGM
YAATIRHREGVFYVICEYLGVGGRDAGVLFRATDPFDDAAWSDALTFAAPKIDPDLF
WDDDGTAYVATQGVQVQRMDLDTGAIGPPVPLWNGTGGVWPEGPHIYRRADHFY
LMIAEGGTAEDHAITIARSDRLTGPYVSCPHNPILTNRGTDEYFQTVGHGDLFQDAA
GNWWGVALATRSGPEYRVYPMGRETVLFPVTWREGDWPVLQPVRGAMSGWPLPP
PTRDLPGDGPFNADPDVKAMPRNLVHWRVPREGAFATTARGLRVALGRNRLDGWP
GGAEPAARAVSFVGRRQTDSLFTFSEAGVTAFLTQLANLQLGLVLPGRRASCGSASS
RRATSRDTAVPEDCTDVGSCDGGDDGGDGGYRFAAMLASDPDPDRTRIEVGTAPAE
LLSGGSGSFVGTLLGVYATCNGAGEGIDCPAGTPDAYFTRWRYTGEGQFYTETDLVP
PDEGQGKGKGKGNGKGKGNGNGNGKAAKRSRFPRWTPGLNGVVIPPLWIMEDDPE
TRWPAQKRAGAGGQSYVFRHGNLHTVRDENDAFKGASLCVPYHTYLAKVIQALTL
NFAHLFGAWRLTV (SEQ ID NO:94)

WGTTAAAAAAVARVGAGAAANSTYYNPILPGWHSDPSCVQVEGIFYCVTSTFISFPG
LPIYASRDLINWKHVSHVWNRESQLPGYSWATEGQQEGMYAATIRHREGVFYVICE
YLGVGGRDAGVLFRATDPFDDAAWSDALTFAAPKIDPDLFWDDDGTAYVATQGVQ
VQRMDLDTGAIGPPVPLWNGTGGVWPEGPHIYRRADHFYLMIAEGGTAEDHAITIA
RSDRLTGPYVSCPHNPILTNRGTDEYFQTVGHGDLFQDAAGNWWGVALATRSGPEY
RVYPMGRETVLFPVTWREGDWPVLQPVRGAMSGWPLPPPTRDLPGDGPFNADPDV
KAMPRNLVHWRVPREGAFATTARGLRVALGRNRLDGWPGGAEPAARAVSFVGRR
QTDSLFTFSEAGVTAFLTQLANLQLGLVLPGRRASCGSASSRRATSRDTAVPEDCTD
VGSCDGGDDGGDGGYRFAAMLASDPDPDRTRIEVGTAPAELLSGGSGSFVGTLLGV
YATCNGAGEGIDCPAGTPDAYFTRWRYTGEGQFYTETDLVPPDEGQGKGKGKGNG
KGKGNGNGNGKAAKRSRFPRWTPGLNGVVIPPLWIMEDDPETRWPAQKRAGAGGQ
SYVFRHGNLHTVRDENDAFKGASLCVPYHTYLAKVIQALTLNFAHLFGAWRLTV
(SEQ ID NO:95)

BXyl45310:

ATGGGGCGCCTAAACGATCTCATAGCCCTCCTTGCACTGTTGAGCGGCAGTGCCA
CATCCGCTGCCGTAAGAAACACGGCTTCTCAGGCTCGCGCGGCGGAATTCAACA
ACCCGGTGCTCTGGGAGGACTATCCGGACCTGGACGTGTTCCGGGTCGGGTCGAC
CTTCTACTACTCCTCCTCCACGTTCGCCTACTCCCCGGGGGCTCCGGTGCTCAAGT
CGTACGACCTGGTGAACTGGACCCCCGTCACCCACTCGGTCCCGACGCTCAACTT
TGGGGACCGCTACAACCTCACGGGCGGCACGCCGGCCGGCTACGTCAAGGGCAT
CTGGGCGTCGACGCTGCGGTACCGGCCCTCCAACGACAAGTTCTACTGGTACGGC
TGCGTCGAGTTCGGCAAGACGTACATCTGGACCAGCTCCGGCACGCGCGCGGGC
GACAGGGACGGCGAGGTGGACCCCGCCGACTGGGTCTGGGAGCCGCACCCGCCC
ATCGACCGGTGCTACTACGACAGCGGCCTGTTGATCGACGACGACGACAAGATG
TACATCGCGTACGGCAACCCCAAGATCGAGGTCGCCGAGCTGTCCGACGACGGG
CTCACCGAGGTCTCCTCCCGGGTCGTCTACACCCCGCCGGCCGGCACCACCATCG
AGGGCTCGCGCATGTACAAGGTCGGCGACGCCTACTACATCCTGGTGACGCGGC
CGGCCGACGCCGAGTGGGTGCTCCGGTCGACGTCCGGGCCCTTTCGGCCCGGCG
GCATGGTCGACACCCCGGACGGCCGCAGCTGGTACTACGTCGCCTTCATGGACGC
GTACCCGGGGGGCCGCATCCCCGTGGTCGCGCCGCTGCGCTGGACGGACGACGG
GTGGCCCGAGGTGGTGACGGACGCGCAGGGCGGCTGGGGCGCCAGCTACCCGGT
CCCCGTGGAGACGGGCAAGACGGTGCCGGACGACGGCTGGGAGCTGGACGAGTT

CAGGGGCGGCCGGCTGAGCCACCACTGGGAGTGGAACCACAACCCGGACCCGGC
CCGCTTCGCGCTCGCGGGCGGGGACGAGGGCGGGCTGGTGCTGCAGGCGGCGAC
GGTGACGGAGGACCTGTTCGCGGCCAGGAACACGCTCACGCGGAGGATCAGGGG
CCCCAAGTCGAGCGGCACGTTCCGGCTGGACGTCAGCAGGATGCGCGACGGCGA
CCGGGCCGGGGCCGTGCTGTTCCGGGACACGGCGGCGTATATCGGCGTGTGGAA
GCAAGGGGACGAGGCCACCATCGTCGTAGTCGACGGCCTTGAGCTGGCTCTGAG
CTCCTGGACGACCGTCTCGACCGGGAGGGTGGCCGAGACGGGCCCGACCCTGAG
CAGCACGCAGGATGTCTGGCTCCGGATCGAGGCCGACATCACGCCCGCGTTCGG
GACCAACACGGCAAGGACCACGACTTTCTCGTACAGTGTGGACGGCGGGAAGAC
CTTTGTCCGTCTTGGCCCGGCCTTCTCGATGAGCAATACTTGGCAATACTTTACGG
GCTACAGGTTCGGAGTCTTCAACTTTGCCACCAAGGAGCTTGGGGGCGAAGTCA
AGGTCAAGAGCTTCCAGATGCAGCCTCTGTGA (SEQ ID NO:96)

MGRLNDLIALLALLSGSATSAAVRNTASQARAAEFNNPVLWEDYPDLDVFRVGSTF
YYSSSTFAYSPGAPVLKSYDLVNWTPVTHSVPTLNFGDRYNLTGGTPAGYVKGIWA
STLRYRPSNDKFYWYGCVEFGKTYIWTSSGTRAGDRDGEVDPADWVWEPHPPIDRC
YYDSGLLIDDDDKMYIAYGNPKIEVAELSDDGLTEVSSRVVYTPPAGTTIEGSRMYK
VGDAYYILVTRPADAEWVLRSTSGPFRPGGMVDTPDGRSWYYVAFMDAYPGGRIP
VVAPLRWTDDGWPEVVTDAQGGWGASYPVPVETGKTVPDDGWELDEFRGGRLSH
HWEWNHNPDPARFALAGGDEGGLVLQAATVTEDLFAARNTLTRRIRGPKSSGTFRL
DVSRMRDGDRAGAVLFRDTAAYIGVWKQGDEATIVVVDGLELALSSWTTVSTGRV
AETGPTLSSTQDVWLRIEADITPAFGTNTARTTTFSYSVDGGKTFVRLGPAFSMSNTW
QYFTGYRFGVFNFATKELGGEVKVKSFQMQPL (SEQ ID NO:97)

VRNTASQARAAEFNNPVLWEDYPDLDVFRVGSTFYYSSSTFAYSPGAPVLKSYDLV
NWTPVTHSVPTLNFGDRYNLTGGTPAGYVKGIWASTLRYRPSNDKFYWYGCVEFG
KTYIWTSSGTRAGDRDGEVDPADWVWEPHPPIDRCYYDSGLLIDDDDKMYIAYGNP
KIEVAELSDDGLTEVSSRVVYTPPAGTTIEGSRMYKVGDAYYILVTRPADAEWVLRS
TSGPFRPGGMVDTPDGRSWYYVAFMDAYPGGRIPVVAPLRWTDDGWPEVVTDAQG
GWGASYPVPVETGKTVPDDGWELDEFRGGRLSHHWEWNHNPDPARFALAGGDEG
GLVLQAATVTEDLFAARNTLTRRIRGPKSSGTFRLDVSRMRDGDRAGAVLFRDTAA
YIGVWKQGDEATIVVVDGLELALSSWTTVSTGRVAETGPTLSSTQDVWLRIEADITP
AFGTNTARTTTFSYSVDGGKTFVRLGPAFSMSNTWQYFTGYRFGVFNFATKELGGE
VKVKSFQMQPL (SEQ ID NO:98)

Bxyl20937:

ATGACGATGCTCAAGTCGGCCCTCCCCGCGGCGCTGGCCCTCCTCCTAACGGCGG
CCAACGGCCACCCTTCCAGGACCCCGGCGGCGGCGGCGGCGGGGGGATGGGCAC
CGCTGGCGAATGGGACATTCCGGAACCCGATCCTGTACGAGGACTTCCCGGACA
ACGACGTGTCGGTCGGGCCGGACGGGGCCTTCTACCTGTCGGCGTCCAACTTCCA
CTTCAGCCCCGGGGCGCCCATCCTGCGGTCTTACGACCTGGTCGACTGGGAGTTT
GTGGGCCACTCGATCCCGCGCCTCGACTTCGGCGCCGGCTACGACCTGCCGCCGA
CGGGCGAGCGGGCGTACCGCGCGGGCACGTGGGCGTCGACGCTGCGGTACCGCG
AGAGCACGGGGCTCTGGTACTGGATCGGGTGCACCAACTTCTGGCGCACCTGGG
TCTTCACCGCCCCGGCGCCCGAGGGGCCCTGGACCCGGGCGGGCGACTTCGGCG
ACGGCGTGTGCTTCTACGACAACGGCCTGCTGGTCGACGACGACGACACCATGT
ACGTCGTCTACACCCACGACGGCGGCAAGCGGGTCCACGTGACCCAGCTGAGCG

CGGACGGGCTGAGCGCCGTCCGCACCGAGACCGTCCTGGTGCCGGAGCAGGCCG
GCGTCGACGCCCTCGAGGGCAACCGCATGTACAAGATCGACGGCCGCTACTACA
TCCTCAACGACCACCCGGGCACCACCGCCTACGTCTGGAAGTCCGACTCGCCCTG
GGGTCCCTACGAGGGCAAGGCGCTGGCCGACAACGTCGCCAGCCCCCTGCCCGG
CGGCGGCGCCCCGCACCAGGGCAGCCTGGTGCCCACGCCCTCGGGCGCCTGGTA
CTTTATGTCCTTCACCTGGGCCTACCCGTCCGGCCGCCTGCCCGTGCTGGCCCCGA
TCGAGTTCCAGCCGGACGGGTTCCCGACCCTCGGCGCCTGGTACTTTATGTCCTT
CACCTGGGCCTACCCGTCCGGCCGCCTGCCCGTGCTGGCCCCGATCGAGTTCCAG
CCGGACGGGTTCCCGACCCTCGTCACCGCCAAGGACAACAACAACAACAACAAC
AACAACGCCTGGGGCGCCAGCTACCCGCTGCCGCCGCTACCGCGCCGGCCGCTG
GGCTACCCGTGGTCGCGGGCGCGGTACGACTTCAGCGCGCTCGCCGAACTGCCG
CCCGCGTTCGAGTGGAACCACAACCCGGACGCGAGCAACTACACGCTGGGAGGG
AACGGCGCTGCCGGCCTGATCCTGCGGGCCGCCACCGTCGCGCCCGACGACGAC
CTGTACTCGGCGCGCAACACGCTGACGCACCGCGCCCACGGGCCCTTCCCCTCGG
CCACGCTGGTCCTCGACGTCGCGGACATGGCCGACGGCGACCGCGCCGGGCTGG
CCGCCTTCCGCGACCGCAGTGCCTACATCGGCATCCACTGCTCCTCCTCCTCTGAT
GAGAAGAAGAAGAAGACGTACGAGGTGGTGGCGCGATTCAACATGACGCTGGA
CGAGTGGGGCAGCGGCGAGACGCTCGATCTGGGCGAGGTGGTGGAGCGGGTCGA
GCTGGCCTCGGGCGTGACGCGCGTGTGGCTGCGGGCGAGCATGGACGCGCGGCC
CGACGGCGAGCGGACGGCCCGGTTCGGGTACAGCGTCGACGGGGGCGAGACCTT
TGCCGGCCTGGGGCCCGCCTACCAACTCTACGCCGGGTGGCCCTTCTTTGTCGGC
TACCGCTTCGCCGTCTTCAACTACGCCACCAAGGCCCTCGGCGGGAGCGTCACCG
TCCTGAGCCTCGAGACCGACTCGGGCGAGGGTGAGCGCGATGCCGAGCAAGCGT
GA (SEQ ID NO:99)

MTMLKSALPAALALLLTAANGHPSRTPAAAAAGGWAPLANGTFRNPILYEDFPDND
VSVGPDGAFYLSASNFHFSPGAPILRSYDLVDWEFVGHSIPRLDFGAGYDLPPTGERA
YRAGTWASTLRYRESTGLWYWIGCTNFWRTWVFTAPAPEGPWTRAGDFGDGVCFY
DNGLLVDDDDTMYVVYTHDGGKRVHVTQLSADGLSAVRTETVLVPEQAGVDALE
GNRMYKIDGRYYILNDHPGTTAYVWKSDSPWGPYEGKALADNVASPLPGGGAPHQ
GSLVPTPSGAWYFMSFTWAYPSGRLPVLAPIEFQPDGFPTLGAWYFMSFTWAYPSGR
LPVLAPIEFQPDGFPTLVTAKDNNNNNNNNAWGASYPLPPLPRRPLGYPWSRARYDF
SALAELPPAFEWNHNPDASNYTLGGNGAAGLILRAATVAPDDDLYSARNTLTHRAH
GPFPSATLVLDVADMADGDRAGLAAFRDRSAYIGIHCSSSSDEKKKKTYEVVARFN
MTLDEWGSGETLDLGEVVERVELASGVTRVWLRASMDARPDGERTARFGYSVDGG
ETFAGLGPAYQLYAGWPFFVGYRFAVFNYATKALGGSVTVLSLETDSGEGERDAEQ
A (SEQ ID NO:100)

HPSRTPAAAAAGGWAPLANGTFRNPILYEDFPDNDVSVGPDGAFYLSASNFHFSPGA
PILRSYDLVDWEFVGHSIPRLDFGAGYDLPPTGERAYRAGTWASTLRYRESTGLWY
WIGCTNFWRTWVFTAPAPEGPWTRAGDFGDGVCFYDNGLLVDDDDTMYVVYTHD
GGKRVHVTQLSADGLSAVRTETVLVPEQAGVDALEGNRMYKIDGRYYILNDHPGTT
AYVWKSDSPWGPYEGKALADNVASPLPGGGAPHQGSLVPTPSGAWYFMSFTWAYP
SGRLPVLAPIEFQPDGFPTLGAWYFMSFTWAYPSGRLPVLAPIEFQPDGFPTLVTAKD
NNNNNNNNAWGASYPLPPLPRRPLGYPWSRARYDFSALAELPPAFEWNHNPDASNY
TLGGNGAAGLILRAATVAPDDDLYSARNTLTHRAHGPFPSATLVLDVADMADGDR
AGLAAFRDRSAYIGIHCSSSSDEKKKKTYEVVARFNMTLDEWGSGETLDLGEVVER
VELASGVTRVWLRASMDARPDGERTARFGYSVDGGETFAGLGPAYQLYAGWPFFV
GYRFAVFNYATKALGGSVTVLSLETDSGEGERDAEQA (SEQ ID NO:101)

Xyl5:

ATGGTTACCCTCACTCGCCTGGCGGTCGCCGCGGCGGCCATGATCTCCAGCACTG
GCCTGGCTGCCCCGACGCCCGAAGCTGGCCCCGACCTTCCCGACTTTGAGCTCGG
GGTCAACAACCTCGCCCGCCGCGCGCTGGACTACAACCAGAACTACAGGACCAG
CGGCAACGTCAACTACTCGCCCACCGACAACGGCTACTCGGTCAGCTTCTCCAAC
GCGGGAGATTTTGTCGTCGGGAAGGGCTGGAGGACGGGAGCCACCAGAAACATC
ACCTTCTCGGGATCGACACAGCATACCTCGGGCACCGTGCTCGTCTCCGTCTACG
GCTGGACCCGGAACCCGCTGATCGAGTACTACGTGCAGGAGTACACGTCCAACG
GGGCCGGCTCCGCTCAGGGCGAGAAGCTGGGCACGGTCGAGAGCGACGGGGGC
ACGTACGAGATCTGGCGGCACCAGCAGGTCAACCAGCCGTCGATCGAGGGCACC
TCGACCTTCTGGCAGTACATCTCGAACCGCGTGTCCGGCCAGCGGCCCAACGGCG
GCACCGTCACCCTCGCCAACCACTTCGCCGCCTGGCAGAAGCTCGGCCTGAACCT
GGGCCAGCACGACTACCAGGTCCTGGCCACCGAGGGCTGGGGCAACGCCGGCGG
CAGCTCCCAGTACACCGTCAGCGGC (SEQ ID NO:102)

MVTLTRLAVAAAAMISSTGLAAPTPEAGPDLPDFELGVNNLARRALDYNQNYRTSG
NVNYSPTDNGYSVSFSNAGDFVVGKGWRTGATRNITFSGSTQHTSGTVLVSVYGWT
RNPLIEYYVQEYTSNGAGSAQGEKLGTVESDGGTYEIWRHQQVNQPSIEGTSTFWQY
ISNRVSGQRPNGGTVTLANHFAAWQKLGLNLGQHDYQVLATEGWGNAGGSSQYTV
SG (SEQ ID NO:103)

APTPEAGPDLPDFELGVNNLARRALDYNQNYRTSGNVNYSPTDNGYSVSFSNAGDF
VVGKGWRTGATRNITFSGSTQHTSGTVLVSVYGWTRNPLIEYYVQEYTSNGAGSAQ
GEKLGTVESDGGTYEIWRHQQVNQPSIEGTSTFWQYISNRVSGQRPNGGTVTLANHF
AAWQKLGLNLGQHDYQVLATEGWGNAGGSSQYTVSG (SEQ ID NO:104)

BXyl7:

ATGTTCTTCGCTTCTCTGCTGCTCGGTCTCCTGGCGGGCGTGTCCGCTTCACCGGG
ACACGGGCGGAATTCCACCTTCTACAACCCCATCTTCCCCGGCTTCTACCCCGAT
CCGAGCTGCATCTACGTGCCCGAGCGTGACCACACCTTCTTCTGTGCCTCGTCGA
GCTTCAACGCCTTCCCGGGCATCCCGATTCATGCCAGCAAGGACCTGCAGAACTG
GAAGTTGATCGGCCATGTGCTGAATCGCAAGGAACAGCTTCCCCGGCTCGCTGA
GACCAACCGGTCGACCAGCGGCATCTGGGCACCCACCCTCCGGTTCCATGACGA
CACCTTCTGGTTGGTCACCACACTAGTGGACGACGACCGGCCGCAGGAGGACGC
TTCCAGATGGGACAATATTATCTTCAAGGCAAAGAATCCGTATGATCCGAGGTCC
TGGTCCAAGGCCGTCCACTTCAACTTCACTGGCTACGACACGGAGCCTTTCTGGG
ACGAAGATGGAAAGGTGTACATCACCGGCGCCCATGCTTGGCATGTTGGCCCAT
ACATCCAGCAGGCCGAAGTCGATCTCGACACGGGGGCCGTCGGCGAGTGGCGCA
TCATCTGGAACGGAACGGGCGGCATGGCTCCTGAAGGGCCGCACATCTACCGCA
AAGATGGGTGGTACTACTTGCTGGCTGCTGAAGGGGGGACCGGCATCGACCATA
TGGTGACCATGGCCCGGTCGAGAAAAATCTCCAGTCCTTACGAGTCCAACCCAA
ACAACCCCGTGTTGACCAACGCCAACACGACCAGTTACTTTCAAACCGTCGGGCA
TTCAGACCTGTTCCATGACAGACATGGGAACTGGTGGGCAGTCGCCCTCTCCACC
CGCTCCGGTCCAGAATATCTTCACTACCCCATGGGCCGCGAGACCGTCATGACAG

CCGTGAGCTGGCCGAAGGACGAGTGGCCAACCTTCACCCCCATATCTGGCAAGA
TGAGCGGCTGGCCGATGCCTCCTTCGCAGAAGGACATTCGCGGAGTCGGCCCCTA
CGTCAACTCCCCCGACCCGGAACACCTGACCTTCCCCCGCTCGGCGCCCCTGCCG
GCCCACCTCACCTACTGGCGATACCCGAACCCGTCCTCCTACACGCCGTCCCCGC
CCGGGCACCCCAACACCCTCCGCCTGACCCCGTCCCGCCTGAACCTGACCGCCCT
CAACGGCAACTACGCGGGGGCCGACCAGACCTTCGTCTCGCGCCGGCAGCAGCA
CACCCTCTTCACCTACAGCGTCACGCTCGACTACGCGCCGCGGACCGCCGGGGAG
GAGGCCGGCGTGACCGCCTTCCTGACGCAGAACCACCACCTCGACCTGGGCGTC
GTCCTGCTCCCTCGCGGCTCCGCCACCGCGCCCTCGCTGCCGGGCCTGAGTAGTA
GTACAACTACTACTAGTAGTAGTAGTAGTCGTCCGGACGAGGAGGAGGAGCGCG
AGGCGGGCGAAGAGGAAGAAGAGGGCGGACAAGACTTGATGATCCCGCATGTG
CGGTTCAGGGGCGAGTCGTACGTGCCCGTCCCGGCGCCCGTCGTGTACCCGATAC
CCCGGGCCTGGAGAGGCGGGAAGCTTGTGTTAGAGATCCGGGCTTGTAATTCGA
CTCACTTCTCGTTCCGTGTCGGGCCGGACGGGAGACGGTCTGAGCGGACGGTGGT
CATGGAGGCTTCGAACGAGGCCGTTAGCTGGGGCTTTACTGGAACGCTGCTGGG
CATCTATGCGACCAGTAATGGTGGCAACGGAACCACGCCGGCGTATTTTTCGGAT
TGGAGGTACACACCATTGGAGCAGTTTAGGGAT (SEQ ID NO:105)

MFFASLLLGLLAGVSASPGHGRNSTFYNPIFPGFYPDPSCIYVPERDHTFFCASSSFNA
FPGIPIHASKDLQNWKLIGHVLNRKEQLPRLAETNRSTSGIWAPTLRFHDDTFWLVTT
LVDDDRPQEDASRWDNIIFKAKNPYDPRSWSKAVHFNFTGYDTEPFWDEDGKVYIT
GAHAWHVGPYIQQAEVDLDTGAVGEWRIIWNGTGGMAPEGPHIYRKDGWYYLLA
AEGGTGIDHMVTMARSRKISSPYESNPNNPVLTNANTTSYFQTVGHSDLFHDRHGN
WWAVALSTRSGPEYLHYPMGRETVMTAVSWPKDEWPTFTPISGKMSGWPMPPSQK
DIRGVGPYVNSPDPEHLTFPRSAPLPAHLTYWRYPNPSSYTPSPPGHPNTLRLTPSRLN
LTALNGNYAGADQTFVSRRQQHTLFTYSVTLDYAPRTAGEEAGVTAFLTQNHHLDL
GVVLLPRGSATAPSLPGLSSSTTTTSSSSSRPDEEEEREAGEEEEEGGQDLMIPHVRFR
GESYVPVPAPVVYPIPRAWRGGKLVLEIRACNSTHFSFRVGPDGRRSERTVVMEASN
EAVSWGFTGTLLGIYATSNGGNGTTPAYFSDWRYTPLEQFRD (SEQ ID NO:106)

SPGHGRNSTFYNPIFPGFYPDPSCIYVPERDHTFFCASSSFNAFPGIPIHASKDLQNWKL
IGHVLNRKEQLPRLAETNRSTSGIWAPTLRFHDDTFWLVTTLVDDDRPQEDASRWD
NIIFKAKNPYDPRSWSKAVHFNFTGYDTEPFWDEDGKVYITGAHAWHVGPYIQQAE
VDLDTGAVGEWRIIWNGTGGMAPEGPHIYRKDGWYYLLAAEGGTGIDHMVTMARS
RKISSPYESNPNNPVLTNANTTSYFQTVGHSDLFHDRHGNWWAVALSTRSGPEYLH
YPMGRETVMTAVSWPKDEWPTFTPISGKMSGWPMPPSQKDIRGVGPYVNSPDPEHL
TFPRSAPLPAHLTYWRYPNPSSYTPSPPGHPNTLRLTPSRLNLTALNGNYAGADQTFV
SRRQQHTLFTYSVTLDYAPRTAGEEAGVTAFLTQNHHLDLGVVLLPRGSATAPSLPG
LSSSTTTTSSSSSRPDEEEEREAGEEEEEGGQDLMIPHVRFRGESYVPVPAPVVYPIPR
AWRGGKLVLEIRACNSTHFSFRVGPDGRRSERTVVMEASNEAVSWGFTGTLLGIYA
TSNGGNGTTPAYFSDWRYTPLEQFRD (SEQ ID NO:107)

[0235] The following sequences comprise additional xylanase (Xyl), beta-xylosidase (BXyl), and alpha-xylosidase (AXyl) sequences of interest. The first sequence provided in each set below comprises the cDNA sequence, the second sequence is the polypeptide sequence with no signal sequence predicted.

Xyl8836:

ATGCTGAACCTATCCCACACCGAGCACACTCTCTTTCGCCCTCTCCCCCTTTCCCT
CCCTCATCACCACCACCACCACCACTTCATTGTCGGCCGCCGCCCGCCCGAGGCG
CTGCGCGGCGCCATCACGCGCCACATCCGCGCCGTCGCCGGCTACTACCGCGGCC
GCTGCTACGCCTGGGACGTGGTCAACGAGGCGCTCGACGAGGACGGCACCTACC
GCAAGAGCCTCTTCTACAACGTCCTCGGCGACGAGTACATCCGCATCGTCAAGAC
CTTCGAGAAGCTGATCCGCGAGAAGCCAAAGCCGGGCTTCAAGCGCAAGAGGAA
AACCGTAGCAGCAAACTAA (SEQ ID NO:108)

MLNLSHTEHTLFRPLPLSLPHHHHHHHFIVGRRPPEALRGAITRHIRAVAGYYRGRCY
AWDVVNEALDEDGTYRKSLFYNVLGDEYIRIVKTFEKLIREKPKPGFKRKRKTVAAN
(SEQ ID NO:109)

AXyl267:

ATGGAGGAGGAAGCGACTCCAAGACCCCAATCGAGTATCGTGCAGATGCAGAGG
CACATGCTCAACTCGCGCTGGCATGCCAGGCGTTTGGCCAACAAACCCCACGGC
GTCTTCCCAAGCTTGGATGGACATCTAAGGACCTACACCAAGGATATCCGACCAG
CCCCGACCTGGCGGGTCGGACAATGGCTCGTGGCCGAGGGCGTACAAGTCCAAT
ACGCCGAGGAAGTATACCGAATCACTCCCACGGCCTCGGGCAAGGGAATCAGCC
TCTTGTGCCCGACGCGCAAGATCTTGAACCGTGGGAACACTCTGAACCTGGCAAC
GCTCAGCATCGACATCGAGCCGGCTTTTGATGGCGTCCTCTCTGTCGAGACCACC
CACTGGCAAGGCGCCGTCCGTCGCGGACCCGACTTCGACCTCTTCCCCGCCGGCC
GGCCCGAGGTGGACGCCAAGGTGACCAAGACGGAGAGCGGCACCACCCTGTCGT
CCGGGACGCTCTCGGCGACAGTCAGCGGCAAGCCGCACGAGTTCGAGATCGCCT
TCCATCCGACCGGGGGCAAGAAGCCCCTGACCACCCTGCTCAACCGGTCAGTCG
GCCTGGCCTACACGCCCGCCCCGAGCACGCCCATGCAGCTGGCCGACATGCGCA
ACTTCCGCCACTACATCTTCACCCAGACCACCCTCGCCGTCGGCGAGTCCATCCA
CGGGCTCGGCGAGCGCTTCGGGCCCTTCAACAAGGTCGGCCAGAGGGTCGAGCT
GTGGAACGCGGACGGGGGCACCTCGTCCGACCAGGCGTACAAGAACGTGGGCTT
CTGGATGAGCTCGCGCGGCTACGGTGTCTTCGTCGACACTCCCGGGCGCGTCGAG
CTCGAGATCGGGAGCGAGCGGTGCTGCCGGCTCCAGACGAGCGTCGAGGGGCAG
CGGCTCCGCTGGTTCATCATCTACGGGCCCTCCCCGCGCGACATCCTGCGCCGGT
ACTCGGTCCTCACCGGAGCCCCCGGCAGCGTGCCCAGCTGGTCCTTCGGCCTGTG
GCTCAGCACGTCCTTCACCACCTCGTACGACGAGGAGACGGTCAACAGCTTCCTG
GCCGGCATGAGGGCGCGCGACATACCCGTCGAGGTCTTCCACTTCGACTGCTTCT
GGCTCAAGGCGTTCCAGTGGTGCGACTTCGAGTTCGACCGCGACATGTTCCCGGA
CCCGAGGGGCCAGATCGGGCGCCTCAAGGCCGGCGGCCTCGTCAAGAAGGTCTG
CGTCTGGACGAACCCGTACCTGGGCCAGGCGTCCCCCGTCTTCGCCGAGGCCGCG
GCCAGGGGCTACCTGCTCCGGCGCAGGAACGGCGACGTCTTCCAGTGGGACCTG
TGGCAGACGGGCATGGGCATCGTCGACTTCACCAACCCGGACGCCCGCGCCTGG
TTCGCCGCCTGTCTCGACCGCCTCTTCGACACGGGCGTCGACTGCATCAAGACCG
ACTTTGGCGAGCGCATCCCCTCCGAGGATGTGCAGTGGTTCGACCCTTCGGTCGA
CCCGGAGCGGATGCACAACTACTACGCCTTCATCTACAACAAGCTCGTCTACGAG
GCCCTGCAGAGGCGTTACGGCGCCAACGAGGCCGTCCTGTTCGCCCGCGCCGCC

ACCGCCGGCTGCCAGCGGTTCCCCCTCACCTGGGGCGGCGACTGCGAGTCGACCC
CCGAGGCCATGGCCGAGTCGCTACGCGGTGGTTTGTCCCTCGGCCTGTCCGGGTT
CGCCTTCTGGAGCGTCGACATTGGCGGCTTCGAGGGGTCGCCGCCTCCCTGGATC
TACAAGCGCTGGGTCGCCTTCGGCCTCCTCTGCTCCCACTCGCGCCTGCACGGCT
CCAACTCGTACCGGGTCCCCTGGACGGTCGACGGCGACGACCAGTCCGAGGAGG
GATGCTCCGCCACGCTGCGCAAGTGGACCCATCTCAAGGCTCGCCTGATGCCCTA
CCTCTTCTCCCAGGCGCAGGAGAGCGTCCGGGGCGGGCTCCCGCTCAGCCTGAG
GGCCATGTGCATCGAGTTCCCCGACGACCCGACCGCCTGGACCCTCGATCGCCAG
TTCATGCTCGGCGACGGCCTCCTCGTCGCCCCCGTCTTCGAGGAGGACGGCACCG
TCGAGTTCTACCTGCCCAGGGGCAAGTGGACCAACTTCTTCACCGGCGAGGTCAA
GGAGGGCCCCGGCTGGTTCGCCGAGACCCACGGGTTCGGCACCCTGCCGCTCTAC
GTCCGGCCCAACACGCTCCTGGTTCTGGGCAAGGAAGGAGAGACGAGGACCGTG
TACGACTACACGAGCGACGTCGAGGTGAGGGCGTATTTTGCCAGTGACAGCGCC
AGCGCCGTGCTGGTCGACGCCGAGGGCAAGACTGTAGGTACCCTGCGTGTCAAG
GACGGGGAGATTATCGGAAAGGAACTGCTATCTGGCAACTCGGTCATCAATGTC
GTGAGCTCCTGA (SEQ ID NO:110)

MEEEATPRPQSSIVQMQRHMLNSRWHARRLANKPHGVFPSLDGHLRTYTKDIRPAP
TWRVGQWLVAEGVQVQYAEEVYRITPTASGKGISLLCPTRKILNRGNTLNLATLSIDI
EPAFDGVLSVETTHWQGAVRRGPDFDLFPAGRPEVDAKVTKTESGTTLSSGTLSATV
SGKPHEFEIAFHPTGGKKPLTTLLNRSVGLAYTPAPSTPMQLADMRNFRHYIFTQTTL
AVGESIHGLGERFGPFNKVGQRVELWNADGGTSSDQAYKNVGFWMSSRGYGVFVD
TPGRVELEIGSERCCRLQTSVEGQRLRWFIIYGPSPRDILRRYSVLTGAPGSVPSWSFG
LWLSTSFTTSYDEETVNSFLAGMRARDIPVEVFHFDCFWLKAFQWCDFEFDRDMFP
DPRGQIGRLKAGGLVKKVCVWTNPYLGQASPVFAEAAARGYLLRRRNGDVFQWDL
WQTGMGIVDFTNPDARAWFAACLDRLFDTGVDCIKTDFGERIPSEDVQWFDPSVDP
ERMHNYYAFIYNKLVYEALQRRYGANEAVLFARAATAGCQRFPLTWGGDCESTPE
AMAESLRGGLSLGLSGFAFWSVDIGGFEGSPPPWIYKRWVAFGLLCSHSRLHGSNSY
RVPWTVDGDDQSEEGCSATLRKWTHLKARLMPYLFSQAQESVRGGLPLSLRAMCIE
FPDDPTAWTLDRQFMLGDGLLVAPVFEEDGTVEFYLPRGKWTNFFTGEVKEGPGWF
AETHGFGTLPLYVRPNTLLVLGKEGETRTVYDYTSDVEVRAYFASDSASAVLVDAE
GKTVGTLRVKDGEIIGKELLSGNSVINVVSS (SEQ ID NO:111)

AXyl6158:

ATGGCCAGCAGCCGGTACCGGTACACGTTCCCGAGGAATCCGAAGGCCAATCCG
AAGGCCGTCGTGACAGGCGGCAAGGGATCCTCTTACTATCGCTTCACCCTCCTCA
CCGAACGGTTGATCCGTTACGAGTGGTCCGAGGACGGAGGCTTCGAGGATCGCG
CGTCCACGTTCGCGGTATTCAGATACTTTGATGCCCCGCAGTACCGCGTTGTCGA
GACAAACGACAGTCTCGAGATCATCACGGACTACTTTCACCTCACCTATGACAAG
AAGAAGTTCTCATCGGAAGGACTTTCCGTCAGAGTCGGCTCCGACCTCTGGAATT
ACGACGGCAAGAGTTATGGAGACCTGGGCGGCACCGCCCGGACCCTAGACGGCG
CCTATGGCCGCGTGGACCTGGAACCGGGTGTGCTCTCGCGCAAAGCTTATGCGGT
TCTCGACGACAGCAAGTCTATGCTCTTTGACGACGACGGGTGGATTGCCATTCGC
GAGCCGGGCCGCATTGACGGTTACGTGTTTGCCTACAGCGGCGAGCACAAGGCC
GCCATCAGGGACTTCTACCGCCTCTCCGGGCGTCAGCCGGTGCTCCCCCGCTGGG
TGCTGGGGAACTGGTGGTCCAGGTACCACGCATACTCGGCCGACGAATACATCG
AGCTTATGGACCACTTCAAACGCGAAGGAATCCCGCTCACGACGAGCATCGTGG

ATATGGACTGGCACCGGGTTGACGACGTCCCGCCCAAGTACGGCTCAGGATGGA
CGGGCTACAGCTGGAACCGCAAGCTGTTCCCGGACCCCGAGGGGTTCCTGCAGG
AGCTGCGTAATCGGAACCTGAAAGTGGCCCTCAACGACCACCCGGCGGACGGCA
TCCGGGCGTATGAGGATCTGTACCCGGCGGTGGCCAAGGCCCTGAATCACGACA
CGTCGCGAGAGGAACCGATCAAGTTTGACTGCACCGATCGCAAGTTCATGGACG
CCTACTTCGACGTTCTGAAGCTCAGCCTTGAGAAGCAGGGCGTCATGTTCTGGTG
GATCGACTGGCAGCAAGGCACCGGCAGCAAGCTCCCCAGCGTAGACCCGCTGTG
GGTGCTCAATCACTACCACTACCTCACCAGTAAGCGCAACGCGAAAGACATCCA
ACGTCCCATCACATTCTCCCGCTACGCCGGCGCCGGTGCCCATCGGTACCCGATC
GGCTTCTCGGGCGACACGCAGACGACTTGGGAAGGTCTCGAGTTCCAGCCCGAG
TTTACCGCAACGGCATCCAACATCGGCTATGGCTGGTGGAGCCACGACATCGGC
GGGCATTGGGGCGGCGTCCGCTCCAACCAGCTGACGGTCCGCTGGGTCCAGCTG
GGCTGCTTCTCCCCGATCCTGCGGCTGCACTCGAACAAGAGCCCGTGGAACTCGA
GAGAGCCGTGGAACTACGAGGACGAGGCGCACAGGATCATGAAGGACTTCCTCA
TCCTGCGCCACCGCCTCATCCCCTTCCTCTACACCATGAACATCCGGGCCAGCTA
CGAGAGCGAGCCGCTCATCCAGCCCATGTACTGGAATCACCCGAAGGACGAAGA
GGCCTACACGGTGCCGACGCAGTACTACTTCGGGCCGGACCTCCTCGTGGCCCCC
ATCACGTCTCCCAACAGCACCGTCACCCTGATGGGCCGCGTGCGCGCCTGGCTGC
CGCCGGGCCGGTACGTCGACCTGTTCTACCCGCACCTGGTCTACGACGGCGGCCG
GTACATGCACCTGCACCGCGACCTGTCGCAGATCCCCGTGCTCGCGCGGGAGGG
CACCATCGTGCCGCTGGACACGACGCCCAGGACGGGCCACGGCGCCGCGCGGCC
GACCGAGATCACCCTCCTCCTCGTCGTCGGCCGGGACGCGCACTTTGAGCTGGTC
GAGGAGCCGGAGCAGCAGGACCACCATCGCCACGGCGGCGGCGACGACGGCGA
TGACCAACCCCCCGCTCAGCGCGTTCGCCCGGACCCCCATCTCGTGGTCGCAGGCG
GACGGCGTGCTCACCATCGGGCCGGAGTGGAACGGCGCCGGGGCCCGCCGCTGG
CGGCAGTGGAACGTCAAGCTGGTCGGGCACACCAACACGGACGTGCAGGCGCAG
GTGCCCGGGTTCCGGGTCACGCGCGACGTCGAGGGCGGGTGCACGACGGTGGCG
CTCGGCAACGTGCACCGGTGGCAGCAGCCGCACCAGCGGGACGGCGGCGGGTTC
GAGATCTCGCTGGGGCGCGACCTGCAGCTGGACGTGGTGGACGTGCGCGCGCGC
GCCTTCGAGGTCCTGCACCGGGCCGAGATGGGGTACGAGGCCAAGGACCCCGTC
TGGGACGTCTTCACGTCCGGCGACGCGGTGCAGACGCGGGTGCAGCGGCTGGCG
GCGCTCGACGTCGACGCCGCGCTCAAGAACGCCCTCATGGAGGTCTGGGCGGCC
GACGGGCGGGCCGAGGGCAGCGCGGCGGGCTACGAGACCTGGGTGGACGTGAA
GGCGTGCGCGGGAGACGCGGTCGAGGAGGCGCTCAAGGAGTACGTTATCGTGTG
A (SEQ ID NO:112)

MASSRYRYTFPRNPKANPKAVVTGGKGSSYYRFTLLTERLIRYEWSEDGGFEDRAST
FAVFRYFDAPQYRVVETNDSLEIITDYFHLTYDKKKFSSEGLSVRVGSDLWNYDGKS
YGDLGGTARTLDGAYGRVDLEPGVLSRKAYAVLDDSKSMLFDDDGWIAIREPGRID
GYVFAYSGEHKAAIRDFYRLSGRQPVLPRWVLGNWWSRYHAYSADEYIELMDHFK
REGIPLTTSIVDMDWHRVDDVPPKYGSGWTGYSWNRKLFPDPEGFLQELRNRNLKV
ALNDHPADGIRAYEDLYPAVAKALNHDTSREEPIKFDCTDRKFMDAYFDVLKLSLE
KQGVMFWWIDWQQGTGSKLPSVDPLWVLNHYHYLTSKRNAKDIQRPITFSRYAGA
GAHRYPIGFSGDTQTTWEGLEFQPEFTATASNIGYGWWSHDIGGHWGGVRSNQLTV
RWVQLGCFSPILRLHSNKSPWNSREPWNYEDEAHRIMKDFLILRHRLIPFLYTMNIRA
SYESEPLIQPMYWNHPKDEEAYTVPTQYYFGPDLLVAPITSPNSTVTLMGRVRAWLP
PGRYVDLFYPHLVYDGGRYMHLHRDLSQIPVLAREGTIVPLDTTPRTGHGAARPTEI
TLLLVVGRDAHFELVEEPEQQDHHRHGGGDDGDDQPPLSAFARTPISWSQADGVLTI
GPEWNGAGARRWRQWNVKLVGHTNTDVQAQVPGFRVTRDVEGGCTTVALGNVH
RWQQPHQRDGGGFEISLGRDLQLDVVDVRARAFEVLHRAEMGYEAKDPVWDVFTS

GDAVQTRVQRLAALDVDAALKNALMEVWAADGRAEGSAAGYETWVDVKACAGD
AVEEALKEYVIV (SEQ ID NO:113)

BXyl323:

ATGCCGCAGGTTCGAAACCCCATCCTCCCCGGCTTCAACCCCGACCCTTCCATCC
TCCGGGTTGGGGATGACTACTACATCGCCACTTCAACCTTTGAGTGGTACCCGGG
TGTTCAGATCCACCACTCCATGGACCTCGCAAACTGGGAACTTGTCACCCGTCCC
CTAAACCGCAAGAGCCAACTGGATATGCGAGGAGATCCGGACAGCTGCGGCATC
TGGGCTCCCTGCCTGACGCATGACGGCGACAGGTTCTGGCTGGTATACACGGACG
TCAAACGCAAGGACGGCTCGTTCAAGGACGCACACAACTACATCGTCAGTGCGC
CCGCCATCGAGGGTCCCTGGTCGGACCCCTTCTATGTCAACTCGTCCGGGTTCGA
CCCCTCGCTCTTCCATGACGACGACGGCCGGAAGTGGTTCGTCAACATGATGTGG
GACCACCGCAGCCGCCCGCGAACCTTTGCCGGCATCGCGCTGCAAGAGTTCGAC
CCCAAGGCCGGGAAGCTGGTTGGGCCGCGCAAGAACATTTACCAAGGCACCGAC
CTGGGCCTCGTCGAGGGCCCGCACTTGTACAAGCGCAACGGGTGGTACTATCTCC
TGACAGCAGAGGGCGGGACTGGCTATGAGCATGCCTGCACCCTCGCCCGGTCTC
GGAACATCTGGGGCCCGTACGAAGATCACCCGCAGAAGTACATCTTGACGTCTA
AGGACCACCCGCACGCAGCCCTGCAGCGAGCCGGCCACGGCGACATCGTCGACA
CCCCCGACGGGCGTACCTACGTCGTTCACCTGACCGGCCGGCCCATCACGCAGTT
CCGCCGCTGTGTCTTGGGGCGCGAGACGGCCATCCAGGAGGCCTACTGGGGCGA
CGACGACTGGCTCTACGTCAAGAACGGCCCTGTGCCCAGCCTGTTCGTGGACCTC
CCGGCCGCCCGCAACGACGACGACTACTGGGCCGAGAAGAGGTACACGTTCGAG
GCGGGCCTGCACAAGGACTTCCAGTGGCTGCGCACGCCCGAGACGGACCGCATC
TTCAGGACGGACAACGGGAAGTTGACGCTCATCGGCCGCGAGTCCATCGGCTCC
TGGTTCGAGCAGGCCCTGGTCGCCCGGCGCCAGACGCACTTCTCGTACGACGCCG
AGACCGTCATCGACTTCAAGCCTGCCGACGAGCGCCAGTTCGCCGGCCTGACGG
CCTATTACTGCCGCTACAACTTCTTCTACCTGACCGTCACGGCCCACTCGGACGG
CCGGCGGGAGCTGCTCATCATGGCCTCCGAGGCCTCCTGGCCCCTCGGCGCCCTC
CGGTCCCCTTATCCGGGACCCGTCCAGATCCCCAACGAGGGCAAGGTCCGGCTCG
CGCTCAAGATCAGGGGCAAGGAGCTGCAGTTCTACTACGCTCTCGAGGGCGAAG
AGCTAAAACAGATTGGGCCCGTATTCGACGCTAGCATCGTTTCTGACGAGTGCGG
CGGCCACCAGAAGCACGGCAGCTTCACGGGCGCCTTCGTCGGCGTGGCTGCTTCC
GACATCAACGGTACTGCTGCCGAGGCGACCTTTGACTACTTTGTGTACAAGCCCG
TGCACCATGAGAGTGACCGGTACGAGATTTAA (SEQ ID NO:114)

MPQVRNPILPGFNPDPSILRVGDDYYIATSTFEWYPGVQIHHSMDLANWELVTRPLN
RKSQLDMRGDPDSCGIWAPCLTHDGDRFWLVYTDVKRKDGSFKDAHNYIVSAPAIE
GPWSDPFYVNSSGFDPSLFHDDDGRKWFVNMMWDHRSRPRTFAGIALQEFDPKAG
KLVGPRKNIYQGTDLGLVEGPHLYKRNGWYYLLTAEGGTGYEHACTLARSRNIWGP
YEDHPQKYILTSKDHPHAALQRAGHGDIVDTPDGRTYVVHLTGRPITQFRRCVLGRE
TAIQEAYWGDDDWLYVKNGPVPSLFVDLPAARNDDDYWAEKRYTFEAGLHKDFQ
WLRTPETDRIFRTDNGKLTLIGRESIGSWFEQALVARRQTHFSYDAETVIDFKPADER
QFAGLTAYYCRYNFFYLTVTAHSDGRRELLIMASEASWPLGALRSPYPGPVQIPNEG
KVRLALKIRGKELQFYYALEGEELKQIGPVFDASIVSDECGGHQKHGSFTGAFVGVA
ASDINGTAAEATFDYFVYKPVHHESDRYEI (SEQ ID NO:115)

BXyl6880:

ATGGCGCCCCTCATCACCAACATCTTCACGGCCGACCCGTCGGCCCACGTCTTCG
AGGGCAAGCTCTTCATATACCCGTCGCACGATCGCGAGACGGACATCAAGTTCA
ACGACGACGGCGACCAGTACGACATGGTCGACTACCACGTATTCAGCACCGAGT
CGCTGGACCCGGCCGCCCCCGTGACCGACCACGGCGTCGTGCTCCGGGCCGAAG
ACGTCCCCTGGGTGTCCAAGCAGCTCTGGGCCCCCGACGCCGCCTACAAGGACG
GCAGGTACTACCTCTACTTCCCCGCCCGCGACAAGCAGGGCGTCTTCCGCATCGG
CGTCGCCGTCGGCGACCGCCCCGAGGGCCCCTTCACCCCCGACCCGGAGCCCATC
CGGGACAGCTACAGCATCGACCCGGCCGTCTTCGTCGACGACGACGGCCGGGCC
TACATGTACTTTGGCGGGCTCTGGGGCGGCCAGCTGCAGTGCTACCAGAAGGGC
AACGGCATCTTCGACCCCGAGTGGCTGGGGCCCAGGGAGCCCTCGGGCGAGGGC
GTCCGGGCGCTGGGGCCGCGCGTCGCCCGGCTGGCGGACGACATGCGCCAGTTC
GCCAGCGAGGTGAAGGAGATTTCGATCCTGGCGCCCGAGACGGGCGAGCCGATC
GCGGCCGACGACCACGACCGCCGCTTCTTCGAGGCCGCCTGGATGCACAAGTAC
GACGGCAAGTACTACTTCAGCTACTCCACCGGCGACACCCACTACCTCGTCTACG
CCGTCGGCGACAGCCCCTACGGGCCCTTCACCTACGCCGGCCGCATCCTCGAGCC
CGTCCTCGGCTGGACCACGCACCACTCCATCGTCGAGTTCCACGGCCGCTGGTGG
CTCTTCCACCACGACTGCGAGCTCAGCGGCGGAGTCGACCACCTGCGCTCCGTCA
AGGTCAAGGAGATCTTCTACGACAAGGACGGCAAGATTGTCACTGAAAAGCCCG
AATAG (SEQ ID NO:116)

MAPLITNIFTADPSAHVFEGKLFIYPSHDRETDIKFNDDGDQYDMVDYHVFSTESLDP
AAPVTDHGVVLRAEDVPWVSKQLWAPDAAYKDGRYYLYFPARDKQGVFRIGVAV
GDRPEGPFTPDPEPIRDSYSIDPAVFVDDDGRAYMYFGGLWGGQLQCYQKGNGIFDP
EWLGPREPSGEGVRALGPRVARLADDMRQFASEVKEISILAPETGEPIAADDHDRRFF
EAAWMHKYDGKYYFSYSTGDTHYLVYAVGDSPYGPFTYAGRILEPVLGWTTHHSIV
EFHGRWWLFHHDCELSGGVDHLRSVKVKEIFYDKDGKIVTEKPE (SEQ ID NO:117)

EXAMPLE 1

Construction, Cloning and Plasmid Preparation of Beta-Xylosidase Variant Libraries

[0236] In this Example, experiments conducted to construct and prepare plasmids for use in xylosidase expression libraries are described. The expressed sequence of the wild-type xylosidase was cloned from genomic DNA into pYTSEC72-trc vector. Figure 1 provides the map of this plasmid.

[0237] For production of "round 1" libraries, the QuikChange Lightening Multi Site-Directed mutagenesis (QCLM) kit (Stratagene) was used in accordance with the manfacturer's instructions to produce the following substitions in isolation: G322A; M280L; P31G; G770P; P362V; G134S; V495L; P454A; F221L; A72V; V567I; A694P; R41T; M435L; W783L; A729T; E496A; S192D; D58L; A314T; D488T; G204P; A819T; S107Y; S211A; G320A; D441N; V155I; R583D; M329T; I236V; F456Y; V369L; T96V; S590A; N106S;

N281Q; A457L; V785L; K764P; T42S; A736L; D444T; N787S; A532D; V738L; S754G; H652S; N219Y; I75L; L561N; S167Q; P443T; F438P; S146A; E45P; F559L; F200L; T620G; F325L; T331L; M113Q; D294G; R843V; S256A; P417S; G516A; H791V; K178S; E821T; Q466G; P653D; A264S; A515E; G669C; G449N; L610M; A528G; V734A; L761I; L584I; P207L; F654Y; F447Q; S586T; N289G; Y16R; K30N; V505I; G565N; G101S; S193P; I341Y; R432S; M522T; E666R; N571G; E189Q; G218A; V779A; V174P; S338G; W352E; S202G; Y742A; A380R; D664V; L510I; A588K; S332D; V307I; K445Q; P824E; A93S; S67N; L781V; T523S; W469L; E480N; T695S; W572Y; V774G; I798V; D551G; C324A; V544L; V473Y; E334G; R257K; D412G; D667A; P493A; V129I; K672L; H230Y; L718F; T696Y; M184L; N411D; G286D; T777K; P44S; V261I; E810V; L115I; H379Y; A247P; P102G; E725T; L553V; V478G; G347Q; L446M; G429K; I675A; R703S; A832Q; S673K; E302D; P710D; E485S; S197R; A692S; S377Q; S339T; E536Q; A514V; L62I; S108A; R389T; G657P; F349L; A461L; Y150F; R208K; I185V; G130A; T556S; R689A; I36N; T321A; A769T; E648S; Y25F; G763P; A758L; R176H; T227V; V355L; P717G; T631P; and S308Y.

[0238] The following were also introduced as combinations of one or more substitutions: G322A; S211A; N219Y; A264S;N571G; W572Y; L115I; S108A; M280L; G320A; E45P; V174P; A247P; F438P; A694P; G763P; K764P; G770P; P102V; F105T; V268G; R398N; T695P; I75L; A515E; E189Q; H379Y; R389T; P31G; D441N; L561N; G669C; G134S; G218A; I798V; A247P; G657K; G770P; V155I; S167Q; G449N; V779A; D551G; P102G; F349L; P632V; R583D; P443T; L610M; V174P; C324A; E725T; A461L; G130A; G134S; G218A; G320A; G322A; G347Q; G429K; G449N; G565N; and G763P.

[0239] QCLM reaction products were transformed into DH10B T1 competent cells (Invitrogen). Plasmids were isolated and transformed into yeast (InvSc1) competent cells using a Miniprep kit (Qiagen). Subsequently, 1.5-4μl of the reaction was used to transform 50μl of DH10B-T1 *E.coli* (Invitrogen) electro-competent cells. The cells were plated on LB agar containing carbenicillin (50ug/ml). Colonies were picked, grown in liquid medium containing carbenicillin (50 ug/ml) and plasmids containing the variants were isolated using a Miniprep kit (Qiagen).

[0240] Competent *S. cerevisiae* cells were generated and transformed with libraries of beta-xylosidase variants using standard methods known in the art to generate and transform yeast cells.

EXAMPLE 2

Construction, Cloning and Plasmid Preparation of Beta-Xylosidase Libraries for Expression in *M. thermophila*

[0241] Three generation of libraries of variant beta-xylosidases were identified for additional characterization. These variants in these libraries included one or more of the following substitutions: G322A, S211A, N219Y, A264S, N571G, W572Y, L115I, S108A, M280L, G320A, E45P, V174P, A247P, F436P, A694P, G763P, K764P, G770P, P102V, F105T, V268G, R398N and T695P. The primers in Table 2-1 were used to introduce the target mutations by PCR amplification using the wild-type xylosidase cloned into pC1DX10PhR vector (See, Figure 2). PCR amplification was performed using QCLMS (QuickChange Lightning Multi Site-Directed Mutagenesis, Stratagene). Briefly, each reaction was set up with 17 µl of water, 2.5 µl of 10x QuickChange Multi reaction buffer , 1 µl supplied dNTP mix, 1 µl of 100 ng/ul plasmid DNA template, 2 µl of 20 µM pooled oligos (See, Table 2-1) and 1 µl of QuickChange Lightning Multi enzyme blend. Thermocylcer conditions were: 95°C 2', 24 cycles of 95°C 20", 65°C 7", with a final extension at 65°C 7'. *E. coli* DH10B-T1 phage resistant electrocompetent cells (Invitrogen) were transformed with 2 µl of the QCLMS PCR product according to the electroporation protocol provided by the manufacturer. Cells were plated onto LB agar plates containing 1% v/v glucose and 100 mg/L carbenicillin for positive selection of clones. After overnight incubation at 37°C, colonies were picked onto a Costar 96-deepwell plates filled with 500 µl of LB containing 1% v/v glucose and 100 mg/L carbenicillin. Plates were allowed to grow overnight for 18-20 hours in Kuhner shaker (200 rpm, 37°C, and 85% relative humidity). Cells were collected by centrifugation at 3500 x *g* for 10 minutes. Plasmid DNA was collected using QIAprep Miniprep Turbo96 (Qiagen).

Table 2-1. Primers

Primer Sequence (5'-3')	Amino Acid Modifications	SEQ ID NO:
AACTCGGCGACGTCGTTCCCGATGCCGATTCTGA TGGCCGCCGCCTTCGACGAC	L115I(ATT); S108A(GCG)	118
AACTCGTCCACGTCGTTCCCGATGCCGATTCTGAT GGCCGCCGCCTTCGACGAC	L115I(ATT); S108S(TCC)	119
AACTCGGCGACGTCGTTCCCGATGCCGCTGCTGA TGGCCGCCGCCTTCGACGAC	L115L(CTG); S108A(GCG)	120
TCATCGCGACCTGCAAGCACTACGCCGGCTATGA CTTTGAGGACTGGAACGGCACG	N219Y(TAT); S211A(GCG)	121

Table 2-1. Primers		
Primer Sequence (5'-3')	**Amino Acid Modifications**	**SEQ ID NO:**
TCATCTCGACCTGCAAGCACTACGCCGGCTATGACTTTGAGGACTGGAACGGCACG	N219Y(TAT); S211S(TCG)	122
TCATCGCGACCTGCAAGCACTACGCCGGCAACGACTTTGAGGACTGGAACGGCACG	N219N(AAC); S211A(GCG)	123
GCGCCAACTCGTACCTCCTGAACACGATCCTGCGCGGGCACTGG	M280L(CTG)	124
ACACCAACGCCGAGGCGACCGCGCTCTGCTTCGAGGCCGGCATGGAC	G322A(GCG); G320A(GCG)	125
ACACCAACGCCGAGGGCACCGCGCTCTGCTTCGAGGCCGGCATGGAC	G322A(GCG); G320G(GGC)	126
ACACCAACGCCGAGGCGACCGGCCTCTGCTTCGAGGCCGGCATGGAC	G322G(GGC); G320A(GCG)	127
GCCGTCCTGTGGGCCGGCTATCCGGGCCAGGACGGCGGCACGGCC	W572Y(TAT); N571G(GGC)	128
GCCGTCCTGTGGGCCAACTATCCGGGCCAGGACGGCGGCACGGCC	W572Y(TAT); N571N(AAC)	129
GCCGTCCTGTGGGCCGGCTGGCCGGGCCAGGACGGCGGCACGGCC	W572W(TGG); N571G(GGC)	130

Fungal High Throughput Transformation

[0242] In a 50-ml tube, 16 ml of CF-410 protoplasts were gently mixed with 400 µl of ATA (0.5M aurintricarboxylic acid). The protoplast-ATA mixture were dispensed into a 96-well PCR plate at 170 µl volume per well. Plasmid DNAs representing the xylosidase library were dispensed at 5 µl volume per well in Costar 96-deepwell plates. The protoplast-ATA mixture was added into the Costar 96-deepwell plates at 20 µl volume per well and incubated at room temperature for 25 minutes. Then, 150 µl of PEG 4000 solution (60% PEG4000, 50 mM $CaCl_2 \cdot H_2O$, 35 mM NaCl, 10 mM Tris-HCl) was added per well, mixed and incubated at room temperature for 20 minutes. Next, 600 ul of STC (NaCl 2.05 g/L, $CaCl_2 \cdot 2H_20$ 7.36 g/L, sorbitol 218.64 g/L, 10 ml of 1M Tris-HCl buffer pH 7.50) was added per well and mixed. The plates were centrifuged at 1500 x *g* for 10 minutes. This STC wash step was performed twice. The supernatants were decanted and the cell pellets were resuspended in the residual fluid. Then, 80 µl of cell resuspension were aspirated onto 24-well minimal medium agar plates containing 20 mg/ml phleomycin. The plates were sealed with VWR adhesive film (Cat. Number 60941-086) and incubated for 9 days at 35°C.

Colony Pooling and Growth

[0243] First, 4 x 1 mm sterile glass beads were dispensed into each of well of the 24-well agar plates described above, using Qiagen bead dispenser (Qiagen). Then, 1.6 ml of sterile

water was dispensed into each well. The plates were heat-sealed and agitated in an orbital shaker for 10 minutes at level 7 setting to resuspend the spores from the agar. Then, 300 ul of spore suspension were dispensed into 24-well plates containing 1.8 ml of fermentation media (F1-02 pH 5.15). The plates were sealed with VWR sterile airpore and incubated for 7 days at 35°C, 250 rpm (2" throw) and 85% relative humidity.

Analysis of Variants Expressed in Yeast

[0244] To evaluate thermostability improvement of the generated beta-xylosidase variants produced as described in Example 1, 160 µl of the supernatant from HTP yeast culture was added to 40 µl of 900 mM sodium acetate buffer (pH 6.0) in a 96-deep well plate and incubated at 55°C or 57°C for 24 hours. After 24 hours, the mixtures were centrifuged for ~5 min at 4000 rpm, 4°C, and tested for activity using following *p*NPX assay: 180 µl of the supernatant-buffer mixture was added to 60 µl of 25 mM *p*NPX (*p*-nitrophenyl-beta-xylanopyranoside) in water and 60 µl 900 mM sodium acetate buffer (pH 6.0), and the reactions were incubated at 55°C for 2 hours. After 2 hours, the reaction mixture was centrifuged for ~5 min at 4000 rpm, 4°C, and 10 µl was transferred to 190 µl of 1 M Na_2CO_3 in a flat-bottom clear plate to terminate the reaction. The plate was mixed gently, then centrifuged for 1 min, and absorbance was measured at $\lambda = 405$ nm with a Spectramax M2 (Molecular Devices). Duplicate plates were created to calculate residual activity after the 2 hour thermal challenge where one copy of the plate was assayed without preincubation while the other copy was incubated at 55°C before assaying. Both copies were assayed using the same *p*NPX assay as described above. Residual activity (in percentage) was calculated as a ratio of fluorescence after and before the thermal challenge multiplied by 100.

Xylo-oligosaccharide (XOS) Activity Assay

[0245] This assay was used to determine the activity of the variants on xylose-containing oligosaccharides. In a total volume of 300 µl, 40 µl of HTP yeast culture supernatant containing secreted protein of a beta-xylosidase variant was added to 40 ul 200 g/L XOS (Xylo-oligosaccharides (Cascade) in 160 ul water and 60 ul 900 mM sodium acetate buffer (pH6.0). The reaction was incubated for 24 hours at 55°C or 57°C. After 24 hours, the reaction mixture was centrifuged for ~5 min at 4000 rpm, 4°C, and 100 µl was transferred into 100 µl of water in a round-bottom 96-well plate. The plate was mixed gently, then centrifuged for 1 min and subjected to sugar analysis using standard HPLC methods known in the art.

[0246] The beneficial mutations for stability and/or activity found in a set of variants are shown in Table 2-2. The improvements are shown in comparison with wild-type *M. thermophila* beta-xylosidase.

Table 2-2. Variants with Improved Thermostability and Activity

Variant	Fold Improvement Over Wild-Type Beta-Xylosidase at 55°C	
	Thermostability on *p*NPX	Activity on XOS
N219Y/N571G	+++	+
S211A	++	+
S108A/S211A/M280L/L761I	++	+
N219Y (g291a)	+++	0
S211A/N219Y (g291a/g510a)	+++	0
P31G/H379Y	+	+
I798V	+	+
G347Q	+	+
G347Q/G763P	+	+
G347Q/G449N	++	+

"0"—indicates less than 1 fold improvement
"+" --indicates 1-2 fold improvement
"++"—indicates >2 and <3 fold improvement
"+++"—indicates >3 and <4 fold improvement
"++++"—indicates >4 fold improvement

Table 2-3 Variants with Improved Thermostability and Activity

Variant	Fold Improvement over Wild-Type Beta-Xylosidase at 57°C	
	Thermostability on *p*NPX	Activity on XOS
S345L	++++	0
V235I	++++	0
A499S	++++	0
V209I	++++	0
A499K	++++	+
V235L	++++	+

"0"—indicates less than 1 fold improvement
"+" --indicates 1-2 fold improvement
"++"—indicates >2 and <3 fold improvement
"+++"—indicates >3 and <4 fold improvement
"++++"—indicates >4 fold improvement

Analysis of Variants Expressed in CF-410

[0247] The beta-xylosidase variants were analyzed using various assays, such as those described below. In addition, some of the thermostable variants were sequenced, as described below.

A. *p*NP-X Thermostability Assays

[0248] To assess the thermostability of beta-xylosidase variants expressed in CF-410, broth supernatants were diluted 1:9 in 100 mM MES pH 6.0 and heated to either 22 ˚C or 55 ˚C for 20 h. Samples were diluted 1:1 with water, and 10 uL of diluted sample was added to 90 ul of 5 mM *p*NPX in 100 mM MES pH 6.0. Samples were incubated for 15 minutes at 37 ˚C, quenched with 150 uL of 1 M Na_2CO_3, and absorbance was measured at 400 nm. The results are shown in the graph presented in Figure 3. The best variants lie above the diagonal defined by the positive control.

B. XOS Thermoactivity Assays

[0249] Assessment of the thermoactivity of the best variants on xylose oligosaccharides was performed. Reactions were set up containing 10 mg/ml xylooligosaccharides, 100 mM sodium acetate pH 6.0, and 1% CF-410 supernatant from the best strains. Reactions were incubated at 37 ˚ C or 57 ˚C for 4 h, heated to 95 ˚C for 5 minutes to inactivate the enzymes, and then analyzed by HPLC, using standard methods known in the art. The results are shown in Figure 4.

C. Sequencing of Heat Stable Variants

[0250] The corresponding plasmid samples transformed into CF-410 that produced heat-stable xylosidase activities were sequenced. Briefly, the gene encoding the wild-type *M. thermophila* beta- xylosidase were PCR amplified from the plasmid templates using the following oligos; 5'-tgtgctgatcctcttccgtcatgaaggcctctgtatcatgcct (SEQ ID NO:131) and 5'-gaggttcgtttacttacttattacctgtgcctccccctggc (SEQ ID NO:132). Each PCR reaction was set up using 16.8 ul water, 5 ul of 5x Kapa buffer B (Kappa Biosystems), 0.5 ul of dNTP (CleanAmp™ 7-deaza-dGTP mix, TriLink), 1.25 ul of each oligo indicated above (20 μM stock concentration), 1ul of plasmid DNA and 0.2 ul of Kapa 2G robust hot start polymerase (Kapa Biosystems). Thermocycler conditions were: 95°C 3', 35 cycles of 95°C 30", 72°C 2", with a final extension at 72°C 5'. After the PCR reaction was completed, 8 ul of ExoSAP-IT (USB) was added into each sample and incubated at 37°C for 20 minutes

followed by enzyme denaturation at 80°C for 15 minutes. The oligonucleotides shown in Table 2-3 were used to sequence the variants. Ten variants were sequenced and the amino acid modifications are shown in Table 2-6 (as compared with the wild-type).

Table 2-3. Sequencing Oligonucleotides

Oligo Name	Sequence (5'-3')
2290-75-Fwd	ACCCCGACTGCACCAAGC (SEQ ID NO:133)
2290-211-Rev	CGC ATA CAT ACC TGA CCA GG (SEQ ID NO:134)
2290-465-Fwd	CGATGCCGCTGCTGATGG (SEQ ID NO:135)
2290-695-Rev	CGA GCC CGC GGA TCA TGG (SEQ ID NO:136)
2290-870-Fwd	TGGCGCCGTTCCAGCAGTG (SEQ ID NO:137)
2290-1059-Rev	CCG AGA CGT CGA GGA C (SEQ ID NO:138)
2290-1275-Fwd	TGGGCTGGGCCGACGTCAA (SEQ ID NO:139)
2290-1476-Rev	CCG CCA AAC AGC TTG TCC (SEQ ID NO:140)
2290-1690-Fwd	CAAGGACCGGATGACGATCG (SEQ ID NO:141)
2290-1887-Rev	TGA GCA GCC GGA CCA C (SEQ ID NO:142)
2290-2060-Fwd	ACCTTCCGGGCCGAGTTCG (SEQ ID NO:143)
2290-2373-Fwd	GAT CAA GAC GCT GGT CTC G (SEQ ID NO:144)
2290-2392-Rev	CGA GAC CAG CGT CTT GAT C (SEQ ID NO:145)

Table 2-4. Variants and Their Substitutions (Compared to Wild-Type *M. thermophila* Beta-Xylosidase SEQ ID NO:2)

Variant	Mutations
985	M280L
873	L115I/S211A
016	L115I/N219Y/W572Y
914	L115I/N219Y/W572Y
989	S108A/S211A/M280L/W572Y
902	S211A/N219Y/M280L/G322A
983	S211A/N219Y/M280L/W572Y
920	S211A/N219Y/M280L/W572Y
963	L115I/S211A/M280L/G322A/W572Y
949	N219Y/G322A/W572Y

[0251] While particular embodiments of the present invention have been illustrated and described, it will be apparent to those skilled in the art that various other changes and modifications can be made without departing from the spirit and scope of the present

invention. Therefore, it is intended that the present invention encompass all such changes and modifications with the scope of the present invention.

[0252] The present invention has been described broadly and generically herein. Each of the narrower species and subgeneric groupings falling within the generic disclosure also form part(s) of the invention. The invention described herein suitably may be practiced in the absence of any element or elements, limitation or limitations which is/are not specifically disclosed herein. The terms and expressions which have been employed are used as terms of description and not of limitation. There is no intention that in the use of such terms and expressions, of excluding any equivalents of the features described and/or shown or portions thereof, but it is recognized that various modifications are possible within the scope of the claimed invention. Thus, it should be understood that although the present invention has been specifically disclosed by some preferred embodiments and optional features, modification and variation of the concepts herein disclosed may be utilized by those skilled in the art, and that such modifications and variations are considered to be within the scope of the present invention.

CLAIMS

We claim:

1. A recombinant beta-xylosidase variant and/or biologically active recombinant beta-xylosidase variant fragment comprising at least one amino acid sequence comprising at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% sequence identity to SEQ ID NO:2 and comprising at least one mutation at position 31, 108, 115, 174, 177, 203, 209, 211, 219, 235, 264, 280, 309, 320, 322, 345, 347, 375, 379, 389, 394, 398, 431, 438, 449, 475, 482, 484, 499, 525, 539, 560, 565, 571, 572, 589, 662, 727, 761, 763, 798, and/or 842, wherein the positions are numbered with reference to SEQ ID NO:2.

2. The recombinant beta-xylosidase variant and/or biologically active recombinant beta-xylosidase variant fragment of Claim 1, comprising:

(i) at least one amino acid sequence comprising at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% sequence identity to SEQ ID NO:2 and comprising at least one mutation at position P31, S108, L115, V174, L177, G203, V209, S211, N219, V235, A264, M280, A309, G320, G322, S345, G347, A354, P375, H379, R389, E394, R398, R431, F438, G449, G475, D482, D484, A499, G525, R539, E560, G565, N571, N572, W572, E589, D662, E727, L761, G763, I798 , and/or G842, wherein the positions are numbered with reference to SEQ ID NO:2;

(ii) at least one amino acid sequence comprising at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% sequence identity to SEQ ID NO:2 and comprising at least one mutation at position P31G, S108A, L115I, V174P, L177G, G203C, V209I, S211A, N219Y, V235I, V235L, V235R, A264S, M280L, A309L, G320A, G322A, S345L, G347Q, A354V, P375E, P375S, H379Y, R389T, E394L, R398N, R431W, F438P, G449N, G475T, D482G, D484P, A499S, A499K, G525R, R539G, R539H, R539Q,

R539S, R560D, G565N, N571G, W572Y, R589K, D662N, E727D, E727T, L761I, G763P, I798V, and/or G842A, wherein the positions are numbered with reference to SEQ ID NO:2; or

(iii) at least one amino acid sequence comprising at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% sequence identity to SEQ ID NO:2, wherein said amino acid sequence comprises SEQ ID NO:5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, and/or 59.

3. The recombinant beta-xylosidase variant or biologically active beta-xylosidase variant fragment of Claim 1, wherein said beta-xylosidase variant is a *Myceliophthora thermophila* beta-xylosidase variant.

4. An enzyme composition comprising at least one beta-xylosidase variant and/or at least one biologically active beta-xylosidase variant fragment of any of Claims 1-3, optionally further comprising:

(i) at least one additional enzyme;

(ii) one or more enzymes selected from cellulases, hemicellulases, xylanases, amylases, glucoamylases, proteases, esterases, and lipases; and/or

(iii) one or more enzyme(s) selected from endoglucanases (EG), β-glucosidases (BGL), Type 1 cellobiohydrolases (CBH1), Type 2 cellobiohydrolases (CBH2), GH61 enzymes, and/or xylanases.

5. A recombinant organism comprising at least one beta-xylosidase variant and/or at least one biologically active beta-xylosidase variant fragment of any of Claims 1-3.

6. A recombinant fungal organism comprising at least one polynucleotide comprising at least one nucleic acid sequence encoding at least one beta-xylosidase variant and/or at least one biologically active fragment of any of Claims 1-3, and/or at least one polynucleotide that hybridizes under stringent hybridization conditions to the polynucleotide and/or a complement of a polynucleotide that encodes a polypeptide comprising the amino acid sequence provided in any of Claims 1-3, optionally wherein said polynucleotide comprises a sequence that has least 70%, at least 75%, at least 80%, at least 85%, at least

90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identity to SEQ ID NOS:1, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, and/or 58.

7. A recombinant nucleic acid construct comprising at least one polynucleotide sequence, wherein the polynucleotide is selected from:

(a) a polynucleotide that encodes a polypeptide comprising an amino acid sequence comprising at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identity to SEQ ID NO: 2, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, and/or 59;

(b) a polynucleotide that hybridizes under stringent hybridization conditions to at least a fragment of a polynucleotide that encodes a polypeptide comprising the amino acid sequence of SEQ ID NO: 2, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, and/or 59; and/or

(c) a polynucleotide that hybridizes under stringent hybridization conditions to the complement of at least a fragment of a polynucleotide that encodes a polypeptide comprising the amino acid sequence of SEQ ID NO: 2, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55 and/or 59.

8. The recombinant nucleic acid construct of Claim 7, wherein:

(i) the polynucleotide sequence is at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO:1,4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, and/or 58;

(ii) the polynucleotide sequence is operably linked to a promoter, optionally wherein said promoter is a heterologous promoter; and/or

(iii) said nucleic acid sequence is operably linked to at least one additional regulatory sequence.

9. A recombinant host cell that expresses at least one polynucleotide sequence

encoding at least one beta-xylosidase variant and/or at least one biologically active beta-xylosidase fragment of any of Claims 1-3.

10. The recombinant host cell of Claim 9, wherein:

(i) said host cell comprises at least one nucleic acid construct as provided in any of Claim 7 and/or 8;

(ii) said host cell comprises at least one polypeptide sequence set forth in SEQ ID NO: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, and/or 59;

(iii) said host cell comprises at least one polynucleotide sequence set forth in SEQ ID NO: 1,4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, and/or 58;

(iv) at least beta-xylosidase variant and/or at least one biologically active beta-xylosidase fragment is produced by said cell, optionally wherein at least one beta-xylosidase variant and/or at least one biologically active beta-xylosidase fragment is secreted from the host cell;

(v) said host cell further produces at least one enzyme selected from endoglucanases (EG), β-glucosidases (BGL), Type 1 cellobiohydrolases (CBH1), Type 2 cellobiohydrolases (CBH2), GH61 enzymes, and xylanases;

(vi) said cell produces at least two recombinant cellulases;

(vii) said cell produces at least three, at least four, or at least five recombinant cellulases;

(viii) said cell is a prokaryotic or eukaryotic cell, optionally wherein said cell is a yeast cell or filamentous fungal cell, such as wherein the filamentous fungal cell is a *Myceliophthora*, a *Thielavia*, a *Trichoderma*, and/or an *Aspergillus* cell; and/or

(ix) said cell is selected from *Saccharomyces* and *Myceliophthora*, optionally wherein the filamentous fungal cell is a *Myceliophthora thermophila* or wherein the yeast cell is *Saccharomyces cerevisiae*.

11. A method for producing at least one fermentable sugar from a feedstock, comprising contacting the feedstock with at least one recombinant beta xylosidase and/or biologically active beta xylosidase fragment of Claim 1 and/or the recombinant host cell set forth in any of Claims 8 to 10, and/or the enzyme composition provided in Claim 4, under culture conditions whereby fermentable sugars are produced.

12. The method of Claim 11:

(i) wherein the enzyme composition and/or recombinant host cell further comprises at least one enzyme selected from endoglucanases (EG), β-glucosidases (BGL), Type 1 cellobiohydrolases (CBH1), Type 2 cellobiohydrolases (CBH2), GH61s, and xylanases, such as wherein said at least one enzyme is a recombinant enzyme and/or wherein said at least one enzyme is a heterologous enzyme;

(ii) further comprising pretreating the feedstock prior to said contacting;

(iii) wherein the feedstock comprises wheat grass, wheat straw, barley straw, sorghum, rice grass, sugarcane, sugar beet, bagasse, switchgrass, corn stover, corn fiber, grains, or a combination thereof;

(iv) wherein the fermentable sugar comprises glucose and/or xylose;

(v) further comprising recovering at least one fermentable sugar;

(vi) further comprising contacting the at least one fermentable sugar with a microorganism under conditions such that said microorganism produces at least one fermentation end product, optionally wherein said fermentation end product is selected from alcohols, fatty acids, lactic acid, acetic acid, 3-hydroxypropionic acid, acrylic acid, succinic acid, citric acid, malic acid, fumaric acid, succinic acid, amino acids, 1,3-propanediol, ethylene, glycerol, and β-lactams, such as wherein said fermentation product is an alcohol selected from ethanol and butanol, preferably wherein said alcohol is ethanol; and/or

(vii) wherein the feedstock is a cellulosic and/or lignocellulosic feedstock.

13. A method of producing an end product from a feedstock, comprising:

a) contacting the feedstock with at least one recombinant beta-xylosidase and/or at least one biologically active beta-xylosidase fragment of Claim 1 and/or the recombinant host cell set forth in any of Claims 8 to 10, and/or the enzyme composition provided in Claim 4, under conditions whereby at least one fermentable sugar is produced from the substrate; and

b) contacting the fermentable sugar with a microorganism under conditions such that the microorganism uses the fermentable sugar to produce an end-product.

14. The method of Claim 13, wherein:

(i) the recombinant organism and/or recombinant host cell further comprises at least one enzyme selected from endoglucanases (EG), β-glucosidases (BGL), Type 1

cellobiohydrolases (CBH1),Type 2 cellobiohydrolases (CBH2), GH61 enzymes, and xylanases;

(ii) said at least one enzyme is a recombinant enzyme and/or said at least one enzyme is a heterologous enzyme;

(iii) the method comprises a simultaneous saccharification and fermentation reactions (SSF) or wherein the method comprises separate saccharification and fermentation reactions (SHF); and/or

(iv) the feedstock is a cellulosic and/or lignocellulosic feedstock.

15. A method of producing a fermentation end product from a feedstock, comprising:

a) obtaining at least one fermentable sugar produced according to the method of Claim 11 and/or 12; and

b) contacting the fermentable sugar with a microorganism in a fermentation to produce at least one fermentation end product,

optionally wherein:

(i) said fermentation end product is selected from alcohols, fatty acids, lactic acid, acetic acid, 3-hydroxypropionic acid, acrylic acid, succinic acid, citric acid, malic acid, fumaric acid, succinic acid, amino acids, 1,3-propanediol, ethylene, glycerol, and β-lactams, such as wherein said fermentation end product is at least one alcohol selected from ethanol and butanol;

(ii) wherein the microorganism is a yeast; and/or

(iii) further comprising recovering the fermentation end product.

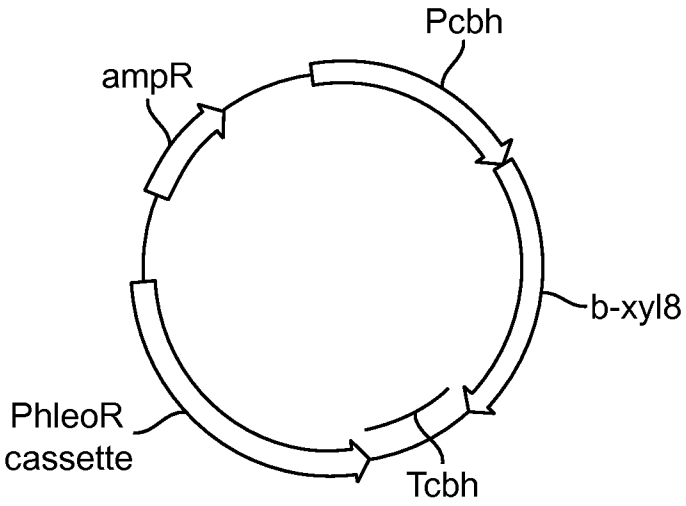
Pcbh
ampR
b-xyl8
PhleoR
cassette
Tcbh

FIG. 1

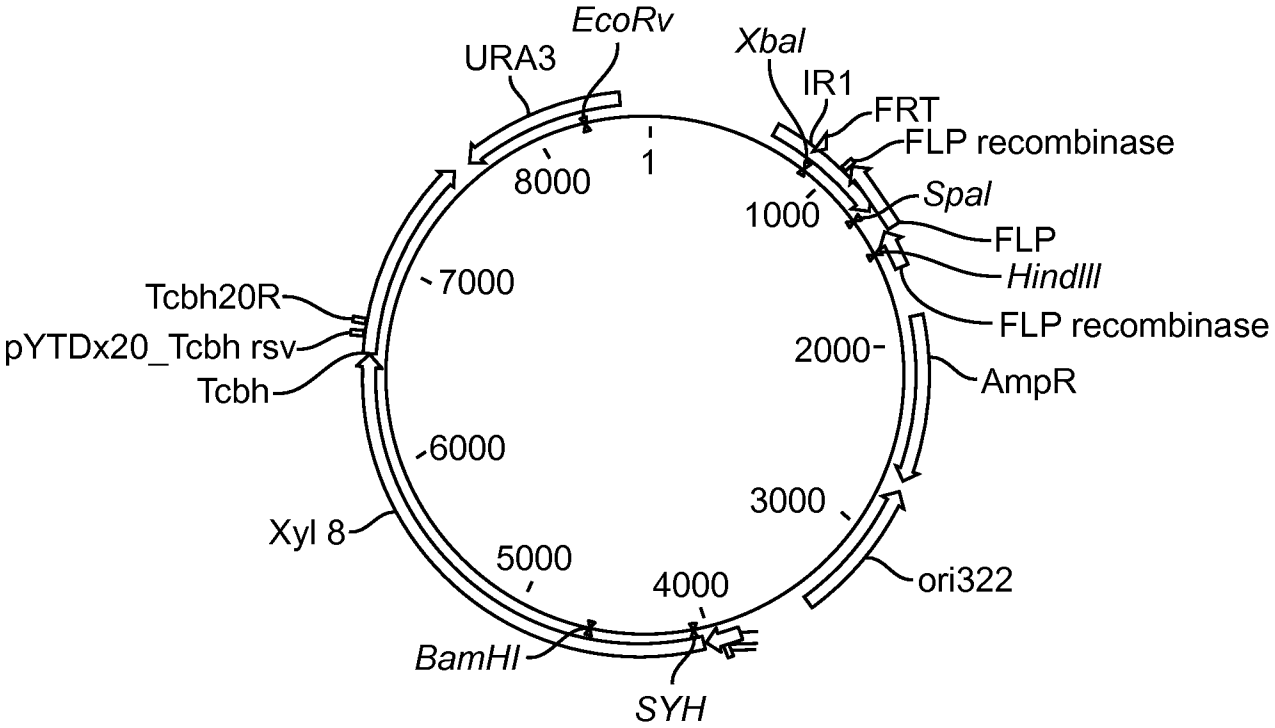
EcoRv
XbaI
URA3
IR1
FRT
FLP recombinase
8000
1
1000
SpaI
FLP
HindIII
FLP recombinase
Tcbh20R
7000
pYTDx20_Tcbh rsv
2000
AmpR
Tcbh
6000
3000
Xyl 8
5000
4000
ori322
BamHI
SYH

FIG. 2

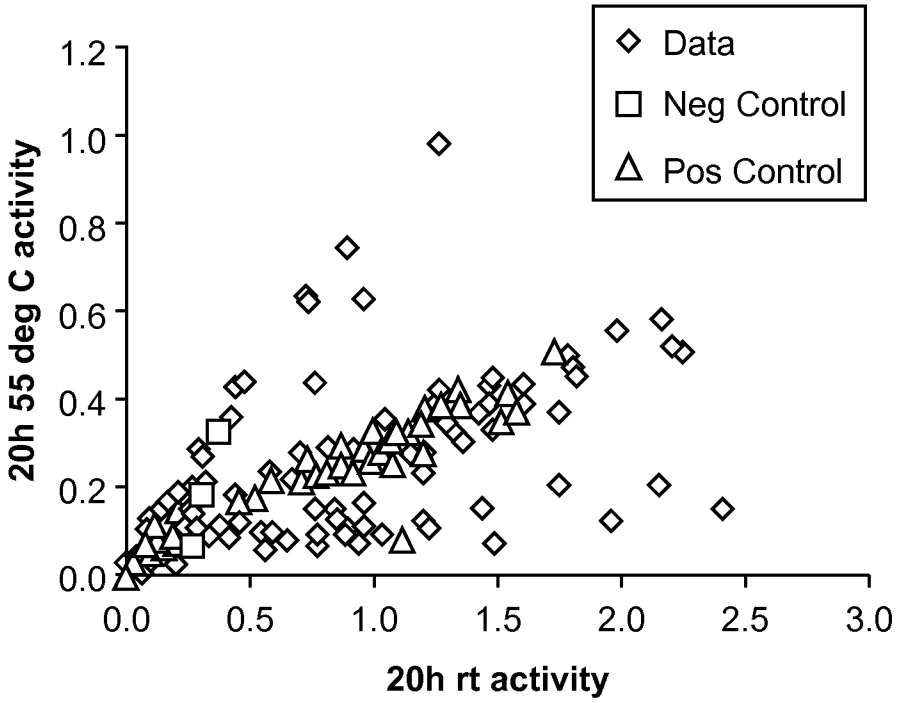
Data
Neg Control
Pos Control
20h 55 deg C activity
20h rt activity
0.0
0.2
0.4
0.6
0.8
1.0
1.2
0.0
0.5
1.0
1.5
2.0
2.5
3.0

FIG. 3

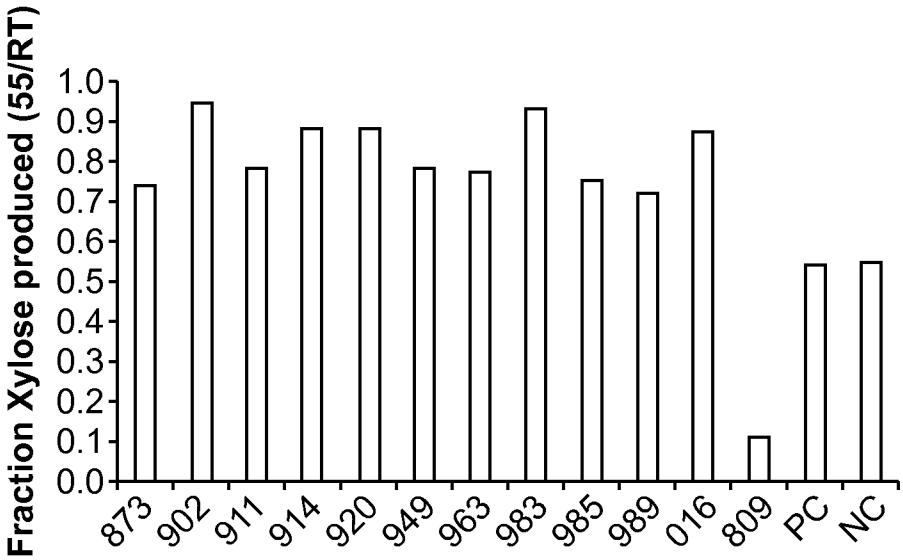
Fraction Xylose produced (55/RT)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
873
902
911
914
920
949
963
983
985
989
016
809
PC
NC

FIG. 4

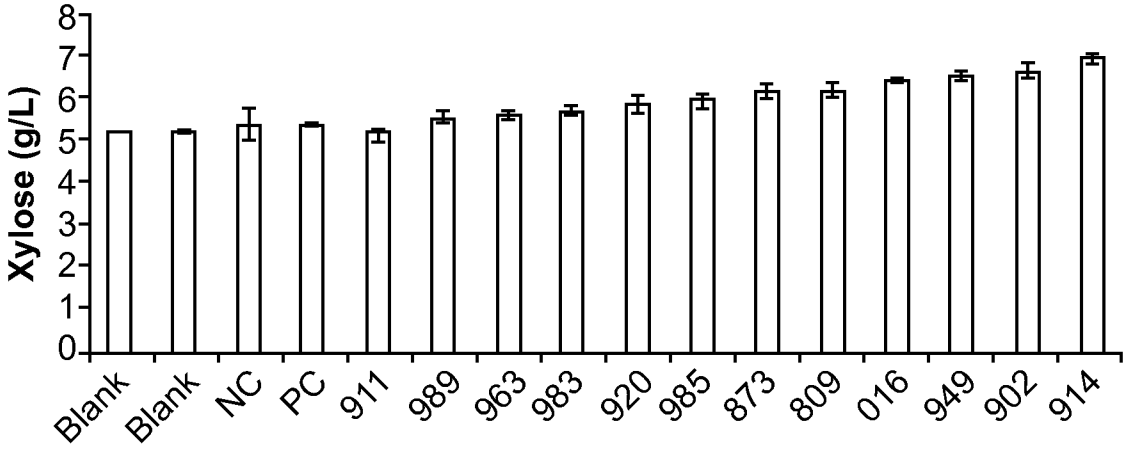
Xylose (g/L)
0
1
2
3
4
5
6
7
8
Blank
Blank
NC
PC
911
989
963
983
920
985
873
809
016
949
902
914

FIG. 5