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

(19) World Intellectual Property Organization
International Bureau

PCT

(43) International Publication Date
23 February 2012 (23.02.2012)

(10) International Publication Number
WO 2012/024662 A2

(51) International Patent Classification:
C12P 21/00 (2006.01) *C12N 1/15* (2006.01)
C12N 15/113 (2010.01) *C12N 1/19* (2006.01)
C12N 15/80 (2006.01)

(21) International Application Number:
PCT/US2011/048560

(22) International Filing Date:
22 August 2011 (22.08.2011)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

61/375,760	20 August 2010 (20.08.2010)	US
61/375,755	20 August 2010 (20.08.2010)	US
61/375,753	20 August 2010 (20.08.2010)	US
61/375,745	20 August 2010 (20.08.2010)	US
61/375,702	20 August 2010 (20.08.2010)	US

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

(72) **Inventors; and**
(75) **Inventors/Applicants** *(for US only)*: **SHOCK, Jennifer, L.** [US/US]; 200 Penobscot Drive, Redwood City, California 94063 (US). **CLARK, Louis** [US/US]; 200 Penobscot Drive, Redwood City, California 94063 (US).

(74) **Agents**: **MACKNIGHT, Kamrin, T.** et al.; Codexis, Inc., 200 Penobscot Drive, Redwood City, CA 94063 (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, 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, KM, 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, PE, PG, PH, PL, PT, QA, RO, RS, RU, SC, 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, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, 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, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: EXPRESSION CONSTRUCTS COMPRISING FUNGAL PROMOTERS

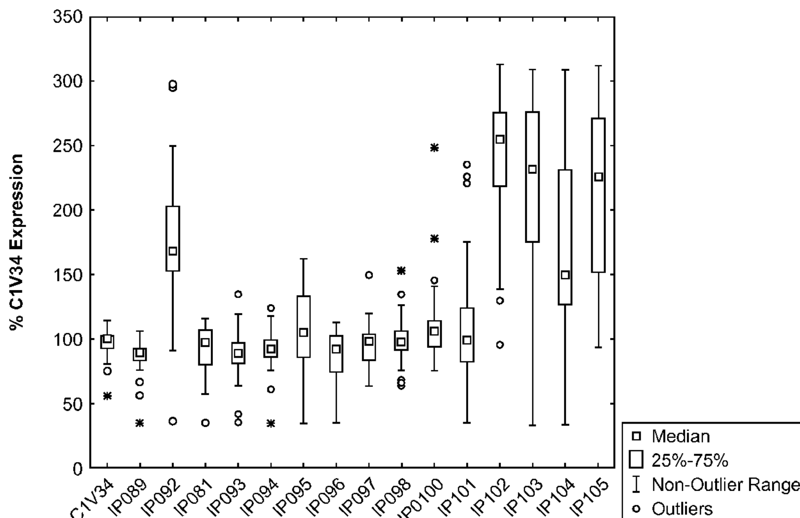

FIG. 2

(57) Abstract: The present invention provides promoters derived from a filamentous fungus. These promoters have application in the fields of molecular biology, microbiology, fungal genetics and production of biofuels and other products.

WO 2012/024662 A2

Published:

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

EXPRESSION CONSTRUCTS COMPRISING FUNGAL PROMOTERS

[0001] The present application claims priority to US Prov. Patent Appln. Ser. Nos. 61/375,702, 61/375,745, 61/375,753, 61/375,755, and 61/375,760, all of which were filed on August 20, 2010, and are hereby incorporated by reference in their entireties.

FIELD OF THE INVENTION

[0002] The present invention provides promoters derived from a filamentous fungus. These promoters have application in the fields of molecular biology, microbiology, fungal genetics and production of biofuels and other products.

BACKGROUND

[0003] In many commercial applications using recombinant host cells, strong promoters are required to express commercially useful amounts of desired proteins in the cell. Although numerous promoters are known in the art, only a limited number of promoters from filamentous fungi have been characterized.

SUMMARY OF THE INVENTION

[0004] The present invention provides promoters derived from a filamentous fungus. These promoters have application in the fields of molecular biology, microbiology, fungal genetics and production of biofuels and other products.

[0005] The present invention provides isolated polynucleotides having promoter activity comprising: (a) a nucleotide sequence set forth in any of SEQ ID NOS:1-40; (b) a subsequence of (a) comprising at least about 75 contiguous nucleotides set forth in any of SEQ ID NOS:1-40; (c) a nucleotide sequence having at least about 90% sequence identity to (a) or (b); and/or (d) a nucleotide sequence that hybridizes to any of SEQ ID NOS:1-40, and/or the complement thereof. Also provided are expression constructs comprising the isolated promoter operably linked to a heterologous DNA sequence encoding a protein. Also provided are expression constructs comprising the isolated promoter operably linked to a heterologous DNA sequence encoding a protein. It is contemplated that any suitable protein will find use in the present invention. In some embodiments, the protein is an enzyme. In some embodiments, the enzyme is a cellulase (*e.g.*, an endoglucanase, a cellobiohydrolase or a β-glucosidase), a glucoamylase, a protease, an alpha amylase, a hemicellulase, a xylanase, an esterase, a cutinase, a phytase, a lipase, an oxidoreductase, a laccase, an isomerase, a pullulanase, a phenol oxidizing enzyme, a

mannase, a catalase, a glucose oxidase, a transferase or a lyase. In some embodiments, the protein comprises a signal peptide fused to a secreted protein sequence, which may be signal peptide not associated with the secreted protein in nature.

[0006] The present invention also provides isolated polynucleotides having C1 promoter activity comprising: (a) a nucleotide sequence set forth in any of SEQ ID NOS:1-40; (b) a subsequence of (a) comprising at least about 75 contiguous nucleotides set forth in any of SEQ ID NOS:1-40; (c) a nucleotide sequence having at least about 90% sequence identity to (a) or (b); and/or (d) a nucleotide sequence that hybridizes to any of SEQ ID NOS:1-40, and/or the complement thereof. Also provided are expression constructs comprising the isolated promoter operably linked to a heterologous DNA sequence encoding a protein. Also provided are expression constructs comprising the isolated promoter operably linked to a heterologous DNA sequence encoding a protein. It is contemplated that any suitable protein will find use in the present invention. In some embodiments, the protein is an enzyme. In some embodiments, the enzyme is a cellulase (*e.g.*, an endoglucanase, a cellobiohydrolase or a β-glucosidase), a glucoamylase, a protease, an alpha amylase, a hemicellulase, a xylanase, an esterase, a cutinase, a phytase, a lipase, an oxidoreductase, a laccase, an isomerase, a pullulanase, a phenol oxidizing enzyme, a mannase, a catalase, a glucose oxidase, a transferase or a lyase. In some embodiments, the protein comprises a signal peptide fused to a secreted protein sequence, which may be signal peptide not associated with the secreted protein in nature.

[0007] In some embodiments, an expression cassette comprising at least one expression construct comprising at least one of the promoters described herein is provided. In some embodiments, an expression vector is provided. In some additional embodiments, cells comprising the expression cassette are provided. In some embodiments, the cell is a yeast cell or a filamentous fungal cell (*e.g.*, C1 cell).

[0008] The present invention also provides processes for producing at least one protein in a host cell by culturing a cell containing the nucleic acid, expression cassette and/or expression vector.

[0009] In some embodiments, the present invention provides isolated and/or recombinant polynucleotides having C1 promoter activity comprising: (a) nucleotide sequence set forth in any of SEQ ID NOS:1-40; (b) a subsequence of (a) comprising at least about 75, about 100, about 200, about 300, about 400, or about 500 contiguous nucleotides set forth in any of SEQ ID NOS:1-40; (c) a nucleotide sequence having at least about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% sequence identity to (a) or (b); and/or (d) a nucleotide sequence that hybridizes to any of SEQ ID NOS:1-40, and/or the complement thereof.

[0010] The present invention provides isolated and/or recombinant polynucleotide with C1 promoter activity comprising: (a) a nucleotide sequence of SEQ ID NO:1, 9, 17, 25 and/or 33; (b) a

subsequence of (a) comprising at least about 75 contiguous nucleotides of SEQ ID NO:1, 9, 17, 25 and/or 33; (c) a nucleotide sequence having at least about 90% sequence identity to (a) or (b); and/or (d) a nucleotide sequence that hybridizes to the complement of SEQ ID NO:1, 9, 17, 25 and/or 33.

[0011] The present invention also provides isolated and/or recombinant polynucleotide with C1 promoter activity comprising: (a) a nucleotide sequence of SEQ ID NO:1, 9, 17, 25 and/or 33; (b) a subsequence of (a) comprising at least about 75, about 150, about 200, about 300, about 400, about 500, about 600, about 700, about 800, about 900, about 1000, about 1100, about 1200, about 1300, or about 1400 contiguous nucleotides of SEQ ID NO:1, 9, 17, 25 and/or 33; (c) a nucleotide sequence having at least about 90% sequence identity to (a) or (b); and/or (d) a nucleotide sequence that hybridizes to the complement of SEQ ID NO:1, 9, 17, 25 and/or 33.

[0012] The present invention also provides isolated and/or recombinant polynucleotides with C1 promoter activity comprising: (a) a nucleotide sequence of SEQ ID NO:1, 9, 17, 25 and/or 33; (b) a subsequence of (a) comprising at least about 75 contiguous nucleotides of SEQ ID NO:1, 9, 17, 25 and/or 33; (c) a nucleotide sequence having at least about 95% sequence identity to (a) or (b); and/or (d) a nucleotide sequence that hybridizes to the complement of SEQ ID NO:1, 9, 17, 25 and/or 33.

[0013] The present invention also provides isolated and/or recombinant polynucleotides with C1 promoter activity comprising: (a) a nucleotide sequence of SEQ ID NO:2-8, 10-16, 18-24, 26-32, and/or 34-40; (b) a subsequence of (a) comprising at least about 75 contiguous nucleotides of SEQ ID NO:2-8, 10-16, 18-24, 26-32, and/or 34-40; (c) a nucleotide sequence having at least about 90% sequence identity to (a) or (b), and/or (d) a nucleotide sequence that hybridizes to the complement of SEQ ID NO: 2-8, 10-16, 18-24, 26-32, and/or 34-40.

[0014] In some embodiments, the present invention also provides isolated and/or recombinant polynucleotides with C1 promoter activity comprising: (a) a nucleotide sequence of SEQ ID NO:2-8, 10-16, 18-24, 26-32, and/or 34-40; (b) a subsequence of (a) comprising at least about 75, about 150, about 200, about 300, about 400, or about 500 contiguous nucleotides of SEQ ID NO:2-8, 10-16, 18-24, 26-32, and/or 34-40; (c) a nucleotide sequence having at least about 90% sequence identity to (a) or (b), and/or (d) a nucleotide sequence that hybridizes to the complement of SEQ ID NO: 2-8, 10-16, 18-24, 26-32, and/or 34-40.

[0015] In some embodiments, the present invention also provides isolated and/or recombinant polynucleotides with C1 promoter activity comprising: (a) a nucleotide sequence of SEQ ID NO:2-8, 10-16, 18-24, 26-32, and/or 34-40; (b) a subsequence of (a) comprising at least about 75 contiguous nucleotides of SEQ ID NO:2-8, 10-16, 18-24, 26-32, and/or 34-40; (c) a nucleotide sequence having at least about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%,

about 98%, or about 99% sequence identity to (a) or (b), and/or (d) a nucleotide sequence that hybridizes to the complement of SEQ ID NO: 2-8, 10-16, 18-24, 26-32, and/or 34-40.

[0016] In some embodiments, the present invention also provides isolated and/or recombinant polynucleotides with C1 promoter activity comprising: (a) a nucleotide sequence set forth in any of SEQ ID NO:3, 11, 19, 27, and/or 35; (b) a subsequence of (a) comprising at least about 75, about 100, about 200, about 300, about 400, about 500, or about 600 contiguous nucleotides set forth in any of SEQ ID NO:3, 11, 19, 27, and/or 35; (c) a nucleotide sequence having at least about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% sequence identity to (a) or (b); or (d) a nucleotide sequence that hybridizes to any of SEQ ID NO:3, 11, 19, 27, and/or 35, and/or a complement thereof.

[0017] The present invention also provides isolated and/or recombinant polynucleotides having C1 promoter activity comprising: (a) a nucleotide sequence set forth in any of SEQ ID NO:4, 12, 20, 28, and/or 36; (b) a subsequence of (a) comprising at least about 75, about 100, about 200, about 300, about 400, or about 500 contiguous nucleotides set forth in any of SEQ ID NO:4, 12, 20, 28, and/or 36; (c) a nucleotide sequence having at least about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% sequence identity to (a) or (b); or (d) a nucleotide sequence that hybridizes to any of SEQ ID NO:4, 12, 20, 28, and/or 36, and/or a complement thereof.

[0018] The present invention also provide isolated and/or recombinant polynucleotides having C1 promoter activity comprising: (a) a nucleotide sequence set forth in any of SEQ ID NO:5, 13, 21, 29, and/or 37; (b) a subsequence of (a) comprising at least about 75, about 100, about 200, about 300, or about 400 contiguous nucleotides set forth in any of SEQ ID NO:5, 13, 21, 29, and/or 37; (c) a nucleotide sequence having at least about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% sequence identity to (a) or (b); or (d) a nucleotide sequence that hybridizes to any of SEQ ID NO:5, 13, 21, 29, and/or 37, and/or a complement thereof.

[0019] The present invention also provides isolated and/or recombinant polynucleotides having C1 promoter activity comprising: (a) a nucleotide sequence set forth in any of SEQ ID NO:6, 14, 22, 30, and/or 38; (b) a subsequence of (a) comprising at least about 75, about 100, about 200, about 300 or about 400 contiguous nucleotides set forth in any of SEQ ID NO:6, 14, 22, 30, and/or 38; (c) a nucleotide sequence having at least about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% sequence identity to (a) or (b); or (d) a nucleotide sequence that hybridizes to any of SEQ ID NO:6, 14, 22, 30, and/or 38, and/or a complement thereof.

[0020] The present invention also provides isolated and/or recombinant polynucleotides having C1 promoter activity comprising: (a) a nucleotide sequence set forth in any of SEQ ID NO:7, 15, 23, 31, and/or 39; (b) a subsequence of (a) comprising at least about 75, about 100, about 200, or about 300 contiguous nucleotides set forth in any of SEQ ID NO:7, 15, 23, 31, and/or 39; (c) a nucleotide sequence having at least about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% sequence identity to (a) or (b); or (d) a nucleotide sequence that hybridizes to any of SEQ ID NO:7, 15, 23, 31, and/or 39, and/or a complement thereof.

[0021] The present invention also provides isolated and/or recombinant polynucleotides having C1 promoter activity comprising: (a) a nucleotide sequence of set forth in any of SEQ ID NO:8, 16, 24, 32, and/or 40; (b) a subsequence of (a) comprising at least about 75, about 100, or about 200 contiguous nucleotides set forth in any of SEQ ID NO:8, 16, 24, 32, and/or 40; (c) a nucleotide sequence having at least about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% sequence identity to (a) or (b); or (d) a nucleotide sequence that hybridizes to any of SEQ ID NO:8, 16, 24, 32, and/or 40, and/or a complement thereof.

[0022] The present invention also provides expression constructs comprising at least one isolated polynucleotide having C1 promoter activity comprising: polynucleotides having C1 promoter activity comprising: (a) a nucleotide sequence set forth in any of SEQ ID NOS:1-40; (b) a subsequence of (a) comprising at least about 75 contiguous nucleotides set forth in any of SEQ ID NOS:1-40; (c) a nucleotide sequence having at least about 90% sequence identity to (a) or (b); and/or (d) a nucleotide sequence that hybridizes to any of SEQ ID NOS:1-40, and/or the complement thereof. In some embodiments, the expression constructs comprise at least one isolated promoter operably linked to at least one heterologous DNA sequence encoding at least one protein. It is contemplated that any suitable protein will find use in the present invention. In some embodiments, the protein is an enzyme. In some embodiments, the protein comprises at least one recombinant protein. In some embodiments, the at least one recombinant protein is at least one recombinant enzyme. In some embodiments, the enzyme is a cellulase (*e.g.*, an endoglucanase, a cellobiohydrolase or a β-glucosidase), a glucoamylase, a protease, an alpha amylase, a hemicellulase, a xylanase, an esterase, a cutinase, a phytase, a lipase, an oxidoreductase, a laccase, an isomerase, a pullulanase, a phenol oxidizing enzyme, a mannase, a catalase, a glucose oxidase, a transferase or a lyase. In some embodiments, the protein comprises a signal peptide fused to a secreted protein sequence, which may be signal peptide not associated with the secreted protein in nature.

[0023] The present invention also provides expression constructs comprising at least one isolated polynucleotide having C1 promoter activity comprising: (a) nucleotide sequence set forth in any of SEQ ID NOS:1-40; (b) a subsequence of (a) comprising at least about 75, about 100, about 200, about 300,

about 400, or about 500 contiguous nucleotides set forth in any of SEQ ID NOS:1-40; (c) a nucleotide sequence having at least about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% sequence identity to (a) or (b); and/or (d) a nucleotide sequence that hybridizes to any of SEQ ID NOS:1-40, and/or the complement thereof. In some embodiments, the expression constructs comprise at least one isolated promoter operably linked to at least one heterologous DNA sequence encoding at least one protein. It is contemplated that any suitable protein will find use in the present invention. In some embodiments, the protein is an enzyme. In some embodiments, the protein comprises at least one recombinant protein. In some embodiments, the at least one recombinant protein is at least one recombinant enzyme. In some embodiments, the enzyme is a cellulase (*e.g.*, an endoglucanase, a cellobiohydrolase or a β-glucosidase), a glucoamylase, a protease, an alpha amylase, a hemicellulase, a xylanase, an esterase, a cutinase, a phytase, a lipase, an oxidoreductase, a laccase, an isomerase, a pullulanase, a phenol oxidizing enzyme, a mannase, a catalase, a glucose oxidase, a transferase or a lyase. In some embodiments, the protein comprises a signal peptide fused to a secreted protein sequence, which may be signal peptide not associated with the secreted protein in nature.

[0024] The present invention also provides expression constructs comprising at least one isolated polynucleotide having C1 promoter activity comprising: (a) a nucleotide sequence set forth in any of SEQ ID NO:1, 9, 17, 25, and/or 33; (b) a subsequence of (a) comprising at least about 75, about 100, about 200, about 300, about 400, about 500, about 600, about 700, about 800, about 900, about 1000, about 1100, about 1200, about 1300, or about 1400 contiguous nucleotides set forth in any of SEQ ID NO:1, 9, 17, 25, and/or 33; (c) a nucleotide sequence having at least about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% sequence identity to (a) or (b); or (d) a nucleotide sequence that hybridizes to any of SEQ ID NO:1, 9, 17, 25, and/or 33, and/or a complement thereof. In some embodiments, the expression constructs comprise at least one isolated promoter operably linked to at least one heterologous DNA sequence encoding at least one protein. It is contemplated that any suitable protein will find use in the present invention. In some embodiments, the protein is an enzyme. In some embodiments, the protein comprises at least one recombinant protein. In some embodiments, the at least one recombinant protein is at least one recombinant enzyme. In some embodiments, the enzyme is a cellulase (*e.g.*, an endoglucanase, a cellobiohydrolase or a β-glucosidase), a glucoamylase, a protease, an alpha amylase, a hemicellulase, a xylanase, an esterase, a cutinase, a phytase, a lipase, an oxidoreductase, a laccase, an isomerase, a pullulanase, a phenol oxidizing enzyme, a mannase, a catalase, a glucose oxidase, a transferase or a lyase. In some embodiments, the protein comprises a signal peptide fused to a secreted protein sequence, which may be signal peptide not associated with the secreted protein in nature.

[0025] The present invention also provides expression constructs comprising at least one isolated polynucleotide having C1 promoter activity comprising: (a) a nucleotide sequence set forth in any of SEQ ID NO:2, 10, 18, 26, and/or 34; (b) a subsequence of (a) comprising at least about 75, about 100, about 200, about 300, about 400, about 500, about 600, about 700, about 800, or about 900 contiguous nucleotides set forth in any of SEQ ID NO:2, 10, 18, 26, and/or 34; (c) a nucleotide sequence having at least about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% sequence identity to (a) or (b); or (d) a nucleotide sequence that hybridizes to any of SEQ ID NO:2, 10, 18, 26, and/or 34, and/or a complement thereof. In some embodiments, the expression constructs comprise the isolated promoter operably linked to a heterologous DNA sequence encoding a protein. It is contemplated that any suitable protein will find use in the present invention. In some embodiments, the protein is an enzyme. In some embodiments, the protein comprises at least one recombinant protein. In some embodiments, the at least one recombinant protein is at least one recombinant enzyme. In some embodiments, the enzyme is a cellulase (*e.g.*, an endoglucanase, a cellobiohydrolase or a β-glucosidase), a glucoamylase, a protease, an alpha amylase, a hemicellulase, a xylanase, an esterase, a cutinase, a phytase, a lipase, an oxidoreductase, a laccase, an isomerase, a pullulanase, a phenol oxidizing enzyme, a mannase, a catalase, a glucose oxidase, a transferase or a lyase. In some embodiments, the protein comprises a signal peptide fused to a secreted protein sequence, which may be signal peptide not associated with the secreted protein in nature.

[0026] The present invention also provides expression constructs comprising at least one isolated polynucleotide having C1 promoter activity comprising: (a) a nucleotide sequence set forth in any of SEQ ID NO:3, 11, 19, 27, and/or 35; (b) a subsequence of (a) comprising at least about 75, about 100, about 200, about 300, about 400, about 500 or about 600 contiguous nucleotides set forth in any of SEQ ID NO:3, 11, 19, 27, and/or 35; (c) a nucleotide sequence having at least about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% sequence identity to (a) or (b); or (d) a nucleotide sequence that hybridizes to any of SEQ ID NO:3, 11, 19, 27, and/or 35, and/or a complement thereof. In some embodiments, the expression constructs comprise at least one isolated promoter operably linked to at least one heterologous DNA sequence encoding at least one protein. It is contemplated that any suitable protein will find use in the present invention. In some embodiments, the protein is an enzyme. In some embodiments, the protein comprises at least one recombinant protein. In some embodiments, the at least one recombinant protein is at least one recombinant enzyme. In some embodiments, the enzyme is a cellulase (*e.g.*, an endoglucanase, a cellobiohydrolase or a β-glucosidase), a glucoamylase, a protease, an alpha amylase, a hemicellulase, a xylanase, an esterase, a cutinase, a phytase, a lipase, an oxidoreductase, a laccase, an isomerase, a

pullulanase, a phenol oxidizing enzyme, a mannase, a catalase, a glucose oxidase, a transferase or a lyase. In some embodiments, the protein comprises a signal peptide fused to a secreted protein sequence, which may be signal peptide not associated with the secreted protein in nature.

[0027] The present invention also provides expression constructs comprising at least one isolated polynucleotide having C1 promoter activity comprising: (a) a nucleotide sequence set forth in any of SEQ ID NO:4, 12, 20, 28, and/or 36; (b) a subsequence of (a) comprising at least about 75, about 100, about 200, about 300, about 400, or about 500 contiguous nucleotides set forth in any of SEQ ID NO:4, 12, 20, 28, and/or 36; (c) a nucleotide sequence having at least about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% sequence identity to (a) or (b); or (d) a nucleotide sequence that hybridizes to any of SEQ ID NO:4, 12, 20, 28, and/or 36, and/or a complement thereof. In some embodiments, the expression constructs comprise at least one isolated promoter operably linked to at least one heterologous DNA sequence encoding at least one protein. It is contemplated that any suitable protein will find use in the present invention. In some embodiments, the protein is an enzyme. In some embodiments, the protein comprises at least one recombinant protein. In some embodiments, the at least one recombinant protein is at least one recombinant enzyme. In some embodiments, the enzyme is a cellulase (*e.g.*, an endoglucanase, a cellobiohydrolase or a β-glucosidase), a glucoamylase, a protease, an alpha amylase, a hemicellulase, a xylanase, an esterase, a cutinase, a phytase, a lipase, an oxidoreductase, a laccase, an isomerase, a pullulanase, a phenol oxidizing enzyme, a mannase, a catalase, a glucose oxidase, a transferase or a lyase. In some embodiments, the protein comprises a signal peptide fused to a secreted protein sequence, which may be signal peptide not associated with the secreted protein in nature.

[0028] The present invention also provides expression constructs comprising at least one isolated polynucleotide having C1 promoter activity comprising: (a) a nucleotide sequence set forth in any of SEQ ID NO:5, 13, 21, 29, and/or 37; (b) a subsequence of (a) comprising at least about 75, about 100, about 200, about 300, or about 400 contiguous nucleotides set forth in any of SEQ ID NO:5, 13, 21, 29, and/or 37; (c) a nucleotide sequence having at least about 90%, about 91%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% sequence identity to (a) or (b); or (d) a nucleotide sequence that hybridizes to any of SEQ ID NO:5, 13, 21, 29, and/or 37, and/or a complement thereof. In some embodiments, the expression constructs comprise at least one isolated promoter operably linked to at least one heterologous DNA sequence encoding at least one protein. It is contemplated that any suitable protein will find use in the present invention. In some embodiments, the protein is an enzyme. In some embodiments, the protein comprises at least one recombinant protein. In some embodiments, the at least one recombinant protein is at least one recombinant enzyme. In some embodiments, the enzyme is a

cellulase (*e.g.*, an endoglucanase, a cellobiohydrolase or a β-glucosidase), a glucoamylase, a protease, an alpha amylase, a hemicellulase, a xylanase, an esterase, a cutinase, a phytase, a lipase, an oxidoreductase, a laccase, an isomerase, a pullulanase, a phenol oxidizing enzyme, a mannase, a catalase, a glucose oxidase, a transferase or a lyase. In some embodiments, the protein comprises a signal peptide fused to a secreted protein sequence, which may be signal peptide not associated with the secreted protein in nature.

[0029] The present invention also provides expression constructs comprising at least one isolated polynucleotide having C1 promoter activity comprising: (a) a nucleotide sequence set forth in any of SEQ ID NO:6, 14, 22, 30, and/or 38; (b) a subsequence of (a) comprising at least about 75, about 100, about 200, about 300, or about 400 contiguous nucleotides set forth in any of SEQ ID NO:6, 14, 22, 30, and/or 38; (c) a nucleotide sequence having at least about 90%, about 91%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% sequence identity to (a) or (b); or (d) a nucleotide sequence that hybridizes to any of SEQ ID NO:6, 14, 22, 30, and/or 38, and/or a complement thereof. In some embodiments, the expression constructs comprise at least one isolated promoter operably linked to at least one heterologous DNA sequence encoding at least one protein. It is contemplated that any suitable protein will find use in the present invention. In some embodiments, the protein is an enzyme. In some embodiments, the protein comprises at least one recombinant protein. In some embodiments, the at least one recombinant protein is at least one recombinant enzyme. In some embodiments, the enzyme is a cellulase (*e.g.*, an endoglucanase, a cellobiohydrolase or a β-glucosidase), a glucoamylase, a protease, an alpha amylase, a hemicellulase, a xylanase, an esterase, a cutinase, a phytase, a lipase, an oxidoreductase, a laccase, an isomerase, a pullulanase, a phenol oxidizing enzyme, a mannase, a catalase, a glucose oxidase, a transferase or a lyase. In some embodiments, the protein comprises a signal peptide fused to a secreted protein sequence, which may be signal peptide not associated with the secreted protein in nature.

[0030] The present invention also provides expression constructs comprising at least one isolated polynucleotide having C1 promoter activity comprising: (a) a nucleotide sequence set forth in any of SEQ ID NO:7, 15, 23, 31, and/or 39; (b) a subsequence of (a) comprising at least about 75, about 100, about 200, or about 300 contiguous nucleotides set forth in any of SEQ ID NO:7, 15, 23, 31, and/or 39; (c) a nucleotide sequence having at least about 90%, about 91%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% sequence identity to (a) or (b); or (d) a nucleotide sequence that hybridizes to any of SEQ ID NO:7, 15, 23, 31, and/or 39, and/or a complement thereof. In some embodiments, the expression constructs comprise at least one isolated promoter operably linked to a heterologous DNA sequence encoding a protein. It is contemplated that any suitable protein will find use in the present invention. In some embodiments, the protein is an enzyme. In some embodiments, the

expression constructs comprise at least one isolated promoter operably linked to at least one heterologous DNA sequence encoding at least one protein. It is contemplated that any suitable protein will find use in the present invention. In some embodiments, the protein is an enzyme. In some embodiments, the protein comprises at least one recombinant protein. In some embodiments, the at least one recombinant protein is at least one recombinant enzyme. In some embodiments, the enzyme is a cellulase (*e.g.*, an endoglucanase, a cellobiohydrolase or a β-glucosidase), a glucoamylase, a protease, an alpha amylase, a hemicellulase, a xylanase, an esterase, a cutinase, a phytase, a lipase, an oxidoreductase, a laccase, an isomerase, a pullulanase, a phenol oxidizing enzyme, a mannase, a catalase, a glucose oxidase, a transferase or a lyase. In some embodiments, the protein comprises a signal peptide fused to a secreted protein sequence, which may be signal peptide not associated with the secreted protein in nature.

[0031] The present invention also provides expression constructs comprising at least one isolated polynucleotide having C1 promoter activity comprising: (a) a nucleotide sequence of set forth in any of SEQ ID NO:8, 16, 24, 32, and/or 40; (b) a subsequence of (a) comprising at least about 75, about 100, or about 200 contiguous nucleotides set forth in any of SEQ ID NO:8, 16, 24, 32, and/or 40; (c) a nucleotide sequence having at least about 90%, about 91%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% sequence identity to (a) or (b); or (d) a nucleotide sequence that hybridizes to any of SEQ ID NO:8, 16, 24, 32, and/or 40, and/or a complement thereof. In some embodiments, the expression constructs comprise at least one isolated promoter operably linked to at least one heterologous DNA sequence encoding at least one protein. It is contemplated that any suitable protein will find use in the present invention. In some embodiments, the protein is an enzyme. In some embodiments, the protein comprises at least one recombinant protein. In some embodiments, the at least one recombinant protein is at least one recombinant enzyme. In some embodiments, the enzyme is a cellulase (*e.g.*, an endoglucanase, a cellobiohydrolase or a β-glucosidase), a glucoamylase, a protease, an alpha amylase, a hemicellulase, a xylanase, an esterase, a cutinase, a phytase, a lipase, an oxidoreductase, a laccase, an isomerase, a pullulanase, a phenol oxidizing enzyme, a mannase, a catalase, a glucose oxidase, a transferase or a lyase. In some embodiments, the protein comprises a signal peptide fused to a secreted protein sequence, which may be signal peptide not associated with the secreted protein in nature.

[0032] The present invention also provides expression constructs comprising at least one isolated polynucleotide having C1 promoter activity comprising: (i) a subsequence comprising at least about 100 contiguous nucleotides of SEQ ID NOS:3, 11, 19, 27, and/or 35; or (ii) a nucleotide sequence having at least about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% sequence identity to (i).

[0033] The present invention also provides expression constructs comprising at least one isolated polynucleotide having C1 promoter activity comprising: (i) a subsequence comprising at least about 150 contiguous nucleotides of SEQ ID NO:3, 11, 19, 27, and/or 35; or (ii) a nucleotide sequence having at least about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% sequence identity to (i).

[0034] The present invention also provides expression constructs comprising at least one isolated polynucleotide having C1 promoter activity comprising: (i) a subsequence comprising at least about 300 contiguous nucleotides of SEQ ID NO:3, 11, 19, 27, and/or 35; or (ii) a nucleotide sequence having at least about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% sequence identity to (i).

[0035] The present invention also provides expression constructs comprising at least one isolated polynucleotide having C1 promoter activity comprising: (i) a subsequence comprising at least about 400 contiguous nucleotides of SEQ ID NO:3, 11, 19, 27, and/or 35; or (ii) a nucleotide sequence having at least about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% sequence identity to (i).

[0036] The present invention also provides host cells comprising at least one expression construct as provided herein. In some embodiments, the host cell is a yeast or filamentous fungal cell. In some embodiments, the host cell is *Myceliophthora thermophila*. In some embodiments, the expression construct is integrated into the genome of said host cell.

[0037] The present invention also provides methods for producing a protein in a host cell, comprising culturing the host cell comprising at least one expression construct, under conditions such that the protein is produced by the host cell. In some embodiments, the methods further comprise the step of isolating the protein produced by the host cell.

BRIEF DESCRIPTION OF THE FIGURES

[0038] Figure 1 shows the sequence of the C1 *bgl1* gene (SEQ ID NO:41) and β-glucosidase protein (SEQ ID NO:42) from copending application No. 601/247379 (filed September 30, 2009) which is incorporated herein by reference for all purposes.

[0039] Figure 2 is a bar graph showing the effect of different promoters on the level of expression of a C1 cellulase in a C1 strain (UV18#100.f [Δ]pyr5 [Δ]alpl).

DESCRIPTION OF THE INVENTION

[0040] The present invention provides promoters derived from a filamentous fungus. These promoters have application in the fields of molecular biology, microbiology, fungal genetics and production of biofuels and other products.

[0041] Promoters from the fungal strain C1 have been identified and characterized, and can be used for the expression of heterologous genes and recombinant protein production in host cells and particularly in fungal host cells. DNA constructs, vectors, cells and methods for protein production are also provided herein.

Definitions and Methods

[0042] Unless defined otherwise, technical and scientific terms used herein generally have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. 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 preferred methods and materials are described. 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.

[0043] Unless indicated otherwise, the techniques and procedures described or referred to herein are generally performed according to conventional methods well known in the art. Texts disclosing general methods and techniques in the field of recombinant genetics are widely available and known to those in the art. DNA sequences can be obtained by cloning, or by chemical synthesis.

[0044] 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.

[0045] 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).

[0046] 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.

[0047] 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.

[0048] As used herein, the term "promoter" refers to a polynucleotide sequence, particularly a DNA sequence, that initiates and facilitates the transcription of a target gene sequence in the presence of RNA polymerase and transcription regulators. In some embodiments, promoters include DNA sequence elements that ensure proper binding and activation of RNA polymerase, influence where transcription will start, affect the level of transcription and, in the case of inducible promoters, regulate transcription in response to environmental conditions. Promoters are located 5' to the transcribed gene and, as used herein, include the sequence 5' from the translation start codon (*i.e.*, in some embodiments, including the 5' untranslated region of the mRNA, typically comprising 100-200 bp). Most often, the core promoter sequences lie within 1-2 kb of the translation start site, more often within 1 kbp and often within 750 bp, 500 bp or 200 bp of the translation start site. By convention, the promoter sequence is usually provided as the sequence on the coding strand of the gene it controls.

[0049] As used herein, the term "inducible promoter" refers to a promoter that initiates transcription only when the host cell comprising the inducible promoter is exposed to particular environmental factors (*e.g.*, temperature or light responsive promoters), chemical factors (*e.g.*, promoters induced by small molecules, such as IPTG or tetracycline), metabolic factors (*e.g.*, promoters induced or repressed by glucose or metabolites; promoters active during exponential growth phase), physical factors and the like.

[0050] As used herein, the term "constitutive promoter" refers to a promoter that drives transcription at about the same level under a variety of environmental or growth conditions. In some embodiments, the term "constitutive promoter" refers to a promoter that is glucose-independent (*i.e.*, is not induced by or repressed by glucose levels). Glucose independent regulation can be determined as

described in the Examples below. In additional embodiments, the term "constitutive promoter" refers to a promoter that is not growth dependent (*i.e.*, drives transcription during both exponential and non-exponential growth phases).

[0051] As used herein, the term "promoter activity" refers to the level of expression or activity of the gene and/or polypeptide operably linked to the promoter of interest. Any suitable method for determining/measuring promoter activity finds use in the present invention. For example, promoter activity can be measured by estimating the levels of expression of transcript, production of protein, or protein activity by one of ordinary skill in the art by well known methods, including but not limited to quantitative real-time PCR (qRT-PCR), Northern blot hybridization, SDS-PAGE analysis, and/or enzyme activity assays.

[0052] As used herein, the term "C1C promoter" refers to any of the promoters encompassed by the invention, including but not limited to promoters comprising nucleic acid sequences provided in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:33, SEQ ID NO:34, and/or SEQ ID NO:35, and subsequences (e.g., SEQ ID NOS:4-8, 12-16, 20-24, 28-32, and 36-40) and variants thereof. The term "C1Ca promoter" refers to the various promoters comprising a nucleic acid sequence of SEQ ID NOS:1, 2 and/or 3, and subsequences (*e.g.*, SEQ ID NOS:4-8) and variants thereof. The term "C1Cb promoter" refers to the various promoters comprising a nucleic acid sequence of SEQ ID NOS:9,10, and/or 11, and subsequences (*e.g.*, SEQ ID NOS:12-16) and variants thereof. The term "C1Cc promoter" refers to the various promoters comprising a nucleic acid sequence of SEQ ID NOS:17, 18 and/or 19, and subsequences (*e.g.*, SEQ ID NOS:20-24) and variants thereof. The term "C1Cd promoter" refers to the various promoters comprising a nucleic acid sequence of SEQ ID NOS:25, 26, and/or 27, and subsequences (*e.g.*, SEQ ID NOS:28-32) and variants thereof. The term "C1Ce promoter" refers to the various promoters comprising a nucleic acid sequence of SEQ ID NOS:33, 34, and/or 35, and subsequences (*e.g.*, SEQ ID NOS:36-40) and variants thereof.

[0053] As used herein, the term "variant" used in reference to a promoter means a C1C promoter that comprises one or more modifications such as substitutions, additions or deletions of one or more nucleotides relative to a wild-type sequence or another "starting" or "parent sequence." In some embodiments, the term refers to a C1Ca promoter with one or more modifications such as substitutions, additions or deletions of one or more nucleotides relative to a wild-type sequence. In other embodiments this refers to C1Cb, C1Cc, C1Cd and/or C1Ce promoters with one or more modifications relative to the wild-type sequence. Such variants retain the ability to drive expression of a protein-encoding polynucleotide to which the promoter is operably linked.

[0054] As used herein, the terms "reference promoter" and “reference sequence” when used in the context of promoters, refer to a promoter to which a variant promoter of the present invention is compared in order to determine the presence of an improved property in the variant promoter being evaluated (*e.g.*, expression). In some embodiments, a reference promoter is a wild-type promoter (*e.g.*, SEQ ID NOS:1, 9, 17, 25, and/or 33). In some embodiments, a reference promoter is another variant promoter (*e.g.*, a variant of a wild-type promoter, of the present invention).

[0055] As used herein, the term "reference enzyme" refers to an enzyme to which a variant enzyme of the present invention is compared in order to determine the presence of an improved property in the variant enzyme being evaluated, including but not limited to improved thermoactivity, improved thermostability, or improved stability. In some embodiments, a reference enzyme is a wild-type enzyme, while in some other embodiments, a reference enzyme is another variant enzyme.

[0056] As used herein, the term “wild-type promoter sequence” refers a promoter sequence that is found in nature (*e.g.*, any one of SEQ ID NOS:1, 9, 17, 25, and/or 33), as well as functional fragments of such promoter sequences.

[0057] As used herein, the terms “modifications” and “mutations” when used in the context of substitutions, deletions, insertions and the like with respect to polynucleotides and polypeptides are used interchangeably herein and refer to changes that are introduced by genetic manipulation to create variants from a wild-type sequence.

[0058] As used herein, the term “functional fragment” refers to a promoter that contains a subsequence, usually of at least about 25, about 50, about 75, about 100, about 150, about 200, about 250, about 300, or about 350, or more, contiguous nucleotides relative to a reference sequence such as one of SEQ ID NOS:1-40 that has promoter activity to drive expression of a polynucleotide encoding a protein to which the promoter is operably linked. Functional fragments typically comprise at least about 20%, 25%, 30%, 35%, 40%, 45%, 50%, about 60%, about 70%, about 75%, about 80%, about 85%, about 90%, about 95% or more, of the promoter activity relative to the 1.5 kb promoter sequence of SEQ ID NO:1, SEQ ID NO:9, SEQ ID NO:17, SEQ ID NO:25, and/or SEQ ID NO:33.

[0059] In some embodiments, a variant or functional fragment of a wild-type promoter set forth in any of SEQ ID NOS:1-40, is evaluated for promoter activity in *M. thermophila* (*e.g.*, in a suitable medium comprising complex sources of nitrogen, salts, and carbon), and assessing the level of at least one protein or RNA transcript that is produced from an expression construct comprising the variant promoter operably linked to a polynucleotide sequence encoding at least one protein. In some embodiments, a variant or functional fragment of a promoter is considered to have promoter activity if the promoter is able to produce at least about 25%, about 50%, about 60%, about 70%, about 80%, about

90%, or about 95% or greater, of the protein or RNA produced using a promoter having the sequence of SEQ ID NO:1, 9, 17, 25, and/or 33, when operably linked to a polynucleotide encoding a protein, as compared to a wild-type promoter under the same expression and testing conditions.

[0060] 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. A recombinant cell contains a recombinant polynucleotide or polypeptide. Thus, the term “recombinant” 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.

[0061] As used herein, the term “recombinant nucleic acid”, or equivalently, “recombinant polynucleotide”, is one that is inserted into a heterologous location such that it is not associated with nucleotide sequences that normally flank the nucleic acid as it is found in nature (*e.g.*, a nucleic acid inserted into a vector). Likewise, a nucleic acid sequence that does not appear in nature, for example a variant of a naturally occurring gene, is recombinant. Thus, in the context of the present invention, an example of a recombinant nucleic acid includes a protein-encoding DNA sequence that is operably linked to a promoter of the present invention.

[0062] As used herein, the terms “reporter,” “reporter protein,” and “reporter sequence” refer to any polypeptide gene expression product that is encoded by a heterologous gene operably linked to a promoter set forth herein.

[0063] As used herein, a "signal sequence" is a DNA sequence that encodes a signal peptide. A signal peptide directs the polypeptide with which is associated through a secretory pathway of a cell in which it is synthesized. In some embodiments, the signal peptide is removed during transit through the secretory pathway.

[0064] As used herein, the term “vector,” refers to a recombinant nucleic acid designed to carry a coding sequence of interest to be introduced into a host cell. In the present invention, vectors comprise at least one promoter sequence and at least one heterologous polynucleotide encoding at least one protein of interest to be expressed by the host cell. This term encompasses many different types of vectors, such as cloning vectors, expression vectors, shuttle vectors, plasmids, phage or virus particles, and the like. Vectors include PCR-based, as well as plasmid vectors. Vectors include an origin of replication and

usually include a multicloning site and a selectable marker. In some embodiments, expression vectors also include, in addition to a coding sequence of interest, elements that direct the transcription and translation of the coding sequence (*e.g.*, promoters, enhancers, and termination/polyadenylation sequences). In some embodiments, vectors comprising a promoter of the present invention are used as integration vectors, such that the promoter is integrated into the host cell genome.

[0065] As used herein, the term “expression cassette” refers to a nucleic acid molecule containing a protein coding sequence and nucleic acid elements that permit transcription of the sequence in a host cell (*e.g.*, promoter and termination/polyadenylation sequences). In some embodiments, expression cassettes are components of expression vectors.

[0066] As used herein, the term “expression construct” refers to a polynucleotide comprising a promoter sequence operably linked to a protein encoding sequence. Expression cassettes and expression vectors are examples of “expression constructs.” The term also encompasses PCR constructions for targeting DNA to direct integration into the host cell genome at a desired site.

[0067] As used herein, the term “expression” of a gene means transcription of the gene. In some embodiments, the term is used in reference to the production of a polypeptide encoded in the gene sequence.

[0068] 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.

[0069] As used herein, the term “increased expression” means at least about a 1.2-fold increase in the level of transcript for a given gene in the modified organism is observed as compared to the level of transcript for the same gene in the parental organism, when grown under identical or nearly identical conditions of medium composition, temperature, pH, cell density and age of culture. For example, the transcript level of a given gene in the modified organism can be increased by at least about 1.2-, about 1.5-, about 2.0-, about 2.5-, about 3.0-, about 3.5, about 4.0-, about 4.5, about 5.0-, about 5.5-, or 10-fold, or more, relative to the transcript level of the same gene in the parental organism when grown or cultured under essentially the same culture conditions. The modulation of expression of genes also can be measured by one of ordinary skill in the art through analysis of selected mRNA or transcript levels by well-known means, for example, quantitative real-time PCR (qRT-PCR), Northern blot hybridization, or global gene expression profiling using cDNA or oligo array hybridization.

[0070] As used herein, the terms “nucleic acid,” “nucleotide,” and “polynucleotide,” refer to deoxyribonucleotides or ribonucleotides and polymers thereof in either single-stranded or double-stranded

form. Except were specified or otherwise clear from context, reference to a nucleic acid sequence (*e.g.,* any of SEQ ID NOS:1-40) encompasses a double stranded molecule.

[0071] 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).

[0072] As used herein, the terms "identical" and "percent identity," in the context of describing two or more polynucleotide sequences, refer to two or more sequences that are the same or have a specified percentage of nucleotides that are the same when compared and aligned for maximum correspondence over a comparison window or designated region, as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection.

[0073] For sequence comparison, typically one sequence acts as a "reference sequence," to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be 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. For sequence comparison of a wild-type promoter sequence (*e.g.*, SEQ ID NOS:1, 9, 17, 25, and/or 33), with its variants, the BLAST and BLAST 2.0 algorithms and the default parameters discussed below may be used.

[0074] A "comparison window" as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 20 to 500, usually about

50 to about 300, also about 50 to 250, and also about 100 to about 200 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. Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted using any suitable method known in the art. Indeed, various methods are well-known to those of skill in the art, including manual alignment and visual inspection, as well as other methods, including computerized methods (*See e.g.*, Smith and Waterman, Adv. Appl. Math., 2:482 [1981]; Needleman and Wunsch, J. Mol. Biol., 48:443 [1970]; Pearson and Lipman, Proc. Nat'l. Acad. Sci. USA 85:2444 [1988], and Wisconsin Genetics Software Package, Genetics Computer Group [*e.g.*, GAP, BESTFIT, FASTA, and TFASTA]).

[0075] Examples of algorithms that are suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms (*See e.g.*, Altschul *et al.*, J. Mol. Biol., 215: 403-410 [1990]; and Altschul *et al.*, Nucleic Acids Res. 25: 3389-3402 [1977]). Software for performing BLAST analyses is publicly available at the National Center for Biotechnology Information website. The BLAST algorithm also performs a statistical analysis of the similarity between two sequences. One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001.

[0076] As used herein, the term "heterologous" refers to a nucleic acid or polypeptide that is cloned and or expressed in a context different from how it is present in nature. The term "heterologous," when used to describe a promoter and an operably linked coding sequence, means that the promoter and the coding sequence are not associated with each other in nature. In some embodiments, a promoter and a heterologous coding sequence are from two different organisms. In some alternative embodiments, a promoter and a heterologous coding sequence are from the same organism, provided the particular promoter does not direct the transcription of the coding sequence in the wild-type organism.

[0077] When two elements (*e.g.*, a promoter and a coding sequence), are said to be "operably linked," it is meant that the juxtaposition of the two elements allows them to be in a functionally active relationship. In other words, a promoter is "operably linked" to a coding sequence when the promoter controls the transcription of the coding sequence.

[0078] As used herein, the term "gene" refers to a segment of DNA that is transcribed. In some embodiments, it includes regions preceding and following the protein coding region (5' and 3'

untranslated sequence), as well as intervening sequences (introns) between individual coding segments (exons).

[0079] The terms “polypeptide,” “peptide,” and “protein” are used interchangeably herein to refer to a polymer of amino acid residues.

[0080] As used herein, the term “isolated” refers to a compound, protein, cell, nucleic acid sequence and/or an amino acid sequence that is removed from at least one component with which it is naturally associated.

[0081] 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," or "BG"). 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.

[0082] 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. The term “EG2" is used in reference 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, he EG2 may be functionally linked to a carbohydrate binding module (CBM) with a high affinity for crystalline cellulase, such as a Family 1 cellulose binding domain.

[0083] As used herein, the terms "cellobiohydrolase" and "CBH" refers to a category of cellulases (EC 3.2.1.91) that hydrolyze glycosidic bonds in cellulose. In some embodiments, the cellobiohydrolase is a "type 2 cellobiohydrolase," which is a cellobiohydrolase belonging to the glycoside hydrolase family 6 (GH6) family of cellulases and which is also commonly called "the Cel6 family." Cellobiohydrolases of the GH6 family are described, for example, in the Carbohydrate Active Enzymes (CAZY) database.

[0084] As used herein, the term "β-glucosidase," "cellobiase," or "BGL" refers to a category of cellulases (EC 3.2.1.21) that catalyze the hydrolysis of cellobiose to glucose.

[0085] As used herein, the term "glycoside hydrolase 61" or "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]).

[0086] A "hemicellulase" as used herein, refers to a polypeptide that can catalyze hydrolysis of hemicellulose into small polysaccharides such as oligosaccharides, or monomeric saccharides. Hemicellulloses include xylan, glucuonoxylan, arabinoxylan, glucomannan and xyloglucan. Hemicellulases include, for example, the following: endoxylanases, b-xylosidases, a-L-arabinofuranosidases, a-D-glucuronidases, feruloyl esterases, coumarolyl esterases, a-galactosidases, b-galactosidases, b-mannanases, and b-mannosidases. In some embodiments, the present invention provides enzyme mixtures that comprise EG1b and one or more hemicellulases.

[0087] 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, cysteine proteases including pepsin, papain and serine proteases including chymotrypsins, carboxypeptidases and metalloendopeptidases.

[0088] 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.

[0089] 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 mutagensis). 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.).

[0090] An effective way to generate a large collection of functional variants is to use a random mutation strategy. There are numerous texts available known to those skilled in the art that describe techniques employing chemical mutagenesis, cassette mutagenesis, degenerate oligonucleotides, mutually priming oligonucleotides, linker-scanning mutagenesis, alanine-scanning mutagenesis, error-prone PCR, etc. Indeed, mutagenesis may be performed in accordance with any of the techniques known in the art, including random and site-specific mutagenesis. Directed evolution can be performed with any of the techniques known in the art to screen for production of variants including shuffling. 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).

[0091] 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 know 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.,* such as 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 EG1b 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, or ribozyme technology finds use in inhibiting gene expression.

[0092] As used herein, the term “recombinant host cell” refers to a cell into which at least one heterologous polynucleotide, gene and promoter set forth herein (*e.g.*, an expression vector) has been introduced. The term is also used in reference to cells having at least one heterologous gene or polynucleotide integrated into their genomes.

[0093] 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.

[0094] 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.

[0095] As used herein, the term "C1" refers to a *Chrysosporium lucknowense* fungal strain described by Garg (*See*, Garg, Mycopathol., 30: 3-4 [1966]). "*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, 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 thermophilia*. Other C1 strains include organisms deposited under accession numbers ATCC 44006, CBS (Centraalbureau voor Schimmelcultures) 122188, CBS 251.72, CBS 143.77, CBS 272.77, 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 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, incorporated herein by reference.

[0096] Methods for recombinant expression of proteins in fungi and other organisms are well known in the art, and a number of suitable expression vectors are available or can be constructed using routine methods. Protocols for cloning and expression in fungal hosts and other organisms are well known in the art (*See e.g.*, Zhu *et al.*, Plasmid 6:128-33 [2009]). Standard references for techniques and protocols are widely available and known to those in the art (*See e.g.*, U.S. Pat. Nos. 6,015,707, 5,811,381 and 6,573,086; US Pat. Pub. Nos. US 2003/0187243, US 2007/0238155, US 2008/0194005, US 2009/0099079; WO 2008/073914 and WO 98/15633, each of which is incorporated by reference herein for all purposes).

[0097] As used herein, the terms "culturing" and "cultivating" refer to growing a population of microbial cells (*e.g.*, host cells and/or recombinant host cells) under suitable conditions in a liquid or solid medium. In some embodiments, culturing refers to fermentative bioconversion of a carbon substrate to an end-product. Suitable conditions for culturing and producing cells are well-known in the art. Any suitable methods and compositions for culturing find use in the present invention, including defined and undefined media, as well as rich and minimal media, as known in the art and are desired for the particular use of the present invention.

[0098] In some embodiments, the recombinant microorganisms (*i.e.*, recombinant host cells) comprising a promoter of the present 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. In some embodiments, fermentations are carried out a temperature of about 10°C to about 60°C, about 15°C to about 50°C, about 20°C to about 45°C, about 20°C to about 40°C, about 20°C to about 35°C, or about 25°C to about 45°C. In some embodiments, the fermentation is carried out at a temperature of about 28°C and/or about 30°C. It will be understood that, in certain embodiments where thermostable host cells are used, fermentations are often carried out at higher temperatures. In some embodiments, the fermentation is carried out for a time period of about 8 hours to about 240 hours, about 8 hours to about 165 hours, about 8 hours to 145 hours, about 16 hours to about 120 hours, or about 24 hours to about 72 hours. In some embodiments, the fermentation will be carried out at a pH of about 3 to about 8, about 4.5 to about 7.5, about 5 to about 7, or about 5.5 to about 6.5. The temperature, time period, pH and other factors will vary depending upon the desired protein production system, and are well-known to those of skill in the art.

[0099] As used herein, the terms "biomass," "biomass substrate," "cellulosic biomass," "cellulosic feedstock," and "cellulosic substrate" refer to materials that contain cellulose. Biomass can be derived from any suitable material, including but not limited to plants, animals, or microorganisms (*e.g.,* agricultural, industrial, and forestry residues, industrial and municipal wastes, and terrestrial and aquatic crops grown for energy purposes). Particular examples of cellulosic 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 straw, switchgrass, waste paper, paper and pulp processing waste, woody or herbaceous plants, fruit or vegetable pulp, distillers grain, rice hulls, cotton, hemp, flax, sisal, sugar cane bagasse, sorghum, soy, components obtained from milling of grains, trees, branches, roots, leaves, wood chips, sawdust, shrubs and bushes, vegetables, fruits, and flowers and mixtures thereof. 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.

[00100] 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).

[00101] 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.

[00102] 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.

[00103] As used herein, the term “lignocellulose biomass” refers to any plant biomass comprising cellulose and hemicellulose, bound to lignin.

[00104] As used herein, the term “lignocellulosic feedstock” refers to any type of plant biomass such as, but 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, baggase (*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, 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 lignocellulosic feedstock comprises cellulosic waste material and/or forestry waste materials, including but not limited to, paper and pulp processing waste, newsprint, cardboard and the like. The biomass may also comprise transgenic plants that express ligninase and/or cellulase enzymes (US 2008/0104724 A1). In some embodiments, the lignocellulosic feedstock comprises one species of fiber, while in some alternative embodiments, the lignocellulosic feedstock comprises a mixture of fibers that originate from different lignocellulosic feedstocks. In some other embodiments, the lignocellulosic feedstock comprises 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 materal 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. 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).

C1C Promoters of the Invention

[00105] Promoter regions from the filamentous fungus designated C1 were identified and are set forth in SEQ ID NOS: 1, 9, 17, 25 and 33. These promoter sequences, designated C1Ca (SEQ ID NO:1), C1Cb (SEQ ID NO:9), C1Cc (SEQ ID NO:17), C1Cd (SEQ ID NO:25) and C1Ce (SEQ ID NO:33) are strong drivers of expression in C1, and function in a glucose-independent fashion. In some embodiments, the C1Ca, C1Cb, C1Cc, C1Cd and/or C1Ce promoter sequences are operably linked to at least one sequence encoding a heterologous protein, such that the heterologous protein is expressed by a host cell.

[00106] In some embodiments, the C1Ca promoter of the invention comprises the 1.5 kb sequence set forth in SEQ ID NO:1. In some embodiments, the C1Ca promoter comprises a subsequence of SEQ ID NO:1, or a variant thereof, as discussed below. In some embodiments, the C1Ca promoter of the invention comprises SEQ ID NO:2, which is the 3' (3-prime) 1 kb of SEQ ID NO:1 (*i.e.*, comprising bases 501-1500 of SEQ ID NO:1). In some embodiments the C1Ca promoter of the invention comprises SEQ ID NO:3, which is the 3' (3-prime) 0.7 kb of SEQ ID NO:1 (*i.e.*, comprising bases 801-1500 of SEQ ID NO:1). The sequences of SEQ ID NOS:1-3 are provided below.

CTTGGGTAATGTAATCCGTCCAGCACATGTCTTGACATGCTTGATGTATGTACTCTATATATT
GAGTACAGTACCTGTCTTGAAGAGCAGTACAGTATACTCTCTGGATGGACTTCGGTGCTTGG
ACGGGAATAATAACTTAGGGTTTGCGCTAAGGTAGACTAGCAGCCAATGAAATCGGCTCAG
TGGCACGAGATACTTTAAAGTGTACGGAGTACTCCGTACGGAACACAGACTCAACAAATTG
CCATGACCATGCAGCTCTCTCTCGCAGCGAGATGAGAGACCCACATAGCATTCCAACCATGT
GAGGAAGTCTATTAGTACGGCTTATGAGTGTACTCCGGACCACTCTAGTTAGTTGTATGCAC
GATCTGTCGAAGAGATCGGTGGTAACTCGCCAGCATGGGCATACCATTTCAAACCAAGATTT
TTATGCGAGCATCAAACTCCACAACTCAGCGTCCGCAGCAGGCATCGGAGAGTTGGCACCCT
CAGCCCTTTTACACTAGAACTAGATTTCTGCAAAAGCTCTTGTGCAAGGCTGCATACAGATA
CACTAGAGTAAGGTGCTTAGCTAATATACTAAAGCGTATTAGTCCCAATCCCATGCAAGCTG
CGAGAGGTCGCAATTTCCCGACTCGGCGCCGCATGCGCTAGTGTGGACTGGGAATCGCGGCC
GAGCGCCTGCCGTAGCTCTTCGCCGCTTGAGAAGGAGACGTGAAGCAGCTGGAACAAACCC
TTTGGAACACCGGGCAGCTTAAAGTTGGCCAAGTAGGAGTTGAGCAACGACGAGGTTCCCG
TGTAGATTACATTGTTAATGTTGGTACTATTGTGCCACAAATAGGCCGCCGTTCCGAGCGAC
AAGAGAGCTATCGAGACCCATCGAACCTCAAATTCCAAGACCGGGACCCCGGCCAGCGTTC
CTTGTCTAGCCCTGGTCGCCCAGAGCTTCCGCCACGCCGCTTCCCCATTGCAACACTGATCTG
GGGTGCGCTAACAGGCGATAATTCTTTCTAGGCGTCCTATTGGCGTGACCCGGGAGAGAAGT
CACTCTCCCATGCCCTCATCGCGTTCCTACTACGGAGTACTCCGTAGTAGCTGCCCATTCCAG
GTCGGCTATTTTGGGTCCAGACGTGCCGCTGCCTTCTTCTTCCGCTCTTCCTTCTTCCTTCCCC
CCTTCCCCAGCCGTAGCTCACACCACACATCCGGCCTGACTGGCATTTCTCTGCCATACTAAT
TAACACTATCCCAACTTCTCCACGGTCGTTCATCTTTGGTTTTTCATTCTCAACTAGACTAATT
AATTACCGCATCGGCGCAGCCAATTCACTTAGGCAAGCTTTGGTAAGTCGCCCCTGTCTGCT
TTGCAGAGTTTCCCCGGCCTTTTCTCCTTGTGACAATCCGGTCCCGAATGGGTTTGTTCTCAA
GTGCATTGACTCACCGCCGAAAGTCGTTCATAGCACTCGCTTTAGATATCAAGCATAAATCC
GGTCCACA (SEQ ID NO:1)

CTTTTACACTAGAACTAGATTTCTGCAAAAGCTCTTGTGCAAGGCTGCATACAGATACACTA
GAGTAAGGTGCTTAGCTAATATACTAAAGCGTATTAGTCCCAATCCCATGCAAGCTGCGAGA
GGTCGCAATTTCCCGACTCGGCGCCGCATGCGCTAGTGTGGACTGGGAATCGCGGCCGAGCG
CCTGCCGTAGCTCTTCGCCGCTTGAGAAGGAGACGTGAAGCAGCTGGAACAAACCCTTTGGA
ACACCGGGCAGCTTAAAGTTGGCCAAGTAGGAGTTGAGCAACGACGAGGTTCCCGTGTAGA
TTACATTGTTAATGTTGGTACTATTGTGCCACAAATAGGCCGCCGTTCCGAGCGACAAGAGA
GCTATCGAGACCCATCGAACCTCAAATTCCAAGACCGGGACCCCGGCCAGCGTTCCTTGTCT
AGCCCTGGTCGCCCAGAGCTTCCGCCACGCCGCTTCCCCATTGCAACACTGATCTGGGGTGC
GCTAACAGGCGATAATTCTTTCTAGGCGTCCTATTGGCGTGACCCGGGAGAGAAGTCACTCT
CCCATGCCCTCATCGCGTTCCTACTACGGAGTACTCCGTAGTAGCTGCCCATTCCAGGTCGGC
TATTTTGGGTCCAGACGTGCCGCTGCCTTCTTCTTCCGCTCTTCCTTCTTCCTTCCCCCCTTCC

CCAGCCGTAGCTCACACCACACATCCGGCCTGACTGGCATTTCTCTGCCATACTAATTAACA
CTATCCCAACTTCTCCACGGTCGTTCATCTTTGGTTTTTCATTCTCAACTAGACTAATTAATTA
CCGCATCGGCGCAGCCAATTCACTTAGGCAAGCTTTGGTAAGTCGCCCCTGTCTGCTTTGCA
GAGTTTCCCCGGCCTTTTCTCCTTGTGACAATCCGGTCCCGAATGGGTTTGTTCTCAAGTGCA
TTGACTCACCGCCGAAAGTCGTTCATAGCACTCGCTTTAGATATCAAGCATAAATCCGGTCC
ACA (SEQ ID NO:2)

CCGTGTAGATTACATTGTTAATGTTGGTACTATTGTGCCACAAATAGGCCGCCGTTCCGAGC
GACAAGAGAGCTATCGAGACCCATCGAACCTCAAATTCCAAGACCGGGACCCCGGCCAGCG
TTCCTTGTCTAGCCCTGGTCGCCCAGAGCTTCCGCCACGCCGCTTCCCCATTGCAACACTGAT
CTGGGGTGCGCTAACAGGCGATAATTCTTTCTAGGCGTCCTATTGGCGTGACCCGGGAGAGA
AGTCACTCTCCCATGCCCTCATCGCGTTCCTACTACGGAGTACTCCGTAGTAGCTGCCCATTC
CAGGTCGGCTATTTTGGGTCCAGACGTGCCGCTGCCTTCTTCTTCCGCTCTTCCTTCTTCCTTC
CCCCCTTCCCCAGCCGTAGCTCACACCACACATCCGGCCTGACTGGCATTTCTCTGCCATACT
AATTAACACTATCCCAACTTCTCCACGGTCGTTCATCTTTGGTTTTTCATTCTCAACTAGACT
AATTAATTACCGCATCGGCGCAGCCAATTCACTTAGGCAAGCTTTGGTAAGTCGCCCCTGTC
TGCTTTGCAGAGTTTCCCCGGCCTTTTCTCCTTGTGACAATCCGGTCCCGAATGGGTTTGTTC
TCAAGTGCATTGACTCACCGCCGAAAGTCGTTCATAGCACTCGCTTTAGATATCAAGCATAA
ATCCGGTCCACA (SEQ ID NO:3)

[00107] In some embodiments, the C1Ca promoter of the invention comprises a subsequence of SEQ ID NO:1, SEQ ID NO:2, and/or SEQ ID NO:3 that retains promoter activity, referred to herein as a "short" promoter sequence. Subsequences that retain promoter activity are identified using any suitable routine methods. Provided with SEQ ID NO:1, 2, and/or 3, any of a number of different functional deletion mutants of the starting sequence can be readily prepared. In some embodiments, the subsequences retain constitutive promoter activity (*e.g.*, glucose-independent promoter activity).

[00108] In some embodiments, constructs with short promoter sequences are prepared using any of a variety of routine molecular biological techniques. For illustration and not limitation, SEQ ID NO:1 may be cloned into an expression vector so that it is 5' to and operably linked to a sequence encoding a reporter protein. One or a series of deletion constructs may be made to produce one or a library of expression vectors with subsequences of SEQ ID NO:1 operably linked to the sequence encoding the reporter protein. Deletions may be made from the 5' end, the 3' end or internally. Methods for making deletions include, for illustration and not limitation, using restriction and ligation to remove a portion of SEQ ID NO:1 from the vector, using exonucleases to trim the end(s) of the parent sequence, randomly fragmenting the parent sequence and preparing a library of clones containing fragments, etc. In some embodiments, PCR techniques find use. The expression vector(s) is then introduced into a host cell and the cell is cultured under conditions in which the protein is produced, with the presence and level of production being indicative of promoter activity. The reporter protein may be one frequently used to assess promoter strength and properties, such as luciferase. An advantage of using "standard" reporters is

the ready availability of commercial assays and materials for assessing transcription. Alternatively the reporter may be another protein, such as a fungal protein and/or a protein product for which high expression is desired.

[00109] In some embodiments, the C1Ca promoter sequence comprises at least about 1000 contiguous nucleotides of SEQ ID NO:1 or 2; at least about 900 contiguous nucleotides of SEQ ID NO:1 or 2; at least about 800 contiguous nucleotides of SEQ ID NO:1 or 2; at least about 700 contiguous nucleotides of SEQ ID NO:1, 2, or 3; at least about 600 contiguous nucleotides of SEQ ID NO:1, 2, or 3; at least about 500 contiguous nucleotides of SEQ ID NO:1, 2, or 3; at least about 450 contiguous nucleotides of SEQ ID NO:1, 2, or 3; at least about 400 contiguous nucleotides of SEQ ID NO:1, 2, or 3; at least about 350 contiguous nucleotides of SEQ ID NO:1, 2, or 3; at least about 300 contiguous nucleotides of SEQ ID NO:1, 2, or 3; at least about 250 contiguous nucleotides of SEQ ID NO:1, 2, or 3; at least about 200 contiguous nucleotides of SEQ ID NO: 1, 2, or 3; at least about 150 contiguous nucleotides of SEQ ID NO: 1, 2, or 3; at least about 100 contiguous nucleotides of SEQ ID NO: 1, 2, or 3; or at least about 75 contiguous nucleotides of SEQ ID NO: 1, 2, or 3.

[00110] In some embodiments, the C1Ca promoter sequence comprises a subsequence of SEQ ID NO:1 comprising about 75 to about 1000 contiguous nucleotides of SEQ ID NO:1 or 2. In some other embodiments, the C1Ca promoter sequence comprises a subsequence of SEQ ID NO:1, 2, or 3 comprising about 50 to about 700, about 50 to about 600, about 50 to about 500, about 50 to about 400, about 50 to about 300, about 50 to about 200, about 75 to about 700, about 75 to about 600, about 75 to about 500, about 75 to about 400, about 75 to about 300, about 75 to about 200, about 100 to about 700, about 100 to about 600, about 100 to about 500, about 100 to about 400, about 100 to about 300, or about 100 to about 200 contiguous nucleotides.

[00111] In some embodiments, the subsequence comprises at least about 25, at least about 50, at least about 100, at least about 150, or at least about 200 contiguous nucleotides of SEQ ID NO:3. In some further embodiments, the C1Ca promoter sequence comprises at least one of SEQ ID NOS:4-8, as provided herein. SEQ ID NO:4 comprises bases 871 - 1430 of SEQ ID NO:1 (0.56 kb), while SEQ ID NO:5 comprises bases 941 - 1430 of SEQ ID NO:1 (0.49 kb), SEQ ID NO:6 comprises bases 1001 - 1450 of SEQ ID NO:1 (0.45 kb), SEQ ID NO:7 comprises bases 1001 - 1400 of SEQ ID NO:1 (0.40 kb), and SEQ ID NO:8 comprises bases 1201 – 1450 of SEQ ID NO:1 (0.25 kb).

AGCTATCGAGACCCATCGAACCTCAAATTCCAAGACCGGGACCCCGGCCAGCGTTCCTTGTC
TAGCCCTGGTCGCCCAGAGCTTCCGCCACGCCGCTTCCCCATTGCAACACTGATCTGGGGTG
CGCTAACAGGCGATAATTCTTTCTAGGCGTCCTATTGGCGTGACCCGGGAGAGAAGTCACTC
TCCCATGCCCTCATCGCGTTCCTACTACGGAGTACTCCGTAGTAGCTGCCCATTCCAGGTCGG

CTATTTTGGGTCCAGACGTGCCGCTGCCTTCTTCTTCCGCTCTTCCTTCTTCCTTCCCCCCTTC
CCCAGCCGTAGCTCACACCACACATCCGGCCTGACTGGCATTTCTCTGCCATACTAATTAAC
ACTATCCCAACTTCTCCACGGTCGTTCATCTTTGGTTTTTCATTCTCAACTAGACTAATTAATT
ACCGCATCGGCGCAGCCAATTCACTTAGGCAAGCTTTGGTAAGTCGCCCCTGTCTGCTTTGC
AGAGTTTCCCCGGCCTTTTCTCCTTGTGACAATCCGGTCCCGAATGGGTTTGTTCTCAA (SEQ
ID NO:4)

GTCGCCCAGAGCTTCCGCCACGCCGCTTCCCCATTGCAACACTGATCTGGGGTGCGCTAACA
GGCGATAATTCTTTCTAGGCGTCCTATTGGCGTGACCCGGGAGAGAAGTCACTCTCCCATGC
CCTCATCGCGTTCCTACTACGGAGTACTCCGTAGTAGCTGCCCATTCCAGGTCGGCTATTTTG
GGTCCAGACGTGCCGCTGCCTTCTTCTTCCGCTCTTCCTTCTTCCTTCCCCCCTTCCCCAGCCG
TAGCTCACACCACACATCCGGCCTGACTGGCATTTCTCTGCCATACTAATTAACACTATCCCA
ACTTCTCCACGGTCGTTCATCTTTGGTTTTTCATTCTCAACTAGACTAATTAATTACCGCATCG
GCGCAGCCAATTCACTTAGGCAAGCTTTGGTAAGTCGCCCCTGTCTGCTTTGCAGAGTTTCCC
CGGCCTTTTCTCCTTGTGACAATCCGGTCCCGAATGGGTTTGTTCTCAA (SEQ ID NO:5)

CGCCGCTTCCCCATTGCAACACTGATCTGGGGTGCGCTAACAGGCGATAATTCTTTCTAGGC
GTCCTATTGGCGTGACCCGGGAGAGAAGTCACTCTCCCATGCCCTCATCGCGTTCCTACTAC
GGAGTACTCCGTAGTAGCTGCCCATTCCAGGTCGGCTATTTTGGGTCCAGACGTGCCGCTGC
CTTCTTCTTCCGCTCTTCCTTCTTCCTTCCCCCCTTCCCCAGCCGTAGCTCACACCACACATCC
GGCCTGACTGGCATTTCTCTGCCATACTAATTAACACTATCCCAACTTCTCCACGGTCGTTCA
TCTTTGGTTTTTCATTCTCAACTAGACTAATTAATTACCGCATCGGCGCAGCCAATTCACTTA
GGCAAGCTTTGGTAAGTCGCCCCTGTCTGCTTTGCAGAGTTTCCCCGGCCTTTTCTCCTTGTG
ACAATCCGGTC (SEQ ID NO:6)

CAGGCGATAATTCTTTCTAGGCGTCCTATTGGCGTGACCCGGGAGAGAAGTCACTCTCCCAT
GCCCTCATCGCGTTCCTACTACGGAGTACTCCGTAGTAGCTGCCCATTCCAGGTCGGCTATTT
TGGGTCCAGACGTGCCGCTGCCTTCTTCTTCCGCTCTTCCTTCTTCCTTCCCCCCTTCCCCAGC
CGTAGCTCACACCACACATCCGGCCTGACTGGCATTTCTCTGCCATACTAATTAACACTATCC
CAACTTCTCCACGGTCGTTCATCTTTGGTTTTTCATTCTCAACTAGACTAATTAATTACCGCAT
CGGCGCAGCCAATTCACTTAGGCAAGCTTTGGTAAGTCGCCCCTGTCTGCTTTGCAGAGTTTC
CCCGGCCTTTTCTCCTTGTGA (SEQ ID NO:7)

CTTCCTTCTTCCTTCCCCCCTTCCCCAGCCGTAGCTCACACCACACATCCGGCCTGACTGGCA
TTTCTCTGCCATACTAATTAACACTATCCCAACTTCTCCACGGTCGTTCATCTTTGGTTTTTCA
TTCTCAACTAGACTAATTAATTACCGCATCGGCGCAGCCAATTCACTTAGGCAAGCTTTGGT
AAGTCGCCCCTGTCTGCTTTGCAGAGTTTCCCCGGCCTTTTCTCCTTGTGACAATCCGGTC
(SEQ ID NO:8)

[00112] In some embodiments, the C1Cb promoter of the present invention comprises SEQ ID NO:9 (1.5 kb). In some embodiments the C1Cb promoter comprises a subsequence of SEQ ID NO:9, or a variant thereof, as provided herein. In some embodiments, the C1Cb promoter of the invention comprises SEQ ID NO:10, which is the 3' (3-prime) 1 kb of SEQ ID NO:9, comprising bases 501-1500 of SEQ ID NO:9. In some embodiments, the C1Cb promoter of the invention comprises SEQ ID NO:11, which is the 3' (3-prime) 0.7 kb of SEQ ID NO:9, comprising bases 801-1500 of SEQ ID NO:9.

ACAGGCTTGTTAAAGGAAGTCTTCACGGTCGGCATACAAATCGGCCAAGACGTTGAAGATG
CCATAATACATAATATGTATGTACAATAACCGGGCACATACGGTGCAACTGCTAGTCAGCAA
TTGCGCCTCGCTTTTTGACACAGAGGGCAATGCAGAACGTCGAATCGACCGAGTCAGTTGCT
GTTGGCGTCATGATGGCTCTAGGGTCAGACAGAAGGTCAGAGAAAGCATGCCTTAGCTCGA
AGCCGCTCAGATGTAATTACGCAACGCTCGGCGTTCGAGTTTACGGAGGACGACGGCTACA
AGATGGGGCTGCTTAAAGTTACCTTAAATAGAAAATAGTGCCTGGCTTAAGAGATCATGTCC
GCGGGGCTAGCAAGGATGTCGGGTCTTAACTCGACGGCTCGCCTAGATTTCGTGAAAAGGG
AACTCACTCCCCGACAGGCCCGCAAGTGAATATGTAATTACTCAATGGAAGTTCTCGAAACG
GAGTCCAGAAATGATGTGGTTCTGTGGGAATGCGGCAAGAGGCGACGTTGCCGTGAATGCG
TGAACATTCCCGCCTCTTCTTCTTCTCGTCTTCTTCCTTCTTCTTCTTTCGGGTCGCGGATGGT
TGACGGCCAGCGTGCGCACGGCTGCGTGTTATCGAGCGTCGGTACGTCTAGCCAACATCCCG
TAGACACGACGACCAAGCGTCTTGAGAATGCAACAACGTCTCGGAACCTGGCACGCATCTTC
CGCCGCAGGTCGGCAGACGCCGCCTGGGCAATACCACCCCTGTCCAGGCCCTTTCCCCGCAG
GCAGAGCCGCGCTCTTCCTTTCATGGTTATTCAGGAACGTGGCTTCCGAGATTCTCGCCTGTT
CTCCCCCAGTCAACCTGCCGACCGTAACCCGGTTCCACCACCGCGGACTGTCCGCAAAACCT
GGTTCGCCCGAGATTAATATGCTATTTCCGGACTAAGTGCACAACACACAAGCACCCCTTCC
GCCTCGCGCTCTAGAATCTGCTTTCTAACCCGGTTCTCGGGCCCTTCCCTTTCGCGACGCCTC
CGCTCTCCTTACCAGGCACCATCCGCAATAGGTAAGGTAGCCAACCGTTTTGGAGCGTGATT
CTGCCAAGGACCGCATCCTTGCATTCGCCATCTGGTCAAGGACCCCTCTTTCCCGCTCCATTC
TGGTGGCTCTATCGGGACGGCGTTCCCCATGGCTCTCCAGGAGAGTGATGTGCGAGTCTGGA
GAGCCGGGGTTGGCGTCACGATGCTGCCCACCTAGGGCCGGCCAGCCCGGCACTGCGCTCCC
GTTGATCCGTCTATCCCCGTCAAGAGCACCAGCCCCGGCGCTCGTGAATTTTCGACTTGTTCG
ACTTGCTACAGGTGATAAAGAGGATGCACGCCGCCCTCGATCGGCCTGTGTGGTTTCTCTCC
CTCGTGCCAAACCACTCCCACCTCCCGCCCCGAGATAGTTGCTTGTTTCGCTCCGTGAGAGG
GACACACACCA (SEQ ID NO:9)

AAATGATGTGGTTCTGTGGGAATGCGGCAAGAGGCGACGTTGCCGTGAATGCGTGAACATT
CCCGCCTCTTCTTCTTCTCGTCTTCTTCCTTCTTCTTCTTTCGGGTCGCGGATGGTTGACGGCC
AGCGTGCGCACGGCTGCGTGTTATCGAGCGTCGGTACGTCTAGCCAACATCCCGTAGACACG
ACGACCAAGCGTCTTGAGAATGCAACAACGTCTCGGAACCTGGCACGCATCTTCCGCCGCAG
GTCGGCAGACGCCGCCTGGGCAATACCACCCCTGTCCAGGCCCTTTCCCCGCAGGCAGAGCC
GCGCTCTTCCTTTCATGGTTATTCAGGAACGTGGCTTCCGAGATTCTCGCCTGTTCTCCCCCA
GTCAACCTGCCGACCGTAACCCGGTTCCACCACCGCGGACTGTCCGCAAAACCTGGTTCGCC
CGAGATTAATATGCTATTTCCGGACTAAGTGCACAACACACAAGCACCCCTTCCGCCTCGCG
CTCTAGAATCTGCTTTCTAACCCGGTTCTCGGGCCCTTCCCTTTCGCGACGCCTCCGCTCTCCT
TACCAGGCACCATCCGCAATAGGTAAGGTAGCCAACCGTTTTGGAGCGTGATTCTGCCAAGG
ACCGCATCCTTGCATTCGCCATCTGGTCAAGGACCCCTCTTTCCCGCTCCATTCTGGTGGCTC
TATCGGGACGGCGTTCCCCATGGCTCTCCAGGAGAGTGATGTGCGAGTCTGGAGAGCCGGG
GTTGGCGTCACGATGCTGCCCACCTAGGGCCGGCCAGCCCGGCACTGCGCTCCCGTTGATCC
GTCTATCCCCGTCAAGAGCACCAGCCCCGGCGCTCGTGAATTTTCGACTTGTTCGACTTGCTA
CAGGTGATAAAGAGGATGCACGCCGCCCTCGATCGGCCTGTGTGGTTTCTCTCCCTCGTGCC
AAACCACTCCCACCTCCCGCCCCGAGATAGTTGCTTGTTTCGCTCCGTGAGAGGGACACACA
CCA (SEQ ID NO:10)

CAGGCAGAGCCGCGCTCTTCCTTTCATGGTTATTCAGGAACGTGGCTTCCGAGATTCTCGCCT
GTTCTCCCCCAGTCAACCTGCCGACCGTAACCCGGTTCCACCACCGCGGACTGTCCGCAAAA

CCTGGTTCGCCCGAGATTAATATGCTATTTCCGGACTAAGTGCACAACACACAAGCACCCCT
TCCGCCTCGCGCTCTAGAATCTGCTTTCTAACCCGGTTCTCGGGCCCTTCCCTTTCGCGACGC
CTCCGCTCTCCTTACCAGGCACCATCCGCAATAGGTAAGGTAGCCAACCGTTTTGGAGCGTG
ATTCTGCCAAGGACCGCATCCTTGCATTCGCCATCTGGTCAAGGACCCCTCTTTCCCGCTCCA
TTCTGGTGGCTCTATCGGGACGGCGTTCCCCATGGCTCTCCAGGAGAGTGATGTGCGAGTCT
GGAGAGCCGGGGTTGGCGTCACGATGCTGCCCACCTAGGGCCGGCCAGCCCGGCACTGCGC
TCCCGTTGATCCGTCTATCCCCGTCAAGAGCACCAGCCCCGGCGCTCGTGAATTTTCGACTTG
TTCGACTTGCTACAGGTGATAAAGAGGATGCACGCCGCCCTCGATCGGCCTGTGTGGTTTCT
CTCCCTCGTGCCAAACCACTCCCACCTCCCGCCCCGAGATAGTTGCTTGTTTCGCTCCGTGAG
AGGGACACACACCA (SEQ ID NO:11)

[00113] In some embodiments, the C1Cb promoter of the present invention comprises a subsequence of SEQ ID NO:9, SEQ ID NO:10, and/or SEQ ID NO:11 that retains promoter activity, referred to herein as a "short" promoter sequence. Provided with SEQ ID NO:9, 10, and/or 11, any of a number of different functional deletion mutants of the starting sequence can be readily prepared as described above. In some embodiments, the subsequences retain constitutive promoter activity (*e.g.*, glucose-independent promoter activity).

[00114] In some embodiments, the C1Cb promoter sequence comprise at least about 1000 contiguous nucleotides of SEQ ID NO:9 or 10; at least about 900 nucleotides of SEQ ID NO:9 or 10; at least about 800 nucleotides of SEQ ID NO:9 or 10; at least about 700 nucleotides of SEQ ID NO:9, 10, or 11; at least about 600 nucleotides of SEQ ID NO:9, 10, or 11; at least about 500 nucleotides of SEQ ID NO:9, 10, or 11; at least about 450 nucleotides of SEQ ID NO:9, 10, or 11; at least about 400 nucleotides of SEQ ID NO:9, 10, or 11; at least about 350 nucleotides of SEQ ID NO:9, 10, or 11; at least about 300 nucleotides of SEQ ID NO:9, 10, or 11, at least 250 nucleotides of SEQ ID NO:9, 10, or 11; at least about 200 nucleotides of SEQ ID NO:9, 10, or 11; at least about 150 nucleotides of SEQ ID NO:9, 10, or 11; at least about 100 nucleotides of SEQ ID NO:9, 10, or 11; or at least about 75 contiguous nucleotides of SEQ ID NO:9, 10, or 11.

[00115] In some embodiments, the C1Cb promoter sequence comprises a subsequence of SEQ ID NO:9, comprising about 75 to about 1000 contiguous nucleotides of SEQ ID NO:9 or 10. In some other embodiments, the C1Cb promoter sequence comprises a subsequence of SEQ ID NO:9, 10, or 11, comprising about 50 to about 700, about 50 to about 600, about 50 to about 500, about 50 to about 400, about 50 to about 300, about 50 to about 200, about 75 to about 700, about 75 to about 600, about 75 to about 500, about 75 to about 400, about 75 to about 300, about 75 to about 200, about 100 to about 700, about 100 to about 600, about 100 to about 500, about 100 to about 400, about 100 to about 300, or about 100 to about 200 contiguous nucleotides.

[00116] In some embodiments, the subsequence comprises at least about 25, at least about 50, at least about 100, at least about 150 or at least about 200 contiguous nucleotides of SEQ ID NO:11. In some embodiments, the C1Cb promoter sequence comprises at least one of SEQ ID NOs:12-16, as provided below. SEQ ID NO:12 comprises bases 871 - 1430 of SEQ ID NO:9 (0.56 kb), while SEQ ID NO:13 comprises bases 941 - 1430 of SEQ ID NO:9 (0.49 kb), SEQ ID NO:14 comprises bases 1001 - 1450 of SEQ ID NO:9 (0.45 kb), SEQ ID NO:15 comprises bases 1001 - 1400 of SEQ ID NO:9 (0.40 kb), and SEQ ID NO:16 comprises bases 1201 – 1450 of SEQ ID NO:9 (0.25 kb).

CCCAGTCAACCTGCCGACCGTAACCCGGTTCCACCACCGCGGACTGTCCGCAAAACCTGGTT
CGCCCGAGATTAATATGCTATTTCCGGACTAAGTGCACAACACACAAGCACCCCTTCCGCCT
CGCGCTCTAGAATCTGCTTTCTAACCCGGTTCTCGGGCCCTTCCCTTTCGCGACGCCTCCGCT
CTCCTTACCAGGCACCATCCGCAATAGGTAAGGTAGCCAACCGTTTTGGAGCGTGATTCTGC
CAAGGACCGCATCCTTGCATTCGCCATCTGGTCAAGGACCCCTCTTTCCCGCTCCATTCTGGT
GGCTCTATCGGGACGGCGTTCCCCATGGCTCTCCAGGAGAGTGATGTGCGAGTCTGGAGAGC
CGGGGTTGGCGTCACGATGCTGCCCACCTAGGGCCGGCCAGCCCGGCACTGCGCTCCCGTTG
ATCCGTCTATCCCCGTCAAGAGCACCAGCCCCGGCGCTCGTGAATTTTCGACTTGTTCGACTT
GCTACAGGTGATAAAGAGGATGCACGCCGCCCTCGATCGGCCTGTGTGGTTTCTCTCCCTC
(SEQ ID NO:12)

ATTAATATGCTATTTCCGGACTAAGTGCACAACACACAAGCACCCCTTCCGCCTCGCGCTCT
AGAATCTGCTTTCTAACCCGGTTCTCGGGCCCTTCCCTTTCGCGACGCCTCCGCTCTCCTTAC
CAGGCACCATCCGCAATAGGTAAGGTAGCCAACCGTTTTGGAGCGTGATTCTGCCAAGGACC
GCATCCTTGCATTCGCCATCTGGTCAAGGACCCCTCTTTCCCGCTCCATTCTGGTGGCTCTAT
CGGGACGGCGTTCCCCATGGCTCTCCAGGAGAGTGATGTGCGAGTCTGGAGAGCCGGGGTT
GGCGTCACGATGCTGCCCACCTAGGGCCGGCCAGCCCGGCACTGCGCTCCCGTTGATCCGTC
TATCCCCGTCAAGAGCACCAGCCCCGGCGCTCGTGAATTTTCGACTTGTTCGACTTGCTACA
GGTGATAAAGAGGATGCACGCCGCCCTCGATCGGCCTGTGTGGTTTCTCTCCCTC (SEQ ID
NO:13)

CTAGAATCTGCTTTCTAACCCGGTTCTCGGGCCCTTCCCTTTCGCGACGCCTCCGCTCTCCTT
ACCAGGCACCATCCGCAATAGGTAAGGTAGCCAACCGTTTTGGAGCGTGATTCTGCCAAGG
ACCGCATCCTTGCATTCGCCATCTGGTCAAGGACCCCTCTTTCCCGCTCCATTCTGGTGGCTC
TATCGGGACGGCGTTCCCCATGGCTCTCCAGGAGAGTGATGTGCGAGTCTGGAGAGCCGGG
GTTGGCGTCACGATGCTGCCCACCTAGGGCCGGCCAGCCCGGCACTGCGCTCCCGTTGATCC
GTCTATCCCCGTCAAGAGCACCAGCCCCGGCGCTCGTGAATTTTCGACTTGTTCGACTTGCTA
CAGGTGATAAAGAGGATGCACGCCGCCCTCGATCGGCCTGTGTGGTTTCTCTCCCTCGTGCC
AAACCACTCCCACCT (SEQ ID NO:14)

CTAGAATCTGCTTTCTAACCCGGTTCTCGGGCCCTTCCCTTTCGCGACGCCTCCGCTCTCCTT
ACCAGGCACCATCCGCAATAGGTAAGGTAGCCAACCGTTTTGGAGCGTGATTCTGCCAAGG
ACCGCATCCTTGCATTCGCCATCTGGTCAAGGACCCCTCTTTCCCGCTCCATTCTGGTGGCTC
TATCGGGACGGCGTTCCCCATGGCTCTCCAGGAGAGTGATGTGCGAGTCTGGAGAGCCGGG
GTTGGCGTCACGATGCTGCCCACCTAGGGCCGGCCAGCCCGGCACTGCGCTCCCGTTGATCC

GTCTATCCCCGTCAAGAGCACCAGCCCCGGCGCTCGTGAATTTTCGACTTGTTCGACTTGCTA
CAGGTGATAAAGAGGATGCACGCCGCC (SEQ ID NO:15)

TTCCCCATGGCTCTCCAGGAGAGTGATGTGCGAGTCTGGAGAGCCGGGGTTGGCGTCACGAT
GCTGCCCACCTAGGGCCGGCCAGCCCGGCACTGCGCTCCCGTTGATCCGTCTATCCCCGTCA
AGAGCACCAGCCCCGGCGCTCGTGAATTTTCGACTTGTTCGACTTGCTACAGGTGATAAAGA
GGATGCACGCCGCCCTCGATCGGCCTGTGTGGTTTCTCTCCCTCGTGCCAAACCACTCCCACC
T (SEQ ID NO:16)

[00117] In some embodiments, the C1Cc promoter of the invention comprises SEQ ID NO:17 (1.5 kb). In some embodiments, the C1Cc promoter comprises a subsequence of SEQ ID NO:17, or a variant thereof, as provided herein. In some embodiments, the C1Cc promoter of the invention comprises SEQ ID NO:18, which is the 3' (3-prime) 1 kb of SEQ ID NO:17, which comprises bases 501-1500 of SEQ ID NO:17. In some other embodiments, the C1Cc promoter of the invention comprises SEQ ID NO:19, which is the 3' (3-prime) 0.7 kb of SEQ ID NO:17, which comprises bases 801-1500 of SEQ ID NO:17.

ATGCGCTCCGCCTTGTGCGCCTCGTGCAGCCGCAGCACCTCGATCAGCACGCGCACCGAGCC
GATCGGGTGGTACAGGTACTCGGACGCCGGCGGGAGCATGTCGGCGAAGGGGGACGTCCTC
TGGAACGAGAGGGCGAACGTGTAGATGGCCAGCCCCGTGAGCGTGCCCTGTTGTTTTGTTTT
TTTGTTTTGTTTTGTTTCCCGTGGTTAGAGGAGGTTTTTTTCGTTTTAGAATGGAAGGAAGA
AAGGAAAGAAGGAAACTTTGGGGCCGGGTAATGAGGGAGGGAAACGTACGACGGTGATGA
CAACATGGATCCAGCGGCTATGCAGGAACCAGTGGGCGAACGTGTTTGGCGGCGGCGGCAG
GCCCGGGTAGTCGCGGGAGGCTTGCGCGCTGGCCTCCTCGAAGCTGAGGTCGCCGCCGTAGT
GGCGCGGGATGCTCTTCTTCGGCAGGCGGGAGCCGTGTGAGGGGGGATTGAAGCGCTCCGG
CTTGGCGAGCACCCGCTGGCCGGACGGAGTGCCGGCGTTGCTGCTGCTGCTGCTGCTGCTGC
TGTTCCTGGCGAGCCGCAGGAATACGCGGCTCCGGGTCGAGACGGAGGCCGGAGCCGAGGC
CGCCGGAGCCGCCCGGCGCCATATCAGGCCGGTCTGCGGGAGGAGGAAGCGGGGGATGATG
GGTGCTGCTGGTGCCGGCGTCATGACTACTATAACTAACTGCCGCCGCGCGTGTCCGCACAA
ATTTCGAGTGAGCGAGGAATGAATTCGGATTGAGGTAATCCGTAGTGTACGAGCGAGATCC
CTCGAAACGAGGGGAGGCAATCAAAGATTCTTTGTCTCCTCTCTCCTCCCTCTTTTTTTTTGC
TTATCCCCGGTTCCTCTCGGCGACAGAAATGCAACTCGGTTTTTCTGGGTGCCCGATCGGGG
GTCCCTCGGCGTCGGGGCAACAAGGCAATTCGCAGGGTCGCGGACGTTGCGGTGCGGCTCA
ATCAGGCGATATGCGAGTGGTCAGAAAATTCGCCTGCGTCAAGTTGCTGCAGGTTTCTGCTG
CTATCCCATTCCGGCTAGCGCTTCTCTTCTGCTGTGCAGTACTCCGTACACTATAGTAGCTCG
CGGTCCTCGGGCCAAGGCGCGTCTTTGGGTTGCCCGGGGGGGGGGGTGGCGGCGCGCCAAC
AGTGCCGGTCGCTCCCGAATTTGCCCGGGGCGACTGACTAACAGTCGAAACATGATTGGCAC
AAGTTAGAAAATAGGTGGGTCATTTTTCCACGGATTACCATGGCTCGCTCGTTGGATGATCA
AGGCTTGGCAGTGTTCATCGATGCAAAAAATCCGGCGCGCGGACCTGGCACGGCGATTGCA
GCAAACTAACACCTCATTCCGAAATTTTTCTTGAACTCTTTCCTACTTCCCTTCACATCCGAC
CTTGCTTCGCAATATCTGCTCTTCCTCACCAACACCGACTCCTCTCAGACACTCAATCCTCTC
ACTACCCCAACCGTCAAG (SEQ ID NO:17)

GCACCCGCTGGCCGGACGGAGTGCCGGCGTTGCTGCTGCTGCTGCTGCTGCTGCTGTTCCTG
GCGAGCCGCAGGAATACGCGGCTCCGGGTCGAGACGGAGGCCGGAGCCGAGGCCGCCGGA

GCCGCCCGGCGCCATATCAGGCCGGTCTGCGGGAGGAGGAAGCGGGGGATGATGGGTGCTG
CTGGTGCCGGCGTCATGACTACTATAACTAACTGCCGCCGCGCGTGTCCGCACAAATTTCGA
GTGAGCGAGGAATGAATTCGGATTGAGGTAATCCGTAGTGTACGAGCGAGATCCCTCGAAA
CGAGGGGAGGCAATCAAAGATTCTTTGTCTCCTCTCTCCTCCCTCTTTTTTTTTGCTTATCCCC
GGTTCCTCTCGGCGACAGAAATGCAACTCGGTTTTTCTGGGTGCCCGATCGGGGGTCCCTCG
GCGTCGGGGCAACAAGGCAATTCGCAGGGTCGCGGACGTTGCGGTGCGGCTCAATCAGGCG
ATATGCGAGTGGTCAGAAAATTCGCCTGCGTCAAGTTGCTGCAGGTTTCTGCTGCTATCCCA
TTCCGGCTAGCGCTTCTCTTCTGCTGTGCAGTACTCCGTACACTATAGTAGCTCGCGGTCCTC
GGGCCAAGGCGCGTCTTTGGGTTGCCCGGGGGGGGGGGTGGCGGCGCGCCAACAGTGCCGG
TCGCTCCCGAATTTGCCCGGGGCGACTGACTAACAGTCGAAACATGATTGGCACAAGTTAGA
AAATAGGTGGGTCATTTTTCCACGGATTACCATGGCTCGCTCGTTGGATGATCAAGGCTTGG
CAGTGTTCATCGATGCAAAAAATCCGGCGCGCGGACCTGGCACGGCGATTGCAGCAAACTA
ACACCTCATTCCGAAATTTTTCTTGAACTCTTTCCTACTTCCCTTCACATCCGACCTTGCTTCG
CAATATCTGCTCTTCCTCACCAACACCGACTCCTCTCAGACACTCAATCCTCTCACTACCCCA
ACCGTCAAG (SEQ ID NO:18)

TCGAAACGAGGGGAGGCAATCAAAGATTCTTTGTCTCCTCTCTCCTCCCTCTTTTTTTTTGCTT
ATCCCCGGTTCCTCTCGGCGACAGAAATGCAACTCGGTTTTTCTGGGTGCCCGATCGGGGGT
CCCTCGGCGTCGGGGCAACAAGGCAATTCGCAGGGTCGCGGACGTTGCGGTGCGGCTCAAT
CAGGCGATATGCGAGTGGTCAGAAAATTCGCCTGCGTCAAGTTGCTGCAGGTTTCTGCTGCT
ATCCCATTCCGGCTAGCGCTTCTCTTCTGCTGTGCAGTACTCCGTACACTATAGTAGCTCGCG
GTCCTCGGGCCAAGGCGCGTCTTTGGGTTGCCCGGGGGGGGGGGTGGCGGCGCGCCAACAG
TGCCGGTCGCTCCCGAATTTGCCCGGGGCGACTGACTAACAGTCGAAACATGATTGGCACAA
GTTAGAAAATAGGTGGGTCATTTTTCCACGGATTACCATGGCTCGCTCGTTGGATGATCAAG
GCTTGGCAGTGTTCATCGATGCAAAAAATCCGGCGCGCGGACCTGGCACGGCGATTGCAGC
AAACTAACACCTCATTCCGAAATTTTTCTTGAACTCTTTCCTACTTCCCTTCACATCCGACCTT
GCTTCGCAATATCTGCTCTTCCTCACCAACACCGACTCCTCTCAGACACTCAATCCTCTCACT
ACCCCAACCGTCAAG (SEQ ID NO:19)

[00118] In some embodiments, the C1Cc promoter of the invention comprises a subsequence of SEQ ID NO:17, SEQ ID NO:18, and/or SEQ ID NO:19 that retains promoter activity, referred to herein as a “short” promoter sequence. Subsequences that retain promoter activity are identified using routine methods well known to those in the art. Provided with SEQ ID NO:17, 18, and/or 19, any of a number of different functional deletion mutants of the starting sequence can be readily prepared. In some embodiments, the subsequences retain constitutive promoter activity (*e.g.,* glucose-independent promoter activity).

[00119] In some embodiments, the C1Cc promoter sequence comprises at least about 1000 contiguous nucleotides of SEQ ID NO:17 or 18; at least about 900 nucleotides of SEQ ID NO:17 or18; at least about 800 nucleotides of SEQ ID NO:17 or 18; at least about 700 nucleotides of SEQ ID NO:17, 18, or 19; at least about 600 nucleotides of SEQ ID NO:17, 18, or 19; at least about 500 nucleotides of SEQ ID NO:17, 18, or 19; at least about 450 nucleotides of SEQ ID NO:17, 18, or 19; at least about 400 nucleotides of SEQ ID NO:17, 18, or 19; at least about 350 nucleotides of SEQ ID NO:17, 18, or 19; at

least about 300 nucleotides of SEQ ID NO:17, 18, or 19; at least about 250 nucleotides of SEQ ID NO:17, 18, or 19; at least about 200 nucleotides of SEQ ID NO:17, 18, or 19; at least about 150 nucleotides of SEQ ID NO:17, 18, or 19; at least about 100 nucleotides of SEQ ID NO:17, 18, or 19; or at least about 75 contiguous nucleotides of SEQ ID NO:17, 18, or 19.

[00120] In some embodiments, the C1Cc promoter sequence comprises a subsequence of comprising about 75 to about 1000 contiguous nucleotides of SEQ ID NO:17 or 18. In some other embodiments the C1Cc promoter sequence comprises a subsequence of SEQ ID NO:17, 18, or 19 comprising about 50 to about 700, about 50 to about 600, about 50 to about 500, about 50 to about 400, about 50 to about 300, about 50 to about 200, about 75 to about 700, about 75 to about 600, about 75 to about 500, about 75 to about 400, about 75 to about 300, about 75 to about 200, about 100 to about 700, about 100 to about 600, about 100 to about 500, about 100 to about 400, about 100 to about 300, or about 100 to about 200 contiguous nucleotides.

[00121] In some embodiments the subsequence comprises at least about 25, at least about 50, at least about 100, at least about 150 or at least about 200 contiguous nucleotides of SEQ ID NO:19. In some embodiments, the C1Cc promoter sequence comprises at least one of SEQ ID NOS:20-24. SEQ ID NO:20 comprises bases 871 - 1430 of SEQ ID NO:17 (0.56 kb), while SEQ ID NO:21 comprises bases 941 - 1430 of SEQ ID NO:17 (0.49 kb), SEQ ID NO:22 comprises bases 1001 - 1450 of SEQ ID NO:17 (0.45 kb), SEQ ID NO:23 comprises bases 1001 - 1400 of SEQ ID NO:17 (0.40 kb), and SEQ ID NO:24 comprises bases 1201 – 1450 of SEQ ID NO:17 (0.25 kb).

GGTTCCTCTCGGCGACAGAAATGCAACTCGGTTTTTCTGGGTGCCCGATCGGGGGTCCCTCG
GCGTCGGGGCAACAAGGCAATTCGCAGGGTCGCGGACGTTGCGGTGCGGCTCAATCAGGCG
ATATGCGAGTGGTCAGAAAATTCGCCTGCGTCAAGTTGCTGCAGGTTTCTGCTGCTATCCCA
TTCCGGCTAGCGCTTCTCTTCTGCTGTGCAGTACTCCGTACACTATAGTAGCTCGCGGTCCTC
GGGCCAAGGCGCGTCTTTGGGTTGCCCGGGGGGGGGGGTGGCGGCGCGCCAACAGTGCCGG
TCGCTCCCGAATTTGCCCGGGGCGACTGACTAACAGTCGAAACATGATTGGCACAAGTTAGA
AAATAGGTGGGTCATTTTTCCACGGATTACCATGGCTCGCTCGTTGGATGATCAAGGCTTGG
CAGTGTTCATCGATGCAAAAAATCCGGCGCGCGGACCTGGCACGGCGATTGCAGCAAACTA
ACACCTCATTCCGAAATTTTTCTTGAACTCTTTCCTACTTCCCTTCACATCCGACCTTGCTTCG
CA (SEQ ID NO:20)

GCAACAAGGCAATTCGCAGGGTCGCGGACGTTGCGGTGCGGCTCAATCAGGCGATATGCGA
GTGGTCAGAAAATTCGCCTGCGTCAAGTTGCTGCAGGTTTCTGCTGCTATCCCATTCCGGCTA
GCGCTTCTCTTCTGCTGTGCAGTACTCCGTACACTATAGTAGCTCGCGGTCCTCGGGCCAAGG
CGCGTCTTTGGGTTGCCCGGGGGGGGGGGTGGCGGCGCGCCAACAGTGCCGGTCGCTCCCG
AATTTGCCCGGGGCGACTGACTAACAGTCGAAACATGATTGGCACAAGTTAGAAAATAGGT
GGGTCATTTTTCCACGGATTACCATGGCTCGCTCGTTGGATGATCAAGGCTTGGCAGTGTTCA
TCGATGCAAAAAATCCGGCGCGCGGACCTGGCACGGCGATTGCAGCAAACTAACACCTCAT

TCCGAAATTTTTCTTGAACTCTTTCCTACTTCCCTTCACATCCGACCTTGCTTCGCA (SEQ ID NO:21)

AGTGGTCAGAAAATTCGCCTGCGTCAAGTTGCTGCAGGTTTCTGCTGCTATCCCATTCCGGCT
AGCGCTTCTCTTCTGCTGTGCAGTACTCCGTACACTATAGTAGCTCGCGGTCCTCGGGCCAAG
GCGCGTCTTTGGGTTGCCCGGGGGGGGGGGTGGCGGCGCGCCAACAGTGCCGGTCGCTCCC
GAATTTGCCCGGGGCGACTGACTAACAGTCGAAACATGATTGGCACAAGTTAGAAAATAGG
TGGGTCATTTTTCCACGGATTACCATGGCTCGCTCGTTGGATGATCAAGGCTTGGCAGTGTTC
ATCGATGCAAAAAATCCGGCGCGCGGACCTGGCACGGCGATTGCAGCAAACTAACACCTCA
TTCCGAAATTTTTCTTGAACTCTTTCCTACTTCCCTTCACATCCGACCTTGCTTCGCAATATCT
GCTCTTCCTCACCA (SEQ ID NO:22)

AGTGGTCAGAAAATTCGCCTGCGTCAAGTTGCTGCAGGTTTCTGCTGCTATCCCATTCCGGCT
AGCGCTTCTCTTCTGCTGTGCAGTACTCCGTACACTATAGTAGCTCGCGGTCCTCGGGCCAAG
GCGCGTCTTTGGGTTGCCCGGGGGGGGGGGTGGCGGCGCGCCAACAGTGCCGGTCGCTCCC
GAATTTGCCCGGGGCGACTGACTAACAGTCGAAACATGATTGGCACAAGTTAGAAAATAGG
TGGGTCATTTTTCCACGGATTACCATGGCTCGCTCGTTGGATGATCAAGGCTTGGCAGTGTTC
ATCGATGCAAAAAATCCGGCGCGCGGACCTGGCACGGCGATTGCAGCAAACTAACACCTCA
TTCCGAAATTTTTCTTGAACTCTTTCCT (SEQ ID NO:23)

GCGACTGACTAACAGTCGAAACATGATTGGCACAAGTTAGAAAATAGGTGGGTCATTTTTCC
ACGGATTACCATGGCTCGCTCGTTGGATGATCAAGGCTTGGCAGTGTTCATCGATGCAAAAA
ATCCGGCGCGCGGACCTGGCACGGCGATTGCAGCAAACTAACACCTCATTCCGAAATTTTTC
TTGAACTCTTTCCTACTTCCCTTCACATCCGACCTTGCTTCGCAATATCTGCTCTTCCTCACCA
(SEQ ID NO:24)

[00122] In some embodiments, the C1Cd promoter of the invention comprises SEQ ID NO:25 (1.5 kb). In some embodiments, the C1Cd promoter comprises a subsequence of SEQ ID NO:25, or a variant thereof, as provided herein. In some embodiments, the C1Cd promoter of the invention comprises SEQ ID NO:26, which is the 3' (3-prime) 1 kb of SEQ ID NO:25, comprising bases 501-1500 of SEQ ID NO:25. In some embodiments, the C1Cd promoter of the invention comprises SEQ ID NO:27, which is the 3' (3-prime) 0.7 kb of SEQ ID NO:25, comprising bases 801-1500 of SEQ ID NO:25.

GCTCGAGGATGTCAAAGTCCACATTTCATTTTTTTTCTTTTTCTTTTTTTTTTTCCGGATAGGC
ATCTTCCATTTGGGAATCATAGATGTCTGCTTACGGGGCGGTGTGCGAGCCGTGGTGCTCGG
CTTTTTGGGGTAAGAATGGCCCATGACGGAGCGATGGCGGTTTTAGTTCAAGGTGCTCGTGT
CCTGATGATAGATGATATTGGTGTGACGTGGTGTGTTCTGCAGATTTTTGAAGCTTGGGGGA
TGTAACTCGGCCCAAGAGGAAAGTGCGGAGATGTGTGCTCAATCGAGGTACTTAATGTTCCG
TATTCTTTCCTCCTCAGGCTATTATCTCGTGCAGTGGCGAATCTGAAGAGTGTCACGGGTACT
TCGTAGGTACCAAAGCTTCACCGTCTTTGGACACGTCGATAGCGACGTACCTGCAGCAGTGG
TCTGATGTGTCTGAAATTTGTCTTGATTCCCGATGGCGACAGGTGTCTGTTATGAGAACCTAC
CTGACCGAGAGTGGCCAGGCAAGAGAACCAATAGCTATATTTTCAAACTCGCTATTTCAAGC

TTGACCTCGAAATGGAAAACCGACTATCAGCAGTGACAATCAATCACGGGCCAGAGTGCAT
TAAATGGATGTACTGTGGGATGCGGAAAGCGAACTATAGTATCTTCGTTTAACTGCTACTGC
TGCTGGTAAGTGGTGGTCGAAGGAAGCGAAGGCTGACTGGGGCCACCGTGCAGGAAGATAT
GGGTGGCTGTAACCCTGGTGGGGCCGGGAGCCCTAGTGGGGCCCGACCACCACAAGTATGT
ACTGTGTACGTCCGTATATACGGATTACATACATACCTACACAGTATAATTATCTGCGCATTG
ATTTCCGGAGAAACTACTCCGTACCTAGGTATACAGAAAAGAACCGCCAACGAAAAGTAAT
TAATTACGTACGCATCACGACTCGCACTCCTTTCCAGCGTACAAGGATTGTTTTGATTCCCTT
TGAGGATGACATTCATTCCACGATACCAATGAGATAGCGGGTTTGGACATTTTTGACTCGAA
CGGAAATGATGAACAGCAAGCAGTATTAGTCGGCTCTCACACGCACACTGGCATCAAGCAG
CAATCGAACACTTGCCGACTCAACGCATCATGACGGCAAAAACCCACGTGGGCATGATGTC
CAAGTCCCTATATTCAGGAACCCCCCGGACCAGATGACGCATGGTACGGTACCTATGTGACA
TCAGGCTCGCCACCAGTTGTCTGTTCCCTATTATAATCCGCCTATTAATTAATTAGTAGCTCT
GATTTGTAAAAGTGCAAGCCTGTTCTGATCATCTTCATGACCTCTACTCTGCAAGTCCGAACA
AGAGATCACCATCAATTGCATATTATTTGATAATTAATACACTGTATCTGTACAAAGAACAG
CACAAAACATATTTTTCTCCTGAATAATTATTACTACTCCCCCAGACCCAAACAAAAAAGTC
TAATTACACC (SEQ ID NO:25)

CTGACCGAGAGTGGCCAGGCAAGAGAACCAATAGCTATATTTTCAAACTCGCTATTTCAAGC
TTGACCTCGAAATGGAAAACCGACTATCAGCAGTGACAATCAATCACGGGCCAGAGTGCAT
TAAATGGATGTACTGTGGGATGCGGAAAGCGAACTATAGTATCTTCGTTTAACTGCTACTGC
TGCTGGTAAGTGGTGGTCGAAGGAAGCGAAGGCTGACTGGGGCCACCGTGCAGGAAGATAT
GGGTGGCTGTAACCCTGGTGGGGCCGGGAGCCCTAGTGGGGCCCGACCACCACAAGTATGT
ACTGTGTACGTCCGTATATACGGATTACATACATACCTACACAGTATAATTATCTGCGCATTG
ATTTCCGGAGAAACTACTCCGTACCTAGGTATACAGAAAAGAACCGCCAACGAAAAGTAAT
TAATTACGTACGCATCACGACTCGCACTCCTTTCCAGCGTACAAGGATTGTTTTGATTCCCTT
TGAGGATGACATTCATTCCACGATACCAATGAGATAGCGGGTTTGGACATTTTTGACTCGAA
CGGAAATGATGAACAGCAAGCAGTATTAGTCGGCTCTCACACGCACACTGGCATCAAGCAG
CAATCGAACACTTGCCGACTCAACGCATCATGACGGCAAAAACCCACGTGGGCATGATGTC
CAAGTCCCTATATTCAGGAACCCCCCGGACCAGATGACGCATGGTACGGTACCTATGTGACA
TCAGGCTCGCCACCAGTTGTCTGTTCCCTATTATAATCCGCCTATTAATTAATTAGTAGCTCT
GATTTGTAAAAGTGCAAGCCTGTTCTGATCATCTTCATGACCTCTACTCTGCAAGTCCGAACA
AGAGATCACCATCAATTGCATATTATTTGATAATTAATACACTGTATCTGTACAAAGAACAG
CACAAAACATATTTTTCTCCTGAATAATTATTACTACTCCCCCAGACCCAAACAAAAAAGTC
TAATTACACC (SEQ ID NO:26)

AGTATGTACTGTGTACGTCCGTATATACGGATTACATACATACCTACACAGTATAATTATCTG
CGCATTGATTTCCGGAGAAACTACTCCGTACCTAGGTATACAGAAAAGAACCGCCAACGAA
AAGTAATTAATTACGTACGCATCACGACTCGCACTCCTTTCCAGCGTACAAGGATTGTTTTG
ATTCCCTTTGAGGATGACATTCATTCCACGATACCAATGAGATAGCGGGTTTGGACATTTTTG
ACTCGAACGGAAATGATGAACAGCAAGCAGTATTAGTCGGCTCTCACACGCACACTGGCAT
CAAGCAGCAATCGAACACTTGCCGACTCAACGCATCATGACGGCAAAAACCCACGTGGGCA
TGATGTCCAAGTCCCTATATTCAGGAACCCCCCGGACCAGATGACGCATGGTACGGTACCTA
TGTGACATCAGGCTCGCCACCAGTTGTCTGTTCCCTATTATAATCCGCCTATTAATTAATTAG
TAGCTCTGATTTGTAAAAGTGCAAGCCTGTTCTGATCATCTTCATGACCTCTACTCTGCAAGT
CCGAACAAGAGATCACCATCAATTGCATATTATTTGATAATTAATACACTGTATCTGTACAA
AGAACAGCACAAAACATATTTTTCTCCTGAATAATTATTACTACTCCCCCAGACCCAAACAA
AAAAGTCTAATTACACC (SEQ ID NO:27)

[00123] In some embodiments, the C1Cd promoter of the invention comprises a subsequence of SEQ ID NO:25, SEQ ID NO:26, and/or SEQ ID NO:27 that retains promoter activity, referred to herein as a “short” promoter sequence. Subsequences that retain promoter activity are identified using routine methods well known to those in the art. Provided with SEQ ID NO:25, 26, and/or 27, any of a number of different functional deletion mutants of the starting sequence can be readily prepared. In some embodiments, the subsequences retain constitutive promoter activity (*e.g.*, glucose-independent promoter activity).

[00124] In some embodiments, the C1Cd promoter sequence comprises at least about 1000 contiguous nucleotides of SEQ ID NO:25 or 26; at least about 900 nucleotides of SEQ ID NO:25 or 26; at least about 800 nucleotides of SEQ ID NO:25 or 26; at least about 700 nucleotides of SEQ ID NO:25, 26, or 27; at least about 600 nucleotides of SEQ ID NO:25, 26, or 27; at least about 500 nucleotides of SEQ ID NO:25, 26, or 27; at least about 450 nucleotides of SEQ ID NO:25, 26, or 27; at least about 400 nucleotides of SEQ ID NO:25, 26, or 27; at least about 350 nucleotides of SEQ ID NO:25, 26, or 27; at least about 300 nucleotides of SEQ ID NO:25, 26, or 27; at least about 250 nucleotides of SEQ ID NO:25, 26, or 27; at least about 200 nucleotides of SEQ ID NO:25, 26, or 27; at least about 150 nucleotides of SEQ ID NO:25, 26 or 27; at least about 100 nucleotides of SEQ ID NO:25, 26, or 27; or at least about 75 contiguous nucleotides of SEQ ID NO:25, 26, or 27.

[00125] In some embodiments, the C1Cd promoter sequence comprises a subsequence comprising about 75 to about 1000 contiguous nucleotides of SEQ ID NO:25 or 26. In some other embodiments, the C1Cd promoter sequence comprises a subsequence of SEQ ID NO:25, 26, or 27 comprising about 50 to about 700, about 50 to about 600, about 50 to about 500, about 50 to about 400, about 50 to about 300, about 50 to about 200, about 75 to about 700, about 75 to about 600, about 75 to about 500, about 75 to about 400, about 75 to about 300, about 75 to about 200, about 100 to about 700, about 100 to about 600, about 100 to about 500, about 100 to about 400, about 100 to about 300, or about 100 to about 200 contiguous nucleotides. In some embodiments, the subsequence comprises at least about 25, at least about 50, at least about 100, at least about 150, or at least about 200 contiguous nucleotides of SEQ ID NO:27. In some embodiments the C1Cd promoter sequence comprises at least one of SEQ ID NOS:28-32. SEQ ID NO:28 comprises bases 871 - 1430 of SEQ ID NO:25 (0.56 kb), while SEQ ID NO:29 comprises bases 941 - 1430 of SEQ ID NO:25(0.49 kb), SEQ ID NO:30 comprises bases 1001 - 1450 of SEQ ID NO:25 (0.45 kb), SEQ ID NO:31 comprises bases 1001 - 1400 of SEQ ID NO:25 (0.40 kb), and SEQ ID NO:32 comprises bases 1201 – 1450 of SEQ ID NO:25 (0.25 kb).

ATTTCCGGAGAAACTACTCCGTACCTAGGTATACAGAAAAGAACCGCCAACGAAAAGTAAT
TAATTACGTACGCATCACGACTCGCACTCCTTTCCAGCGTACAAGGATTGTTTTGATTCCCTT
TGAGGATGACATTCATTCCACGATACCAATGAGATAGCGGGTTTGGACATTTTTGACTCGAA
CGGAAATGATGAACAGCAAGCAGTATTAGTCGGCTCTCACACGCACACTGGCATCAAGCAG
CAATCGAACACTTGCCGACTCAACGCATCATGACGGCAAAAACCCACGTGGGCATGATGTC
CAAGTCCCTATATTCAGGAACCCCCCGGACCAGATGACGCATGGTACGGTACCTATGTGACA
TCAGGCTCGCCACCAGTTGTCTGTTCCCTATTATAATCCGCCTATTAATTAATTAGTAGCTCT
GATTTGTAAAAGTGCAAGCCTGTTCTGATCATCTTCATGACCTCTACTCTGCAAGTCCGAACA
AGAGATCACCATCAATTGCATATTATTTGATAATTAATACACTGTATCTGTACAAAGAACAG
CA (SEQ ID NO:28)

ACGCATCACGACTCGCACTCCTTTCCAGCGTACAAGGATTGTTTTGATTCCCTTTGAGGATGA
CATTCATTCCACGATACCAATGAGATAGCGGGTTTGGACATTTTTGACTCGAACGGAAATGA
TGAACAGCAAGCAGTATTAGTCGGCTCTCACACGCACACTGGCATCAAGCAGCAATCGAAC
ACTTGCCGACTCAACGCATCATGACGGCAAAAACCCACGTGGGCATGATGTCCAAGTCCCTA
TATTCAGGAACCCCCCGGACCAGATGACGCATGGTACGGTACCTATGTGACATCAGGCTCGC
CACCAGTTGTCTGTTCCCTATTATAATCCGCCTATTAATTAATTAGTAGCTCTGATTTGTAAA
AGTGCAAGCCTGTTCTGATCATCTTCATGACCTCTACTCTGCAAGTCCGAACAAGAGATCAC
CATCAATTGCATATTATTTGATAATTAATACACTGTATCTGTACAAAGAACAGCA (SEQ ID
NO:29)

TGACATTCATTCCACGATACCAATGAGATAGCGGGTTTGGACATTTTTGACTCGAACGGAAA
TGATGAACAGCAAGCAGTATTAGTCGGCTCTCACACGCACACTGGCATCAAGCAGCAATCG
AACACTTGCCGACTCAACGCATCATGACGGCAAAAACCCACGTGGGCATGATGTCCAAGTC
CCTATATTCAGGAACCCCCCGGACCAGATGACGCATGGTACGGTACCTATGTGACATCAGGC
TCGCCACCAGTTGTCTGTTCCCTATTATAATCCGCCTATTAATTAATTAGTAGCTCTGATTTGT
AAAAGTGCAAGCCTGTTCTGATCATCTTCATGACCTCTACTCTGCAAGTCCGAACAAGAGAT
CACCATCAATTGCATATTATTTGATAATTAATACACTGTATCTGTACAAAGAACAGCACAAA
ACATATTTTTCTCCTG (SEQ ID NO:30)

TGACATTCATTCCACGATACCAATGAGATAGCGGGTTTGGACATTTTTGACTCGAACGGAAA
TGATGAACAGCAAGCAGTATTAGTCGGCTCTCACACGCACACTGGCATCAAGCAGCAATCG
AACACTTGCCGACTCAACGCATCATGACGGCAAAAACCCACGTGGGCATGATGTCCAAGTC
CCTATATTCAGGAACCCCCCGGACCAGATGACGCATGGTACGGTACCTATGTGACATCAGGC
TCGCCACCAGTTGTCTGTTCCCTATTATAATCCGCCTATTAATTAATTAGTAGCTCTGATTTGT
AAAAGTGCAAGCCTGTTCTGATCATCTTCATGACCTCTACTCTGCAAGTCCGAACAAGAGAT
CACCATCAATTGCATATTATTTGATAAT (SEQ ID NO:31)

CCCCGGACCAGATGACGCATGGTACGGTACCTATGTGACATCAGGCTCGCCACCAGTTGTCT
GTTCCCTATTATAATCCGCCTATTAATTAATTAGTAGCTCTGATTTGTAAAAGTGCAAGCCTG
TTCTGATCATCTTCATGACCTCTACTCTGCAAGTCCGAACAAGAGATCACCATCAATTGCATA
TTATTTGATAATTAATACACTGTATCTGTACAAAGAACAGCACAAAACATATTTTTCTCCTG
(SEQ ID NO:32)

[00126] In some embodiments, the C1Ce promoter of the invention comprises SEQ ID NO:33 (1.5 kb). In some embodiments, the C1Ce promoter comprises a subsequence of SEQ ID NO:33, or a

variant thereof, as provided herein. In some embodiments, the C1Ce promoter of the invention comprises SEQ ID NO:34, which is the 3' (3-prime) 1 kb of SEQ ID NO:33, comprising bases 501-1500 of SEQ ID NO:33. In some embodiments, the C1Ce promoter of the invention comprises SEQ ID NO:35, which is the 3' (3-prime) 0.7 kb of SEQ ID NO:33, comprising bases 801-1500 of SEQ ID NO:33.

CCAACCGAACCCCCATCGTCGCAGCCCCTCTCGCTTTTGAAACGGCTCCCAAGCCACTTAAA
CCCGCTAGAGCACTCTCTCAAGCCAGCGCGGTGGGTCTAGCCTACGTACGATACACCCCCAC
CCCCAAACAACCGTGACAGGATACACGACTCCCACAACACAATGGATAGGATGCACGATGG
ATGAACCGAGGACGGAGGCACACAAGAAATGCAATGTGGCCCTTCTCGGCGGAAGCACACG
GGCTGTAGGGAGCGGGGGGAAAAGGAGACAGACAGGCGTCCTTGCAGCAGAGGGTTTGA
AGTCGACCCACACACACCGATGAAGCCGCCTTTTGCAGCCTCTCTCTCGTCATCACCGCCCTC
CTCGAGTTCAGGCAGTTTGCCGCTTTGCCTCGTGCACACATAGGGCCCGGCATTTTTCCTGGG
CATGGAATTCTGGAATGAAGACCAGGACATCAATCTGCGTCGGGCGAGGCAACGCCGAAGA
GCTGTTGTATTCCGGACACTCGTATACTGACCCTAAACGTGTATGTATGCACAGTACAATGA
AAGATTCGAAATGGGGAAAAAGTAGCCAATGCATACGTACGTACGAAGGAACGCATTGAGC
ATTCGACACTAGTTCTGACCTTCAAAAAACCGTCATTCGAATCTGGACCTGGCTTGTCAGTTA
TGGTGTTTGACTGAGTGCCAGGTGAGTCGAGCACTAACGAAGGGAGTGCTTTAAAACCCTTC
CAGGCTGCTCCAGACACCCTGATTCTGGGGCTGCTGCAGGAATCGACACGGGGAAGAAGCA
GCATTGTTTCAGAATGTAGACATCAAGCGGGTCCGGAAAGCACATGTATGGAAAGTAAGTA
CCTCCGTACGGAGTACTGCATGTCCATCCGTACTTGAGGAAACGCTGAGGTAACATGGAGGT
AGAAGAAACCACGAGAGACTATGGGTTACACCTGCTCAAACCCACTGCACCTCTCCTCTGGG
GATTTTCCGACTCTTCCCCTGCTTGAATGCACAGACAGCTGTGTCTTTGGTACACTTTACTAA
AGACCACAGCCAAGCGGGAGAAAACGGGGACGATGAGTCACGTCCGGGAGATTCCGGCCTG
CTGTGTCGGAAGCAATCAGCTGAGCTGCTCAATCCTGAACTTTCAGTACACGGCTGCCAACT
GAGTTGCAGCGGTCCGAAGCGTTCCGAGTCCTTTGTCAGTTGCTTTTTCGCTTAGTTTATTCC
TTGGACCAGCTCTTGTCAAAGACCGAAAATGCTTTCGGGAGTAAGAGCGCTTGGGATTTGGG
GTTATGTCATAGACCGGATGAAGTCGGCCTGTGGTCCCTTCGTTTTCCTGCCTTCCCCTTGAG
GTGGCTTGCGAACAAACGTATAAATGATGTATCACTCTCCAAGACATGCCCGTACTCGCCTT
GATGAATCTCAGACTCGTGATCCATCGCGACAAGACAGTATATTGGCAGCCATCTGTCTGTT
GAAGCTTTCAACCCC (SEQ ID NO:33)

TATTCCGGACACTCGTATACTGACCCTAAACGTGTATGTATGCACAGTACAATGAAAGATTC
GAAATGGGGAAAAAGTAGCCAATGCATACGTACGTACGAAGGAACGCATTGAGCATTCGAC
ACTAGTTCTGACCTTCAAAAAACCGTCATTCGAATCTGGACCTGGCTTGTCAGTTATGGTGTT
TGACTGAGTGCCAGGTGAGTCGAGCACTAACGAAGGGAGTGCTTTAAAACCCTTCCAGGCT
GCTCCAGACACCCTGATTCTGGGGCTGCTGCAGGAATCGACACGGGGAAGAAGCAGCATTG
TTTCAGAATGTAGACATCAAGCGGGTCCGGAAAGCACATGTATGGAAAGTAAGTACCTCCG
TACGGAGTACTGCATGTCCATCCGTACTTGAGGAAACGCTGAGGTAACATGGAGGTAGAAG
AAACCACGAGAGACTATGGGTTACACCTGCTCAAACCCACTGCACCTCTCCTCTGGGGATTT
TCCGACTCTTCCCCTGCTTGAATGCACAGACAGCTGTGTCTTTGGTACACTTTACTAAAGACC
ACAGCCAAGCGGGAGAAAACGGGGACGATGAGTCACGTCCGGGAGATTCCGGCCTGCTGTG
TCGGAAGCAATCAGCTGAGCTGCTCAATCCTGAACTTTCAGTACACGGCTGCCAACTGAGTT
GCAGCGGTCCGAAGCGTTCCGAGTCCTTTGTCAGTTGCTTTTTCGCTTAGTTTATTCCTTGGA
CCAGCTCTTGTCAAAGACCGAAAATGCTTTCGGGAGTAAGAGCGCTTGGGATTTGGGGTTAT
GTCATAGACCGGATGAAGTCGGCCTGTGGTCCCTTCGTTTTCCTGCCTTCCCCTTGAGGTGGC
TTGCGAACAAACGTATAAATGATGTATCACTCTCCAAGACATGCCCGTACTCGCCTTGATGA

ATCTCAGACTCGTGATCCATCGCGACAAGACAGTATATTGGCAGCCATCTGTCTGTTGAAGC
TTTCAACCCC (SEQ ID NO:34)

CAGCATTGTTTCAGAATGTAGACATCAAGCGGGTCCGGAAAGCACATGTATGGAAAGTAAG
TACCTCCGTACGGAGTACTGCATGTCCATCCGTACTTGAGGAAACGCTGAGGTAACATGGAG
GTAGAAGAAACCACGAGAGACTATGGGTTACACCTGCTCAAACCCACTGCACCTCTCCTCTG
GGGATTTTCCGACTCTTCCCCTGCTTGAATGCACAGACAGCTGTGTCTTTGGTACACTTTACT
AAAGACCACAGCCAAGCGGGAGAAAACGGGGACGATGAGTCACGTCCGGGAGATTCCGGC
CTGCTGTGTCGGAAGCAATCAGCTGAGCTGCTCAATCCTGAACTTTCAGTACACGGCTGCCA
ACTGAGTTGCAGCGGTCCGAAGCGTTCCGAGTCCTTTGTCAGTTGCTTTTTCGCTTAGTTTAT
TCCTTGGACCAGCTCTTGTCAAAGACCGAAAATGCTTTCGGGAGTAAGAGCGCTTGGGATTT
GGGGTTATGTCATAGACCGGATGAAGTCGGCCTGTGGTCCCTTCGTTTTCCTGCCTTCCCCTT
GAGGTGGCTTGCGAACAAACGTATAAATGATGTATCACTCTCCAAGACATGCCCGTACTCGC
CTTGATGAATCTCAGACTCGTGATCCATCGCGACAAGACAGTATATTGGCAGCCATCTGTCT
GTTGAAGCTTTCAACCCC (SEQ ID NO:35)

[00127] In some embodiments, the C1Ce promoter of the invention comprises a subsequence of SEQ ID NO:33, SEQ ID NO:34, and/or SEQ ID NO:35 that retains promoter activity, referred to herein as a “short” promoter sequence. Subsequences that retain promoter activity are identified using routine methods well known to those in the art. Provided with SEQ ID NO:33, SEQ ID NO:34, and/or SEQ ID NO:35, any of a number of different functional deletion mutants of the starting sequence can be readily prepared. In some embodiments, the subsequences retain constitutive promoter activity (*e.g.,* glucose-independent promoter activity).

[00128] In some embodiments, the C1Ce promoter sequence comprises at least about 1000 contiguous nucleotides of SEQ ID NO:33 or 34; at least about 900 nucleotides of SEQ ID NO:33 or 34; at least about 800 nucleotides of SEQ ID NO:33 or 34; at least about 700 nucleotides of SEQ ID NO:33, SEQ ID NO:34, or SEQ ID NO:35; at least about 600 nucleotides of SEQ ID NO:33, SEQ ID NO:34, or SEQ ID NO:35; at least about 500 nucleotides of SEQ ID NO:33, SEQ ID NO:34, or SEQ ID NO:35; at least about 450 nucleotides of SEQ ID NO:33, SEQ ID NO:34, or SEQ ID NO:35; at least about 400 nucleotides of SEQ ID NO:33, SEQ ID NO:34, or SEQ ID NO:35; at least about 350 nucleotides of SEQ ID NO:33, SEQ ID NO:34, or SEQ ID NO:35; at least about 300 nucleotides of SEQ ID NO:33, SEQ ID NO:34, or SEQ ID NO:35; at least about 250 nucleotides SEQ ID NO:33, SEQ ID NO:34, or SEQ ID NO:35; at least about 200 nucleotides of SEQ ID NO:33, SEQ ID NO:34, or SEQ ID NO:35; at least about 150 nucleotides of SEQ ID NO:33, SEQ ID NO:34, or SEQ ID NO:35; at least about 100 nucleotides of SEQ ID NO:33, SEQ ID NO:34, or SEQ ID NO:35; or at least about 75 contiguous nucleotides of SEQ ID NO:33, SEQ ID NO:34, or SEQ ID NO:35.

[00129] In some embodiments, the C1Ce promoter sequence comprises a subsequence comprising about 75 to about 1000 contiguous nucleotides of SEQ ID NO:33 or 34. In some other embodiments, the C1Ce promoter sequence comprises a subsequence of SEQ ID NO:33, SEQ ID NO:34, or SEQ ID NO:35, comprising about 50 to about 700, about 50 to about 600, about 50 to about 500, about 50 to about 400, about 50 to about 300, about 50 to about 200, about 75 to about 700, about 75 to about 600, about 75 to about 500, about 75 to about 400, about 75 to about 300, about 75 to about 200, about 100 to about 700, about 100 to about 600, about 100 to about 500, about 100 to about 400, about 100 to about 300, or about 100 to about 200 contiguous nucleotides. In some embodiments, the subsequence comprises at least about 25, at least about 50, at least about 100, at least about 150, or at least about 200 contiguous nucleotides of SEQ ID NO:35. In some embodiments, the C1Ce promoter sequence comprises at least one of SEQ ID NOS:36-40. SEQ ID NO:36 comprises bases 871 - 1430 of SEQ ID NO:33 (0.56 kb), while SEQ ID NO:37 comprises bases 941 - 1430 of SEQ ID NO:33 (0.49 kb), SEQ ID NO:38 comprises bases 1001 - 1450 of SEQ ID NO:33(0.45 kb), SEQ ID NO:39 comprises bases 1001 - 1400 of SEQ ID NO:33 (0.40 kb), and SEQ ID NO:40 comprises bases 1201 – 1450 of SEQ ID NO:33 (0.25 kb).

ACGGAGTACTGCATGTCCATCCGTACTTGAGGAAACGCTGAGGTAACATGGAGGTAGAAGA
AACCACGAGAGACTATGGGTTACACCTGCTCAAACCCACTGCACCTCTCCTCTGGGGATTTT
CCGACTCTTCCCCTGCTTGAATGCACAGACAGCTGTGTCTTTGGTACACTTTACTAAAGACCA
CAGCCAAGCGGGAGAAAACGGGGACGATGAGTCACGTCCGGGAGATTCCGGCCTGCTGTGT
CGGAAGCAATCAGCTGAGCTGCTCAATCCTGAACTTTCAGTACACGGCTGCCAACTGAGTTG
CAGCGGTCCGAAGCGTTCCGAGTCCTTTGTCAGTTGCTTTTTCGCTTAGTTTATTCCTTGGAC
CAGCTCTTGTCAAAGACCGAAAATGCTTTCGGGAGTAAGAGCGCTTGGGATTTGGGGTTATG
TCATAGACCGGATGAAGTCGGCCTGTGGTCCCTTCGTTTTCCTGCCTTCCCCTTGAGGTGGCT
TGCGAACAAACGTATAAATGATGTATCACTCTCCAAGACATGCCCGTACTCGCCTTGATGAA
T (SEQ ID NO:36)

AGACTATGGGTTACACCTGCTCAAACCCACTGCACCTCTCCTCTGGGGATTTTCCGACTCTTC
CCCTGCTTGAATGCACAGACAGCTGTGTCTTTGGTACACTTTACTAAAGACCACAGCCAAGC
GGGAGAAAACGGGGACGATGAGTCACGTCCGGGAGATTCCGGCCTGCTGTGTCGGAAGCAA
TCAGCTGAGCTGCTCAATCCTGAACTTTCAGTACACGGCTGCCAACTGAGTTGCAGCGGTCC
GAAGCGTTCCGAGTCCTTTGTCAGTTGCTTTTTCGCTTAGTTTATTCCTTGGACCAGCTCTTGT
CAAAGACCGAAAATGCTTTCGGGAGTAAGAGCGCTTGGGATTTGGGGTTATGTCATAGACC
GGATGAAGTCGGCCTGTGGTCCCTTCGTTTTCCTGCCTTCCCCTTGAGGTGGCTTGCGAACAA
ACGTATAAATGATGTATCACTCTCCAAGACATGCCCGTACTCGCCTTGATGAAT (SEQ ID
NO:37)

TTCCCCTGCTTGAATGCACAGACAGCTGTGTCTTTGGTACACTTTACTAAAGACCACAGCCA
AGCGGGAGAAAACGGGGACGATGAGTCACGTCCGGGAGATTCCGGCCTGCTGTGTCGGAAG

CAATCAGCTGAGCTGCTCAATCCTGAACTTTCAGTACACGGCTGCCAACTGAGTTGCAGCGG
TCCGAAGCGTTCCGAGTCCTTTGTCAGTTGCTTTTTCGCTTAGTTTATTCCTTGGACCAGCTCT
TGTCAAAGACCGAAAATGCTTTCGGGAGTAAGAGCGCTTGGGATTTGGGGTTATGTCATAGA
CCGGATGAAGTCGGCCTGTGGTCCCTTCGTTTTCCTGCCTTCCCCTTGAGGTGGCTTGCGAAC
AAACGTATAAATGATGTATCACTCTCCAAGACATGCCCGTACTCGCCTTGATGAATCTCAGA
CTCGTGATCCATCG (SEQ ID NO:38)

TTCCCCTGCTTGAATGCACAGACAGCTGTGTCTTTGGTACACTTTACTAAAGACCACAGCCA
AGCGGGAGAAAACGGGGACGATGAGTCACGTCCGGGAGATTCCGGCCTGCTGTGTCGGAAG
CAATCAGCTGAGCTGCTCAATCCTGAACTTTCAGTACACGGCTGCCAACTGAGTTGCAGCGG
TCCGAAGCGTTCCGAGTCCTTTGTCAGTTGCTTTTTCGCTTAGTTTATTCCTTGGACCAGCTCT
TGTCAAAGACCGAAAATGCTTTCGGGAGTAAGAGCGCTTGGGATTTGGGGTTATGTCATAGA
CCGGATGAAGTCGGCCTGTGGTCCCTTCGTTTTCCTGCCTTCCCCTTGAGGTGGCTTGCGAAC
AAACGTATAAATGATGTATCACTCTC (SEQ ID NO:39)

GTCCTTTGTCAGTTGCTTTTTCGCTTAGTTTATTCCTTGGACCAGCTCTTGTCAAAGACCGAA
AATGCTTTCGGGAGTAAGAGCGCTTGGGATTTGGGGTTATGTCATAGACCGGATGAAGTCGG
CCTGTGGTCCCTTCGTTTTCCTGCCTTCCCCTTGAGGTGGCTTGCGAACAAACGTATAAATGA
TGTATCACTCTCCAAGACATGCCCGTACTCGCCTTGATGAATCTCAGACTCGTGATCCATCG
(SEQ ID NO:40)

[00130] In some embodiments, the present invention provides sequences comprising tandem duplication of at least two promoters, derived sub-sequences and/or “short” sequences, comprising C1Ca, C1Cb, C1Cc, C1Cd and/or C1Ce, in any combination.

Promoter Variants

[00131] It is not intended that the present invention be limited specifically to SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3; SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:33, SEQ ID NO:34, and/or SEQ ID NO:35 or subsequences thereof (*e.g.*, short promoters provided herein). It will be appreciated by a person of skill guided by this disclosure that promoter regions can tolerate considerable variation and without diminution of activity. Thus, the invention encompasses C1C promoters that (a) hybridize under stringent conditions to SEQ ID NO:1; SEQ ID NO:9; SEQ ID NO:17; SEQ ID NO:25 and/or SEQ ID NO:33; or (b) comprise a sequence substantially identical to SEQ ID NO:1; SEQ ID NO:9; SEQ ID NO:17; SEQ ID NO:25 and/or SEQ ID NO:33 and/or a subsequence thereof. Variants are further described herein. Any suitable method for introducing variation into a promoter sequence and identifying variants that retains promoter activity find use in the present invention, including but not limited to the specific methods set forth herein.

[00132] In some embodiments, promoters of the invention include sequences (*e.g.*, variant sequences) having promoter activity and with at least about 20%, at least about 25%, at least about 30%, at least about 35%, about 40%, at least about 45%, at least about 50%, at least about 55%, about 60%, at least about 65%, 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%, and/or at least about 99% sequence identity to any of the subsequences provided herein (*i.e.*, SEQ ID NOS:2-8, 10-16, 18-24, 26-32, and 34-40). In some embodiments, the variant sequences comprise about 50 to about 700, about 50 to about 600, about 50 to about 500, about 50 to about 400, about 50 to about 300, about 50 to about 200, about 75 to about 700, about 75 to about 600, about 75 to about 500, about 75 to about 400, about 75 to about 300, about 75 to about 200, about 100 to about 700, about 100 to about 600, about 100 to about 500, about 100 to about 400, about 100 to about 300, or about 100 to about 200 contiguous nucleotides of SEQ ID NOS:2-8, 10-16, 18-24, 26-32, and/or 34-40. In some embodiments the promoter comprises a sequence of at least about 100 nucleotides that differs from the corresponding subsequence of SEQ ID NOS:2-8, 10-16, 18-24, 26-32, and/or 34-40, at one or more (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 25, 30, or 40) nucleotides.

[00133] The promoters provided in the present invention can be characterized using any suitable method known in the art. For example, in some embodiments, the specific C1C promoters (*e.g.*, C1Ca, C1Cb, C1Cc, C1Cd and/or C1Ce) of the present invention are characterized by their ability to hybridize under high stringency hybridization conditions to SEQ ID NO:1; SEQ ID NO:9; SEQ ID NO:17; SEQ ID NO:25 and/or SEQ ID NO:33 or the complement of, SEQ ID NO:1; SEQ ID NO:9; SEQ ID NO:17; SEQ ID NO:25 and/or SEQ ID NO:33. In some embodiments, the C1Ca promoter of the invention can be characterized by its ability to hybridize under high stringency hybridization conditions to SEQ ID NO:1 and/or the complement of SEQ ID NO:1. In some embodiments, the C1Cb promoter of the invention can be characterized by its ability to hybridize under high stringency hybridization conditions to SEQ ID NO:9 and/or the complement of SEQ ID NO:9. In some embodiments, the C1Cc promoter of the invention can be characterized by its ability to hybridize under high stringency hybridization conditions to SEQ ID NO:17 and/or the complement of SEQ ID NO:17. In some embodiments, the C1Cd promoter of the invention can be characterized by its ability to hybridize under high stringency hybridization conditions to SEQ ID NO:25 and/or the complement of SEQ ID NO:25 . In some embodiments, the C1Ce promoter of the invention can be characterized by its ability to hybridize under high stringency hybridization conditions to SEQ ID NO:33 and/or the complement of SEQ ID NO:33. As used herein, the phrase “high stringency hybridization conditions” refers to hybridization at about 5°C to 10°C below the

T_M of the respective C1Ca promoter (*e.g.*, SEQ ID NO:1), C1Cb (*e.g.*, SEQ ID NO:9), C1Cc (*e.g.*, SEQ ID NO: 17), C1Cd (*e.g.*, SEQ ID NO 25), and/or C1Ce (*e.g.*, SEQ ID NO:33) melting temperature (TM) of the hybridized duplex sequence, followed by washing at 0.2 x SSC/0.1% SDS at 37°C for 45 minutes. The melting temperature of the nucleic acid hybrid can be calculated using methods well known to those in the art.

[00134] In some embodiments, C1Ca promoters of the invention are characterized based on alignment with SEQ ID NO:1 (*e.g.,* C1Ca promoter sequences less than 1.5 kb are aligned to the corresponding subsequence of SEQ ID NO:1.) C1Ca promoters of the invention include sequences with at least about 80%, 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%, and at least about 99% sequence identity to SEQ ID NO: 1, 2, or 3 or to short promoters described herein having promoter activity. Thus, in some embodiments, the C1Ca promoter has a sequence that has at least about 80%, 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%,and/or at least 99% sequence identity to a subsequence of SEQ ID NO:1 comprising about 75 to about 1000 contiguous nucleotides of SEQ ID NO:1 or 2, and/or a subsequence of SEQ ID NO:1, 2 or 3 comprising about 50 to about 700; about 50 to about 600, about 50 to about 500; about 50 to about 400, about 50 to about 300, about 50 to about 200, about 75 to about 700, about 75 to about 600, about 75 to about 500, about 75 to about 400, about 75 to about 300, about 75 to about 200, about 100 to about 700, about 100 to about 600, about 100 to about 500, about 100 to about 400, about 100 to about 300, or about 100 to about 200 contiguous nucleotides. For example, in some embodiments, the C1Ca promoter sequence has at least about 90% or at least about 95% sequence identity to SEQ ID NO:3, or a subsequence of at least about 100, at least about 200, at least about 300, at least about 400 or at least about 500 bases of SEQ ID NO:3. In another example, the C1Ca promoter sequence may have at least 90% or at least 95% sequence identity to SEQ ID NO:2, or a subsequence of at least about 100, at least about 200, at least about 300, at least about 400 or at least about 500 bases of SEQ ID NO:2. In some embodiments, the C1Ca promoter comprises a sequence of at least about 100 bases that differs from the corresponding subsequence of SEQ ID NO:1 at one or more (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more) bases.

[00135] In some embodiments, the C1Cb promoters of the invention can be characterized based on alignment with SEQ ID NO:9 (*i.e.*, C1Cb promoter sequences less than about 1.5 kb are aligned to the corresponding subsequence of SEQ ID NO:9.) In some embodiments, C1Cb promoters of the invention include sequences with at least about 80%, 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%, and at least about 99% sequence identity to SEQ ID NO: 9, 10, and/or 11, and/or to short promoters described herein having promoter activity. Thus, in some embodiments the C1Cb promoter has a sequence that has at least about 80%, 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%, and/or at least about 99% sequence identity to a subsequence comprising about 75 to about 1000 contiguous nucleotides of SEQ ID NO:9 or 10, or a subsequence of SEQ ID NO:9, 10 or 11 comprising about 50 to about 700, about 50 to about 600, about 50 to about 500, about 50 to about 400, about 50 to about 300, 50 to 200, about 75 to about 700, about 75 to about 600, about 75 to about 500, about 75 to about 400, about 75 to about 300, about 75 to about 200, about 100 to about 700, about 100 to about 600, about 100 to about 500, about 100 to about 400, about 100 to about 300, or about 100 to about 200 contiguous nucleotides SEQ ID NO:9, 10 and/or 11. For example, in some embodiments, the C1Cb promoter sequence comprises at least about 90%, or at least about 95% sequence identity to SEQ ID NO:11, and/or a subsequence of at least about 100, at least about 200, at least about 300, at least about 400, or at least about 500 bases of SEQ ID NO:11. In some embodiments, the C1Cb promoter sequence comprises at least about 90% or at least about 95% sequence identity to SEQ ID NO:10, or a subsequence of at least about 100, at least about 200, at least about 300, at least about 400, or at least about 500 bases of SEQ ID NO:10. In some embodiments, the C1Cb promoter comprises a sequence of at least about 100 bases that differs from the corresponding subsequence SEQ ID NO:9 at one or more (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more) bases.

[00136] In some embodiments, C1Cc promoter of the invention are characterized based on alignment with SEQ ID NO:17 (*i.e.*, C1Cc promoter sequences less than 1.5 kb are aligned to the corresponding subsequence of SEQ ID NO:17.) C1Cc promoters of the invention include sequences with at least about 80%, 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%, and/or at least about 99% sequence identity to SEQ ID NOS: 17, 18, and/or 19 and/or to short promoters described hereinabove having promoter activity. Thus, in some embodiments, the C1Cc promoters have sequence that are at least about 80%, 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%, and/or at least about 99% sequence identity to a subsequence comprising about 75 to about 1000 contiguous nucleotides of SEQ ID NO: 17 and/or 18, and/or a subsequence of SEQ ID NO:17, 18, and/or 19, comprising about 50 to about 700, about 50 to about 600, about 50 to about 500, about 50 to about 400, about 50 to about 300, about 50 to about 200, about 75 to about 700, about 75 to about 600, about 75 to about 500, about 75 to about 400, about 75 to about 300, about 75 to about 200, about 100 to about 700,

about 100 to about 600, about 100 to about 500, about 100 to about 400, about 100 to about 300, or about 100 to about 200 contiguous nucleotides. In some embodiments, the C1Cc promoter sequences comprise at least 90% or at least 95% sequence identity to SEQ ID NO:19, or a subsequence of at least about 100, at least about 200, at least about 300, at least about 400 or at least about 500 bases of SEQ ID NO:19. In some other embodiments, the C1Cc promoter sequences comprise at least about 90% or at least about 95% sequence identity to SEQ ID NO:18, or a subsequence of at least about 100, at least about 200, at least about 300, at least about 400 or at least about 500 bases of SEQ ID NO:18. In some embodiments the C1Cc promoter comprises a sequence of at least about 100 bases that differs from the corresponding subsequence SEQ ID NO:17 at one or more (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more) bases.

[00137] In some embodiments, C1Cd promoters of the present invention are characterized based on alignment with SEQ ID NO:25 (*i.e.*, C1Cd promoter sequences less than about 1.5 kb are aligned to the corresponding subsequence of SEQ ID NO:25.) C1Cd promoters of the invention include sequences with at least about 80%, 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%, and/or at least about 99% sequence identity to SEQ ID NO: 25, 26, and/or 27, and/or to short promoters described herein having promoter activity. Thus, in some embodiments, the C1Cd promoter has a sequence that has 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%, and/or at least about 99% sequence identity to a subsequence comprising about 75 to about 1000 contiguous nucleotides of SEQ ID NO:25 and/or 26, and/or a subsequence of SEQ ID NO:25, 26, and/or 27, comprising about 50 to about 700, about 50 to about 600, about 50 to about 500, about 50 to about 400, about 50 to about 300, about 50 to about 200, about 75 to about 700, about 75 to about 600, about 75 to about 500, about 75 to about 400, about 75 to about 300, about 75 to about 200, about 100 to about 700, about 100 to about 600, about 100 to about 500, about 100 to about 400, about 100 to about 300, or about 100 to about 200 contiguous nucleotides. In some embodiments, the C1Cd promoter sequences have at least about 90% or at least about 95% sequence identity to SEQ ID NO:27, and/or a subsequence of at least about 100, at least about 200, at least about 300, at least about 400, or at least about 500 bases of SEQ ID NO:27. In some other embodiments, the C1Cd promoter sequences have at least about 90% or at least about 95% sequence identity to SEQ ID NO:26, and/or a subsequence of at least about 100, at least about 200, at least about 300, at least about 400, or at least about 500 bases of SEQ ID NO:26. In some embodiments, the C1Cd promoter comprises a sequence of at least about 100 bases that differs from the corresponding subsequence SEQ ID NO:25 at one or more (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more) bases.

[00138] In some embodiments, the C1Ce promoters of the present invention are characterized based on alignment with SEQ ID NO:33 (*i.e.,* C1Ce promoter sequences less than about 1.5 kb are aligned to the corresponding subsequence of SEQ ID NO:33.) C1Ce promoters of the invention include sequences with 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%, and at least about 99% sequence identity to SEQ ID NOS:33, 34, and/or 35, and/or to short promoters described herein having promoter activity. Thus, in some embodiments, the C1Ce promoter has a sequence that has at least about 80%, 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%, and at least about 99% sequence identity to a subsequence comprising about 75 to about 1000 contiguous nucleotides of SEQ ID NOS:33 and/or 34, and/or a subsequence of SEQ ID NOS: 33, 34, and/or 35 comprising about 50 to about 700, about 50 to about 600, about 50 to about 500, about 50 to about 400, about 50 to about 300, about 50 to about 200, about 75 to about 700, about 75 to about 600, about 75 to about 500, about 75 to about 400, about 75 to about 300, about 75 to about 200, about 100 to about 700, about 100 to about 600, about 100 to about 500, about 100 to about 400, about 100 to about 300, or about 100 to about 200 contiguous nucleotides. For example, in some embodiments, the C1Ce promoter sequences have at least about 90% or at least about 95% sequence identity to SEQ ID NO:35, and/or a subsequence of at least about 100, at least about 200, at least about 300, at least about 400, or at least about 500 bases of SEQ ID NO:35. In some other embodiments, the C1Ce promoter sequences have at least about 90% or at least about 95% sequence identity to SEQ ID NO:34, and/or a subsequence of at least about 100, at least about 200, at least about 300, at least about 400, or at least about 500 bases of SEQ ID NO:34. In some embodiments, the C1Ce promoter comprises a sequence of at least about 100 bases that differs from the corresponding subsequence SEQ ID NO:33 at one or more (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more) bases.

[00139] Any of the sequences provided herein can be prepared and screened for activity utilizing standard methods known in the art. For example, any of SEQ ID NOS:1, 2, 3, 9, 10, 11, 17, 18, 19, 25, 26, 27, 33, 34 and 35, and short promoter sequences disclosed herein, as well as any of a number of different functional variant sequences can be readily prepared and screened for function. For example, mutagenesis and, optionally, directed evolution methods can be readily applied to polynucleotides such as, for example, the wild-type C1Ca promoter sequence (*e.g.*, SEQ ID NO: 1), C1Cb (*e.g.*, SEQ ID NO:9), C1Cc (*e.g.*, SEQ ID NO:17), C1Cd (*e.g.*, SEQ ID NO:25), and/or the wild-type C1Ce promoter sequence (*e.g.*, SEQ ID NO: 33). Mutagenesis may be performed in accordance with any of the techniques known in the art, including random and site-specific mutagenesis. Directed evolution can be

performed with any of the techniques known in the art to screen for improved promoter variants including shuffling. 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).

[00140] One targeted method for preparing mutagenized promoters relies upon the identification of putative regulatory elements within the target sequence by, for example, comparison with promoter sequences known to be expressed in a similar manner. Sequences that are shared are likely candidates for the binding of transcription factors and are thus likely elements which confer expression patterns. Confirmation of such putative regulatory elements can be achieved by deletion analysis of each putative regulatory region followed by functional analysis of each deletion construct by assay of a reporter gene which is functionally attached to each construct.

[00141] To produce a vector such as an expression cassette utilizing the C1C promoters of this invention for gene expression, a variety of methods well known in the art find use in obtaining the polynucleotide sequences for the promoter and the coding sequence of interest, and join the two sequences so that they are operably linked for gene expression. In some embodiments, the polypeptide

coding sequence encodes at least one detectable protein (*e.g.,* proteins of interest for production and conventional reporter proteins for routine screening for promoter activity).

[00142] In some embodiments, the protein is an or at least one enzyme, including but not limited to cellulases, hemicellulases, glucoamylases, amylases (*e.g.*, alpha amylases and/or abeta amylases), proteases (*e.g.*, acid proteases, alkali proteases, neutral proteases, pepsin, peptidases, trypsin, chymosin, or subtilisin), phytases, lipases, esterases, xylanases, reductases, oxidoreductases, laccases, cutinases, isomerases (*e.g.*, glucose isomerase or xylose isomerase), pullulanases, phenol oxidizing enzymes, starch hydrolyzing enzymes, mannanases, mannases, catalases, glucose oxidases, transferases, lyases (*e.g.*, pectate lyase or acetolactate decarboxylase), cellobiohydrolases, endoglucanases, beta-glucoidases, alpha-glucosidases, aminopeptidases, carbohydrases, carboxypeptidases, catalases, chitinases, cutinases, cyclodextrin glycosyltransferases, deoxyribonucleases, alpha-galactosidases, beta-galactosidases, glucocerebrosidases, invertases, mannosidases, mutanases, oxidases, pectinolytic enzymes, peroxidases, phospholipases, phytases, polyphenoloxidases, ribonucleases, and trans-glutaminases. In some embodiments, the enzyme is a bacterial enzyme, while in some other embodiments, the enzyme is a fungal enzyme. Indeed, it is not intended that the present invention be limited to the enzyme source, as any suitable prokaryotic and/or eukaryotic enzymes find use in the present invention. Furthermore, in some embodiments, the protein and/or enzyme is a variant of a wild type or naturally occurring enzyme. In some embodiments, the recombinant expression constructs comprise a protein-coding sequence that is an endogenous gene operably linked to a promoter of the present invention.

[00143] It is not intended that the present invention be limited to any particular enzyme family or class. However, it is intended that the promoters of the present invention will find use in driving the expression of numerous families or classes of enzymes, including, but not limited to oxidoreductases (E.C.1); transferases (E.C.2); hydrolyases (E.C.3); lyases (E.C.4); isomerases (E.C. 5) and ligases (E.C. 6). More specific, but non-limiting subgroups of oxidoreductases include dehydrogenases (*e.g.,* alcohol dehydrogenases (carbonyl reductases), xylulose reductases, aldehyde reductases, farnesol dehydrogenase, lactate dehydrogenases, arabinose dehydrogenases, glucose dehyrodgenase, fructose dehydrogenases, xylose reductases and succinate dehyrogenases), oxidases (*e.g.*, glucose oxidases, hexose oxidases, galactose oxidases and laccases), monoamine oxidases, lipoxygenases, peroxidases, aldehyde dehydrogenases, reductases, long-chain acyl-[acyl-carrier-protein] reductases, acyl-CoA dehydrogenases, ene-reductases, synthases (*e.g.*, glutamate synthases), nitrate reductases, mono and di-oxygenases, and catalases. More specific, but non-limiting subgroups of transferases include methyl, amidino, carboxyl, and phosphotransferases, transketolases, transaldolases, acyltransferases, glycosyltransferases, transaminases, transglutaminases and polymerases. More specific, but non-limiting subgroups of

hydrolases include invertases, ester hydrolases, peptidases, glycosylases, amylases, cellulases, hemicellulases, xylanases, chitinases, glucosidases, glucanases, glucoamylases, acylases, galactosidases, pullulanases, phytases, lactases, arabinosidases, nucleosidases, nitrilases, phosphatases, lipases, phospholipases, proteases, ATPases, and dehalogenases. More specific, but non-limiting subgroups of lyases include decarboxylases, aldolases, hydratases, dehydratases (*e.g.*, carbonic anhydrases), synthases (*e.g.*, isoprene, pinene and farnesene synthases), pectinases (*e.g.*, pectin lyases) and halohydrin dehydrogenases. More specific, but non-limiting subgroups of isomerases include racemases, epimerases, isomerases (*e.g.*, xylose, arabinose, ribose, glucose, galactose and mannose isomerases), tautomerases, and mutases (*e.g.* acyl transferring mutases, phosphomutases, and aminomutases. More specific, but non-limiting subgroups of ligases include ester synthases.

[00144] In some embodiments, the enzyme is a cellulase such as an endoglucanase (E.C. 3.2.1.4 also called β -1,4 endoglucanases), an exoglucanases (E.C. 3.2.1.91, also called cellobiohydrolases), or beta-glucosidase (E.C. 3.2.1.21 also called cellobiases). Additional cellulases include, but are not limited to invertases, xylanases, and GH61s. Numerous cellulases are known and described in the literature (*See e.g.*, US Pat Nos. 6,287,839 and 6,562,612; Jung *et al.*, Appl. Environ. Microbiol. 59:3032-3043 [1993]; and Lao *et al*, J. Bacteriol. 173:3397-3407 [1991]).

[00145] In some embodiments, the coding sequence encodes for a protein such as an enzyme, a therapeutic protein, a receptor protein, etc. In some further embodiments, the protein is a protease, such as a metallo, thiol, acid or serine protease (*e.g.*, subtilisin). In some additional embodiments, the coding sequence which is operably linked to any of the C1C promoters of the invention encodes a protein other than an enzyme (*e.g.*, hormones, receptors, growth factors, antigens and antibodies, including single chain antibodies and antibody heavy and light chains). In some additional embodiments, the protein coding sequences operably linked to any of the C1C promoters are chimeric or fusion proteins. For example, in some embodiments, the proteins include epitope tags (*e.g.*, c-myc, HIS_6 or maltose-binding protein) to aid in purification. It is not intended that the present invention be limited to any specific protein, protein coding sequence, and/or epitope tag.

[00146] In some embodiments, the protein comprises a conventional or commercially available reporter protein such as but not limited to beta-galatosidase (lacZ), beta-glucuronidase (GUS), fluorescent protein (GFP), luciferase, chloramphenicol, or acetyl transferase (CAT). Any protein for which expression can be measured (*e.g.*, by enzymatic, immunological or physical methods) finds use as a reporter. Although conventional reporters tend to be better suited to high throughput screening, production of any protein can be assayed by immunological methods, mass spectroscopy, etc.

Alternatively, expression can be measured at the level of transcription by assaying for production of specific RNAs.

[00147] In some embodiments, the reporter protein is a fungal enzyme, such as a fungal cellulase, such as a C1 cellulase or variant thereof. For example, in some embodiments, the promoter(s) find use in driving expression of a C1 beta-glucosidase and expression of the enzyme measured using p-nitrophenyl-beta-D-glucoside-based assay. In some embodiments, the promoters are used to drive expression of a C1 cellobiohydrolase and expression of the cellobiohydrolase measured using a 4-methylumbelliferyl betaD-lactopyranoside (MUL) assay known in the art. In some embodiments, the promoters are used to drive expression of C1 endoglucanse (EG) and expression of EG measured using a azo-CMC assay known in the art.

Signal Peptides

[00148] In some embodiments, a polypeptide product is secreted extracellularly by host cells comprising vectors comprising at least one promoter provided herein. In some embodiments, a signal sequence is used to produce a protein comprising a signal peptide at the amino-terminus of the protein. The signal sequence may be endogenous or exogenous to the host organism. If an endogenous signal peptide is used, it may or may not be naturally associated with the protein. For example, the promoter may be operably linked to a sequence encoding (i) a C1 Cbh1a signal sequence fused to a C1 Cbh1a mature protein sequence (*e.g*, the wildtype C1 Cbh1a), (ii) a C1 Bgl1 signal sequence fused to a C1 Cbh1a mature protein sequence, or (iii) an exogenous, non-C1, signal peptide fused to a a C1 Cbh1a mature protein sequence. A signal sequence not naturally associated with the coding sequence may be required when the coding sequence does not normally contain a signal peptide coding region, or may replace the natural signal peptide coding region to enhance secretion of the protein relative to secretion achieved with the natural signal peptide.

[00149] Exemplary non-C1 signal peptides which may be used in C1 or other filamentous fungi include, for illustration and not limitation, a signal sequence from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* neutral amylase, *Aspergillus niger* glucoamylase, *Rhizomucor miehei* aspartic proteinase, *Humicola insolens* cellulase, and *Humicola lanuginosa* lipase. However, any signal peptide coding region capable of directing the expressed polypeptide into the secretory pathway of a host cell of choice may be used in the present invention, and one skilled in the art is well aware of numerous signal sequences that may be used depending on the protein being produced and secreted in a host organism.

Expression Construct

[00150] The promoter sequence of the present invention and any coding sequence of interest may be operably linked in an expression construct (*e.g.*, an expression vector). A number of known methods are suitable for the purpose of ligating the two sequences, such as ligation methods based on PCR and ligation methods mediated by various ligases (*e.g.*, bacteriophase T4 ligase). The promoter used to direct expression of a heterologous sequence is optionally positioned about the same distance from the heterologous translation start site as it is from the translation start site in its natural setting. However, as is known in the art, some variation in this distance can be accommodated without loss of promoter function. Where there is a 3' or internal deletion relative to the promoter sequence, this can be accomplished by inserting a number of nucleotides approximately equal to the number deleted (*e.g.*, inserting from about 70 to about 130% of the number deleted, sometimes about 80 to about 120%, and sometimes about 90 to about 110%). It will be appreciated that the vector may comprise flanking sequences (*i.e.*, additional nucleotides) 5' to the C1C promoter sequences and 3' to the protein coding sequence.

[00151] In some embodiments, when the C1C promoter sequence of the invention is not truncated at the 3'end (for example, SEQ ID NOS:1-3,9-11,17-19,25-27,and 33-35) the promoter sequence is linked to the protein coding sequence at or close to the start codon (*e.g.*, the 5'-UTR of the heterologous gene is deleted). In another approach, all or a portion of the 5'-UTR of the heterologous gene is retained and a 3' portion of any of the C1C promoter is deleted to maintain the same spacing between upstream promoter elements and the translation start site. However, in general, it may be preferred to retain the complete promoter sequence.

[00152] In addition to the promoter, the expression construct optionally contains all the additional elements required for the expression of the heterologous sequence in host cells, such as signals required for efficient polyadenylation of the transcript, translation termination, and optionally enhancers. If genomic DNA is used as the heterologous coding sequence, introns may also be included.

[00153] The expression construct is typically contained in an expression vector that also includes a replicon that functions in bacterial and/or yeast and/or other host cells, and may contain a gene encoding a selectable marker to permit selection of microorganisms that harbor recombinant vectors. Selectable markers are well known and widely used in the art and include antibiotic resistance genes, metabolic selection markers, and the like. Exemplary selectable markers for use in filamentous fungi include amdS, argB, bar, hygB, niaD, pyrG, sC, trpC and the like.

[00154] In addition to episomal DNA based expression, the expression construct comprising any of the C1C promoter sequences (*e.g.*, SEQ ID NOS: 1-40) and a polypeptide coding sequence may be

integrated into the host genome such as by homologous recombination. In some alternative embodiments, the expression construct is randomly integrated into the host chromosome (*e.g.*, by non-homologous recombination). In some embodiments, a promoter of the invention is introduced into a plasmid harboring a DNA fragment encoding a protein sequence of interest (*e.g.*, a cellulase), for targeted integration into the genome at a desired site. Methods of targeted integration are known (*See e.g.*, Gaillardin and Ribet, Curr. Genet., 11: 369-375 [1987]).

Host Cells

[00155] The present invention also provides a recombinant cell comprising a promoter of the invention operably linked to a protein coding sequence. In some embodiments, the host cell is a eukaryotic cell, such as fungal cells, algal cells, insect cells, and plant cells. In some other embodiments the host cell is a fungal cell such as a yeast cell or a filamentous fungus cell. In some embodiments, the host cell is filamentous fungal cell, including but not limited to *Aspergillus*, *Chrysosporium*, *Corynascus*, *Fusarium*, *Humicola*, *Hypocrea*, *Myceliophthora*, *Mucor*, *Neurospora*, *Penicillium*, *Rhizomucor*, *Rhizopus*, *Talaromyces*, *Thermoascus*, *Thielavia*, *Trametes*, *Trichoderma*, or teleomorphs, or anamorphs, and synonyms or taxonomic equivalents thereof. In some embodiments, the host cell is a *Myceliophthora* species, such as *Myceliophthora thermophilia.* Indeed, cross-species use of promoters is well known (*See e.g.*, Punt *et al.*, J. Biotechnol., 17:19-33 [1991]; and Roberts *et al.,* Curr Genet., 15:177-80 [1989]).

[00156] As used herein, the term "C1" refers to *Myceliophthora thermophilia*, including a 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.

[00157] Strains that find use in the practice of the invention include both prokaryotic and eukaryotic strains obtainable 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). However it is not intended that only strains available from these collections will find use with the present invention.

[00158] Promoters of the invention may be employed in host cells (*e.g.,* a *Myceliopthera* strain such as *Myceliophthora thermophilia*) or any other suitable filamentous fungal host cells, that are genetically modified to have characteristics that improve expression (*e.g.,* improve protein secretion, protein stability or other properties desirable for expression and/or secretion of a protein). Genetic modification can be achieved by any suitable known genetic engineering techniques and/or using classical microbiological techniques, such as chemical or UV mutagenesis and subsequent selection. A combination of recombinant modification and classical selection techniques may be used to produce the organism of interest. For example, using recombinant technology, nucleic acid molecules can be introduced, deleted, inhibited or otherwise modified in a host cell. In some embodiments, a host cell is modified to reduce or eliminate expression of an endogenous gene. For example, a genetically modified host cell may have a modification introduced into the cell using homologous recombination to specifically suppress expression a targeted gene. In an alternative approach, interfering RNA, antisense, or ribozyme technology can be used to inhibit gene expression in a host cell.

[00159] Thus, in some embodiments, the host cells are 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, knock out of alp1 function results in a cell that is protease deficient, while knock out 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 enzymes is inhibited to increase production of cellulases of interest. Genetic modification can be achieved by any suitable method, including but not limited to genetic engineering techniques or using classical microbiological techniques, such as chemical or UV mutagenesis and subsequent selection. In one genetic engineering approach, homologous recombination can be used to

induce targeted gene modifications by specifically targeting a gene in vivo to suppress expression of the encoded protein. In one alternative approach, siRNA, antisense, or ribozyme technology can be used to inhibit gene expression.

[00160] Standard transformation methods find use in producing recombinant host cells harboring the expression vector of the present invention. For example, introduction of a vector into a host cell can be achieved by any suitable method, including but not limited to calcium phosphate transfection, DEAE-dextran mediated transfection, electroporation, PEG-mediated protoplast transformation, and/or other common techniques well known to those in the art.

[00161] The recombinant or engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters and selecting transformants. Culture conditions, such as temperature, pH and the like will be apparent to those skilled in the art. In addition, some culture conditions may be obtained from the American Type Culture Collection (ATCC).

Promoter Activity Assays

[00162] It will be understood that not every subsequence of, or variant of the promoter sequence (*e.g.,* SEQ ID NO:1, 9, 17, 25 or 33), will have C1 promoter activity, or constitutive C1 promoter activity. However, routine screening may be used to identify those variants and subsequences with the desired properties. For example, starting with a given sequence (*e.g.*, SEQ ID NO: 1, 9, 17, 25 or 33) fragments of different sizes can be operatively linked to a reporter protein and the ability of the fragment to drive expression can be measured. If desired, a deletion series can be made beginning at the 5' end of the sequence using exo- or endo-nucleases and the effect of truncation on promoter activity can be determined. Alternatively, random fragments of the target sequence (*e.g.*, SEQ ID NO: 1, 9, 17, 25 or 33) can be generated, cloned in front of the reporter protein coding sequence, and expression of the reporter measured. Controls, such as for differences in gene copy number, are typically employed. In addition, high-throughput methods may be used to assay promoter activity. Indeed various methods known in the art find use in the present invention (*See e.g.*, McNabb *et al.*, Eukary. Cell, 4:1539-49 [2005]; Bell *et al.*, Yeast 15:1747-59 [1999]; Bron *et al.*, Appl. Environ. Microbiol., 70:310-17 [2004]; Alper *et al.*, Proc. Nat'l. Acad. Sci. USA, 102:12678-83 [2005]).

[00163] In some embodiments, the C1 promoter activity of a polynucleotide (*i.e.*, "promoter sequence") can be determined using any suitable methods. For example, the promoter sequence is cloned into an expression vector encoding C1 beta-glucosidase (SEQ ID NO:41; *See,* FIG. 1), using any suitable method known in the art. Then, the expression construct is introduced into a C1 cell to produce a "host cell." Introduction of a vector or a DNA construct into a host cell using any suitable method known in the

art (including but not limited to calcium phosphate transfection, DEAE-dextran mediated transfection, electroporation, or other common techniques). Next, the host cell is cultured under normal conditions. Suitable culture conditions for C1 host cells are well known in the art (*See e.g.*, US Pat. Appln. Publ. Nos. 2008/0194005 and US 20030187243; WO 08/073914 and WO 01/79507, each of which is incorporated herein by reference). Wild-type or recombinant host cells can be cultured in conventional nutrient media as known in the art and described in the scientific literature. After a suitable period of incubation (*e.g.*, 24-72 hours), the culture medium (*i.e.*, broth) is removed and the β-glucosidase activity is measured. Several β-glucosidase activity assays are well known in the art. One suitable assay is the colorimetric pNPG (p-nitrophenyl-β-D-glucopyranoside)-based assay. Briefly, in a total volume of 100 μL, 20 μL clear media supernatant containing β-glucosidase enzyme is added to 4 mM pNPG (Sigma-Aldrich) solution in 50 mM sodium phosphate buffer at pH 5. The reactions are incubated at pH 5, 50°C for 1.5 hours. The reaction mixture is quenched with 100 μL of 1M sodium carbonate pH 11 solution. The absorbance of the solution is measured at 405 nm to determine the conversion of pNPG to p-nitrophenol. The release of p-nitrophenol (ε = 17,700 M-1 cm-1) is measured at 405 nm to calculate β-glucosidase activity. Detectable β-glucosidase activity is observed under high throughput screening conditions (pH 7, 50°C) (*See e.g.*, Breves *et al.*, Appl. Environmental Microbiol., 63:3902 [1997], incorporated herein by reference). Alternatively, the level of C1 β-glucosidase protein can be measured directly using well known methods. For example, immunological methods such as radioimmunoassays and/or quantitative "immunoblotting" may be used. In addition, promoter activity can be compared to expression of endogenous C1 beta-glucosidase. Promoter activity can be measured relative to the activity of wild-type promoter sequences (*e.g.*,SEQ ID NO:1, 9, 17, 25 or 33).

[00164] Constitutive (*i.e.*, glucose-independent) promoter activity can be assayed generally as above, but expression in the presence or absence of 10 g/L glucose can be compared. A promoter may be considered constitutive if the expression level in the presence of 10 g/L glucose for 1 hour differs from expression in the absence of glucose by no more than 50% of the non-glucose levels. In some embodiments, the difference is less than about 25%, while in some other embodiments, it is sometimes less than about 10%.

[00165] In some embodiments, short and/or variant C1C promoters have at least the same level of promoter activity as the activity exhibited by wild-type C1C promoter sequences (*e,g.*,SEQ ID NO:1, 9, 17, 25 or 33) under equivalent conditions (*e.g.*, expression in C1). In some embodiments, the promoter activity is greater, for example at least about 5%, at least about 10%, at least about 15%, at least about 20%, at least about 25%, at least about 30%, at least about 40%, or at least about 50%.

[00166] Host cells comprising a C1C promoter of the invention operatively linked to a sequence encoding a cellulase or other cellulose degrading enzyme may be used, for illustration and not for limitation, in saccharification of cellulosic biomass. In some embodiments, cells comprising a C1C promoter and expressing a coding sequence are grown under batch, fed-batch or continuous culture conditions, well known in the art.

[00167] In some embodiments, the cells of the invention are used to produce a recombinant protein of interest, such as a therapeutic protein. In some embodiments the protein is isolated from the cells or culture media.

EXPERIMENTAL

[00168] The following examples are provided by way of illustration only and not by way of limitation. Those of skill in the art will readily recognize a variety of non-critical parameters that could be changed or modified to yield essentially the same or similar results. 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.

[00169] 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); cm (centimeters); mm (millimeters); um and μm (micrometers); sec. (seconds); min(s) (minute(s)); h(s) and hr(s) (hour(s)); U (units); MW (molecular weight); rpm (rotations per minute); °C (degrees Centigrade); DNA (deoxyribonucleic acid); RNA (ribonucleic acid); bp (base pair); kbp (kilo base pair); CDW (cell dry weight); HPLC (high pressure liquid chromatography); ARS (ARS Culture Collection or NRRL Culture Collection, Peoria, IL); and Sigma-Aldrich (Sigma-Aldrich, St. Louis, MO).

EXAMPLE 1

Identification of Fungal Promoters from C1 Cells

[00170] Oligonucleotide microarrays were designed based on genes identified by bioinformatics analysis of the C1 genomic sequence. Each predicted gene was represented by at least four different oligonucleotides spanning the entire gene sequence.

[00171] C1 cells (wild-type and an *Alp*1-deficient strain) were grown in separate shake flasks in the presence of lactose. Cells were cultured for 3 days and then collected ("uninduced cells"), or cultured

for an additional hour in the presence of 10 g/L glucose ("induced cells"). RNA was collected from induced and uninduced cells and used to prepare cDNA probes for hybridization in the microarrays.

[00172] Twenty-two genes regulated by constitutive promoters were identified. For each, a fragment containing approximately 1.5 kb of sequence upstream from the translation start site was cloned into an expression vector so that the upstream sequence was operably linked to a sequence encoding an endogenous C1 cellulase protein. Transformants were selected and a cellulase activity assay was used to determine amount of cellulase secreted into culture medium (*See*, FIG. 2). Five promoters (IP092, IP102, IP103, IP104 and IP105) were identified as strong constitutive promoters showing cellulase overexpression relative to expression of the endogenous cellulase (designated "C1V34" in FIG. 2). The C1Ca promoter corresponds to IP104. The C1Cb promoter corresponds to IP092. The C1Cc promoter corresponds to IP102. The C1Cd promoter corresponds to IP105. The C1Ce promoter corresponds to IP103.

[00173] While the present invention has been described with reference to the specific embodiments thereof, it should be understood by those skilled in the art that various changes can be made and equivalents can be substituted without departing from the scope of the invention. In addition, many modifications can be made to adapt a particular situation, material, composition of matter, process, process step or steps, to achieve the benefits provided by the present invention without departing from the scope of the present invention. All such modifications are intended to be within the scope of the claims appended hereto. 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 this invention.

[00174] All publications and patent documents cited herein are incorporated herein by reference as if each such publication or document was specifically and individually indicated to be incorporated herein by reference. Citation of publications and patent documents is not intended as an indication that

any such document is pertinent prior art, nor does it constitute any admission as to the contents or date of the same.

CLAIMS

WHAT IS CLAIMED IS:

1. An isolated polynucleotide having promoter activity comprising

(a) a nucleotide sequence set forth in any of SEQ ID NOS:1-40;

(b) a subsequence of (a) comprising at least 75 contiguous nucleotides set forth in any of SEQ ID NOS:1-40;

(c) a nucleotide sequence having at least 90% sequence identity to (a) or (b); or

(d) a nucleotide sequence that hybridizes to any of SEQ ID NOS:1-40, and/or the complement thereof.

2. The isolated polynucleotide having promoter activity of Claim 1, comprising

(a) a nucleotide sequence set forth in any of SEQ ID NO:1, 9, 17, 25, and/or 33;

(b) a subsequence of (a) comprising at least 75 contiguous nucleotides set forth in any of SEQ ID NO:1, 9, 17, 25, and/or 33;

(c) a nucleotide sequence having at least 90% sequence identity to (a) or (b); or

(d) a nucleotide sequence that hybridizes to any of SEQ ID NO:1, 9, 17, 25, and/or 33, and/or a complement thereof.

3. The isolated polynucleotide having promoter activity of Claim 1, comprising

(a) a nucleotide sequence set forth in any of SEQ ID NO:2, 10, 18, 26, and/or 34;

(b) a subsequence of (a) comprising at least 75 contiguous nucleotides set forth in any of SEQ ID NO:2, 10, 18, 26, and/or 34;

(c) a nucleotide sequence having at least 90% sequence identity to (a) or (b); or

(d) a nucleotide sequence that hybridizes to any of SEQ ID NO:2, 10, 18, 26, and/or 34, and/or a complement thereof.

4. The isolated polynucleotide having promoter activity of Claim 1, comprising

(a) a nucleotide sequence set forth in any of SEQ ID NO:3, 11, 19, 27, and/or 35;

(b) a subsequence of (a) comprising at least 75 contiguous nucleotides set forth in any of SEQ ID NO:3, 11, 19, 27, and/or 35;

(c) a nucleotide sequence having at least 90% sequence identity to (a) or (b); or

(d) a nucleotide sequence that hybridizes to any of SEQ ID NO:3, 11, 19, 27, and/or 35, and/or a complement thereof.

5. The isolated polynucleotide having promoter activity of Claim 1, comprising

(a) a nucleotide sequence set forth in any of SEQ ID NO:4, 12, 20, 28, and/or 36;

(b) a subsequence of (a) comprising at least 75 contiguous nucleotides set forth in any of SEQ ID NO:4, 12, 20, 28, and/or 36;

(c) a nucleotide sequence having at least 90% sequence identity to (a) or (b); or

(d) a nucleotide sequence that hybridizes to any of SEQ ID NO:4, 12, 20, 28, and/or 36, and/or a complement thereof.

6. The isolated polynucleotide having promoter activity of Claim 1, comprising

(a) a nucleotide sequence set forth in any of SEQ ID NO:5, 13, 21, 29, and/or 37;

(b) a subsequence of (a) comprising at least 75 contiguous nucleotides set forth in any of SEQ ID NO:5, 13, 21, 29, and/or 37;

(c) a nucleotide sequence having at least 90% sequence identity to (a) or (b); or

(d) a nucleotide sequence that hybridizes to any of SEQ ID NO:5, 13, 21, 29, and/or 37, and/or a complement thereof.

7. The isolated polynucleotide having promoter activity of Claim 1, comprising

(a) a nucleotide sequence set forth in any of SEQ ID NO:6, 14, 22, 30, and/or 38;

(b) a subsequence of (a) comprising at least 75 contiguous nucleotides set forth in any of SEQ ID NO:6, 14, 22, 30, and/or 38;

(c) a nucleotide sequence having at least 90% sequence identity to (a) or (b); or

(d) a nucleotide sequence that hybridizes to any of SEQ ID NO:6, 14, 22, 30, and/or 38, and/or a complement thereof.

8. The isolated polynucleotide having promoter activity of Claim 1, comprising

(a) a nucleotide sequence set forth in any of SEQ ID NO:7, 15, 23, 31, and/or 39;

(b) a subsequence of (a) comprising at least 75 contiguous nucleotides set forth in any of SEQ ID NO:7, 15, 23, 31, and/or 39;

(c) a nucleotide sequence having at least 90% sequence identity to (a) or (b); or

(d) a nucleotide sequence that hybridizes to any of SEQ ID NO:7, 15, 23, 31, and/or 39, and/or a complement thereof.

9. The isolated polynucleotide having promoter activity of Claim 1, comprising

(a) a nucleotide sequence of set forth in any of SEQ ID NO:8, 16, 24, 32, and/or 40;

(b) a subsequence of (a) comprising at least 75 contiguous nucleotides set forth in any of SEQ ID NO:8, 16, 24, 32, and/or 40;

(c) a nucleotide sequence having at least 90% sequence identity to (a) or (b); or

(d) a nucleotide sequence that hybridizes to any of SEQ ID NO:8, 16, 24, 32, and/or 40, and/or a complement thereof.

10. The isolated polynucleotide of any of Claims 1 through 9, wherein said promoter activity is C1 promoter activity.

11. An expression construct comprising the isolated polynucleotide having promoter activity set forth in any of Claims 1-10, operably linked to at least one heterologous DNA sequence encoding at least one protein.

12. The expression construct of Claim 11, wherein the protein comprises a signal peptide fused to a secreted protein sequence.

13. The expression construct of Claim 12, wherein the signal peptide is not associated with the secreted protein in nature.

14. The expression construct of any of Claims 10-13, wherein the protein comprises at least one enzyme.

15. The expression construct of Claim 14, wherein the enzyme comprises at least one cellulase.

16. The expression construct of Claim 15, wherein the at least one cellulase comprises an endoglucanase, a cellobiohydrolase and/or a β-glucosidase.

17. The expression construct of Claim 14, wherein said enzyme comprises at least one glucoamylase, protease, alpha amylase, cellulase, hemicellulase, xylanase, esterase, cutinase, phytase, lipase, oxidoreductase, reductase, dehydrogenase, synthase, invertase, laccase, isomerase, pullulanase, phenol oxidizing enzyme, mannanase, mannase, catalase, glucose oxidase, transferase, and/or lyase.

18. The expression construct of any of Claims 14-17, wherein said at least one enzyme comprises at least one recombinant enzyme.

19. The expression construct set forth in any of Claims 11-18, comprising (i) a subsequence of (a) comprising at least 100 contiguous nucleotides of SEQ ID NOS:3, 11, 19, 27, and/or 35; or (ii) a nucleotide sequence having at least 90% sequence identity to (i).

20. The expression construct of Claim 19, comprising (i) a subsequence of (a) comprising at least 150 contiguous nucleotides of SEQ ID NO:3, 11, 19, 27, and/or 35; or (ii) a nucleotide sequence having at least 90% sequence identity to (i).

21. The expression construct set forth in any of Claim 19 and/or 20, comprising (i) a subsequence of (a) comprising at least 300 contiguous nucleotides of SEQ ID NO:3, 11, 19, 27, and/or 35; or (ii) a nucleotide sequence having at least 90% sequence identity to (i).

22. The expression construct set forth in any of Claims 19-21, comprising (i) a subsequence of (a) comprising at least 400 contiguous nucleotides of SEQ ID NO:3, 11, 19, 27 and/or 35; for (ii) a nucleotide sequence having at least 90% sequence identity to (i).

23. A host cell comprising the expression construct of any of Claims 11-22.

24. The host cell of Claim 23, wherein said cell is a yeast or filamentous fungal cell.

25. The host cell of Claim 23 and/or 24, wherein said cell is *Myceliophthora thermophila*.

26. The host cell of any of Claims 23-25, wherein said expression cassette is integrated into the genome of said host cell.

27. A method for producing a protein in a host cell, comprising culturing a cell of any of Claim 23-26, under conditions such that said protein is produced by said host cell.

28. The method of Claim 27, further comprising the step of isolating said protein produced by said host cell.

FIGURE 1

ATGAAGGCTGCTGCGCTTTCCTGCCTCTTCGGCAGTACCCTTGCCGTTGCAGGCGCCATTGAATCGAGAAAGGTTCACCAGAAGCCCCTC
M K A A A L S C L F G S T L A V A G A I E S R K V H Q K P L
123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
1 2 3 4 5 6 7 8 9
GCGAGATCTGAACCTTTTTACCCGTCGCCATGGATGAATCCCAACGCCGACGGCTGGGCGGAGGCCTATGCCCAGGCCAAGTCCTTTGTC
A R S E P F Y P S P W M N P N A D G W A E A Y A Q A K S F V
123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
10 11 12 13 14 15 16 17 1
TCCCAAATGACTCTGCTAGAGAAGGTCAACTTGACCACGGGAGTCGGCTGGGGGGCTGAGCAGTGCGTCGGCCAAGTGGGCGCGATCCCT
S Q M T L L E K V N L T T G V G W G A E Q C V G Q V G A I P
123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
8 19 20 21 22 23 24 25 26 2
CGCCTTGGACTTCGCAGTCTGTGCATGCATGACTCCCCTCTCGGCATCCGAGGAGCCGACTACAACTCAGCGTTCCCCTCTGGCCAGACC
R L G L R S L C M H D S P L G I R G A D Y N S A F P S G Q T
123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
7 28 29 30 31 32 33 34 35 3
GTTGCTGCTACCTGGGATCGCGGTCTGATGTACCGTCGCGGCTACGCAATGGGCCAGGAGGCCAAAGGCAAGGGCATCAATGTCCTTCTC
V A A T W D R G L M Y R R G Y A M G Q E A K G K G I N V L L
123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
6 37 38 39 40 41 42 43 44 4
GGACCAGTCGCCGGCCCCCTTGGCCGCATGCCCGAGGGCGGTCGTAACTGGGAAGGCTTCGCTCCGGATCCCGTCCTTACCGGCATCGGC
G P V A G P L G R M P E G G R N W E G F A P D P V L T G I G
123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
5 46 47 48 49 50 51 52 53 5
ATGTCCGAGACGATCAAGGGCATTCAGGATGCTGGCGTCATCGCTTGTGCGAAGCACTTTATTGGAAACGAGCAGGAGCACTTCAGACAG
M S E T I K G I Q D A G V I A C A K H F I G N E Q E H F R Q
123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
4 55 56 57 58 59 60 61 62 6
GTGCCAGAAGCCCAGGGATACGGTTACAACATCAGCGAAACCCTCTCCTCCAACATTGACGACAAGACCATGCACGAGCTCTACCTTTGG
V P E A Q G Y G Y N I S E T L S S N I D D K T M H E L Y L W
123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
3 64 65 66 67 68 69 70 71 7

```
CCGTTTGCCGATGCCGTCCGGGCCGGCGTCGGCTCTGTCATGTGCTCGTACCAGCAGGTCAACAACTCGTACGCCTGCCAGAACTCGAAG
 P  F  A  D  A  V  R  A  G  V  G  S  V  M  C  S  Y  Q  Q  V  N  N  S  Y  A  C  Q  N  S  K
123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
2        73        74        75        76        77        78        79        80        8
CTGCTGAACGACCTCCTCAAGAACGAGCTTGGGTTTCAGGGCTTCGTCATGAGCGACTGGCAGGCACAGCACACTGGCGCAGCAAGCGCC
 L  L  N  D  L  L  K  N  E  L  G  F  Q  G  F  V  M  S  D  W  Q  A  Q  H  T  G  A  A  S  A
123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
1        82        83        84        85        86        87        88        89        9
GTGGCTGGTCTCGATATGTCCATGCCGGGCGACACCCAGTTCAACACTGGCGTCAGTTTCTGGGGCGCCAATCTCACCCTCGCCGTCCTC
 V  A  G  L  D  M  S  M  P  G  D  T  Q  F  N  T  G  V  S  F  W  G  A  N  L  T  L  A  V  L
123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
0        91        92        93        94        95        96        97        98        9
AACGGCACAGTCCCTGCCTACCGTCTCGACGACATGGCCATGCGCATCATGGCCGCCCTCTTCAAGGTCACCAAGACCACCGACCTGGAA
 N  G  T  V  P  A  Y  R  L  D  D  M  A  M  R  I  M  A  A  L  F  K  V  T  K  T  T  D  L  E
123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
9        100       101       102       103       104       105       106       107       1
CCGATCAACTTCTCCTTCTGGACCGACGACACTTATGGCCCGATCCACTGGGCCGCCAAGCAGGGCTACCAGGAGATTAATTCCCACGTT
 P  I  N  F  S  F  W  T  D  D  T  Y  G  P  I  H  W  A  A  K  Q  G  Y  Q  E  I  N  S  H  V
123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
08       109       110       111       112       113       114       115       116       1
GACGTCCGCGCCGACCACGGCAACCTCATCCGGGAGATTGCCGCCAAGGGTACGGTGCTGCTGAAGAATACCGGCTCTCTACCCCTGAAC
 D  V  R  A  D  H  G  N  L  I  R  E  I  A  A  K  G  T  V  L  L  K  N  T  G  S  L  P  L  N
123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
17       118       119       120       121       122       123       124       125       1
AAGCCAAAGTTCGTGGCCGTCATCGGCGAGGATGCTGGGTCGAGCCCCAACGGGCCCAACGGCTGCAGCGACCGCGGCTGTAACGAAGGC
 K  P  K  F  V  A  V  I  G  E  D  A  G  S  S  P  N  G  P  N  G  C  S  D  R  G  C  N  E  G
123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
26       127       128       129       130       131       132       133       134       1
ACGCTCGCCATGGGCTGGGGATCCGGCACAGCCAACTATCCGTACCTCGTTTCCCCCGACGCCGCGCTCCAGGCCCGGGCCATCCAGGAC
 T  L  A  M  G  W  G  S  G  T  A  N  Y  P  Y  L  V  S  P  D  A  A  L  Q  A  R  A  I  Q  D
123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
35       136       137       138       139       140       141       142       143       1
GGCACGAGGTACGAGAGCGTCCTGTCCAACTACGCCGAGGAAAAGACAAAGGCTCTGGTCTCGCAGGCCAATGCAACCGCCATCGTCTTC
 G  T  R  Y  E  S  V  L  S  N  Y  A  E  E  K  T  K  A  L  V  S  Q  A  N  A  T  A  I  V  F
123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
44       145       146       147       148       149       150       151       152       1
GTCAATGCCGACTCAGGCGAGGGCTACATCAACGTGGACGGTAACGAGGGCGACCGTAAGAACCTGACTCTCTGGAACAACGGTGATACT
```

```
 V  N  A  D  S  G  E  G  Y  I  N  V  D  G  N  E  G  D  R  K  N  L  T  L  W  N  N  G  D  T
123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
53        154       155       156       157       158       159       160       161       1
CTGGTCAAGAACGTCTCGAGCTGGTGCAGCAACACCATCGTCGTCATCCACTCGGTCGGCCCGGTCCTCCTGACCGATTGGTACGACAAC
 L  V  K  N  V  S  S  W  C  S  N  T  I  V  V  I  H  S  V  G  P  V  L  L  T  D  W  Y  D  N
123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
62        163       164       165       166       167       168       169       170       1
CCCAACATCACGGCCATTCTCTGGGCTGGTCTTCCGGGCCAGGAGTCGGGCAACTCCATCACCGACGTGCTTTACGGCAAGGTCAACCCC
 P  N  I  T  A  I  L  W  A  G  L  P  G  Q  E  S  G  N  S  I  T  D  V  L  Y  G  K  V  N  P
123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
71        172       173       174       175       176       177       178       179       1
GCCGCCCGCTCGCCCTTCACTTGGGGCAAGACCCGCGAAAGCTATGGCGCGGACGTCCTGTACAAGCCGAATAATGGCAATGGTGCGCCC
 A  A  R  S  P  F  T  W  G  K  T  R  E  S  Y  G  A  D  V  L  Y  K  P  N  N  G  N  G  A  P
123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
80        181       182       183       184       185       186       187       188       1
CAACAGGACTTCACCGAGGGCGTCTTCATCGACTACCGCTACTTCGACAAGGTTGACGATGACTCGGTCATCTACGAGTTCGGCCACGGC
 Q  Q  D  F  T  E  G  V  F  I  D  Y  R  Y  F  D  K  V  D  D  D  S  V  I  Y  E  F  G  H  G
123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
89        190       191       192       193       194       195       196       197       1
CTGAGCTACACCACCTTCGAGTACAGCAACATCCGCGTCGTCAAGTCCAACGTCAGCGAGTACCGGCCCACGACGGGCACCACGGCCCAG
 L  S  Y  T  T  F  E  Y  S  N  I  R  V  V  K  S  N  V  S  E  Y  R  P  T  T  G  T  T  A  Q
123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
98        199       200       201       202       203       204       205       206       2
GCCCCGACGTTTGGCAACTTCTCCACCGACCTCGAGGACTATCTCTTCCCCAAGGACGAGTTCCCCTACATCTACCAGTACATCTACCCG
 A  P  T  F  G  N  F  S  T  D  L  E  D  Y  L  F  P  K  D  E  F  P  Y  I  Y  Q  Y  I  Y  P
123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
07        208       209       210       211       212       213       214       215       2
TACCTCAACACGACCGACCCCCGGAGGGCCTCGGCCGATCCCCACTACGGCCAGACCGCCGAGGAGTTCCTCCCGCCCCACGCCACCGAT
 Y  L  N  T  T  D  P  R  R  A  S  A  D  P  H  Y  G  Q  T  A  E  E  F  L  P  P  H  A  T  D
123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
16        217       218       219       220       221       222       223       224       2
GACGACCCCCAGCCGCTCCTCCGGTCCTCGGGCGGAAACTCCCCCGGCGGCAACCGCCAGCTGTACGACATTGTCTACACAATCACGGCC
 D  D  P  Q  P  L  L  R  S  S  G  G  N  S  P  G  G  N  R  Q  L  Y  D  I  V  Y  T  I  T  A
123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
25        226       227       228       229       230       231       232       233       2
GACATCACGAATACGGGCTCCGTTGTAGGCGAGGAGGTACCGCAGCTCTACGTCTCGCTGGGCGGTCCCGAGGATCCCAAGGTGCAGCTG
 D  I  T  N  T  G  S  V  V  G  E  E  V  P  Q  L  Y  V  S  L  G  G  P  E  D  P  K  V  Q  L
```

```
12345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
34        235       236       237       238       239       240       241       242       2
CGCGACTTTGACAGGATGCGGATCGAACCCGGCGAGACGAGGCAGTTCACCGGCCGCCTGACGCGCAGAGATCTGAGCAACTGGGACGTC
 R  D  F  D  R  M  R  I  E  P  G  E  T  R  Q  F  T  G  R  L  T  R  R  D  L  S  N  W  D  V
12345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
43        244       245       246       247       248       249       250       251       2
ACGGTGCAGGACTGGGTCATCAGCAGGTATCCCAAGACGGCATATGTTGGGAGGAGCAGCCGGAAGTTGGATCTCAAGATTGAGCTTCCT
 T  V  Q  D  W  V  I  S  R  Y  P  K  T  A  Y  V  G  R  S  S  R  K  L  D  L  K  I  E  L  P
12345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
52        253       254       255       256       257       258       259       260       2
tga
 *
1234
```

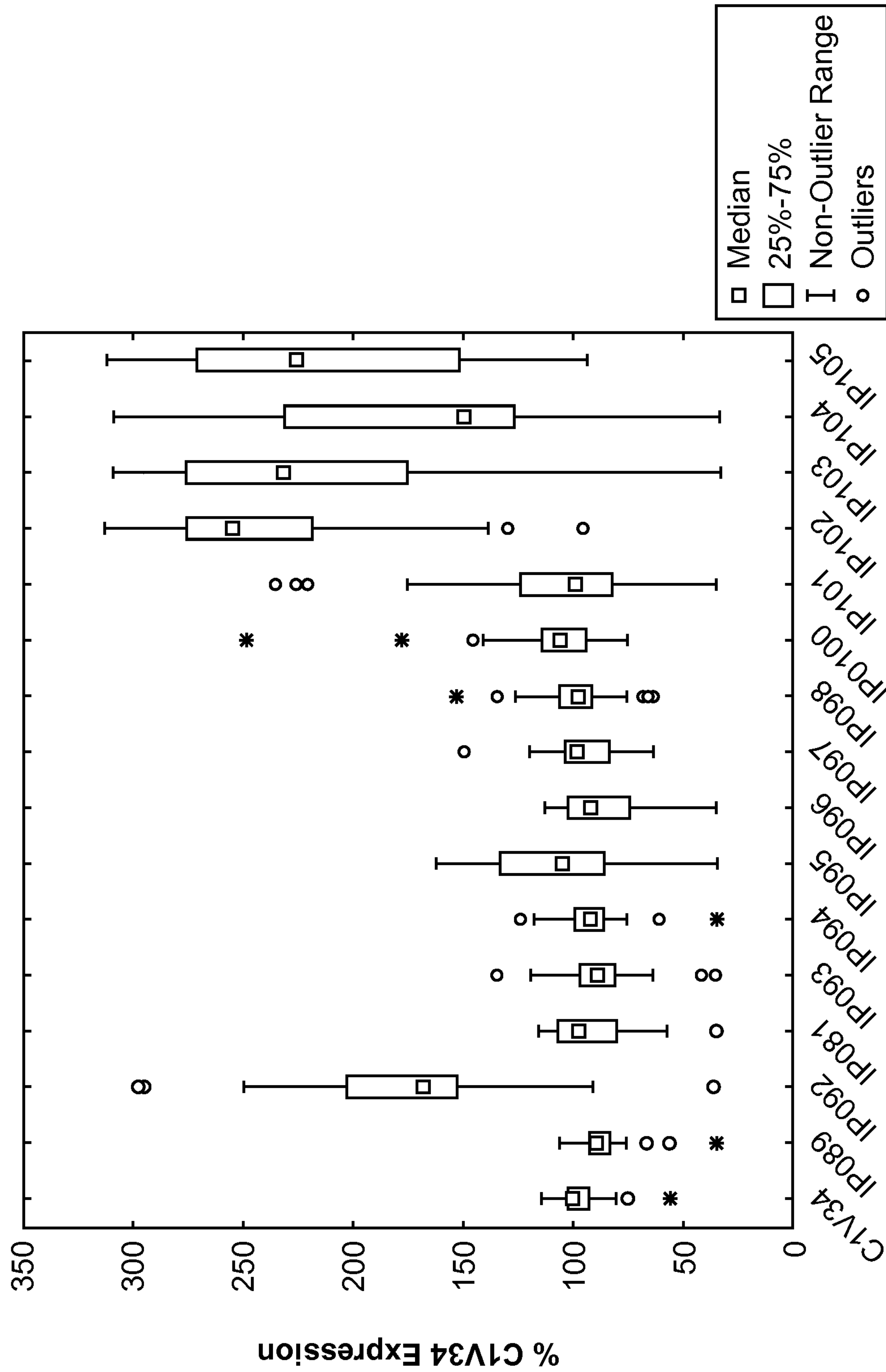

FIG. 2