

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
11 May 2006 (11.05.2006)

PCT

(10) International Publication Number
WO 2006/049777 A2

- (51) **International Patent Classification:**
C12P 21/06 (2006.01) *C12N 9/06* (2006.01)
- (21) **International Application Number:**
PCT/US2005/035325
- (22) **International Filing Date:**
30 September 2005 (30.09.2005)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
60/614,874 30 September 2004 (30.09.2004) US
- (71) **Applicants (for all designated States except US):**
TRUSTEES OF DARTMOUTH COLLEGE [US/US];
Technology Transfer Office, 11 Rope Ferry Road, Hanover,
NH 03755-1404 (US). **REGENTS OF THE UNIVERSITY OF MINNESOTA** [US/US]; Office of Patent and
Technology Marketing, 450 University Gateway, 200 Oak
Street S.E., Minneapolis, MN 55455-2070 (US).
- (72) **Inventors; and**
- (75) **Inventors/Applicants (for US only):** **DUNLAP, Jay, C.**

[US/US]; 545 Academy Road, Thetford, VT 05074 (US).
LOROS, Jennifer, J. [US/US]; 545 Academy Road, Thetford, VT 05074 (US). **MEHRA, Arun** [CA/US]; 22 East Wheelock, Unit B, Hanover, NH 03755 (US). **GOOCH, Van, D.** [US/US]; 208 West 5th Street, Morris, MN 56267 (US).

(74) **Agents:** **LICATA, Jane, Massey** et al.; Licata & Tyrell P.C., 66 E. Main Street, Marlton, NJ 08053 (US).

(81) **Designated States (unless otherwise indicated, for every kind of national protection available):** AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) **Designated States (unless otherwise indicated, for every kind of regional protection available):** ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM,

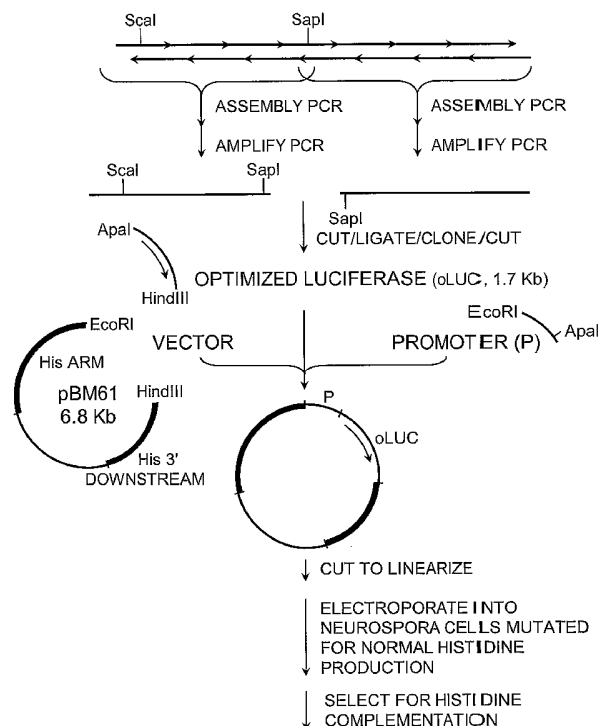
[Continued on next page]

(54) **Title:** NUCLEIC ACID SEQUENCES ENCODING LUCIFERASE FOR EXPRESSION IN FILAMENTOUS FUNGI

OPTIMIZE FIREFLY LUCIFERASE SEQUENCE USING CODON BIAS TABLE
(415 CHANGES IN 1619 NUCLEOTIDES => 26% CHANGE)

OBTAIN 86 OLIGONUCLEOTIDE OF 40 bp EACH AND 20 bp OVERLAP

GENERATE THE MOLECULE IN TWO UNITS WITH OVERLAP AT SapI SITE.



(57) **Abstract:** The present invention relates to genetic reporters. Specifically, the present invention is directed to a modified gene encoding a luciferase for high level expression in an organism with a bias for cytosine (C) or guanine (G) in the third position of the codon.



ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Published:

- *without international search report and to be republished upon receipt of that report*

**NUCLEIC ACID SEQUENCES ENCODING LUCIFERASE FOR EXPRESSION
IN FILAMENTOUS FUNGI**

5 Introduction

This application claims the benefit of priority from U.S. Provisional Patent Application Serial No. 60/614,874, filed September 30, 2004, the content of which is incorporated herein by reference in its entirety.

10 This invention was made in the course of research sponsored by the National Institute of General Medical Sciences (Grant No. R37 GM34985) and the National Institute of Mental Health (Grant No. R01 MH44651). The U.S. government may have certain rights in this invention.

15

Background of the Invention

Bioluminescence is the light produced in certain organisms as a result of luciferase-mediated oxidation reactions. The luciferase genes, e.g., the genes from
20 luminous beetle and, in particular, the luciferase from *Photinus pyralis* (the common firefly of North America), are very useful luminescent reporter genes. See, e.g., Bronstein, et al. ((1994) *Cal. Biochem.* 219:19-181) for a review of luminescent reporter gene assays.

25 The gene encoding luciferase was cloned from *Photinus pyralis*, and demonstrated to produce active enzyme in *E. coli* (de Wet, et al. (1987) *Mol. Cell Biol.* 7:725). Firefly luciferase has become highly valuable as a genetic reporter due to the convenience, sensitivity and linear range of the
30 luminescence assay. Luciferase is used in virtually every experimental biological system, including prokaryotic and eukaryotic cell culture, transgenic plants and animals, and cell-free expression systems.

Engineering of luciferases has been reported. Sung and
35 Hang ((1998) *Photochem. Photobiol.* 68(5):749-53) disclose that the N-terminal amino acid sequences of the firefly

-2-

luciferase are important for the stability of the enzyme. Branchini et al. ((1998) *Biochemistry* 37(44):15311-9) performed site-directed mutagenesis of His²⁴⁵ in firefly luciferase. White et al. ((1996) *Biochem. J.* 319(Pt 2):343-50) showed an improvement in the thermostability of the North American firefly luciferase by saturation mutagenesis at residue 354. Kajiyama and Nakano ((1994) *Biosci. Biotechnol. Biochem.* 58(6):1170-1) constructed firefly luciferase mutants from *Luciola lateralis* in which Ala at position 217 was replaced by each of three hydrophobic amino acid residues (Ile, Leu, and Val). U.S. Pat. No. 5,401,629 describes assay methods and compositions useful for measuring the transduction of an intracellular signal. U.S. Patent No. 5,670,356 teaches modified luciferase gene sequences with the peroxisomal translocation sequence removed to yield a cytoplasmic form of the enzyme interfering restriction sites and genetic regulatory sites from the gene removed, and improved codon usage for mammalian cells.

Despite its utility as a reporter, however, native luciferase is not necessarily optimized for the wide variety of host organisms, e.g., filamentous fungi, in which it can be used. The present invention meets this need in providing nucleic acid sequences encoding luciferase for high levels of expression in filamentous fungi.

Summary of the Invention

The present invention relates to a variant nucleic acid molecule encoding a luciferase for expression in an organism with a bias for cytosine (C) or guanine (G) in the third position of the codon, wherein the nucleic acid molecule has a G and C content of at least 50%. In particular embodiments, the nucleic acid molecule further contains at least one intron. Vectors, host cells, and kits

-3-

containing the variant nucleic acid molecule are also provided.

The present invention also relates to a method for enhancing the level of luciferase enzyme in an organism with a bias for C or G in the third position of the codon. The method involves introducing into the organism a nucleic acid molecule encoding a luciferase having a G and C content of at least 50% so that the level of luciferase enzyme produced in the organism is enhanced.

10

Brief Description of the Drawings

Figure 1 illustrates the protocol for generating a novel G/C-rich nucleic acid molecule encoding luciferase which provides high levels of luciferase expression in organisms with a bias for cytosine (C) or guanine (G) in the third position of the codon.

Figure 2A-2C is an alignment of native *Photinus pyralis* nucleic acid sequences encoding luciferase and optimized nucleic acid sequences encoding luciferase which provide high levels of luciferase expression in organisms with a bias for C or G in the third position of the codon. "*" denotes nucleotides which are identical between the coding sequences.

Detailed Description of the Invention

A novel nucleic acid molecule encoding a luciferase enzyme has now been generated which has a higher level of *in vivo* expression in organisms with a bias for cytosine (C) or guanine (G) in the third position, and for some amino acids C in the first position, of the codon than has been achievable with native firefly luciferase reporter genes. This improved expression results in greater reliability and convenience as a genetic reporter in various microbial systems including filamentous fungi

(e.g., *Neurospora*, *Aspergillus*, *Phytophthora*, etc.), Gram-positive bacteria (e.g., *Arthrobacter*, *Streptomyces*, *Mycobacterium*, *Pseudomonas* and the like), trypanosomes, and select vertebrates. Such applications include the
5 detection, localization and measurement of filamentous fungi; measurement of protein expression, intracellular signaling and other turnover reactions in cells or fluids; DNA and RNA binding assays; and immunoassay and other protein assays.

10 To improve the general suitability of luciferase as a genetic reporter, a variant of the common *Photinus pyralis* luciferase gene has been developed. For reference, the nucleic acid and amino acid sequence of native *Photinus pyralis* luciferase is presented herein as SEQ ID NO:2 and
15 SEQ ID NO:3, respectively. For the purposes of the present invention, the terms modified luciferase, variant luciferase, or optimized luciferase refer to a nucleic acid molecule encoding luciferase, wherein the nucleic acid molecule has been modified from its wild-type or native
20 counterpart so that increased levels of expression in organisms with a bias for G or C in the third position of the codon is achieved. Desirably, the variant luciferase has a G + C content of at least 50%, 60% or more and encodes for a functional luciferase protein. In particular
25 embodiments, the C content of the variant luciferase is in the range of about 30% to 50%, and most desirably about 40%. However, it should be understood that the modified, variant, or optimized luciferase encodes a protein with an amino acid sequence identical to functional wild-type
30 luciferase (i.e. SEQ ID NO:3) having an apparent molecular weight of 62 kilodaltons (kD) and emitting light in the presence of ATP, Mg^{2+} , molecular oxygen, and luciferin.

Codon usage generally refers to a bias toward preferred codons that generally correspond to the most

abundant tRNA species (Kemura, et al., eds. (1992) In: *Transfer RNA in Protein Synthesis*, CRC, Boca Raton, FL). A positive correlation exists between codon usage bias and the level of gene expression (Gouy and Gautier (1982) *Nucleic Acids Res.* 10:7055-7074; Sharp and Li (1986) *J. Mol. Evol.* 24:28-38). This bias provides optimal translational efficiency, and is most pronounced in highly expressed genes in species whose effective population size is large (Bulmer (1991) *Genetics* 129:897-907; Li (1987) *J. Mol. Evol.* 24:337-345). As used in the context of the instant invention, an organism with a bias for G or C in the third position is intended to mean that of all the possible codons available for a particular amino acid, the organism efficiently translates mRNAs with amino acid codons having a G or C in the third position (i.e., Ala, Arg, Asn, Asp, Cys, Glu, Gln, Gly, His, Ile, Leu, Lys, Phe, Pro, Ser, Thr, Tyr and Val). As will be appreciated, a bias for G or C in the third position of the codon is generally accompanied by a bias for C in the first position (i.e., Arg and Leu). The inventive nucleic acid molecule was generated by making base substitutions in the native *Photinus pyralis* luciferase gene in accord with the codon usage bias for organisms with a bias for C or G in the third position of the codon, or more particularly a bias for C in the third position of the codon, to improve the expression level of luciferase in such organisms, without altering the amino acid sequence of the resulting enzyme. Using *Neurospora crassa* as an exemplary organism with a bias for C or G in the third position of the codon, the codon bias of native luciferase was modified based on the frequency of C or G in the first and third position of the codon of *Neurospora crassa* genes. Unlike *Photinus pyralis* which generally has a bias for adenine (A) or uracil (U), filamentous fungi generally have a propensity for cytosine

C or guanine G (most notably C) in the third position of the codon and C in the first position, while codons containing A or U are less often used. (see Table 1).

TABLE 1

Residue	Codon	N.c.	A.f.	P.i.	A.a.	M.g.	F.o.	P.p.
Arg	CGA	7.0	9.4	3.4	8.8	8.2	13.0	6.4
	CGC	17.7	15.4	16.0	11.7	18.9	13.7	4.3
	CGG	8.5	10.8	1.8	5.8	8.5	3.9	2.1
	CGU	8.9	11.1	13.8	10.0	8.9	9.0	4.3
	AGA	7.9	7.6	1.6	7.0	5.7	7.0	17.0
	AGG	11.8	6.0	1.6	6.2	10.8	3.8	0.0
Leu	CUA	5.9	6.6	5.7	9.4	5.3	5.8	8.5
	CUC	26.9	23.2	18.9	25.5	25.9	27.4	9.6
	CUG	18.2	24.5	33.5	15.7	24.7	13.6	14.9
	CUU	14.2	15.8	10.2	19.5	13.5	20.8	15.2
	UUA	2.7	4.2	1.1	3.8	2.8	3.4	21.6
	UUG	14.9	15.8	9.2	12.4	13.2	9.6	22.3
Ser	UCA	9.2	10.2	4.9	11.4	10.0	9.8	3.5
	UCC	20.1	18.2	11.7	17.7	15.2	13.4	11.3
	UCG	14.5	14.9	26.6	12.1	18.2	6.4	7.8
	UCU	12.0	13.8	7.3	18.3	11.0	18.3	16.7
	AGC	17.4	15.7	16.3	15.5	19.3	15.2	3.5
	AGU	8.6	9.6	6.0	6.8	6.2	7.9	12.0
Thr	ACA	10.7	12.0	8.5	14.0	10.6	12.0	17.7
	ACC	24.9	20.6	27.8	22.2	23.5	23.8	13.5
	ACG	13.6	11.9	30.9	11.1	15.1	7.3	10.6
	ACU	11.2	14.0	13.1	15.4	10.5	18.2	12.8
Pro	CCA	12.3	11.4	7.9	10.5	11.4	11.9	13.1
	CCC	22.5	17.4	15.4	14.1	18.5	18.2	12.0
	CCG	14.5	12.6	20.8	10.4	15.3	5.8	12.0
	CCU	15.1	16.6	12.8	14.4	11.5	23.4	15.9
Ala	GCA	12.5	15.1	12.3	17.6	14.9	15.5	14.9
	GCC	36.2	28.5	38.3	33.1	37.5	28.9	19.8
	GCG	17.3	17.9	20.0	13.8	16.3	10.2	19.8
	GCU	21.2	22.7	32.5	30.8	17.3	31.9	17.7
Gly	GGA	13.6	14.4	10.9	16.5	13.0	15.5	31.5
	GGC	29.2	25.1	37.8	25.6	34.0	26.9	18.8
	GGG	10.9	10.2	3.4	8.0	9.4	4.4	16.3
	GGU	18.4	19.1	24.0	21.2	16.9	24.2	17.7
Val	GUA	5.3	5.6	3.5	8.8	5.7	5.5	17.0
	GUC	24.9	25.4	19.7	28.9	28.3	27.4	17.7
	GUG	15.4	16.5	38.7	9.8	16.4	7.3	18.1
	GUU	13.9	14.8	7.6	19.9	13.6	19.1	31.5
Lys	AAA	11.5	14.5	6.4	12.8	11.8	14.1	49.6
	AAG	40.4	34.4	44.2	39.9	35.5	46.3	23.4
Asn	AAC	27.1	23.4	30.6	25.2	30.8	34.9	15.6
	AAU	10.3	14.3	7.6	10.5	9.4	11.3	19.5
Gln	CAA	16.9	14.6	6.6	15.0	15.0	16.2	15.9
	CAG	26.1	25.0	25.1	22.7	26.7	24.2	13.5

-7-

His	CAC	14.8	12.8	17.3	13.3	15.8	13.4	11.3
	CAU	9.4	11.3	3.8	9.1	7.2	11.1	14.5
Glu	GAA	22.3	23.5	13.0	21.6	17.1	22.1	41.1
	GAG	42.7	36.8	51.2	37.7	40.1	35.3	17.7
Asp	GAC	32.5	29.6	41.7	36.8	35.9	26.9	23.7
	GAU	23.9	26.3	11.8	21.5	19.5	28.1	35.1
Tyr	UAC	17.4	18.1	30.8	17.5	20.7	23.5	17.7
	UAU	8.4	11.5	5.6	8.6	7.3	11.4	17.0
Cys	UGC	7.7	7.7	11.3	12.1	9.8	8.6	2.8
	UGU	3.3	4.7	4.6	7.1	3.4	5.5	3.9
Phe	UUC	22.2	26.1	28.7	26.3	23.5	26.6	19.8
	UUU	11.7	13.1	6.8	11.6	14.4	10.5	30.1
Ile	AUA	4.0	4.6	1.2	8.8	6.1	3.5	16.3
	AUC	26.6	27.9	27.0	27.0	28.7	29.7	16.7
	AUU	14.0	17.4	13.0	15.6	15.3	17.0	33.7
Met	AUG	21.7	21.2	21.6	19.8	24.4	25.3	22.7
Trp	UGG	13.1	14.5	11.9	13.4	13.3	13.6	6.4
Ter	UAA	0.6	0.5	1.6	1.2	0.8	0.7	1.4
	UAG	0.5	0.6	0.7	0.6	0.4	0.6	0.0
	UGA	0.8	0.8	0.4	0.6	0.7	0.7	0.7

-Numbers are presented as the frequency per thousand.

-Data obtained from the codon page of kazusa.or.jp of the world-wide web.

- 5 -N.c., *Neurospora crassa*; A.f., *Aspergillus fumigatus*; P.i., *Phytophthora infestans*; A. a., *Alternaria alternata*; M. g., *Magnaporthe grisea*; F. o., *Fusarium oxysporum*; P. p., *Photinus pyralis*.

The luciferase variant of the invention was generated using assembly PCR in accordance with standard methods, reagents, and PCR conditions (Stemmer, et al. (1995) Gene 164(1):49-53) using primers specific for the gene encoding native *Photinus pyralis* luciferase and a GC-Rich Taq such as FASTSTART™ Taq DNA Polymerase (Roche Molecular Biochemicals) or TAKARA LA TAQ™ (Takara Bio Inc.). See Figure 1. As will be appreciated by the skilled artisan, however, any well-known method for altering genetic sequences can be used, e.g., oligonucleotide-mediated, site-specific mutagenesis directed against single-stranded plasmid templates (Lewis and Thompson (1990) *Nuc. Acids Res.* 18:3439-3443), and need not be further elaborated herein.

Figure 2 shows the substitutions made (i.e., approximately 415 changes in 378 codons) in the native

luciferase (540 codons) to generate the variant luciferase with optimized codon usage bias for improved expression in organisms with a bias for C or G in the third position of the codon. In general, the codons selected were as follows:

5 Arg, CGC; Leu, CUC; Ser, UCC; Thr, ACC; Pro, CCC; Ala, GCC; Gly, GGC; Val, GUC; Lys, AAG; Asn, AAC; Gln, CAG; His, CAC; Glu, GAG; Asp, GAC; Tyr, UAC; Cys, UGC; Phe, UUC; and Ile, AUC. In cases where the first codon position could be C versus A or U (*i.e.*, Arg and Leu), C was selected. Further,

10 U was selected over A for the first codon position of Serine. The resulting overall base composition of the native versus variant luciferase is listed in Table 2.

TABLE 2

Base	Native		Variant	
	Occurrences	Percentage	Occurrences	Percentage
A	479	29.0	313	18.9
C	346	20.9	705	42.6
G	394	23.8	372	22.5
T	434	26.3	263	15.9

15 As will be appreciated by one of skill in the art, all or a portion of the 387 codons modified as disclosed herein can be modified to create a variant luciferase. In other words, instead of the 415 codon modifications, a subset, *e.g.*, 380 or 390 of the codon modifications, can be made to

20 provide a variant luciferase with enhanced expression in an organism with a bias for C or G in the third position of the codon. In certain embodiments at least 60%, 65%, or 70%, (*i.e.*, 324, 351, or 378 codons) of the wild-type 540 codons are modified to create a variant luciferase with

25 enhanced expression in an organism with a bias for C or G in the third position of the codon. The nucleotide sequence of a variant luciferase with optimized expression in filamentous fungi is provided as SEQ ID NO:1.

-9-

To illustrate the utility of the inventive luciferase nucleic acid molecule, *ccg-2* (GenBank accession no. (CAA47754.1) and *frq* (Aronson, et al. (1994) *Science* 263:1578-1584) promoters were independently cloned upstream of the variant luciferase nucleic acid molecule. The resulting fusions were inserted into suitable expression vectors and subsequently transformed into *Neurospora crassa* in accordance with standard methods. Unexpectedly, when the variant luciferase nucleic acid molecule were fused to the *ccg-2* promoter, light emission from transformed cells could be detected with the naked eye in a darkened room. Transformed cells harboring the *ccg-2*- and *frq*-luciferase variant nucleic acid molecule were found to express levels of luciferase which were observable using either Roper or Hamamatsu cameras; a Roper camera has a 50X to 100X sensitivity, whereas a Hamamatsu camera has a 1X sensitivity. Under similar assay conditions, cells expressing native luciferase could not be visualized by available camera systems (standard scintillation counter, Turner luminometer, or Packard Top Count luminescence detector) and an N-terminally modified luciferase (Morgan, et al. (2003) *Fungal Genet. Biol.* 38(3):327-32) could be detected only when the strong *ccg-2* promoter was used to drive luciferase and only when a Roper camera or similar camera was used.

Introduction of an intron into a gene has been found to enhance gene expression, stabilize the transcribed mRNA, as well as enhance the rate of mRNA export. To broaden the applicability and utility of the optimized luciferase, the first intron of the *ccg-2* gene, a highly expressed *Neurospora crassa* transcript, was introduced into the nucleic acid molecule encoding the optimized luciferase in a position analogous to that of the intron found in the wild-type firefly gene first intron (i.e., at approximately

-10-

nucleotide 127 of the coding sequence). The nucleic acid sequence encoding this intron containing transcript is provided herein as SEQ ID NO:4. The intron-containing, variant luciferase transcript was appropriately spliced by *Neurospora*, yielding a fully functional luciferase. Accordingly, in particular embodiments of the instant invention, a nucleic acid molecule encoding a variant luciferase further contains at least one intron. In certain embodiments, the intron is located at, or within 50 nucleotides, of the position of the intron in the wild-type firefly luciferase gene (i.e., position 127 of the coding sequence). Desirably, the intron is of a suitable length and contains suitable recognition sequences for efficient splicing in the organism being employed. The intron can be derived from the organism in which use is intended or can be from another organism so long as the organism in which use is intended can recognize and splice the intron from the coding sequence of the variant luciferase. It is also contemplated that at least two, three, or more introns can be introduced into the variant luciferase nucleic acid molecule of the instant invention. Intron and intron recognition sequences for genes of the organisms disclosed herein are well-known to those of skill in the art and can be readily identified using established gene prediction programs such as ORF Finder at NCBI, Gene Finder, Generation, GenScan, Glimmer, and the like available on the world-wide web. The nucleic acid sequence of an exemplary intron is disclosed herein as SEQ ID NO:5.

Assays employing the disclosed variant luciferase indicated that an increase in luciferin resulted in a corresponding increase in light emission. While, luciferin concentrations were not found to be saturating, it is contemplated that at very high levels, luciferin may become toxic. In addition, sprayed luciferin effectively generated

-11-

light. Moreover, in this fungal system, the luciferin-luciferase reaction was strongly oxygen sensitive and luciferin was stable, lasting days to weeks.

Individual *Neurospora* hyphal cells expressing *ccg-2*-variant luciferase emitted light, and actively growing cells (e.g., those in the growth front of a race tube) and cells in high glucose growth conditions (e.g., 3.0% glucose) produced the most light. Further, high levels of variant luciferase expression were not found to be dependent on any particular strain or clone of *Neurospora*.

With regard to promoter-specific regulation of variant luciferase, the *ccg-2* promoter revealed the following expression patterns: interband area darkening and conidial band areas brightened and left a lasting light in race tubes experiments; circadian behavior at the growth front disappeared under conditions where conidial bands do not form; and no rhythm-to-a slight rhythm was observed in lasting bands. *frq* promoter revealed the following expression patterns: strong peak-to-trough oscillations of a *Neurospora* colony can be seen on a sorbose plate, wherein the oscillation could persist for days; bands in a race tube continued to show light with no apparent diminishment in activity; and oscillatory activity was demonstrated in liquid media.

Accordingly, a variant luciferase as disclosed herein can enhance the level of luciferase produced in organisms (e.g., a filamentous fungus such as *Neurospora*) with a bias for C or G in the third position of the codon. As used herein, the level of luciferase is said to be enhanced when there is a 10- to 100-fold higher level of luciferase present in cells expressing the variant luciferase as compared to similar cells expressing a luciferase which has not been modified for expression in organisms with a bias for C or G in the third position of the codon (e.g., native

-12-

luciferase, or a luciferase gene with modifications in the first 21 amino acids (Morgan, et al. (2003) *supra*)). In general, organisms in which the variant luciferase finds use include those organisms with a genomic G/C content of greater than 50%, 60%, 70% or more. Organisms, and host cells derived therefrom, generally having a bias for C or G in the third position of the codon, and typically also a bias for C in the first position, include, but are not limited to, filamentous fungi (e.g., *Neurospora*, *Aspergillus*, *Phytophthora*, *Alternaria*, *Magnaporthe*, etc.), Gram-positive bacteria (e.g., *Arthrobacter*, *Streptomyces*, *Mycobacterium*, *Pseudomonas*, and the like), trypanosomes, and vertebrates such as chicken, bovine, and human, which have been shown to utilize either G or C at the third position of the codon (Marin, et al. (1989) *Nucl. Acid Res.* 17:6181-6189).

To further expand the utility of the luciferase encoded by the variant nucleic acid molecule disclosed herein, it is contemplated that additional modifications and changes can be made to the nucleic acid molecule while still retaining a luciferase enzyme having the desired characteristic of high levels of expression in organisms with a bias for C or G in the third position of the codon. For example, certain amino acids can be substituted for other amino acids in the protein structure without appreciable loss of bioluminescence. As will be appreciated by one of skill in the art, such changes create luciferase enzymes which are functionally equivalent (i.e., emit light in the presence of the appropriate substrates) but have altered protein thermostability (see, e.g., U.S. Patent No. 6,265,177), enzyme efficiency, binding affinity of substrates (see, e.g., U.S. Patent No. 6,265,177), or emission wavelength (see, e.g., U.S. Patent Nos. 6,265,177; 6,495,355; and 5,330,906). So long as a mutation or change

-13-

maintains the suitable codon bias of organisms with a bias for C or G in the third position of the codon, the resultant protein is considered functionally equivalent for the purposes of the invention. Accordingly, in particular
5 embodiments, the nucleic acid molecule of the instant invention encodes a luciferase of SEQ ID NO:3 or a functional equivalent thereof.

The polypeptide sequence can be further modified, either chemically or by genetic engineering, to enable the
10 luciferase to be targeted into a specific subcellular compartment. For example, a suitable sequence at the N-terminus will locate the bioluminescent protein in the mitochondria, while others will target the protein to the endoplasmic reticulum, optionally with a sequence at the C-
15 terminus to retain it there. Genetically engineering the gene such that the protein contains a signal peptide can locate the protein to the inner or outer surface of the plasma membrane or within a particular intracellular organelle (e.g., peroxisome, mitochondrion, endoplasmic
20 reticulum, golgi, secretory vesicle, nucleus or endosome).

The functional luciferase encoded by the inventive variant luciferase nucleic acid molecule is useful in a wide variety of assays including fusion proteins, gene expression studies (i.e., reporter gene) or screening
25 purposes. Therefore, the present invention further relates to vectors and host cells expressing the variant luciferase nucleic acid molecule.

Insertion of the variant luciferase nucleic acid molecule into a vector can be carried out using standard
30 molecular biology methods (e.g., PCR or restriction enzyme digestion and ligation). Advantageously, the modification of native luciferase nucleic acid molecule resulted in the elimination of *AflIII*, *ApoI*, *ClaI*, *EcoRI*, *EcoRV*, *MluI*, *PacI*, *SphI*, and *XbaI* from the native luciferase sequence.

-14-

Therefore, these enzymes can be added to the repertoire of enzymes useful for cloning the variant luciferase nucleic acid molecule into vectors as disclosed herein.

As used herein, a vector includes a plasmid for
5 amplifying, and not necessarily expressing, the luciferase variant nucleic acid molecule. Such vectors include, but are not limited to, pBluescript, pSP72, and pGEM vectors. A vector also includes an expression vector which is any type of genetic construct containing a luciferase variant
10 nucleic acid molecule in which the nucleic acid molecule is capable of being transcribed and translated in a host cell. An expression vector of the invention should also direct translation into luciferase protein. Expression vectors will generally include restriction enzyme cleavage sites
15 and other initial, terminal, and intermediate DNA sequences that are usually employed in vectors to facilitate their construction and use. Suitable expression vectors will be dependent on the host cell being transformed and are well-known to the skilled artisan.

20 Expression vectors for use in accordance with the present invention ordinarily include an origin of replication (as necessary), a selectable marker and a promoter operably linked to the gene(s) to be expressed. A polyadenylation site and transcriptional terminator
25 sequences are desirably included on genes to be expressed in target cells. Ribosome binding sites, internal ribosome entry sites (IRES) and RNA splice sites can also be included. A vector can be integrated into the host cell chromosome or be maintained as an episome (e.g., in
30 bacteria having a bias for G or C at the third position of the codon). Vectors will usually include an origin of replication functional in bacteria and a typical antibiotic resistance gene, allowing propagation and selection in transformed bacterial cells.

-15-

Specific initiation signals can also be included for efficient translation. These signals include the ATG initiation codon and adjacent sequences. Translational control signals within or outside of the luciferase coding sequence, including the ATG initiation codon, can additionally be provided in native or modified forms. One of ordinary skill in the art would readily be capable of determining this and providing the necessary signals. These exogenous translational control signals and initiation codons can be of a variety of origins, both natural and synthetic. The efficiency of expression is enhanced by the inclusion of appropriate transcription elements and transcription terminators. An appropriate polyadenylation site can also be included. Typically, the poly A addition site is placed about 30 to 2000 nucleotides downstream of the termination site of the protein at a position prior to transcription termination.

A number of selection systems can be used, including, but not limited to, the herpes simplex virus thymidine kinase (TK); hygromycin resistance; auxotrophic markers; antimetabolite resistance such as dihydrofolate reductase (dhfr) gene to confer resistance to methotrexate; neomycin resistance (bacterial in origin) for resistance to the aminoglycoside G-418; and the like.

Expression vectors will contain the variant luciferase nucleic acid molecule operably attached to a promoter. The promoter is generally positioned at the 5'-end of the transcription initiation site of the luciferase nucleic acids molecule such that the upstream promoter stimulates transcription of the DNA and promotes expression of the encoded luciferase protein. A promoter is operably attached when it is in the correct location and orientation in relation to the variant luciferase nucleic acid molecule to control RNA polymerase initiation and expression of the

-16-

luciferase. The promoter used to express the variant luciferase is not critical to the present invention. The use of promoters is well-known in the art.

When a host cell or organism is used to replicate
5 and/or express the luciferase from the nucleic acid molecule of the invention, the host cell or organism is within the scope of the invention. A host cell is a cell which is engineered or genetically modified with variant nucleic acid molecule of the present invention encoding the
10 luciferase. The host cell can be isolated, in a tissue, in an organ, or in an organism. Alternatively, established cell lines that grow continuously in culture from a single cell or group of cells can be used in connection with the present invention

15 A host cell can be transformed, transfected or transduced with a vector containing a variant luciferase nucleic acid molecule, e.g., by permeabilizing the cell membrane either chemically or physically or with viral- or *Agrobacterium*-mediated transfer. Calcium phosphate
20 precipitation, DEAE-dextran, electroporation, and direct microinjection are examples of such methods. Alternatively, liposomes or protein conjugates formed with certain lipids and amphophilic peptides can be used for *in vivo* and *in vitro* transfection.

25 The variant luciferase nucleic acid molecule of the present invention can be used in a variety of assays as a reporter gene, e.g., to identify transformed cells; to measure gene expression *in vitro*, *in vivo* and *ex vivo*; to label specific microorganisms or cells of multicellular
30 organisms; to label and locate fusion proteins; to study intracellular trafficking and the like. The variant luciferase nucleic acid molecule can be used in combination with other light emitting reporter genes, e.g., fluorescent proteins, luciferases that emit light at a different

-17-

wavelength or convert substrates that emit light, change color and/or fluoresce, etc. Different colored reporter genes can be used simply to identify multiple cell populations in a mixed cell culture or to track multiple
5 cell types, enabling differences in cell movement or migration to be visualized in real-time. Other embodiments include tracking and determining the location of cells or multiple proteins within a single cell, tissue or organism; differential promoter analyses in which gene expression
10 from two different promoters is determined in the same cell, tissue or organism; imaging, and FACS sorting of mixed cell populations.

Using the variant luciferase nucleic acid molecule, a range of promoters can be tested for their suitability for
15 use with a given gene, cell, or system, including *in vitro* uses (e.g., identifying a suitable promoter for use in recombinant expression, high level protein production, etc.) and *in vivo* uses (e.g. pre-clinical testing, gene therapy in subjects, etc).

20 For promoter analysis, the variant luciferase-coding sequence is operably linked to the promoter to be tested. Alternatively, the variant luciferase can be introduced, without a promoter, into the chromosome with the intent of identifying a previously uncharacterized promoter element
25 (the technique referred to as promoter trapping). The production of light is related to the expression levels of the luciferase, which are controlled by the promoter element. The results are generally compared to a control gene, cell or system. Optimizing the combination of a given
30 promoter and a given cell type in recombinant expression and protein production can often be necessary to ensure that the highest possible levels are achieved.

A further development of using promoters along with the variant luciferase nucleic acid molecule of the present

-18-

invention is its use in screening protocols. In these embodiments, which are generally conducted *in vitro*, a genetically engineered cell is used to identify the presence of a particular compound or agent in a composition. The variant luciferase nucleic acid molecule is positioned downstream of a promoter that is known to be inducible by the agent that one wishes to identify. Expression of luciferase in the cells will normally be silent, and will be switched on by exposing the cell to a composition that contains the selected agent. In using a promoter that is responsive to, for example, a heavy metal, a toxin, a hormone, or other defined molecule, the presence of a heavy metal, toxin, cytokine or such like can readily be determined. Alternatively, constitutive expression of the reporter by genetically engineered cells can be useful in determining repressors of particular promoters.

Kits containing the variant luciferase nucleic acid molecule of the present invention are another embodiment of the present invention. Such kits will generally contain a variant luciferase nucleic acid molecule or a vector capable of expressing a variant luciferase nucleic acid molecule. The variant luciferase nucleic acid molecule or vector containing the same will be provided in a suitable formulation or as a lyophilized powder, and the kits can also contain, or be packaged with, one or more further molecular biological reagents, such as restriction enzymes, and with instructions for use.

What is claimed is:

1. A nucleic acid molecule encoding a luciferase for
expression in an organism with a bias for cytosine (C) or
5 guanine (G) in the third position of the codon, wherein the
nucleic acid sequence has a G and C content of at least
50%.

2. The nucleic acid molecule of claim 1, further
10 comprising at least one intron.

3. A vector comprising the nucleic acid molecule of
claim 1.

15 4. A host cell containing the vector of claim 3.

5. A method for enhancing the level of a luciferase
enzyme in an organism with a bias for C or G in the third
position of the codon comprising introducing into an
20 organism with a bias for C or G in the third position of
the codon a nucleic acid molecule of claim 1, thereby
enhancing the level of luciferase enzyme produced in the
organism.

25 6. A kit for expressing a luciferase enzyme in an
organism with a bias for C or G in the third position of
the codon comprising the nucleic acid molecule of claim 1.

1/4

OPTIMIZE FIREFLY LUCIFERASE SEQUENCE USING CODON BIAS TABLE
(415 CHANGES IN 1619 NUCLEOTIDES => 26% CHANGE)

OBTAIN 86 OLIGONUCLEOTIDE OF 40 bp EACH AND 20 bp OVERLAP

GENERATE THE MOLECULE IN TWO UNITS WITH OVERLAP AT SapI SITE.

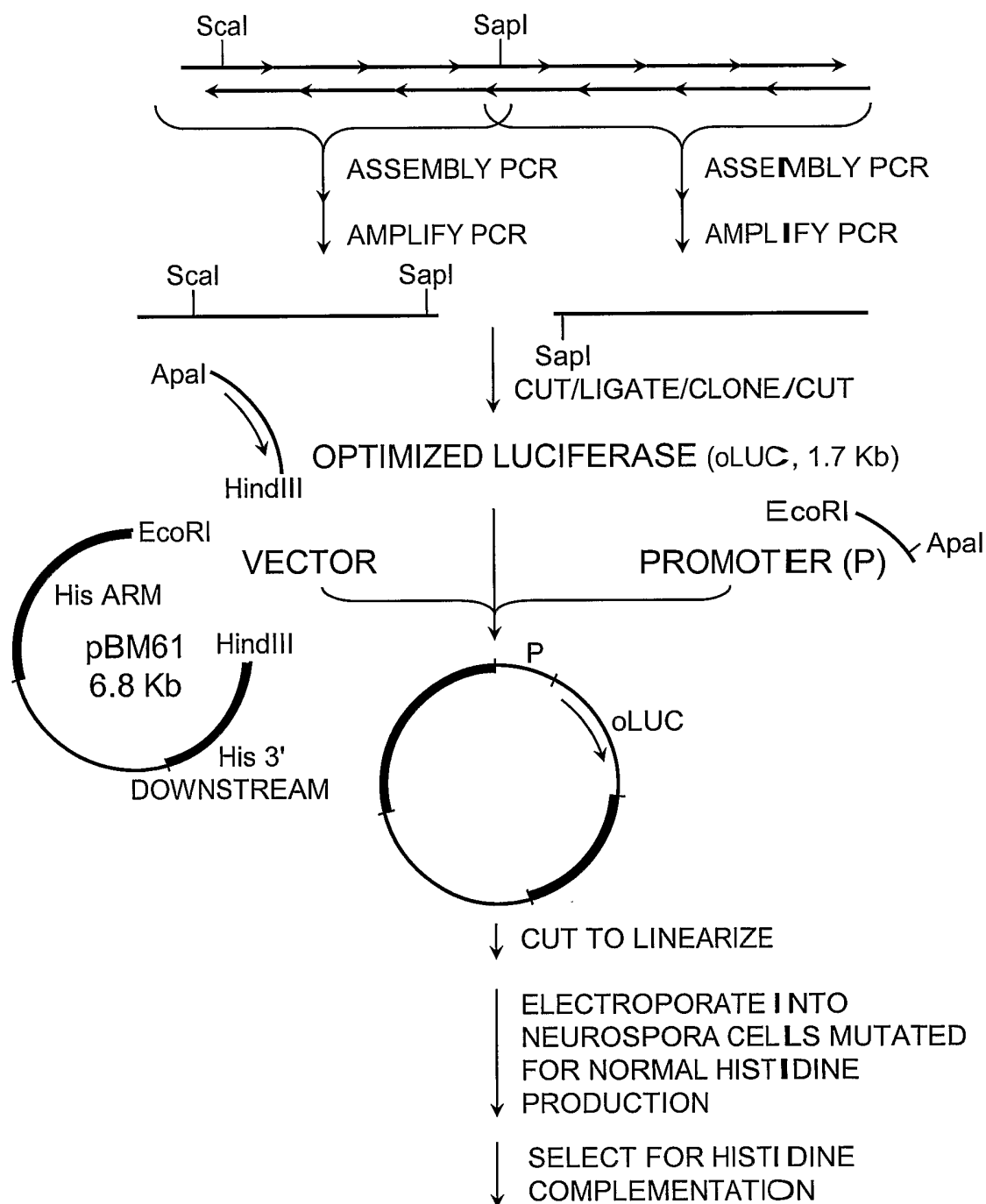


FIG. 1

NATIVE	ATGGAAGACGCCCAAAAACATAAAGAAAGGCCCGGCCCATTTCTATCCTCTAGAGGATGGA	60
VARIANT	ATGGAGGACGCCCAAGAACATCAAGAAAGGCCCGGCCCTTTCTACCCCTCGAGGACGGC	60

NATIVE	ACCGCTGGAGAGCAACTGCATAAGGCTATGAAGAGATACGCCCTGGTTCCCTGGAAACAATT	120
VARIANT	ACCGCCGGCGAGAGCTCCACAAAGGCCATGAAGCGCTACGCCCTCGTCCCGGCACCATC	120

NATIVE	GCTTTTACAGATGCACATATCGAGGTGAACATCACGTACCGGGAATACTTCGAAATGTCC	180
VARIANT	GCCTTCACCGACGCCCAATCGAGGTCAACATCACCTACGCCGAGTACTTCGAGATGTCC	180

NATIVE	GTTCGGTTGGCAGAAGCTATGAAACGATATGGGCTGAATACAAATCACAGAATCGTCGTA	240
VARIANT	GTCCGCCCTCGCCGAGGCCATGAAGCGCTACGGCTCAACACCAACCCGCGCATCGTCGTC	240

NATIVE	TGCAGTGAAAACTCTCTTCAATTCTTTATGCCGGTGTGGGCGCGTTATTATCGGAGTT	300
VARIANT	TGCTCCGAGAACTCCCTCCAGTTCTTTCATGCCCCGTCTCGGCGCCCTCTTCATCGCGTC	300

NATIVE	GCAGTTGCGCCCGCGAACGACATTTATAATGAACGTGAATTGCTCAACAGTATGAACATT	360
VARIANT	GCCGTGCCCCCGCCAAACGACATCTAACAGCGCGAGCTCCTCAACTCCATGAACATC	360

NATIVE	TCGCAGCCTACCGTAGTGTTTTCCAAAAAGGGGTTGCAAAAAAATTTGAACGTGCAA	420
VARIANT	TCCAGCCCAACCGTCGTCTCGTCTCCAAAGAGGGCTCCAGAAAGATCCTCAACGTCCAG	420

NATIVE	AAAAAATTACCAATAATCCAGAAAAATTATTATCATGGATTCTAAAACGGATTACCAGGGA	480
VARIANT	AAGAAGTCCCCCATCATCCAGAAAGATCATCATCATGGACTCCAAGACCGACTACCAGGC	480

NATIVE	TTTCAGTCGATGTACACGTTTCGTACATCTCATCTACCTCCCGGTTTAAATGAATACGAT	540
VARIANT	TTCCAGTCCATGTACACCTTCGTACCTCCACCTCCCGGCTTCAACGAGTACGAC	540

FIG. 2A

3/4

NATIVE	TTTGTACCAAGAGTCCTTTGATCGTGACAAAACAATTGCACTGATAAATGAATTCCTCTGGA	600
VARIANT	TTCGTCCCCGAGTCCTTCGACCGCGACAAGACCATCGCCCTCATCATGAATCCTCCGGC	600
	** ** *	
NATIVE	TCTACTGGGTACCTAAGGGTGTGGCCCTTCCGCATAGAACTGCCCTGCGTCAGATTCTCG	660
VARIANT	TCCACCGGCTCCCAAGGGCTCGCCCTCCCCACCGCACCGCTGCGTCCGCTTCTCC	660
	** ** *	
NATIVE	CATGCCAGAGATCCCTATTTTGGCAATCAAAATCATCCGGATACTGCGATTTAAAGTGTT	720
VARIANT	CACGCCCGGACCCCATCTTCGGCAACAGATCATCCCCGACACCGCCATCCTCTCCGTC	720
	** ** *	
NATIVE	GTTCCATTCCATCACGGTTTGGAAATGTTTACTACACTCGGATATTTGATATGTGGATTT	780
VARIANT	GTCCCTTCCACCAACGGCTTCGGCATGTTACCAACCTTCGGCTACCTCATCTGCGGCTTC	780
	** ** *	
NATIVE	CGAGTCGTCTTAATGTATAGATTGAAGAAAGAGCTGTTTACGATCCCTTCAGGATTAC	840
VARIANT	CGCGTCGTCTCATGTACCGCTTCGAGGAGAGCTCTTCCCTCCGCTCCCTCCAGGACTAC	840
	** ** *	
NATIVE	AAAATTCAAAGTGCCTTGCTAGTACCAACCCCTATTTTCATTTCCGCCAAAAGCACTCTG	900
VARIANT	AAGATCCAGTCCGCCCTCCTCGTCCCAACCCCTCTTCTCTTCTTCCCAAGTCCACCCCTC	900
	** ** *	
NATIVE	ATTGACAAATACGATTATCTAAATTTACACGAAATGCTTCTGGGGGGCGCACCTCTTTCG	960
VARIANT	ATCGACAAAGTACGACCTCTCCAACCTCCACGAGATCGCTCCGGGGCGCCCCCTCTCC	960
	** ** *	
NATIVE	AAAGAAGTCGGGGAAGCGGTTGCAAAAACCGCTTCCATCTTCCAGGGATACGACAAGGATAT	1020
VARIANT	AAGGAGGTCGGCGAGGCGGTCGCCAAGCGCTTCCACCTCCCGGCATCCGCCAGGGCTAC	1020
	** ** *	
NATIVE	GGGCTCACTGAGACTACATCAGCTATTCTGATTACACCCGAGGGGATGATAAACCGGGC	1080
VARIANT	GGCCTCACCGAGACCACCTCCGCCATCTCATACCCCGAGGGCGACGACAAGCCCGGC	1080
	** ** *	

FIG. 2B

NATIVE	GCGGTCGGTAAAGTTGTTCCATTTTTGAAGCGAAAGGTTGTGGATCTGGATACCGGGAAA	1140
VARIANT	GCCGTCGGCAAGGTCGTCCCTTCTTCGAGGCCAAGGTCGTGACCTCGACACCGGCAAG	1140
	** ** ** ** **	
NATIVE	ACGCTGGGCGTTAATCAGAGAGGCGGAATTATGTGTGACAGGACCTATGATTATGTCCGGT	1200
VARIANT	ACCCTCGGCGTCAACCAGCGCGGAGCTCTGCGTCCGCGGCCCATGATCATGTCCGGC	1200
	** ** ** ** * ** ** * ** ** * ** ** * ** ** * ** ** * **	
NATIVE	TATGTAAACAATCCGGAAGCGACCAACGCCCTTGATTGACAAAGGATGGCTACATTCT	1260
VARIANT	TACGTCAACAACCCCGAGGCCACCAACGCCCTCATCGACAAAGGACGGCTGGCTCCACTCC	1260
	** ** ** ** * ** ** * ** ** * ** ** * ** ** * ** ** * **	
NATIVE	GGAGACATAGCTTACTGGGACGAAGACGAACACTTCTTCATAGTTGACCGCTTGAAGTCT	1320
VARIANT	GGGACATCGCTACTGGGACGAGGACGAGCACTTCTTCATCGTCGACCGCTCAAGTCC	1320
	** ** ** ** * ** ** * ** ** * ** ** * ** ** * ** ** * **	
NATIVE	TTAATTAAATACAAAGGATATCAGGTGGCCCGCCCGCTGAATTGGAATCGATATTGTTACAA	1380
VARIANT	CTCATCAAGTACAAAGGCTACCAGGTGCCCCCGCCGAGCTCGAGTCCATCCTCCTCCAG	1380
	* ** ** ** * ** ** * ** ** * ** ** * ** ** * ** * ** *	
NATIVE	CACCCCAACATCTTCGACGCGGGCGTGGCAGGTCTTCCGACGATGACGCCGGTGAACTT	1440
VARIANT	CACCCCAACATCTTCGACGCGGGCGTGGCCGGCTTCCCGACGACGACGCCGGCGAGCTC	1440
	** ** ** ** * ** ** * ** ** * ** ** * ** ** * ** ** * **	
NATIVE	CCCGCCGCGGTTGTTGTTGGAGCACGGAAAGACGATGACGGGAAAAGAGATCGTGGAT	1500
VARIANT	CCCGCCGCGTCGTCTGTCCTCGAGCACGGCAAGACCATGACCGAGAAGGAGATCGTCGAC	1500
	** ** ** ** * ** ** * ** ** * ** ** * ** ** * ** * ** *	
NATIVE	TACGTCGCCAGTCAAGTAACAACCGGAAAAGTTGCGCGGAGGAGTTGTTGTGGAC	1560
VARIANT	TACGTCGCCCTCCAGGTCAACCAACCGCAAGAGCTCCGCGGGCGGTCTGTTGTGGAC	1560
	** ** ** ** * ** ** * ** ** * ** ** * ** ** * ** * ** *	
NATIVE	GAAGTACCGGAAAGGTCTTACCGGAAAACCTCGACGCAAGAAAATCAGAGAGATCCTCATA	1620
VARIANT	GAGGTCCCAAGGGCTCACCGGCAAGCTCGACGCCCCGCAAGATCCGCGAGATCCTCATC	1620
	** ** ** * ** * ** * ** ** * ** * ** ** * ** * ** * ** * **	
NATIVE	AAGGCCAAGAAGGGCGGAAAGTCCAAATTGTAA	1653 (SEQ ID NO:2)
VARIANT	AAGGCCAAGAAGGGCGGCAAGTCCAAGCTCTGA	1653 (SEQ ID NO:1)
	** ** ** * ** * ** * ** ** * ** * ** ** * ** * ** * ** *	

FIG. 2C

SEQUENCE LISTING

<110> Trustees of Dartmouth College
University of Minnesota
Dunlap, Jay C.
Loros, Jennifer J.
Mehra, Arun
Gooch, Van D.

<120> NUCLEIC ACID SEQUENCES ENCODING LUCIFERASE FOR EXPRESSION IN
FILAMENTOUS FUNGI

<130> DC0280WO

<150> US 60/614,874
<151> 2004-09-30

<160> 5

<170> PatentIn version 3.3

<210> 1
<211> 1653
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic luciferase nucleic acid

<400> 1
atggaggacg ccaagaacat caagaagggc cccgccccct tctaccccct cgaggacggc 60
accgccggcg agcagctcca caaggccatg aagcgctacg ccctcggtccc cggcaccatc 120
gccttcaccg acgcccacat cgagggtcaac atcacctacg ccgagtactt cgagatgtcc 180
gtccgcctcg ccgaggccat gaagcgctac ggctcaaca ccaaccaccg catcgctcgtc 240
tgctccgaga actccctcca gttcttcatg cccgtcctcg ggcgcctctt catcggcgtc 300
gccgtcggcc ccgccaacga catctacaac gagcgogagc toctcaactc catgaacatc 360
tcccagccca ccgtcgtctt cgtctccaag aagggcctcc agaagatcct caacgtccag 420
aagaagctcc ccatcatcca gaagatcatc atcatggact ccaagaccga ctaccagggc 480
ttccagtcga tgtacacctt cgtcacctcc cacctccccc ccggcttcaa cgagtacgac 540
ttcgtccccc agtccttcga ccgcgacaag accatcgccc tcatcatgaa ctctccggc 600
tccaccggcc tcccaaggc cgtcgccctc cccacacgca ccgcctgggt ccgcttctcc 660
cacgcccggc acccatctt cggcaaccag atcatccccg acaccgcat cctctccgtc 720
gtcccttccc accacggctt cggcatgttc accaccctcg gctacctcat ctggcgcttc 780
cgcgtcgtcc tcatgtaccg cttcgaggag gagctcttcc tccgtccct ccaggactac 840
aagatccagt ccgcctcct cgtccccacc ctcttctcct tcttcgcaa gtccaccctc 900

- 2 -

```

atcgacaagt acgacctctc caacctccac gagatcgctt ccggcgggcg cccctctctc 960
aaggaggtcg gcgaggccgt cgccaagcgc ttccacctcc ccggcatccg ccaggggtac 1020
ggcctcaccg agaccacctc cgccatcctc atcacccccg agggcgacga caagcccggc 1080
gccgtcgga aggctgtccc cttcttcgag gccaaggtcg tcgacctga caccggcaag 1140
accctcggcg tcaaccagcg cggcgagctc tgcgtccgcg gcccctgat catgtccggc 1200
tacgtcaaca accccgaggc caccaacgcc ctcatcgaca aggacggctg gctccactcc 1260
ggcgacatcg cctactggga cgaggacgag cacttcttca tcgtcgaccg cctcaagtcc 1320
ctcatcaagt acaagggcta ccaggctgcc cccgccgagc tcgagtccat cctcctccag 1380
caccccaaca tcttcgacgc cggcgctgcc ggccctcccg acgacgacgc cggcgagctc 1440
cccgccgccg tcgtcgtcct cgagcacggc aagaccatga ccgagaagga gatcgtcgac 1500
tacgtcgctt cccaggtcac caccgccaag aagctccgcg gcggcgctgt cttcgtcgac 1560
gaggtcccca agggcctcac cggcaagctc gacgcccga agatccgga gatcctcatc 1620
aaggccaaga agggcgggcaa gtccaagctc tga 1653

```

```

<210> 2
<211> 1653
<212> DNA
<213> Photinus pyralis

```

```

<400> 2
atggaagacg ccaaaaacat aaagaaaggc ccggcgccat tctatcctct agaggatgga 60
accgctggag agcaactgca taaggctatg aagagatacg ccctgggttc tggaacaatt 120
gcttttacag atgcacatat cgaggngaac atcacgtacg cggaatactt cgaaatgtcc 180
gttcgggttg cagaagctat gaaacgatat gggctgaata caaatcacag aatcgtcgta 240
tgcagtgaag actctcttca attctttatg ccggtgttgg gcgcgttatt tatcggagtt 300
gcagttgcgc ccgcgaacga catttataat gaacgtgaat tgctcaacag tatgaacatt 360
tcgcagccta ccgtagtggt tgtttccaaa aaggggttgc aaaaaatttt gaacgtgcaa 420
aaaaaattac caataatcca gaaaattatt atcatggatt ctaaacgga ttaccaggga 480
tttcagtcga tgtacacgtt cgtcacatct catctacctc ccggttttaa tgaatacgat 540
tttgtaccag agtcctttga tcgtgacaaa acaattgcac tgataatgaa ttctcttgga 600
tctactgggt tacctaaggg tgtggccctt ccgcatagaa ctgcctgcgt cagattctcg 660
catgccagag atcctatctt tggcaatcaa atcattccgg atactgcgat ttttaagtgt 720
gttccattcc atcacggttt tggaatgttt actacactcg gatatttgat atgtggattt 780

```

- 3 -

```

cgagtcgtct taatgtatag atttgaagaa gagctgtttt tacgatccct tcaggattac      840
aaaattcaaa gtgcgttgct agtaccaacc ctattttcat tcttcgccaa aagcactctg      900
attgacaaaat acgatttatc taattttacac gaaattgctt ctgggggcg c acctctttcg      960
aaagaagtcg gggaagcggg tgcaaaacgc ttccatcttc cagggatagc acaaggatat    1020
gggctcactg agactacatc agctattctg attacacccg agggggatga taaaccgggc    1080
gcggtcggta aagttgttcc attttttgaa gcgaagggtt tggatctgga taccgggaaa    1140
acgctgggcg ttaatcagag aggcgaatta tgtgtcagag gacctatgat tatgtccggg    1200
tatgtaaaca atccggaagc gaccaacgcc ttgattgaca aggatggatg gctacattct    1260
ggagacatag cttactggga cgaagacgaa cacttcttca tagttgaccg cttgaagtct    1320
ttaattaaat acaaaggata tcaggtggcc cccgctgaat tggaatcgat attgttacia    1380
caccccaaca tcttcgacgc gggcgtggca ggtcttcccg acgatgacgc cgggtgaactt    1440
cccgccgccg ttgttgtttt ggagcacgga aagacgatga cggaaaaaga gatcgtggat    1500
tacgtcgcca gtcaagtaac aaccgcgaaa aagttgcgcg gaggagtgtg gtttgtggac    1560
gaagtaccga aaggtcttac cggaaaactc gacgcaagaa aaatcagaga gatcctcata    1620
aaggccaaga agggcggaag gtccaaattg taa                                1653

```

<210> 3
 <211> 550
 <212> PRT
 <213> *Photinus pyralis*

<400> 3

```

Met Glu Asp Ala Lys Asn Ile Lys Lys Gly Pro Ala Pro Phe Tyr Pro
1              5              10              15

```

```

Leu Glu Asp Gly Thr Ala Gly Glu Gln Leu His Lys Ala Met Lys Arg
                20              25              30

```

```

Tyr Ala Leu Val Pro Gly Thr Ile Ala Phe Thr Asp Ala His Ile Glu
                35              40              45

```

```

Val Asn Ile Thr Tyr Ala Glu Tyr Phe Glu Met Ser Val Arg Leu Ala
                50              55              60

```

```

Glu Ala Met Lys Arg Tyr Gly Leu Asn Thr Asn His Arg Ile Val Val
65              70              75              80

```

```

Cys Ser Glu Asn Ser Leu Gln Phe Phe Met Pro Val Leu Gly Ala Leu
                85              90              95

```

- 4 -

Phe Ile Gly Val Ala Val Ala Pro Ala Asn Asp Ile Tyr Asn Glu Arg
 100 105 110

Glu Leu Leu Asn Ser Met Asn Ile Ser Gln Pro Thr Val Val Phe Val
 115 120 125

Ser Lys Lys Gly Leu Gln Lys Ile Leu Asn Val Gln Lys Lys Leu Pro
 130 135 140

Ile Ile Gln Lys Ile Ile Ile Met Asp Ser Lys Thr Asp Tyr Gln Gly
 145 150 155 160

Phe Gln Ser Met Tyr Thr Phe Val Thr Ser His Leu Pro Pro Gly Phe
 165 170 175

Asn Glu Tyr Asp Phe Val Pro Glu Ser Phe Asp Arg Asp Lys Thr Ile
 180 185 190

Ala Leu Ile Met Asn Ser Ser Gly Ser Thr Gly Leu Pro Lys Gly Val
 195 200 205

Ala Leu Pro His Arg Thr Ala Cys Val Arg Phe Ser His Ala Arg Asp
 210 215 220

Pro Ile Phe Gly Asn Gln Ile Ile Pro Asp Thr Ala Ile Leu Ser Val
 225 230 235 240

Val Pro Phe His His Gly Phe Gly Met Phe Thr Thr Leu Gly Tyr Leu
 245 250 255

Ile Cys Gly Phe Arg Val Val Leu Met Tyr Arg Phe Glu Glu Glu Leu
 260 265 270

Phe Leu Arg Ser Leu Gln Asp Tyr Lys Ile Gln Ser Ala Leu Leu Val
 275 280 285

Pro Thr Leu Phe Ser Phe Phe Ala Lys Ser Thr Leu Ile Asp Lys Tyr
 290 295 300

Asp Leu Ser Asn Leu His Glu Ile Ala Ser Gly Gly Ala Pro Leu Ser
 305 310 315 320

Lys Glu Val Gly Glu Ala Val Ala Lys Arg Phe His Leu Pro Gly Ile
 325 330 335

- 5 -

Arg Gln Gly Tyr Gly Leu Thr Glu Thr Thr Ser Ala Ile Leu Ile Thr
 340 345 350

Pro Glu Gly Asp Asp Lys Pro Gly Ala Val Gly Lys Val Val Pro Phe
 355 360 365

Phe Glu Ala Lys Val Val Asp Leu Asp Thr Gly Lys Thr Leu Gly Val
 370 375 380

Asn Gln Arg Gly Glu Leu Cys Val Arg Gly Pro Met Ile Met Ser Gly
 385 390 395 400

Tyr Val Asn Asn Pro Glu Ala Thr Asn Ala Leu Ile Asp Lys Asp Gly
 405 410 415

Trp Leu His Ser Gly Asp Ile Ala Tyr Trp Asp Glu Asp Glu His Phe
 420 425 430

Phe Ile Val Asp Arg Leu Lys Ser Leu Ile Lys Tyr Lys Gly Tyr Gln
 435 440 445

Val Ala Pro Ala Glu Leu Glu Ser Ile Leu Leu Gln His Pro Asn Ile
 450 455 460

Phe Asp Ala Gly Val Ala Gly Leu Pro Asp Asp Asp Ala Gly Glu Leu
 465 470 475 480

Pro Ala Ala Val Val Val Leu Glu His Gly Lys Thr Met Thr Glu Lys
 485 490 495

Glu Ile Val Asp Tyr Val Ala Ser Gln Val Thr Thr Ala Lys Lys Leu
 500 505 510

Arg Gly Gly Val Val Phe Val Asp Glu Val Pro Lys Gly Leu Thr Gly
 515 520 525

Lys Leu Asp Ala Arg Lys Ile Arg Glu Ile Leu Ile Lys Ala Lys Lys
 530 535 540

Gly Gly Lys Ser Lys Leu
 545 550

<210> 4

<211> 1741

- 6 -

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic luciferase nucleic acid

<400> 4

```

atggaggacg ccaagaacat caagaagggc cccgccccct tctacccccct cgaggacggc      60
accgcccggcg agcagctcca caaggccatg aagcgctacg ccctcgcccc eggcaccatc      120
gccttcgtag gtttcctcca gctctcgctt ccagcaccgg aggcacatct cgggcatctt      180
cacaacaaca gacactgaca tctcattctc acagaccggc gccacatctg aggtcaacat      240
cacctacgcc gagtacttcg agatgtccgt ccgcctcgcc gaggccatga agcgctacgg      300
cctcaacacc aaccaccgca tcgtcgtctg ctccgagAAC tccctccagt tcttcatgcc      360
cgtcctcggc gccctcttca tcggcgctgc cgtcgccccg gccaacgaca tctacaacga      420
gcgcgagctc ctcaactcca tgaacatctc ccagcccacc gtcgtcttcg tctccaagaa      480
gggcctccag aagatcctca acgtccagaa gaagctcccc atcatccaga agatcatcat      540
catggactcc aagaccgact accagggctt ccagtccatg tacaccttcg tcacctccca      600
cctccccccc ggcttcaacg agtacgactt cgtccccgag tccttcgacc ggcacaagac      660
catcgccctc atcatgaact cctccggctc caccggcctc cccaagggcg tcgcccctcc      720
ccaccgcacc gcctgcgtcc gcttctccca cgcgcgcgac cccatcttcg gcaaccagat      780
catccccgac accgccatcc tctccgtcgt ccccttccac cagggcttcg gcatgttcac      840
caccctcggc tacctcatct gcggcttcgg cgtcgtcctc atgtaccgct tcgaggagga      900
gctcttcctc cgctccctcc aggactacaa gatccagtcc gccctcctcg tccccacct      960
cttctccttc ttgcgaagt ccacctcat cgacaagtac gacctctcca acctccacga      1020
gatcgccctc ggcggcggcc cctctccaa ggaggtcggc gaggccgtcg ccaagcgctt      1080
ccacctcccc ggcatccgcc agggctacgg cctcaccgag accacctccg ccatcctcat      1140
cacccccgag ggcgacgaca agcccggcgc cgtcggcaag gtcgtccctt tcttcgaggg      1200
caaggtcgtc gacctcgaca ccggcaagac cctcggcgtc aaccagcgcg gcgagctctg      1260
cgtcgcggcg cccatgatca tgtcgggcta cgtcaacaac cccgaggcca ccaacgccct      1320
catcgacaag gacggctggc tccactccgg cgacatcgcc tactgggacg aggacgagca      1380
cttcttcctc gtogaccgcc tcaagtccct catcaagtac aagggtacc aggtcgcccc      1440
cgccgagctc gagtccatcc tctccagca cccaacatc ttcgacgccg gcgtcgccgg      1500
cctccccgac gacgacgccg gcgagctccc cgcgcgcgtc gtcgtcctcg agcacggcaa      1560
gacatgacc gagaaggaga tcgtcgaata cgtcgcctcc caggtcacca ccgccaagaa      1620

```

- 7 -

gctccgcggc ggcgctcgtct tcgtcgacga ggtccccaag ggcctcaccg gcaagctcga 1680
cgcccgaag atccgcgaga tcctcatcaa ggccaagaag ggcggcaagt ccaagctctg 1740
a 1741

<210> 5
<211> 88
<212> DNA
<213> *Neurospora crassa*

<400> 5
gtaggtttcc tccagctctc gcctccagca cccgaggcac atctcgggca tcttcacaac 60
aacagacact gacatctcat tctcacag 88