

(CMV), Rous sarcoma virus (RSV), adenovirus (ADV), and bovine papilloma virus (BPV).

[0192] In addition, the control element may also be comprised of a termination sequence and poly(A) addition sequences which are operably linked to nucleotide sequences encoding *D. v. virgifera* protein homologues or fragments thereof. The control element may also be comprised of an enhancer sequence which increases the expression of *D. v. virgifera* protein homologues or fragments thereof.

[0193] Furthermore, the control element may also be comprised of an enhancer, which is any regulatory DNA sequence that can stimulate transcription up to 1000-fold when linked to endogenous or heterologous promoters, with synthesis beginning at the normal mRNA start site. Enhancers are also active when they are placed upstream or downstream from the transcription initiation site, in either normal or flipped orientation, or at a distance of more than 1000 nucleotides from the promoter (Maniatis et al. *Science*, 236:1237 (1987); Alberts et al., *Molecular Biology of the Cell*, 2nd ed. (1989)). Enhancers derived from viruses may be particularly useful, because they typically have a broader host range. Examples include the SV40 early gene enhancer (Dijkema et al, *EMBO J.*, 4:761 (1985)) and the enhancer/promoters derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus (Gorman et al., *Proc. Natl. Acad. Sci.* 79:6777 (1982b)) and from human cytomegalovirus (Boshart et al., *Cell*, 41:521 (1985)). Additionally, some enhancers are regulatable and become active only in the presence of an inducer, such as a hormone or metal ion (Sansone-Corsi and Borelli, *Trends Genet.* 2:215 (1986); Maniatis et al. *Science*, 236:1237 (1987)).

[0194] Where selection is intended, sequences which encode selectable markers may also be included in the vector. Selectable markers for mammalian cells are known in the art, and include for example, thymidine kinase, dihydrofolate reductase (together with methotrexate as a DHFR amplifier), aminoglycoside phosphotransferase, hygromycin B phosphotransferase, asparagine synthetase, adenosine deaminase, metallotionin, and antibiotic resistant genes such as neomycin.

[0195] For homologous recombination, constructs can be prepared where the amplifiable gene will be flanked, normally on both sides with DNA homologous with the DNA of the target region. Depending upon the nature of the integrating DNA and the purpose of the integration, the homologous DNA will generally be within 100 kb, usually 50 kb, preferably about 25 kb, of the transcribed region of the target gene, more preferably within 2 kb of the target gene. Where modeling of the gene is intended, homology will usually be present proximal to the site of the mutation. By gene is intended the coding region and those sequences required for transcription of a mature mRNA. The homologous DNA may include the 5'-upstream region outside of the transcriptional regulatory region or comprising any enhancer sequences, transcriptional initiation sequences, adjacent sequences, or the like. The homologous region may include a portion of the coding region, where the coding region may be comprised only of an open reading frame or combination of exons and introns. The homologous region may comprise all or a portion of an intron, where all or a portion of one or more exons may also be present. Alternatively, the homologous region may comprise the 3'-region, so as to comprise all or a portion of the transcriptional termination region, or the region 3' of this region. The homologous regions may extend over all or a

portion of the target gene or be outside the target gene comprising all or a portion of the transcriptional regulatory regions and/or the structural gene.

[0196] The integrating constructs may be prepared in accordance with conventional ways, where sequences may be synthesized, isolated from natural sources, manipulated, cloned, ligated, subjected to in vitro mutagenesis, primer repair, or the like. At various stages, the joined sequences may be cloned, and analyzed by restriction analysis, sequencing, or the like. Usually during the preparation of a construct where various fragments are joined, the fragments, intermediate constructs and constructs will be carried on a cloning vector comprising a replication system functional in a prokaryotic host, e.g., *E. coli*, and a marker for selection, e.g., biocide resistance, complementation to an auxotrophic host, etc. Other functional sequences may also be present, such as polylinkers, for ease of introduction and excision of the construct or portions thereof, or the like. A large number of cloning vectors are available such as pBR322, the pUC series, etc. These constructs may then be used for integration into the primary mammalian host.

[0197] The mammalian cell expression vectors described herein may be synthesized by techniques well known to those skilled in this art. Other appropriate expression vectors of which numerous types are known in the art for mammalian expression can also be used for this purpose.

[0198] Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC). Exemplary mammalian host cells include particularly primate cell lines and rodent cell lines, including transformed cell lines. Suitable cell lines include, but are not limited to, Chinese hamster ovary (CHO) cells, HeLa cells, baby hamster kidney (BHK) cells, monkey kidney cells (COS-1), human hepatocellular carcinoma cells (e.g., Hep G2), human adenovirus transformed 293 cells, mouse L-929 cells, HaK hamster cell lines, murine 3T3 cells derived from Swiss, Balb-c or NIH mice and a number of other cell lines.

[0199] The DNA can be introduced into the expression host by a variety of techniques that include calcium phosphate/DNA co-precipitates, microinjection of DNA into the nucleus, electroporation, yeast protoplast fusion with intact cells, transfection, polycations, e.g., polybrene, polyornithine, etc., or the like. The DNA may be single or double stranded DNA, linear or circular. The various techniques for transforming mammalian cells are well known (see Keown et al., *Methods Enzymol.* (1989), Keown et al., *Methods Enzymol.* 185:527-537 (1990); Mansour et al., *Nature* 336:348-352, (1988)).

[0200] (h) Computer Media

[0201] The nucleotide sequence provided in SEQ ID NO:1, through SEQ ID NO:9112 or fragment thereof, or complement thereof, or a nucleotide sequence at least 90% identical, preferably 95%, identical even more preferably 99% or 100% identical to the sequence provided in SEQ ID NO:1 through SEQ ID NO:9112 or fragment thereof, or complement thereof, can be "provided" in a variety of mediums to facilitate use. Such a medium can also provide a subset thereof in a form that allows a skilled artisan to examine the sequences.

[0202] In one application of this embodiment, a nucleotide sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium that can be read and accessed directly by a computer. Such media include, but are not lim-