

(12) **FASCÍCULO DE PATENTE DE INVENÇÃO**

|                                                                                    |                                                                                                                                                                                 |
|------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| (22) Data de pedido: <b>2007.01.05</b>                                             | (73) Titular(es):<br><b>TRANSGENE SA</b><br><b>PARC D'INNOVATION BOULEVARD GONTHIER</b><br><b>D'ANDERNACH 67400 ILLKIRCH</b><br><b>GRAFFENSTADEN</b><br><b>FR</b>               |
| (30) Prioridade(s): <b>2006.01.05 EP 06360001</b><br><b>2006.10.13 EP 06360047</b> |                                                                                                                                                                                 |
| (43) Data de publicação do pedido: <b>2008.10.15</b>                               |                                                                                                                                                                                 |
| (45) Data e BPI da concessão: <b>2010.08.11</b><br><b>205/2010</b>                 | (72) Inventor(es):<br><b>PHILIPPE ERBS</b><br><b>JEAN-MARC BALLOUL</b><br><b>MARINA KAPFER</b><br><b>NATHALIE SILVESTRE</b><br><b>FR</b><br><b>FR</b><br><b>FR</b><br><b>FR</b> |
|                                                                                    | (74) Mandatário:<br><b>MANUEL GOMES MONIZ PEREIRA</b><br><b>RUA ARCO DA CONCEIÇÃO, N.º 3, 1º ANDAR 1100-028</b><br><b>LISBOA</b><br><b>PT</b>                                   |

(54) Epígrafe: **TRANSCRIPTASE INVERSA DA TELOMERASE AVIÁRIA**

(57) Resumo:

O PRESENTE INVENTO REFERE-SE NOTAVELMENTE A NOVAS TRANSCRIPTASES INVERSAS DA TELOMERASE RECOMBINANTES, A MOLÉCULAS DE ÁCIDOS NUCLEICOS QUE AS CODIFICAM, A CÉLULAS COMPREENDENDO AS DITAS MOLÉCULAS DE ÁCIDOS NUCLEICOS E AO USO DAS CÉLULAS PARA A PRODUÇÃO DE SUBSTÂNCIA DE INTERESSE.

## RESUMO

### **TRANSCRIPTASE INVERSA DA TELOMERASE AVIÁRIA**

O presente invento refere-se notavelmente a novas transcriptases inversas da telomerase recombinantes, a moléculas de ácidos nucleicos que as codificam, a células compreendendo as ditas moléculas de ácidos nucleicos e ao uso das células para a produção de substância de interesse.

## DESCRIÇÃO

### **TRANSCRIPTASE INVERSA DA TELOMERASE AVIÁRIA**

O presente invento refere-se notavelmente a novas transcriptases inversas da telomerase recombinantes, a moléculas de ácidos nucleicos que as codificam, a células compreendendo as ditas moléculas de ácidos nucleicos e ao uso das células para a produção de substância de interesse.

Em 1965 L. Hayflick descobriu que as células têm um momento programado de morte. Como uma explicação para o envelhecimento, ele sugeriu que o número de vezes que uma célula humana se pode dividir seja limitado (Exp Cell Res. 1965 Mar; 37: 614-36). Tal é conhecido por ser causado pelo encurtamento de telómeros à medida que a célula se divide. Os cromossomas são nivelados por telómeros que consistem em sequência de ADN hexamérico, conservado, com repetições em tandem, não-codificante associados a proteínas de ligação de cadeia simples e dupla. Os telómeros são responsáveis por funções de estabilidade do genoma e em particular replicação do terminal do cromossoma. A replicação da extremidade do cromossoma com sucesso requer tanto a estrutura do telómero única e a enzima especializada transcriptase inversa da telomerase, que é uma nucleoproteína que tem uma actividade enzimática de transcriptase inversa. A transcriptase inversa da telomerase é capaz de alongar o vector de repetição do telómero, permitindo replicação estendida da cadeia filha complementar. Nas células a que falta a transcriptase inversa da telomerase, o ADN telomérico encurta em sucessivas divisões pois as enzimas de síntese de ADN são incapazes de replicar completamente a extremidade dos cromossomas assim que o iniciador de ARN é removido. Numerosos trabalhos reportaram a evidência que o designado "relógio de telómero"

é uma característica importante do período de vida da célula humana. A hipótese do telómero do envelhecimento celular propõe que a redução do telómero está relacionada com uma falta de actividade da transcriptase inversa da telomerase durante e desencadeia instabilidade cromossómica, conduzindo, a senescência, apoptose. A actividade da transcriptase inversa da telomerase apresenta regulação negativa em linhagens de células somáticas durante o desenvolvimento *in vivo* e em células primárias *in vitro* correlacionando com o encurtamento do telómero. Ao contrário, a regulação positiva ou desregulação da actividade da transcriptase inversa da telomerase ocorre em células transformadas e tumores. Os ADNcs da transcriptase inversa da telomerase (TERT) de vários mamíferos e um anfíbio foram clonados e estudados (Nakamura et al. 1997. Science 277. 955 - 959; Greenberg et al. 1998. Oncogene 16, 1723-1730).

Uma sobreexpressão de TERT numa célula conduz à imortalização da dita célula, foi proposto o uso de TERT para a produção de linhas celulares (McSharry et al., 2001 J Gen Virol. 82, 855-63). Contudo, a actividade de TERT é restringida à espécie. Por exemplo, a TERT humana é incompatível com o aparelho de manutenção do telómero aviário (Michailidis et al. 2005. Biochem Biophys Res Commun. 335 (1), 240-6). Assim, para desenvolver linhas celulares aviárias e mais particularmente de *anatidae*, existe uma necessidade de TERT que actue nestas células particulares.

Delany et al (Delany ME, Daniels LM. Gene. 2004 Sep 15;339:61-9) divulga um polipéptido TERT isolado de galinha *Gallus domesticus* (i.e., chTERT) que pertence à família *Fasianidae* e a sequência de ácidos nucleicos correspondente. Os autores mostram que polipéptidos TERT de vertebrados (i.e.

galinha, humano, *Xenopus*, ratinho, ratazana e hamster) apresentam conservação em regiões específicas. Contudo, os autores deste artigo não divulgam nem sugerem que a molécula de ácidos nucleicos que codifica o polipéptido chTERT pode ser utilizada para a imortalização de células, muito menos para a produção de vírus ou de moléculas de interesse. Delany et al nunca referem o interesse de clonar proteínas TERT aviárias adicionais.

Fragnet et al (Fragnet L, Kut E, o Rasschaert D. J Biol Chem. 2005 Jun 24;280(25):23502-15) é um artigo de revisão relativo à correlação entre vírus e activação da actividade da telomerase. Os autores mencionam que a actividade de telomerase pode estar associada com a carcinogénese, e apresenta a actividade da telomerase como um factor entre muitos outros no processo multifactor da tumorigénese, que envolve o desregulamento da expressão de várias proteínas celulares de ciclo e de sistemas de reparo de ADN.

As linhas celulares eucarióticas são fundamentais para o fabrico de vacinas virais e muitos produtos da biotecnologia. Os biológicos produzidos em culturas celulares incluem enzimas, hormonas, imunobiológicos (anticorpos monoclonais, interleucinas, linfocinas), e agentes anti-cancro. Embora possam ser produzidas muitas proteínas mais simples utilizando células bacterianas, proteínas mais complexas que são glicosiladas, devem ser actualmente feitas em células de eucarióticas.

As linhas celulares aviárias são particularmente úteis uma vez que vários vírus usados em composição farmacêutica são capazes de se reproduzir nelas. Mais notoriamente, vários vírus são apenas capazes de crescer em células aviárias. Tal

é por exemplo o caso do Vírus Ankara Modificado (MVA) que não é capaz de crescer na maioria das células de mamífero. Este poxvírus, que derivou de um Vírus Vaccinia por mais de 500 passagens em CEF foi usado nos anos setenta para vacinar pessoas com imunodeficiência contra Varíola. Actualmente, o MVA é principalmente utilizado como um vector para terapia génica e fins de imunoterapia. Por exemplo, o MVA é utilizado como um vector para o gene MUC1 para vacinar pacientes contra tumor que expressa este antigénio (Scholl et al., 2003, J Biomed Biotechnol., 2003, 3, 194-201). Os MVA que transportam o gene que codificam antigénios HPV são também utilizados como vector para o tratamento terapêutico de lesões cervicais de elevado grau. Mais recentemente, o MVA foi o vector de escolha para preparar tratamento profiláctico contra doenças emergentes recentes ou armas biológicas prováveis tais como o vírus do Nilo ocidental e antraz.

Assim, existe uma necessidade para uma nova Transcriptase Inversa da Telomerase capaz de imortalizar células aviárias e mais particularmente células *Anatidae*.

Tal como é usado ao longo do pedido, os termos “um” e “uma” são utilizados no sentido que querem dizer “pelo menos um”, “pelo menos um primeiro”, “um ou mais” ou “uma pluralidade” dos componentes ou passos referenciados, a menos que o contexto indique claramente de outro modo. Por exemplo, o termo “uma célula” inclui uma pluralidade de células, incluindo suas misturas.

O termo “e/ou” onde quer que seja aqui utilizado inclui o significado de “e”, “ou” e “todos ou qualquer outra combinação dos elementos ligados pelo dito termo.”

Tal como é aqui utilizado, pretende-se que os termos “compreendendo” e “compreende” signifiquem que os produtos, composições, e métodos incluam os componentes ou passos referenciados, mas não excluam outros. “Que consiste essencialmente de” quando utilizado para definir produtos, composições e métodos, significará excluindo outros componentes ou passos de qualquer significado essencial. Assim, uma composição que consiste essencialmente dos componentes recitados não excluirá restos de contaminantes e portadores farmacêuticamente aceitáveis. “Que consiste em” deverá significar excluir mais do que restos de elementos de outros componentes ou passos.

O presente invento refere-se a um polipéptido recombinante ou isolado compreendendo uma sequência de aminoácidos que tem pelo menos 80% de identidade de sequência de aminoácidos à sequência de aminoácidos da transcriptase inversa da telomerase de *Cairina moschata* SEQ ID N°: 1. Num modo de execução mais preferido do invento, o polipéptido do invento compreende uma sequência de aminoácidos que tem pelo menos 90% de identidade da sequência de aminoácidos a SEQ ID N°: 1. Num modo de execução mais preferido, o polipéptido do invento compreende a sequência de aminoácido exposta em SEQ ID N°: 1.

O polipéptido do invento tem uma actividade TERT e permite a imortalização de uma célula que pertence à família *Anatidae*.

Tal como é aqui utilizado, o termo “isolado” e/ou “recombinante” significa que a molécula de ácidos nucleicos, ADN, ARN, polipéptidos ou proteínas assim designados foram produzidos em tal forma pela mão do homem, e assim estão separados do seu ambiente celular *in vivo* nativo. Como resultado desta intervenção humana, os ADNs, ARNs,

polipéptidos e proteínas recombinantes do invento são úteis de modos aqui descritos que os ADNs, ARNs, polipéptidos ou proteínas tal como ocorrem naturalmente o não são.

Noutro modo de execução, o presente invento refere-se a uma molécula de ácidos nucleicos recombinante ou isolada que codifica o polipéptido do invento.

Num modo de execução preferido do presente invento, a molécula de ácidos nucleicos que codifica o polipéptido do invento compreende uma sequência de ácidos nucleicos que apresenta pelo menos 90% de identidade de sequência de ácidos nucleicos à SEQ ID N°: 2. De preferência, a dita molécula de ácidos nucleicos apresenta pelo menos 95% de identidade da sequência de ácidos nucleicos exposta em SEQ ID N°: 2. As moléculas de ácidos nucleicos preferidas que codificam o polipéptido do invento compreendem a mesma sequência nucleotídica exposta em SEQ ID N°: 2.

A molécula de ácidos nucleicos do invento pode ser um ARN, um ADNc ou uma sequência genómica de um tipo misto. Pode, onde apropriado, conter um ou mais intrões, sendo este de origem nativa, heteróloga (por exemplo o intrão dos genes da  $\beta$ -globina de coelho etc.) ou sintética, para aumentar a expressão nas células hospedeiras.

O presente invento refere-se também a um vector que transposta uma molécula de ácidos nucleicos isolada ou recombinante de acordo com o invento.

Tal como é aqui utilizado, o termo "vector" é entendido significar um vector de origem plasmídica ou viral, e facultativamente um tal vector combinado com uma ou mais

substâncias melhora a eficácia de transfecção e/ou a estabilidade do dito vector e/ou a protecção do dito vector *in vivo* para o sistema imunitário do organismo hospedeiro. Estas substâncias encontram-se amplamente documentadas na literatura que está acessível a especialistas na técnica (ver por exemplo Felgner et al., 1987, Proc. West. Farmacol. Soc. 32, 115-121; Hodgson e Solaiman, 1996, Nature Biotechnology 14, 339-342; Remy et al., 1994, Bioconjugate Chemistry 5, 647-654). A título de ilustração mas sem limitação, podem ser polímeros, lípidos, em particular lípidos catiónicos, lipossomas, proteínas nucleares ou virais ou lípidos neutros. Estas substâncias podem ser usadas sozinhas ou em combinação. Exemplos de tais compostos encontram-se disponíveis em particular nos pedidos de patente WO 98/08489, WO, 98/17693, WO 98/34910, WO 98/37916, WO 98/53853, EP 890362 ou WO 99/05183. Uma combinação que pode ser considerada é um vector plasmídico recombinante combinado com lípidos catiónicos conhecidos de um especialista na técnica e uma série deles encontra-se comercialmente disponível, mas é também possível construí-los ou modificá-los por técnicas de engenharia genética. Pode-se mencionar, a título de exemplo, os plasmídeos derivados do pBR322 (Gibco BRL), pUC (Gibco BRL), pBluescript (Stratagene), pREP4, pCEP4 (Invitrogene) ou pPoly (Lathe et al., 1987, Gene 57, 193-201). De preferência, um plasmídeo usado no contexto do presente invento contém uma origem de replicação que assegura a iniciação da replicação num célula produtora e/ou numa célula hospedeira (por exemplo, a origem CoIE1 pode ser seleccionada para um plasmídeo pretendido ser produzido em *E. coli* e o sistema oriP/EBNA1 pode ser seleccionado se for desejado que seja auto-replicativo numa célula hospedeira de mamífero, Lupton e Levine, 1985, Mol. Cell. Biol. 5, 2533-2542; Yates et al., Nature 313, 812-815). Pode compreender elementos adicionais

que melhoram a sua manutenção e/ou a sua estabilidade numa determinada célula (sequência cer que promove a manutenção monomérica de um plasmídeo (Summers e Sherrat, 1984, Cell 36, 1097-1103, sequências para integração, no genoma da célula).

Relativamente a um vector viral, é possível considerar um vector derivado de um poxvírus (vírus *vaccinia*, em particular MVA, canarypox e semelhantes), de um adenovírus, de um retrovírus, de um herpesvírus, de um alfavírus, de um vírus espumoso ou de um vírus adeno-associado. Será utilizado de preferência um vector não-replicativo. A este respeito, os vectores adenovirais são mais particularmente adequados para levar a cabo o presente invento.

De acordo com um modo de execução preferido do invento, o vector de acordo com o invento compreende adicionalmente os elementos necessários para a expressão da molécula de ácidos nucleicos do invento numa célula hospedeira.

Os elementos necessários para a expressão consistem no conjunto de elementos que permitem a transcrição da sequência nucleotídica para o ARN e a tradução do ARNm a um polipéptido, em particular as sequências do promotor e/ou sequências reguladoras que são eficazes na dita célula, e facultativamente as sequências necessárias para permitir a excreção ou a expressão à superfície das células alvo para o dito polipéptido. Estes elementos podem ser reguláveis ou constitutivos. Evidentemente que o promotor é adaptado ao vector seleccionado e à célula hospedeira. Pode-se mencionar, a título de exemplo, os promotores eucarióticos dos genes PGK (Fosfo-Glicerato Cinase), MT (metalotioneína; Mclvor et al., 1987, Mol. Cell Biol. 7, 838-848),  $\alpha$ -1 antitripsina, CFTR, os promotores do gene que codifica a creatina cinase do músculo,

o tensioactivo pulmonar actina, imunoglobulina ou  $\beta$ -actina (Tabin et al., 1982, Mol. Cell Biol. 2, 416-436), SR $\alpha$  (Takebe et al., 1988, Mol. Cell. 8, 466-472), o promotor precoce do vírus SV40 (Vírus de Símio), o LTR do RSV (Vírus do Sarcoma de Rous), o Promotor de MPSV, o promotor de TK-HSV-1, o promotor precoce do vírus CMV (Citomegalovírus), os promotores do vírus vaccinia, p7.5K pH5R, pKIL, p28, p 11 e os promotores adenovirais E1A e MLP ou uma combinação dos ditos promotores. O promotor precoce do Citomegalovírus (CMV) é mais particularmente preferido.

De acordo com um modo de execução preferido, o vector do invento compreende adicionalmente duas sequências que são homólogas com porções contidas numa região de uma sequência de ADN alvo nativa para o genoma de uma célula. A presença das ditas sequências homólogas permite a inserção específica num locus da molécula de ácidos nucleicos do invento na sequência de ADN alvo através de recombinação homóloga.

O termo "recombinação homóloga" refere-se à troca de fragmentos de ADN entre duas moléculas de ADN no locus de sequências de nucleótidos essencialmente idênticas. De preferência, as sequências homólogas no vector são cem por cento homólogas à região da sequência alvo. Contudo, pode ser utilizada menor homologia de sequência. Assim, pode ser utilizada sequência de homologia tão baixo quanto aproximadamente 80%.

As sequências homólogas no vector compreendem pelo menos 25pb, são preferidas regiões mais compridas, de pelo menos 500 pb e mais preferivelmente pelo menos 5000 pb.

De acordo com um modo de execução mais preferido do invento, a molécula de ácidos nucleicos é rodeada pelas sequências homólogas no vector.

Tal como é aqui utilizado “rodeado” significa que uma das sequências homólogas encontra-se localizada a montante da molécula de ácidos nucleicos do invento e que uma das sequências homólogas fica localizada a jusante da molécula de ácidos nucleicos do invento. Tal como é aqui utilizado, “rodeado” não significa necessariamente que as duas sequências homólogas estão directamente ligadas às extremidades 3' ou 5' da molécula de ácidos nucleicos do invento, a molécula de ácidos nucleicos do invento e as sequências homólogas podem ser separadas por um número ilimitado de nucleótidos.

Tal como é aqui utilizado, “uma sequência de ADN alvo” é uma região dentro do genoma de uma célula que é marcada para modificação por recombinação homóloga com o vector. Sequências de ADN alvo incluem genes estruturais (i.e., sequências de ADN que codificam polipéptidos incluindo no caso de eucariotas, intrões e exões), sequências reguladoras tais como sequências facilitadoras, promotores e semelhantes e outras regiões no genoma de interesse. Uma sequência de ADN alvo também pode ser uma sequência que, quando marcada por um vector não tem efeito na função do genoma do hospedeiro.

Tal como é aqui utilizado, “inserido numa sequência de ADN alvo” significa amplamente que o processo de recombinação homóloga que conduz à inserção da molécula de ácidos nucleicos do invento introduz uma deleção ou um rompimento na sequência de ADN marcado.

O especialista na técnica é capaz de escolher as sequências homólogas apropriadas com vista a marcar uma sequência de ADN específica no genoma de uma célula. Por exemplo, uma sequência homóloga pode ser homóloga a uma parte da sequência de ADN marcado, em que a outra sequência homóloga é homóloga a uma sequência de ADN localizada a montante ou a jusante da sequência marcada. De acordo com outro exemplo, uma das sequências homólogas pode ser homóloga a uma sequência de ADN localizada a montante da sequência de ADN marcada, em que a outra sequência homóloga é homóloga a uma sequência de ADN localizada a jusante da sequência de ADN alvo. Noutro exemplo, ambas as sequências homólogas são homólogas a sequências localizadas na sequência de ADN alvo.

De acordo com um modo de execução preferido do invento, a sequência de ADN alvo é o gene HPRT (Hipopoxantina fosforibosil transferase).

A sequência genómica compreendendo o promotor do HPRT e o gene HPRT de *cairina moschata* encontra-se exposto em SEQ ID N°: 3. A sequência que codifica o HPRT começa no codão ATG na posição 8695 da sequência de ácidos nucleicos exposta em SEQ ID N°: 3, a sequência a montante deste codão de ATG é a sequência do promotor do HPRT.

O especialista na técnica é capaz de escolher as sequências homólogas necessárias para a integração da molécula de ácidos nucleicos do invento no gene HPRT. Tal como entre os vários membros de uma família, as sequências genómicas que codificam HPRT são altamente homólogas entre aviários, o especialista na técnica é assim capaz de projectar as sequências homólogas necessárias para marcar o gene HPRT de outras células aviárias.

De acordo com um modo de execução mais preferida do invento, as sequências homólogas são parametrizadas para inserir a molécula de ácidos nucleicos do invento a jusante do promotor do HPRT. Neste modo de execução particular, a molécula de ácidos nucleicos do invento é ligada de modo operacional ao promotor do HPRT endógeno da célula. No contexto do presente invento, pretende-se que “ligado de modo operacional” signifique que a molécula de ácidos nucleicos seja ligada ao promotor de um modo que permite a sua expressão na célula.

De acordo com este modo de execução particular, a sequência homóloga, a montante da molécula de ácidos nucleicos do invento, tem de preferência uma sequência de ácidos nucleicos que é homóloga com pelo menos 500 pb contíguos e mais preferencialmente pelo menos 5000 pb contíguos da sequência de ácidos nucleicos que inicia no nucleótido na posição 1 e termina com o nucleótido na posição 8694 da sequência de ácidos nucleicos exposta em SEQ ID N°: 3, com a condição que a dita sequência homóloga não é homóloga com a sequência de ácidos nucleicos que inicia no nucleótido na posição 8695 e termina com o nucleótido na posição 26916 da sequência de ácidos nucleicos exposta em SEQ ID N°: 3. Além disso, esta sequência homóloga a montante está de preferência directamente ligada ao codão de iniciação da molécula de ácidos nucleicos de acordo com este invento. De acordo com um modo de execução do invento ainda mais preferido, a sequência homóloga a montante da molécula de ácidos nucleicos do invento consiste na sequência de ácidos nucleicos que inicia no nucleótido na posição 1383 e termina com o nucleótido na posição 8694 da sequência de ácidos nucleicos exposta em SEQ ID N°: 3. Por exemplo, o vector de acordo com o invento compreende a sequência de ácidos nucleicos que inicia no

nucleótido na posição 1 e termina com o nucleótido na posição 11227 da sequência de ácidos nucleicos exposta em SEQ ID N°: 4. A sequência homóloga, a jusante da molécula de ácidos nucleicos do invento, tem de preferência uma sequência de ácidos nucleicos que é homóloga com pelo menos 500 pb contíguos e mais preferivelmente pelo menos 5000 pb contíguos da sequência de ácidos nucleicos que inicia no nucleótido na posição 10581 e termina com o nucleótido na posição 17800 da sequência de ácidos nucleicos exposta em SEQ ID N°: 3. E mais preferencialmente, a dita sequência homóloga, a jusante da molécula de ácidos nucleicos do invento, consiste na sequência de ácidos nucleicos que inicia no nucleótido na posição 10581 e termina com o nucleótido na posição 17800 da sequência de ácidos nucleicos exposta em SEQ ID N°: 3.

De acordo com um modo de execução preferido, o vector do invento compreende um primeiro marcador de selecção, em que este primeiro marcador de selecção é um marcador de selecção positivo e em que o dito primeiro marcador de selecção e a molécula de ácidos nucleicos do invento são colocados na mesma secção do vector, sendo a dita secção delimitada pelas sequências homólogas.

Tal como é aqui utilizado, o termo "marcador de selecção positivo" refere-se notavelmente a um gene que codifica um produto que permite que apenas as células que transportam o gene sobrevivam e/ou cresçam sob determinadas condições. Marcadores de selecção típicos codificam proteínas que conferem resistência a antibióticos ou outras toxinas, e.g., ampicilina, neomicina, metotrexato, ou tetraciclina, deficiências auxotóficas de complemento, ou facultar nutrientes críticos não disponíveis de meio complexo. Num modo de execução preferido de acordo com o invento, o

primeiro marcador de selecção codifica uma proteína que confere resistência a antibióticos.

De acordo com um modo de execução do invento mais preferido, o primeiro marcador de selecção, no vector, é rodeado por sequências que permitem a sua supressão. As ditas sequências que permitem a supressão do primeiro marcador de selecção não rodeiam a molécula de ácidos nucleicos do invento. Quando o vector é circular, as sequências que permitem a supressão do primeiro marcador de selecção, o primeiro marcador de selecção e a molécula de ácidos nucleicos do invento são colocados na mesma secção do vector de transferência, sendo a dita secção delimitada pelas sequências homólogas.

As sequências que permitem a supressão de um fragmento de ácido nucleico são bem conhecidas do especialista na técnica (Nunes-Duby, S. et al (1998) *Nucleic Acids Res.* 26:391-406). Estas sequências podem ser reconhecidas por uma ou mais enzimas específicas que induzem a supressão do ácido nucleico compreendido entre as ditas sequências, estas enzimas são designadas "recombinases". Por exemplo, três recombinases bem conhecidas que permitem a supressão de um fragmento de ácido nucleico são as recombinases FLP, ISCEI e Cre.

Uma recombinase locus-específica típica é a recombinase Cre. A Cre é um produto de 38-kDa do gene cre (recombinação por ciclização) do bacteriófago PI e é uma recombinase de ADN locus-específica da família Int. Stemberg, N., et al. (1986) *J. Mol. Biol.* 187: 197-212. A Cre reconhece um locus de 34-pb no genoma PI designado loxP (locus de X-over do PI) e catalisa de modo eficaz a recombinação de ADN conservativa recíproca entre pares de locus loxP. O locus loxP consiste em duas repetições invertidas de 13-pb que flanqueiam uma região

do núcleo 8-pb não palindrômico. A recombinação mediada por Cre entre dois locus loxP directamente repetidos resulta na excisão de ADN entre eles como um círculo covalentemente fechado. A recombinação mediada por Cre entre pares de locus loxP em orientação invertida resultará em inversão do ADN interveniente em vez de excisão. A quebra e junção de ADN são limitadas a posições discretas dentro da região do núcleo e prossegue em cadeia a uma altura por via do acoplamento de ADN fosfotirosina-proteína transiente com a enzima.

Outra recombinase locus-específica é a I-SceI. Outra endonuclease intrão-retorno, por exemplo I-TliI, I-CeuI, I-CreII, I-PpoI e PI-PspI, também podem ser substituídas por I-SceI. Muitos são listados por Belfort e Roberts ((1997) Nucleic Acid Research 25:3379-3388). Muitas destas endonucleases derivam de genomas de organelos nos quais o uso do códon difere do uso do códon nuclear padrão. Usar tais genes para expressão nuclear das suas endonucleases pode ser necessário para alterar a sequência de codificação para emparelhar com os genes nucleares. A I-SceI é uma endonuclease de cadeia dupla que cliva o ADN no seu locus de reconhecimento. A I-SceI gera um corte alternado de 4 pb com saliências 3'OH.

A enzima I-SceI tem um locus de reconhecimento conhecido. O locus de reconhecimento da I-SceI é uma sequência não-simétrica que se estende por mais de 18 pb.

5' TAGGGATAACAGGGTAAT3'  
3' ATCCCTATTGTCCATTAA5

Assim, num modo de execução preferido do invento, as sequências que permitem a supressão do primeiro marcador de selecção compreendem o locus de reconhecimento de I-SceI.

Outra recombinase locus-específica é a recombinase FLP. A recombinase FLP reconhece um locus mínimo de 34-pb distinto que tolera apenas degenerescência limitada da sua sequência de reconhecimento (Jayaram, 1985; Senecoff et al., 1988). A interação entre a recombinase FLP e uma sequência FRT foi examinada (Panigrahi et al., 1992). Exemplos de sequências de FRT variantes são facultadas por Jayaram (1985) e Senecoff et al. (1988), e um ensaio para recombinação mediada por FLP em diferentes substratos é descrito em Snaith et al. (1996).

No modo de execução particular, onde o vector do invento compreende sequências que permitem a supressão do primeiro marcador de selecção, o dito vector pode vantajosamente compreender uma primeira sequência de homologia A e uma segunda sequência de homologia B, em que as sequências de homologia A e B têm um comprimento suficiente e uma homologia suficiente que permite recombinação homóloga entre eles. Em relação às sequências de homologia A e B, "homologia suficiente," refere-se de preferência a sequências com pelo menos 70%, de preferência 80%, ainda preferencialmente pelo menos 90%, especialmente preferível pelo menos 95%, bastante especialmente preferível pelo menos 99%, mais preferível 100%, de homologia nestas sequências de homologia por um comprimento de pelo menos 20 pares de bases, de preferência pelo menos 50 pares de bases, especialmente preferível pelo menos 100 pares de bases, bastante especialmente preferível pelo menos 250 pares de bases, mais preferível pelo menos 500 pares de bases. Neste modo de execução, o vector do invento compreende nas orientações 5'- para 3'- como se segue a molécula de ácidos nucleicos do invento, a primeira sequência de homologia A, uma sequência que permite a supressão do primeiro marcador de selecção, o primeiro marcador de

selecção, uma sequência que permite a supressão do primeiro marcador de selecção e a sequência de homologia B.

De acordo com um modo de execução preferido, o vector do invento compreende um segundo marcador de selecção que não é rodeado pelas ditas sequências homólogas, em que o dito segundo marcador de selecção é um marcador de selecção negativo. O dito segundo marcador de selecção é particularmente útil quando o vector do invento é circular. Quando o vector for circular, o facto que o segundo marcador de selecção não é rodeado pelas ditas sequências homólogas significa que o segundo marcador de selecção e a molécula de ácidos nucleicos do invento não são colocados na mesma secção do vector de transferência, sendo a dita secção delimitada pelas sequências homólogas.

De acordo com um modo de execução preferido do invento, o vector do invento compreende um terceiro marcador de selecção em que o dito terceiro marcador de selecção é um marcador de selecção negativo e em que o dito terceiro marcador de selecção encontra-se situado entre as sequências permitindo a supressão do primeiro marcador de selecção. Quando o vector for circular, o facto de o terceiro marcador de selecção estar situado entre as sequências que permitem a supressão do primeiro marcador de selecção significa que o terceiro marcador de selecção e o primeiro marcador de selecção estão colocados na mesma secção do vector de transferência, sendo a dita secção delimitada pelas sequências permitindo a supressão do primeiro marcador de selecção.

Tal como é aqui utilizado, o termo "marcador de selecção negativo" refere-se notavelmente a um gene que codifica um produto que mata as células que transportam o gene sob certas

condições. Estes genes compreendem notavelmente “gene suicida”. Os produtos codificados por estes genes são capazes de transformar um pró-fármaco num composto citotóxico. Diversos pares de gene suicida/pró-fármaco encontram-se actualmente disponíveis. Podem-se mencionar mais particularmente os pares:

- vírus herpes simplex tipo I timidina cinase (HSV-1 TK) e aciclovir ou ganciclovir (GCV) (Caruso et al., 1993, Proc. Natl. Acad. Sci. USA 90, 7024-7028; Culver et al., 1992, Science 256, 1550-1552; Ram et al., 1997, Nat. Med. 3, 1354-1361,);
- citocromo p450 e ciclofosfofamide (Wei et al., 1994, Human Gene Therapy 5, 969-978);
- purina nucleósido fosforilase de *Escherichia coli* (*E. coli*) e 6-metilpurina desoxirribonucleósido (Sorscher et al., 1994, Gene Therapy 1, 233-238);
- guanina fosforibosiltransferase de *E. coli* e 6-tioxantina (Mozz e Moolten, 1993, Human Gene Therapy, 4, 589-595) e
- citosina desaminase (CDase) e 5-fluorocitosina (5FC).
- FCU1 e 5-fluoro-citosina (5FC) (W09954481).
- FCUI-8 e 5-fluoro-citosina (5FC) (W02005007857)

O primeiro, segundo e terceiro marcadores de selecção podem ser usados separadamente. Por exemplo, o vector do invento pode compreender o primeiro e o terceiro marcador de selecção

mas não o segundo, ou o segundo e o terceiro marcador selecção mas não o primeiro.

De acordo com um modo de execução preferido do invento, o primeiro, o segundo e/ou o terceiro marcador de selecção são colocados sob controlo dos elementos necessários para a sua expressão numa célula hospedeira.

Os elementos necessários para a expressão consistem no conjunto de elementos que permitem a transcrição da sequência nucleotídica para o ARN e a tradução do ARNm a um polipéptido, em particular as sequências do promotor e/ou sequências reguladoras que são eficazes na dita célula, e facultativamente as sequências necessárias para permitir a excreção ou a expressão à superfície das células hospedeiras do dito polipéptido. Estes elementos podem ser reguláveis ou constitutivos. Claro que o promotor é adaptado ao vector seleccionado e à célula hospedeira. Pode-se mencionar, a título de exemplo, os promotores eucariotas dos genes PGK (Fosfo Glicerato Cinase), MT (metalotioneína; Mclvor et al., 1987, Mol. Cell Biol. 7, 838-848),  $\alpha$ -1 antitripsina, CFTR, os promotores do gene que codifica creatina cinase do músculo, tensioactivo pulmonar actina, imunoglobulina ou  $\beta$ -actina (Tabin et al., 1982, Mol. Cell Biol. 2, 416-436), SR $\alpha$  (Takebe et al., 1988, Mol. Cell. 8, 466-472), o promotor precoce do vírus SV40 (Vírus de Símio), o LTR do RSV (Vírus do Sarcoma de Rous), o Promotor de MPSV, o promotor de TK-HSV-1, o promotor precoce do vírus CMV (Citomegalovírus), os promotores do vírus vaccinia, p7.5K p115R, pK1L, p28, p 11 e os promotores adenovirais E1A e MLP ou uma combinação dos ditos promotores. O promotor precoce do Citomegalovírus (CMV) é mais particularmente preferido.

O presente invento refere-se também a uma célula aviária transfectada por uma molécula de ácidos nucleicos ou um vector de acordo com o invento e células aviárias suas derivadas. Tal como é aqui utilizado, o termo “derivado” refere-se a células que desenvolvem ou diferenciam de ou têm como antepassado uma célula transfectada por uma molécula de ácidos nucleicos de acordo com o invento. De preferência, a dita molécula de ácidos nucleicos é inserida no gene HPRT das ditas células primárias.

O presente invento refere-se também ao uso dos polipéptidos, moléculas de ácidos nucleicos e a vectores de acordo com o invento para a imortalização de uma célula aviária.

Também é descrita uma célula aviária que compreende a molécula de ácidos nucleicos do invento, em que a dita molécula de ácidos nucleicos encontra-se ligada de modo operacional ao promotor do HPRT endógeno da célula aviária. Pretende-se que “ligado de modo operacional” signifique que a molécula de ácidos nucleicos está ligada ao promotor de um modo que permite a sua expressão na célula. Num modo de execução preferido, a célula aviária de acordo com o invento compreende a sequência de ácidos nucleicos exposta em SEQ ID N°: 4.

O presente invento refere-se também a um processo para imortalizar uma célula aviária compreendendo o passo de transfectar um vector de acordo com o invento na dita célula aviária.

Uma célula imortalizada, tal como é aqui utilizado, refere-se a uma célula capaz de crescer em cultura durante mais de 35 passagens.

O termo número de passagens refere-se ao número de vezes que uma população celular foi removida do recipiente de cultura e foi submetida um processo de subcultura (passagem) para manter as células numa densidade suficientemente baixa para estimular crescimento adicional.

Tal como é aqui utilizado, o termo “transfectado” refere-se à transfecção estável ou à transfecção transiente da célula aviária do invento.

O termo “transfecção estável” ou “transfectado de modo estável” refere-se à introdução e integração de ADN estranho no genoma da célula transfectada. O termo “transfectante estável” refere-se a uma célula que integrou de modo estável ADN estranho no ADN genómico.

O termo “transfecção transiente” ou “transientemente transfectado” refere-se à introdução de ADN estranho numa célula onde o ADN estranho falha na integração do genoma da célula transfectada. O ADN estranho persiste no núcleo da célula transfectada durante vários dias. O termo “transfectante transiente” refere-se a células que captaram ADN estranho mas falham na integração deste ADN.

De acordo com um modo de execução preferido do invento, a célula aviária do invento deriva de uma célula da família Anatidae ou da família Phasianidae. Entre a Anatidae, as células que pertencem ao género Cairina ou Anas são particularmente preferidas. Ainda mais preferencialmente as células de acordo com o invento pertencem às espécies Cairina moschata ou Anas platyrhynchos.

De preferência, a célula aviária de acordo com o invento deriva de um organismo embrionário. Os métodos que permitem o isolamento de células de um organismo vivo são bem conhecidos do especialista na técnica. Por exemplo, podem ser usados métodos descritos no exemplo 2. De acordo com um modo de execução preferido do invento, a célula aviária primária é isolada de um embrião que pertence à família Anatidae que tem entre 0 e 20 dias de idade, mais preferivelmente entre 5 e 15 dias de idade e ainda mais preferencialmente entre 11 e 14 dias de idade.

O vector usado no processo do invento pode compreender um primeiro marcador de selecção. A integração do primeiro marcador de selecção permite a selecção das células que incorporaram a molécula de ácidos nucleicos do invento. Em conformidade, o processo de acordo com o invento pode compreender adicionalmente um passo em que as ditas células aviárias são cultivadas num meio que apenas permite o crescimento das células que incorporaram o primeiro marcador de selecção. Por exemplo num meio que compreende um antibiótico.

Quando o vector usado no processo do invento compreende sequências que permitem a supressão do primeiro marcador de selecção, o processo de acordo com o invento pode compreender adicionalmente um passo que consiste em suprimir o primeiro marcador de selecção do genoma da dita célula aviária primária. Para suprimir o dito primeiro marcador de selecção, a célula aviária é transfectada pelo gene que codifica a recombinase específica para as sequências que permitem a supressão do primeiro marcador de selecção. Métodos e vectores capazes de transferir o dito gene na célula são bem

conhecidos do especialista na técnica, por exemplo, pode ser usado o método descrito no exemplo 4 do presente pedido.

Quando o vector usado no processo do invento compreende um segundo marcador de selecção, o processo de acordo com o invento pode adicionalmente compreender um passo em que as células aviárias são cultivadas num meio que apenas permite o crescimento das células que não incorporaram o segundo marcador de selecção. O dito passo pode ser realizado simultaneamente com ou separadamente do passo em que as ditas células aviárias são cultivadas num meio que apenas permite o crescimento das células que incorporaram o primeiro marcador de selecção.

O dito segundo marcador de selecção é particularmente útil quando o vector, usado no processo de acordo com o invento, é circular. A presença do dito segundo marcador de selecção permite a destruição das células aviárias nas quais o processo de recombinação homóloga conduz à introdução da secção do vector de transferência que não compreende a molécula de ácidos nucleicos do invento.

Quando o vector usado no processo do invento compreende um terceiro marcador de selecção, o processo de acordo com o invento pode adicionalmente compreender um passo no qual a dita célula aviária é cultivada num meio que não permite o crescimento das células que compreendem o terceiro marcador de selecção. Por exemplo, um meio que não permite o crescimento das células que compreendem FCU1 como um terceiro marcador de selecção, compreende 5-Fluorocitosina.

Este passo permite a selecção das células aviárias nas quais ocorreu a supressão do primeiro marcador de selecção. Tal

significa que o passo que consiste na supressão do primeiro marcador de selecção também conduzirá à supressão do terceiro marcador de selecção. A presença do terceiro marcador de selecção permite a destruição das células aviárias nas quais o primeiro marcador de selecção está presente.

O presente invento refere-se mais particularmente, mas não está limitado a um processo para imortalizar uma célula aviária compreendendo os passos de:

- transferir para a célula aviária um vector compreendendo:
  - uma molécula de ácidos nucleicos de acordo com o invento rodeada por sequências homólogas.
  - Um primeiro marcador de selecção em que o dito primeiro marcador de selecção é um marcador de selecção positivo e em que o dito primeiro marcador de selecção é rodeado pelas ditas sequências homólogas.
  - Sequências que permitem a supressão do primeiro marcador de selecção.
  - Um segundo marcador de selecção que não é rodeado pelas ditas sequências homólogas, em que o dito marcador de selecção é um marcador de selecção negativo.
  - Um terceiro marcador de selecção em que o dito terceiro marcador de selecção é um marcador de selecção negativo e em que o dito terceiro marcador de selecção encontra-se localizado entre as sequências que permitem a supressão do primeiro marcador de selecção.

- cultivar as ditas células aviárias num meio que apenas permite o crescimento das células que incorporaram o primeiro marcador de selecção.
- cultivar as ditas células aviárias num meio que não permite o crescimento das células que incorporaram o segundo marcador de selecção.
- excluir o primeiro marcador de selecção do genoma da dita célula aviária.
- cultivar a dita célula aviária num meio que não permite o crescimento das células que compreendem o terceiro marcador de selecção.

Num modo de execução particularmente preferida, o invento refere-se a uma célula aviária imortalizada que deriva duma célula de um animal que pertence à espécie *Cairina moschata* e que compreende a transcriptase inversa da telomerase de *Cairina moschata*, sob controlo do promotor de HPRT de *Cairina moschata*, inserido no gene HPRT da célula.

A célula aviária de acordo com o invento pode adicionalmente compreender uma ou mais sequências de ácidos nucleicos que permitem a propagação de um vírus defectivo. "Vírus defectivo" refere-se a um vírus no qual um ou mais genes virais necessários para a sua replicação são deletados ou tornam-se não funcionais. O termo "sequência de ácidos nucleicos que permite a propagação de um vírus defectivo" refere-se a uma sequência de ácidos nucleicos que faculta em trans a(s) função(ões) que permite a replicação do vírus defectivo. Noutras palavras, a(s) dita(s) sequência(s) de ácidos nucleicos codifica a(s) proteína(s) necessária(s) para

a replicação e encapsidação do dito vírus defectivo. A título de ilustração, para a produção de um vector adenoviral, a que falta a maior parte da região El, a célula aviária de acordo com o invento pode ser transfectada transientemente ou permanentemente com uma sequência de ácidos nucleicos que codifica a região El.

A célula aviária de acordo com o invento também pode compreender uma sequência de ácidos nucleicos que codifica uma substância de interesse. Neste caso, a dita célula aviária de acordo com o invento pode ser usada para a produção da dita substância de interesse. Tal como é aqui utilizado, uma substância de interesse pode incluir, mas não está limitada a, uma proteína farmacologicamente activa, por exemplo factores de crescimento, reguladores de crescimento, anticorpos, antigénios, os seus derivados úteis para imunização ou vacinação e semelhantes, interleucinas, insulina, G-CSF, GM-CSF, hPG-CSF, M-CSF ou combinações dos mesmos, interferões, por exemplo, interferão- $\alpha$ , interferão- $\beta$ , interferão  $\gamma$ , factores de coagulação do sangue, por exemplo, Factor VIII, Factor IX, ou tPA ou combinações dos mesmos. "Substância de interesse" refere-se também a enzimas industriais, por exemplo para uso na pasta de papel ou papel, modificação têxtil, ou produção de etanol. Finalmente, "substância de interesse" refere-se também a suplemento proteico ou um produto de valor acrescentado para alimentação animal.

As células aviárias obtidas pelo processo de acordo com o invento, a célula aviária do invento e as células aviárias suas derivadas são notavelmente úteis para a replicação de um vírus vivo, atenuado, ou recombinante. Mais preferivelmente, as ditas células são particularmente úteis para a replicação

de poxvírus (vírus vaccinia, em particular MVA, poxvírus de canário, etc.), um adenovírus, um retrovírus, um herpesvírus, um alfavírus, um vírus espumoso ou de um vírus associado ao adenovírus.

Os retrovírus têm a propriedade de infectar, e na maioria dos casos integrar em células em divisão e a este respeito são particularmente apropriados para uso em relação ao cancro. Um retrovírus recombinante de acordo com o invento contém geralmente as sequências LTR, uma região de encapsidação e a sequência de nucleótidos de acordo com o invento, que é colocado sob controlo do LTR retroviral ou de um promotor interno tais como os abaixo descritos. Um vector retroviral pode conter modificações, em particular nos LTRs (substituição da região do promotor com um promotor eucariótico) ou a região de encapsidação (substituição com uma região de encapsidação heteróloga, por exemplo o tipo VL30) (ver pedidos Franceses 94 08300 e 97 05203).

Ao vector adenoviral pode faltar a totalidade ou parte de pelo menos uma região que é essencial para replicação e que é seleccionada das regiões E1, E2, E4 e L1 L5. Uma deleção da região E1 é preferida. Contudo, pode ser combinada com outra(s) modificação(ões)/deleção(ões) afectando, em particular, a totalidade ou parte das regiões E2, E4 e/ou L1 L5. A título de ilustração, a deleção da parte principal da região E1 e da unidade de transcrição E4 é bastante particularmente vantajosa. Com a finalidade de aumentar as capacidades de clonagem, ao vector adenoviral pode adicionalmente faltar a totalidade ou parte da região não essencial E3. De acordo com outra alternativa, é possível fazer uso de um vector adenoviral mínimo que retém as sequências que são essenciais para encapsidação,

designadamente os ITRs (Repetição Terminal Invertida) 5' e 3', e a região de encapsidação. Os vários vectores adenovirais, e as técnicas para os preparar, são conhecidos (ver, por exemplo, Graham e Preveet, 1991, em *Methods in Molecular Biology*, Vol 7, p 109 128; Ed: E. J. Murey, The Human Press Inc).

A família dos poxvirus compreende vírus das subfamílias Chordopoxvirus e Entomopoxvirus. Entre estes, o poxvirus de acordo com o invento é preferivelmente escolhido do grupo que compreende Orthopoxviruses, Parapoxviruses, Avipoxviruses, Capripoxviruses, Leporipoxviruses, Suipoxviruses, Molluscipoxviruses, Yatapoxviruses. De acordo com um modo de execução mais preferida, o poxvirus do invento é um orthopoxvirus.

O Orthopoxvirus é de preferência um vírus vaccinia e mais preferivelmente um vírus vaccinia Ankara modificado (MVA) em particular MVA 575 (ECACC V00120707) e MVA-BN (ECACC V00083008).

O termo "vírus recombinante" refere-se a um vírus que compreende uma sequência exógena inserida no seu genoma. Tal como é aqui utilizado, uma sequência exógena refere-se a um ácido nucleico que não está naturalmente presente no vírus progenitor.

Num modo de execução, a sequência exógena codifica uma molécula que tem uma função citotóxica directa ou indirecta. Por citotóxica "directa ou indirecta", queremos dizer que a molécula codificada pela sequência exógena pode ela própria ser tóxica (por exemplo ricina, factor de necrose tumoral, interleucina-2, interferão-gama, ribonuclease, desoxirri-

bonuclease, Exotoxina A de *Pseudomonas*) ou pode ser metabolizada para formar um produto tóxico, ou pode actuar em qualquer outra coisa para formar um produto tóxico. A sequência do ADNc da ricina é divulgada em Lamb et al (Eur. J. Biochem., 1985, 148, 265-270).

Num modo de execução preferido do invento, a sequência exógena é um gene suicida. Um gene suicida codifica uma proteína capaz de converter um pró-fármaco relativamente não-tóxico a um fármaco tóxico. Por exemplo, a enzima citosina desaminase converte 5-fluorocitosine (5FC) a 5-fluorouracil (5FU) (Mullen et al (1922) PNAS 89, 33); a enzima de herpes simplex timidina cinase sensibiliza células a tratamento com o agente antiviral ganciclovir (GCV) ou aciclovir (Moolten (1986) Cancer Res., 46, 5276; Ezzedine et al (1991) New Biol 3, 608). Pode ser utilizada a citosina desaminase de qualquer organismo, por exemplo, *E. Coli* ou *Saccharomyces cerevisiae*.

Assim, num modo de execução mais preferido do invento, o gene codifica uma proteína que tem uma actividade de citosina desaminase e ainda mais preferivelmente uma proteína como descrito nos pedidos de patente W02005007857 e W09954481

Num modo de execução adicional o gene exógeno codifica uma ribozima capaz de clivar ARN ou ADN marcado. O ARN ou ADN marcado para ser clivado pode ser ARN ou ADN que é essencial à função da célula e a sua clivagem resulta em morte celular ou o ARN ou ADN a ser clivado podem ser ARN ou ADN que codifica uma proteína indesejável, por exemplo um produto de oncogene, e a clivagem deste ARN ou ADN pode impedir que a célula se torne cancerosa.

Ainda num modo de execução adicional o gene exógeno codifica um ARN anti-sentido.

Por “ARN anti-sentido” queremos dizer uma molécula de ARN que hibridiza a, e interfere com a expressão de uma molécula de ARNm que codifica uma proteína ou a outra molécula de ARN dentro da célula tal como pre-ARNm ou ARNt ou ARNr, ou hibridiza, e interfere com a expressão de um gene.

Noutro modo de execução do invento, a sequência exógena substitui a função de um gene defectivo numa célula alvo. Existem diversos milhares de doenças genéticas hereditárias de mamíferos, incluindo humanos, que são causadas por genes defectivos. Exemplos de tais doenças genéticas incluem fibrose quística, onde se sabe haver uma mutação no gene CFTR; distrofia muscular de Duchenne onde se sabe haver uma mutação no gene distrofina; drepanocitose onde se sabe haver uma mutação no gene HbA. Muitos tipos de cancro são causados por genes defectivos, especialmente protooncogenes, e anti-oncogenes que sofreram mutação.

Exemplos de protooncogenes são o ras, src, bcl etc; exemplos de anti-oncogenes são o p53 e Rb.

Num modo de execução adicional do invento, a sequência exógena codifica Antigénio Associado a Tumor (TAA). TAA refere-se a uma molécula que é detectada a uma frequência ou densidade mais elevada em células tumorais do que em células não-tumorais do mesmo tipo de tecido. Exemplos de TAA incluem mas não estão limitadas a CEA, MART-1, MAGE-1, MAGE-3, GP-100. MUC-1, MUC-2, oncogene ras mutado pontualmente, p53 normal ou mutado pontualmente, p53 sobre-expresso, CA-125, PSA, C-erb/B2, BRCA I, BRCA II, PSMA, tirosinase, TRP-1, TRP-

2, NY-ESO-1, TAG72, KSA, HER-2/neu, bcr-abI, pax3-fkhr, ews-fli-1, sobrevivente e LRP. De acordo com um modo de execução mais preferido o TAA é MUC 1.

O vírus recombinante pode compreender mais do que uma sequência exógena e cada sequência exógena pode codificar mais do que uma molécula. Por exemplo, pode ser útil para associar num mesmo poxvírus recombinante, uma sequência exógena que codifica um TAA com uma sequência exógena que codifica uma citocina.

Noutro modo de execução do invento, o gene exógeno codifica um antígeno. Tal como é aqui utilizado, "antígeno" refere-se a um ligando que pode ser ligado por um anticorpo; um antígeno não necessita de ser imunogénico.

De preferência o antígeno é derivado de um vírus tal como por exemplo VIH-1, (tal como gp 120 ou gp 160), qualquer Vírus da imunodeficiência felino, humano ou vírus de herpes animal, tais como gD ou seus derivados ou proteína precoce imediata tal como ICP27 de HSV1 ou HSV2, citomegalovírus (tal como gB ou seus derivados), Vírus Varicela Zoster (tais como gpI, II ou III), ou de um vírus da hepatite tal como vírus da hepatite B por exemplo antígeno de Superfície da Hepatite B ou um seu derivado, vírus da hepatite A, vírus da hepatite C (de preferência proteína não estrutural do genótipo 1b estirpe ja) e vírus da hepatite E, ou de outros patógenos virais, tais como Vírus Sincicial Respiratório, Vírus do papiloma humano (de preferência a proteína E6 e E7 da estirpe HPV16) ou vírus influenza, ou derivados de patógenos bacterianos tais como Salmonela, Neisseria, Borrelia (por exemplo OspA ou OspB ou seus derivados), ou Chlamidia, ou

Bordetella por exemplo P.69, PT e FHA, ou derivados de parasitas tais como plasmódio ou Toxoplasma.

Figura 1: Vector que compreende um gene que codifica a transcriptase inversa da telomerase.

Figura 2: Representação esquemática da inserção locus específica do gene que codifica a transcriptase inversa da telomerase no gene HPRT.

Figura 3: Representação esquemática da eliminação do primeiro e terceiro marcadores de selecção do genoma da célula imortalizada obtida pelo processo do invento

### **Exemplos:**

#### **Exemplo 1: Sistema de expressão da telomerase**

Inserção aleatória

Utilizou-se para esta finalidade um plasmídeo que não partilha qualquer sequência específica de homologia com o genoma de pato (Figura 1).

Inserção marcada

Construiu-se um plasmídeo compreendendo dois fragmentos de 5kb homólogos ao gene HPRT de *Cairina moschata* que rodeia o gene da transcriptase inversa da telomerase de *cairina moschata* e dois marcadores de selecção. Seleccionou-se o gene HPRT que codifica a hipoxantina guanina fosforil transferase como um sítio adequado para a expressão constitutiva da telomerase de *cairina moschata*.

Estes dois marcadores de selecção são o gene FCU1 (Erbs et al. Cancer Res. 2000. 15. 60: 3813-22) sob controlo de um promotor do CMV (Thomsen et al. P.N.A.S. 1984. 81. 3: 659-63) e o gene de resistência à Neomicina colocado sob controlo de um promotor do SV40. A resistência à Neomicina e a cassete de expressão de FCU-1 são rodeados por locus de clivagem SclI que permitem a eliminação das cassetes de selecção da linha celular final. Fora dos braços do gene HPRT é inserido um marcador de selecção que codifica o HSVTK conduzido por um promotor do RSV (Figura 2).

Exemplo 2: Preparação de lote de CEC de ovos de Cairina moschata com 12 dias e descrição de subpopulações.

Incubam-se 25 ovos SPF fertilizados a 37,5°C. Abrem-se os ovos após 12 dias de incubação seguindo protocolo disponível.

Picam-se 23 embriões, lavam-se uma vez em Tampão de Fosfato da Dulbecco (PBS) e dissocia-se em TrypLE Select (Invitrogen) 5 horas à temperatura ambiente.

Após centrifugação a baixa velocidade ressuspendem-se as células em Meio Basal da Eagle (MBE) suplementado com 10% de soro fetal de bezerro (FCS), gentamicina 0,04 g/L, semeadas em frascos triplos de 500 cm<sup>2</sup> e incubam-se a 37°C 5% CO<sub>2</sub>.

Após 24h removem-se as células confluentes dos frascos usando TrypLE Select (5 mL/frasco triplo), volta-se a semear parte das células em frascos de 175 cm<sup>2</sup> para segunda passagem. Concentraram-se as restantes células a 10<sup>7</sup> células/mL em meio adequado (60% BME, 30% FCS e 10% DMSO) e congelam-se num recipiente regulado com álcool isopropílico (Recipiente NALGENE®, "Mr. Frosty" 1°C freezing) a -80°C antes de

transferir para azoto líquido para armazenamento a longo prazo, constituindo o banco celular inicial ( $50 \times 1,5 \cdot 10^7$  células/frasco,  $44 \times 1 \cdot 10^7$  células/frasco).

As células que permanecem em cultura são submetidas classicamente até 18 passagens, durante as 3 primeiras passagens as células não ligadas são recolhidas por centrifugação a baixa rotação do meio condicionado, novamente semeadas e passam adicionalmente do mesmo modo que a cultura inicial.

As subpopulações que exibem aspectos morfológicos diferentes característicos, foram reprodutivelmente isoladas durante a vida da cultura.

### Exemplo 3: Métodos de transfecção.

São conhecidos na arte um grande número de métodos de transfecção para introduzir um vector capaz de dirigir expressão de uma sequência de nucleótidos de interesse. Uma lista não limitativa destes métodos é apresentada a seguir: precipitação por  $\text{CaPO}_4$ , electroporação, método de transfecção por lipofectina. Um exemplo facultado é baseado no procedimento de precipitação por  $\text{CaPO}_4$ .

As células devem ter uma confluência de cerca de 80-90%. O meio é alterado duas horas antes da adição de  $\text{CaPO}_4/\text{ADN}$ . As 30  $\mu\text{g}$  de ADN são ressuspendidas em 31  $\mu\text{l}$  2M  $\text{CaCl}_2$  - 161,3 mM Tris pH 7,6. Adiciona-se  $\text{H}_2\text{O}$  para perfazer um volume final de 0,5 mL.

Por transfecção, distribuem-se 0,5 mL de 2X HEBS em tubos Falcon esterilizados de 15 mL e adiciona-se a solução de ADN

gota a gota enquanto se submete cuidadosamente a vortex ou borbulhação a solução de ADN. A solução deve ficar leitosa. Deixa-se que a solução repouse à temperatura ambiente durante 10-30 min. Em seguida pipeta-se uma vez para dentro e para fora com pipeta esterilizada em cultura de tecido para quebrar flocos e aplica-se gota a gota às células. Incubam-se em seguida as células entre 6 horas até durante a noite a 37°C. Um precipitado fino deve cobrir a superfície celular. Para completar o procedimento de transfecção aquecer até 37°C a solução choque de glicerol. Aspira-se o meio, adicionam-se 5 mL de BME para lavar a camada celular, aspira-se em seguida o meio e adiciona-se 1 mL de solução de choque de glicerol durante 2 minutos ou menos. Subsequentemente adicionam-se lentamente 10 mL de BME para diluir o glicerol e remove-se completamente o BME-glicerol. Adiciona-se em seguida 10 mL de meio desejado e incubam-se placas à temperatura adequada.

#### Exemplo 4: Métodos de selecção.

Inserção aleatória:

Aplica-se pressão de selecção 48 horas após transfecção: dissociam-se as células com TrypLE select, centrifugam-se a baixa rotação e voltam-se a semear em BME com FCS 10%, e G418 800 µg/mL.

Passam-se as células em série até poderem ser isolados clones individuais em crescimento. Isolam-se os focos em multiplicação e amplificam-se antes da quantificação de actividade da telomerase com o estojo de detecção de telomerase TRAPeze@XL (S7707, Chemicon) e análises de manchas a sul para estabelecer a integração no locus específico marcado.

Inserção marcada:

Aplica-se pressão de selecção 48 horas após transfecção: dissociam-se as células com TrypLE select, centrifugam-se a baixa rotação e voltam-se a semear em BME com FCS 10%, Ganciclovir 25 µg/mL, e G418 800 µg/mL.

Passam-se as células em série até poderem ser isolados clones individuais em crescimento. Isolam-se os focos em multiplicação e amplificam-se antes da quantificação de actividade da telomerase com o estojo de detecção de telomerase TRAPeze®XL (S7707, Chemicon) e análises de manchas a sul de oligos para estabelecer a integração no locus específico marcado.

Os clones celulares com actividade em telomerase restaurada detectada e integração do locus HPRT marcado são subsequentemente transfectados um plasmídeo de expressão da meganuclease I-SceI seguindo o método a seguir em baixo.

Para seleccionar a eliminação dos marcadores de selecção aplica-se 5-fluorocitosina (5-FC) 48 horas após transfecção: dissociam-se as células com TrypLE select, centrifugam-se a baixa rotação e voltam-se a semear em meio com 5-FC a concentração que varia de  $10^{-3}$  a  $10^{-7}$  M mantém-se selecção de G418 (BME com FCS 10%, 5-FC, e G418800 µg/mL).

#### LISTAGEM DAS SEQUÊNCIAS

<110> TRANSGENE SA

ERBS Philippe

BALLOUL Jean-Marc

KAPFER Marina

SILVESTRE Nathalie

<120> Transcriptase inversa da telomerase aviária

<130> TG167PCT

<150> EP 06360001.9

<151> 2006-01-05

<150> EP 06360047.2

<151> 2006-10-13

<160> 4

<170> Patent In versão 3.3

<210> 1

<211> 1304

<212> PRT

<213> Cairina moschata

<400> 1

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Ala | Gly | Ala | Glu | Pro | Phe | Gly | Ala | Val | Leu | Gly | Ala | Leu | Arg | Asp | 1   | 5   | 10  | 15  |
| Cys | Tyr | Ala | Gln | Ala | Ala | Pro | Leu | Glu | Thr | Phe | Leu | Arg | Gly | Leu | Gly | 20  | 25  | 30  |     |
| Glu | Ser | Gly | Ala | Glu | Glu | Ala | Glu | Val | Val | Arg | Asp | Asp | Asp | Ala | Ala | 35  | 40  | 45  |     |
| Cys | Tyr | Arg | Thr | Phe | Val | Ser | Gln | Cys | Val | Val | Cys | Val | Pro | His | Gly | 50  | 55  | 60  |     |
| Ala | Arg | Asp | Ile | Pro | Arg | Pro | Phe | Ser | Leu | Glu | Gln | Leu | Ser | Ser | Gln | 65  | 70  | 75  | 80  |
| Ser | Glu | Val | Ile | Ser | Arg | Val | Met | Gln | Arg | Leu | Cys | Gly | Lys | Lys | Lys | 85  | 90  | 95  |     |
| Lys | Asn | Ile | Leu | Thr | Tyr | Gly | Tyr | Ser | Leu | Leu | Asp | Glu | Asn | Ser | Ser | 100 | 105 | 110 |     |
| His | Phe | Gln | Ile | Met | Pro | Leu | Ser | Asn | Val | Tyr | Ser | Tyr | Leu | Pro | Asn | 115 | 120 | 125 |     |
| Thr | Ala | Thr | Glu | Thr | Met | Arg | Ile | Ser | Gly | Leu | Trp | Glu | Thr | Leu | Leu | 130 | 135 | 140 |     |
| Ser | Arg | Ile | Gly | Asp | Asp | Val | Met | Met | Tyr | Leu | Leu | Glu | His | Cys | Ala | 145 | 150 | 155 | 160 |
| Ile | Phe | Met | Leu | Val | Pro | Pro | Ser | Asn | Cys | Tyr | Gln | Val | Cys | Gly | Gln | 165 | 170 | 175 |     |
| Pro | Ile | Tyr | Glu | Leu | Ile | Ser | Gln | Asn | Val | Glu | Ser | Ala | Pro | Ala | Phe | 180 | 185 | 190 |     |
| Val | Lys | Gln | Arg | Leu | Ser | Lys | His | Lys | Arg | Ser | Ser | Leu | Leu | Lys | Tyr | 195 | 200 | 205 |     |
| Thr | Gln | Lys | Arg | Leu | Thr | Phe | His | Arg | Gln | Tyr | Leu | Ser | Lys | Ser | Arg | 210 | 215 | 220 |     |
| Gln | Ser | Lys | Arg | Arg | Gln | Arg | Leu | Glu | Ala | Asn | Val | Ser | Ser | Met | Arg | 225 | 230 | 235 | 240 |
| Asn | Lys | Thr | Ser | Asn | Asn | Ile | Gln | Ser | Leu | Gly | Ser | Ala | Ala | Leu | Glu |     |     |     |     |



Lys Asn Met Leu Phe Tyr Tyr Arg Lys Cys Ile Trp Ala Lys Leu Gln  
 740 745 750  
 Asp Ile Gly Ile Arg Lys His Phe Ala Lys Val Gln Leu Arg Pro Leu  
 755 760 765  
 Thr Ala Glu Glu Met Glu Ala Ile His Gln Lys Lys Tyr Leu Pro Met  
 770 775 780  
 Ala Ser Lys Leu Arg Phe Ile Pro Lys Val Ser Gly Leu Arg Pro Ile  
 785 790 795 800  
 Val Arg Met Ser Gly Val Val Glu Ala Gln Thr Leu Ser Lys Glu Ser  
 805 810 815  
 Arg Ala Lys Lys Met Asn His Tyr Asn Met Gln Leu Lys Asn Leu Phe  
 820 825 830  
 Ser Val Leu Asn Tyr Glu Arg Thr Ile Asn Thr Ser Tyr Ile Gly Ser  
 835 840 845  
 Ser Val Phe Gly Arg Asp Asp Ile Tyr Lys Lys Trp Lys Thr Phe Val  
 850 855 860  
 Lys Lys Val Leu Lys Ser Asp Gly Glu Ile Pro His Phe Tyr Tyr Val  
 865 870 875 880  
 Lys Ala Asp Val Ser Arg Ala Phe Asp Ser Ile Pro His Asp Lys Leu  
 885 890 895  
 Val Glu Val Ile Ser Gln Val Leu Lys Pro Glu Lys Lys Thr Val Tyr  
 900 905 910  
 Cys Ile Arg Arg Tyr Ala Val Val Met Ile Thr Gly Ser Gly Lys Thr  
 915 920 925  
 Arg Lys Leu Tyr Arg Arg His Val Ser Thr Phe Lys Asp Phe Met Pro  
 930 935 940  
 Asp Met Lys Gln Phe Val Ser Arg Leu His Glu Ser Thr Ser Leu Arg  
 945 950 955 960  
 Asp Ala Ile Ile Val Glu Gln Ser Leu Thr Phe Asn Glu Thr Ser Ala  
 965 970 975  
 Ser Leu Phe Asn Phe Phe Leu Gln Met Leu Asn Asn Asn Ile Leu Glu  
 980 985 990  
 Ile Glu Arg Ser Tyr Tyr Leu Gln Cys Ser Gly Ile Pro Gln Gly Ser  
 995 1000 1005  
 Leu Leu Ser Thr Leu Leu Cys Ser Leu Cys Tyr Gly Asp Met Glu  
 1010 1015 1020  
 Asn Lys Leu Phe Ser Gly Val Gln Lys Asp Gly Val Leu Ile Arg  
 1025 1030 1035  
 Leu Ile Asp Asp Phe Leu Leu Val Thr Pro His Leu Thr His Ala  
 1040 1045 1050  
 Arg Thr Phe Leu Arg Thr Leu Ala Met Gly Ile Pro Glu Tyr Gly  
 1055 1060 1065  
 Phe Leu Ile Asn Pro Lys Lys Thr Val Val Asn Phe Ser Val Asp  
 1070 1075 1080  
 Asp Ile Pro Glu Cys Ser Glu Phe Lys Gln Leu Pro Asn Cys Arg  
 1085 1090 1095  
 Leu Ile Pro Trp Cys Gly Leu Leu Leu Asp Thr Gln Thr Leu Glu  
 1100 1105 1110  
 Val Tyr Cys Asp Tyr Ser Ser Tyr Ser Cys Thr Ser Ile Arg Ser  
 1115 1120 1125  
 Ser Leu Ser Phe Asn Ser Asn Arg Thr Ala Gly Lys Asn Met Lys  
 1130 1135 1140  
 His Lys Leu Val Ala Val Leu Lys Leu Lys Cys His Gly Leu Phe  
 1145 1150 1155  
 Leu Asp Leu Gln Ile Asn Ser Val Lys Thr Val Phe Ile Asn Val  
 1160 1165 1170  
 Tyr Lys Ile Phe Leu Leu Gln Ala Tyr Arg Phe His Ala Cys Val  
 1175 1180 1185  
 Ile Gln Leu Pro Phe Asn Gln Lys Val Arg Asn Asn Pro Asp Phe  
 1190 1195 1200  
 Phe Leu Arg Val Ile Ala Glu Asn Ala Ser Cys Cys Tyr Ser Met

|         |                     |                 |                 |      |
|---------|---------------------|-----------------|-----------------|------|
| 1205    |                     | 1210            |                 | 1215 |
| Leu Lys | Ala Lys Asn Pro Gly | Phe Thr Leu Gly | Asn Arg Gly Ala |      |
| 1220    |                     | 1225            |                 | 1230 |
| Ser Gly | Met Phe Pro Ser Glu | Ala Ala Glu Trp | Leu Cys Tyr His |      |
| 1235    |                     | 1240            |                 | 1245 |
| Ala Phe | Thr Val Lys Leu Ser | Asn His Lys Val | Val Tyr Lys Cys |      |
| 1250    |                     | 1255            |                 | 1260 |
| Leu Leu | Lys Pro Leu Lys Phe | Cys Met Thr Gln | Leu Phe Arg Lys |      |
| 1265    |                     | 1270            |                 | 1275 |
| Ile Pro | Lys Asp Thr Lys Ala | Leu Leu Lys Thr | Val Thr Glu Pro |      |
| 1280    |                     | 1285            |                 | 1290 |
| Ser Ile | Cys Gln Asp Phe Lys | Ala Ile Leu Asp |                 |      |
| 1295    |                     | 1300            |                 |      |

<210> 2

<211> 3915

<212> ADN

<213> Cairina moschata

<400> 2

|             |             |             |             |             |              |      |
|-------------|-------------|-------------|-------------|-------------|--------------|------|
| atggcgggcg  | cggagccctt  | cggcgccgtg  | ctggggcgcc  | tgcgggactg  | ctacgcgcag   | 60   |
| gcggccccgc  | tggagacctt  | cctccggggg  | ctggggggaga | gcgggcgccga | ggaagccgag   | 120  |
| gtggtgcggg  | acgacgacgc  | cgcttgcctac | cgcaccttgc  | tgtcccagtg  | cgtgggtgtg   | 180  |
| gtccccccacg | gcgcccgcga  | catcccccg   | cccttcagct  | tggagcagtt  | atctagtcag   | 240  |
| agcgaagtca  | tctcaagagt  | catgcagagg  | ctgtgtggga  | aaaagaagaa  | gaacatcctc   | 300  |
| acatatggat  | actccttgct  | ggatgaaaac  | agttctcact  | tccaaatcat  | gccgctctca   | 360  |
| aacgtgtaca  | gctacctgcc  | caacaccgca  | acagaaacca  | tgcgtatcag  | tggcctctgg   | 420  |
| gaaacgctgc  | tgagcaggat  | aggggatgac  | gtgatgatgt  | atctatttga  | gcactgtgca   | 480  |
| atcttttatgc | tggttccccc  | tagtaactgt  | taccaagtct  | gtgggcaacc  | aatttatgaa   | 540  |
| cttatcttcgc | aaaatgtaga  | atcagcccca  | gcgtttgtta  | aacaacggct  | ttcaaaagcac  | 600  |
| aaacgtagta  | gcttgcttaa  | gtatacgcag  | aaaaggctaa  | cgtttcacag  | acagtatctt   | 660  |
| tcaaagtcac  | gtcagtcgaa  | acgcaggcaa  | agacttgaag  | ctaattgtct  | cagcatgaga   | 720  |
| aataaaacca  | gcaataatat  | acaaagccta  | gggtcccgctg | ctctggaaaa  | acagagtagc   | 780  |
| tccaatgcag  | gtttgtcagc  | tacagcacca  | tccttaaaaa  | ggaagcttgc  | tagggaaacaa  | 840  |
| ctggaagtca  | cggctaagag  | agcaagatta  | gaagagaaag  | agaggaggga  | acaggcttgt   | 900  |
| aatactgctc  | ctaattgtaa  | ccagagtatt  | cccaagaggt  | atggaaccag  | ctgtgtagca   | 960  |
| tcacgttctg  | taagtcttat  | taaagaaaaa  | tacatttctc  | aaagaagtta  | cagtgtatag   | 1020 |
| tctcgtectt  | cttttagttca | caattctcat  | catgggaaga  | agtctgtggc  | agacaaaagc   | 1080 |
| tctttcctgc  | aaggagctga  | gagtaacaga  | catttaaaagc | ccagcattga  | aatgcaagca   | 1140 |
| ggatccagca  | ggaagagagt  | agagatacac  | aggectatac  | ctcggttgga  | ttggatacca   | 1200 |
| atcgaaccgg  | cggaaagtag  | ttcttcagga  | cacaaaaaagc | aggaaagtcc  | cctagctcat   | 1260 |
| ctggcagagg  | agttaccaaa  | tagggttttg  | ccatctacaa  | tatacattga  | caggaaagttt  | 1320 |
| cttctgtatt  | ctcgcaggta  | ctggggggaa  | cgtttcccaa  | aatccttctc  | attgaatcgc   | 1380 |
| ctgaagggtta | gtgaggcagg  | tgtaaagcga  | ctaatagaaa  | cgatattctt  | aagccaaaaat  | 1440 |
| ccgtttgggc  | aaaagcgcaa  | ccaaggctctg | ccacagaaaa  | aatggagaaa  | gaagaagctt   | 1500 |
| cccaaacgct  | tctggagaat  | gagaagtacg  | tttcaaaaaac | tcttaaaagaa | tcattggaaaag | 1560 |
| ttcccttacg  | tagctttctt  | gagacaaaat  | tgccctcttc  | ggatatctga  | aaccattttg   | 1620 |
| ggaaaagcca  | agctgtctcag | tcggggcaect | ttgcctgggc  | aagcagaggc  | tcacaagcaa   | 1680 |
| gcagaacagc  | ttgggaagga  | gcctgctaag  | cgtgtggcaa  | gcagcagatg  | cgaatctggt   | 1740 |
| cacaccaacg  | tgcccagcag  | cgtacgcgct  | cctctcgagc  | catctgcgtg  | cgtggagcca   | 1800 |
| gggggggagg  | agcagatccc  | tgagaggcg   | tctgattcag  | tcttcaggga  | gcttctcaag   | 1860 |
| gagcactgca  | gccactteca  | ggtgtacctc  | tttgtgaggg  | agtgcgtgga  | gcgggtgatc   | 1920 |
| cccgcgagc   | tctgggggtc  | aaaccataac  | aagcgccggt  | tcttcaagaa  | cgtgaaagca   | 1980 |
| ttcatttcca  | tggggaagta  | cgttaagctt  | tccttgacag  | tggtgatgtg  | gaagatgaga   | 2040 |
| gtaaatgact  | gcattgtggt  | tcgtctggcc  | aaaggtaatc  | actttgttcc  | tgccctctgaa  | 2100 |
| caccgttacc  | gtgaagaaat  | tttggctaaa  | ttcctatact  | ggctgatgga  | tacgtatggt   | 2160 |
| gttgagttgc  | tcagatcatt  | tttctatata  | accgagacca  | tgttccagaa  | aaatatgctt   | 2220 |
| ttctactacc  | gaaagtgtat  | ttgggccaag  | ttacaggaca  | ttggaattag  | aaagcattttt  | 2280 |
| gccaaagtac  | agctacgtcc  | tttaactgca  | gaggagatgg  | aagcgatcca  | tcagaaaaaa   | 2340 |

|             |             |             |            |             |             |      |
|-------------|-------------|-------------|------------|-------------|-------------|------|
| tacettctcta | tggeatcaaa  | gtcccggttc  | atccccaaag | tcagtggact  | aagaccatc   | 2400 |
| gtcagaatga  | gcggtgttgt  | tgaagcacaa  | acgttgagca | aggaaagcag  | agcaaagaag  | 2460 |
| atgaatcact  | acaacatgca  | actgaaaaat  | ctatttagtg | tgtaaatta   | tgaacgaact  | 2520 |
| ataaacacca  | gttacatcgg  | ctcttcagtg  | tttgggagag | atgatattcta | caagaagtgg  | 2580 |
| aagacatttg  | ttaaaaaggt  | tcttaaatca  | gatggtgaaa | ttcttcattt  | ctactatgta  | 2640 |
| aaggccgatg  | tgtccagggc  | ttttgatagc  | attcctcacg | ataaacttgt  | ggaagtgatt  | 2700 |
| tcacaggtct  | taaaacctga  | gaaaaaaaact | gtctactgca | tacggcgcta  | tgcagtgggt  | 2760 |
| atgatcactg  | gaagtggaaa  | aaccaggaag  | ttatacagga | gacatgttct  | tactttcaag  | 2820 |
| gatttttatgc | cagacatgaa  | gcagtttgtg  | tcccggcttc | atgagagtac  | ctcattgcga  | 2880 |
| gatgcaataa  | tagttgaaca  | gagcttaact  | ttcaatgaga | caagtgccag  | tctattttaat | 2940 |
| ttttttcttc  | aaatgctaaa  | taataacatc  | ctggaaattg | agcgagcta   | ctacttacag  | 3000 |
| tgctctggaa  | ttccacaggg  | ctcccttttg  | tcaacettgc | tttgagctt   | gtgctatgga  | 3060 |
| gacatggaaa  | acaaattatt  | cagtggggta  | cagaaggatg | gagtctgat   | cgtctcatt   | 3120 |
| gatgactttt  | tgtttgttac  | accacactta  | acgcattgca | gaactttcct  | aaggactcta  | 3180 |
| gcaatgggca  | ttcctgagta  | tggctttttg  | ataaacccca | aaaagacggg  | gggtgaatttt | 3240 |
| tctgttgacg  | atatcccaga  | gtgttccgaa  | tttaaacagc | tgccaaactg  | tcgtttgate  | 3300 |
| ccatgggtgtg | gcttatttatt | ggatacacag  | acacttgagg | tttactgtga  | ttactccagc  | 3360 |
| tattcctgta  | cttctatcag  | atcaagtctt  | tccttcaatt | caaacagaa   | agctgggaaa  | 3420 |
| aacatgaaac  | acaaattggg  | tgcagtcctt  | aaactgaaat | gccatggctt  | gtttcttgat  | 3480 |
| ttacagatca  | atagcgttaa  | aacagttttc  | attaatgtct | acaagatatt  | tttacttcag  | 3540 |
| gcttacaggt  | tccatgcctg  | tgttattcaa  | cttccattca | accagaaagt  | taggaacaat  | 3600 |
| cctgattttct | tcctcagagt  | catcgctgag  | aatgcacgtg | gctgctatct  | tatgctgaaa  | 3660 |
| gctaaaaatc  | cagggtttac  | tttaggtaac  | agaggtgcat | ctggcatgtt  | cccttctgag  | 3720 |
| gcagcagagt  | ggctctgcta  | tcatgccttc  | actgtcaaac | tgtcaaacca  | caaagttgtt  | 3780 |
| tacaaatgct  | tgcttaagcc  | cctgaagtgc  | tgtatgacac | agctattccg  | gaagatccca  | 3840 |
| aaggatacta  | aggcactact  | gaagacagtg  | acagaaccat | ctatttgtca  | agattttcaa  | 3900 |
| gctatcctgg  | actga       |             |            |             |             | 3915 |

<210> 3

<211> 26916

<212> ADN

<213> Cairina moschata

<400> 3

|            |             |             |              |            |             |      |
|------------|-------------|-------------|--------------|------------|-------------|------|
| gctatcactc | catttcaagg  | aagggcaaaa  | ggccgggttca  | ataccaacat | ctgtgtagct  | 60   |
| aagtaggatg | aaatagatta  | ggtgaccaa   | atatctgctt   | attcagcagg | tgttgatcca  | 120  |
| caggagggtt | tacgataaag  | ctccagtagg  | agttgtacca   | gcgtaattcc | tggagggcag  | 180  |
| caataagtc  | accatctaca  | ctccacaaag  | tctgatcttt   | aggtacaaag | agctgtgctg  | 240  |
| atctagaatg | atgtgggttag | actgaatctg  | gctctaaatt   | tctgttatta | gagtattatg  | 300  |
| tattattaag | taagaggcct  | tgtttcttac  | tgtgtcctaa   | atgaagatta | actttacagt  | 360  |
| gaagcagagt | gagagatgaa  | gggggtgataa | tgtgtttgaa   | atatcagact | taatctggca  | 420  |
| ggtgataatc | agttcattcc  | aaaaacgtta  | atgtgctgaa   | tcagattgac | aagaaaatga  | 480  |
| atcctaatta | gttgaaatta  | atctgaagag  | aactgaaagt   | acatcatagc | aaatagtaac  | 540  |
| tcttcttcca | catcacaagg  | aacagattcc  | actgtgttag   | tgatggaggt | ttatacaaaa  | 600  |
| aaaaaaaaaa | aagaaaaacc  | agaacagctc  | ttaatttata   | tacttattga | agcagattct  | 660  |
| atatatgcat | atgtatgcac  | attcattaat  | cagcagtaca   | tctcaacct  | taaaagatcc  | 720  |
| tggagaactg | ttgcttgtca  | aaaagcactc  | aggttaggct   | aacatttgca | gcttttgctc  | 780  |
| tagcatacat | ttacattagg  | tgtaaagaca  | aggaatgtgt   | aaaagagggt | aaatgcattc  | 840  |
| ctatctacaa | tgttatgtat  | atcttggtcc  | tgttatatattg | gctttacttc | caaagtgtga  | 900  |
| gtttgcagta | ttagtttctt  | tttatgtgta  | tccttgcata   | tatatatatt | caataaatga  | 960  |
| gttatgaatt | tcatatattgc | atatctgtcc  | tttttctgaa   | agtagaactt | taaaactgac  | 1020 |
| acttcagtg  | ataaaagcag  | tgttccctat  | tgatgttana   | gtacctcgaa | agttttacca  | 1080 |
| gaagacaagt | gagtgcatac  | ttaaagtatg  | ctacttactt   | ggtggtaact | gcgtattttgt | 1140 |
| gcattgttcc | tgactttgtg  | taatgggtaa  | gttgctttgt   | atttaccact | tgctaataat  | 1200 |
| gtgaagtggc | tggctcacta  | aaggttgcca  | ttctcacctt   | tcaaggactt | cttgaaggct  | 1260 |
| gtgtcataat | aatttagagt  | aagacagata  | ctgagagcgc   | agagtgtaat | atagttatgt  | 1320 |
| tgagaagag  | cggatagaca  | caggatgcat  | gttagaacca   | tataaatgtg | gtgtaaaaga  | 1380 |
| tcagcaaaac | gaagaattgc  | ttggaagatg  | cttgggtttg   | tccaaataaa | tagctgctga  | 1440 |
| gagtgagaga | gtgagtgagt  | aatattcagt  | atcttttaaag  | tgtcgagaaa | tgtaatgggg  | 1500 |

|             |             |             |             |             |             |      |
|-------------|-------------|-------------|-------------|-------------|-------------|------|
| aaaaatttaa  | aataaatgtt  | ttttgtttgt  | tttaaaagga  | gctgtgtgct  | tttatgttga  | 1560 |
| catgttgatt  | taattactgc  | ggtaggtatt  | tactgtgccc  | tgggaaacgg  | tatccccagt  | 1620 |
| caacacagac  | ttatttgtca  | ggaaaaaat   | tgctatagat  | catcttcaaa  | taagagttga  | 1680 |
| caaataatca  | aatctttagt  | gagtataatt  | attaatactg  | tgattacata  | tccatctttg  | 1740 |
| caagggtttc  | ctgaaaagg   | cagttttaag  | tcttctattg  | tcaaactctga | tgtttggagt  | 1800 |
| agttcgctac  | atgggtgttg  | atgccatgta  | aatagttacc  | agatagacgt  | tttattttat  | 1860 |
| gtgctgtatg  | tttttgtttt  | tcattcagtc  | agctggaatc  | attgaaacag  | agaagggttc  | 1920 |
| tcaagaaata  | tcctgaacct  | gttttgtgga  | tgcttctatt  | gcaatagctg  | gtcactgaaa  | 1980 |
| tcttgggaata | acgacagagg  | aatccatagg  | cacaggacca  | aacacctttc  | attgtcctca  | 2040 |
| tatagcgtga  | tgtaggggca  | gagtggtag   | tacttcagtg  | gctcctgtgt  | ccatgcttta  | 2100 |
| atgaactcta  | cttaactctac | catatgtaag  | agacttgacg  | cacaggcaaa  | acaagggaag  | 2160 |
| ttatctttgt  | gcctagataa  | tgtaccacat  | atggtaaac   | attttcaagc  | ctcagagaca  | 2220 |
| aaagaaggat  | ctgtgtaaagt | ctcaagtctc  | agcttctgtc  | tctgtttcca  | cctttctgtt  | 2280 |
| ctttctgttc  | agacctctgg  | gccagtaatt  | taaactggaa  | aattaatgga  | acagagaaac  | 2340 |
| tgtttgtgag  | cctatagaaa  | gatcaaattg  | tgctagagta  | atgttgctcc  | ttctacaccc  | 2400 |
| agataaactt  | tttgagcaga  | ggagaagcag  | tgaacagagc  | tttactttca  | tgactttgca  | 2460 |
| taagaaaaca  | tgggaatgtg  | gtccgagacc  | aatttataaaa | tagagggttt  | gaaaaactgt  | 2520 |
| ttggaaaaaca | aaacttgctt  | ggaaaacctc  | attcagagct  | gtgaatcatt  | cacagacaa   | 2580 |
| ctcttagggg  | tgtagccact  | catcagctga  | atctgattca  | acgatatgct  | gaaaaaaaaa  | 2640 |
| aacacaaaaca | ttgttgtgga  | atgtatgaac  | aggaatgtag  | tctgtgaagat | gcgtgtcttg  | 2700 |
| gtccttcttc  | tttaggtaat  | agtagaacac  | ctttctggga  | aaatgggtatc | cagctgttgt  | 2760 |
| aggcactgca  | ctccatggaa  | aatgtgaaga  | aagtggaaat  | aatccagaga  | aaagcaatga  | 2820 |
| gagggatgac  | aggataaaaa  | atgagccttg  | ctttaagggtg | atgagagaaa  | agtgtttttc  | 2880 |
| aaatctgtta  | gagagccaca  | gagaagaaag  | aaagaaacag  | ttctcatcat  | tgtaggttag  | 2940 |
| acaagaaaaa  | taggtttaag  | ttgctgcagg  | aagcctttag  | cttagacatc  | agaatgggta  | 3000 |
| attgtccttg  | acagtaggaa  | gaattaagct  | gtagcatgca  | cttctctggag | aggtcgtgga  | 3060 |
| atctcaggaa  | ttaggtttta  | aagggaatat  | ttgtaggcat  | cttttggggc  | tggtgcaatt  | 3120 |
| gtaggttgga  | ggtgctgggg  | ataaggctaa  | tggaataccc  | taccagccac  | tctaaggcct  | 3180 |
| tctgcaagga  | gtcagaacca  | cttttctggaa | gtaaaaactct | gtactggcag  | aggtcctgtg  | 3240 |
| cttgacctta  | aggctcagag  | ccatgcattt  | ccgttttctat | cttttccactt | aaaatagcac  | 3300 |
| ttgtggtagt  | aataacatgc  | tcttgccggag | tggcacagggt | ctgcatggtt  | atgaactttc  | 3360 |
| tgccaccacc  | cagcaaatgg  | cattttccac  | tcttctgttt  | cagattttctg | ggagtacttc  | 3420 |
| cttactccca  | aattcctgat  | tgaatacgac  | tgaataacttc | aacacgtttt  | tagtgacaga  | 3480 |
| agtgtacttt  | atacaaatgt  | gtgggactat  | tgacacacaac | ttactgattt  | tcttcaactgt | 3540 |
| gtgcacatgt  | gctctgtgat  | gatacagagt  | ttgggggtgac | tgaactgtta  | cccagtcttt  | 3600 |
| accgaattag  | ggcagcgatc  | ttaaaccttc  | atctgaactt  | ttgctagaga  | tgatcttcta  | 3660 |
| ttttgttttg  | acaggggtct  | ctgcttgctt  | gcatttgttc  | taaaacgaca  | gtctggatga  | 3720 |
| gaagaaaacca | accagggtctg | gcactgccct  | accttttatt  | cctgggatac  | tttacttgge  | 3780 |
| acatcacttg  | gcacacatgc  | aataaccttt  | tttcccccgg  | gtacccaaag  | aaaagacgtg  | 3840 |
| ctcctgcaca  | gactcagcca  | aattcctgcc  | tgccagtggg  | gggctgtaca  | gaggcttggg  | 3900 |
| gcagaggggag | gaaactcact  | cccaaagcat  | atactaaaaa  | tagggcaatg  | cattgcaaac  | 3960 |
| agctttgtct  | tgagagcttc  | cttctggcat  | gtagccatgc  | agctgcctta  | tagttttgag  | 4020 |
| cctagatacc  | tccaaaacaa  | aacaaaaagt  | cgctaaaagt  | aaatgcagcc  | agcagtgtct  | 4080 |
| atagtcttga  | taatccctac  | agtgtatcaa  | tttttagttt  | ttccagagtg  | caggtagttt  | 4140 |
| actcaattta  | acttcaaatg  | tctgagttac  | accattgaca  | ccattgtcaa  | gactagtgtc  | 4200 |
| ggcaatccct  | gaagtctgcc  | aagttcaaa   | cagggaggag  | gccttttaa   | ttgaaatact  | 4260 |
| cgggtgacgt  | gtccaaaaag  | ttttctctgt  | gctttctctg  | ggagtaggaa  | ctgcagttagc | 4320 |
| taggttgcaa  | aaagttttta  | aactctccaa  | attgctgcta  | actaaacatc  | cttaaactct  | 4380 |
| gacatgagag  | tatgaaatga  | gaaagtcaag  | tccttgattt  | agaactagctt | gaaaagaata  | 4440 |
| aactttttgc  | acaggactcc  | ctggacagtt  | gggctggata  | atagaagtga  | ttcattcagg  | 4500 |
| tcttgatgcc  | tacaagtttc  | ttgggttttg  | ccatattttc  | cagtgatttg  | tgacatcatt  | 4560 |
| tttgccctaa  | aataaatcgc  | atcatgtcct  | gccgtatgtt  | ttgactatga  | aaatggagggt | 4620 |
| ttcggtttga  | ataccaaaact | tcgttgttac  | ttctttcaga  | ggctgttgat  | aaaggaaaaac | 4680 |
| gtgaagcttt  | tatgtaaatg  | atcatgattt  | ctcatttgca  | tcaagaatat  | tttccaaaat  | 4740 |
| agtagttaa   | tccctccaat  | gattttcagc  | tatgtcacat  | aaatcaccca  | aaacacggta  | 4800 |
| attttgcacg  | ggtgtttttt  | ctcctttttt  | ttattgtcta  | atactcagct  | tatatttgag  | 4860 |
| tcttctgtat  | ccaattacct  | gatttctgtg  | gcaattaaac  | agaagatcag  | cgctcattgta | 4920 |
| cttcacattc  | tttttaaaact | catcagtagt  | atgaccacac  | tagttgggtc  | tgtaagtttg  | 4980 |
| taatttaaga  | tgtgcctgga  | aggggttcaa  | agagacagga  | gcagttagtg  | gtctcacttc  | 5040 |
| cctttggaaa  | gcctttctga  | tactgtcaag  | caatatttac  | atcaaaaggg  | aactgaattc  | 5100 |
| tggtaaaactt | ttccagactc  | caagtaaac   | gaacaacctg  | agatttagat  | aaaccttggt  | 5160 |

|             |             |            |             |             |             |      |
|-------------|-------------|------------|-------------|-------------|-------------|------|
| ggatttaaaa  | acttgatggt  | atcactgtct | aagactttgt  | ttagcccatc  | agatattgca  | 5220 |
| gacgtatatt  | ttgaggaaa   | gcttttcagt | aattgataaa  | attaagagca  | gagttttggc  | 5280 |
| aaaaaaaaac  | gccatctctg  | ctcagtaact | catggcatct  | tgaattttct  | cagtttagagg | 5340 |
| tacaggtgct  | ctgtttctga  | attaaccctg | gctgttaatt  | aaaatgtact  | gtttgtctaaa | 5400 |
| tatgattagt  | gtctggagcc  | aggtagcttg | aaaaagcttc  | aagattattt  | ctttttacca  | 5460 |
| ggaccatcaa  | gttttttagc  | ctccctttag | aggaagggtc  | agaggagggtc | ctatttaata  | 5520 |
| tgtcccagaa  | gaaaggctat  | ttctccctct | agaatggggc  | ttggcagtc   | ttttcagagc  | 5580 |
| acttttcatc  | cttgtctcat  | agggattttt | aaggacagat  | ttctgaaagg  | attttccacc  | 5640 |
| gggtaaagcg  | ggatcctgaa  | tggttgacag | ataatacgt   | ggaattcttg  | cttccctact  | 5700 |
| ccaggttggt  | tatcttgaga  | tcttaaaaa  | ggacttattt  | ccctttttaa  | gtgtaaactg  | 5760 |
| agttcttgag  | gcaccttttc  | cactgaagta | tctgactttg  | cccttactgc  | ccactctcag  | 5820 |
| cttgacacatt | taacccaaat  | tcagatccag | ggtgtgggtg  | tactcctaga  | acagctctta  | 5880 |
| ctgtttgagg  | atctgaatga  | ctattttgcc | agcttgtgcc  | atacaactcg  | agaggacctg  | 5940 |
| caaaggggag  | gggagtcctg  | atgggcagtt | ccctccagc   | ctggcatcg   | cttcgtgtct  | 6000 |
| ccgcttccct  | tataatctgc  | ccagaaaaga | tttctgaaat  | acttgagcag  | ctgccacaat  | 6060 |
| ctgcattggg  | ggcagcagcc  | ttggcatgtg | agctctgcca  | cctgatgcc   | ggtctgtact  | 6120 |
| gaggcagtg   | cagctggatc  | tgcaagcctc | agcgaatata  | gccttcgtca  | gcgagttctc  | 6180 |
| tgatcagatt  | gctgccttgg  | agccttttgc | tggtcttctc  | cagctcttct  | cctgttactc  | 6240 |
| agatagcctt  | accacctcct  | ctcatcatct | tttgagtttc  | tatcctcag   | ttacagccag  | 6300 |
| gtaagcccat  | ttacccatgc  | tgatatcaac | aggttccaga  | ttttttaaaa  | aacaatcctg  | 6360 |
| tctggctcac  | agctagggag  | cctctagtga | cacagctaa   | tcccttgcc   | ccattttaga  | 6420 |
| atgtttctgg  | agggatgcag  | gtacgtcttg | cactcaggtg  | cctaataaaa  | gcagattgac  | 6480 |
| ctgttcattc  | cttcgaaaag  | aactaagcaa | agtggggaat  | cccatgtctc  | tgccccactt  | 6540 |
| ctgtgttctc  | gacggcaaga  | cctcctgcca | gcaccaaccg  | cctgtgtttg  | gagtcagcca  | 6600 |
| gccttccggt  | ctgactgaag  | ttgaagggtc | agtacaaact  | gctggctaaa  | agaagctgct  | 6660 |
| tgacttgctt  | ttttaggagg  | cttgtcttgt | ggggaaaaaa  | aaaaaaagta  | cagataaatt  | 6720 |
| ctgctgatcc  | cttctgagtc  | acccagcagc | gatggcaaa   | ctaggagaaa  | aaatgggatt  | 6780 |
| acagagggat  | tgatgaatgg  | acagtgagtg | acctgaagat  | ctaactcaaa  | ggacgttact  | 6840 |
| tcaagtgcga  | gtgtcaacat  | ataaatgctg | agttgttgtc  | tccaagtacc  | ccactgttgg  | 6900 |
| ggtatcctgg  | ttaccaagca  | aggttttagt | aagctctgtc  | gaaaagttac  | atacagctga  | 6960 |
| cattcatgca  | ggccatgatt  | tgccaggcct | gagctctgag  | gtattcctgg  | acgggttagag | 7020 |
| aggataaaat  | aatttagaag  | tcaggaaaac | actcagagac  | aaaccattat  | ttataaaatt  | 7080 |
| attttgctct  | ctaaaagatg  | tgcttccag  | agaaatcaca  | acgttggtct  | caaccttttt  | 7140 |
| gcacatctta  | gctagaaaat  | gtgcagcttc | cagataccgc  | cagatcacct  | cccttgacct  | 7200 |
| tgccctggaa  | acatcaatgg  | ctccataacc | agtaagaaca  | aaactgggtc  | tctactaaat  | 7260 |
| ccgttgctga  | tctcagctaa  | ccaacttctc | gtaatatcag  | caaataattc  | tgcttttgta  | 7320 |
| aatctgttct  | tcttccata   | atgggggtgc | agcaggaatg  | cttgggatcc  | agaaatgcgc  | 7380 |
| tggttgccag  | gaagaacaaa  | gaaattgttc | acccagaagg  | gcaaaaaagg  | actaatgctt  | 7440 |
| tcgtccgctt  | gggagaagtt  | gcacagaact | tatgccaaac  | acttttgcac  | gcaatcagaa  | 7500 |
| agatgcgcgt  | ggatgttact  | tttaatatga | gacgtttaata | tcagttatta  | attagaaatt  | 7560 |
| tcttcagtaa  | ccagactaaa  | agcagatcct | gaaacactcc  | tggtggtgaa  | cagtctcttg  | 7620 |
| acagacactg  | cccactgtcc  | agtaatgtca | ggcgctctct  | gaacttgaca  | gggcagctgc  | 7680 |
| tggttttccc  | agccttctgg  | aaataggcca | gctctgacat  | gtttctgata  | ttagctgggt  | 7740 |
| gtatttcatt  | ctgctgcctt  | aggcagtttg | ataaaggctc  | cttccctctg  | ctccagagct  | 7800 |
| aaccacccca  | aactgtagcc  | gagcacctcg | ctgcaacaaa  | actgcactcg  | ctagggttct  | 7860 |
| gcctgtttgc  | ttcatttaaga | tctgcttaaa | ttgtttcgta  | caaagggaaca | ctcaactga   | 7920 |
| tttctgagcc  | caaagtagca  | gtgctagggt | tacatcagga  | gttggtttgg  | atgaagaaac  | 7980 |
| attgcatggg  | cactgtatga  | ataaagttat | ttttaagaat  | cattatccct  | tcttgataac  | 8040 |
| caagtcttta  | tgcggcagaa  | aatcaaaact | ggtctccacc  | cttacagaaa  | gcagagggaat | 8100 |
| gctttcagct  | gatagttgct  | taagctagaa | tataagaaac  | catgaatttc  | tgtgtgcaact | 8160 |
| ggggcattgc  | ccttcattcc  | agacctacag | aaaaaaacga  | cacttttggt  | actatttttt  | 8220 |
| tcttcccat   | atgagaccag  | gggagctacc | caggcatttc  | cattcttata  | attttacctc  | 8280 |
| aagatcaaat  | tttctccagg  | cagttaaagg | cagctgcacc  | cggagacctc  | gctcagcctc  | 8340 |
| cccttgcatc  | ccacggagct  | gcgttttagt | agaaacctcc  | cccagggtga  | cgggctgcag  | 8400 |
| gggaacctct  | cccaacggcg  | tcccgttccc | tttctcagt   | gcaaacgcag  | ccaccgccc   | 8460 |
| ttgaaccttc  | ctccgggttc  | tctcggttcg | ggggaggcag  | gagggggcgg  | tgcccgccgc  | 8520 |
| ccgggagctc  | ctcacagggc  | ccgggccccg | gggcgggagc  | gccggggcca  | tggtgagggc  | 8580 |
| ggggagcgcg  | gaaggcgggc  | ccggggcgcg | tgcgggggcg  | ggcgccctcc  | cgtcccccgc  | 8640 |
| cgtctctcgc  | cgcgcgcagc  | cgaacacccg | gccccggccc  | cggcgggccc  | cgcctatggc  | 8700 |
| acccccagcc  | cctgcacgt   | ggtgagtgcg | ggccccggcg  | tgccctctcg  | gcctcccgct  | 8760 |
| gggccccgcg  | tgcggtgaac  | ggggaggggg | gagggcgctg  | cggggagccg  | gcggggggcg  | 8820 |

|            |             |             |             |             |             |       |
|------------|-------------|-------------|-------------|-------------|-------------|-------|
| cccgccccg  | tggggcgccc  | cgcgtggcgt  | ggagtcacct  | tgaggggctc  | gggctccccg  | 8880  |
| tgccctgcac | ggcccccgccc | ggccccccagc | gccccctcctc | ggctgcccggc | tcgggtctcgc | 8940  |
| ctcctgcccg | gtgtcctgcc  | atgggggcggg | cgggggtgttg | gcaatttcggg | ggcgctgctt  | 9000  |
| ggggcggtcg | cgggtgggag  | agccgcgggt  | gctgcctcag  | aaggcgccca  | gaaataaactt | 9060  |
| ccaccagaaa | taacttccac  | caggcaccga  | cccggcgtgg  | cacaacttcc  | tattcctgtt  | 9120  |
| gctcctcaca | gcgrgctcgg  | gaggtcgtgc  | tccgtgtacc  | aaaacctgcg  | agctcagcac  | 9180  |
| ccgaactggg | atgccttttg  | gaggaagacc  | ccgctgaggg  | tgtatttttc  | atcatcagtg  | 9240  |
| tgctctgcta | agaatccttt  | tttttttaat  | tattattttt  | tttaaaccca  | gccccgggtg  | 9300  |
| ccgtgagcgg | tagcgtgaca  | cgcctgtggc  | gcctcgccggc | ccacggccat  | gcctgcccgtc | 9360  |
| aggctgtgcc | cgcgcgcctg  | cctgtacttt  | ctcgctaaaa  | atccatctgt  | aaaaccacc   | 9420  |
| gaacagcccg | ctcagtacca  | aacgtagcct  | gcaaggctta  | gtggatgtgc  | tagatgtagt  | 9480  |
| tatggggctg | aaccgtagga  | aatttccgta  | gctgctgaga  | gggaccagc   | tcctgacacg  | 9540  |
| cagcaggtca | cgggtcgggt  | gcctggcctg  | ctgcgagcag  | gggtgtgtga  | ggggctgctg  | 9600  |
| tggctgggag | ggcaggccctg | gctcccggcc  | aaaatcctca  | ttaccgcttg  | cgcaacccaaa | 9660  |
| aaacatgtaa | acagctgggt  | gatcacaggg  | aaggtaagtg  | cttccttggg  | cactgaaatg  | 9720  |
| gaagcgtgat | ctgcaccggc  | agcggagctg  | aggtggatgc  | ttaccatgcc  | tgccccggtg  | 9780  |
| cattgccagc | agcaggcttc  | agtgttcccc  | acaaacttct  | ttttacactg  | agcagccttc  | 9840  |
| agctgcgtaa | ccaacaagtg  | aagcgattca  | aaatatattg  | tcattttcaa  | gtggcagcgg  | 9900  |
| ggctaagggt | cagggggaagt | tttttcttgg  | tgtttttgga  | acaatatgcc  | ctgaaccaca  | 9960  |
| gcagtgggtg | tataattaag  | tacataggca  | aaccccttgt  | cctaattaat  | acactcctgc  | 10020 |
| taagtgttga | agctgtttca  | tgtggaaaat  | ctgcccgtgc  | ggaacagggt  | gctgaaacag  | 10080 |
| ggctttaatt | tctgattcta  | aagatcctga  | acctccactg  | tttccagatg  | gtggtcagag  | 10140 |
| aaggcttgtt | tcataggcac  | ataacttgtc  | ccgttaatat  | agctgtcttg  | tcttgtattg  | 10200 |
| acagctgggt | tgtatgtttt  | gttgtgatcc  | gggtgtgtgg  | agatttggaa  | gatgtccctg  | 10260 |
| ttacaaggct | tgcttactta  | gttgcaagg   | caggttttgc  | attacactgc  | tgaaatcaga  | 10320 |
| gatgtttgaa | agctgaaaag  | tgcttgatcc  | ttgtttattt  | atttattttt  | aaatgaaaata | 10380 |
| agattatttg | aggtgtagta  | atttcatttt  | ccccgtcttc  | caatgggggg  | aaatgttgtt  | 10440 |
| agggcttgaa | atcatttccc  | ctccacatcg  | tgctgagcga  | cagctcatgt  | agcatttcag  | 10500 |
| acaactcttg | tgaactgtgg  | gttgtgcaag  | cttccacttt  | gtgctctgtt  | cagaaggaaa  | 10560 |
| ataaggtagt | cccaactagt  | gtaatggatt  | tggtagttaa  | ttgagtaaa   | caaaggattg  | 10620 |
| gcagtttctc | actacaggct  | ttctataaga  | ctttgtagaa  | atctcacctt  | atttccctttt | 10680 |
| cagattgacg | atgatgaaca  | aggttacgac  | ctggacttgt  | tctgcatacc  | taaacattat  | 10740 |
| gcagatgatt | tggaaaaagt  | ctatatcctt  | catgggctca  | tcattggacag | gtttgtttga  | 10800 |
| cttcagacag | tacactgctc  | cagctgatcc  | catgacactg  | gaaaaaacaa  | tcttccagtg  | 10860 |
| atagttcttg | tgccctagtga | ctgtctaaac  | agattacatt  | taattagaga  | ctaagaaata  | 10920 |
| cacatgttaa | tttaactctc  | cttgtttggc  | ttcaaaagag  | ttgtacattt  | gcagttacgc  | 10980 |
| tattgttttg | aaatttgtca  | attctcaaag  | aaatttgtgg  | tacgtagcag  | tctgtgactt  | 11040 |
| tctttacagt | gtttctttga  | tgttttactt  | aaagttaatta | gaacacatta  | cttggtgtgc  | 11100 |
| tagttcaatc | taaaaacagt  | tatagtctct  | caaacatctt  | taggatatta  | ataagagtag  | 11160 |
| attattaatc | acattatgat  | aagacttttt  | catgttcctg  | tggtagataa  | ctaaccacca  | 11220 |
| attttccctc | actgtctgcg  | aaagacatta  | gcctttgcaa  | ataccaatct  | gtcacttgtg  | 11280 |
| gttgctgaaa | tgtatgattt  | tctctggaag  | ttttatctcg  | tgatgagaaa  | tgggtacatg  | 11340 |
| aacctataag | gtgtttttgt  | tttactttgt  | gtaagttaaag | tggaggagtt  | gctggagaca  | 11400 |
| cagaaccact | gaagagcggg  | tctgagtaga  | tctgtgtaat  | aggaatgctt  | ctagattttg  | 11460 |
| ctgggtgctg | tttgatcaga  | ttattacagt  | atttataata  | aaatgttttt  | taactttaca  | 11520 |
| ctgaaagacc | ctatatagga  | aagcattgga  | caaagtacag  | gtcattaaat  | agctgatgta  | 11580 |
| aagtttgtta | tggcaggcat  | tctctgagaa  | acctgctgtc  | agctgctata  | ctgtaaatac  | 11640 |
| ataccatgct | ttctgaatta  | aattgcaaga  | taaatattaga | aacaatgatc  | actgaaaaac  | 11700 |
| tgttcagtg  | tctcttgctc  | tgctttattg  | gcattatatt  | ttgcagtcag  | gacaaattta  | 11760 |
| ttcagcaaaa | cataacgcta  | tagttgataa  | tttgagagtt  | ttcttgctcc  | tcagttagtg  | 11820 |
| agagctgttg | tctttttggg  | tgtgctgatt  | tattttgctc  | tctgcatgga  | agctgaacct  | 11880 |
| atctttggaa | gaagaaaaaca | cccttatgtc  | tcttatctga  | cagtaaaaaca | attcaggggtg | 11940 |
| ttcagatttg | ctttggctga  | gtatgatgta  | tgaanaaaaa  | gaagtttggc  | agtgttactg  | 12000 |
| ttagattaac | cttgggaacgc | aaaactttgt  | tgaccaatag  | tgggttaaag  | tgactgaagc  | 12060 |
| attaggcaaa | tatttctgag  | caaaaatagc  | ttccgagttt  | gcattgtgtt  | gctgttgttt  | 12120 |
| tttgcaatac | aaaataactgc | tgccatagta  | agcaaaactaa | atgtgttaca  | acagctaaact | 12180 |
| ctcttttttt | tttttttgta  | taggacagag  | agactggcac  | gagaaaattat | gaagggcatg  | 12240 |
| ggaggacatc | acattgtagc  | tctctgtgta  | ctcaagggtg  | gctataaatt  | ttttgctgat  | 12300 |
| ttattagact | acatcaaagc  | actgaacaga  | aacagtgaca  | aatcaatccc  | catgactgta  | 12360 |
| gacttcatta | ggttgaagag  | ttactgtgta  | agtatctctg  | caataccatg  | caatttttct  | 12420 |
| gtaaaattga | ctaacttcaa  | attaacaaca  | gggatgattg  | agaattgcca  | acaaatgttg  | 12480 |

|             |             |             |            |             |             |       |
|-------------|-------------|-------------|------------|-------------|-------------|-------|
| caaaagcttt  | gcctaagtac  | tgcctaaatt  | gtgctaattt | tatacaata   | gttaagacaa  | 12540 |
| ttaaggggaa  | aaaagcagtc  | acaagctaac  | ttgttctttg | tctatcttat  | atgatctggg  | 12600 |
| ttctttcaga  | ctttatctcc  | tcggccagtc  | aaaatccaga | gcaaagagac  | cccttccatg  | 12660 |
| tcctttactt  | cttaaaccaat | ttcctctctg  | ccccctgtcc | cacctctaaa  | actgtgggtc  | 12720 |
| tagaataata  | cagaggatag  | tectacaaat  | cttattacaa | aaacttaact  | ctaggaattt  | 12780 |
| tcatgtggct  | tagacatcac  | tcaccagata  | ggaaaacttg | aaaactgtga  | gcatggttat  | 12840 |
| atgtgggttc  | cctacctcat  | cttttgtttg  | gctcaactga | acaatgaatg  | gaatgaattg  | 12900 |
| catttggcca  | tgaggaaata  | ctagattcat  | aaaaacagat | ttatgggtag  | cggccactga  | 12960 |
| atatgggtgc  | acctttttaa  | gcctaggtct  | aggttgcctt | ggctttatgc  | ttttggaaaa  | 13020 |
| gaattttatt  | tcattttgca  | cacaagtatt  | taactttaca | gggaaatgga  | gcatgaggta  | 13080 |
| gtatgtaaga  | tttttataag  | ggaagcatta  | gattacattg | tgcagggtcaa | aggagggaagc | 13140 |
| agatgtatgt  | tagtactgta  | tgttctctga  | ctcagcaaca | gcctagtcta  | gtgttttcaa  | 13200 |
| gacagtattt  | gtgggtaatt  | tagatttttt  | ccataagatc | acctgtagt   | tggaaataat  | 13260 |
| taagtttttt  | gatttcaatt  | tacggtgtca  | aaataaattg | cttcttaagg  | tcatgctgtt  | 13320 |
| agttctgtgt  | ctcgatgact  | gtgttttcca  | gtcagtagta | taagctactt  | gtcatgggtt  | 13380 |
| tggagtagtg  | cctgctgcag  | tagaaaaaaa  | tagattacta | agttggaaac  | tcaactgcat  | 13440 |
| tgttcattta  | attaaaccag  | agataggggt  | acttgcagat | ttaggttgsa  | ctgtagaaga  | 13500 |
| gcttcaaaaa  | tgcgctgctg  | ctgttgagcg  | atctgtaaaa | atctcacctc  | gttgagtggtc | 13560 |
| aaagctcaga  | ccagtcagag  | tttggccagg  | cctgttttga | gaaaagctgt  | gccagctccc  | 13620 |
| aaggtgtgtg  | tgcagtgtat  | tcttggggct  | acctggaaac | caccgtcatg  | tctgcagagc  | 13680 |
| gtgctgcagt  | gactcagctg  | cactctttta  | agagaaggca | agggcaggca  | gaaagggaca  | 13740 |
| tgccagcctt  | tctgtacatc  | ttgcaagggg  | agagctgctt | gacacacttt  | gtttctgtctc | 13800 |
| cccacaaaag  | attcaggtct  | cctctaccag  | tccattcaac | cacaaagcca  | ctacgtgccc  | 13860 |
| acatacgtgt  | tctgtaaaag  | ccacaacaga  | aaggtgaatg | tctctattag  | aggataatat  | 13920 |
| ttgaaagaat  | aggaaataac  | tcccaccttt  | tttggcgagg | aaatttgtat  | tttctgtctt  | 13980 |
| cttcaaggaa  | cacatccaat  | aactttggtt  | tatttataat | ataaggacta  | caccttctgt  | 14040 |
| ctctctccac  | cttccgagta  | acttggaaac  | agctatggga | taagtgttgt  | gaaacacctg  | 14100 |
| gaaaaaatac  | tacattgcaa  | agcaaggctt  | attgtacagc | atgtttttat  | tgttcatctt  | 14160 |
| tatgcattag  | gcatgtattg  | gttctgtctg  | ttttattttt | tacaccgccc  | ccccaaacta  | 14220 |
| tagtgctttt  | atattaatag  | gaaactgaac  | agcaaaaata | ttacgggaagt | tttaattctc  | 14280 |
| tttttgtagt  | aagaaggagt  | aaacaaagag  | aaaagcagag | aaatttacac  | agaaaggagt  | 14340 |
| gtaagcagaa  | tgttttccat  | gaggtgttgg  | cttcagggtt | ttggttttgt  | cagccaatc   | 14400 |
| tgtaaaaggaa | tgttttcttg  | ttctatatga  | ccgaggagct | ccagttactt  | gatggcccca  | 14460 |
| ctagcagatt  | attacgagca  | tttcatttct  | gtcaactctc | taaggtagtt  | attaatatcc  | 14520 |
| agacttacgt  | tctttagatt  | gcataacttg  | atagactggg | taagactgtc  | ctaaccacaa  | 14580 |
| catgcgtgta  | attgattcac  | ttagttttcc  | attattaatg | gtgctggaat  | gttaaatctg  | 14640 |
| ctctgtagtt  | ttcattcgtg  | ctttactaaa  | cggaggagtc | agcattctta  | atgatgggaa  | 14700 |
| aaccatcaca  | aaccgtaatg  | gacagttgaa  | agtagagcta | gctctcttgt  | gattgttgtt  | 14760 |
| ggatgcgaatt | actagatcac  | gctgcctgcc  | tttcacagat | gggtgaaaga  | tgtaaaaaat  | 14820 |
| accgtcgttg  | cataaacaaa  | gcagctgaac  | tcgtataatg | gttctaaaat  | tgatgcgatt  | 14880 |
| gatgttctat  | tcctatttga  | aattgatcat  | ttaatatttt | caagcttatg  | ggttatttgt  | 14940 |
| ttcataatgc  | aaagttcata  | ctatgattgc  | acccaaaaat | aattctctaa  | gaatactgag  | 15000 |
| atgatgaagc  | tgtgattaaa  | tgttgcttat  | aattcctgct | ttcggagtcc  | ttacaaatga  | 15060 |
| gttttaaatga | tctggcaagc  | tgtaagaaa   | ttgaaagata | gtgtttgtga  | acagaaccag  | 15120 |
| atctctgtat  | tcttgacaaa  | tagtttattt  | ttctgtgtct | agaagaatgg  | cttttcacca  | 15180 |
| gtttcctgca  | tgtaaagtta  | gagtttaagca | aaagtgtgct | atgtagtcac  | aaagggcagc  | 15240 |
| caaaacacaa  | tggacgtgtt  | gcattgtcac  | tcagggaat  | agggcttttc  | ggaggagtac  | 15300 |
| tgcagcttta  | aggcaaaatc  | tagttaggct  | ggttgagaca | gtattctgaa  | agatctcatg  | 15360 |
| ttaaactgagc | cttatagcaa  | ataccatgag  | gaacttaaaa | tgccttgta   | ggggttacga  | 15420 |
| tcttgagca   | tgttattatt  | ttccagcaat  | gacaatgagg | cagtgaatac  | agatctaaac  | 15480 |
| agtaaatctc  | taatgtgtta  | aatcacaaag  | cattaaagtc | ttaccagtac  | tgaatcaga   | 15540 |
| aataatgctc  | ctgaagataa  | ttaaagtgtt  | gcttaattat | taaacaaggt  | tctttccaat  | 15600 |
| ggaaatctgg  | ttaaaaccag  | tgttggaat   | ttgttaacca | tttttgatcc  | tttaattgct  | 15660 |
| cagaaactca  | gtacctgcat  | aggcgtgtag  | gtttatatgc | cagcgaaggt  | ttatgtttgt  | 15720 |
| tgtaaaaaaa  | tgtgcaatt   | tgcagggtgc  | ttggtgggaa | gaagaagggc  | agattttgtt  | 15780 |
| ggtttagattg | tttttctctg  | ctgatcagtg  | tacttgagta | tgcataacce  | acagggttct  | 15840 |
| gtagtgtctt  | gtagtctgt   | gttggtattg  | tttgattata | gtgtgtgtgt  | ttgtagccag  | 15900 |
| agagatcacc  | aggaaactga  | tctgcatgca  | tgtactgaga | ggactgggtc  | atcatctgtg  | 15960 |
| catggtgttt  | taatggaaca  | aatatctgta  | acatctttct | ccatcttctt  | tctgtaagaa  | 16020 |
| atggtgcttt  | ttacatcctt  | ccatggaaaa  | aggacgaat  | tcttattagt  | gtaggaatag  | 16080 |
| ataagagaga  | atcctgcaac  | ataaacaggg  | aaaagcttaa | agctgtcatt  | taagataatt  | 16140 |

|             |             |             |             |             |             |       |
|-------------|-------------|-------------|-------------|-------------|-------------|-------|
| ttaagatatg  | atgaacgcag  | gaaatgcccc  | tcagcattac  | aaacacgggtg | cttaactgaa  | 16200 |
| tcccatTTTT  | gtgattaatc  | taaccagtat  | ctctggcatt  | cagtagcagt  | ctgctgttta  | 16260 |
| ctgggtcaag  | cagcagacat  | tttggacatc  | ctatTTTTtg  | actTTTTcaa  | cacagtggta  | 16320 |
| ttggttctgt  | tcactatagt  | tttctttaaa  | acatggggta  | aaactgtaaa  | ggcgcttttt  | 16380 |
| ctaagtgtgt  | gtgcttgcc   | ttttttta    | gttttgatc   | gggaatctaa  | aatagaaagt  | 16440 |
| atctgttagt  | ctcgttatct  | gcttttacat  | acctagagct  | gtcagcactt  | ctttatccta  | 16500 |
| gagcattgac  | gtcactgtcc  | aagggtttt   | aatacatttt  | ctctttactg  | taatcccagt  | 16560 |
| gtcacttatg  | gtgtgctcag  | tgtgtgtttc  | tctctaaaac  | atttaccctt  | tttttgcagt  | 16620 |
| tgataacact  | tctacttttt  | gtctgcaact  | ggtagcagt   | gaatatgtgt  | gtcagtgtaaa | 16680 |
| acaggaagct  | gtatgtggcc  | gtctgttccc  | ctgtgtgtc   | ctctctattg  | tatctgccat  | 16740 |
| tttggggcgt  | ggggaaggaa  | ctgctctttt  | ctgctttgtc  | tgcatgggtc  | tgcatgaagg  | 16800 |
| ttgggcttca  | gactgaagcg  | gcaaaataat  | aaagggtttc  | taaagggaag  | atatgaagtg  | 16860 |
| aagagctgtg  | agaatgaagt  | ttgaaaagt   | gtatgaccct  | tcaataaat   | gattaaattg  | 16920 |
| taggataatg  | acatcatctg  | taagtagcac  | agtaatgagt  | atagcacaat  | taataagaac  | 16980 |
| taggagacaa  | tgtcaagggt  | taaagtcagc  | tgaatgactt  | taataaaact  | gcagagatct  | 17040 |
| ggaggtagcc  | caggagttgt  | ttttaacttt  | cttatataac  | aaacagttca  | ctgatgtcta  | 17100 |
| catcaaaagc  | tagttttact  | aaagcaaaat  | gctgtacttc  | agtgatagga  | ttgtgtagct  | 17160 |
| gtgtcgctgt  | tttaattacca | caatgagtaa  | tttatcacta  | tttgtaaaat  | aggcaaaccc  | 17220 |
| aagcttttct  | ggctattctt  | gtagcttata  | tttcttgttc  | ttaatctgtg  | ccttgttttg  | 17280 |
| tggctatttt  | attgtttgta  | tgattgctta  | tttttattca  | gagtaaggct  | ttctaaggca  | 17340 |
| tcaaagttta  | aaatgtaat   | taccagctat  | atagcagcgt  | acagcttggg  | aacttttttt  | 17400 |
| ttttcctgtt  | ttccataacct | tcaatgaata  | tgtaggaaat  | aataattttt  | taataacat   | 17460 |
| gtgtgtatat  | ataatgggtga | cagatgagaa  | gtgggtgcctg | tatttatgta  | aataattttt  | 17520 |
| caaacatctt  | tttatctgaa  | ataccatctt  | tcagtttgat  | tacttcta    | ttcttagtta  | 17580 |
| atagattgtg  | ctgggtgcaat | gcttctgaaa  | gctttgggtat | ttgtgggtta  | tgtactttgc  | 17640 |
| agtttttgata | tttattagaa  | gaaaaattac  | tttaattctg  | gtaatactaa  | tgcttaacct  | 17700 |
| gctctttctt  | gctagaatga  | tcagtcact   | ggagatataa  | aagtcattgg  | tggggatgac  | 17760 |
| ctttcaacct  | tgactggaaa  | ggtatggcat  | gtcatatgtt  | taaagcaacc  | ccctcccccg  | 17820 |
| cagaaaaaaa  | aatcctaaaa  | aatcctaaaa  | aaaaacaaaa  | caaaacccaa  | aaaaacaaaa  | 17880 |
| caaaacaaaa  | caaaaaaaa   | aaaaaacacc  | aaagaaccaa  | ataattagag  | ttttagaaag  | 17940 |
| gaaaaaaaaa  | aaaaaagaaa  | gaaaactgta  | gtgtttcatt  | taaaatttag  | gttagagggt  | 18000 |
| ggacttgatg  | atcttgaggt  | ctcttccaac  | ctagaaattc  | tgtgattctg  | tgattctgta  | 18060 |
| aaatatctac  | cattgttagag | tggtgggggtg | cataagcccc  | accagagcca  | tcagtttagt  | 18120 |
| ttgaacttca  | taacatatgt  | atgtatatct  | actgtcttca  | gtatcaggaa  | gtgttagagt  | 18180 |
| tggtttttctg | cattgcctaa  | gaatagtctt  | gtcttgggtg  | ttgttaacaa  | agtaccttta  | 18240 |
| aaagagagta  | gcgttactta  | ttactgtgtt  | ggtgtttcat  | tgcaatagtg  | aacagagttg  | 18300 |
| gtactgtcta  | gacatataac  | ttgcagggtta | gggaataagg  | gtcctgttac  | aggcaatgaa  | 18360 |
| gagacagaaa  | tgacagagaat | gtaaattggg  | aatctgcttt  | tcttcagtg   | ctgtgtagag  | 18420 |
| atcattgagc  | tagatatctg  | tgggtaaaaac | atggatgcct  | cctgttctgt  | cagaacacat  | 18480 |
| ggctttcccc  | ttctgattac  | acataaacag  | tctctgtccc  | atttcagtg   | aagtgtatga  | 18540 |
| gggtgtgacaa | gacagtacat  | aactttatag  | tatgacattc  | tctgcattat  | aagaatttaa  | 18600 |
| aggccacttt  | gtactcagca  | tgagatgttg  | aggcaagtct  | cattgtcaac  | taattgtcat  | 18660 |
| tagagcatca  | gaaatgttaa  | ctaaatagaa  | ttcactgtta  | tgtgtcctaa  | catacataaa  | 18720 |
| ccatgggtaca | aatattttac  | tgcatgtgaa  | aataaattat  | taagtatatag | gagacaatat  | 18780 |
| atgtaagaaa  | tgctgtattc  | tttaaatgct  | aatggattct  | tctttttttt  | tttttcttct  | 18840 |
| caccagaatg  | ttttaattgt  | agaagtaagt  | attatgtttc  | actttgaagt  | tetaacgtgt  | 18900 |
| gactggcaaa  | agaccttaaa  | cagaagttgg  | ttttaagtca  | gaaatcctac  | agtttctgg   | 18960 |
| ttcaagggtt  | ttagtttctg  | actctaggcc  | tgaattatgc  | aagtgggtct  | tgaatttgaa  | 19020 |
| attacctgtg  | ttttgaattc  | tcacatagca  | ttcaactata  | ccaaatttaa  | caaggaaaaat | 19080 |
| atctgtgtaa  | aattgcttgc  | tgtttgcatt  | tgatttcagt  | ttctgtgat   | tctgtcaatt  | 19140 |
| cttaaatgtc  | tttttttttt  | ttttaataat  | tcttattaag  | aagttagcctg | agcaaaagtgg | 19200 |
| tactaagtgt  | gttattgttc  | tgagtttata  | acagcaaaact | aggaaattaa  | actggaggggg | 19260 |
| tagaaagaag  | ttcctagtag  | gaatatccta  | aaatacatgg  | gaaaacagag  | ctaaggccaa  | 19320 |
| catctctgtg  | tacagtttct  | ttacaggcat  | ggagaatgtt  | gattccta    | ttgcattttg  | 19380 |
| ataatgctta  | attcccattt  | ctctatcttg  | tcgtgggtct  | ttgatactag  | gtacaatata  | 19440 |
| cttcagtctg  | atttccctagt | agctactgtt  | gcttaaatct  | tgcttccata  | cctctttgat  | 19500 |
| gaacgatgtt  | gcttacattc  | agtaagtatg  | acttttacia  | aggatgatct  | agaataagat  | 19560 |
| cgctaacttg  | aggttggcaa  | agcagaagta  | ctaaacagga  | aaaatgaatg  | tatagaagta  | 19620 |
| ataatgtctt  | cagactccaa  | atagtggaaa  | tgggtttgtc  | cataagctat  | ggcaactgat  | 19680 |
| tcagacacca  | aatgaaagga  | tgtttccctc  | tgtcattaaa  | tgtgtttttt  | ttatgtacgt  | 19740 |
| cttttttttaa | tttgagaagg  | ttatattatt  | ccataggaaa  | cacttttgac  | ctaaactgtt  | 19800 |

|             |             |             |            |             |             |       |
|-------------|-------------|-------------|------------|-------------|-------------|-------|
| ttagttcagt  | tctaaaaactc | ctacccaaact | agtatgtata | cagcagtcac  | ctgtgtgacc  | 19860 |
| ggtggcgctcc | tttgaagtgt  | atttgcttca  | gataggcaga | tgcatttaac  | tacgtatctg  | 19920 |
| ctccatttca  | gatgaatgta  | tgtcttcaaa  | gtcccaggca | gctaagtagg  | ttagtttctc  | 19980 |
| atctgttcag  | tgaatcatct  | ctccagtgc   | ttatatataa | ttatacattg  | tcaaatatga  | 20040 |
| accttttgac  | cagtcaccgt  | tttaactgtc  | aaacaatggg | ggaaaaacaa  | caaacaaaac  | 20100 |
| caaaaaacat  | tatttatagge | accagagcac  | ttagggtttc | catacagaat  | ccctttgatg  | 20160 |
| tataaaaaaca | aacaaacgaa  | aaagaagaaa  | agaatcccat | ccctttctgt  | gcaggggatg  | 20220 |
| ggattaagat  | ggtgtgtcaa  | acatcatttt  | agtgtatata | aatagaaact  | ttagaactgc  | 20280 |
| accagatat   | tgccagtgc   | gcttctttgt  | ctttggattt | ttttgggtct  | gtaatgtaag  | 20340 |
| ttgcacttct  | ttttccctta  | attgcattga  | tgggtgtgtc | tgaatgtttt  | cccgctctat  | 20400 |
| tttctttctg  | ttgtatacat  | ctctctctga  | gttcaactgc | aagactaatg  | ttccatttat  | 20460 |
| aatactggaa  | ttaaacaatg  | tttgtcaaat  | agctacaagg | aacacaagcc  | ataaactctt  | 20520 |
| cagaggggagc | ctagaagggtc | atagttcagt  | aactgctttc | aaagtagtag  | ttatctacag  | 20580 |
| ttttgactgt  | tgtgatcact  | ttactgagca  | taattgctta | agcacaaat   | agcgttatgt  | 20640 |
| atcacctgta  | tttgattatg  | aacattatct  | tgtcatatct | ttgtgtctat  | acggatgagc  | 20700 |
| aattttctgca | gtaaaaaactt | gaaaatgcta  | ctctgaaaat | tttaatatc   | tcacttacat  | 20760 |
| gacactttta  | ttttcaggat  | ataattgata  | ctggtaaaac | aatgaaaaa   | ttgctgtctc  | 20820 |
| tactcaagca  | gtacaatcca  | aagatgggtg  | aagtagccag | gtaaacttct  | aaatgtgatt  | 20880 |
| atttgtgtctg | tattttcaact | tcagaatgac  | acacagtggg | ataaaaaat   | aatatagaat  | 20940 |
| atatctccat  | agaagagaaa  | caggaaggaa  | catttctgca | ggaatcatat  | atgatggagt  | 21000 |
| tagagtctcag | tatttggttt  | aactcaactt  | ctgcttgttg | tgggagcata  | tgttcattaa  | 21060 |
| ggctaacaga  | aatgctcttc  | tttccagtcc  | tattgcttga | aaaatatcag  | tgccttatat  | 21120 |
| tctagtaata  | tttgtggggt  | tttaccactg  | atctttaact | taggtttaaa  | agcatatgga  | 21180 |
| agaattttaa  | tattttgggc  | ataaaaaaac  | acaacaacct | ctcaagggtg  | caggggtatt  | 21240 |
| ttcaacacaa  | atattagtcca | taaatataaa  | ctttataaat | ataaagtcca  | actagagatt  | 21300 |
| gggatttaaaa | aaaaaagtca  | gtgggtgacg  | tctatggaca | tacagtatct  | gaacatgttt  | 21360 |
| cagtttgcaa  | cttcatcttt  | tttcagtggg  | gtatctgcaa | tggagacctt  | tggctgtcca  | 21420 |
| gctttacttg  | tagaattaaa  | ttagtgttga  | ctaaacatta | ccgtttatat  | agagtatata  | 21480 |
| gtgtagctga  | gcttcagagg  | tcattctgtg  | agtcaccacg | ccatagtctc  | ttttaggctt  | 21540 |
| ttctacttag  | tgcactcttt  | aatcagttag  | ttatgaaacg | ttttctaact  | aaataactta  | 21600 |
| gatttagtaa  | gaaatgttga  | gggtaatat   | aaccaatata | aacattccctc | ctacttcttt  | 21660 |
| ttaaattgca  | tcttttatgc  | tccatgtctg  | tgatttgtga | ttttacaata  | aaagcataat  | 21720 |
| gatttgctggg | gatttagagat | aaaccagaa   | tgtgtgtaaa | atctaatcag  | ttgtgttgcc  | 21780 |
| ataacaatgt  | caaggcagtt  | gttagatgta  | ctagagcagc | tcttcttgat  | gataatgtgg  | 21840 |
| cgatgacctg  | ctgaagggtga | gggttttggg  | ttttttttct | ccgtggcctg  | cttaaatgta  | 21900 |
| ccagattttaa | aatattgtgc  | actgaatagt  | accttaacgc | tacagtgtac  | atcacagata  | 21960 |
| gaatgcagca  | tgaacaaagc  | ctagcttccg  | attgtttaaa | cctgttgacg  | tttgggtctg  | 22020 |
| ccgaagacct  | gagcaagtaa  | gcgtgatcgt  | gctggatagc | tttaagtgtc  | atgctcttag  | 22080 |
| ttcacttgcc  | tttcttttct  | taattttgtt  | ttctttctct | ctgacaatta  | agcagttgta  | 22140 |
| gcagggagat  | gtgaattaat  | gctaattgtga | ttttgggttg | gtttgatcaa  | gttttatgcc  | 22200 |
| actactcaaa  | ttaagactct  | caactagaag  | tttcagaaac | tgaatgtcta  | aagacagaca  | 22260 |
| aatgtcagga  | tgatggggtc  | atcttctgaa  | agtgtgcaac | tgaataatga  | gtgagtcaaa  | 22320 |
| ttaaactaga  | aagggtatttc | attctgtctac | agtaccccca | actctgcattg | tctatgcttt  | 22380 |
| ttacacctct  | gactattaat  | tctgtgaaag  | tggagctggg | ttatatcagc  | ttatgatcat  | 22440 |
| caaaaccaat  | tacattttct  | gtatgggggt  | aattagtact | accagtggaa  | aaacagctga  | 22500 |
| ggacatgttg  | cacaagtgc   | attgaggtta  | catttaacca | gctgaacttt  | tgttccataa  | 22560 |
| cactaacgct  | tttgaaaaga  | cctcagctgg  | tctggctaga | tcccaagttg  | tcacctcttt  | 22620 |
| acctgcgagc  | taaactgatt  | gcacctgagg  | tttctaaaaa | cacataaaac  | ctaattaaat  | 22680 |
| cacttctggc  | aatgaaaaac  | cttccagtta  | agtatctcac | agcctctttg  | ctgcaaacat  | 22740 |
| tttgtttatc  | ccttaggctg  | taaaccattc  | atgattctag | gaagtcttga  | gtttgtctat  | 22800 |
| atgatcagct  | gcacctgcaa  | atgaataata  | tgtttaaaat | gtatatagtt  | atgctgtcaa  | 22860 |
| ttatacacat  | gtagtctcac  | tgaagtggaa  | ttcagcta   | cttctgtgtt  | catataagtg  | 22920 |
| aacctgggtg  | ttttccacgt  | ttcagtttta  | cattgttttt | atttgacagt  | ttgttggtaa  | 22980 |
| aaagaactcc  | tgaagcgtg   | ggatatcggc  | cagactgtaa | gtgacttttt  | gacacccatt  | 23040 |
| atcagttttt  | aaaaatatgt  | tgtatatatg  | ttgatgtatt | gtgtactctt  | tgggtcaaagt | 23100 |
| ccaaggggtg  | tccaagtaag  | tttccatgaa  | tgggttttga | tttattttgc  | tottatttgg  | 23160 |
| gaagttttaa  | tatttgggaat | gtattttgac  | agagttaata | cagatttggg  | ctagtccttt  | 23220 |
| ttttcttaac  | tttttgggtg  | attttattat  | acaaaaagga | aactttgata  | ttttagtcca  | 23280 |
| agttaatccc  | tcttcatatg  | gaaattctcc  | tagcttgcac | gttaatcaaa  | tgggttgggtc | 23340 |
| agcagaagtg  | tcattcttga  | cacttttggg  | agtgtccag  | gtgcctaaac  | atagccaccc  | 23400 |
| agtgcacctc  | ggatgtcccc  | agaaatgtaa  | tgggaccaca | cagggtaaca  | gatctgatca  | 23460 |

|            |             |             |             |            |             |       |
|------------|-------------|-------------|-------------|------------|-------------|-------|
| aagtttacag | tggttaagatg | ttcaaagtag  | cacaaacccc  | atcttgttta | agcatctgga  | 23520 |
| ttactcctgt | catgtcttct  | tgtctgtccc  | aagaccataa  | aggaaggaag | agcatccaaa  | 23580 |
| gcacacgggg | gtgtcagtgga | tagccctctt  | tggtctggaa  | agggctacgt | gtgctatccc  | 23640 |
| tgtgccagag | gttatttctt  | gttcattgtg  | actgtgcgt   | ttcacaatgg | aagtttaaca  | 23700 |
| gaattgcatt | ttctttcccc  | tcctttccaa  | gtttgcacag  | tattctttaa | ttgaaatttc  | 23760 |
| tcagtgcctt | gaaagcttaa  | ttcaggtgtt  | tcctctactc  | gaggatgcct | ctctttccctg | 23820 |
| agatcttatt | atgggtttag  | ctgggttaac  | agtctgcctt  | tctgagcatt | ttgcagcaga  | 23880 |
| tgacttggtt | tgccattctgt | ccagagacga  | cctgttcagt  | gcagaacccg | tgagctctga  | 23940 |
| aggcaagaga | caacgcagca  | gcagagcttg  | ggcctgcctt  | ctgtccttgt | atcatgaggc  | 24000 |
| caagtagact | ctaaacattg  | ccttattttt  | tcacagaagcg | tacacagtct | ttcactggaa  | 24060 |
| ttgtgtgatt | tatctccgtt  | tttttcatta  | ctgagcctcg  | ctgtaaccag | ggagagagaa  | 24120 |
| acagcacagc | tcggggagtct | ttcatgttta  | gcataagggg  | agaaagcctt | ctggaaacgt  | 24180 |
| ccttaggttg | tcttacagtg  | tgtgtggatg  | gcttggaactg | atacttcttg | actatccaa   | 24240 |
| gactgttctg | ctagagccat  | gtccacttgg  | actatctgtg  | gaacagctga | tgtgtttttt  | 24300 |
| ttaggcagct | tctatgggaa  | aaaaaaaaaa  | cgaacaaaaa  | cacaacaaaa | aacccgcctg  | 24360 |
| tttctgaaaa | gtgcaagtga  | caccgtaaat  | aagagagcgc  | ctttaccctt | gtggtcttca  | 24420 |
| ctattctcta | taccaagggc  | ttggcactta  | cagcgtgttg  | ctctgcagag | ctgtcccgc   | 24480 |
| gtcagagtat | gtcctggtag  | accaggttgt  | tgtccacgag  | cagccagagg | ggagtgaatt  | 24540 |
| ttgtttggca | ctgccctcaa  | gacaaaaggac | tggttccctt  | gcgttaactc | tgttctctcc  | 24600 |
| tggtgttgac | gtgttccatg  | cttactctcc  | cagctcttca  | caaagagccc | agttgggctt  | 24660 |
| gcacacctgt | agttttgagg  | gagctgagag  | ctgaacctat  | ctgctcggtt | tcccagaggt  | 24720 |
| gaccttactt | atagctgtgc  | ccaggctgag  | ggcacactca  | cacaggtgct | accttttgaa  | 24780 |
| taatactgtg | agggtgacaa  | ctaaaaacaa  | ataaaatgct  | actgcttggg | aaccagtaag  | 24840 |
| taaatctatt | tcatttttgt  | ttagttgaat  | taaatcaaca  | aataagagtt | gaccattaac  | 24900 |
| tagcttgaat | attaatctcg  | ggatacttga  | atggaagaaa  | tgaaatgtta | tgtttatatg  | 24960 |
| tctataaaat | tctagattct  | ctctgtaaaa  | aggtctagtg  | aactctagga | agttcttttc  | 25020 |
| taacatactg | atgacttatg  | ttctttttca  | gttgttggtg  | ttgaagtgc  | agacaaaatt  | 25080 |
| gttgtgggat | acgccttaga  | ttacaatgaa  | tacttcagag  | atttgaatgt | aagtaatttt  | 25140 |
| tcccgtatgt | ttttgtccat  | ttaaaaactag | ggaggaggaa  | ggaaaggaag | tctcctcgca  | 25200 |
| tcactctttc | tcttctgtgt  | aaatcacagc  | tcatgtgcaa  | gcctgttcca | ttagtgtatg  | 25260 |
| aacgacgaga | gcaaaggcta  | gaccagcaca  | acagtgtttt  | ggagctcata | tgaaactaac  | 25320 |
| ttttgttttg | cttcatgcag  | aatggaggta  | ttctgcatga  | agctgtcagc | tttgaaatac  | 25380 |
| tgtgtgcagt | tgtatgaagt  | ttcttatctt  | accagacata  | tctgcaagaa | acaagtgcat  | 25440 |
| ttctctagtt | acatcaaaat  | tagagaggag  | agaccttgga  | tgttgagatg | tttgactttt  | 25500 |
| atgctaggtc | attattaatt  | tttcatctgg  | gtctcactaa  | ccttataaca | gacatagggg  | 25560 |
| aggagagaga | gagaaagtcc  | agtgttgaga  | tctgtagaca  | aaaaaaaaaa | gaagctaaag  | 25620 |
| ttttcttgca | taaacaaat   | cacaaggctc  | ttcggtagctg | aaatgtgtat | agcatgcttt  | 25680 |
| tatttttttc | agaagtgcac  | tattgaatac  | atctgtaaac  | atcatcctac | attactttct  | 25740 |
| aacaggagct | ttttattttg  | ttttgttaca  | gcattctctg  | gtgatcagtg | agacggggaa  | 25800 |
| gcagaaatac | aaagcatgaa  | agcatagtcc  | aagtgtcttg  | atagcagagc | tttgaaatgt  | 25860 |
| tttgtttact | gagtctctac  | tttcacaggc  | ttcaactctg  | gtacaccagc | taaaattgt   | 25920 |
| gaattatcca | cccccttgt   | catatgctta  | ctataactta  | ttgcatgtac | aatatacaaa  | 25980 |
| tcttgccttg | ttcattttata | ttttagaaat  | gtaaacaatt  | agtgcatttc | tgcactcctt  | 26040 |
| attttgattt | gcactatgac  | tctaccgact  | attgctgccc  | ctggttgggt | tgtgctgttt  | 26100 |
| gtgagctccg | gagtctaaat  | cttgcagtg   | taaatgtgctt | aaacctctac | aaccagaaac  | 26160 |
| tgaaatagtt | caagtactgt  | aaatgtaaaa  | cactttatga  | taggggaagt | tattagtaat  | 26220 |
| atttttaaaa | tctgtaattt  | aagttttata  | ttttcatgaa  | caagtgtgat | tgtgattaat  | 26280 |
| ggatagctgc | acctttgggt  | gttataaaac  | atgaggagca  | gccagttaca | gtatctgtaa  | 26340 |
| tcaccaagga | gctgattcaa  | gcctcctgga  | gttgcccttat | cgtgttaaaa | gtccgggttaa | 26400 |
| aaaaaatctt | ttttttttta  | atttttttta  | tttttttgagc | taaaaggaga | tggaaatgaca | 26460 |
| aagcagaatg | tttaaaatct  | agtttaagtt  | gagctattca  | tttcttttgg | attttttctt  | 26520 |
| taaaaatacc | catcagacag  | tgaatttgcc  | tttttaattt  | aaacaatttt | aatatatatt  | 26580 |
| tgaagaagta | ttgtaattgt  | tactttataa  | gatctaaaaa  | acaaaaagta | ctttaaataa  | 26640 |
| aggctgtctc | tttaaaataa  | gcccataaca  | tctattgccac | tcattctgta | tattaaagtg  | 26700 |
| ctcttgaaat | tttcttagta  | atatttttca  | taagtgtttc  | tgacagtga  | ggcagactct  | 26760 |
| acgtttgtta | tctttgtaga  | gccttttagtc | agtattgtgc  | catctacaag | agatgatata  | 26820 |
| ggcctgaggg | aaagctatcg  | aaatacttgc  | gtattctata  | accaaactgg | atttcagaga  | 26880 |
| aaatgcacta | agtagtgtcc  | cgtattctgt  | tttggc      |            |             | 26916 |

<210> 4

<211> 18700

<212> ADN

<213> Cairina moschata

<400> 4

|             |             |             |             |             |             |      |
|-------------|-------------|-------------|-------------|-------------|-------------|------|
| agcaaaacga  | agaattgctt  | ggaagatgct  | tggttttgtc  | caaataaata  | gctgctgaga  | 60   |
| gtgagagagt  | gagtgaagtaa | tattcagtat  | ttttaaaagt  | tcgagaaatg  | taatggggaa  | 120  |
| aaatttaaaa  | taaatgtttt  | ttgtttgttt  | taaaaggagc  | tggtgtgctt  | tatgttgaca  | 180  |
| tggttgattta | attactgcgg  | taggtattta  | ctgtgccttg  | ggaaacggta  | ccccagtc    | 240  |
| acacagactt  | atttgtcagg  | aaaaaaattg  | tcatagatca  | tcttcaaata  | agagttgaca  | 300  |
| aataatcaaa  | tcttgatgga  | gtataattat  | taatactgtg  | attacatata  | catctttgca  | 360  |
| aggggttccct | gaaaaggta   | gttttaagtc  | ttctattgtc  | aaatctgatg  | tttgagtag   | 420  |
| ttcgctacat  | gggtgttgat  | gccatgtaaa  | tagttaccag  | atagacgttt  | tattttatgt  | 480  |
| gctgtatgtt  | tttgtttttc  | attcagtcag  | ctggaatcat  | tgaaacagag  | aagggtttct  | 540  |
| aagaaatata  | ctgaacctgt  | tttgtggatg  | tcttcattgc  | aatagctggg  | cactgaaatc  | 600  |
| ttggaataac  | gacagaggaa  | tccataggca  | caggaccaaa  | cacctttcat  | tgctctcata  | 660  |
| tagcgtgatg  | ttagggcaga  | gtgggtagta  | cttcagtggt  | tctgtgtccc  | atgctttaat  | 720  |
| gaactctact  | taactctacca | tatgtaagag  | acttgcagca  | caggcaaaac  | aagggaaggt  | 780  |
| atctttgtgc  | ctagataatg  | taccacatat  | ggtaaaacaat | tttcaagcct  | cagagacaaa  | 840  |
| agaaggatgc  | tgtaaaagtct | caagtctcag  | cttgtgtctc  | tgtttccacc  | ttctgttct   | 900  |
| ttctgttcag  | acctctgggc  | cagtaattta  | aactggaaaa  | ttaatgggaa  | agagaaactg  | 960  |
| tttgtgagcc  | tatagaaaaga | tcaaattgtg  | tcagagtaat  | gttgctcctt  | ctacaccag   | 1020 |
| ataaactttt  | tgagcagagg  | agaagcagtg  | aacagagctt  | tactttcatg  | actttgcata  | 1080 |
| agaaaacatg  | ggaatgtggt  | cagagaccaa  | tttaaaaaata | gagggtttga  | aaacttggtt  | 1140 |
| ggaaaacaaa  | acttgcttgg  | aaaacctat   | tcagagctgt  | gaatcattca  | cagacaacct  | 1200 |
| cttaggggtg  | tagccactca  | tcagctgaat  | atgattcaac  | gatatgctga  | aaaaaaaaaa  | 1260 |
| cacaaacatt  | gttgtggaat  | gtatgaacag  | gaatgtagtc  | tgtaagatgc  | gtgtcttggt  | 1320 |
| cttcttctct  | taggtaatat  | tagaacacct  | ttctgggaaa  | atggtatcca  | gctgttgtag  | 1380 |
| gcactgcact  | ccatggaaaa  | tgtgaagaaa  | gtggaataaa  | tccagagaaa  | agcaatgaga  | 1440 |
| gggatgacag  | gataaaaaat  | gagccttgct  | ttaagggtgat | gagagaaaaag | tgtttttcaa  | 1500 |
| atctgtaaga  | gagccacaga  | gaagaagaa   | agaaacagtt  | ctcatcattg  | caggtaggac  | 1560 |
| aagaaaaata  | ggtttaagtt  | gctgcaggaa  | gcctttagct  | tagacatcag  | aatgggtaat  | 1620 |
| tgctcttgac  | agtaggaaga  | attaaagctg  | agcatgcact  | tctgggagag  | gtcgtggaat  | 1680 |
| ctcaggaatt  | aggtttttaa  | gggaatatct  | gtaggcatct  | tttggggctg  | gtgcaattgt  | 1740 |
| aggttgaggg  | tgctggggat  | aaggctaata  | gaaatcccta  | ccagccactc  | taaggccttc  | 1800 |
| tgcaaggagt  | cagaaccact  | tttcggaagt  | aaaactctgt  | actggcagag  | gtcctgtgct  | 1860 |
| tgaccttaag  | gctcagagcc  | atgcatttcc  | gtttctatct  | tttacttaa   | aatagcactt  | 1920 |
| gtggtagtaa  | taacatgctc  | ttgcggagtg  | gcacaggtct  | gcattgggtat | gaactttctg  | 1980 |
| tcaccaccca  | gcaaatggca  | ttttccactc  | ttctgtttca  | gattttcggt  | agtaacttct  | 2040 |
| tactcccaaa  | ttcctgattg  | aatacgactg  | aaaacttcaa  | cacgttttta  | gtgcacgaag  | 2100 |
| tgtactttat  | acaaatgtgt  | gggactattg  | cacacaactt  | actgattttc  | ttcactgtgt  | 2160 |
| gcacatgtgc  | tctgtgatga  | tacagagttt  | gggttgactg  | aactgttacc  | cagtctttac  | 2220 |
| cgaattaggg  | cagcgatctt  | aaaccttcat  | ctgaactttt  | gctagagatg  | atcttctatt  | 2280 |
| ttgttttagac | agggtttctt  | gcttgcttgc  | atttgttcta  | aaacgacagt  | ctggatgaga  | 2340 |
| agaaaccaac  | cagggtctgg  | actgccttac  | cttttatctc  | ctggataact  | tacttggcac  | 2400 |
| atcacttggc  | acacatgcaa  | tacctttttt  | tcaccccggt  | acccaaagaa  | aagacgtgct  | 2460 |
| cctgcacaga  | ctcagccaaa  | ttcctgcttg  | ccagtggagg  | gcgttacaga  | ggcttggggc  | 2520 |
| agagggagga  | aactcactcc  | caaagcatat  | actaaaaata  | ggccaatgca  | ttgcaaacag  | 2580 |
| ctttgctgtg  | agagcttctt  | tctggcatgt  | agccatgcag  | ctgccttata  | gttttgagcc  | 2640 |
| tagatacctc  | caaaaacaaa  | caaaaagtgc  | ctaaagttaa  | atgcagccag  | cagtgtttat  | 2700 |
| agtcttgata  | atccctacag  | tgtatcaatt  | ttagtttctt  | ccagagtgc   | ggtagtttac  | 2760 |
| tcaatttaac  | ttcaaatgtc  | tgagttacac  | cattgacacc  | attgtcaaga  | ctagtgtctg  | 2820 |
| caatccctga  | agtctgcca   | gttcaaaagca | gggaggaggg  | ctttaaagt   | gaaatactcg  | 2880 |
| ggtagcgtgt  | ccaaaaagtt  | ttcctgtggc  | ttcctgtggg  | agtaggaact  | gcagtagcta  | 2940 |
| ggttgcaaaa  | agttttttaa  | ctctccaaat  | tgctgttaac  | taaacaacct  | taaatccctga | 3000 |
| catgagagta  | tgaaatgaga  | aagtcagttc  | cttgatttag  | actagcttga  | aaagaataaa  | 3060 |
| ctttttgcac  | aggactccct  | ggacagttgg  | gctggataat  | agaagtgaat  | cattcaggtc  | 3120 |
| ttgatgccta  | caagtttctt  | gggttttggc  | atattttcca  | gtgatttgtg  | acatcatttt  | 3180 |
| tgccctaaaa  | taaatcgcat  | catgtcctgc  | cgtatgtttt  | gactatgaaa  | atggaggttt  | 3240 |
| cgggttgaaat | accaaacttc  | gttgttactt  | ctttcagagg  | ctgttgataa  | aggaaaaact  | 3300 |

|             |             |             |             |             |             |      |
|-------------|-------------|-------------|-------------|-------------|-------------|------|
| gaagctttta  | tgtaaatgat  | catgatttct  | catttgcac   | aagaatattt  | tccaaaatag  | 3360 |
| tagttaaatc  | cctccaatga  | ttttcagcta  | tgtcacataa  | atcacccgaa  | acacggtaat  | 3420 |
| tttgcacggg  | tgttttttct  | cccttttttt  | attgtctaata | actcagctta  | tatttgagtc  | 3480 |
| ttctgtatcc  | aattacctga  | tttgtgtggc  | aattaaacag  | aagatcagcg  | tcattgtact  | 3540 |
| tcacattctt  | tttaaaactca | tcagtagtat  | gaccacacta  | gttggctctg  | taagtttgta  | 3600 |
| attttaagatg | tgcctggaag  | gggttcaaag  | agacaggagc  | agtagtgtgt  | ctcacttccc  | 3660 |
| tttggaaagc  | ctttctgata  | ctgtcaagca  | atattttacat | caaaaaggaa  | ctgaattctg  | 3720 |
| gtaaaactttt | ccagactcca  | agtaaaccca  | acaacctgag  | atttagataa  | accttggtgg  | 3780 |
| atttaaaaaac | ttgatgggat  | cactgtctaa  | gacttgtgtt  | agcccatcag  | atattgcaga  | 3840 |
| cgtatatttt  | gaggaaaggc  | tttcagttaa  | ttgataaaat  | taagagcaga  | gttttgga    | 3900 |
| aaaaaacccg  | catctctgct  | cagtacttca  | tggcatcttg  | aatttctgca  | gttagaggtg  | 3960 |
| caggtgctct  | ttctctgcaat | taaccctgct  | tgttaattaa  | aatgtactgt  | ttgctaaata  | 4020 |
| tgattagtgt  | ctggagccag  | gtagcttgaa  | aaagcttcaa  | gattatttct  | ttttaccagg  | 4080 |
| accatcaagt  | ttttagccct  | cccttttagag | gaaggttcag  | aggaggtcct  | atttaactctg | 4140 |
| tcccaagaaga | aaggctattt  | ctcccttag   | aatggggtct  | ggcagtcctt  | ttcagagcac  | 4200 |
| ttttcatcct  | tgtctcatag  | ggatttttaa  | ggacagattt  | ctgaaggat   | tttccaccgg  | 4260 |
| gtaaaagcggg | atcctgaatg  | ggtagacagat | aatacgtagg  | aattctgggt  | tcctactgcc  | 4320 |
| aggttggttta | tcttgagatc  | ttaaaaatgg  | acttatttcc  | ctttttaaagt | gtaaaactgag | 4380 |
| ttcttgaggc  | accttttcca  | ctgaagtatc  | tgaacttggc  | cttactgccc  | atctcagctt  | 4440 |
| tgcacatttta | acccaaatcc  | agatccaggg  | tgtgggtgta  | ctcctagaac  | agctcttact  | 4500 |
| gtttgaggat  | ctgaatgact  | attttgcag   | cttgtgccat  | acaactcgag  | aggacctgca  | 4560 |
| aaggggaggg  | gagtcctgat  | gggcagttcc  | ctccagcct   | ggcatgcgct  | tcgtgtctcc  | 4620 |
| gcttccctta  | taactctgccc | agaaaagatt  | cttgaaatac  | ttgagcagct  | gccacaatct  | 4680 |
| gcattggagg  | cagcagcctt  | ggcatgtgag  | ctctgccacc  | tgatgcccgg  | tctgtactga  | 4740 |
| ggcagtgcca  | gctggatctg  | caagcctcag  | cgaatacagc  | cttcgtcagc  | gagttctctg  | 4800 |
| atcagattgc  | tgccttggag  | cccttttctg  | gtcttctcca  | gtcttctccc  | tgttactcag  | 4860 |
| atagccttac  | cacctctctt  | catcatcttt  | tgagtttcta  | tcctcacgtt  | acagccaggt  | 4920 |
| aagcccatctt | acccatgctg  | atatcaacag  | gttccagatt  | ttttaaaaaa  | caatcctgtc  | 4980 |
| tggctcacag  | ctagggagcc  | tctagtgaac  | cagctaaagc  | ccctgcagcc  | atttagaaat  | 5040 |
| gtttctggag  | ggatgcaggt  | acgtcttgca  | ctcaggtgcc  | taatgaaagc  | agattgacct  | 5100 |
| gttcattcct  | tcgaaagcaa  | ctaagcaaaag | tggggaatcc  | catgtctctg  | ccccacttct  | 5160 |
| tggttctctga | cggcaagacc  | tcctgccagc  | accaaccgcc  | tgtgtttgga  | gtcagccagc  | 5220 |
| cttcggttct  | gactgaagtt  | gaaggtacag  | tacaaaactgc | tggctaaaag  | aagctgcttg  | 5280 |
| acttgctttt  | ttaggaggct  | tgtcttgtgg  | ggaaaaaaaa  | aaaaagtaca  | gataaattct  | 5340 |
| gttgattcct  | tctgagtcac  | ccagcagcga  | tggcaaagct  | aggagaaaaa  | atgggatcac  | 5400 |
| agagggattg  | atgaatggac  | agttagtgac  | ctgaagatct  | aactcaagg   | acgttacttc  | 5460 |
| aagtgcgaagt | gtcaacatat  | aaatgctgag  | ttgtgtcttc  | caagtacccc  | actgttgggg  | 5520 |
| tatcctgggt  | accaagcaag  | gttttagtaa  | gtctgttcga  | aaagttacat  | acagctgaca  | 5580 |
| ttcatgcagg  | ccatgatttg  | ccaggcctga  | gctctgaggt  | attcctggac  | ggttagagag  | 5640 |
| gataaaataa  | tttagaagtc  | aggaaccac   | tcagagacaa  | accattattt  | ataaaattat  | 5700 |
| tttgcttctt  | aaaagatgtg  | cgttccagag  | aaatcacaa   | gttggctcca  | acctttttgc  | 5760 |
| atcatctagc  | tagaaaaatgt | gcagcttcca  | gataccgcca  | gacacacctc  | cttgacctg   | 5820 |
| ccctggaaac  | atcaatggct  | ccataaccag  | taagaacaaa  | actggctcct  | tactaaatcc  | 5880 |
| gttgctgac   | tcagctaacc  | aacttctctg  | aatatcagca  | aatatttctg  | cttttgtaaa  | 5940 |
| tctgttcttc  | cttccataat  | gggtgcccag  | caggaaatgt  | tgggatccag  | aaatgcgctg  | 6000 |
| gttggcagga  | agaacaaaga  | aattgttcac  | ccagaagggc  | aaaaaaggac  | taatgctttc  | 6060 |
| gtccgcttgg  | gagaagttgc  | acagaactta  | tgcacaaca   | ttttgcacgc  | aatcagaaag  | 6120 |
| atgcgcgtgg  | atgttacttt  | taatagcaga  | cgtaatatc   | agttattaat  | tagaaatgtc  | 6180 |
| ttcagtaacc  | agactaaaag  | cagatcctga  | aacactcctg  | tgggtgaaca  | gtctcttgac  | 6240 |
| agacactgcc  | cactgtccag  | taatgtcagg  | cgctctctga  | acttgacagg  | gcagctgctg  | 6300 |
| tttttcccag  | ccttctggaa  | ataggccagc  | tctgacatgt  | ttctgatatt  | agctgggtgt  | 6360 |
| atttcattct  | gtgcctctag  | gcagtttgat  | aaaggtcctt  | tcctcctgct  | ccagagctaa  | 6420 |
| cccacccaaa  | ctgtagccga  | gcacctcgct  | gcaacaaaac  | tgcactcgct  | aggggtctgc  | 6480 |
| ctgtttgctt  | cattaaagtc  | tgetttaaatt | gtttcgtaca  | aaggaaacact | caaactgatt  | 6540 |
| tctgagccca  | aagtagcagt  | gctaggtgta  | catcaggagt  | tgtttggcat  | gaagaaacat  | 6600 |
| tgccatggca  | ctgtatgaat  | aaagttattt  | ttagaatca   | ttatcccttc  | cttgatacca  | 6660 |
| agtctttatg  | cggcagaaaa  | tcaaaacttg  | tctccacctt  | tacagaaagc  | agaggaatgc  | 6720 |
| tttcagctga  | tagttgctta  | agctagaata  | taagaaacca  | tgaatttctg  | tgtgcactgc  | 6780 |
| ggcattgcc   | ttcattccag  | acctacagaa  | aaaaacgaca  | cttttggttac | tatttttttc  | 6840 |
| cttcccatat  | gagaccaggg  | gagctaccca  | ggcatttcca  | ttcttataat  | tttacctcaa  | 6900 |
| gatcaaatct  | tctccaggca  | gttaaaggca  | gctgcaccgc  | gagacctcgc  | tcagcctccc  | 6960 |

|             |             |             |            |            |             |       |
|-------------|-------------|-------------|------------|------------|-------------|-------|
| cttgcattccc | acggagctgc  | gttttagtgag | aaacctcccc | cgaggtgacg | ggctgcaggg  | 7020  |
| gaccccttcc  | cacacgcgtc  | cccgtccctt  | ttctcagtg  | aaacgcagcc | accgcccctt  | 7080  |
| gaacccctcc  | ccgggtcttc  | tcgggtccgc  | ggagggcagg | ggggggccgt | cccgcgcgcc  | 7140  |
| gggagctcct  | cacagggccc  | gggccccggg  | gcggagcggc | cgccggccat | ttgagggcgg  | 7200  |
| ggagcgcgga  | agggcggccc  | ggggcccgct  | cggggcgcgg | cgccctcccc | ctccccgcgg  | 7260  |
| ctcctgcgcg  | ccgcgcagcc  | caacacccgg  | cccggccccc | gcggggccgc | ccatggcggg  | 7320  |
| cgcgagagccc | ttcgggcgcg  | tgctggggcgc | cctgcgggac | tgctacgcgc | aggcgggccc  | 7380  |
| gctggagagcc | ttcctccggg  | ggctggggga  | gagcggcggc | gaggaagccg | agggtggtgcg | 7440  |
| ggacgacgac  | gcccgcctgct | accgcacctt  | cgtgtccccg | tgctggtgtg | gtgtccccc   | 7500  |
| cgcgccccc   | gacatccccc  | ggcccttcag  | cttggagcag | ttatctagtc | agagcgaagt  | 7560  |
| catctcaaga  | gtcatgcaga  | ggctgtgtgg  | gaaaaagaag | aagaacatcc | tcacatattgg | 7620  |
| atactccttg  | ctggatgaaa  | acagttctca  | cttccaaatc | atgccgctct | caaacgtgta  | 7680  |
| cagctacctg  | cccaacaccg  | caacagaaac  | catgcgtatc | agtggcctct | gggaaacgct  | 7740  |
| gctgagcagg  | ataggggatg  | acgtgatgat  | gtatttattg | gagcactgtg | caatctttat  | 7800  |
| gctggttccc  | cctagtaact  | gttaccaggt  | ctgtgggcaa | ccaatttatg | aacttatttc  | 7860  |
| gcaaaatgta  | gaatcagccc  | cagcgtttgt  | taaacacacg | ctttcaaagc | acaaacgtag  | 7920  |
| tagcttgctt  | aagtatacgc  | agaaaaggct  | aacgtttcac | agacagtatc | tttcaaagtc  | 7980  |
| acgtcagtcg  | aaacgcaggg  | aaagacttga  | agctaattgt | tccagcatga | gaaataaaa   | 8040  |
| cagcaataat  | atacaaacgc  | tagggctcgc  | tgctctggaa | aaacagagta | gctccaatgc  | 8100  |
| aggtttgtca  | gctacagcac  | catccttaaa  | agggagcctt | gctagggaac | aactggaagt  | 8160  |
| cacggctaag  | agagcaagat  | tagaagagaa  | agagagggag | gaacaggcct | gtaatactgc  | 8220  |
| tcctaattgta | aaccagagta  | ttcccaagag  | gtatggaaac | agctgtgtag | catcacgttc  | 8280  |
| tgtaagtctt  | attaaagaaa  | aatacatttc  | tcaaagaagt | aacagtgata | tgctctgctc  | 8340  |
| ttcttttagtt | cacaattctc  | atcatgggaa  | gaagtcctgt | gcagacaaaa | gctctttcct  | 8400  |
| gcaaggagct  | gagagtaaca  | gacattttaa  | cccagcattt | gaaatgcaag | caggatccag  | 8460  |
| caggaagaga  | gtagagatac  | acaggcctat  | acctcggttg | gattggatac | caatcgaacc  | 8520  |
| ggcggaaggt  | agttcttcag  | gacacaaaaa  | gcaggaaagt | cccctagctc | atctggcaga  | 8580  |
| ggagttacca  | aatagggttt  | tgccatctac  | aataacattt | gacaggaagt | ttcttctgta  | 8640  |
| ttcttcgcagg | tactgggggg  | aacgtttccc  | aaaatccttc | ctattgaatc | gcctgaaggg  | 8700  |
| tagtgaggca  | ggtgtaaacg  | gactaataga  | aacgatattc | ttaagccaaa | atccgtttgg  | 8760  |
| gcaaaagcgc  | aaccaaggtc  | tgccacagaa  | aaaatggaga | aagaagaagc | ttcccaaacg  | 8820  |
| cttctggaga  | atgagaagta  | cgtttcaaaa  | actcttaaa  | aatcatggaa | agttccctta  | 8880  |
| cgtagctttc  | ttgagacaaa  | attgccctct  | tcggatatct | gaaaccattt | tgggaaaagc  | 8940  |
| caagctgctc  | agtcgggcac  | ctttgcctgg  | gcaagcagag | gctcacaage | aagcagaaca  | 9000  |
| gcttgggaag  | gagcctgcta  | agcgtgtggc  | aagcagcaga | tgcgaaatct | gtcacacca   | 9060  |
| cgtgcccagc  | agcgtacgcg  | ctcctctcgc  | agcatctgcg | tgctggagc  | caggggggga  | 9120  |
| ggagcagatc  | cctgcagagg  | cgtctgattc  | agtcctcagg | gagcttctca | aggagcactg  | 9180  |
| cagccacttc  | caggtgtacc  | tctttgtgag  | ggagtgccgt | gagcgggtga | tccccgcga   | 9240  |
| gctctggggg  | tcaaacata   | acaagcgcg   | gttcttcaag | aacgtgaaag | cattcatttc  | 9300  |
| catggggaag  | tacgctaagc  | tttcccttga  | ggtgttgatg | tggaagatga | gagtaaatga  | 9360  |
| ctgcatgtgg  | cttcgtctgg  | ccaaaggtaa  | tcactttgtt | cctgcctctg | aacaccgtta  | 9420  |
| ccgtgaagaa  | attttggcta  | aattcctata  | ctggctgatg | gatacgtatg | ttgttgagtt  | 9480  |
| gctcagatca  | tttttctata  | tcaccgagac  | catgttccag | aaaaatagtc | ttttctacta  | 9540  |
| ccgaaagtgt  | atttggggca  | agttacagga  | cattggaatt | agaaagcatt | ttgccaaagt  | 9600  |
| acagctacgt  | ccttttaactg | cagaggagat  | gggaagcgtc | catcagaaaa | aataaccttc  | 9660  |
| tatggcatca  | aagctccgtt  | tcattcccaa  | agtcagtgga | ctaagaccca | tcgtcagaat  | 9720  |
| gagcgggtgt  | gttgaagcac  | aaacgttgag  | caaggaaaag | agagcaaaag | agatgaatca  | 9780  |
| ctacaacatg  | caactgaaaa  | atctatttag  | tgtgttaaat | tatgaacgaa | ctataaacac  | 9840  |
| cagttacatc  | ggctcttcag  | tggttggggg  | agatgatate | tacaagaagt | ggaagacatt  | 9900  |
| tgtaaaaaag  | gttctttaa   | cagatgggtg  | aattcctcat | ttctactatg | taaaggccga  | 9960  |
| tggtgtccagg | gcttttgata  | gcattcctca  | cgataaaact | gtggaagtga | tttcacaggt  | 10020 |
| cttaaaacct  | gagaaaaaaa  | ctgtctactg  | catacggcgc | tatgcagtgg | ttatgatcac  | 10080 |
| tggaaagtgg  | aaaaccagga  | agttatacag  | gagacatgtt | tctactttca | aggattttat  | 10140 |
| gccagacatg  | aagcagtttg  | tgctccggct  | tcattgagag | acctcattgc | gagatgcaat  | 10200 |
| aatagttgaa  | cagagcttaa  | ctttcaatga  | gacaagtgcc | agtcctattt | atttttttct  | 10260 |
| tcaaatgcta  | aataataaca  | tcctggaaat  | tgagcgcagc | tactacttac | agtgtctctg  | 10320 |
| aattccacag  | ggctcccttt  | tgtaaacctt  | gctttgcagc | ttgtgctatg | gagacatgga  | 10380 |
| aaacaaatta  | ttcagtgggg  | tacagaagga  | tggagtcctg | atccgtctca | ttgatgactt  | 10440 |
| tttgcttggt  | acaccacact  | taacgcagtc  | aagaactttc | ctaaggactc | tagcaatggg  | 10500 |
| cattcctgag  | tatggctttt  | tgataaaccc  | caaaaagacg | gtggtgaatt | tttctgttga  | 10560 |
| cgatatccca  | gagtggtccg  | aatttaaac   | gctgccaaac | tgctgtttga | tcccatgggtg | 10620 |

|             |             |             |             |             |             |       |
|-------------|-------------|-------------|-------------|-------------|-------------|-------|
| tggtctatta  | ttggatacac  | agacacttga  | ggtttactgt  | gattactcca  | gctattcctg  | 10680 |
| tacttctate  | agatcaagtc  | tttcttcaa   | ttcaaacaga  | acagctggga  | aaaacatgaa  | 10740 |
| acacaaattg  | gttgcaagtc  | ttaaactgaa  | atgccatggc  | ttgtttcttg  | atttacagat  | 10800 |
| caatagcggt  | aaaacagttt  | tcattaatgt  | ctacaagata  | tttttacttc  | aggcttacag  | 10860 |
| gttccatgcc  | tgtgttatte  | aacttccatt  | caaccagaaa  | gttaggaaca  | atcctgattt  | 10920 |
| cttcttcaga  | gtcatcgctg  | agaatgcate  | gtgctgctat  | tctatgctga  | aagctaaaaa  | 10980 |
| tccagggttt  | actttaggta  | acagagggtc  | atctggcatg  | ttcccttctg  | aggcagcaga  | 11040 |
| gtggctctgc  | tatcatgctt  | tcactgtcaa  | actgtcaaac  | cacaaagtgt  | tttacaattg  | 11100 |
| cttgcttaag  | cectgaagt   | tctgtatgac  | acagctattc  | cggaagatcc  | caaaggatac  | 11160 |
| taaggcacta  | ctgaagacag  | tgacagaacc  | atctatttgt  | caagatttca  | aagctatcct  | 11220 |
| ggactgatct  | agagtctcta  | gtctagacta  | gagctcgctg  | atcagcctcg  | actgtgcctt  | 11280 |
| ctagttgcc   | gccatctgtt  | gtttgccctt  | cccccgctgc  | ttccttgacc  | ctggaagggt  | 11340 |
| ccactccac   | tgctcttctc  | taataaaatg  | aggaaattgc  | atcgcatctg  | ctgagttagt  | 11400 |
| gtcattctat  | tctggggggg  | gggggtgggg  | aggacagcaa  | gggggaggat  | tgggaagaca  | 11460 |
| atagcaggca  | tgtctggggg  | gtaatggatt  | tggtagttaa  | ttgagttaa   | caaaggattg  | 11520 |
| gcagtttctc  | actacaggct  | ttctataaga  | ctttgttagaa | atctcacctt  | atttctcttt  | 11580 |
| cagattgacg  | atgatgaaca  | aggttacgac  | ctggacttgt  | tctgcatacc  | taaacattat  | 11640 |
| gcagtgatt   | tggaaaaagt  | ctatattcct  | ctatggctca  | tcattggacg  | gtttgtttga  | 11700 |
| cttcagacag  | tgcactgtct  | cagctgattc  | catgacactg  | gaaaaaacaa  | tcttccagtg  | 11760 |
| atagttttgc  | tgcttagtga  | ctgtctaaac  | agattacatt  | taattagaga  | ctaagaaata  | 11820 |
| cacatgttaa  | ttaaactctc  | cttgttttgc  | ttcaaagagt  | ttgtacattt  | gcagttacgc  | 11880 |
| tattgttttg  | aaatttgc    | atttctcaa   | aaatttgttg  | tacgtagcag  | tctgtgactt  | 11940 |
| tctttacagt  | gtttctttga  | tgttttactt  | aaagtaatta  | gaacacatta  | cttgtgtgtg  | 12000 |
| tagttcaate  | taaaaacagt  | tatagctctc  | caaacatctt  | taggatatta  | ataagagtag  | 12060 |
| attattaate  | acattatgat  | aagacttttt  | catgttcatg  | tggtagataa  | ctaacaccca  | 12120 |
| attttccctc  | actgtctgct  | aaagacatta  | gcctttgcaa  | ataccaatct  | gtcactgtgt  | 12180 |
| gttgcagaaa  | tgtatgattt  | tctctggaag  | ttttatctcg  | tgatgagaaa  | tgggtacatg  | 12240 |
| aacctataag  | gtgtttttgt  | tttactttgt  | gtaagtaaa   | tggaggagtt  | gctggagaca  | 12300 |
| cagaaccact  | gaagagcggt  | tctgagtaga  | tcttgtgaat  | aggaaatgct  | ctagattttg  | 12360 |
| catggtgctg  | tttgatcaga  | ttattacagt  | atttataata  | aaatgttttt  | taactttaca  | 12420 |
| ctgaagacc   | ctatatagga  | aagcattgga  | caaagtacag  | gtcattaagt  | agctgatgta  | 12480 |
| aagtttctga  | tggcaggcat  | tctctgagaa  | acctgctgtc  | agctgtcata  | ctgtaaatat  | 12540 |
| ataccatgct  | ttctgaatta  | aattgcaaga  | taaattttaga | aacaatgate  | actgaaaaac  | 12600 |
| tgttcagtg   | tctcttgctc  | tgctttattg  | gcatttatatt | ttgcagtcag  | gacaaattta  | 12660 |
| ttcagcaaaa  | cataacgcta  | tagttgataa  | tttgagagtt  | ttcttgctcc  | tcagttagtg  | 12720 |
| agagctgttg  | tcttttttgt  | tgtgctgatt  | tattttgctc  | tctgcatgga  | agctgaacct  | 12780 |
| atctttggaa  | gaagaaaaaca | cccttatgtc  | tcttatctga  | cagtaaaaaa  | attcagggtg  | 12840 |
| ttcagatttg  | ctttggctga  | gtatgatgta  | tgaaaaacaaa | gaagtgtggc  | agtgttactg  | 12900 |
| ttagattaac  | cttggaaacgc | aaaactttgt  | tgaccaatag  | tgggttaaag  | tgactgaagc  | 12960 |
| attaggcaaa  | tatttctgag  | caaaatatgc  | ttccagagtt  | gcatgtgttt  | gctgtgtgtg  | 13020 |
| tttgcaatac  | aaaataactgc | tgccatagta  | agcaaaactaa | atgtgttaca  | acagctaact  | 13080 |
| ctcttttttt  | tttttttgta  | taggacagag  | agactggcac  | gagaaattat  | gaagggcag   | 13140 |
| ggaggacatc  | acattgtagc  | tctctgtgta  | ctcaagggtg  | gctataaatt  | ttttgctgat  | 13200 |
| ttattagact  | acatcaaagc  | actgaacaga  | aacagtga    | aatcaatccc  | catgactgta  | 13260 |
| gacttcatta  | ggttgaagag  | ttactgtgta  | agtatctctg  | caataccatg  | caatttttct  | 13320 |
| gtaaatttga  | ctaacttcaa  | attaacaaca  | gggatgattg  | agaattgcca  | acaaatgttg  | 13380 |
| caaaagcttt  | gcctaagtac  | tgcttaaatt  | gtgctaattt  | tatacaataa  | gttaagacaa  | 13440 |
| ttaaaggggaa | aaaagcagtc  | acaagctaac  | ttgttctttg  | tctatcttat  | atgatctggt  | 13500 |
| ttctttcaga  | ctttatctcc  | tgggcccagt  | aaaatccaga  | gcaaagagac  | cctttccatg  | 13560 |
| tccttttact  | cttaaacat   | ttctctctct  | ccccctgtcc  | cacctctaaa  | actgtggttc  | 13620 |
| tagaataata  | cagaggatag  | tcctacaaat  | cttattacaa  | aaacttaact  | ctaggaattt  | 13680 |
| tcattgtggc  | tagacatcac  | tcaccagata  | ggaaaaacttg | aaaactgtga  | gcatgggttat | 13740 |
| attttgggttc | cctacctcat  | cttttgtttg  | gtcctaactga | acaatgaatg  | gaatgaattg  | 13800 |
| catttgggcca | tgaggaaata  | ctagattcat  | aaaaacagat  | ttatgggttag | cggccactga  | 13860 |
| atatgggtgct | acctttttaa  | gcctaggtct  | agggttgcct  | ggctttatgc  | ttttggaaaa  | 13920 |
| gaattttatt  | tcatttttga  | cacaagattt  | taactttaca  | gggaaatgga  | gcatgaggta  | 13980 |
| gtatgtaaga  | tttttataag  | ggaagcatta  | gattacattg  | tgaggtcaa   | aggaggaagc  | 14040 |
| agatgtatgt  | tagtactgta  | tgcttctctga | ctcagcaaca  | gcctagtcta  | gtgttttcaa  | 14100 |
| gcagtatatt  | gtgggttaatt | tagatttttt  | ccataagatc  | acctgtagtg  | tggaaataat  | 14160 |
| taagtttttt  | gatttcaatt  | tacgggtgtca | aaataaattg  | cttcttaagg  | tcattgctgtt | 14220 |
| agttctgtgt  | ctcgtatgact | gtgttttcca  | gtcagttagta | taagctactt  | gtcatgggtg  | 14280 |

|             |             |             |             |             |             |       |
|-------------|-------------|-------------|-------------|-------------|-------------|-------|
| tggagtagtg  | octgctgcag  | tagaaaaaaa  | tagattacta  | agttggaaac  | tcaactgcac  | 14340 |
| tggttcattta | attaaaccag  | agataggggt  | acttgccagat | tragggttga  | ctgtagaaga  | 14400 |
| gcttcacaaaa | tgccgtgctg  | ctggtgagcg  | atctgtataaa | atttcacctc  | gttgagtgge  | 14460 |
| aaagctcaga  | ccagtcagag  | tttggccagg  | cctggtttgca | gaaaagctgt  | gccagctccc  | 14520 |
| aaggtgtgtg  | tgccagtgatg | tccgtggggct | acctggaaac  | caccgctcatg | tctgcagagc  | 14580 |
| gtgctgcagt  | gactcagctg  | cactctttta  | agagaaggca  | aggccaggca  | gaaagggaca  | 14640 |
| tgccagcctt  | tcgttacatc  | ttgcaagggg  | agagctgctt  | gacacacttt  | gtttctgctc  | 14700 |
| cccacaaaaag | attcagggctc | cctctaccag  | tccattcaac  | cacaaagcca  | ctacgtgccc  | 14760 |
| acatacgcctg | tctgtaaaaag | ccacaacaga  | aaggtgaatg  | tctctattag  | aggataatat  | 14820 |
| ttgaaagaat  | aggaaataac  | tcccaccttt  | tttggcgagg  | aaatttgtat  | tttctgcttt  | 14880 |
| cttcacaggaa | cacatccaat  | aactttgggt  | tatttlaaaat | ataaggacta  | caccctttctg | 14940 |
| ctctctccac  | cttcagagta  | acttggaaca  | agctatggga  | taagtgttgt  | gaaacacctg  | 15000 |
| gaaaaaatac  | tacattgcaa  | agcaaggctt  | attgtacagc  | atgtttttat  | tgttcatctt  | 15060 |
| tatgcattag  | gcattgtattg | gttctgctgc  | ttttattttt  | tacacgcgcc  | ccccaacta   | 15120 |
| tagtgccttt  | atattcaatag | gaaactgaac  | agcaaaaataa | ttacgggaagt | tttaattctc  | 15180 |
| tttttgtagt  | aagaaggagt  | aaacaaagag  | aaaagcagag  | aaatttacac  | agaaaggagt  | 15240 |
| gtaagcagaa  | tgctttccat  | gaggtgttgg  | cttcagggtt  | ttggttttgt  | cagccaatc   | 15300 |
| tgtaaaaggaa | tgctttcttg  | ttctatatga  | ccagggagct  | ccagttactt  | gatggcccca  | 15360 |
| ctagcagatt  | attacagagca | tttcatttct  | gtcaactctc  | taaggtagtt  | attaatatcc  | 15420 |
| agacttacgt  | tcttttagatt | gcataacttg  | atagactggg  | taagactgtc  | ctaacacaaa  | 15480 |
| catgcgtgta  | attgattcac  | ttagttttcc  | attattaatg  | gtgctggaat  | gttaaatctg  | 15540 |
| ctctgtagtt  | ttcattcgtg  | ctttactaaa  | cggaggagtc  | agcattctta  | atgatgggaa  | 15600 |
| aacctacaca  | aaccgtaatg  | gacagttgaa  | agtagagcta  | gctctcttgt  | gattgttgtt  | 15660 |
| ggatgcatt   | actagatcac  | gctgcctgcc  | tttacagact  | gggtgaaaga  | tggtaaaaat  | 15720 |
| accgtcgttg  | cataaacaaa  | gcagctgaac  | togtataatg  | gttctaaaat  | tgatgcgatt  | 15780 |
| gatgttctat  | tccattttga  | aattgatcat  | ttaatatttt  | caagcttatg  | gggtatttgt  | 15840 |
| ttcataatgc  | aaagttcata  | ctatgattgc  | acccaaaaat  | aattctctaa  | gaatactgag  | 15900 |
| atgatgaagc  | tggtgattaaa | tggtgcttat  | aattcctgct  | ttcggagtc   | ttacaaatga  | 15960 |
| gttttaatga  | tctggcaagc  | tgtaagaaa   | ttgaaagata  | gtgttttgtga | acagaaccag  | 16020 |
| atctctgtat  | tcttgacaaa  | tagtttattt  | ttctgtgtct  | agaagaatgg  | cttttcacca  | 16080 |
| gtttctctgca | tgtaaaagta  | gagtttaagca | aaagttgctt  | atgtagtcat  | aaagggcagc  | 16140 |
| caaaacacaa  | tggaagtggt  | gcattgtcac  | tcagggaat   | agggtctttc  | ggaggagtac  | 16200 |
| tgccagcttta | aggcaaaatc  | tagtttaggt  | gggttgagaca | gtattctgaa  | agatctctag  | 16260 |
| ttactcgagc  | cttatagcaa  | ataccatgag  | gaacttaaaa  | tgctttgtca  | ggggttacga  | 16320 |
| tcttgccagca | tggtattatt  | ttccagcaat  | gacaatgagg  | cagtgaanaac | agatctaaac  | 16380 |
| agtaaatctc  | taagtgtgta  | aatcacaaag  | cattaaagtc  | ttaccagtac  | tgaaatcaga  | 16440 |
| aataatgctc  | ctgaagataa  | ttaaagtggt  | gcttaattat  | taaacaaagt  | tctttccaat  | 16500 |
| ggaaatctg   | ttaaaaccag  | tggtggaat   | ttgttaacca  | tttttgatcc  | tttaatttgt  | 16560 |
| cagaaactca  | gtacctgcac  | aggcgtgtag  | gttttatatgc | cagcgaaggt  | ttatgttttg  | 16620 |
| tgtaaaaaaa  | tgtagcaatt  | tgccagggtgc | ttggtgggaa  | gaagaagggc  | agatttttgt  | 16680 |
| gggttagattg | ttttctctctg | ctgatcagtg  | tacttgagta  | tgcataacc   | acagggttct  | 16740 |
| gtagtgatct  | gtagttctgt  | gttggtattg  | tttgattata  | gtgtgttgtc  | ctgtagccag  | 16800 |
| agagatcacc  | aggaaactga  | tctgcattga  | tgtaactgaga | ggactggctc  | atcatctgtg  | 16860 |
| catggtgttt  | taatggaaaca | aactcttctc  | ccatcttctc  | ctgttaagaa  | tctgttaagaa | 16920 |
| atggtgtctt  | ttacatcttt  | ccatggaaaa  | aggacgaat   | tcttattagt  | gtaggaatag  | 16980 |
| ataagagaga  | atcctgcaac  | ataaacaggg  | aaaagctcaa  | agctgtcatt  | taagataatt  | 17040 |
| ttaaagatatg | atgaacgcag  | gaaatgcccc  | tcagcattac  | aaacacgggtg | cttaactgaa  | 17100 |
| tcccattttc  | gtgattaatc  | taaccagtat  | ctctggcatt  | cagtagcagt  | ctgctgttta  | 17160 |
| ctgggtcaag  | cagcagacat  | tttggacatc  | ctatttttgg  | actttttcaa  | cacagtggta  | 17220 |
| ttggttctgt  | tcactatagt  | tttcttttaa  | acatggggta  | aaactgtaaa  | ggcgcttttt  | 17280 |
| ctaagtgtgt  | gtgcttgctc  | tttttttaat  | gttttgtcat  | gggaatctaa  | aatagaaagt  | 17340 |
| atctgttagt  | ctcgttatct  | gcttttacat  | acctagagct  | gtcagcactt  | ctttatccca  | 17400 |
| gagcattgac  | gtcactgtcc  | aagggatttt  | aatacatctt  | ctctttactg  | taatcccagt  | 17460 |
| gtcacttatg  | gtgtgctcag  | tgtgtgtttc  | tctctaaacc  | atttaccctt  | tttttgcatg  | 17520 |
| tgataacact  | cttacttttt  | gtctgcaact  | ggtagagtat  | gaatatgtgt  | gtcagtgaaa  | 17580 |
| acagggaaget | gtatgtggcc  | gtctgttccc  | ctgtggtctc  | ctctctattg  | tatctgccat  | 17640 |
| tttggggcgt  | ggggaaggaa  | ctgctctttt  | gctgcttgtc  | tgcatggta   | tgcatgaagg  | 17700 |
| ttgggcttca  | gactgaagcg  | gcaaaataat  | aaaggttttc  | taaggggaag  | atatgaagtg  | 17760 |
| aagagctgtg  | agaatgaagt  | ttgaaaagtg  | gtatgacctt  | tcaaatcaat  | gattaaattg  | 17820 |
| taggataaatg | acatcatctg  | taagtagcac  | agtaatgagt  | atagcacaat  | taataagaac  | 17880 |
| taggagacaa  | tgtaaggtt   | taaaagtacg  | tgaatgactt  | taataaaact  | gcagagatct  | 17940 |

|            |             |             |            |            |            |       |
|------------|-------------|-------------|------------|------------|------------|-------|
| ggaggtaccc | caggagtgt   | ttttaacttt  | cttatataac | aaacagttca | ctgatgctta | 18000 |
| catcaaagcc | tagttttact  | aaagcaaaaat | gctgtacttc | agtgatagga | ttgtgtagct | 18060 |
| gtgtcgctgt | ttaattacca  | caatgagtaa  | tttatcacta | tttgtaaaat | aggcaaacc  | 18120 |
| aagcttttct | ggctattctt  | gtagcttata  | tttcttggtc | ttaatctgtg | ccttggttgc | 18180 |
| tggctatttc | attgttggtt  | tgattgctta  | tttttattca | gagtaaggct | ttctaaggca | 18240 |
| tcaaagtcca | aaatgtaaat  | taccagctat  | atagcagcgt | acagcttgga | aacttttttt | 18300 |
| ttttcctgtt | ttccatacct  | ttaatgaata  | tgtaggaaat | aataattttt | taataacat  | 18360 |
| gtgtgtatat | ataatgggtga | cagatgagaa  | gtggtgcctg | tatttatgta | aataattttt | 18420 |
| caaacaatct | tttatctgaa  | ataccatctt  | tcagtctgat | tacttgtaat | ttcttagtta | 18480 |
| atagatgttg | ctggtgcaat  | gcttctgaaa  | gcttgggtat | ttgtgggtta | tgtactttgc | 18540 |
| agttttgata | tttattagaa  | gaaaaattac  | tttaattctg | gtaatactaa | tgtttaccct | 18600 |
| gctctttctt | gctagaatga  | tcagtcaact  | ggagatataa | aagtcattgg | tggggatgac | 18660 |
| ctttcaacct | tgactggaaa  | ggtatggcat  | gtcatatggt |            |            | 18700 |

15-10-2010

## **REIVINDICAÇÕES**

1. Polipéptido isolado ou recombinante compreendendo uma sequência de aminoácidos que tem pelo menos 80% de identidade de sequência de aminoácidos com a sequência de aminoácido SEQ ID N°: 1 da transcriptase inversa da telomerase de *Cairina moschata*, e tendo actividade TERT.
2. Molécula de ácidos nucleicos isolada ou recombinante que codifica um polipéptido de acordo com a reivindicação 1.
3. Molécula de ácidos nucleicos isolada ou recombinante de acordo com a reivindicação 2 em que a dita molécula de ácidos nucleicos compreende uma sequência de ácidos nucleicos que tem pelo menos 90% de identidade de sequência de ácidos nucleico com a SEQ ID N°: 2.
4. Vector que transporta uma molécula de ácidos nucleicos isolada ou recombinante de acordo com qualquer uma das reivindicações 2 ou 3.
5. Vector de acordo com a reivindicação 4, em que o dito vector compreende duas sequências homólogas a uma sequência de ADN alvo.
6. Vector de acordo com a reivindicação 5, em que a dita molécula de ácidos nucleicos é rodeada pelas ditas sequências homólogas.
7. Vector de acordo com qualquer uma das reivindicações 5 a 6, em que o dito vector compreende adicionalmente um primeiro marcador de selecção em que o dito primeiro marcador de selecção é um marcador de selecção positivo e em que o dito

primeiro marcador de selecção é rodeado pelas ditas sequências homólogas.

8. Vector de acordo com a reivindicação 7, em que o dito primeiro marcador de selecção é rodeado por sequências que permitem a sua supressão.

9. Vector de acordo com qualquer uma das reivindicações 5 a 8, em que o dito vector compreende um segundo marcador de selecção que não é rodeado pelas ditas sequências homólogas e em que o dito marcador de selecção é um marcador de selecção negativo.

10. Vector de acordo com qualquer uma das reivindicações 7 a 9, em que o dito vector compreende um terceiro marcador de selecção em que o dito terceiro marcador de selecção é um marcador de selecção negativo e em que o dito terceiro marcador de selecção fica situado entre as sequências que permitem a supressão do primeiro marcador de selecção.

11. Célula aviária transfectada por uma molécula de ácidos nucleicos de acordo com qualquer uma das reivindicações 2 ou 3.

12. Célula aviária de acordo com a reivindicação 11, em que a dita molécula de ácidos nucleicos é inserida no gene HPRT da dita célula aviária.

13. Célula aviária de acordo com qualquer uma das reivindicações 11 a 12 em que a dita célula aviária deriva de um animal que pertence à família *Anatidae*, de preferência às espécies *cairina moschata* ou *Anas platyrhynchos*.

14. Célula aviária de acordo com qualquer uma das reivindicações 11 a 13, que compreende adicionalmente uma sequência de ácidos nucleicos que codifica uma substância de interesse.

15. Célula aviária de acordo com qualquer uma das reivindicações 11 a 14, que compreende adicionalmente uma cassette de complementação que permite a propagação de um vírus defectivo.

16. Uso de um polipéptido de acordo com a reivindicação 1, de uma molécula de ácidos nucleicos de acordo com qualquer uma das reivindicações 2 ou 3 ou de um vector de acordo com qualquer uma das reivindicações 4 a 10 para a imortalização de uma célula aviária.

17. Uso de uma célula aviária de acordo com qualquer uma das reivindicações 1 a 14 para a replicação de um vírus vivo, atenuado ou recombinante, escolhido de preferência do grupo que consiste em poxvírus, adenovírus, retrovírus, herpesvírus, alfavírus, vírus espumoso, e vírus adenovírus-associado.

18. Uso de um célula aviária de acordo com a reivindicação 17, em que o poxvírus é um vírus vaccinia, em particular um vírus vaccinia modificado Ankara (MVA).

19. Uso de uma célula aviária de acordo com a reivindicação 14 para a produção de uma substância de interesse.

20. Uso de uma célula aviária de acordo com a reivindicação 19, em que a dita substância de interesse é seleccionada do grupo que consiste em factores de crescimento, reguladores de

crescimento, anticorpos, antigénios, interleucinas, insulina, G-CSF, GM-CSF, hPG-CSF, M-CSF, interferões, factores de coagulação do sangue, e suas combinações.

21. Processo para imortalizar uma célula aviária compreendendo o passo de transfectar na dita célula um vector de acordo com qualquer uma das reivindicações 4 a 10.

22. Processo de acordo com a reivindicação 21, **caracterizado por** compreender adicionalmente um passo em que as ditas células aviárias são cultivadas num meio que apenas permite o crescimento das células que incorporaram o primeiro marcador de selecção.

23. Processo de acordo com qualquer uma das reivindicações 21 a 22 **caracterizado por** compreender adicionalmente um passo em que as ditas células aviárias são cultivadas num meio que não permite o crescimento das células que incorporaram o segundo marcador de selecção.

24. Processo de acordo com qualquer uma das reivindicações 21 a 23, em que o dito processo compreende adicionalmente um passo que consiste em excluir o primeiro marcador de selecção do genoma da dita célula aviária.

25. Processo de acordo com a reivindicação 24 em que células aviárias, obtidas após o dito passo que consiste em excluir o primeiro marcador de selecção do genoma das ditas células aviárias, são cultivadas num meio que não permite o crescimento das células que compreendem o terceiro marcador de selecção.

26. Processo de acordo com qualquer uma das reivindicações 21 a 25, em que a dita célula aviária é retirada de um organismo que pertence à família Anatidae, de preferência às espécies *cairina moschata* ou *Anas platyrhynchos*.

27. Processo de acordo com qualquer uma das reivindicações 21 a 26, em que a dita molécula de ácidos nucleicos é inserida numa sequência de ADN alvo da dita célula, de preferência no gene HPRT.

15-10-2010

## **REIVINDICAÇÕES**

1. Polipéptido isolado ou recombinante compreendendo uma sequência de aminoácidos que tem pelo menos 80% de identidade de sequência de aminoácidos com a sequência de aminoácido SEQ ID N°: 1 da transcriptase inversa da telomerase de *Cairina moschata*, e tendo actividade TERT.
2. Molécula de ácidos nucleicos isolada ou recombinante que codifica um polipéptido de acordo com a reivindicação 1.
3. Molécula de ácidos nucleicos isolada ou recombinante de acordo com a reivindicação 2 em que a dita molécula de ácidos nucleicos compreende uma sequência de ácidos nucleicos que tem pelo menos 90% de identidade de sequência de ácidos nucleico com a SEQ ID N°: 2.
4. Vector que transporta uma molécula de ácidos nucleicos isolada ou recombinante de acordo com qualquer uma das reivindicações 2 ou 3.
5. Vector de acordo com a reivindicação 4, em que o dito vector compreende duas sequências homólogas a uma sequência de ADN alvo.
6. Vector de acordo com a reivindicação 5, em que a dita molécula de ácidos nucleicos é rodeada pelas ditas sequências homólogas.
7. Vector de acordo com qualquer uma das reivindicações 5 a 6, em que o dito vector compreende adicionalmente um primeiro marcador de selecção em que o dito primeiro marcador de selecção é um marcador de selecção positivo e em que o dito

primeiro marcador de selecção é rodeado pelas ditas sequências homólogas.

8. Vector de acordo com a reivindicação 7, em que o dito primeiro marcador de selecção é rodeado por sequências que permitem a sua supressão.

9. Vector de acordo com qualquer uma das reivindicações 5 a 8, em que o dito vector compreende um segundo marcador de selecção que não é rodeado pelas ditas sequências homólogas e em que o dito marcador de selecção é um marcador de selecção negativo.

10. Vector de acordo com qualquer uma das reivindicações 7 a 9, em que o dito vector compreende um terceiro marcador de selecção em que o dito terceiro marcador de selecção é um marcador de selecção negativo e em que o dito terceiro marcador de selecção fica situado entre as sequências que permitem a supressão do primeiro marcador de selecção.

11. Célula aviária transfectada por uma molécula de ácidos nucleicos de acordo com qualquer uma das reivindicações 2 ou 3.

12. Célula aviária de acordo com a reivindicação 11, em que a dita molécula de ácidos nucleicos é inserida no gene HPRT da dita célula aviária.

13. Célula aviária de acordo com qualquer uma das reivindicações 11 a 12 em que a dita célula aviária deriva de um animal que pertence à família *Anatidae*, de preferência às espécies *cairina moschata* ou *Anas platyrhynchos*.

14. Célula aviária de acordo com qualquer uma das reivindicações 11 a 13, que compreende adicionalmente uma sequência de ácidos nucleicos que codifica uma substância de interesse.

15. Célula aviária de acordo com qualquer uma das reivindicações 11 a 14, que compreende adicionalmente uma cassette de complementação que permite a propagação de um vírus defectivo.

16. Uso de um polipéptido de acordo com a reivindicação 1, de uma molécula de ácidos nucleicos de acordo com qualquer uma das reivindicações 2 ou 3 ou de um vector de acordo com qualquer uma das reivindicações 4 a 10 para a imortalização de uma célula aviária.

17. Uso de uma célula aviária de acordo com qualquer uma das reivindicações 1 a 14 para a replicação de um vírus vivo, atenuado ou recombinante, escolhido de preferência do grupo que consiste em poxvírus, adenovírus, retrovírus, herpesvírus, alfavírus, vírus espumoso, e vírus adenovírus-associado.

18. Uso de um célula aviária de acordo com a reivindicação 17, em que o poxvírus é um vírus vaccinia, em particular um vírus vaccinia modificado Ankara (MVA).

19. Uso de uma célula aviária de acordo com a reivindicação 14 para a produção de uma substância de interesse.

20. Uso de uma célula aviária de acordo com a reivindicação 19, em que a dita substância de interesse é seleccionada do grupo que consiste em factores de crescimento, reguladores de

crescimento, anticorpos, antigénios, interleucinas, insulina, G-CSF, GM-CSF, hPG-CSF, M-CSF, interferões, factores de coagulação do sangue, e suas combinações.

21. Processo para imortalizar uma célula aviária compreendendo o passo de transfectar na dita célula um vector de acordo com qualquer uma das reivindicações 4 a 10.

22. Processo de acordo com a reivindicação 21, **caracterizado por** compreender adicionalmente um passo em que as ditas células aviárias são cultivadas num meio que apenas permite o crescimento das células que incorporaram o primeiro marcador de selecção.

23. Processo de acordo com qualquer uma das reivindicações 21 a 22 **caracterizado por** compreender adicionalmente um passo em que as ditas células aviárias são cultivadas num meio que não permite o crescimento das células que incorporaram o segundo marcador de selecção.

24. Processo de acordo com qualquer uma das reivindicações 21 a 23, em que o dito processo compreende adicionalmente um passo que consiste em excluir o primeiro marcador de selecção do genoma da dita célula aviária.

25. Processo de acordo com a reivindicação 24 em que células aviárias, obtidas após o dito passo que consiste em excluir o primeiro marcador de selecção do genoma das ditas células aviárias, são cultivadas num meio que não permite o crescimento das células que compreendem o terceiro marcador de selecção.

26. Processo de acordo com qualquer uma das reivindicações 21 a 25, em que a dita célula aviária é retirada de um organismo que pertence à família Anatidae, de preferência às espécies *cairina moschata* ou *Anas platyrhynchos*.

27. Processo de acordo com qualquer uma das reivindicações 21 a 26, em que a dita molécula de ácidos nucleicos é inserida numa sequência de ADN alvo da dita célula, de preferência no gene HPRT.

15-10-2010