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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
01 October 2020 (01.10.2020)

WIPO | PCT

(10) International Publication Number
WO 2020/197518 A1

(51) International Patent Classification:
C12N 15/63 (2006.01) *C07K 14/44* (2006.01)
C12N 15/00 (2006.01)

(21) International Application Number:
PCT/TR2019/051165

(22) International Filing Date:
20 December 2019 (20.12.2019)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
TR 2019/04291 22 March 2019 (22.03.2019) TR

(71) Applicant: **ANADOLU UNIVERSITESI** [TR/TR]; Yunus Emre Kampüsü, Eskisehir (TR).

(72) Inventors: **ARSLANYOLU, Muhittin**; Eskisehir Teknik Universitesi Fen Fakültesi Biyoloji Bölümü, Eskisehir (TR). **USTUNTANIR DEDE, Ayca Fulya**; Eskisehir Teknik Universitesi Fen Fakültesi Biyoloji Bölümü, Eskisehir (TR). **CALISEKI, Mehmet**; 71 Evler Mah. Onaran Sok. Melih Yapi Kooperatifleri Solmaz Sitesi 3A Daire :7, Odunpazari/Eskisehir (TR).

(74) Agent: **INVOKAT INTELLECTUAL PROPERTY SERVICES**; Kartaltepe Mh. Yildiztepe Sk. No:6, 34145 Bakirköy/Istanbul (TR).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

(54) Title: TETRAHYMENA THERMOPHILA ARTIFICIAL CHROMOSOME 2 (TTAC2) AND ITS USE IN THE RECOMBINANT PROTEIN PRODUCTION

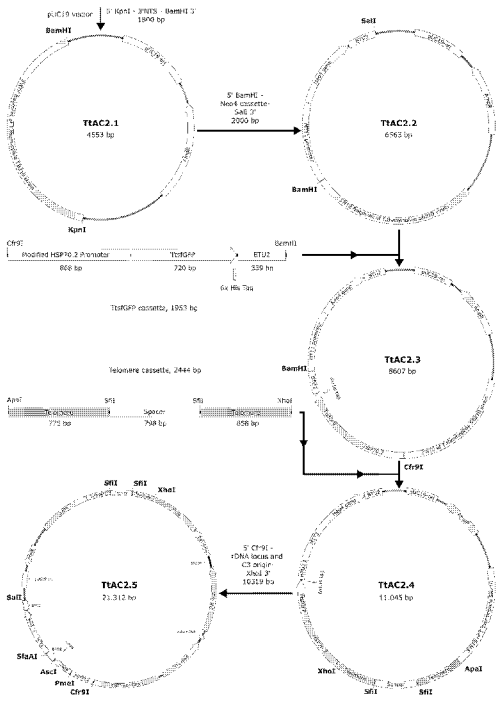


FIGURE 1

(57) Abstract: The present invention relates to Tetrahymena thermophila artificial chromosome (TtAC2) having a nucleotide sequence that is at least 80% homologous to SEQ ID NO: 4 and comprising a modified HSP70.2 promoter having SEQ ID NO: 1, 3´NTS sequence having SEQ ID NO: 2, "5´NTS containing C3 origin, rDNA genes and 3´NTS" sequence having SEQ ID NO: 3. Tetrahymena thermophilaartificial chromosome, which is formed by biomimetics of the Tetrahymena thermophilamacronucleus rDNA minichromosome further comprises a telomere sequence, Neo4 cassette and TtsfGFP marker gene cassette. Within the scope of the present invention, also provided a method for obtaining said Tetrahymena thermophila artificial chromosome. In another aspect, the present invention provides a method for recombinantly producing homologous and/or heterologous genomic DNA, RNA and/or proteins and/or derivatives thereof by using Tetrahymena thermophilaartificial chromosome as a production system/platform. In another aspect, the present invention provides the use of the Tetrahymena thermophila artificial chromosome as a production system/platform for recombinantly producing homologous and/or heterologous genomic DNA, RNA and/or proteins and/or derivatives thereof.

WO 2020/197518 A1

[Continued on next page]

Published:

— *with international search report (Art. 21(3))*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
— *with sequence listing part of description (Rule 5.2(a))*
— *in black and white; the international application as filed contained color or greyscale and is available for download from PATENTSCOPE*

TETRAHYMENA THERMOPHILA ARTIFICIAL CHROMOSOME 2 (TtAC2) AND ITS USE IN THE RECOMBINANT PROTEIN PRODUCTION

Technical Field

The present invention relates to *Tetrahymena thermophila* artificial chromosome (TtAC2) having a nucleotide sequence that is at least 80% homologous to SEQ ID NO: 4 and comprising a modified HSP70.2 promoter having SEQ ID NO: 1, 3´NTS sequence having SEQ ID NO: 2, "5´NTS containing C3 origin, rDNA genes and 3´NTS" sequence having SEQ ID NO: 3. *Tetrahymena thermophila* artificial chromosome, which is formed by biomimetics of the *Tetrahymena thermophila* macronucleus rDNA minichromosome further comprises a telomere sequence, Neo4 cassette and TtsfGFP marker gene cassette. More specifically, the present invention relates to a *Tetrahymena thermophila* artificial chromosome which can be used in linear or circular forms by virtue of the telomeric sequence it contains, including"5´NTS containing C3 origin, rDNA genes and 3´NTS" sequence on one arm, and 3´NTS sequence, Neo4 cassette and a marker gene cassette on the other arm, and which can be used in recombinant protein production by means of heat-shock induction owing to modified HSP70.2 promoter it contains.

Background of the Invention

Biomimetics refers to the imitation of models and systems in nature for the purpose of solving a problem or improving an existing system. By the virtue of this field, many technologies inspired by biological solutions at the macro or nano level have been developed. By constructing expression vectors with the principles of biomimetics, it is possible to obtain vectors with high transformation efficiency and intracellular stability. The aim of biomimetics applications in the field of molecular biology is to form vectors that are similar to the cell's chromosome, therefore cells cannot distinguish between their wild-type structures and biomimetics products.

Tetrahymena thermophila is a ciliated eukaryotic unicellular organism which is one of the best-characterized unicellular eukaryotes. *Tetrahymena thermophila* bears two nuclei: a transcriptionally silent, diploid germline nucleus (MIC) and transcriptionally active, polyploid somatic macronucleus (MAC). Use of *Tetrahymena thermophila* as an alternative model organism for recombinant protein expression for biological and biotechnological studies has been known for many years. The organism has been on the focus based on its advantages in the field. These advantages include short cell division time (under optimal conditions, the generation time is about 1.5-3 hours), the ability to introduce post-translational modifications to the protein of interest, and easy/cost-effective laboratory handling *(A.P. Turkewitz, E. Orias, G. Kapler, 2002.*

Functional genomics: the coming of age for Tetrahymena thermophila, Trends Genet. 18 35–40; *Weide, T., Herrmann, L., Bockau, U., Niebur, N., Aldag, I., Laroy, W., Contreras, R., Tiedtke, A. ve Hartmann, M. W., 2006. Secretion of functional human enzymes by Tetrahymena thermophila. BMC Biotechnol, 6: 19)*. Therefore, it is of great interest to use *Tetrahymena thermophila* as a host for recombinant protein production in a way that is productive and efficient. There have been various methods used for transforming the organism with the gene(s) of interest, including microinjection, electroporation and biolistic bombardment *(Cassidy-Hanley, D., et. al. 1997. Germline and somatic transformation of mating Tetrahymena thermophila by particle bombardment. Genetics, 146(1), 135-147)*.

It is possible to grow Tetrahymena cells with a batch tank, a fed-batch tank and continuous feed bioreactors by approaching to values of a cell density of approximately $2x10^7$ cells/ml, a dry cell weight of 80 g/L and a tank scale-up volume of 1000 L. A typical reactor time takes about 3-4 days from inoculation to harvest.

Tetrahymena ribosomal RNA genes (rDNA) are one of the best characterized eukaryotic genes *(Yu, G. L., Hasson, M., & Blackburn, E. H., 1988. Circular ribosomal DNA plasmids transform Tetrahymena thermophila by homologous recombination with endogenous macronuclear ribosomal DNA. Proceedings of the National Academy of Sciences of the United States of America, 85(14), 5151-5)*. The rDNA genes, along with 5´NTS bearing the Tetrahymena origin region, 3´NTS containing the subtelomeric regions, and the telomere sequence, form the 21 kbp chromosomes (rDNA chromosomes) of Tetrahymena called minichromosomes which exist as two palindromic copies. These rDNA chromosomes reach approximately 9000 copies extrachromosomally in the macronucleus of the organism. Thanks to their advantages, Tetrahymena rDNA genes and the rDNA origin sequence containing vectors are frequently used as tools in the transformation of desired genes into the organism *(Blomberg P, Randolph C, Yao CH, Yao MC., 1997. Regulatory sequences for the amplification and replication of the ribosomal DNA minichromosome in Tetrahymena thermophila. Mol Cell Biol.; 17:7237–47; Engberg J, Nielsen H., 1990. Complete sequence of the extrachromosomal rDNA molecule from the ciliate Tetrahymena thermophila strain B1868VII. Nucleic Acids Res, 18:6915-6919)*. However, many of these vectors do not contain all of the elements such as recombinant gene expression cassette, antibiotic resistance gene, origin and telomere at the same time, and these vectors in use are not formed with a combination suitable for effective transformation and long term storage thereafter.

A cloning vector can be defined as a fragment of DNA that can be used as a tool to carry another fragment of DNA into a host organism. Cloning vectors can be formed by combining the DNA sequences that are obtained from a virus, a prokaryotic bacterial cell or a cell of a higher organism.

Cloning vectors are intermediary DNA sequences used to transfer a nucleic acid sequence to the host organism and to quantitatively increase the DNA sequence in the host. Some of the features that an ideal cloning vector possesses could be regarded as: not containing DNA elements that can cause unwanted changes by recombination or mutation in the host organism, containing a gene for selection after transformation, being designed as to contain a region that a DNA fragment can be cloned, having an origin region that enables controlled and stable copy numbers, etc.

Artificial chromosomes are used both as a cloning vector and expression vector. They can carry DNA inserts larger than the capacity of plasmids or lambda-phage-derived vectors. Artificial chromosomes, like other cloning vectors, contain the nucleic acid elements that are necessary for replication and stability of the vector and its product in the host cell. These elements ensure the continuity of the vector copy number in the daughter cells upon cell division. Artificial chromosomes are ideal vectors for stable, controlled and high-level production of proteins that necessitate the coordinated expression of several genes or that are encoded by large genes. US 8,288,610 B2 patent document relates to preparation of plant cell lines that comprise artificial chromosomes. Methods for heterologous nucleic acid insertion into the artificial chromosomes, directing the artificial chromosomes to the selected cells and isolation of the said artificial chromosomes are also provided in the said patent application. Said artificial chromosomes were directed to plant cells.

Antibiotic resistance cassettes are important tools in cloning and vector constitution. Antibiotic resistance allows the positive transformants to be selected easily and effectively. The transformant cells that contain the vector can grow in the presence of selected antibiotic, which proves that the transformation was successful and vector functions in the cell.

Recombinant protein production in *Tetrahymena thermophila* has been widely used. Numerous proteins from different species, including functional human enzymes were produced using Tetrahymena as an expression system *(Weide T., Herrmann L., Bockau U., Niebur N., Aldag I., Laroy W., Contreras R., Tiedtke A., Hartmann M. W. (2006). Secretion of functional human enzymes by Tetrahymena thermophila. BMC Biotechnol. 6:19)*. Therefore, various vectors and promoters specific to the organism have been developed. However, current techniques and tools has its downsides along with the advantages described above. First of all, vectors used are mostly circular. Although being easy to be used and be transformed, circular vectors limit the transformable base pairs of nucleic acids. This in turn, makes it hard to produce large proteins that are coded by large genes by vectors. Also, metabolic pathways require several genes working in cooperation. Therefore, expression systems which can bear larger gene sequences are required. Another problem to be solved about current vectors is that the efficiency of transformation is not

at high levels. Efficiency and stability are important features a vector should have, especially if the protein needs to be produced in large quantities.

Out of the promoters that are studied for recombinant protein production in *Tetrahymena thermophila*, one of most commonly used promoters is MTT1 promoter. MTT1 is a inducible-repressible promoter for driving high-level expression of heterologous or homologous genes in *Tetrahymena thermophila*. It belongs to a Cd-inducible metallothionein gene (MTT1) of *Tetrahymena thermophila*. The activation and deactivation of the promoter are achieved by simply adding or depleting cadmium. Although it's an efficient promoter to induce production, this promoter is not suitable for good manufacturing practise, since Cd is a heavy metal. Heavy metals are toxic and restrain the cell growth. Also, the disposal of contaminated media takes much time and resource, since heavy metals are a threat to environment and to human health. Use of a different promoter with less negative impact on the host, humans and environment would be favourable.

Heat shock proteins (HSPs) are stress response proteins that are found in all eukaryotic organisms studied to date. The genes that code these proteins are inducible by heat, as well as other stress factors or external agents. That means the cells respond to a stress condition, such as high growth temperature, by producing high levels of HSPs. Regulation of HSPs has been studied extensively and HSP promoters were found to be strong and effective promoters. Also, the production process can be simply activated and deactivated by the changes in temperature, without a need for chemical agents. Therefore, heat shock promoters are suitable for recombinant studies and can be used for controlled heterologous gene expression at high levels. WO 2007/006812 A1 relates to the use of heat-inducible HSP90 promoter that belongs to the family of *Tetrahymena thermophila* heat-shock proteins. The use of the promoter for the expression of homologous and/or heterologous proteins in the ciliate *Tetrahymena thermophila* is also claimed within the scope of WO 2007/006812 A1 patent document. In another patent application, CN 101586119 A, HSP70.2 promoter sequence is used in the expression of GFP marker gene. HSP70.2 that is defined in this patent application consists of 1132 bp sequence. Said sequence contains 1105 bp HSP70.2 promoter and 5´UTR region of the mRNA of the gene, 21 bp sequence of HSP70.2 gene that codes protein starting from ATG start codon and 6 bp BamHI restriction enzyme cloning site.

In consideration of the points mentioned above, an improvement is needed in recombinant expression of proteins in *Tetrahymena thermophila*. It is an object of the invention to promote the use of *Tetrahymena thermophila* organism as an organismal factory, in particular, to develop a vector system for constructing a protein expression platform in DNA/RNA/protein-derived drug-enzyme-hormone production studies.

Brief Description of the Figures

Figure 1 shows the elements and sequences that construct TtAC2, the vectors that these elements are derived, and circular form of TtAC1.

Figure 2 shows A. Circular and linear structures of TtAC2, B. SDS-PAGE image of TtsfGFP gene expression, C. Western blot image of TtsfGFP gene expression, 1: ThermoFisher Scientific PageRuler™ Plus Prestained Protein Ladder, 10 to 250 kDa (26619) is used as a marker DNA, it also contains 6xHis-TtsfGFP (~27 kDa), 2: cell total protein containing Par+ TtAC2 circular artificial chromosome, 3: cell total protein containing Par+ TtAC2 linear artificial chromosome, 4: cell total protein containing Par- TtAC2 circular artificial chromosome 5: cell total protein containing the Par- TtAC2 linear artificial chromosome, D. fluorescence microscopy analysis of the expression of the TtsfGFP gene in cells having TtAC2 at 5th month post-selection, E. fluorescence microscopy analysis of the expression of the TtsfGFP gene in cells having TtAC2 at 24th month post-selection.

Detailed Description of the Invention

In this detailed description, the steps of obtaining the *Tetrahymena thermophila* artificial chromosome of the present invention and use thereof for the recombinant protein production is described for better understanding of the subject matter.

Said artificial chromosome was designed for unicellular eukaryotic organism *Tetrahymena thermophila* and other ciliate host cells of the Tetrahymenidae family. The object of the invention is to produce recombinant proteins by forming an artificial chromosome that maintains intracellular stability for a long time and to increase the efficiency of *Tetrahymena thermophila* as an expression system. This artificial chromosome (TtAC2) of the present invention was developed by performing biomimetics of the *Tetrahymena thermophila* rDNA minichromosome.

Tetrahymena thermophila is a well studied organism in the field of biology. Reaching high cell densities in short reproduction time, being able to grow easily and inexpensively in the laboratory and performing post-translational modifications in the produced proteins are some of the reasons for the widespread use of this organism. This latter feature is particularly necessary in the production of proteins of eukaryotic origin, since post-translational modifications play an important role in protein function in eukaryotes. Therefore, eukaryotic organism *Tetrahymena thermophila*, provides some advantages in recombinant protein production over prokaryotic host organisms.

Artificial chromosomes are constructed in laboratory by amplifying and ligating the necessary elements like origin sequence, telomere sequence, antibiotic resistance cassette for selection, and an inducible promoter. This "vector backbone" is then added the gene sequence of protein or enzyme of interest. In the present invention, vectors used as source of the backbone elements are preferred as pUC19 *(Yanisch-Perron C, Vieira J, Messing J., 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33:103–119)* and pNeo4 *(Mochizuki, K., 2008. High efficiency transformation of Tetrahymena using a codon-optimized neomycin resistance gene. Gene, 425(1), 79-83)*.

In an aspect, the artificial chromosome of the present invention comprises a modified HSP70.2 promoter having SEQ ID NO: 1, a 3'NTS sequence having SEQ ID NO: 2 and "5´NTS containing C3 origin, rDNA genes and 3´NTS" sequence having SEQ ID NO: 3.

A specific HSP70.2 promoter of the invention is contained in the TtsfGFP cassette and has the nucleotide sequence set forth in SEQ ID NO: 1. In this sequence, G (guanine) base of the ATG codon of the gene region encoding the artificial 11 amino acid HSP70 protein sequence, which is involuntarily added to the N-terminus of the recombinant proteins, is changed by mutation with T (thymine) base, so the ATG codon was transformed to ATT, resulting in modification of the 11 amino acids encoding sequence into a non-coding form. In addition, an A (adenine) base was added prior to the PmeI enzyme restriction sequence at the 3´ end of the HSP70.2 promoter sequence.

Many vectors and promoters have been developed for the use in *Tetrahymena thermophila*. MTT1 promoter which belongs to Cd-inducible metallothionein gene of *Tetrahymena thermophila* is widely used in recombinant protein production. However, Cd is a heavy metal and the induction with Cd harms both the organism and environment. Therefore Cd-inducible promoters are not favorable in the terms of good manufacturing practices. Heat shock promoters on the other hand, are induced by heat or other stress conditions. They belong to heat shock genes, which are found in all eukaryotic cells. The induction and termination of transcription can simply be carried out by increasing or decreasing the growth temperature. Therefore, specific modified HSP70.2 promoter of the present invention is important and advantageous in that it can be induced by increasing the growth temperature in line with good manufacturing practices.

The artificial chromosome of the present invention produced for *Tetrahymena thermophila* comprises a 3´NTS sequence (SEQ ID NO: 2). In the context of the present invention, said 3´NTS sequence can be amplified with specific primers from the genomic DNA of the *Tetrahymena thermophila* SB210 strain and can be inserted into a vector backbone sequence to obtain the

artificial chromosome. The specific 3´NTS sequence of the invention comprises subtelomeric sequences of the rDNA minichromosome in *Tetrahymena thermophila* and enables the stable operation of telomeres and telomeric genes. Thus, it helps to increase the life time of the artificial chromosome in the cell. In addition, the 3´NTS sequence is located between the gene cassettes and the telomere sequences necessary for the production of recombinant protein, thereby preventing possible telomere pressure in the transcription of the genes.

The *Tetrahymena thermophila* artificial chromosome of the present invention further comprises "5´NTS containing C3 origin, rDNA genes and 3´NTS" sequence (SEQ ID NO: 3). The specific rDNA locus and C3 ori sequence of the present invention can be amplified with specific primers from the rDNA minichromosome of *Tetrahymena thermophila* C3.368.1 strain and can be inserted into the vector backbone of the present invention. The rDNA locus and C3 ori sequence, which form an arm of the TtAC2 artificial chromosome, are important for the biomimetics of the *Tetrahymena thermophila* rDNA minichromosome and for long-term biological maintenance after transformation.

In an aspect of the present invention, said vector backbone is pUC19 vector. The specific vector backbone of the invention is advantageous in that it has diversity of restriction enzymes in the cloning site, the vector size is small (2.6 kbp) and it has an intermediate copy origin (100+).

In an aspect of the present invention, artificial chromosome for *Tetrahymena thermophila* further comprises a telomere sequence. In the context of the present invention, said telomere sequence can be cut from a pPXV-GFP vector, amplified and inserted into a backbone vector sequence in order to obtain the artificial chromosome. The telomere cassette of the invention is important and advantageous in that it enables the use of TtAC2 artificial chromosome in circular or linear form. The use of linear vector presents the advantage of removing the limitation about the size of the insertion. Therefore, the linear form of the artifical chromosome enables vector to be usable when more gene cassettea are inserted. By this aspect, it is useful for investigating the genes encoding relatively large proteins and/or metabolic pathways, which include sets of genes working in coordination. The telomere cassette sequence of the present invention includes a sequence which is cut from both ends and released with SfiI enzyme in order to convert the circular vector to a linear vector.

The *Tetrahymena thermophila* artificial chromosome of the present invention further comprises a Neo4 neomycin resistance cassette. The specific Neo4 cassette of the present invention can be amplified with specific primers from the pNeo4 vector and be inserted into pUC19 vector backbone of the present invention. Neo4 cassette which is $CdCl_2$-inducible via the MTT1 promoter

is suitable for enabling efficient selection of positive transformants with increased antibiotic suppression and reduced cadmium induction by the *Tetrahymena thermophila* codon-optimized neomycin resistance cassette it contains.

The *Tetrahymena thermophila* artificial chromosome of the present invention further comprises a marker gene cassette. In order to demonstrate that the obtained artificial chromosome can direct recombinant protein production in the host organism, the expression cassette containing the TtsfGFP-6xHis gene was taken from the pVTtsfGFP vector and inserted into TtAC2 as a marker gene with some sequence changes *(Yilmaz, G., & Arslanyolu, M. 2015. Efficient expression of codon-adapted affinity tagged superfolders for fluorescent protein localization and affinity purification studies in Tetrahymena thermophila BMC biotechnology, 15 (1), 22)*. The modified HSP70.2 promoter (SEQ ID NO: 1) contained in the cassette is heat-inducible, thereby enabling a favorable production by the production of the marker gene without the need for chemicals.

The ability of the *Tetrahymena thermophila* artificial chromosome of the present invention (TtAC2) to produce TtsfGFP-6xHis protein proves that chromosome construction and transformation is successful and that TtAC2 works effectively in the cell. The sfGFP protein emits fluorescent light, making it easier to trace. The marker gene used in the invention may be replaced by a DNA sequence encoding homologous and/or heterologous genomic DNA, RNA and/or proteins and/or derivatives thereof. Thus, the artificial chromosome of the present invention can be used for recombinant production of any desired sequence in *Tetrahymena thermophila*.

Thus, the artificial chromosome of the present invention is constructed as to contain "5´NTS containing C3 origin, rDNA genes and 3´NTS" sequence (SEQ ID NO: 3) on one arm and a 3´NTS sequence (SEQ ID NO: 2), Neo4 cassette and TtsfGFP marker gene cassette on the other arm. The direction of the transcription of the gene cassettes on both arms is configured from the center to the telomeres, thereby reducing the potential encounter of DNA replication and transcription complexes with biomimetics.

In a preferred embodiment of the present invention, *Tetrahymena thermophila* artificial chromosome of the present invention (TtAC2) has a nucleotide sequence that is at least 80% homologous to SEQ ID NO: 4.

The *Tetrahymena thermophila* artificial chromosome of the present invention may optionally be in circular or linear form by virtue of the telomere cassette it contains. The specific telomere cassette sequence of the present invention comprises a fragment sequence that can be cut at both ends with the SfiI enzyme and released to convert the circular chromosome into a linear

chromosome. The conservation of the telomere sequences at the ends of the artificial chromosome in linear form is ensured by 3´NTS sequences.

In another aspect, the present invention provides a method for producing the *Tetrahymena thermophila* artificial chromosome (TtAC2) described above, comprising the following steps:

- introducing the vector backbone for the *Tetrahymena thermophila* artificial chromosome,
- inserting a modified HSP70.2 promoter (SEQ ID NO: 1), a telomere sequence, a 3´NTS sequence (SEQ ID NO: 2), a Neo4 cassette, and a "5´NTS containing C3 origin, rDNA genes and 3´NTS" sequence (SEQ ID NO: 3) to the vector backbone (pUC19) by biomimetics of the rDNA minichromosome, and
- obtaining the artificial chromosome.

In a preferred embodiment of the method of the present invention, said method further comprises the step of inserting the TtsfGFP marker gene cassette or a DNA sequence encoding homologous and/or heterologous genomic DNA, RNA and/or proteins and/or derivatives thereof.

In an embodiment of the method of the present invention, said vector backbone is pUC19, so that the *Tetrahymena thermophila* artificial chromosome obtained by the method has a nucleotide sequence that is at least 80% homologous to SEQ ID NO: 4.

In another aspect, the present invention provides a method for recombinantly producing homologous and/or heterologous genomic DNA, RNA and/or proteins and/or derivatives thereof by using the *Tetrahymena thermophila* artificial chromosome as a production system/platform. Said method comprises the following steps:

- introducing a ciliate host cell of *Tetrahymena thermophila* or Tetrahymenidae family member that contains a recessive allele other than the C3 origin allele on the rDNA minichromosome of the macronucleus,
- extrachromosomally integrating the circular and/or linear *Tetrahymena thermophila* artificial chromosome (TtAC2) containing the dominant C3 origin in the macronucleus rDNA minichromosome into a ciliate host cell of *Tetrahymena thermophila* or Tetrahymenidae family member containing a recessed allele other than the C3 origin allele, and
- producing DNA, RNA and/or protein and/or derivatives thereof encoded by the DNA, RNA and/or protein encoded by a homologous and/or heterologous sequence carried on the *Tetrahymena thermophila* artificial chromosome by *Tetrahymena thermophila* or by a ciliate host cell member of the Tetrahymenidae family.

Microinjection, electroporation, and biolistic bombardment methods are usually preferred for transforming the Tetrahymena cells with the newly created vector. In preferred embodiments of the method of the present invention, biolistic gun method is used for the transformation of the artificial chromosome. *Tetrahymena thermophila* artificial chromosome can be directed to the macronucleus of *Tetrahymena thermophila* or Tetrahymenidae family member ciliate host cell by biolistic gun method.

In preferred embodiments of the method of the present invention, said *Tetrahymena thermophila* or Tetrahymenidae family member ciliate host cell is a cultured vegetative ciliate cell or two conjugating ciliate cells.

In another aspect, the present invention provides a transformant Tetrahymenidae family member ciliate host cell that is formed by transformation of TtAC2 bearing at least one homologous and/or heterologous sequence which is obtained as a result of recombinant production method of homologous and/or heterologous genomic DNA, RNA and/or proteins and/or derivatives thereof by using the *Tetrahymena thermophila* artificial chromosome of the invention as a production system/platform.

In another aspect, the present invention provides the use of the *Tetrahymena thermophila* artificial chromosome as a production system/platform for recombinantly producing homologous and/or heterologous genomic DNA, RNA and/or proteins and/or derivatives thereof.

Therefore, the present invention is performed by inserting a modified HSP70.2 promoter (SEQ ID NO: 1), a 3´NTS sequence (SEQ ID NO: 2), and "5´NTS containing C3 origin, rDNA genes and 3´NTS" sequence (SEQ NO: 3) into the pUC19 vector backbone. In the scope of the present invention, the vector backbone was also added with a telomere sequence, Neo4 neomycin resistance cassette and TtsfGFP marker gene cassette. With this structure, the present invention is a novel artificial chromosome vector that can be used for the production of recombinant protein in *Tetrahymena thermophila* or in a cilate host cell of the Tetrahymenidae family member. Said *Tetrahymena thermophila* artificial chromosome shows at least 80% homology with SEQ ID NO: 4.

A) Construction of *Tetrahymena Thermophila* Artificial Chromosome (TtAC2)

In the construction the TtAC2 artificial chromosome of the invention, the vector pUC19 was used as the backbone vector for *E. coli* cloning *(Yanisch-Perron, C., Vieira, J., & Messing, J. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mpl8 and pUC19 vectors, Gene, 33 (1), 103-119)*. This vector is chosen because of the restriction enzyme diversity in the cloning region, the small size of the vector (2.6 kbp), and the intermediate copy origin

(100+). According to the biomimetics design, the 3´NTS region (1.8 kbp), Neo4 cassette (2 kbp), TtsfGFP cassette (2 kbp), telomere region (2.4 kbp) and rDNA gene locus containing C3 origin (10.6 kbch) were respectively cloned into the pUC19 vector (Figure 1). As a result of these steps, artificial chromosome TtAC2 was obtained. The elements that make up *Tetrahymena thermophila* artificial chromosome TtAC2, their sequence and the size of the artificial chromosome has been simulated with biomimetics as to resemble to *Tetrahymena thermophila* macronucleus rDNA chromosome (Figure 1).

The wild-type Tetrahymena rDNA chromosome does not contain centromere, similarly TtAC2 artificial chromosome does not contain centromere. In the artificial chromosome, the direction of transcription is configured from the center to the telomeres. Thus, an artificial chromosome was obtained with the rDNA locus on one arm and the marker gene (TtsfGFP) and Neo4 cassettes on the other arm. The TtAC2 artificial chromosome, which can be used in circular form, can also be used in linear form by cutting from the SfiI cutting region located between the telomere regions. The circular and linear forms of the artificial chromosome TtAC2 and the conversion of the circular form to the linear form are shown in Figure 2-A.

1. Insertion of 3´NTS Element (3´non-transcribed spacer)

The *Tetrahymena thermophila* macronucleus rDNA chromosomes have a 3´NTS region as a copy in the region between the telomere and the rDNA gene sequence at both ends. The 3´NTS region contains the telomere associated DNA sequence. One of the tasks of these subtelomeric sites is to ensure replication and maintenance of the end regions of chromosomes containing the DNA locus *(Challoner, PU, Amine, AA, Pearlman, RE, & Blackburn, EH 1985. Conserved arrangements of repeated DNA sequences in nontranscribed spacers of ciliate ribosomal RNA genes: evidence for molecular coevolution, Nucleic acids research, 13 (7), 2661-2680)*. Another task is to include elements consisting of repetitive transcription and translation terminating DNA sequences. Finally, these regions are responsible for the formation of nucleosomes in the telomere region and help to pack the chromosome into the DNA-protein complex *(Budarf, ML, Blackburn, EH 1986. Chromatin structure of the telomeric region and 3´-nontranscribed spacer of Tetrahymena ribosomal RNA genes. Biological Chemistry, 261 (1), 363-369)*.

Purified genomic DNA from the *Tetrahymena thermophila* SB210 strain was used as the source DNA to produce the 1.8 kbp long 3´NTS sequence by Polymerase Chain Reaction (PCR). The 3´NTS region was generated using the relevant forward and reverse primers. With the ends of the produced 3´NTS sequence of 1.8 kbp long, the pUC19 vector was cut as to have sticky ends by restriction with the restriction enzymes KpnI (Thermo, #ER0521) and BamHII (Thermo,

#ER0052). The cut pUC19 vector and 3´NTS sequence containing DNA bands were visualized on a 0.6% agarose gel and cut from the gel under a UV cabinet at 365 nm wavelength. The respective sequence and vector were then purified using gel purification kit (Thermo, GeneJet Gel Extraction Kit, #K0691). The sticky end 3'NTS sequence and pUC19 vector were ligated by incubating with T4 DNA ligase (NEB, M0202S) at 16 °C for 16 hours and transformed into *E. coli* XL1Blue competent cells with heat shock of 42 °C. After incubation at 37 °C for 16 hours, the master plate (containing ampicillin-LB agar) was made from transformant competent cells and the selected colonies were screened by PCR. PCR products were run on 1% agarose gel and positive colonies were detected by imaging. Selected positive colonies were inoculated on LB broth supplemented with ampicillin and incubated, after 16 hours the plasmids were purified with Thermo GeneJET plasmid miniprep kit (#K0502). As a result, pUC19-3´NTS vector (TtAC2.1) was obtained (Figure 1). Successful production of the pUC19-3´NTS vector was confirmed by double restriction fingerprint analysis with KpnI-BamHI restriction enzymes and single restriction with SalI (Thermo, #ER0642) enzyme. The subsequent cloning steps were similarly performed.

2. Insertion of Neo4 Gene Cassette

The Neo4 cassette contains the paramomycin-resistant Neo4 gene, which contains a heavy metal cadmium-inducible MTT1 promoter and a BTU2 termination sequence. This cassette is added to the artificial chromosome structure for the selection of positive *Tetrahymena thermophila* cells containing the vector.

In the PCR reaction, pNeo4 vector *(Mochizuki, K., 2008. High efficiency transformation of Tetrahymena using a codon-optimized neomycin resistance gene. Gene, 425(1), 79-83)* was used as the source DNA and 2 kbp long DNA was amplified by the PCR reaction. Said 2 kbp long DNA band was purified from the agarose gel. The vector pUC19-3´NTS (TtAC2.1) and the Neo4 gene cassette were made to become sticky-ended using restriction enzymes SalI (Thermo, #ER0642) and BamHII (Thermo, #ER0052) and then ligated. The resulting vector was transformed into *E. coli* XL1Blue competent cells with a 42 °C heat shock. Positive colonies were detected by colony PCR. After plasmid isolation, it was confirmed that pUC19-3´NTS-Neo4 (TtAC2.2) vector was successfully generated with double restriction SalI (Thermo, #ER0642) and BamHII (Thermo, #ER0052) as well as single restrictions with HindIII and EcoRI (Figure 1).

3. Insertion of TtsfGFP Gene Cassette

Within the scope of the invention, the GFP gene cassette was used as a marker gene to demonstrate the reproducibility of a recombinant protein, therapeutic drug or enzymes using the

Tetrahymena thermophila artificial chromosome TtAC2. The use of toxic substances or heavy metals as promoter stimulators in the production of recombinant proteins is not in line with the "Good Manufacturing Practices" (GMP) principles. Therefore, the modified HSP70.2 promoter is used in the GFP gene cassette, which can be physically triggered by heat shock and is suitable for good manufacturing practices. The transcription and translation terminating BTU2 termination sequence is also included in this cassette. Said cassette that contains HSP70.2 promoter (Heat Shock Promoter) *(Barchetta, S., La Terza, A., Ballarini, P., Pucciarelli, S., & Miceli, C. 2008. Combination of two regulatory elements in the Tetrahymena thermophila HSP70-1 Eukaryotic cell, 7 (2), 379-386; Shang, Y., Song, X., Bowen, J., Corstanje, R., Gao, Y., Gaertig, J., & Gorovsky. , MA 2002. A robust inducible-repressible promoter greatly facilitates gene knockouts, conditional expression, and overexpression of homologous and heterologous genes in Tetrahymena thermophila. Proceedings of the National Academy of Sciences, 99 (6), 3734-3739)* and BTU2 termination sequence was synthetically manufactured to Shinegene Company.

HSP70.2 promoter and BTU2 sequences which were synthetically generated and cloned into the pUC57 vector were amplified by PCR reaction. With this reaction; The AscI restriction sequence was added to the 3´ end of the HSP70.2 promoter sequence and the PmeI restriction sequence was added before the BTU2 termination sequence. To add the TtsfGFP gene cassette to the artificial chromosome design, the TtsfGFP gene from the pVTtsfGFP vector was amplified with primers containing the AscI and PmeI restriction sequence *(Yilmaz, G., & Arslanyolu, M. 2015. Efficient expression of codon-adapted affinity tagged super folder green fluorescent protein for synchronous protein localization and affinity purification studies in Tetrahymena thermophila. BMC biotechnology, 15(1), 22)*. The TtsfGFP gene was inserted between the HSP70.2 promoter and the BTU2 sequences by the overlapping PCR method, and the 6xHis affinity tag encoding sequence was also added to the C-terminus of TtsfGFP during the ligation. DNA of the vector containing the Cfr9I-HSP70.2/TtsfGFP-6xHis/BTU-BamHI gene expression cassette which was ligated with the overlapping PCR technique was transformed to E. coli. The vector DNA purified from the positive E. coli transformants and the sticky-ended TtsfGFP-6xHis cassette were purified from agarose gel. The TtsfGFP-6xHis cassette which was prepared with the vector pUC19-3´NTS-Neo4 (TtAC2.2) which was purified by restriction with BamHI (Thermo, #ER0052) and Cfr9I (Thermo, #ER0171) enzymes was ligated and transformed into E. coli. The plasmid isolated from the positive clones detected by colony PCR procedures was taken to double enzyme restriction with BamHI and Cfr9I enzymes and single enzyme restriction with EcoRI and HindIII enzymes, and the presence of TtsfGFP-6xHis gene cassette was confirmed. This newly created vector was named pUC19-3´NTS-Neo4-TtsfGFP (TtAC2.3) (Figure 1).

4. Insertion of Telomere Sequence

The designed artificial chromosome is planned to be used both circularly and linearly. In order to protect the ends of the chromosome in linear form, telomeres must also be added to the ends of the chromosome in accordance with biomimetics. Telomere sequences generally assume roles of maintaining the stability of linear chromosomal ends, completing DNA replication at the ends, and performing nuclear transport during meiosis and karyokinesis *(Kirk, K. E., Blackburn, E. H. 1995. An unusual sequence arrangement in the telomeres of the germ-line micronucleus in Tetrahymena thermophila. Genes & development, 9(1), 59-71).*

Tetrahymena thermophila telomere cassette is produced as a 2.4 kbp long product with a PCR reaction in which pPXV-GFP vector *(Hauser, K., Haynes, W. J., Kung, C., Plattner, H., & Kissmehl, R. (2000). Expression of the green fluorescent protein in Paramecium tetraurelia. European journal of cell biology, 79(2), 144-149)* was used as a source DNA, and the primers that attach the ApaI and XhoI restriction enzymes to the appropriate positions were used. The purified vector pUC19-3´NTS-Neo4-TtsfGFP (TtAC2.3) and the telomere cassette sequence were cut with restriction enzymes ApaI (Thermo, #ER1411) and XhoI (Thermo, #ER0691) to make the sticky ends and ligated. The structural positioning is confirmed by XhoI and ApaI double restriction enzyme digestion. pUC19-3´NTS-Neo4-TtsfGFP-Telomere (TtAC2.4) vector that could be linerazed by SfiI single enzyme restriction was finally formed after cloning the sequence of the telomere cassette (Figure 1).

5. Insertion of *Tetrahymena thermophila* "5´NTS containing C3 origin, rDNA genes and 3´NTS" Sequence

Wild type 21 kbp long *Tetrahymena thermophila* macronuclear rDNA chromosome contains an origin region (5´NTS) *(Reischmann, KP, Zhang, Z., & Kapler, GM 1999. Long range cooperative interactions regulate the initiation of the Tetrahymena thermophila rDNA minichromosome. Nucleic acids research, 27 (15), 3079-3089)*, the locus containing the rDNA genes and the 3´NTS sequence as 2 copies in the palindromic orientation and is protected by telomere at the ends. The 5´NTS alleles containing the *Tetrahymena thermophila* rDNA origin sequence are differentiated and the B type allele is recessive compared to the C3 type allele due to a 42 base deletion. The rDNA replication superiority or dominance order of alleles is as C3>B>C3 rmm1 or C3 rmm4 *(Yaeger, PC, Orias, E., Shaiu, WL, Larson, DD, & Blackburn, EH 1989. The replication advantage of a free linear rRNA gene is restored by somatic recombination in Tetrahymena thermophila, Molecular and cellular biology, 9 (2), 452-460)*. Transformant cells obtained from circular vectors formed by the presence of only one copy of this rDNA locus (5´NTS-rDNA genes-3´NTS), which are found in two palindromic copies, have been shown to maintain stable copies of this vector within the

macronucleus genome *(Yu, G. L., & Blackburn, E. H. 1989. Transformation of Tetrahymena thermophila with a mutated circular ribosomal DNA plasmid vector. Proceedings of the National Academy of Sciences, 86(21), 8487-8491)*. Furthermore, the vector pMND-1 containing the entire palindromic rDNA gene locus without deletion reached 9,000 copies *(Blomberg, P., Randolph, C., Yao, C. H., & Yao, M. C. 1997. Regulatory sequences for the amplification and replication of the ribosomal DNA minichromosome in Tetrahymena thermophila. Molecular and cellular biology, 17(12), 7237-7247).*

In the light of the above information, using biomimetics principles, at least one copy of "5´NTS containing C3 origin, rDNA genes and 3´NTS" sequences for each was required to be used. Genomic DNA from wild type *Tetrahymena thermophila* C3.368.1 strain (Stock ID: SD0026, Cornell University Tetrahymena Stock Center) containing C3 origin was purified and used as source to provide allelic dominance during DNA replication following macronucleus division. Artificial chromosomes having C3 origin were transformed into *Tetrahymena thermophila* CU428 strain having B origin. Thus, the transformants were directed by C3> B allelic predominance antibiotic selection, and the number of artificial chromosome copies in the vegetative cell macronucleus increased.

Since the large size 10.3 kbp long "5´NTS containing C3 origin-rDNA genes-3´NTS" sequence is to be amplified, PCR reaction was performed using Herculase Fusion DNA Polymerase II enzyme (Agilent, #600675) and the genomic DNA purified from wild-type *Tetrahymena thermophila* C3368.1 strain as the source DNA. The steps of the PCR reaction were optimized as follows: initial denaturation at 92 °C for 2 minutes, 10 cycles of 92 °C denaturation for 10 cycles, hybridization for 20 seconds at 56 °C, elongation at 68 °C for 7 minutes, 20 cycles of denaturation at 20 °C for 10 seconds, hybridization at 56 °C for 20 seconds, 68 °C elongation for 7 minutes (so as to increase 20 seconds in each cycle) and one-time 68 °C final elongation. Thus, the first stage product of the two-stage PCR reaction was produced. This PCR product was used as the source of DNA in the second reaction, and the amount was increased.

After the second PCR step, the PCR product containing 10.3 kbp sequence of "5´NTS containing C3 origin-rDNA genes-3´NTS" was purified from the gel, and was integrated into the pJET1.2/ blunt-ended cloning vector with CloneJet PCR Cloning Kit (Thermo, #K1231). The size of the cloned region was confirmed by PCR and the DNA sequence was verified by DNA sequence analysis. The *E. coli* clone containing the pJET1.2/blunt vector carrying the DNA fragment containing the "5´NTS containing C3 origin, rDNA genes and 3´NTS" sequence was then produced, and the vector was purified and the DNA fragment which was cut by XhoI-Cfr9I (Thermo, #ER0691 - #ER1171) double enzyme restriction was purified from the gel. Then, artificial

chromosome intermediate vector (TtAC2.4) which was cut by the same enzymes and purified and "5´NTS containing C3 origin, rDNA genes and 3´NTS" sequence were ligated. The vector formed after ligation was named as TtAC2.5. After the cloning steps, cloning was confirmed by both the double XhoI-Cfr9I restriction and the single ApaI and XbaI restrictions.

In order to facilitate the industrial use of the TtAC2.5 vector, deletion of the BamHI restriction enzyme site (which is found between the Neo4 and TtsfGFP cassettes seen in the TtAC2.3 vector construct) was initiated by Xma-SalI enzyme restriction of the TtsfGFP and Neo4 cassettes. Some changes to the HSP70.2, TtsfGFP gene and BTU2 sequence in the TtsfGFP gene cassette were also performed at this stage. The first is the convertion of the G (guanine) base of the ATG start codon of the region encoding the artificial 11 amino acid HSP70 protein sequence, which is involuntarily added to the N-terminus of a produced protein and could function as a possible IRES element sequence residing at the 3´ end of the HSP70.2 promoter, to the T (thymine) base by mutation. This was carried out by PCR using a specific primer set. Furthermore, during amplification with PCR primers, an A (adenine) base was added prior to the PmeI enzyme restriction sequence at the 3´ end of the HSP70.2 promoter sequence. PmeI sequence was inserted on the forward primer of TtsfGFP and AscI sequence was inserted on the reverse primer. The SfaAI restriction enzyme sequence was also inserted on the reverse primer of the BTU2 sequence. The TtsfGFP gene cassette was regenerated by ligating the modified HSP70.2 promoter, TtsfGFP gene and BTU2 sequences with overlapping PCR. No changes were performed on the Neo4 cassette sequence, but the SfaAI restriction enzyme recognition sequence instead of BamHI was added to the forward primer. TtsfGFP and Neo4 cassettes were ligated with overlapping PCR and integrated into the vector with XmaI-SalI enzymes. These steps yielded the artificial chromosome TtAC2 (SEQ ID NO: 4). Sequential accuracy of the TtAC2 artificial chromosome was confirmed by DNA sequence analysis (Figure 1).

B) Transformation of *Tetrahymena thermophila* Artificial Chromosome (TtAC2) to Vegetative Cells by Biolistic Gun Method

The gold particles are coated with designed chromosome DNAs and sent into the cell with the help of a biolistic gun under certain pressure and the artificial chromosome (TtAC2) vector DNAs are transferred to the nucleus while passing through the cell.

After the vectors to be transformed are prepared in a linear or circular form as to contain 2 µl DNA in 1 µl, the gold particles must be coated with DNA. To the tubes containing gold particles for plating, respectively 1 µl of linearized or circular vector DNA, 10 µl 2.5 M $CaCl_2$ and 4 µL of spermidine were rapidly added. After each addition, it was vortexed for about 3 seconds and all

components were shaken for 10 minutes at 4 °C. At the end of shaking, the samples were centrifuged at 4 °C and 10,000 g for 6 seconds to collect DNA-coated gold particles. The supernatant was gently removed with a micropipette and the washing steps proceeded. Chromosome DNA and DNA-bound gold particles were washed and centrifuged with 100 µl of 70% ethanol. After the supernatant was removed by pipetting again, 100 µl of 100% ethanol was added to the pellet and the same process was repeated. Subsequently, 10 µl of 100% ethanol was added and dissolved on gold particles bound to the vector DNA and this mixture was transferred to the middle part of the macrocarrier chamber of the biolistic gun. The macrocarrier was stored in the desiccator to remove alcohol.

Following these processes, the stage of preparing the cells to be transformed was started. For this purpose, Tetrahymena thermophila CU428 cells of B origin were grown in SPP medium, and when they reached a density of $2x10^5$ cells/ml, they were fasted in 50 ml fasting buffer (10 mM Tris-HCl pH 7.4). On the day of transformation, starving cells at a density of $2x10^5$ cells/ml and a volume of 50 ml were collected by centrifugation at 1100 g for 3 minutes in a conical bottom tube to obtain 10 million cells for each firing. After centrifugation, the supernatant was removed and 10 million cells for each firing were dissolved in 500 µl of 10 mM Tris-HCl (pH 7.5). Sterilized circular "whatmann 50" filter paper placed in a Petri dish (100 mm) was moistened with 1 ml of 10 mM Tris-HCl (pH 7.5). 500 µl of the cells dissolved in fasting buffer were spread to 3/4 of the filter paper starting from the middle.

When the DNA samples in the cells and the macrocarrier were ready, the firing step was started with the biolistic gun. A 900 psi rupture disc wetted with isopropanol was placed in the holding apparatus. The holding apparatus is mounted by rotating it onto the threaded helium port in the particle bombardment apparatus. After the macrocarrier was placed in place, the petri dish containing the cells was placed on the lowest shelf of the device to provide the necessary conditions for firing. Firing was performed when the conditions were ready. 10 ml of SPP was added to the transformed cells at 30 °C and allowed to incubate at 30 °C until further firing was complete. All biolistic gun transformation was performed in a sterile cabinet.

C) Selection of Positive Transformants by Antibiotic Suppression

After transformation, cells on sterile filter paper were transferred to the medium by discarding the paper into 1XSPP medium. The medium was incubated for 3 hours at 30 °C without agitation, allowing the cells to recover. At the end of this period, 0.5 µg/ml $CdCl_2$ and 100 µg/ml paromomycin were added to the medium. Cells were diluted and 100 µl of cells is transferred into each well of a 96-well microplate. While these cells were monitored for 6 days at 30 °C, the

antibiotic amount was increased to 300 µg/ml on the 3rd day and to 600 µg/ml on the 5th day. $CdCl_2$ was added at a reduced concentration during this time. These processes were also applied to bulk cultures. On the 10th day of antibiotic selection, the antibiotic amount was increased to 1000 µg/ml and the selection was continued and positive transformant cells showing dynamic viability were transferred to glass tubes. During the monthly passaging of the tubed cells, the paromomycin suppression applied to the "antibiotic experimental group (Par+)" was reduced by 100 µg/ml and the transformant cells were maintained at this suppression for months. In addition to this experimental group, "non-antibiotic experiment group (Par-)" was also tubed, but antibiotic addition to the cells was not performed during monthly passaging. During the first 2 months, paromomycin and cadmium were added to the antibiotic-suppression tubes every 3 days. Once a week Pen/Strep and Amphotericin were added to the cells in all tubes and these cells were transferred to NEFF medium and passagings were performed. After the first 2 months, the cells in the tube were monitored for a long term by monthly passaging.

As a part of the invention, 2 main experimental groups are used: with antibiotic-suppression and without antibiotic-supression. These groups are as follows:

<u>Transformant cells that are paromomycin-suppressed (Par+):</u>

1. Circular TtAC2
2. Linear TtAC2

<u>Transformant cells that are not paromomycin-suppressed (Par-):</u>

3. Circular TtAC2
4. Linear TtAC2

D) Recombinant TtsfGFP Production by Using TtAC2 Artificial Chromosome as Protein Expression Vector

Untransformed *Tetrahymena thermophila* CU428 cells were used as negative control of the production assay of recombinant TtsfGFP. Modified HSP70.2 promoters of transformant cell lines were induced by heat shock from weekly or monthly samples, production of TtsfGFP-6xHis recombinant protein was monitored by microscopic methods and SDS-PAGE and western blot analyzes. The cells to be monitored were grown in shaker incubator in 15 ml or 25 ml NEFF medium until the logarithmic growth phase is reached. After incubation, the cells were pelleted by centrifugation at 1100 g for 5 minutes in a swing rotor centrifuge. The medium, previously added Pen/Strep and amphotericin, was heated at 38 °C, and a volume sufficient to dissolve the pellet was added very slowly onto the cell pellets to dissolve the cells. The dissolved cells were

transferred to larger surface flasks containing sterile 15 ml or 25 ml medium. The cells were subjected to heat shock for 3 hours in a 38 °C water bath. Light and fluorescence microscope analyzes of the cells were performed after this 3 hour TtsfGFP generation period. For this purpose, the Olympus Ix53 microscope with fluorescent apparatus was used with a GFP filter and a Black-and-white 10x-IR camera (for imaging at the end of 24 months) and the Olympus Bx50 microscope with the GFP filter and the DP72 camera (for imaging at the end of the 5th month) (40X magnification). *In-vivo* microscopic cell analyzes showed that the cells had strong GFP fluorescence at 5 months (Figure 2-D) and at 24 months (Figure 2-E).

SDS-PAGE *(Shapiro, A. L., Viñuela, E., & Maizel Jr, J. V. 1967. Molecular weight estimation of polypeptide chains by electrophoresis in SDS-polyacrylamide gels. Biochemical and biophysical research communications, 28(5), 815-820)* and western blot analyzes were performed in order to observe the GFP fluorescence of the total proteins which are purified from some of the cells that were grown to show TtsfGFP-6xHis recombinant protein production by TtAC2. Monthly grown cells were pelleted by centrifugation at 1100 g for 5 min, washed with Tris-HCl pH 7.5, supernatant is removed by centrifugation at 8000 rpm for 5 min, and cell pellets were stored at -86 °C for protein analysis. Cell pellets were taken into protein isolation using T100B (25 mM Tris-HCl pH 7.5, 300 mM NaCl, 1% Triton-X-100, 2 mM PMSF, 1x Protease Inhibitor Cocktail) buffer. When purified total proteins were taken to GFP fluorescence analysis with SDS-PAGE gel (5-12%) without boiling; 5-month-old cells containing Par+ circular TtAC2 (Figure 2-B, well 2) and Par+ linear TtAC2 (Figure 2-B, well 3) were found to have strong GFP fluorescence. In the transformant cell experimental group that is not antibiotic-suppressed, 5-month-old cells containing Par- circular TtAC2 (Figure 2-B, well 4) and Par- linear TtAC2 (Figure 2-B, well 5) had a weakened level of GFP fluorescence. Said purified total protein samples were also taken for western blot analysis with mouse monoclonal anti-GFP primary antibody (1:1000 dilution) (Roche 11814460001) and anti-mouse secondary antibodies (1:10000 dilution) (Sigma, Cat.No:A5225). Imaging was performed by TMB colorimetric method (Figure 2-C). The purified total proteins were boiled and loaded onto SDS-PAGE gel (5-12%). The presence of GFP protein was high in 5-month-old cells containing Par+ circular TtAC2 (Figure 2-C, well 2) and Par+ linear TtAC2 (Figure 2-C, well 3). In the 5-month-old transformant cell experimental group that was not under antibiotic suppression on the other hand, GFP was found to be weak in cells containing Par- circular TtAC2 (Figure 2-C, well 4) and Par- linear TtAC2 (Figure 2-C, well 5). 6XHis-TtsfGFP (~ 31 kDa) produced in *E. coli* was used as a positive control in SDS-PAGE and Western blot analyzes, by addition into the protein marker (Figure 2-B, well 1 and Figure 2-C, well 1, blue protein band below the 35 kDa line). As negative control, untransformed Tetrahymena CU428 total protein was used but no fluorescence or protein bands were seen (data not provided).

E) Determination of *Tetrahymena thermophila* TtAC2 Artificial Chromosome Copy Number by Real-Time PCR Method

The neomycin resistance gene, which is not present in the *Tetrahymena thermophila* genome but is found in the artificial chromosomes of the invention, was used as marker gene to determine the number of copies of TtAC2 vectors. For real-time PCR, Taqman probe and primers which are suitable for the neomycin resistance-marker gene were designed. In the same examples, Taqman probes and primers were also designed for the catalase gene (GenBank: XP_001026590.2), which would help to detect the number of cells per nanogram of DNA due to its presence in the Tetrahymena genome as a single copy. DNA isolation from 1 week, 2 month and 7 month cell pellets for Par+ and Par- experimental groups was performed using DNA isolation kit (Promega, Wizard Genomic DNA Purification Kit, A1120) to determine the number of copies of artificial chromosome vectors. In this way, the quality and concentration of the Real-Time PCR experiments were investigated and the nanograms of the DNAs were determined. Purified DNAs were run on agarose gel to confirm their usability in Real-Time PCR.

Taqman Probe and primers designed according to neomycin resistance gene using Primer3 program (http://bioinfo.ut.ee/primer3-0.4.0/) are as follows:

Taqman Probe: 5' FAM TCTGGTTTCATCGACTGTGG TAMRA 3'

Forward primer: 5' CTGCTTACCCAATATCATG 3'

Reverse primer: 5' CAAGTTCTTCAGCAATATCA 3'

The robe and primers designed for the catalase gene are as follows:

Taqman Probe: 5' FAM AATCCCTGAAAGAGTCGTCCACGC TAMRA 3'

Forward primer: 5' TGACAACCAAAATTCCGTCACTGCA 3'

Reverse primer: 5' CCCCAAGCACCAGCACCTTTG 3'

For the real time PCR assay, the probe and primer concentrations described above were diluted to the final concentration of 10,000 nM. For the reaction they were used at a concentration of 500 nm. In order to constitute a “DNA Copy Number Standard Curve” which will be used as a base in Real-Time PCR for TtAC2 copy number calculations, TtAC2 vector whose base content and number are known was used. To generate a standard curve, three independent dilutions from 267 ng/µl TtAC2 vector were performed. The R^2 value of the plotted standard curve is 0.989 and the productivity is found to be 101.7%. The number of vector copies at each dilution of 1:10 is calculated as (Standard 1: 5.270.000.000, Standard 2: 527.000.000, Standard 3: 52.700.000,

Standard 4: 5.270.000, Standard 5: 527.000, Standard 6: 52.700, Standard 7: 5.270 and Standard 8: 527 copies). The curve equation to be used in the experimental groups was determined as y = 3,282x + 39,144 (Table 1).

In order to constitute the "DNA Copy Number Standard Curve" of the catalase gene, pVGF-catalase vector whose vector size and base sequence are known (267 ng/µl) were prepared, and 3 independent dilutions were performed and used in Real-time PCR for curve formation. Catalase gene standard copy numbers are determined as; Standard 1: 7,690,000,000, Standard 2: 769,000,000, Standard 3: 76,900,000, Standard 4: 7,690,000, Standard 5: 769,000, Standard 6: 76,900, Standard 7: 7,690 and Standard 8: 769. The R^2 value of the standard curve is 0.997 and the productivity is found to be 89.2%. These values are in a reliable and usable range. Thus, the equation to be used to determine the number of cells during sampling of experimental groups is determined as y = 3,612x + 43,889 (Table 1).

The real-time PCR reaction was established with the Brilliant III Ultra-Fast QPCR Master Mix (Catalog #600880, Agilent Technologies) kit. The real-time PCR reaction was carried out with a program of 40 cycles of 15 sec at 95 °C and 30 sec at 60 °C after the 3 minutes of pre-denaturation at 95 °C. Genomic DNAs that were purified from cells containing antibiotic-suppressed Par+ circular TtAC2 and Par+ linear TtAC2 and Par- circular TtAC2 and Par-linear TtAC2 were used along with their non-template reaction controls for real-time PCR reaction. Genomic DNAs of the experimental group were diluted to 50 ng/µl and the template genomic DNA was used as 100 ng/µl in the reaction. When necessary, reactions containing genomic DNA template amounts of 50 ng/µl were also established for control purposes. Reactions were carried out using separate probes and primers for the Neo4 and catalase genes. The reactions were carried out in 3 replicates each. After determining the number of cells for each sample with the catalase gene, the experimental group neomycin gene copy number, ie TtAC2 copy number, was calculated using standard curve equations as the mean value (Table 1).

Table 1

	Experimental Groups	**1st Week**	**2nd Month**	**7th Month**
Par+	Circular TtAC2	416	745	1929
	Linear TtAC2	5	601	2735
Par-	Circular TtAC2	1243	725	607
	Linear TtAC2	1933	138	52

TtAC2 artificial chromosome vector is maintained in the macronucleus genome in circular and linear form of the Par+ and Par- experimental groups (Table 1). The circular and linear forms of TtAC2 differ in the number of copies between Par+ and Par- experimental groups. While the experimental groups without antibiotic suppression showed a rapid increase in the number of copies in the first weeks, it was observed that the number of copies decreased in time. While the experimental groups under antibiotic suppression had a slow increase in the number of copies in the first weeks, the increase in the number of copies gained momentum in time. This suggests that instead of applying antibiotic suppression immediately after transformation, the cells should be suppressed after the first week based on the natural predominance of the C3 allele on the B allele. Thus, it is possible to achieve a much higher number of TtAC2 copies. In summary, the data in Table 1 show that TtAC2 has an artificial chromosome properties and can be maintained by the cell for a long time without paromomycin suppression, which is of great importance for industrial use.

Sequence Listing

<110> Anadolu University

<120> Tetrahymena thermophila Artificial Chromosome 2

<130> 4

<160> 4

<170> PatentIn version 3.5

<210> 1

<211> 868

<212> DNA

<213> Artificial Sequence

<220>

<223> Modified HSP70.2 promoter

<400> 1

cccgggtaaa ttaagtgcaa gcaaatctta aaaatattat aaattgaaaa taactaaaat 60

taaacaattc cttaaattta tattgtcaga tattttagct atgtaatttg caagcaatat 120

ttgtttacaa aacatgaatt aataaaacta aaaaatattc attatataat taaaaaagga 180

acttttaaac aaaatatctt aatttctata ttaatttatt aattatttac tttaaagaaa 240

aagagaaaac ttctaaatca aattaatttt aatagaagtt tttaatctat ctaaatagta 300

ataattgtta tatttgaata atataaattt atcaataaat aaatacattt tagcaaaata 360

tttatttta tatttctgct gataattaag tgagtgttat agctaatttt ttaattttaa 420

aatatttta aaaaaatttg gttgtttggt ttgtgataga taattaatta tttgtaagtt 480

ttagtttata atcaattttt tttaatcttt ctaaaattaa taataagatt ttttaatctt 540

tccttttct tatttgctat taataatcaa tttattcata aaggatattt aataatttag 600

aaacaattta ctgatttcta aaattctttt gttattaaaa aatacacttc aaattaaaaa 660

aaataaaatt tccaacaatt ctatcagaaa tttttttcgc tttaaacagc taaaaagaaa 720

tttgatcaaa aataaataat caattaattc aaaaaaacaa aaaaaacact ctcaacaaaa 780

atcttcaaaa agcgactcaa acaaataatt aaaaaagttt acaaaaattt ctggaaaata 840

aaaagctgaa ggaattggta gtttaaac 868

<210> 2

<211> 1877

<212> DNA

<213> Artificial Sequence

<220>

<223> 3´NTS sequence

<400> 2

ggtaccgaat ttctacgatc tgctgagatt acgcccgtct ccttagattt atctcatctc 60

cctttatttt ttacttctgc tggggttgtt aacctcttta agaaattttt tttattttga 120

tttgtttaat ttaattttgt ttaactttag taaatttttt tccttttttc actccactgg 180

ggttattaaa tacttagaga ttttttattt ttatcaaata atttatgaac tcattaaaac 240

aaaacaaaac aaagttaaaa aaaactcaat agatttgcta ataagatgca aagcagctat 300

gaggcaattt ttctcatttg gaaagcttca atcaactgct tataaataaa ttataaatca 360

atattaaaaa tgttaaagtt ttatgttatt tgttagtaaa aaaattgaat agttgtgttt 420

aagctgatat aagtctttat gcatgatatg ttaaaagtta cgcttaaaat tatgcttttt 480

acgcagaatg agcttagcta aatttttttct caaagtaaat ttttttttaat gcaaaatgaa 540

tgaaaaaatt ttagtatttt ataaaaattc ttcattcaaa ttttacccac ttatcaattt 600

atttttttttt gtgactaaag cagtcccaga gccttttctc taaaagttga attttattaa 660

caatgccact tttatagaaa attttgcatg gatttcctgg ggctccaatg gaaaaattgc 720

gaaagtggat tttgaatgaa aaagtgaatg taaaaattaa agtaaaattt tgctttataa 780

aatgaatgaa aatttaaaac aagcaaaatg aatgaaaaaa tctttgcatt ttaacaaaat 840

ttttcattca aaatttcacc cactttatca ataacttttt ttttccccca gcccgtcaaa 900

agaccctatt atttgttttc ctaaaagttg acttttatta agaaacaaaa tgaatgaaaa 960

aatctttgta ttttaacaaa atttttcatt caaaatttca cccacttatc aatatatttt 1020

ttttaagtcc tgcctgcagg ctaatattct ttcctaaaag ttgactttta ttaagaagca 1080

aaatgaatga aaaaatcttt gtattttaac caaatttttc attcaaaatt tcacccactt 1140

atcaatattt tttttgggac caaaccaaaa gacctaaaga tttgatttaa aaagttgact 1200

tttttcaaga aaaccacttt attagataaa tctctttttt accacggctt gtccaatgaa 1260

taatttgcta aagtggattt gaataaaaat ttttttgcgt gtaaaaatgc gctaaactac 1320

gcttagattt taactttatc ccactttaat ttcaagcgta aaaataaaaa tcccacacaa 1380

aaattaagtg gaaattgatg caaaaatttc actaaaattt aattcaataa atatgtaaaa 1440

atggttgatc tctataattt atgagatttg cattatttaa ggcttataag aaattttaaa 1500

ttaacgcgga agcttcattt ttagataaaa tttattaatc atcattaatt tcttgaaaaa 1560
cattttattt attgatcttt tataacaaaa aacccttcta aaagtttatt tttgaatgaa 1620
aaacttataa aaatttatga aaactacaaa aaataaaatt tttaattaaa ataattttga 1680
taagaacttc aatctttgac tagcttagtc atttttgaga tttaattaat attttatgtt 1740
tattcatata taaactattc aaaatattat agaatttaaa cattttaaca tcttaatcat 1800
tgataaataa ccaaaaatca aagtattaca tcaataaata acttttactc aatgtcaaag 1860
aaattattat agggccc 1877

<210> 3
<211> 10356
<212> DNA
<213> Artificial Sequence

<220>
<223> 5´NTS containing C3 origin, rDNA genes and 3´NTS sequence

<400> 3
cccgggtcat cgtacacgta tagctcgtac ggtttttttg ctttttgttg ttagttttat 60
agccttcagc actatgttgt tgaattttac cttcgaaaat aacttaaaat tgagtaataa 120
ttgggtttaa aatttaaatt tgagtagata agaattagat gtttatattc tgctaatttc 180
actggtgaaa atgtagcaaa tagaaattat tttaatctaa taaactagca aatagtattt 240
aaaacaaaaa tatttgtttt ttatgttgta aaatgtttta aattagataa aatttacaaa 300
tttacaaatt ttcaagcaaa ataggttcta aaaaatgaga aaatattaca tattttagct 360
atttgactac tttaatgcta gtaaattaaa atgaatttaa ttcattttca ctttaaaaca 420
cttattttaa taaaatatat gattttaaaa tgataaaata ttttttaaga ggtaaattta 480
agaaattagt taaattttaa agaaaaagca tctaaaaatg gacaaaaatg aagtatttcc 540
tttttttata catttaaatg ctagaaaatt taagtaaaac atttataaat aaaagtaaaa 600
tagttttagg aataagagta aatagttttt tttatgtaaa aaacatttta tcaatttcat 660
ttattcattt tagttaaatt tttcattcac aaaaaacttt ttttggtatg gtaaaataaa 720
gactttataa agataactta aagaaaaagt ttatctagaa ttaaaaatat tgattttgaa 780
aattgctcat tagatatttt tttggcaaaa aaaaaaacaa aaatagtaaa aaatcacttt 840
ttttgagagt tgaaaaaaag acttagaaaa aattttaaaa gtgtaaaaaa agacttagag 900
aaaaaatcaa aaagagataa aaagacttag agaaaattta taaattaaaa atgatagaaa 960

agtaaaattt attttatatt ttttaatcat ttaaatgcta gtaaatttaa ataaaacata 1020
tataaaaaaa cataaaacaa ttttaacagc atgcgtatat cattttttat atgtaaaaaa 1080
cattttatca atttcattta ttcattttag ttaaatttta cattcaaaat aaatttttt 1140
tgattaaata aagagttata aagagaactt aaagaaaaag tttatctaga attaaaaata 1200
ttgattttga aaattgctca ttagaaattt ttttggcaaa aaaaaaaaca aaaatagtaa 1260
acttagaaaa atttttgaaa aatgaaaaaa aaagacttag agaaaaaaat caaaaagtga 1320
aaaaagactt agaaaaattt ttaaaaatga aaaaatgatt taggagaaat tttgagattg 1380
cgcttagatt ttgtgtgaag ccacttacaa aaaatgagcg gactcgctca aatatttaag 1440
tggactcgca taaaaatgag tggagtcact aaaaaaatta agtgaactca cttaaaaatg 1500
agtggagcca ctcaaaaaat taagcggact cgcttaatat tcgcggagtt aaacaaaaat 1560
aagtggactc acagaaaaat taagcggatt gcgctaaaaa atgagtggac tcactcaaaa 1620
atgaagcaga ctcgcttaaa aatgagtgga gccactcaaa aagttcagca gagccactta 1680
aaaatttagc ttaaaatcag ctctaaatta aattagactt agtgaaaaat agcgaaaatg 1740
aaaaaaatga aaaaatgaat gaaaactgaa aaatttacaa gggattgaaa attttggcag 1800
agtctttttt ttggcaaaaa aaaaaacaaa aatagtaaac cttccgaact tttttgactt 1860
tgagaaaaat tctttggcaa aaaaaataaa aataatatca ggggggtaaa aatgcatatt 1920
taagaagggg aaacatctcc ggatcaaaaa taaaatatca gctcgatttg agcttcagta 1980
agatttcctt ttgggcaacc aaggataacg ataatgaagc gctaactgag cagacgtttt 2040
tctctatggc ttcggctttt agtcgatggc cgctgagggt ctgttgaagg tttttctgga 2100
ttaaggctcg tattagagca aatggcctga ctgaaatttt catgaaggct gtaaattcac 2160
tgcaaagctt cgcagaaact tttcccagtg acacttgttg tatcgatatc tatgcagata 2220
ttgttacaaa taacgcaaca cgctagtact gttataaatc ggtgaaatcg cagatgttat 2280
taacagctag caacaaagtt gactagagtc gaagagatgc gatagagttt tctcattgtg 2340
ccttcgaaga ttttagcaac tagaagaaac taatagtaaa cgaaacgatg cgggatctat 2400
gtataaagct taatctaacg atatagctga gtactgatct attacaacgc gtcagttctc 2460
gatgaactat taatcttttg tgaaccaacc ttttggaaca ctattcaaaa aatgagcaag 2520
ctgttggaag atgcaaatcg gaaaatagcg agcaaatttt gaggatagta acctggttga 2580
tcctgccagt tacatatgct tgtcttaaat attaacccat gcatgtgcca gttcagtatt 2640
gaacagcgaa actgcgaatg gctcattaaa acagttatag tttatttgat aattaaagat 2700
tacatggata accgagctaa ttgttgggct aatacatgct taaaattccg tgtcctgcga 2760
ccggaacgta tttattagat attagaccaa tcgcagcaat gtgattgaga tgaatcaaag 2820
taactgatcg gatcgaggtt tacctcgata aatcatctaa gtttctgccc tatcagctct 2880
cgatggtagt gtattggact accatggcag tcacgggtaa cggagaatta gggttcgatt 2940

ccggagaagg agcctgagaa acggctacta caactacggt tcggcagcag ggaagaaaat 3000
tggccaatcc taattcaggg agccagtgac aagaaatagc aagctgggaa acttacgttt 3060
ctacggcatt gaaatgagaa cagtgtaaat ctcttagcga ggaacaattg gagggcaagt 3120
catggtgcca gcagccgcgg taattccagc tccaatagcg tatattaaag ttgttgcagt 3180
taaaaagctc gtagttgaac ttctgttcag gttcatttcg attcgtcgtg tgaaactgga 3240
catacgtttg caaactaaaa tcggccttca ctggttcgac ttagggagta aacattttac 3300
tgtgaaaaaa ttagagtgtt ccaggcaggt ttagcccga atacattagc atggaataat 3360
ggaataggac taagtccatt ttattggttc ttggatttgg taatgattaa tagggacagt 3420
tgggggcatt agtatttaat agtcagaggt gaaattcttg gatttattaa ggactaacta 3480
atgcgaaagc atttgccaaa gatgttttca ttaatcaaga acgaaagtta ggggatcaaa 3540
gacgatcaga taccgtcgta gtcttaacta taaactatac cgactcggga tcggctggaa 3600
taaatgtcca gtcggcaccg tatgagaaat caaagtcttt gggttctggg ggaagtatgg 3660
tacgcaagtc tgaaacttaa aggaattgac ggaacagcac accagaagtg gaacctgcgg 3720
cttaatttga ctcaacacgg ggaaactcac gagcgcaaga cagagaaggg attgacagat 3780
tgagagctct ttcttgattc tttgggtggt ggtgcatggc cgttcttagt tggtggagtg 3840
atttgtctgg ttaattccgt taacgaacga gaccttaacc tgctaactag tctgcttgta 3900
aataacaggt tgtacttctt agagggacta ttgtgcaata agccaatgga agtttaaggc 3960
aataacaggt ctgtgatgcc cctagacgtg ctcggccgca cgcgcgttac aatgactggc 4020
gcaaaaagta tttcctgtcc tgggaaggta cgggtaatct tattaatacc agtcgtgtta 4080
gggatagttc tttggaattg tggatcttga acgaggaatt tctagtaagt gcaagtcatc 4140
agcttgcgtt gattatgtcc ctgccgtttg tacacaccgc ccgtcgcttg tagtaacgaa 4200
tggtctggtg aaccttctgg actgcgacag caatgttgcg gaaaaataag taaaccctac 4260
catttggaac aacaagaagt cgtaacaagg tatctgtagg tgaacctgca gatggatcat 4320
taacacaatt aacaaacctt aacttatgta ctttcgaaga gaacttcggt tttcttcgag 4380
gttttattgt cacacctagt gtgaataaaa atttttcata tgtctaagat ctggataaca 4440
tccaaaacga aaagaaaact ttcaacggtg gatatctagg ttcccgtgac gatgaagaac 4500
gcagcgaaat gcgatacgca atgcgaattg cagaaccgcg agtcatcaga tctttgaacg 4560
caagtggtgg aggtgtaaaa accttcatgt ttgtttcagt gtggaaagga atcacgcatc 4620
ttaatgcgat tgaagccgtc aaaagcttct ctcgttaaac gtgatgggtg gtcgagcaat 4680
cgccgccaga acgaagtagt cacattccag taatgtgaac attcgttcag gcatcaaggc 4740
gaatgctcac tatgctactc atagaaaaat tacatttttc tcactacacc tgaaacaagc 4800
aagattaccc gctgaactta agcatatcag taagcggagg aaaagaaact aactaggata 4860
gccccagtaa tggcgaatga acaggctaaa gctcaaagtg aaaatctgaa gtggtcaaca 4920

caacagaatt gtaatctaaa gggtcaacct gaaactaagc tcctctcata agttccttgg 4980
gacaggacgt caaagagggt gacaaccccg tagtcggaga ggaaggctgg tgtaagggag 5040
atttcaaaga gtcgggttgt ttggaattgc agccctaagt gggagataaa cttcttctaa 5100
agctaaatat acacgggaga ccgatagcga acaagtactg cgaaggaaag atgaaaagaa 5160
ctttgaaaag agggctaaaa gacttgaaac cgttgagaag gaagctgtag aagagcaata 5220
aactggacgg cgcataaggg ggaagtacta atcactgcag agtcgatacg taaaaggtcg 5280
atgagtaagg aaatggtaca gaacttgcta caccggtcag aagacaaaat gggttcagat 5340
tgaaggagtc acctgagatc gggcagcaat gcagatcaaa aggaaaactt caaactggac 5400
tgaggggcct aagggcgatt ttgtcaaaat ggcttctact gacccgtctt gaaacacgga 5460
ccaaggagtc tatcaattaa gcgagtgata gggtggaaaa acccgtccgc gaaacgaaag 5520
tgagtacaag gtgccaagcc gcaaggtagc agcatcaccc gccttgagtc tccgcgaagg 5580
gttcgaggaa gagcttaatt gttaggaccc gaaagatggt gaactacgct tgaatagggt 5640
gaagccaggg gaaactctgg tggaagctcg tagcgatact gacgtgcaaa tcgttcgtca 5700
aatttgagtg taggggcgaa agactaatcg aaccatctag tagctggttc cctccgaagt 5760
ttctctcagg atagcaagag caagtacgca gttttattag gtaaagcgaa tgattgaagg 5820
actcgggagt cctaagaact tcgacctatt ctcaaacttt aaattggtaa gagccgcgga 5880
gtttacttaa atgaactctc gggaagaacg cagtgctctt gagttgggcc atttttggta 5940
agcagaactg gcgatgaggg atgaacctaa cgttgagata aggcgcccaa atgcacgctc 6000
atcagatacc acaaaaggtg ttggttcata cggacagcag gacggtggct atggaagtta 6060
gaatccgcta aggagtgtgt aacaactcac ctgccgaatg aactagccct gaaaatgaat 6120
ggcgctgaag cgtgttgccg atactcaacc atcagagcaa atgcgaggct ttgatgagta 6180
ggagggcgtg atcgttgcct agaagtattg gcgtgagcct atatggagca gcgattagtg 6240
agatcttggt ggtagtagca aatattcaaa tgagaacttt gaagaccgaa gtggagaagg 6300
gttccatgag aacagcaatt gttcatgggt tactcgatcc taagacatag gttaactcct 6360
tgcaatacaa gaagacattc gttttcgttg tcaaaaggga atgaggttaa tattcctcaa 6420
gctggacgtg gtataggcgg taacgcaaag aaacccggaa acgtcagcag gtgtcactgg 6480
aagagttatc ttttcttttt aacatactat ggccatgaaa ttggattatc cagagatatc 6540
ggctgtatgt atggcagagc agctcaccct aagagctgtc agttgcgcgc ctgatgaccc 6600
ttgaaaatct gggggagaca taatttcacg ccagttcgta cccataaccg catcaggtct 6660
ccaaggttag cagcctctgg tccatagaac aatgtagata agggaagtcg gcaaattgga 6720
tccgtaactt cgggataagg attggctctg aggatcgggt ataaaggctt tgtaatgata 6780
tccaagcttg tttgttagtg tggcaacatg ctgatagact tgcgaacgat gaatttgcaa 6840
ggtaggtttc ggccgtcttt atacaattaa cgatcaactc agaactgaag cggacaaagg 6900

taatccgact gtttaataaa aacaaagcat tgtgacggcc tcaacaggtg atgacacaat 6960
gtgatttctg cccagtgctc tgaatgtcaa agtgacgcaa ttcaaccaag cgcgggtaaa 7020
cggcgggagt aactatgact ctctaaatag caatatttac ctttggaggg aaaagttatc 7080
aggcatgcac ctggtagcta gtctttaaac caatagattg catcggttta aaaggcaaga 7140
ccgtcaaatt gcgggaaagg ggtcaacagc cgttcagtac caagtctcag gggaaacttt 7200
gagatggcct tgcaaagggt atggtaataa gctgacggac atggtcctaa ccacgcagcc 7260
aagtcctaag tcaacagatc ttctgttgat atggatgcag ttcacagact aaatgtcggt 7320
cggggaagat gtattcttct cataagatat agtcggacct ctccttaatg ggagctagcg 7380
gatgaagtga tgcaacactg gagccgctgg gaactaattt gtatgcgaaa gtatattgat 7440
tagttttgga gtactcgtaa ggtagccaaa tgcctcgtca tctaattagt gacgcgcatg 7500
aatggattaa tgagattacc actgtcccta tctactatct agcgaaccca cagctaaggg 7560
aacgggctta gaataatcag cggggaaaga agaccctgtt gagcttgact ctagtctaac 7620
tttgtgaaat ggcacgtggg gtatagccta ggtgggagag caatcgatcc tgtaaaacca 7680
ctacccacgt agtcattttg cttatttcgt gaagaaaaga ctggtgcaaa ccagttctaa 7740
gattaaggtc atttattgac tgattttgcg aagacatggt taggggggga gtttgtctgg 7800
ggcggaatgc ctgttaaacc ataacgcagg cgtcctaagt gtagctcagt gagaacggaa 7860
atctcacgta gaacaaaagg gtaaaagcta cattgatttt gattttcagt aggaatacaa 7920
accgcgaaag cgtggcctat cgatccttta actttacaag ttttaagcta gaggtgtcag 7980
aaaagttacc acagggataa ctggcttgtg gcagccaaga gttcatatcg acgttgcttt 8040
ttgatccttc gatgtcggct cttcctatca ttgtgaagca gaattcacaa cggtgtcgga 8100
ttgttcaccc gctaataggg aacgtgagct gggcttagac cgtcgtgaga caggttagtt 8160
ttaccctact gatgaaacga tgttgcgaca gtaatttaag ttagtacgag aggaacactt 8220
aaatcagata attggtaaat acggttgtct gaaaagacaa tgccgtgaag ctaccatctg 8280
ttggattatg actgaaggcc tctaagtcag aatccatgct ggaaagcaat gtctaagtgt 8340
gatgataaac gaaaaaaaat aaaaattaag ttcgaaaggt agagcgggga agagcgaaaa 8400
agcttgacct taactgctaa tcgtattcca aattatcatc tacgtaaatc ttttgtagac 8460
gacttaacat ggaacgggta ttgtaagcat gagagtagaa tttctacgat ctgctgagat 8520
tcagcccgtc tccttagatt tatctcatct ccctttattt tttacttctg ctggggttgt 8580
taacctcttt aagaaatttt ttatgttttg atttgtttaa tttaattttg tttaacttta 8640
gtaaattttt ttccttttt cactcactgg gttattaaat acttagagat tttacatttt 8700
atcaaataat ttatgaactc attaaaacaa aacaaaacaa agttaaaaaa aactcaatag 8760
atttgctaat aagatgcaaa gcagctatga ggcaattttt ctcatttgga aagcttcaat 8820
caactgctta taaataaatt ataaatcaat attaaaaatg ttaaagtttt atgttatttg 8880

ttagtaaaaa aattgaatag ttgtgtttaa gctgatataa gtctttatgc atgatatgtt 8940
aaaagttacg cttaaaatta tgctttttac gcagaatgag cttagctaaa ttttttctca 9000
aagtaaattt tttttaatgc aaaatgaatg aaaaaatttt agtattttat aaaaattctt 9060
cattcaaatt ttacccactt atcaatttat tttttttgt gactaaagca gtcccagagc 9120
ctttctctaa aagttgaatt ttattaacaa tgccactttt atagaaaatt ttgcatggat 9180
ttcctggggc tccaatggaa aaattgcgaa agtggatttg aatgaaaaag tgaatgtaaa 9240
aattaaagta aaattttgct ttataaaatg aatgaaaatt taaaacaagc aaaatgaatg 9300
aaaaaatctt tgcattttaa caaaattttt cattcaaaat ttcacccact tatcaatact 9360
ttttttttctc cagccgtcaa agacctatat tgttttccta aaagttgact tttattaaga 9420
aacaaaatga atgaaaaaat ctttgtattt taacaaaatt tttcattcaa aatttcaccc 9480
acttatcaat atatttttttt taagtcctgc ctgcaggcta atattctttc ctaaaagttg 9540
acttttatta agaagcaaaa tgaatgaaaa aatctttgta ttttaaccaa atttttcatt 9600
caaaatttca cccacttatc aatatttttt ttgggaccaa accaaaagac ctaaagattt 9660
gatttaaaaa gttgactttt ttcaagaaaa ccactttatt agataaatct ctttttttacc 9720
acggcttgtc caatgaataa tttgctaaag tggatttgaa taaaaatttt tttgcgtgta 9780
aaaatgcgct aaactacgct tagattttaa ctttatccca ctttaatttc aagcgtaaaa 9840
ataaaaatcc cacacaaaaa ttaagtggaa attgatgcaa aaatttcact aaaatttaat 9900
tcaataaata tgtaaaaatg gttgatctct ataatttatg agatttgcat tatttaaggc 9960
ttataagaaa ttttaaattt aacgcggaag cttcattttt agataaaatt tattaatcat 10020
cattaatttc ttgaaaaaca ttttatttat tgatctttta taacaaaaaa cccttctaaa 10080
agtttatttt tgaatgaaaa acttataaaa atttatgaaa actacaaaaa ataaaatttt 10140
taattaaaat aattttgata agaacttcaa tctttgacta gcttagtcat ttttgagatt 10200
taattaatat tttatgttta ttcatatata aactattcaa aatattatag aatttaaaca 10260
ttttaacatc ttaatcattg ataaataacc aaaaatcaaa gtattacatc aataaataac 10320
ttttactcaa tgtcaaagaa attattgggg ctcgag 10356

<210> 4
<211> 21312
<212> DNA
<213> Artificial Sequence

<220>
<223> TtAC2

<400> 4

ttgcacttgt tcgctcagtt cagccataat atgaaatgct tttcttgttg ttcttacgga 60

ataccacttg ccacctatca ccacaactaa ctttttcccg ttcctccatc tcttttatat 120

ttttttttctc gatcgagttc aagagaaaaa aaaagaaaaa gcaaaaagaa aaaaggaaag 180

cgcgcctcgt tcatagaaaa agaaaattgc gggaaaggac tgtgttatga cttccctgac 240

taatgccgtg ttcaaacgat acctggcagt gactcctagc gctcaccaag ctcttaaaac 300

gagaattaag aaaaagtcgt catctttcga taagtttttc ccacagcaaa gcaatagtag 360

aaaaaaacaa tgggaaacgt tgaatgaaga caaagcgtcg tggtttaaaa ggaaatacgc 420

tcacgtacat gctagggaac aggaccgtgc agcggatcgg cctaggcggc cgcgatcccg 480

cgcatcaaca atattttcac ctgaatcagg atattcttct aatacctgaa tgctgttttc 540

ccaccccaac cccaacccca accccaaccc caaccccaac cccaacccca accccaaccc 600

caaccccaac cccaacccca accccaaccc caaccccaac cccaacccca accccaaccc 660

caaccccaac cccaacccca accccaaccc caaccccaac cccaacccca accccaaccc 720

caaccccaac cccaacccca accccaaccc caaccccaac cccaacccca accccaaccc 780

caaccccaac cccaacccca accccaaccc caataatttc tttgacattg agtaaaagtt 840

atttattgaa tgtaatactt tgatttttag ttatttatga atgattaaga tgttaaaatg 900

tttaaattct ataatattga aaaaggaaga gtatgagtat tcaacatttc cgtgtcgccc 960

ttattccctt ttttgcggca ttttgccttc ctgtttttgc tcacccagaa acgctggtga 1020

aagtaaaaga tgctgaagat cagttgggtg cacgagtggg ttacatcgaa ctggatctca 1080

acagcggtaa gatccttgag agttttcgcc ccgaagaacg ttttccaatg atgagcactt 1140

ttaaagttct gctatgtggc gcggtattat cccgtattga cgccgggcaa gagcaactcg 1200

gtcgccgcat acactattct cagaatgact tggttgagta ctcaccagat actcgagccc 1260

caataatttc tttgacattg agtaaaagtt atttattgat gtaatacttt gatttttggt 1320

tatttatcaa tgattaagat gttaaaatgt ttaaattcta taatattttg aatagtttat 1380

atatgaataa acataaaata ttaattaaat ctcaaaaatg actaagctag tcaaagattg 1440

aagttcttat caaaattatt ttaattaaaa attttatttt ttgtagtttt cataaatttt 1500

tataagtttt tcattcaaaa ataaactttt agaagggttt tttgttataa aagatcaata 1560

aataaaatgt ttttcaagaa attaatgatg attaataaat tttatctaaa aatgaagctt 1620

ccgcgttaaa tttaaaattt cttataagcc ttaaataatg caaatctcat aaattataga 1680

gatcaaccat ttttacatat ttattgaatt aaattttagt gaaatttttg catcaatttc 1740

cacttaattt ttgtgtggga tttttatttt tacgcttgaa attaaagtgg gataaagtta 1800

aaatctaagc gtagtttagc gcatttttac acgcaaaaaa atttttattc aaatccactt 1860

tagcaaatta ttcattggac aagccgtggt aaaaaagaga tttatctaat aaagtggttt 1920

tcttgaaaaa agtcaacttt ttaaatcaaa tctttaggtc ttttggtttg gtcccaaaaa 1980
aaatattgat aagtgggtga aattttgaat gaaaaatttg gttaaaatac aaagattttt 2040
tcattcattt tgcttcttaa taaaagtcaa cttttaggaa agaatattag cctgcaggca 2100
ggacttaaaa aaaatatatt gataagtggg tgaaattttg aatgaaaaat tttgttaaaa 2160
tacaaagatt ttttcattca ttttgtttct taataaaagt caacttttag gaaaacaata 2220
taggtctttg acggctggag aaaaaaaagt attgataagt gggtgaaatt ttgaatgaaa 2280
aattttgtta aaatgcaaag atttttttcat tcattttgct tgttttaaat tttcattcat 2340
tttataaagc aaaattttac tttaattttt acattcactt tttcattcaa atccactttc 2400
gcaatttttc cattggagcc ccaggaaatc catgcaaaat tttctataaa agtggcattg 2460
ttaataaaat tcaactttta gagaaaggct ctgggactgc tttagtcaca aaaaaaaata 2520
aattgataag tgggtaaaat ttgaatgaag aatttttata aaatactaaa atttttttcat 2580
tcattttgca ttaaaaaaaa tttactttga gaaaaaattt agctaagctc attctgcgta 2640
aaaagcataa ttttaagcgt aacttttaac atatcatgca taaagactta tatcagctta 2700
aacacaacta ttcaattttt ttactaacaa ataacataaa actttaacat ttttaatatt 2760
gatttataat ttatttataa gcagttgatt gaagctttcc aaatgagaaa aattgcctca 2820
tagctgcttt gcatcttatt agcaaatcta ttgagttttt tttaactttg ttttgttttg 2880
ttttaatgag ttcataaatt atttgataaa atgtaaaatc tctaagtatt taataaccca 2940
gtgagtgaaa aaaggaaaaa aatttactaa agttaaacaa aattaaatta aacaaatcaa 3000
aacataaaaa atttcttaaa gaggttaaca accccagcag aagtaaaaaa taaagggaga 3060
tgagataaat ctaaggagac gggctgaatc tcagcagatc gtagaaattc tactctcatg 3120
cttacaatac ccgttccatg ttaagtcgtc tacaaaagat ttacgtagat gataatttgg 3180
aatacgatta gcagttaagg tcaagctttt tcgctcttcc ccgctctacc tttcgaactt 3240
aatttttatt ttttttcgtt tatcatcaca cttagacatt gctttccagc atggattctg 3300
acttagaggc cttcagtcat aatccaacag atggtagctt cacggcattg tcttttcaga 3360
caaccgtatt taccaattat ctgatttaag tgttcctctc gtactaactt aaattactgt 3420
cgcaacatcg tttcatcagt agggtaaaac taacctgtct cacgacggtc taagcccagc 3480
tcacgttccc tattagcggg tgaacaatcc gacaccgttg tgaattctgc ttcacaatga 3540
taggaagagc cgacatcgaa ggatcaaaaa gcaacgtcga tatgaactct tggctgccac 3600
aagccagtta tccctgtggt aacttttctg acacctctag cttaaaactt gtaaagttaa 3660
aggatcgata ggccacgctt tcgcggtttg tattcctact gaaaatcaaa atcaatgtag 3720
cttttaccct tttgttctac gtgagatttc cgttctcact gagctacact taggacgcct 3780
gcgttatggt ttaacaggca ttccgcccca gacaaactcc ccccctaacc atgtcttcgc 3840
aaaatcagtc aataaatgac cttaatctta gaactggttt gcaccagtct tttcttcacg 3900

aaataagcaa aatgactacg tgggtagtgg ttttacagga tcgattgctc tcccacctag 3960
gctatacccc acgtgccatt tcacaaagtt agactagagt caagctcaac agggtcttct 4020
ttccccgctg attattctaa gcccgttccc ttagctgtgg gttcgctaga tagtagatag 4080
ggacagtggt aatctcatta atccattcat gcgcgtcact aattagatga cgaggcattt 4140
ggctacctta cgagtactcc aaaactaatc aatatacttt cgcatacaaa ttagttccca 4200
gcggctccag tgttgcatca cttcatccgc tagctcccat taaggagagg tccgactata 4260
tcttatgaga agaatacatc ttccccgacc gacatttagt ctgtgaactg catccatatc 4320
aacagaagat ctgttgactt aggacttggc tgcgtggtta ggaccatgtc cgtcagctta 4380
ttaccatacc ctttgcaagg ccatctcaaa gtttcccctg agacttggta ctgaacggct 4440
gttgacccct ttcccgcaat ttgacggtct tgccttttaa accgatgcaa tctattggtt 4500
taaagactag ctaccaggtg catgcctgat aacttttccc tccaaaggta aatattgcta 4560
tttagagagt catagttact cccgccgttt acccgcgctt ggttgaattg cgtcactttg 4620
acattcagag cactgggcag aaatcacatt gtgtcatcac ctgttgaggc cgtcacaatg 4680
ctttgttttt attaaacagt cggattacct ttgtccgctt cagttctgag ttgatcgtta 4740
attgtataaa gacggccgaa acctaccttg caaattcatc gttcgcaagt ctatcagcat 4800
gttgccacac taacaaacaa gcttggatat cattacaaag cctttatacc cgatcctcag 4860
agccaatcct tatcccgaag ttacggatcc aatttgccga cttcccttat ctacattgtt 4920
ctatggacca gaggctgcta accttggaga cctgatgcgg ttatgggtac gaactggcgt 4980
gaaattatgt ctcccccaga ttttcaaggg tcatcaggcg cgcaactgac agctcttagg 5040
gtgagctgct ctgccataca tacagccgat atctctggat aatccaattt catggccata 5100
gtatgttaaa aagaaaagat aactcttcca gtgacacctg ctgacgtttc cgggtttctt 5160
tgcgttaccg cctataccac gtccagcttg aggaatatta acctcattcc cttttgacaa 5220
cgaaaacgaa tgtcttcttg tattgcaagg agttaaccta tgtcttagga tcgagtaacc 5280
catgaacaat tgctgttctc atggaaccct tctccacttc ggtcttcaaa gttctcattt 5340
gaatatttgc tactaccacc aagatctcac taatcgctgc tccatatagg ctcacgccaa 5400
tacttctagg caacgatcac gccctcctac tcatcaaagc ctcgcatttg ctctgatggt 5460
tgagtatcgg caacacgctt cagcgccatt cattttcagg gctagttcat tcggcaggtg 5520
agttgttaca cactccttag cggattctaa cttccatagc caccgtcctg ctgtccgtat 5580
gaaccaacac cttttgtggt atctgatgag cgtgcatttg ggcgccttat ctcaacgtta 5640
ggttcatccc tcatcgccag ttctgcttac caaaaatggc ccaactcaag agcactgcgt 5700
tcttcccgag agttcattta agtaaactcc gcggctctta ccaatttaaa gtttgagaat 5760
aggtcgaagt tcttaggact cccgagtcct tcaatcattc gctttaccta ataaaactgc 5820
gtacttgctc ttgctatcct gagagaaact tcggagggaa ccagctacta gatggttcga 5880

ttagtctttc gcccctacac tcaaatttga cgaacgattt gcacgtcagt atcgctacga 5940
gcttccacca gagtttcccc tggcttcacc ctattcaagc gtagttcacc atctttcggg 6000
tcctaacaat taagctcttc ctcgaaccct tcgcggagac tcaaggcggg tgatgctgct 6060
accttgcggc ttggcacctt gtactcactt tcgtttcgcg gacgggtttt tccaccctat 6120
cactcgctta attgatagac tccttggtcc gtgtttcaag acgggtcagt agaagccatt 6180
ttgacaaaat cgcccttagg cccctcagtc cagtttgaag ttttcctttt gatctgcatt 6240
gctgcccgat ctcaggtgac tccttcaatc tgaacccatt ttgtcttctg accggtgtag 6300
caagttctgt accatttcct tactcatcga ccttttacgt atcgactctg cagtgattag 6360
tacttccccc ttatgcgccg tccagtttat tgctcttcta cagcttcctt ctcaacggtt 6420
tcaagtcttt tagccctctt ttcaaagttc ttttcatctt tccttcgcag tacttgttcg 6480
ctatcggtct cccgtgtata tttagcttta gaagaagttt atctcccact tagggctgca 6540
attccaaaca acccgactct ttgaaatctc ccttacacca gccttcctct ccgactacgg 6600
ggttgtcacc ctctttgacg tcctgtccca aggaacttat gagaggagct tagtttcagg 6660
ttgacccttt agattacaat tctgttgtgt tgaccacttc agattttcac tttgagcttt 6720
agcctgttca ttcgccatta ctggggctat cctagttagt ttcttttcct ccgcttactg 6780
atatgcttaa gttcagcggg taatcttgct tgtttcaggt gtagtgagaa aaatgtaatt 6840
tttctatgag tagcatagtg agcattcgcc ttgatgcctg aacgaatgtt cacattactg 6900
gaatgtgact acttcgttct ggcggcgatt gctcgaccac ccatcacgtt taacgagaga 6960
agcttttgac ggcttcaatc gcattaagat gcgtgattcc tttccacact gaaacaaaca 7020
tgaaggtttt tacacctcca ccacttgcgt tcaaagatct gatgactcgc ggttctgcaa 7080
ttcgcattgc gtatcgcatt tcgctgcgtt cttcatcgtc acgggaacct agatatccac 7140
cgttgaaagt tttcttttcg ttttggatgt tatccagatc ttagacatat gaaaaatttt 7200
tattcacact aggtgtgaca ataaaacctc gaagaaaacc gaagttctct tcgaaagtac 7260
ataagttaag gtttgttaat tgtgttaatg atccatctgc aggttcacct acagatacct 7320
tgttacgact tcttgttgtt ccaaatggta gggtttactt atttttccgc aacattgctg 7380
tcgcagtcca gaaggttcac cagaccattc gttactacaa gcgacgggcg gtgtgtacaa 7440
acggcaggga cataatcaac gcaagctgat gacttgcact tactagaaat tcctcgttca 7500
agatccacaa ttccaaagaa ctatccctaa cacgactggt attaataaga ttacccgtac 7560
cttcccagga caggaaatac tttttgcgcc agtcattgta acgcgcgtgc ggccgagcac 7620
gtctaggggc atcacagacc tgttattgcc ttaaacttcc attggcttat tgcacaatag 7680
tccctctaag aagtacaacc tgttatttac aagcagacta gttagcaggt taaggtctcg 7740
ttcgttaacg gaattaacca gacaaatcac tccaccaact aagaacggcc atgcaccacc 7800
acccaaagaa tcaagaaaga gctctcaatc tgtcaatccc ttctctgtct tgcgctcgtg 7860

agtttccccg tgttgagtca aattaagccg caggttccac ttctggtgtg ctgttccgtc 7920
aattccttta agtttcagac ttgcgtacca tacttccccc agaacccaaa gactttgatt 7980
tctcatacgg tgccgactgg acatttattc cagccgatcc cgagtcggta tagtttatag 8040
ttaagactac gacggtatct gatcgtcttt gatcccctaa ctttcgttct tgattaatga 8100
aaacatcttt ggcaaatgct ttcgcattag ttagtcctta ataaatccaa gaatttcacc 8160
tctgactatt aaatactaat gcccccaact gtccctatta atcattacca aatccaagaa 8220
ccaataaaat ggacttagtc ctattccatt attccatgct aatgtattcg ggctaaaacc 8280
tgcctggaac actctaattt tttcacagta aaatgtttac tccctaagtc gaaccagtga 8340
aggccgattt tagtttgcaa acgtatgtcc agtttcacac gacgaatcga aatgaacctg 8400
aacagaagtt caactacgag ctttttaact gcaacaactt taatatacgc tattggagct 8460
ggaattaccg cggctgctgg caccatgact tgccctccaa ttgttcctcg ctaagagatt 8520
tacactgttc tcatttcaat gccgtagaaa cgtaagtttc ccagcttgct atttcttgtc 8580
actggctccc tgaattagga ttggccaatt ttcttccctg ctgccgaacc gtagttgtag 8640
tagccgtttc tcaggctcct tctccggaat cgaaccctaa ttctccgtta cccgtgactg 8700
ccatggtagt ccaatacact accatcgaga gctgataggg cagaaactta gatgatttat 8760
cgaggtaaac ctcgatccga tcagttactt tgattcatct caatcacatt gctgcgattg 8820
gtctaatatc taataaatac gttccggtcg caggacacgg aattttaagc atgtattagc 8880
ccaacaatta gctcggttat ccatgtaatc tttaattatc aaataaacta taactgtttt 8940
aatgagccat tcgcagtttc gctgttcaat actgaactgg cacatgcatg ggttaatatt 9000
taagacaagc atatgtaact ggcaggatca accaggttac tatcctcaaa atttgctcgc 9060
tattttccga tttgcatctt ccaacagctt gctcattttt tgaatagtgt tccaaaaggt 9120
tggttcacaa aagattaata gttcatcgag aactgacgcg ttgtaataga tcagtactca 9180
gctatatcgt tagattaagc tttatacata gatcccgcat cgtttcgttt actattagtt 9240
tcttctagtt gctaaaatct tcgaaggcac aatgagaaaa ctctatcgca tctcttcgac 9300
tctagtcaac tttgttgcta gctgttaata acatctgcga tttcaccgat ttataacagt 9360
actagcgtgt tgcgttattt gtaacaatat ctgcatagat atcgatacaa caagtgtcac 9420
tgggaaaagt ttctgcgaag ctttgcagtg aatttacagc cttcatgaaa atttcagtca 9480
ggccatttgc tctaatacga gccttaatcc agaaaaacct tcaacagacc ctcagcggcc 9540
atcgactaaa agccgaagcc atagagaaaa acgtctgctc agttagcgct tcattatcgt 9600
tatccttggt tgcccaaaag gaaatcttac tgaagctcaa atcgagctga tattttattt 9660
ttgatccgga gatgtttccc cttcttaaat atgcattttt acccccctga tattattttt 9720
attttttttg ccaaagaatt tttctcaaag tcaaaaaagt tcggaaggtt tactattttt 9780
gtttttttttt ttgccaaaaa aaagactctg ccaaaatttt caatcccttg taaatttttc 9840

agttttcatt cattttttca tttttttcat tttcgctatt tttcactaag tctaatttaa 9900
tttagagctg attttaagct aaatttttaa gtggctctgc tgaacttttt gagtggctcc 9960
actcattttt aagcgagtct gcttcatttt tgagtgagtc cactcatttt ttagcgcaat 10020
ccgcttaatt tttctgtgag tccacttatt tttgtttaac tccgcgaata ttaagcgagt 10080
ccgcttaatt ttttgagtgg ctccactcat ttttaagtga gttcacttaa tttttttagt 10140
gactccactc atttttatgc gagtccactt aaatatttga gcgagtccgc tcatttttg 10200
taagtggctt cacacaaaat ctaagcgcaa tctcaaaatt tctcctaaat cattttttca 10260
tttttaaaaa tttttctaag tctttttca cttttgatt tttttctcta agtcttttt 10320
tttcattttt caaaaatttt tctaagttta ctatttttgt ttttttttt gccaaaaaaa 10380
tttctaatga gcaattttca aaatcaatat ttttaattct agataaactt tttctttaag 10440
ttctctttat aactctttat ttaatcaaaa aaaatttatt ttgaatgtaa aatttaacta 10500
aaatgaataa atgaaattga taaaatgttt tttacatata aaaaatgata tacgcatgct 10560
gttaaaattg ttttatgttt ttttatatat gttttattta aatttactag catttaaatg 10620
attaaaaaat ataaaataaa ttttacttt ctatcatttt taatttataa attttctcta 10680
agtcttttta tctcttttg attttttctc taagtctttt tttacacttt taaaattttt 10740
tctaagtctt tttttcaact ctcaaaaaaa gtgatttttt actatttttg ttttttttt 10800
tgccaaaaaa atatctaatg agcaattttc aaaatcaata tttttaattc tagataaact 10860
ttttctttaa gttatcttta taaagtcttt attttaccat accaaaaaaa gttttttgtg 10920
aatgaaaaat ttaactaaaa tgaataaatg aaattgataa aatgtttttt acataaaaaa 10980
aactatttac tcttattcct aaaactattt tacttttatt tataaatgtt ttacttaaat 11040
tttctagcat ttaaatgtat aaaaaaagga aatacttcat ttttgtccat ttttagatgc 11100
tttttcttta aaatttaact aatttcttaa atttacctct taaaaaatat tttatcattt 11160
taaaatcata tattttatta aaataagtgt tttaaagtga aaatgaatta aattcatttt 11220
aatttactag cattaaagta gtcaaatagc taaaatatgt aatattttct catttttag 11280
aacctatttt gcttgaaaat ttgtaaattt gtaaatttta tctaatttaa aacattttac 11340
aacataaaaa acaaatattt ttgttttaaa tactatttgc tagtttatta gattaaaata 11400
atttctattt gctacatttt caccagtgaa attagcagaa tataaacatc taattcttat 11460
ctactcaaat ttaaatttta aacccaatta ttactcaatt ttaagttatt ttcgaaggta 11520
aaattcaaca acatagtgct gaaggctata aaactaacaa caaaaagcaa aaaaaccgta 11580
cgagctatac gtgtacgatg acccgggtaa attaagtgca agcaaatctt aaaaatatta 11640
taaattgaaa ataactaaaa ttaaacaatt ccttaaattt atattgtcag atattttagc 11700
tatgtaattt gcaagcaata tttgtttaca aaacatgaat taataaaact aaaaaatatt 11760
cattatataa ttaaaaaagg aacttttaaa caaaatatct taatttctat attaatttat 11820

taattattta ctttaaagaa aaagagaaaa cttctaaatc aaattaattt taatagaagt 11880
ttttaatcta tctaaatagt aataattgtt atatttgaat aatataaatt tatcaataaa 11940
taaatacatt ttagcaaaat atttatttt atatttctgc tgataattaa gtgagtgtta 12000
tagctaattt tttaatttta aaatattttt aaaaaaattt ggttgtttgg tttgtgatag 12060
ataattaatt atttgtaagt tttagtttat aatcaatttt ttttaatctt tctaaaatta 12120
ataataagat tttttaatct ttccttttttc ttatttgcta ttaataatca atttattcat 12180
aaaggatatt taataattta gaaacaattt actgatttct aaaattcttt tgttattaaa 12240
aaatacactt caaattaaaa aaaataaaat ttccaacaat tctatcagaa atttttttcg 12300
ctttaaacag ctaaaaagaa atttgatcaa aaataaataa tcaattaatt caaaaaaaca 12360
aaaaaaacac tctcaacaaa aatcttcaaa aagcgactca aacaaataat taaaaaagtt 12420
tacaaaaatt tctggaaaat aaaaagctga aggaattggt agtttaaaca tgtctaaagg 12480
tgaagaatta ttcactggtg ttgttcctat tttagttgaa ttagacggtg acgttaacgg 12540
tcacaaattc tctgttagag gtgaaggtga aggtgacgct actaacggta aattaacttt 12600
aaaattcatt tgcactactg gtaaattacc tgttccttgg cctactttag ttactacttt 12660
aacttacggt gttcaatgct tctctagata ccctgaccac atgaaaagac acgacttctt 12720
caaatctgct atgcctgaag gttacgttca agaaagaact atttctttca aagacgacgg 12780
tacttacaaa actagagctg aagttaaatt cgaaggtgac actttagtta acagaattga 12840
attaaaaggt attgacttca aagaagacgg taacatttta ggtcacaaat tagaatacaa 12900
cttcaactct cacaacgttt acattactgc tgacaaacaa aaaaacggta ttaaagctaa 12960
cttcaaaatt agacacaacg ttgaagacgg ttctgttcaa ttagctgacc actaccaaca 13020
aaacactcct attggtgacg gtcctgtttt attacctgac aaccactact tatctactca 13080
atctgtttta tctaaagacc ctaacgaaaa aagagaccac atggttttat tagaattcgt 13140
tactgctgct ggtattactc acggtatgga cgaattatac aaacaccacc accaccacca 13200
ctgatgaggc gcgccgatcc ttaaattaaa aattcaatat atatttacaa actttcatat 13260
aaaataaata tattatataa aattaatttt tagtgtatta tattaacatt aaagcaccaa 13320
aaaaacgtgt taatatacta ctataaaata taatttattc caaattgact aaaatcatta 13380
ttttacaact catttgtata tatattttat gtcaattatt ttttttaact ttctaaaaaa 13440
aaaattcctc ttcacataca tgttagctct taaaaatttg tctgcaaatc caataataat 13500
attttttttt gccattaaat tttcaaattt ttactggaaa aatgcagcga tcgcattggc 13560
cggccatttc gatatacaat ttatttctaa aaaatattta aaaataaaaa ataataaggg 13620
ttttgaataa ctccttttaat ttaaatacac atttttaaat tttttttttag ctctttaaat 13680
attcataaaa ataaaaaata actaactaaa aataaataaa aagataataa tgattaaagg 13740
tataatactg aataagaaaa aacataatag agtacttatt ttttatatca ctatttttaa 13800

tatcttgaaa gcaaaacttt tttatatatc ttaaaatata ttgtatcgtt tattcaatta 13860
ttttctttaa atttcaaata tattgataaa aaagatgaca tgtttttttaa agaaaacatg 13920
aaatataaaa tagataaata tcaattattt tatttattaa atatataagc tgctcaaaac 13980
atagctcatt catcaattat aatatgtgaa tcattaattt tcaaaatatt actcattatt 14040
taggctatca tttatttttt attttcaatt atccgtttct attatatttt aatattaagt 14100
tgtgattctt gaattttgtg tcatgaatta tttgtaaatc ttttttatttc tgataaaaaa 14160
tataaattga ttgactcatg atttaaatca tgagtcaacc taactaattt tcaaaattct 14220
tctattctaa aatatagatg tgattcttga atctctcttg aatataaagt aatttttttat 14280
atttctgata taattcttag ctacgtgatt cacgatttat acaatgatcc atataaaata 14340
atgtaaatag tgtatatata tatattcgtc ttttttattc tttatataat ttaaaaaaat 14400
taaaaaaatt taataaagct ctaataaaat aaataataat actaaactta aaataatgat 14460
tgaacaagat ggtttacacg ctggttctcc cgccgcttgg gtcgaaagac ttttcggtta 14520
tgactgggct caacaaacca tcggttgctc tgatgccgcc gtcttccgtc tttctgctca 14580
aggtcgtcct gttcttttcg tcaagaccga cctttctggt gcccttaatg aacttcaaga 14640
tgaagctgcc cgtctttctt ggcttgccac caccggtgtt ccttgcgctg ctgtccttga 14700
cgttgtcact gaagccggta gagactggct tcttttaggt gaagtccccg gtcaagatct 14760
tctttcttct caccttgctc ctgccgaaaa agtttctatc atggctgatg ctatgcgtcg 14820
tcttcatacc cttgatcccg ctacctgccc tttcgaccac caagccaaac atcgtatcga 14880
acgtgctcgt actcgtatgg aagccggtct tgtcgatcaa gatgatcttg acgaagaaca 14940
tcaaggtctt gcccctgccg aacttttcgc cagacttaag gcccgtatgc ccgacggtga 15000
agatcttgtc gtcacccatg gtgatgcctg cttacccaat atcatggttg aaaatggtcg 15060
tttttctggt ttcatcgact gtggtcgtct tggtgtcgcc gaccgttatc aagatattgc 15120
cttagctacc cgtgatattg ctgaagaact tggtggtgaa tgggctgacc gtttccttgt 15180
cctttacggt atcgccgctc ccgattctca acgtatcgcc ttctatcgtc ttcttgacga 15240
attcttctga gatccttaaa ttaaaaattc aatatatatt tacaaacttt catataaaat 15300
aaatatatta tataaaatta atttttagtg tattatatta acattaaagc accaaaaaaa 15360
cgtgttaata tactactata aaatataatt tattccaaat tgactaaaat cattatttta 15420
caactcattt gtatatatat tttatgtcaa ttatttttt taactttcta aaaaaaaaaa 15480
ttcctcttca catacatgtt agctcttaaa aatttgtctg caaatccaat aataatattt 15540
ttttttgcca ttaaattttc aaatttttac tggaaaaatg cagtcgacta tctgcaggca 15600
tgcaagcttg gcgtaatcat ggtcatagct gtttcctgtg tgaaattgtt atccgctcac 15660
aattccacac aacatacgag ccggaagcat aaagtgtaaa gcctggggtg cctaatgagt 15720
gagctaactc acattaattg cgttgcgctc actgcccgct ttccagtcgg gaaacctgtc 15780

gtgccagctg cattaatgaa tcggccaacg cgcggggaga ggcggtttgc gtattgggcg 15840
ctcttccgct tcctcgctca ctgactcgct gcgctcggtc gttcggctgc ggcgagcggt 15900
atcagctcac tcaaaggcgg taatacggtt atccacagaa tcaggggata acgcaggaaa 15960
gaacatgtga gcaaaaggcc agcaaaaggc caggaaccgt aaaaaggccg cgttgctggc 16020
gtttttccat aggctccgcc cccctgacga gcatcacaaa aatcgacgct caagtcagag 16080
gtggcgaaac ccgacaggac tataaagata ccaggcgttt ccccctggaa gctccctcgt 16140
gcgctctcct gttccgaccc tgccgcttac cggatacctg tccgcctttc tcccttcggg 16200
aagcgtggcg ctttctcata gctcacgctg taggtatctc agttcggtgt aggtcgttcg 16260
ctccaagctg ggctgtgtgc acgaaccccc cgttcagccc gaccgctgcg ccttatccgg 16320
taactatcgt cttgagtcca acccggtaag acacgactta tcgccactgg cagcagccac 16380
tggtaacagg attagcagag cgaggtatgt aggcggtgct acagagttct tgaagtggtg 16440
gcctaactac ggctacacta gaagaacagt atttggtatc tgcgctctgc tgaagccagt 16500
taccttcgga aaaagagttg gtagctcttg atccggcaaa caaaccaccg ctggtagcgg 16560
tggtttttt gtttgcaagc agcagattac gcgcagaaaa aaaggatctc aagaagatcc 16620
tttgatcttt tctacggggt ctgacgctca gtggaacgaa aactcacgtt aagggatttt 16680
ggtcatgaga ttatcaaaaa ggatcttcac ctagatcctt ttaaattaaa aatgaagttt 16740
taaatcaatc taaagtatat atgagtaaac ttggtctgac agttaccaat gcttaatcag 16800
tgaggcacct atctcagcga tctgtctatt tcgttcatcc atagttgcct gactccccgt 16860
cgtgtagata actacgatac gggagggctt accatctggc cccagtgctg caatgatacc 16920
gcgagaccca cgctcaccgg ctccagattt atcagcaata aaccagccag ccggaagggc 16980
cgagcgcaga agtggtcctg caactttatc cgcctccatc cagtctatta attgttgccg 17040
ggaagctaga gtaagtagtt cgccagttaa tagtttgcgc aacgttgttg ccattgctac 17100
aggcatcgtg gtgtcacgct cgtcgtttgg tatggcttca ttcagctccg gttcccaacg 17160
atcaaggcga gttacatgat cccccatgtt gtgcaaaaaa gcggttagct ccttcggtcc 17220
tccgatcgtt gtcagaagta agttggccgc agtgttatca ctcatggtta tggcagcact 17280
gcataattct cttactgtca tgccatccgt aagatgcttt tctgtgactg gtgagtactc 17340
aaccaagtca ttctgagaat agtgtatgcg gcgaccgagt tgctcttgcc cggcgtcaat 17400
acgggataat accgcgccac atagcagaac tttaaaagtg ctcatcattg gaaaacgttc 17460
ttcggggcga aaactctcaa ggatcttacc gctgttgaga tccagttcga tgtaacccac 17520
tcgtgcaccc aactgatctt cagcatcttt tactttcacc agcgtttctg ggtgagcaaa 17580
aacaggaagg caaaatgccg caaaaaaggg aataagggcg acacggaaat gttgaatact 17640
catactcttc ctttttcaat attattgaag catttatcag ggttattgtc tcatgagcgg 17700
atacatattt gaatgtattt agaaaaataa acaaataggg gttccgcgca catttccccg 17760

aaaagtgcca cctgacgtct aagaaaccat tattatcatg acattaacct ataaaaatag 17820
gcgtatcacg aggccctttc gtctcgcgcg tttcggtgat gacggtgaaa acctctgaca 17880
catgcagctc ccggagacgg tcacagcttg tctgtaagcg gatgccggga gcagacaagc 17940
ccgtcagggc gcgtcagcgg gtgttggcgg gtgtcggggc tggcttaact atgcggcatc 18000
agagcagatt gtactgagag tgcaccatat gcggtgtgaa ataccgcaca gatgcgtaag 18060
gagaaaatac cgcatcaggc gccattcgcc attcaggctg cgcaactgtt gggaagggcg 18120
atcggtgcgg gcctcttcgc tattacgcca gctggcgaaa gggggatgtg ctgcaaggcg 18180
attaagttgg gtaacgccag ggttttccca gtcacgacgt tgtaaaacga cggccagtga 18240
attcgagctc ataggtaccg aatttctacg atctgctgag attacgcccg tctccttaga 18300
tttatctcat ctccctttat tttttacttc tgctggggtt gttaacctct ttaagaaatt 18360
ttttttattt tgatttgttt aatttaattt tgtttaactt tagtaaattt ttttcctttt 18420
ttcactccac tggggttatt aaatacttag agatttttta tttttatcaa ataatttatg 18480
aactcattaa aacaaaacaa aacaaagtta aaaaaaactc aatagatttg ctaataagat 18540
gcaaagcagc tatgaggcaa tttttctcat ttggaaagct tcaatcaact gcttataaat 18600
aaattataaa tcaatattaa aaatgttaaa gttttatgtt atttgttagt aaaaaaattg 18660
aatagttgtg tttaagctga tataagtctt tatgcatgat atgttaaaag ttacgcttaa 18720
aattatgctt tttacgcaga atgagcttag ctaaattttt tctcaaagta aatttttttt 18780
aatgcaaaat gaatgaaaaa attttagtat tttataaaaa ttcttcattc aaattttacc 18840
cacttatcaa tttatttttt tttgtgacta aagcagtccc agagcctttt ctctaaaagt 18900
tgaattttat taacaatgcc acttttatag aaaattttgc atggatttcc tggggctcca 18960
atggaaaaat tgcgaaagtg gattttgaat gaaaaagtga atgtaaaaat taaagtaaaa 19020
ttttgcttta taaaatgaat gaaaatttaa aacaagcaaa atgaatgaaa aaatctttgc 19080
attttaacaa aatttttcat tcaaaatttc acccacttta tcaataactt tttttttccc 19140
ccagcccgtc aaaagaccct attatttgtt ttcctaaaag ttgactttta ttaagaaaca 19200
aaatgaatga aaaaatcttt gtattttaac aaaatttttc attcaaaatt tcacccactt 19260
atcaatatat tttttttaag tcctgcctgc aggctaatat tctttcctaa aagttgactt 19320
ttattaagaa gcaaaatgaa tgaaaaaatc tttgtatttt aaccaaattt ttcattcaaa 19380
atttcaccca cttatcaata tttttttttgg gaccaaacca aaagacctaa agatttgatt 19440
taaaaagttg actttttttca agaaaaccac tttattagat aaatctcttt tttaccacgg 19500
cttgtccaat gaataatttg ctaaagtgga tttgaataaa aatttttttg cgtgtaaaaa 19560
tgcgctaaac tacgcttaga ttttaacttt atcccacttt aatttcaagc gtaaaaataa 19620
aaatcccaca caaaaattaa gtggaaattg atgcaaaaat ttcactaaaa tttaattcaa 19680
taaatatgta aaaatggttg atctctataa tttatgagat ttgcattatt taaggcttat 19740

aagaaatttt aaattaacgc ggaagcttca tttttagata aaatttatta atcatcatta 19800
atttcttgaa aaacatttta tttattgatc ttttataaca aaaaaccctt ctaaaagttt 19860
atttttgaat gaaaaactta taaaaattta tgaaaactac aaaaaataaa atttttaatt 19920
aaaataattt tgataagaac ttcaatcttt gactagctta gtcatttttg agatttaatt 19980
aatattttat gtttattcat atataaacta ttcaaaatat tatagaattt aaacatttta 20040
acatcttaat cattgataaa taaccaaaaa tcaaagtatt acatcaataa ataactttta 20100
ctcaatgtca aagaaattat tatagggccc gtaactgctg ctggaattac acatggcatg 20160
gatgaattat acaaaggtac cgtcttctga ggtaccctta tctaaatgaa caacacttaa 20220
ataaatataa attaaagaac gtacttatat ttaaataaat ctcgaaattt tctaattact 20280
attattttta ttttttatac ttatgtcaaa caatacatag acatcatatg gattcacaac 20340
ataaagtatt ggagtaataa acactacgcg ccccagggcc gtcgttttca acgtcgtgac 20400
tgggcaaacc ctggcgttac ccaacttaat cgccttgcag cacatccccc tttcgccagc 20460
tggcgtaata gcgaagaggc ccgcaccgat cgcccttccc aacagttgcg cagcctgaat 20520
ggcgaatgga aattgtaagc gttaatatta tagaatttaa acattttaac atcttaatca 20580
ttcataaata actaaaaatc aaagtattac attcaataaa taacttttac tcaatgtcaa 20640
agaaattatt ggggttgggg gttggggttg gggttggggt tggggttggg gttggggttg 20700
gggttggggt tggggttggg gttggggttg gggttggggt tggggttggg gttggggttg 20760
gggttggggt tggggttggg gttggggttg gggttggggt tggggttggg gttggggttg 20820
gggttggggt tggggttggg gttggggttg gggttggggt tggggttggg gttggggttg 20880
gggttggggt tggggttggg gttggggttg gggttggggt gggaaaacag cattcaggta 20940
ttagaagaat atcctgattc aggtgaaaat attgttgatg cgcgggatcg gcctaggcgg 21000
ccgcgatcct cggggacacc aaatatggcg atctcggcct tttcgtttct tggagctggg 21060
acatgtttgc catcgatcca tctaccacca gaacggccgt tagatctgct gccaccgttg 21120
tttccaccga agaaccaccg ttgccgtaac caccacgacg gttgttgcta aagaagctgc 21180
caccgccacg gccaccgttg tagccgccgt tgttgttatt gtagttgctc atgttatttc 21240
tggcacttct tggttttcct cttaagtgag gaggaacata accattctcg ttgttgtcgt 21300
tgatgcttaa at 21312

CLAIMS

1. A *Tetrahymena thermophila* artificial chromosome having SEQ ID NO: 4, comprising a modified HSP70.2 promoter having SEQ ID NO: 1, 3´NTS sequence having SEQ ID NO: 2, "5´NTS containing C3 origin, rDNA genes and 3´NTS" sequence having SEQ ID NO: 3.

2. A *Tetrahymena thermophila* artificial chromosome according to claim 1, further comprising pUC19 vector backbone.

3. A *Tetrahymena thermophila* artificial chromosome according to claim 1, further comprising a telomere sequence.

4. A *Tetrahymena thermophila* artificial chromosome according to claim 1, further comprising Neo4 antibiotic resistance cassette.

5. A *Tetrahymena thermophila* artificial chromosome according to claim 1, further comprising TtsfGFP marker gene cassette or a DNA sequence encoding homologous and/or heterologous genomic DNA, RNA and/or proteins and/or derivatives thereof.

6. A *Tetrahymena thermophila* artificial chromosome according to claim 1 to claim 5, comprising "5´NTS containing C3 origin, rDNA genes and 3´NTS" sequence (SEQ ID NO: 3) on one arm, and 3´NTS sequence (SEQ ID NO: 2), Neo4 cassette and TtsfGFP marker gene cassette on the other arm.

7. A *Tetrahymena thermophila* artificial chromosome according to claim 6, wherein direction of transcription on both arms is configured from the center to the telomeres.

8. A *Tetrahymena thermophila* artificial chromosome according to any one of the preceding claims, wherein said *Tetrahymena thermophila* artificial chromosome has a nucleotide sequence that is at least 80% homologous to SEQ ID NO: 4.

9. A *Tetrahymena thermophila* artificial chromosome according to any one of the preceding claims, wherein said *Tetrahymena thermophila* artificial chromosome is in a circular or in a linear form.

10. A *Tetrahymena thermophila* artificial chromosome according to claim 9, wherein the conservation of the telomere sequences in the linear form is provided by 3´NTS sequences.

11. A method for producing a *Tetrahymena thermophila* artificial chromosome as defined in claim 1, comprising the steps of:

- introducing the vector backbone for the *Tetrahymena thermophila* artificial chromosome,
- inserting a modified HSP70.2 promoter (SEQ ID NO: 1), a telomere sequence, a 3´NTS sequence (SEQ ID NO: 2), a Neo4 cassette, and a "5´NTS containing C3 origin, rDNA genes and 3´NTS" sequence (SEQ ID NO: 3) to the vector backbone by biomimetics of the rDNA minichromosome, and
- obtaining the artificial chromosome.

12. A method according to claim 11, wherein said method further comprises the step of also inserting the TtsfGFP marker gene cassette or a DNA sequence encoding homologous and/or heterologous genomic DNA, RNA and/or proteins and/or derivatives thereof into the vector backbone.

13. A method according to claim 11, wherein said vector backbone is pUC19, thereby the artificial chromosome has a nucleotide sequence that is at least 80% homologous to SEQ ID NO: 4.

14. A method for recombinantly producing homologous and/or heterologous genomic DNA, RNA and/or proteins and/or derivatives thereof by using the *Tetrahymena thermophila* artificial chromosome as defined in claim 1 as a production system/platform, comprising the steps of:

- introducing a ciliate host cell of *Tetrahymena thermophila* or Tetrahymenidae family member that contains a recessive allele other than the C3 origin allele on the rDNA minichromosome of the macronucleus,
- extrachromosomally integrating the circular and/or linear *Tetrahymena thermophila* artificial chromosome (TtAC2) containing the dominant C3 origin in the macronucleus rDNA minichromosome into a ciliate host cell of *Tetrahymena thermophila* or Tetrahymenidae family member containing a recessed allele other than the C3 origin allele, and
- producing DNA, RNA and/or protein and/or derivatives thereof encoded by the DNA, RNA and/or protein encoded by a homologous and/or heterologous sequence carried on the *Tetrahymena thermophila* artificial chromosome by *Tetrahymena thermophila* or by a ciliate host cell member of the Tetrahymenidae family.

15. A method according to claim 14, wherein the process of extrachromosomally integrating the *Tetrahymena thermophila* artificial chromosome (TtAC2) into a ciliate host cell of *Tetrahymena thermophila* or Tetrahymenidae family member containing a recessed allele other than the C3 origin allele in step two is performed by biolistic gun method.

16. A method according to claim 14, wherein said *Tetrahymena thermophila* or Tetrahymenidae family member ciliate host cell is a cultured vegetative ciliate cell or two conjugating ciliate cells.

17. A transformant Tetrahymenidae family member ciliate host cell that is formed by transformation of TtAC2 bearing at least one homologous and/or heterologous sequence according to the method defined in claim 14.

18. Use of the *Tetrahymena thermophila* artificial chromosome as defined in claim 1 as a production system/platform for recombinantly producing homologous and/or heterologous genomic DNA, RNA and/or proteins and/or derivatives thereof.

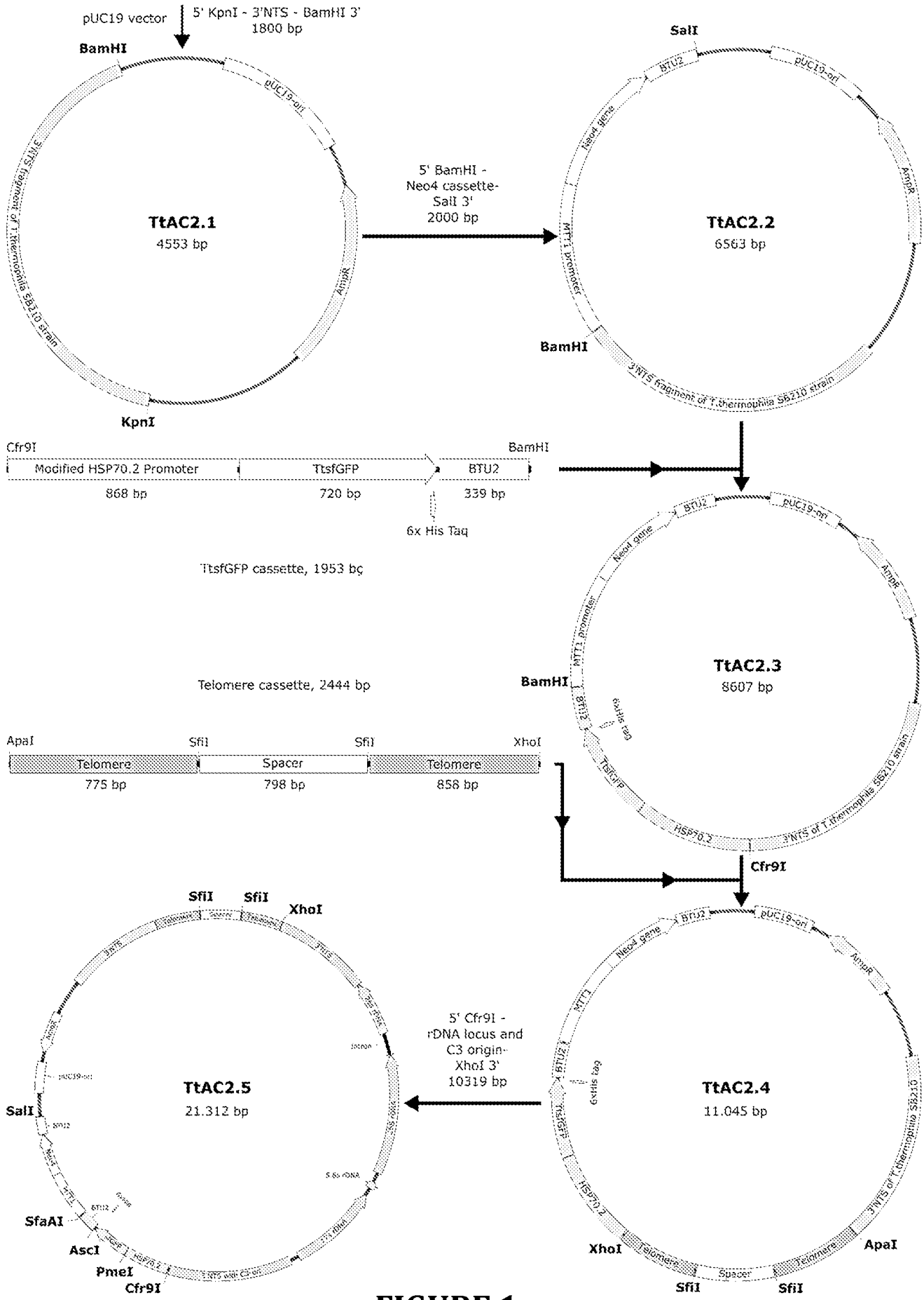
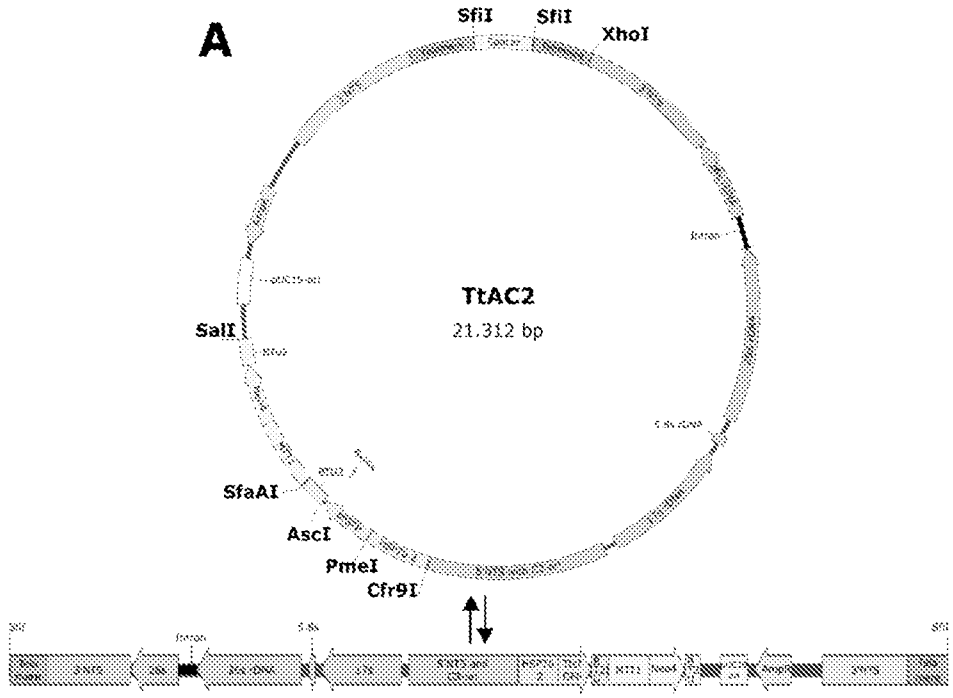
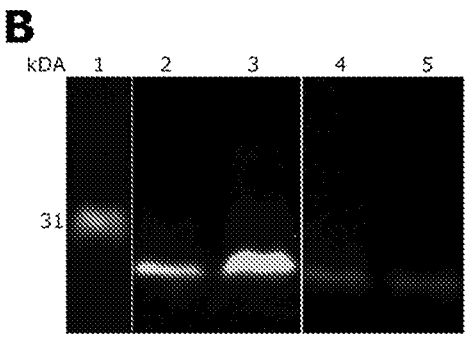
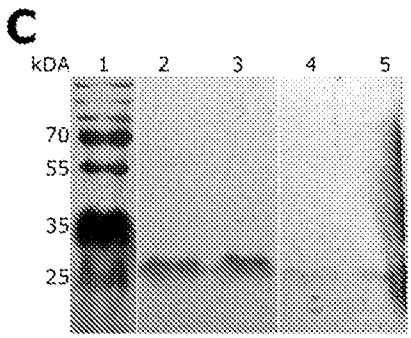
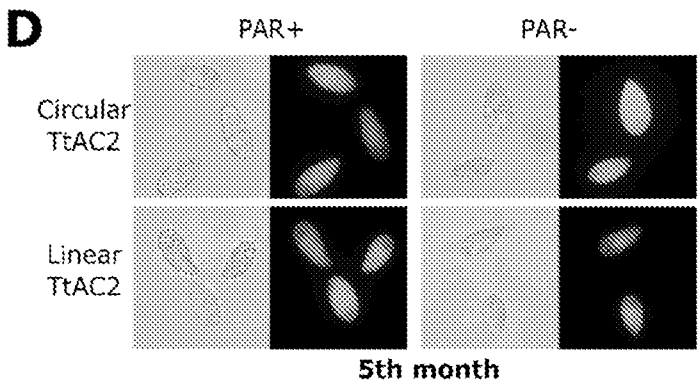

pUC19 vector
5' KpnI - 3'NTS - BamHI 3'
1800 bp
BamHI
KpnI
TtAC2.1
4553 bp
5' BamHI - Neo4 cassette- SalI 3'
2000 bp
SalI
BamHI
TtAC2.2
6563 bp
Cfr9I
BamHI
Modified HSP70.2 Promoter
TtsfGFP
BTU2
868 bp
720 bp
339 bp
6x His Taq
TtsfGFP cassette, 1953 bç
Telomere cassette, 2444 bp
ApaI
SfiI
SfiI
XhoI
Telomere
Spacer
Telomere
775 bp
798 bp
858 bp
BamHI
TtAC2.3
8607 bp
Cfr9I
SfiI
SfiI
XhoI
5' Cfr9I - rDNA locus and C3 origin- XhoI 3'
10319 bp
TtAC2.5
21.312 bp
SalI
SfaAI
AscI
PmeI
Cfr9I
TtAC2.4
11.045 bp
XhoI
ApaI
SfiI
SfiI

FIGURE 1


A
SfiI
SfiI
XhoI
TtAC2
21.312 bp
SalI
SfaAI
AscI
PmeI
Cfr9I


B
kDA 1 2 3 4 5
31


C
kDA 1 2 3 4 5
70
55
35
25


D
PAR+
PAR-
Circular TtAC2
Linear TtAC2
5th month

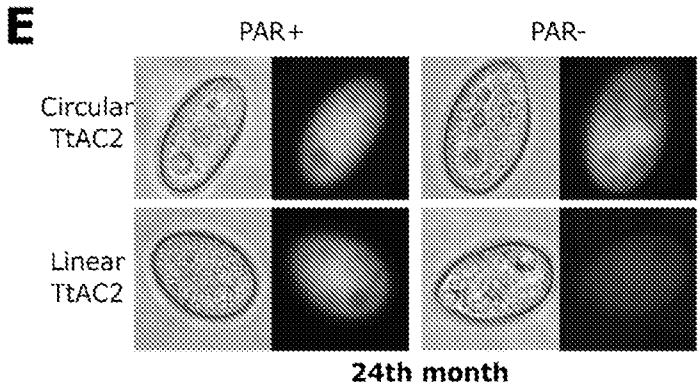

E
PAR+
PAR-
Circular TtAC2
Linear TtAC2
24th month

FIGURE 2

INTERNATIONAL SEARCH REPORT

International application No
PCT/TR2019/051165

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N15/63 C12N15/00 C07K14/44
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N C40B C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, BIOSIS, CHEM ABS Data, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2010/108183 A1 (TETRAGENETICS INC; CLARK THEODORE G [US] ET AL.) 23 September 2010 (2010-09-23)	17
Y	paragraphs [0068], [0101], [0132]; claims; sequence 1	1-18

X	ENGBERG J ET AL: "COMPLETE SEQUENCE OF THE EXTRACHROMOSOMAL RECOMBINANT DNA MOLECULE FROM THE CILIATE TETRAHYMENA-THERMOPHILA STRAIN B1868VII", NUCLEIC ACIDS RESEARCH, vol. 18, no. 23, 1990, pages 6915-6920, XP009521488, ISSN: 0305-1048 cited in the application	17
Y	Abstract; figures 1,2	1-18

	-/--	

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
14 July 2020	24/07/2020

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Burkhardt, Peter

INTERNATIONAL SEARCH REPORT

International application No
PCT/TR2019/051165

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 101 586 119 A (INST OF AQUATIC LIFE CHINESE A [CN]) 25 November 2009 (2009-11-25)	17
Y	claims; sequence 1	1-18

X	AZUSA HAYASHI ET AL: "Targeted Gene Disruption by Ectopic Induction of DNA Elimination in Tetrahymena", GENETICS, vol. 201, no. 1, 12 September 2015 (2015-09-12), pages 55-64, XP055712132, DOI: https://doi.org/10.1534/genetics.115.178525	17
Y	figures	1-18

X	TING YU ET AL: "A Novel Robust Heat-inducible Promoter for Heterologous Gene Expression in", PROTIST, FISCHER, JENA, DE, vol. 163, no. 2, 28 June 2011 (2011-06-28), pages 284-295, XP028452581, ISSN: 1434-4610, DOI: 10.1016/J.PROTIS.2011.07.003 [retrieved on 2011-07-11]	17
Y	Abstract; figure 3; table 2	1-18

X	G?RKAN YILMAZ ET AL: "Efficient expression of codon-adapted affinity tagged super folder green fluorescent protein for synchronous protein localization and affinity purification studies in Tetrahymena thermophila", BMC BIOTECHNOLOGY, BIOMED CENTRAL LTD. LONDON, GB, vol. 15, no. 1, 25 March 2015 (2015-03-25), page 22, XP021221189, ISSN: 1472-6750, DOI: 10.1186/S12896-015-0137-9	17
Y	Abstract; figures 1,2	1-18

T	AYÇA FULYA ÜSTÜNTANIR DEDE ET AL: "Construction and dynamic characterization of a Tetrahymena thermophila macronuclear artificial chromosome", GENE., vol. 748, 1 July 2020 (2020-07-01), page 144697, XP055712280, NL ISSN: 0378-1119, DOI: 10.1016/j.gene.2020.144697 the whole document	1-18

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/TR2019/051165

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2010108183 A1	23-09-2010	CA 2755978 A1	23-09-2010
		EP 2408911 A1	25-01-2012
		US 2012129217 A1	24-05-2012
		US 2015240248 A1	27-08-2015
		WO 2010108183 A1	23-09-2010
CN 101586119 A	25-11-2009	NONE	