Elsevier Editorial System(tm) for Marine Genomics
Manuscript Draft

Manuscript Number: MARGEN-D-09-00005R1

Title: Characterization of the pheromone gene family of an Antarctic and Arctic protozoan ciliate, Euplotes nobilii

Article Type: Full Length Article

Keywords: Psychrophilic microorganisms; gene-sized DNA molecules; water-soluble signal proteins; sub-chromosomic ciliate genomes.

Corresponding Author: Dr Adriana Vallesi,

Corresponding Author's Institution: University of Camerino

First Author: Adriana Vallesi

Order of Authors: Adriana Vallesi; Claudio Alimenti; Antonietta La Terza ; Graziano Di Giuseppe; Fernando Dini; Pierangelo Luporini

Abstract: Allelic genes encoding water-borne signal proteins (pheromones) were amplified and sequenced from the somatic (macronuclear) sub-chromosomic genome of Antarctic and Arctic strains of the marine ciliate, Euplotes nobilii. Their open reading frames appeared to be specific for polypeptide sequences of 83 to 94 amino acids identifiable with cytoplasmic pheromone precursors (pre-pro-pheromones), requiring two proteolytic steps to remove the pre- and pro-segments and secrete the mature pheromones. Differently from most of the macronuclear genes that have so far been characterized from Euplotes and other hypotrich ciliates, the 5' and 3' non-coding regions of all the seven E. nobilii pheromone genes are much longer than the coding regions (621 to 700 versus 214 to 285 nucleotides), and the 5' regions in particular show nearly identical sequences across the whole set of pheromone genes. These structural peculiarities of the non-coding regions are likely due to the presence of intron sequences and provide presumptive evidence that they are

site of basic, conserved activities in the mechanism that regulates the expression of the E. nobilii pheromone genes.

Characterization of the pheromone gene family of an Antarctic and Arctic protozoan ciliate, *Euplotes nobilii*

Keywords: psychrophilic microorganisms, gene-sized DNA molecules, water-soluble signal proteins, sub-chromosomic ciliate genomes.

Adriana Vallesi[a*], Claudio Alimenti[a], Antonietta La Terza[a], Graziano Di Giuseppe[b], Fernando Dini[b], Pierangelo Luporini[a]

[a]*Dipartimento di Biologia Molecolare Cellulare e Animale, University of Camerino, 62032 Camerino (MC), Italy*

[b]*Dipartimento di Biologia, Università di Pisa, Via Volta 4, 56126 Pisa, Italy*

*Corresponding author: Tel.: +39-0737403256; fax: +39-0737403290.

E-mail address: adriana.vallesi@unicam.it

Abstract

Allelic genes encoding water-borne signal proteins (pheromones) were amplified and sequenced from the somatic (macronuclear) sub-chromosomic genome of Antarctic and Arctic strains of the marine ciliate, *Euplotes nobilii*. Their open reading frames appeared to be specific for polypeptide sequences of 83 to 94 amino acids identifiable with cytoplasmic pheromone precursors (pre-pro-pheromones), requiring two proteolytic steps to remove the pre- and pro-segments and secrete the mature pheromones. Differently from most of the macronuclear genes that have so far been characterized from *Euplotes* and other hypotrich ciliates, the 5' and 3' non-coding regions of all the seven *E. nobilii* pheromone genes are much longer than the coding regions (621 to 700 versus 214 to 285 nucleotides), and the 5' regions in particular show nearly identical sequences across the whole set of pheromone genes. These structural peculiarities of the non-coding regions are likely due to the presence of intron sequences and provide presumptive evidence that they are site of basic, conserved activities in the mechanism that regulates the expression of the *E. nobilii* pheromone genes.

Introduction

Life thrives also in the coolest environments that abound on the surface of our planet. Among the cold-loving (psychrophilic) organisms that appear to be experimentally more promising to investigate the genetic and molecular basis of this biological success, protozoan ciliates deserve much more attention than they received so far. In them we find an unique combination of ecological and biological properties. In addition to representing a major microbial component of every aquatic system also in the polar regions [1,2], they are facultative sexual organisms. Their life cycles in fact involve an optional alternation between periods of vegetative growth (mitotic multiplication) associated with availability of energetic resources from the environment, and sexual events associated with famine and other stressing conditions [3]. Like most fungi and other facultative sexual organisms [4,5], ciliates are thus presumably able to modulate (accelerate?) their adaptive evolution to extremely varied environmental conditions more efficiently than asexual, or obligate sexual microorganisms.

The ciliate of interest in this context is *Euplotes nobilii*. Of this species, we are studying Antarctic and Arctic populations (represented in laboratory by wild-type strains) in relation to the biology of the genetic mechanism of the mating types on which cells rely to shift between the growth and sexual stages of their life cycle. As in other species of *Euplotes* [6], these mating types appear to be under the control of a series of alleles co-dominant at the Mendelian locus *mat* of the chromosomic genome (transcriptionally silent) of germ-line nucleus (micronucleus) [7, and unpublished results], and become manifest through the activity of specific and diffusible signal proteins usually referred to as “pheromones”. Cells grow in response to the autocrine (homologous) binding of their secreted (self) pheromones, and temporarily arrest their growth and unite into mating pairs in response to the paracrine (heterologous) binding of foreign (non-self) pheromones [8,9].

Previously we determined the conformations of a set of these *E. nobilii* pheromones by analysis of nuclear magnetic resonance of native protein preparations, and showed the relationships

of structural homology that make these molecules able to bind their target cell receptors in competition with one another [10,11]. Now we have determined the full-length sequences of the functional genes that, in form of individual and multiple "gene-sized DNA molecules" [12, as a review], develop from the chromosomic locus *mat* of the micronucleus and determine the synthesis of these pheromones in the transcriptionally active genome of the somatic nucleus (macronucleus). In this work, we present a comparative structural analysis of a complex of seven *E. nobilii* macronuclear pheromone genes. Three genes (designated *en-ant1*, *en-ant2* and *en-ant6*) encode the pheromones E*n*-1, E*n*-2, and E*n*-6 of Antarctic origin, that have already been isolated from preparations of culture supernatant and structurally characterized [10,11]; the four other genes (designated *en-arc1*, *en-arc2*, *en-arc3*, and *en-arc4*) specify the pheromones E*n*-A1, E*n*-A2, E*n*-A3, and E*n*-A4 of Arctic origin, that have not yet been purified and analyzed chemically.

2. Materials and methods

2.1 Strain origin and DNA preparation

Each one of the four *E. nobilii* strains (AC-1, 5QAA15, 2QAN1 and 4Pyrm4) used in this work was expanded starting from a single individual isolated from samples of sediment and seawater. Strain AC-1 was collected from the surroundings of the Italian Research Station "Mario Zucchelli" (Terra Nova Bay, Weddel Sea, Antarctica); strains 5QAA15 and 2QAN1 from two different sites of the Qaanaaq-Thule area in Western Greenland (Nares Strait, Baffin Bay, Arctic Glacial Sea); strain 4Pyrm4 from the surroundings of Pyramiden (Spitzbergen Island, Svalbard Archipelago, Eastern Greenland Sea). All the strains were grown at 4-7 °C, with the green alga *Dunaliella tertiolecta* as food.

DNA preparations were obtained by the guanidine hydrochloride extraction method. Cell pellets were lysed by incubation in 4 volumes of 8 M guanidine hydrochloride for at least 1h at room temperature, and equal volumes of 0.5 M EDTA pH 8.0 were then added to these mixtures. DNA was precipitated with two volumes of cooled absolute ethanol, washed twice with 70%

ethanol, suspended in distilled water, and incubated with RNase A (50 μg ml^{-1}), at 37 °C for 1h. Finally, DNA was extracted with phenol/chloroform and precipitated with an equal volume of 20% PEG containing 2.5 M NaCl, at room temperature for 1h.

2.2 DNA amplification by Polymerase Chain Reaction (PCR) and molecular cloning

The two oligonucleotides (5L-FW and 3T-RV) used as primers in all the DNA amplifications were synthesized by Sigma-Aldrich. Their sequences are the following: 5'-ATGGCACCTATTTCCGATTGC-3' (5L-FW); 5'-GTGAATGGTAGGAGATGTTGA-3' (3T-RV). Aliquots (0.5 μg) of DNA preparations were added to 50 μl-reaction mixtures containing 0.5 μM of each primer, 0.25 mM dNTP, 1.5 mM $MgCl_2$, and 1 U of Platinum high-fidelity Taq polymerase (Invitrogen), and amplified following a standard program (30s 94°C, 40s 58°C, 60s 72°C for 35 cycles). All PCR amplifications were run in a GeneAmp 9700 thermal cycler (PE Applied Biosystems). Amplified products were cloned into the pCR 2.1-TOPO vector of the TOPO TA Cloning Kit following the procedures suggested by the supplier (Invitrogen), and not less than five clones were analyzed for each gene sequence determination. Reactions were carried out with the ABIPrism sequence analyzer, Model 373A, using the Big Dye Terminator methodology (PE Applied Biosystems).

3. Results and discussion

3.1 Pheromone gene identification and cloning

Of the complex of seven pheromone gene sequences of *E. nobilii* analyzed in this work, six sequences (i. e., *en-ant1*, *en-ant2*, *en-arc1*, *en-arc2*, *en-arc3*, and *en-arc4*) represent new determinations. All of them were obtained by a PCR strategy based on: (i) the knowledge of the sequence of the pheromone gene *en-ant6* that, as reported elsewhere [13], was cloned earlier using a genetic approach developed in relation to the structural characterization of the pheromone E*n*-6 specified by this gene [10]; (ii) the assumption that the sequences of the *E. nobilii* pheromone

genes, like those of the *E. octocarinatus* pheromone genes [14,15], are closely similar at level of their sub-telomeric regions. Thus, in practice, every new *E. nobilii* pheromone gene was cloned using as PCR primers the same two oligonucleotides (5L-FW and 3T-RV, see Material & Methods) equivalent to the sub-telomeric sequences of the known pheromone gene *en-ant6*.

After electrophoresis on agarose gels, all the products of PCR amplification of every DNA preparation were revealed as single bands of dimensions varying from 900 to 1000 bp (data not shown). They were cloned and their full-length nucleotide sequences were determined. Two distinct gene sequences were obtained in relation to DNA amplified from each one of the strains AC-1 and 5QAA15, (i. e., the sequences *en-ant1*and *en-ant2* from strain AC-1, and the sequences *en-arc2* and *en-arc3* from strain 5QAA15), thus implying that these strains were heterozygous at their micronuclear locus *mat* and consistently carried the respective macronuclear gene-sized versions of the two *mat* alleles. Instead, a unique sequence was determined in relation to DNA amplified from the strains 4Pyrm4 and 2QAN1 (i. e., the sequence *en-arc1* from strain 4Pyrm4 and the sequence *en-arc4* from strain 2QAN1), thus implying that these strains were *mat*-homozygous and consistently carried only the macronuclear gene-sized version of this allele, or alternatively, if *mat*-heterozygous, they carried the macronuclear gene-sized versions of the two *mat* alleles in a very eccentric ratio.

3.2 Nucleotide sequences

The seven pheromone genes that have been cloned extend from 900 (*en-ant1*) to 961 bp (*en-arc4*), the two telomeric $(C_4A_4)_3CCCC$ motifs excluded (Fig. 1). Their sequences include multiple ATG start codons and multiple TAA and TAG stop codons, which (it is worth recalling) some ciliates may use unconventionally to specify glutamine or glutamic acid so as, *Euplotes* in particular, may use TGA to specify cysteine instead of signaling stop to translation [16-19]. Nevertheless, the open reading frame (ORF) that spans from the ATG codon uniformly carried in position 362 (361 in *en-ant6*) to the TAA codon lying between the positions 613 (*en-ant1*) to 645

(*en-ant6*) is clearly the best candidate to be expressed. The amino acid sequences of 83 to 94 residues predicted by this ORF all appear identifiable with the pheromone cytoplasmic precursors. They show the canonical functional subdivision into a hydrophobic signal (or pre) peptide, a pro segment, and the region of the "mature" (secreted) protein. In the case of the *en-ant1*, *en-ant2* and *en-ant6* genes, the predicted mature sequences fully match the sequences of the E*n*-1, E*n*-2 and E*n*-6 pheromones that have previously been determined by chemical analysis of the native proteins[10,11,20,21]. This matching thus strongly corroborates the identification of the ATG codon in position 362 with the effective site of initiation of transcription.

Two indications emerge from the comparative analysis of the seven *E. nobilii* pheromone gene sequences that have been determined (Fig. 1), and the association of these sequences into a phylogenetic tree (Fig. 2). The first one is that among the three genes of Antarctic origin there is a degree of sequence identity (80 to 84%) higher than among the four genes of Arctic origin (65 to 80%), and that the extent of sequence divergence among the Arctic genes is closely similar to the extent revealed by the comparisons between the Antarctic and Arctic genes (64 to 72%). Overall, these different extents of sequence variations likely reflect the fact that the Antarctic genes have been cloned from cells of the same local population, while the Arctic ones derive from cells of two distinct populations whose wide geographic separation (one inhabits the Svalbard coasts and one inhabits the Western Greenland coasts) has most likely imposed also a separation (at least partial) of genetic nature.

The second indication is that all the seven *E. nobilii* pheromone gene sequences have maximal levels of identity in coincidence with their 5' and 3' non-coding regions, and between these two regions the former is more uniformly conserved (88 to 97% of identity) than the latter (43 to 92% of identity). They bear the motif GAAAA (41 bp upstream the starting ATG codon) as presumptive site of initiation of transcription in the place of the more conventional TATA (or CAAT) boxes [22] and, as occurs in other ciliates[23-25], lack a polyadenylation signal represented

by the canonical AATAAA motif. Good candidates to replace this motif are the sequences TTATTT and AATAA/G, that are shared by all the sequences in equivalent positions.

This tight structural conservation of the non-coding regions implies that they are site of specific and strictly conserved functions in the mechanism of expression of the pheromone genes. As is the case in pheromone genes of other species of *Euplotes* [14,26,27], these functions are presumably correlated with the inclusion of intron sequences the splicing of which results in the expression of new ORF's and the synthesis of new and functionally diversified pheromone isoforms. This possibility is supported by the common conservation in the 5' regions of, (i) multiple ATG start codons, (ii) A + T rich sequences, and (iii) consensus donor GT (5') and acceptor AG (3') splice-site junctions [28,29].

3.3 Predicted aminoacid sequences

The polypeptide precursors that are predicted by the ORF's identified in the pheromone gene sequences (Fig. 1) vary in length from 83 (*en-ant1*) to 94 (*en-ant6*) amino acid residues. As shown by the multiple sequence alignment of these precursors (Fig. 3), they all conform with the same organization based on three functionally distinct domains: (i) a 19-amino acid signal peptide identified by an initial positively charged motif followed by a hydrophobic middle region and a terminal polar sequence, (ii) a 12-amino acid pro segment rich in basic residues, and (iii) the domain of the mature protein spanning from 52 (E*n*-1) to 63 (E*n*-6) amino acids in length.

As is the case in *E. raikovi* and *E. octocarinatus* [15,30,31], also these *E. nobilii* pheromone precursors are clearly characterized by a gradient of sequence identity, that drastically drops passing from the signal peptide and pro segment to the secreted protein. There are only six variable positions over the 31 ones that overall form the signal peptide and pro segment (Val/Met and Ser/Thr in the former; Arg/Lys, Lys/Arg/Asn, Lys/Gln, and Glu/Xxx in the latter). On the other hand, there are only eight positions in the secreted proteins (over a total of 52 to 63), in which a cysteine residue is fully conserved in specific relation with a stabilizing function of the pheromone

secondary structures, all of which are characterized by a three-helix bundle core, as described elsewhere [10,11].

To process these pheromone precursors and release the mature protein into the medium, two proteolytic steps are required, i. e., the co-translational removal of the signal peptide and the subsequent post-translational cleavage of the pro segment. The putative Ala-Phe cleavage site of the signal peptide appears to be identical in all the precursor molecules, while the Ser-Ser/Thr-Ala sequence preceding the cleavage site agrees well with the consensus sequence for known signal peptidases [32,33]. In contrast, the putative cleavage site of the pro segment for producing the mature protein appears to vary in the different gene sequences. It would be represented by Gly-Asn/Asp, Ala-Tyr, Ser-Thr, or Glu-Thr/Asp combinations. Thus, while the signal peptide would be processed by a single endopeptidase common to every *E. nobilii* cell type, the removal of the pro segment for the pheromone secretion would require either multiple cell type-specific enzymes, or enzymes with multiple specificities.

4. Conclusions

Cloning and full-length sequencing of some dozens of genes from different hypotrich ciliates have strongly supported the concept that the typical organization of the macronuclear gene-sized DNA molecules includes coding regions that are markedly more extended (500 to 20,000 bp) than their flanking 5' and 3' non-coding regions (usually, less than 200 bp) [34,35]. Further credit to this concept now derives also from results of genome sequencing in *Oxytricha trifallax* [29,36] and *Nyctotherus ovalis* [25].

The observations described in this study thus raise an odd case. The *E. nobilii* pheromone genes appear to be all formed by 5' and 3' non-coding regions that extend for nearly three-fold the length of the coding regions and, in addition, show a tight sequence conservation throughout the entire gene family. This peculiar organization of the *E. nobilii* pheromone genes probably reflects a conserved inclusion of intron sequences, whose splicing might have a fundamental role in the

mechanism of expression of these genes. This hypothesis receives support in particular from previous functional analysis of *E. raikovi* pheromone genes [26,27], showing that the primary transcripts of these genes undergo intron splicing (at not canonical sites) to generate membrane-bound pheromone isoforms that cells use as autocrine pheromone receptors [9]. Its assessment, in addition to being crucial to elucidate the mechanism of expression of the *E. nobilii* pheromone gene family, may also contribute insightful information on the more general problem of the functional organization of the sub-chromosomic, gene-sized macronuclear genomes of the hypotrich ciliates.

Acknowledgements: Financial support to this study was provided by the "Programma Nazionale di Ricerche in Antartide" (PNRA) and by the Italian Ministero dell'Università e della Ricerca**.**

References

[1] W. Petz, Ciliates, in: F.J. Scott, H.J. Marchant (Eds.), Antarctic marine protists, Australian Biological Resources Study, Canberra, 2005, pp. 347-448.

[2] W. Petz, A. Valbonesi, U. Schiftner, A. Quesada, J. Cynan Ellis-Evans, Ciliate biogeography in Antarctic and Arctic freshwater ecosystems: endemism or global distribution of species?, FEMS Microbiol. Ecol. 59 (2007) 396-408.

[3] D.L. Nanney, Experimental ciliatology, Wiley, New York, 1980.

[4] M.R. Goddard, Why bother with sex? Answers from experiments with yeast and other organisms, in: J. Heitman, J. Kronstad, J. Taylor, L. Casselton (Eds.), Sex in fungi: molecular determination and evolutionary implications, American Society for Microbiology Press, 2007, pp. 489-506.

[5] M.R. Goddard, H.J.C. Godfray, A. Burt, Sex increases the efficiency of natural selection in experimental yeast populations, Nature 434 (2005) 636-640.

[6] F. Dini, D. Nyberg, Sex in ciliates, in: J.G. Jones (Ed.), Advances in microbial ecology, Plenum press, New York, 13, 1993, pp. 85-153.

[7] A. Felici, C. Alimenti, C. Ortenzi, P. Luporini, Purification and initial characterization of two pheromones from the marine Antarctic ciliate, *Euplotes nobilii,* Italian J. Zool. 66 (1999) 355-360.

[8] A. Vallesi, G. Giuli, R.A. Bradshaw, P. Luporini, Autocrine mitogenic activity of pheromones produced by the protozoan ciliate *Euplotes raikovi*, Nature 4 (1995) 1221-1227.

[9] A. Vallesi, P. Ballarini, B. Di Pretoro, C. Alimenti, C. Miceli, P. Luporini, The autocrine, mitogenic pheromone-receptor loop of the ciliate *Euplotes raikovi*: pheromone-induced receptor internalization, Eukaryot. Cell 4 (2005) 1221-1227.

[10] B. Pedrini, W.J. Placzek, E. Koculi, C. Alimenti, A. La Terza, P. Luporini, K. Wüthrich, Cold-adaptation in sea-water-borne signal proteins: sequence and NMR structure of the pheromone E*n*-6 from the Antarctic ciliate *Euplotes nobilii*, J. Mol. Biol. 372 (2007) 277-286.

[11] W.J. Placzek, T. Etezady-Esfarjani, T. Herrmann, B. Pedrini, W. Peti, C. Alimenti, P. Luporini, K. Wüthrich, Cold-adapted signal proteins: NMR structures of pheromones from the Antarctic ciliate *Euplotes nobilii*, IUBMB Life 59 (2007) 578-585.

[12] C.L. Jahn, L.A. Klobutcher, Genome remodelling in ciliated protozoa, Ann. Rev. Microbiol. 56 (2002) 489-520.

[13] A. La Terza, N. Dobri, C. Alimenti, A. Vallesi, P. Luporini, The water-borne protein signals (pheromones) of the Antarctic ciliated protozoan, *Euplotes nobilii*: structure of the gene coding for the E*n*-6 pheromone, Can. J. Microbiol. 55 (2009) 57-62.

[14] C. Brünen-Nieveler, J.C. Weiligmann, B. Hansen, H.W. Kuhlmann, M. Möllenbeck, K. Heckmann, The pheromones and pheromone genes of new stock of *E. octocarinatus* species complex, Eur. J. Protistol. 34 (1998) 124-132.

[15] M. Möllenbeck, K. Heckmann, Characterization of two genes encoding a fifth so far unknown pheromone of *Euplotes octocarinatus*, Europ. J. Protistol. 35 (1999) 225-230.

[16] F. Meyer, H.J. Schmidt, E. Plumber, A. Hasilik, G. Mersmann, H. Meyer, A. Engstrom, K. Heckmann, UGA is translated as cysteine in pheromone 3 of *Euplotes octocarinatus*, Proc. Natl. Acad. Sci. U.S.A 88 (1991) 3758-3761.

[17] R.D. Knight, S.J. Freeland, L.F. Landweber, Rewiring the keyboard: evolvability of the genetic code, Nat. Rev. Genet. 2 (2001) 49-58.

[18] R. Sánchez-Silva, E. Villalobo, L. Morin, A. Torres, A new noncanonical nuclear genetic code: translation of UAA into glutamate, Curr. Biol. 13 (2003) 442-447.

[19] A.A. Turanov, A.V. Lobanov, D.E. Fomenko, H.G. Morrison, M.L. Sogin, et al., Genetic code supports targeted insertion of two amino acids by one codon, Science 323 (2009) 259-261.

[20] C. Alimenti, C. Ortenzi, V. Carratore, P. Luporini, Structural characterization of a protein pheromone from a cold-adapted (Antarctic) single-cell eukaryote, the ciliate *Euplotes nobilii*, FEBS Lett. 514 (2002) 329-332.

[21] C. Alimenti, C. Ortenzi, V. Carratore, P. Luporini, Structural characterization of E*n*-1, a cold-adapted protein pheromone isolated from the Antarctic ciliate *Euplotes nobilii*, Biochim. Biophys. Acta. 1621 (2003) 17-21.

[22] S. Ghosh, J.W. Jarakzewsky, L.A. Klobutcher, C.L. Jahn, Characterization of transcription initiation, translation initiation, and poly(A) addition sites in the gene-sized macronuclear DNA molecules of *Euplotes*, Nucleic Acids Res. 22 (1994) 214-221.

[23] M. Tan, K. Heckmann, The two *γ*-tubulin-encoding genes of the ciliate *Euplotes crassus* differ in their sequences, codon usage, transcription initiation sites and poly(A) addition sites, Gene 210 (1998) 53-60.

[24] N.R. McEwan, S.C.P. Eschenlauer, R.E.R. Calza, J. Wallace, C. J. Newbold, The 3' untranslated region of messages in the rumen protozoan *Entodinium caudatum*, Protist 151 (2000) 139-146.

[25] G. Ricard, R.M. de Graaf, B.S. Dutilh, I. Duarte, T.A. van Alen, et al., Macronuclear genome structure of the ciliate *Nyctotherus ovalis*: single-gene chromosomes and tiny introns, BMC Genomics 9 (2008) 587-601.

[26] C. Miceli, A. La Terza, R.A. Bradshaw, P. Luporini, Identification and structural characterization of a cDNA clone encoding a membrane-bound form of the polypeptide pheromone E*r*-1 in the ciliate protozoan *Euplotes raikovi*, Proc. Natl. Acad. Sci. U.S.A. 89 (1992) 1988-1992.

[27] G. Di Giuseppe, C. Miceli, R. Zahn, F. Damberger, K. Wüthrich, P. Luporini, A structurally deviant member of the *Euplotes raikovi* pheromone family: E*r*-23, J. Eukaryot. Microbiol. 49 (2002) 86-92.

[28] D.M. Prescott, J.D. Prescott, R.M. Prescott, Coding properties of macronuclear DNA molecules in *Sterkiella nova* (*Oxytricha nova*), Protist 153 (2002) 71–77.

[29] A.R.O. Cavalcanti, N.A. Stover, L. Orecchia, T.G. Doak, L.F. Landweber, Coding properties of *Oxytricha trifallax* (*Sterkiella histriomuscorum*) macronuclear chromosomes: analysis of a pilot genome project, Chromosoma 113 (2004) 69-76.

[30] C. Miceli, A. La Terza, R.A. Bradshaw, P. Luporini, Structural characterization of mating pheromone precursors of the ciliate protozoan *Euplotes raikovi*, Europ. J. Biochem. 202 (1991) 759-764.

[31] P. Luporini, C. Alimenti, C. Ortenzi, A. Vallesi, Ciliate mating types and their specific protein pheromones, Acta Protozool. 44 (2005) 89-101.

[32] G. von Heijne, SIGPEP: a sequence database for secretory signal peptides, Protein Seq. Data Anal. 1 (1987) 41-42.

[33] H. Nielsen, J. Engelbrechtl, S. Brunak, G. von Heijne, Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites, Protein Eng. 10 (1997) 1-6.

[34] D.M. Prescott, The DNA of ciliated protozoa, Microbiol. Rev. 58 (1994) 233-267.

[35] W.J. Chang, V.M. Addis, A.J. Li, E. Axelsson, D.H. Ardell, L.F. Landweber, Intron evolution and information processing in the DNA polymerase α gene in spirotrichous ciliates: a hypothesis for interconversion between DNA and RNA deletion, Biol. Direct (2007) 2:6, doi:10.1186/1745-6150-2-6.

[36] A.R.O. Cavalcanti, D.M. Dunn, R. Weiss, G. Herrick, L.F. Landweber, T.G. Doak, Sequence features of *Oxytricha trifallax* (class Spirotrichea) macronuclear telomeric and subtelomeric sequences, Protist 155 (2004) 311–322.

[37] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Res. 22 (1994) 4673-4680.

[38] K. Tamura, M. Nei, Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees, Mol. Biol. Evol. 10 (1993) 512–526.

[39] N. Saitou, M. Nei, The neighbor-joining method: A new method for reconstructing phylogenetic trees, Mol. Biol. Evol. 4 (1987) 406-425.

[40] D.L. Swofford, PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods), Sinauer Associates, Sunderland, MA, 2002.

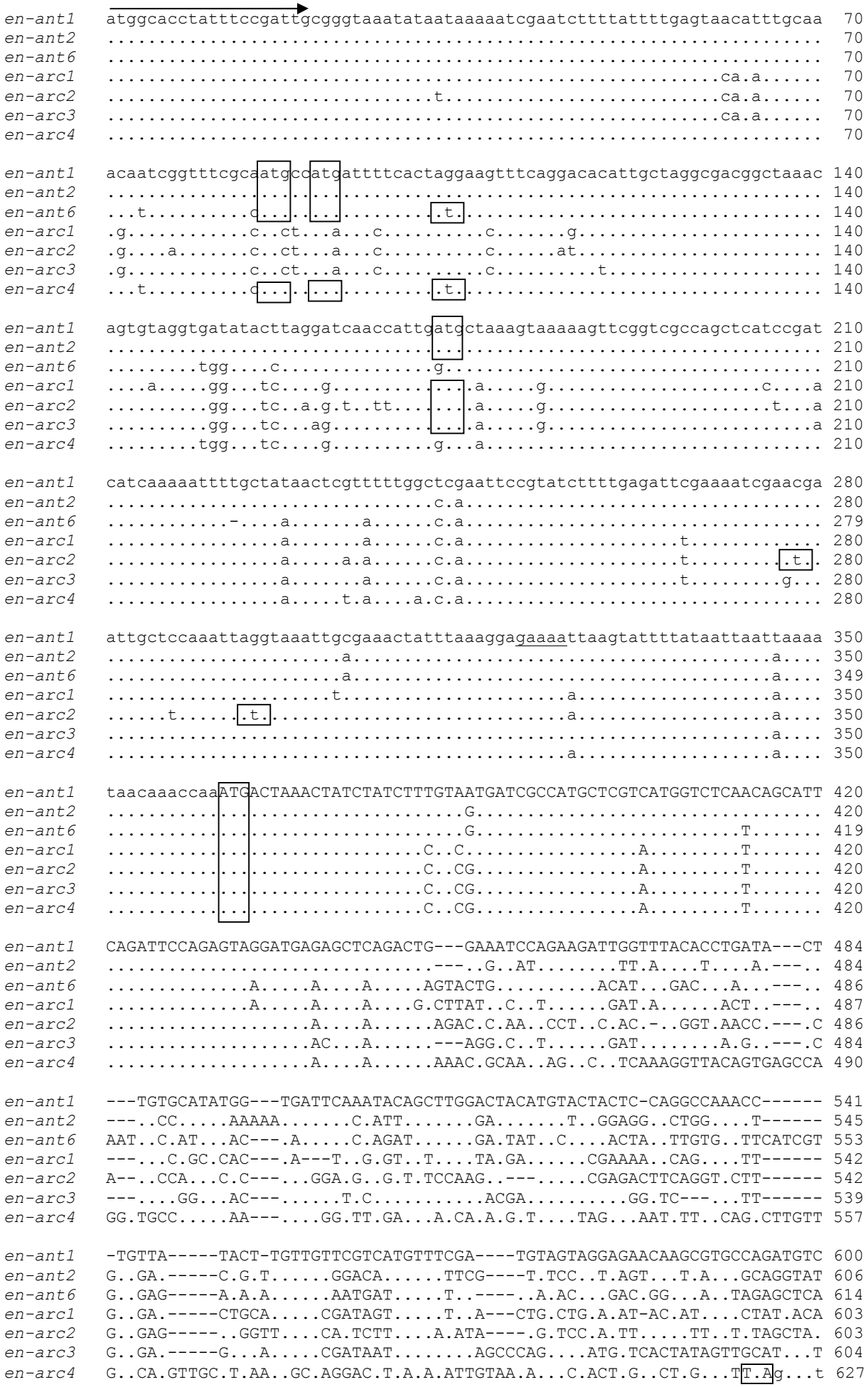
Figure legends

Fig. 1. Sequence alignment of *E. nobilii* pheromone genes, generated using Clustal W [37]. Gaps were inserted to maximize homologies, while dots stand for identical nucleotides. The ATG start codon and the TAA stop codon, that delimit the open reading frames (capitol letters) specific for the cytoplasmic pheromone precursors, are boxed. Other potential ATG start codons are also indicated in bold. Positions of oligonucleotides used as primers in PCR reactions are indicated by arrows. The sequence motif for the initiation of transcription in the 5' non-coding region is underlined. Nucleotide positions are numbered on the right. The accession numbers of the sequences deposited at the National Centre for Biotechnology Information are the following: *en-ant1*, **FJ645718**; *en-ant2*, **FJ645719**; *en-ant6*, **EF030059**; *en-arc1*, **FJ645720**; *en-arc2*, **FJ645721**; *en-arc3*, **FJ645722**; *en-arc4*, **FJ645723**.

Fig. 2. **Fig. 2**. Dendrogram of the phylogenetic relationships of the *E. nobilii* pheromone genes. The dendrogram was calculated using the Tamura and Nei model [38] for the genetic distance and the Neighbor-Joining method [39] in PAUP v4.0b10 [40]. Numbers at the nodes refer to the bootstrap support after 1000 resamplings of the dataset. The scale bar corresponds to 5 nucleotide substitutions per 100 positions.

Fig. 3. Multiple sequence alignment of *E. nobilii* pheromone precursors. The alignment is generated by Clustal W [37] and maximized by gap insertions. The pre, pro e secreted regions are indicated. Filled and light dots mark positions occupied by an identical amino acid residue, or two similar residues, respectively. The pre and pro signal peptidase sites are indicated by white arrows. The number of residues is reported at the right.

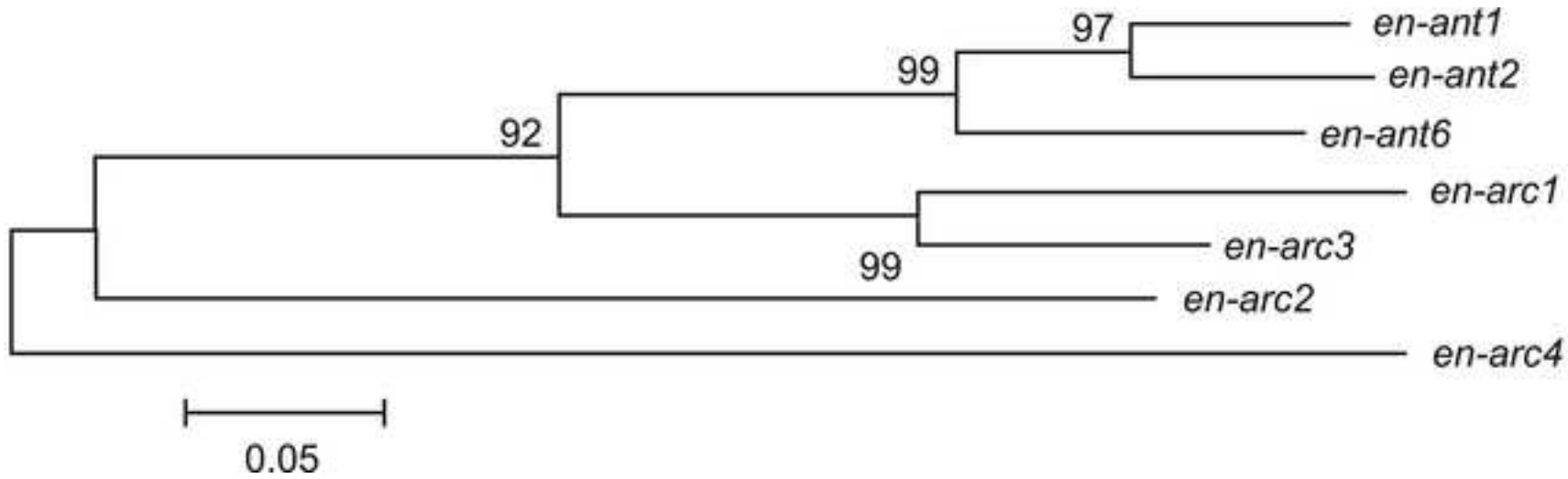
Figure


en-ant1 atggcacctatttccgattgcgggtaaatataataaaaatcgaatcttttattttgagtaacatttgcaa 70
en-ant2 70
en-ant6 70
en-arc1 .....ca.a..... 70
en-arc2 .....t.....ca.a..... 70
en-arc3 .....ca.a..... 70
en-arc4 70
en-ant1 acaatcggtttcgcaatgccatgattttcactaggaagtttcaggacacattgctaggcgacggctaaac 140
en-ant2 140
en-ant6 ...t.....c.....t..... 140
en-arc1 .g.....c..ct...a...c.....c.....g..... 140
en-arc2 .g...a.....c..ct...a...c.....c.....at..... 140
en-arc3 .g.....c..ct...a...c.....c.....t..... 140
en-arc4 ...t.....c.....t..... 140
en-ant1 agtgtaggtgatatacttaggatcaaccattgatgctaaagtaaaaagttcggtcgccagctcatccgat 210
en-ant2 210
en-ant6 .....tgg....c.....g..... 210
en-arc1 ....a.....gg...tc....g.....a.....g.....c....a 210
en-arc2 .....gg...tc..a.g.t..tt.....a.....g.....t...a 210
en-arc3 .....gg...tc...ag.....a.....g.....a 210
en-arc4 .....tgg...tc....g.....g...a..... 210
en-ant1 catcaaaaattttgctataactcgtttttggctcgaattccgtatcttttgagattcgaaaatcgaacga 280
en-ant2 .....c.a..... 280
en-ant6 .....-....a.....a.....c.a..... 279
en-arc1 .....a.....a.....c.a.....t..... 280
en-arc2 .....a.....a.a.....c.a.....t.....t. 280
en-arc3 .....a.....a.....c.a.....t.....g... 280
en-arc4 .....a.....t.a....a.c.a..... 280
en-ant1 attgctccaaattaggtaaattgcgaaactatttaaaggagaaaattaagtattttataattaattaaaa 350
en-ant2 .....a.....a.... 350
en-ant6 .....a.....a.... 349
en-arc1 .....t.....a.....a.... 350
en-arc2 .....t.....t.....a.....a.... 350
en-arc3 .....a.....a.... 350
en-arc4 .....a.....a.... 350
en-ant1 taacaaaccaaATGACTAAACTATCTATCTTTGTAATGATCGCCATGCTCGTCATGGTCTCAACAGCATT 420
en-ant2 .....G..... 420
en-ant6 .....G.....T..... 419
en-arc1 .....C..C.....A.....T..... 420
en-arc2 .....C..CG.....A.....T..... 420
en-arc3 .....C..CG.....A.....T..... 420
en-arc4 .....C..CG.....A.....T..... 420
en-ant1 CAGATTCCAGAGTAGGATGAGAGCTCAGACTG---GAAATCCAGAAGATTGGTTTACACCTGATA---CT 484
en-ant2 .....---..G..AT.....TT.A....T....A.---.. 484
en-ant6 .....A.....A....A.....AGTACTG.....ACAT...GAC...A...---.. 486
en-arc1 .....A.....A....A....G.CTTAT..C..T.....GAT.A.....ACT..---.. 487
en-arc2 .....A....A.....AGAC.C.AA..CCT..C.AC.-..GGT.AACC.---.C 486
en-arc3 .....AC...A.....---AGG.C..T.....GAT.....A.G..---.C 484
en-arc4 .....A....A.....AAAC.GCAA..AG..C..TCAAAGGTTACAGTGAGCCA 490
en-ant1 ---TGTGCATATGG---TGATTCAAATACAGCTTGGACTACATGTACTACTC-CAGGCCAAACC------ 541
en-ant2 ---..CC.....AAAAA.....C.ATT.....GA.....T..GGAGG..CTGG....T------ 545
en-ant6 AAT..C.AT...AC---.A.....C.AGAT.....GA.TAT..C....ACTA..TTGTG..TTCATCGT 553
en-arc1 ---...C.GC.CAC---.A---T..G.GT..T....TA.GA.....CGAAAA..CAG....TT------ 542
en-arc2 A--..CCA...C.C---...GGA.G..G.T.TCCAAG..---.....CGAGACTTCAGGT.CTT------ 542
en-arc3 ---....GG...AC---.....T.C.....ACGA.....GG.TC---...TT------ 539
en-arc4 GG.TGCC.....AA---....GG.TT.GA...A.CA.A.G.T....TAG...AAT.TT..CAG.CTTGTT 557
en-ant1 -TGTTA-----TACT-TGTTGTTCGTCATGTTTCGA----TGTAGTAGGAGAACAAGCGTGCCAGATGTC 600
en-ant2 G..GA.-----C.G.T.....GGACA.....TTCG----T.TCC..T.AGT...T.A...GCAGGTAT 606
en-ant6 G..GAG-----A.A.A.....AATGAT.....T..----..A.AC...GAC.GG...A..TAGAGCTCA 614
en-arc1 G..GA.-----CTGCA.....CGATAGT.....T..A---CTG.CTG.A.AT-AC.AT....CTAT.ACA 603
en-arc2 G..GAG-----..GGTT....CA.TCTT....A.ATA----.G.TCC.A.TT.....TT..T.TAGCTA. 603
en-arc3 G..GA.-----G...A.....CGATAAT.....AGCCCAG....ATG.TCACTATAGTTGCAT...T 604
en-arc4 G..CA.GTTGC.T.AA..GC.AGGAC.T.A.A.ATTGTAA.A...C.ACT.G..CT.G...TT.Ag...t 627

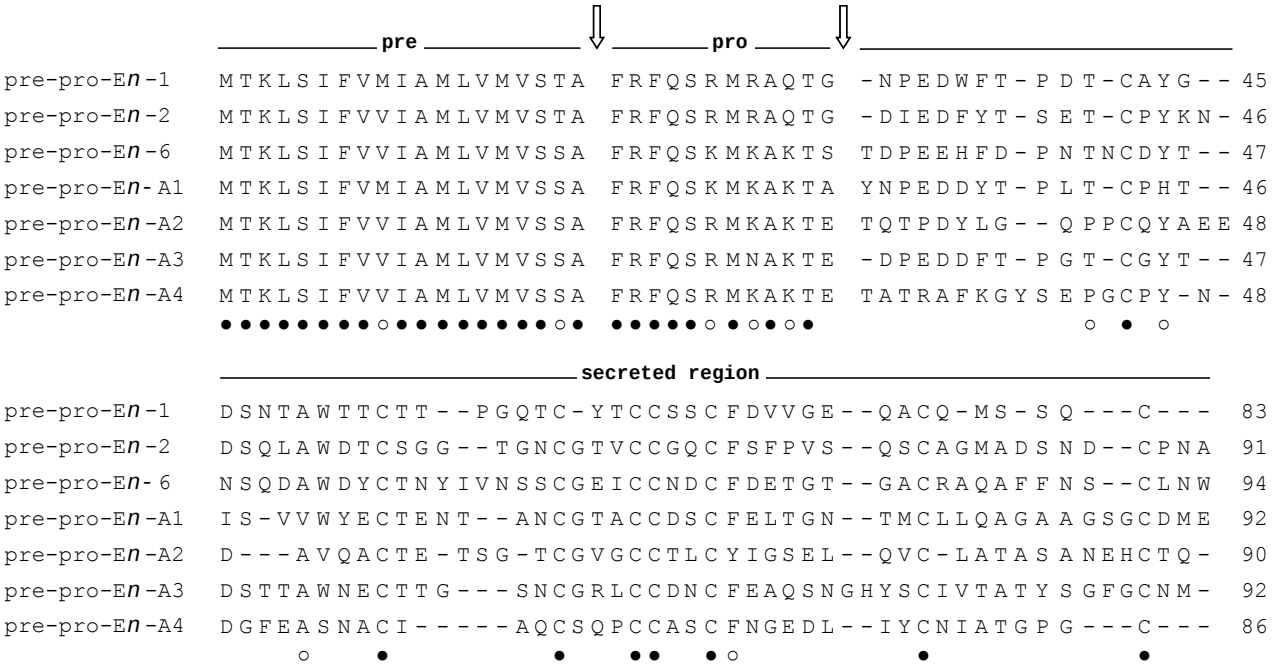

en-ant1 G----------GCCCAATGCTAAagctaaa-tggctagctgc---tcag-----aagccgctggt 646
en-ant2 .----------GCC.ATTCTAATG.CT..CCT.A..CG..A...g---....-----...a..... 656
en-ant6 .----------GCATTTGG.AATAGTT..CTC.A.T.G..Aa...---....-----..... 664
en-arc1 A------GCAGGTGCA..AGG.A..GG.T.TG.C.TG.AA..A..a.atac...ctatt.....aac... 667
en-arc2 T----------GCTAGTGC.AATG..CAT.GT.C.CAA..Ag..actag.g.ttcag.....at.... 661
en-arc3 AC----TGCAACTTATT.TGGT.T.GG.T.TA.C.TGTAA.....atac...ctattg....a.ca.. 670
en-arc4 agcggttcctagtcactaggt..t.ag.g.ta.tcggt.t.t...a.ac-.tg.acgac..t.tttcaa. 696
en-ant1 gtgtgctaaggt--tagcctgtgcaccacc------attattttgttaacaaaataacaacagttcaaat 708
en-ant2 ....--c.............-----c...........gt..--.............. 717
en-ant6 ....--c.....c.....-----c...........gt..--...........t... 725
en-arc1 ....a.--c.a...c.....-----c...........gt..--..g............ 728
en-arc2 ....ca--a..a..c.....tcaccc...........gt..--..........a.... 727
en-arc3 ....--c.............-----c...........gt..--.............. 731
en-arc4 t.ca.a.....ctcc.t.tga....a....taaccc...........gt..--..........t.... 764
en-ant1 atctactgtctaagtttgtgt---ggttgtacctgg-ttaggtaggagggaga---gggtgtaggttgga 771
en-ant2 ....---a..........-....a..........---.............. 780
en-ant6 ....t........---...........-.............---.--g..gta--..t 784
en-arc1 ....tt.......ga-..ga.....a..-.....c.........g---.ttg...taa...t 793
en-arc2 ..a.cg.t.....a...a.a.ggtt.acc..t....g...t...t.cta.gttt--..t.t.g..ca..g 795
en-arc3 ..a.gt...........a..----..g.....t...-.............---.--g..-t.a..tg 790
en-arc4 .cactg.......a...a..cga-aa.ccc.tt.aa-....t.taa.cat.tgctg.....a..a..... 832
en-ant1 agagcgaagggctggggtagactagtttaggggtagttcttagctttaaatatgttacaaatataaattt 841
en-ant2 ....gg.......t..........ta.............................------..tt.... 844
en-ant6 t.---....a.---..aag.g..t.gg..aac-.....tag..g.agttc.tatc.tt.......t.... 847
en-arc1 ta---.gt.t.t.a.a.a.t.t.t..gcta.a-.....a.aga.ca.t.gat.....g.g.c..t..... 859
en-arc2 tatgtaggaa.a..tttg..gt..a.g.tta.aa.ta.a.at.tg..tt.g.a..ggg.gg....c.... 865
en-arc3 ..---..g.t.t.a..ag.t.t.t.cg.ta..-....ga.aga.ca.t.ggc.....g..ta...t.... 856
en-arc4 t.t.t.ggta.gacatt.ta.t..t...ttt.ag.ag.tggtag.gggtga.g.cggtggggt..tt..a 902
en-ant1 aattatcgaacttttttaaattttactctgaataatattcaacatctcctaccattcac 900
en-ant2 903
en-ant6 ....a..... 906
en-arc1 ....g.a.............a..... 918
en-arc2 ....-..t.a.............ga..... 923
en-arc3 ....a..... 915
en-arc4 ....a.............a..... 961

Figure

Click here to download high resolution image



Figure


pre
pro
pre-pro-En-1 MTKLSIFVMIAMLVMVSTA FRFQSRMRAQTG -NPEDWFT-PDT-CAYG-- 45
pre-pro-En-2 MTKLSIFVVIAMLVMVSTA FRFQSRMRAQTG -DIEDFYT-SET-CPYKN- 46
pre-pro-En-6 MTKLSIFVVIAMLVMVSSA FRFQSKMKAKTS TDPEEHFD-PNTNCDYT-- 47
pre-pro-En-A1 MTKLSIFVMIAMLVMVSSA FRFQSKMKAKTA YNPEDDYT-PLT-CPHT-- 46
pre-pro-En-A2 MTKLSIFVVIAMLVMVSSA FRFQSRMKAKTE TQTPDYLG--QPPCQYAEE 48
pre-pro-En-A3 MTKLSIFVVIAMLVMVSSA FRFQSRMNAKTE -DPEDDFT-PGT-CGYT-- 47
pre-pro-En-A4 MTKLSIFVVIAMLVMVSSA FRFQSRMKAKTE TATRAFKGYSEPGCPY-N- 48
secreted region
pre-pro-En-1 DSNTAWTTCTT--PGQTC-YTCCSSCFDVVGE--QACQ-MS-SQ---C--- 83
pre-pro-En-2 DSQLAWDTCSGG--TGNCGTVCCGQCFSFPVS--QSCAGMADSND--CPNA 91
pre-pro-En-6 NSQDAWDYCTNYIVNSSCGEICCNDCFDETGT--GACRAQAFFNS--CLNW 94
pre-pro-En-A1 IS-VVWYECTENT--ANCGTACCDSCFELTGN--TMCLLQAGAAGSGCDME 92
pre-pro-En-A2 D---AVQACTE-TSG-TCGVGCCTLCYIGSEL--QVC-LATASANEHCTQ- 90
pre-pro-En-A3 DSTTAWNECTTG---SNCGRLCCDNCFEAQSNGHYSCIVTATYSGFGCNM- 92
pre-pro-En-A4 DGFEASNACI-----AQCSQPCCASCFNGEDL--IYCNIATGPG---C--- 86