



Data Article

A new dataset of pheromone and pheromone-gene structures from the ciliate, *Euplotes crassus*



Graziano Di Giuseppe^{a,b,*}, Claudio Alimenti^c, Pierangelo Luporini^c,
Adriana Vallesi^{c,*}

^a Department of Biology, University of Pisa, Italy

^b MARinePHARMA Center, University of Pisa, Italy

^c Laboratory of Eukaryotic Microbiology and Animal Biology, School of Bioscience and Veterinary Medicine, University of Camerino, Italy

ARTICLE INFO

Article history:

Received 5 April 2023

Revised 9 May 2023

Accepted 14 July 2023

Available online 20 July 2023

Dataset link: [Pheromone and pheromone-gene sequences from the marine ciliate, *Euplotes crassus* \(Original data\)](#)

Keywords:

Water-borne protein pheromones

Ciliate multiple mating-type systems

Macronuclear gene-size DNA molecules

Ciliate conjugation

Sex in protozoan ciliates

ABSTRACT

Like many other organisms, ciliates communicate and interact socially via diffusible chemical signals, named pheromones, that are functionally associated with a genetic mating-type mechanism of cell self/not-self recognition. In *Euplotes* species, pheromones form species-specific families of small, globular, and disulfide-rich proteins folding into exclusively helical secondary structures. Each is specified by one of a series of high-multiple alleles that are inherited in Mendelian fashion with relationships of co-dominance at the so-called *mat* genetic locus of the cell transcriptionally inert micronuclear genome, and expressed in the transcriptionally active macronuclear genome as individual DNA molecules in which the central coding region is flanked by 5'-leader and 3'-trailer noncoding regions ending with C₄A₄/T₄G₄ telomeric repeats. In *E. crassus*, a cosmopolitan marine species with a long tradition in the study of ciliate mating systems and breeding patterns, oligonucleotides specific to amino acid sequences of pheromones Ec-1 and Ec- α were previously used to clone and sequence a first set of four structurally dis-

* Corresponding authors.

E-mail addresses: graziano.di.giuseppe@unipi.it (G. Di Giuseppe), adriana.vallesi@unicam.it (A. Vallesi).

tinct macronuclear (*mac*) pheromone coding genes, *mac-ec- α* , *mac-ec-1*, *mac-ec-2* and *mac-ec-3*, from two interbreeding strains, L-2D and POR-73. The use of these oligonucleotides in PCR amplifications of macronuclear DNA preparations from three other *E. crassus* interbreeding strains, ES10, Fava4 and MN4, has now resulted in the characterization of a second set of eight new pheromone coding genes, *mac-ec- β* , *mac-ec- γ* , *mac-ec- δ* , *mac-ec- ϵ* , *mac-ec- μ* , *mac-ec-4*, *mac-ec-5* and *mac-ec-6*. Multiple alignment between previously and newly determined pheromone-gene sequences reinforces the concept that the *E. crassus* pheromone-gene family includes two sub-families, which likely reflect a duplication of the micronuclear *mat* gene locus and represent an apomorphic trait of the *E. crassus* clade. Members of one sub-family (each identified with a Greek letter) show a 500-bp 5'-leader noncoding region rich in AGGA/AGGGA repetitions, and encode 56-amino acid pheromones with eight conserved Cys residues. Members of the other sub-family (each identified with an Arabic numeral) show an 800-bp 5'-leader noncoding region without AGGA/AGGGA repetitions, and encode 45-amino acid pheromones with ten conserved Cys residues.

© 2023 The Author(s). Published by Elsevier Inc.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

Specifications Table

Subject	Biological Sciences
Specific subject area	Eukaryotic Microbiology, Protistology/Protozoology, Molecular/Cell Biology
Type of data	Two tables and three figures
How the data were acquired	Data were obtained from Sanger sequencing of polymerase chain reaction (PCR)-purified products.
Data format	Raw Analyzed
Description of data collection	<i>Euplotes</i> strains used in this study were taxonomically identified as <i>E. crassus</i> on the basis of morphological and genetic criteria [1]. Macronuclear DNA preparations were amplified using primers designed on conserved sequences adjacent to the telomeric ends and previously determined from other <i>E. crassus</i> pheromone coding genes [2]. Amplicons were cloned, plasmids of five distinct clones were Sanger-sequenced in both directions, and partial sequences were overlapped to produce the complete gene sequences.
Data source location	Living cell samples and DNA preparations from <i>E. crassus</i> strains source of the analyzed pheromone coding genes are available from the G.D.G. laboratory at the Department of Biology of Pisa University, and the A.V. laboratory at the School of Biosciences and Veterinary Medicine of Camerino University.
Data accessibility	Repository name: Mendeley data https://data.mendeley.com/datasets/63m8bp8zxc
Related research article	A. Vallesi, C. Alimenti, S. Federici, G. Di Giuseppe, F. Dini, G. Guella, P. Luporini, Evidence for gene duplication and allelic codominance (not hierarchical dominance) at the mating-type locus of the ciliate, <i>Euplotes crassus</i> . J. Eukaryot. Microbiol. 61 (2014), 620-629. https://doi.org/10.1111/jeu.12140

Value of the Data

- The mating-type mechanism that controls ciliate conjugation is basically a phenomenon of cell self/not-self recognition driven by cell interactions with diffusible protein pheromones. The determination of new pheromone and pheromone-gene sequences from *E. crassus* improves the knowledge of the molecular basis of this mechanism.
- In *E. crassus*, pheromones are secreted in tiny amounts making it hard to determine their complete amino acid sequences in a significant number via direct chemical analysis of native protein preparations. The PCR primers reported in the dataset are designed on conserved stretches of the pheromone-gene sequences, and help to circumvent the chemical way by cloning and sequencing other pheromone coding genes from new *E. crassus* strains.
- An upgraded dataset of pheromone and pheromone-gene sequences may be used for a more reliable identification of which pheromone structural specificities are functionally more relevant in diversifying conspecific cells in their spectrum of mating interactions, and which nucleotide sequence stretches turn out to be useful in designing new primer combinations for qPCR analyses of pheromone-gene amplification and expression.
- The knowledge of more *E. crassus* pheromone amino acid sequences greatly validates the application of the AlphaFold AI system in scouting their three-dimensional folding, and the knowledge of this folding pattern may add a relevant brick to the general picture of how the *Euplotes* pheromone molecular structures—so far determined only from *E. nobilii*, *E. petzi* and *E. raikovi* [3,4]—evolve in concert with speciation and habitat colonization.

1. Objective

Results from traditional Mendelian analyses of the *E. crassus*, *E. minuta* and *E. vannus* mating systems have for long accounted for an eccentric model of serial dominance relationships between single-locus multiple *mat*-alleles [5-7], and the synthesis of insoluble (membrane-bound) protein pheromones [8]. Results from two original research articles, based on molecular approaches to the *E. crassus* mating system, conflicted with this model [2,9]. They provided evidence that (i) *E. crassus* cells conform with the more common (basic) ‘*E. patella*-model’ in relation to *mat*-allele relationships of co-dominance and the synthesis of diffusible pheromones [10], and (ii) the difference resides in the evolution of two structurally distinct pheromone-coding gene sub-families likely reflecting a phenomenon of *mat*-gene locus duplication [2]. In addition to reinforcing evidence that a *mat*-gene locus duplication is a distinctive (apomorphic) feature of the *E. crassus* clade, the knowledge of new *E. crassus* pheromone and pheromone-gene sequences may help (i) planning *in silico* analyses of whether and how pheromones from the two sub-families interact in binding target cells, and (ii) exploring whether and how these interactions can account for the serial-dominance relationships that were originally presumed to govern *mat*-allele inheritance and expression in the species of the *E. crassus* clade.

2. Data Description

Table 1 reports the denominations of all the pheromone-coding genes that have been cloned from *E. crassus* and analyzed for their sequences, while Table 2 enumerates the respective GenBank sequence numbers. Figs. 1 and 2 illustrate the multiple alignments of the nucleotide

sequences that are overall known from *E. crassus* for each of the two pheromone-gene subfamilies: one (Fig. 1) including genes (labelled with progressive Arabic numbers) that show closer orthologous relationships with pheromone genes of other *Euplotes* species; the second (Fig. 2) including genes (labelled with Greek letters) with closer *E. crassus* specificity. Fig. 3 illustrates the multiple amino acid sequence alignments of the cytoplasmic pheromone precursor forms (pre-pro-pheromones) that are specified by the structurally determined sequences of genes from the two sub-families.

Table 1

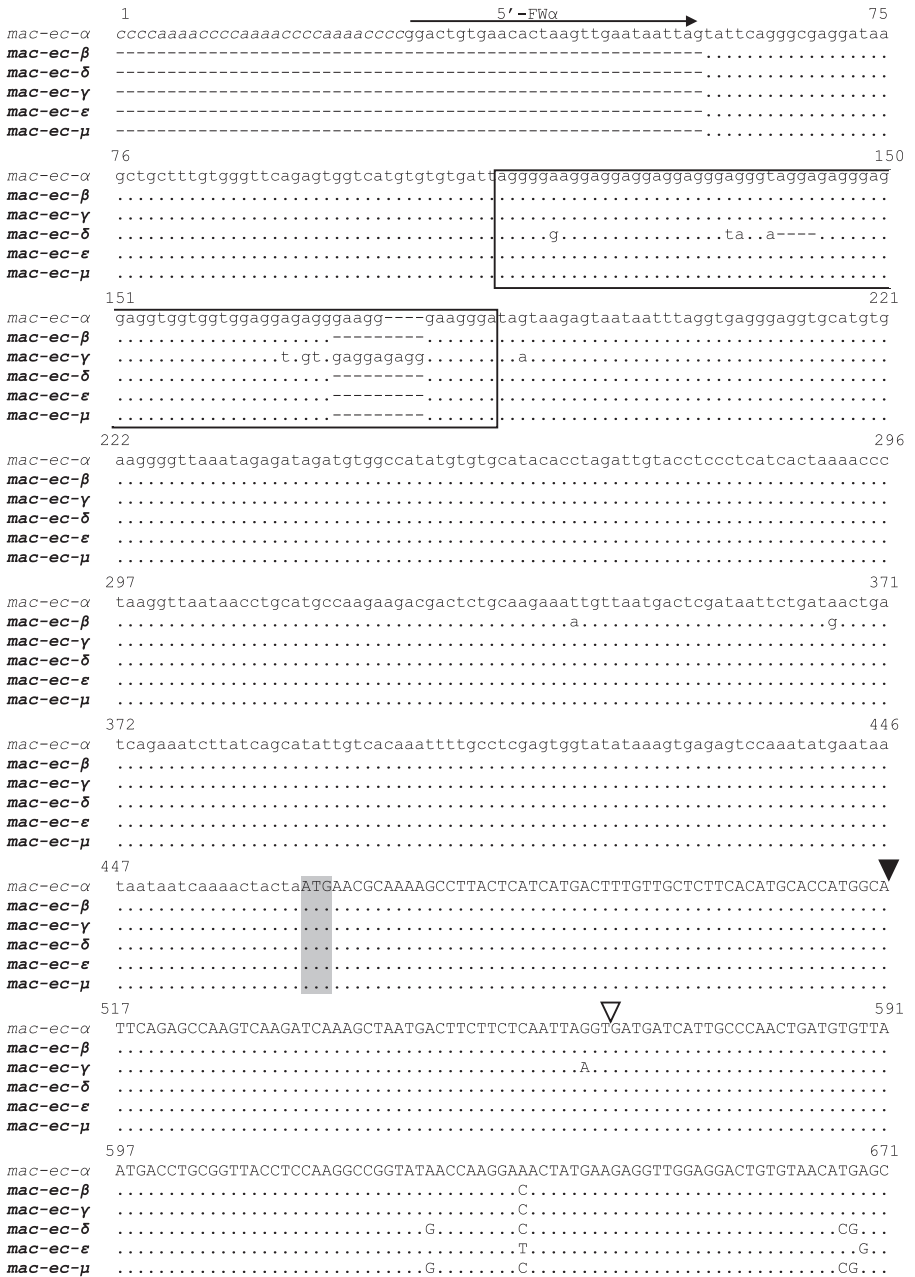
Denominations, encoded pheromones and source strains of previously (light letters) and newly (bold letters) characterized *E. crassus* macronuclear pheromone genes.

<i>Pheromone genes</i>	<i>Encoded pheromones</i>	<i>Source strains</i>	<i>Collection sites</i>	<i>Latitude/Longitude</i>	<i>Year</i>
<i>mac-ec-α, mac-ec-1</i>	<i>Ec-α, Ec-1</i>	L-2D	Leghorn (IT)	N 43°33'/E 10°19'	1978
<i>mac-ec-α, mac-ec-2, mac-ec-3</i>	<i>Ec-α, Ec-2, Ec-3</i>	POR-73	Porto Recanati (IT)	N 43°26'/E 13°40'	1978
<i>mac-ec-β, mac-ec-4</i>	<i>Ec-β, Ec-4</i>	MN4	Montecristo Island (IT)	N 42°20'/E 10°19'	1978
<i>mac-ec-γ, mac-ec-δ, mac-ec-5</i>	<i>Ec-γ, Ec-δ, Ec-5</i>	Fava4	Favone, Corsica Island (FR)	N 41°46'/E 9°23'	2002
<i>mac-ec-ε, mac-ec-μ, mac-ec-6</i>	<i>Ec-ε, Ec-μ, Ec-6</i>	ES10	Motril (ES)	N 36°44'/W 3°31'	1986

Table 2

GenBank accession numbers of *E. crassus* macronuclear pheromone genes. Genes with previously determined sequences are in light letters, and genes with newly determined sequences are in bold letters.

<i>Gene denominations</i>	<i>Accession numbers</i>
<i>mac-ec-α</i>	HM072429
<i>mac-ec-1</i>	HM072430
<i>mac-ec-2</i>	KF905230
<i>mac-ec-3</i>	KF905231
<i>mac-ec-β</i>	OQ630543
<i>mac-ec-4</i>	OQ630548
<i>mac-ec-γ</i>	OQ630544
<i>mac-ec-δ</i>	OQ630545
<i>mac-ec-5</i>	OQ630549
<i>mac-ec-ε</i>	OQ630546
<i>mac-ec-μ</i>	OQ630547
<i>mac-ec-6</i>	OQ630550



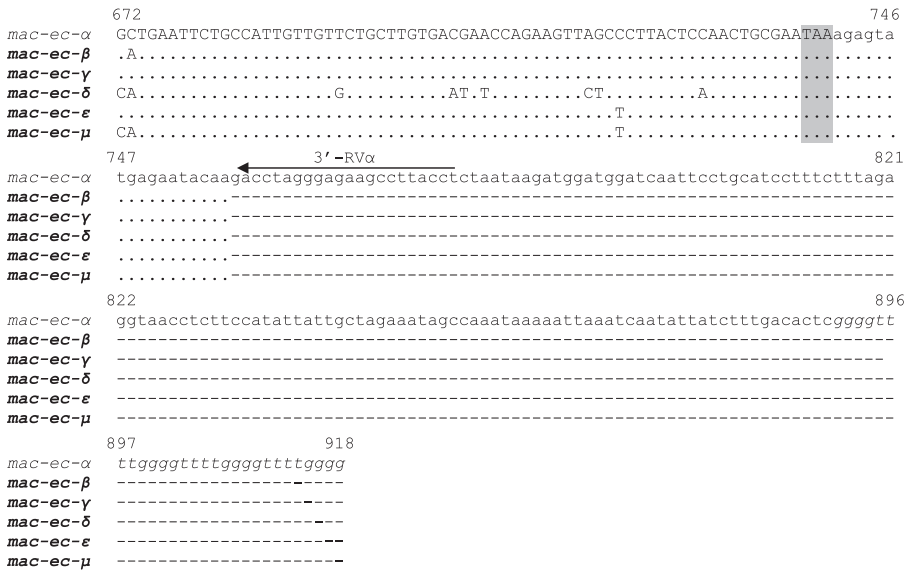


Fig. 1. Multiple nucleotide sequence alignment of *E. crassus* pheromone genes of the subfamily (distinguished by Greek letters) encoding pheromone sequences of 56 amido acids with eight Cys residues. Gaps were inserted to optimize the alignment, and dots indicate identical nucleotides. A previously determined sequence is in light letters, and the newly determined sequences are in bold letters. The 5'-leader and 3'-trailer non-coding regions are in lower case letters, the coding region is in capital letters, and the telomeric repetitions are in italics. The ATG start codon and the TAA stop codon are shadowed. Filled and light arrowheads delimit the sequence segments specifying the signal peptide and the pro segment, respectively, of the cytoplasmic (immature) pheromone precursor form (pre-pro-pheromone). In the 5'-leader region, a box delimits the domain characterized by AGGA/AGGA repetitions.

```

1                                     75
mac-ec-1  ccccaaaacccccaaaacccccaaaaccccgatataccaattatatatagaagctgtctagaaatgtccgtttcgtaa
mac-ec-2  .....
mac-ec-3  .....
mac-ec-4  .....
mac-ec-5  .....
mac-ec-6  .....

76                                     150
mac-ec-1  attatattatatggttgcgctcaacctcacctccaaaagttaggttaatcgcttggaaaattcactaaatcatta
mac-ec-2  .....c.g.....
mac-ec-3  .....c.g.....
mac-ec-4  .....c.....
mac-ec-5  .....c.....
mac-ec-6  .....c.g.....

151                                     225
mac-ec-1  ggaatactactccagggttatagtgaccaatttgccattttatgaggtcattaagaagtctctagctgaaagc
mac-ec-2  .....a.....c.....
mac-ec-3  .....a.....c.....
mac-ec-4  .....g.....g.....c.....
mac-ec-5  .....g.....c.....
mac-ec-6  .....gg.....c.....

226                                     300
mac-ec-1  aaaggcttagataaaaagcattttatctgtggttgaataaaatctttgcaattttagaagccataagggcacgaaa
mac-ec-2  .....
mac-ec-3  .....
mac-ec-4  .....
mac-ec-5  .....
mac-ec-6  .....

301                                     375
mac-ec-1  aattctaatttgagtaccggagtgtgtagaagcttcagcttcctaacacgatgaaagaagaggaggtataa
mac-ec-2  .....g.....
mac-ec-3  .....g.....
mac-ec-4  .....g.....
mac-ec-5  .....g.....
mac-ec-6  .....g.....

376                                     449
mac-ec-1  tgtaatattactctttcgaataatcggaagaaaaaa-tatagaaaatttagaattttatcaagcctcacgag
mac-ec-2  .....---...t.-...
mac-ec-3  .....ata.....
mac-ec-4  .....---.....
mac-ec-5  .....---.....
mac-ec-6  .....a...---.....

450                                     524
mac-ec-1  gagcttctcctaagaaaacaataataccagcaaccagaaagatcaatcggatattttaggggtgctctcgctct
mac-ec-2  .....
mac-ec-3  .....
mac-ec-4  .....g.....
mac-ec-5  .....
mac-ec-6  .....g.....

525                                     599
mac-ec-1  ctgtatgactcgtcgattttccgaattatttcgaaattcccaaaactaattataatggttagtcttagttagcaaa
mac-ec-2  .....
mac-ec-3  .....
mac-ec-4  .....
mac-ec-5  .....
mac-ec-6  .....

600                                     674
mac-ec-1  ttggaaataccgcaaaaggcatgatgccgaagaagaatgtggcaagaaattctcatcaaaaacaattttcccttc
mac-ec-2  .....t.....
mac-ec-3  .....t.....
mac-ec-4  .....t.....
mac-ec-5  .....t.....
mac-ec-6  .....g.....t.....

```

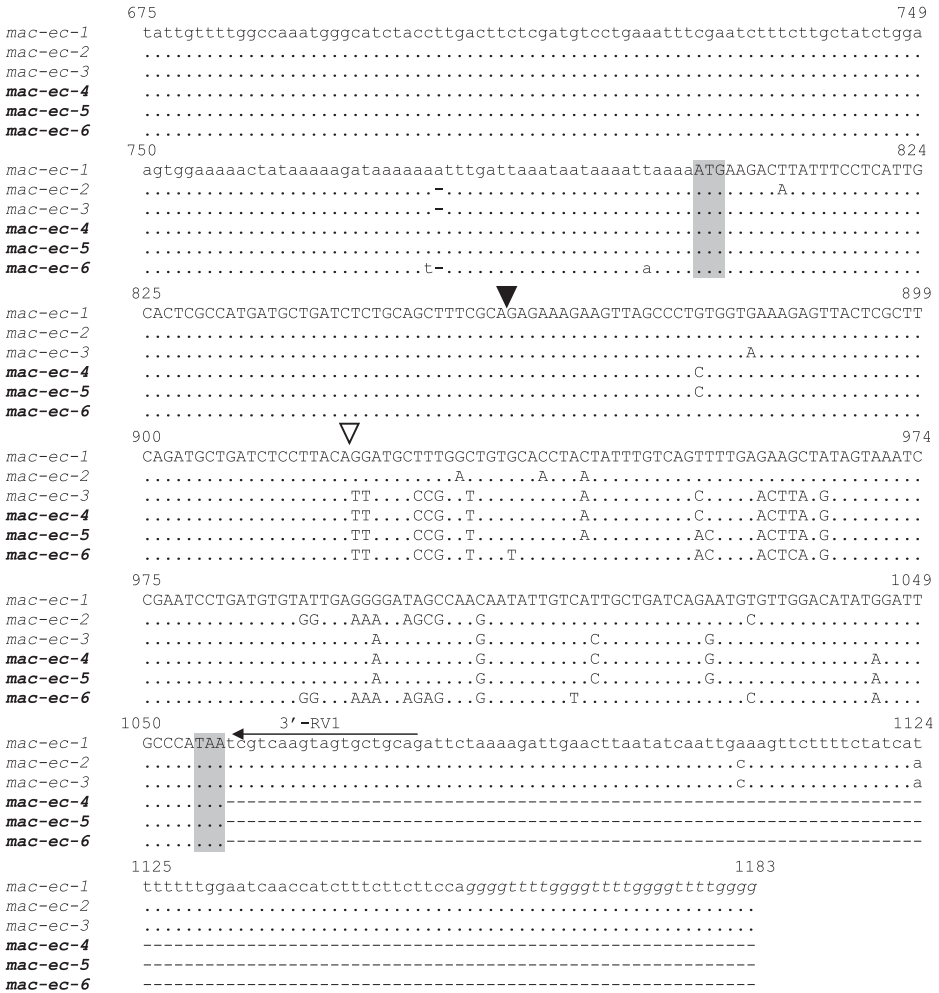


Fig. 2. Multiple nucleotide sequence alignment of *E. crassus* pheromone genes of the subfamily (distinguished by Arabic numerals) encoding pheromone sequences of 45 amido acids with ten Cys residues. Gaps were inserted to optimize the alignment, and dots indicate identical nucleotides. Three previously determined sequences are in light letters, and the newly determined sequences are in bold letters. The 5'-leader and 3'-trailer non-coding regions are in lower case letters, the coding region is in capital letters, and the telomeric repetitions are in italics. The ATG start codon and the TAA stop codon are shadowed. Filled and light arrowheads delimit the sequence segments specifying the signal peptide and the pro segment, respectively, of the cytoplasmic (immature) pheromone precursor form (pre-pro-pheromone).

	1	pre	20	pro	40	secreted	60	80	91
pre-pro-Ec- α	MNAKALLIMTLLLF	CTMAFRAKRSRSLMTSSQLGDDHCPTDVLMT	CGYLQGRYNQGNIEYEVGGLCNMSAEFCH	CCSAC	DEPEVSPYSNCE				
pre-pro-Ec- β									
pre-pro-Ec- γ									
pre-pro-Ec- δ									
pre-pro-Ec- ϵ									
pre-pro-Ec- μ									

	1	pre	20	pro	40	secreted	60	80	84
pre-pro-Ec-1	MKTYFLIALAMMLI	SAAFAEKEVSPVVKELLASDADLLTG	CFGCAPT	CGFCE	EAIVNPNPDVYCGDSQQYCH	CCSECV	GHMDCP		
pre-pro-Ec-2									
pre-pro-Ec-3									
pre-pro-Ec-4									
pre-pro-Ec-5									
pre-pro-Ec-6									

Fig. 3. Multiple alignment of pre-pro-pheromone amino acid sequences specified by the two *E. crassus* pheromone-gene sub-families. Bars span the pre, pro and secreted regions, Cys residues are highlighted in bold, and dots indicate identical residues. Previously and newly determined sequences are distinguished by denominations in light and bold letters, respectively.

3. Experimental Design, Materials and Methods

3.1. Cell Cultures

The *E. crassus* strains used for the pheromone-gene sequence determination were expanded each starting from a single specimen isolated from samples of shallow water and sediment. They were cultivated in sterilized natural, or artificial seawater, at 20–22°C under a light-dark cycle following routine procedures [2] and using the green algae *Dunaliella tertiolecta* as standard food source.

3.2. DNA Preparation and PCR Amplification

DNA was prepared, following a standard protocol [1,2], from *E. crassus* cultures previously deprived of food for 2 days and resuspended in fresh seawater for 1 day. Amplifications were run in a GeneAmp 9700 thermal cycler (PE Applied Biosystems) following a standard program (30s 94°C, 40s 58°C, 60s 72°C for 35 cycles), and using oligonucleotides synthesized by Invitrogen (Van Allen Way, Carlsbad, CA). The two oligonucleotides, 5'-FW1 (GCTCAACCTCACCTCAA) and 3'-RV1 (CTGCAGCACTACTGACGA), were used as forward and reverse primers, respectively, to amplify the *mac-ec-4*, *mac-ec-5* and *mac-ec-6* genes. The *mac-ec- β* , *mac-ec- γ* , *mac-ec- δ* , *mac-ec- ϵ* , and *mac-ec- μ* genes were instead amplified with the two oligonucleotides, 5'-FW α (GGACTATGAACCTAAGTTGAATAATTAG) and 3'-RV α (AGGTAAGGCTTCTCCCTAGGTC), used as forward and reverse primers, respectively. The 50- μ l reaction mixtures contained 0.5 μ g of DNA, 2 mM MgCl₂, 250 mM of dNTP, one unit of Taq DNA polymerase (Polymed, Florence, Italy), and 0.2 mM of each primer. Amplicons were cloned into the pCR 2.1-TOPO vector of the TOPO TAcloning Kit (Invitrogen), and five clones were analyzed for each gene sequence determination.

Ethics Statements

This work did not involve experimental analyses of animals or humans.

Data Availability

Pheromone and pheromone-gene sequences from the marine ciliate, *Euplotes crassus* (Original data) (Mendeley Data)

CRediT Author Statement

Graziano Di Giuseppe: Conceptualization, Investigation, Data curation, Writing – original draft; **Claudio Alimenti:** Investigation, Formal analysis; **Pierangelo Luporini:** Conceptualization, Supervision, Writing – review & editing; **Adriana Vallesi:** Conceptualization, Investigation, Data curation, Supervision, Writing – original draft.

Declaration of Competing Interest

The authors declare no competing financial interest.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

- [1] G. Di Giuseppe, F. Dini, F. Erra, P. Luporini, A. Vallesi, Analysis of autapomorphic point mutations provides a key for the tangled taxonomic distinction of the closely related species, *Euplotes crassus*, *E. minuta* and *E. vannus* (Ciliophora, Euplotida), *Eur. J. Protistol.* 86 (2022) 125917, doi:10.1016/j.ejop.2022.125917.
- [2] A. Vallesi, C. Alimenti, S. Federici, G. Di Giuseppe, F. Dini, G. Guella, P. Luporini, Evidence for gene duplication and allelic codominance (not hierarchical dominance) at the mating-type locus of the ciliate, *Euplotes crassus*, *J. Eukaryot. Microbiol.* 61 (2014) 620–629.
- [3] P. Luporini, B. Pedrini, C. Alimenti, A. Vallesi, Revisiting fifty years of research on pheromone signaling in ciliates, *Eur. J. Protistol.* 55 (2016) 26–38.
- [4] C. Alimenti, F. Buonanno, G. Di Giuseppe, G. Guella, P. Luporini, C. Ortenzi, A. Vallesi, Bioactive molecules from ciliates: Structure, activity, and applicative potential, *J. Eukaryot. Microbiol.* 69 (2022) e12887, doi:10.1111/jeu.12887.
- [5] K. Heckmann, Experimentelle untersuchungen an *Euplotes crassus*: paarungssystem, konjugation und determination der paarungstypen, *Z. Vererbungsl.* 95 (1964) 114–124.
- [6] K.G. Grell, *Protozoology*, Springer-Verlag, Berlin, 1973.
- [7] R. Nobili, P. Luporini, F. Dini, Breeding system, species relationships and evolutionary trends in some marine species of Euplotidae (Ciliata Hypotrichida), in: B. Battaglia, J. Beardmore Marine Organisms, Genetics, Ecology and Evolution, Plenum Press, New York, 1978, pp. 591–616.
- [8] A. Miyake, Fertilization and sexuality in ciliates, in: K. Hausmann, P.C. Bradbury (Eds.), *Ciliates: Cells as Organisms*, Fischer Verlag, Stuttgart, 1996, pp. 243–290.
- [9] C. Alimenti, A. Vallesi, S. Federici, G. Di Giuseppe, F. Dini, V. Carratore, P. Luporini, Isolation and structural characterization of two water-borne pheromones from *Euplotes crassus*, a ciliate commonly known to carry membrane-bound pheromones, *J. Eukaryot. Microbiol.* 58 (2011) 234–241.
- [10] R.F. Kimball, The nature and inheritance of mating types in *Euplotes patella*, *Genetics* 27 (1942) 269–285.