Insight into the morphology and genetic diversity of the Chaetoceros tenuissimus (Bacillariophyta) species complex

Daniel Grzebyk, Vanina Pasqualini, Marie Garrido, Yann Quilichini, Clément Pereto, Philippe Cecchi

Insight into the morphology and genetic diversity of the *Chaetoceros tenuissimus* (Bacillariophyta) species complex

Daniel Grzebyk [a], Vanina Pasqualini [b], Marie Garrido [c], Yann Quilichini [b], Clément Pereto [b], Philippe Cecchi [a]

[a] MARBEC, Univ Montpellier, CNRS, Ifremer, IRD, Montpellier, France

[b] UMR CNRS Sciences for the Environment / UMS CNRS Stella Mare, University of Corsica, BP 52, 20250 Corte, France

[c] Environmental Agency of Corsica, 7 avenue Jean Nicoli, 20250 Corte, France

Corresponding author:

Vanina Pasqualini *pasqualini_v@univ-corse.fr*; ORCID 0000-0001-5919-4805

Daniel Grzebyk: daniel.grzebyk@umontpellier.fr; ORCID 0000-0002-1130-7724

Marie Garrido: marie.garrido@oec.fr; ORCID 0000-0002-5411-5060

Yann Quilichini: quilichini_y@univ-corse.fr; ORCID 0000-0003-0739-0155

Clément Pereto: clement.pereto@gmail.com; ORCID 0000-0003-3031-8960

Philippe Cecchi: philippe.cecchi@ird.fr; ORCID 0000-0001-5562-6877

ABSTRACT

Among the marine planktonic diatoms, *Chaetoceros* is among the most species-rich genera, and many *Chaetoceros* species are considered important primary producers. However, little is known about the ecology and distribution of few small solitary species within this genus, including *Chaetoceros tenuissimus*. This article describes a minute *Chaetoceros* strain, identified as *C. tenuissimus* and named CT16ED, that was isolated at a coastal lagoon in Corsica Island, Western Mediterranean. The strain was characterized by light microscopy and scanning and transmission electron microscopy, with a specific focus on the fine structure and construction of setae, and by studying its behaviour in culture. Then, the CT16ED strain was compared with other strains we isolated from the species type locality (Ostend harbour, North Sea) by sequencing a fragment of the nuclear ribosomal DNA (rDNA) spanning from the 18S rDNA to the D3 region of the 28S rDNA, and the plastid *rbc*L gene that codes the large RuBisCO subunit. On the basis of the literature and the available sequencing data, the analysed strains were similar to *C. tenuissimus*, but the phylogenetic analysis evidenced a *C. tenuissimus* species complex that contained several clades. The current taxonomical status of *C. tenuissimus* is discussed. The comparison with the available rDNA and *rbc*L sequencing data of strains assigned to species considered as synonyms of *C. tenuissimus*, including *Chaetoceros simplex* var. *calcitrans*, *Chaetoceros calcitrans* and *Chaetoceros calcitrans* f. *pumilus*, suggested that these taxa are paraphyletic in the genus *Chaetoceros*.

KEYWORDS *Chaetoceros tenuissimus*; rDNA; internal transcribed spacer (ITS); morphology; cell division; ultrastructure; phylogeny

Introduction

Species in the planktonic diatom genus *Chaetoceros* Ehrenberg are among the key primary producers in the world's oceans and coastal seas (Rines & Hargraves, 1988; Leblanc *et al.*, 2012). The genus is highly diverse with more than 200 accepted species (Rines & Hargraves, 1988; Hernández-Becerril, 1996; Shevchenko *et al.*, 2006; Guiry, 2019), and new species are continuously described (e.g. Li *et al.*, 2013; Chen *et al.*, 2018; Kaczmarska *et al.*, 2019; Xu *et al.*, 2019a, 2019b). The species appear to be widely distributed (Malviya *et al.*, 2016; De Luca *et al.*, 2019a). Many of morphologically recognized species show remarkable cryptic diversity (Chamnansinp *et al.*, 2013, 2015; Balzano *et al.*, 2017; Gaonkar *et al.*, 2017; Li *et al.*, 2017; Xu *et al.*, 2018, 2019 a, 2019b, 2020).

Within the genus *Chaetoceros*, some solitary species have been described and were placed in the section Simplicia that poses "*some of the greatest problems*" for species diagnosis, compared with chain-forming species that have distinctive colony characteristics (Rines & Hargraves, 1988). Few small species have raised taxonomic issues due to the "*lack of distinctive features in the vegetative cells*", as highlighted by Rines & Hargraves (1988). Most of these species have remained poorly documented (i.e. without using electron microscopy descriptions) and this may have led to misidentifications.

Among these small solitary species, *Chaetoceros tenuissimus* Meunier was described from samples collected in an artificial oyster basin in Ostend harbour, North Sea (Belgium), in August 1912. None of the original samples is available today for re-analysis. The three drawn specimens in Meunier (1913), from low magnification microscopy observations (Fig. S1), show very minute cylindrical cells with setae in the sagittal plane (i.e. the apical plane, following Rines & Hargraves, 1988), without any indication of the estimated cell size. Cell shape is square to rectangular in girdle view. The straight setae are oriented with a ~45° angle relative to the pervalvar and apical axes. They are approximately three times and five times longer than the diameter of a square cell and of a rectangular, narrower cell, respectively. The presence of spores was not reported. A subsequent description was made by Hustedt (1930). In more recent studies, *C. tenuissimus* samples have been described as very small cells (3-5 µm in

diameter, 3 to 7-12 µm in length), among the smallest in the genus, and have been observed as solitary cells in natural phytoplankton samples (Rines & Hargraves, 1988; Hasle & Syvertsen, 1997; Bérard-Therriault *et al.*, 1999). The thin setae can be up to 10 times longer than the cell diameter, as illustrated in Bérard-Therriault *et al.* (1999). *Chaetoceros tenuissimus* cells contain a single chloroplast (Rines & Hargraves, 1988; Bérard-Therriault *et al.*, 1999; Sar *et al.*, 2002; Shirai *et al.*, 2008). Other small solitary *Chaetoceros* species, such as *Chaetoceros simplex* var. *calcitrans* Paulsen (1905), *Chaetoceros galvestonensis* Collier & Murphy (1962), *Chaetoceros simplex* Ostenfeld Hustedt (1930) and *Chaetoceros calcitrans* f. *pumilus* Takano (1968), have been considered synonyms of *C. tenuissimus* (Rines & Hargraves, 1988; Hasle & Syvertsen, 1997). However, Rines & Hargraves (1988) cautiously asked for the reinvestigation of the type material to validate this conclusion and acknowledged that "*the possible existence of cryptic taxa should, however, be considered*", particularly for *C. tenuissimus*, *C. galvestonensis* and *C. calcitrans* f. *pumilus*. Nevertheless, the species name *C. tenuissimus* was recognized as valid by Hasle & Syvertsen (1997). In the first scanning electron microscopy (SEM) study of specimens assigned to the *C. tenuissimus* species, Sar *et al.* (2002) described the helical structure of setae. Despite these recent descriptions, *C. tenuissimus* is still lacking a declared epitype characterized by high-magnification electron microscopy and DNA sequence data (e.g. ribosomal DNA (rDNA) and/or *rbc*L gene).

Phylogenetic analyses of the genus *Chaetoceros* inferred from D1-D3 sequences of the 28S rDNA divided the sampled species in four major clades, and *C. tenuissimus* strains were placed into a large composite clade (Li *et al.*, 2015; Gaonkar *et al.*, 2018; De Luca *et al.*, 2019b; Xu *et al.*, 2019a). However, the closest phylogenetically related species in this clade have not been well documented, due to the incomplete availability of adequate gene sequences (including the differently variable rDNA regions and the *rbc*L gene) across these species.

Therefore, *C. tenuissimus* remains poorly known, although it is considered a cosmopolitan species that thrives in coastal waters, including brackish lagoons (Hasle & Syvertsen, 1997). High cell concentrations of *C. tenuissimus* (up to 2.4 x 10^7 cells l^{-1}) have been detected in phytoplankton blooms (Rines & Hargraves, 1988; Tomaru *et al.*, 2018). The species has been reported in the North-eastern and South-western Atlantic

Ocean (Rines & Hargraves, 1988; Bérard-Therriault *et al.*, 1999; Sar *et al.*, 2002), Pacific Ocean around Japan (Toyoda *et al.*, 2010; Tomaru *et al.*, 2018), Indian Ocean (Härnström *et al.*, 2009; Deasi *et al.*, 2010) and Mediterranean and Black Seas (Kooistra *et al.*, 2010; Montresor *et al.*, 2013; Baytut *et al.*, 2013). Few studies have detected *C. tenuissimus* in Mediterranean coastal lagoons, by microscopy observations (Sakka Hlaili *et al.*, 2007) and molecular detection (Grzebyk *et al.*, 2017). However, due to the taxonomic uncertainties, it is difficult to firmly establish *C. tenuissimus* distribution, as highlighted by Hasle & Syvertsen (1997). Nevertheless, 18S rDNA metagenomic barcodes assigned as *C. tenuissimus* have been found in samples from all around the world between tropical and temperate waters (De Luca *et al.*, 2019a).

This article describes a minute *Chaetoceros* strain, named CT16ED, isolated from a coastal lagoon of Corsica Island (Western Mediterranean Sea). The morphology and biological features of cultured cells were thoroughly investigated. Then, due to the uncertainties raised by the basic original description and the lack of a truly characterized epitype (i.e. including genetic data), the CT16ED strains and nine strains collected at Ostend harbour (North Sea), close to where Meunier identified the original specimen in 1913, were sequenced. A phylogenetic study was carried out using these sequencing data and the Japanese strain NIES-3715 assigned to *C. tenuissimus* as reference material (Shirai *et al.*, 2008; Toyoda *et al.*, 2010). Sequences of the *C. simplex* var. *calcitrans* CCAP1085/3 strain were also included to challenge the hypothesis that it is a synonym species of *C. tenuissimus*. The obtained morphological and sequencing data allowed reviewing the current taxonomical status of *C. tenuissimus*.

Material and methods

Sampling, strain isolation and culture methods

The *Chaetoceros* strain CT16ED was isolated from the Diana Lagoon in May 2016 (42°07’28”N, 9°31’05”E; Corsica Island, France, Mediterranean Sea). This oval-shaped euryhaline coastal lagoon has a surface area of 5.7 km² and a maximum depth of 11 m, and is permanently connected northward to the sea by a regularly maintained channel (Bec *et al.*, 2011). The Diana Lagoon is privately-owned and mainly used for shellfish culture and fishery. This polymictic lagoon is characterized by important annual variations in temperature (7-28°C) and to a lower extent, in salinity (35-39). On

sampling day, the water temperature was 22°C and salinity was 37. A 50-L volume of water was concentrated through a Apstein 20 µm screen plankton net and the retained material (100 ml) was collected in polyethylene containers and transported at 4°C to the laboratory (< 2 h). To establish unialgal cultures, each single *Chaetoceros* cell was isolated under a Olympus CKX41 inverted microscope by micro-pipetting using a sharpened Pasteur pipette. Each cell was washed several times in sterile seawater (Andersen, 2005), before inoculation in a well of a 24-well culture plate with f/2 culture medium at a salinity of 25, as recommended by Guillard (1975) for optimal growth. Non-axenic unialgal cultures were grown in 100 ml Erlenmeyer flasks (containing ca. 20 ml of culture), incubated at 20 ± 2°C under 40 µmol photons m^{-2} s^{-1} (Spherical Micro Quantum Sensor US-SQS/L Walz) with a 20:4 h light-dark cycle.

To monitor cell chain formation during the exponential growth of the CT16ED strain, cells were grown in f/2 with 40 g l^{-1} of sodium metasilicate and salinity of 25, at 18.1 ± 0.2°C under 190 ± 21 µmol photons m^{-2} s^{-1}, provided by cool-white fluorescent tubes, with a 12:12h light-dark photoperiod, and continuous bubbling of air with 1% of CO_2. Triplicate cultures were set up in Erlenmeyer flasks containing 800 ml culture medium by inoculation of 10^5 cells ml^{-1}. Cell growth was monitored by daily sampling (20 ml) one hour after the onset of the light phase. Samples were fixed in 2% formaldehyde (final concentration). Cells were counted under a microscope in a Malassez chamber. Each sample was counted at least three times, at least 400 cells in total when possible, in order to guarantee an accuracy of ± 10% (Lund *et al.*, 1958). The growth rate, k, was calculated as the number of divisions per day (div d^{-1}) according to Guillard's method (Guillard, 1973).

To identify the putative type material of *C. tenuissimus*, samples were collected also in Ostend harbour (Belgium), at the Vuurtoren dock (51°14'15''N, 2°55'54''E), on 29 and 30 July, 2020. Water temperature was 21-22°C, and salinity was 34. Seawater was pumped through a 100-µm sieve at a ~0.5 m depth, filtered through a 20-µm sieve and collected in a 5-µm sieve. The 5-20 µm phytoplankton collected in the 5-µm sieve were used for isolating solitary *Chaetoceros* cells that matched Meunier's *C. tenuissimus* description. Approximately 140 cells were individually picked with a thinned Pasteur pipette under a Leica DMIL LED inverted microscope, and inoculated into wells of 96-well plates with enriched seawater medium. Cells were grown at 18.5°C in a climate

chamber at the Marine Station of Ostend with a 12:12 h light-dark photoperiod for 5-6 days. Then, well-developed colonies were selected and transferred to culture tubes (Nunc) containing 5 ml of culture medium for travelling to Montpellier University (France). There, isolates were transferred in 50-ml suspension culture flasks (Greiner) with f/2 medium at a salinity of 34, and grown at 16°C and a 12:12 h light-dark photoperiod.

Morphological analyses

The cultured CT16ED strain (i.e. cells grown in 24-well plates and in Erlenmeyer flasks) was examined in detail by light microscopy (LM), SEM, and transmission electron microscopy (TEM). The LM observations were made using live material under an inverted microscope equipped with an Olympus E-620 camera. For SEM analysis, cultured cells were first fixed in formaldehyde (2.5% final concentration). Then, 1-ml drops of cell suspension were deposed on poly-L-lysine-coated slides. After settling for two hours, slides were rinsed in Milli-Q water and dehydrated in 30%, 50%, 70%, 90% and 100% ethanol (30 minutes for each step and twice for 100% ethanol). After critical point drying in an Emitech K850 instrument, samples were mounted on aluminium studs using double-sided carbon adhesive, and coated with gold/palladium in a Quorum Technologies SC7640 sputter coater. Samples were examined under a Hitachi S-3400-N scanning electron microscope operated at an accelerating voltage of 5 and 10 kV. For TEM, a droplet of fixed cell suspension was deposited on a formvar/carbon-coated 100 mesh cooper grid, air dried, and examined under a Hitachi H-7650 transmission electron microscope operated at 80 kV accelerating voltage. TEM and SEM analyses were carried out at the "Service d'Etude et de Recherche en Microscopie Electronique" of the University of Corsica (Corte, France).

Sequencing and phylogenetic analyses

Pelleted cells (1-2 ml of culture) of the Corsica strain CT16ED, the Japanese strain NIES-3715, the nine strains collected in Ostend harbour, and the *C. simplex* var. *calcitrans* strain CCAP1085/3 were resuspended in 100 µl of lysis buffer from the PureLink RNA Mini Kit (Ambion), sonicated on ice for 1 min using an UP-100H ultrasonic processor (Hielscher Ultrasonics, Germany) equipped with a 0.5 mm

diameter sonotrode with the time cycle set to 0.8 (0.8 s sonication and 0.2 s relaxation cycle) and 80% power amplitude. After addition of 400 µl of lysis buffer, lysates were incubated at room temperature for 1 hour and then centrifuged at 12,000 g for 10 min for pelleting cell debris. After a purification step with chloroform:isoamyl alcohol (v:v 24:1), genomic DNA was recovered by ethanol precipitation.

For all analysed strains, an assembled rDNA sequence that included the 18S rRNA gene, the internal transcribed spacer (ITS) region (ITS1, the 5.8S rDNA, and ITS2), and the D1-D3 region of the 28S rRNA gene was constructed with sequencing data obtained from two overlapping PCR-amplified rDNA fragments. The 18S rDNA fragment was amplified and sequenced first. The second rDNA fragment, which included the 18S rDNA end, the ITS region and the beginning of the 28S rDNA, was obtained with a species-specific primer that hybridizes to the 3' end of 18S rDNA (~200-bp overlap with the 18S rDNA amplicon sequence). A nearly full-length fragment of the plastid *rbc*L gene also was amplified and sequenced. The PCR primers are described in supplementary material (Table S1). The PCR reactions were performed using a Mastercycler Ep Gradient S thermal cycler (Eppendorf) and the PrimeSTAR GXL DNA Polymerase Kit (Takara Bio Inc., Japan) with a high-fidelity enzyme. The PCR programme consisted of 40 cycles: 98°C for 15 s, 52°C for 15 s, and 68°C for 2 min, followed by a final elongation period at 68°C for 2 min. The amplicons were purified with the QIAquick PCR Purification Kit (Qiagen), and sequenced using the appropriate sequencing primers (Table S1), the Big Dye Terminator V3.1 and an ABI 3500XL Genetic Analyzer (Applied Biosystem, Foster City, CA, USA) at the ISEM-Labex CEMEB sequencing facility (Montpellier University, France). The sequence chromatograms were checked by eye and the DNA fragments were assembled using the BioEdit v7.2.6.0 program (Hall, 1999). The strain information and the accession numbers of the obtained DNA sequences are provided in Table 1.

Due to the heterogeneous distribution of reference sequence data for the genus *Chaetoceros* concerning the 18S rDNA, the ITS region and the D1-D3 region of 28S rDNA, their sequencing data were used in separate phylogenetic analyses. The 28S rDNA phylogenetic analysis focused on the D1-D2 region because many reference sequences lack the D3 domain, and due to the importance of the D1-D2 rDNA barcode for phylogenetic studies (Grzebyk *et al.*, 2017). Reference sequences were selected by

BLASTN similarity analyses (Altschul *et al.*, 1990) using the web interface NCBI BLAST (Johnson *et al.*, 2008) and the GenBank nucleotide database. Alignments were generated with CLUSTAL X 2.1 (Larkin *et al.*, 2007) and were refined by eye using the BioEdit program (Hall, 1999). Phylogenetic analyses were performed with the online application Phylogeny.fr (Dereeper *et al.*, 2008) run by the ATGC bioinformatics facility (http://www.atgc-montpellier.fr/). The "A la Carte" mode was used with the corrected alignment, in which the phylogenetic analysis pipeline implemented PHYML 3.0 (Guindon *et al.*, 2010), using the HKY85 substitution model and four categories of substitution rates, with the Gamma distribution parameter, the proportion of invariable sites and the transition/transversion ratio estimated by the program. The estimation of branch support in the phylogenetic tree was statistically tested with the approximate likelihood-ratio test (Anisimova & Gascuel, 2006).

Results

Cell morphology and ultrastructure

The description of strain CT16ED isolated from Diana Lagoon was based entirely on cultured material. Solitary cells were the dominant form in culture. The shape of solitary cells was square to rectangular in girdle view, and the pervalvar axis was longer than the apical axis (Fig. 1). The girdle was composed of several stacked copulae (Fig. 2). Valves were circular to slightly elliptical, with a diameter of 3-6 µm in valve view (Figs 3-6). Valves had a central rimoportula that looked like a short flattened tubular process that extended a slit-shaped opening through the valve wall (Figs 2-5), or sometimes, showing a longer tube protruding from the valve (Fig. 6). The valve wall was ribbed with a dendritic structure made of radial, slightly branched costae that extended into the narrow hyaline valve rim on the marginal ridge (Figs 4-5) and in the mantle (Fig. 2). Valve costae could become thicker and nearly join (Fig. 3).

Thin setae, up to 25-30 µm in length, emerged at the valve corners and were oriented with a 45° angle relative to the pervalvar and apical axes, although an angle deviation could sometimes be observed (Fig. 1). In valve view, setae laid in the apical plane (Fig. 4). Setae (Fig. 7) were circular in cross-section (i.e. somehow cylindrical) with a ~320 nm diameter that narrowed towards the tip (Fig. 1). Setae were composed of six thin

longitudinal, helically twisted, silica costae shaped like strings (Fig. 8) but sometimes, their number varied from five to seven (Figs 11-12). These strings were ~50 nm in diameter. They were separated by a gap of ~100 nm, and were interconnected by tiny transverse costae perpendicularly to the seta longitudinal axis (Figs 8-9). These transverse costae were separated by a gap of ~45 nm, thus forming a slit-shaped poroid structure; the thickness of these costae was about half the slit width (Fig. 9). Setae were decorated by small, shark fin-shaped 60-80 nm-long spines attached to strings and pointed towards the seta tip (Figs 7-8). Setae emerged from the cell valve as full smooth silica tubes that slightly narrowed before forming the structure made of strings joined by transverse costae (Figs 7, 10). The seta diameter increased due to the string formation and the slit-shaped poroid structure (Fig. 10). Strings were tightly twisted upon formation (Figs 7-10) with the helical winding becoming less and less tight towards the seta tip. Towards the seta tip, the seta structure looked like fraying (Fig. 1). The disconnection of the string assemblage and transverse costae was observed (Figs 11-12).

Sometimes, a new, supernumerary seta was observed (Fig. 13). This seta formed through a budding process that took place in the hyaline rim of the valve marginal ridge, extending inside a sheath (Figs 14-15).

Growth, cell division, and chain formation

Chain formation was observed in culture during the growth phase. Inside chains, the cells exhibited variable shapes. Terminal cells, generally, had a well-shaped cylindrical body, shaped after the terminal valve. The body shape of internal cells varied between well-shaped cylindrical and deformed (Figs 16-17). The aperture between two cylindrical cells was elliptical with a width up to 1.5 µm, and the intercalary sibling valves generally lacked a central rimoportula (Fig. 18). However, apertures with a central rimoportula on intercalary sibling valves were also observed (Fig. 19).

Daughter cell production through frustule separation has been observed (Figs 20-22). After the generation of new setae, the transverse shearing of the mantle showed the formation and separation of two new valves (Figs 20, 24). Then, the widening of the aperture by the separation of the newly formed sibling valves ending in the central area, resulted in the parting of the two daughter cells (Fig. 21-22).

In the middle of an elongated cell, the formation of a transverse groove was sometimes observed and the emergence of a new bifid seta (Fig. 23). In other elongated cells, a constriction appeared in the soft girdle in the middle of the cell length (Fig. 24). The splitting of a mother cell was observed by shearing across the girdle (Fig. 25, also in Figs 16-17). Solitary cells were often observed with a single valve and pair of setae (Figs 26-27). Similar solitary cells with a single valve showed a regenerated seta from the side of the missing valve (Fig. 28). Some cells showed a soft and rounded valve without rimoportula (Fig 29). In some elongated cells, a cell extremity was terminated by a valve and with a regular cylindrical and smooth shape whereas the opposite half-cell looked shrivelled and shrinked (Figs 30-31).

In growth monitoring experiments, the daily growth rate of the CT16ED strain increased from day 2 to day 5, and peaked (1.2-1.4 div d^{-1}) at day 5-6 (Fig. 32A). Concomitantly with the growth acceleration, the percentage of single cells decreased from 80% at day 0 to 43% at day 4-6, when growth rate was highest, whereas the percentage of two-cell chains nearly doubled (up to 30.6%) (Fig. 32B). On inoculation day, the longest chains had only three cells. Then, chains gradually elongated and the part of cells in chains reached 26.3%. The longest chains observed at day 6 included 11 cells (1% of the whole cell population). Moreover, during strain CT16ED culture maintenance, chains with >20 cells were commonly observed in the growth phase. Conversely, in stationary-phase cultures, the cell population included mainly solitary cells.

At Ostend harbour, the phytoplankton natural community sampled for the isolation of *C. tenuissimus* strains mostly contained solitary cells with a small proportion of two-cell chains. In the growth phase in culture, the isolated strains displayed a variety of phenotypic behaviours (Table 1): solitary cells with rare cell pairs, mostly solitary cells with rare short chains (often up to four cells), and large proportion of chains of variable lengths (sometimes longer than 20 cells).

Molecular identification and phylogenetic analysis

In all analysed strains putatively assigned to *C. tenuissimus* (i.e. the Corsican strain CT16ED, the Japanese strain NIES-3715, and the nine strains from Ostend), the size of the rDNA sequence comprising the 18S rDNA, the ITS region and the D1-D3 region of

the 28S rDNA was 3346 base pairs (bp) between the two external PCR primer-binding sites. This was shorter than the size (3497 bp) of the same sequence from the *C. simplex* var. *calcitrans* strain CCAP1085/3 (accession number MK331990). This difference was due to numerous stretches of inserted nucleotides, mostly localized in the ITS1 and ITS2 regions, in the *C. simplex* var. *calcitrans* sequence. A preliminary phylogenetic analysis of the genus *Chaetoceros* using the sequencing data of the D1–D2 region of 28S rDNA (Fig. S2) indicated that the *C. simplex* var. *calcitrans* strain CCAP1085/3 sequence belonged to a clade that contained also *Chaetoceros gracilis* Pantocsek (accession number JQ217338), and that both sequences were suitable as outgroup for the phylogenetic analysis of *C. tenuissimus* strains and related *Chaetoceros* sequences.

In the phylogenetic trees obtained using the sequencing data of the three rDNA regions (18S, ITS region, and D1-D2 region) (Figs 33-35), the sequences generated in the present study gathered with other strains previously identified as *C. tenuissimus* and with unidentified strains from various tropical and temperate marine locations. In the three phylogenetic analyses, the clade with sequences obtained from strains identified as *C. tenuissimus* was identified as a sister clade to a large clade that included the *Chaetoceros neogracilis* (F.Schütt) VanLandingham species complex. This large clade included a third group with the strains CCMP189 and CCMP190 (renamed as *Chaetoceros* sp. in Balzano *et al.*, 2017) in the trees based on the D1-D2 and ITS region sequences (Figs 34-35), and the X85390/EU090012 cluster in the 18S rDNA-based tree (Fig. 33).

The 18S rDNA analysis (Fig. 33) revealed that the sequences of the analysed strains and the reference sequences belonging to the *C. tenuissimus* complex were almost all identical, with rare differences (e.g. nucleotide substitutions) in some reference sequences (Fig. S3), similar to what observed in the close *C. neogracilis* complex.

The phylogenetic analysis based on the D1-D2 region sequences (Fig. 34) divided the *C. tenuissimus* complex in three clades (I, II and III). The nine strains from Ostend harbour were categorized in Clade I (4 strains) and Clade II (5 strains). Clade I was subdivided in three groups among which two groups harboured one specific nucleotide substitution (at positions 435 and 513) (Fig. S4). These three groups contained sequences from numerous strains from the Gulf of Naples (Italy; Mediterranean Sea) and a set of environmental barcodes from four French Mediterranean lagoons (accession

number MK193876). The D1-D2 sequence of strain CT16ED (accession number MK331989) harboured two unique nucleotide substitutions and differed from other sequences in clade I by 2-3 substitutions (i.e. the two unique substitutions and one additional substitution at position 435 or 513), sharing > 99.4% of identity (Fig. S4). Clade II comprised five strains isolated from Ostend harbour with identical sequences. They differed by 16-18 nucleotides from the Clade I sequences (2.7-3.1% of difference), and by 14 nucleotides (2.4% of difference) with the NIES-3715 sequence in Clade III. Although included in Clade II, the sequence EF423470 presented 17 polymorphic positions that corresponded, but for one, to nucleotide substitutions between Clade II and Clade I (Fig. S4). Clade III sequences, represented by the Japanese strain NIES-3715, differed from Clade I sequences by 9-10 nucleotides (~1.5-1.7% of difference). Two Black Sea strains (V2 and V5) identified as *C. tenuissimus* were grouped in Clade III, but had shorter sequences resulting in a large number of deleted regions compared with the other *Chaetoceros* sequences (Fig. S4). Except for the two Black Sea strains, the D1-D2 sequence structure remained similar in the three main species clusters (*C. tenuissimus*, *C. neogracilis*, and *Chaetoceros* sp.), with a similar length and few inserted/deleted bases (not shown).

In the ITS hypervariable region, the aligned ITS1 and ITS2 sequences showed many cluster-specific indels and variable regions in the *C. neogracilis*, *Chaetoceros* sp. and *C. tenuissimus* clusters. In the phylogenetic analysis (Fig. 35), the *C. tenuissimus* complex was subdivided in the same three clades as in the D1-D2 analysis (5, 10 and 6 clade-specific nucleotide substitutions, respectively) (Fig. S5). Clade I exhibited again genetic variations among the included strains, even within the four Ostend strains. The Corsican CT16ED strain showed the highest genetic divergence within this clade. Conversely, the ITS sequences of the five Ostend strains in Clade II were identical. In Clade I, the partial sequences (5.8S and ITS2) of the *C. simplex* CCMP200 (Persian Gulf) and CCMP199 (Sargasso Sea) strains were identical to those of several strains (RCC4812, RCC4821 and RCC4826) from France (Fig. S5).

Overall, the amount of genetic variation was very low among the *C. tenuissimus* complex 18S rDNA sequences and was similar to that observed in the close *C. neogracilis* complex. Only in the 28S rDNA (D1-D2) and the ITS regions, genetic variation was higher compared with the close *C. neogracilis* complex (Figs. 33-35).

The phylogenetic analysis based on the *rbc*L gene (1427 bp in length between the PCR primer-binding sites) clearly separated the identical sequences of Ostend strains in Clade II (the only clade with five clade-specific substitutions in this gene) from a clade gathering the other strains that belonged to Clade I and Clade III and showed greater genetic variability (Fig. 36; Fig. S6). In Clade I, the four strains from Ostend harbour formed two genetically different pairs of identical sequences. The sequences of the four Mediterranean strains, including the Corsican CT16ED strains, were all different by few nucleotide changes, and one was identical to that of the *C. simplex* CCMP200 strain (whole plastid genome sequence KJ958479). Despite all these nucleotide substitutions, the *rbc*L amino acid sequences were identical in most *C. tenuissimus* strains, with the exception of the two Ostend strains in Clade I showing one amino acid change (Fig. S6).

Interestingly, our study revealed that the 18S rDNA and *rbc*L sequences (accession numbers AB246746 and AB246745) attributed to the diatom endosymbiont hosted by the dinoflagellate *Blixaea quinquecornis* (T.H.Abé) Gottschling (formerly *Peridinium quinquecorne*) were nearly identical to the 18S rDNA and *rbc*L sequences of strains belonging to Clade I of the *C. tenuissimus* complex (Figs. 33 and 36; Figs S3, S6).

Discussion

In the last decades, several studies have investigated the morphology and genetics of solitary *Chaetoceros* cells identified as *C. tenuissimus*, but without a comparison with a reference material from the type locality of this species, so that the modern definition of *C. tenuissimus* species after these studies could remain uncertain. In the continuity of these studies, the light microscopy observations of our CT16ED strain were also consistent with the rudimentary description of *C. tenuissimus* by Meunier (1913). The additional morphological observations and DNA sequences obtained from the CT16ED strain matched with material described in the literature and with reference DNA sequences (deposited in public databases) of various field samples and cultured strains that have been identified as *C. tenuissimus* (e.g. Sar *et al.*, 2002; Shirai *et al.*, 2008; Kooistra *et al.*, 2010). Subsequently, we demonstrated that, genetically, the CT16ED

strain is close to strains we isolated from the type locality (Ostend harbour) and that might include representatives of the specimens described by Meunier in 1913.

Morphology and biological observations

The solitary cell status of *C. tenuissimus* was established from field observations and was confirmed during sampling at Ostend harbour. However, culture of clonal strains demonstrated that single cells can form chains in this species. Other morphologically similar *Chaetoceros* species in the same size range exhibit solitary cell behaviour and also chain formation (with variable numbers of cells), including *C. salsugineus* Takano (Takano, 1983; Orlova & Selina, 1993; Trigueros *et al.*, 2002; Shevchenko *et al.*, 2006), *Chaetoceros fallax* Proshkina-Lavrenko (Takano, 1983), *C. neogracilis* (Balzano *et al.*, 2017) and *Chaetoceros similis* Cleve (Hernández-Becerril, 2009). The elongation of cell chains and the decreased proportion of solitary cells during the exponential growth phase of culture has been documented also in *C. salsugineus* (Orlova & Aizdaicher, 2000). Furthermore, even typically chain-forming species can have solitary stages, for instance *Chaetoceros socialis* Lauder (Meunier, 1913). Phylogenetically, the dominant solitary cell status is shared with *Chaetoceros* sp. (e.g. strain AnM002, Choi *et al.*, 2008) and *C. neogracilis*, the two clades close to *C. tenuissimus*, as documented for strains ArM004 and ArM005 (Choi *et al.*, 2008), and in clades I and II of the *C. neogracilis* species complex, in which short colonies have been observed in culture (Balzano *et al.*, 2017). As observed in our study, chain formation is shared by the three *C. tenuissimus* clades and cannot be used as a discriminating taxonomical feature.

The morphological comparison of *C. tenuissimus* CT16ED and *C. salsugineus* (Takano, 1983; Orlova & Selina, 1993; Trigueros *et al.*, 2002; Shevchenko *et al.*, 2006; Ichimi *et al.*, 2012) revealed many similar features. In both species, the seta structure is built with transverse costae connecting six twisted strings, originally named “ribs” by Takano (1983) in *C. salsugineus* then “costae” by Sar *et al.* (2003). The valve costae show a similar dendritic pattern. A rimoportula is generally present on both valves of solitary cells and on terminal valves in chains, and rimoportula can be present on sibling valves in the apertures between daughter cells inside chains, although seemingly less frequently than observed in *C. salsugineus* (Takano, 1983; Trigueros *et al.*, 2002). The rimoportula on terminal valves is also a feature in *C. neogracilis*, whereas this structure

has not been observed on intercalary valves inside chains (Balzano *et al.*, 2017). Variations in the aperture shape and width (in the apical plane) have been previously mentioned in descriptions of chain-forming species, including *Chaetoceros sporotruncatus* Gaonkar, Kooistra & Lange, *Chaetoceros dichatoensis* Gaonkar, Montresor & Sarno, *Chaetoceros cinctus* Gran and *Chaetoceros radicans* Schütt (Gaonkar *et al.*, 2017), and *C. neogracilis* (Balzano *et al.*, 2017). As observed in the present study, these variations might be related to the moment captured by the observation before the separation of daughter cells.

Our analysis indicated that in the *C. tenuissimus* strain CT16ED, the frustule is weakly silicified, as observed in *C. salsugineus* (Trigueros *et al.*, 2002), compared with other species. Our observations in culture suggested that cells might divide even before the reconstruction of a complete frustule. This explains the presence of soft cells in the middle of extending chains (Figs. 16-17) and might lead to extremely high experimental growth rates, as reported for *C. salsugineus* (Ichimi et al., 2012). The valve dendritic pattern was very weakly reticulated with thin costae, compared with the more densely reticulated valves described in other *Chaetoceros* species, for example in *Chaetoceros contortus* Schütt and *Chaetoceros debilis* Cleve (Kooistra *et al.*, 2010), *C. neogracilis* and *Chaetoceros decipiens* Cleve (Balzano *et al.*, 2017), *Chaetoceros gelidus* Chamnansinp, Y.Li, Lundholm & Moestrup (Chamnansinp *et al.*, 2013) and *Chaetoceros hirtisetus* (Rines & Hargreaves) Chamnansinp, Moestrup & Lundholm (Chamnansinp *et al.*, 2015), and *C. sporotruncatus*, *C. dichatoensis* and *C. cinctus* (Gaonkar *et al.*, 2017).

It has been proposed that some seta ultrastructural features can be used to identify and distinguish *Chaetoceros* species (Lee *et al.*, 2014a; 2014b). Consistently with the phylogeny results, the seta structure of the *C. tenuissimus* strain CT16ED resembled that of *C. neogracilis* (Balzano *et al.*, 2017) and also of *C. salsugineus* (Takano, 1983; Trigueros *et al.*, 2002), including the circular section with a similar diameter, the longitudinal helical string structure, the transverse costae and similar short spines, spirally arranged along the helical strings. Our observations suggested that setae are elongated from the extension of the basal tube, followed by the opening of slit-shaped poroids giving birth to strings and transverse costae that form the helical structure (Fig. 10), then the structure gradually loosens away from the base (Fig. 7). Towards the tip of

the seta, the aging of the structure can lead to its disintegration by the separation of the strings and transverse costae (Figs. 11-12). Observations also suggested that seta duplication during cell division begins from the extremities (from the bifid seta seen in Fig. 23) and progresses towards the tubular basis before separating, which would explain the finding of fused and crossed bases of setae (Fig. 21) before their complete separation to allow cell separation.

Compared with other species in the genus, *C. tenuissimus* setae are among the less silicified. Their thinness (low diameter) and the fine helical winding structure that loosens towards its extremity with the increasing spacing between the strings and transverse costae (Fig. 7) confer structural flexibility. Moreover, the setae are rather smooth, with a low number of small spines the spacing of which increases with the seta extension. Conversely, typically chain-forming species generally have thick and rigid setae (e.g. Kooistra *et al.*, 2010; Balzano *et al.*, 2017; Gaonkar *et al.*, 2017; Li *et al.*, 2017). These features might make setae less cohesive and therefore, chained *C. tenuissimus* cells (formed after cell division) can easily come off due to the shearing stress generated by small-scale turbulence in the water mass, leading to solitary cells in natural populations. This would explain why long chains were observed mainly in cultures grown in wells or without agitation.

Supernumerary setae, which have been observed in *C. similis* (Hernández-Becerril, 2009), might be related to cell division. Our microscopic analysis suggested two mechanisms of cell division and proliferation during blooms that might produce cells with two valves or with only one valve. Cell division in which the valve face regeneration precedes the separation of daughter cells would allow the proliferation of typical solitary cells with two well-shaped and silicified valves bearing a pair of setae (Fig. 1). Conversely, cells with only one valve and a pair of setae (as shown in Fig. 26) could result from cell division by splitting across the soft and weakly silicified girdle, and these cells could subsequently regenerate the missing setae (Fig. 28) and the lost opposite valve (Fig. 29) to reform regular solitary cells. This division process might be faster during blooms and we hypothesize that it could favour bloom developments. Pairs of sibling cells without a terminal valve (Figs 20-22), which might be formed after this second division process (Fig. 25), have been observed in several species (Hernández-Becerril, 1996). This raises the question of whether these observations might be

consistent with the detection of solitary cells in chain-forming *Chaetoceros* species that might represent a transient stage in the life history of these populations (Rines & Hargraves, 1988). Other peculiar cell shapes (Figs 28-29) were previously described in the solitary species *C. similis* (Hernández-Becerril, 2009). These shapes might represent seta or frustule regeneration in daughter cells following the shear splitting of mother cells.

Spores have not been detected in the *C. tenuissimus* species complex (Rines & Hargraves, 1988; Kooistra *et al.*, 2010) and in the close species *C. neogracilis* (Balzano *et al.*, 2017). Some of our observations suggested that a process of cellular retraction could take place in elongated cells (Figs 30-31). This might be an alternative to spore formation in conditions of nutrient starvation, by which cellular material from the shrinking part of the cell is recycled and stored within a smaller cell, possibly including silicon because frustule and setae appeared collapsing in the process. This might enhance the cell surviving capacity.

Phylogenetics, and genetic diversity of the C. tenuissimus complex

The three phylogenetic analyses highlighted similar topologies for the three rDNA regions (Figs 33-35), showing similar proportionally long genetic distances with high-latitude strains forming the *C. neogracilis* species complex from Arctic Ocean, and with a taxonomically poorly defined taxon identified as *Chaetoceros* sp. (following Balzano *et al.*, 2017) from the Southern Ocean.

Rines & Hargraves (1988) hypothesized the existence of cryptic species that correspond to the description of *C. tenuissimus* by Meunier (1913). Genetically, the (nearly) identical 18S rDNA sequences of the analysed strains could support a *C. tenuissimus* species complex, similarly to what found for other cosmopolitan species, including diatoms (for a review: Amato *et al.*, 2019), and for the *C. neogracilis* species complex. However, the genetic variations observed in the 28S rDNA D1-D2 and ITS sequences between the three clades within the *C. tenuissimus* complex were greater than variations observed between the four intraspecific clades of the *C. neogracilis* species complex. A larger set of strains representative of the *C. tenuissimus* complex must be studied to determine its overall genetic diversity and the worldwide distribution of different clades.

563 The range of intraspecific genetic variation in the D1-D3 region of the 28S rDNA
564 gene from the three clades was similar to what described for other *Chaetoceros* species,
565 such as *Chaetoceros elegans* Y.Li, Boonprakob, Moestrup & Lundholm and
566 *Chaetoceros laevisporus* Y.Li, Boonprakob, Moestrup & Lundholm (Li *et al.*, 2017).
567 Conversely, the same variation range observed in three *C. tenuissimus* clades was
568 sufficient to distinguish different *Chaetoceros* species in other studies, for example *C.*
569 *dichatoensis* and *C. sporotruncatus* the 18S rDNA sequences of which also differed by
570 10 nucleotides (Gaonkar *et al.*, 2017). Other cryptic species have been genetically
571 distinguished on the basis of small numbers of nucleotide differences in both D1-D3
572 and 18S rDNA sequences (Chen *et al.*, 2018; Xu *et al.*, 2019b). The possible existence
573 of numerous cryptic species in the genus *Chaetoceros* has been suggested by the genetic
574 differentiation of morphospecies from different geographic origins, while other widely
575 distributed species appear genetically identical (i.e. with identical rDNA sequences)
576 (Gaonkar *et al.*, 2018). Within Clade I, due to the ITS region and *rbc*L sequence
577 similarities of the *C. simplex* strain CCMP200 with various *C. tenuissimus* strains,
578 additional morphological and genetic studies are necessary to provide new evidence for
579 the debate about the synonymy between *C. tenuissimus* and *C. simplex*.
580 In the phylogenies inferred from each of the three rDNA markers, the *C. tenuissimus*
581 sequence data were clearly different from those of *C. simplex* var. *calcitrans*
582 CCAP1085/3 that was used as outgroup. The ITS sequence of strain CCAP1085/3 was
583 similar to that of other strains also identified as *C. simplex* var. *calcitrans* (e.g. strain
584 CSIRO CS251, accession number DQ358114) and as *C. calcitrans* (strain Arg13,
585 accession number DQ897644). Among the deposited ITS sequences, the reference
586 sequences of strains assigned to *C. calcitrans* f. *pumilus* are more genetically divergent
587 (data not shown), for instance from the strain CCAP1010/11 (accession number
588 DQ358117) or the strain CCMP1315 (accession DQ358111) sequences. Consequently,
589 the genetic data obtained for *C. simplex* var. *calcitrans*, *C. calcitrans*, and *C. calcitrans*
590 f. *pumilus*, which are currently considered as synonymous of *C. tenuissimus*, suggest
591 that these taxa are paraphyletic in the genus *Chaetoceros*, if the analysed strains were
592 correctly identified. A similar observation can be made when comparing the plastid
593 *rbc*L gene sequence data for strains assigned to the same taxa.

The analysis also identified *C. tenuissimus* as the putative origin of the dinoflagellate *B. quinquecornis* endosymbiont, the position of which within the genus *Chaetoceros* could not be accurately established due to the lack of close 18S rDNA and *rbc*L reference sequences (Horiguchi & Takano, 2006; Gottschling *et al.*, 2017). However, the ultrastructural analysis of *B. quinquecornis* showed a single endosymbiotic nucleus with about ten chloroplasts distributed in the periphery of the dinoflagellate cell (Horiguchi & Pienaar, 1991), whereas *C. tenuissimus* contains a single chloroplast. Furthermore, the endosymbiont chloroplast structure is different from that of a *C. tenuissimus* cell reported in Shirai *et al.* (2008). Hence, it cannot be excluded that both 18S rDNA and *rbc*L sequences have been accidentally obtained from an unknown contamination of the dinoflagellate sample by a *C. tenuissimus* cell. Therefore, additional genetic analyses of the endosymbiont nature of *B. quinquecornis* are needed. Alternatively, the nearly identical sequences between the endosymbiont and *C. tenuissimus* raise the question of whether the endosymbiont is permanently or only temporarily established inside the dinoflagellate host cell. In the latter case, the engulfment of the *C. tenuissimus* cell could be facilitated by its biological features (small size, weakly silicified frustule, and a possible transient cell stage that allows the protoplasm capture).

Finally, our genetic analysis stressed the advantage of sequencing different rDNA regions (18S, ITSs and 28S rDNA) using a strategy based on the assembly of overlapping PCR-amplified fragments, particularly for algal cultures the species purity of which is difficult to verify (e.g. chain-forming *Chaetoceros* species). Indeed, the BLAST analysis of *C. tenuissimus* 18S rDNA showed a 100% identity with sequences assigned to *Chaetoceros dayaensis* Y.Li & S.Zhu (accession numbers KM401853 and KM401854; release PLN 26-JUN-2015) (Li *et al.*, 2015). This result is inconsistent with the phylogenetic positioning of this species within another *Chaetoceros* clade on the basis of the 28S (D1-D3) rDNA sequences (accession numbers KM401851 and KM401852) obtained from the same two strains (Li *et al.*, 2015). This suggests that the two 18S rDNA sequences (accession numbers KM401853 and KM401854) might be from *C. tenuissimus*.

In conclusion, our study demonstrated that the morphological species *C. tenuissimus* genetically represents a species complex based on identical (or nearly identical)

sequences of the conserved 18S rDNA. This species complex contains several clades (at least three in this study) defined by the more variable rDNA regions (the D1-D2 region of 28S rDNA and the ITS region). Due to the morphological ultrastructural similarities of the specimens analysed here and on the basis of the available literature data, *C. tenuissimus* is hardly distinguishable from other species, such as *C. salsugineus* and *C. neogracilis*, including the fluctuation between solitary cells and chains. Therefore, it is uncertain that the different *C. tenuissimus* clades is morphologically distinguishable, and the Corsican strain CT16ED from the different strains from Ostend harbour. Nevertheless, due to the rDNA sequence differences between the CT16ED strain and the Ostend strains, it would be better to select an epitype of the species among strains from Ostend (North Sea), although it was very recently proposed a Spanish Mediterranean strain as an epitype of *C. tenuissimus* (Arin *et al.*, 2021). The question remains of which of the two genetic clades found in the type locality best represents the species described by Meunier.

Acknowledgments

Thanks to the Vlaams Institute of Marine Sciences (VLIZ) for its welcome and support at the Marine Station of Ostend (MSO) facility; thanks to MSO staff members for their help, especially to D. Cattrijsse and M. De Rijcke. Thanks to S. Pantani for assistance in the genetic analysis of the Ostend strains during a technician internship. The authors gratefully acknowledge S. Van Pé, M. Soleil and C. André (University of Corsica, UMS Stella Mare) and R. Pradelles (Microphyt) for assistance with the isolation and culture of microalgae.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was supported by the European Funds for Economic and Regional Development, the State-region program with the Corsican regional council “Collectivité de Corse” and the University of Corsica (Gerhyco program). The genetic work was carried out with the support of LabEx CeMEB, an ANR "*Investissements d'Avenir*"

programme (ANR-10-LABX-04-01), through the use of the GenSeq sequencing facility. The research leading to some of these results received partial funding from the European Union Horizon 2020 research and innovation programme under grant agreement number 730984, for the European consortium ASSEMBLE Plus project: specifically, the sampling in Ostend (Belgium) was supported by the Transnational Access programme through the project CHAETOTYPE. We acknowledge the anonymous reviewers for their helpful comments.

Supplementary information

Additional Supplementary Information may be found in the online version of this article at the publisher's web site:

Table S1: PCR and sequencing primers used in this study.

Fig. S1: Photography of the *C. tenuissimus* description drawings in Figure 55 from the article by Meunier (1913). The coin (diameter of 23mm) was used as a size scale of the published drawing. Photograph obtained by courtesy of the botanical library of the Muséum National d'Histoire Naturelle, Paris (France).

Fig. S2: Phylogenetic tree of the genus *Chaetoceros* inferred by maximum likelihood analysis, based on the sequences of the D1–D2 region in the 28S rDNA gene, with *Leptocylindrus aporus* and *Leptocylindrus convexus* (Chaetocerotophycidae, Leptocylindrales) as outgroups. The branch length is proportional to the number of substitutions per site (see scale bar at the bottom). Branch support (if posterior probability value > 0.5) is provided at the main nodes.

Fig. S3: Alignment of the *C. tenuissimus* 18S rDNA sequences and from *Chaetoceros* sp. strains. All sequence fragments begin after the forward PCR primer 18S-F used in this study and end at the binding site of the reverse PCR primer 18S-R (3' end of 18S rDNA, see the Table S1). The shown alignment is a subset of the one used for the 18S rDNA phylogenetic analysis (Fig. 33). All sequences, but one (*Chaetoceros* sp. CHMS01), are

identical, but for some insertions or deletions located in homopolymer repeats that might be sequencing errors. The *Chaetoceros* sp. CHMS01 sequence (accession number AF145226) differs only by one T substitution at position 543 in the alignment. The last sequence (accession AB246746), which was attributed to the diatom endosymbiont of the dinoflagellate *B. quinquecornis* (formerly *P. quinquecorne*), is identical to the *C. tenuissimus* sequences, except at the 3' end where the A substitution (circled in red) corresponds to a mismatch in the reverse PCR primer SR12b (rectangular black box) used by Horiguchi & Takano (2006).

Fig. S4: Alignment of the *C. tenuissimus* 28S (D1-D2) rDNA sequences. The alignment spans from the binding site of the forward primer D1R (5' end of 28S rDNA) to that of the reverse primer D2C (3' end of the D2 domain of 28S rDNA), used for sequencing in this study (see Table S1). The red arrows and circles at positions 109 and 476 indicate two unique substitutions in the CT16ED sequence. The alignment is a subset of the one used for the 28S rDNA phylogenetic analysis (Fig. 34).

Fig. S5: Alignment of the *C. tenuissimus* ITS region sequences and from *Chaetoceros* spp. strains. The alignment spans from the binding site of the reverse PCR primer 18S-R (3' end of 18S rDNA) to that of the forward primer D1R (5' end of 28S rDNA) used in this study (Table S1). The alignment is a subset of the one used for the ITS-based phylogenetic analysis (Fig. 35).

Fig. S6: RuBisCO large subunit (*rbc*L) gene. Alignments of the nucleotide sequences (sequence included between the used PCR primers), and the corresponding amino acid sequences. The sequences including the 1311 nucleotide long fragment (positions 59-1369) were used for the phylogenetic analysis (Fig. 36). The sequence with accession number AB246745, attributed to the diatom endosymbiont of the dinoflagellate *B. quinquecornis* (formerly *P. quinquecorne*), is also included. Most sequences code identical proteins, with the exception of two strains from Ostend harbour in Clade I (accession numbers: MZ189415 and MZ198421). The red arrows and circles indicate the A/T nucleotide substitution at position 1065 in the nucleotide alignment of these two

strains and the corresponding phenylalanine (F) to tyrosine (Y) substitution at position 355 of the amino acid alignment.

Author contributions

V. Pasqualini conceptualized the project, isolated the Corsican strain and performed cultures. D. Grzebyk isolated the Ostend strains in Belgium, performed the molecular work and the phylogenetic analyses, and contributed to the morphological description. M. Garrido microscopically determined the Corsican strain and described its morphology. Y. Quilichini performed scanning electron and transmission electron microscopy studies. C. Pereto performed growth experiments. P. Cecchi organized and supervised the activities. All authors equally contributed to the draft preparation and to its redaction.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, **215**: 403–410.

Amato, A., Kooistra, W.H.C.F. & Montresor, M. (2019). Cryptic diversity: a long-lasting issue for diatomologists. *Protist*, **170**: 1–7.

Andersen, R.A. (2005). *Algal culturing techniques*. Phycological Society of America, Elsevier Academic Press. Amsterdam. 578 pp.

Anisimova, M. & Gascuel, O. (2006). Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Systematic Biology*, **55**: 539–52.

Arin, L., Reñé, A., Gallisai, R., Sarno, D., Garcés, E, Estrada, M. (2021). Taxonomic relationship between two small-sized *Chaetoceros* species (Bacillariophyta): *C. tenuissimus* and *C. salsugineus*, and comparison with *C. olympicus* sp. nov. from Catalan coastal waters (NW Mediterranean). *Eur. J. Phycol.*, 1-19.

753 Balzano, S., Percopo, I., Siano, R., Gourvil, P., Chanoine, M., Marie, D., Vaulot, D. &
754 Sarno, D. (2017). Morphological and genetic diversity of Beaufort Sea diatoms with
755 high contributions from the *Chaetoceros neogracilis* species complex. *Journal of*
756 *Phycology*, **53**:161–187.
757
758 Baytut, Ö., Moestrup, Ø., Lundholm, N. & Gönülol, A. (2013). Contributions to the
759 Diatom flora of the Black Sea from ultrastructural and molecular studies: new
760 records of *Skeletonema marinoi*, *Pseudo-nitzschia pungens* var. *aveirensis* and
761 *Chaetoceros tenuissimus* for the marine flora of Turkey. *Nova Hedwigia*, **96**: 427–
762 444.
763
764 Bec, B., Collos, Y., Souchu, P., Vaquer, A., Lautier, J., Fiandrino, A., Benau, L.,
765 Orsoni, V. & Laugier, T. (2011). Distribution of picophytoplankton and
766 nanophytoplankton along an anthropogenic eutrophication gradient in French
767 Mediterranean coastal lagoons. *Aquatic Microbial Ecology*, **63**: 29–45.
768
769 Bérard-Therriault, L., Poulin, M. & Bossé, L. (1999). Guide d'identification du
770 phytoplancton marin de l'estuaire et du golfe du Saint-Laurent incluant également
771 certains protozoaires. Publication spéciale canadienne des sciences halieutiques et
772 aquatiques, Conseil National de Recherches du Canada, Ottawa, **128**: 1–387.
773
774 Chamnansinp, A., Li, Y., Lundholm, N. & Moestrup, Ø. (2013). Global diversity of two
775 widespread, colony-forming diatoms of the marine plankton, *Chaetoceros socialis*
776 (syn. *C. radians*) and *Chaetoceros gelidus* sp. nov. *Journal of Phycology*, **49**: 1128–
777 1141.
778
779 Chamnansinp, A., Moestrup, Ø. & Lundholm, N. (2015). Diversity of the marine
780 diatom *Chaetoceros* (Bacillariophyceae) in Thai waters – revisiting *Chaetoceros*
781 *compressus* and *Chaetoceros contortus*. *Phycologia*, **54**: 161–175.
782

Chen, Z.Y., Lundholm, N., Moestrup, Ø., Kownacka, J. & Li, Y. (2018). *Chaetoceros pauciramosus* sp. nov. (Bacillariophyceae), a widely distributed brackish water species in the *C. lorenzianus* complex. *Protist*, **169**: 615–631.

Choi, H.G., Joo, H.M., Jung, W., Hong, S.S., Kang, J.S. & Kang, S.H. (2008). Morphology and phylogenetic relationships of some psychrophilic polar diatoms (Bacillariophyta). *Nova Hedwigia*, **133**: 7–30.

Collier, A. & Murphy, A. (1962). Very small diatoms: preliminary notes and description of *Chaetoceros galvestonensis*. *Science*, **136**: 780–781.

De Luca, D., Kooistra, W.H.C.F, Sarno, D., Gaonkar C.C. & Piredda, R. (2019a). Global distribution and diversity of *Chaetoceros* (Bacillariophyta, Mediophyceae): integration of classical and novel strategies. *PeerJ*, **7**: e7410.

De Luca, D., Sarno, D., Piredda, R. & Kooistra, W.H.C.F. (2019b). A multigene phylogeny to infer the evolutionary history of Chaetocerotaceae (Bacillariophyta). *Molecular Phylogenetics and Evolution*, **140**: 106575.

Deasi, S.R., Verlecar, X.N., Ansari, Z.A., Jagtap, T.G., Sarkar, A., Vashistha, D. & Dalal, S.G. (2010). Evaluation of genotoxic responses of *Chaetoceros tenuissimus* and *Skeletonema costatum* to water accommodated fraction of petroleum hydrocarbons as biomarker of exposure. *Water Research*, **44**: 2235–2244.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.F., Guindon, S., Lefort, V., Lescot, M., Claverie, J.M. & Gascuel, O. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research*, **36**: W465–W469.

Gaonkar, C.C., Kooistra, W.H.C.F., Lange, C.B., Montresor, M. & Sarno, D. (2017). Two new species in the *Chaetoceros socialis* complex (Bacillariophyta):

C. sporotruncatus and C. *dichatoensis*, and characterization of its relatives, *C. radicans* and *C. cinctus*. *Journal of Phycology*, **53**: 889–907.

Gaonkar, C.C., Piredda, R., Minucci, C., Mann, D.G., Montresor, M., Sarno, D. & Kooistra, W.H.C.F. (2018). Annotated 18S and 28S rDNA reference sequences of taxa in the planktonic diatom family Chaetocerotaceae. *PLoS ONE*, **13**: e0208929.

Gottschling, M., Zerdoner Calasan, A., Kretschmann, J. & Gu, H. (2017). Two new generic names for dinophytes harbouring a diatom as an endosymbiont, *Blixaea* and *Unruhdinium* (Kryptoperidiniaceae, Peridiniales). *Phytotaxa*, **306**: 296–300.

Grzebyk, D., Audic, S., Lasserre, B., Abadie, E., de Vargas, C. & Bec, B. (2017). Insights into the harmful algal flora in northwestern Mediterranean coastal lagoons revealed by pyrosequencing metabarcodes of the 28S rRNA gene. *Harmful Algae*, **68**: 1–16.

Guillard, R.R.L. (1973). Division rates. In *Handbook of Phycological Methods. Cultures methods and growth measurements* (Stein, J. R., editor), 290-311. Cambridge University Press.

Guillard, R.R.L. (1975). Culture of phytoplankton for feed marine invertebrates. In *Culture of Marine Invertebrate Animals* (Smith, W. L. & Chanley, M. H., editors), 29–60. Plenum Press, New York.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, **59**: 307–321.

Guiry, M.D. (2019). Genus *Chaetoceros*. In AlgaeBase (Guiry, M.D., & Guiry, G.M., editors). Available at: http://www.algaebase.org (last accessed 12 June 2019).

Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**: 95–98.

Härnström, K., Karunasagar, I. & Godhe, A. (2009). Phytoplankton species assemblages and their relationship to hydrographic factors - a study at the old port in Mangalore, coastal Arabian Sea. *Indian Journal of Marine Sciences*, **38**: 224–234.

Hasle, G.R. & Syversten, E.E. (1997). Marine diatoms. In *Identifying Marine Phytoplankton* (Tomas, C.R., editor), 5–385. Academic Press, San Diego.

Hernández-Becerril, D.U. (1996). A morphological study of *Chaetoceros* species (Bacillariophyta) from the plankton of the Pacific Ocean of Mexico. *Bulletin of natural History Museum of London (Botany)*, **26**: 1–73.

Hernández-Becerril, D.U. (2009). Morphological variability of the marine planktonic diatom *Chaetoceros similis* (Bacillariophyceae). *Cryptogamie, Algologie*, **30**: 125–134.

Horiguchi, T. & Pienaar, R.N. (1991). Ultrastructure of a marine dinoflagellate, *Peridinium quinquecorne* Abé (Peridiniales) of South Africa with particular reference to its chrysophyte endosymbiont. *Botanica Marina*, **34**: 123–131.

Horiguchi, T. & Takano, Y. (2006). Serial replacement of a diatom endosymbiont in the marine dinoflagellate *Peridinium quinquecorne* (Peridiniales, Dinophyceae). *Phycological Research*, **54**: 193–200.

Hustedt, F. (1930). Die Kieselalgen Deutschlands, Österreichs und der Schweiz. In *Dr. L. Rabenhorst's Kryptogamen-Flora von Deutschlands, Österreichs und der Schweiz* 7. Akademische Verlags-gesellschaft, Leipzig, pp. 1–920.

Ichimi, K., Kawamura, T., Yamamoto, A., Tada, K. & Harrison, P.J. (2012). Extremely high growth rate of the small diatom *Chaetoceros salsugineum* isolated from an estuary in the eastern Seto Inland Sea, Japan. *Journal of Phycology*, **48**: 1284–1288.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S. & Madden, T.L. (2008). NCBI BLAST: a better web interface. *Nucleic Acids Research*, **36**: W5–W9.

Kaczmarska, I., Samanta, B., Ehrman, J.M. & Porcher, E.M.A. (2019). Auxosporulation in *Chaetoceros acadianus* sp. nov. (Bacillariophyceae), a new member of the Section *Compressa*. *European Journal of Phycology*, **54**: 206–221.

Kooistra, W.H.C.F., Sarno, D., Hernández-Becerril, D.U., Assmy, P., Di Prisco, C. & Montresor, M. (2010). Comparative molecular and morphological phylogenetic analyses of taxa in the Chaetocerotaceae (Bacillariophyta). *Phycologia*, **49**: 471–500.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. & Higgins, D.G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**: 2947–2948.

Leblanc, K., Arístegui, J., Armand, L., Assmy, P., Beker, B., Bode, A., Breton, E., Cornet, V., Gibson, J., Gosselin, M.-P., Kopczynska, E., Marshall, H., Peloquin, J., Piontkovski, S., Poulton, A. J., Quéguiner, B., Schiebel, R., Shipe, R., Stefels, J., van Leeuwe, M. A., Varela, M., Widdicombe, C. & Yallop, M. (2012). A global diatom database – abundance, biovolume and biomass in the world ocean. *Earth System Science Data*, **4**: 149–165.

Lee, S.D., Park, J.S., Yun, S.M. & Lee, J.H. (2014a). Critical criteria for identification of the genus *Chaetoceros* (Bacillariophyta) based on setae ultrastructure. I. Subgenus *Chaetoceros*. *Phycologia*, **53**: 74–87.

907 Lee, S.D., Joo, H.M. & Lee, J.H. (2014b). Critical criteria for identification of the genus
908 *Chaetoceros* (Bacillariophyta) based on setae ultrastructure. II. Subgenus
909 *Hyalochaete*. *Phycologia*, **53**: 614–638.
910
911 Li, Y., Lundholm, N., Moestrup, Ø. (2013). *Chaetoceros rotosporus* sp. nov.
912 (Bacillariophyceae), a species with unusual resting spore formation. *Phycologia*, **52**:
913 500-508.
914
915 Li, Y., Zhu, S., Lundholm, N. & Lü, S. (2015). Morphology and molecular phylogeny
916 of *Chaetoceros dayaensis* sp. nov. (Bacillariophyceae), characterized by two 90°
917 rotations of the resting spore during maturation. *Journal of Phycology*, **51**: 469–479.
918
919 Li, Y., Boonprakob, A., Gaonkar, C.C., Kooistra, W.H.C.F., Lange, C.B., Hernández-
920 Becerril, D., Chen, Z., Moestrup, Ø. & Lundholm, N. (2017). Diversity in the
921 globally distributed diatom genus *Chaetoceros* (Bacillariophyceae): three new
922 species from warm-temperate waters. *PLoS ONE*, **12**: e0168887.
923
924 Lund, J.W.G., Kipling, C. & Le Cren, E.D. (1958). The inverted microscope method of
925 estimating algal numbers and the statistical basis of estimations by counting.
926 *Hydrobiologia*, **11**: 143-170.
927
928 Malviya, S., Scalco, E., Audic, S., Vincent, F., Veluchamy, A., Poulain, J., Wincker, P.,
929 Iudicone, D., de Vargas, C., Bittner, L., Zingone, A., Bowler, C., 2016. Insights into
930 global diatom distribution and diversity in the world's ocean. *Proceedings of the*
931 *National Academy of Sciences*, **113**: E1516-E1525.
932
933 Meunier, A. (1913). Microplankton de la Mer Flamande. 1ère partie: Le genre
934 "*Chaetoceros*" Ehr. *Mémoires du Musée Royal d'Histoire Naturelle de Belgique* 7:1–
935 58. (Available for download at the Library of the Royal Belgian Institute of Natural
936 Sciences: biblio.naturalsciences.be/rbins-publications/memoires/tome-7/1913-
937 0265ddf.pdf)
938

939 Montresor, M., Di Prisco, C., Sarno, D., Margiotta, F. & Zingone, A. (2013). Diversity
940 and germination patterns of diatom resting stages at a coastal Mediterranean site.
941 *Marine Ecology Progress Series*, **484**: 79–95.
942
943 Orlova, T.Y. & Aizdaicher, N.A. (2000). Development in culture of the diatom
944 *Chaetoceros salsugineus* from the Sea of Japan. *Russian Journal of Marine Biology*,
945 **26**: 8–11.
946
947 Orlova, T.Y. & Selina, M.S. (1993). Morphology and ecology of the bloom-forming
948 planktonic diatom *Chaetoceros salsugineus* Takano in the Sea of Japan. *Botanica*
949 *Marina*, **36**:123-130.
950
951 Paulsen, O. (1905). On some Peridineae and plankton diatoms. *Meddelelser fra*
952 *Kommissionen for Havundersøgelser Series Plankton*, **1**: 1–7.
953
954 Rines, J.E.B. & Hargraves, P.E. (1988). The *Chaetoceros* Ehrenberg
955 (Bacillariophyceae) flora of Narragansett Bay, Rhode Island, USA. *Bibliotheca*
956 *Phycologica*, **79**: 1–196. J. Cramer, Berlin.
957
958 Sakka Hlaili, A., Grami, E.B., Mabrouk, H.H., Gosselin, M. & Hamel, D. (2007).
959 Phytoplankton growth and microzooplankton grazing rates in a restricted
960 Mediterranean lagoon (Bizerte Lagoon, Tunisia). *Marine Biology*, **151**: 767–783.
961
962 Sar, E.A., Hernández-Becerril, D.U. & Sunesen, I. (2002). A morphological study of
963 *Chaetoceros tenuissimus* Meunier, a little-known planktonic diatom, with a
964 discussion of the section Simplicia, subgenus Hyalochaete. *Diatom Research*, **17**:
965 327–335.
966
967 Shevchenko, O.G., Orlova, T.Y. & Hernandez-Becerril, D.U. (2006). The genus
968 *Chaetoceros* (Bacillariophyta) from Peter the Great Bay, Sea of Japan. *Botanica*
969 *Marina*, **49**: 236–258.
970

971 Shirai, Y., Tomaru, Y., Takao, Y., Suzuki, H., Nagumo, T. & Nagasaki, K. (2008).
972 Isolation and characterization of a single-stranded RNA virus infecting the marine
973 planktonic diatom *Chaetoceros tenuissimus* Meunier. *Applied and Environmental*
974 *Microbiology*, **74**: 4022–4027.
975
976 Takano, H. (1968). On the diatom *Chaetoceros calcitrans* (Paulsen) emend. and its
977 dwarf form *pumilus* forma nov. *Bulletin of Tokai Regional Fisheries Research*
978 *Laboratory*, **55**: 1–7.
979
980 Takano, H. (1983). New and rare diatoms from Japanese marine waters, 10. A new
981 *Chaetoceros* common in estuaries. *Bulletin of Tokai Regional Fisheries Research*
982 *Laboratory*, **110**: 1–11.
983
984 Tomaru, Y., Kimura, K. & Yamaguchi, H. (2014). Temperature alters algicidal activity
985 of DNA and RNA viruses infecting *Chaetoceros tenuissimus*. *Aquatic Microbial*
986 *Ecology*, **73**: 171–183.
987
988 Tomaru, Y., Toyoda, K. & Kimura, K. (2018). Occurrence of the planktonic bloom-
989 forming marine diatom *Chaetoceros tenuissimus* Meunier and its infectious viruses
990 in western Japan. *Hydrobiologia*, **805**: 221–230.
991
992 Toyoda, K., Nagasaki, K. & Tomaru, Y. (2010). Application of real-time PCR assay for
993 detection and quantification of bloom-forming diatom *Chaetoceros tenuissimus*
994 Meunier. *Plankton Benthos Research*, **5**: 56–61.
995
996 Trigueros, J.M., Orive, E. & Arriluzea, J. (2002). Observations on *Chaetoceros*
997 *salsugineus* (Chaetocerotales, Bacillariophyceae): first record of this bloom-forming
998 diatom in a European estuary. *European Journal of Phycology*, **37**: 571–578.
999
1000 Werner, D. (1977). Introduction with a note on taxonomy. In *The biology of diatoms,*
1001 *vol. 13* (Werner, D., editor), 1–23. Blackwell, Oxford, UK.
1002

Xu, X.J., Chen, Z.Y., Lundholm, N., Li, Y. (2018). Revisiting *Chaetoceros subtilis* and *C. subtilis* var. *abnormis* (Bacillariophyceae), reinstating the latter as *C. abnormis*. *Phycologia*, **57**: 659-673.

Xu, X.J., Chen, Z.Y., Lundholm, N. & Li, Y. (2019a). Diversity in the section *Compressa* of the genus *Chaetoceros* (Bacillariophyceae), with description of two new species from Chinese warm waters. *Journal of Phycology*, **55**: 104–117.

Xu, X.J., Lundholm, N., Chen, Z.Y., Li, Y. (2019b). Revisiting section *Compressa* of *Chaetoceros* (Bacillariophyceae), with descriptions of *C. brevispinosus* sp. nov. and *C. ornatus* comb. nov. *Phycologia*, **58**: 614-627.

Xu, X., Lundholm, N., Li, Y. (2020). A study of *Chaetoceros debilis* sensu lato species (Bacillariophyceae), with ememdation of *C. debilis* and description of *C. galeatus* sp. nov. *Journal of Phycology*, **56**: 784-797.

Figure captions

Figs 1-6. Solitary cells of the *Chaetoceros tenuissimus* strain CT16ED. SEM (Figs 1-3, 6) and TEM (Figs 4-5) micrographs. **Fig. 1.** Elongated solitary cell in girdle view; note the seta flexibility that allows bending of the seta extremities. **Fig. 2.** Girdle view showing the stacked copulae (white arrows); note the protruding rimoportula (white arrowhead) and the striation in the mantle due to the extension of the valve costae (black arrowhead). **Fig. 3.** Valve with a central slit-shaped rimoportula (white arrowhead); radial costae appear thickened and nearly joining. **Fig. 4.** The valve dendritic structure made of slightly branched costae that radiate from the central rimoportula (white arrowhead). **Fig. 5.** Typical rimoportula structure consisting of a flattened tube (black arrowhead). **Fig. 6.** Long rimoportula tube (white arrowhead) protruding from the centre of the valve. Scale bars: Fig. 1, 10 µm; Figs 2-6, 1 µm.

Figs 7-15. Seta structure in the *Chaetoceros tenuissimus* strain CT16ED. TEM (Figs 7-14) and SEM (Fig. 15) photographs. **Fig. 7.** Seta made of thin longitudinal, helically twisted costae, like silica strings. The winding decreases towards the seta tip, increasing the distance between spines. **Fig. 8.** Seta structure with six silica strings that bear shark fin-shaped spines (white arrowheads) pointing towards the seta tip. **Fig. 9.** Detail of the seta structure with twisted strings connected by thin transverse costae perpendicularly to the seta axis that form slit-shaped poroids. Their regular spacing gives a grid appearance. **Fig. 10.** Tubular basis of a seta from which the helically twisted structure originates and where transverse costae start to form with the opening of slit-shaped poroids. **Fig. 11.** Disassembled seta with seven twisted strings lacking transverse costae. **Fig. 12.** Disassembled seta extremity with five twisted strings. **Fig. 13.** Valve with a supernumerary seta. Two setae are close to each other on the same valve side (white arrows). **Fig. 14.** A new, supernumerary seta emerging by budding (white arrowhead), close to another seta on the same valve side. **Fig. 15.** Extending regenerated seta (arrow); note the whip-like extremity made of free longitudinal strings (white arrowhead). Scale bars: Figs 9-10, 100 nm; Figs 8, 11-12, 200 nm; Fig. 7, 500 nm; Figs 13-15, 1 µm.

Figs 16-22. Chain formation in the *Chaetoceros tenuissimus* strain CT16ED; SEM photographs. **Fig. 16.** Short chain in which both terminal cells (arrowheads) have a terminal valve with a rimoportula and a cylindrical, well-silicified mantle. The two cells within the chain have a soft body cell. In the lower part of the image, the terminal cell seems to split in two by shearing of the soft girdle (double arrow). **Fig. 17.** Short chain in which intercalary cells with a well-shaped valve (arrowheads) alternate with cells with a soft body. In the centre, a soft cell undergoing division with the formation of a segmentation groove (arrow). The cell underneath seems to split in two by shearing of the soft girdle (double arrow). **Fig. 18.** Aperture between daughter cells in which sibling intercalary valves lack a rimoportula. White arrow, fusion point of the sibling setae. **Fig. 19.** Aperture between daughter cells in which both sibling intercalary valves have a rimoportula (arrowheads). **Fig. 20.** During cell division at the early stage of construction of new sibling valves, shearing of the mantle (arrow) is observed as the daughter cells start to separate. **Fig. 21.** Intermediate stage in the separation of daughter cells showing the beginning of the aperture formation. Two setae are duplicated and diverge from the same tubular base (arrow). **Fig. 22.** Late stage in the separation of daughter cells showing the yet unachieved opening of the aperture and regenerated sibling valves still attached in their central area. The setae of daughter cells remain fused at their base (arrow). Scale bars: Figs 16-17, 5 µm; Figs 18-22, 1 µm.

Figs 23-31. *Chaetoceros tenuissimus* strain CT16ED. SEM (Figs 23-26, 28-31) and TEM (Fig. 27) photographs. **Fig. 23.** Early cell division: formation of a segmentation groove (arrow), and a bifid seta during regeneration of the new setae in the daughter cells (arrowhead). **Fig. 24.** Sequential segmentation of a very elongated cell: in its middle (arrows), cell division is undergoing with construction of two sibling valve faces and the regeneration of new setae. In the middle of the two budding elongated cells, the soft girdle (double arrows) is constricted, preparing for splitting by shearing across the girdle. **Fig. 25.** Splitting of an elongated cell into daughter cells with the nearly achieved shearing of the soft girdle (double arrow), to free a solitary cell with a single pair of setae. **Fig. 26.** Solitary cell with a single pair of setae, formed from the mother cell after splitting by shearing across the girdle. **Fig. 27.** View of the valve of a solitary cell formed after splitting: the girdle shearing occurred along the copula suture. **Fig. 28.**

Solitary cell, formed by splitting, with a regenerated seta (arrow). **Fig. 29.** In valve view, a solitary cell with a soft and rounded terminal valve without rimoportula and with three setae **Fig. 30.** Elongated solitary cell in which one half (chevron arrow) is shrinking, whereas the opposite half displays a well-shaped, cylindrical valve. **Fig. 31.** Small chain showing cells with a cylindrical valve on one side, and either with a softened body (white arrows) or with a shrunken aspect (chevron arrow) on the opposite side.
Scale bars: Figs 23, 27, 29, 1 µm; Figs 24-26, 28, 30, 5 µm; Fig. 31, 3 µm.

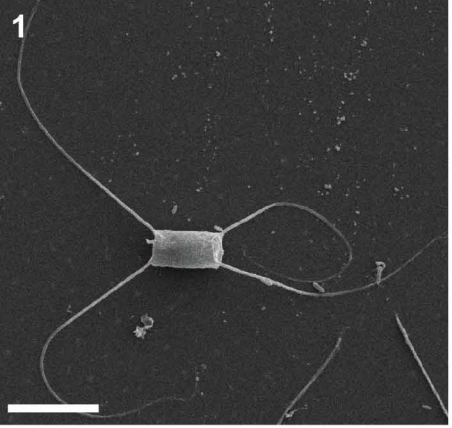
Fig. 32. *In vitro* growth of CT16ED cultures under a 12:12 h light-dark photoperiod. (A) Growth curve (full line) and variation in daily growth rate k (dashed line); error bars represent the standard deviation of triplicate experiments. (B) Variation in the proportion of solitary cells and chain-forming cells during the growth monitoring experiment.

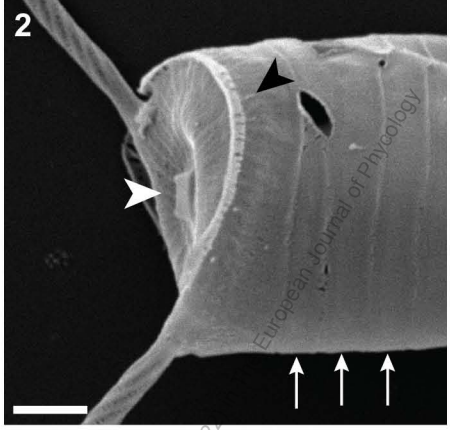
Fig 33. Phylogenetic tree based on the 18S rDNA sequence showing the position of the *C. tenuissimus* species complex relative to the closest taxa within the *Chaetoceros* genus, using *C. gracilis* and *C. simplex* var. *calcitrans* (strain CCAP1085/3) as near outgroups. The 18S rDNA sequences are identified by their accession number; species name and strain name, and geographic origin (when known) are shown. The strains sequenced in this study are in bold. The branch length is proportional to the number of substitutions per site (the scale bar represents the number of nucleotide substitutions per site). Branch support values (if posterior probability value > 0.5) are provided at the main nodes.

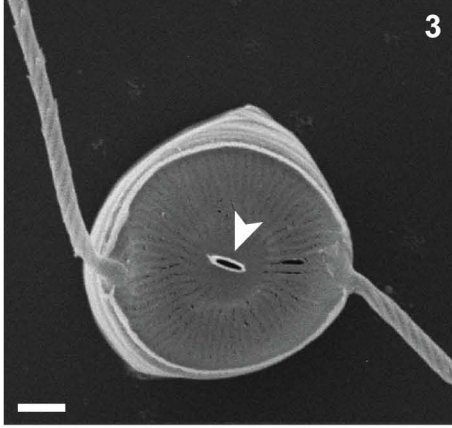
Fig. 34. Phylogenetic tree based on the sequence of the D1-D2 region of 28S rDNA. Information on the strains, branch length and support, are as described in the legend to Fig. 33.

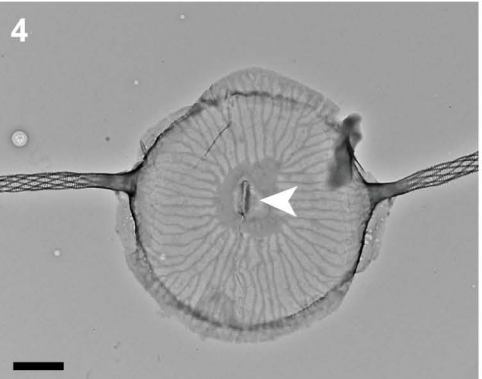
Fig. 35. Phylogenetic tree based on the ITS region sequence (ITS1, 5.8S rDNA and ITS2). In the box, an unrooted tree for *C. tenuissimus* complex alone is provided to

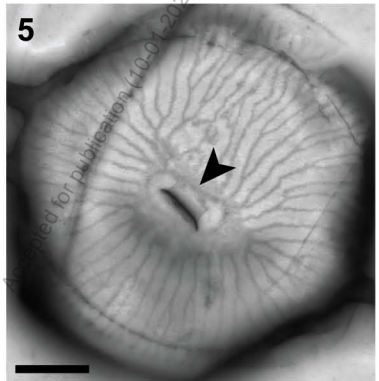
1113 better show the genetic distances between clades and strains. Strain information, and
1114 branch length and support are as described in the legend to Fig. 33.
1115
1116 **Fig. 36.** Phylogenetic tree based on the sequence of the *rbc*L gene in the *C. tenuissimus*
1117 species complex (partial sequences for 1311 nucleotides). Strain information, and
1118 branch length and support are as described in the legend to Fig. 33.

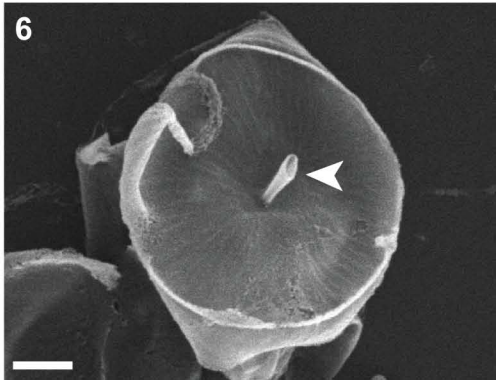


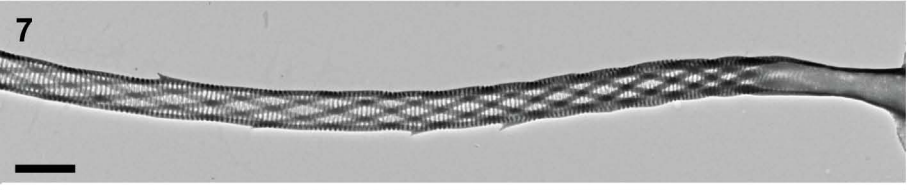


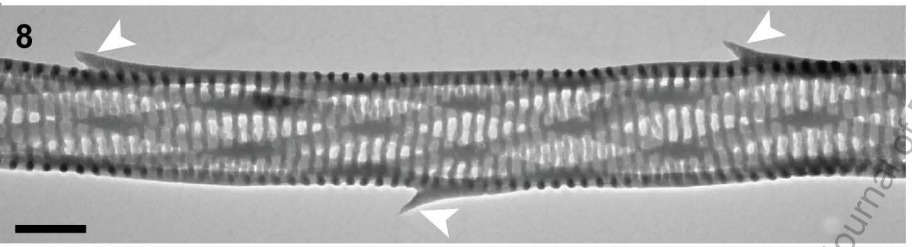


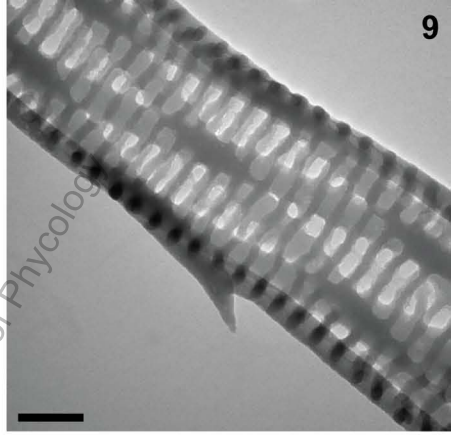


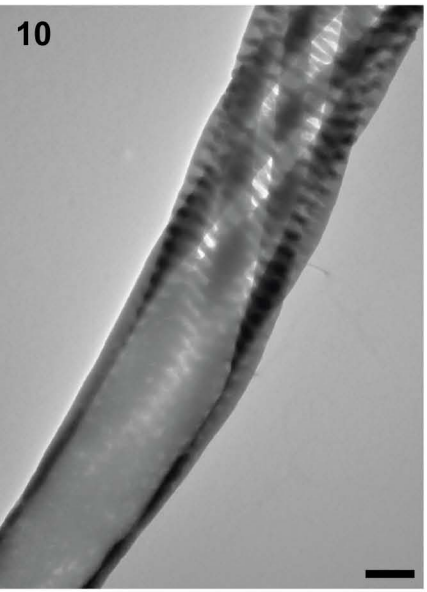


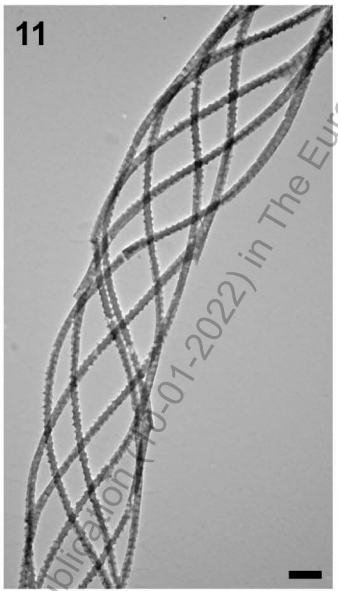


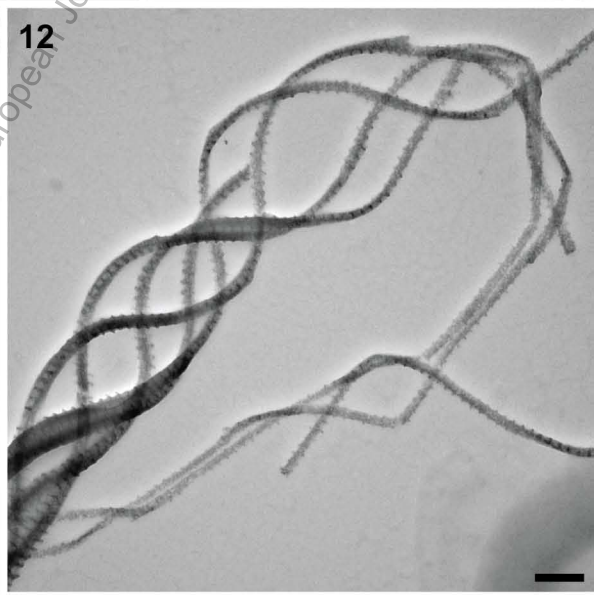


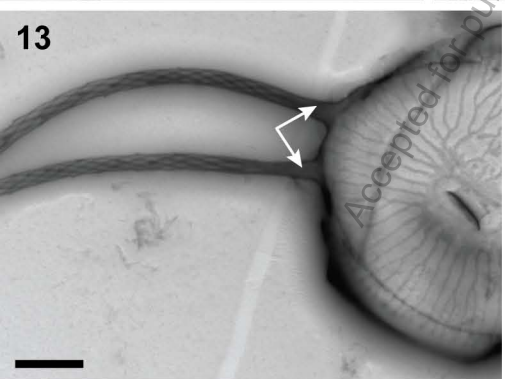


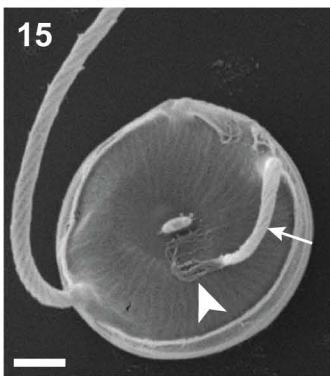


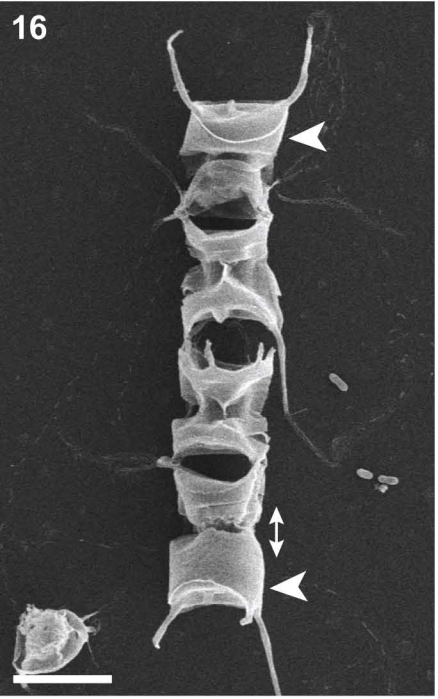


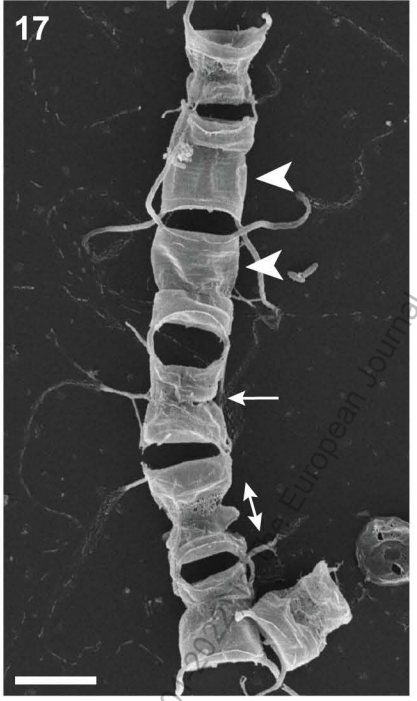


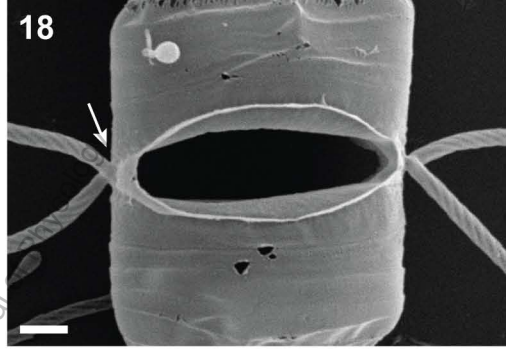


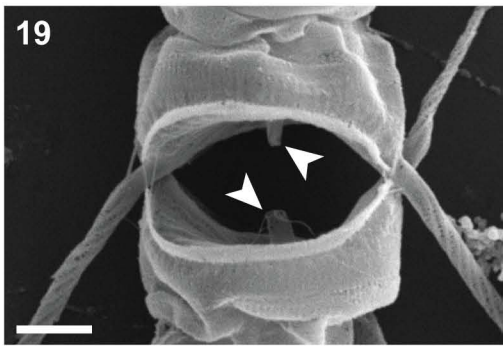


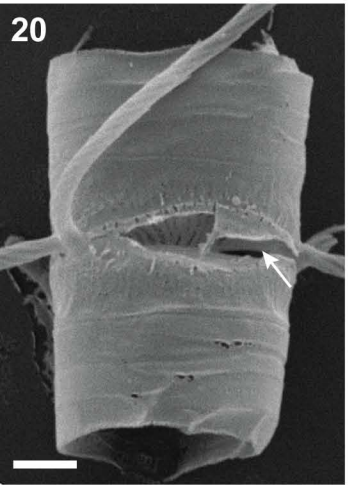


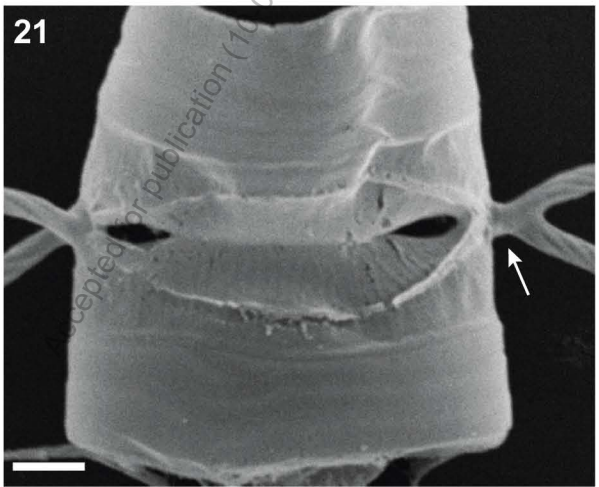


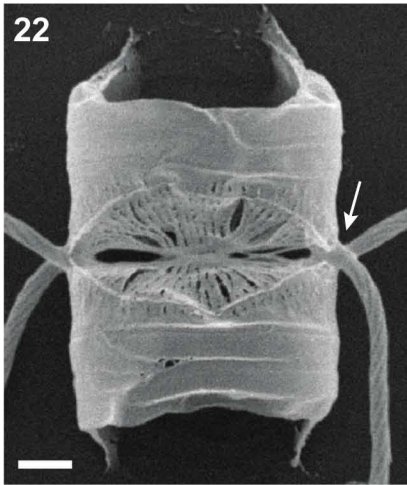


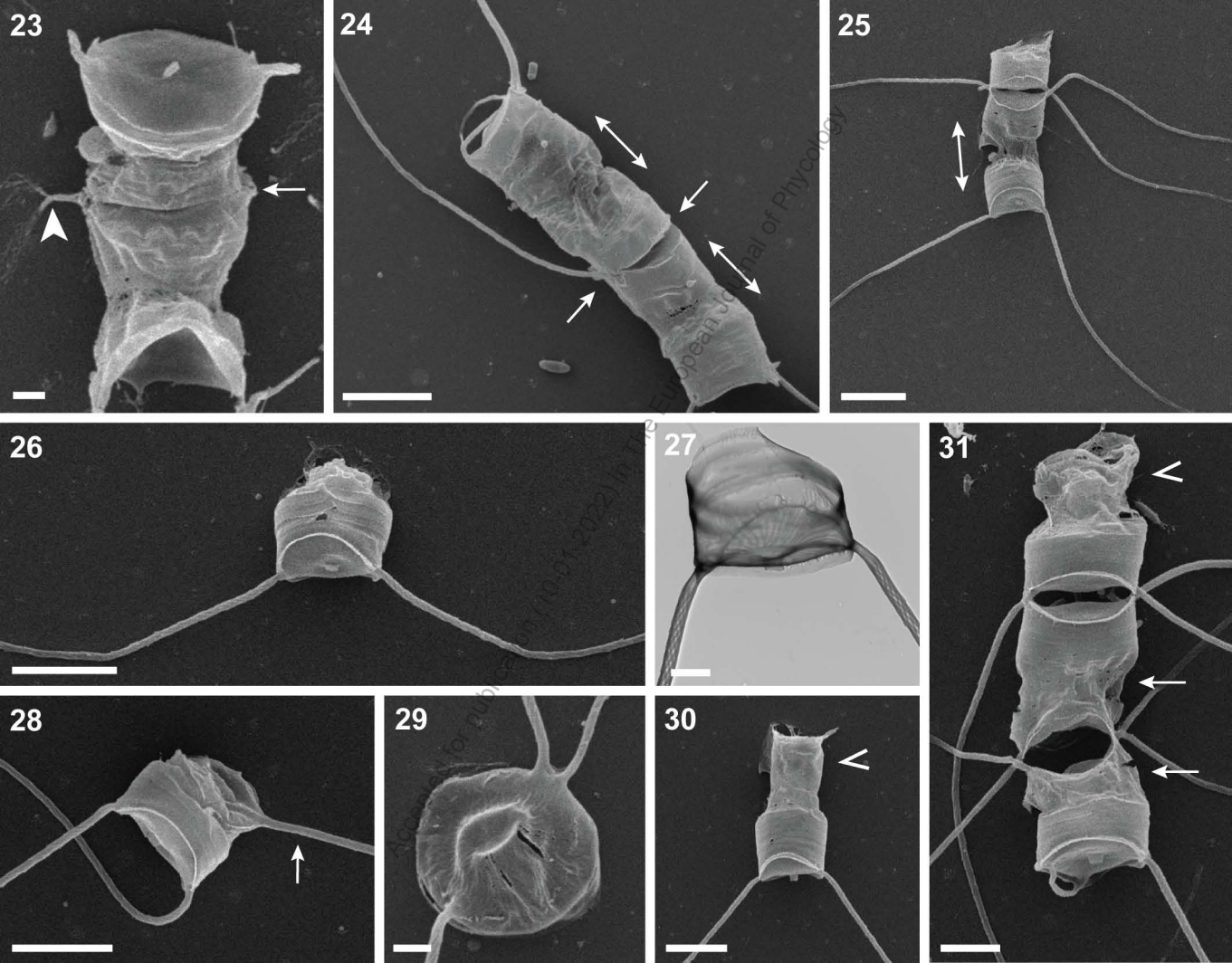


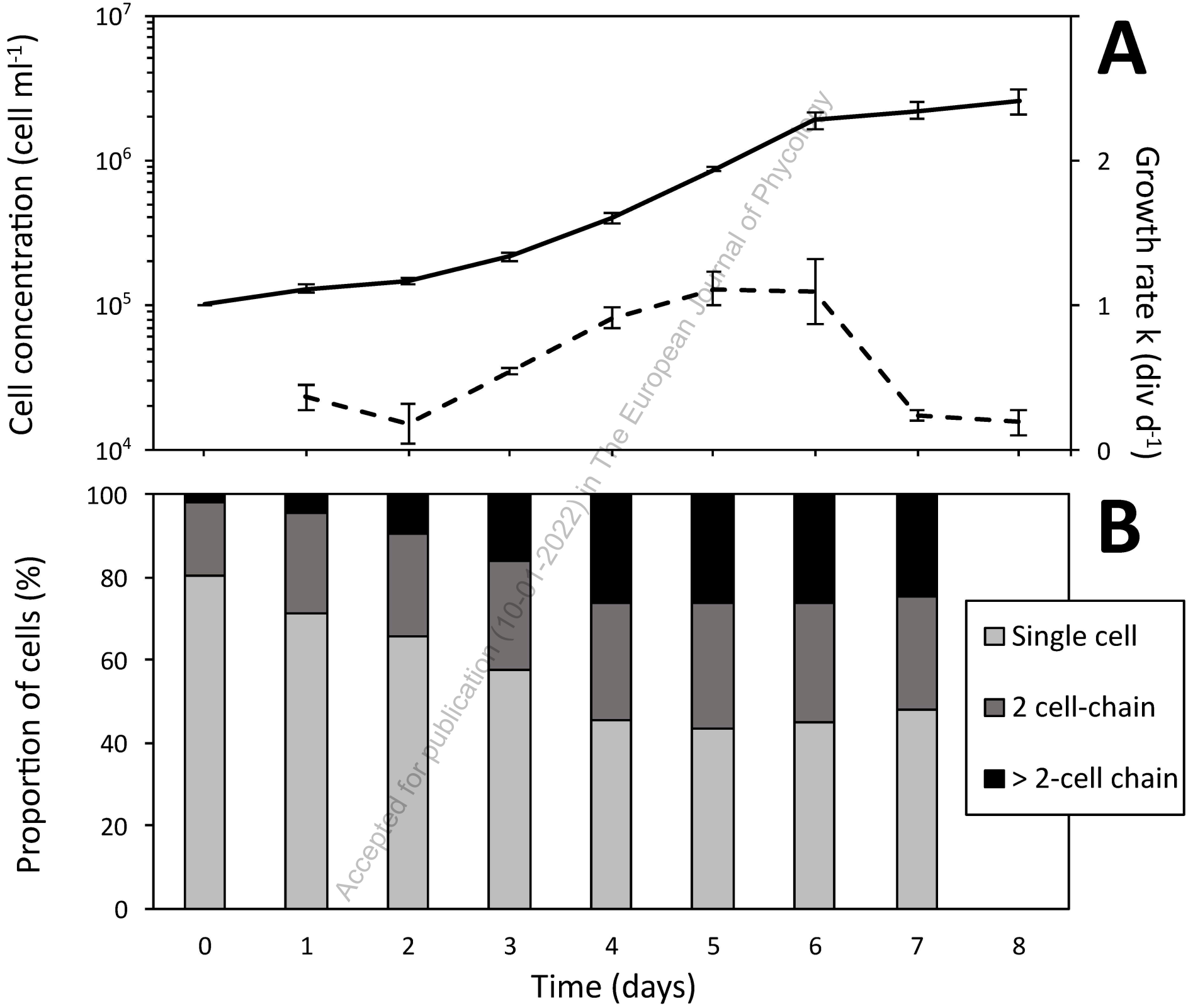


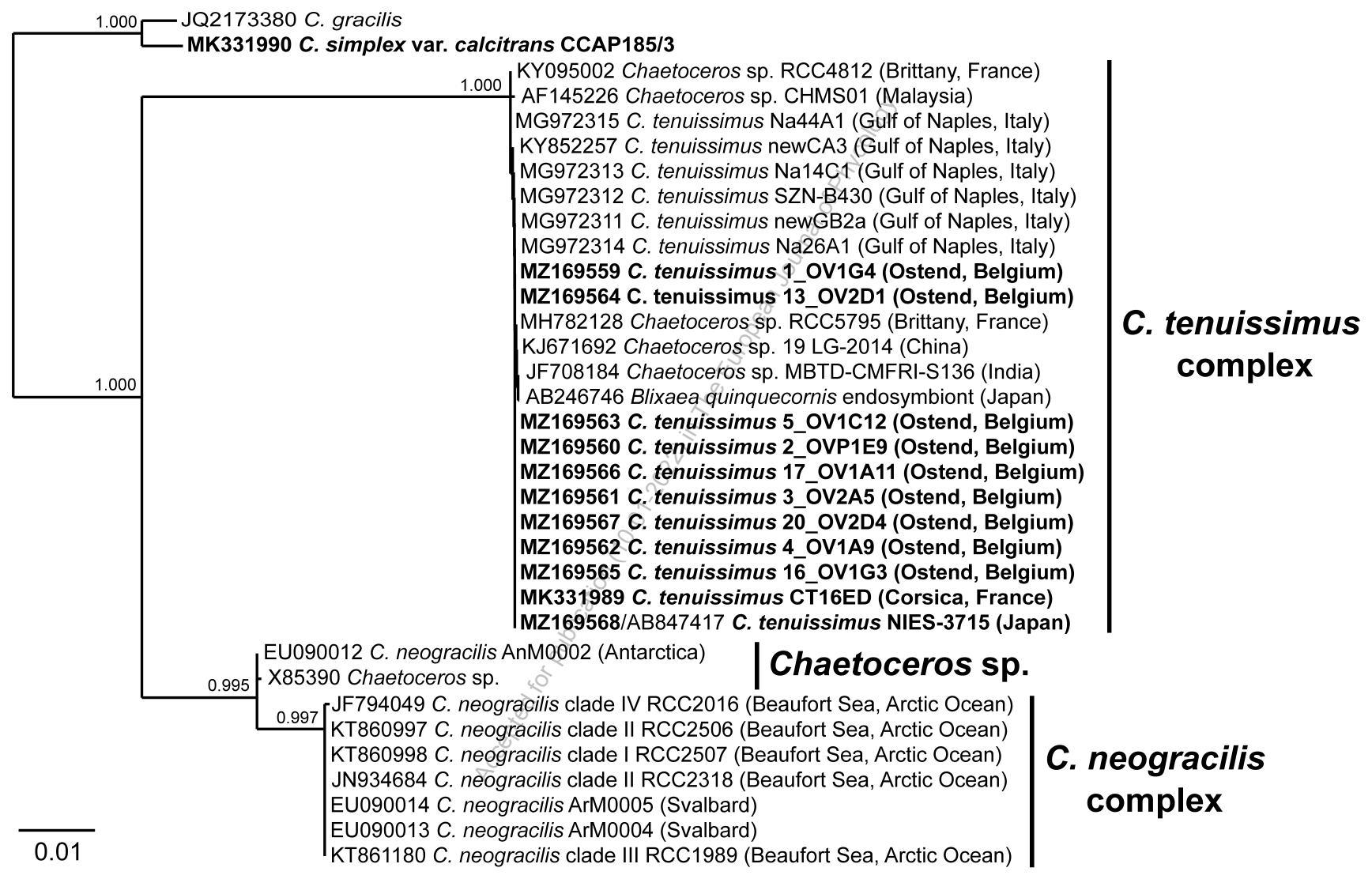


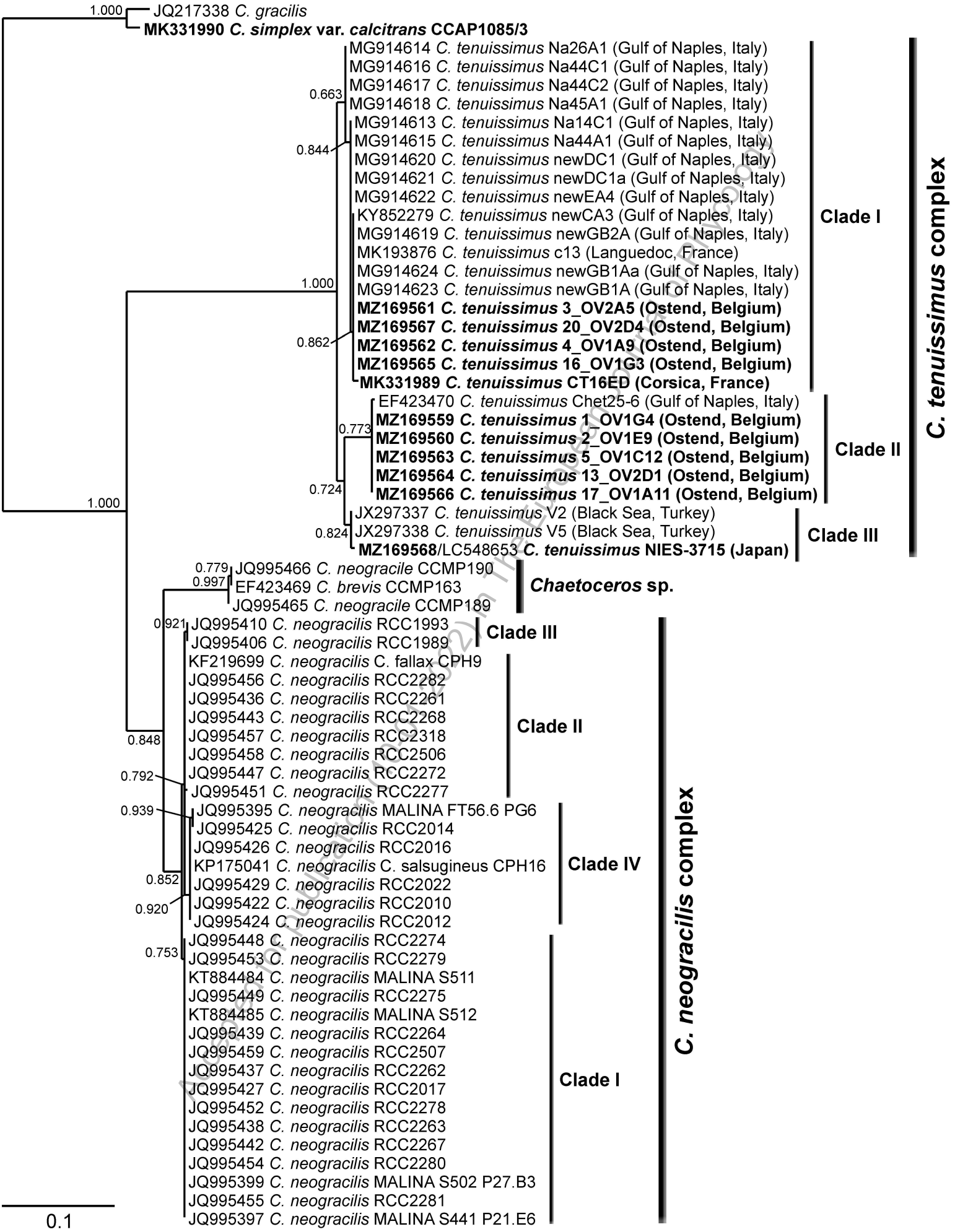


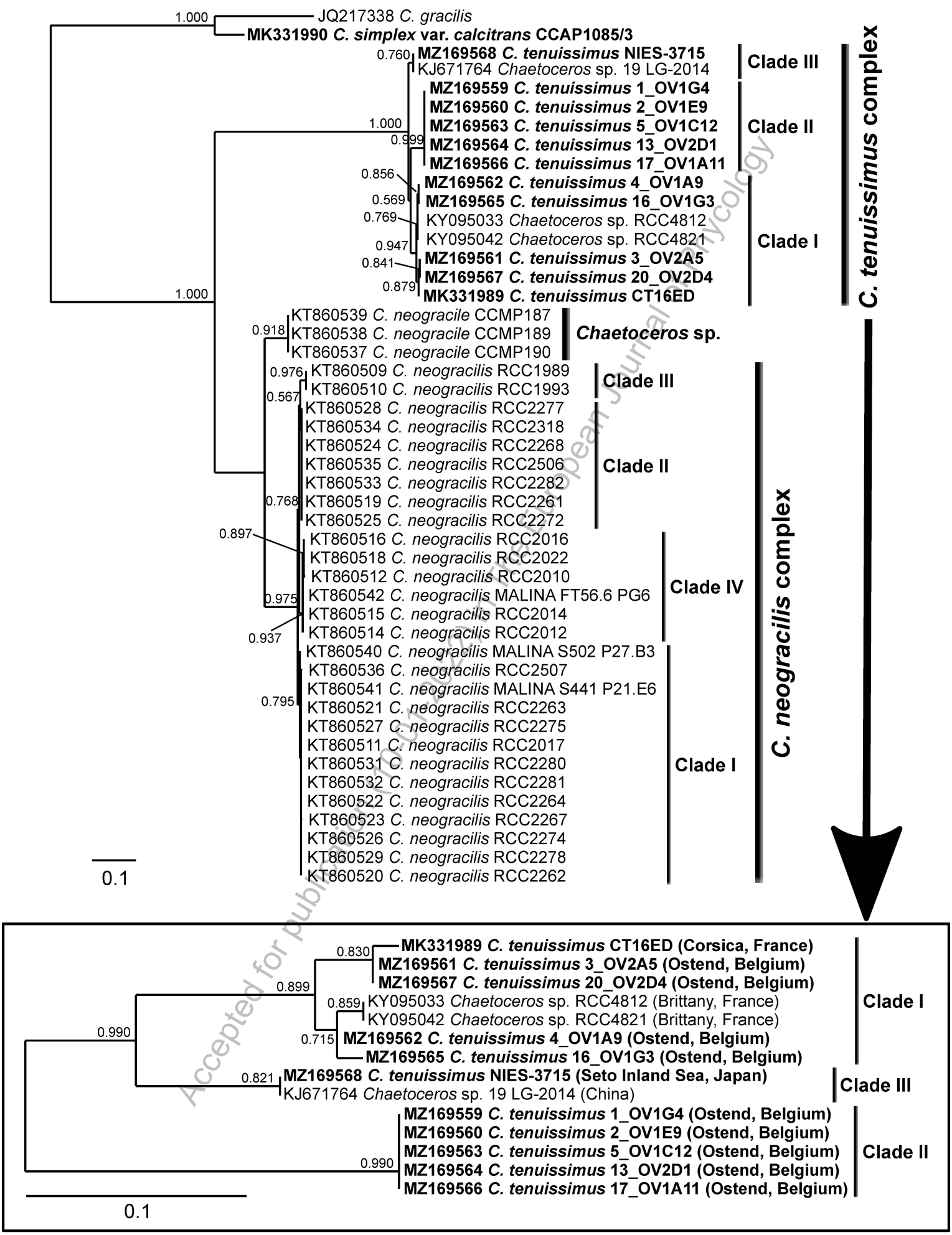





A
B
Cell concentration (cell ml-1)
Growth rate k (div d-1)
Proportion of cells (%)
Time (days)
Single cell
2 cell-chain
> 2-cell chain
Accepted for publication (10-01-2022) in The European Journal of Phycology







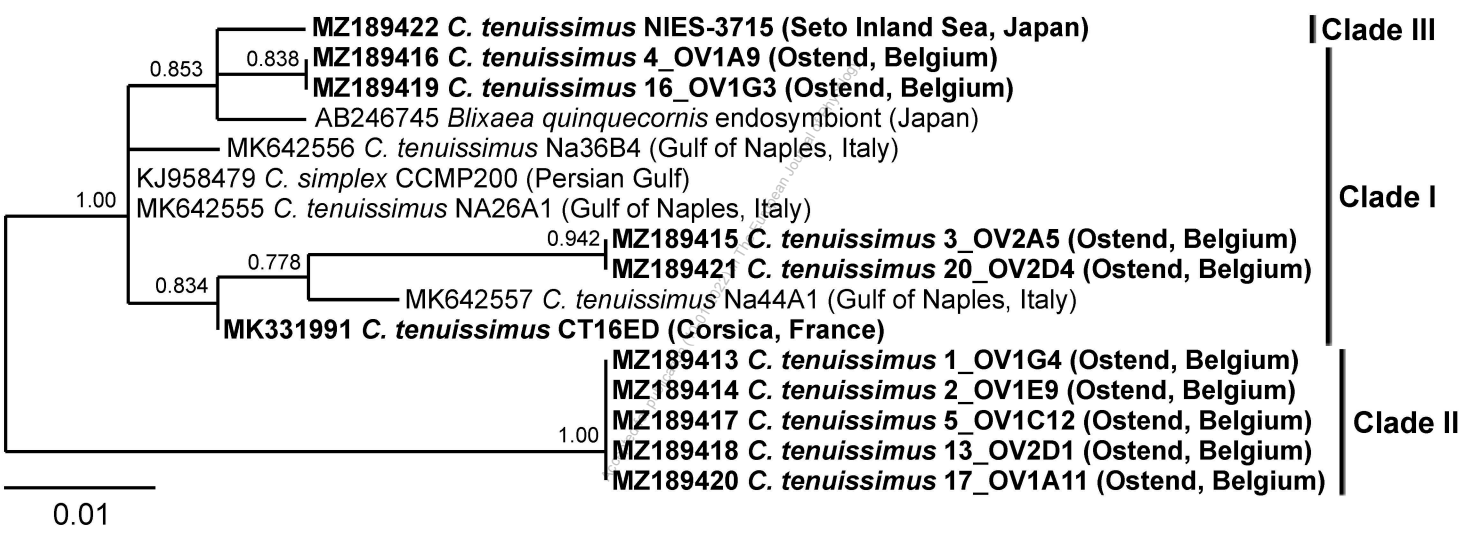

MZ189422 C. tenuissimus NIES-3715 (Seto Inland Sea, Japan)
MZ189416 C. tenuissimus 4_OV1A9 (Ostend, Belgium)
MZ189419 C. tenuissimus 16_OV1G3 (Ostend, Belgium)
AB246745 Blixaea quinquecornis endosymbiont (Japan)
MK642556 C. tenuissimus Na36B4 (Gulf of Naples, Italy)
KJ958479 C. simplex CCMP200 (Persian Gulf)
MK642555 C. tenuissimus NA26A1 (Gulf of Naples, Italy)
MZ189415 C. tenuissimus 3_OV2A5 (Ostend, Belgium)
MZ189421 C. tenuissimus 20_OV2D4 (Ostend, Belgium)
MK642557 C. tenuissimus Na44A1 (Gulf of Naples, Italy)
MK331991 C. tenuissimus CT16ED (Corsica, France)
MZ189413 C. tenuissimus 1_OV1G4 (Ostend, Belgium)
MZ189414 C. tenuissimus 2_OV1E9 (Ostend, Belgium)
MZ189417 C. tenuissimus 5_OV1C12 (Ostend, Belgium)
MZ189418 C. tenuissimus 13_OV2D1 (Ostend, Belgium)
MZ189420 C. tenuissimus 17_OV1A11 (Ostend, Belgium)
0.853
0.838
1.00
0.942
0.778
0.834
1.00
0.01
Clade III
Clade I
Clade II

Table 1. List of *Chaetoceros* strains analyzed in this study, with the accession numbers of obtained rDNA and rbc*L* gene sequences.

***Chaetoceros* species**	**Strain name**	**Origin and isolation date**	**Morphology**	**Accession numbers**	
			Solitary vs. forming chain	**18S to 28S D3 rDNA**	***rbc*L**
C. tenuissimus	CT16ED	Diana Lagoon, Corsica, France; May 2016	Can form chains	MK331989	MK331991
	1_OV1G4	Vuurtoren Dock, Ostend, Belgium; 29 July 2020	Solitary	MZ169559	MZ189413
	2_OV1E9	Vuurtoren Dock, Ostend, Belgium; 30 July 2020	Solitary, rare 4-cell chains	MZ169560	MZ189414
	3_OV2A5	Vuurtoren Dock, Ostend, Belgium; 30 July 2020	Mostly solitary	MZ169561	MZ189415
	4_OV1A9	Vuurtoren Dock, Ostend, Belgium; 30 July 2020	Solitary	MZ169562	MZ189416
	5_OV1C12	Vuurtoren Dock, Ostend, Belgium; 30 July 2020	Solitary	MZ169563	MZ189417
	13_OV2D1	Vuurtoren Dock, Ostend, Belgium; 30 July 2020	Can form chains	MZ169564	MZ189418
	16_OV1G3	Vuurtoren Dock, Ostend, Belgium; 29 July 2020	Can form chains	MZ169565	MZ189419
	17_OV1A11	Vuurtoren Dock, Ostend, Belgium; 30 July 2020	Can form chains	MZ169566	MZ189420
	20_OV2D4	Vuurtoren Dock, Ostend, Belgium; 30 July 2020	Can form chains	MZ169567	MZ189421
	NIES-3715	Seto Inland Sea, Japan; August 2002	Can form chains	MZ169568	MZ189422
C. simplex var. *calcitrans*	CCAP1085/3	Not available	Mostly solitary	MK331990	MK331992