1 **Bioinformatic prediction of putative metallothioneins in non-ciliate protists**

2 Sergio Balzano[1,2*], Angela Sardo[1,3]

3 [1] Stazione Zoologica Anton Dohrn Napoli (SZN), Naples, Italy

4 [2] NIOZ Royal Netherlands Institute for Sea Research, Den Burg, Netherlands

5 [3] Istituto di Scienze Applicate e Sistemi Intelligenti – CNR, via Campi Flegrei 34, 80078 Pozzuoli,

6 Naples, Italy.

7

8 * Corresponding author: sergio.balzano@szn.it

9

Running title: metallothioneins in non-ciliate protists

Abstract

Intracellular ligands that bind heavy metals (HMs) minimising their detrimental effects to cellular metabolism, are attracting great interest for a number of applications including bioremediation and development of HM-biosensors. Metallothioneins (MTs) are short, cysteine-rich, genetically encoded proteins involved in intracellular metal-binding and play a key role in HMs detoxification. We searched ~700 genomes and transcriptomes of non-ciliate protists for novel putative MTs by similarity and structural analyses and found 21 unique proteins playing a potential role as MTs. Most putative MTs derive from heterokonts and dinoflagellates and share common features such as (1) a putative metal-binding domain in proximity of the N-terminus, (2) two putative MT-specific domains near the C-terminus and (3) one to three CTCGXXCXCGXXCXCXXC patterns. Although the biological function of these proteins has not been experimentally proven, knowledge on their genetic sequences adds useful information on proteins potentially involved in HM-binding hypothetical and can contribute designing future biomolecular assays on HM-microbe interactions and MT-based biosensors.

Keywords: Heavy metals, pollution, metallothioneins, non-ciliate protists

Introduction

Microorganisms inhabiting heavy metal (HM)-contaminated environments, eventually incorporating contaminants within the cell, are biotechnologically interesting because of their potential use for bioremediation (Kumar et al. 2015). Passive adsorption of cations onto cell walls and transport across cell membrane are the two major mechanisms of HM uptake by living cells (Das et al. 2008, Blaby-Haas and Merchant 2012). Subsequently, intracellular polypeptides such as enzymatically-produced phytochelatins and genetically-encoded metallothioneins (MTs), limit the detrimental effect of HMs by complexing and transporting them towards vacuoles, chloroplast, or mitochondria (Cobbett and Goldsbrough 2002, Perales-Vela et al. 2006).

MTs are low-molecular weight proteins exhibiting a low content of aromatic amino acids and high proportions of cysteine residues (≥10%), they have been characterised in deep details in multicellular organisms (Blindauer and Leszczyszyn 2010, Leszczyszyn et al. 2013) as well as bacteria (Blindauer 2011), yeasts, and ciliates (Gutierrez et al. 2011) and are currently classified in 15 families that are not phylogenetically related but are likely to result from convergent evolution (Capdevila and Atrian 2011). Ciliate MTs are generally longer than average and, along with MTs from metazoans and fungi, contain greater proportions of cysteine than MTs from plants and bacteria (Ziller et al. 2017). In addition to classified proteins, MTs isolated and characterised experimentally from the brown macroalga *Fucus vesiculosus* (Morris et al. 1999), the excavate *Trichomonas vaginalis* (Capdevila and Atrian 2011), different orthopterans, annelids, tunicates, ascomycetes, and basidiomycetes (Ziller and Fraissinet-Tachet 2018), as well as HM-contaminated soils (Lehembre et al. 2013) could not be classified and were suggested to make up novel MT families (Ziller and Fraissinet-Tachet 2018).

The broad genetic diversity spanning living organisms (Baldauf 2008, Keeling 2013) and the scarcity of known MTs in microbial eukaryotes other than fungi and ciliates (Balzano et al. 2020), suggest that the real diversity of MTs as well as the number of distinct families is likely to be broader than that currently known. For example, < 1% proteins annotated on GenBank as MTs

belong to microbial eukaryotes, they are mostly associated with parasitic genera (*Babesia*, *Entoamoeba*, *Plasmodium*, *Trichomonas*), and other microorganisms including microalgae are highly underrepresented (Balzano et al. 2020). MTs from both eukaryotic and prokaryotic microbes have been recently reviewed by Gutiérrez et al. (2019) and while MTs from ciliates and fungi have been characterised in details and classified in different families, little is known on MTs from other non-ciliate protists. Although some common features, such as a prevalence of CXC motifs were observed, MTs from non-ciliate protists do not share a common evolutionary origin and are likely to result from convergent evolution of different genes (Gutiérrez et al. 2019). Overall, very little is known to date on proteins from microalgae and, in general, from protists different than ciliates. Here we predicted, through a bioinformatic approach, novel potential MTs from eukaryotic microbial genomes and transcriptomes.

Materials and Methods

We searched 44 genomes (Blaby-Haas and Merchant 2019) and 636 transcriptomes (Keeling et al. 2014) for novel MTs of non-ciliate protists. The amino acid sequences of the proteins predicted from the genomes were downloaded from GenBank (Supplementary Table S1), whereas a re-assembled version of the proteins predicted from the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP) database (Supplementary Table S2) was downloaded from iMicrobe (Johnson et al. 2019, Youens-Clark et al. 2019). We carried out structural analyses of the proteins predicted from the abovementioned databases using Interproscan (Jones et al. 2014) with default parameters (https://interproscan-docs.readthedocs.io/en/latest/HowToRun.html); proteins found to possess regions identified as MT-domains with a score (e-value) < $5x10^{-5}$ were retained for downstream analyses (Supplementary Table S3). GPS-prot software (Fahey et al. 2011) was used to plot the position of the different domains within each protein. The resulting proteins were aligned using MAFFT-linsy (Katoh and Standley 2013) and analyses revealed the presence of one to three highly conserved CTCGXXCXCGXXCXCXXC patterns in most proteins. We then searched for other proteins possessing the CTCGXXCXCGXXCXCXXC pattern within the abovementioned databases and results were then added to the previous alignments (Supplementary Figure S1). A sequence logo of the abovementioned pattern was generated using WebLogo (Crooks et al. 2004).

Results and Discussion

Functional analyses of genomes and transcriptomes sequenced from non-ciliate protists yielded 10 unique proteins possessing putative MT-specific domains (Table 1, Supplementary Table S4). *AlanMT* protein (*Armaparvus languidus,* amoebozoan, excavate) possesses a region sharing similarities with a domain present in yeast MTs (IPR035715), whereas all the other proteins found here contain two adjacent regions sharing similarities with known MT domains from molluscs (IPR001008). Most (8 out of 10) proteins also contain a putative HM-associated domain (HMA, IPR006121) located in proximity of the N-terminus (Fig. 1), one to three conserved cysteine-rich patterns 18 AA long (CTCGXXCXCGXXCXCXXC), and have been originally isolated from species affiliated to the Stramenopile-Alveolata-Rhizaria (SAR) supergroup. Twelve additional unique proteins containing the same 18 AA pattern were subsequently found in other SAR species (Supplementary Table S5). Overall, structural analyses and pattern search allowed the identification of 21 unique proteins (Table 1), 19 of which derive from SAR species and possess a highly conserved cysteine-rich pattern, that are likely to play a role as MTs (Fig. 2). Thirteen putative MTs are present in more than one transcriptome of the MMETSP database being thus very unlikely to result from contaminations or sequencing errors. Interestingly, in many cases, our putative MTs derive from transcriptomes sequenced out of specimens collected under stress conditions such as high light irradiance (> 300 µE m^{-2} s^{-1}) or under nitrogen (<2 µM) or phosphorus (<0.5 µM) limitation (Table 1). Both high light irradiance and nutrient starvation can generate oxidative stress (Niyogi 1999, Zhang et al. 2013) that has been reported to induce MT biosynthesis (Cobbett and Goldsbrough 2002, Ruttkay-Nedecky et al. 2013). Current data thus suggest that the proteins found here are more likely to be expressed while microorganisms thrive under oxidative stress conditions, coherently with a potential role as MTs.

Little is known on metal-binding mechanisms in microalgal MTs. MTs are generally known to have affinity with monovalent and divalent ions, with cation coordinated by 3 to 4 cysteine residues, and each residue coordinating one or two cations (Korkola et al. 2020, Leszczyszyn et al.

2013, Scheller and Irvine 2018). The number of monovalent or divalent metal cations that can be coordinated by the putative MTs found here cannot be predicted *in silico* but needs to be evaluated experimentally. It has been suggested that an MT is able to chelate a number of monovalent cations slightly higher than half of its cysteine residues and a number of divalent cations lower than 50% cysteine residues (Leszczyszyn et al. 2013, Scheller et al. 2018, Korkola et al. 2020). Short putative MTs such as *AlanMT*, *CrotMT* or *EspiMT* can coordinate around 5-10 cations whereas the longest proteins found here such as *AplaMT* (258 AA), *AstoMT* (264), and *SyneMT* (255) can coordinate up to 30 cations.

Current results strongly suggest that, at least the proteins found here from SAR representatives, that possess an HMA domain along with two adjacent MT domains (Fig. 1), are likely to play a role as MTs. HMA domain has been previously found in proteins involved in HM transport and detoxification in mammals (Bull and Cox 1994, Gitschier et al. 1998), and two adjacent MT-domains typically occur in known MTs from plants (Leszczyszyn et al. 2013), mammals (Nielsen et al. 2007), and ciliates (Zahid et al. 2018). Proteins found here from SAR representatives are longer than most known MTs (Table 1), ranging from 189 (*DbriMT*) to 320 AA (*OaurMT*). The presence of multiple, conserved cysteine-rich patterns (Fig. 2), and the fact that such proteins are longer than average, suggest that putative SAR MTs might have resulted from gene duplication of shorter MTs, similarly to what has been hypothesised for very long MTs in fungi (Iturbe-Espinoza et al. 2016), molluscs (Pedrini-Martha et al. 2020), and *T. vaginalis* (Capdevila and Atrian 2011).

The cysteine content found in our putative MTs is lower than that of most known MTs, ranging from 8% (*AlanMTs*) to 19% (*CrotMT* and *CwaiMT*) and was highly variable even within SAR-derived proteins (Table 2). Histidine content is very low (<2%) in all proteins except *CrotMT* (3.6 %) and *SyneMT* (3.9%); aromatic amino acids account for < 5% in most proteins, whereas lysine contribution ranges from 0.9% (*CrotMT*) to 10% (*AlanMTs*). Overall, putative SAR MTs found here, along with the known MT *AuanMT2*, exhibit a similar domain distribution (Fig. 1), contain cysteine residues mostly clustered in CXC motifs, and share one to three conserved 18 AA pattern

(Fig. 2). Gutierrez et al. (2019) observed a predominance of CXC motifs, especially CKC, in MTs from non-ciliate protists. However, while some known MTs like *BlasMT, CowcMT,* and *TvagMT* are indeed rich (>8) in CKC motifs, this does not seem to be a common feature among the putative MTs found here in non-ciliate protists. For example, *AuanMTs* and *MconMT* do not contain such motif, whereas only one CKC motif occurs in *BbigMT*, *CsorMT*, and *TpseMT* (Table 2). Similarly, among our putative MTs, such motif is present 5 times in *OaurMT*, but it is repeated three times or less in the other proteins (Table 3). In general CTC and CQC motifs are more common than CKC motifs in our putative SAR MTs (Table 2). Current data indicate that both proteins with an experimentally proven HM-binding activity and putative MTs found here via bioinformatic analyses, exhibit a highly variable content in CKC, CTC, and CQC motifs.

The possible role of our SAR proteins as MTs is further suggested by the presence of a region slightly different from our 18 AA pattern, in known MTs from some metazoans, amoebozoans, fungi and higher plants. In this case, the threonine residue on the second position is replaced by other polar or positively charged amino acids (Supplementary Figure S2). Beside this difference, putative SAR MTs share the same number and position of cysteine residues with metal-binding domains in Type 1 MTs from plants (Leszczyszyn et al. 2013), copper and cadmium MTs in snails (Dvorak et al. 2018), and silver MTs in fungi (Sácký et al. 2014).

In spite of the similarities found, even putative SAR MTs, possessing the shared 18 AA pattern, exhibit great differences among each other, and we could not construct a meaningful (i.e. bootstrap support > 30%, using neighbour joining or maximum likelihood algorithms) phylogenetic tree from the alignment of such sequences. This variability is likely to reflect the broad genetic diversity of non-ciliate protists and suggests that, although SAR species share a common evolutionary origin (Keeling 2013), their MTs are likely to result from convergent evolution of different genes, in spite of the shared 18 AA pattern.

Although the putative SAR MTs found here possess two regions related to metal-binding domains of mollusc MTs (Fig. 1) and a conserved 18 AA cysteine-rich region (Fig. 2) that can be

found, in part, in MTs from different organisms (Supplementary Figure S2), none of the putative SAR MTs found here possess the motifs previously described for the 15 MT families (Capdevila and Atrian 2011, Ziller et al. 2017), and thus do not belong to any family described to date. Despite ciliates are part of the SAR supergroup, MTs from ciliates (Family 7) are shorter, contain greater cysteine proportions, and differ in their amino acid sequence from the putative SAR MTs found here (Gutierrez et al. 2011). In addition, except *AuanMT2*, known unclassified MTs from SAR species (*BbigMT, BlasMT, EsilMT, FvesMT, TpseMT*) do not possess the conserved 18AA pattern observed here (Fig. 2), suggesting great differences even within SAR MTs.

MTs can contribute to the development of more efficient HM-sensors. Whole cell MT-based biosensors have been developed in different microbes (Shetty et al. 2004, Shitanda et al. 2005, Amaro et al. 2011) and ciliates are currently considered as the most suitable candidates because of the absence of cell wall (Gutierrez et al. 2009, Gutierrez et al. 2015). However, testing the potential of MTs from other microbes for the development of whole cell biosensors might yield some more efficient candidates. Microalgae can be cultured autotrophically in simple seawater or freshwater enriched with basic nutrients and several green algae, diatoms, dinoflagellates, and Eustigmatophyceae are commonly used for genetic editing. In particular, lightly silicified diatoms, unarmoured dinoflagellates, and *Chlorella* spp. are known for their weak cell walls (Dunker and Wilhelm 2018) and cell-wall free mutants of *Chlamydomonas* spp. are currently available (https://www.chlamycollection.org). Diatoms and dinoflagellates typically dominate shallow benthic communities (Forster et al. 2016) including heavy metal contaminated sediments (Gu et al. 2020) and might thus reveal suitable for the development of MT-based sensors.

Bioinformatic mining of eukaryotic genomes and transcriptomes thus contributed to predict putative MTs of 21 species, 19 of which deriving from SAR representative a sharing an 18 amino acid long cysteine-rich motif. The biological function of these proteins is to be experimentally proven for a complete structural and functional *in vivo* characterisation as well as for the quantification of MT expression in polluted environments and in laboratory microcosms by real

time PCR, and, finally, for the development of MT-based biosensors. Furthermore, physiological assays of species tolerance to HMs can be combined to gene expression determination to improve our understanding on microbes-HM interactions.

Acknowledgements

The authors are grateful to M. Miralto and L. Ambrosino (RIMAR, SZN) for their support in bioinformatic data processing, and to G. Lanzotti (RIMAR, SZN) for graphical assistance. Analyses were performed by using the SZN bioinformatics server Falkor available at SZN (https://bioinfo.szn.it). The authors received no financial support for the research, and authorship of this article.

Author contribution

AS and SB conceived the study, SB carried out bioinformatic analyses, both SB and AS drafted the manuscript. Both AS and SB agree be held accountable for the content therein and approve the final version of the manuscript.

208 **Table and Figure Legend**

209

210 **Table 1.** List of protein sequences predicted from eukaryotic genomes and transcriptomes likely to
211 play a role as MTs, as revealed by Interproscan analyses or motif search[1]

212

213 **Table 2.** Main features, proportion of amino acids potentially involved in metal chelation, and
214 occurrence of cysteine-rich motifs in known and putative MTs from non-ciliate protists

215

216 **Figure 1**. Proteins, from different microbial eukaryotes, containing MT-specific domains as found
217 by structural analyses using Interproscan (Jones et al. 2014). Numbers indicate protein length and
218 the position of the different domains. Domains specific for MTs are in black (Mollusc MTs,
219 IPR001008; crustacean MTs, IPR002045; eukaryotic MT, PF12809), whereas heavy metal
220 associated domains (HMA, IPR006121) are in grey. Species name and sequence identifiers are
221 indicated on the left of each putative MTs, whereas class names are on the right.

222

223 **Figure 2.** Alignment of the putative MTs from heterokonts (Labyinthulids, Pelagophyceae, and
224 diatoms) and dinoflagellates and sequence logo of the highly conserved motif
225 CTCGXXCXCGXXCXCXXC. Underlined sequence IDs correspond to putative MTs found in the
226 present study, whereas IDs which are not underlined are related known MTs from previous studies.
227 Numbers reflect the amino acid position with respect to the longest protein found here (SyneMT
228 from *Synedropsis* sp. CCMP1620). Cysteine residues are highlighted in black while histidine
229 residues, that might also be involved in HM binding, are in grey. Only the regions corresponding to
230 the heavy metal associated domains (HMA, IPR006121, positions 197 to 242) and those exhibiting
231 the cysteine-rich motif CTCGXXCXCGXXCXCXXC are shown for clarity, whereas the full
232 alignment is shown on Supplementary Figure S1. MTs predicted in this study are underlined,

whereas MT-activity has been previously proven or predicted in the other proteins. The species, strain, and treatment associated with each protein abbreviated here are reported on Table 1. Sequence logo was created using WebLogo (weblogo.berkeley.edu/logo.cgi).

Supplementary Figure S1. Full length alignment of the putative MTs from heterokonts and dinoflagellates. Numbers correspond to the amino acid position with respect to the known MT from *Aureococcus anophagefferens* (XP_009037419).

Supplementary Figure S2. Comparison of the most conserved cysteine-rich region of previously described MTs and putative MTs found in this study.

References

Amaro, F., A. P. Turkewitz, A. Martin-Gonzalez and J. C. Gutierrez (2011). "Whole-cell biosensors for detection of heavy metal ions in environmental samples based on metallothionein promoters from Tetrahymena thermophila." Microbial Biotechnology **4**(4): 513-522.

Baldauf, S. L. (2008). "An overview of the phylogeny and diversity of eukaryotes." Journal of Systematics and Evolution **46**(3): 263-273.

Balzano, S., A. Sardo, M. Blasio, T. B. Chahine, F. Dell'Anno, C. Sansone and C. Brunet (2020). "Microalgal Metallothioneins and Phytochelatins and Their Potential Use in Bioremediation." Frontiers in Microbiology **11**.

Blaby-Haas, C. E. and S. S. Merchant (2012). "The ins and outs of algal metal transport." Biochimica Et Biophysica Acta-Molecular Cell Research **1823**(9): 1531-1552.

Blaby-Haas, C. E. and S. S. Merchant (2019). "Comparative and Functional Algal Genomics." Annual Review of Plant Biology **70**(1): 605-638.

Blindauer, C. A. (2011). "Bacterial metallothioneins: past, present, and questions for the future." Journal of Biological Inorganic Chemistry **16**(7): 1011-1024.

Blindauer, C. A. and O. I. Leszczyszyn (2010). "Metallothioneins: unparalleled diversity in structures and functions for metal ion homeostasis and more." Natural Product Reports **27**(5): 720-741.

Bull, P. C. and D. W. Cox (1994). "Wilson disease and Menkes disease: new handles on heavy-metal transport." Trends in genetics : TIG **10**(7): 246-252.

Capdevila, M. and S. Atrian (2011). "Metallothionein protein evolution: a miniassay." Journal of Biological Inorganic Chemistry **16**(7): 977-989.

Cobbett, C. and P. Goldsbrough (2002). "Phytochelatins and metallothioneins: Roles in Heavy Metal Detoxification and Homeostasis." Annual Review of Plant Biology **53**(1): 159-182.

Crooks, G. E., G. Hon, J. M. Chandonia and S. E. Brenner (2004). "WebLogo: a sequence logo generator." Genome Res **14**(6): 1188-1190.

Das, N., R. Vimala and P. Karthika (2008). "Biosorption of heavy metals - An overview." Indian Journal of Biotechnology **7**(2): 159-169.

Dunker, S. and C. Wilhelm (2018). "Cell Wall Structure of Coccoid Green Algae as an Important Trade-Off Between Biotic Interference Mechanisms and Multidimensional Cell Growth." Front Microbiol **9**: 719.

Dvorak, M., R. Lackner, M. Niederwanger, C. Rotondo, R. Schnegg, P. Ladurner, V. Pedrini-Martha, W. Salvenmoser, L. Kremser, H. Lindner, M. García-Risco, S. Calatayud, R. Albalat, Ò. Palacios, M. Capdevila and R. Dallinger (2018). "Metal binding functions of metallothioneins in the slug Arion vulgaris differ from metal-specific isoforms of terrestrial snails." Metallomics **10**(11): 1638-1654.

Fahey, M. E., M. J. Bennett, C. Mahon, S. Jäger, L. Pache, D. Kumar, A. Shapiro, K. Rao, S. K. Chanda, C. S. Craik, A. D. Frankel and N. J. Krogan (2011). "GPS-Prot: A web-based visualization platform for integrating host-pathogen interaction data." BMC Bioinformatics **12**(1): 298.

Forster, D., M. Dunthorn, F. Mahé, J. R. Dolan, S. Audic, D. Bass, L. Bittner, C. Boutte, R. Christen, J.-M. Claverie, J. Decelle, B. Edvardsen, E. Egge, W. Eikrem, A. Gobet, W. H. C. F. Kooistra, R. Logares, R. Massana, M. Montresor, F. Not, H. Ogata, J. Pawlowski, M. C. Pernice, S. Romac, K. Shalchian-Tabrizi, N. Simon, T. A. Richards, S. Santini, D. Sarno, R. Siano, D. Vaulot, P. Wincker, A. Zingone, C. de Vargas and T. Stoeck (2016). "Benthic protists: the under-charted majority." FEMS Microbiology Ecology **92**(8).

Gitschier, J., B. Moffat, D. Reilly, W. I. Wood and W. J. Fairbrother (1998). "Solution structure of the fourth metal-binding domain from the Menkes copper-transporting ATPase." Nature structural biology **5**(1): 47-54.

Gu, R., P. Sun, Y. Wang, F. Yu, N. Jiao and D. Xu (2020). "Genetic Diversity, Community Assembly, and Shaping Factors of Benthic Microbial Eukaryotes in Dongshan Bay, Southeast China." Frontiers in microbiology **11**: 592489-592489.

Gutiérrez, J.-C., P. de Francisco, F. Amaro, S. Díaz and A. Martín-González (2019). Structural and Functional Diversity of Microbial Metallothionein Genes. Microbial Diversity in the Genomic Era. S. Das and H. R. Dash, Academic Press**:** 387-407.

Gutierrez, J. C., F. Amaro, S. Diaz, P. de Francisco, L. L. Cubas and A. Martin-Gonzalez (2011). "Ciliate metallothioneins: unique microbial eukaryotic heavy-metal-binder molecules." Journal of Biological Inorganic Chemistry **16**(7): 1025-1034.

Gutierrez, J. C., F. Amaro and A. Martin-Gonzalez (2009). "From heavy metal-binders to biosensors: Ciliate metallothioneins discussed." Bioessays **31**(7): 805-816.

Gutierrez, J. C., F. Amaro and A. Martin-Gonzalez (2015). "Heavy metal whole-cell biosensors using eukaryotic microorganisms: an updated critical review." Frontiers in Microbiology **6**: 8.

Iturbe-Espinoza, P., S. Gil-Moreno, W. Y. Lin, S. Calatayud, O. Palacios, M. Capdevila and S. Atrian (2016). "The Fungus Tremella mesenterica Encodes the Longest Metallothionein Currently Known: Gene, Protein and Metal Binding Characterization." Plos One **11**(2).

Johnson, L. K., H. Alexander and C. T. Brown (2019). "Re-assembly, quality evaluation, and annotation of 678 microbial eukaryotic reference transcriptomes." Gigascience **8**(4).

Jones, P., D. Binns, H. Y. Chang, M. Fraser, W. Z. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, S. Pesseat, A. F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S. Y. Yong, R. Lopez and S. Hunter (2014). "InterProScan 5: genome-scale protein function classification." Bioinformatics **30**(9): 1236-1240.

Katoh, K. and D. M. Standley (2013). "MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability." Molecular Biology and Evolution **30**(4): 772-780.

Keeling, P. J. (2013). The Number, Speed, and Impact of Plastid Endosymbioses in Eukaryotic Evolution. Annual Review of Plant Biology, Vol 64. S. S. Merchant. **64:** 583-607.

Keeling, P. J., F. Burki, H. M. Wilcox, B. Allam, E. E. Allen, L. A. Amaral-Zettler, E. V. Armbrust, J. M. Archibald, A. K. Bharti, C. J. Bell, B. Beszteri, K. D. Bidle, C. T. Cameron, L. Campbell, D. A. Caron, R. A. Cattolico, J. L. Collier, K. Coyne, S. K. Davy, P. Deschamps, S. T. Dyhrman, B. Edvardsen, R. D. Gates, C. J. Gobler, S. J. Greenwood, S. M. Guida, J. L. Jacobi, K. S. Jakobsen, E. R. James, B. Jenkins, U. John, M. D. Johnson, A. R. Juhl, A. Kamp, L. A. Katz, R. Kiene, A. Kudryavtsev, B. S. Leander, S. Lin, C. Lovejoy, D. Lynn, A. Marchetti, G. McManus, A. M. Nedelcu, S. Menden-Deuer, C. Miceli, T. Mock, M. Montresor, M. A. Moran, S. Murray, G. Nadathur, S. Nagai, P. B. Ngam, B. Palenik, J. Pawlowski, G. Petroni, G. Piganeau, M. C. Posewitz, K. Rengefors, G. Romano, M. E. Rumpho, T. Rynearson, K. B. Schilling, D. C. Schroeder, A. G. B. Simpson, C. H. Slamovits, D. R. Smith, G. J. Smith, S. R. Smith, H. M. Sosik, P. Stief, E. Theriot, S. Twary, P. E. Umale, D. Vaulot, B. Wawrik, G. L. Wheeler, W. H. Wilson, Y. Xu, A. Zingone and A. Z. Worden (2014). "The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans through Transcriptome Sequencing." Plos Biology **12**(6).

Korkola, N. C., P. M. Scarrow and M. J. Stillman (2020). "pH dependence of the non-cooperative binding of Bi³⁺ to human apo-metallothionein 1A: kinetics, speciation, and stoichiometry." Metallomics : integrated biometal science **12**(3): 435-448.

Kumar, K. S., H. U. Dahms, E. J. Won, J. S. Lee and K. H. Shin (2015). "Microalgae - A promising tool for heavy metal remediation." Ecotoxicology and Environmental Safety **113**: 329-352.

Lehembre, F., D. Doillon, E. David, S. Perrotto, J. Baude, J. Foulon, L. Harfouche, L. Vallon, J. Poulain, C. Da Silva, P. Wincker, C. Oger-Desfeux, P. Richaud, J. V. Colpaert, M. Chalot, L. Fraissinet-Tachet, D. Blaudez and R. Marmeisse (2013). "Soil metatranscriptomics for mining eukaryotic heavy metal resistance genes." Environmental Microbiology **15**(10): 2829-2840.

Leszczyszyn, O. I., H. T. Imam and C. A. Blindauer (2013). "Diversity and distribution of plant metallothioneins: a review of structure, properties and functions." Metallomics **5**(9): 1146-1169.

Morris, C. A., B. Nicolaus, V. Sampson, J. L. Harwood and P. Kille (1999). "Identification and characterization of a recombinant metallothionein protein from a marine alga, Fucus vesiculosus." The Biochemical journal **338 (Pt 2)**(Pt 2): 553-560.

Nielsen, A. E., A. Bohr and M. Penkowa (2007). "The Balance between Life and Death of Cells: Roles of Metallothioneins." Biomarker insights **1**: 99-111.

Niyogi, K. K. (1999). "Photoprotection revisited: Genetic and Molecular Approaches." Annu Rev Plant Physiol Plant Mol Biol **50**: 333-359.

Pedrini-Martha, V., S. Koll, M. Dvorak and R. Dallinger (2020). "Cadmium Uptake, MT Gene Activation and Structure of Large-Sized Multi-Domain Metallothioneins in the Terrestrial Door Snail Alinda biplicata (Gastropoda, Clausiliidae)." International Journal of Molecular Sciences **21**(5): 21.

Perales-Vela, H. V., J. M. Pena-Castro and R. O. Canizares-Villanueva (2006). "Heavy metal detoxification in eukaryotic microalgae." Chemosphere **64**(1): 1-10.

Ruttkay-Nedecky, B., L. Nejdl, J. Gumulec, O. Zitka, M. Masarik, T. Eckschlager, M. Stiborova, V. Adam and R. Kizek (2013). "The role of metallothionein in oxidative stress." Int J Mol Sci **14**(3): 6044-6066.

Sácký, J., T. Leonhardt, J. Borovička, M. Gryndler, A. Briksí and P. Kotrba (2014). "Intracellular sequestration of zinc, cadmium and silver in Hebeloma mesophaeum and characterization of its metallothionein genes." Fungal Genet Biol **67**: 3-14.

Scheller, J. S., G. W. Irvine and M. Stillman (2018). "Unravelling the mechanistic details of metal binding to mammalian metallothioneins from stoichiometric, kinetic, and binding affinity data." Dalton transactions **47 11**: 3613-3637.

Shetty, R. S., S. K. Deo, Y. Liu and S. Daunert (2004). "Fluorescence-based sensing system for copper using genetically engineered living yeast cells." Biotechnol Bioeng **88**(5): 664-670.

Shitanda, I., K. Takada, Y. Sakai and T. Tatsuma (2005). "Amperometric Biosensing Systems Based on Motility and Gravitaxis of Flagellate Algae for Aquatic Risk Assessment." Analytical Chemistry **77**(20): 6715-6718.

Youens-Clark, K., M. Bomhoff, A. J. Ponsero, E. M. Wood-Charlson, J. Lynch, I. Choi, J. H. Hartman and B. L. Hurwitz (2019). "iMicrobe: Tools and data-driven discovery platform for the microbiome sciences." GigaScience **8**(7).

Zahid, M. T., F. R. Shakoori, S. Zulfiqar, K. A. Al-Ghanim and A. R. Shakoori (2018). "Growth Characteristics, Metal Uptake and Expression Analysis of Copper Metallothionein in a Newly Reported Ciliate, Tetrahymena farahensis." Pakistan Journal of Zoology **50**(3): 1171-1181.

Zhang, Y.-M., H. Chen, C.-L. He and Q. Wang (2013). "Nitrogen starvation induced oxidative stress in an oil-producing green alga Chlorella sorokiniana C3." PloS one **8**(7): e69225-e69225.

Ziller, A. and L. Fraissinet-Tachet (2018). "Metallothionein diversity and distribution in the tree of life: a multifunctional protein." Metallomics **10**(11): 1549-1559.

Ziller, A., R. K. Yadav, M. Capdevila, M. S. Reddy, L. Vallon, R. Marmeisse, S. Atrian, O. Palacios and L. Fraissinet-Tachet (2017). "Metagenomics analysis reveals a new metallothionein family: Sequence and metal-binding features of new environmental cysteine-rich proteins." Journal of Inorganic Biochemistry **167**: 1-11.

Table 1. List of known and predicted MTs from non-ciliate protists as revealed by Interproscan analyses or motif search [1]

Protein ID	Species	Class	Supergroup	Strain ID	Transcriptome ID	Database ID	No identical transcripts[2]	Stress condition[3]	Interproscan domain code[4]
EinvMT	***Entoamoeba invadens***	**Archamoeboe**	**Amoebozoans**	**IP1**	**NA**	**XP_004259069**	**1**	**NA**	
AlanMT	*Armaparvus languidus*	Vannellids	Amoebozoans	PRA-29	MMETSP0420	Tr3694	1	HL	IPR035715
CsorMT	***Chlorella sorokiniana***	**Green algae**	**Archaeplastida**	**1602**	**NA**	**PRW44601.1**	**1**	**NA**	**IPR002045**
MconMT	***Micractinium condutrix***	**Green algae**	**Archaeplastida**	**SAG 241.80**	**NA**	**PSC70917**	**1**	**NA**	
TvagMT	***Trichomonas vaginalis***	**Parabasalids**	**Discoba**	**ATCC PRA-98**	**NA**	**XP_001321197**	**1**	**NA**	
CrotMT	*Chrysochromulina rotalis*	Haptophytes	Hacrobians	UIO044	MMETSP0287	Tr26136	1	HL	IPR001008
CowcMT	***Capsaspora owczarzaki***	**Filozoa**	**Opisthokonts**	**ATCC 30864**	**NA**	**XP_011270693**	**1**	**NA**	
BbigMT	***Babesia bigemina***	**Apicomplexa**	**SAR**[5]		**NA**	**XP_012768823**	**1**	**NA**	
BlasMT	***Blastocystis* sp.**	**Bigyra**	**SAR**	**ATCC 50177**	**NA**	**OAO13187**	**1**	**NA**	
EsilMT	***Ectocarpus siliculosus***	**Brown algae**	**SAR**		**NA**	**CBJ32637**			**IPR001008**
FvesMT	***Fucus vesiculosus***	**Brown algae**	**SAR**		**NA**	**CAA06729**			**IPR001008**
AglaMT1	*Asterionellopsis glacialis*	Diatoms	SAR	CCMP134	MMETSP0708	Tr19519	3	N-/P-	IPR001008
AglaMT2	*Asterionellopsis glacialis*	Diatoms	SAR	CCMP1581	MMETSP1394	Tr220	1	N-/P-	
CwaiMT	*Coscinodiscus wailesii*	Diatoms	SAR	CCMP2513	MMETSP1066	Tr41518	1	HL	IPR001008
DbriMT	*Ditylum brightwellii*	Diatoms	SAR	GSO105	MMETSP0998	Tr22984	8	HL/No/N-/P-	IPR001008
EspiMT	*Extubocellulus spinifer*	Diatoms	SAR	CCMP396	MMETSP0697	Tr10701	1	Si-/No/HL	
MpolMT	*Minutocellus polymorphus*	Diatoms	SAR	NH13	MMETSP1070	Tr24663	2	No/HL	
OaurMT	*Odontella aurita*	Diatoms	SAR	Is-1302-5	MMETSP0015	Tr34634	2	HL	
PdubMT	*Pseudodictyota dubia*	Diatoms	SAR	CCMP147	MMETSP1175	Tr24667	1	HL	IPR001008
SyneMT	*Synedropsis* sp.	Diatoms	SAR	CCMP1620	MMETSP1176	Tr28518	2	HL	IPR001008
TpseMT	***Thalassiosira pseudonana***	**Diatoms**	**SAR**	**CCMP1335**	**NA**	**XP_002296843**	**1**	**NA**	
AcatMT	*Alexandrium catenella*	Dinoflagellate	SAR	OF101	MMETSP0790	Tr99632	1	No	
AmonMT	*Alexandrium monilatum*	Dinoflagellate	SAR	CCMP3105	MMETSP0096	Tr45933	4	HL/P-	
AzspMT	*Azadinium spinosum*	Dinoflagellate	SAR	3D9	MMETSP1037	Tr93697	2	HL	IPR001008
GspiMT	*Gonyaulax spinifera*	Dinoflagellate	SAR	CCMP409	MMETSP1439	Tr79705	1	HL	
LpolMT	*Lingulodinium polyedrum*	Dinoflagellate	SAR	CCMP1738	MMETSP1032	Tr14667	4	No/HL	
AplaMT	*Aplanochytrium* sp.	Labyrinthulids	SAR	PBS07	MMETSP0956	Tr7261	4	NA	IPR001008
AstoMT	*Aplanochytrium stocchinoi*	Labyrinthulids	SAR	GSBS06	MMETSP1349	Tr9377	4	NA	IPR001008
AuanMT1	*Aureococcus anophagefferens*	Pelagophyceae	SAR	CCMP1850	MMETSP0917	Tr30268	3	N-	
AuanMT2	***Aureococcus anophagefferens***	**Pelagophyceae**	**SAR**	**CCMP1984**	**NA**	**XP_009037419**	**1**	**NA**	**IPR001008**
PsubMT	*Pelagococcus subviridis*	Pelagophyceae	SAR	CCMP1429	MMETSP0883	Tr17315	3	N-	
PcalMT	*Pelagomonas calceolata*	Pelagophyceae	SAR	RCC969	MMETSP1328	Tr480	4	NA	

[1] known MTs identified in previous studies are in bold

[2] In many cases 2 or more identical proteins possessing MT-specific domain or resulting from keyword searches were found from different transcriptomes of the same strain

[3] Stress condition at which the strain was maintained prior to transcriptome sequencing. "No" refers to transcriptomes derived from strains cultured at standard conditions. In some cases identical sequences were obtained from different transcriptomes reflecting either different stress treatments or both stress and non-stress conditions.

Abbreviations: N-, nitrogen deprivation (<2 µM) ; P-, phosphorus deprivation (< 0.5 µM);Si-, silica deprivation (< 0.5 µM) HL, high light (> 300 $\mu Em^{-2}s^{-1}$)

[4] The sequences without an Interproscan code were identified by keyword search of the conserved CTCGXXCXCGXXCXCXXC motif.

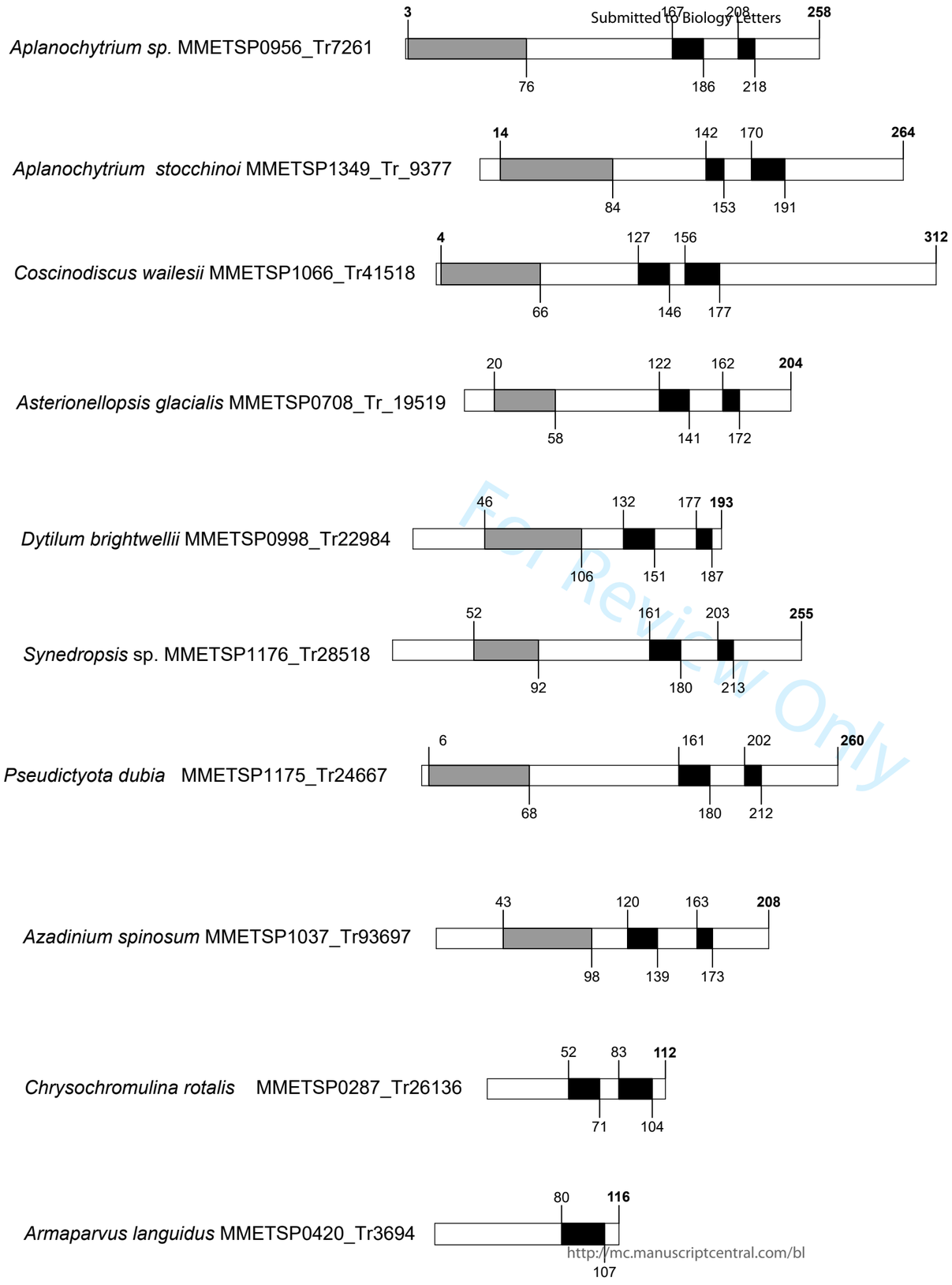
[5] Stramenopiles-Alveolata-Rhizaria

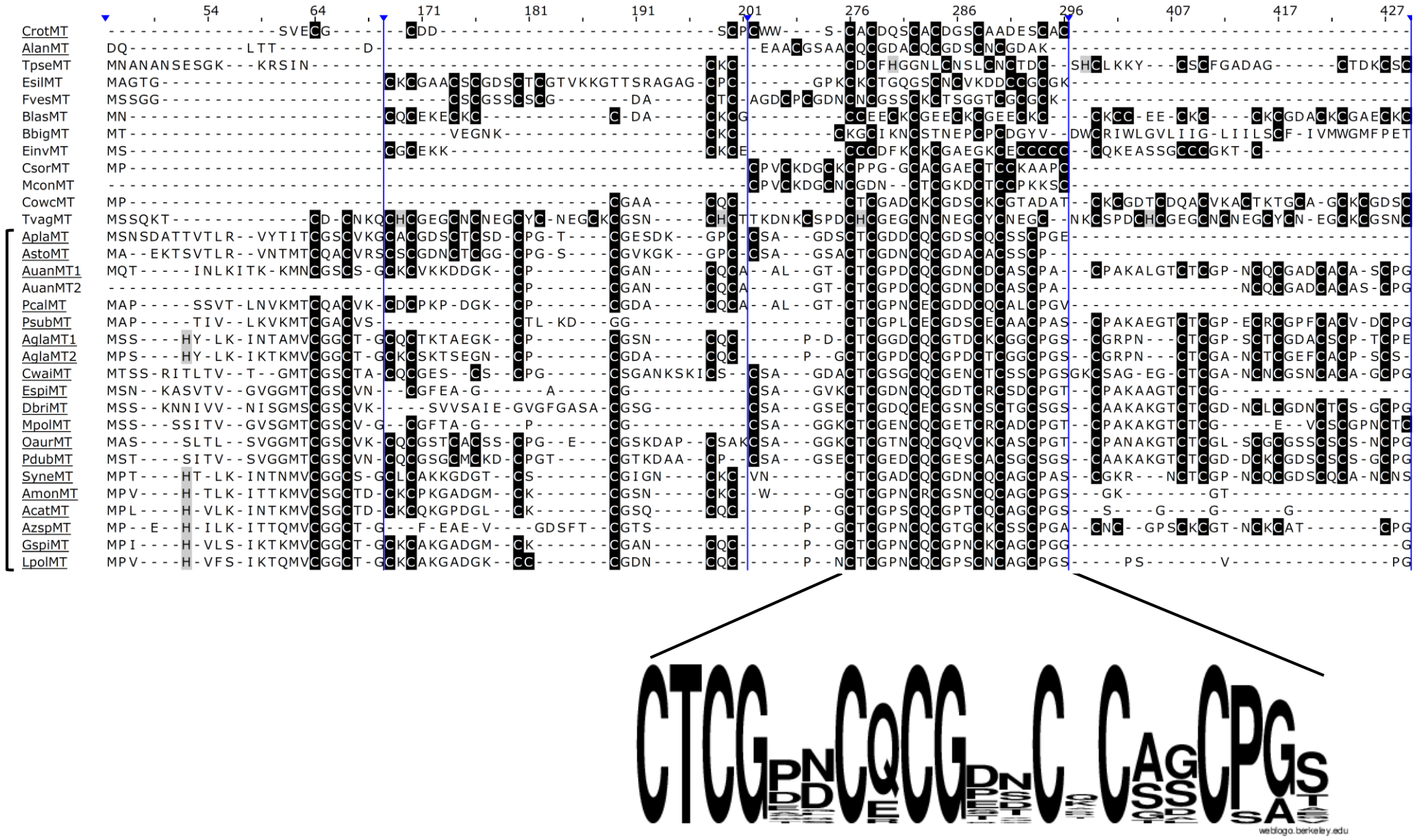
Table 2. Main features and proportion of amino acid potentially involved in metal chelation, for the putative MTs found in the present study

Species	Protein ID	Length	Amono acid residues			Specific motifs			
			Cysteine[1]	Histidine	Aromatic AA	CXC	CKC	CTC	18 AA[2]
Alexandrium catenella	AcatMT	189	9%	1.6%	3.2%	6	2	1	1
Alexandrium monilatum	AmonMT	196	10%	1.0%	5.1%	6	3	1	1
Aplanochytrium sp.	AplaMT	258	15%	0.4%	1.9%	12	0	5	3
Aplanochytrium stocchinoi	AstoMT	264	14%	1.1%	1.5%	12	2	6	3
Armaparvus languidus	AlanMT	116	8%	1.7%	7.7%	3	0		
Asterionellopsis glacialis	AglaMT1	204	13%	1.5%	2.5%	9	1	3	2
Asterionellopsis glacialis	AglaMT2	196	14%	1.0%	2.6%	9	1	4	1
Aureococcus anophagefferens	AuanMT1	232	15%	0.0%	1.3%	11	1	4	1
Azadinium spinosum	AzspMT	208	11%	0.5%	4.8%	6	3	1	1
Chrysochromulina rotalis	CrotMT	112	19%	3.6%	6.3%	6	1		
Coscinodiscus wailesii	CwaiMT	312	19%	0.0%	0.0%	18	0	7	4
Ditylum brightwellii	DbriMT	193	11%	1.0%	1.6%	6	0	3	1
Extubocellulus spinifer	EspiMT	129	12%	0.8%	1.6%	4	0	2	1
Gonyaulax spinifera	GspiMT	164	10%	0.6%	3.7%	6	3	1	
Lingulodinium polyedrum	LpolMT	196	10%	1.0%	3.6%	6	1	1	1
Minutocellus polymorphus	MpolMT	203	11%	1.0%	1.5%	6	0	3	1
Odontella aurita	OaurMT	320	17%	0.0%	1.3%	17	5	6	3
Pelagococcus subviridis	PsubMT	207	11%	0.5%	2.4%	6	3	2	1
Pelagomonas calceolata	PcalMT	160	11%	0.6%	2.5%	6	0	1	2
Pseudodictyota dubia	PdubMT	260	19%	0.0%	0.8%	15	3	5	3
Synedropsis sp.	SyneMT	255	11%	3.9%	4.7%	9	1	2	1
Aureococcus anophagefferens	**AuanMT2**	**171**	**18.%**	**0.0%**	**1.2%**	**12**	0	3	1
Babesia bigemina	**BbigMT**	**214**	**12%**	**1.9%**	**7.9%**	**2**	1		
***Blastocystis* sp.**	**BlasMT**	**207**	**40%**	**0.0%**	**0.0%**	**33**	29		
Capsaspora owczarzaki	**CowcMT**	**176**	**27%**	**0.0%**	**0.0%**	**15**	11	1	
Chlorella sorokiniana	**CsorMT**	**56**	**32%**	**0.0%**	**0.0%**	**6**	1	1	
Entoamoeba invadens	**EinvMT**	**103**	**35%**	**0.0%**	**1.9%**	**13**	4		
Micractinium condutrix	**MconMT**	**59**	**30%**	**0.0%**	**0.0%**	**6**	0	3	
Thalassiosira pseudonana	**TpseMT**	**141**	**13%**	**1.4%**	**5.7%**	**6**	1		
Trichomonas vaginalis	**TvagMT**	**308**	**30%**	**6.2%**	**2.3%**	**41**	9		

[1] Values refer to the number of aminoacid in the sequence

[2]Specific, 18 amino acid motif (CTCGXXCXCGXXCXCXXC) identified in the putative MTs found in the present study




54
64
171
181
191
201
276
286
296
407
417
427
CrotMT
AlanMT
TpseMT
EsilMT
FvesMT
BlasMT
BbigMT
EinvMT
CsorMT
MconMT
CowcMT
TvagMT
AplaMT
AstoMT
AuanMT1
AuanMT2
PcalMT
PsubMT
AglaMT1
AglaMT2
CwaiMT
EspiMT
DbriMT
MpolMT
OaurMT
PdubMT
SyneMT
AmonMT
AcatMT
AzspMT
GspiMT
LpolMT
weblogo.berkeley.edu