Marine Genomics xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Marine Genomics

journal homepage: www.elsevier.com/locate/margen

Method paper

Novel subfamilies of actin-regulating proteins

Morozov A.A.*, Bedoshvili Ye.D., Popova M.S., Likhoshway Ye.V.

Cell Ultrastructure Dept., Limnological Institute SB RAS, Irkutsk, Russia

A B S T R A C T

Ability of actin to polymerise and depolymerise makes it essential to key functions of eukaryotic cell. The functioning of actin is controlled by a host of regulatory proteins, the repertoire of which in diatoms is known to remarkably differ from other organisms. We have performed a phylogenetic analysis of 521 actin and actin-related proteins' aminoacid sequences, as well as 190 sequences of gelsolin family proteins from various genomic and transcriptomic datasets. Based on the results of this analysis, as well as on the presence of clade-specific indels in some of the actin-related proteins, we describe a novel ARP subfamily, dubbed ARP12, which is specific to heterokonts and related organisms. We also describe two novel diatom-specific subfamilies, dGRC1 and dGRC2, among short gelsolin repeat-containing proteins.

1. Introduction

Actin is one of the most important proteins of the eukaryotic cell, essential to all types of cellular activity. Actin filaments are involved in cell division, endocytosis, exocytosis, secretion, *et cetera* (Lodish et al., 2007). In some organisms they also take part in more specialised processes, such as maintaining cell shape and morphogenesis of cells or cell structures (Mathur et al., 1999, Higaki et al., 2010). Other than actin itself, maintaining functional actin filaments requires a host of proteins that perform actin filament assembly, disassembly and movement. In total, about 150 proteins are known to bind globular actin or actin filaments (dos Remedios et al., 2003).

Diatoms are one of the most successful groups of aquatic organisms, adapted to a broad range of environments (Round et al., 1990). Their most remarkable feature is the species-specific siliceous cell wall, which is formed inside the cell; cell wall elements are created in the specialised vesicle called SDV under cytoskeleton control (Drum and Pankratz, 1964, Reimann, 1964). Data from fluorescent and confocal microscopy show that actin is localized near the forming thecae during their morphogenesis and that inhibiting actin filament functions can lead to cell wall abnormalities, as shown on *Proboscia alata* (Van de Meene and Pickett-Heaps, 2002), Rhizosolenia sotigera (Van de Meene and Pickett-Heaps, 2004), and a range of other species (Tesson and Hildebrand, 2010).

It was recently discovered that repertoire of proteins involved in actin filament functions in diatoms is remarkably different from the one in other organisms (Aumeier et al., 2015). In particular, diatom genomes don't seem to encode proteins known to be key elements of cytoskeleton regulation, such as profilin, fascin, thymosin, and others (Aumeier, 2012; Aumeier et al., 2015). On the other hand, it's reasonable to suggest that diatoms have their own unique actin-binding proteins, which are not present in human or yeast. Annotating diatom- or heterokont-specific subfamilies in known families involved in the functioning of the actin filaments can be the first step in the search for unique actin filament regulators.

Earlier works of this type were limited to a few published diatom genomes and Sanger sequencing. The latter is complicated for non-model organisms and protein subfamilies because primers designed using a reference protein may not be suitable for its distant homologues, and the former obviously wasn't taxonomically representative. Publication of the transcriptomic data by the MMETSP project allowed the researchers to analyse many previously unsequenced marine eukaryotes, including diatoms (Keeling et al., 2014). In addition, using RNA sequences instead of DNA, while not a conclusive evidence of expression, decreases the probability of the sequence in question being a pseudogene. This work's aim, thus, was to search for novel families of actin-related proteins and proteins of the gelsolin family using sequence data from as broad range of species as possible, with a particular interest in diatom- or heterokont-specific proteins.

The family of actin-related proteins (ARP) is formed by 11 subfamilies of the distant homologues of actin that are similar to it in sequence and structure, but do not form filaments. Instead, they perform various functions in the cell. ARP1, ARP10 and ARP11 take part in binding vesicles to the cytoskeleton for the vesicular transport; ARP2 and ARP3 are necessary for the formation of the branching actin filaments; the rest are involved in chromatin remodeling. As for their

* Corresponding author.
E-mail address: morozov@lin.irk.ru (A.A. Morozov).

http://dx.doi.org/10.1016/j.margen.2017.10.001
Received 15 March 2017; Received in revised form 28 July 2017; Accepted 11 October 2017

Table 1
The known ARP classes.

ARP subfamily	Function	Organisms	First proposed by
ARP1	As a component of the dynactin complex, binds vesicles to the cytoskeleton (Eckley and Schroer, 2003)	Universal	Lees-Miller et al., 1992
ARP2	Form the complex responsible for nucleating a branching filament (Skau and Waterman, 2015)	Universal with the few exceptions.	Schwob and Martin, 1992
ARP3			Lees-Miller et al., 1992
ARP4	As a component of several nuclear complexes, takes part in histone metilation and chromatin remodeling (Oma and Harato, 2011, Katoh et al., 2011). Binds histones (Georgieva et al., 2015)	Universal	Goffeau et al., 1996
ARP5	As a component of the INO80 complex, takes part in chromatin remodeling. Involved in transcription, replication and reparation (Oma and Harato, 2011)	Patchy distribution; cooccurs with ARP8	
ARP6	As a component of the SWR1 complex, takes part in ATP-dependent replacement of H2A histone with H2A.Z histone variant (Seo et al., 2016)	Universal	
ARP7	As a component of SWI and RSC complexes, takes part in chromatin remodeling (Oma and Harato, 2011)	Fungi	
ARP8	As a component of the INO80 complex, takes part in chromatin remodeling. Hypothetically binds double-stranded breaks (Osakabe et al., 2014)	Patchy distribution; cooccurs with ARP5	
ARP9	As a component of SWI and RSC complexes, takes part in chromatin remodeling (Oma and Harato, 2011)	Fungi	
ARP10	Functionally analogous to ARP11 (Clark and Rose, 2006)	Yeast	
ARP11	As a component of the dynactin complex, takes part in binding vesicles to the cytoskeleton (Eckley and Schroer, 2003)	Patchy distribution	Eckley et al., 1999
ARP12	Unknown	Heterokonts, haptophytes	This work

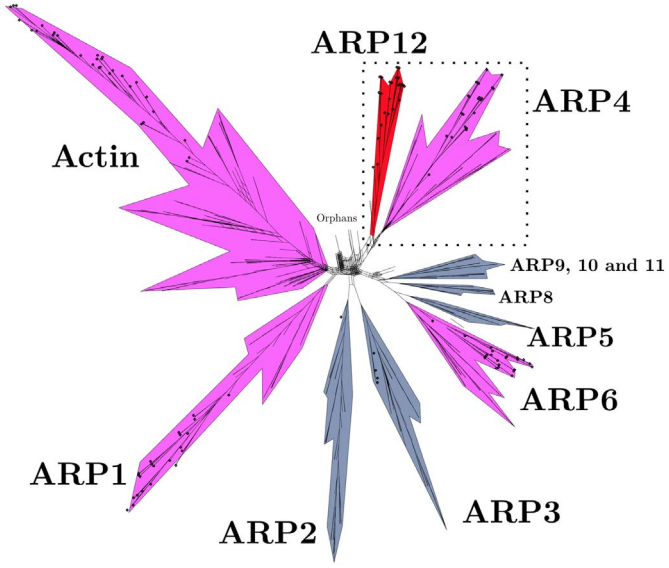

Fig. 1. The consensus phylogenetic network of 256 ML-trees of actins and ARP. Diatom sequences are marked by dots. Clades that do not include diatom sequences have blue background (see Discussion for ARP2/3), and clades where most diatoms are present and the topology suggests that the corresponding gene was present in their MRCA are coloured magenta. The heterokont-specific ARP12 is shown in red. The area within a dotted rectangle is shown at Fig. 2 in higher resolution. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

distribution among organisms, actin, ARP1, ARP4 and ARP6 are present in all eukaryotes; ARP5, ARP8 and ARP11 show patchy distribution on the Tree of life and the history of their acquisition and loss is yet to be documented. Finally, ARP7, ARP9 and ARP10 are only present in fungi and haven't been found in other organisms. The functions and distribution of ARP subfamilies are summarised in Table 1. Authors of the most recent ARP classification (Muller et al., 2005) also note that their analysis has recovered several “orphan” ARP sequences which cannot be placed in any subfamily. According to their hypothesis, those may be representatives of subfamilies specific to non-model groups from which only a single gene has been sequenced. Although Muller et al. have postulated the possibility of expanding ARP classification, no works have been published in this area and no new subfamilies were described.

Gelsolin family of actin-binding proteins is characterised by the presence of multiple gelsolin repeats consisting of three beta-sheets and two alpha-helices (Smirnov et al., 2007). Proteins of this family consist most often of either 6 (gelsolin, vilin) or 3 (fragmin, severin, capG) such repeats. There are also proteins with other numbers of domains, from 2 to 7, thought to arise by multiple duplications and domain losses (Ghoshdastider et al., 2013). However, only genes encoding three-domain proteins were found in diatom genomes (Aumeier, 2012). The ability to bind both globular actin and actin filaments gives gelsolin family proteins a broad range of functions. For example, gelsolin binds to the side of the actin filament and cuts it. After cutting gelsolin remains bound to a filament's barbed end, thus capping it and preventing elongation. But the same protein can also bind two molecules of globular actin, leading to their binding and filament nucleation (Burtnick et al., 1997; Nag et al., 2013). Most proteins of this family, both three- or six-repeat-containing, can perform all three functions: they initiate actin filament nucleation, cut the filament and cap it.

There are three major groups of three-domain-containing proteins in the gelsolin family: capG, severins and fragmins. CapG, also known as the macrophage capping protein, was originally sequenced from the animal macrophages, hence the name (Dabiri et al., 1992). Later it was also shown to be important for the motility of the fibroblastic and endothelial cells (Sun et al., 1995; Pellieux et al., 2003). Unlike most of the other gelsolin family proteins, it is unable to cut actin filaments, but takes part both in the filament nucleation and barbed end capping (Southwick and DiNubile, 1986). Fragmin and severin were documented in *Physarum polycephalum* and *Dictyostelium discoideum* respectively (Ampe and Vandekerchkove, 1987; Andre et al., 1988), and they are both capable of filament nucleation and calcium-dependent filament cleaving (McGough et al., 2003; Hasegawa et al., 1980).

In this work we propose two new subfamilies of gelsolin-family proteins and a subfamily of ARPs based on the results of the phylogenetic analysis and the presence of synapomorphic indels.

2. Materials and methods

2.1. Dataset preparation and phylogenetic analysis

Dataset for a phylogenetic analysis of ARPs consisted of two major parts: a reference set of annotated actins and ARPs, and a collection of unsorted diatom (and heterokont) actin/ARP sequences. We have used