Gene 593 (2016) 64–75

Contents lists available at ScienceDirect

Gene

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

Research paper

Nonagonal cadherins: A new protein family found within the Stramenopiles

Kyle I.G. Fletcher [a,b,c,1], Pieter van West [b], Claire M.M. Gachon [c,*]

[a] Oceanlab, University of Aberdeen, Newburgh, Aberdeenshire AB41 6AA, United Kingdom
[b] Aberdeen Oomycete Laboratory, Institute of Medical Sciences, University of Aberdeen, Foresterhill, Aberdeen AB25 2ZD, United Kingdom
[c] Culture Collection for Algae and Protozoa, Scottish Marine Institute, Scottish Association for Marine Science, Dunstaffnage, Oban, Argyll PA37 1QA, United Kingdom

ARTICLE INFO

Article history:
Received 6 April 2016
Received in revised form 30 July 2016
Accepted 3 August 2016
Available online 04 August 2016

Keywords:
Oomycete
Peronosporomycetes
Oomycota
Labyrinthulomycete
Labyrinthulea
Thraustochytrids
Aureococcus anophagefferens
Blastocystae
Blastocystis hominis
Carbohydrate binding
WSC domain
Cell adhesion

ABSTRACT

Cadherins, a group of molecules typically associated with planar cell polarity and Wnt signalling, have been little reported outside of the animal kingdom. Here, we identify a new family of cadherins in the Stramenopiles, termed Nonagonal after their 9 transmembrane passes, which contrast to the one or seven passes found in other known cadherin families. Manual curation and experimental validation reveal two subclasses of nonagonal cadherins, depending on the number of uninterrupted extracellular cadherin (EC) modules presented. Firstly, shorter mono-exonic, unimodular, protein models, with 3 to 12 EC domains occur as duplicate paralogs in the saprotrophic Labyrinthulomycetes *Aurantiochytrium limanicum* and *Schizochytrium aggregatum*, the gastrointestinal *Blastocystis hominis* (Blastocystae) and as a single copy gene in the autotrophic Pelagophyte *Aureococcus anophagefferens*. Larger, single copy, multi-exonal, tri-modular protein models, with up to 72 EC domain in total, are found in the Oomycete genera *Albugo*, *Phytophthora, Pythium* and *Eurychasma*. No homolog was found in the closely related autotrophic Phaeophyceae (brown algae) or Bacillariophyceae (diatoms), nor in several genera of plant and animal pathogenic oomycetes (*Aphanomyces*, *Saprolegnia* and *Hyaloperonospora*). This potential absence was further investigated by synteny analysis of the genome regions flanking the cadherin gene models, which are found to be highly variable. Novel to this new cadherin family is the presence of intercalated laminin and putative carbohydrate binding in tri-modular oomycete cadherins and at the N-terminus of thraustochytrid proteins. As we were unable to detect any homologs of proteins involved in signalling pathways where other cadherin families are involved, we present a conceptual hypothesis on the function of nonagonal cadherin based around the presence of putative carbohydrate binding domains.

1. Introduction

Cadherins have been reported throughout the Holozoa, from complex multicellular vertebrates (Usui et al., 1999; Yonemura et al., 1995), through to the closest living relatives to the Metazoa, the Choanoflagellates (or Choanomonada, e.g. *Monosiga brevicollis* and *Salpingoeca rosetta*) and the more distant Filastera (e.g. *Capsaspora owczarzaki,* (Abedin and King, 2008; King et al., 2003)). Classic cadherins were originally discovered during mouse embryogenesis and chick development where they were shown to be responsible for Ca^{2+}-dependent homophillic cell adhesion (Gallin et al., 1983; Peyriéras et al., 1983; Takeichi, 1988). Since these initial records, five structural classes of cadherins have been recognised within animals (Fig. 1A); all contain calcium binding extracellular cadherin (EC) domains (Nollet et al., 2000; Pettitt, 2005). Atypical cadherins, Fat, Dachsous and Flamingo (Fmi) are longer, though no variation is presently noted in the number of EC domains (Halbleib and Nelson, 2006), in contrast with canonical cadherins (Classic, Protocadherin) which display variability in their respective number of EC domains (Hulpiau and van Roy, 2011). In comparison, the lesser studied cadherin proteins of the choanoflagellates do not appear to fall under the same architecture. Within protein families here, a large amount of variation with regards to the number of EC domains has been reported and additional extracellular domains can be identified, these additional domains appear to be the delineating factor between protein families (Fig. 1b; Nichols et al., 2012). Present thinking suggests that the choanoflagellate proteins are closely related to Fmi and Protocadherin protiens (King et al., 2003). Holozoan cadherins present only a single transmembrane pass, except for the metazoan Fmi (Fig. 1, (Halbleib and Nelson, 2006; Tissir and Goffinet, 2013; Usui et al., 1999)).

Functionally, animal cadherins have been linked with pathways such as Wnt planar cell polarity (PCP) and HIPPO-mediated contact inhibition (Harumoto et al., 2010; Kim et al., 2011; Tissir and Goffinet,

* Corresponding author.
E-mail addresses: kfletcher@ucdavis.edu (K.I.G. Fletcher), p.vanwest@abdn.ac.uk (P. van West), claire.gachon@sams.ac.uk (C.M.M. Gachon).
[1] Present address: Genome Center, University of California-Davis, CA 95616, USA.

http://dx.doi.org/10.1016/j.gene.2016.08.003
0378-1119/

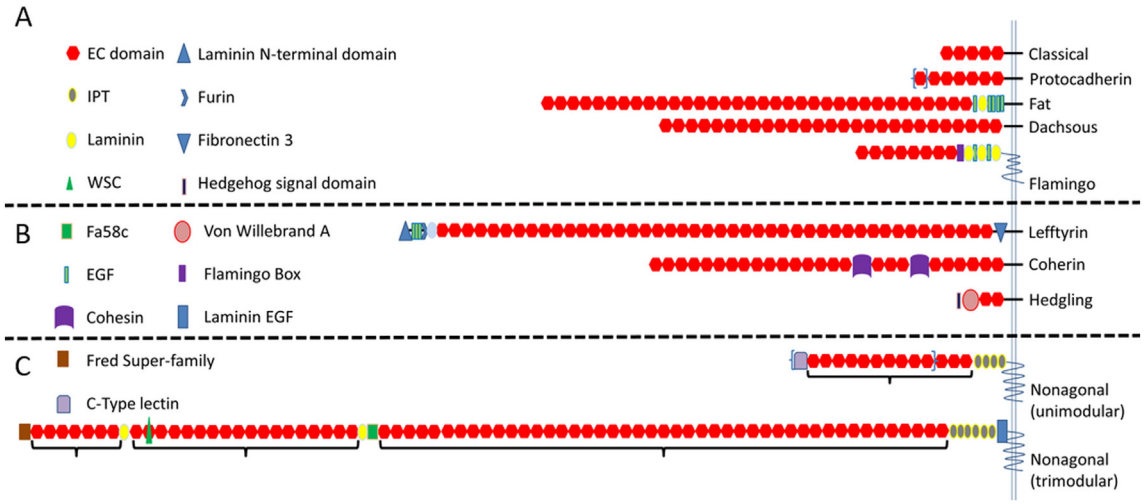

Fig. 1. Overview of current cadherin protein models. A. Classical cadherins found throughout the metazoa. B. Non-classical cadherins more recently defined from the choanoflagellates, also present in a limited number of metazoan species. C. Two types of nonagonal cadherins defined in this study. Section A adapted from Halbleib and Nelson (2006), section B adapted from Nichols et al. (2012).

2013; van Roy and Berx, 2008; Zhao et al., 2007). In choanoflagellates, where the PCP pathway is seemingly absent (Fairclough et al., 2013), the function of cadherins have been speculated to include: bacterial prey capture, attachment to extracellular matrix/environmental substrates or gamete recognition (Nichols et al., 2012).

Distantly removed from the Holozoa (Fig. 2), a small cluster of four cadherin-encoding genes was reported in the genome of *Pythium ultimum* var. ultimum, with additional EC domains detected in *Phytophthora infestans* but interestingly none of these cadherins were predicted as transmembranar and no functional analysis was performed (Levesque et al., 2010). Other oomycete proteins with EC domains have been automatically annotated (Table SI1-1), though none have been reported or curated.

Here, we combine detailed genome structural annotation and experimental validation, through RNAseq and cDNA Sanger sequencing, to describe a new class of cadherin (EC-containing proteins) with nine transmembrane passes, restricted to the Stramenopiles. Gene models were experimentally verified through transcriptome data mining and Sanger sequencing of *Eurychasma dicksonii* cDNA. In order to identify the pathway(s) in which nonagonal cadherins might be involved, a search for candidate orthologs of known cadherin interactants and downstream signalling proteins was conducted. Additional information is provided for the exon architecture on the scaffolds of these organisms and the flanking proteins present on each scaffold were investigated for purposes of synteny.

2. Methods

2.1. In silico curation of cadherin protein models

The cadherin sequences previously identified in *P. ultimum* var. ultimum (Table SI1-1; Levesque et al., 2010) and cadherins from animals (*Drosophila melangostar* Flamingo; BAA84069.1, *Homo sapiens* Fat; NP_005236) were used to query (blastp, tblastn) a wide variety of publically available genomes and predicted proteomes. All sequences retrieved were manually checked for EC domains using CDD (Marchler-Bauer et al., 2015) and transmembrane regions using TMHMM v2.0 (Krogh et al., 2001; Sonnhammer et al., 1998). Protein fragments (i.e. without stop codons) were aligned against a more complete protein model and were investigated on the nucleotide level to identify missing or mispredicted coding regions. Additionally, several contigs from a de novo-assembled HiSeq transcriptome of the basal oomycete *E. dicksonii* CCAP 4018/1 (Gachon et al., in prep.) were included in the study.

Whenever possible, protein models were mapped on genome scaffolds and confronted to transcriptome data to establish intron/exon boundaries. Annotated scaffold sections encoding multiexon protein models are provided in Suppl. Mat. 2–3. The protein models were annotated by querying CDD online (Marchler-Bauer et al., 2015) and TMHMM v2.0; the isoelectric point of the intracellular tail of each protein predicted using the ExPASY compute pI tool (Gasteiger et al., 2005). Further secondary structure prediction was conducted using Jpred (Drozdetskiy et al., 2015). Average EC positions were estimated through multiples alignments or orthologous proteins allowing the investigation of conserved motifs within these domains. Protein models and annotations are provided as unaligned fasta files with corresponding annotations (Suppl. Mat. 4 & 5); an overview of the corrections made to previous released models is given in Suppl. Mat. 6. Additional potential pathway orthologs were queried against the genomes of several Stramenopiles. The *Eurychasma* sequence determined experimentally was deposited in NCBI under accession KU997633.

Synteny was used to investigate the absence of protein models not so far predicted. The *P. ultimum* genome browser (http://pythium.plantbiology.msu.edu/cgi-bin/gbrowse/pug1/) and *A. candida* genome scaffold (CAIX01000017.1) were used to establish flanking proteins of the cadherin genome. Homologs of these proteins were then identified in all genomes surveyed (tblastn), including those which did not present the cadherin protein model, and mapped on to corresponding genome scaffolds. The cadherin protein model could only be established as absent if flanking proteins were found on a single scaffold, with no cadherin sequence in between.

2.2. Experimental validation of E. dicksonii cadherin structural prediction

Total RNA was extracted from female gametophytes of *Macrocystis pyrifera* (strain CCAP 1323/1) infected with *E. dicksonii* (strain CCAP 4018/1) and cDNA was synthesised as described in (Zambounis et al., 2013) except that random hexamers were used. EST contigs were aligned against other oomycete cadherins and gaps in the *E. dicksonii* sequence were predicted. PCR amplification of these missing regions was attempted with primers flanking each gap (listed in Table SI1-2), using the following program: 95 °C – 5 min 40×(95 °C - 30s, 47–60 °C 45 s, 72 °C - 60s) 72 °C – 5 min, using Qiagen Taq PCR master mix Kit (201445) with 0.75 pmol of each primer, 1× master mix and 1 μl of cDNA. Likewise, a primer pair (WWF & WWR, Table SI1-2) was designed to directly confirm the presence of a WW domain amongst a stretch of acidic aminoacid residues.