Infection, Genetics and Evolution 39 (2016) 201–211

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Research paper

Characterization of parasite-specific indels and their proposed relevance for selective anthelminthic drug targeting

Qi Wang [a,1], Esley Heizer [a,1], Bruce A. Rosa [a], Scott A. Wildman [b], James W. Janetka [c], Makedonka Mitreva [a,d,e,*]

[a] McDonnell Genome Institute, Washington University in St. Louis, St. Louis, MO, USA
[b] Carbone Cancer Center, School of Medicine and Public Health, University of Wisconsin, Madison, WI, USA
[c] Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, 660 South Euclid Ave., St. Louis, MO, USA
[d] Department of Internal Medicine, Washington University School of Medicine, St. Louis, MO, USA
[e] Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA

ARTICLE INFO

Article history:
Received 6 October 2015
Received in revised form 29 December 2015
Accepted 28 January 2016
Available online 30 January 2016

Keywords:
Insertions and deletions
Adaptation
Spatial structures of proteins
Drug targets
Selective targeting
Parasites
Nematodes

ABSTRACT

Insertions and deletions (indels) are important sequence variants that are considered as phylogenetic markers that reflect evolutionary adaptations in different species. In an effort to systematically study indels specific to the phylum Nematoda and their structural impact on the proteins bearing them, we examined over 340,000 polypeptides from 21 nematode species spanning the phylum, compared them to non-nematodes and identified indels unique to nematode proteins in more than 3000 protein families. Examination of the amino acid composition revealed uneven usage of amino acids for insertions and deletions. The amino acid composition and cost, along with the secondary structure constitution of the indels, were analyzed in the context of their biological pathway associations. Species-specific indels could enable indel-based targeting for drug design in pathogens/parasites. Therefore, we screened the spatial locations of the indels in the parasite's protein 3D structures, determined the location of the indel and identified potential unique drug targeting sites. These indels could be confirmed by RNA-Seq data. Examples are presented illustrating the close proximity of some indels to established small-molecule binding pockets that can potentially facilitate selective targeting to the parasites and bypassing their host, thus reducing or eliminating the toxicity of the potential drugs. This study presents an approach for understanding the adaptation of pathogens/parasites at a molecular level, and outlines a strategy to identify such nematode-selective targets that remain essential to the organism. With further experimental characterization and validation, it opens a possible channel for the development of novel treatments with high target specificity, addressing both host toxicity and resistance concerns.

1. Introduction

The phylum Nematoda is one of the largest and most diverse phyla on the planet. At least 25,000 distinct nematode species have been described and it is estimated that the actual species count may go well into the millions (Hugot et al., 2001). Members of this phylum are found in hot springs, polar ice, and almost everywhere in between, and the lifestyles of these organisms vary from free-living to parasitic organisms (which are found in plants, vertebrates, insects, and even other nematodes). Plant and animal parasitic nematodes are of special concern because of their detrimental effect on the economy and global health. It is estimated by the WHO that 2.9 billion people are infected with parasitic nematodes (Hotez et al., 2007). In addition parasitic nematodes cost the agricultural industry more than $80 billion per year in crop treatment and lost product (Nicol et al., 2011). Currently anthelminthic drugs utilized to treat and prevent nematode infections are becoming less effective as drug resistance increases among populations (e.g., (Wolstenholme et al., 2004; Wrigley et al., 2006)). As resistance increases, drugs with novel mechanisms of action and/or alternate therapeutic approaches for control are needed to combat these parasites.

In the past decade, fast-evolving DNA and RNA sequencing technology has greatly enriched our understanding of many organisms (including many nematodes) from a genomic perspective. The rapid growth of genome information for nematodes has led to many in-depth studies of their genetics, genomics and functional evolution (Brindley et al., 2009; Dieterich and Sommer, 2009; Mitreva et al., 2007; Sommer and Streit, 2011). This genomic data can also be exploited to better understand parasite adaptations at a molecular level, and to facilitate the pursuit of novel treatments for prevention and/or control. Parasite genes or proteins are often examined in terms of their potential to serve as targets of

Abbreviations: PFC, protein family cluster; KO, KEGG orthology pathway; NemFam, PFC with at least 1 nematode species; RefFam, PFC with at least 1 non-nematode species; PDB, RCSB Protein Data Bank.

* Corresponding author at: McDonnell Genome Institute, Washington University in St. Louis, St. Louis, MO, USA.
E-mail address: mmitreva@genome.wustl.edu (M. Mitreva).

[1] Equal contribution.

http://dx.doi.org/10.1016/j.meegid.2016.01.025
1567-1348/

new treatments for parasite control. There are two main groups of proteins that can be exploited for these purposes: i) proteins that are specific to the parasite and ii) proteins that are highly homologous between the parasite and the host, but have diverged sufficiently to enable selective targeting in the parasite. These two groups of potential targets are non-overlapping and potentially provide promising targets for the development of drugs with low toxicity to the host.

Previous studies have examined drug targets unique to the target organism in order to minimize or eliminate toxic effects to the host (e.g. (Galperin and Koonin, 1999)), but if conserved (i.e. non-unique) essential proteins are eliminated from the target pool, then only a small fraction of the proteins are left for further exploration. For example, in a study of *Bacillus subtilis*, 96% of essential genes were found to be conserved in other bacteria and nearly 70% were found to be conserved in Archaea and Eukaryotes (Kobayashi et al., 2003). This indicates that if only the proteins that are unique to *B. subtilis* are examined for potential drugs, most of the proteome would need to be excluded. The majority of the nematode-specific proteins remain so distinct from host proteins that extrapolation of distant homology based on protein folds can only be used to infer putative functions for ~10% of the novel proteins (Yin et al., 2009). Therefore, selecting nematode-specific proteins as drug targets requires extensive experimental characterizations of their functions. This is also reflected in the fact that few of the current anthelmintics are targeted against species-specific or nematode-specific proteins.

On the other hand, proteins that are essential and conserved in multiple species are likely to be involved in core cellular processes (Kobayashi et al., 2003). A set of 458 core proteins shared among most eukaryotes has been previously defined (Parra et al., 2007), and these could prove to be more effective targets than species-unique proteins. However, unless differentiated regions are identified in order to facilitate specific targeting, there is a possibility of high toxicity to the host. The differential regions within these proteins can range from single amino acid changes to the insertion or deletion (indels) of multiple amino acids (Thorne, 2000). Indels have been shown to have a greater effect on protein structure and function than single amino acid changes that result from substitutions (Hormozdiari et al., 2009; Salari et al., 2008), and can also create a unique ligand binding site on the protein surface (Studer et al., 2013). It has been shown that indels rarely affect the structural scaffold of a protein, but much more often alter peripheral elements (Studer et al., 2013), which may lead to changes in binding sites that facilitate specific ligand binding.

A study (Wang et al., 2009) identified important roles of indels in nematode adaptation, but the focus was on the relevance of the indels for evolutionary adaptations, so many aspects related to the structural impact of the indels were not investigated. Comparisons of the homologous protein structures in proteins in the Protein Data Bank, has shown that the location of indels in a protein occur in a non-random manner; specifically, they tend to be located in loop regions more frequently than elsewhere (Fechteler et al., 1995). In one study up to 85% of indels were found in coiled regions of proteins (Pascarella and Argos, 1992). Indels have also been shown to vary in composition from other sections of proteins (Hsing and Cherkasov, 2008), and tend to be enriched in amino acids with small side chains and flanked by highly structured regions (Wrabl and Grishin, 2004). In a large scale study of bacterial/human homologs, sizeable indels were shown to exist in 5–10% of bacterial proteins with human homologs, and this number is even larger (~25%) for some protozoan pathogens (Cherkasov et al., 2006). A

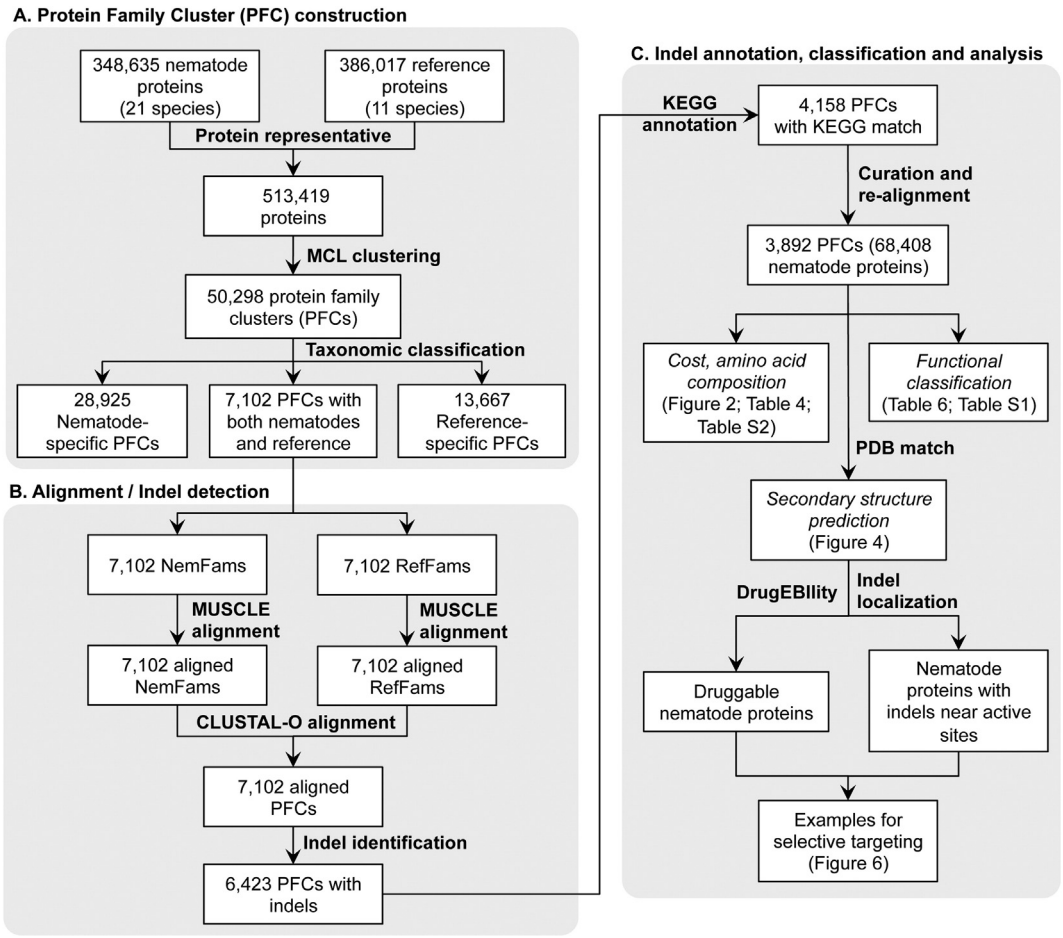

Fig. 1. Systematic identification, analysis and evaluation of nematode specific indels.