Molecular Phylogenetics and Evolution xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Molecular signatures that are distinctive characteristics of the vertebrates and chordates and supporting a grouping of vertebrates with the tunicates ☆

Radhey S. Gupta

Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, Ontario L8N 3Z5, Canada

ARTICLE INFO

Article history:
Received 29 April 2015
Revised 16 September 2015
Accepted 18 September 2015
Available online xxxx

Keywords:
Chordates
Vertebrates
Cephalochordates/urochordates
Evolutionary relationships
Conserved signature indels
Molecular signatures

ABSTRACT

Members of the phylum *Chordata* and the subphylum *Vertebrata* are presently distinguished solely on the basis of morphological characteristics. The relationship of the vertebrates to the two non-vertebrate chordate subphyla is also a subject of debate. Analyses of protein sequences have identified multiple conserved signature indels (CSIs) that are specific for *Chordata* or for *Vertebrata*. Five CSIs in 4 important proteins are specific for the *Vertebrata*, whereas two other CSIs are uniquely found in all sequenced chordate species including *Ciona intestinalis* and *Oikapleura dioica* (Tunicates) as well as *Brachiostoma floridae* (Cephalochordates). The shared presence of these molecular signatures by all vertebrates/chordate species, but in no other animal taxa, strongly indicates that the genetic changes represented by the identified CSIs diagnose monophyletic groups. Two other discovered CSIs are uniquely shared by different vertebrate species and by either one (*Ciona intestinalis*) or both tunicate (*Ciona and Oikapleura*) species, but they are not found in *Branchiostoma* or other animal species. Specific presence of these CSIs in different vertebrates and either one or both tunicate species provides strong independent evidence that the vertebrate species are more closely related to the urochordates (tunicates) than to the cephalochordates.

1. Introduction

The phylum *Chordata* comprises *Vertebrata* as well as two other non-vertebrate taxa *Urochordata* and *Cephalochordata* (Nielsen, 1995; Gee, 1996; Jefferies, 1986). The chordates in turn are part of the superphylum *Deuterostomes*, which also includes the phyla *Echinodermata*, *Hemichordata* and a recently described phylum *Xenoturbellida* (Nielsen, 1995; Gee, 1996; Jefferies, 1986; Bourlat et al., 2006; Blair and Hedges, 2005). Because *Vertebrata* contains all known vertebrate species, an understanding of its evolutionary relationship to the other chordates and deuterostomes is of central importance to zoology (Blair and Hedges, 2005; Philippe and Telford, 2006; Edgecombe et al., 2011; Springer et al., 2004; Bourlat et al., 2006; Delsuc et al., 2006). Of the two non-vertebrate chordate taxa, cephalochordates are morphologically more similar to the vertebrates than to the adult urochordates (tunicates); thus, they are traditionally considered to be the closest relatives of vertebrates (Nielsen, 1995; Gee, 1996; Jefferies, 1986). A grouping of cephalochordates with vertebrates to the exclusion of urochordates is also observed in a number of phylogenetic studies based primarily on small subunit (SSU) and large subunit (LSU) rRNA gene sequences (Cameron et al., 2000; Mallatt and Winchell, 2007; Winchell et al., 2002). Additionally, the genomic organization of the *Hox* genes in cephalochordates also suggests that cephalochordates are more similar phylogenetically to the vertebrates than to the tunicates, whose genomes are highly divergent and not informative in this regard (Pascual-Anaya et al., 2013; Swalla and Smith, 2008). In contrast to these studies, the interrelationships among different deuterostomes and metazoan phyla have been examined in detail based on large datasets of sequences for nuclear proteins (Delsuc et al., 2006, 2008; Bourlat et al., 2006; Blair and Hedges, 2005). Surprisingly, the results of these studies strongly indicate that the urochordates and not cephalochordates are the sister taxon to the vertebrates. The distal branching of the tunicates from vertebrates in earlier studies was shown to be an artifact of long-branch attraction attributed to rapid evolution within the tunicates (Tsagkogeorga et al., 2010; Delsuc et al., 2006). In some of these studies, monophyly of the phylum Chordata was ambiguous using various phylogenetic methods (Delsuc et al., 2006; Winchell et al., 2002; Cameron et al., 2000; Glenner et al., 2004).

The inference from recent studies that tunicates are the closest relatives of vertebrates is of much importance for understanding

☆ This paper was edited by the Associate Editor A. Larson.

E-mail address: gupta@mcmaster.ca

http://dx.doi.org/10.1016/j.ympev.2015.09.019
1055-7903/

the origin and evolution of vertebrates (Delsuc et al., 2006; Blair and Hedges, 2005; Bourlat et al., 2006; Swalla and Smith, 2008). Currently, the evidence that tunicates are more closely related to the vertebrates is entirely based on molecular phylogenetic studies (Delsuc et al., 2006; Blair and Hedges, 2005; Bourlat et al., 2006). However, several recent studies show that the inferences from molecular phylogenetic studies, even when they are based on large datasets involving multiple proteins, are sensitive to multiple confounding factors including differences in evolutionary rates among species, composition biases in sequences, on sampling of taxa, conflict in phylogenetic signal contained within the different amino acid sequences, and long-branch length attraction (Rokas et al., 2003; Jeffroy et al., 2006; Nosenko et al., 2013; Delsuc et al., 2008). Due to these factors, inferences from independent phylogenetic studies are often contradictory (Delsuc et al., 2006; Bourlat et al., 2006; Teeling and Hedges, 2013; Song et al., 2012; Philippe and Telford, 2006). Although in studies that group tunicates and vertebrates as a clade precautions were taken to guard against these artifacts (Delsuc et al., 2006; Blair and Hedges, 2005; Bourlat et al., 2006), it is important to confirm the relationship of vertebrates to the other chordate taxa by independent means.

The chordate as well as vertebrate clades are presently distinguished from other animals only on the basis of a limited number of morphological characteristics (Nielsen, 1995; Gee, 1996; Jefferies, 1986; Swalla and Smith, 2008). Besides the morphological characteristics, no other reliable molecular or biochemical property is known that is specifically shared by either all chordates or all vertebrates and can be used to distinguish these important groups of animals from all others. The availability of genome sequences from large numbers of animal species covering the diversity of metazoan taxa now provides a valuable resource for identifying novel molecular markers that are diagnostic for different animal taxa. Conserved signature indels (CSIs) in protein sequences constitute one type of rare genetic changes (RGCs) that provide very useful markers for this purpose, and they have been used extensively for evolutionary and systematic studies (Rokas and Holland, 2000; Springer et al., 2004; Baldauf and Palmer, 1993; Rivera and Lake, 1992; Gupta et al., 1994; Gupta, 1998, 2014; Bhandari et al., 2012). Although the shared presence of CSIs in protein sequences in some cases can represent homoplasy or lateral gene transfers (Bapteste and Philippe, 2002; Gupta, 2012), in general, when a conserved indel of a definite length is found uniquely in a phylogenetically related group of organisms, its most parsimonious explanation is inheritance from the most recent common ancestor (Nielsen, 1995; Gee, 1996; Jefferies, 1986). Thus, conserved signature indel(s) provide powerful means to support or refute a given phylogenetic hypothesis.

In the present work, I have examined sequence alignments of >3000 proteins from different metazoan species to identify conserved signature indels that are specific for either the *Vertebrata* or groupings of vertebrates with other animal taxa (particularly the other chordate lineages). These studies have identified seven CSIs in 6 widely distributed proteins that are uniquely found in either all sequenced vertebrate species, or all sequenced chordate species, but which are not present in any other animal groups/phyla. The unique shared presence of these CSIs (synapomorphies) in these animal groups provides evidence that these groups are monophyletic, and the identified characteristics provide novel molecular means for distinguishing members of these groups from other animal taxa. Additionally, the present study has identified 2 other CSIs in widely distributed proteins that are uniquely shared by all sequenced vertebrate species and the urochordate species (*Ciona intestinalis* and *Oikopleura dioica*), but which are not found in *Branchiostoma* (cephalochordate) or other deuterostome species. The specific presence of these CSIs in these two chordate lineages provides strong and independent confirmation that the vertebrate species are more closely related to the tunicates (or *Urochordata* subphylum) than to the cephalochordates.

2. Materials and methods

Identification of conserved signature indels that are specific for the chordates or vertebrates was performed as described in earlier work (Gupta, 2014, 1998; Gupta and Golding, 1996). Briefly, for these studies, Blastp searches were performed on >2000 proteins from the genome of *Ciona intestinalis* (Satou et al., 2008). For each protein for which high-scoring homologs were found in assorted vertebrates as well as non-vertebrate species, sequences for 20–25 homologs from divergent chordates and other animal taxa were retrieved, and their multiple sequence alignments were created using Clustal X 1.83 (Jeanmougin et al., 1998). Additionally, multiple sequence alignments for large numbers (>1000) of other proteins from diverse eukaryotic lineages were also utilized in this work. The alignments were visually inspected to identify any conserved insert or deletion (indel) that was flanked on both sides by at least 5–6 identical/conserved residues in the neighboring 30–40 amino acids, and which was uniquely found in either different chordates or vertebrate species. The indels that were not flanked on both sides by conserved regions, or those limited to other clades of animals were not further evaluated in this work. The selected indels of interest were further evaluated by performing repeat Blastp searches on the indels and their flanking conserved regions. In most cases, top 500 hits were examined to determine the taxon specificity of the observed indels (Gupta, 2014; Naushad et al., 2014). The results of these Blast searches were processed using SIG_CREATE and SIG_STYLE programs to construct signature files (Gupta, 2014). In the main figures, the results for the presence or absence of the indels in different groups are shown for only a limited number of representative species. However, unless otherwise indicated, the CSIs described here are specific for the indicated clades of animals, and similar CSIs were not present in other animals within the 500 blast hits.

3. Results

The aim of this study is to identify novel molecular markers in the form of conserved signature indels in protein sequences specific for either the Vertebrata or other chordate subphyla. The presence or absence of conserved indels in gene/protein sequences is generally not affected by factors such as differences in the evolutionary rates among species, composition biases, long-branch attraction, etc., which significantly affect phylogenetic analyses of base substitutions (Rokas and Holland, 2000; Gupta, 1998, 2014; Springer et al., 2004). Thus, the presence of such molecular markers as synapomorphies for a given group of species generally provides strong evidence that the species harboring the given signature form a monophyletic group (Baldauf and Palmer, 1993; Bhandari et al., 2012; Gupta, 2014; Rivera and Lake, 1992). To identify conserved indels, multiple sequence alignments of >3000 proteins from assorted animal taxa were created and then examined for the presence of conserved inserts or deletions that were restricted to members of the groups of interest. These studies have identified a number of conserved indels that are specific for either all sequenced vertebrate species, or which provide information regarding the relationship of vertebrates to other chordate lineages.

3.1. Molecular signatures that are specific for the vertebrates

Of the molecular signatures identified in this work, five are specific for the subphylum *Vertebrata*. Two examples of the

Please cite this article in press as: Gupta, R.S. Molecular signatures that are distinctive characteristics of the vertebrates and chordates and supporting a grouping of vertebrates with the tunicates. Mol. Phylogenet. Evol. (2015), http://dx.doi.org/10.1016/j.ympev.2015.09.019