Molecular Phylogenetics and Evolution 68 (2013) 471–481

Contents lists available at SciVerse ScienceDirect

Molecular Phylogenetics and Evolution

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

Elongation factor-1α, a putative single-copy nuclear gene, has divergent sets of paralogs in an arachnid

Ronald M. Clouse [a,*], Prashant P. Sharma [a], Gonzalo Giribet [b], Ward C. Wheeler [a]

[a] American Museum of Natural History, Division of Invertebrate Zoology, Central Park West at 79th St., New York City, NY 10024, USA
[b] Museum of Comparative Zoology, Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA

ARTICLE INFO

Article history:
Received 11 September 2012
Revised 21 March 2013
Accepted 22 April 2013
Available online 10 May 2013

Keywords:
Concerted evolution
Phylogenetics
Intrachromosomal gene conversion
Opiliones
Cyphophthalmi

ABSTRACT

Identification of paralogy in candidate nuclear loci is an important prerequisite in phylogenetics and statistical phylogeography, but one that is often overlooked. One marker commonly assumed to be a single-copy gene and claimed to harbor great utility for inferring recent divergences is elongation factor-1α (EF-1α). To test this hypothesis, we systematically cloned EF-1α in three disjunct populations of the harvestman *Metasiro americanus*. Here we show that EF-1α has a large number of paralogs in this species. The paralogs do not evolve in a concerted manner, and the paralogs diverged prior to the population divergence. Moreover, the paralogs of *M. americanus* are not comparable to the highly divergent EF-1α paralogs found in bees and spiders, which are easily recognized and separated through the use of specific primers. We demonstrate statistically that our detection of paralogs cannot be attributed to amplification error. The presence of EF-1α paralogs in *M. americanus* prevents its use in statistical phylogeography, and the presence of out-paralogs argues against its use in phylogenetic inference among recently diverged clades. These data contradict the common assumption that EF-1α is for most or all taxa a single-copy gene, or that it has a small number of paralogs that are homogenized through gene conversion, unequal crossing over, or other processes.

1. Introduction

Paralogy is a commonly observed genomic phenomenon whereby a gene undergoes duplication and the duplicated copies are retained (Fig. 1). Often, a gene duplication event occurs after a speciation event, with resultant paralogs restricted to individual species, forming what are termed "in-paralogs" (Sonnhammer and Koonin, 2002). However, genes can also duplicate before speciation events, engendering copies called "out-paralogs" in multiple extant lineages (Dolinski and Botstein, 2007; Sonnhammer and Koonin, 2002; Studer and Robinson-Rechavi, 2009). The evolutionary history of out-paralogs is of great biological interest, particularly when out-paralogs undergo sub-functionalization, the division of the original gene's function into multiple functions, or neo-functionalization, the acquisition of new functions altogether (Duarte et al., 2006; Hurles, 2004; Prince and Pickett, 2002). Examples of well-studied out-paralogs include the metazoan Hox gene cluster (Escriva et al., 2006; Roth et al., 2007; Taylor and Raes, 2004), the ABC gene family in angiosperms (Higgins, 1992; Holland and Blight, 1999), and the opsin family (Feuda et al., 2012; Oakley et al., 2007; Spady et al., 2006).

Although they are a topic of great research value from the perspective of gene family evolution, out-paralogs can be problematic occurrences for molecular phylogenetics, since phylogenetic reconstruction is based on analyzing fixed mutations in homologous structures (in this case genetic loci) that have become separated through population isolation or speciation, what in genetics are called "orthologs" (Bailey et al., 2003; Koonin, 2005). If paralogy is known to occur in phylogenetic markers, failure to correctly assign orthology to out-paralogs can result in fallacious species trees, particularly if gene copies are lost in a clade-specific manner. Ideally, out-paralogs acquire sufficient mutations to render them easily distinguishable from one another, becoming important and identifiable elements of the genome in their own right. If originating far enough back in time, they can be deployed as useful and independent markers of historical speciation and population genetic events (Carraro et al., 2012; Michez et al., 2009). Another ideal condition for phylogenetics is the other extreme: molecular mechanisms can maintain homogeneity among out-paralogs, spreading or eliminating new variations in the genomic collection of copies (the paralog "array") more quickly than speciation events, a process called "concerted evolution" (Hood et al., 1975; Liao, 1999; Zimmer et al., 1980). Homogenization renders orthology impossible to determine but also irrelevant for phylogeneticists; since Sanger sequencing calls nucleotide identities based on the most common nucleotide at that site in the array, rare variants

* Corresponding author. Fax: +1 (212) 769 5277.
E-mail address: rclouse@amnh.org (R.M. Clouse).

1055-7903/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ympev.2013.04.025

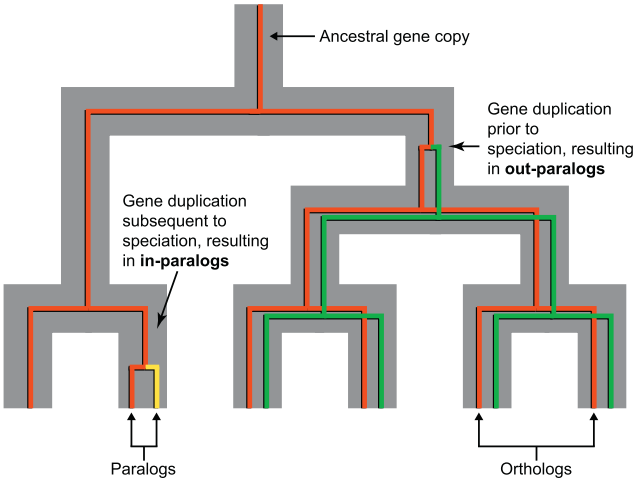

Fig. 1. Diagram showing orthologs and paralogs, the latter of which can be "in-paralogs" or "out-paralogs," depending on their origin relative to speciation events. Here, the thick, gray cladogram represents species histories (bifurcations being speciation events), and the thin red, green, and yellow lines show gene histories. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

are never seen. This is why, for example, the nuclear ribosomal array (18S, 5.8S, and 28S rRNAs, and the internal transcribed spacers ITS1 and ITS2) has been so commonly used in phylogenetics despite coming in hundreds (if not thousands) of copies; highly homogenized, the added copies actually enhance the ease with which the loci can be amplified and sequenced—but exceptions obviously exist (Carranza et al., 1996; Telford and Holland, 1997).

For this reason, molecular phylogeneticists have historically sought new molecular markers from the nuclear genome that are single- (or low-) copy or, alternatively, markers wherein any incident paralogs undergo concerted evolution. Such markers would provide at least three utilities: (a) constitute independent inferences of species trees, due to being unlinked (Friedlander et al., 1992); (b) be a powerful addition to traditionally used mitochondrial sequences in statistical phylogeography (Avise et al., 1987; Hare, 2001; Knowles, 2004); and (c) be easier to align than ribosomal genes due to the external alignment criterion of conceptual amino acid translations in exon sequences (Friedlander et al., 1992). One of the markers most commonly encountered in the literature as a putative single-copy gene is elongation factor-1α (EF-1α).

An essential and highly conserved member of the eukaryotic transcriptional apparatus, EF-1α has been characterized explicitly in many taxa as a single-copy nuclear gene (Aguileta et al., 2008; Danforth et al., 2004; Esseghir and Ready, 2000; Perlman et al., 2008; Regier et al., 2008, 2010), such that its copy number is rarely evaluated or the supporting evidence rarely cited. The common assumption of singly occurring or few copies of EF-1α has immediate implications for many downstream applications, not just molecular phylogenies and population genetics. For example, Faure et al. (2007) reported evidence of purifying selection on polymorphisms of the intron of EF-1α in a deep-sea bivalve genus but used an abundance of caution in accepting sequence variants, based on the presumption of low gene copy. As a consequence, many minor variants were attributed to PCR error and simply removed from the analysis. Additionally, Wilhelm et al. (2003) described a method for estimating genome size that uses real-time-PCR measures of amplicon product of *bona fide* single-copy genes. *Apropos*, Jeyaprakash and Hoy (2009) applied this method to estimate the size of a mite genome using EF-1α without testing for the incidence of paralogs.

Paralogy in EF-1α has in fact been reported in some taxa, although additional copies are typically believed to be few in number and/or divergent enough to be easily recognized or avoided. Two copies of the gene were discovered early in *Drosophila* ("F1" and "F2", not to be confused with elongation factor-2, often written as "EF-2"; Hovemann et al., 1988; Walldorf et al., 1985), and two copies have been found in representatives of all major bee families (Danforth and Ji, 1998). In both *Drosophila* and bees, the different copies are divergent enough to be easily recognized and selectively amplified with specific primers. In the spider genus *Habronattus*, two copies of EF-1α were found that differ considerably in size; the short version appears to be a pseudogene that lacks introns, and it, too, can be avoided with copy-specific primers (Hedin and Maddison, 2001). In *Artemia* (brine shrimp), EF-1α has been estimated to have 2–4 copies per haploid genome, but this was based on hybridization signals of a particular exon fragment (Lenstra et al., 1986), and further investigation is needed with updated technologies. Most recently, recent advents in transcriptomic data have revealed transcribed EF-1α paralogs in two distantly related protostomes, a nemertean and an earthworm (Riesgo et al., 2012).

Such hints of paralogy in EF-1α portend methodological and analytical challenges for its use in systematic biology, particularly given its widespread use in molecular phylogenies (Danforth et al., 2004; Hines et al., 2006; Kjer et al., 2006; Monteiro and Pierce, 2001; Regier et al., 2010; Sharma and Giribet, 2011). One advantageous group for testing the incidence of paralogs in EF-1α is Opiliones, the arachnid order commonly known as harvestmen and our taxon of interest. This order of arthropods has an ancient evolutionary history extending to the Devonian (Dunlop et al., 2004; Giribet et al., 2010, 2012b), numerous species easily distinguished by morphological characters, and a well-circumscribed phylogeny (Boyer and Giribet, 2007; Clouse and Giribet, 2010; de Bivort et al., 2010; Giribet et al., 1999, 2010, 2012a; Hedin and Thomas, 2010; Murienne and Giribet, 2009; Sharma and Giribet, 2009, 2011). Testing for paralogy in EF-1α is especially important for this group, because it has been recently reported that this gene is single- or low-copy in all Opiliones and thus of great utility for resolving ancient and recent divergences (using the exonic and intronic regions of the gene, respectively) (Hedin et al., 2010). Amplifiable regions of EF-1α are also well annotated, and thus exon–intron boundaries are known from multiple exemplars of the group (Hedin et al., 2010).

In order to test the incidence of paralogy in EF-1α and its effect on inferences of recent divergences, we sampled a species of cyphophthalmid Opiliones, *Metasiro americanus* (Davis, 1933), from all three known localities of its range in the southeastern U.S. Cyphophthalmi are well-known for being poor dispersers (Giribet, 2000; Juberthie, 1988), and we would expect these populations to have been separated for many millions of years, perhaps even being cryptic species. We amplified EF-1α using published and designed primers, then cloned amplicons, compared the number of sequence discrepancies to the empirically measured amplification error rate, and gauged the effectiveness of the resulting alignment to distinguish populations. We thereby tested the claim that EF-1α is a phylogenetically useful single-copy nuclear gene in arachnids.

2. Materials and methods

2.1. Specimen collection

During March, 2010, R.M.C. and P.P.S. collected specimens from known localities of *M. americanus* in Florida Caverns State Park in Jackson Co. (both a hillside and an upland glades area), Florida, Sassafras Mountain in Pickens Co., South Carolina, and Savannah