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

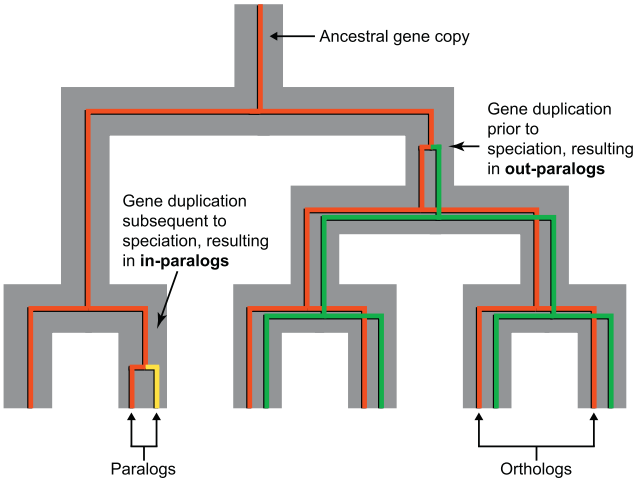

Fig. 1. Diagram showing orthologs and paralogs, the latter of which can be "in-paralogs" or "out-paralogs," depending on their origin relative to speciation events. Here, the thick, gray cladogram represents species histories (bifurcations being speciation events), and the thin red, green, and yellow 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

Table 1
Taxon names, collection and specimen numbers, and locality information for all specimens from which EF-1α paralogs were obtained in this study.

Taxon	Collection	Specimen	Location	Latitude	Longitude
Metasiro americanus	105643	7168, 7173	Jocosse Gorge, Sassafras Mt., off F. van Clayton Highway	35.06228	−82.79500
Metasiro americanus	105644	7175	Jocosse Gorge, Sassafras Mt., off F. van Clayton Highway	35.06231	−82.79500
Metasiro americanus	105645	7204, 7206	Kingfisher Pond, Savannah National Wildlife Refuge, Jasper Co., SC	32.18923	−81.08000
Metasiro americanus	105655	7392–95, 7397	Florida Caverns State Park, Jackson Co., FL	30.82119	−85.24639
Metasiro americanus	105662	7431	Florida Caverns State Park, upland glade, Jackson Co., FL	30.81433	−85.24667
Metasiro americanus	105663	7455	Florida Caverns State Park, upland glade, Jackson Co., FL	30.81433	−85.24667

Wildlife Refuge in Jasper Co., South Carolina (Table 1). Specimens were collected by sifting leaf-litter into white pans, which were then examined for live animals. These were then preserved in 95% EtOH and kept cool. Individual specimens are reported in alignments and trees by the Museum of Comparative Zoology DNA collection number and a Cyphophthalmi Biota database (Colwell, 2004–2011) specimen number, such as "105655-7393", meaning MCZ accession number DNA105655 and Biota specimen SPM007393. Specimen information can be obtained at the Museum of Comparative Zoology specimen database (mczbase.mcz.harvard.edu).

2.2. DNA extraction and amplification

DNA was extracted using the Qiagen DNeasy tissue kit (Qiagen, Valencia, CA, USA) and amplified using illustra™ Ready-To-Go™ PCR Beads (GE Healthcare, Little Chalfont, UK; manufacturer reported *Taq* error rate of 10^{-5}–10^{-4} errors/base/duplication). Shultz and Regier (2001) obtained exon-only EF-1α sequences for the Cyphophthalmi *Siro acaroides* (Ewing, 1923) and *Parasiro coiffaiti* Juberthie, 1956 using reverse transcription PCR (RT-PCR) (Regier and Shultz, 1997). From these sequences, primers were developed by Hedin et al. (2010)—EF1-OP3, hereafter referred to as "1f" for simplicity, and EF1-OPRC4LEIO, "3r"—which amplified the region around and including Intron III in another Cyphophthalmi, *S.* cf. *kamiakensis* (Newell, 1943) (Fig. 2). We initially used primers 1f/3r to amplify the Intron III region in *M. americanus*, and we used the resulting sequence to design more effective primers in this species (Table 2).

2.3. Cloning

We generated amplicons using three different primer pairs, 1f/1r, 2f/2r, and 2f/3r, the resulting fragments of which are called here Fragments *1*, *2*, and *3*, respectively (Fig. 2; Table A.1). These PCR amplicons were cloned using the Invitrogen TA Cloning® kit and One Shot® Competent Cells, with a modified protocol to conserve vector and maximize transformation. The ligation reaction

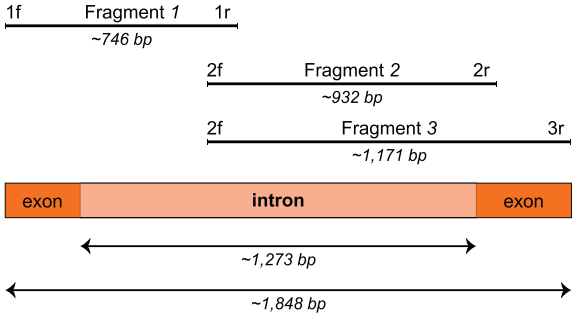

Fig. 2. Schematic of the EF-1α gene fragments amplified and sequenced in this study. Lengths are approximate due to the occurrence of various indels in the intron. Primers "1f" and "3r" correspond to EF1-OP3 and F1-OPRC4LEIO of Hedin et al. (2010).

Table 2
Primers used to target EF-1α gene fragments in this study (simplified/as originally reported or designed), amplified using temperature profiles "TD6"[1] and "TD53"[2] from Chenuil et al. (2010).

Primer name	Sequence (5′–3′)
1f/EF1-OP3[3]	TTT GAR GAA ATC CAR AAR GAA GT
2f/EF1aSNP2F	TGG AGA TAA GCC CCT TTA AAC A
1r/EF1aSNP3R	GAA AAA TGC ATA AAA GGG AAT CTT
2r/EF1-Ma-3R	ACG GGA GCA AAT GTA ACC AC
3r/EF1-OPRC4LEIO[3]	GAA CTT GCA AGC AAT GTG AGC

[1] TD6: [94 °C (1:00 min), 58–45 °C (1:00), 72 °C (1:00)], 25 × [94 °C (0:30), 58 °C (0:45), 72 °C (0:45)].
[2] TD53: [94 °C (1:00 min), 63–53 °C (1:00), 72 °C (1:00)], 4 × [94 °C (1:00), 53 °C (1:00), 72 °C (1:00)], 25 × [94 °C (0:30), 58 °C (0:45), 72 °C (0:45)].
[3] Hedin et al. (2010).

consisted of the following mixture: 0.4 μL PCR product, 0.2 μL 10X Ligation Buffer, 0.4 μL pCR®2.1 vector (25 ng/μL), 0.8 μL water, and 0.2 μL T4 DNA Ligase (4.0 Weisse units). The ligation mixture was refrigerated for 48 h and then added to 25 μL of competent cells. After transformation, 125 μL of S.O.C. medium was added, and the cells were placed in a shaking incubator at 37 °C for 3–4 h. Two colony plates were then made for each transformation, using 50 μL and 100 μL of solution, which were then incubated overnight at 37 °C. Approximately 12 randomly picked colonies were then amplified and sequenced with the M13 primer pair according to the manufacturer's protocol.

2.4. Sequence analysis

Sequences were proofread in Sequencher 5.0, inspected in SeaView 4.3.2 (which uses MUSCLE and ClustalW for alignment) (Edgar, 2004; Gouy et al., 2010; Larkin et al., 2007) and BioEdit 7.0.5.3 (Hall, 2007), aligned using MAFFT 6.847 (Katoh et al., 2002), and unique haplotypes were identified using DnaSP v. 5 (Librado and Rozas, 2009).

2.5. Phylogenetic analysis

For phylogenetic analyses of EF-1α paralogs, we first combined sequences of Fragments *2* and *3* taken during our initial survey from all localities. Then the alignment was trimmed to only to the 2r primer region for all sequences so there were no missing data, since the Sassafras Mt., Savannah R., and upland glades are of Florida Caverns only had sequence to the 2r primer. These unaligned sequence data (ranging from 895 to 937 bp) were analyzed in POY 4.1.3 (Varón et al., 2009) under the parsimony criterion with the cost scheme "3211" (gap openings cost 3, transversions 2, and transitions and gap extensions 1). To estimate the phylogenetic content of the sequences outside of the large (>20 bp) indels, which were locality-specific and would thus necessarily define large, biogeographically relevant clades under most cost schemes, the implied alignment from the first analysis was used to remove these indel regions, and the data (871 bp) were read as prealigned and analyzed again in POY under the same 3211 cost scheme. The

sequences without large indels were also read into the program MEGA4 (Tamura et al., 2007), which output a pairwise matrix of transition and transversion differences among the sequences.

We also created an alignment of all *M. americanus* EF-1α sequences generated by us (Fragments *1–3*) in combination with EF-1α sequences from other Opiliones available on GenBank. For those outgroups that had sequence for Intron III, it was so divergent from *M. americanus*, we coded all intron sequence for the outgroups as missing data, thus allowing rooting to be determined solely by the more conserved exonic regions. A phylogenetic analysis of this data set was run on RAxML (Stamatakis et al., 2008) on the CIPRES computing cluster (Miller et al., 2010) and POY (prealigned and using the 3211 transformation cost matrix, described above).

2.6. PCR error analysis

To measure the number of discrepancies introduced by *Taq* polymerase when amplifying DNA, we amplified a known homogenous template and used cloning to separate and inspect individual amplicons. The homogenous template came from one clone each of two different types of EF-1α sequences (distinguished by a 28-bp indel, discussed below). We amplified the EF-1α fragment from these clones using the Fragment *3* primer pair (2f/3r) and temperature profile (TD53, Table 2) originally used with genomic DNA, and we then cloned and sequenced the resulting amplicons. We compared the number and distribution of sequence discrepancies derived from amplifying this homogenous template to sequences of the same fragment from the same animal (105655-7397) amplified from genomic DNA.

3. Results

3.1. General description of cloning products

From our initial survey of EF-1α in different individuals and populations of *M. americanus*, we cloned and sequenced 133 amplicons from 12 individuals taken in six collections from the three main areas of its distribution. Most paralog sequences (97) were generated from individuals in the Florida Caverns collection 105655, which consisted of animals sifted from less than one square meter of leaf litter alongside a hillside log. For animals of this collection, Fragment *1* of EF-1α was 745–747 bp long, and we successfully sequenced 32 clones of it from five different animals. Of these clones, 25 were unique, and one was shared between two individuals. Specimen 7393 had ten unique clones of Fragment *1* sequenced, which covered three different lengths (Table 3).

Fragment *2* from collection 105655 was 904–932 bp long and was successfully cloned and sequenced from five different animals; 28 clones were recovered, 25 of which were unique. Specimen 7394 had 8 different copies of Fragment *2* sequenced, which included one with a single indel and another with a synonymous substitution in the exon region.

Fragment *3* from collection 105655 ranged in length from 1170 to 1199 bp, and from four different animals we initially cloned and sequenced 37 copies of it, 31 of which were unique. Specimen 7393 produced ten different sequences, none of which was recovered more than once and six of which had mutations in the exon; five of the exon substitutions were non-synonymous. Specimen 7397 originally produced nine different sequences, all of which were unique and covered four different lengths; after we picked and sequenced additional clones to assess *Taq* error, we ultimately had 32 sequences of Fragment *3* from specimen 7397, 24 of which were unique.

The six paralogs from the upland glade area of Florida Caverns, which is about 750 m from the hillside collection 105655, were all unique but not generally distinct from those from the hillside; four unique sequences were recovered from the upland glade specimen 7455.

Of the 26 paralog sequences recovered from four Sassafras Mt. specimens, 24 were unique, and one was shared between two different specimens. All 11 sequences recovered from specimen 7168 were unique, as were 11 of the 12 sequences recovered from specimen 7170.

All four paralog sequences recovered from two Savannah River individuals were unique.

GenBank accession numbers for unique sequences from each specimen are shown in Table A.1.

3.2. EF-1α paralog sets

Intron III of EF-1α from the Florida Caverns population had a 28-bp indel, and five positions in various locations of the intron region had nucleotide differences that correlated with this indel; all copies with the 28 bp had G, C, T, A, and A at those sites, and all sequences missing this indel had A, T, C, T, and C at the same positions, respectively. We refer to paralogs that have the same condition for this indel (and, apparently, other sites in the intron)

Table 3
The distribution of indels among sequences recovered for EF1α from the Florida Caverns collection locality 105655. The number of sequences of different lengths is noted, as well as the specimens from which they were cloned. Asterisks indicate specimens that have three or more sequence lengths for the same fragment. Indel positions (i–vi) and the number of bases lost at each are indicated to the right.

	Length (bp)	Specimens	*n*	Position and bp loss			
Fragment *1*				i			
	747	7393*	2	–			
	746	7392, 7393*, 7394, 7395	29	(1)			
	745	7393*	1	(2)			
Fragment *2*				i	ii	iii	iv
	932	7393, 7394, 7395	17	–	–	–	–
	931	7397	1	(1)	–	–	–
	931	7395	1	–	(1)	–	–
	931	7394	1	–	–	(1)	–
	904	7392	7	–	–	–	(28)
Fragment *3*				iv	v	vi	
	1199	7393, 7394, 7397*	21	–	–	–	
	1198	7397*	1	–	–	(1)	
	1171	7392, 7397*	15	(28)	–	–	
	1170	7397*	1	(28)	(1)	–	

as belonging to the same "set," to differentiate them from the whole collection of paralogs in the population (the paralog "array") or alternative versions of a single homologous locus ("alleles").

We subsequently tested for the independent assortment of these sets by assessing their occurrence in a larger sample of individuals. This was done by amplifying a 108-bp conserved region flanked by primers 2f and 1r. This region contained one of the variable nucleotides consistently associated with the 28-bp indel, and we genotyped 40 individuals from the same small hillside area as collection 105655 in Florida Caverns State Park for this polymorphism (specimens 7369–7398 and 7406–7415 from collections 105654–105655, and 105657). Four individuals were homozygous TT, 23 homozygous CC, and 13 heterozygous; this differed little from the expected values (3, 22, and 15, respectively) if these two variants sort independently, and the population is in Hardy–Weinberg equilibrium ($p = 0.60$, χ^2 test).

3.3. EF-1α phylogenetic analysis

The phylogenetic analysis of EF-1α clones from three different populations resulted in 14 equally parsimonious trees of length 326, the consensus of which recovered three distinct clades: all sequences from Sassafras Mountain, all Florida caverns sequences with the 28-bp indel, and all Florida Caverns sequences without the 28-bp indel (Fig. 3a). Three of the four sequences from the Savannah River Wildlife Refuge formed a clade, the most derived two of which were distinguished by a 38-bp indel, but the relationship among this clade, the remaining Savannah sequence, and the Sassafras Mt. clade was not resolved. Nucleotide substitution differences were also low between the short Florida paralogs and those from Sassafras Mt. and Savannah R. (Table 4). Analyzing the same fragment as prealigned and with the 28-bp and 38-bp indels regions removed resulted in a two equally parsimonious trees of length 209, which, when used to make a strict consensus with zero-length branches and ambiguously reconstructed nodes collapsed, recovered two clades: Sassafras Mountain and the Florida Caverns long sequences (Fig. 3b). The relationship among these clades and all but one of the Florida Caverns short sequences plus those from Savannah River was not resolved. Analyzing these data, with the indels, under maximum likelihood in RAxML recovered a tree very similar to that of the one from POY with the large indels removed, since RAxML treats gaps as missing data (Fig. 3c).

Analysis of the EF-1α sequences for *M. americanus* with 34 other Opiliones in RAxML resulted in a phylogeny of *M. americanus* EF-1α paralogs rooted in the short set found in Florida Caverns (Fig. 4). The other three types of paralogs—the long set from Florida, those from the Savannah River, and those from Sassafras Mt.—were recovered as clades that diversified independently out of the short set. The alignment analyzed in POY naturally separated the sequences by locality and short- and long-set paralogs, since they are easily distinguished by large indels, which contribute to the tree cost in POY (Fig. A.1). Nonetheless, it recovered the long Florida paralogs as diversifying before the populations separated, and the short paralogs and the sequences from Sassafras Mt. and Savannah R. as diversifying out of them.

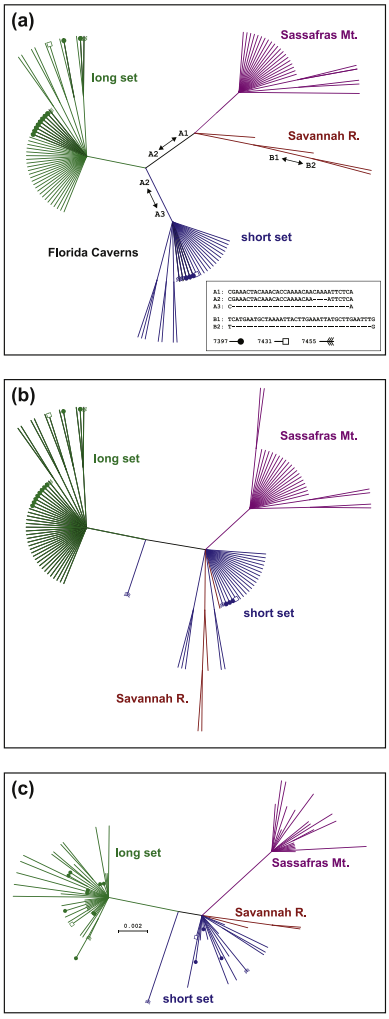

Fig. 3. Unrooted phylogenetic hypotheses for EF-1α sequences recovered from three different populations of *Metasiro americanus*, analyzed using three different methods: (a) unaligned sequences between primer sites 2f and 2r analyzed using dynamic homology in POY 4.1.3 (Varón et al., 2009) under the parsimony criterion, (b) the same fragment, prealigned and with the large indel regions removed, analyzed in POY under parsimony, and (c) the same fragment, prealigned and analyzed under the maximum likelihood criterion in RAxML (which treats gaps as missing data, Stamatakis et al., 2008). Inset in panel (a) shows states for two large indels recovered in the intron of Florida and Savannah sequences, and a key to individuals that carried both long and short paralogs.

Table 4
The number of transversion and transition differences (min.–max.) between EF-1α paralogs analyzed phylogenetically. Minimal distances of zero indicate that the most similar paralogs differed only by indels.

	Sassafras Mountain	Savannah River Wildlife Refuge	Florida Caverns (short)	Florida Caverns (long)
Sassafras Mt.	0–7	–	–	–
Savannah R.	6–11	1–5	–	–
FL Caverns (short)	4–13	2–8	0–8	–
FL Caverns (long)	6–16	4–11	2–11	0–8

3.4. PCR and cloning error

From clones derived from amplification of genomic DNA from specimen 7397, we had 21 sequences of the long version of Fragment 3 and 11 sequences of the short version. The distribution of discrepancies among these cloned sequences did not follow a Poisson distribution using the average number of observed differences as the distribution spread ($\lambda = 2.05$ for long paralogs and 3.55 for short paralogs; replicated *G*-test for goodness-of-fit: $p = 0.03$, df 10, G-statistic = 19.65; Fig. 5). The distribution was even more different from a hypothetical distribution based on an error rate of 1 base per 1000-bp fragment (reported by Palumbi and Baker (1994) and referenced by Faure et al. (2007); $p < 0.001$; Fig. 5). There exist various methods for estimating amplification error, and a brief discussion of them is provided in Appendix B.

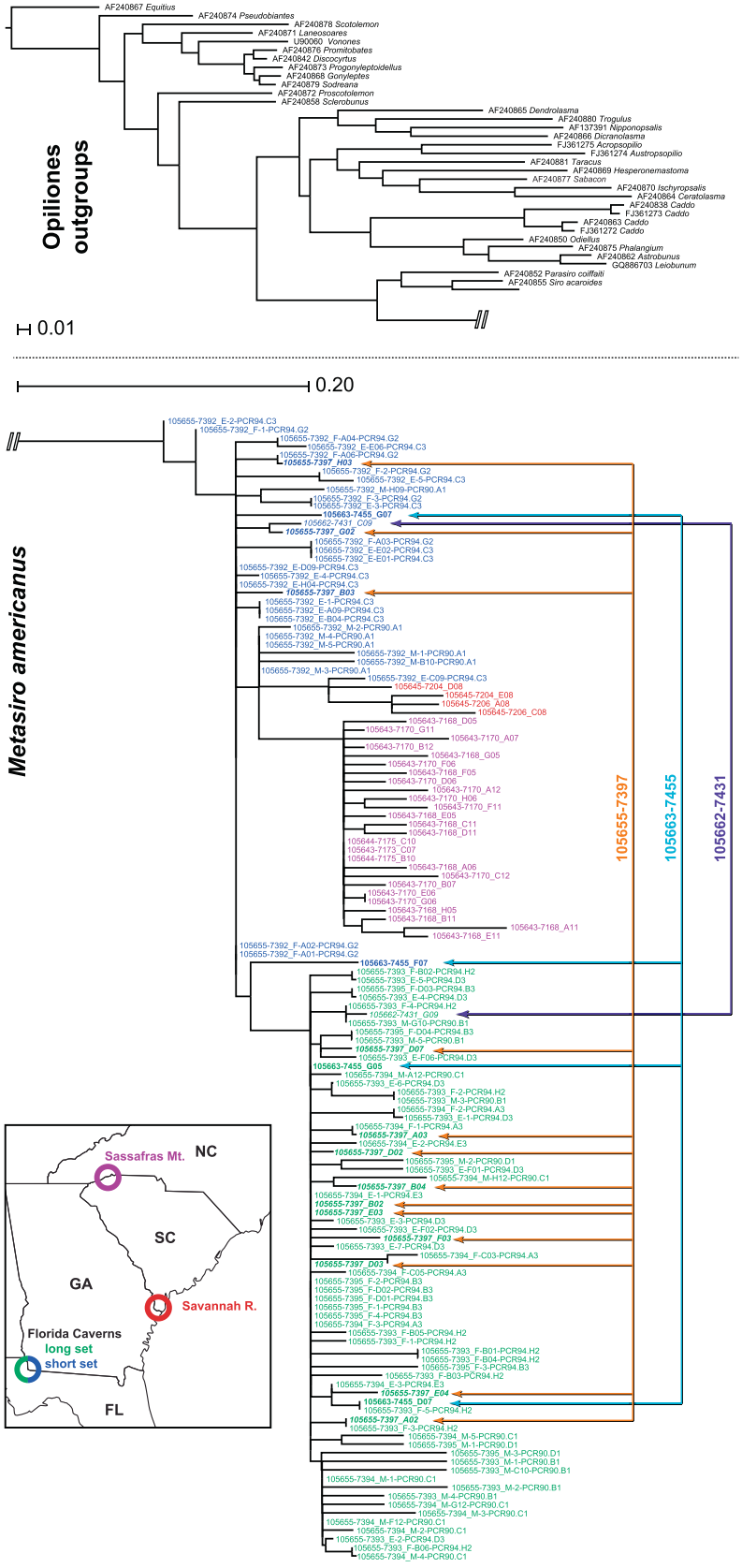

Fig. 4. Phylogeny of *Metasiro americanus* EF-1α sequences, rooted using outgroup sequences available on GenBank for other Opiliones and representing all four suborders, and analyzed under maximum likelihood. Paralog sequences for *M. americanus* are colored according to their locality: light purple for Sassafras Mt., red for Savannah R., and blue and green for Florida (depending on whether they are from the short or long set, respectively). Long and short paralogs from the same individual are indicated to the right, orange for 105655-7397, light blue for 105663-7455, and purple for 105662-7431. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

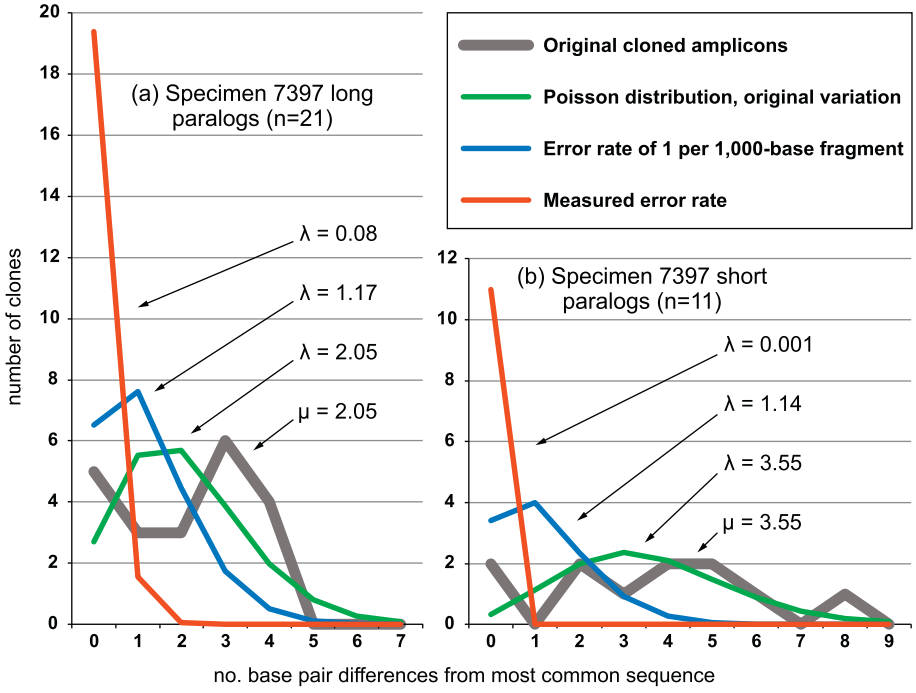

Fig. 5. Actual and expected distributions of nucleotide discrepancies among clones from Florida Caverns specimen 105655-7397, with long (a) and short (b) clones shown separately. The actual distributions are shown as a thick gray line, and the average number of discrepancies (μ) is noted. A Poisson distribution using the same level sample size and discrepancy level (μ becoming λ) is shown in green. A Poisson distribution based on the error rate of 1 per 1000-bp sized fragment (in this case, approximately 1171 bp) is shown in blue, and the same distribution using the measured error rate of 0.07 per 1000 (or 0.08 per 1171 for this fragment) is shown in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Amplifying homogenous DNA template, in this case taken from single clonal colonies containing a known paralog each, resulted in almost no discrepancies and thus an estimate *Taq* error several times lower than that commonly reported or advised by the manufacturer (1 error per 10^4–10^5 nucleotides per duplication, Clarke et al., 2001; Tindall and Kunkel, 1988): 35 of 37 long paralog sequences were identical to each other and to the original sequence obtained from that colony, and all 25 short paralog sequences were identical to each other and to the original sequence obtained from that colony. The discrepancies among the long paralogs amplified from homogenous DNA (differing by one and two nucleotides) were not significantly different from a Poisson distribution ($p = 0.12$; *G*-test for goodness-of-fit; not appropriate for the short paralogs, since no discrepancies were detected). The calculated error rate of 0.068 per 1000 bp for the long paralogs (and effectively zero for the short paralogs but put at 0.001 for calculation purposes) generated expected Poisson distributions of discrepancies extremely different from the one observed from clones derived from genomic template ($p < 0.001$; Fig. 5).

4. Discussion

4.1. Multiple paralogs of EF-1α occur in Metasiro americanus

We detected a large number of EF-1α paralogs in *M. americanus* by cloning amplicons of the gene. These paralogs had a wide array of nucleotide discrepancies, from single base substitutions to large indels. There were no idels in the exons, but there were a number of nonsynonymous substitutions. Specimens from which we successfully sequenced several clones in our initial survey gave us as many as 11 different versions of EF-1α from one individual, and after collecting additional clonal sequences from specimen 7397 to asses *Taq* error, we ultimately recovered 24 unique EF-1α sequences (out of 32 total) from that animal. In general, EF-1α sequences from single individuals, and even from single collections (which cover individuals living within 1 m of each other), were mostly unique.

The recovery of nonsynonymous substitutions in the exons of some paralogs does not permit inference of which EF-1α copies are functional, although this could be tested by cloning the same region from cDNA to see which EF-1α copies are transcribed. As the focus of our investigation was on the informativeness of the gene for population-level divergences (i.e., the intronic regions), we did not pursue this avenue of inquiry here. However, we note that paralogies affecting the exonic regions—the parts of the gene used for inferring interspecific relationships—may in turn affect reconstructions of species trees, a possibility that has not been explored in empirical Opiliones phylogenies that have used EF-1α (Derkarabetian et al., 2010; Sharma and Giribet, 2011).

4.2. Detected paralogous copies cannot be attributed to PCR amplification errors

A common explanation for sequence variations among cloned amplicons in empirical datasets is PCR error (Faure et al., 2007; Song et al., 2008; Thornhill et al., 2007; Whittall et al., 2006). However, our analyses demonstrate that the discrepancies we detected among various EF-1α sequences cannot be attributed to mis-incorporation of nucleotides by *Taq* DNA polymerase during amplification. We have three lines of evidence for this. (1) From single individuals we recovered three or more paralogs that differed by indels, not just nucleotide substitutions (Table 3). Indels in non-repetitive sequences (Clarke et al., 2001) have been measured as occurring 5–11 times less frequently than nucleotide substitutions (Kunkel, 1984; Tindall and Kunkel, 1988). (2) The distributions of discrepancies among different sequences from the same individual did not follow a Poisson distribution, which a random process like mis-incorporation of nucleotides must produce. (3) The mean number of nucleotide discrepancies among our EF-1α paralogs

was over 25 times greater than what would be expected from the actual *Taq* mis-incorporation error rate measured empirically. The distributions of sequence discrepancies derived from amplified genomic DNA had 0.01% probability of resulting from the same process as produced the two discrepancies found when amplifying homogenous template. Moreover, they had a similar probability (0.04%) of resulting from a *Taq* error rate commonly assumed to give one error per 1 kb of amplified and cloned sequence (Palumbi and Baker, 1994).

4.3. EF-1α copies are out-paralogs that disrupt inference of population divergence

In Florida Caverns, the EF-1α paralog array is divided into divergent sets, leading to incomplete homogenization of the entire array. For the region we amplified, these sets are distinguished by a 28-bp indel in the intron and five consistent nucleotide polymorphisms. Unrooted phylogenetic analyses of sequences with the large indel region removed recover a close relationship among the set of short Florida paralogs and the paralogs from Savannah R. and Sassafras Mt. This close relationship can also be seen in the low number of nucleotide substitutions between the short Florida paralogs and those of the other localities relative to the long Florida paralogs (Table 4). A rooted phylogenetic hypothesis recovered using maximum likelihood indicates that the short paralogs diverged before the populations of *M. americanus* separated, and the hypothesis recovered under parsimony and counting indels suggests the long paralogs diversified before all other paralogs in this species evolved. In either case, they are out-paralogs. Moreover, with the Sassafras Mt. and Savannah R. paralogs diversifying separately from within one set of short Florida paralogs, small samples of EF-1α sequences from the Florida population are likely to produce conflicting hypotheses of historical relationships among the populations.

The large number of paralogs that exist in *M. americanus*, together with the possibility that a modest indel alone can be sufficient to isolate paralogs from gene conversion (discussed below), suggest that paralogy assessment through cloning is an ineluctable prerequisite to using EF-1α in other Opiliones and arthropods generally. This is especially true for statistical phylogeographic studies, which are based on distributions of ortholog variants and are thus completely undermined by paralogy. Recent Opiliones studies have employed EF-1α in phylogenetics at both species and populations levels (Derkarabetian et al., 2010; Sharma and Giribet, 2011; Shultz and Regier, 2001, 2009), but a variety of factors may inadvertently mitigate the effects of any EF-1α out-paralogs: consensus effects of Sanger sequencing, removal of intron sequence, analysis in combination with other markers, use to recover deep nodes, and low sequencing success relative to other markers. However, as analyses become more detailed and branch lengths are used in molecular dating and ancestral area reconstruction, the pernicious effects of using out-paralogs instead of orthologs become compounded.

4.4. Length variability of paralogs may inhibit interchromosomal gene conversion

The two paralog sets discovered in the Florida Cavers population are different from the long and short EF-1α paralogs discovered in bees and spiders; those are proposed to be a single pair of paralogs that reside in the same chromosome, not different sets of paralogs on different homologous chromosomes. The sets in *M. americanus* do appear to homogenize internally, often differing by only single base-pair substitutions and 1–2 bp indels. This contrasts sharply with the extensive differences between intron III among sequences for different Cyphophthalmi species, which bear almost no resemblance.

Linkage disequilibrium has been detected in paralogous arrays before (Arnheim et al., 1982; Parkin and Butlin, 2004); this has usually been taken as evidence for the importance of rapid intrachromosomal gene conversion (Bollag and Liskay, 1988; Copenhaver and Pikaard, 1996; Klein and Petes, 1981; Polanco et al., 1998; Scherer and Davis, 1980; Schlötterer and Tautz, 1994). If new mutants are able to engage rapidly in the non-reciprocal transfer of genetic information along their chromosome (either via sister chromatid exchange or on the same strand) faster than they can convert copies on a homologous chromosome, then identical sets of paralogs can appear on single chromosomes (Klein and Petes, 1981). Explored in various models, the different mechanisms and explanations for patterns of concerted evolution depend not only on relative rates of intra- and interchromosomal gene conver-

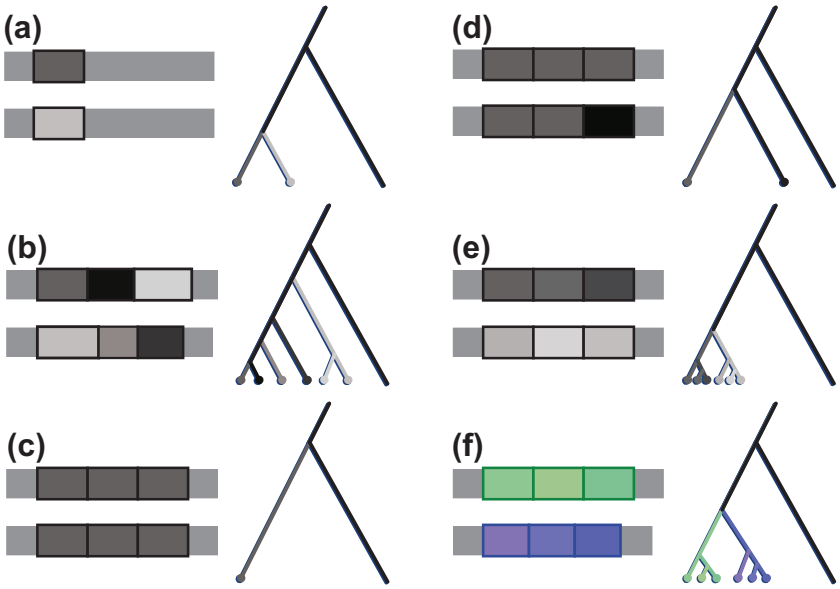

Fig. 6. Variation of nuclear markers: (a) a single locus with two alleles, (b) an array of poorly homogenized paralogs, (c) an array of homogenized paralogs which appear to be evolving "in concert;" (d) an array of homogenized paralogs that contain some divergent members, (e) an array of paralogs that is divided into different, homogenized sets on separate chromosomes, and (f) an array where different sets vary by size (e.g., EF-1α in *M. americanus*).

sion, but also selection advantages and population dynamics (Nagylaki, 1984; Nagylaki and Petes, 1982; Slatkin, 1986). In addition, models need to factor in the primary explanation for sequence homogenization besides gene conversion: unequal crossing over (Hillis et al., 1991; Smith, 1974, 1976). Still, length variation has long been suggested as an important factor in gene conversion (Zimmer et al., 1980), and it seems more likely that the indel in the Florida Caverns EF-1α paralog sets has simply blocked interchromosomal gene conversion through alignment inhibition, not that intrachromosomal gene conversion is especially quick or efficient. Indeed, at the rate we found novel EF-1α sequences, it seems that intrachromosomal gene conversion is a rather sloppy or slow process.

A variety of outcomes are possible after gene duplication and the emergence of a paralog array, each of which should produce certain gene trees, some of which are shown in Fig. 6. Should the paralogs not undergo concerted evolution, they will evolve separately and accumulate mutations that will reveal deep histories (Fig. 6b), but if they are homogenized, they will appear as one copy and one allele (Fig. 6c). Even with concerted evolution, gene conversion can miss some copies, which, as they diverge over time, become more difficult to homogenize with the remainder of the array (Fig. 6d). If intrachromosomal gene conversion is far more effective than interchromosomal gene conversion, paralog sets can be sorted onto different chromosomes (Fig. 6e), and should these sets acquire substantial length differences, intrachromosomal gene conversion can be forevermore blocked (Fig. 6f), as may be the case with EF-1α in *M. americanus*. The paralog phylogenies we recovered are consistent with the latter scenario, with two of the most distinct groupings of paralogs being the two length-distinguished paralog sets in Florida, which appear to be sorting independently on different chromosomes. Moreover, their histories are deep, suggesting they have not engaged in gene conversion for a considerable time.

5. Conclusion

The nuclear gene elongation factor-1 α (EF-1α) has many paralogs in *Metasiro americanus* (Davis, 1933), a harvestman (Opiliones) that lives in three disjunct populations in the Southeastern US. In one population, at least, the paralogs exist in different sets which have diverged before the populations separated, a phylogenetic analysis of the paralogs recovering them as out-paralogs. This positions EF-1α to give different phylogenetic hypotheses of the populations depending on which copies are sampled, and it prevents the use of this marker in statistical phylogeographic analyses in *M. americanus*. Assessing EF-1α paralogy in other species of interest should be done before using this marker, the true nature of which may vary widely among taxa. This can be done using basic cloning techniques and comparing the results to expectations from *Taq* polymerase error, which can be measured empirically by amplifying homogenous template.

Future studies should take advantage of such established tools as fluorescent *in situ* hybridization and northern blotting in order to enumerate the duplications of EF-1α in Opiliones genomes. Recent advances in the understanding of chromosomal number evolution in Opiliones, as well as gene expression techniques, are anticipated to aid determination of locations of duplicated gene copies and identification of functional EF-1α variants (Sharma et al., 2012; Šťáhlavský et al., 2012).

Acknowledgments

Molecular work was conducted at the American Museum of Natural History (AMNH) and was funded by a grant from the Defense Advanced Research Projects Agency (DARPA, W911NF-05-1-0271). Collecting *Metasiro americanus* in Florida and South Carolina was funded by a grant from the French National Research Council to P. Grandcolas and G. Giribet (BIONEOCAL—Endemism in New Caledonia: a phylogenetic and population-level study of its origins). Collecting permits for Florida were facilitated by Tova Spector and Arthur Stiles, and for Savannah River National Wildlife Refuge by Chuck Hayes, to whom we are grateful. Dennis Chastain helped us find *M. americanus* populations on Sassafras Mountain, where we also received the assistance of Greg Lucas of the South Carolina Department of Natural Resources. Chris Hawthorne and Steve Cutshaw provided important information leading to productive collecting sites at Florida Caverns State Park. Additional support was provided by Isabella Kappner, Mohammed Faiz, and Mercer Brugler. Miquel Arnedo and two anonymous reviewers provided useful comments that improved this paper.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2013.04.025.

References

Aguileta, G., Marthey, S., Chiapello, H., Lebrun, M.H., Rodolphe, F., Fournier, E., Gendrault-Jacquemard, A., Giraud, T., 2008. Assessing the performance of single-copy genes for recovering robust phylogenies. Syst. Biol. 57, 613–627.

Arnheim, N., Treco, D., Taylor, B., Eicher, E.M., 1982. Distribution of ribosomal gene length variants among mouse chromosomes. Proc. Natl. Acad. Sci. USA 79, 4677–4680.

Avise, J.C., Arnold, J., Ball, R.M., Bermingham, E., Lamb, T., Neigel, J.E., Reeb, C.A., Saunders, N.C., 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annu. Rev. Ecol. Syst. 18, 489–522.

Bailey, C.D., Carr, T.G., Harris, S.A., Hughesa, C.E., 2003. Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes. Mol. Phylogenet. Evol. 29, 435–455.

Bollag, R.J., Liskay, R.M., 1988. Conservative intrachromosomal recombination between inverted repeats in mouse cells: association between reciprocal exchange and gene conversion. Genetics 119, 161–169.

Boyer, S.L., Giribet, G., 2007. A new model Gondwanan taxon: systematics and biogeography of the harvestman family Pettalidae (Arachnida, Opiliones, Cyphophthalmi), with a taxonomic revision of genera from Australia and New Zealand. Cladistics 23, 337–361.

Carranza, S., Giribet, G., Ribera, C., Baguñà, J., Riutort, M., 1996. Evidence that two types of 18S rDNA coexist in the genome of *Dugesia* (*Schmidtea*) *mediterranea* (Platyhelminthes, Turbellaria, Tricladida). Mol. Biol. Evol. 13, 824–832.

Carraro, N., Tisdale-Orr, T.E., Clouse, R.M., Knöller, A.S., Spicer, R., 2012. Diversification and expression of the PIN, AUX/LAX, and ABCB families of putative auxin transporters in *Populus*. Fron. Plant Sci. 3, 17.

Chenuil, A., Hoareau, T.B., Egea, E., Penant, G., Rocher, C., Aurelle, D., Mokhtar-Jamai, K., Bishop, J.D.D., Boissin, E., Diaz, A., Krakau, M., Luttikhuizen, P.C., Patti, F.P., Blavet, N., Mousset, S., 2010. An efficient method to find potentially universal population genetic markers, applied to metazoans. BMC Evol. Biol. 10, 276.

Clarke, L.A., Rebelo, C.S., Gonçalves, J., Boavida, M.G., Jordan, P., 2001. PCR amplification introduces errors into mononucleotide and dinucleotide repeat sequences. Mol. Pathol. 54, 351–353.

Clouse, R.M., Giribet, G., 2010. When Thailand was an island—the phylogeny and biogeography of mite harvestmen (Opiliones, Cyphophthalmi, Stylocellidae) in Southeast Asia. J. Biogeogr. 37, 1114–1130.

Colwell, R.K., 2004–2011. Biota 2: The Biodiversity Database Manager. Sunderland: Sinauer Associates.

Copenhaver, G.P., Pikaard, C.S., 1996. Two-dimensional RFLP analyses reveal megabase-sized clusters of rRNA gene variants in *Arabidopsis thaliana*, suggesting local spreading of variants as the mode for gene homogenization during concerted evolution. Plant J. 9, 273–282.

Danforth, B.N., Brady, S.G., Sipes, S.D., Pearson, A., 2004. Single-copy nuclear genes recover Cretaceous-age divergences in bees. Syst. Biol. 53, 309–326.

Danforth, B.N., Ji, S., 1998. Elongation factor-1α occurs as two copies in bees: implications for phylogenetic analysis of EF-1α sequences in insects. Mol. Biol. Evol. 15, 225–235.

de Bivort, B.L., Clouse, R.M., Giribet, G., 2010. A morphometrics-based phylogeny of the temperate Gondwanan mite harvestmen (Opiliones, Cyphophthalmi, Pettalidae). J. Zool. Syst. Evol. Res. 48, 294–309.

Derkarabetian, S., Steinmann, D.B., Hedin, M., 2010. Repeated and time-correlated morphological convergence in cave-dwelling harvestmen (Opiliones, Laniatores) from montane western North America. PLoS One 5, e10388.

Dolinski, K., Botstein, D., 2007. Orthology and functional conservation in eukaryotes. Annu. Rev. Genet. 41, 465–507.
Duarte, J.M., Cui, L., Wall, P.K., Zhang, Q., Zhang, X., Leebens-Mack, J., Ma, H., Altman, N., dePamphilis, C.W., 2006. Expression pattern shifts following duplication Indicative of subfunctionalization and neofunctionalization in regulatory genes of *Arabidopsis*. Mol. Biol. Evol. 23, 469–478.
Dunlop, J.A., Anderson, L.I., Kerp, H., Hass, H., 2004. A harvestman (Arachnida: Opiliones) from the Early Devonian Rhynie cherts, Aberdeenshire, Scotland. Trans. Royal Soc. Edinb. Earth Sci. 94, 341–354.
Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucl. Acids Res. 32, 1792–1797.
Escriva, H., Bertrand, S., Germain, P., Robinson-Rechav, M., Umbhauer, M., Cartry, J., Duffraisse, M., Holland, L., Gronemeyer, H., Laudet, V., 2006. Neofunctionalization in vertebrates: the example of retinoic acid receptors. PLoS Genet. 2, e102.
Esseghir, S., Ready, P.D., 2000. Speciation of *Phlebotomus* sandflies of the subgenus *Larroussius* coincided with the late Miocene–Pliocene aridification of the Mediterranean subregion. Biol. J. Linn. Soc. 70, 189–219.
Faure, B., Bierne, N., Tanguy, A., Bonhomme, F., Jollivet, D., 2007. Evidence for a slightly deleterious effect of intron polymorphisms at the EF1α gene in the deep-sea hydrothermal vent bivalve *Bathymodiolus*. Gene 406, 99–107.
Feuda, R., Hamilton, S.C., McInerney, J.O., Pisani, D., 2012. Metazoan opsin evolution reveals a simple route to animal vision. Proc. Natl Acad. Sci. USA 109, 18868–18872.
Friedlander, T.P., Regier, J.C., Mitter, C., 1992. Nuclear gene sequences for higher level phylogenetic analysis: 14 promising candidates. Syst. Biol. 41, 483–490.
Giribet, G., 2000. Catalogue of the Cyphophthalmi of the world (Arachnida, Opiliones). Rev. Iber. Aracnol. 2, 49–76.
Giribet, G., Rambla, M., Carranza, S., Baguñà, J., Riutort, M., Ribera, C., 1999. Phylogeny of the arachnid order Opiliones (Arthropoda) inferred from a combined approach of complete 18S and partial 28S ribosomal DNA sequences and morphology. Mol. Phylogenet. Evol. 11, 296–307.
Giribet, G., Sharma, P.P., Benavides, L., Boyer, S., Clouse, R., de Bivort, B.L., Dimitrov, D., Kawauchi, G.Y., Murienne, J., Schwendinger, P., 2012a. Evolutionary and biogeographic history of the harvestman suborder Cyphophthalmi (Arachnida, Opiliones)—an ancient and global group of arachnids. Biol. J. Linn. Soc. 105, 92–130.
Giribet, G., Tourinho, A.L., Shih, C., Ren, D., 2012b. An exquisitely preserved harvestman (Arthropoda, Arachnida, Opiliones) from the Middle Jurassic of China. Org. Divers. Evol. 12, 51–56.
Giribet, G., Vogt, L., González, A.P., Sharma, P.P., Kury, A.B., 2010. A multilocus approach to harvestman (Arachnida: Opiliones) phylogeny with emphasis on biogeography and the systematics of Laniatores. Cladistics 26, 408–437.
Gouy, M., Guindon, S., Gascuel, O., 2010. SeaView Version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol. 27, 221–224.
Hall, T., 2007. BioEdit. Ibis Biosciences, Carlsbad, CA.
Hare, M.P., 2001. Prospects for nuclear gene phylogeography. Trends Ecol. Evol. 16, 700–706.
Hedin, M., Derkarabetian, S., McCormack, M., Richart, C., Shultz, J.W., 2010. The phylogenetic utility of the nuclear protein-coding gene EF-1α for resolving recent divergences in Opiliones, emphasizing intron evolution. J. Arachnol. 38, 9–20.
Hedin, M., Thomas, S.M., 2010. Molecular systematics of eastern North American Phalangodidae (Arachnida: Opiliones: Laniatores), demonstrating convergent morphological evolution in caves. Mol. Phylogenet. Evol. 54, 107–121.
Hedin, M.C., Maddison, W.P., 2001. Phylogenetic utility and evidence for multiple copies of elongation factor-1α in the spider genus *Habronattus* (Araneae: Salticidae). Mol. Biol. Evol. 18, 1512–1521.
Higgins, C.F., 1992. ABC transporters: from microorganisms to man. Annu. Rev. Cell Biol. 8, 67–113.
Hillis, D.M., Moritz, C., Porter, C.A., Baker, R.J., 1991. Evidence for biased gene conversion in concerted evolution of ribosomal DNA. Science 251, 308–310.
Hines, H.M., Cameron, S.A., Williams, P.H., 2006. Molecular phylogeny of the bumble bee subgenus Pyrobombus (Hymenoptera: Apidae: Bombus) with insights into gene utility for lower-level analysis. Invertebr. Syst. 20, 289–303.
Holland, I.B., Blight, M.A., 1999. ABC-ATPases, adaptable energy generators fuelling transmembrane movement of a variety of molecules in organisms from bacteria to humans. J. Mol. Biol. 293, 381–399.
Hovemann, B.T., Richter, S., Walldorf, U., Cziepluch, C., 1988. Two genes encode related cytoplasmic elongation factors 1α (EF-1α) in *Drosophila melanogaster* with continuous and stage specific expression. Nucl. Acids Res. 16, 3175–3194.
Hurles, M., 2004. Gene duplication: the genomic trade in spare parts. PLoS Biol. 2, 900–904.
Jeyaprakash, A., Hoy, M.A., 2009. The nuclear genome of the phytoseiid *Metaseiulus occidentalis* (Acari: Phytoseiidae) is among the smallest known in arthropods. Exp. Appl. Acarol. 47, 263–273.
Juberthie, C., 1988. Les Opilions Cyphophthalmes: biogéographie, vitesse d'évolution, périodes de colonisation du milieu souterrain. TUB-Dokumentation Kongresse und Tagungen, Berlin 38, 303–308.
Katoh, K., Misawa, K., Kuma, K., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucl. Acids Res. 30, 3059–3066.
Kjer, K., Carle, F.L., Litman, J., Ware, J., 2006. Molecular phylogeny of Hexapoda. Arthropod Syst. Phylogeny 64, 34–44.
Klein, H.L., Petes, T.D., 1981. Intrachromosomal gene conversion in yeast. Nature 289, 144–148.
Knowles, L.L., 2004. The burgeoning field of statistical phylogeography. J. Evol. Biol. 17, 1–10.
Koonin, E.V., 2005. Orthologs, paralogs, and evolutionary genetics. Annu. Rev. Genet. 39, 309–338.
Kunkel, T.A., 1984. Mutational specificity of depurination. Proc. Natl. Acad. Sci. USA 81, 1494–1498.
Hood, L., Campbell, J.H., Elgin, S.C.R., 1975. The organization, expression, and evolution of antibody genes and other multigene families. Annu. Rev. Genet. 9, 305–353.
Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948.
Lenstra, J.A., Van Vliet, A., Arnberg, A.C., Van Hemert, F.J., Möller, W., 1986. Genes coding for the elongation factor EF-1α in *Artemia*. Eur. J. Biochem. 155, 475–483.
Liao, D., 1999. Concerted evolution: molecular mechanism and biological implications. Am. J. Hum. Genet. 64, 24–30.
Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451–1452.
Michez, D., Patiny, S., Danforth, B., 2009. Phylogeny of the bee family Melittidae (Hymenoptera: Anthophila) based on combined molecular and morphological data. Syst. Entomol. 34, 574–597.
Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees. Gateway Computing Environments Workshop (GCE), New Orleans, Louisiana.
Monteiro, A., Pierce, N.E., 2001. Phylogeny of *Bicyclus* (Lepidoptera: Nymphalidae) inferred from COI, COII, and EF-1α gene sequences. Mol. Phylogenet. Evol. 18, 264–281.
Murienne, J., Giribet, G., 2009. The Iberian Peninsula: ancient history of a hot spot of mite harvestmen (Arachnida: Opiliones: Cyphophthalmi: Sironidae) diversity. Zool. J. Linn. Soc. 156, 785–800.
Nagylaki, T., 1984. The evolution of multigene families under intrachromosomal gene conversion. Genetics 106, 529–548.
Nagylaki, T., Petes, T.D., 1982. Intrachromosomal gene conversion and the maintenance of sequence homogeneity among repeated genes. Genetics 100, 315–337.
Oakley, T.H., Plachetzki, D.C., Rivera, A.S., 2007. Furcation, field-splitting, and the evolutionary origins of novelty in arthropod photoreceptors. Arthropod Struct. Dev. 36, 386–400.
Palumbi, S.R., Baker, C.S., 1994. Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. Mol. Biol. Evol. 11, 426–435.
Parkin, E.J., Butlin, R.K., 2004. Within- and between-individual sequence variation among ITS1 copies in the meadow grasshopper *Chorthippus parallelus* indicates frequent intrachromosomal gene conversion. Mol. Biol. Evol. 21, 1595–1601.
Perlman, S.J., Kelly, S.E., Hunter, M.S., 2008. Population biology of cytoplasmic Incompatibility: maintenance and spread of *Cardinium* symbionts in a parasitic wasp. Genetics 178, 1003–1011.
Polanco, C., Gonzalez, A.I., de la Fuente, A., Dover, G.A., 1998. Multigene family of ribosomal DNA in *Drosophila melanogaster* reveals contrasting patterns of homogenization for IGS and ITS spacer regions: a possible mechanism to resolve this paradox. Genetics 149, 243–256.
Prince, V.E., Pickett, F.B., 2002. Splitting pairs: the diverging fates of duplicated genes. Nat. Rev. Genet 3, 827–837.
Regier, J.C., Shultz, J.W., 1997. Molecular phylogeny of the major arthropod groups indicates polyphyly of crustaceans and a new hypothesis for the origin of hexapods. Mol. Biol. Evol. 14, 902–913.
Regier, J.C., Shultz, J.W., Ganley, A.R., Hussey, A., Shi, D., Ball, B., Zwick, A., Stajich, J.E., Cummings, M.P., Martin, J.W., Cunningham, C.W., 2008. Resolving arthropod phylogeny: exploring phylogenetic signal within 41 kb of protein-coding nuclear gene sequence. Syst. Biol. 57, 920–938.
Regier, J.C., Shultz, J.W., Zwick, A., Hussey, A., Ball, B., Wetzer, R., Martin, J.W., Cunningham, C.W., 2010. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. Nature 463, 1079–1083.
Riesgo, A., Andrade, S.C.S., Sharma, P.P., Novo, M., Pérez-Porro, A.R., Vahtera, V., González, V.L., Kawauchi, G.Y., Giribet, G., 2012. Comparative description of ten transcriptomes of newly sequenced invertebrates and efficiency estimation of genomic sampling in non-model taxa. Front. Zool. 9, 33.
Roth, C., Rastogi, S., Arvestad, L., Dittmar, K., Light, S., Ekman, D., Liberles, D.A., 2007. Evolution after gene duplication: models, mechanisms, sequences, systems, and organisms. J. Exp. Zool. Part B Mol. Dev. Evol. 306B, 58–73.
Scherer, S., Davis, R.W., 1980. Recombination of dispersed repeated DNA sequences in yeast. Science 209, 1380–1384.
Schlötterer, C., Tautz, D., 1994. Chromosomal homogeneity of *Drosophila* ribosomal DNA arrays suggests intrachromosomal exchanges drive concerted evolution. Curr. Biol. 4, 777–783.
Sharma, P., Giribet, G., 2009. Sandokanid phylogeny based on eight molecular markers—The evolution of a southeast Asian endemic family of Laniatores (Arachnida, Opiliones). Mol. Phylogenet. Evol. 52, 432–447.
Sharma, P.P., Giribet, G., 2011. The evolutionary and biogeographic history of the armoured harvestmen – Laniatores phylogeny based on ten molecular markers, with the description of two new families of Opiliones (Arachnida). Invertebr. Syst. 25, 106–142.
Sharma, P.P., Schwager, E.E., Extavour, C.G., Giribet, G., 2012. Hox gene expression in the harvestman *Phalangium opilio* reveals divergent patterning of the chelicerate opisthosoma. Evol. Dev. 14, 450–463.

Shultz, J.W., Regier, J.C., 2001. Phylogenetic analysis of Phalangida (Arachnida, Opiliones) using two nuclear protein-encoding genes supports monophyly of Palpatores. J. Arachnol. 29, 189–200.
Shultz, J.W., Regier, J.C., 2009. *Caddo agilis* and *C. pepperella* (Opiliones, Caddidae) diverged phylogenetically before acquiring their disjunct, sympatric distributions in Japan and North America. J. Arachnol. 37, 238–240.
Slatkin, M., 1986. Interchromosomal biased gene conversion, mutation and selection in a multigene family. Genetics 112, 681–698.
Smith, G.P., 1974. Unequal crossover and the evolution of multigene families. Cold Spring Harbor Symp. Quant. Biol. 38, 507–513.
Smith, G.P., 1976. Evolution of repeated DNA sequences by unequal crossover. Science 191, 528–535.
Song, H., Buhay, J.E., Whiting, M.F., Crandall, K.A., 2008. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. Proc. Natl. Acad. Sci. USA 105, 13486–13491.
Sonnhammer, E.L., Koonin, E.V., 2002. Orthology, paralogy and proposed classification for paralog subtypes. Trends Genet. 18, 619–620.
Spady, T.C., Parry, J.W.L., Robinson, P.R., Hunt, D.M., Bowmaker, J.K., Carleton, K.L., 2006. Evolution of the cichlid visual palette through ontogenetic subfunctionalization of the opsin gene arrays. Mol. Biol. Evol. 23, 1538–1547.
Šťáhlavský, F., Boyer, S.L., Harvey, M.S., Giribet, G., 2012. First cytogenetic study of a member of the harvestman family Pettalidae (Opiliones: Cyphophthalmi). Aust. J. Entomol. 51, 299–302.
Stamatakis, A., Hoover, P., Rougemont, J., 2008. A rapid bootstrap algorithm for the RAxML web servers. Syst. Biol. 57, 758–771.
Studer, R.A., Robinson-Rechavi, M., 2009. How confident can we be that orthologs are similar, but paralogs differ? Trends Genet. 25, 210–216.
Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24, 1596–1599.
Taylor, J.S., Raes, J., 2004. Duplication and divergence: the evolution of new genes and old ideas. Annu. Rev. Genet. 38, 615–643.
Telford, M.J., Holland, P.W.H., 1997. Evolution of 28S ribosomal DNA in chaetognaths: duplicate genes and molecular phylogeny. J. Mol. Evol. 44, 135–144.
Thornhill, D.J., Lajeunesse, T.C., Santos, S.R., 2007. Measuring rDNA diversity in eukaryotic microbial systems: how intragenomic variation, pseudogenes, and PCR artifacts confound biodiversity estimates. Mol. Ecol. 16, 5326–5340.
Tindall, K.R., Kunkel, T.A., 1988. Fidelity of DNA Synthesis by the *Thermus aquaticus* DNA Polymerase. Biochemistry 27, 6008–6013.
Varón, A., Sy Vinh, L., Wheeler, W.C., 2009. POY version 4: phylogenetic analysis using dynamic homologies. Cladistics 26, 72–85.
Walldorf, U., Hovemann, B., Bautz, E.K.F., 1985. F1 and F2: two similar genes regulated differently during development of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA 82, 5795–5799.
Whittall, J.B., Medina-Marino, A., Zimmer, E.A., Hodges, S.A., 2006. Generating single-copy nuclear gene data for a recent adaptive radiation. Mol. Phylognet. Evol. 39, 124–134.
Wilhelm, J., Pingoud, A., Hahn, M., 2003. Real-time PCR-based method for the estimation of genome sizes. Nucl. Acids Res. 31, e56.
Zimmer, E.A., Martin, S.L., Beverley, S.M., Kan, Y.W., Wilson, A.C., 1980. Rapid duplication and loss of genes coding for the α chains of hemoglobin. Proc. Natl. Acad. Sci. USA 77, 2158–2162.