PRIMER NOTE

Evolutionary conservation and versatility of a new set of primers for amplifying the ribosomal internal transcribed spacer regions in insects and other invertebrates

YA-JIE JI, DE-XING ZHANG and LI-JUN HE

State Key Laboratory of Integrated Management of Pest Insects & Rodents, Institute of Zoology, Chinese Academy of Sciences, 25 Beisihuan-Xi Road, Beijing 100080, P. R. China

Abstract

The evolutionary conservation and versatility of a new set of nuclear primers for the amplification of the ribosomal internal transcribed spacer regions in insects and other invertebrates have been studied. These primers, conserved across Insecta and other invertebrates, are based on a comprehensive taxonomic survey of the current DNA sequence databases. Their versatility was demonstrated by extensive polymerase chain reaction assays in 16 species from two arachnid orders, eight insect orders, three invertebrate and vertebrate chordate orders and by direct sequencing of the amplified products. The availability of these primers to the insect research community should facilitate the use of internal transcribed spacer regions in intraspecific studies as well as phylogenetic analysis of closely related taxa.

Keywords: Arachnida, *Buthus* scorpion, Insecta, ITS1 and ITS2 regions, nuclear markers, ribosomal DNA

Received 6 February 2003; revision received 10 April 2003; accepted 30 July 2003

Ribosomal DNA (rDNA) has been the most widely employed nuclear sequence in evolutionary analyses, primarily because concerted evolution occurs among its repeated copies and conserved primers are available, which facilitate the isolation of the target sequences by polymerase chain reaction (PCR) (Hillis & Davis 1988; White *et al.* 1990; Hillis & Dixon 1991). Due to their rapid rates of evolution, the internal transcribed spacer (ITS) regions between the 18S, 5.8S and 28S genes (or their homologues), ITS1 and ITS2, have been popular in phylogenetic inference for closely related taxa (e.g. Miller *et al.* 1996), cryptic species identification (for review see Collins & Paskewitz 1996) and phylogeographical and other population genetic studies (e.g. Mukabayire *et al.* 1999; Wörheide *et al.* 2002).

Studies in insects with ITS sequences often start with conserved primers based on fungal rDNA genes (White *et al.* 1990) or the general primers of Hillis & Dixon (1991). The applicability of these primers to insects has not been systematically examined and variable results have often

been obtained in different species tested. These primers were designed over a decade ago based on the limited number of sequences available at that time and, therefore, they unavoidably have some limitations. We have carried out a more comprehensive taxonomic survey of the current DNA sequence databases and designed a new set of conserved primers for the amplification of these regions in insects in particular and invertebrates in general. Here we report the outcome of such studies.

The EMBL database was extensively searched for nuclear ribosomal 18S/5.8S/28S (or their homologues) gene sequences (complete or partial). Taxonomically representative sequences were retrieved and multiple alignments of homologue sequences carried out using the CLUSTAL W (Thompson *et al.* 1994)-based EMBOSS 'EMMA' program implemented by the HGMP Bioinformatics Service (Cambridge, UK). In designing these primers, we not only located them to a DNA region highly conserved among invertebrate organisms but paid much attention to their overall structure, particularly their internal stability. The pattern of internal stability of a primer sequence greatly affects the effectiveness of specific priming: primers with relatively less stable 3'-ends tend to avoid false priming

Correspondence: De-Xing Zhang. Fax: (+ 86) 10 6261 2962; E-mail: dxzhang@panda.ioz.ac.cn

```
A). Information on the five ribosomal ITS primers

Name            Position        Sequence (5' to 3')             Length (bases)
CAS18sF1        1843            TACACACCGCCCGTCGCTACTA          22
CAS5p8sB1d      2805            ATGTGCGTTCRAAATGTCGATGTTCA      26
CAS5p8sFt       2830            TGAACATCGACATTTYGAACGCATAT      26
CAS5p8sFc       2830            TGAACATCGACATTTYGAACGCACAT      26
CAS28sB1d       3328            TTCTTTTCCTCCSCTTAYTRATATGCTTAA  30
```

```
B). Primer CAS18sF1

Access/ID       Sequence (5' to 3')             Taxonomic status
                TACACACCGCCCGTCGCTACTA          CAS18sF1
                ||||||||||||||||||||||CCGATTG   18d (Hillis & Dixon 1991)
DMRG            ||||||||||||||||||||||          Insecta, Diptera
AF012476        ||||||||||||||||||||||          Insecta, Coleoptera
X89482          ||||||||||||||||||||||          Insecta, Neuroptera
X89487          ||||||||||||||||||||||          Insecta, Mecoptera
X89492          ||||||||||||||||||||||          Insecta, Hymenoptera
X89494          ||||||||||||||||||||||          Insecta, Raphidoptera
X89497          ||||||||||||||||||||||          Insecta, Megaloptera
U09207          ||||||||||||||||||||||          Insecta, Hemiptera
U68400          ||||||||||||||||||||||          Insecta, Plecoptera
X89488          ||||||||||||||||||||||          Insecta, Phasmida
U29493          ||||||||||||||||||||||          Myriapoda (centipedes)
Z74479          ||||||||||||||||||||||          Arachnida, Acari
AF005438        ||||||||||||||||||||||          Pycnogonida (sea spiders)
M29931          ||||||||||||||||||||||          Pentastomida (tongue worms)
U49912          ||||||||||||||||||||||          Tardigrada (water bears)
M91179          ||||||||||||||||||||||          Chondrichthyes (fish)
M32847          ||||||||||||||||||||||          Amphibia
M10098          ||||||||||||||||||||||          Human
```

```
C). Primer CAS5p8sFt

Access/ID       Sequence (5' to 3')             Taxonomic status
                TGAACATCGACATTTYGAACGCATAT      CAS5p8sFt
                TGAACATCGACATTTYGAACGCACAT      CAS5p8sFc
                T|||T|||||||||C||C|||||||C      5.8d (Hillis & Dixon 1991)
DMRG            A|||||||||||||||T|||||||||      Insecta, Diptera
L22060          A|||||C||||||CG|T|||||||||      Insecta, Diptera
BMRRN58S        T|||||||||||||C|||||||C||      Insecta, Lepidoptera
NG02955         A|||||||||||||C|||||||C||      Insecta, Hymenoptera
AF007264        GC|||||||||AG|C|||||||C||      Insecta, Hemiptera
U93226          T||||||||||||||T|||||||||      Mollusca (molluscs)
LVRRNA01        T|||T|||||||||C|||||||C||      Echinoidea (sea urchin)
EH58101         T|||||||||||C|T|||||||||      Trematoda (parasitic flukes)
U89871          T|||||||||||C||C|||||||C|      Trematoda (parasitic flukes)
AF040936        T|||||||||||||C|||||||||C      Trematoda (parasitic flukes)
AF047853        C|||T|||T||||C||||||||A||      Turbellaria(flatworms)
ALU82735        CC||T|||||||TCC||C|||||A||     Cnidaria
CEF17B5         A|TTC|T|||||||C||G|||||||CT|   Nematoda
                ↑                         ↑
```

```
D). Primer CAS28sB1d

Access/ID       Sequence (5' to 3')             Taxonomic status
                TTCTTTTCCTCCSCTTAYTRATATGCTTAA  CAS28sB1d
DMRG            |||||||||||C||A|T|A|||||||||||  Insecta, Diptera
AAINSP          ||||C|||||||C||||T|A|||||||||   Insecta, Diptera
LCU417709       |||||||||||C|C|T|A|||||||||||   Insecta, Diptera
AB052895        ||||||||||||G|||G|A|||||||||    Insecta, Hymenoptera
S50426          ||G|||||||||G|||T|A|||||||||    Insecta, Hemiptera
EAM303091       |||||||||C|||G|A|||||||||||     Insecta, Thysanoptera
AF321244        |||||||||||G|C||T|A|||||||||    Insecta, Blattaria
AF212167        |||||||||||G||||T|T|||||||||    Merostomata, Xiphosura
AF084942        ||||||||||G|||T|T|||||||||||    Arachnida, Araneae
AF548530        ||||||||||G|||G|G|||||||||||    Arachnida, Acari
CAL297569       |||||||||||G|||A|A|||||||||     Arachnida, Acari
X54004          |||||||||||G|||C|G|||||||||     Ciliophora (Ciliates)
Z49904          |||||||||||G|||C|G|||||||||     Ciliophora (Ciliates)
M98363          |||||||||||G|||G|A|||||||||     Ciliophora (ciliates)
X16108          |||||A|||||G|||C|T|||||||||     Dinophyceae (dinoflagellates)
AF042814        |||||A|||||G|||C|T|||||||||     Dinophyceae (dinoflagellates)
U63484          |C|||||||||G|||C|T|||||||||     Dinophyceae (corals)
X57261          ||||||||||||G|||T|G|||||||||    Porifera (sponges)
```

more effectively and thus perform better in sequencing and PCR (Rychlik 1992). Primers were designed using ᴏʟɪɢᴏ™ 4 primer analysis software (National Biosciences Inc.) and synthesized by Sangon Biotech Ltd-China. Information on the five primers studied here is given in Fig. 1(A).

The PCR were carried out on GeneAmp 9700 systems (Perkin-Elmer) using *Taq* DNA polymerase (Promega), with the following general conditions (annealing temperature varies and depends on the organism involved, see Table 1): Some 5–10 ng of genomic DNA was used in a 10-µL reaction with 0.5–1.0 U of *Taq* DNA polymerase, 200 µᴍ of each dNTP and 0.1–0.2 µᴍ of each primer. Following an initial denaturation at 94 °C for 4 min, 35 cycles were performed, each consisting of melting at 95 °C for 20 s, annealing at the appropriate temperatures (Table 1) for 40 s and extension at 72 °C for 20 s. A final extension was run at 72 °C for 2 min. A volume (5 µL) of the reaction was then examined on 0.8% agarose gel to check the quality and quantity of the PCR product. Negative PCR control was carried out along with normal experiments to check against cross contamination. The 'Touchdown' programming technique (Don *et al*. 1991) was employed for some PCR (Table 1). After purification, PCR products were directly sequenced with ABI BigDye™ Terminators Cycle Sequencing Kit (version 2.0) in an ABI PRISM3100 automated sequencer.

Amplicon CAS18sF1/CAS5p8sB1d corresponds to the ITS1 region between the 18S and 5.8S genes. Both amplicons CAS5p8sFt/CAS28sB1d and CAS5p8sFc/CAS28sB1d cover the ITS2 region between the 5.8S and 28S genes. The primer combination CAS18sF1/CAS28sB1d allows the amplification of the entire region between the 18S and 28S genes (comprising ITS1, 5.8S gene and ITS2). Figure 1(B–D) shows the status of sequence conservation of these primers in a wide range of organisms (mostly invertebrates). It can be seen from Fig. 1(B) that CAS18sF1 is completely conserved in all animals compared, from fruit fly to man. This primer overlaps partially with the primer 18d of Hillis & Dixon (1991) but is 5 bases shorter and with a better internal stability profile.

Primer CAS5p8sB1d is the complement of CAS5p8sFt and CAS5p8sFc. These primers are located at a position similar to the 5.8c and 5.8d primers of Hillis & Dixon (1991) but have some key modifications that ensure their versatility. First, the most 3' base of the backward primer of Hillis & Dixon (1991) (5.8c, which is the complement of 5.8d)

is at a highly variable position where all four bases were observed, as shown in Fig. 1(C) (indicated by the first arrow in the figure). In Insecta, 'T' was observed in Diptera and Hymenoptera, 'A' in Lepidoptera and 'C' in Hemiptera. Therefore, 5.8c will not work in many insect species. Second, the forward primer (5.8d) of Hillis & Dixon (1991) ends with the base 'C'. However, in dipteran insects and several other invertebrates, 'T' is observed at this position (Fig. 1C, indicated by the second arrow). Hence, 5.8d also has a limited applicability. These limitations were overcome in our primers by modifying their location and length. In addition, a degenerate base was introduced to increase their versatility. The backward primer CAS5p8sB1d should work in many invertebrate species. [Note that the most 3' base should be changed to 'G' for hemipteran insects according to Fig. 1(C) but we suspect that there may be some sequencing error here with the AF007264 sequence. In fact, our empirical PCR study showed that CAS5p8sB1d works well in an aphid (Table 1).] Two slightly different forward primers were designed, with CAS5p8sFt having the nucleotide 'T' at the third base next to the 3' end and CAS5p8sFc having the nucleotide 'C' instead. CAS5p8sFt should work better in dipteran insects and many noninsect invertebrates while CAS5p8sFc is better for most non-dipteran insects and echinoderms (Fig. 1C). Due to some exceptional variation observed at the 3' end positions in some flatworms and flukes, these primers may not work well in these organisms (Fig. 1C). Empirically, we found that both forms of the forward primer can work in many insects tested, albeit with different efficiencies (see below, Table 1).

The primer CAS28sB1d does not have any significant overlap with any published primer that we know. It is highly conserved in invertebrates (Fig. 1D). Three degenerate bases were introduced in the central part of the primer where pronounced nucleotide variation is observed. The complement of this primer, as well as that of CAS18sF1, can be used as primers for PCR amplification if needed.

Overall, the above analysis suggests that the five ITS primers that we designed are highly conserved in sequences among insects and other invertebrate organisms allowing, in theory, the isolation of ribosomal ITS regions in many invertebrate organisms by simple PCR amplification.

The amplifiability of these primers was tested empirically by PCR and direct sequencing of PCR products in 16 invertebrate and vertebrate species, including two

Fig. 1 Nuclear ribosomal internal transcribed spacer (ITS) primers studied here and their evolutionary conservation. (A) Position refers to the nucleotide position of the 3' base of primers with respect to the *Drosophila melanogaster* sequence (sequence identification in EMBL database is DMRG). F, forward primer; B, backward primer; d, degenerate primer; t, third base from the 3' end being T; c, third base from the 3' end being C. The following ambiguous symbols in sequences are used: R = A/G, S = G/C and Y = C/T. (B–D) Access/ID, sequence identification or the Accession no. of the corresponding sequence in the EMBL sequence database. In Taxonomic status, a class-order affiliation is given for insects and arachnid species but, for other organisms, only a generally well-known taxonomic category is listed with their common names given in brackets. Vertical bars in the sequences signify nucleotides which are identical to those in the primer sequence.

Table 1 Universality of the internal transcribed spacer primers in Insecta

Organism	Taxonomy	Common name	CAS18sF1/ CAS5p8sB1d		CAS5p8sFt/ CAS28sB1d		CAS5p8sFc/ CAS28sB1d	
			T_a (°C)	Yield	T_a (°C)	Yield	T_a (°C)	Yield
Thermobia domestica	Thysanura	Firebrat	67	S	55	w	62	S
Calopteryx splendens	Odonata	Damselfly	67	S	55	w	55	S
Locusta migratoria	Orthoptera	Migratory locust	55	S	55	w	55	S
Calliptamus abbreviatus	Orthoptera	Grasshopper	67	S	55	w	55	S
Periplaneta americana	Blattaria	American cockroach	67	S	55	w	55	S
Acyrthosiphon pisum	Hemiptera	Pea aphid	60	S	55	w	55	S
Tetramorium sp.	Hymenoptera	Ant	67	S	62	S	62	w
Helicoverpa armigera	Lepidoptera	Cotton bollworm	65	S	62	S	62	S
Dendrolimus punctatus	Lepidoptera	Pine moth	67	S	60	S	55	S
Drosophila melanogaster	Diptera	Fruit fly	67	S	55	S	55	S
Bombus lapidarius	Hymenoptera	Bumble bee	62	S	62	S	62	S
Tegenaria domesticus	Araneae	House spider	60	S	55	?	55	S
Buthus martensii	Scorpiones	Chinese scorpion	50	S	50	?	50	S
Acipenser sinensis	Acipenseriformes	Chinese sturgeon	67	S	55	w	60	S
Gadus morhua	Gadiformes	Atlantic cod	67→ 60*	S	55	?	55	S
Branchiostoma belcheri	Cephalochordata	Amphioxus	60	S	—	—	—	—

Numbers given in the table are annealing temperatures (T_a) used in successful polymerase chain reaction (PCR) amplification for the corresponding taxon/amplicon combinations. 'Successful' means a main band, which is specific and reproducible, has been obtained in the PCR conditions tested. Bands amplified have been verified to be ribosomal DNA by direct sequencing of the PCR products. Where no T_a is given, it means that the corresponding amplification was not satisfactory or no product was obtained after several tries at different conditions. S, bright and specific band; w, specific band but (much) weaker compared with bands denoted S; ?, very weak band; –, no specific amplification.
*T_a for 'Touchdown' PCR.

chelicerate species from two arachnid orders, 11 insect species from eight orders and three chordate organisms (see Table 1 for details). The insects studied cover the main divisions of Insecta, with one primitive wingless insect (firebrat), a primitive winged insect (damselfly) and nine other winged insects (two grasshoppers, a cockroach, an aphid, a fruit fly, an ant, a bee and two moths). Several invertebrate samples were used previously in the study of Zhang & Hewitt (1997). Table 1 summarizes the test results. Bands amplified have been verified to be ribosomal DNA by direct sequencing of the PCR products (GenBank Accession nos of the related sequences, AJ577252–AJ577262). The versatility of these ribosomal primers in insects is clearly demonstrated. Also, our experimental studies confirmed the theoretical expectations based on sequence conservation about the relative performance of the two 5.8S forward primers (discussed above): CAS5p8sFc performs better than CAS5p8sFt in most nondipteran insects (Table 1; density of bands amplified with CAS5p8sFc is normally much stronger, denoted 'S' in the Table). It is also worth noting that the primers reported here amplify quite well in three vertebrate and invertebrate chordates tested (fishes and amphioxus), indicating their probable applicability beyond invertebrate organisms (Table 1).

We have examined the intraspecific variability of the ITS1 region in the Chinese scorpion *Buthus martensii* using the primer combination CAS18sF1/CAS5p8sB1d. Sixty-one individuals from four geographical localities (across a 700-km area) were sequenced. Four haplotypes (named after their polymorphic sites as ΔGAC, ΔGGC, AAGT and AAGC, respectively) were identified, forming two distinct groups congruent with the local geography. The estimated average number of segregating sites per nucleotide site (θ) is 0.00623 [with the polymorphicity *Pm* (Zhang & Hewitt 2003) being 1.143%], which is a rather high value for nuclear DNA sequences (Zhang & Hewitt 2003). This shows that ITS regions amplified with the primers reported here can serve as valuable nuclear DNA markers for the phylogeographical study of scorpions in China.

Acknowledgements

References

Collins FH, Paskewitz SM (1996) A review of the use of ribosomal DNA (rDNA) to differentiate among cryptic *Anopheles* species. *Insect Molecular Biology*, **5**, 1–9.

Don RH, Cox PT, Wainwright BJ, Baker K, Mattich JS (1991) 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Research*, **19**, 4008.

Hillis DM, Davis SK (1988) Ribosomal DNA — intraspecific polymorphism, concerted evolution, and phylogeny reconstruction. *Systematic Zoology*, **37**, 63–66.

Hillis DM, Dixon MT (1991) Ribosomal DNA — molecular evolution and phylogenetic inference. *Quarterly Review of Biology*, **66**, 410–453.

Miller BR, Crabtree MB, Savage HM (1996) Phylogeny of fourteen *Culex* mosquito species, including the *Culex pipiens* complex, inferred from the internal transcribed spacers of ribosomal DNA. *Insect Molecular Biology*, **5**, 93–107.

Mukabayire O, Boccolini D, Lochouarn L, Fontenille D, Besansky NJ (1999) Mitochondrial and ribosomal internal transcribed spacer (ITS2) diversity of the African malaria vector *Anopheles funestus*. *Molecular Ecology*, **8**, 289–297.

Rychlik W (1992) *Reference manual.* OLIGO™ *version 4.0 for Macintosh.* National Biosciences, Plymouth, MN.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple alignment through sequence weighting, position-specific gap penalities and weight matrix choice. *Nucleic Acids Research*, **22**, 4673–4680.

White TJ, Bruns T, Lee S, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: *PCR Protocols: a Guide to Methods and Applications* (eds Innis MA, Gelfand DH, Sninsky JJ, White TJ), pp. 315–322. Academic Press, San Diego.

Wörheide G, Hooper JNA, Degnan BM (2002) Phylogeography of western Pacific *Leucetta 'chagosensis'* (Porifera: Calcarea) from ribosomal DNA sequences: implications for population history and conservation of the Great Barrier Reef World Heritage Area (Australia). *Molecular Ecology*, **11**, 1753–1768.

Zhang DX, Hewitt GM (1997) Assessment of the universality and utility of a set of conserved mitochondrial COI primers in insects. *Insect Molecular Biology*, **6**, 143–150.

Zhang DX, Hewitt GM (2003) Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular Ecology*, **12**, 563–584.