Modification of Rapid Amplication of cDNA Ends for the Extension of cDNA libraries.

Barkwan SS[2], Fokunang CN [1, 2], Tomkins PT[1], Tembe-Fokunang EA[1]

[1]Faculty of Medicine and Biomedical Sciences University of Yaoundé1, Cameroon

[2]Centre for Biopolymer and Biomolecular Research, Athlone Institute of Technology, Republic of Ireland.

Correspondence to:Dr Charles Fokunang

Faculty of Medicine & Biomedical Sciences, University of Yaoundé 1.

Tel: +237 70902446; email: charlesfokunang@yahoo.co.uk

ABSTRACT

Rapid Amplification of cDNA Ends (RACE) is a technique used in molecular biology to obtain full length sequence of an RNA transcript found within cell. RACE results in the production of a copy of the RNA sequence of interest, which is converted through reverse transcription to cDNA, followed by PCR amplification of the cDNA copies.

A cDNA capture method was devised to construct full-length cDNA libraries. In this method, a library of cDNA fragments generated from substractive hybridiazation studies (full-length cDNA library RAD), allowed the respective full-length cDNA cognates to be amplified by PCR and then cloned.

Five clones from subtractive hydridization studies, identified as containing a high proportion of differentially expressed genes (coded for hypothetical proteins), were chosen as templates for the preparation of full length cDNA. Primers were designed using Genefisher software (Bioinformatics Bielefeld) .Typically, 30-35 cycles of PCR were employed. As a target for the capture reaction; double-stranded cDNA was amplified by PCR using the modification of SMART II cDNA library construction kit (Clontech) with the same RNA sample that had served as the tester sample for the respective RAD reaction

This modified RACE technique resulted in the production of a copy of RNA sequence of interest, produced through reversed transcriptase followed by a successful PCR amplification of cDNA copies. The cDNA libraries were successfully cloned and sequenced, leading to a successful extension of two unknown sequences by 200 base pairs.

Key words*:* cDNA ends, RNA transcripts, RACE, subtractive hybridization, cloning.

RESUME

L'amplification rapide de terminale cADN (RACE) est une technique utilise biologie moléculaire pour obtenir une longueur complet de l'ARN qui se trouve parmi les cellules. RACE aboutir dans la production d'une copie de séquence l'ARN d'intérêt.

Une méthode de capture cADN était établi pour construire la bibliothèque de grande longueur de cADN. Par cette méthode une bibliothèque de fragment de cADN dériver de l'étude de *subtractive hybridization* (*full-length cDNA library RAD*), a permit the chaque fragment cADN grand longueur pour être amplifier et puis cloner par PCR.

Cinq clone de l'étude de subtrative hybridization, identifier comme contenant une grand proportion de gène diferentiellement exprime (code pour des protéines hypothétiques) , étaient choisit comme modèle pour la préparation de cADN a grand échelle de longueur. Les amorces étaient construire en utilisant les logiciels (Bio-informatics Bielefield). Comme d'habitude, 30-35 cycles de PCR étaient utilisés. Comme point de repère pour la réaction de capture : l'ADN double fragment était amplifie en utilisant la modification de la bibliothèque cADN SMART II logiciel (Clontech) avec le même échantillon l'ARN qui a servi de échantillon d'essai pour la réaction de RAD.

Le modification de technique de RACE nous a permis a augmente deux séquence inconnu par 200 pair de bases. Le PCR colonie était utilisé pour identifier the les colonies positif en utilisant l'amorce M13 avant et derrière. La taille de bande au dessus de 200 pairs de bases.

Les Mots clés : terminal cADN, transcrits L'ARN, RACE, *subtractive hybridization,* clonage.

INTRODUCTION

Rapid Amplification of cDNA Ends (RACE) is a technique used in molecular biology to obtain full length sequence of an RNA transcript found within cell (Vingsbo-Lundberg et al., 1998; Tillet et al., 2000; Xianan and Baird, 2001). RACE results in the production of a copy of the RNA sequence of interest, produced through reverse transcription, followed by PCR amplification of the cDNA copies. The amplified cDNA copies are then sequenced and if long enough can map to a unique mRNA that has been described, the full sequence of which is known (Iwahana and Itakura, 1997; Dai et al., 2000; Cheng et al., 2009). To expand our understanding of the structural and functional characteristic of these sequences it is important to obtain full-length sequences. RACE is a technique based on PCR, which facilitates the cloning of full-length cDNA sequences when only a partial cDNA sequence is available (Eyal et al., 1999; Jurecic and Belmont, 2000; Kim et al., 2008). Using a simplified technique, the 3' end of the target sequence was amplified. The 3' end is generally considered an easier task than the 5' because of the cap structure found on the intact 5' (Dai et al., 2000). Many methods have been described to amplify the 5' and the 3' ends of cDNA (Hubank and Schatz, 1999; Livesey, 2000; Rong et al., 2008).

This modification is a fast and easy method to isolate the 3' ends genes only; the RACE primer sequence used as in usual PCR amplification reaction. The (oligo dT) is used as a second primer in the reaction after creating the second strand cDNA to the cDNA strand (Figure 1). Other competitor RACE technologies require ligation of a single-strand adaptor to RNA which is a very inefficient process (Noirdquist et al., 2000; Xianan and Baird 2001; Chenq et al., 2008; Messina et al., 2008).

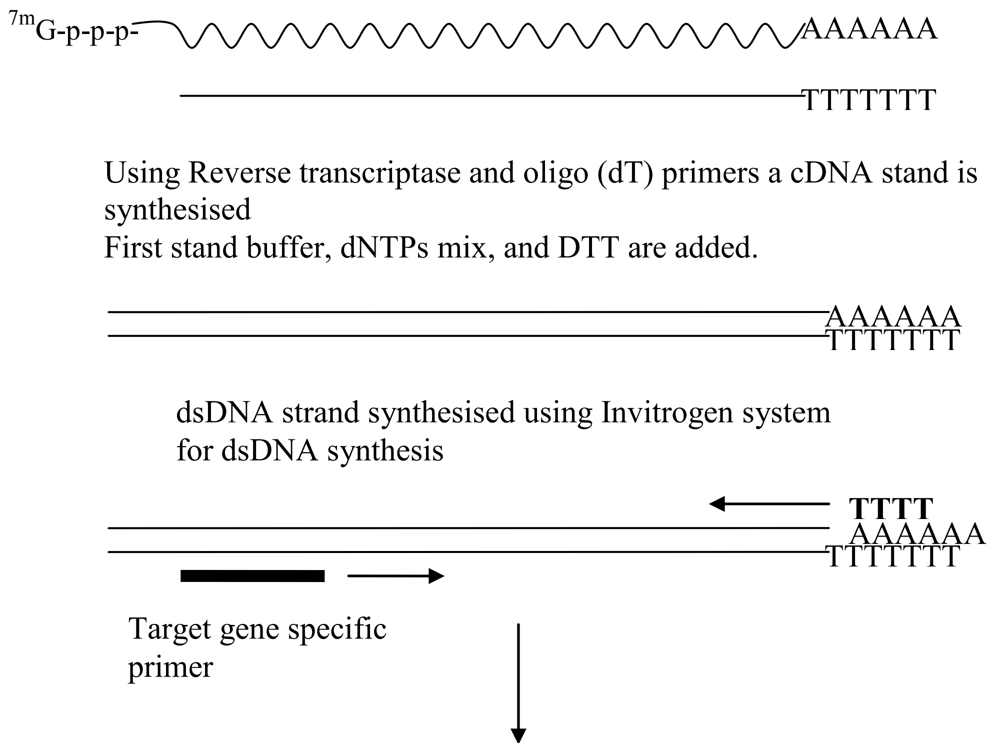

Fig 1: An overview of the modified RACE cDNA procedure. The diagram shows procedures for PCR reaction using target sequence –specific primers with an Oligo (dT) as second primers.

MATERIALS AND METHODS

Primer Design
Gene specific primers for the RACE were designed from the original clones produced from the RAD using the Genefisher Software (Bioinformatics Bielefeld) as shown in table 1.

RNA Isolation
RNA was isolated from rat endothelial cell supplied by the Athlone Institute of Technology Cell Culture and Toxicology Laboratory. Medium and supplement were ordered from Clonetics, BioWhittaker Inc, UK. The growth medium with all supplements and growth factors in separate, frozen aliquots. Total RNA was isolated using Trizol reagents from (Invitrogen) as described (Hubank and Schatz, 1999; Livesey, 2000). Reverse transcription (RT) reactions were performed using cDNA specific primers (Table1), a 200 U SuperScript II reverse transcriptase (Invitrogen). The RT reactions were performed in 50 µl of 2 mM dNTP, 10 pmol of the cDNA specific primers and between 0.5 and 1 µg of total RNA. The reaction was performed at 45ºC for 45 min. Two parallel reactions were performed with the reverse transcriptase omitted from one reaction as a negative control.

Table 1: List of the primers for amplification of the RACE products.
Modified RACE technique used to amplify the 3' end of the target sequence

Primer No	Forward Primer	Reverse Primer	Product size (base pair)
0.22s	GAATGTCCCAAAGGA	ACAGAGGTCTTTCCTCCA	143
0.46s	GGGAACACTGCCAGAGA	GAGCACTGACTAAGTGG A	150
55s	ACCTAGTAAAGGGGGTCA	CCATGCCAGTTCATAGC A	110
158s	CTCCCATCAACCTAGCAA	CCATGCCAGTTCATAGC A	119
Aligo d(T)	TTTTTTTTTTTTTTTTTTTT TTTT	-------------------------	
0.22rat	GTGCTGGCTATACACAGA	GTCAGGATTAGGCAGCA	EMBL:CA513489=500
0.55rat	CTGGCTGCTTTGAGTCA	TTGGTCAACCTGGTA	EMBL:BI301290=457
0.46rat	TAGTGGGAAGGAAAGGA	GGTCCTACGGCATCCA	EMBL:CA510864=498
0.25 sample	ATGCTCCACTTCTCCTGA	CCGGGGGATCAAAGCA	293

RACE method used was a simple modification of Tillett et al, 2000, with a gene specific primer from our known sequence and Oligo (dT) as a second primer in the reaction. A primer design program was used (Bioinformatics Bielefeld) to design the nested primer. The primers used are shown in Table 1. If the distance from the primers to the 3' end of the target was larger than 1 kb or unknown, an extended range *Taq* polymerase was used to improve the outcome.
Figure 2 shows 0.22, 0.25, 0.46, 0.55, 158 gene specific primer with the Oligo dT as a second primer. Different optimisation conditions resulted in different amplification patterns.

RESULTS.

On the gels of the non-specific amplification for the gene of clone 0.22 and 0.25, Lane 1 shows two intense bands amplified using the primers for clone 0.25 and annealing temperature of 50˚C, one with a size of nearly 310 base pair (bp), and the second of nearly 90 bp. Both were excised from the gel and

cloned. Lanes 2, 3, 4 and 5 were amplified with primer 0.22 with different annealing temperatures.

The gel in Figure 3 shows the non-specific amplification of an unknown gene of clone 0.55. Lane 1 employed annealing temperature of 49 ˚C while lane 2, 3 were at 50 ˚C, 51 ˚C and 52 ˚C respectively. Most lanes gave two very distinctive products at around 480 and 700 bp. Both of these bands were excised and inserted in a topo vector for cloning and sent for sequencing.

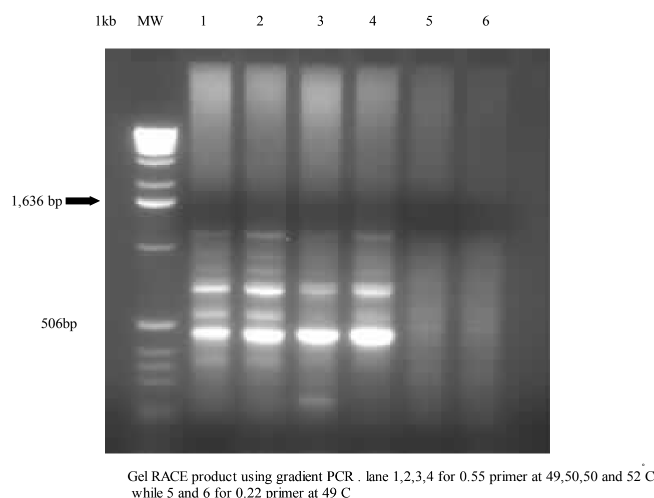

Figure 3: Gel shows the non-specific amplification of unknown gene of clone 0.55
at different annealing temperatures with a product of 0.22 amplification.

Cloning Race products

The PCR fragments regenerated from the reactions (see Plates 4) were run on TBE gels and bands excised and extracted using QIAquick Gel Extraction Kit (QIAGEN Ltd). PCR product was cloned using a "quick PCR cloning" Kit which employed Topoisomerase (T/A overhangs) (Invitrogen) following the manufacturer instruction. The vector with insert was transformed in a one shot chemically competent *E.*coli and mixed gently. Bacteria were grown on a selective medium. Positive clones were picked up and colony PCR used for positive clone detection.

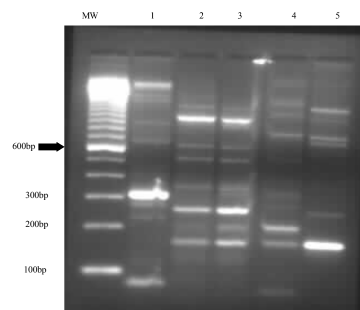

Fig 2: Gel shows non-specific amplification of unknown fragment of clone 0.22, 0.25 at different annealing temperatures. Lane 1 (primer for clone 0.25 and annealing T=50˚C), Lanes 2,3,4 and 5 (primer for clone 0.22 with different annealing temperature).

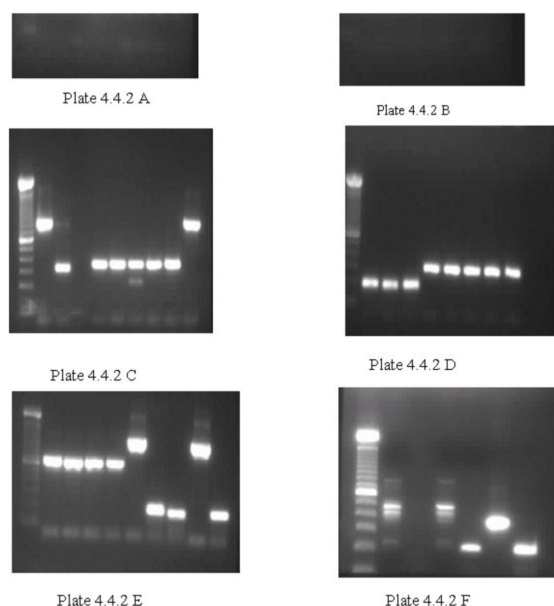

Fig 4 (A, B, C, D, and E): Colony PCR was employed to identify positive clones using M13 primers forward and reverse. Band size more than 200 base pair indicating a positive clone for target RACE sequence while bands, with less than 200 bp, were false positives which amplified only the vector region between the M13 primer. (MW-100bp molecular weight ladder), as shown in Table 2.

Table 2: Key – identification of bands of Fig 4 (A, B, C, D, and E).

Plate 4.4.2	Arrows indicates
A	Inserts amplified ± 300 bp
B	Inserts amplified with size 300 bp
C	Bands around 250 bp carry no inserts , while the other are positive
D	Arrow pointed to positive inserts
E	Arrow pointed to band size around 250 bp carry no inserts
F	Arrow pointed to band size around 250 bp carry no inserts

Analysis of RACE product

Around 50 bands from RACE products of the different reactions were selected and sent for sequencing. Bands from PCR product representing positive clones were extracted using QIAquick Gel Extraction Kit (QIAGEN Ltd). Samples were sent to MWG Germany for sequencing. Sequences of the clone were aligned with the original sequence of the fragment under investigation. Multiple alignment software was used to detect any possibility of matching between the RACE product and our sequences under investigation. Figure 4 (A, B,C,D and E) depicts the entire product of the RACE method for the target sequences that have been aligned together with the sequences under study using Clustal multiple alignment programme. The result of the alignment showed that only one (0.46 clone) of the target sequences had a strong homology with five of the Race products. The RACE products as shown in Fig 4 (A, B, C, D, and E) are C4, E1, F2, E8 and D2. The E1 and F2 products extended the 0.46 clones by nearly 200 base pairs. On the other hand none of the other RACE product gave any match to the other target sequences.

The RACE products, which had a high similarity to the 0.46 clone sequence were aligned together with the 0.46 clone sequence and are depicted in the above Figure 4 (A, B, C, D and E).

DISCUSSION

The 5'RACE technique was used to amplify nucleic acid sequences between a known sequence site and unknown sequence towards the 5' end of the mRNA. The technique was modified to make it easier and simple; the diagram in figure 1 illustrates the approach. The target sequence was selected for amplification and many different optimisation and temperature conditions were used in this study to achieve the amplification. Preliminary sequencing of clones 0.22, 0.25, 0.46, 0.55 and 158 and comparison with database sequences suggested that those sequences represent hypothetical proteins and some gene tools identified them as a conserved non-coding region, and therefore RACE was utilized to obtain further sequence of these clones.

We identified no clear new sequence for clone 0.46 using RACE, and the result was unable to give more confirmatory information on what the sequence might be. Other sequences did not produce any reliable product, which was a weakness of the technique. The developed sequence of 0.46 RACE represented an additional 263 bps. The 0.46 RACE products were investigated to look for homology in the database. BlASTN found a sequence with 80% homology, which coded for *Mus musculus* ADP-ribosylation factor 3 mRNA, with accession number BC024935.1, submitted by Strausberg, on the 6-oct-2003. It is a putative *Rattus norvegicus* ADP-ribosylation factor (cDNA clone MGC: 24115 IMAG: 3969152). The Vista plots from the result Fig 4(A, B, C, D, and E) showed the global alignment between the 0.46 RACE products with the BlASTN results. The plot implied that the sequence covered only a small region of the human sequence and was coloured blue to indicate it was an exon.

The EST database produced one sequence for *Rattus norvegicus* with 99% homology with 0.46 RACE with Accession number of CB724757.

The mapping of the 5' and 3' terminus of mRNA can be difficult if it is rare and / or if there is only a small amount of total RNA (Rudi et al., 1999; Tillett et al., 2000; Xianan and Baird, 2001; Brigaud et al., 2009). A variety of methods are available for mapping the 5' and 3' ends of RNA molecules, including RNase protection, S1 mapping and primer extension (Fehr et al., 1999; Kilstrup and Kristiansen, 2000; Wang et al., 2009). It was clear that these procedures were subject to various limitations (Yohko et al., 2005; Rong et al., 2008; Cheong et al., 2009). The approach was adopted in this study to amplify from cDNA using a single 5' custom-designed primer and a 3' oligo-dA primer. This method needed a lot of optimisation using different annealing temperatures to add non-specific amplification. Success was achieved in amplifying only an extra 200 base pair. Different annealing temperatures for the specific 5' primer bound to target sequence to allow preferential annealing of oligo dA to that sequence. A limitation of this method was that it was restricted to amplification of polyadenylated mRNA (Iwahana and Itakura, 1997; Hubank & Schatz 1999; Fehr et al., 1999; Messina et al., 2008; Wang et al., 2009), as well as allowing amplification of a sequence with the same 5' end but different 3' end. Technically, RACE is a time-consuming technique and sometimes the expected results were unachievable but in reality no alternative was available.

CONCLUSION

The clones were found to belong to hypothetical proteins according to the blast search in the Gene Bank database. More Bioinformatics work was performed using different bioinformatics tools to gather more information about what kind of protein and function might be for those fragments. Gene expression studies are necessary to determine whether genes isolated are truly expressed in cell samples from which they are derived

ACKNOWLEDGEMENT

We wish to acknowledge the Centre for Biopolymer and Biomolecular Research, Athlone Institute of Technology, Republic of Ireland for funding this project. The Ministry of Higher Education is also thanked for the special research grant for lecturers.

REFERENCES

Brigaud I, Grosmaître X, François MC, Jacquin-Joly E. Cloning and expression pattern of a putative octopamine/tyramine receptor in antennae of the noctuid moth Mamestra brassicae; Cell. Tissue. Research., 335(2):455-63, 2009.

Chen J, Xiao Y, Di P, Yu X, Chen W, Zhang L. Molecular cloning and characterization of a 2C-methyl-D: -erythritol 2, 4-cyclodiphosphate synthase gene from Cephalotaxus harringtonia. Molecular Biology and Reproduction; 36(7):1749-56, 2009.

Chenq Q, Cao Y, Pan H, Wang M, Huang M. Isolation and characterization of two genes encoding polygalacturonase-inhibiting protein from Populus deltoides. .Genetic & Genomics. 35(10):631-8, 2008.

Cheong N, Ramos JD, Tang CY, Chng HH, Yao R, Liang Z, Lee BW, Chua KY. Mite amylase from Blomia tropicalis (Blo t 4): differential allergenicity linked to geographical regions. International. Archive on Allergy and Immunology; 149(1):25-32, 2009.

Dai SM, Chen HH, Chang AD, Riggs DC and Flanagan SD. Ligation-mediated PCR for quantitative in vivo foot printing. Nature. Biotechnology 18: 1108-1111, 2000.

Eyal YH, Neumann E, Or H and Frydman A. Inverse single strand RACE: an adapter-independent method of 5' RACE. BioTechniques 27: 656-658.

Fehr C, Fickova M, Hiemke C, Dahmen N. Rapid cloning of cDNA ends polymerase chain reaction of G-protein-coupled receptor kinase 6: an improved method to determine 5'-and 3'-cDNA ends; Brain Research Proteins; 3: 242-251, 1999.

Hubank M, and Schatz DG. cDNA representational difference analysis: a sensitive and flexible method for identification of differentially expressed genes. Meth. Enzymol. 303: 325-349, 1999.

Iwahawa H, and Itakura M. An end trimming method and its application to amplify adjacent cDNA and genomic DNA fragments by PCR. Methods Molecular Biology 67: 247-260, 1997.

Jurecic R, and Belmont JW. Long distance DD-PCR and cDNA microarrays. Curr Opin. Microbiol. 3: 316-321, 2000.

Kilstrup M and Kristiansen KN. Rapid genome walking: a simplified oligo-cassette mediated polymerase chain reaction using a single genome-specific primer. Nucleic Acids Research 28: 655-662, 2000.

Kim M, Ahn IY, Cheon J, Park H. Molecular cloning and thermal stress-induced expression of a pi-class glutathione S-transferase (GST) in the Antarctic bivalve Laternula elliptica. Comp Biochem Physiol A Mol Integrated Physiology. 152(2):207-13, 2009.

Kim Y, Yeo S, Song HG, Choi HT. Enhanced expression of laccase during the degradation of endocrine disrupting chemicals in Trametes versicolor; Journal of Microbiology. 46(4):402-7, 2008.

Livesey SP. Functional Genomics. Oxford Univ Press, Oxford.p.513, 2000.

Messina A, Chirulli V, Gervasi PG, Longo V. Purification, molecular cloning, heterologous expression and characterization of pig CYP1A2. Xenobiotica; 38(12):1453-70, 2008.

Nordquist OP, Vinsbo-Lundberg C, Petterson U, Holmdahl R. Related Articles complex genetic control in a rat model for Rheumatoid Arthritis. Journal of Autoimmunity. 28:448-456, 2000.

Rong S, Cai JH, Andrews J. Cloning and apoptosis-inducing activities of canine and feline Trail. Mol Cancer Ther. 7(7):2181-91, 2008.

Rudi K, Fossheim T and Jakobsen KS. Restriction cutting independent method for cloning genomic DNA segments outside the boundaries of known sequences BioTechniques 27: 1170-1174, 1999.

Tillett D, Burns BP, and Neilan BA. Optimized rapid amplification of cDNA ends (RACE) for mapping bacterial mRNA transcripts. BioTechiques. 28: 448-452, 2000.

Vingsbo-Lundberg NN, Olofsson P, Sundvall M, Saxne T, Peterson U, Holmdahl R. OMIC Genetic control of arthritis onset, severity and chronicity in a model for rheumatoid arthritis in rats. Nature & Genetics. 20: 401-404, 1998.

Wang Z, Wu Z, Jian J, Lu Y. Cloning and expression of heat shock protein 70 gene in the haemocytes of pearl oyster (Pinctada fucata, Gould 1850) responding to bacterial challenge. Fish Shellfish Immunology, 26(4):639-45, 2009.

Xianan L and Baird WV. Rapid Amplification of Genomic DNA End by Nla III Partial Digestion and Ploynucleotide Tailing. Plant. Molecular. Biology. Reporter. 19:261-267, 2001.

Yohko MY, Matsubara Y, Tetsuya E, Yoh-ichi K, Kiyoko OG, Isobe N, Hattori S, Irie S. Cloning and nucleotides sequence of a novel 28-kDa protein from the mantle muscle of the squid Todarodes pacificus with homology to tropomyosin. Comparative. Biochemistry &. Physiology ; 141 : 3-12, 2005.