Molecular Characterization and Phylogeny Analysis Based on Sequences of Cytochrome Oxidase gene From *Hemiscorpius lepturus* of Iran

Abbas Jolodar

Department of Basic Sciences, Biochemistry and Molecular Biology Section, Faculty of Veterinary Medicine, Shahid Chamran University of Ahvaz, Ahvaz, Iran

**Abstract:**

**BACKGROUND:** *Hemiscorpius lepturus* is a medically important scorpion found along the Iranian borders, especially near to Khuzestan Province in the south-west of Iran. This is the only non-buthid scorpion which is potentially lethal in southern Iran and is responsible for severe dermonecrotic scorpionism.

**OBJECTIVES:** In this study, DNA fragment of the mitochondrial cytochrome c oxidase subunit I (COXI) gene of *H. Lepturus* for the molecular phylogenetic analysis was amplified.

**METHODS:** We amplified a 624 bp gene fragment of cytochrome C oxidase subunit I (COXI) from *H. lepturus* collected from Khuzestan, Ahvaz by PCR. After sequencing of the PCR products, the phylogenetic analysis was performed using the neighbor-joining method with 1000 replicates of bootstrapping using the MEGA7 software.

**RESULTS:** The results of phylogenetic analysis revealed four distinct clusters (A1, A2, B and C) belonging to the family of Hemiscorpiionidae that were grouped together with bootstrap score between 77-96%. The gene fragment of Hl-Kh formed a cluster relative to the only scorpion of *H. Lepturus* (HL) from Izeh, Iran with a good bootstrap score of 96. These two samples isolated from the Khuzestan province of Iran are closely related, as they clustered together as Cluster B. The genetic distances of Hl-Kh among the cluster A1, A2 and C ranged from 16 to 19% and the lowest interspecific distance was in cluster B between Hl-Kh and Hl (6%).

**CONCLUSIONS:** Hl-Kh isolated from the Khuzestan province of Iran are closely related to Hl, as they showed the greatest interspecific variation observed in Hemiscorpius genus in this study. Although the sample size in Cluster B is not large enough to draw a final conclusion, the percentage of sequence divergence was high enough for interspecific comparisons to provide separation of species.

**Keywords:**

*Hemiscorpius lepturus*, Scorpion, phylogenetic, Cytochrome C oxidase subunit 1, Venom
Introduction

Hemiscorpius lepturus (Scorpions: Hemiscorpionidae) is a medically important scorpion found along the Iranian borders, especially adjacent to Khuzestan province in the south-west of Iran. *H. lepturus* venom is known to have extremely toxic effects on the central nervous system and cardiovascular system (Radmanesh, 1990, 1998). This is the only dangerous and potentially deadly non-Buthid scorpion which is responsible for severe ulceration of the skin and also hemolysis of blood cells during the warm months of the year. Although only 10-12% of total scorpion stings during the hot season in Khuzestan province are due to this species, it is responsible for more than 95% of deaths (Manouchehirifar et al., 2013). *H. lepturus* seems to be restricted to Iraq and the western and south western regions of Iran. However, the distribution area of this species also includes the west of Pakistan (Fet, 2000). Three species belong to the family Hemiscorpidae that are known to exist in in Iran are *H. gaillardi*, *H. lepturus* and *H. persicus*. These species are morphologically very close to each other and difficult to distinguish for a non-skilled person, therefore we assume that *H. lepturus* is probably not the only species responsible for significant scorpionism problems in the southern provinces of Iran.

DNA sequences of the mitochondrial cytochrome c oxidase I (COXI) gene have been one of the most widely used molecular markers for identifying all kinds of animals. This is a key enzyme in aerobic metabolism. In fact, the evolution of this gene is rapid enough to allow the discrimination of not only closely related species, but also phylogenetic variations inside a particular species (Hebert et al., 2003). Phylogenetic analysis using COXI gene sequences were extensively carried out by several workers in different groups of organisms like the genus members of Puntius belongs to Cyprinidae family (Garg et al., 2017), 42 species of Culicoides (Diptera, Ceratopogonidae) from three continents (Augot et al., 2017 and Linto et al., 2002), white backed plant hopper Sogatella furcifera (Sreejith and Sebastian, 2014) and Commercially Available Pangasius hypophthalmus in Italy (Bellagamba et al., 2015). This gene is widely accepted as a DNA barcode for the accurate and easy identification of species.

In this study, DNA sequence of the mitochondrial cytochrome c oxidase subunit I (COXI) gene of *H. lepturus* for phylogenetic analysis was amplified.

Materials and Methods

Scorpion samples: Iranian scorpions *H. lepturus* were collected from Khuzestan Province in Iran. The specimens were obtained from the reference laboratory of the Razi Vaccine and Serum Research Institute in 95% ethanol and were frozen at -20 °C. An individual scorpion of Mesobuthus eupeus served as outgroup for phylogenetic analyses.

Genomic DNA extraction: Scorpion tissue (0.5-1.0 g) was homogenized in 10 ml of RSB buffer (10 mM Tris–HCl pH 7.4, 10 mM NaCl/25 mM EDTA, SDS 1%). The DNA was extracted with an equal volume of phenol/chloroform, and the aqueous phase was then extracted with chloroform. Precipitated Nucleic Acid is pelleted out by centrifuging at 12000 rpm for 15min before the pelleted DNA was washed with 70% ethanol, air dried and resuspended in 50μl
TE buffer. DNA concentration and purity were determined by calculating the absorbance ratio A260/280.

**PCR amplification and sequencing:**
PCR amplification was carried out on aliquots of the *H. Lepturus* genomic DNA as template with initial denaturation for 5 min at 95 °C, followed by 35 cycles of 45 s at 94 °C, 1 min at 48 °C, and 1 min at 72 °C and, finally, 7 min of incubation at 72 °C in a final reaction volume of 25 µl containing 50 mM KCl, 10mM Tris-HCl (pH 8.3), 20 pmol of each primer, deoxynucleotides (each at 220 µM), 1.5 mM MgCl2, 100 ng genomic DNA, and 1 U Taq polymerase. The primers were COXI-F 5’-GGTCAACAAATCAAGATATTGG and COXI-R 5’-TA-AACTCCAGGGTGACCAAAAAATCA (Folmer et al., 1994). The Amplified PCR products were electrophoresed through a 1% agarose gel, and stained with DNA Safe Stain (Sinaclon, Iran) before detection by UV transillumination. The amplified DNA fragments were extracted from agarose gel prior to performing DNA sequencing according to the dideoxy termination method using an automated Applied Biosystems 373 DNA sequencer. The sequence was determined for both strands by using overlapping fragments.

**DNA analysis:** The DNA sequences comparisons and putative functions of the DNA fragments were done using the blastn and blastx algorithms programs in the Nartional Center for Biotechnology Information (NCBI) GenBank database (Altschul et al., 1997). Similar sequences are downloaded from the NCBI database. The sequence obtained after sequencing was translated and its secondary structure of protein was predicted using the Prediction of Transmembrane Regions and Orientation server (em-bnet.vital-it.ch/software/TMPRED_form.html). The phylogenetic analysis was performed using the neighbor-joining method with 1000 replicates of bootstrapping using the MEGA7 software. A COXI sequence of Mesobuthus eupeus was used as outgroup.

**Results**

Amplification and sequence analysis. Starting with 0.05 g of fresh tissue, 5 µg of genomic DNA was obtained. The results obtained by PCR analysis confirmed the presence of COXI in *H. lepturus* since the expected 624 bp cDNA fragment was amplified (Fig. 1). No amplification product was obtained in the control samples where template DNA was excluded. According to the sequencing results (Fig. 2), the DNA contained a single open reading frame of 208 amino acids. The Blastx search indicated that the sequenced segment is closely related to different scorpion species from the order chelicerata. Domain analysis of HI-Kh fragment showed e-value of 6.81e-93 with conserved domain of cytochrome c oxidase subunit I (accession number MTH00142) with good query coverage. This multiple-domain contains Heme-copper oxidase subunit I Superfamily with e-value 3.20e-50. The PSIPRED protein sequence analysis method suggested the secondary structure of the protein. Each amino acid is allocated values for alpha helix, beta sheet and coils using a window of 7 amino acids. The
likelihood of a given amino acid calculated, and the secondary structure with the largest conformation probability is determined by the amino acid. We obtained a prediction of 60% alpha-helical for the coding sequence of COXI from *H. lepturus* (Fig. 3). The secondary structure data (Table 1) were used to enable a consensus confirmation of predicted transmembrane (TM) helices.

**Phylogenetic analysis:** By comparing the sequence of HI-Kh with the NCBI database, a total of 13 sequences including *Hemiscorpius lepturus* cytochrome oxidase subunit I (COI) gene (Hl; KU341987), *Diplocentrus tehuacanus* voucher AMCC:LP2044 cytochrome oxidase subunit I (COI) gene (Dt; K514661), *Brachistosternus ferrugineus* isolate LP_7632 cytochrome c oxidase subunit I (COI) gene (Bf; KT446962), *Scorpius maurus* fuliginosus isolate Sc51 cytochrome oxidase subunit I (COI) gene (Smf; FJ198060), *Brachistosternus paposo* isolate S027 cytochrome c oxidase subunit I (COI) gene (Bp; KX517218), *Brachistosternus roigalsinai* isolate S021 cytochrome c oxidase subunit I (COI) gene (Br; KX517215), *Pandinus cavimanus* cytochrome oxidase subunit I (COI) gene (Pc; AB156508), *Brachistosternus simoneae* isolate LP_10708 cytochrome c oxidase subunit I (COI) gene (Bs; KT447023), *Brachistosternus ehrenbergii* isolate LP_9279 cytochrome c oxidase subunit I (COI) gene (Be; KT446957), *Diplocentrus bereai* voucher AMMC:LP6532 cytochrome oxidase subunit I (COI) gene (Db; K514638), *Diplocentrus reddelli* voucher AMMC:LP10981 cytochrome oxidase subunit I (COI) gene (Dr; K514638), *Diplocentrus cozumel* voucher AMMC:LP4102 cytochrome oxidase subunit I (COI) gene (Dc; K514642), *Scorpio fuscus* isolate 12228H cytochrome oxidase subunit I (COI) gene (Sf; KT188243) and *Mesobuthus eupeus* cytochrome c oxidase subunit I gene (Me-Kh) were retrieved. The nucleotide sequence of HI-Kh with the retrieved sequence data was subjected to multiple sequence alignment and phylogenetic analysis using MEGA7 software (Fig. 4). Almost all the internal branches showed high bootstrap value bigger than 80. Branches corresponding to partitions reproduced

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>HI-Kh</td>
<td>0.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Hl</td>
<td>0.16</td>
<td>0.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Dt</td>
<td>0.16</td>
<td>0.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Bf</td>
<td>0.18</td>
<td>0.17</td>
<td>0.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Smf</td>
<td>0.17</td>
<td>0.15</td>
<td>0.16</td>
<td>0.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Bp</td>
<td>0.18</td>
<td>0.17</td>
<td>0.14</td>
<td>0.13</td>
<td>0.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Br</td>
<td>0.19</td>
<td>0.16</td>
<td>0.16</td>
<td>0.14</td>
<td>0.19</td>
<td>0.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Pc</td>
<td>0.19</td>
<td>0.17</td>
<td>0.17</td>
<td>0.18</td>
<td>0.16</td>
<td>0.21</td>
<td>0.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Bs</td>
<td>0.19</td>
<td>0.18</td>
<td>0.17</td>
<td>0.06</td>
<td>0.19</td>
<td>0.15</td>
<td>0.15</td>
<td>0.15</td>
<td>0.19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Be</td>
<td>0.18</td>
<td>0.16</td>
<td>0.16</td>
<td>0.14</td>
<td>0.18</td>
<td>0.12</td>
<td>0.10</td>
<td>0.19</td>
<td>0.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Db</td>
<td>0.16</td>
<td>0.15</td>
<td>0.11</td>
<td>0.18</td>
<td>0.16</td>
<td>0.14</td>
<td>0.14</td>
<td>0.16</td>
<td>0.17</td>
<td>0.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Dr</td>
<td>0.18</td>
<td>0.16</td>
<td>0.11</td>
<td>0.18</td>
<td>0.18</td>
<td>0.14</td>
<td>0.15</td>
<td>0.18</td>
<td>0.18</td>
<td>0.14</td>
<td>0.10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Dc</td>
<td>0.18</td>
<td>0.16</td>
<td>0.12</td>
<td>0.17</td>
<td>0.18</td>
<td>0.14</td>
<td>0.15</td>
<td>0.19</td>
<td>0.17</td>
<td>0.17</td>
<td>0.12</td>
<td>0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Sf</td>
<td>0.18</td>
<td>0.18</td>
<td>0.18</td>
<td>0.19</td>
<td>0.09</td>
<td>0.17</td>
<td>0.17</td>
<td>0.17</td>
<td>0.19</td>
<td>0.18</td>
<td>0.17</td>
<td>0.17</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Me-Kh</td>
<td>0.25</td>
<td>0.25</td>
<td>0.24</td>
<td>0.25</td>
<td>0.24</td>
<td>0.27</td>
<td>0.25</td>
<td>0.26</td>
<td>0.26</td>
<td>0.22</td>
<td>0.22</td>
<td>0.24</td>
<td>0.24</td>
<td>0.26</td>
</tr>
</tbody>
</table>
in less than 50% bootstrap replicates are collapsed. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to conclude the phylogenetic tree. *M. eupeus* (Me) used as an out group was clearly branched separately, which signified the credibility of constructed phylogram. This phylogeny allows addressing several important taxonomic issues relating to the Scorpionidae family. The constructed phylogram showed four distinct clusters (A1, A2, B and C). The phylogenetic tree shows that all the species belonging to Cluster A are joined together with a reasonable bootstrap score of 77. This major cluster was further subdivided into two sub-clusters with a high bootstrap value of 98 (Cluster A1; Brachistosternus) and 88 (Cluster A2; Diplocentrus). The species belonging to the genus Hemiscorpius were grouped together with a good bootstrap score of 96. The Khuzestan’s sample Hl-Kh formed a cluster (Cluster B) relative to the only retrieved sample of Genbank *H. Lepturus* (Hl) from Iran. Within this cluster, there is a strong support for species Hl and Hl-Kh (96%). The similarity of the species belonging to Pandinus cavimanus (Pc) and *Scorpio fuscus* (Sf) formed a single cluster with a score of 83 (Cluster C). As it was expected Hl-Kh used in this study was perfectly arranged in cluster B showing higher genetic similarity with Hl and showed higher genetic distance particularly with species in Cluster A.

**Genetic distances:** The genetic distances between Hemiscorpionidae family scorpions (Cluster A1, A2, B and C) and the outgroup species *M. eupeus* were calculated (Table 2). The genetic variations of Hl-Kh with cluster A1, A2 and C ranged from 16 to 19%. The highest distance was observed with cluster C (18-19%) and the lowest distance was calculated between Hl-Kh and Hl (6%). Finally, as expected, the distances between Hl-Kh and the outgroup species *M. eupeus* was 25%.

**Discussion**

In common with other protein coding genes, COXI appears to have a greater range of phylogenetic signal than any other mitochondrial gene. The third-position nucleotides of this gene show a high frequency rate of base alterations. This makes it about three times more valuable than that of 12S or 16S rDNA in molecular phylogeny (Knowlton and Weigt, 1998).

This study was undertaken with the aim to sequence the mitochondrial Cytochrome C Oxidase subunit 1 for the species *H. lepturus*. Comparison of 624 bp of the encoded COX I gene provided good resolution of the species that were examined in this study. TMs were found as shown in Fig. 3. The aim of this TM search step was to confirm our sequence database based hypothesis of the putative sequence being identi-
Identified as a cytochrome C oxidase subunit I. Cytochrome C oxidases are known to have these TM regions and the identification of these regions served as a confirmatory step in determining the validity of the predicted structure (Mohamed et al., 2003). Although COXI gene is somehow conserved among species, both at the level of nucleotide and amino acid sequence, the presence of those TM extents predicted to form amphipathic alpha-helical secondary structures are relatively conserved (Saraste et al., 1994).

The Khuzestan’s *Hemiscorpius* scorpions in cluster B including Hl-Kh and Hl (isolated from Izeh in Khuzestan) are very divergent from the other Hemiscorpionidae family members (Cluster A1, A2, and C) which are isolated from North Africa, Middle east, South America and Brazil. From the phylogenetic studies we also understand that the
Hl-Kh sequence was clustered with the only entry in the Genbank database corresponding to *H. lepturus* (HI) with a relatively high distance (6%) among Hemiscorpius genus in Cluster B. Hl-Kh isolated from the Khuzestan province of Iran are closely related to Hl, as they showed the greatest interspecific variation observed in Hemiscorpius genus in this study. Although the sample size in this cluster is not large enough to draw a final conclusion, the percentage of sequence divergence was high enough for interspecific comparisons to provide separation of species. But, further studies need to be conducted to prove this concept by analyzing total mitochondrial DNA sequence.

**Acknowledgments**

This work was supported by a grant from the Shahid Chamran University of Ahvaz, Ahvaz, Iran.

**Conflicts of interest**

The author declared no conflict of interest.

**References**


tochondrial cytochrome oxidase subunit I gene of five species of the *Culicoides imicola* species complex. Med Vet Entomol, 16(2), 139-46. https://doi.org/10.1046/j.1365-2915.2002.00356.x PMID: 12109707


شناسایی ملکولی و آنالیز فیلوژنی عقرب همیسکورپیوس لپچروس ایران بر اساس توالی زن سیتوکروم C اکسیداز

عباس جلودار
گروه علوم پایه، بخش بیوشیمی و بیولوژی مولکولی، دانشکده دامپزشکی، دانشگاه شهید چمران اهواز، ایران

چکیده
زمینه مطالعه: همیسکورپیوس لپچروس یک عقرب با اهمیت از نظر پزشکی است که بموازات مرز ایران بویژه در استان خوزستان واقع در جنوب غرب ایران یافت می‌شود. این از نظر گونه عقرب از خانواده غیر بوتیده است. می‌تواند عامل اصلی درماتیت شدید ناشی از گزش عقرب‌ها در استان گردید.
هدف از این مطالعه شناسایی ملکولی و آنالیز فیلوژنیک عقرب همیسکورپیوس لپچروس با استفاده از زن سیتوکروم C اکسیداز بود.

روش کار: قطعه زن سیتوکروم C با استفاده از آغازگرهای اختصاصی بوسیله PCR موجود در توالی‌های سایر عقرب‌های بندری و BA1 و BA2 در مرحله یک و آغازگرهای مشابه آن در مرحله دوم از زنجیره‌ی اس اس بالا به کار رفتند. تکثیر و توالی‌یابی گردید و با استفاده از نرم‌افزار ClustalW نتایج فیلوژنیک بدست آمد.
نتایج: توالی‌های مربوط به همیسکورپیوس لپچروس گروه دومی از نظر بسیار مشابه بودند. در کل گروه برابر با نتایج فیلوژنیک مشخص شد که این گروه در بین گروه‌های اصلی قطعه سیتوکروم C با بسیاری از نتایج مشابه بود.

نتیجه گیری: که با توجه به گروه‌بندی موجود در گروه B قطعه سیتوکروم C با بسیاری از نتایج مشابه بود.

واژه‌های کلیدی: همیسکورپیوس لپچروس، عقرب، فیلوژنیک، سیتوکروم C اکسیداز

Email: jolodara@scu.ac.i