Indian Journal of Biotechnology
Vol 5, April 2006, pp 195-199

Cytochrome oxidase I sequence of *Helicoverpa* (Noctuidae: Lepidoptera) species in India—Its utility as a molecular tool

S Kranthi, K R Kranthi, A A Bharose, S N Syed, C S Dhawad, R M Wadaskar, G T Behere and E K Patil
Central Institute for Cotton Research, Post Bag No 2, Shankar Nagar P.O. Nagpur 440 010

Received 2 November 2004; revised 25 May 2005; accepted 20 July 2005

The genus *Helicoverpa* comprises of 2 species in India—*H. armigera* and *H. assulta*. This paper compares partial CO-1 sequences of the two-field-collected species with a laboratory strain of *H. armigera* and *Drosophila yakuba* whose entire mitochondrial genome has been sequenced. The region sequenced in the study corresponds to 2111 to 2601 bp sequence of *D. yakuba* mitochondrial genome, i.e mid to near terminal segment of the CO-I region. When analyzed, at least 18 nucleotide and 8 amino acid substitutions were observed between the two species. Using this information, a specific PCR-RFLP tool was designed that distinguishes between the two species at the egg stage itself, thus, influencing pest control options significantly, especially in areas and crops where the two may occur simultaneously.

Keywords: CO-I sequence, *Drosophila yakuba*, *Helicoverpa assulta*, *H. armigera*

IPC Code: Int. Cl.[8] C12N15/10, 15/52

Introduction

The genus *Helicoverpa* comprises of only two species in India—*H. armigera* and *H. assulta* that are visually different as adults. *H. armigera* (Huebner) is a polyphagous pest, recorded on about 181 host plants and considered the most important pest of field crops in India[1]. Whereas *H. assulta* (Guenee) is an oligophagus pest that feeds largely on solanaceous crops, but not exclusively on food plants, such as tobacco (*Nicotiana tobaccum)* and *Datura* spp.[2]. It is common to encounter both the species together in vegetable ecosystems, such as tomato. *H. assulta* is a pest of predominantly dicots, while *H. armigera* is found on both dicots and monocots. The service of trained entomologist is required to differentiate the two species in the adult stage, applying taxonomic keys as described by Hardwick[3] and Mathews[4,5]. However, eggs and neonates are virtually indistinguishable in mixed populations.

Studies on inter and intra specific variations in insects are often attempted using taxonomic, biochemical or molecular markers[6,7]. The taxonomic tools are often laborious, time-consuming and require considerable skill; whereas biochemical markers, such as esterase isozymes, are influenced by the environment, food plants (hosts) and stage of the insect. In contrast, molecular methods are relatively recent, reasonably accurate and could be complimented with other methods.

Mitochondrial genome is maternally inherited[8]. Any change in the mitochondrial (mt) DNA is transmitted to the entire progeny. Evolutionary changes in conserved regions of the mt DNA spread rapidly within populations. If the changes have a functional significance in evolutionary adaptation, the change itself along with the extent of spread represents the adapted strains. Such changes, which can be micro evolutionary in nature, can be a function of selection pressures induced by suitability or unsuitability of host plants, insecticides, predators and parasitoids or other biotic and abiotic stresses. *H. armigera* (cotton bollworm) has been subjected to such pressures over the past decades. Though changes were speculated in terms of race or geographical variability, there has so far been no genetic evidence to indicate the same

The cytochrome oxidase I (CO-I) region of mt DNA is the most studied region of the insect mitochondrial genome. CO-I is the terminal catalyst in mitochondrial respiratory chain and its biochemical role has been clearly elucidated. It is also the largest of the three mitochondrial encoded cytochrome oxidase subunits, which is one of the largest protein coding genes of metazoan mitochondrial genome[9].

*Author for correspondence:
Tel: 91-7103-275549; Fax: 91-7103-275529
E-mail: Sandhya.Kranthi@gmail.com

The present study was undertaken to compare the partial CO-I sequences of *H. assulta* and *H. armigera* field collected populations. Sequences were also compared with a laboratory-reared strain of *H. armigera* and *Drosophila yakuba,* an insect whose entire mitochondrial genome has been sequenced[10]. Consistent base pair substitutions between the two species of *Helicoverpa* and specific restriction enzymes that can cleave at the point of single nucleotide polymorphism were identified. Further, a molecular tool that distinguishes the two species in the morphologically indistinguishable stage, such as eggs, larvae and pupae, was presented.

Materials and Methods

Insects

Populations of *H. armigera* were obtained from cotton fields of Nagpur district, Maharashtra, India. During the same period, *H. assulta* were collected from the plants of *Datura* sp. The laboratory-reared population of *H. armigera* was an exclusive insecticide susceptible strain. Larvae were reared on a semi synthetic diet[11] up to F_1 generation. The freshly emerged female moths were used for the study.

Isolation of Genomic DNA

DNA was isolated from ovaries unmated female moths of F_2 generation[12].

PCR Protocols

Primers were designed to amplify the mid region of the CO-I gene based on conserved sequences[8]. The mid CO-I region is reported to have a high functional significance and was therefore chosen for the study. The forward (C1-J-2090) and reverse (C1-N-2659) primers have the following sequence: 5′-AGT TTT AGC AGG AGC AAT TAC TAT-3′ (C1-J-2090) and 5′-GCT AAT CCA GTA AAT AAA GG-3′ (C1-N-2659). Desalted and HPLC purified primers were obtained from Thermo Hybaid (Germany) and *Taq* polymerase was obtained from Genetix USA. The dissolved genomic DNA (1 µL) was diluted in 9 µL of 10 m*M* Tris buffer (*p*H 8.3). The diluted sample (1 µL) was used for amplification in a 50 µL PCR mixture. The PCR program adopted was as follows: 94°C (30 sec), 50°C (30 sec) 72°C (1.5 min) for 38 cycles, 72°C (10 min) and finally at 4°C on a MJ Thermal Cycler (PTC 100$^{TM)}$. Amplified fragments were visualized on 1% agarose gel.

Sequencing

Amplified products were purified using gel elution columns (Sigma). They were quantified by agarose gel electrophoresis and sequenced on Beckmann 8 Capillary System, CEQ 2000, following the protocol described in the instruction manual. Field-collected *H. armigera* (4 moths), *H. assulta* (4 moths) and an exclusive lab reared *H. armigera* strain (2 moths) were sequenced using both forward and reverse primers. Multiple sequence alignment was done using Clustal X (1.81)[3].

Restriction Digestion

The PCR product (9 µL) was directly digested at 37°C for 6 h with *Alu* 1 in a 20-µL-reaction mixture, containing 2U restriction enzyme, 2 µL 10X RE buffer and distilled water. Digested DNA bands were separated on 2% agarose gel and were visualized on Kodak EDAS 230 gel documentation system.

Results and Discussion

Primers amplified the expected fragment size of DNA. Blank DNA extraction, sequencing controls and negative PCR controls were carried out along with normal experiments to check against cross-contamination. *H. armigera* (AY 264944) and *H. assulta* (AY 264943) sequences are available on the NCBI site (http://www.ncbi.nlm.nih.gov). Using primers C1-J-2090 and CI-N-2659, a region with an approximate size of 560 bp, demonstrating a high fidelity of both forward and reverse reads, was sequenced (Fig. 1). Translation of the nucleotide sequence into amino acid sequences, with the invertebrate mitochondrial genetic code[10] in each of the populations, resulted in an amino sequence with a high level of consensus between the three populations of *Helicoverpa* species, which were rather different from the sequence of *D. yakuba* (Fig. 2). The CO-I region in *D. yakuba* extends between 1470 to 3009 bp[10]. The region sequenced in this study corresponds to the 2111 to 2601 bp sequence of *D. yakuba*, i.e the mid to near terminal region of the CO-I region.

The molecular diagnostic tool that is being proposed involves isolation of the genomic DNA, amplifying it with the primers C1-J-2090 and CI-N-2659, restricting the amplified fragment with either *Rsa*1 or *Alu* 1 and checking the restricted digests by agarose gel electrophoresis. The restricted-digest pattern obtained by restriction with *Alu* 1 (from

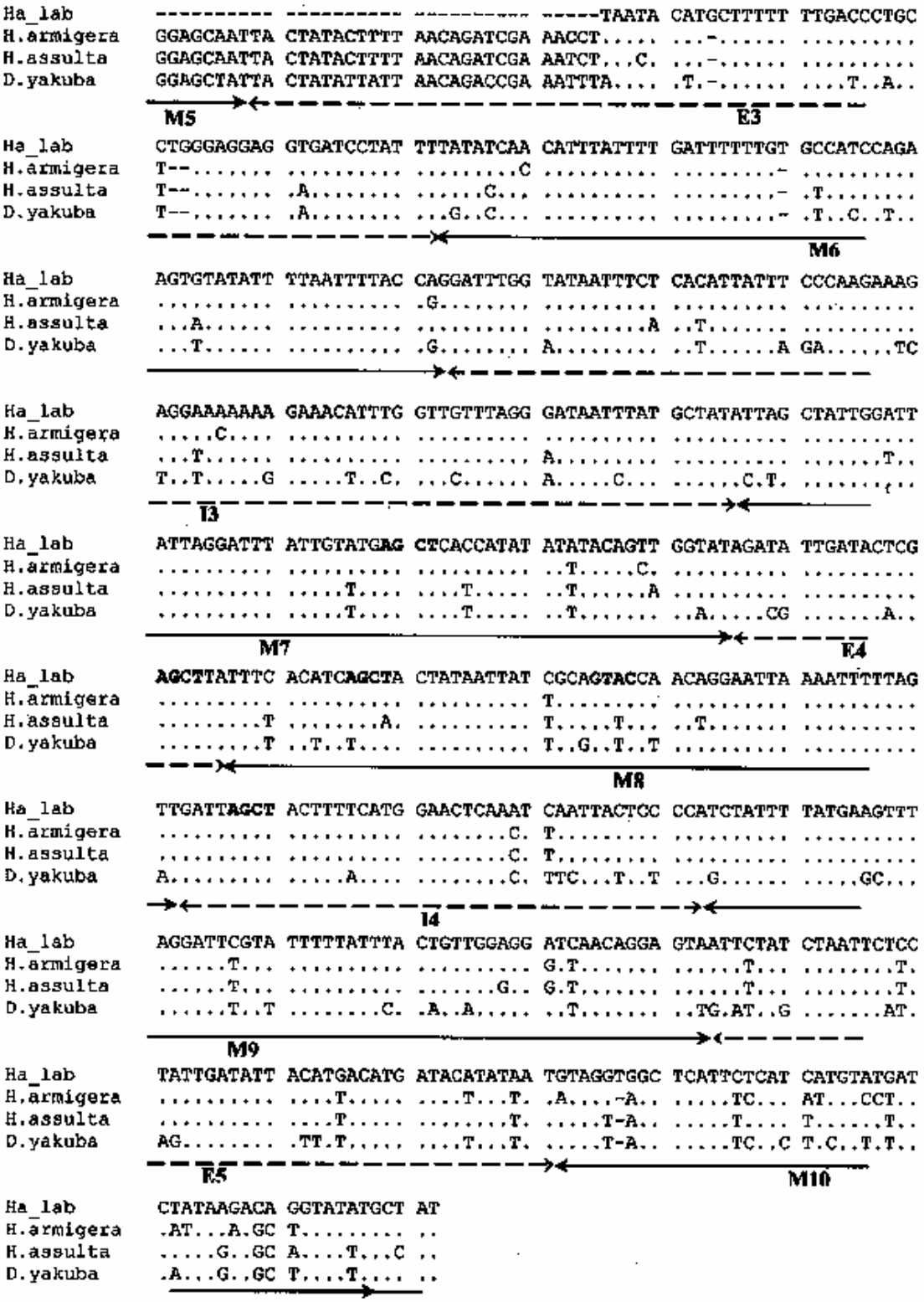

Fig. 1—Multiple sequence alignment of partial nucleotide sequence of cytochrome oxidase subunit 1 (CO-I) of *Helicoverpa armigera* (H. armigera) GenBank Accession No. AY264944, *Helicoverpa armigera* Laboratory strain (Ha_lab), *Helicoverpa assulta* (H. assulta) GenBank Accession No. AY264943 and *Drosophila yakuba* (D. yakuba) GenBank Accession No. X03240. Structural regions are indicated (M1, 2, Transmembrane region; E1, 2, External loops; I1, 2, Internal loops).

Bangalore Genie, India) demonstrated a banding pattern as shown in Fig. 3. PCR-RFLP with *Alu* 1 clearly distinguished *H. armigera* from *H. assulta*.

The sequencing of amplified products showed that 570 bp region of laboratory strain of *H. armigera* contained 68.9% A-T and 30.6% G-C contents. While 557 bp sequenced region of field strains of both *H. armigera* and *H. assulta* contained identical A-T and G-C contents and that were 70.2 and 30.8%, respectively. The laboratory strain demonstrated deletion of base pairs in a few positions.

The partial CO-I sequences were analyzed using Jellyfish Version 3.0.1 (Lab Velocity Inc. USA 2001) for the restriction sites. Two restriction enzymes were identified that could selectively restrict the sequenced region of CO-I at single nucleotide polymorphic sites and, hence, could differentiate between the two species.

The partial CO-I sequences of *H. armigera* and *H. assulta* indicate that although the GC content is similar, 18 nucleotide substitutions occur between the two. Further, amino acid sequences show eight substitutions. Most of the substitutions involve the replacement of one hydrophobic amino acid with another. The CO-I amino acid sequence is divided into twenty-five regions comprising of five structural

```
H.assulta     GAITMLLTDR NLNTSFFDPA GGGDPILYQH LFW-FFGHPE VYILILPGFG MISHIISQES
D.yakuba      .......... .......... .......... .......... .......... ..........
H.arm         .......... .......... ........H. .......... .......... ..........
H.arm-Lab     ---------- --MHA..... WEEVILF.IN IYFD.LC... .......... ..........

H.assulta     GKKETFGCLG MIYAMLAIGL LGFIVWAHHM FTVGMDIDTR AYFTSATMII AVPTGIKIFS
D.yakuba      .......S.. .......... .......... ......V... .......... ..........
H.arm         .T........ .......... .......... ..A....... .......... ..........
H.arm-Lab     .......... .......... .......... Y......... .......... ..........

H.assulta     WLATFHGTQL NYSPSILWSL GFVFLFTVGG LTGVILSNSS IDITLHDTYY VVAHFHYVLS
D.yakuba      ....L..... S...A...A. .......... ....V.A... V..I...... ..........
H.arm         .......... .......... .......... .......... .......... MG....I...
H.arm-Lab     .........I .......... .......... S........P ....W....N .GGS.SSCMI

H.assulta     MGAVFAILEG IIMMP
D.yakuba      .......MA. F.HWY
H.arm         LK..Y..MVA HFHWY
H.arm-Lab     YKTGMCY.-- -----
```

Fig. 2—Multiple sequence alignment of partial amino acid sequence of cytochrome oxidase subunit I (CO-I) of *Helicoverpa armigera* (H. armigera) GenBank Accession No. AAP21834, *Helicoverpa armigera* Lab strain (Ha-lab), *Helicoverpa assulta* (H. assulta) GenBank Accession No. AAP21833 and *Drosophila yakuba* (D.yakuba).

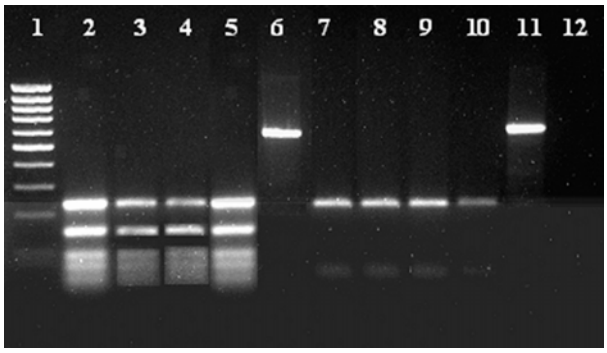

Fig. 3—PCR-RFLP of *H. armigera* and *H. assulta* partial CO-I with *Alu* 1: Lane 1, molecular weight marker 1Kb; Lane 2-5, PCR-RFLP product of *H. assulta*; Lane 6, undigested PCR product of *H. assulta*, Lane 7-10, PCR-RFLP product of *H. armigera*; Lane 11, undigested PCR product of *H. armigera*; and Lane 12: negative control.

classes (twelve transmembrane helices, M1-M12; six external loops, E1-E6; five internal loops; and carboxy and amino terminals)[14]. Our nucleotide sequence corresponds to the two external loops (E3 and E5), five transmembrane helices (M6, M7, M8, M9, M10) and two internal loops (I3 and I4) described in other insects. Of these regions M9 and I4 are reported to be the most variable. Transmembrane helices M2, M6, M7 and M10 provide metal ligands to interact with the two-haeme groups and copper atom that is essential for the activity of CO-I[14]. M8 is supposed to be involved in the cytochrome oxidase proton conduction channel. Thr, Thr and Lys, indicated in bold italics, in the translated amino acid sequence are polar residues that are reported to be completely conserved among all animals so far

studied; the list now also includes *Helicoverpa*. These three amino acids are thought to be essential for translocational activity. Of five internal loops, internal loop 1 is highly conserved in contrast to other internal loops that are relatively variable. E5 is thought to lie very close to haeme A, in the association of which tyrosine (marked in bold) has been suggested to play an important role[15]. Just following this region, the substitution of two consecutive valine (amino acid with hydrophobic R-group) positions with methionine (amino acid with hydrophobic R-group) in *H. assulta* and with glycine (amino acid with uncharged polar R group) in *H. armigera* is therefore significant.

CO-I and CO-II sequences have often been used to distinguish organisms at the species level. Five amino acid substitutions in the CO-I and CO-II regions were found between two budworm species, of which two substitutions were in the CO-I region[16]. Although two of these changes involved non-polar residues that unlikely to have major structural consequences, the other three involved changes to amino acids of different classes, which was reported to reflect naturally selected difference. At the amino acid level, structural and functional similarities in any given protein compared among diverse organisms are reflected in the proportion of hydrophobic versus hydrophilic amino acids in different regions of the gene[8]. A 400 bp region sequenced between 2301 to 2700 bp of the CO-I gene was also used to analyze the phylogenetic relationship between coleopterans[17]. This region closely overlaps the region mentioned in the present study. Differences in partial CO-I nucleotide and amino acid sequences within and between other species of insects have been reported[17-19]. However, the CO-I regions sequenced either overlap or differ from the sequence being reported here.

This paper presents differences observed in the nucleotide and amino acid sequences between two species of *Helicoverpa*. Using the nucleotide sequence, a specific molecular diagnostic kit has been designed to distinguish the two species of *Helicoverpa*.

References

1 Manjunath T M, Bhatnagar V S, Pawar C S & Sithanatham S, Economic importance of *Heliothis* spp. in India and an assessment of their natural enemies and host plants, in *Proc of Workshop on Biological Control of Heliothis: Increasing the effectiveness of natural enemies* (edited by E G King & R D Jackson), heald on 11-15 Nov. 1985, New Delhi, India. Pp 197-228.

2 Jadhav D R & Armes N J, Comparative status of insecticide resistance in the *Helicoverpa* and *Heliothis* species (Lepidoptera: Noctuidae) of South India, *Bull Entomol Res*, 86 (1996), 525-531.

3 Hardwick D F, The corn earworm complex, *Mem Entomol Soc Can*, 40 (1965) 245.

4 Mathews M, Classification of the Heliothinae, in *NRI Bulletin No.44* (Natural Resources Institute, Chatham, UK) 1991, 198.

5 Mathews M, Heliothine moths of Australia: A guide to pest bollworms and related noctuid groups, *Monogr Aust Lepidoptera*, 7 (1999) 320.

6 Pashley P, Host associated differentiation in armyworms (Lepidoptera: Noctuidae): An allozymic and mitochondrial DNA perspective, in *Electrophoretic studies on agricultural pests*, edited by H D Loxdale & J Hollander (Clarendon Press, Oxford), *Syst Assoc Spec Vol*, 39 (1989), 103-114.

7 Harrison R G, Molecular changes at speciation, *Annu Rev Ecol Syst*, 22 (1991) 281-308.

8 Simon C, Frati F, Beckenbach A, Crespi B, Liu H *et al*, Evolution weighting and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers, *Ann Entomol Soc Am*, 87 (1994) 651-701.

9 Lunt D H, Zhang D X, Szymura J M & Hewitt G M, The insect cytochrome oxidase I gene, evolutionary patterns and conserved primers for phylogenetic studies, *Insect Mol Biol,* 5 (1996) 153-165.

10 Clary D O & Wolstenholme D R, The mitochondrial DNA molecule of *Drosophila yakuba*: Nucleotide sequence, gene organization and genetic code, *J Mol Evol,* 22 (1985) 252-271

11 Armes N J, Bond G S & Cooter R J, The laboratory culture and development of *Helicoverpa armigera*, in *NRI Bulletin No. 57* (Natural Resources Institute, Chatham, UK) 1992, 57

12 Zhang D X, Szymura J M & Hewitt G M, Evolution and structural conservation of the control region of insect mitochondrial DNA, *J Mol Evol*, 40 (1995) 382-391.

13 Thompson J D, Gibson T J, Plewniak F, Jeanmougin F & Higgins D G, The Clustal X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools, *Nucleic Acids Res*, 24 (1997) 4876-4882.

14 Gennis R B, Site directed mutagenesis studies on sub unit 1 of the aa 3 type cytochrome C oxidase of *Rhodobacter sphaeroides*: A brief review of progress to date, *Biochem Biophys Acta,* 1101 (1992) 184-187.

15 Holm L, Saraste M & Wilkstrom M, Structural models of the redox centers in cytochrome oxidase, *EMBO,* 6 (1987) 2819-2823.

16 Sperling F A H & Hickey D A, Mitochondrial DNA sequence variation in the spruce budworm species complex (*Choristoneura* spp; Lepidoptera), *Mol Biol Evol,* 11 (1994) 656-665.

17 Howland D E & Hewitt G M, Phylogeny of the Coleoptera based on mitochondrial cytochrome oxidase I sequence data, *Insect Mol Biol*, 4 (1995) 203-215.

18 Kelley S T & Farrell B D, Strong differentiation in mitochondrial DNA of *Dendroctonous brevicomis* on different subspecies of *Ponderosa* pines, *Ann Entomol Soc Am,* 92 (2) **(**1999) 193-197.

19 Shufran K A, Burd J D, Anstead J A & Lushai G, Mitochodrial DNA sequence divergence among green bug biotypes: Evidence for host adapted races, *Insect Mol Biol*, 9 (1999) 179-184.