Heredity (2003) 90, 382–389

www.nature.com/hdy

The *ceratotoxin* gene family in the medfly *Ceratitis capitata* and the Natal fruit fly *Ceratitis rosa* (Diptera: Tephritidae)

M Rosetto[1,3], D Marchini[1], T de Filippis[1], S Ciolfi[1], F Frati[1], S Quilici[2] and R Dallai[1]
[1]*Department of Evolutionary Biology, University of Siena, Siena, Italy;* [2]*CIRAD-FLHOR, Laboratoire d'Entomologie, Saint-Pierre, France*

Ceratotoxins (Ctxs) are a family of antibacterial sex-specific peptides expressed in the female reproductive accessory glands of the Mediterranean fruit fly *Ceratitis capitata*. As a first step in the study of molecular evolution of *Ctx* genes in *Ceratitis*, partial genomic sequences encoding four distinct *Ctx* precursors have been determined. In addition, anti-*Escherichia coli* activity very similar to that of the accessory gland secretion from *C. capitata* was found in the accessory gland secretion from *Ceratitis (Pterandrus) rosa*. SDS–PAGE analysis of the female reproductive accessory glands from *C. rosa* showed a band with a molecular mass (3 kDa) compatible with that of Ctx peptides, also slightly reacting with an anti-Ctx serum. Four nucleotide sequences encoding *Ctx*-like precursors in *C. rosa* were determined. Sequence and phylogenetic analyses show that *Ctxs* from *C. rosa* fall into different groups as *C. capitata Ctxs*. Our results suggest that the evolution of the *ceratotoxin* gene family might be viewed as a combination of duplication events that occurred prior to and following the split between *C. capitata* and *C. rosa*. Genomic hybridization demonstrated the presence of multiple *Ctx*-like sequences in *C. rosa*, but low-stringency Southern blot analyses failed to recover members of this gene family in other tephritid flies.
Heredity (2003) **90,** 382–389. doi:10.1038/sj.hdy.6800258

Keywords: female-specific peptides; antibacterial peptides; insect immunity; molecular evolution; gene duplication

Introduction

The female reproductive accessory glands of the Mediterranean fruit fly *Ceratitis capitata* (family Tephritidae) produce a secretion with antibacterial properties, partly due to the presence of short cationic antibacterial peptides named Ceratotoxins (Ctxs) (Marchini *et al*, 1991, 1993). Three peptides, named Ctxs A, B and C, were previously isolated from the gland secretion, and their amino-acid and cDNA nucleotide sequences were determined (Marchini *et al*, 1993, 1995; Rosetto *et al*, 1996).

Unlike most insect antibacterial peptides, Ctxs are not induced by bacterial infection, but they are expressed in the female reproductive accessory glands of adult insects (Marchini *et al*, 1995; Rosetto *et al*, 1996) in response to juvenile hormone stimulation (Manetti *et al*, 1997). Since Ctxs are produced only after sexual maturity is achieved, their possible physiological role could be related to the protection of the female reproductive tract from bacterial invasion during mating. However, the presence of biologically active Ctx peptides on the laid egg surface suggests at least a function of Ctxs in protecting embryos and early larvae from environmental bacteria (Marchini *et al*, 1997, 2002).

A genomic clone containing four clustered genes encoding different members of the *Ctx* family has been previously isolated (Rosetto *et al*, 1997). One of the genes encoded a second form of Ctx A precursor protein, named *Ctx A2*. Two genes encoding *Ctx C* (*C1* and *C2*) were also found. The other gene in the cluster encoded a novel peptide of the same family, named *Ctx D*. The genes corresponding to the previously sequenced *Ctx A* and *B* cDNAs were not present in the isolated genomic clone. Indeed, genomic Southern blot analysis showed the presence of additional *Ctx* genes that may have included *A* and *B* mapping outside the sequenced cluster (Rosetto *et al*, 1997). The analysis of polytene and mitotic chromosomes by *in situ* hybridization showed that *Ctx* genes map on the X chromosome (Rosetto *et al*, 2000), the first report of a female-specific X-linked gene in *C. capitata*.

In this paper, we report the partial genomic sequences encoding two Ctx A and two Ctx B precursors. We also report the detection of Ctx-like peptides in the female accessory gland secretion of the Natal fruit fly *Ceratitis* (*Pterandrus*) *rosa* and the partial nucleotide sequences encoding four of them. Phylogenetic analysis has been performed to investigate the pattern of evolution of the *Ctx* genes from *C. capitata* and their evolutionary relations with the homologous genes from *C. rosa*.

Materials and methods

Insects

C. capitata flies were reared in standard laboratory conditions, at 23°C, 70% relative humidity and 14:10 light–dark regime (Rabossi *et al*, 1991). Adult specimens

Correspondence: R Dallai, Department of Evolutionary Biology, Via A. Moro 2, University of Siena, I-53100 Siena, Italy. E-mail: dallai@unisi.it
[3]Current address: Department of Environmental Sciences, University of Tuscia, Viterbo, Italy
Received 15 July 2002; accepted 17 January 2003

of *C. rosa* were reared from pupae collected from infested fruits of guava (*Psidium guajava* L.) in two localities in the south of Reunion Island (Bassin Martin, Ravine des Cabris, France), in the Indian Ocean. *Bactrocera (Dacus) oleae* pupae were collected from olive fruits in Puglia (southern Italy). *Rhagoletis pomonella* pupae were collected from *Crataegus mollis* in East Lansing (Michigan State). *Anastrepha fraterculus* pupae were collected from infested fruits of *Psidium guajava* in Louveira, Sao Paulo State, Brazil. *Anastrepha obliqua* pupae were collected from *Psidium guajava* in Bauru, Sao Paulo State, Brazil. Adults were maintained in the laboratory as reported above for *C. capitata*.

DNA isolation, sequencing and Southern blot analysis

Genomic DNA was extracted from adult flies according to the method described by Sambrook *et al* (1989). To isolate *Ctx* sequences, PCR reactions were performed using approximately 300 ng of genomic DNA, and degenerate primers encoding the amino- and carboxyl-terminal ends of Ctx precursor peptides: 5′-TTCAC-CATGGCMAAYMTTAAAGCT-3′ (primer 1) and 5′-MYWYWTTATCCTACAAGH-3′ (primer 2). *C. rosa Ctx* sequences were amplified using primer 1 and 5′-CAATGGGKAWGGCGAYCTTDSCAA-3′ (primer 3). Amplification reactions were performed as follows: five cycles of 94°C for 1 min, 50°C for 2 min, 72°C for 2 min, followed by 40 cycles of 94°C for 1 min, 35°C for 2 min and 72°C for 2 min. An extension at 72°C for 10 min was added at the end of the reaction. The amplification products were inserted into the Bluescript II SK (−) vector (Stratagene). Nucleotide sequences were determined by the MWG-BIOTECH automated DNA Sequencing Service (Ebersberg, Germany). Tubulin-specific primers were used in PCR reactions to control the intactness of genomic DNA. Southern blot analysis was performed with genomic DNA digested with *Eco* RI and *Eco* RV restriction enzymes, run on 1% agarose gel and then transferred to a Protran nitrocellulose transfer membrane (Schleicher & Schuell GmbH, Dassel, Germany). A *Ctx A* cDNA probe (Marchini *et al*, 1995) was ^{32}P-labeled by random priming using the Prime-a-gene kit (Promega). High-stringency hybridization was performed overnight at 65°C in 300 mM NaCl, 30 mM sodium citrate, pH 7.0 (2 × SSC), 1 × Denhardt's solution and 50 μg/ml salmon sperm DNA. The filter was washed twice in 2 × SSC, 0.1% SDS for 5 min at room temperature, and then three times at 65°C for 30 min in the same solution. Low-stringency hybridization and washes were performed at 50°C in 4 × SSC.

Nucleotide sequence analysis

Sequences were aligned with Clustal W (Thompson *et al*, 1994), and phylogenetic analyses were performed with PAUP* (Swofford, 1998), using MP, ME and ML. For ML analysis, an iterative search strategy was adopted (Swofford *et al*, 1996), using parsimony (equal weights) to select initial topologies and evaluating the likelihood score of these topologies under different models of evolution in order to select the model having the best fit (see Frati *et al*, 1997). Robustness of the nodes in all reconstructions was tested with bootstrap, while the statistical confidence of our 'best trees' compared to alternative topologies was tested with the Shimodaira–Hasegawa method (SH test: Shimodaira and Hasegawa, 1999) as implemented in PAUP*, using RELL (10 000 bootstrap replicates).

The possible occurrence of selection was tested by calculating the ratio between the number of nonsynonymous (d_N) and synonymous (d_S) substitutions using the Nei and Gojobori (1986) method as implemented in MEGA, version 2.0 (Kumar *et al*, 2001).

Recovery of female accessory gland secretion and protein assay

Accessory glands from sexually mature *C. capitata* and *C. rosa* females were dissected, homogenized in Eppendorf tubes with Eppendorf micropestles and centrifuged essentially as described in Marchini *et al* (1989) and Rosetto *et al* (1996). The supernatant, containing the secretion in 100 mM Na-phosphate buffer at pH 6.8 (PB) (1 gland equivalent/1–4 μl PB), was freeze-dried and stored at −20°C. The protein content of the accessory gland secretion was determined according to Bradford (1976) using BSA as a standard.

Assay for antibacterial activity

Escherichia coli LE 392, cultured in Luria–Bertani medium (Sambrook *et al*, 1989), was used as a test organism to assay antibacterial activity. Inhibition zone assay was carried out as previously reported (Faye and Wyatt, 1980; Marchini *et al*, 1997), using a chemically synthesized Ctx A 1-36 (Dompé S.p.A., Milano, Italy) as a control of antibacterial activity.

Electrophoresis and Western blot analysis

SDS–PAGE was performed according to Laemmli (1970). Low molecular weight markers were purchased from Promega. Chemically synthesized Ctx A 1-36 and Ctx C 1-32 (Dompé S.p.A., Milano, Italy) were also used as markers. After electrophoresis, proteins were transferred onto a nitrocellulose filter (Bioblot-NC, Costar) as described by Towbin *et al* (1979). Ctxs were detected as previously reported (Marchini *et al*, 1995): filters were soaked for 30 min in 16 mM PB, pH 7.4, 150 mM NaCl containing 3% BSA and 0.1% Triton X-100 (PBSAT) and incubated overnight with an anti-Ctx serum (Marchini *et al*, 1995), diluted 1:200 in PBSAT. The second antibody (goat anti-rabbit IgG, 1:1000 dilution, horseradish peroxidase conjugated, Cappel) was applied after rinsing of the filters with PBSAT. The color reaction was developed by 4-chloro-1-naphthol (Sigma) in 50 mM Tris-HCl, pH 6.8 and stopped with distilled water.

Results

Isolation of novel Ctx sequences in C. capitata

A PCR-based strategy was used to isolate genomic sequences related to the *Ctx* genes previously identified (Marchini *et al*, 1995; Rosetto *et al*, 1996, 1997). The entire *Ctx* coding region, with the exception of the nucleotide sequences encoding six amino-terminal and four carboxyl-terminal amino-acid residues, was amplified using a combination of the primers described in the section 'DNA isolation, sequencing and Southern blot analysis.' PCR products were then subcloned to obtain their nucleotide sequences (Figure 1). A sequence containing a coding region identical to the previously isolated

```
              signal peptide    pro-region            mature peptide
        _______________________            ____________________________________

Ctx A1  MANLKAVFLICIVAFIALQCVVAEPAAEDSVVVKRSIGSALKKALPVAKKIGKIALPIAKAALPVAAGLVG
Ctx A2                   F            I       SA     P
Ctx A3  ------           FH           I      GSA     P A                   ----
Ctx B1  ------           F       T     M  R     F            A L     A   V ----
Ctx B2  ------       AF  FH      T     M  R     F            A L     A   V ----
Ctx C1     I    L    A   FHC     T  D    K  L GVI****SGA KVA V I  G AV   V K
Ctx C2    NI    L    A   FHC     T  D    K  L GVI****SGA KVA V I  G  V   V K
Ctx D    A      L  ILA  AFHC  GAPT    I       TAV  AV IA  VG V I  A  VLS VGQ
CrCtx 1 ------I      A   FTC     T          L  AIR AV IA  ---------------------
CrCtx 2 ------I F SL A   FHC   APT          LAGVI****SGA K---------------------
CrCtx 3 ------I     IA   FTC     T      I  GL  TTRK----------------------------
CrCtx 4 ------IIL    AL  FHC   A T  DS     S   AIR AIPIA  ---------------------
                          5
```

Figure 1 Comparison of Ctx deduced translational products. Ctx A1, A2, A3, B1, B2, C1, C2, D: ceratotoxins from *C. capitata*; CrCtx 1, 2, 3, 4: ceratotoxins from *C. rosa*. Amino-acids differing from upper Ctx A1 sequence are shown. Underlined amino-acid residues indicate silent nucleotide substitutions. Dashes indicate nonsequenced regions. Asterisks indicate amino-acid residue deletions. Hypothetical signal peptide and mature peptide sequences are indicated.

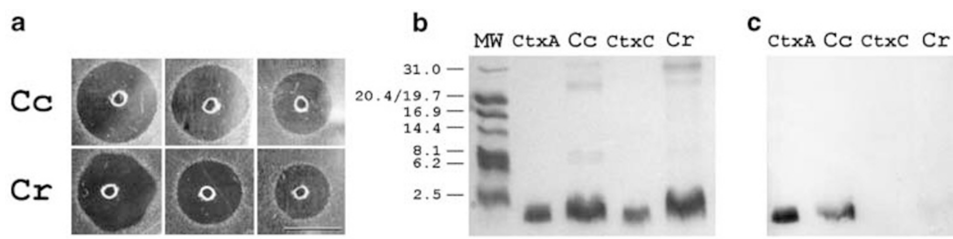

Figure 2 Antibacterial, SDS–PAGE and Western blot analysis of accessory gland secretion from *C. capitata* (Cc) and *C. rosa* (Cr). (**a**) Anti-*E. coli* activity as determined by inhibition zone assay. Photographs show, from left to right in both lines, the inhibition growth haloes produced by samples containing 2.11, 1.05 or 0.53 μg total protein, respectively. Bar corresponds to 10 mm. (**b**) Protein staining (Ponceau) and (**c**) immunoblot of the same filter using an anti-Ctx serum, after SDS–PAGE. The gel (20% polyacrylamide) was loaded with samples of 2.5 accessory gland equivalent, corresponding to 2 and 3.5 μg total protein from *C. capitata* and *C. rosa*, respectively. Chemically synthesized Ctx A and Ctx C: 2.8 μg, respectively. MW: low molecular weight markers, expressed in kDa, 2 μg/each protein.

cDNA clone encoding Ctx A (Marchini *et al*, 1995) was determined, and was named *Ctx A1* (accession number AJ272446). In addition, the partial sequence of a novel gene encoding another Ctx A precursor was found, sharing 92.7% of sequence identity with *Ctx A1* nucleotide sequence. It was named *Ctx A3* (accession number AJ272447), since the sequence of a second form of *Ctx A* (named *Ctx A2*) had been previously determined (Rosetto *et al*, 1997). Two sequences were found to encode Ctx B precursors. One of them matched the sequence of a truncated cDNA clone encoding Ctx B we previously isolated (Marchini *et al*, 1995). This novel sequence was named *Ctx B1* (accession number AJ272448). The other *Ctx B* sequence showed 91.6% sequence identity with *Ctx B1*. For this reason it was named *Ctx B2* (accession number AJ272449).

Search for the presence of Ctx-like genes in other dipteran species

The presence of *Ctx*-like genes in other dipteran species was investigated. First, we studied the Natal fruit fly *C. rosa*, another major agricultural pest closely related to *C. capitata* (Kourti *et al*, 1992). This species is found in many countries of eastern and southern Africa, and in some islands of the Indian Ocean (White and Elson-Harris, 1992).

As a preliminary step, we assayed the female accessory gland secretion from *C. rosa* against *E. coli*, since Ctx peptides are greatly responsible for the antibacterial activity of the accessory gland secretion from *C. capitata* (Marchini *et al*, 1993, 1997; Marri *et al*, 1996; Rosetto *et al*, 1996). As shown in Figure 2a, the inhibition growth haloes produced by the secretion from *C. rosa* are in the same range as those from *C. capitata*.

Thus, we searched for the presence of Ctx peptides by performing SDS–PAGE and Western blot analyses of the accessory glands of the Natal fruit fly in comparison with an equivalent material from the medfly (Figure 2b, c). A major band of 3 kDa was present in the accessory glands of *C. rosa*, compatible with the molecular mass of Ctxs from *C. capitata* (Figure 2b). In addition, the anti-Ctx A 1-29 serum, which strongly recognized the 3 kDa band of *C. capitata* and the synthetic Ctx A 1-29 and Ctx A 1-36 (Marchini *et al*, 1995, 1997; Figure 2c) but not Ctx C 1-32 (Figure 2c), gave a slight positive reaction with the correspondent protein band from *C. rosa* (Figure 2c).

To confirm the presence of sequences encoding Ctx-like peptides in the genome of *C. rosa*, genomic DNA extracted from a single fly was used as a template in PCR reactions with degenerate primers derived from *C. capitata Ctx* sequences. The resulting products were then subcloned and four different *Ctx*-like nucleotide sequences were determined (Figure 1): *CrCtx1* (accession number AJ272450), *CrCtx 2* (accession number AJ313430), *CrCtx 3* (accession number AJ313431) and *CrCtx 4* (accession number AJ313432).

To investigate the presence of *Ctx*-like sequences in the genome of other related species, a Southern blot analysis was performed; other members from the Tephritidae

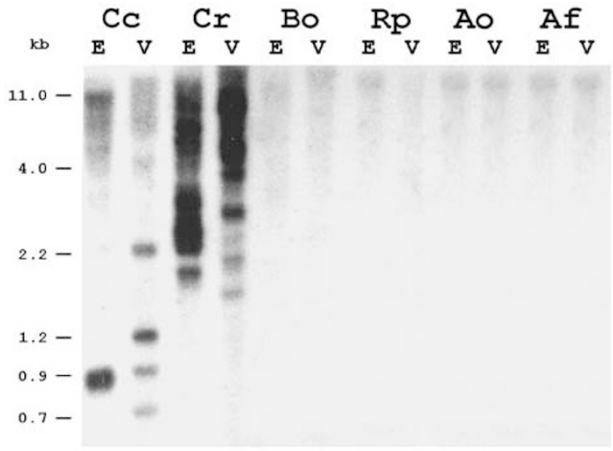

Figure 3 Low-stringency Southern blot analysis of *Ctx*-like sequences in the genomes of *C. capitata* (Cc), *C. rosa* (Cr), *B. oleae* (Bo), *R. pomonella* (Rp), *A. obliqua* (Ao) and *A. fraterculus* (Af). Samples were digested with *Eco* RI (E) or *Eco* RV (V) restriction enzymes. Molecular sizes of the bands are expressed in kilobases (kb).

family were analysed in addition to *C. capitata* and *C. rosa*: the olive fly *B. oleae*, the apple maggot fly *R. pomonella*, the West Indian fruit fly *A. obliqua* and the South American fruit fly *A. fraterculus*. Genomic DNAs were digested with *Eco* RI and *Eco* RV restriction enzymes, and then hybridized with a *Ctx A* cDNA probe. In addition to the expected band pattern in *C. capitata* lanes (Rosetto *et al*, 1996) several bands were detected in *C. rosa* lanes, while no band was detectable in any other species, even at lower stringency (Figure 3). PCR analysis performed using genomic DNA from the same species gave a clear band with control primers but did not allow the amplification of any *Ctx*-like products, thus supporting Southern blot data (data not shown).

Comparison of Ctx genes

All *Ctx* genes share remarkable sequence similarity (Figure 1), especially in their hypothetical signal peptides, as previously reported (Rosetto *et al*, 1996, 1997). *C. capitata Ctx A1* and *Ctx A3* precursors differ from each other for three conservative substitutions in their pre-pro-regions. In addition, the nucleotide sequences encoding their mature peptides are different in five positions, none of which changes the amino-acid sequence. The coding regions of *Ctx B1* and *Ctx B2* PCR fragments are different in three positions, only one of which causes a conservative amino-acid change in the signal peptide. The hypothetical translational products of *Ctxs* from *C. rosa* share a high degree of sequence similarity with *Ctxs* from *C. capitata*, especially in their pre-pro-regions. Nevertheless, they contain some substitutions in residues that are highly conserved in *C. capitata Ctx* precursors. For instance, an arginine residue is present at position 7 in the presumptive mature peptides of CrCtx 1, CrCtx 3, CrCtx 4 from *C. rosa*, where a lysine is present in all Ctxs from *C. capitata* (Figure 1). The hypothetical translational products of *C. rosa CrCtx 1* and *CrCtx 3* display the highest sequence similarity with *Ctx B2* from *C. capitata* (81 and 70%, respectively). Another *C. rosa* precursor (*CrCtx 2*) shares high sequence similarity with *Ctx C* from *C. capitata* (85%), and there is an internal gap of four residues, like in *C. capitata Ctx C* genes, as compared to other *Ctx* genes (Rosetto *et al*, 1996). Finally, *CrCtx 4* precursor shows considerable sequence conservation not only with *C. capitata Ctx B2* (79% similarity) but also with *Ctx D* (77% similarity).

All the sequenced *Ctx* genes share the same exon–intron–exon structure. The introns (59–79 nucleotides in *C. capitata*, 58–90 nucleotides in *C. rosa*) are located at a conserved position, after the codon encoding the residue at position 19 (glutamine in Ctx A1, Ctx A2 and Ctx B2, histidine in Ctx A3, Ctx B1, Ctx C1, Ctx C2, Ctx D and CrCtx 2 and CrCtx 4, threonine in CrCtx 1 and CrCtx 3), in the region encoding the presumptive signal peptide. The introns of *Ctx A1*, *Ctx A2*, *Ctx A3* and *Ctx B1* genes from *C. capitata* and *CrCtx 1* and *CrCtx 3* from *C. rosa* share a high degree of sequence conservation. The intron of *C. rosa CrCtx 2* shares high sequence similarity with the introns of *Ctxs C* from *C. capitata* (88% identity). All *Ctx* introns of *C. capitata* and *C. rosa* contain the GTAAGT-5PyNCAG motif, in accordance with the data by Breathnach and Chambon (1981) on eukaryotic splice junctions.

Phylogenetic analysis of Ctx genes

Evolutionary analysis of *Ctx* genes has been performed separately on the coding and the intronic regions, and on the amino-acid sequences. In all analyses, positions experiencing indels were removed, as it is difficult to establish ancestral states in those sites. The absence of *Ctx* genes in closely related species did not allow us to use outgroups, and all the trees obtained were, therefore, unrooted.

In the coding region of the genes of *C. capitata* and *C. rosa*, almost a half of nucleotide positions are variable (Table 1). Variability is concentrated in third codon positions, which account for slightly less than 50% of variable sites. First codon positions have twice as much variable sites as second positions.

The phylogenetic analysis was performed on nucleotide sequences using the maximum likelihood method with the GTR+Γ model of evolution (Yang, 1994; Gu *et al*, 1995). This model was selected by the likelihood-ratio test (Yang *et al*, 1995) as the one having the best fit with the data, and the analysis was performed by optimizing model parameters during the search. The resulting tree, shown in Figure 4, has the same topology as the tree obtained with MP and with ME on GTR+Γ-corrected distances (Table 2), and it supports the grouping of *CrCtx 2* of *C. rosa* with *Ctxs C* of *C. capitata*. It also shows the affinity between *Ctx A* and *Ctx B* of *C. capitata* and *CrCtx 1* and *CrCtx 3* of *C. rosa*. According to this tree, the distinction of the four *Ctx*

Table 1 Distribution of variable sites in the Ctx sequences

	% of variable sites with respect to total sites	*% of variable sites in each position with respect to total variable sites*
All sites	42.9	
1° pos.	45.7	35.6
2° pos.	22.9	17.8
3° pos.	60.0	46.7

Table 2 Genetic divergence estimates across *Ctx* genes in *C. capitata* and *C. rosa*. In the upper-right triangle, GTR-Γ-corrected genetic distances are indicated. In the lower-left triangle, divergence of the amino-acid sequence is reported. All estimates have been calculated by removing positions experiencing gaps

	C. rosa				*C. capitata*							
	1	*2*	*3*	*4*	*A1*	*A2*	*A3*	*B1*	*B2*	*C1*	*C2*	*D*
1	–	0.179	0.060	0.138	0.168	0.180	0.200	0.152	0.153	0.133	0.133	0.257
2	0.229	–	0.235	0.218	0.315	0.304	0.332	0.303	0.305	0.172	0.172	0.329
3	0.143	0.343	–	0.190	0.223	0.236	0.259	0.205	0.207	0.184	0.184	0.303
4	0.143	0.257	0.286	–	0.191	0.203	0.224	0.174	0.166	0.168	0.168	0.236
A1	0.171	0.371	0.286	0.229	–	0.040	0.064	0.051	0.086	0.180	0.180	0.245
A2	0.171	0.371	0.286	0.229	0.057	–	0.020	0.072	0.109	0.178	0.178	0.213
A3	0.171	0.343	0.286	0.200	0.086	0.029	–	0.099	0.112	0.198	0.198	0.218
B1	0.143	0.343	0.257	0.200	0.114	0.114	0.143	–	0.030	0.163	0.163	0.226
B2	0.143	0.314	0.257	0.171	0.143	0.143	0.114	0.029	–	0.164	0.164	0.210
C1	0.114	0.171	0.228	0.200	0.200	0.200	0.171	0.171	0.143	–	0.000	0.270
C2	0.114	0.171	0.228	0.200	0.200	0.200	0.171	0.171	0.143	0.000	–	0.270
D	0.257	0.343	0.342	0.229	0.257	0.200	0.171	0.229	0.200	0.229	0.229	–

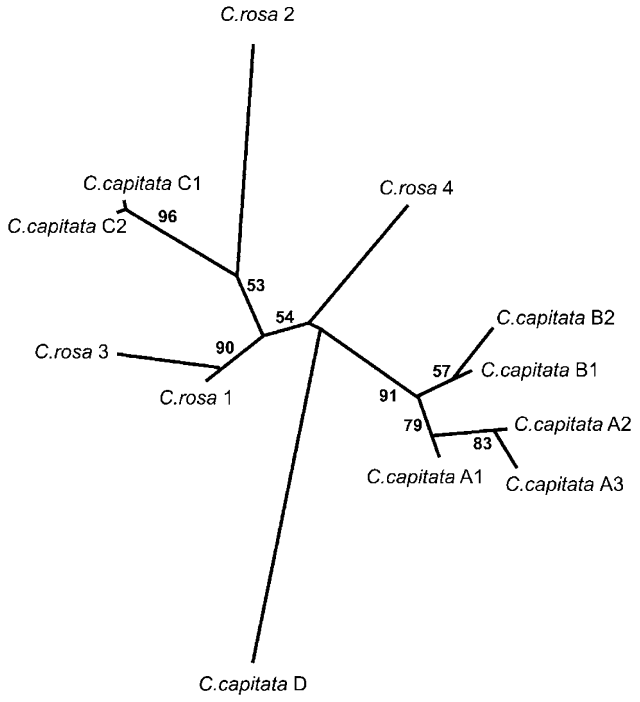

Figure 4 ML tree from nucleotide sequences of the coding region obtained with the GTR+ Γ model (−ln likelihood = 470.16174; α = 1.144136). Branch lengths are ML estimates. Bootstrap values (100 replicates with the parameters estimated during the original search) are shown at the nodes.

genes of *C. capitata* (A, B, C and D) is correct, although little bootstrap support was found for the groupings of *Ctxs B*. Finally, *Ctx A* and *Ctx B* of *C. capitata* are closely related. However, the SH test does not allow one to statistically reject ($P = 0.73$) the best tree where *Ctx* genes from the same species form monophyletic groupings, which has a likelihood score 3.42 higher than our best tree of Figure 4.

The parsimony-based phylogenetic analysis of amino-acid sequences produced a single most parsimonious tree (33 steps; CI = 0.8485; not shown), showing some relation of Ctx D with Ctx A and B.

In order to test for the presence of selection, we estimated the d_N/d_S ratio for all possible pairwise comparisons. The average d_N/d_S ratio was 0.457 ± 0.284, and in only four of the 66 pairwise comparisons this ratio was >1. Three of these four comparisons involved *CrCtx 3* of *C. rosa*, including its comparison with *CrCtx 1* of *C. rosa*, with which it is clustered in the tree of Figure 4. Only one comparison with $d_N/d_S > 1$ involved two *Ctx* sequences from *C. capitata* (A1 and B1).

We also tested the performance of the intronic sequences to reconstruct evolutionary relations among these sequences. The ML tree obtained with the HKY85+Γ model (Hasegawa *et al*, 1985) using these sequences is shown in Figure 5, and it differs from the tree of Figure 4 in that it placed *Ctx D* of *C. capitata* together with *Ctxs C* and *CrCtx 2* of *C. rosa*. This tree groups all the remaining *Ctxs* of *C. rosa* together. Also in this case, the SH test does not reject ($P = 0.39$) the best tree where *Ctx* genes from the same species form monophyletic clades, which has a likelihood score 8.85 higher than our best tree of Figure 5.

Discussion

To date, eight distinct *Ctx* sequences have been identified in *C. capitata* and four in *C. rosa*, either by the isolation of cDNA and genomic clones, or by amplification of genomic DNA fragments. Based on the amino-acid sequences of the purified mature peptides, Ctxs are likely to be processed to yield the mature products by removal of a signal peptide at the amino-terminal end followed by the subsequent removal of a pro-sequence (Marchini *et al*, 1995; Rosetto *et al*, 1996, 1997). Assuming that the same maturation events occur in all the Ctx peptides, only four different mature peptides are produced in *C. capitata* (Ctx A, Ctx B, Ctx C, Ctx D), three of which (A, B and C) have been isolated from the medfly accessory gland secretion (Marchini *et al*, 1993; Rosetto *et al*, 1996).

In *C. capitata*, where PCR was performed on genomic DNA from a pool of individuals, we cannot rule out the hypothesis that some of the sequences represent allelic variants, especially the most similar ones within the same group (*Ctx A2* and *A3*, *Ctx B1* and *B2*). However,

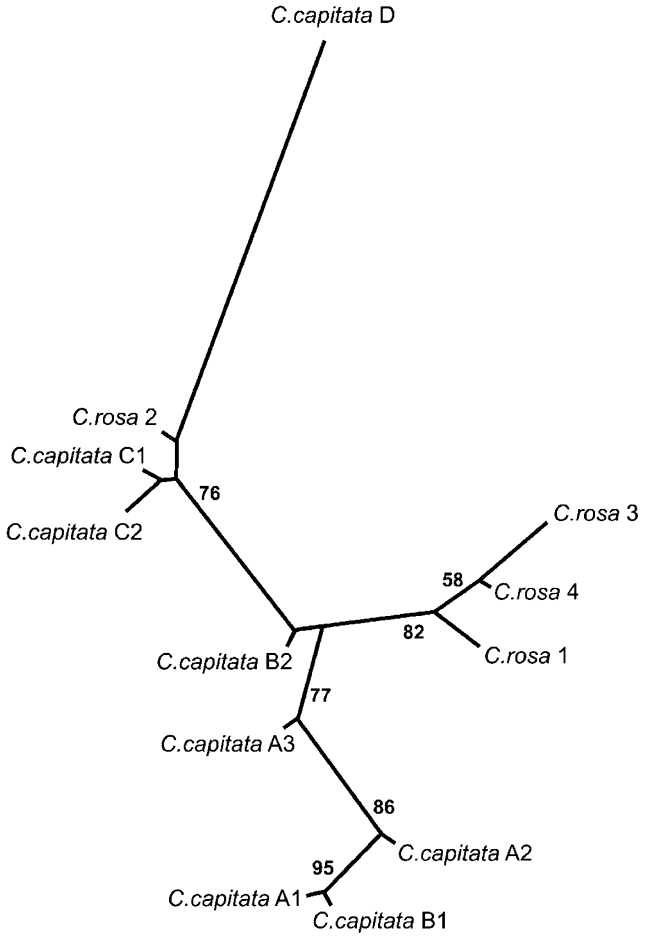

Figure 5 ML tree from nucleotide sequences of the intron obtained with the HKY+Γ model (−ln likelihood = 278.75892; $\alpha = 1.092968$). Branch lengths are ML estimates. Bootstrap values (100 replicates with the parameters estimated during the original search) are shown at the nodes.

although balanced polymorphism may occur, the insects used throughout the experiments, which derived from an inbred population reared in our laboratory for 20 years, should have a very low average heterozygosity, due to the effects of genetic drift in small populations. Moreover, within the Ctx C group, the forms *C1* and *C2* surely represent two separate genes, as they were previously found in the same gene cluster (Rosetto *et al*, 1997). Interestingly, these are also the most similar forms of *Ctx* (Table 2) and the presence of much higher levels of sequence divergence between all other sequences of *Ctx* seems to argue for the fact that they represent different genes rather than allelic variants. The occurrence of allelic variants is even less likely in *C. rosa*, where PCR amplifications were performed on genomic DNA extracted from a single fly.

The clustered organization of *Ctx* genes suggests that they presumably arose as a result of gene duplication events. Multigene families, most likely arisen from gene duplications, are also characteristic of other antibacterial peptides from insects, such as cecropins, attacins, apidaecins (Boman *et al*, 1991; Casteels-Josson *et al*, 1993), and from vertebrates, such as the amphibian caerulein precursor fragment (Richter *et al*, 1986; Moore *et al*, 1991). From an evolutionary point of view, the tree obtained by comparing nucleotide sequences of the coding region (Figure 4) shows that the genes encoding for the same mature peptide (A, B or C) constitute monophyletic groups. This may suggest that the duplication process first led to the differentiation of the four groups (A, B, C and D) and later to the differentiation of each gene within the same group. However, the analysis of the intron sequences places *Ctx B1* within all *Ctx A* sequences (Figure 5), suggesting caution in considering groups A and B as clearly separate. Since introns are more likely to be evolutionary neutral, they may even be providing a more accurate idea of the evolutionary relations. If the rate of evolution is proportional to the divergence time, it can be hypothesized that the most recent duplication has involved the sequences C1 and C2, which show the smallest genetic distances. Interestingly, the intronic sequences of these two genes are very similar, supporting the hypothesis that they have diverged very recently.

Before a phylogenetic tree such as the one in Figure 4 could be considered indicative of the historical pattern, one should consider the possibility of selection, which might be acting here given the antibacterial activity of these peptides. Among the various tests developed to detect the presence of adaptive selection (Yang and Bielawski, 2000), some take into account the number of synonymous and nonsynonymous nucleotide changes, such as in the relatively conserved method of Nei and Gojobori (1986). The test is performed under the assumption that if positive selection is occurring, the d_N/d_S ratio should be >1, while the ratio should be <1 if only purifying selection is acting as it will constrain nonsynonymous sites to a much greater extent than the nearly neutral synonymous sites. This test applied to *Ctx* sequences appears to reject the hypothesis of the occurrence of directional selection. It could be concluded that selection does not significantly influence the evolution of *Ctx* genes in *C. capitata* and *C. rosa*, whose relations should therefore be indicative of a historical pattern.

The picture emerging from Figure 4 seems to suggest that at least some steps of the diversification of *Ctx* genes in the genus *Ceratitis* must have occurred before the split between *C. capitata* and *C. rosa*. In fact, *Ctxs C* of *C. capitata* are grouped with *CrCtx 2* of *C. rosa*, and not with all other *Ctx* genes of *C. capitata* like it should be expected if both species had inherited only one *Ctx* gene from their common ancestor and had independently diversified different *Ctx* genes by duplication events. While the trees presenting *Ctxs* from the same species clustered together are not rejected by rigorous statistical tests, the fact that in both cases the best tree supports the diversification of at least some *Ctx* genes prior to the diversification of the two species points toward our hypothesis. To confirm this, sequence divergence values (Table 2) strongly suggest higher affinity of *CrCtxs 2* with *Ctxs C* of *C. capitata*. In addition, the divergence levels in the coding regions are similar to those estimated for other orthologous genes between *C. capitata* and *C. rosa* (Villablanca *et al*, 1998). Similar duplication events have been recorded for other genes in *C. capitata* and other tephritids, such as the *Adh* gene (Brogna *et al*, 2001). The relations between *CrCtx 2* of *C. rosa* and *Ctxs C* of *C. capitata* are evident even from the analysis of intron sequences (Figure 5), which are even less influenced by

selection (but see Villablanca *et al* (1998) for an example of constraints on substitutions in introns). The absence of *Ctx* genes in other members of the genus where they have been looked for does not give us the possibility to root the tree and to understand the complete historical pattern of these genes. However, it should be noted that anywhere the root is placed in the tree of Figure 4, there is never the possibility of observing two separate clusters dividing the *Ctx* genes from each of the two species. Subsequent duplication events must also have occurred within each species. Since *Ctx A* and *Ctx B* are very similar to each other, it can be suggested that these genes have evolved from a common ancestor, which, in turn, had already differentiated from the other genes. During its evolution, the *ceratotoxin* gene family in *C. capitata* and *C. rosa* might therefore have experienced several duplication events, prior to and following the split between the two species.

Although we cannot exclude the presence of peptides with biological properties comparable to those of *Ctxs* even in other distantly related dipteran species, the presence of Ctx-like peptides in the secretion of the female reproductive accessory glands is presumably restricted to a few closely related species. In fact, Southern blot and PCR analyses showed the absence of *Ctx*-like genes in the genomes of other tephritid flies. Therefore, the evolution of *Ctx* genes might have occurred in relatively recent times, possibly right before the differentiation of the *Ceratitis* genus, or even only in some of its species. In spite of this, antimicrobial peptides sharing considerable sequence similarity with Ctxs were found in vertebrates. In fact, Ctx mature peptides share a remarkable sequence similarity (55.5%) with a caerulein precursor fragment peptide (CPF) from *Xenopus laevis* (Moore *et al*, 1991). In addition, the hypothetical translational product of *Ctx A* cDNA shows considerable amino-acid sequence similarity (70%) with dermaseptin, another antimicrobial peptide secreted by the skin of the amphibian *Phyllomedusa sauvagei* (Mor *et al*, 1991). Moreover, Ctx B mature peptide shows 68% identity with the antimicrobial peptide pleurocidin, purified from the skin mucous secretion of the fish *Pleuronectes americanus* (Cole *et al*, 1997). The similarity between all these peptides also extends to their predicted α-helix secondary structure.

Acknowledgements

This work was supported by Mi.P.A. and M.I.U.R. to RD and the University of Siena (P.A.R.) to DM and FF. The authors are indebted to G Nuzzaci (University of Bari, Italy) for supplying *B. oleae* pupae, J Smith and V Gavrilovic (Michigan State University) for *R. pomonella* pupae, and P Dos Santos (University of California, Berkeley) for *Anastrepha* specimens.

References

Boman HG, Faye I, Gudmundsson GH, Lee J-Y, Lidholm D-A (1991). Cell-free immunity in *Cecropia*. A model system for antibacterial proteins. *Eur J Biochem* **201**: 23–31.

Bradford MM (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding. *Anal Biochem* **72**: 248–254.

Breathnach R, Chambon P (1981). Organization and expression of eukaryotic split genes coding for proteins. *Annu Rev Microbiol* **41**: 103–126.

Brogna S, Benos PV, Gasperi G, Savakis C (2001). The *Drosophila* alcohol dehydrogenase gene may have evolved independently of the functional homologous medfly, olive fly, and flesh fly genes. *Mol Biol Evol* **18**: 322–329.

Casteels-Josson K, Capaci T, Casteels P, Tempst P (1993). Apidaecin multipeptide precursor structure – a putative mechanism for amplification of the insect antibacterial response. *EMBO J* **12**: 1569–1578.

Cole AM, Weis P, Diamond G (1997). Isolation and characterization of pleurocidin, an antimicrobial peptide in the skin secretions of winter flounder. *J Biol Chem* **272**: 12008–12013.

Faye I, Wyatt GR (1980). The synthesis of antibacterial proteins in isolated fat body from *Cecropia* silkmoth pupae. *Experientia (Berlin)* **36**: 1325–1326.

Frati F, Simon C, Sullivan J, Swofford DL (1997). Evolution of the mitochondrial cytochrome oxidase II gene in Collembola. *J Mol Evol* **44**: 145–158.

Gu X, Fu Y, Li W (1995). Maximum likelihood estimation of the heterogeneity of substitution rate among nucleotide sites. *Mol Biol Evol* **12**: 546–557.

Hasegawa M, Kishino M, Yano T (1985). Dating the human–ape split by a molecular clock of mitochondrial DNA. *J Mol Evol* **22**: 160–174.

Kourti A, Loukas M, Sourdis J (1992). Dispersion pattern of the medfly from its geographic centre of origin and genetic relationships of the medfly with two close relatives. *Entomol Exp Appl* **63**: 63–69.

Kumar S, Tamura K, Jakobsen IB, Nei M (2001). *MEGA2: Molecular Evolutionary Genetics Analysis* software. *Bioinformatics* **17**: 1244–1245.

Laemmli UK (1970). Cleavage and structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680–685.

Manetti AGO, Rosetto M, de Filippis T, Marchini D, Baldari CT, Dallai R (1997). Juvenile hormone regulates the expression of the gene encoding ceratotoxin A, an antibacterial peptide from the female reproductive accessory glands of the medfly *Ceratitis capitata*. *J Insect Physiol* **43**: 1161–1167.

Marchini D, Bernini LF, Dallai R (1989). β-*N*-acetylhexosaminidases in the secretion of the female reproductive accessory glands of *Ceratitis capitata* (Diptera). *Insect Biochem* **19**: 549–555.

Marchini D, Bernini LF, Marri L, Giordano PC, Dallai R (1991). The female reproductive accessory glands of the medfly *Ceratitis capitata*: antibacterial activity of the secretion fluid. *Insect Biochem* **21**: 597–605.

Marchini D, Giordano PC, Amons R, Bernini LF, Dallai R (1993). Purification and primary structure of ceratotoxin A and B, two antibacterial peptides from the female reproductive accessory glands of the medfly *Ceratitis capitata* (Insecta: Diptera). *Insect Biochem Mol Biol* **5**: 591–598.

Marchini D, Manetti AGO, Rosetto M, Bernini LF, Telford JL, Baldari CT *et al* (1995). cDNA sequence and expression of the ceratotoxin gene encoding an antibacterial sex-specific peptide from the medfly *Ceratitis capitata* (Diptera). *J Biol Chem* **270**: 6199–6204.

Marchini D, Marri L, Rosetto M, Manetti AGO, Dallai R (1997). Presence of antibacterial peptides on the laid egg chorion of the medfly *Ceratitis capitata*. *Biochem Biophys Res Commun* **240**: 657–663.

Marchini D, Rosetto M, Dallai R, Marri L (2002). Bacteria associated with the oesophageal bulb of the medfly *Ceratitis capitata* (Diptera: Tephritidae). *Curr Microbiol* **44**: 120–124.

Marri L, Dallai R, Marchini D (1996). The novel antibacterial peptide ceratotoxin A alters permeability of the inner and outer membrane of *Escherichia coli* K-12. *Curr Microbiol* **33**: 40–43.

Moore KS, Bevins CL, Brasseur MM, Tomassini N, Turner K, Eck H *et al* (1991). Antimaicrobial peptides in the stomach of *Xenopus laevis*. *J Biol Chem* **266**: 19851–19857.

Mor A, Nguyen VH, Delfour A, Migliore-Samour D, Nicolas P (1991). Isolation, amino-acid sequence, and synthesis of dermaseptin, a novel antimicrobial peptide of amphibian skin. *Biochemistry* **30**: 8824–8830.

Nei M, Gojobori T (1986). Simple methods for estimating the number of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* **3**: 418–426.

Rabossi A, Boccaccio GL, Wappner P, Quesada-Allué LA (1991). Morphogenesis and cuticular markers during the larval–pupal transformation of the medfly *Ceratitis capitata*. *Entomol Exp Appl* **60**: 135–141.

Richter K, Egger R, Kreil G (1986). Sequence of preprocaerulein cDNAs cloned from skin of *Xenopus laevis*. *J Biol Chem* **261**: 3676–3680.

Rosetto M, de Filippis T, Mandrioli M, Zacharopoulou A, Gourzi P, Manetti AGO *et al* (2000). Ceratotoxins, female-specific X-linked genes from the medfly *Ceratitis capitata*. *Genome* **43**: 707–711.

Rosetto M, de Filippis T, Manetti AGO, Marchini D, Baldari CT, Dallai R (1997). The genes encoding the antibacterial sex-specific peptides ceratotoxins are clustered in the genome of the medfly *Ceratitis capitata*. *Insect Biochem Molec Biol* **27**: 1039–1046.

Rosetto M, Manetti AGO, Giordano PC, Marri L, Amons R, Baldari CT *et al* (1996). Molecular characterization of ceratotoxin C, a novel antibacterial female-specific peptide of the ceratotoxin family from the medfly *Ceratitis capitata*. *Eur J Biochem* **241**: 330–337.

Sambrook J, Fritsch EF, Maniatis T (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY.

Shimodaira H, Hasegawa M (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol Biol Evol* **16**: 1114–1116.

Swofford DL (1998). *PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods)*. Version 4.0. Sinauer Associates: Sunderland, MA.

Swofford DL, Olsen GJ, Waddell PJ, Hillis DM (1996). Phylogenetic inference. In: Hillis DM, Moritz C, Mable BK (eds) *Molecular Systematics*, 2nd edn. Sinauer: Sunderland MA. pp 407–514.

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

Towbin H, Stahelin T, Gordon J (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc Natl Acad Sci USA* **76**: 4350–4354.

Villablanca FX, Roderick GK, Palumbi SR (1998). Invasion genetics of the Mediterranean fruit fly: variation in multiple nuclear introns. *Mol Ecol* **7**: 547–560.

White IM, Elson-Harris MM (1992). *Fruit flies of economic significance: their identification and bionomics*. CAB International/ACIAR: Wallingford.

Yang Z (1994). Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: aproximate methods. *J Mol Evol* **39**: 306–314.

Yang Z, Bielawski JP (2000). Statistical methods for detecting molecular adaptation. *TREE* **15**: 496–503.

Yang Z, Goldman N, Friday A (1995). Maximum likelihood trees from DNA sequences: a peculiar statistical estimation problem. *Syst Biol* **44**: 384–399.