engrailed-family gene RT-PCR primer sequences:

Oncopeltus fasciatus EH1 (forward non-degenerate) 5´-GCGATCCTCAGCCCATCCTT-3´;
Forward *invected*-specific domain 5´-GGATTCGCTGYTNWSNGTNGG-3´;
Forward EH2 5´-GGCCCGCCTGGGTNTAYTG-3´;
Reverse EH2 5´-CGGAGTAGCGGGTGCARWANAYCCA-3´;
Reverse EH5 (inner) 5´-GTGGAGTGGTTGTACAGGCCYTGNGCCAT-3´;
Reverse EH5 (outer) 5´-CGGTCATGGGGACGGTGSWRTGRTTRTA-3´

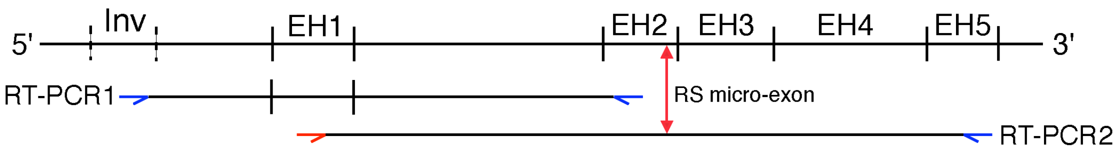
RT-PCR was performed using a touchdown program:
Annealing; 65°C – 55°C dropping 0.5°C each cycle for 20 cycles, then 20 cycles at 55°C.

The *invected*-specific domain and the cloning of *O. fasciatus* 5' sequence

(a)

D. melanogaster	ENSVLSVGQE
A. gambiae	DD-E------
B. mori	DDT-----N-
A. mellifera	TSE-----S-
T. castaneum	DDT-----N-
S. gregaria	AD-L----S-
P. americana	-DEL----S-

(b)



(a) Amino acid alignment of the Invected-specific domain. (b) Using degenerate primers (blue) targeted against EH2 and sequence encoding the Invected-specific domain, we amplified 5´ sequence from an *O. fasciatus engrailed*-family gene (RT-PCR1). To identify which *O. fasciatus* gene this sequence was from, we designed a gene specific primer (red) targeting sequence within EH1, and used it in combination with an EH5 degenerate reverse primer to amplify 3´ sequence (RT-PCR2). Only clones encoding the RS motif were recovered, indicating that the 5´ sequence originates from *O.f.en-r2* (Peterson et al, 1998).

```
CGGCTTACAAGTCAT TTCCAGGCGCCGTCC TGCGTCGCGGCGCCT CCACTCGTCCCAGAC TTCTTTACATTTCCC CGCACGAGGCAGTTT ACTCCGGGTGCTCCT
ATGGTTTACGATGTT CAGTCGGTACTCGAC CGGCCACTTGTACGC GCCTGACATTTTTTT TCGGAACAGAGAATA ACTCGTTTAAGGGAA GGACGCTAGTGCGTG
AGGGAGTAAAGAAGT GGAGGTCTCGGGCTG TCGGATGAGAATGCC ACAAATTGTGCTCTC GCCCAGTTTTTCTGC AACCACTAGCTGAAT AACTGCAAATTGACT
GGGCTCAGCTGTGGT GGCAGACGGGTCGAG AAATAAAAGAGAAAA CGATGGCTCGAGTCT GGGCGCGGCTGTAGC GAGGGGCGGAAGTGC AGTGCTAGGACTGTC GTG
```

```
ATG GCA CTG GAC ACG GAG GGC CGC GGC GAG GCG GCG GGG CTG CCC TTC TCG GTG GCC
 M   A   L   D   T   E   G   R   G   E   A   A   G   L   P   F   S   V   A
AAC ATC CTG CGG CCC GAC TTC GGC CGC CGC TGC GTG CGC CTC TTC CGG CCA CAC GAG
 N   I   L   R   P   D   F   G   R   R   C   V   R   L   F   R   P   H   E
GCG CTG CCC CTC ATT CCC AGC GTC GCC CCG GCC GCC CCC AAG AAG CCG GTG GCA CAG
 A   L   P   L   I   P   S   V   A   P   A   A   P   K   K   P   V   A   Q
CCG CCA CCC CCG GTG CCG CCG CCC CCG GTG CCG CCT CCA CCG ACG CTG CCG ACG CCT
 P   P   P   P   V   P   P   P   P   V   P   P   P   P   T   L   P   T   P
TCG TTG CCA CTG ACC GTG CCG GTG CCG GTT CCG GTG CCG GTA CCA GTG CCG GTG CCG
 S   L   P   L   T   V   P   V   P   V   P   V   P   V   P   V   P   V   P
GTG CCG CCC CCG CAG CAG CCG CCG CCT CCG CAG CCT GCG CCG CCC CCG CCA GCC CCC
 V   P   P   P   Q   Q   P   P   P   P   Q   P   A   P   P   P   P   A   P
GCA CAG GCT CCG ACG GCC ACC GCC GCC ACC ACC GCC GCC GCC ACC CCC GCC TCC CCC
 A   Q   A   P   T   A   T   A   A   T   T   A   A   A   T   P   A   S   P
GCC GCA GCC CAG CAG CTG TGG CCG GCT TGG GTG TAC TGC ACC AGA TAC TCG GAC AGA
 A   A   A   Q   Q   L   W   P   A   W   V   Y   C   T   R   Y   S   D   R
CCC TCT TCA GGT CCT CGG TCG CGA CGG CTG AAG CGC AAA GAC AAG AAG CCG GAA GAG
 P   S   S   G   P   R   S   R   R   L   K   R   K   D   K   K   P   E   E
AAG CGG CCG CGT ACG GCG TTC AGC GGC GAG CAG CTG GCG CGG CTG AAG CAC GAG TTC
 K   R   P   R   T   A   F   S   G   E   Q   L   A   R   L   K   H   E   F
ACC GAG AAC AGA TAC CTG ACT GAG CGC CGC CGC CAG GAG CTG GCC AGG GAG CTG GGC
 T   E   N   R   Y   L   T   E   R   R   R   Q   E   L   A   R   E   L   G
CTC AAC GAG GCG CAG ATC AAG ATC TGG TTC CAG AAC AAG CGC GCC AAG ATC AAG AAG
 L   N   E   A   Q   I   K   I   W   F   Q   N   K   R   A   K   I   K   K
GCC AGC GGC CAG AAG AAC CCG CTG GCG CTG CAG CTG ATG GCG CAG GGC CTC TAC AAC
 A   S   G   Q   K   N   P   L   A   L   Q   L   M   A   Q   G   L   Y   N
CAC AGT ACC GTG CCC ATC GAT GAA GAC GAC GAA GAG ACG ACA GCA CCG CCC CCG CAG
 H   S   T   V   P   I   D   E   D   D   E   E   T   T   A   P   P   P   Q
CAG CAG CAG CCG CCG ACG GCG CCG CAG ACC GTC TCC GGC GTC CTG GCT GCT CCC ACA
 Q   Q   Q   P   P   T   A   P   Q   T   V   S   G   V   L   A   A   P   T
ACA CCG
 T   P
```

```
TGAATCGCCGCGACA CTGTGCAGTGACTTG GACTACTGACAGCAT CAACAAACATTATGC AGTAGCAATAGCAGT GCACCAGCATGTCAG ACAATTCCGCTGTGC
GTCGAGAGGGAGTAA CGAAACGAGAACTAA AAATCCCTCCTGCGC GGACGTTTGAGAATG ATCTCGCATTTAATC CACCGTAGCTACTAA TCGTTGTGGAGTGCA
GGTTGCCATATTCCT ACTGTATGTGGGATT TATCTGAGAAGCTTG TAGCTTAATTATAAG TTTATGACTGGTATC TTGTGTTCTGTTCAC GAGAATAATGCTGAT
AATGAGTGATAGGCA ATTCCAGCAGTTTAA GACGGAAAGTGTCAA ATCTTCGTTTAGTTA ATTTGATTTGCTCGC TGCCTGTTATTTTGC CATTAGCTAAAATTA
CATTATTTATTTTTG GTATGCCAGTAGCTG GCTGAAAATGCTCCA AGAAAAGATAGTAGA ACAGATTAAATCGCA TTCTTGTCTAAATTA CGAGATATATTTTTG
AAATAATATAGCGTT AATTCTGTAATTATC CGAATGAAACAAACA GGTGCTTGTACATAT CAGTAGAACACTTAA GAAAATCAGAAAGAC ACGATGTTCTCAATG
GTAAGCACTGTACCA CTGAAATGTAACAGT GGCCGTTAAACAGTA GATGCAATATTTCCT CGAAGTAAAGGTTTA CCGTTGGAGGTCGCT TCATTGAGAGTATAA
TTGCGTACAGGAATT TTTATTAAAAATAAA TAAACATGTATATTA TATTTCAGTGCAACA TGTTTAATACCAGTA TGTCACAAGATGTCT GTGTCTCCAGGCTTC
TAGTCCATTTCATAA ACAATACGTTACCAC GTTCATCAGCTGGTG CACTTACTGGGAGTA ATTAGTTTCCCTCAG ACTGTTGGAGGAAAT TTCTAAGAAATTAAA
AAAAAAGCTCAAGTT GCTGCAAGTTGTGAT TCTTCTTCAGTAATT TTATTGAGTTTTGTA GTGAAAAAATATCTA TTTCCTTCCAGTTTG TGAAAAAGGAACATA
TTACCAAAAACATTT CGTTTATAACTATAA ACAATCATTTTCGTT TGTCTTATCCGAAAT TATAGACAAATATTT AGTTGATTTAACACA GTGATAAACTTAGAA
ATGCACTCGTGTTAA CAGTAGACAAAATGT TTTAAGAAAAGTGAA TGCATGTCAGCGCTG TAGTATTTATATGTA GCAGAGCAATAATAT ATGGTGATTAGTACA
CTACATCGCTAAAGA ACAATGCTGGGAATA AGGCTAACATTATCA GCAATATAAATGACA GTAAGTGAATATGAT TTTTTCGTGATATCC TTCTATAATGTAAAT
ATAATTAAAGTCAGA TATGAATAAAACATT TCCTAAAAAAAAAAA AAAAAAAA
```

cDNA sequence of *Schistocerca gregaria engrailed-1* (*Sgen-1*). The coding region is in larger font and accompanied by the translated amino acid sequence. The 5′ and 3′ untranslated regions are in small font. An in-frame upstream stop codon is underlined. The *Hin*dIII restriction site in the 3´UTR that was used in the production of the *in situ* hybridisation probe is highlighted in red. *Sgen-1* encodes a 287 amino acid protein. EH1/Groucho-binding domain is in red. EH2/Extradenticle-binding domain is in orange. EH3 is in dark red. EH4 or Homeobox/Homeodomain is in blue. EH5 is in green.

```
                                                CCG CGA TGT GAC ACG GCT GAG
                                                 P   R   C   D   T   A   E
ATG GCG TCG ACG ACG GTG ATG GTG CGG GAC GCC GCG TCG GTC CAC AGC GAC GAC GAC
 M   A   S   T   T   V   M   V   R   D   A   A   S   V   H   S   D   D   D
GAC GCC GAC TCG CTG CTG AGC GTG GGC AGC GAG TCG CTG CCG CCG CCG CCT GTA GCC
 D   A   D   S   L   L   S   V   G   S   E   S   L   P   P   P   P   V   A
GCC GCC GCC GCC GAA TGC AAG ACC TGC TCG GCG TCG GCA TGC TCG CCG TCG GCG CCG
 A   A   A   A   E   C   K   T   C   S   A   S   A   C   S   P   S   A   P
GCG TCG TCG GGC CCG GCG GCG CCG CTC AGC TTC AGC ATC GAG AAT ATC CTG CGG CCG
 A   S   S   G   P   A   A   P   L   S   F   S   I   E   N   I   L   R   P
GAG TTT GGA AAG CGG TCC TCC GTG ACG GCC GCC AAA GCG CCG CAG CGG CCA CCG CAG
 E   F   G   K   R   S   S   V   T   A   A   K   A   P   Q   R   P   P   Q
CCG CAG CCG CAG CAG CCC ACC GCC GCC ACC GCC CCC GGT GCC ACC CCA GGC GCC GCC
 P   Q   P   Q   Q   P   T   A   A   T   A   P   G   A   T   P   G   A   A
CCC AGC GCC GCC GCG CCC GTT GAT CTG TCG CCG GTC GGC GTC TCC GCG TCG TCG CCT
 P   S   A   A   A   P   V   D   L   S   P   V   G   V   S   A   S   S   P
CCC ATG CTG TGG CCC GCG TGG GTG TAC TGC ACG CGA TAC TCG GAC CGC CCG TCG TCC
 P   M   L   W   P   A   W   V   Y   C   T   R   Y   S   D   R   P   S   S
GGA AGA AGT CCG CGG TCG CGC CGA CTG AAG CGC AAC AAG AAG CCC GAG GAG AAG CGG
 G   R   S   P   R   S   R   R   L   K   R   N   K   K   P   E   E   K   R
CCG CGC ACG GCG TTC AGC GGC GAG CAG CTG GCG CGG CTG AAG CAC GAG TTC ACC GAG
 P   R   T   A   F   S   G   E   Q   L   A   R   L   K   H   E   F   T   E
AAC AGA TAC CTG ACT GAG CGC CGC CGC CAG GAG CTG GCC AGG GAG CTG GGC CTC AAC
 N   R   Y   L   T   E   R   R   R   Q   E   L   A   R   E   L   G   L   N
GAG GCG CAG ATC AAG ATC TGG TTC CAG AAC AAG CGC GCC AAG ATC AAG AAG GCC AGC
 E   A   Q   I   K   I   W   F   Q   N   K   R   A   K   I   K   K   A   S
GGC CAG AAG AAC CCG CTT GCG CTG CAG CTG ATG GCG CAG GGC CTC TAC AAC CAC AGC
 G   Q   K   N   P   L   A   L   Q   L   M   A   Q   G   L   Y   N   H   S
ACC GTG CCC ATG ACG CGC GAG GAG GAA GAA CAG GCC ATC GCC TCC GAC AAG
 T   V   P   M   T   R   E   E   E   E   Q   A   I   A   S   D   K
```

```
TGAATCGCTGCGCCT CTCTTCCTTCACTAA TCTAACGGAATTTCG CTGCCGTGACTACTG GGTGCACATTGCACA AAATGAAATTCGGCC AGCGAATAGTGAGTG
CACGCGCCTACGGTT CCATTGTGTGTGCTG ACGGATAATTTGGCC AGTTCCTTACCAGGA CATCTTAGCGCTCGA ACTGAAGTGAGCTCT TCGTTGCTGTTCAAT
GGACAGAAGAAATGA ACTCCAAATGTTTCA GATAGGTTTACGTCT CTATTTCTGATGAAG TGCCACGGGTACCAG ATTTTTTTCAAAAGA CGTCCTAAGCGTATT
AAGGAAAGAAACAAC AGTAGAGGCTGGAAA TTATAAAGAATGTGG CAGACTGCTCTAGCA GCTGAGTAATACCTG TTGTGTTCAATGGTG AAGGAATGCATGTAA
GAAATTCTGTCTACA ACATATCAGTTTACA GTGTAGCAGACTGCT TAAAATGACACTAGA ATCCACAACCGACTA ATGACATGGAGTGTC GGTGTCACTAAAGTG
TGGAAAACAAATGTA GACATTTTGAAACTA TGTACCTCACTGTGA AATCAGTGGCTGAAG TTTGTAAGGGACAGG AGTTTCCTGGATCCA CCGCAAGTCTTTCTT
GGCAATCGCCACCTG CATTCTGCCAACCTT TAAACTGAAAGTTGT ACGCAGTCGACTGTA TTATGTGTAGAGGAG CAACCTCGCATTGTG TTTTATCTGTGTCTC
TTAACTGACAGAATG ACTGTGCAATAATTT AGTACCCATTTTTTC ACATAACAATGAGCA TTTCCTCCAGACGAT CTACCGACATTAGTT TTATACATGTATCAG
TCAACTGCGAACAAC AAGTGCAAAATAAGT ATTTAATCGGATGTA GAACCAAAGGAGCTT TTAAACATCCTGCAA ATAAGCTAAAGCACA AACAAATTGTTTTAA
ACAAAATATGCGTGT GTACAAATGTGTAAC AATGTAAATGTGACT GATTTTTCATTATTA TTATTATTTTATTAT TATTATTTGCAGATG ATGATAGACCTATAT
CGATTCAAACGCGTT GAAGGTATATATTTA TACTGTAGGATCAGA TCAAACGTATGGCGC AACATTTACTACTTA TATCTTTGATCCAAA TATGACGTCACATCA
AGGCCTTGTACATTT TCCAGTACATAACAG TTTTTTTCTCACTTG AGAACTCTTGTTAGC AAGATAATGCTCAAT AAGAAAGTTATTAGT CTTATTCATCATTGC
TTAGAAATGTCGCTT GAGGAACTGGATAAG TGTTTTTTGCTTTTA TTTCCTCTCTGCTAA GAAGCAGCAGAGGAA CCCGTTACCATATAT CTTCTTTTATTGATT
CTACTGAGAAACTTG AACATTTTCACATTG CAGCAGTAAACTTTT TTTAAATAATTCATT TCGGTTGCCCTGTTG TACTCTGAAAGAGAA AGATATAGCAATAAT
TTTTTACTTCGGATG TGCTATAATTAAATA AATATTTAATTATGC AGTCTGCCAAGTTTA AAGTCTTATTTATCA TGGAAATAAGAAGAT ATTACACCAAAGTCT
GATTAATTTATTCTC CATGTGTTAGAAGTA ACTGGAAAGCATTGC TTGATTTGATGACTG CCAGGTACTGTTAGC AGGATGACGTACTTC CACAGCTGTCTCTCT
CTTCAGTAACAAAAA TTGAAAAAGCTGAGA ATATTTAGTGTATAA GGAAGACAAAAATGT ACATATAATTCAGAA ATGCTCTATACTCGT TTAAACGTGTAGAAT
GTACAGAAAAATATC TAGATATAATTTAAT ACGATATAAAATGTA TATACATTGTACCTA CATATTCCTACATGT TCTCTTGTAGCATTG GTAGAGACTAATTAC
TGCATTTATGCATGA GACAAGATTTCTTGC TGAAAACAGCAATTC CATTAAGAGAATTAT GCCCATTTGATTGTG ACTGTAAAAAGTAGT GATTGGTAACATTAA
AATAACTGCAGTATT TTTCCTCATTTTCAT GCAAAGATATGTACT TAAAATATTTATTGA AGTCAGAATAAAGCT TACTTCATTTACTTA AAAAAAAAAAAAAAA
A
```

cDNA sequence of *Schistocerca gregaria engrailed-2* (*Sgen-2*). The coding region is in larger font and accompanied by the translated amino acid sequence. The 3′ untranslated region is in small font. The *Sal*I restriction site in the 3´UTR used in the production of the *in situ* hybridisation probe is highlighted in red. *Sgen-2* encodes a protein at least 264 amino acid long. We are unsure whether the first methionine is the start codon due to the lack of an upstream stop codon and Kozak consensus sequence. The *invected*-specific domain is underlined. EH1/Groucho-binding domain is in red. EH2/Extradenticle-binding domain is in orange. The hexanucleotide micro-exon sequence and RS-motif is in violet. EH3 is in dark red. EH4 or Homeobox/Homeodomain is in blue. EH5 is in green.

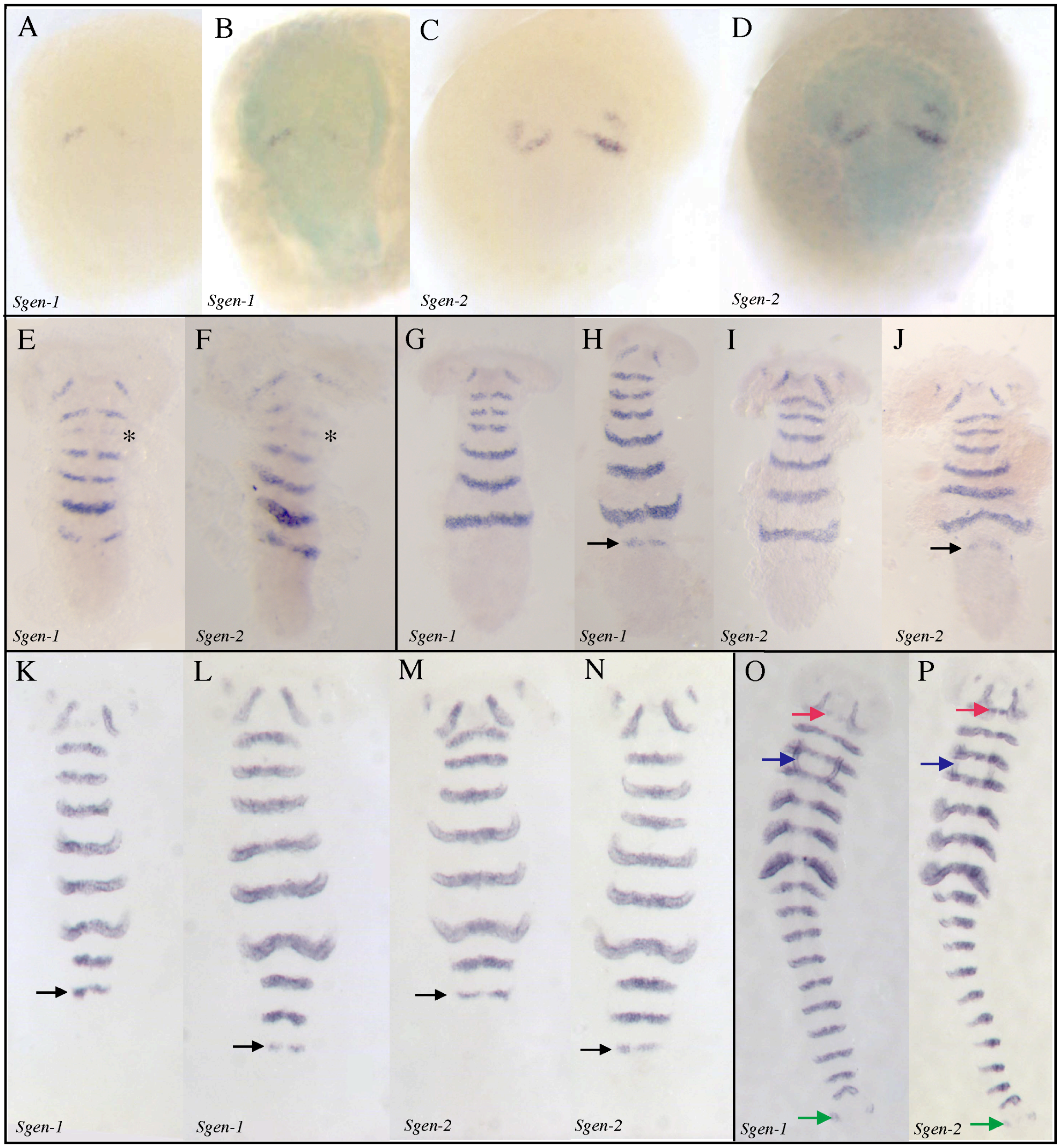
CLUSTAL W (1.82) pairwise alignment of the Sgen-1 and Sgen-2 3'UTR in situ hybridisation probes. Note that the longest contiguous stretch of sequence identity (apart from the polyA tails) is only ten nucleotides in length.

Gap penalties were set at the lowest possible levels, but alignments constructed with higher penalties for gaps had even shorter maximum stretches of contiguous sequence identity. The probes are unlikely to cross hybridise.

```
Sgen-1          --AGCTTGTAGCTTA-AT-TATA--AGT---TTATG-ACTG-GT---ATCTTGTGTTCTG 46
Sgen-2          TCGACT-GTA--TTATGTGTAGAGGAGCAACCT-CGCATTGTGTTTTATCT-GTGT-CTC 54
                    ** ***  ***  * ** *  **     *  * * ** **   **** **** **

Sgen-1          TTCAC-GA--GAATAA-TGCTG--ATAATG-AGTGATAGGCAATTC--CAGCAGTTTAA- 96
Sgen-2          TTAACTGACAGAATGACTG-TGCAATAATTTAGT-ACC--CATTTTTTCA-CA---TAAC 106
                ** ** **  **** * ** **  *****  *** *    ** **   ** **   ***

Sgen-1          GACGGAA-AGTG--TC-A-A--ATCT-TCG-T-TTAGTTAATT-TG-AT-T-T--GCTCG 140
Sgen-2          AATG-AGCATTTCCTCCAGACGATCTACCGACATTAGTT--TTATACATGTATCAG-TCA 162
                 * * *  * *   ** * *  ****  **   ******  ** *  ** * *  * **

Sgen-1          -CTGCCTGTTATTTTGC--CATTAGC-TAAAATTACATTATTTATTTTTGG-TAT-G--C 192
Sgen-2          ACTGC--G--A----ACAACA--AGTGCAAAAT-A-AGTATTTAAT--CGGATGTAGAAC 208
                 ****  *  *     *  **  **   ***** * * ****** *   ** * * *  *

Sgen-1          CA--GTAGCTGGCTGAAA-ATGCTCCAAGAAAAGATA--GTA-GAACAGATT-----AAA 241
Sgen-2          CAAAGGAGCT--TTTAAACATCCTGCAA-ATAAGCTAAAGCACAAACAAATTGTTTTAAA 265
                **  * ****   * *** ** ** *** * *** **  * *  **** ***     ***

Sgen-1          TCGCA-T-T-C-T-TGT-CTAAAT---TA-CGA-G---AT---A-T-ATTTTTGAAATAA 282
Sgen-2          -CAAAATATGCGTGTGTAC-AAATGTGTAACAATGTAAATGTGACTGATTTTTCA--TTA 321
                 *  * * * * * *** * ****   ** * * *   **   * * ****** *  * *

Sgen-1          TATAGCGTTAATTCTGT-A--ATTATC--C-GA--ATGA-A-AC--A-A---A--CAGGT 324
Sgen-2          T-TATTATTA-TTTTATTATTATTATTTGCAGATGATGATAGACCTATATCGATTCAAAC 379
                * **   *** ** * * *  *****   * **  **** * **  * *   *  **

Sgen-1          GC-TTG----TACATATCAGTAGAACACT-TAAGA--AAATCAGA-A-A-GACACGATGT 373
Sgen-2          GCGTTGAAGGTATATATT--TA---TACTGTAGGATCAGATCAAACGTATGGCGCAACAT 434
                ** ***    ** ****   **    *** ** **  * **** *   * * * * *  *

Sgen-1          T--CTCAATGGTA-AGCACTG-TACCAC-TGA-AATGT-A-A-CAGTGGCCGTTA-A-A- 421
Sgen-2          TTACT-ACT--TATATCTTTGAT-CCAAAT-ATGACGTCACATCAA-GGCC-TTGTACAT 487
                *  ** * *  ** * *  ** * ***  * *  * ** * * **  **** **  * *

Sgen-1          ----CAGTAGATG-CAATATTTC-CTCGAAGTAAAGGTTTACCGTTGG--AGGT--CGCT 471
Sgen-2          TTTCCAGTACATAACAGTTTTTTTCTC-ACTTGAGAACTC-TTGTTAGCAAGATAATGCT 545
                    ***** **  ** * ***  *** *  * *    *    *** *  ** *   ***

Sgen-1          TCATTGAGA--GT-AT-AAT--TG--CGT-A---C--AGGAATTTT-TATTAAA-AA-TA 514
Sgen-2          -CAATAAGAAAGTTATTAGTCTTATTCATCATTGCTTAGAAATGTCGC-TTGAGGAACTG 603
                 ** * ***  ** ** * *  *   * * *   *  ** *** *    ** *  ** *

Sgen-1          AATAAACATGTAT-A-TTATATTTCAGTGCAACATGTTTAATACCAGTATGTCACAAGAT 572
Sgen-2          GATAAGTGTTTTTTGCTTTTATTTCC-T-CT-C-TGCT-AAGA--AGCA-G-CAGAGGA- 653
                 ****   * * *   ** ******  * *  * ** * ** *  ** * * ** * **
```

Sgen-1/Sgen-2 3'UTR alignment (cont.)

```
Sgen-1          GTCTGTGTCTCCA-G-G-CTTCTAGTCCATT---TC-A-T---AAAC--AATACGTTACC 619
Sgen-2          ACCCGT-TA-CCATATATCTTCT--TTTATTGATTCTACTGAGAAACTTGA-ACATTTTC 708
                  * ** *  ***     *****  *  ***   ** * *   ****   * ** **  *

Sgen-1          ACGTTCATCAGCTGGTGCACTTACTGGGAGTAATT-A-----GTTTCCCT-CAG-ACTGT 671
Sgen-2          ACATT-G-CAGC-AGTAAACTTTTTTTAAATAATTCATTTCGGTTGCCCTGTTGTACT-C 764
                ** **   ****  **  ****  *   * ***** *     *** ****   * ***

Sgen-1          TG---GAG---GA---------AATTTCTAA----GAA-AT--TA-AA--AAA-AAA-GC 704
Sgen-2          TGAAAGAGAAAGATATAGCAATAATTTTTTACTTCGGATGTGCTATAATTAAATAAATAT 824
                **   ***   **         ***** * *    * *  *  ** **  *** ***

Sgen-1          TCAAGT-TGC----TGC-AAGTTGTGAT-TCTT-CTT--CA-GTAATTTTATTGA-GTT- 751
Sgen-2          TTAATTATGCAGTCTGCCAAGTT-TAAAGTCTTATTTATCATGGAAAT--AA-GAAGATA 880
                * ** * ***    *** ***** * *  ****  **  ** * ** *  *  ** * *

Sgen-1          TTGTAGTGAAAA---AAT--ATCTATTTCCTTCCAGTTTGTGAAAAAGGAACATATTACC 806
Sgen-2          TTACA-CCAAAGTCTGATTAATTTATT-C--TCCA-TGTGTTAGAA-GTAAC----TG-- 928
                **  *   ***     **  ** **** *  **** * *** * ** * ***    *

Sgen-1          AAAAACATT--TCG-TTT-ATAACT---A--TA-----AACA--AT--CAT--TTTCGT- 845
Sgen-2          GAAAGCATTGCTTGATTTGATGACTGCCAGGTACTGTTAGCAGGATGACGTACTTCCACA 988
                 *** ****  * * *** ** ***   *  **     * **  **  * *  ** *

Sgen-1          -TTGTCT-TATC--C-G------AAATT-ATAGA-C--A-AATATTTAGT-TGATTT--A 886
Sgen-2          GCTGTCTCTCTCTTCAGTAACAAAAATTGAAAAAGCTGAGAATATTTAGTGT-ATAAGGA 1047
                  ***** * **  * *      ***** * * * *  * ********** * **    *

Sgen-1          ACACAG---TG---ATA-AACTTAGAAATGC-----ACTCGTGTTAA-CA-GTAGA---- 928
Sgen-2          AGACAAAAATGTACATATAATTCAGAAATGCTCTATACTCGT-TTAAACGTGTAGAATGT 1106
                * ***    **   *** ** * ********     ****** **** *  *****

Sgen-1          -CA---AAAT-----G-T----TTTAA---GA---AAAGTG-A-ATGCAT-GTCAGCGCT 965
Sgen-2          ACAGAAAAATATCTAGATATAATTTAATACGATATAAAATGTATATACATTGT-A-C-CT 1163
                 **   ****     * *    *****   **   *** ** * ** *** ** * * **

Sgen-1          GTAGTATT--TATATGT--------AGCA--G--AGCA-A-TAATATATGGTG-ATT-AG 1007
Sgen-2          ACA-TATTCCTACATGTTCTCTTGTAGCATTGGTAG-AGACTAAT-TAC--TGCATTTA- 1217
                  * ****  ** ****        ****  *  ** * * **** **   ** *** *

Sgen-1          TACACT---ACA-----TC--GCTAAAGA-A-CAATGC--T---GGGAATAAGGCTAACA 1050
Sgen-2          TGCA-TGAGACAAGATTTCTTGCTGAAAACAGCAATTCCATTAAGAGAATTATGCC--CA 1274
                * ** *   ***     **  *** ** * * **** *  *   * **** * **   **

Sgen-1          TT--ATCAGCAA-TATAAA------TGAC-AGTAA-GT--GAATA--TG-A-T-TTTTTC 1092
Sgen-2          TTTGATT-GTGACTGTAAAAAGTAGTGATTGGTAACATTAAAATAACTGCAGTATTTTTC 1333
                **  **  *  * * ****      ***   ****  *   ****  ** * * ******

Sgen-1          GTGATATCCTTC-TATAA-----TGTA---AATATA---ATTAAAGTCAGATATGAA--- 1137
Sgen-2          CTCAT-T--TTCATGCAAAGATATGTACTTAAAATATTTATTGAAGTCAGA-ATAAAGCT 1389
                 * ** *  *** *  **     ****   ** ***   *** ******** ** **

Sgen-1          TAAAACATTT-CCTAAAAAAAAAAAAAAAAAAA 1169
Sgen-2          TACTTCATTTACTTAAAAAAAAAAAAAAAAA-- 1420
                **   ***** * ******************
```



Whole mount *in situ* hybridisation for *Sgen-1* (A-B, E, G-H, K-L, O) and *Sgen-2* (C-D, F, I-J, M-N, P), on embryos from five egg pods (boxes) at different stages in development. Embryos from the same egg pod (boxes) are synchonized in their development to within about 1%. At 15% of development (A-D), *Sgen-2* is typically expressed as two chevrons in the head (C), while *Sgen-1* is typically expressed as only one chevron (A) (B & D: The same embryos stained with methylene blue). By 17% of development (E & F) expression of both genes has appeared in seven stripes, corresponding to the antennal, three gnathal and three thoracic segments. Note that expression in the maxillary segment (*) is particularly weak, perhaps suggesting it has only just appeared. Between 17% and 30% development abdominal stripes appear in an anterior to posterior progression (arrows) (G-N). The within pod variation in expression patterns is the same for *Sgen-1* as for *Sgen-2* (compare G-H with I-J, and K-L with M-N). By 30% of development expression of both genes has also appeared in the intercalary segment (red arrows), dorsal ridge (blue arrows) and cerci (green arrows) (O & P).

```
GTGGTGAACTTATGACCGAGG
```

```
ATG GCG TTC GAG GAC CGC TCC AGC CCC AAC ACC ACC GAC GAT GCC TCC CAG ATC AAA ACC
 M   A   F   E   D   R   S   S   P   N   T   T   D   D   A   S   Q   I   K   T
CCC AAC TCC CCC GAG TCC TCC CGG ACC TCT CCC TAC ACC TGC ACC ACC CTC TCG CAG GAC
 P   N   S   P   E   S   S   R   T   S   P   Y   T   C   T   T   L   S   Q   D
TCC CCC AAG GGC GAC TTC TTC CGC ATC ACC TCC TTC GCC CTG TCC AGC CCC AGG GCC TCC
 S   P   K   G   D   F   F   R   I   T   S   F   A   L   S   S   P   R   A   S
CCA TCC CCG CCG CCC CAG CCC GAG CCC GTC ATC CAG ACG CTC AAA TAC TCC ATC AGG AAC
 P   S   P   P   P   Q   P   E   P   V   I   Q   T   L   K   Y   S   I   R   N
ATA CTA AAG CCC GAG TTC GGG AAA AAC GCC GTT CTT AAA ACA CGC ACC AAG ATT GGC TTC
 I   L   K   P   E   F   G   K   N   A   V   L   K   T   R   T   K   I   G   F
AAG CCC TAT GAG ATT AAA GAG GAC GTC AAA CCC TTC TCG ACG GCG CCG CTG GGC AGC CTG
 K   P   Y   E   I   K   E   D   V   K   P   F   S   T   A   P   L   G   S   L
TGC CAA GCC GTG TCG CAG ATC GGG AGT CCG GAG CCC GTG GCG AGG CCG AAG AGT CCG GTA
 C   Q   A   V   S   Q   I   G   S   P   E   P   V   A   R   P   K   S   P   V
AAA GGG CTG TTG CCC ACT CCG GAC GAT ATC AAG AAG GAC GAG GGG TCG GTG CCC ACC TTG
 K   G   L   L   P   T   P   D   D   I   K   K   D   E   G   S   V   P   T   L
TGG CCC GCT TGG GTC TAC TGC ACG CGG TAT TCG GAC CGG CCC AGT TCA GGG CCG CGC TCG
 W   P   A   W   V   Y   C   T   R   Y   S   D   R   P   S   S   G   P   R   S
CGC CGC ATG AAA AAA CCG TCC AAA CCC AAC GGT GAG GAC AAG AGG CCG CGG ACG GCT TTC
 R   R   M   K   K   P   S   K   P   N   G   E   D   K   R   P   R   T   A   F
TCC AGT GCT CAA CTG GCA CGC CTC AAG CAC GAG TTT AAT GAA AAT CGA TAT TTA ACC GAA
 S   S   A   Q   L   A   R   L   K   H   E   F   N   E   N   R   Y   L   T   E
CGG AGG AGG CAG CAG TTA AGT GCC GAG TTG GGT CTG AAC GAG GCA CAA ATC AAA ATT TGG
 R   R   R   Q   Q   L   S   A   E   L   G   L   N   E   A   Q   I   K   I   W
TTC CAA AAT AAG CGA GCC AAG ATC AAA AAA TCC TCG GGT CAG AAA AAC CCA CTG GCG CTG
 F   Q   N   K   R   A   K   I   K   K   S   S   G   Q   K   N   P   L   A   L
CAG CTG ATG GCG CAG GGA CTC TAC AAC CAC TCG ACA GTG GCG TGC GAC GAG GAC GAC TTG
 Q   L   M   A   Q   G   L   Y   N   H   S   T   V   A   C   D   E   D   D   L
CCC CTC TCG TCC
 P   L   S   S
```

```
TAAAGGACTTCATAG CTTCTAAGCCGATCT CATTGTATTTATTTG TACAAATTATTTAGG TGCATTATTGTTGTG TAAATATTGTAGAGA
GCGAAGGCAAATCTC GTAGTTGCGTAATTT ATAAGCCATTATTGC TTGAAATTTTGCGTG AAAATAACACGAGTA GTTATGTAGATGTAG
ATCGCTGTGTATGAT AAAAGCCTTAAAAAT AGACTGATTAAATAT TTAGATTATAGCAAA TTGTGCAATGTTCTC TTAAGACTTTCTTGT
ATATATGAAATAAAT ATTGTTCTTCCATCG TGACTGTTTTATTGC CTCCTTTCAACAATG ACCATTATTAAATTT TTCCGTCACCACGCT
TTGATTTGTTTCTCA ATTGAAATTCGCTGG CATGCGAAAGAATGA AAGTCGATTAAGATC TTGTAAGGCTTTGCG TTAACGATATTCATA
AACAACTGTCACAGC CCAATAAAGTAATTA ACGGTGACTTAATTG ACTTTATCTTTTTAA CATCTGTAATCTCGT GCTTCTCGAATAAAG
CGCATTTGAACGTTT TATTGCAGTTTACCA ATTAAGTGATTCAAG TTTTGAAAAAATTTA CATTATCTGGAATAC ACGATCTTTCACCAT
TTTCTAGTGCAAATA AATTTGCAATACCGT AAAAGTCGCCGCTAA GTTAGTACCGTACTT AAAAGTAATTCATTT GCTAATTCTCGAAAC
GTGATGCACGTAATC ACTGATGCTCCACTA TTGTCTTTGGAAAAA AATAAAA
```

The genomic DNA sequence (minus introns) of *Tribolium castaneum engrailed.* This second *engrailed*-family gene was not identified by Brown *et al*, 1994. The coding region is in larger font and accompanied by the translated amino acid sequence. The 5′ and 3′ untranslated regions are in small font. An in-frame upstream stop codon is underlined. *Tcen-1* encodes a 284 amino acid protein. EH1/Groucho-binding domain is in red. EH2/Extradenticle-binding domain is in orange. EH3 is in dark red. EH4 or Homeobox/Homeodomain is in blue. EH5 is in green. The position of the EN2 intron is shown with an arrow.

```
ACGTCCGCGGACGTG GGCGAATCGTCCCTA ATTGTATCAGTTGAA GCGGCCTAGCTGTGC CTAAAACCAAGTGGG TGTCTCCCGACAGAA CCTGGAGAGCACCGA
TGGCTCCGTCTCC

ATG GAC TCC AGC GAC CAC TTC GAC CGG GAA TCC CCC AAC ATC GAC CAC AAC AGC TGC AGC
 M   D   S   S   D   H   F   D   R   E   S   P   N   I   D   H   N   S   C   S
AGC GAC GAC ACC GTG CTT TCC GTC GGC AAC GAG AAC CCG CCA CCG GAA GAC ACT CCG CTC
 S   D   D   T   V   L   S   V   G   N   E   N   P   P   P   E   D   T   P   L
TCC TTC AAG AAC ATC GAG AGC CAC CTC AAC GCC ATT TCG CAA ATC ACG AAC AGT ACC TTA
 S   F   K   N   I   E   S   H   L   N   A   I   S   Q   I   T   N   S   T   L
GAC CCG GGC CGC AAG AGC CCC TCT AGT CCC AGG ATA AGC AGT CCG TCG AGT ACG AAG TCC
 D   P   G   R   K   S   P   S   S   P   R   I   S   S   P   S   S   T   K   S
GGC TCG CCC GGA TTT TTG ACC TAC ACG AAG GCG GAC CGG GAC GTG GAC CTG TTC CGG GGA
 G   S   P   G   F   L   T   Y   T   K   A   D   R   D   V   D   L   F   R   G
TCG TCC ACC CCG GAG TCG CCG GAA CAC TAC TAC AAC CAG AAG ACG CTT CAG GCT AAT AAT
 S   S   T   P   E   S   P   E   H   Y   Y   N   Q   K   T   L   Q   A   N   N
AAC GAC GCG AGT AAT GGG AAT CTC AAG TTT TCG ATA GAT AAT ATT CTG AAG GCG GAT TTT
 N   D   A   S   N   G   N   L   K   F   S   I   D   N   I   L   K   A   D   F
GGA CGG AGG ATC ACG GAT CCG ATT AAC ATC CGG AAG TGC AAG CCG AAG AAG GTT GTG CCG
 G   R   R   I   T   D   P   I   N   I   R   K   C   K   P   K   T   V   V   P
GAA GTT GGG GGT GTT GAA GAG GCT AAA GGA CCT GTA GAT TTG AGT AAG AGT GAG CCG GAG
 E   V   G   G   V   E   E   A   K   G   P   V   D   L   S   K   S   E   P   E
AAG AAG ACT GAG AGT CAG CCG ATG TTG TGG CCA GCC TGG GTT TAC TGC ACG AGG TAC AGC
 K   K   T   E   S   Q   P   M   L   W   P   A   W   V   Y   C   T   R   Y   S
GAC CGA CCC AGT TCA GGA CGA AGT CCG CGA ACC AGG CGA GTC AAG AAA CCC GGG GCC AAG
 D   R   P   S   S   G   R   S   P   R   T   R   R   V   K   K   P   G   A   K
CAG GGG GCG CCC ACC GCC GAG GAA AAG CGG CCC CGG ACG GCC TTT TCA GGG GCG CAG TTA
 Q   G   A   P   T   A   E   E   K   R   P   R   T   A   F   S   G   A   Q   L
GCC CGG CTG AAG CAC GAG TTC GCC GAA AAC CGC TAT CTG ACC GAG AGG CGG CGG CAG CAG
 A   R   L   K   H   E   F   A   E   N   R   Y   L   T   E   R   R   R   Q   Q
CTC AGT GCG GAG CTG GGG CTC AAC GAG GCC CAG ATC AAA ATC TGG TTC CAG AAC AAA AGG
 L   S   A   E   L   G   L   N   E   A   Q   I   K   I   W   F   Q   N   K   R
GCC AAA ATC AAA AAG GCC TCC GGC CAG AAA AAC CCC CTA GCC CTC CAG CTG ATG GCG CAG
 A   K   I   K   K   A   S   G   Q   K   N   P   L   A   L   Q   L   M   A   Q
GGG CTG TAC AAC CAC TCC ACC ATT CCG CTG ACG AAG GAG GAG GAA GAG CTG CAG GAG ATG
 G   L   Y   N   H   S   T   I   P   L   T   K   E   E   E   E   L   Q   E   M
CAA GGC ACT AAG AGC CCC GCG
 Q   G   T   K   S   P   A

TAGGCCGCTGGTTTT GTTCGTGATAATTTG ATGCTTCCAGTTGTT TATTGTGATTTGTTG TGATTCGTCCGTGGA TGAGGTGAGAGTGGA TTCTTTGGATTTTTA
CAGTGATTTGCGA-G CGTTTTTGCCTCGTC CAAAGACGAAGTCGG GTAAGCTGTCAATGG TCTACTTAAGAGTTG GGCAACTTTTTGCGG TGATCTCTCGAGTAT
GTTAGCATTGTTTCT TTAATTCTACAATTC TCAGAGCTTCGTGTT ACAAAAAATGTATTA ATGTACCATTCATTT AAACACAATTGCGAG CTGTGTGGTACTAAA
TAACAGTTTTTACAA ACGGTGCCATTATTG TTTAAATCGCATGGG ATCATATTCTTGTTT GGTTTTTCTTGCACG ACTACCAAAGACCGT GCAAGACACGGGATA
TTTGATCCGACCTTG TACATAACACTCTGT AAAAATTACAACTAG GACTAGTTATTTATT GTATTAGCAAAATAA TTCGGAATCGAAAGC AATATTTATTAGGCA
AAACTTGAACCTAAT GTATATTTAAATGAA ACTATTGTAAATATT TATATAATACTGTCT GTGAGTTGAAGCTTA TTTATAATGTATTCG ACCTAAGTAATAGTG
AAAAATATGTACCTA CAGAAATTGTTTATA ATTCAACTGTAAAAT ATAGATTA-TATAAA ATAAATACTATTCAA GCGAAAAAAAAAAAA AAAAAAAAAAAA
```

The genomic sequence (minus introns) of *Tribolium castaneum invected.* The coding region is in larger font and accompanied by the translated amino acid sequence. The 5' & 3' untranslated regions are in small font. In-frame upstream stop codons are underlined. *Tcen-2* encodes a protein 327 amino acid long. Bases and amino acids that differ from the sequence published by Brown *et al*, 1994 are boxed. The *invected*-specific domain is underlined. EH1/Groucho-binding domain is in red. EH2/Extradenticle-binding domain is in orange. The hexanucleotide micro-exon sequence and RS-motif is in violet. EH3 is in dark red. EH4 or Homeobox/Homeodomain is in blue. EH5 is in green. The positions of the EH2 introns are shown with arrows.