

## SI Appendix 1

Bioinformatics analysis of Acp70 proteins from a series of *Drosophila* species (*D. mauritiana*\*, *D. melanogaster*\*, *D. madeirensis*, *D. sechellia*, *D. simulans* and *D. subobscura*) against the *An. gambiae* translated sequences from known or novel gene predictions, as downloaded from Ensembl (database release version 42.3e Feb 2006). A minimum of five best *Anopheles* hits is shown for each Acp70A from the different *Drosophila* species. Reverse searches against the *D. melanogaster* genome were performed for *Anopheles* proteins (shown in color) recurrently retrieved among the best alignments. Only the best alignment and the alignment against Acp70A, when obtained, are provided. The *Anopheles* proteins/genes marked in red (AGAP009352, AGAP012681 and AGAP012682) repeatedly provide the best alignments with Acp70A (also see the phylogenetic tree of these genes at the bottom of the document). AGAP009352 and AGAP012681 correspond to two alleles of the same gene and their alignment with each other and with AGAP012682 is also shown. Blosum 62 was used as scoring matrix in all searches, unless otherwise indicated. Note that AGAP008968 is present in Table 1 as the orthologue of CG31704, and AGAP009355 as the orthologue of Drosocin, based on the matlab analyses reported here (see also SI Appendix 2). The unrooted phylogenetic tree at the end of the document was constructed by the neighbour-joining method based on the sequence alignment of 4 *Drosophila* (Acp70A, Drosocin, Mst57Da, CG31704) and 5 *Anopheles* genes in red (AGAP009352/AGAP012681, AGAP012682, AGAP009355/AGAP009357, AGAP008968 and AGAP009354/AGAP009359/AGAP009356) generated using Clustal\_X and Clustal\_W (1). These genes were selected as they were identified in the forward and reverse searches reported here. The proteins are identified by the last 5 digits of the Ensembl identifier, omitting the AGAP0 code (for instance, AGAP009352 is represented by 09352). The bootstrap values of a 1,000 replicates are reported. The scale bar represents the amino acid divergence.

\*A refined *An. gambiae* data set was used in the analysis reported here for these genes.

### D. MAURITIANA

```
best alignment n: 1
names: sp|P67806|Acp70A - AGAP012681
identities: 15/59 (25%), positives: 32/59 (54%)
score: 6
MK--TSLSLFLVLVCLLGLVQSWWEPWNRKPTKYP I--PSPNPRDKWCRLNLGP--AWGGRC
|| :  :::|:|::|:|:|:| : | : | :| : : | : :| :| : |
MKFYVQMMLLALLALLAFVVGAGEMYVTDKVDPPPPFDPDDESFYNGCLEGTNPHVW---C
```

```
best alignment n: 2
names: sp|P67806|Acp70A - AGAP009352
identities: 16/62 (25%), positives: 32/62 (51%)
score: 5.500000e+000
MK--TSLSLFLVLVCLLGLVQSWWEPWNRKPTKYP I PSPNPRDK----WCRLNLGP--AWGGRC
|| :  :::|:|:|:|:|:|:| : | : | :| : : | : :| :| : |
MKLYVQMMLLALLALLAFVVGAGEMYVTDKVDPPPPF---DPDDESFYDGCLEGTNPHVW---C
```

```
best alignment n: 3
names: sp|P67806|Acp70A - AGAP009357
identities: 16/59 (27%), positives: 29/59 (49%)
score: 2
MK-TLSLFL-VLVCLLGLVQSWWEPWN--RKPTKYP I PSPNPRDKWCRLNLGPAWGGRC
|| :::|:| :|:|:|:|:|:|:| : | : | :| : : | : :| :| : |
MKFFVQVFLFALLALLGLVSAGKMAINPATTTTARPIFFPCFDEDGIFIGCTP----HP
```



names: acp70A - AGAP012682  
identities: 14/57 (24%), positives: 29/57 (50%)  
score: 2.500000e+000  
MKTLA--LFLVLVVCVGLVQAWQWEPWNRKPTKFPIPSPNPRDKWCRLNLGPAWGGRC  
|| : ::|:|:|:|:|:| || | : : | : : | :  
MKLYVQMLLALLLALVAFVVGAREMPVTEKVD----PPPFEPDGY----IHPTDYG-I

best alignment n: 6  
names: acp70A - AGAP009356  
identities: 16/78 (20%), positives: 32/78 (41%)  
score: -2.500000e+000  
MKTLA--LFLVLVCVLG-----LVQSWE-----WPWNRKPTKFPIPSPNPRDKWCRLNLGPAWGGRC  
||| ::|:|:|:|:|:| || : : : || : : : : | : ||  
MKFLATTILFAFIALGQSAEPSRNDISVYEKLCQTFGLEKDCWTTANPPTKRPLADIKQETKQNK-NL-----MR

**D. MADEIRENSIS**

best alignment n: 1  
names: O76860 (Acp70A) - AGAP012681  
identities: 17/61 (27%), positives: 35/61 (57%)  
score: 1.950000e+001  
MKFPISLMLLLLVLGVA-LGV-HWRITRRITSTSSTWGPRD--IQKWCRNFGP-AWGGRGC  
||| ::|:|:|:|:|:| || : : : : | : : : | : : | :  
MKFYVQMLLALLLALLAFVVGAGEMVTDKVDPPP-FDPDDESIFYNGCLEGTNPHVW----C

best alignment n: 2  
names: O76860 (Acp70A) - AGAP009352  
identities: 16/61 (26%), positives: 35/61 (57%)  
score: 16  
MKFPISLMLLLLVLGVA-LGV-HWRITRRITSTSSTWGPRDIQKW--CRLNFGP-AWGGRGC  
||: ::|:|:|:|:|:| || : : | : : : : | : : : | : : | :  
MKLYVQMLLALLLALLAFVVGAGEMVTDKVDPPP-FDPDDESIFYDGCLEGTNPHVW----C

best alignment n: 3  
names: O76860 (Acp70A) - AGAP001574  
identities: 26/93 (27%), positives: 38/93 (40%)  
score: 1.350000e+001  
MKFPISLMLLLL---LLVGVA---LGVHWR-----ITRRITSTSSTWGPRDIQKWCRNLN-----FGPA-----WG-----  
GR--GC---  
: :|:|:|:|: | : : : | || | : || : : | || :| :| |||| ||  
|| ||  
RQ-ALSLVVLVGPQLSTSIARLCTGVLCRQAFAPYAFIT--TERAS---PRT-  
EWLKLPFWDSPPFGPASCCKKFWGHLSPSPGRHHGCRSLV

best alignment n: 4  
names: O76860 - AGAP011482  
identities: 20/81 (24%), positives: 38/81 (46%)  
score: 1.350000e+001  
MKFPISLMLLLLVLGV-ALGVHWRITRRITSTSSTWG-PR-----DIQKW-----CRLN--FGPAWG-----GRGC  
|:| :| :|:|:| || || : : : : | || : : : ||:| :| :| :| :| :| :|  
MRF-LSYINLMVL-GVLALVL-----SASAGPCGCPRSYRPVCGTDLKTYSNQCVLDCRINSNYGRKFLKLLRDLGH-C

best alignment n: 5

names: 076860 - [AGAP009357](#)  
identities: 18/66 (27%), positives: 32/66 (48%)  
score: 12  
MKFPISLML--LLLLVGVALGVHWRITRRITSTSTSWGPRDIQKWCRNLNFGPAWG--GR--GC-----  
||| ::::| || |:|:: : | : |:|:: | | |:: | ||  
MKFFVQVFLFALLALLGLVVSAGKMAINPATTTTA----RPI-----FFPCFDEDGIFIGCTPHP

**D. SECHELLIA/D. SIMULANS**

best alignment n: 1  
names: 018417 (Acp70A) - [AGAP008968](#)  
identities: 16/65 (24%), positives: 37/65 (56%)  
score: 17  
MKTLSVFLVLVC--LLGLVQSWEPWNRQPTRY--PIPSNPRD--KWCRLN--LGPW----GGRC  
|| | :: |:| :|: : ::| | :|: : : : | | : :|| :|:|  
MKLLFLLSFLLCAILAAAGKYSCP--ACPANYLPVCGTDGKTYANECALLECTVAPAVKVARSGEC

best alignment n: 2 female  
names: 018417 (Acp70A) - AGAP006555  
identities: 21/85 (24%), positives: 38/85 (44%)  
score: 1.650000e+001  
MK-TLSVFLVLVCLLGLVQSWEPWNRQPTRY--RYP--SP--NP--RDKWCRNLNLPWGG--RC-----  
--  
|| ||::||::|:: |:|: :| : | : |:| :| | :| | : | | :|  
MKYTLAAFLLVVAVQAQGGCYP--SVAPCSSCQKPVVQYRVNYVAPVCNTCEEKRCPM--  
GYKWDGCKCVPDCACLRQLVSLY

best alignment n: 3 female  
names: 018417 (Acp70A) - AGAP011482  
identities: 20/74 (27%), positives: 36/74 (48%)  
score: 1.450000e+001  
MKTLSVFLVLVCLLGLVQSWEPWNRQPTRY--PIPSNPR-----DKWCRLN--LGPW----GRC  
|: || : :|:| :|:| | : | | :|:| | :| | :| | :|  
MRFLSYINLMVLGVLALVLSASAGPCGC--PRSYRPVCGTDLKTYSNQCVLD--CRINSNYGRKFGLKLLRDGHC

best alignment n: 4 our target  
names: 018417 (Acp70A) - [AGAP009352](#)  
identities: 17/61 (27%), positives: 31/61 (50%)  
score: 14  
MKTLSVFLVLVCLLGLVQSWEPWNRQPTRYPIPSNPRDK----WCRNLNLP--AWGGRC  
|| | | ::|:|:|:|:| : : | | | :| | : | : :| :| |  
MK--LYVQMLLALLALL--AFVAGEMYVTDKVDPPFPDDESFYDGCLEGTNPHVW---C

best alignment n: 5  
names: 018417 - [AGAP012681](#)  
identities: 17/63 (26%), positives: 31/63 (49%)  
score: 13  
MK--TLSVFLVLVCLLGLVQSWEPWNRQPTRYPIPSNPRDK----WCRNLNLP--AWGGRC  
|| : :|:|:|:|:| : | | | :| | : | : :| :| |  
MKFYVQMLLALLALLAFVAGE----MYVTDKVDPPFPDDESFYNGCLEGTNPHVW---C

best alignment n: 6  
names: 018417 - [AGAP003130](#)  
identities: 20/70 (28%), positives: 33/70 (47%)  
score: 1.150000e+001  
MKTLSVFLVLVCLLGLVQSW---EWP--WNRQPTRYPI---PSP----NP---RDKWCRNLNLPWGGRC

```

|| :|||:| | ||: | | :| :| :| | | :| :|
MKLFQVFLIL-C--GLIASTLAVPGPLLLKVKVIVKQPIFVKPAPVIVHRPIVVR-KFA--SRGGWW-----

```

---

**D. SUBOBSCURA**

```

best alignment n: 1
names: O62539* (Acp70A) - ENSANGP00000030793 pep:novel
chromosome:AgamP3:3R:17743980:17751126:1 gene:ENSANGG00000026739
transcript:ENSANGT00000031261
identities: 13/55 (23%), positives: 30/55 (54%)
score: 15
MMVPIISMLLLLLLVGVALGMPNPMPPARKSSTWGPRDIQKWCRNLNFGPAWGGRAC-
| | :| | |:| : : |:| |::| | :| | :| | :|
KSVPRCGVVLCLIVKLLKHFYSPVPAQRA--WA-KDFDT---LRADERWN--SCS

```

---

```

best alignment n: 2
names: O62539* (Acp70A) - AGAP011482
identities: 20/73 (27%), positives: 32/73 (43%)
score: 1.350000e+001
M--MVPISMLLL--LLLVGVALGMPNPMPPARKSSTWGPRDIQKW-----CRLN--FGPAWG-----GRAC
| : |:| | | | | :| | | :| | | :| | :| | :| | :| | :| |
MRFLSYINLMVLGVLALVLSASAGPCGCPRSYRPFVCGT-DLKTYSNQCVLDCRINSNYGRKFGLKLLRDGH-C

```

---

```

best alignment n: 3
names: O62539* (Acp70A) - AGAP009352
identities: 16/66 (24%), positives: 30/66 (45%)
score: 1.150000e+001
MMVPIISMLLLLLLVGVA-LGMP-----NPMPPARKSSTWGPRDIQKW--CRLNFGP-AWGGRAC
| : ::| | | | | :| | :| | | :| :| | :| :| |
MKLYVQMMLLALLALLAFVVGAGEMYVTDKVDPPP-----FDPDESFFYDGCLEGTNPHVW----C

```

---

```

best alignment n: 4
names: O62539* (Acp70A) - AGAP012807
identities: 14/62 (22%), positives: 30/62 (48%)
score: 1.150000e+001
MMVPIISMLLLLLLVGV-----ALGMPNPMPPARKSSTWGPRDIQKW-CRLNFGPAW--GGRAC
| : : |:| | | | :| :| | :| :| | :| | :| | |
MKLLFHLLLVIVLIVATCFVTCEAVTAPSQVVKREAM----RQCVMWEGKLMCPTGWIPGG---

```

---

```

best alignment n: 5
names: O62539* - AGAP012681
identities: 16/66 (24%), positives: 29/66 (43%)
score: 11
MMVPIISMLLLLLLVGVA-LGMP-----NPMPPARKSSTWGPRD--IQKWCRNLNFGP-AWGGRAC
| : ::| | | | | :| :| | :| | | :| :| | :| :| |
MKFYVQMMLLALLALLAFVVGAGEMYVTDKVDPPP-----FDPDESFFYNGCLEGTNPHVW----C

```

---

**REVERSE BLAST**

```

best alignment n: 1
names: AGAP008968-PA - CG31704-PA type=protein; loc=2L:11840712..11840915;
name=CG31704-PA;
dbxref=FlyBase:FBpp0079857,GB_protein:AAN10787.1,FlyBase_Annotation_IDs:CG31704-

```

PA, FlyBase:FBgn0051704; MD5=aa076d77745f58cc4ddc450bf49be164; parent=FBtr0080273; release=r4.3; species=Dmel; length=68; identities: 31/73 (42%), positives: 50/73 (68%) score: 69  
MKLLFLLSFLLCAILA-AAG---KYSCPACPANYLPVCGTDGKTYANECALLECTVAP--AVKVARSGEGCIKG  
|: | ::| | | :|| | :| :||| || || ||||:|: ||:|:|:|:| : :: | ::|:|  
MRCLAFIAFCLSALLAMAVGQVFQYSCP-CPRNYDPVCGSDSVTYSNQCVLDCLIKEGRSITVEKKGRC----

NEVER RECOGNIZES ACP70A, ALSO CHANGING MATRICES AND GAP. SEE BELOW FOR PHYLOGENETIC TREE

-----  
-----

best alignment n: 1  
names: [AGAP003130-PA](#) - CG15065-PA type=protein; loc=2R:13904852..13905037; name=CG15065-PA; dbxref=FlyBase:FBpp0085848, GB\_protein:AAF57705.1, FlyBase\_Annotation\_IDs:CG15065-PA, FlyBase:FBgn0040734; MD5=539fc5367b559235070d259f4dfa0e50; parent=FBtr0086667; release=r4.3; species=Dmel; length=40; identities: 18/60 (30%), positives: 30/60 (50%) score: 11  
MKLFQVFLILCGLIASTLAVPGP LLLKKVIVKQPIFVKPAPVIVHRP IIVVRKFASRGGWW  
|| :: ::|||:| :|| :|| :|: ||:: | | ||  
MK-WMSLVFLCGLLA--MAVASPL-----NPGNVIINGDC--RHCNVRGG--

-----  
-----

best alignment n: 7  
names: [AGAP003130-PA](#) - Acp70A-PA type=protein; loc=3L:13275585..13275814; name=Acp70A-PA; dbxref=FlyBase:FBpp0075572, GB\_protein:AAF49836.1, FlyBase\_Annotation\_IDs:CG17673-PA, FlyBase:FBgn0003034; MD5=a4c18f4fec6a166ea6017651e8d64171; parent=FBtr0075836; release=r4.3; species=Dmel; length=55; identities: 15/60 (25%), positives: 28/60 (46%) score: 6  
MKLFQVFLILCGLIASTLAVPGP LLLKKVIVKQPIFVKPAPVIVHRP IIVVRKFASRGGWW  
|| : :||:| :: : | | : | || | :| : : : | |  
MKTLALFLVLVLCVGLGLVQAWWEPWNRKPT--KFPI---PSPNPRDKWCRLNLGPAWGGRC

-----  
-----

best alignment n: 1  
names: [AGAP009352-PA](#) - CG16836-PA type=protein; loc=2R:13903963..13904256; name=CG16836-PA; dbxref=FlyBase:FBpp0085847, GB\_protein:AAF57706.1, FlyBase\_Annotation\_IDs:CG16836-PA, FlyBase:FBgn0040735; MD5=152fabfb86c1303a129e69ca92445267; parent=FBtr0086666; release=r4.3; species=Dmel; length=76; identities: 24/76 (31%), positives: 39/76 (51%) score: 24  
MK-LYVQMMLLALLALLAFVGA--GEMVYTDK-VD---PPPFDP-----DDESF----Y---DGCLE-GTNPHVWC  
|| | | | : | :||| | :| | :||:| :| | | | | :| | :|:| | | :|  
MKALQVAGTLMMLFCLLAAVNATPGQVYINGKCIDCNKPDNDPGIIPPDHKSAGSMSYTLTSGAIFFGIYHIFS

best alignment n: 5  
names: [AGAP009352-PA](#) - Acp70A-PA type=protein; loc=3L:13275585..13275814; name=Acp70A-PA; dbxref=FlyBase:FBpp0075572, GB\_protein:AAF49836.1, FlyBase\_Annotation\_IDs:CG17673-PA, FlyBase:FBgn0003034; MD5=a4c18f4fec6a166ea6017651e8d64171; parent=FBtr0075836; release=r4.3; species=Dmel; length=55; identities: 17/63 (26%), positives: 32/63 (50%) score: 1.450000e+001  
MKLYVQMMLLALLALLAFVGAEMVYTDKVD----PPPFDPDDESFYDGCLEGTNPHVW---C  
|| : :|:|:|:|:| | | : | | | :| | : | :| :| | :| :| |  
MKTLA--LFLVLVLCVGLGLVQAWWEPWNRKPTKFP IPSP-NPRDK----WCRLNLGP-AWGGRC

best alignment n: 1  
 names: **AGAP012682** - CG15353-PA type=protein; loc=2L:complement(2006907..2007146);  
 name=CG15353-PA;  
 dbxref=FlyBase:FBpp0077532,GB\_protein:AAF51316.1,FlyBase\_Annotation\_IDs:CG15353-  
 PA,FlyBase:FBgn0040718; MD5=8ac71b2e7d00b6d4c2dabf24d700e6c7; parent=FBtr0077863;  
 release=r4.3; species=Dmel; length=53;  
 identities: 16/53 (30%), positives: 23/53 (43%)  
 score: 15  
 MKLYVQMMLLALLLVAFAVGA-----REMPVTEKVDPPPFEPDGYIHPTDYGI  
 | | :| |||:| | : || : | : || | |  
 MLKLVLLLLLAALLAVAMAVPAPGSPNPAPVPQFVYSAGYPAVGYASPYVYYG

NEVER RECOGNIZE ACP70A, ALSO CHANGING MATRICES AND GAP. SEE BELOW FOR PHYLOGENETIC TREE

best alignment n: 1  
 names: **AGAP009357-PA** - Dro-PA type=protein; loc=2R:10260853..10261044; name=Dro-PA;  
 dbxref=FlyBase:FBpp0086566,GB\_protein:AAF58216.1,FlyBase\_Annotation\_IDs:CG10816-  
 PA,FlyBase:FBgn0010388; MD5=1ef7557d7cd0ac3cb9d691e361196bbc; parent=FBtr0087436;  
 release=r4.3; species=Dmel; length=64;  
 identities: 17/64 (26%), positives: 29/64 (45%)  
 score: 1.750000e+001  
 MKFFVQVFLFALLALLGLVSAGK-MAINPATTTTARPIFF---PCFDEDGIF-IGCTPHP----  
 ||| : :| : :|::| | :| | : || | : : | |  
 MKFTIVFLLLACVFAMGVATPGKPRPYSPRPTSHPRP IRVRREALAIEDHLTQAAIRPPPILPA

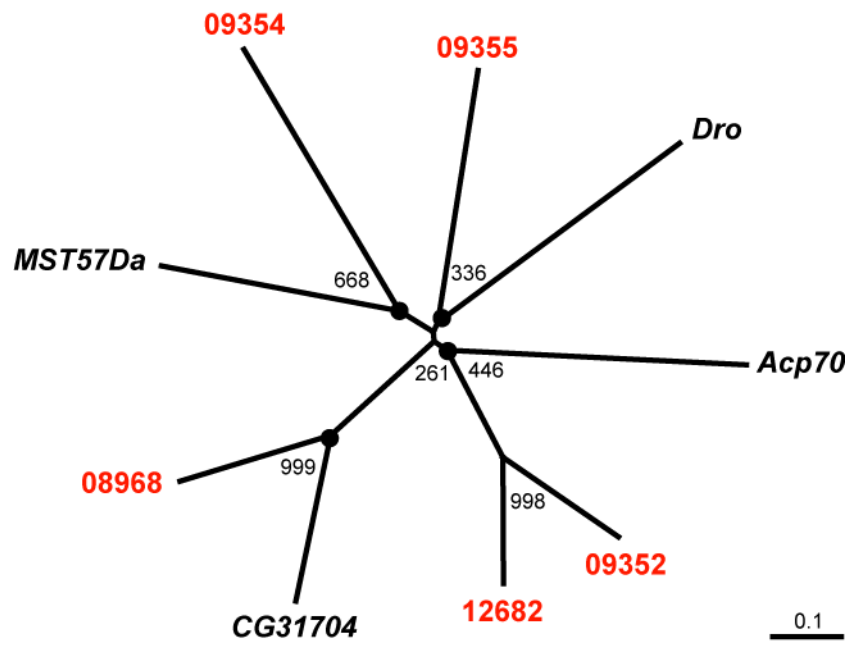
NEVER RECOGNIZE ACP70A, ALSO CHANGING MATRICES AND GAP. SEE BELOW FOR PHYLOGENETIC TREE

**ALIGNMENT BETWEEN AGAP009352-PA, AGAP012681-PA AND AGAP012682-PA**

best alignment n: 1  
 names: **AGAP009352-PA** - **AGAP012681-PA**  
 identities: 54/56 (96%), positives: 56/56 (100%)  
 score: 150  
 MKLYVQMMLLALLALLAFVVGAGEMYVTDKVDPPPFDPDDESFYDGCLEGTNPHVWC  
 ||:|||||||||||||||||||||||||||||||||||||:|||||||||||||  
 MKFYVQMMLLALLALLAFVVGAGEMYVTDKVDPPPFDPDDESFYNGCLEGTNPHVWC

best alignment n: 2  
 names: **AGAP009352-PA** - **AGAP012682-PA**  
 identities: 35/56 (62%), positives: 42/56 (75%)  
 score: 74  
 MKLYVQMMLLALLALLAFVVGAGEMYVTDKVDPPPFDPDDESFYDGCLEGTNPHVWC  
 |||||:|||||:||||| || ||:|||||:|| | :| : :  
 MKLYVQMMLLALLLVAFAVGAAREMPVTEKVDPPPFEPD-----GYIHPTDYGI--

**Phylogenetic relationships for Mst57Da, Dro, Acp70A, CG31704 and their Anopheles homologues.**



1. Thompson JD, Higgins DG, Gibson TJ (1994) *Nucleic Acids Res* 22:4673-4680.