Identification and characterization of accessory gland proteins in the *Glossina* genus as a conceivable vector control strategy in tsetse flies

Muna Fuad Abry,BSc.
Reg.No. I56/68246/2013

College of Biological and Physical Sciences
Centre for Biotechnology and Bioinformatics
University of Nairobi

A thesis submitted in partial fulfillment for the requirement of the degree of
Master of Science in Bioinformatics

November 2015

DECLARATION AND APPROVAL

DEDICATION

To my parents, Mrs. Majidah Taib Bajaber and Mr. Fuad Abry Al-Nahdi, my husband Abubakar Bajaber and my two daughters, Gamar and Aaliyah Bajaber.

ACKNOWLEDGEMENT

Firstly, I would like to thank the Almighty God for His grace and blessings that have allowed me to successfully finish this project, my family for the encouragement and the financial support that has pulled me through, I cannot overstate my gratitude.

Secondly, I would like to thank my supervisors, Dr. Benard Kulohoma for his invaluable teachings, and Dr. Daniel Masiga for the words of scientific wisdom. May the Almighty bless you both with all the goodness of this world and the hereafter.

I would also like to acknowledge my friends and colleagues for the support and suggestions they have put in to pull this project through. Special thanks goes out to Kelvin Muteru for his technical ingenuity and assistance, Anne Owiti and Sharon Towett for being strong and supportive friends when times were rough, my parents-in law Fatma Ali Bajaber and Ahmed Abubakar Bajaber for taking good care of my kids, I cannot thank you all enough.

Finally, many thanks to the staff of the Center for Biotechnology and Bioinformatics for their support and co-operation throughout my study there. Professor James Ochanda for believing in me and Dr. George Obiero for always being positive and pushing me forward, thank you.

ABSTRACT

Accessory gland proteins (ACPs) are reproductive proteins produced by the male accessory glands (MAGs) of most insect species. These proteins are essential for male fertility and play a major role in regulating female reproductive physiology and oviposition in insects like *Drosophila melanogaster* and *Anopheles gambiae*. ACPs therefore present attractive potential vector control entry points. Currently, there is limited information on the identity and organization of ACPs in the recently sequenced *Glossina* genus genomes. This study aimed to identify for the first time the presence of ACPs in the five publicly available genomes of the *Glossina* genus specifically: *G.austeni, G.brevipalpis, G.fuscipes, G.morsitans* and *G.pallidipes* insects. The availability of the *Glossina* genomes in VectorBase was exploited to explore the presence and diversity of ACPs. Orthologous genes in multiple *Glossina* species genomes were identified and their genetic diversity assessed using phylogenetic approaches. ACPs that are well characterized in *Drosophila melanogaster* and *Anopheles gambiae* genomes were used as reference sequences in this analysis. A total of 41 homologous clusters of *Drosophila melanogaster, Anopheles gambiae* and *Glossina* ACPs were identified. Amongst the 41 clusters, 12 clusters were composed of *Drosophila melanogaster, Anopheles gambiae* and *Glossina* orthologs, 7 clusters were exclusively *Anopheles gambiae* and *Glossina* orthologous ACPs while 5 clusters were exclusively *Drosophila melanogaster* and *Glossina* orthologous ACPs. The other 17 clusters were exclusive to either *Drosophila melanogaster* or *Anopheles gambiae*. These findings suggest that most ACPs are conserved in the order *Diptera*, despite its evolutionary radiation into multiple insect species. Further genetic characterization of *Glossina* ACPs will highlight novel strategies for field vector control.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ACPs	Accessory gland proteins
A.gambiae	*Anopheles gambiae*
CAP	Adenylate cyclase associated protein
D.melanogaster	*Drosophila melanogaster*
G.austeni	*Glossina austeni*
G.brevipalpis	*Glossina brevipalpis*
G.fuscipes	*Glossina fuscipes*
G.morsitans	*Glossina morsitans*
G.pallidipes	*Glossina pallidipes*
HAT	Human African Trypanosomiasis
Hsf	Heat shock factor protein
MAG	Male accessory glands
ML	Maximum likelihood
MUSCLE	Multiple sequence comparison by log expectation
SNAP	Synonymous and non-synonymous analysis program
WHO	World health organization

CHAPTER 1

1.0 INTRODUCTION

1.1 Background information

Accessory gland proteins (ACPs) are the major reproductive proteins produced by the male accessory glands (MAGs) of most insect species. These proteins are essential for male fertility and they trigger significant physiological and behavioral changes on the female upon copulation, thus performing fundamental roles in reproduction. They also perform important biological roles as chaperones, lipases and redox proteins (Dottorini et al., 2007). The ACP gene transcription is partly regulated by a heat shock factor (Hsf) protein and Hsf silencing leads to a significant reduction of ACP transcripts (Dottorini, Persampieri, Palladino, Spaccapelo, & Crisanti, 2012). ACPs and other seminal fluid proteins are thought to be under similar stronger or more frequent directional selection which is assumed to be controlled by three selective forces (Mueller et al., 2005). These forces are: female sperm preference (Eberhard & Cordero, 1995) sperm competition (Clark, Aguadé, Prout, Harshman, & Langley, 1995)and sexual conflict (Rice, 1996). Some ACPs have no functional domains with a signal sequence being the primary functional element making it capable for ACP to emerge from small open reading frames present in ancestrally non-coding sequences (Begun, Lindfors, Thompson, & Holloway, 2006).

1.2 The role of ACPs in reproduction

In *Drosophila* and *Anopheles* species, such reproductive proteins are known to be important for female physiology regulation. Experiments focused on normal and forced copulation in *Anopheles gambiae* mosquitoes show (Dottorini et al., 2007; Walker et al., 2006) constituents of MAGs are important to induce refractoriness to subsequent inseminations in females and to stimulate ovulation and oviposition (Dottorini et al., 2007; Tripet, Touré, Dolo, & Lanzaro, 2003; Wolfner, 2002). Females copulated by males with degenerate testes and accessory glands fail to oviposit and readily re-mate (Dottorini et al., 2007). This is in contrast to females mated to males with degenerate testes but fully developed accessory glands who laid unfertilized eggs and did not re-mate (Dottorini et al., 2007). ACPs from male species

influences the female's way of survival and hence the rapid evolution of ACPs (Wolfner, 2002) Interestingly, the male *Drosophila* replenishes his ACPs by resynthesis, which only occurs after the transfer of seminal fluid to the females. Although topical application of juvenile hormone on the male's cuticles also causes in vivo synthesis of ACPs to levels similar to those present before mating (Wolfner et al., 1997). Besides triggering egg laying (Soller, Bownes, & Kubli, 1997) and reduced sexual receptivity, MAGs secretions tend to induce the expression of immune peptides and reduction of female lifespan rendering them important candidates for the biological and genetic control of insect pests (Tracey Chapman & Davies, 2004). Currently, ACPs are not well characterized in the *Glossina* genome compared to *Drosophila melanogaster* and *Anopheles gambiae* and a need for this new knowledge is important to come up with an effective strategy for tsetse control.

1.3 Tsetse Biology

The tsetse fly, *Glossina* species, is the principal vector of the parasite that causes life-threatening human African trypanosomiasis (HAT) and cattle nagana in endemic areas (Mpanya et al., 2012). This trypanosome parasite puts over 60 million people and 80 million cattle in sub-Saharan Africa at risk of contracting disease (Grady, Messina, & McCord, 2011) and agricultural losses amounting to an estimated US$4.75 billion annually (Aksoy, 2010). Tsetses differ from other insects in their blood feeding behavior and reproducing a living young rather than laying eggs (Gooding & Krafsur, 2005; Krafsur, 2009). Both female and male flies feed on blood and are capable of transmitting infection(Franco, Simarro, Diarra, & Jannin, 2014).

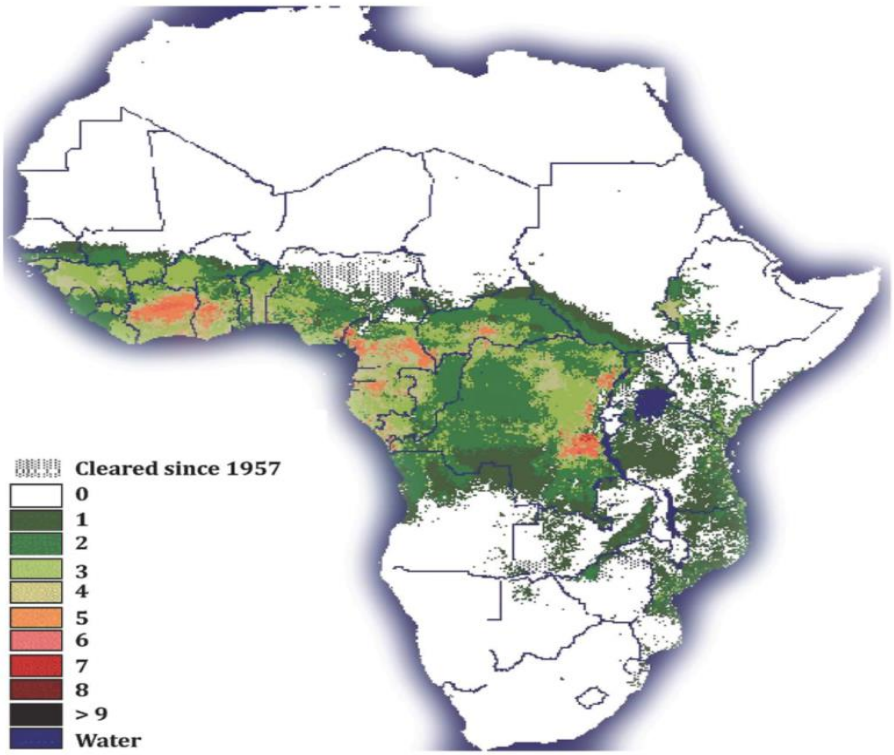


Figure 1: A map of Africa showing the distribution of Glossina in sub-Saharan Africa
The colours in the figure legend correspond to the colours in the figure and show the numbers of different tsetse species present in sub-Saharan African countries. The dots represent areas that have been cleared of *Glossina* since 1957. This image was adapted from (Kariithi et al., 2013)

1.4 Life cycle of the Tsetse fly

The female tsetse flies mate only once during the first few days of their life and store the sperm in their spermathecae which they subsequently use for a lifetime to self fertilise. The male flies on the other hand remain sexually active throughout life. The female flies do not lay eggs, but deliver a mature larva (ready to pupate) after a gestation of about ten days. During gestation, the larva is fed on a secretion in the mother's uterus (Gouteux & Jarry, 1998). This larviparity results in a relatively slow reproductive rhythm (one larva in every 9-11 days) and the pupal period of 25-55 days is similar for all the *Glossina* species. Since the average life-span of a tsetse fly varies between one and two months depending on the climatic conditions, each female will deliver on average between 3 and 6 larvae during her life (Gouteux & Jarry, 1998)(Figure 1). The peculiarity of the *Glossina* female's lifecycle is the absence of oviposition and the development of a single larva in its uterus (Franco et al., 2014). Thus successful tsetse reproduction results in a single larva rather than a mass of eggs and provides a unique vector control target point for trypanosomiasis management.

.

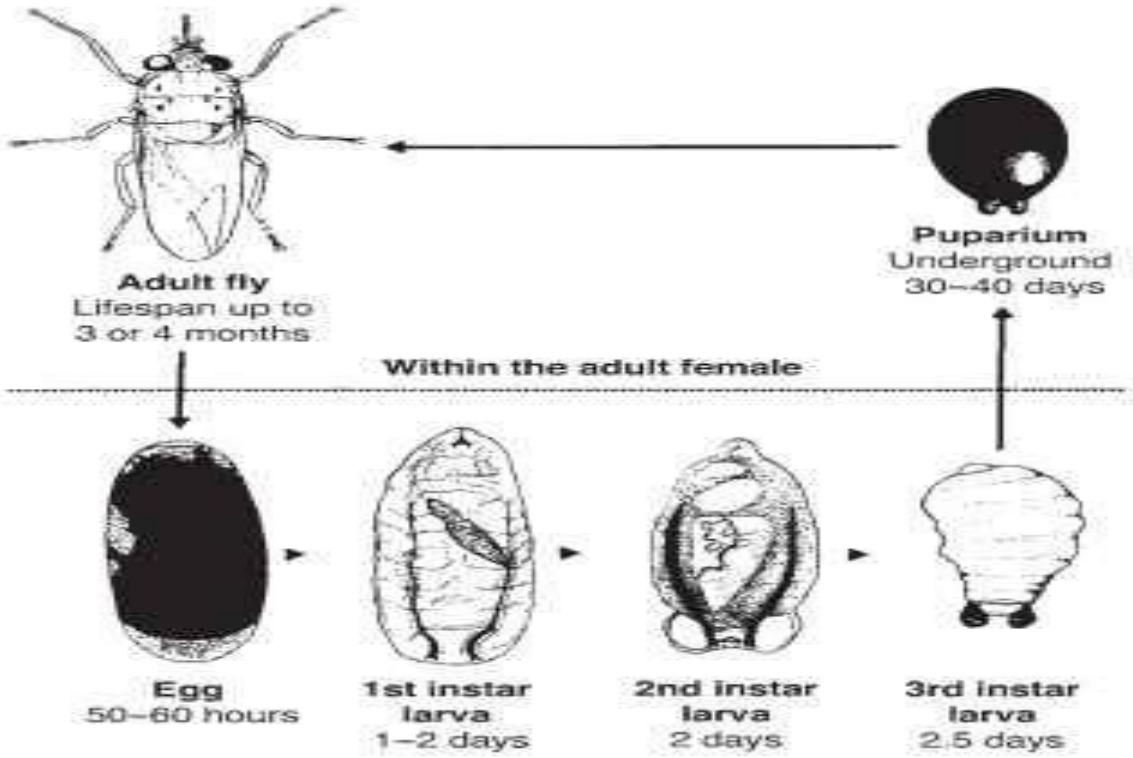


Figure 2: Life cycle of the tsetse fly
The adult female fly does not lay eggs but delivers a mature larva that begins as an egg which develops into a 1st instar larva for 1-2 days then becomes a 2nd instar level for 2 days and finally a 3rd instar larva for 2.5 days where it gets buried as a puparium underground for 30-40 days. During gestation, the larva is fed on a secretion in the mother's uterus. This image was adapted from "Tsetse flies (insects), http://what-when-how.com/insects/tsetse-fly-insects/".

1.5 The effect of evolution on ACP genes function and species diversification

Genes that are involved in reproduction usually show signs of adaptive evolution. Comparisons between *Drosophila simulans* and *Drosophila melanogaster* orthologs suggest that ACPs on average have two times more replacement substitutions than non- ACP genes (Almeida & Desalle, 2009; Swanson, Wong, Wolfner, & Aquadro, 2004). There is strong homology in the molecular structure of the ACP peptides between closely related species and an increase in divergence as the phylogenetic distance increases (Chen, 1996). This rapid evolution rate of ACP genes has made identification of orthologs in other insects challenging in the absence of genomic data (Collins, Caperna, Williams, Garrett, & Evans, 2006; Davies & Chapman, 2006). Furthermore, ACPs evolving more rapidly than other proteins is consistent with the notion that genes coding for male reproductive functions are rich in

lineage-restricted genes (Begun et al., 2006).

Currently, 173 ACPs have been identified in *Drosophila melanogaster* and 56 of these are also present in *Anopheles gambiae* (Dottorini et al., 2007). The importance of the tsetse fly to human and cattle health and the potential for the control displayed by manipulation of its reproduction cycle highlights the importance of characterizing ACP gene orthologs as potential intervention target sites.

1.6 Tsetse fly control and eradication

Tsetse flies have two characteristics that make them suitable for eradication, their low reproduction rate and their limited ability to rebound in areas where their populations have been reduced (Kariithi et al., 2013). In most cases minimizing the number of vectors or minimizing contact between man and fly or cattle and fly is the fastest way of controlling the disease (Hocking, Lamerton, & Lewis, 1963). This is achieved by partial removal of bush cover and frequent surveys of fly infestations since the complete removal is time, labour and money intensive. Other control methods include insecticide application (Allsopp, 2001), insecticide impregnated targets and live bait technologies.

1.7 Aim

This study aimed to examine the presence of ACP genes in five species of the *Glossina* genus.

1.8 Hypothesis

ACP genes are present and there is a low level of genetic diversity at the ACP gene locus in the different *Glossina* species.

1.9 Objectives

To identify the presence of ACP genes in the available *G.austeni, G.brevipalpis, G.fuscipes, G.morsitans* and *G.pallidipes* genomes.

To examine the genetic diversity between the 5 *Glossina* ACPs' homologs to those identified in the *D.melanogaster* and *A.gambiae* genomes.

1.10 Justification

A vector control strategy for the tsetse fly is essential for human and livestock health as well as to agricultural productivity. The interactions between the trypanosome parasite and the tsetse fly vector as well as with the human and animal hosts within a particular environment determines the disease's epidemiology (Franco et al., 2014). The WHO has a target set of eliminating the HAT disease as a public health problem by 2020 ("Control and surveillance of human African trypanosomiasis.," 2013). Trypanosomiasis control strategies targeted towards provoking sexual sterility in natural vector populations are an attractive alternative to the use of insecticides. Potential targets for this novel control strategy are the Accessory gland protein (ACP) genes that are produced by the male accessory glands (MAGs) in most insect species (Almeida & Desalle, 2009).

The ACP gene is an attractive candidate as it plays a significant role in insect reproduction including the tsetse fly. The presence of a single larva and absence of a mass of eggs in the tsetse fly's reproductive cycle is an advantage in the manipulation of its life cycle.

Thus the characterization of ACP gene orthologs in the tsetse's genome will not only pave way towards better understanding of the tsetse fly's genetic diversity but also display potential intervention target sites. This will help provide new avenues to develop control measures targeted towards disease elimination.

CHAPTER 2

2.0 MATERIALS AND METHODS

2.1 Identification of *Glossina* homologs of *Anopheles* and *Drosophila* ACP genes

The complete proteomes of *G. austeni, G. morsitans, G. pallidipes, G.fuscipes* and *G. brevipalpis* were retrieved manually from VectorBase (www.vectorbase.org). The retrieved *Glossina* protein sequences alongside (Dottorini et al., 2007)'s ACP reference sequences from *A.gambiae* (n=56) and *D.melanogaster* (n=173) were subjected to OrthoMCL clustering analysis (Li, Stoeckert, & Roos, 2003) that uses a Markov Cluster algorithm to group orthologs and paralogs. *Glossina* sequences that clustered together with *Anopheles gambiae* and *Drosophila melanogaster* ACPs were grouped into a multifasta file. This multifasta file containing *Glossina*, *Anopheles gambiae* and *Drosophila melanogaster* ACP protein orthologs was cleaned in Unix to obtain individual *Glossina*, *Anopheles gambiae* and *Drosophila melanogaster* ACP clusters that were saved in a folder.

2.2 Identification of ACP domains

The ACP functional domains were identified by performing similarity search queries against the Pfam database (Finn et al., 2014). Individual *Glossina* ACP protein sequences were then copied from the OrthoMCL clusters against the Pfam website (http://pfam.xfam.org/) to view the domain organization of the ACP protein sequence. This was also done for the *Anopheles gambiae* and *Drosophila melanogaster* ACP proteins from the same clusters. The identity of the conserved domain on the *Glossina* ACP was confirmed on being the same as the conserved domain on the *Anopheles gambiae* and *Drosophila melanogaster* ACP sequences that had clustered together according to the OrthoMCL results.

2.3 Sequence alignment and phylogeny reconstruction

Multiple sequence alignments were performed on the OrthoMCL clusters of *Glossina*, *Anopheles gambiae* and *Drosophila melanogaster* ACP protein sequences using Multiple Sequence Comparison by Log Expectation (MUSCLE) (Edgar, 2004). No penalties were assigned for opening or extending a gap during the sequence alignment. Each run of the

application resulted in a list of potential ACP amino acid sequences in the *Glossina* genus genome that were subjected to reciprocal best hit analysis against the *A.gambiae* and *D.melanogaster* genomes using the same parameters used in the initial similarity searches. The maximum-likelihood (ML) phylogenetic trees of the 41 multiple aligned Acp protein sequences with 100 bootstrap replicates were reconstructed using PhyML (version 3.5) (Dereeper et al., 2008).

2.4 Determining the direction and extent of selection pressure

The direction and extent of selection pressure on the ACPs sequences was tested based on the ratio (ω) of non-synonymous to synonymous rates (ω = dN/dS), where dN is the average number of non-synonymous substitutions per non-synonymous sites and dS is the average number of synonymous substitutions per synonymous site. If ω = 1, amino acid substitution are assumed to be largely neutral, ω > 1 indicates positive selection whereas ω < 1 indicates negative or purifying selection. The nucleotide bases of the converted OrthoMCL ACP protein clusters of *A.gambiae*, *D.melanogaster*, *G.austeni*, *G.brevipalpis*, *G.fuscipes*, *G.morsitans* and *G. pallidipes* were uploaded on the Synonymous Non-synonymous Analysis Program (SNAP) (www.hiv.lanl.gov) (Korber, Allen G, & Gerald H, 2000) and the dN/dS value was calculated using default parameters.

2.5 Visualising data using CIRCOS softwares

The *Glossina austeni*, *Glossina brevipalpis*, *Glossina fuscipes*, *Glossina morsitans*, *Glossina pallidipes*, *Anopheles gambiae* and *Drosophila melanogaster* OrthoMCL ACP clusters identified together with their respective functions and quantities in percentages were visually represented in a heatmap using CIRCOS software (Krzywinski et al., 2009). To create the circular ideogram, data from Table 1 was arranged in GFF style and uploaded to the CIRCOS online utility. The created output image in PNG was saved.

CHAPTER 3

3.0 RESULTS

3.1 Identification of ACP genes in *Glossina*

This study aimed to identify accessory gland proteins (ACPs) that are vital for reproduction in tsetse flies. The five recently released *Glossina* genomes were analysed to identify orthologs that were homologous to those previously identified in *Anopheles gambiae* (n=56) and *Drosophila melanogaster (*n=173) ACP orthologs. OrthoMCL searches identified a total of 41 clusters (Table 1).Amongst the identified orthologous groups, 12 homologous ACPs were present in *Anopheles gambiae*, *Drosophila melanogaster* and *Glossina* species. Seven ACPs were present in both *Anopheles gambiae* and *Glossina* species and 5 were present in *Drosophila melanogaster* and *Glossina* species giving a total of 24 ACPs identified in the 5 *Glossina* species. The remaining 17 clusters were individual *Anopheles* and *Drosophila* ACPs that did not have any orthologs in other species under observation. Specifically, 7 were individual *Anopheles gambiae* clusters and 10 were individual *Drosophila melanogaster* clusters (Table.1).

To further analyse the ACPs clusters, we organized the proteins into groups according to their functional class. Groups of 23 functional classes were obtained (Table 1) with macroglobulin (Group 1) and heat shock protein (Group 19) having the most number of *Glossina* species homologs. Macroglobulins are immunity related proteins that help protect the male ejaculate and female reproductive tract from infections. The heat shock protein (Hsf) plays a role in ACP gene regulation and its silencing leads to a down regulation of a significant amount of ACP genes and thus a reduction in insect progeny (Dottorini et al., 2013). Group 2 were cell adhesion proteins that help in the cell to cell and cell to extracellular matrix binding process such as when forming the mucous plug that carries the sperm during mating. Group 3 proteins included the carboxylesterase which are enzymes that catalyze carbohydrate metabolic reactions. Group 4 consisted of 2 chaperones detected in significant levels. These proteins facilitate in the folding of other proteins and sperm-egg interactions (Baldini, Gabrieli, Rogers, & Catteruccia, 2012). Group 5 proteins were the cytochromes that play a role in regulating cell metabolism and respiration. Group 6 included a protease, an enzyme

that accelerates the degradation of other proteins. Amongst the 7 protease inhibitors of Group 7, only AGAP005246 and serpin formed orthologous clusters with *Glossina* species. Group 8 only had isoform B, a possible product of gene duplication that is bound to increase the diversity of its proteome. Group 9 had 2 isomerase proteins that act as enzymes by catalyzing the process of isomerization. Group 10 consisted of adenylate cyclase associated protein (CAP) that is an actin binding protein. Group 11 included the redox proteins that play a role in cellular oxidation-reduction reactions. Group 12 consisted of 2 transport proteins that facilitate the movement of ions and biomolecules within an organism. Group 13 ribonucleases function in the degradation of RNA into smaller components. Group 14 included 2 lipases. These are enzymes that catalyze the hydrolysis of fats and play an important role in the transport of dietary lipids. Group 15 consisted of calnexin a chaperone that acts in protein folding. Group 16 beta-defensin proteins are antimicrobial proteins that prevent microbial colonization on the epithelial surfaces of cells. Group 17 consisted of proline oxidase, which initiates the proline cycle. Group 18 are 9 accessory proteins that play essential roles during mating and reproduction of most species. Group 19 consisted of the heat shock protein (Hsf). Groups 20, 21, 22, and 23 consisted of the sex peptide ACP 70A, the antibacterial component andropin, cysteine rich venom protein AGAP006583 and the embryonic pattern formation component, msopa proteins respectively. These proteins were not detected in the *Glossina* genus.

Group	Groups	new_labels	labels	clusters	G-austeni	G-brevipalpi	G-fuscipes	G-morsitans	G-pallidipes	A-gambiae	D-melanoga	Dn-ds
Group-1	Macroglobulin	A	AGAP008364	C14	6	7	5	3	6	1	3	1.07
Group-2	Cell adhesion	B	AGAP004428	C29	1	1	2	1	1	1	16	1.073
Group-3	Carboxylester	C	AGAP005370	C59	1	2	1	1	1	2	10	1.201
Group-4	Chaperone	D1	AGAP001424	C84	2	2	3	3	3	1	1	1.023
	Chaperone	D2	AGAP004212	C1152	1	1	1	1	1	1	0	0.818
Group-5	Cytochrome	E	AGAP009363	C106	2	5	2	1	2	1	1	1.118
Group-6	Protease	F	AGAP006610	C139	3	2	2	3	2	1	0	1.189
Group-7	Protease inhit	G1	AGAP005246	C225	1	1	1	1	1	5	1	0.951
	Protease inhit	G2	serpin-9	C1550	1	1	1	1	1	0	1	1.221
	Protease inhit	G3	Acp62F	C44	0	0	0	0	0	0	20	0.376
	Protease inhit	G4	Acp63F	C49	0	0	0	0	0	0	19	1.343
	Protease inhit	G5	AGAP006581	C10560	0	0	0	0	0	3	0	1.951
	Serpin protea	G6	AGAP007691	C10561	0	0	0	0	0	3	0	-
	Protease inhit	G7	CG31704	C12522	0	0	0	0	0	0	2	-
Group-8	IsoformB	H	CG6168	C284	2	1	2	2	2	0	1	1.103
Group-9	Isomerase	I1	AGAP008822	C429	1	1	1	1	1	1	3	1.152
	Isomerase	I2	AGAP007088	C654	1	1	1	0	2	1	2	1.12
Group-10	CAP	J	AGAP006418	C653	1	1	1	1	1	1	2	1.098
Group-11	Redox	K	CG4670	C655	1	1	1	1	1	1	2	1.198
Group-12	Transport pro	L1	CG5793	C657	1	1	1	1	1	0	3	1.337
	Transport pro	L2	AGAP009364	C1155	1	1	1	1	1	1	0	1.167
Group-13	Ribonuclease	M	AGAP009842	C868	1	1	1	1	1	1	1	1.221
Group-14	Lipase	N1	CG17097	C869	1	1	1	1	1	0	2	1.076
	Lipase	N2	AGAP003083	C1151	1	1	1	1	1	1	0	1.044
Group-15	Calnexin	O	AGAP005032	C1153	1	1	1	1	1	1	0	1.155
Group-16	Beta-Defensi	P	AGAP007049	C1154	1	0	1	1	0	1	3	1.411
Group-17	Proline oxida	Q	AGAP001271	C1156	1	1	1	1	1	1	0	1.082
Group 18	Accessory gla	R1	CG14770	C1157	1	1	1	1	1	0	1	0.946
	Accessory gla	R2	Acp29AB	C37	0	0	0	0	0	0	21	0.403
	Accessory gla	R3	Acp26a	C140	0	0	0	0	0	0	13	1.045
	Accessory gla	R4	Acp53Ea	C282	0	0	0	0	0	0	10	1.281
	Accessory gla	R5	AGAP009354	C10562	0	0	0	0	0	3	0	0.544
	Accessory gla	R6	AGAP009355	C12518	0	0	0	0	0	2	0	-
	Accessory gla	R7	AGAP012830	C12519	0	0	0	0	0	2	0	-
	Accessory gla	R8	AGAP012706	C12520	0	0	0	0	0	2	0	1.063
	Accessory gla	R9	Acp26Ab	C12521	0	0	0	0	0	0	2	-
Group 19	Heat shock pr	S	AGAP004192	C5	10	5	9	8	8	1	0	1.128
Group 20	Sex peptide	T	Acp70A	C283	0	0	0	0	0	0	10	0.547
Group 21	Antibacterial	U	Andropin	C656	0	0	0	0	0	0	8	0.261
Group 22	Cysteine rich	V	AGAP006583	C12517	0	0	0	0	0	2	0	-
Group 23	Embryonic pa	W	msopa	C16432	0	0	0	0	0	0	2	1.196
										41	160	

Table 1: List of ACPs clusters with functional groups and dN/dS values in *Anopheles gambiae*, *Drosophila melanogaster* and the 5 *Glossina* species

3.2 Orthology patterns between *Anopheles, Drosophila* and the five *Glossina* species ACPs are consistent to the insect's way of reproduction

Putative orthologs of Group 18 (Table 1 and 'R' in Figure 3) are important ACPs to *Drosophila melanogaster* (R3, R4, R9 and T) and *Anopheles gambiae* (R5, R6, R7 and R8) that were not detected in the *Glossina* species under examination. This suggests that such ACPs do not play an important role in the *Glossina* species' reproductive cycle. Two ACPs in this cluster: 53Ea (Figure 3 R4) and 26Ab (Figure 3 R9) function in sperm defense by preventing displacement of an already present ejaculate within the female fly by a second ejaculate. Their absence in the *Glossina* genome is thus consistent with the tsetse female insects' natural way of reproduction since they only mate once in their lifetime and use stored sperm to self-fertilise.

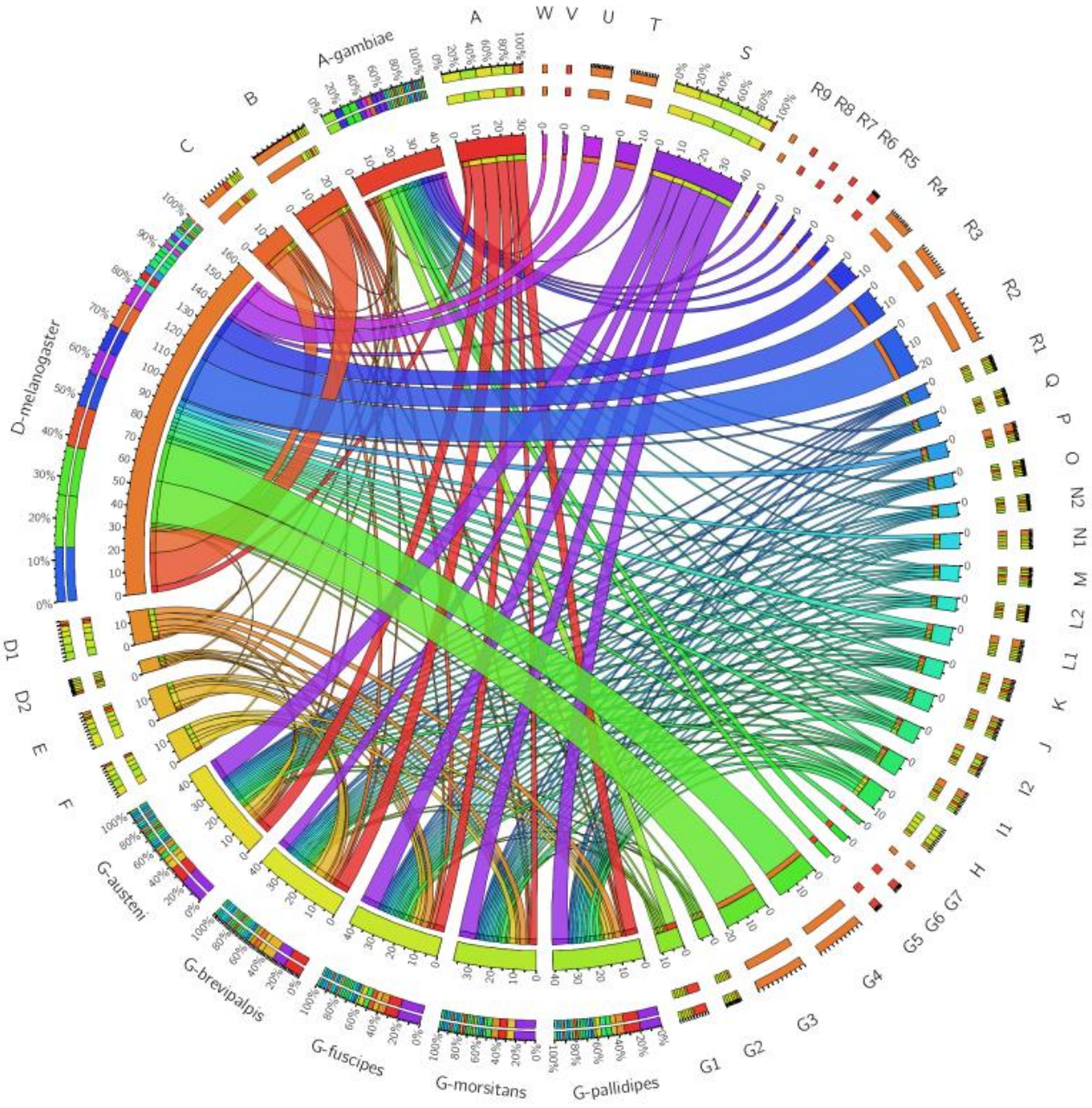


Figure 3: Heatmap showing relationship between *A.gambiae*, *D. melanogaster* , *G.austeni*, *G.brevipalpis*, *G.fuscipes*, *G.morsitans* and *G.pallidipes* ACPs
It highlights the 24 ACP clusters from *A.gambiae*, *D.melanogaster* and the 5 *Glossina*

species. The inner patches indicate the respective ACP functions while the outer patch shows the proteins' abundance in percentage within the respective species. The coloured ribbons indicate the number of ACP clusters obtained from *A.gambiae, D.melanogaster* and the 5 *Glossina* species. Same colour indicates same ACP function. **Key: A:** Macroglobulin **B:** Cell adhesion **C:** Carboxylesterase **D:** Chaperone **E:** Cytochrome **F:** Protease **G:** Protease inhibitor **H:** Isoform **I:** Isomerase **J:** CAP **K:** Redox **L:** Transport protein **M:** Ribonuclease **N:** Lipase **O:** Calnexin **P:** Beta- defensin **Q:** Proline oxidase **R:** Accessory gland protein **S:** Heat shock protein **T:** ACP70A/Sex peptide **U:** Antibacterial **V:** Cysteine rich venom protein **W:** Msopa

3.3 dN/dS values lie more towards positive selection

Majority of the ACPs contain a dN/dS value of slightly more than 1 with 27 clusters specifically, AGAP008364 (Fig.3 A), AGAP004428 (Fig.3 B), AGAP005370 (Fig.3 C) , AGAP001424 (Fig.3 D1), AGAP009363 (Fig.3 E), AGAP006610 (Fig.3 F), Serpin-9 (Fig.3 G2), ACP63F (Fig.3 G4), AGAP006581 (Fig.3 G5), CG6168 (Fig.3 H) , AGAP008822 (Fig.3 I1), AGAP007088 (Fig.3 I2), AGAP006418 (Fig.3 J), CG4670 (Fig.3 K), CG5793 (Fig.3 L1), AGAP009364 (Fig.3 L2), AGAP009842 (Fig.3 M), CG17097 (Fig.3 N1), AGAP003083 (Fig.3 N2), AGAP005032 (Fig.3 O), AGAP007049 (Fig.3 P), AGAP0012718 (Fig.3 Q), ACP26a (Fig.3 R3), ACP53Ea (Fig.3 R4), AGAP0012706 (Fig.3 R8), AGAP004192 (Fig.3 S) and msopa (Fig.3 W). Eight clusters had a dN/dS of less than 1. These were, AGAP004212 (Fig.3 D2), AGAP005246 (Fig.3 G1), ACP62F (Fig.3 G3), CG14770 (Fig.3 R1), ACP29AB (Fig.3 R2), AGAP009354 (Fig.3 R5), ACP70A (Fig.3 T) and andropin (Fig.3 U). The remaining 6 clusters, AGAP007691 (Fig.3 G6), CG31704 (Fig.3 G7), AGAP009355 (Fig.3 R6), AGAP012830 (Fig.3 R7), ACP26Ab (Fig.3 R9) and AGAP006583 (Fig.3 V) showed a negligible value. The *Anopheles gambiae* ACP, AGAP006581 a protease inhibitor (Table 1, G5 of Group 7) has the highest dN/dS of 1.95 and *Drosophila melanogaster* ACP, andropin (Table 1, U of Group 21), has the least dN/dS of 0.26 (Table 1). A value of > 1 is indicative of positive selection while a dN/dS of < 1 indicates negative or purifying selection. This suggests that most of the ACPs in the species are under positive selection and are hence diversifying while the few 8 (Fig.3 D2, G1, G3, R1, R2, R5, T, U) are under negative selection and hence have fewer variants.

3.4 Phylogenetic analysis of ACP sequences from *Anopheles, Drosophila* and the five *Glossina* species identifies *Glossina* ACPs as the most rapidly evolving amongst the three genera

Various ACPs such as the chaperone and calnexin (Figure 4 A and B), the redox and lipase (Figure 6 A and D) the transport protein and ribonuclease (Figure 7 B and C) indicate the *Anopheles gambiae* ACPs as having earlier copies of the chaperone, calnexin, redox, lipase, transport protein and ribonuclease ACPs compared to *G.austeni*, *G.brevipalpis*, *G.fuscipes*, *G.morsitans* and *G.pallidipes* which are more diverse. The Drosophila ACPs CG14770 and serpin (Figure 4 C and D), isoform B (Figure 5 B) and protease inhibitor (Figure 6 B) are ancestral to the respective ACPs of all five species of *Glossina* with the protease inhibitor (Figure 6 B) also bearing ancestry to the *Anopheles gambiae* ACP, AGAP005246. From the branching depth, observations of the ACPs belonging to the three genera implied that *Glossina* ACPs are the most rapidly evolving compared to the deeper branching of the *Drosophila melanogaster* and *Anopheles gambiae* ACPs that seem to be evolving more slowly. This may also indicate that the ACPs are most recent in the *Glossina* genus and are older in the *A.gambiae* and *D.melanogaster* species. As the branching distance increases the diversity also increases indicating a weaker homology in the molecular structure of the ACP between non-related species while a decrease in branching distance means less diversity and hence a stronger homology in the case of related species (Figures 8 and 9).

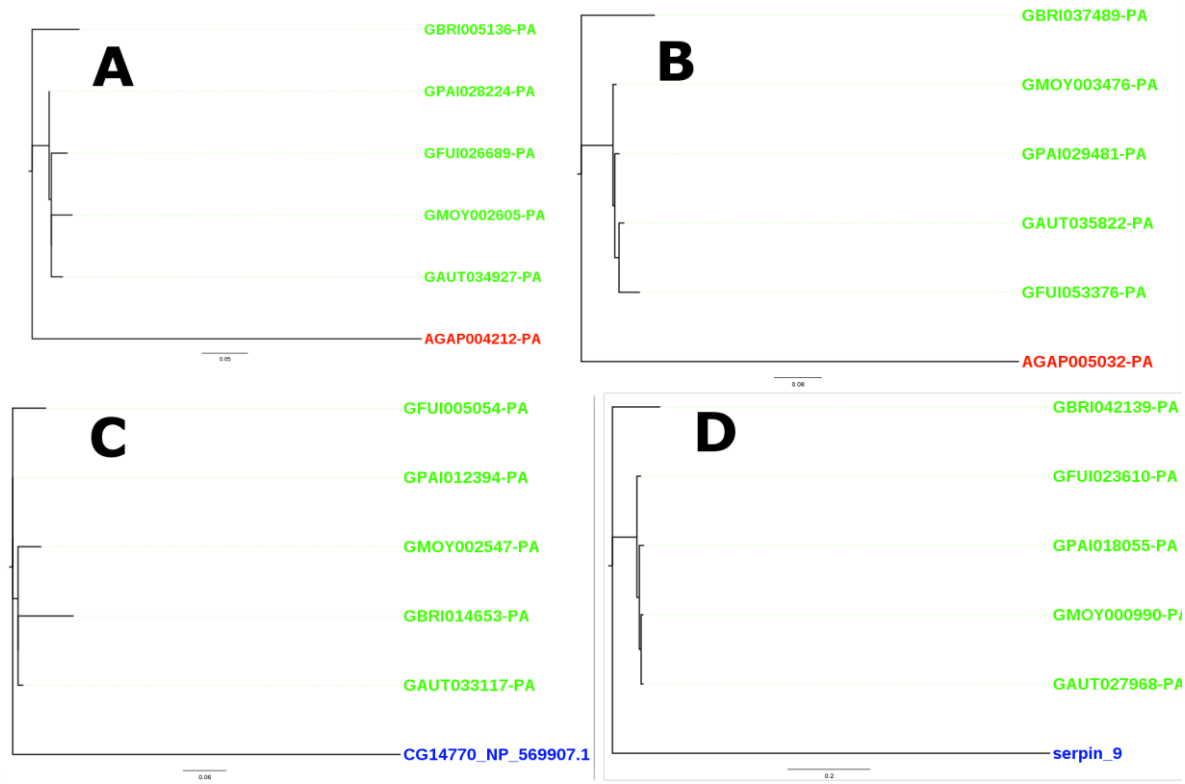


Figure 4: Phylogenetic trees showing evolution and adaptation of ACPs between *Anopheles gambiae*, *Drosophila melanogaster* and the 5 *Glossina* species

A: Chaperone ACP **B:** Calnexin ACP **C:** Accessory gland protein ACP **D:** Serpin ACP

Key:

Green: GAUT: *Glossina austeni*, GBRI: *Glossina brevipalpis*, GFUI: *Glossina fuscipes*, GMOY: *Glossina morsitans*, GPAI: *Glossina pallidipes*

Red: AGAP: *Anopheles gambiae*

Blue: CG14770, serpin 9: *Drosophila melanogaster*

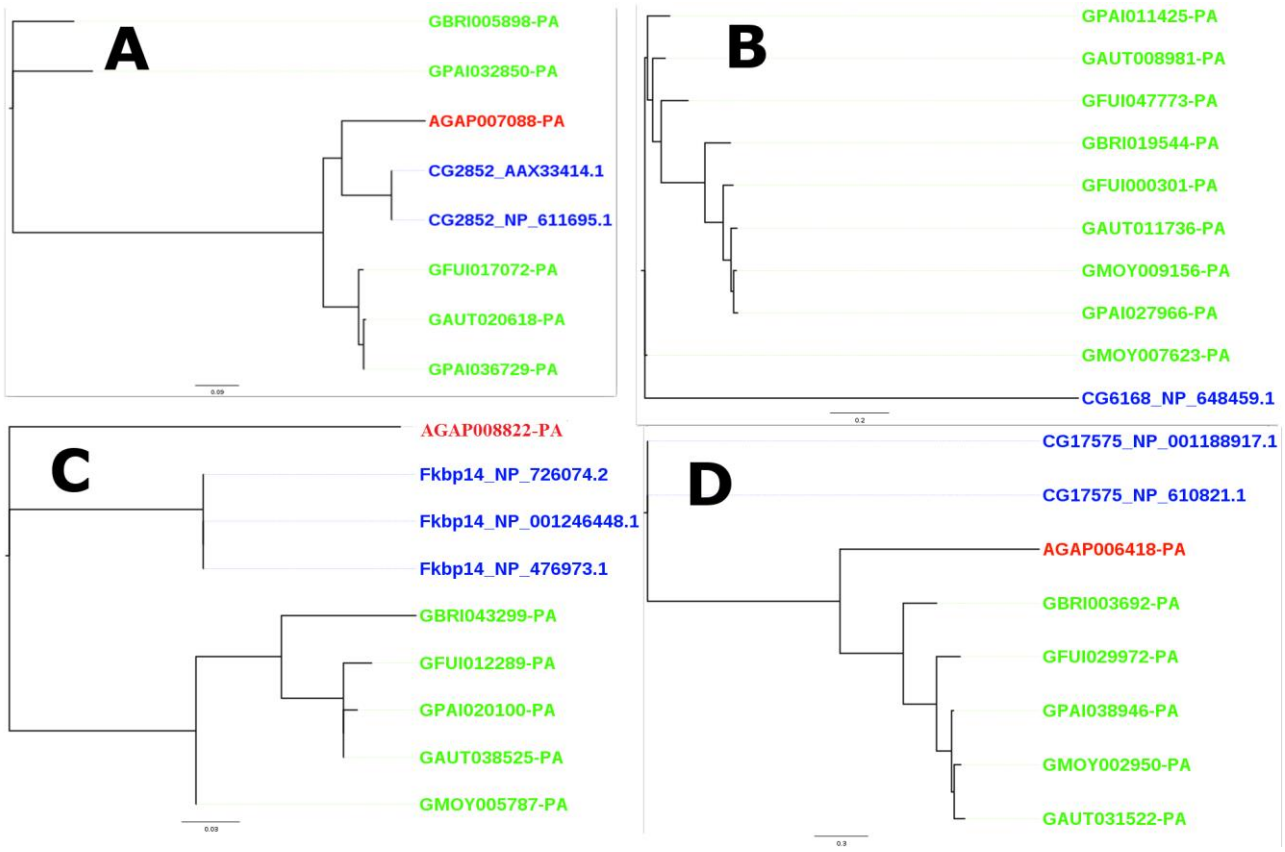


Figure 5: Phylogenetic trees showing adaptation and evolution of ACPs between *Anopheles gambiae*, *Drosophila melanogaster* and the 5 *Glossina* species

A: Isomerase ACP, **B**: Isoform B ACP, **C**: Fkbp14 ACP, **D**: CAP ACP

Key:

Green: GAUT: *Glossina austeni*, GBRI: *Glossina brevipalpis*, GFUI: *Glossina fuscipes*, GMOY: *Glossina morsitans*, GPAI: *Glossina pallidipes*

Red: AGAP: *Anopheles gambiae*

Blue: CG2852, CG6168, Fkbp14, CG17575: *Drosophila melanogaster*

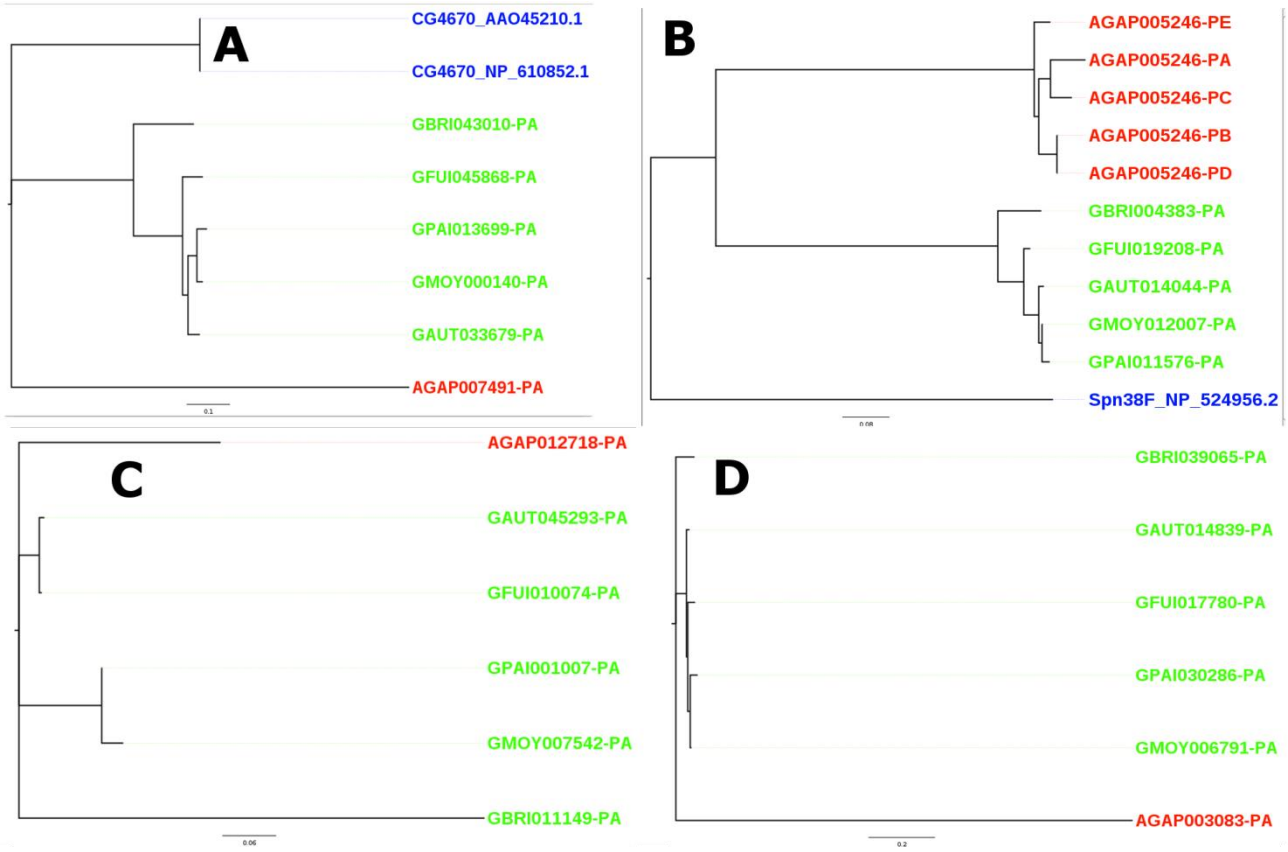


Figure 6: Phylogenetic trees showing evolution and adaptation of ACPs between *Anopheles gambiae*, *Drosophila melanogaster* and the 5 *Glossina* species

A: Redox ACP, **B**: Protease inhibitor ACP, **C**: Proline oxidase ACP, **D**: Lipase ACP

Key:

Green: GAUT: *Glossina austeni*, GBRI: *Glossina brevipalpis*, GFUI: *Glossina fuscipes*, GMOY: *Glossina morsitans*, GPAI: *Glossina pallidipes*

Red: AGAP: *Anopheles gambiae*

Blue: CG4670, Spn38F: *Drosophila melanogaster*

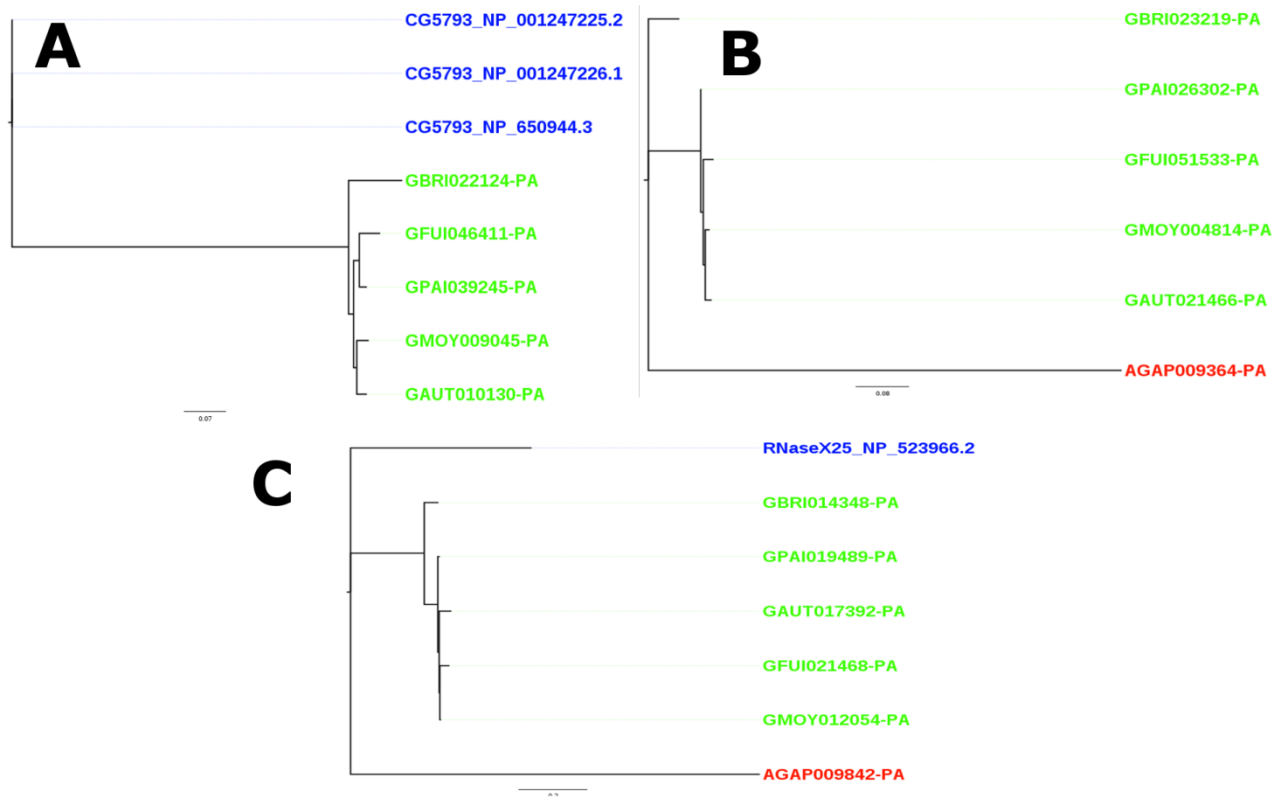


Figure 7: Phylogenetic trees showing evolution and adaptation of ACPs between *Anopheles gambiae, Drosophila melanogaster* and the 5 *Glossina* species

A, **B**: Transport proteins ACPs **C**: Ribonuclease ACP

Key:

Green: GAUT: *Glossina austeni*, GBRI: *Glossina brevipalpis*, GFUI: *Glossina fuscipes*, GMOY: *Glossina morsitans*, GPAI: *Glossina pallidipes*

Red: AGAP: *Anopheles gambiae*

Blue: CG5793, RNaseX25: *Drosophila melanogaster*

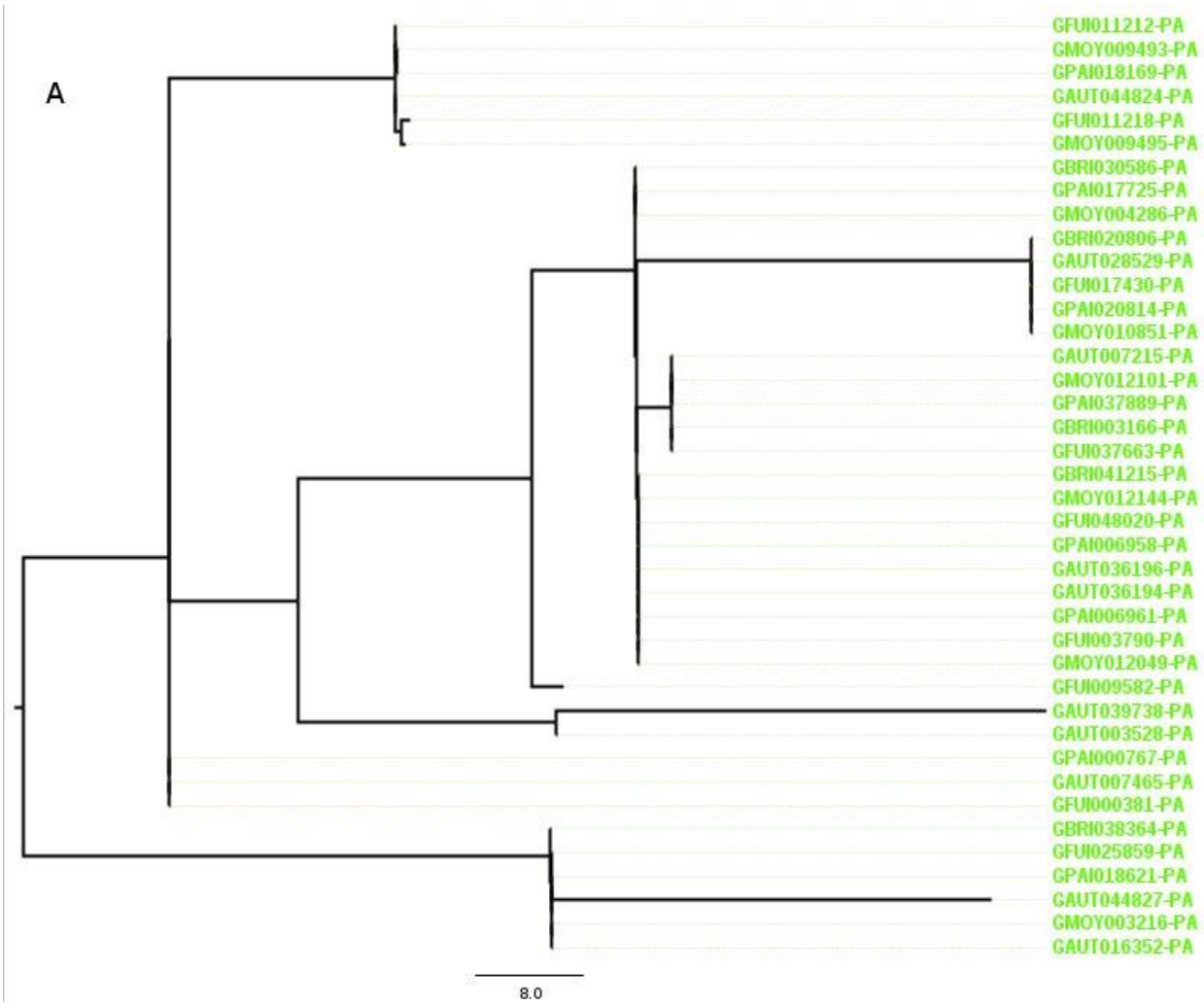


Figure 8: Phylogenetic tree showing homology of ACP sequences between *G.austeni*, *G.brevipalpis*, *G.fuscipes*, *G.morsitans* and *G.pallidipes*

A: Heat shock factor protein ACP

Key:

Green: GAUT: *Glossina austeni*, GBRI: *Glossina brevipalpis*, GFUI: *Glossina fuscipes*, GMOY: *Glossina morsitans*, GPAI: *Glossina pallidipes*

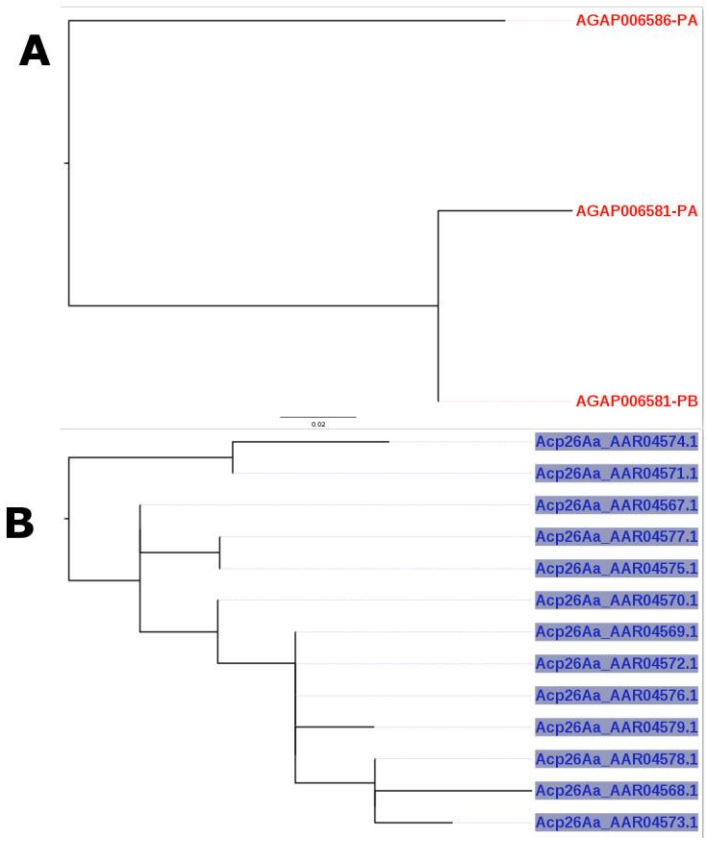


Figure 9: Phylogenetic trees showing homology between ACPs belonging to species of the same genus

A: Protease inhibitor, **B:** Accessory gland protein Acp26Aa

Key:

Red: AGAP: *Anopheles gambiae*

Blue: Acp26Aa: *Drosophila melanogaster*

CHAPTER 4

4.0 DISCUSSION

This study has identified ACP genes from a tsetse vector of human trypanosomiasis and cattle Nagana. It provides evidence that ACPs are typical proteins that function as lipases, hydrolases, chaperones and redox proteins. These proteins play major roles in reproduction ranging from egg stimulation to sperm defense within the female insect (Dottorini et al., 2007). An understanding of the ACPs nature and performance is vital in the control of the insect's reproduction that will further diminish the spread of parasites such as the trypanosome.

A high proportion of these MAGs secretions consist of characteristics such as a dN/dS value of > 1 (Table 1) that makes them attractive targets of positive Darwinian selection (Aguadé, Miyashita, & Langley, 1992; Wolfner, 2002). This positive selection of the ACPs might be the reason behind the tsetse fly's reduced genetic variability within each vector population and hence limited ability to respond to various control methods (Kariithi et al., 2013). The ACP genes in the three insects share a common ancestor (Figures. 4, 5, 6 and 7) and hence they are likely to share common ACP sequences derived from that ancestor. But, when one ancestral species splits into two, differences accumulate as a result of mutations resulting in divergence of that sequence. The greater the amounts of divergence indicate longer duration after the split. Gene duplication is common during evolution. If an extra copy of a gene can be made, then that extra copy is free to mutate and evolve into a separate function.

The orthologous sequence alignments of *A.gambiae*, *D.melanogaster*, *G.austeni*, *G.brevipalpis*, *G.fuscipes*, *G.morsitans* and *G.pallidipes* showed high sequence similarity for ACP genes that code for proteases, macroglobulins and transport proteins but not for ACP genes that code for *Drosophila melanogaster* (26a, 26Ab, 53Ea and 70A) and *Anopheles gambiae* ACPs (AGAP009354, AGAP009355, AGAP012706 and AGAP012830). Such ACPs that are present in the seminal fluids of *D.melanogaster* exhibit high evolutionary changes thus displaying between species divergence and within species polymorphismin (T Chapman, 2001) which affect the female *Drosophila*'s reproductive behavior that is not observed in the *Glossina* genus. This might be due to adaptive evolution since most genes that play a role in reproduction diverge rapidly (Swanson & Vacquier, 2002). Although sequence similarity does not necessarily entail similarity in function, the functional domains of well-aligned sequences did show similar function of ACP genes for all three genera on

Pfam. This relates to the conserved role of ACP genes despite their high evolution rate. An example would be the Anopheles ACP, AGAP006581, with the highest dN/dS value of 1.95. AGAP006581 is a homolog to Drosophila ACP62F (Dottorini et al., 2007) which is important in reproduction as it plays a role in the up-regulation of egg production and muscle development, characteristics that are vital for *Anopheles* and *Drosophila* reproduction but are futile in the *Glossina* and hence absent (Avila, Sirot, LaFlamme, Rubinstein, & Wolfner, 2011). The andropin gene with the lowest dN/dS of 0.261 is present in the seminal fluid of *Drosophila* and has antibacterial properties that protect the male reproductive tract from infection (Samakovlis, Kylsten, Kimbrell, Engström, & Hultmark, 1991). Despite having an important role, the andropin ACP has the lowest dN/dS of 0.261, which suggests that this ACP is under negative or purifying selection. Because a greater number of DNA changes are actually more harmful than advantageous, negative selection plays an important role in balancing the long-term stability of biological structures by removing deleterious mutations (Loewe, 2008).

CHAPTER 5

5.0 CONCLUSION

5.1 Conclusion

The main objective of this study was to identify the presence of ACPs in the five *Glossina* species (*G.austeni*, *G.brevipalpis*, *G.fuscipes*, *G.morsitans* and *G.pallidipes*). This study identified the presence of ACPs in tsetse flies. Bioinformatics analyses indicate that ACPs are rapidly evolving and are under adaptive selection meaning that they are present according to how the insect reproduces. Laboratory experiments should be performed for ACPs to ascertain as the appropriate candidates for vector control.

5.2 Recommendations

The following are recommended for future works:

1) Laboratory analysis should be performed on the tsetse fly to confirm ACP function.
2) Silencing of the Hsf gene in the tsetse fly to determine the effect on the insect's reproducing system and use this observation for vector control.

REFERENCES

Aguadé, M., Miyashita, N., & Langley, C. H. (1992). Polymorphism and divergence in the Mst26A male accessory gland gene region in Drosophila. *Genetics*, *132*(3), 755–70. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1205212&tool=pmcentrez&rendertype=abstract

Aksoy, S. (2010). A proposal for tsetse fly (Glossina) genome projects. Retrieved May 16, 2015, from http://www.genome.gov/Pages/Research/DER/PathogensandVectors/A_proposal_for_tsetse_fly_Apr28.pdf

Allsopp, R. (2001). Options for vector control against trypanosomiasis in Africa. *Trends in Parasitology*, *17*(1), 15–19. doi:10.1016/S1471-4922(00)01828-6

Almeida, F. C., & Desalle, R. (2009). Orthology, function and evolution of accessory gland proteins in the Drosophila repleta group. *Genetics*, *181*(1), 235–45. doi:10.1534/genetics.108.096263

Avila, F. W., Sirot, L. K., LaFlamme, B. A., Rubinstein, C. D., & Wolfner, M. F. (2011). Insect seminal fluid proteins: identification and function. *Annual Review of Entomology*, *56*, 21–40. doi:10.1146/annurev-ento-120709-144823

Baldini, F., Gabrieli, P., Rogers, D. W., & Catteruccia, F. (2012). Function and composition of male accessory gland secretions in Anopheles gambiae: a comparison with other insect vectors of infectious diseases. *Pathogens and Global Health*, *106*(2), 82–93. doi:10.1179/2047773212Y.0000000016

Begun, D. J., Lindfors, H. A., Thompson, M. E., & Holloway, A. K. (2006). Recently evolved genes identified from Drosophila yakuba and D. erecta accessory gland expressed sequence tags. *Genetics*, *172*(3), 1675–81. doi:10.1534/genetics.105.050336

Chapman, T. (2001). Seminal fluid-mediated fitness traits in Drosophila. *Heredity*, *87*(Pt 5), 511–21. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11869341

Chapman, T., & Davies, S. J. (2004). Functions and analysis of the seminal fluid proteins of

male Drosophila melanogaster fruit flies. *Peptides*, *25*(9), 1477–90. doi:10.1016/j.peptides.2003.10.023

Chen, P. S. (1996). The accessory gland proteins in male Drosophila: structural, reproductive, and evolutionary aspects. *Experientia*, *52*(6), 503–10. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8698082

Clark, A. G., Aguadé, M., Prout, T., Harshman, L. G., & Langley, C. H. (1995). Variation in sperm displacement and its association with accessory gland protein loci in Drosophila melanogaster. *Genetics*, *139*(1), 189–201. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1206317&tool=pmcentrez&rendertype=abstract

Collins, A. M., Caperna, T. J., Williams, V., Garrett, W. M., & Evans, J. D. (2006). Proteomic analyses of male contributions to honey bee sperm storage and mating. *Insect Molecular Biology*, *15*(5), 541–9. doi:10.1111/j.1365-2583.2006.00674.x

Control and surveillance of human African trypanosomiasis. (2013). *World Health Organization Technical Report Series*, (984), 1–237. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/24552089

Davies, S. J., & Chapman, T. (2006). Identification of genes expressed in the accessory glands of male Mediterranean Fruit Flies (Ceratitis capitata). *Insect Biochemistry and Molecular Biology*, *36*(11), 846–56. doi:10.1016/j.ibmb.2006.08.009

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., … Gascuel, O. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research*, *36*(Web Server issue), W465–9. doi:10.1093/nar/gkn180

Dottorini, T., Nicolaides, L., Ranson, H., Rogers, D. W., Crisanti, A., & Catteruccia, F. (2007). A genome-wide analysis in Anopheles gambiae mosquitoes reveals 46 male accessory gland genes, possible modulators of female behavior. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(41), 16215–20. doi:10.1073/pnas.0703904104

Dottorini, T., Persampieri, T., Palladino, P., Baker, D. A., Spaccapelo, R., Senin, N., &

Crisanti, A. (2013). Regulation of Anopheles gambiae male accessory gland genes influences postmating response in female. *FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology*, *27*(1), 86–97. doi:10.1096/fj.12-219444

Dottorini, T., Persampieri, T., Palladino, P., Spaccapelo, R., & Crisanti, A. (2012). Silencing of the Hsf gene, the transcriptional regulator of A. gambiae male accessory glands, inhibits the formation of the mating plug in mated females and disrupts their monogamous behaviour. *Pathogens and Global Health*, *106*(7), 405–12. doi:10.1179/2047773212Y.0000000058

Eberhard, W. G., & Cordero, C. (1995). Sexual selection by cryptic female choice on male seminal products - a new bridge between sexual selection and reproductive physiology. *Trends in Ecology & Evolution*, *10*(12), 493–6. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21237123

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. doi:10.1093/nar/gkh340

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., … Punta, M. (2014). Pfam: the protein families database. *Nucleic Acids Research*, *42*(Database issue), D222–30. doi:10.1093/nar/gkt1223

Franco, J. R., Simarro, P. P., Diarra, A., & Jannin, J. G. (2014). Epidemiology of human African trypanosomiasis. *Clinical Epidemiology*, *6*, 257–75. doi:10.2147/CLEP.S39728

Gooding, R. H., & Krafsur, E. S. (2005). Tsetse genetics: contributions to biology, systematics, and control of tsetse flies. *Annual Review of Entomology*, *50*, 101–23. doi:10.1146/annurev.ento.50.071803.130443

Gouteux, J., & Jarry, M. (1998). Tsetse flies , biodiversity and the control of sleeping sickness . Structure of a Glossina guild in southwest C8te d ' Ivoire, *19*(5), 453–471.

Grady, S. C., Messina, J. P., & McCord, P. F. (2011). Population vulnerability and disability in Kenya's tsetse fly habitats. *PLoS Neglected Tropical Diseases*, *5*(2), e957. doi:10.1371/journal.pntd.0000957

Hocking, K. S., Lamerton, J. F., & Lewis, E. A. (1963). Tsetse-fly control and eradication. *Bulletin of the World Health Organization*, *28*(5-6), 811–23. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2554946&tool=pmcentrez&rendertype=abstract

Kariithi, H. M., van Oers, M. M., Vlak, J. M., Vreysen, M. J. B., Parker, A. G., & Abd-Alla, A. M. M. (2013). Virology, Epidemiology and Pathology of Glossina Hytrosavirus, and Its Control Prospects in Laboratory Colonies of the Tsetse Fly, Glossina pallidipes (Diptera; Glossinidae). *Insects*, *4*(3), 287–319. doi:10.3390/insects4030287

Korber, B., Allen G, R., & Gerald H, L. (2000). SNAP: Synonymous Non-synonymous Analysis Program. *Kluwer Academic Publishers*. Retrieved November 3, 2015, from http://www.hiv.lanl.gov/content/sequence/SNAP/SNAP.html

Krafsur, E. S. (2009). Tsetse flies: genetics, evolution, and role as vectors. *Infection, Genetics and Evolution : Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, *9*(1), 124–41. doi:10.1016/j.meegid.2008.09.010

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., … Marra, M. A. (2009). Circos: an information aesthetic for comparative genomics. *Genome Research*, *19*(9), 1639–45. doi:10.1101/gr.092759.109

Li, L., Stoeckert, C. J., & Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research*, *13*(9), 2178–89. doi:10.1101/gr.1224503

Loewe, L. (2008). Negative selection. *Nature Education*, *1*(1), 59. Retrieved from http://www.nature.com/scitable/topicpage/Negative-Selection-1136

Mpanya, A., Hendrickx, D., Vuna, M., Kanyinda, A., Lumbala, C., Tshilombo, V., … Lutumba, P. (2012). Should I get screened for sleeping sickness? A qualitative study in Kasai province, Democratic Republic of Congo. *PLoS Neglected Tropical Diseases*, *6*(1), e1467. doi:10.1371/journal.pntd.0001467

Mueller, J. L., Ravi Ram, K., McGraw, L. A., Bloch Qazi, M. C., Siggia, E. D., Clark, A. G., … Wolfner, M. F. (2005). Cross-species comparison of Drosophila male accessory gland protein genes. *Genetics*, *171*(1), 131–43. doi:10.1534/genetics.105.043844

Rice, W. R. (1996). Sexually antagonistic male adaptation triggered by experimental arrest of female evolution. *Nature*, *381*(6579), 232–4. doi:10.1038/381232a0

Samakovlis, C., Kylsten, P., Kimbrell, D. A., Engström, A., & Hultmark, D. (1991). The andropin gene and its product, a male-specific antibacterial peptide in Drosophila melanogaster. *The EMBO Journal*, *10*(1), 163–9. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=452624&tool=pmcentrez&rendertype=abstract

Soller, M., Bownes, M., & Kubli, E. (1997). Mating and Sex Peptide Stimulate the Accumulation of Yolk in Oocytes of Drosophila Melanogaster. *European Journal of Biochemistry*, *243*(3), 732–738. doi:10.1111/j.1432-1033.1997.00732.x

Swanson, W. J., & Vacquier, V. D. (2002). The rapid evolution of reproductive proteins. *Nature Reviews. Genetics*, *3*(2), 137–44. doi:10.1038/nrg733

Swanson, W. J., Wong, A., Wolfner, M. F., & Aquadro, C. F. (2004). Evolutionary expressed sequence tag analysis of Drosophila female reproductive tracts identifies genes subjected to positive selection. *Genetics*, *168*(3), 1457–65. doi:10.1534/genetics.104.030478

Tripet, F., Touré, Y. T., Dolo, G., & Lanzaro, G. C. (2003). Frequency of multiple inseminations in field-collected Anopheles gambiae females revealed by DNA analysis of transferred sperm. *The American Journal of Tropical Medicine and Hygiene*, *68*(1), 1–5. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12556139

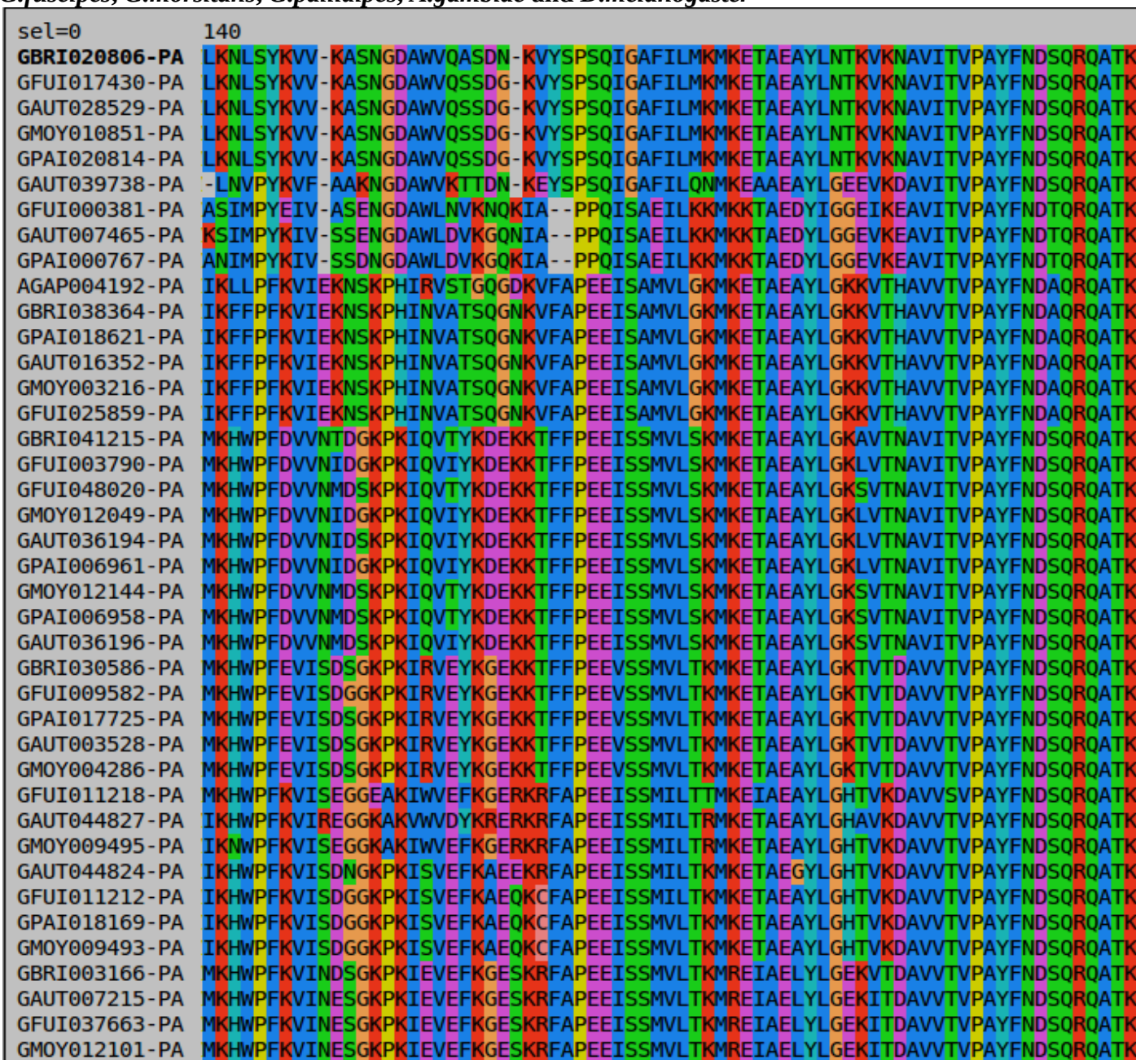
Walker, M. J., Rylett, C. M., Keen, J. N., Audsley, N., Sajid, M., Shirras, A. D., & Isaac, R. E. (2006). Proteomic identification of Drosophila melanogaster male accessory gland proteins, including a pro-cathepsin and a soluble gamma-glutamyl transpeptidase. *Proteome Science*, *4*, 9. doi:10.1186/1477-5956-4-9

Wolfner, M. F. (2002). The gifts that keep on giving: physiological functions and evolutionary dynamics of male seminal proteins in Drosophila. *Heredity*, *88*(2), 85–93. doi:10.1038/sj.hdy.6800017

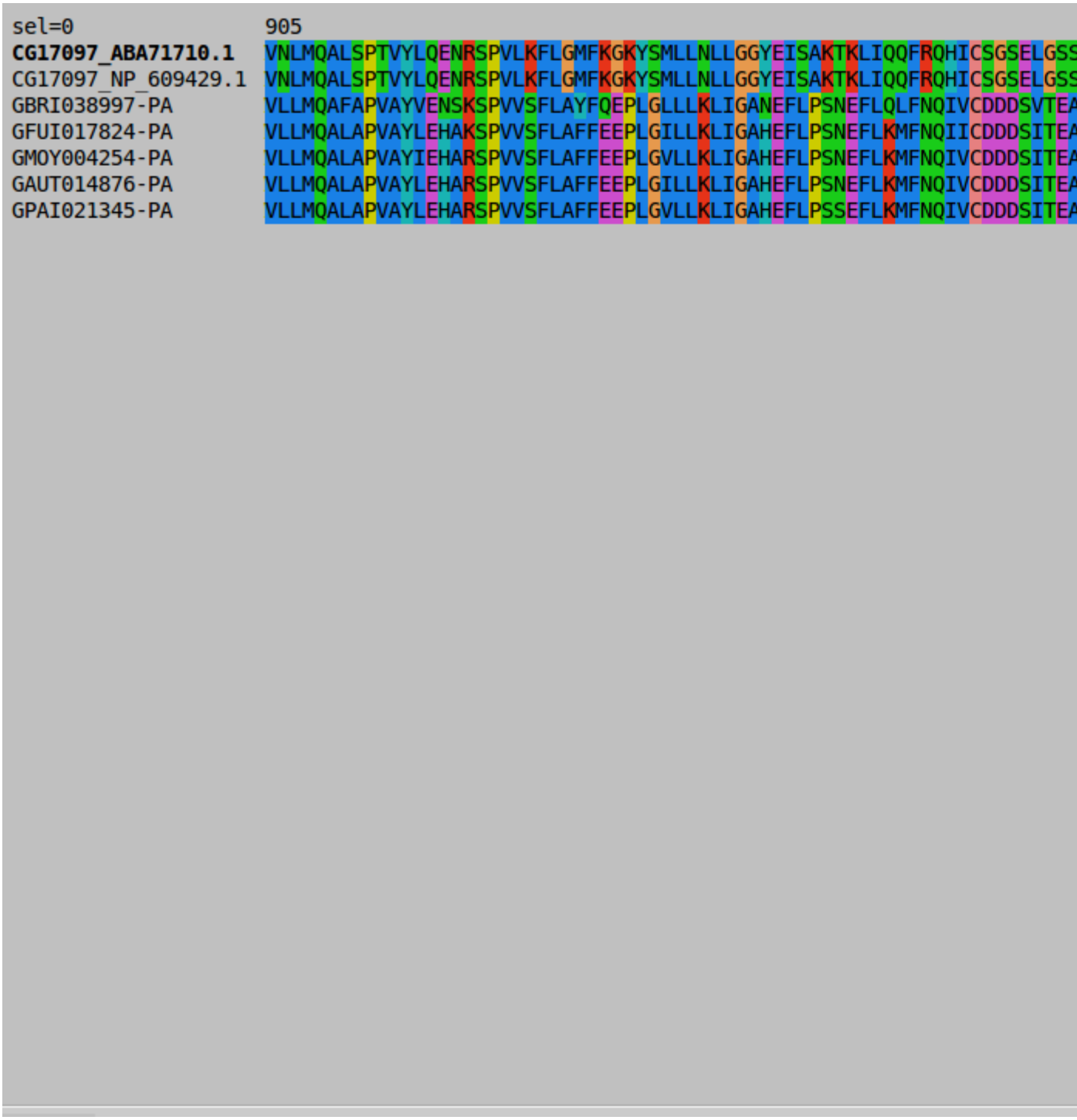
Wolfner, M. F., Partridge, L., Lewin, S., Kalb, J. M., Chapman, T., & Herndon, L. A. (1997). Mating and hormonal triggers regulate accessory gland gene expression in male

Drosophila. *Journal of Insect Physiology*, *43*(12), 1117–1123. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12770484

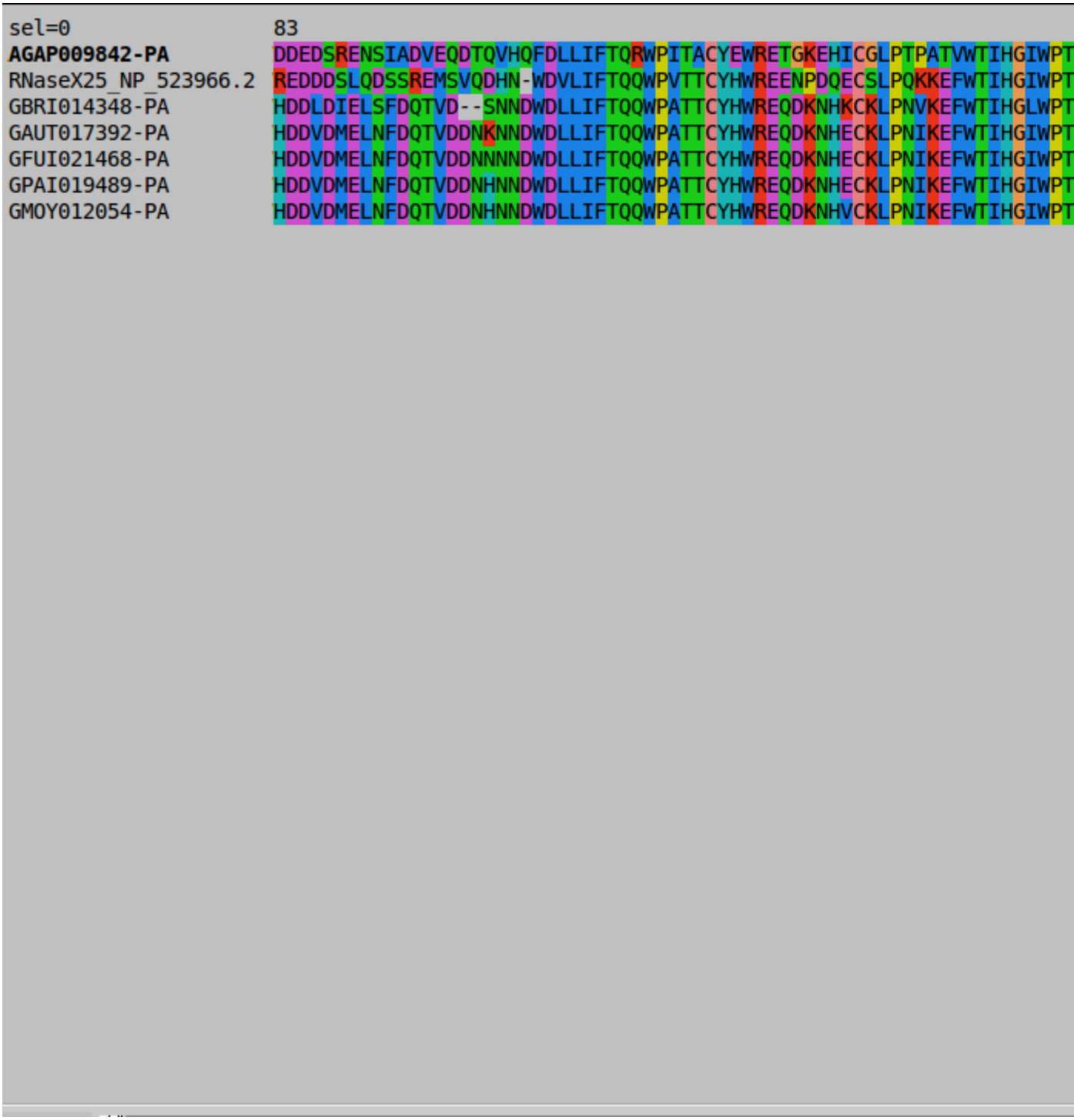
APPENDIX 1: 26 Alignment screenshots of ACPs identified in *G.austeni*, *G.brevipalpis*, *G.fuscipes*, *G.morsitans*, *G.pallidipes*, *A.gambiae* and *D.melanogaster*



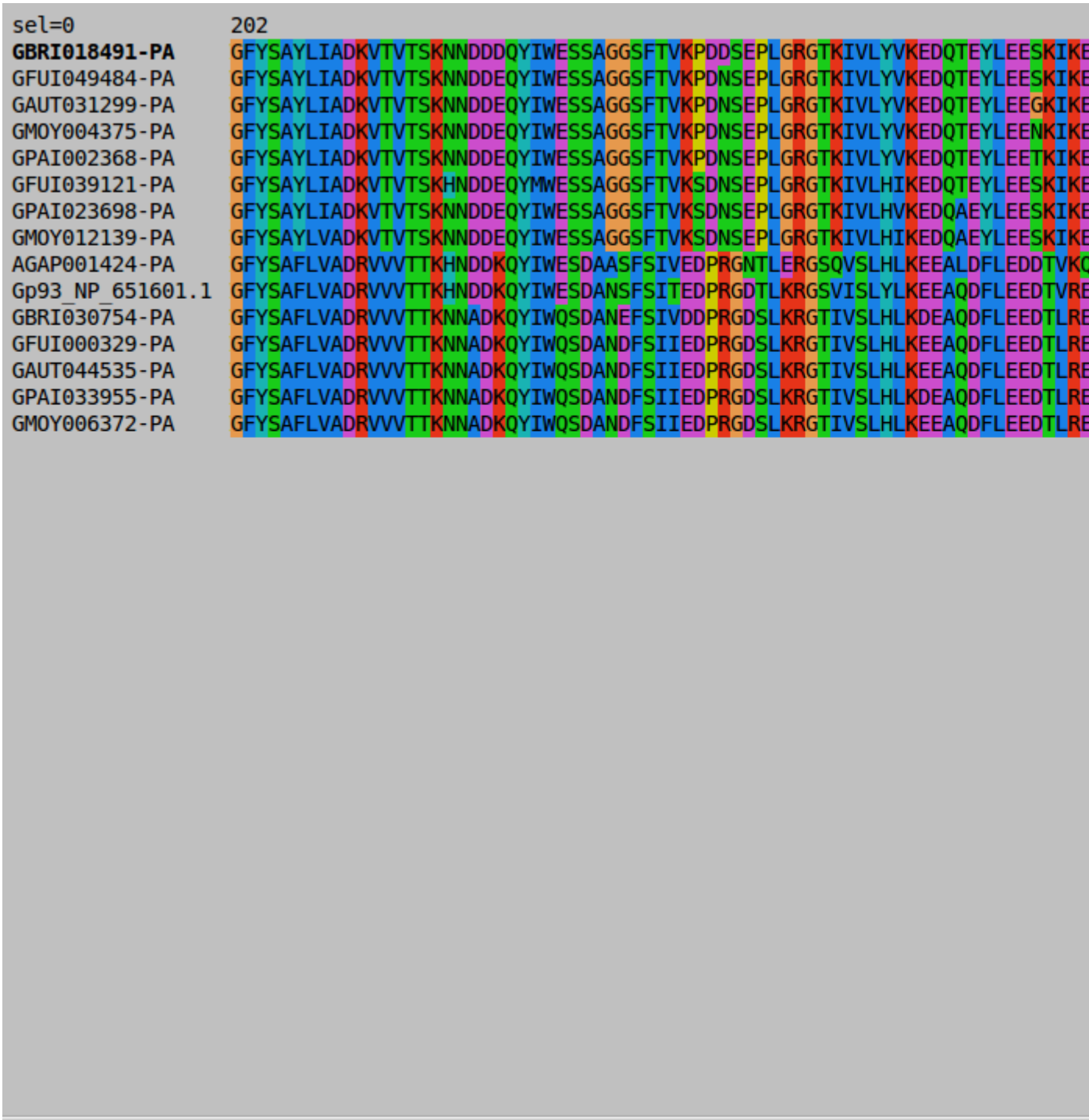
Screenshot displaying multiple alignment of heat shock factor ACP present in *A.gambiae* (AGAP), *G.austeni* (GAUT), *G.brevipalpis* (GBRI), *G.fuscipes* (GFUI), *G.morsitans* (GMOY) and *G.pallidipes* (GPAI)



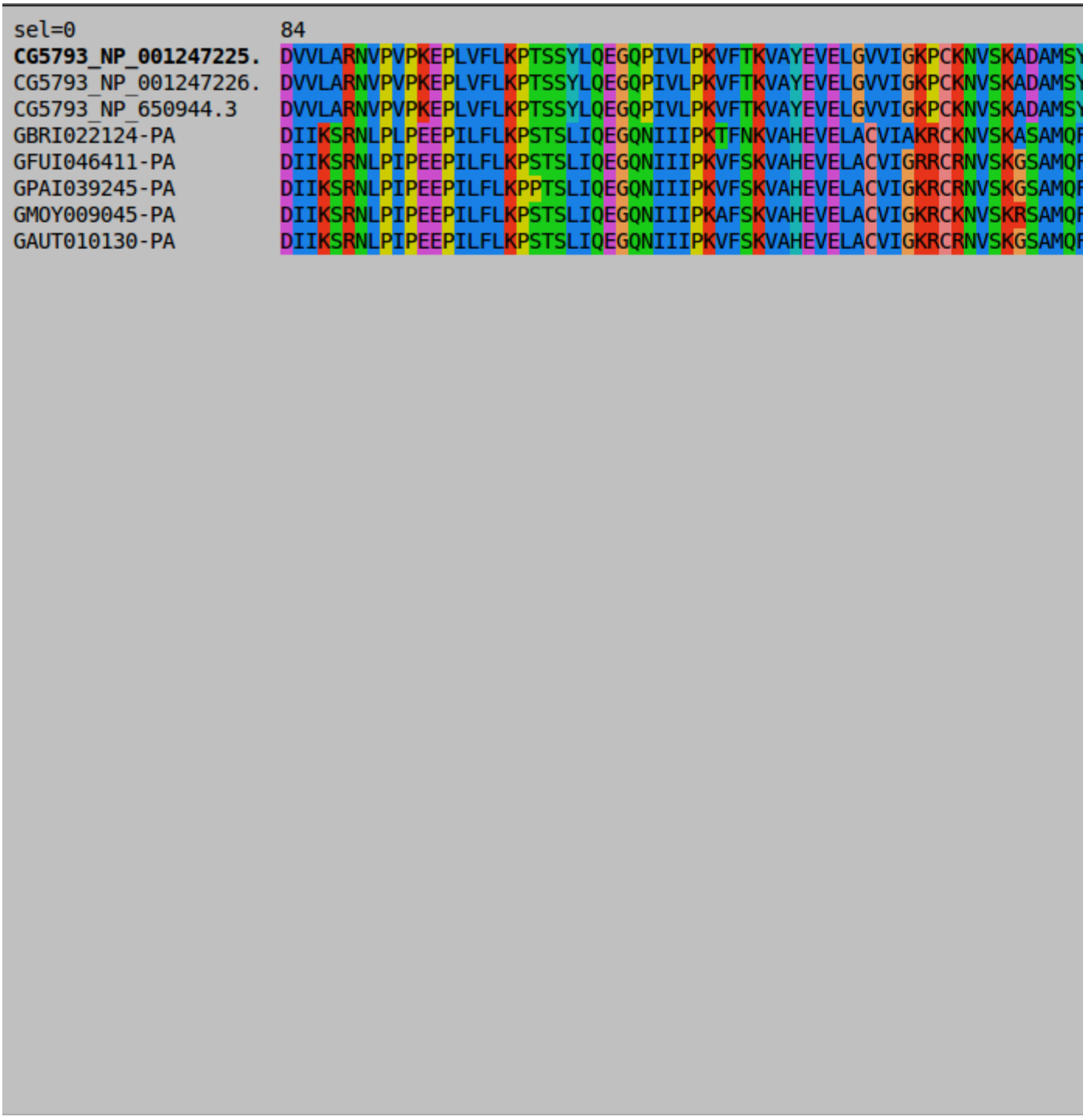
Screenshot displaying multiple alignment of lipase ACP present in *D.melanogaster* (CG17097), *G.austeni* (GAUT), *G.brevipalpis* (GBRI), *G.fuscipes* (GFUI), *G.morsitans* (GMOY)and *G.pallidipes* (GPAI)



Screenshot displaying multiple alignment of ribonuclease ACP present in *A.gambiae* (AGAP), *D.melanogaster* (RNaseX25), *G.austeni* (GAUT), *G.brevipalpis* (GBRI), *G.fuscipes* (GFUI), *G.morsitans* (GMOY) and *G.pallidipes* (GPAI)

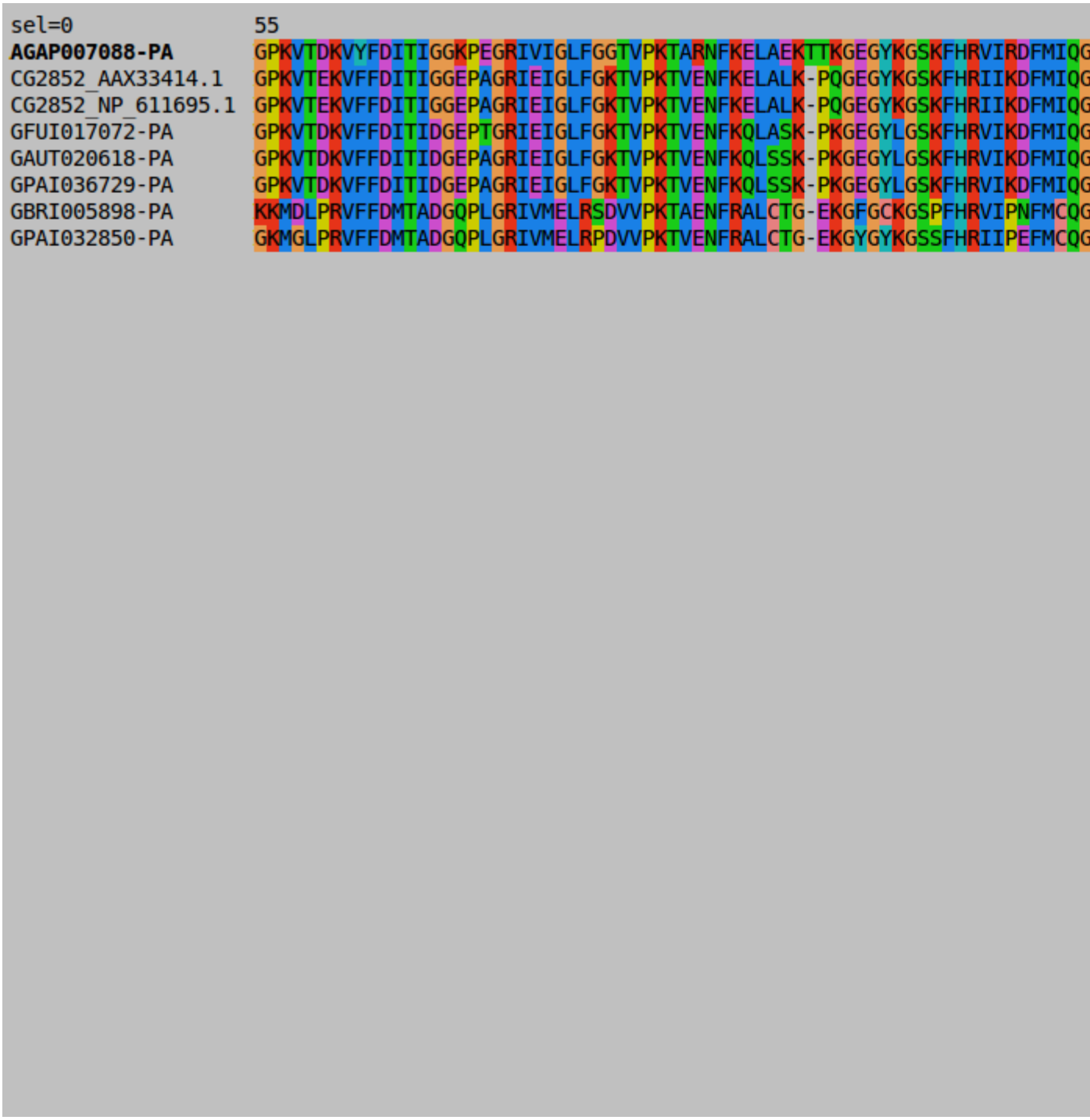


Screenshot displaying multiple alignment of chaperone ACP present in *A.gambiae* (AGAP), *D.melanogaster* (Gp93) , *G.austeni* (GAUT), *G.breviplapis* (GBRI), *G.fuscipes* (GFUI), *G.morsitans* (GMOY) and *G.pallidipes* (GPAI)

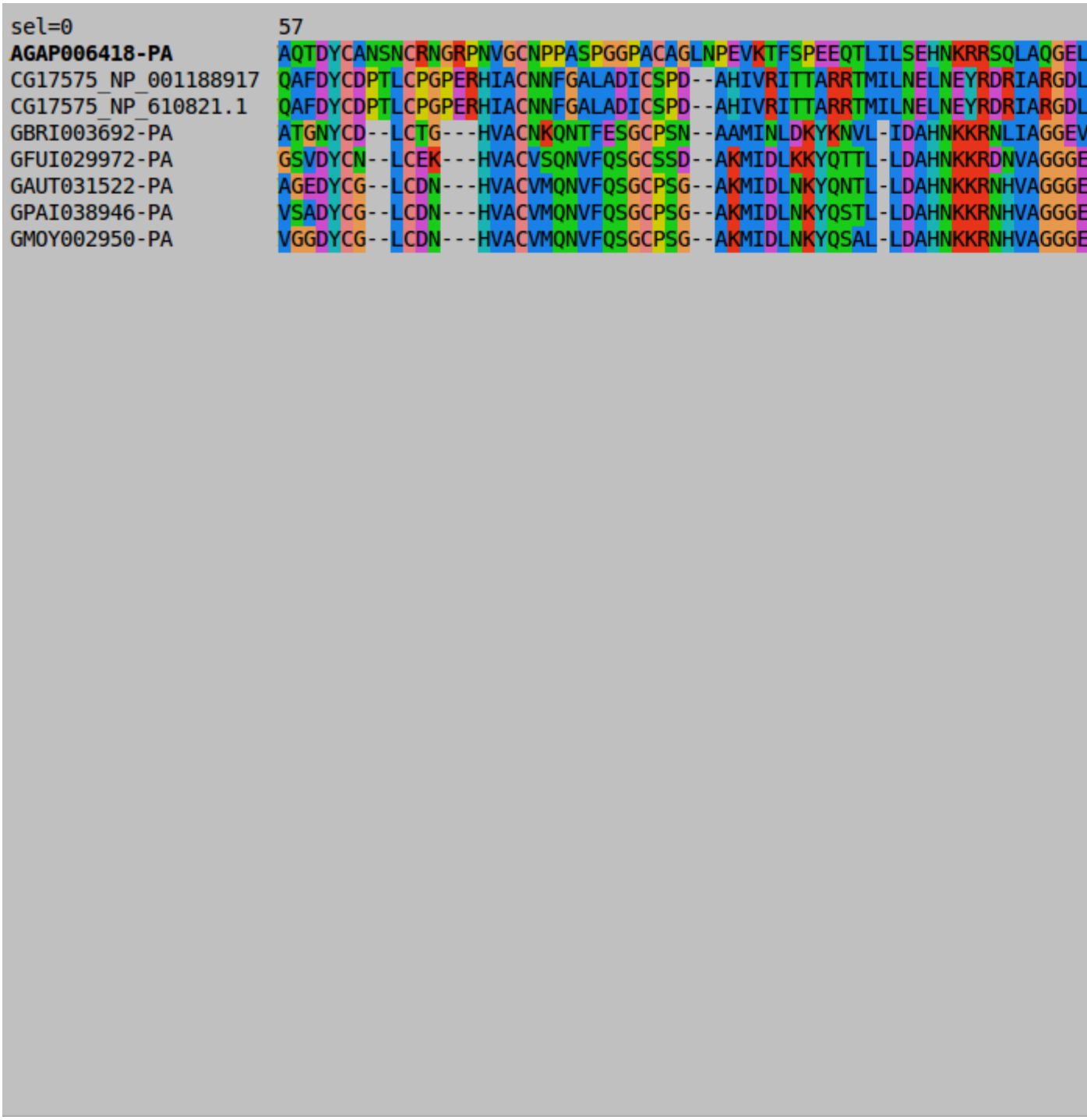


Screenshot displaying multiple alignment of transport protein ACP present in *D.melanogaster* (CG5793), *G.austeni* (GAUT), *G.brevipalpis* (GBRI), *G.fuscipes* (GFUI), *G.morsitans* (GMOY) and *G.pallidipes* (GPAI)

Screenshot displaying multiple alignment of redox ACP present in *A.gambiae* (AGAP), *D.melanogaster* (CG4670), *G.austeni* (GAUT), *G.brevipalpis* (GBRI), *G.fuscipes* (GFUI), *G.morsitans* (GMOY) and *G.pallidipes* (GPAI)

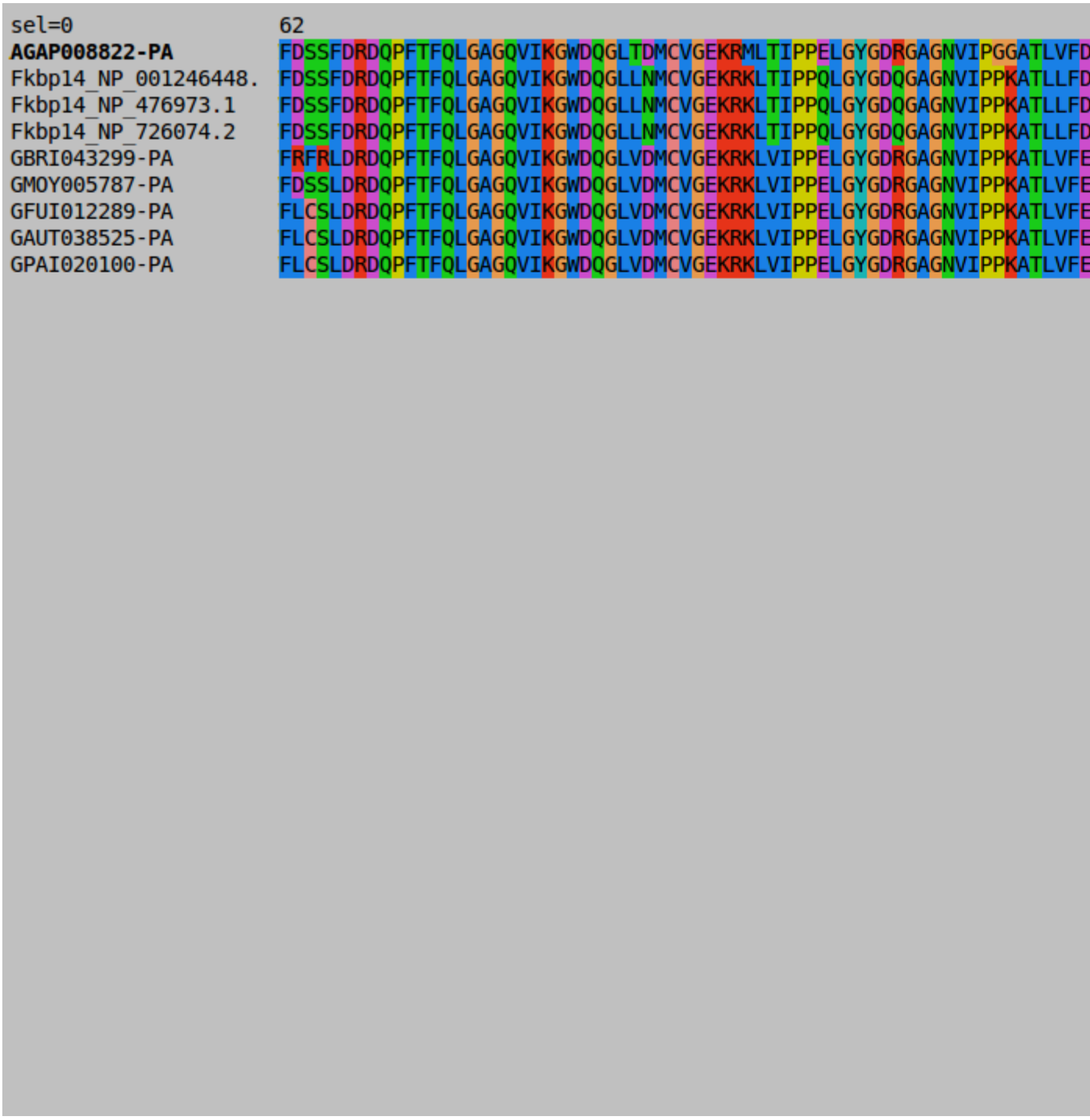


Screenshot displaying multiple alignment of isomerase ACP present in *A.gambiae* (AGAP), *D.melanogaster* (CG2852), *G.austeni* (GAUT), *G.brevipalpis* (GBRI), *G.fuscipes* (GFUI) and *G.pallidipes* (GPAI)

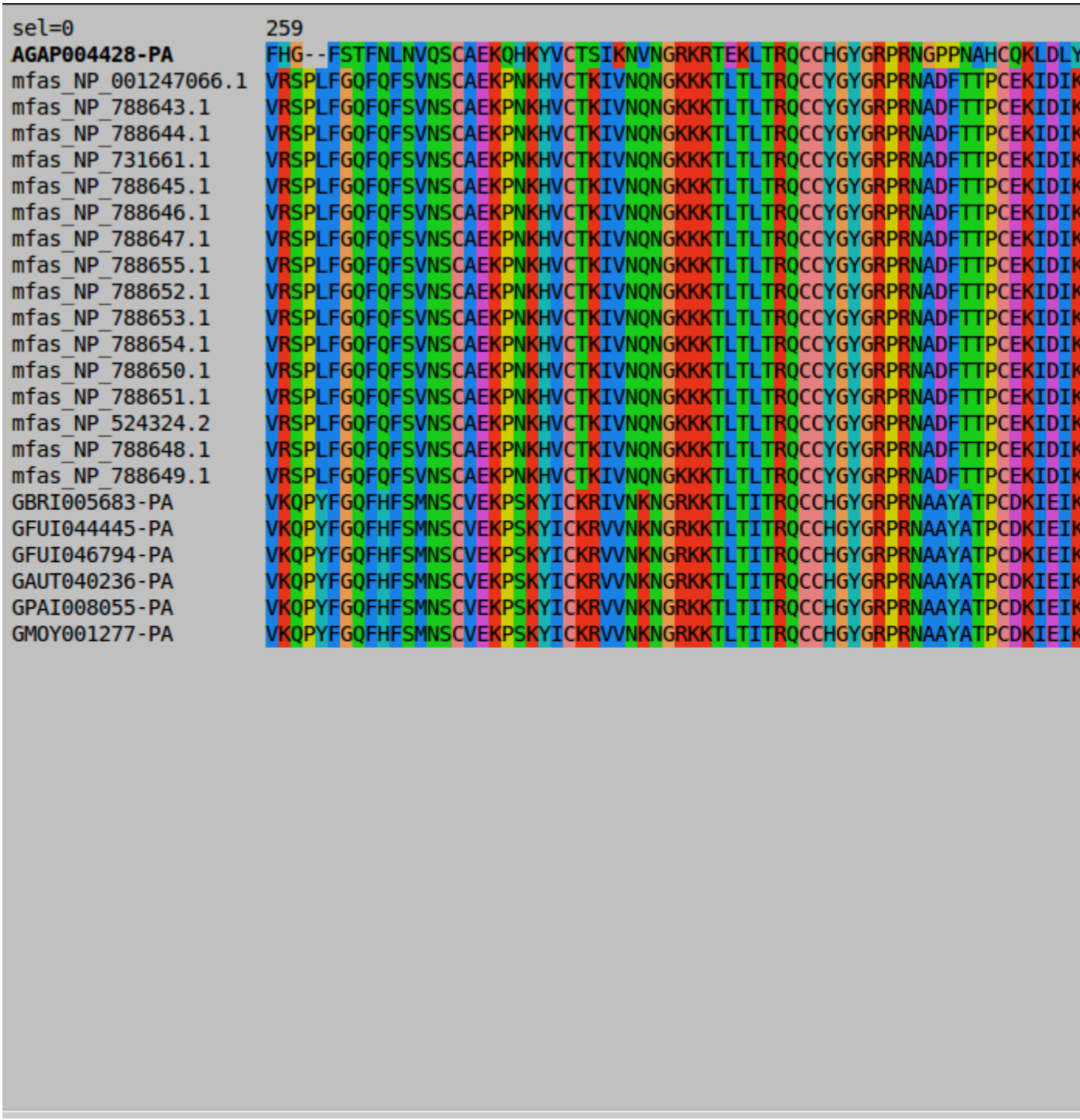


Screenshot diplaying multiple alignment of CAP ACP present in *A.gambiae* (AGAP), *D.melanogaster* (CG17575), *G.austeni* (GAUT), *G.brevipalpis* (GBRI), *G.fuscipes* (GFUI), *G.morsitans* (GMOY) and *G.pallidipes* (GPAI)

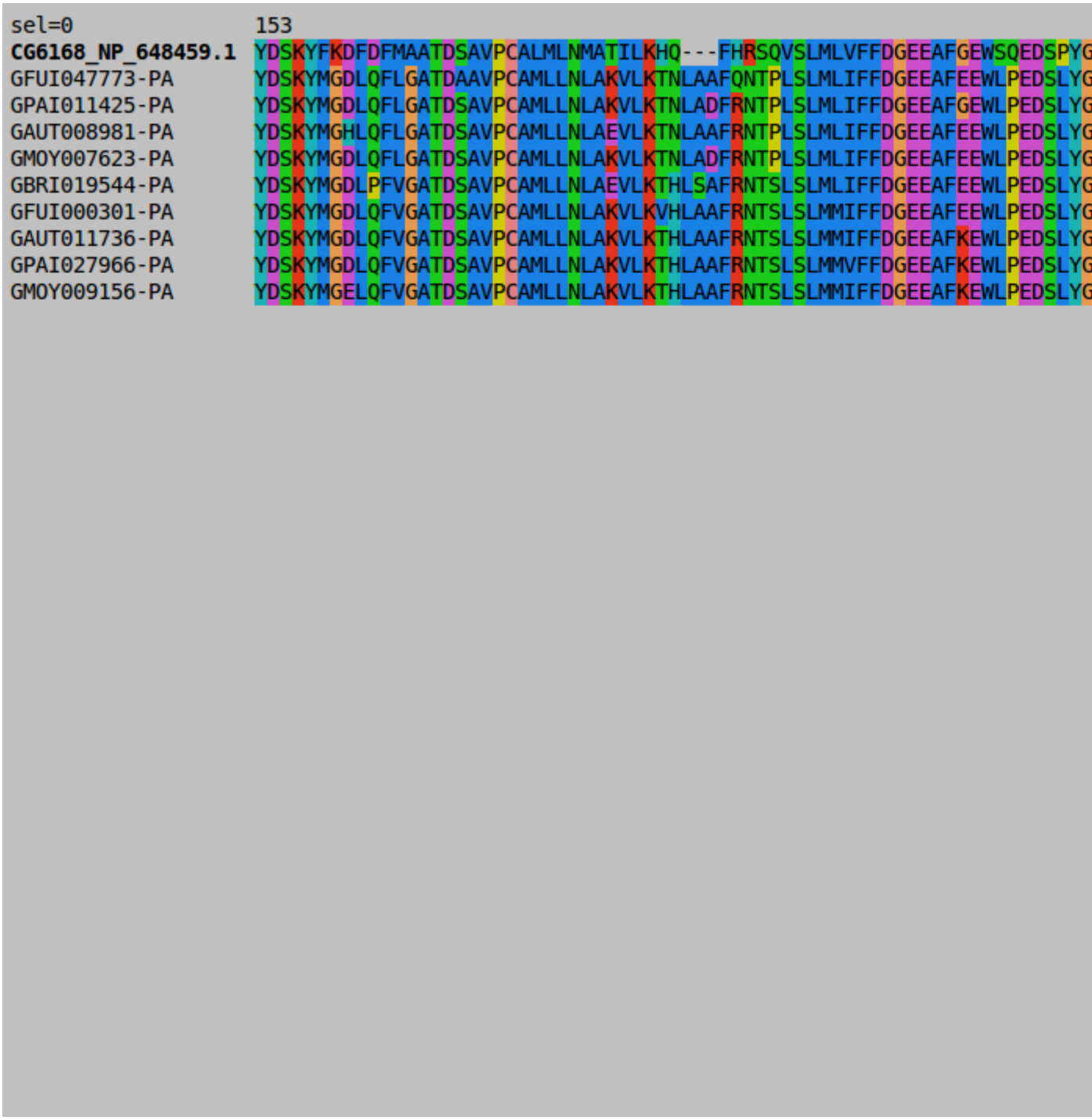
Screenshot displaying multiple alignment of carboxylesterase ACP present in *A.gambiae* (AGAP), *D.melanogaster* (Est), *G.austeni* (GAUT), *G.brevipalpis* (GBRI), *G.fuscipes* (GFUI), *G.morsitans* (GMOY) and *G.pallidipes* (GPAI)



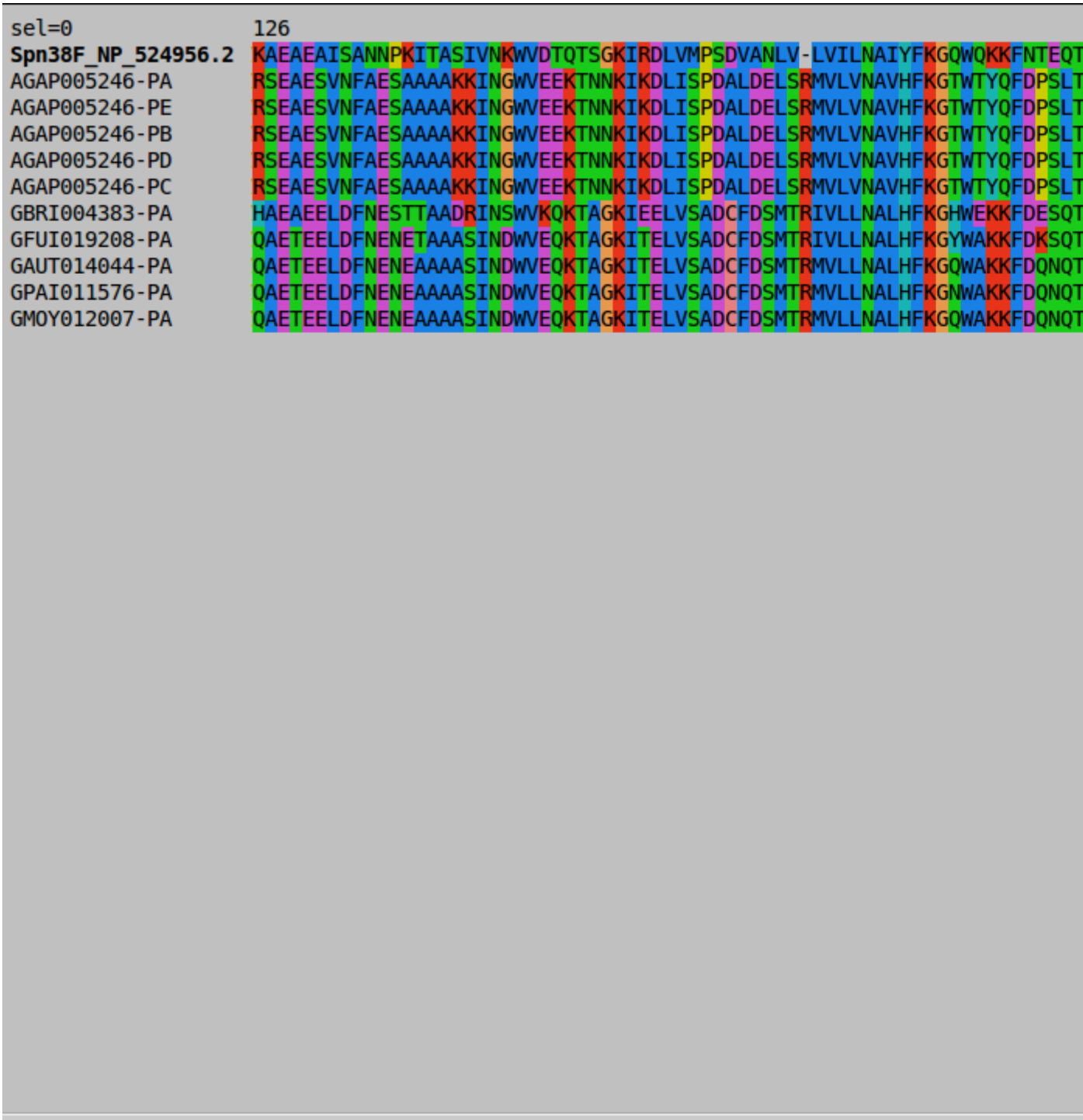
Screenshot displaying multiple alignment of isomerase ACP present in *A.gambiae* (AGAP), *D.melanogaster* (Fkbp_14_NP), *G.austeni* (GAUT), *G.brevipalpis* (GBRI), *G.fuscipes* (GFUI), *G.morsitans* (GMOY) and *G.pallidipes* (GPAI)



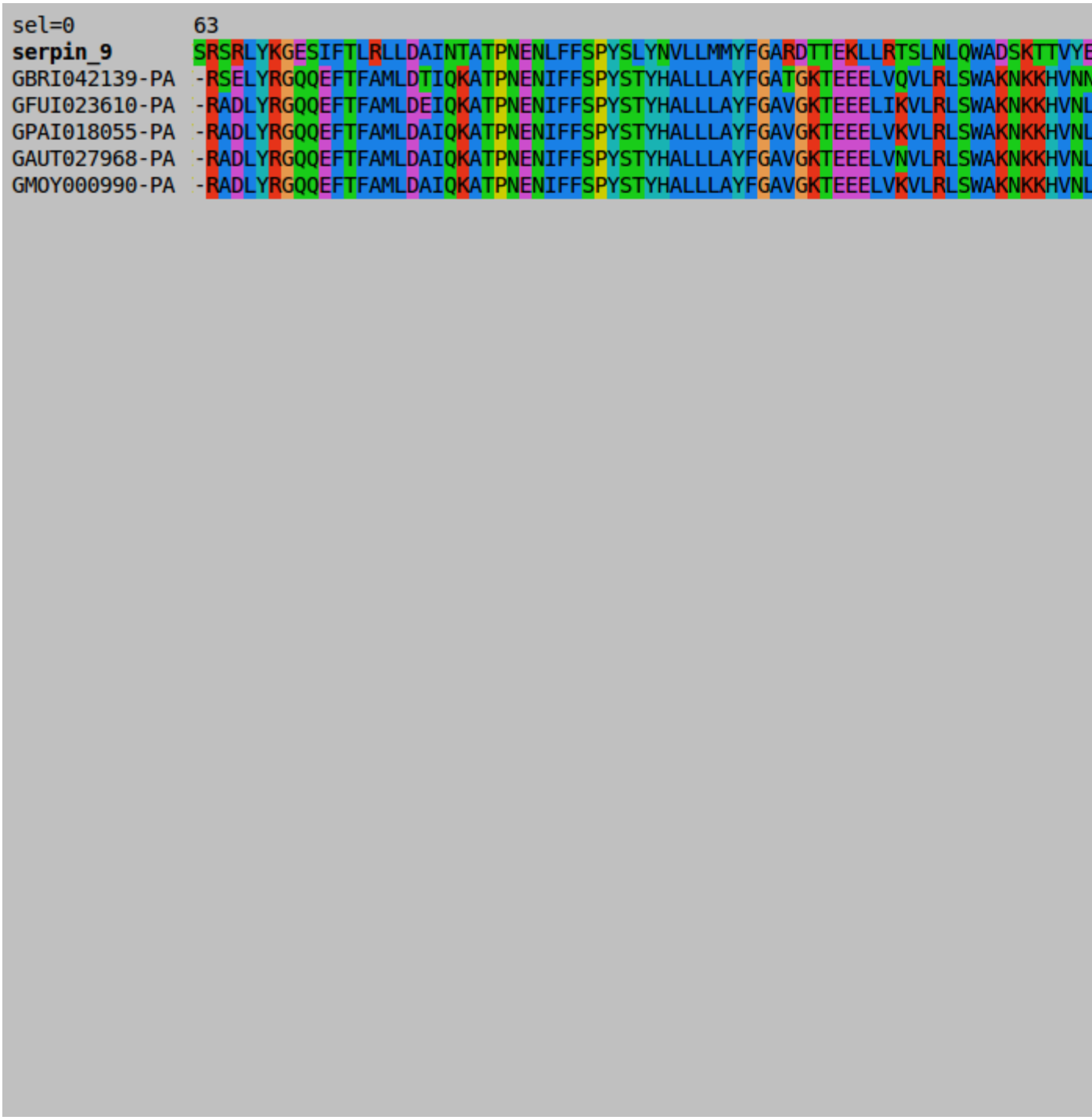
Screenshot displaying multiple alignment of cell adhesion ACP present in *A.gambiae* (AGAP), *D.melanogaster* (mfas_NP) *G.austeni* (GAUT), *G.brevipalpis* (GBRI), *G.fuscipes* (GFUI), *G.morsitans* (GMOY) and *G.pallidipes* (GPAI)



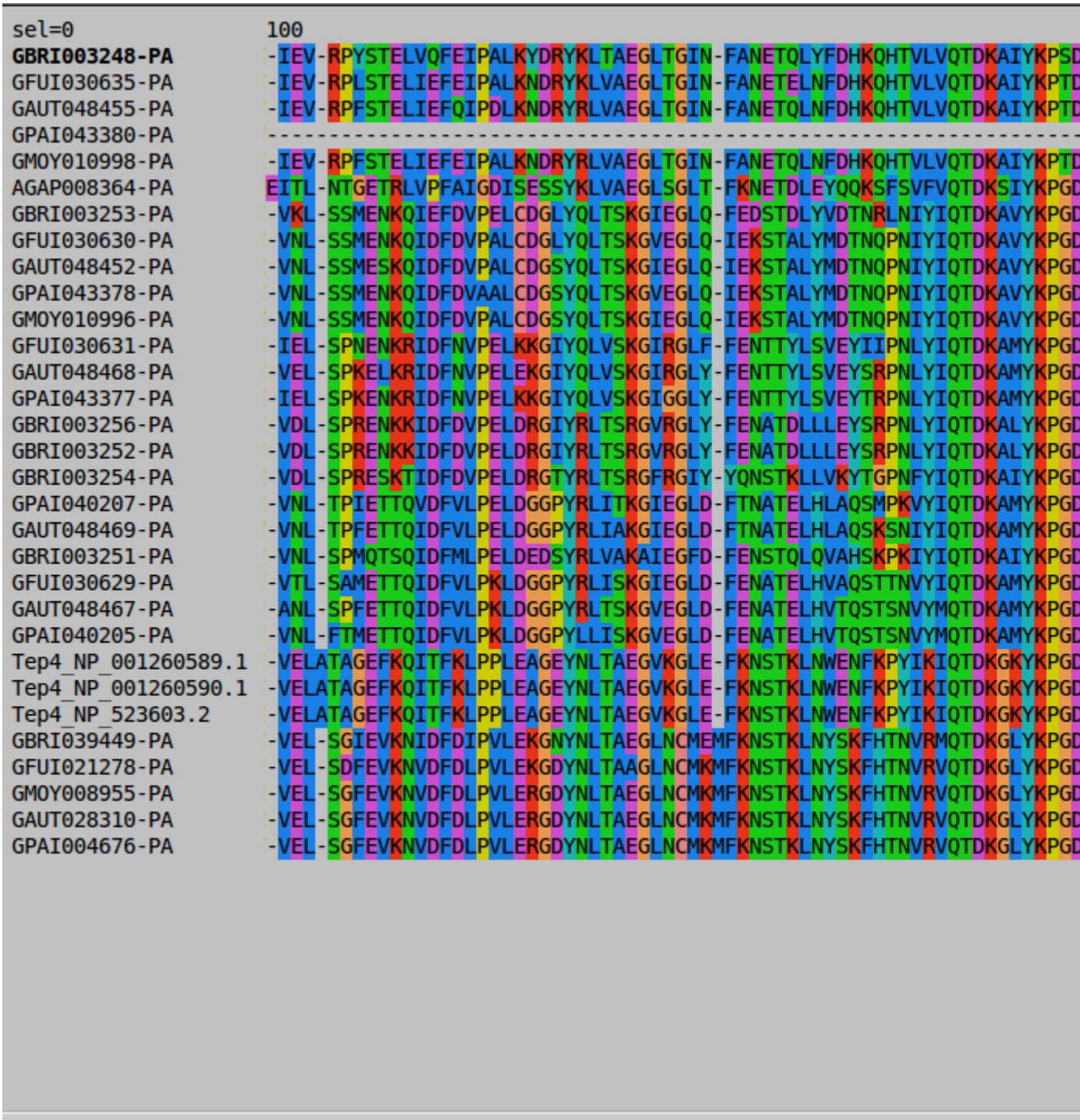
Screenshot displaying multiple alignment of isoform B ACP present in *D.melanogaster* (CG6168), *G.austeni* (GAUT), *G.brevipalpis* (GBRI), *G.fuscipes* (GFUI), *G.morsitans* (GMOY) and *G.pallidipes* (GPAI)



Screenshot displaying multiple alignment of protease inhibitor ACP present in *A.gambiae* (AGAP), *D.melanogaster* (Spn38F_NP), *G.austeni* (GAUT), *G.brevipalpis* (GBRI) , *G.fuscipes* (GFUI), *G.morsitans* (GMOY) and *G.pallidipes* (GPAI)

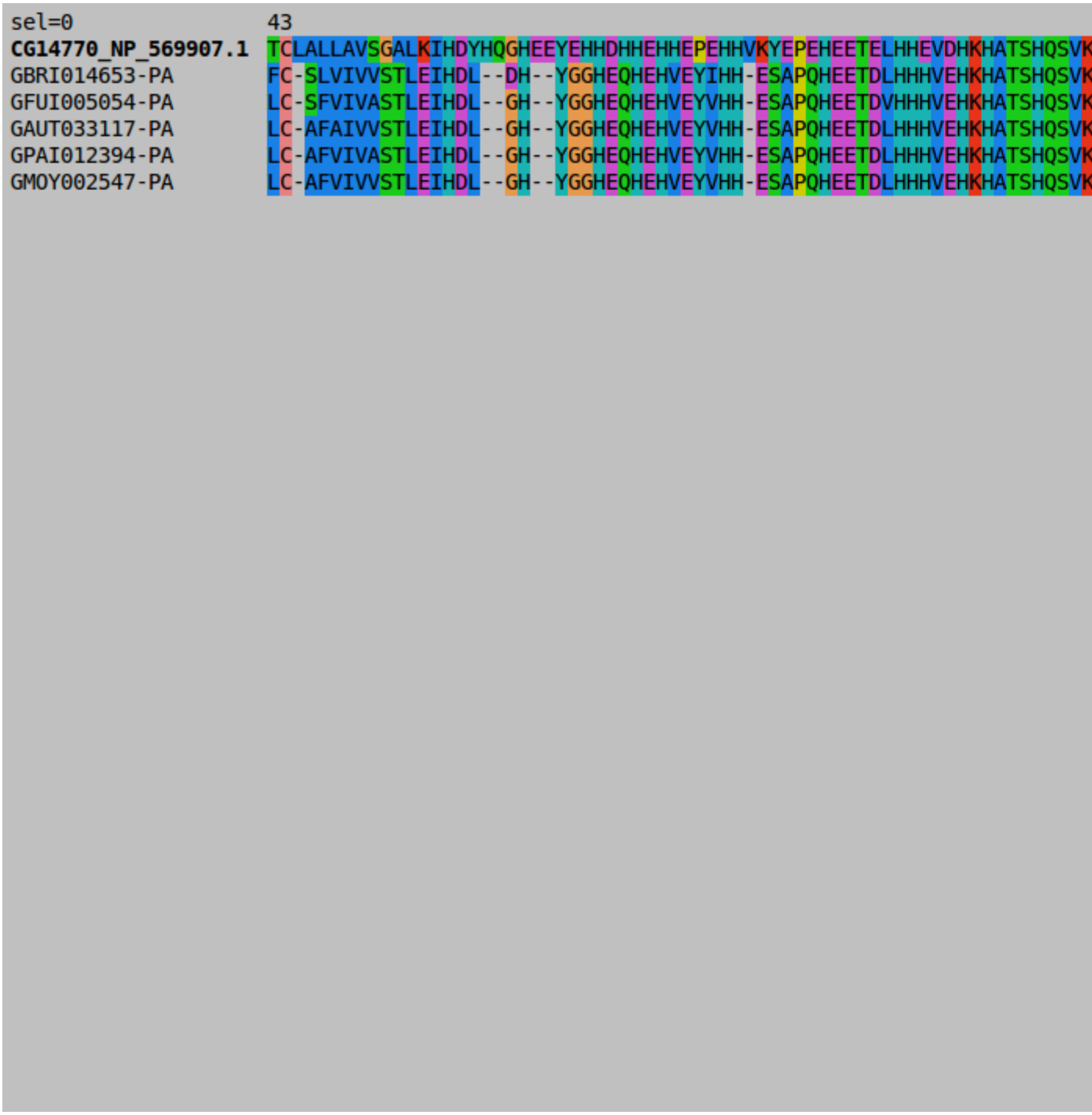


Screenshot displaying multiple alignment of serpin 9 protease inhibitor ACP present in *D.melanogaster* (serpin9), *G.austeni* (GAUT), *G.brevipalpis* (GBRI), *G.fuscipes* (GFUI), *G.morsitans* (GMOY) and *G.pallidipes* (GPAI)

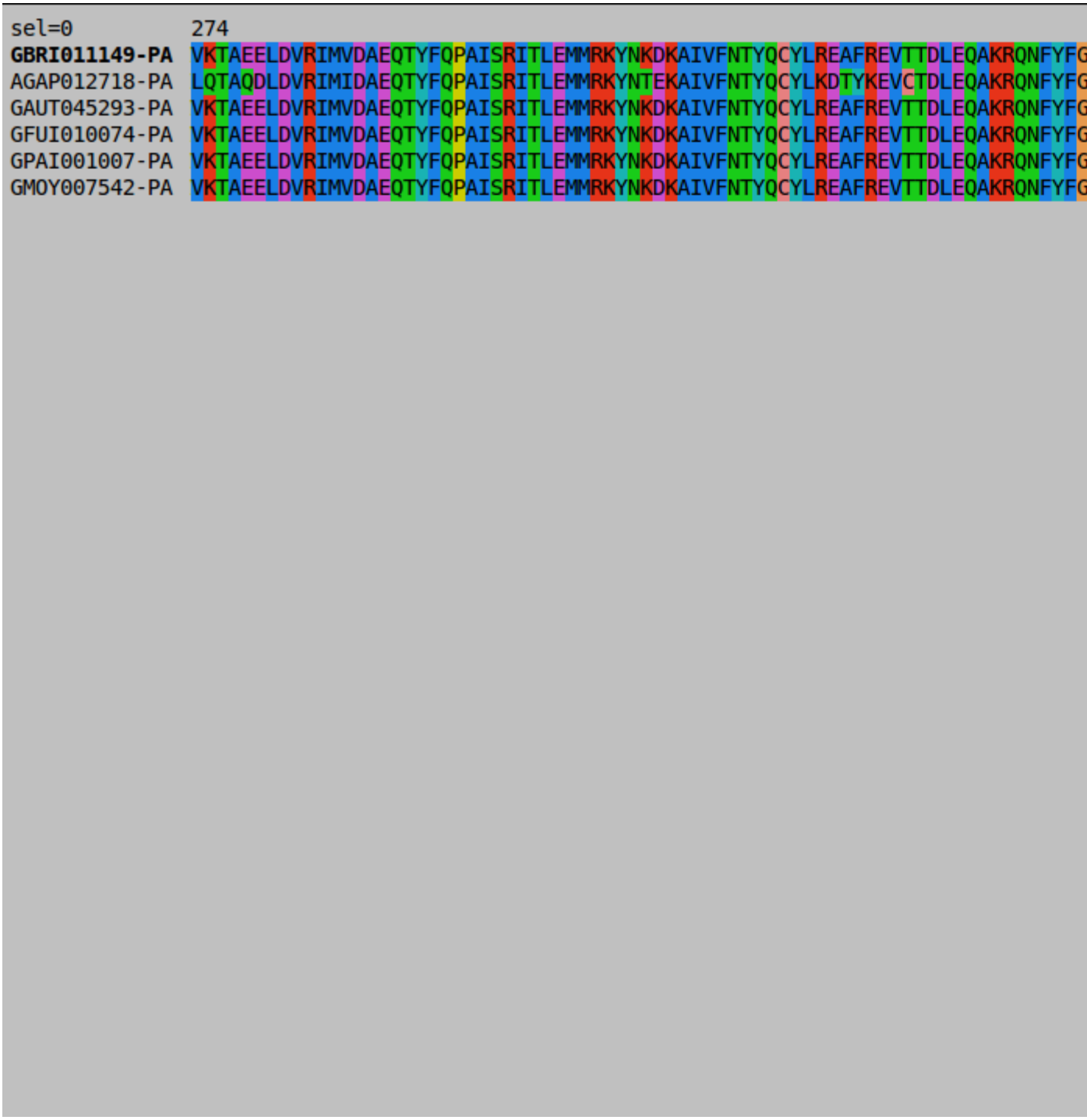


Screenshot displaying multiple alignment of macroglobulin ACP present in *A.gambiae* (AGAP), *D.melanogaster* (Tep4_NP), *G.austeni* (GAUT), *G.brevipalpis* (GBRI), *G.fuscipes* (GFUI), *G.morsitans* (GMOY) and *G.pallidipes* (GPAI)

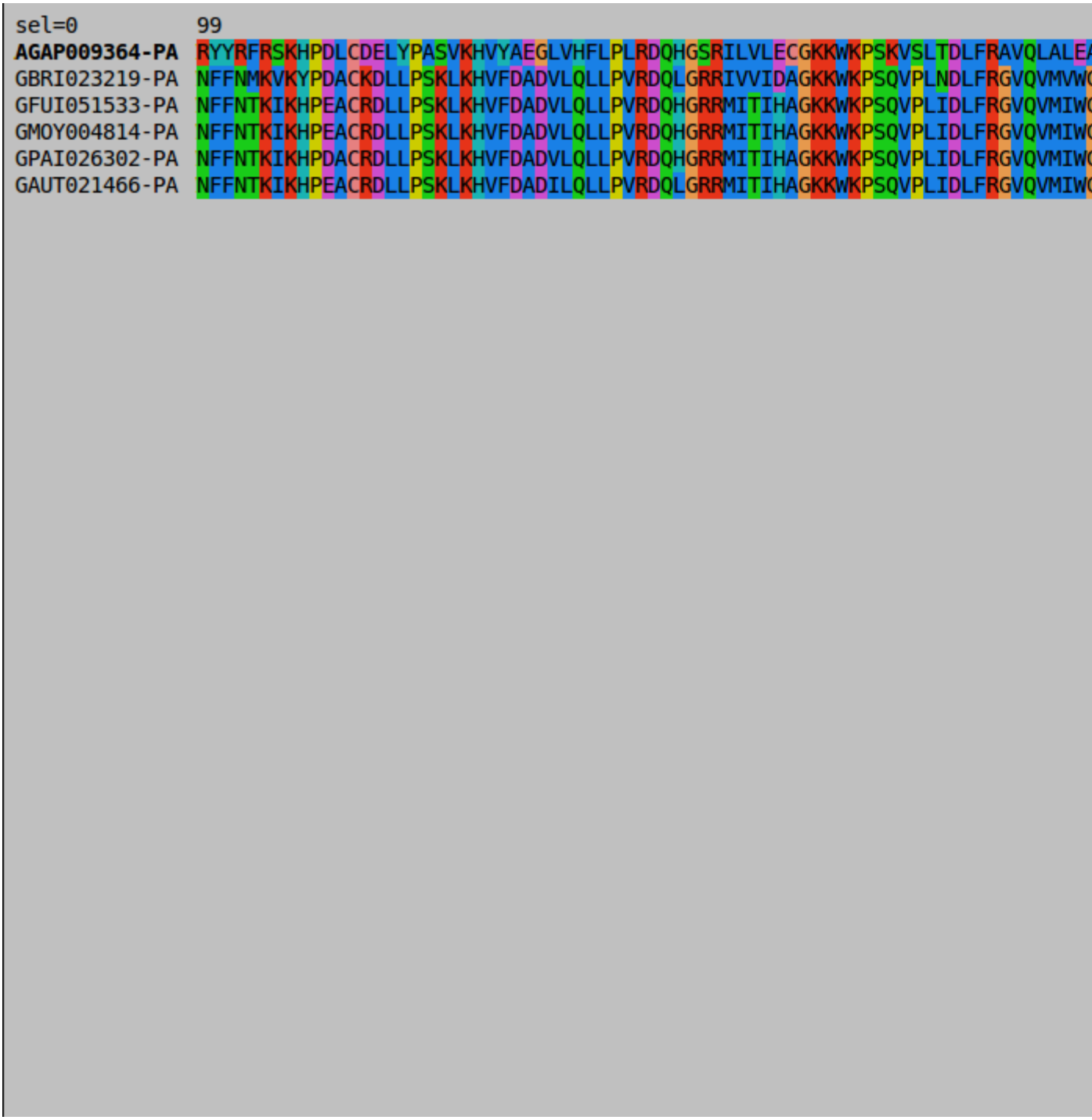
Screenshot displaying multiple alignment of protease ACP present in *A.gambiae* (AGAP), *G.austeni* (GAUT), *G.brevipalpis* (GBRI), *G.fuscipes* (GFUI), *G.morsitans* (GMOY) and *G.pallidipes* (GPAI)



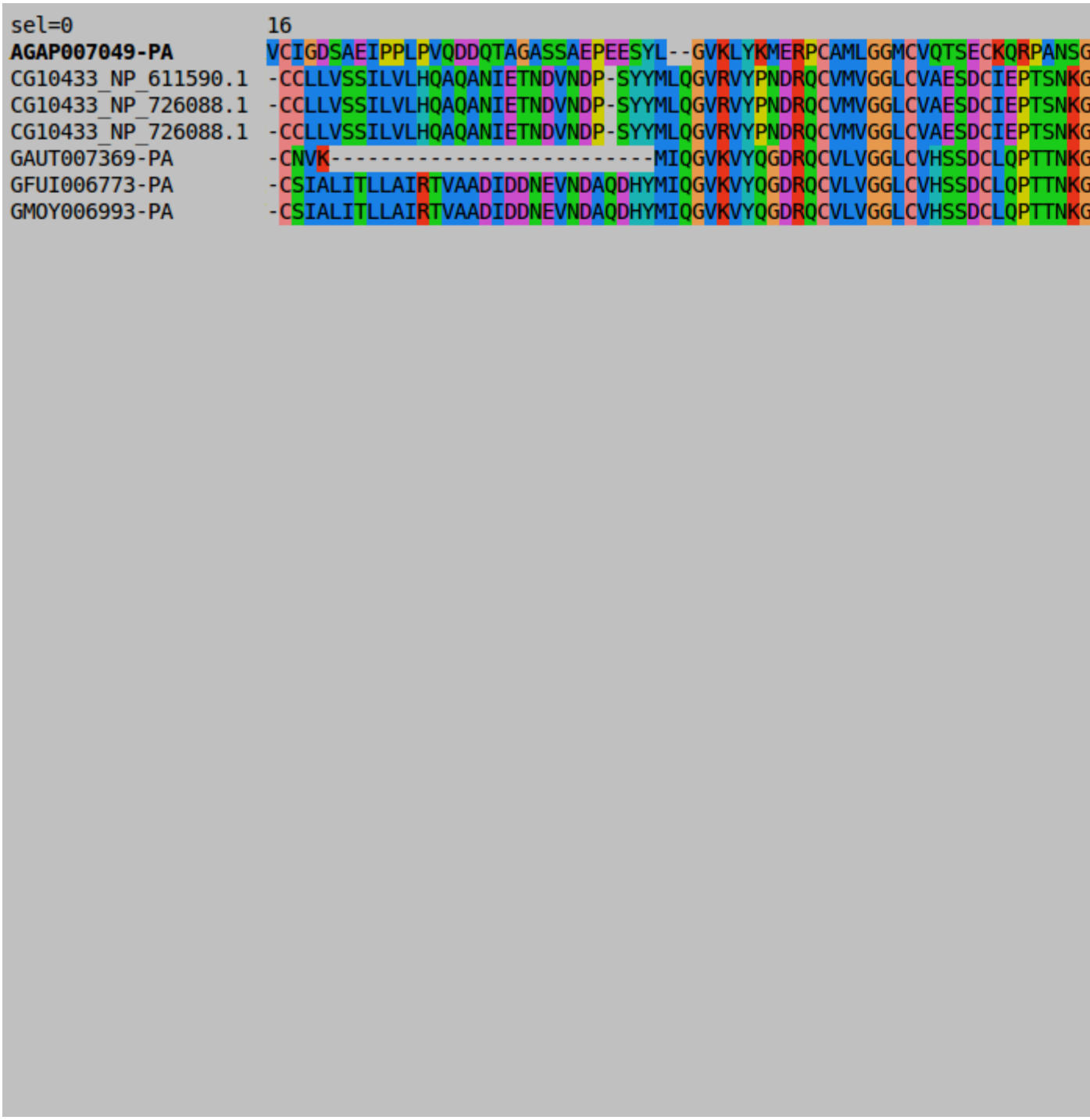
Screenshot displaying multiple alignment of accessory gland ACP present in *D.melanogaster* (CG14770), *G.austeni* (GAUT), *G.brevipalpis* (GBRI), *G.fuscipes* (GFUI), *G.morsitans* (GMOY) and *G.pallidipes* (GPAI)



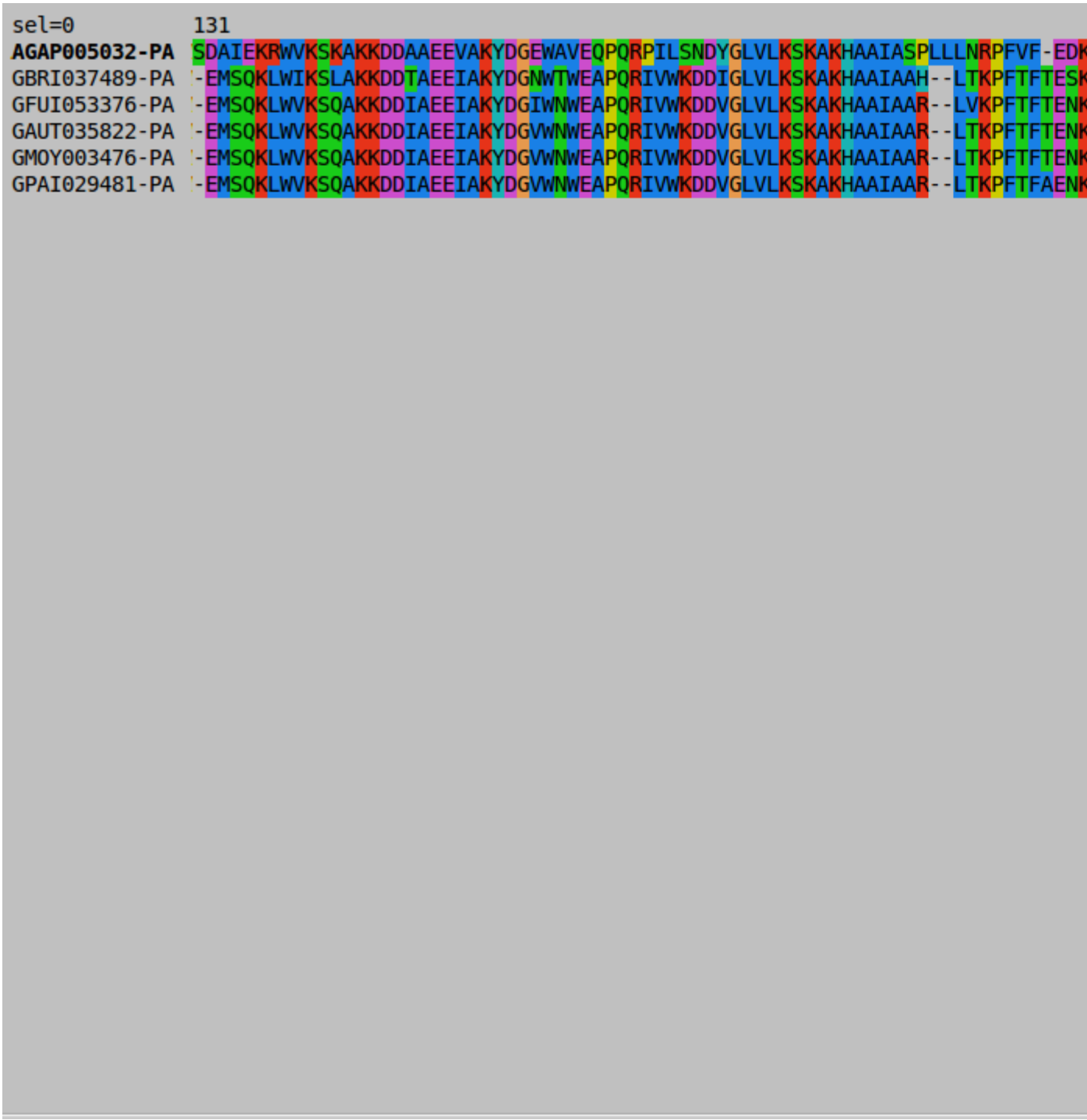
Screenshot displaying multiple alignment of proline oxidase ACP present in *A.gambiae* (AGAP), *G.austeni* (GAUT), *G.brevipalpis* (GBRI), *G.fuscipes* (GFUI), *G.morsitans* (GMOY) and *G.pallidipes* (GPAI)



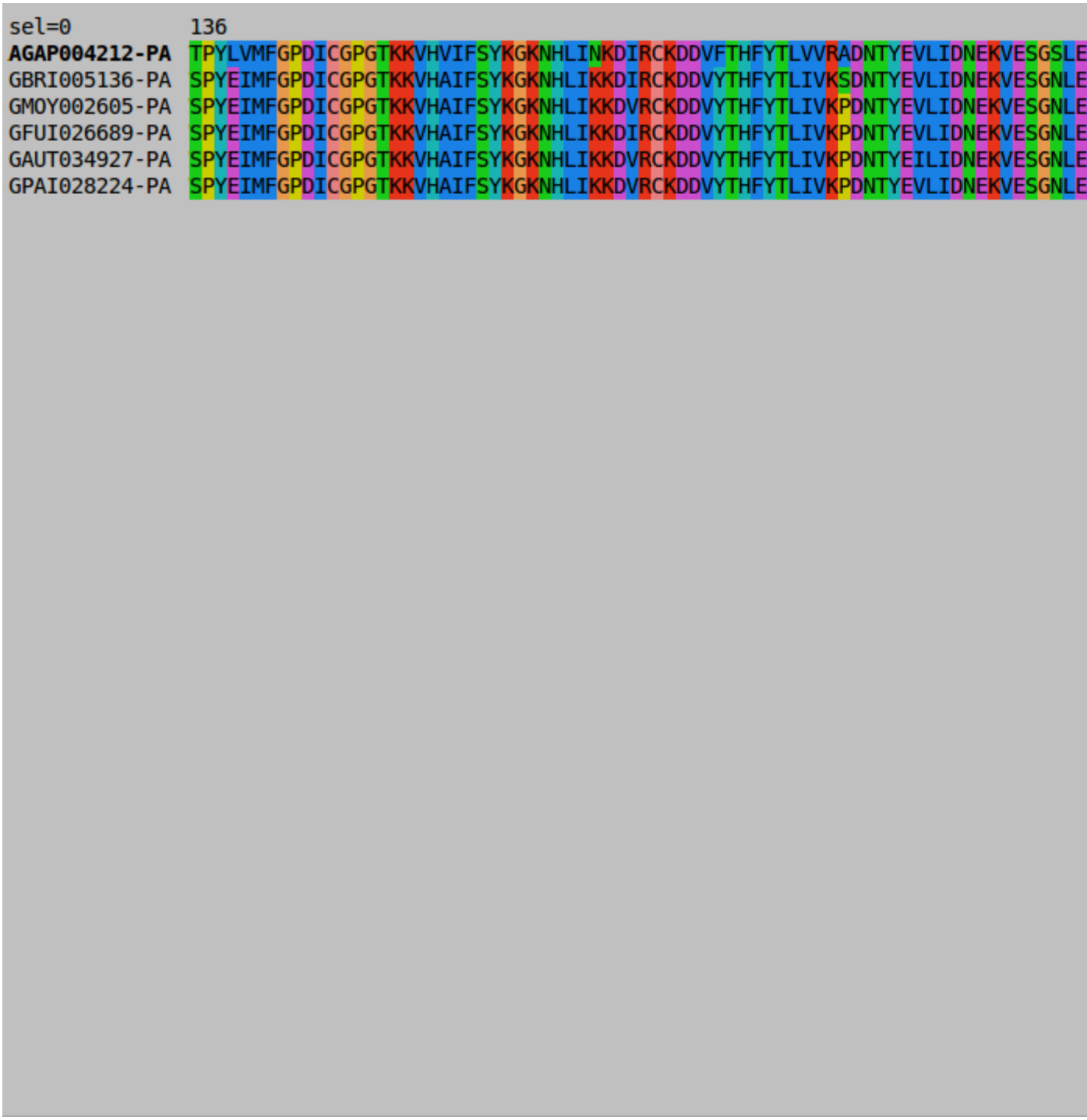
Screenshot displaying multiple alignment of transport protein ACP present in *A.gambiae* (AGAP), *G.austeni* (GAUT), *G.brevipalpis* (GBRI), *G.fuscipes* (GFUI), *G.morsitans* (GMOY) and *G.pallidipes* (GPAI)



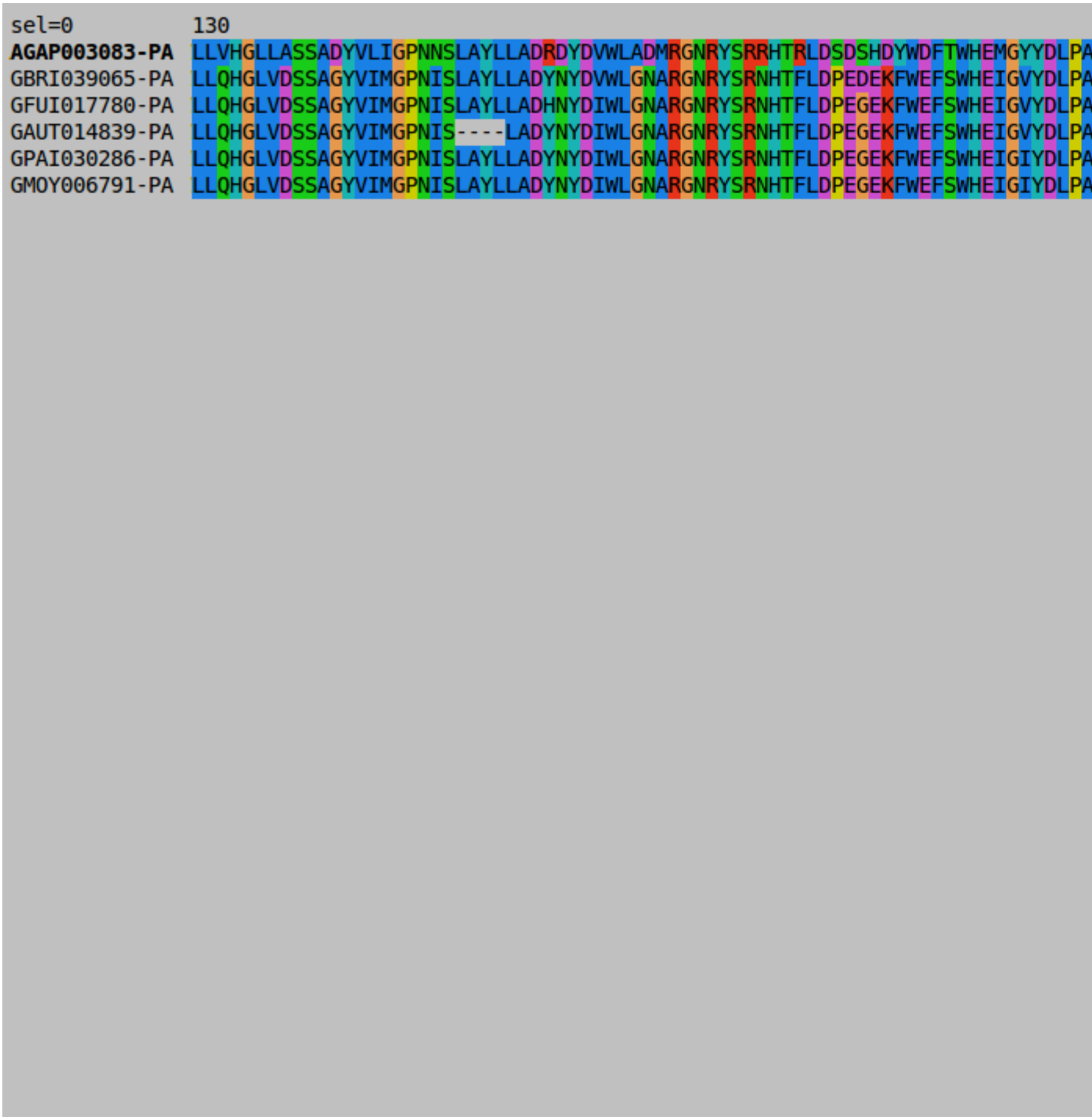
Screenshot displaying multiple alignment of beta-defensin ACP present in *A.gambiae* (AGAP), *D.melanogaster* (CG10433), *G.austeni* (GAUT), *G.fuscipes* (GFUI) and *G.morsitans* (GMOY)



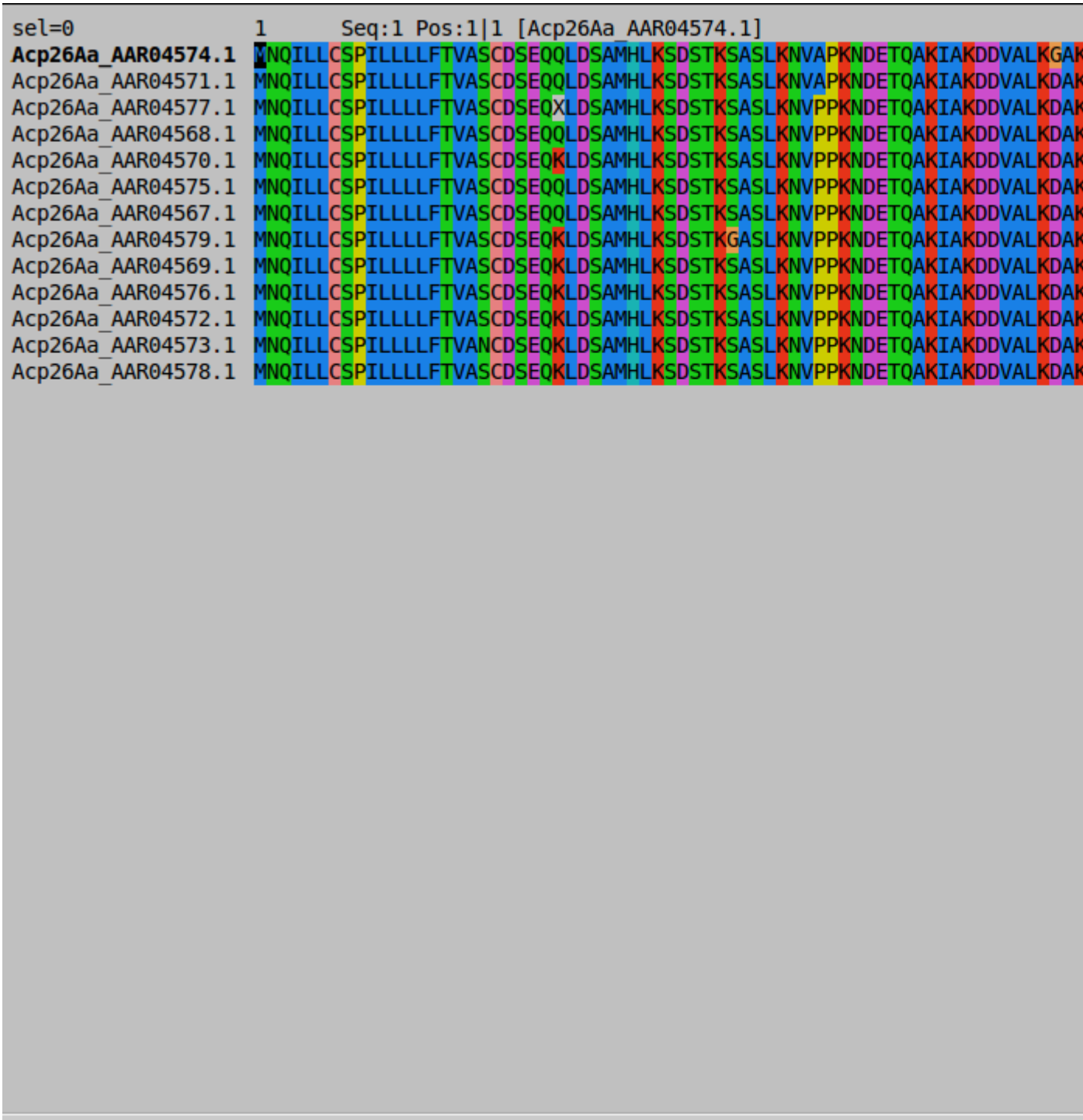
Screenshot displaying multiple alignment of calnexin ACP present in *A.gambiae* (AGAP), G.austeni (GAUT), *G.brevipalpis* (GBRI), *G,fuscipes* (GFUI), *G.morsitans* (GMOY) and *G.pallidipes* (GPAI)



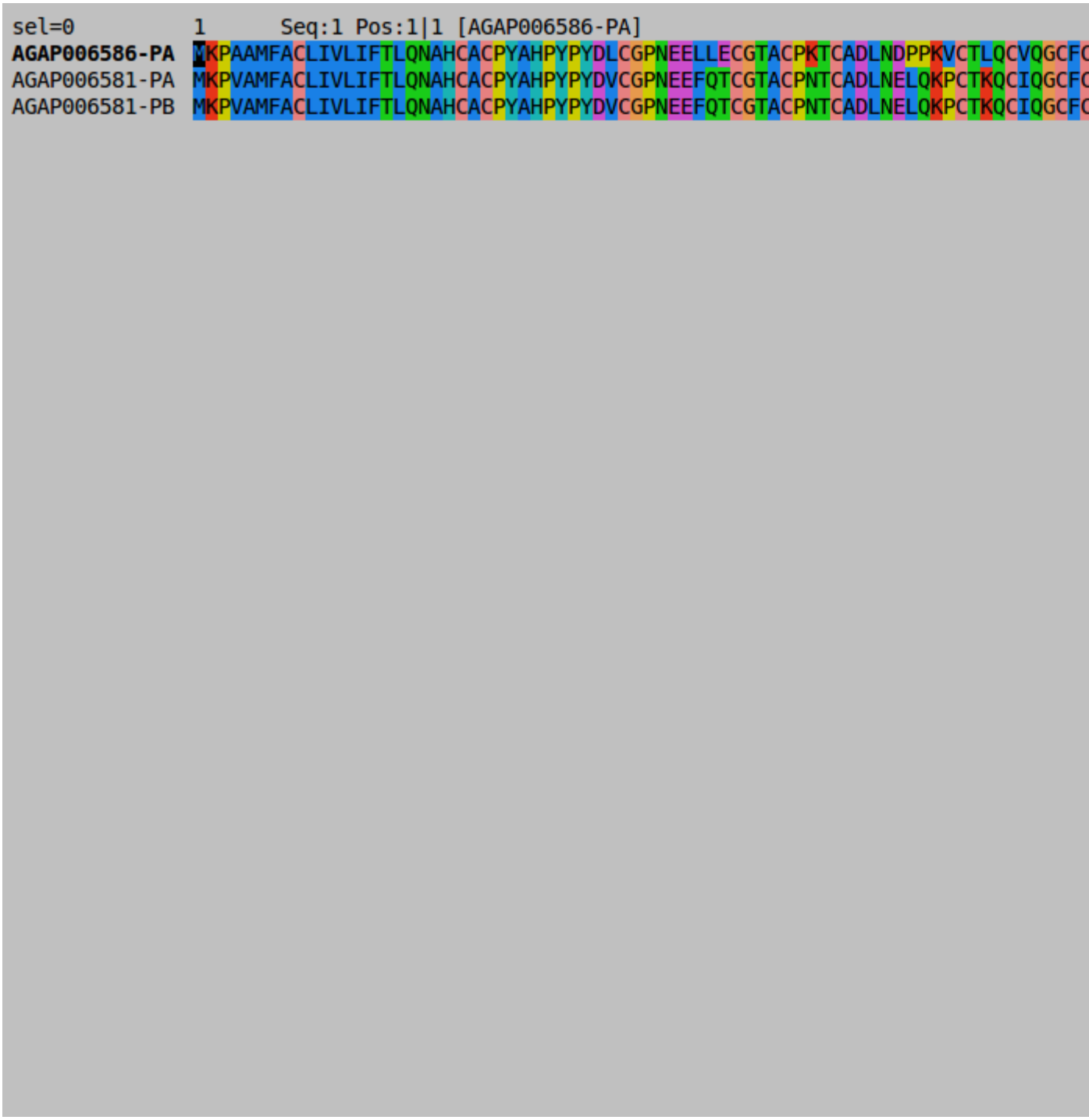
Screenshot displaying multiple alignment of chaperone ACP present in *A.gambiae* (AGAP), *G.austeni* (GAUT), *G.breviplapis* (GBRI), *G.fuscipes* (GFUI), *G.morsitans* (GMOY) and *G.pallidipes* (GPAI)



Screenshot displaying multiple alignment of lipase ACP present in *A.gambiae* (AGAP), *G.brevipalpis* (GBRI), *G.fuscipes* (GFUI), *G.morsitans* (GMOY) and *G.pallidipes (*GPAI)



Screenshot displaying multiple alignment of accessory protein 26Aa ACP present in *D.melanogaster* (AAR0)



Screenshot displaying multiple alignment of protease inhibitor ACP present in *A.gambiae* (AGAP)

Screenshot displaying multiple alignment of cytochrome ACP present in *A.gambiae* (AGAP), *D.melanogaster* (Cyp9f2), *G.austeni* (GAUT), *G.brevipalpis* (GBRI), *G.fuscipes* (GFUI), *G.morsitans* (GMOY) and *G.pallidipes* (GPAI)