RESEARCH NOTE

REVISED Comparative genomics identifies male accessory gland proteins in five *Glossina* species [version 2; referees: 2 approved]

Muna F. Abry[1,2], Kelvin M. Kimenyi[1,2], Daniel Masiga[2], Benard W. Kulohoma[1,2]

[1]Center for Biotechnology and Bioinformatics, University of Nairobi, P.O. Box 30197, Nairobi, 00100, Kenya
[2]International Centre for Insect Physiology and Ecology, P.O. Box 30772, Nairobi, 00100, Kenya

v2 **First published:** 30 Aug 2017, **2**:73 (doi: 10.12688/wellcomeopenres.12445.1)
Latest published: 22 Nov 2017, **2**:73 (doi: 10.12688/wellcomeopenres.12445.2)

Abstract

Accessory gland proteins (ACPs) are important reproductive proteins produced by the male accessory glands (MAGs) of most insect species. These proteins are essential for male insect fertility, and are transferred alongside semen to females during copulation. ACPs are poorly characterized in *Glossina* species (tsetse fly), the principal vector of the parasite that causes life-threatening Human African Trypanosomiasis and Animal trypanosomiasis in endemic regions in Africa. The tsetse fly has a peculiar reproductive cycle because of the absence of oviposition. Females mate once and store sperm in a spermathecal, and produce a single fully developed larva at a time that pupates within minutes of exiting their uterus. This slow reproductive cycle, compared to other insects, significantly restricts reproduction to only 3 to 6 larvae per female lifespan. This unique reproductive cycle is an attractive vector control strategy entry point. We exploit comparative genomics approaches to explore the diversity of ACPs in the recently available whole genome sequence data from five tsetse fly species (*Glossina morsitans, G. austeni, G. brevipalpis, G. pallidipes* and *G. fuscipes*). We used previously described ACPs in *Drosophila melanogaster* and *Anopheles gambiae* as reference sequences. We identified 36, 27, 31, 29 and 33 diverse ACP orthologous genes in *G. austeni, G. brevipalpis, G. fuscipes, G. pallidipes* and *G. morsitans* genomes respectively, which we classified into 21 functional classes. Our findings provide genetic evidence of MAG proteins in five recently sequenced *Glossina* genomes. It highlights new avenues for molecular studies that evaluate potential field control strategies of these important vectors of human and animal disease.

Open Peer Review

Referee Status: ✓ ✓

	Invited Referees 1	2
REVISED **version 2** published 22 Nov 2017		✓ report
version 1 published 30 Aug 2017	✓ report	? report

1 **Emiliano Mancini**, Roma Tre University, Italy

2 **Kristipati Ravi Ram**, Council of Scientific & Industrial Research (CSIR)- Indian Institute of Toxicology Research, India

Discuss this article

Comments (0)

Corresponding author: Benard W. Kulohoma (bkulohoma@iscb.org)

Author roles: Abry MF: Formal Analysis, Methodology, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Kimenyi KM**: Formal Analysis, Methodology, Writing – Review & Editing; **Masiga D**: Conceptualization, Supervision, Writing – Review & Editing; **Kulohoma BW**: Conceptualization, Formal Analysis, Funding Acquisition, Investigation, Methodology, Supervision, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

How to cite this article: Abry MF, Kimenyi KM, Masiga D and Kulohoma BW. **Comparative genomics identifies male accessory gland proteins in five *Glossina* species [version 2; referees: 2 approved]** Wellcome Open Research 2017, **2**:73 (doi: 10.12688/wellcomeopenres.12445.2)

Grant information: The work was supported by the Wellcome Trust [087540], a pump-priming grant to BWK from the Training Health Researchers into Vocational Excellence in East Africa (THRiVE); UK's Department for International Development (DFID); Swedish International Development Cooperation Agency (Sida); the Swiss Agency for Development and Cooperation (SDC); and the Kenyan Government.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

First published: 30 Aug 2017, **2**:73 (doi: 10.12688/wellcomeopenres.12445.1)

REVISED **Amendments from Version 1**

We have revised the Abstract, Introduction and Discussion sections to address the over-emphasis that ACPS play a crucial role in the tsetse fly life reproductive cycle, and thus vector control.

We have revised the Discussion section to show that this is a preliminary study using the initial release of the Glossina genomes, which are publicly available, and future studies using RNASeq/transcriptome datasets of the male accessory glands and testis may identify rapidly evolving Glossina-specific ACPs, which perhaps do not bear a well-known protein domains, and could escape *in silico* investigation dependent on a comparative genomic approach alone.

We have amended the Results and Discussion sections to reflect that we identified possible signatures of selection, since a wider analysis including more closely related taxa was not performed.

We used a different approach (outlined in the Methods section with references 17 & 18) to that used by Scolari *et al.*, 2016. Our approach is able to distinguish functional redundancy (orthologs from paralogs) from BLASTp reciprocal hits; and resolve the many-to-many orthologous relationships inherent in comparisons across multiple genomes (reference 17). Only orthologs with a reciprocal BLASTP E-value cut-off 1e-5 and inflation index 2.5 (default settings) were considered for further analysis. We also provide annotated (gene name and exact genome location) sequences for each of the orthologous ACP clusters in the Supplementary materials section.

Further, we have examined transcriptome results from Scolari *et al.*, 2016 (reference 37), which uses a different approach (reciprocal BLASTp hits). This analysis only identified 4 non-annotated ACPs (GMOY000024, GMOY007757, GMOY009744, GMOY012189) orthologous to *D. melanogaster* not detected in our analysis; and no orthologs to *A. gambiae* ACPs.

See referee reports

Introduction

Accessory gland proteins (ACPs) are important reproductive proteins produced by the male accessory glands (MAGs) of most insect species. These proteins are essential for male insect fertility, and are transferred alongside semen to females during copulation[1]. ACPs trigger significant physiological and behavioral changes in females after copulation, which include: egg laying, reduced sexual receptivity and refractoriness to subsequent inseminations, induce the expression of immune peptides and reduction of female lifespan[1–5]. ACPs are only resynthesized after transfer of seminal fluid to females, but topical application of juvenile hormone on the male's cuticles stimulates *in vivo* re-synthesis to pre-mating levels[2]. Female *Anopheles gambiae* mosquitoes copulated by males with degenerate testes and MAGs fail to oviposit and readily re-mate[1]. Conversely, those copulated by males with degenerate testes but fully developed MAGs lay unfertilized eggs and do not re-mate[1]. This underscores the relevance of ACPs as an entry point for vector borne disease control.

ACPs are poorly characterized in *Glossina* (tsetse fly), compared to *Drosophila* and *Anopheles* species[1,6]. The tsetse fly is the principal vector of the parasite that causes life-threatening human (sleeping sickness) and cattle (nagana) trypanosomiasis in endemic regions in Africa[7]. Over 60 million people and 80 million cattle are at risk of contracting disease[8]. Female tsetse flies only mate once during their lifespan and store the male ejaculate in their spermathecae, which they subsequently use to self-fertilize[9]. They have a peculiar reproductive cycle because of the absence of oviposition, with females producing a single fully developed larva at a time that pupates within minutes of exiting their uterus. This slow reproductive cycle, compared to other insects, significantly restricts reproduction to only 3 to 6 larvae per female lifespan[10]. This unique reproductive cycle is an attractive vector control target. An improved understanding of tsetse fly's reproductive biology, and specifically ACPs that are crucial determinants of successful reproduction in other insect species, may provide valuable possible vector control strategy entry points.

Comparisons between ACP gene orthologs in *Drosophila simulans* and *D. melanogaster* show they are rapidly evolving, relative to non-ACP genes[11–13]. However, there is strong ACP peptide structural homology between closely related species, which decreases as species phylogenetic distances increase[14]. This rapid rate of ACP genes evolution has made it challenging to reliably identify orthologs across insect species in the absence of genomic data[15,16]. The recently available whole genome sequence data from five tsetse fly species (*Glossina morsitans*, *G. austeni*, *G. brevipalpis*, *G. pallidipes* and *G. fuscipes*) has made it possible to revisit detailed examination of ACP gene distribution and genetic diversity in tsetse flies. We exploit comparative genomics approaches to interrogate these genomes, using previously described ACPs in *D. melanogaster* and *A. gambiae* as reference sequences.

Methods

Identification of *Glossina*, *Anopheles* and *Drosophila* species ACP homologs

The 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 previously described ACP sequences from *A. gambiae* (n=57) and *D. melanogaster* (n=173) (Supplementary Table 1)[1], used as references, were assigned to homologous clusters using OrthoMCL with default settings (BLASTP E-value cut-off 1e-5 and inflation index 2.5)[17] (Supplementary File 1). Clusters with singletons were omitted from further processing. Mapped orthologs were subsequently processed using BMX[18].

Sequence alignment and phylogeny reconstruction

Multiple sequence alignments were performed using MUSCLE[19]. Maximum likelihood (ML) phylogenetic analysis of the multiple aligned sequences with bootstrap values of 100 replicates was performed using PHYML version 3.5[20].

Determining the direction and extent of selection pressure

The magnitude and direction of selection pressure on the ACP sequences was tested based on the ratio ($\omega = d_N/d_S$) of the average number of non-synonymous substitutions per non-synonymous site (d_N) to the average number of synonymous substitutions per synonymous site (d_S). If $\omega = 1$, amino acid substitution is assumed to be under neutral selection, $\omega > 1$ is indicative of positive selection whereas $\omega < 1$ is evidence of negative or purifying selection. Sequence alignments of each of the ACP clusters

containing *A. gambiae*, *D. melanogaster*, *G. austeni*, *G. brevipalpis*, *G. fuscipes*, *G. morsitans* and *G. pallidipes* were generated. Each alignment was then uploaded to the SNAP program[21] (www.hiv.lanl.gov), which calculates synonymous and non-synonymous substitution rates to determine the magnitude of selection pressure.

Data visualization

A list of the relative abundance of ACPs with secretory signals identified within each ortholog cluster for *Glossina, Anopheles* and *Drosophila* species was generated (Supplementary Figure 1). The 21 ACPs clusters were visually presented in a single circular ideogram using CIRCOS software[22].

Results

Identification of ACP gene orthologs in *Glossina* species

We analyzed five recently sequenced *Glossina* genomes: *Glossina morsitans*, *G. austeni*, *G. brevipalpis*, *G. pallidipes* and *G. fuscipes*, to examine the presence of orthologs to ACP genes previously identified in *Drosophila melanogaster* (n=173) and *Anopheles gambiae* (n=57)[1,6]. First, we identified 41 ACP ortholog clusters that we broadly classified into 23 groups based on the encoded protein's functional class. These ACPs have a distinct species distribution with: 12 ortholog clusters common to *Glossina, A. gambiae,* and *D. melanogaster* species; and some clusters only present in *A. gambiae* and *Glossina* species (n=7) or *D. melanogaster* and *Glossina* species (n=5) (Supplementary Table 1). The remaining 17 clusters consist of ACP orthologs exclusive to either *A. gambiae* (n=7) or *D. melanogaster* (n=10). Next, we shortlisted genes that encode proteins carrying classical secretory signals[23] to distinguish the matched testes-specific secreted male accessory gland proteins from other insect peptides. We identified 36, 27, 31, 29 and 33 ACP orthologs with secretory signals in *G. austeni*, *G. brevipalpis*, *G. fuscipes, G. pallidipes* and *Glossina morsitans* genomes respectively (Figure 1), across 21 functional class groups. α2-macroglobulins (Group 1) and heat shock proteins (Group 17) are the most abundant ACP orthologs in *Glossina* species (Figure 1). Interestingly, *Glossina* species lack orthologs to Acp70A (Group 18) and andropin (Group 19), which has antimicrobial properties and safeguards the male ejaculate, and stimulation of long-term post mating responses in females respectively[1,24,25].

Most ACP genes are under positive selection

We inferred the direction and magnitude of selection pressure on the identified ACP orthologs using dN/dS ratios. We observed possible signatures of positive selection in all genes except five are evolving under positive selection (Figure 1). We found that α2-macroglobulins, which have been shown to be important in mosquito and *Drosophila* immunity[26,27], display signatures of purifying selection suggesting they are critical for successful reproduction and all deleterious variations are purged. Our analysis was restricted to reference genes present in *A. gambiae* and *D. melanogaster*, and future studies that integrate data from more closely related taxa will highlight evolutionary changes associated with ACPs in more detail. We reconstructed the phylogeny of ACP orthologs within each cluster (Supplementary Figure 1). We failed to identify any pattern associated with the diverse ecological niche and unique reproductive style in *Glossina* species in our analysis.

ACP distribution in *Glossina* species

Distribution of ACP orthologs varies widely between species (Figure 1). *Glossina* species have a disproportionately large number of α2-macroglobublin and heat shock proteins, which are important in immunity in *A. gambiae* and *Glossina* species[26–29]. Our analysis did not detect β-defensin orthologs, which are antimicrobial peptides involved in immune responses[30], in the *G. pallidipes* and *G. brevipalpis* genomes. We also did not detect Acp29AB, Acp70A, and andropin orthologs in this comprehensive catalogue of *Glossina* genes. This raises the possibility that these genes were lost by tsetse flies after evolutionary radiation of insects into multiple taxa, and alternative species-specific proteins might compensate for the same roles.

Discussion

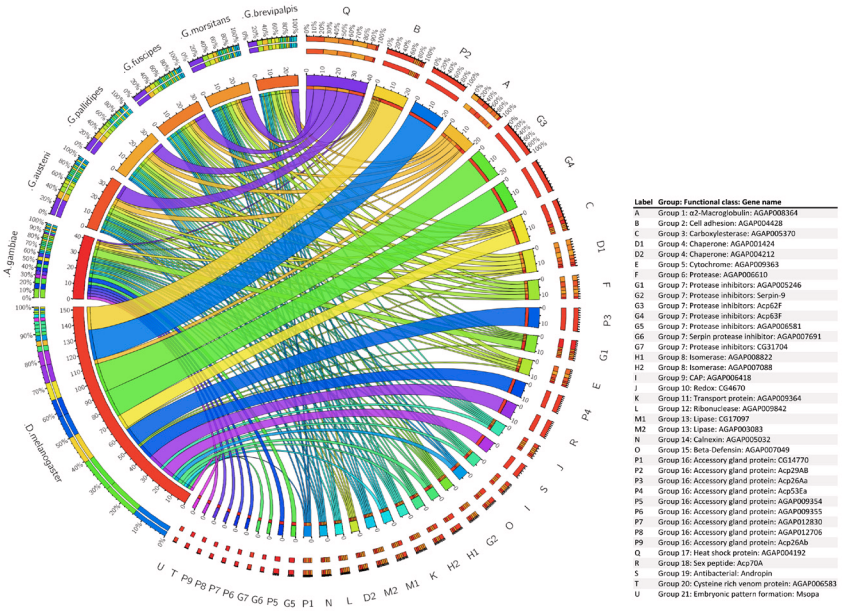
We performed comparative genomics analysis to detect the presence of male accessory gland proteins (ACPs) orthologs previously identified in *A. gambiae* and *Drosophila*[1]. The motivation here was to improve knowledge on the biology of *Glossina* species ACPs given the importance of reproductive molecules in strategic designs of vector control. We identified 21 functional classes of ACP orthologs with secretory signals in five *Glossina* species genomes. We observed genetic signatures of a high rate of ACP protein divergence, supporting similar findings on male reproduction-related genes in *Drosophila*[16]. ACPs exhibit high evolutionary changes, thus displaying between species divergence and within species polymorphism[3,31]. We restricted analysis to reference genes present in *A. gambiae* and *D. melanogaster*, and future studies integrating datasets from more closely related taxa will be useful to understand evolutionary changes associated with ACPs in more detail. α2-macroglobulins and heat shock proteins are the most abundant ACP orthologs in *Glossina* species. α2-macroglobulins are important in mosquito and *Drosophila* immunity[26,27], and display signatures of purifying selection, suggesting they are critical for successful reproduction and all deleterious variations are purged. α2-macroglobulin over-representation, and the absence of other ACP orthologs implicated in immunity in the *Glossina* genomes points to their critical role in ensuring successful tsetse fly reproduction. Heat shock protein silencing in *A. gambiae* down-regulates up to 50% of male accessory gland proteins, half of which are male reproductive tract specific and encode the homologs of 13 known *Drosophila* ACPs that include Acp70A[1]. Interestingly, *Glossina* species lack orthologs to Acp70A, andropin, Acp26Ab, Acp29AB, and Acp62F, which play critical roles in successful reproduction in *Anopheles* and *Drosophila* species[1,24,25].

Acp70A or sex peptide stimulates long-term post mating behavior, resulting in non-receptivity to mating and increased oviposition[1,25,32]. Andropin is an antimicrobial peptide transferred to the female during copulation, and defends the female reproductive tract against microbes[33]. Andropin also protects the male ejaculate from Gram-positive and Gram-negative bacterial infections[1]. Acp26Ab stimulates oviposition in *Drosophila melanogaster* females[34], and together with Acp26Ab protects the male ejaculate from microbial infections, and displacement

A **B** **C** **D**

Group	Functional class	Gene name	ds	dn	dn/ds	G. austeni	G. brevipalpis	G. fuscipes	G. morsitans	G. pallidipes	A. gambiae	D. melanogaster
Group 1	α2-Macroglobulin	AGAP008364	0.496	0.467	0.940	4	3	3	2	4	1	3
Group 2	Cell adhesion	AGAP004428	0.124	0.203	1.636	1	1	1	1	1	1	16
Group 3	Carboxylesterase	AGAP005370	0.243	0.487	2.002	1	2	1	1	1	2	10
Group 4	Chaperone	AGAP001424	0.098	0.271	2.767	2	2	3	3	3	1	1
	Chaperone	AGAP004212	0.043	0.092	2.144	1	1	1		1	1	
Group 5	Cytochrome	AGAP009363	0.180	0.238	1.319	2	3	1		2	1	1
Group 6	Protease	AGAP006610	0.320	0.315	0.984	3	2	2	3	2	1	
Group 7	Protease inhibitors	AGAP005246	0.172	0.349	2.033	1	1	1	1	1	5	1
	Protease inhibitors	Serpin-9	0.137	0.199	1.453	1	1	1	1	1	1	
	Protease inhibitors	Acp62F	NA	NA	NA							20
	Protease inhibitors	Acp63F	0.010	0.014	1.392							19
	Protease inhibitors	AGAP006581	0.107	0.118	1.097						3	
	Serpin protease inhibitor	AGAP007691	0.024	0.028	1.173						2	
	Protease inhibitors	CG31704	NA	NA	NA							2
Group 8	Isomerase	AGAP008822	0.075	0.106	1.426	1			1		1	3
	Isomerase	AGAP007088	0.104	0.079	0.755	1		1		1	1	2
Group 9	CAP	AGAP006418	0.445	0.370	0.831	1		1	1	1	1	2
Group 10	Redox	CG4670	0.235	0.260	1.109	1	1	1	1	1	1	2
Group 11	Transport protein	AGAP009364	0.054	0.160	2.978	1	1	1	1	1	1	
Group 12	Ribonuclease	AGAP009842	0.136	0.234	1.721		1	1	1	1	1	
Group 13	Lipase	CG17097	0.111	0.287	2.580	1	1	1		1		2
	Lipase	AGAP003083	0.150	0.168	1.124	1	1	1	1	1	1	
Group 14	Calnexin	AGAP005032	0.066	0.130	1.979	1		1	1	1	1	
Group 15	Beta-Defensin	AGAP007049	0.121	0.381	3.157	1		1	1		1	3
Group 16	Accessory gland protein	CG14770	0.064	0.118	1.865	1	1		1	1		1
	Accessory gland protein	Acp29AB	0.004	0.005	1.333							21
	Accessory gland protein	Acp26Aa	0.006	0.008	1.407							13
	Accessory gland protein	Acp53Ea	0.010	0.013	1.326							10
	Accessory gland protein	AGAP009354	0.011	0.024	2.259						3	
	Accessory gland protein	AGAP009355	NA	NA	NA						2	
	Accessory gland protein	AGAP012830	0.028	0.019	0.682						2	
	Accessory gland protein	AGAP012706	0.039	0.097	2.482						2	
	Accessory gland protein	Acp26Ab	NA	NA	NA						2	
Group 17	Heat shock protein	AGAP004192	0.099	0.291	2.939	10	5	8	8	8	1	
Group 18	Sex peptide	Acp70A	NA	NA	NA							10
Group 19	Antibacterial	Andropin	0.025	0.026	1.032							8
Group 20	Cysteine rich venom protein	AGAP006583	0.171	0.340	1.993						2	
Group 21	Embryonic pattern formation	Msopa	NA	NA	NA							2
		Orthologous ACPs identified				36	27	31	29	33	42	152

E



Label	Group: Functional class: Gene name
A	Group 1: α2-Macroglobulin: AGAP008364
B	Group 2: Cell adhesion: AGAP004428
C	Group 3: Carboxylesterase: AGAP005370
D1	Group 4: Chaperone: AGAP001424
D2	Group 4: Chaperone: AGAP004212
E	Group 5: Cytochrome: AGAP009363
F	Group 6: Protease: AGAP006610
G1	Group 7: Protease inhibitors: AGAP005246
G2	Group 7: Protease inhibitors: Serpin-9
G3	Group 7: Protease inhibitors: Acp62F
G4	Group 7: Protease inhibitors: Acp63F
G5	Group 7: Protease inhibitors: AGAP006581
G6	Group 7: Serpin protease inhibitor: AGAP007691
G7	Group 7: Protease inhibitors: CG31704
H1	Group 8: Isomerase: AGAP008822
H2	Group 8: Isomerase: AGAP007088
I	Group 9: CAP: AGAP006418
J	Group 10: Redox: CG4670
K	Group 11: Transport protein: AGAP009364
L	Group 12: Ribonuclease: AGAP009842
M1	Group 13: Lipase: CG17097
M2	Group 13: Lipase: AGAP003083
N	Group 14: Calnexin: AGAP005032
O	Group 15: Beta-Defensin: AGAP007049
P1	Group 16: Accessory gland protein: CG14770
P2	Group 16: Accessory gland protein: Acp29AB
P3	Group 16: Accessory gland protein: Acp26Aa
P4	Group 16: Accessory gland protein: Acp53Ea
P5	Group 16: Accessory gland protein: AGAP009354
P6	Group 16: Accessory gland protein: AGAP009355
P7	Group 16: Accessory gland protein: AGAP012830
P8	Group 16: Accessory gland protein: AGAP012706
P9	Group 16: Accessory gland protein: Acp26Ab
Q	Group 17: Heat shock protein: AGAP004192
R	Group 18: Sex peptide: Acp70A
S	Group 19: Antibacterial: Andropin
T	Group 20: Cysteine rich venom protein: AGAP006583
U	Group 21: Embryonic pattern formation: Msopa

Figure 1. Identification of male accessory gland proteins (ACPs) in five *Glossina* species. (**A**) We identified 38 clusters of ACPs, which were classified into 21 functional classes. (**B**) The identified *Glossina* species ACPs are orthologous to well-characterized ACP genes in *Anopheles gambiae* and *Drosophila melanogaster genomes*. (**C**) The magnitude and direction of selection pressure on the ACP sequences was tested based on the ratio ($\omega = d_N/d_S$) of the average number of non-synonymous substitutions per non-synonymous site (d_N) to the average number of synonymous substitutions per synonymous site (d_S). A ratio greater than 1 indicates positive selection, and a ratio less than 1 indicates purifying selection. (**D**) The relative abundance of ACPs identified per ortholog cluster. We identified 36, 27, 31, 29 and 33 ACP orthologs with secretory signals in *G. austeni, G. brevipalpis, G. fuscipes, G. pallidipes* and *Glossina morsitans* genomes, respectively. (**E**) Schematic presentation showing the relative abundance of each ACP functional class in *Glossina* species, *A. gambiae*, and *Drosophila melanogaster*. α2-macroglobulins [A] and, heat shock proteins [Q] are the most abundant ACPs in *Glossina* species.

by a second ejaculate[1]. *Drosophila* Acp29AB and Acp62F up-regulate genes for egg production and muscle development, although Acp29AB or Acp62F null males do not show a reproduction impairment phenotype[35]. Acp62F also protects sperm in the female reproductive tract from protease attack[36].

A limitation in our study was the absence of transcriptome data to measure ACP differential gene expression. Analysis of transcriptome data available to others using a different approach (reciprocal BLASTp hits) identified only 4 non-annotated ACPs (GMOY000024, GMOY007757, GMOY009744, GMOY012189) orthologous to *D. melanogaster* not detected in our analysis; and no orthologs to *A. gambiae* ACPs[37]. Future studies focused on transcriptome datasets of male accessory glands and testis may identify rapidly evolving *Glossina*-specific ACPs, which perhaps do not bear a well-known protein domains, and could escape *in silico* investigation dependent on a comparative genomic approach alone.

Our analysis detected orthologous *Glossina, Anopheles* and *Drosophila* ACPs belonging to the same functional classes, suggesting a conserved role for these proteins across all three genera. However, some ACPs may represent lineage-specific ACPs that may have evolved to perform species-specific reproductive functions. Our findings support evolutionary adaptation to different reproductive styles. Tsetse fly females produce a single fully developed larva at a time that pupates within minutes of exiting their uterus, and may have lost non-essential ACP genes after adaptation.

Conclusions

Our findings provide genetic evidence of male accessory glad proteins in five recently sequenced *Glossina* genomes. It provides new avenues for molecular studies that evaluate potential field control strategies of these important vectors of human and animal disease.

Data availability

Anopheles gambiae and *Drosophila melanogaster* ACPs were obtained from: DOI, 10.1073/pnas.0703904104[1].

Glossina species genome sequences were obtained from VectorBase.

Competing interests

No competing interests were disclosed.

Grant information

The work was supported by the Wellcome Trust [087540], a pump-priming grant to BWK from the Training Health Researchers into Vocational Excellence in East Africa (THRiVE); UK's Department for International Development (DFID); Swedish International Development Cooperation Agency (Sida); the Swiss Agency for Development and Cooperation (SDC); and the Kenyan Government.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgments

We would like to thank Abraham Parwos, Anne Owiti, Edwin Rono, and Dr. George Obiero, and support from the University of Nairobi's Centre for Biotechnology and Bioinformatics, and ICIPE.

This work is published with permission from the University of Nairobi and ICIPE.

Supplementary material

Supplementary Table 1: (A) *Anopheles gambiae* (n=57) and *Drosophila melanogaster* (n=173) ACP orthologs. (B) ACP ortholog clusters identified in *Glossina*, *A. gambiae*, and *D. melanogaster* species prior to signal-peptide filtering.

Click here to access the data.

Supplementary File 1: Amino acid sequences of the ACP ortholog clusters.

Click here to access the data.

Supplementary Figure 1: Reconstructed phylogenies of the ACP orthologs within each cluster.

Click here to access the data.

References

1. Dottorini T, Nicolaides L, Ranson H, *et al.*: **A genome-wide analysis in *Anopheles gambiae* mosquitoes reveals 46 male accessory gland genes, possible modulators of female behavior.** *Proc Natl Acad Sci U S A.* 2007; **104**(41): 16215–16220.
PubMed Abstract | Publisher Full Text | Free Full Text
2. Wolfner MF, Partridge L, Lewin S, *et al.*: **Mating and hormonal triggers regulate accessory gland gene expression in male *Drosophila*.** *J Insect Physiol.* 1997; **43**(12): 1117–1123.
PubMed Abstract | Publisher Full Text
3. Chapman T, Davies SJ: **Functions and analysis of the seminal fluid proteins of male *Drosophila melanogaster* fruit flies.** *Peptides.* 2004; **25**(9): 1477–1490.
PubMed Abstract | Publisher Full Text
4. Walker MJ, Rylett CM, Keen JN, *et al.*: **Proteomic identification of *Drosophila melanogaster* male accessory gland proteins, including a pro-cathepsin and a soluble γ-glutamyl transpeptidase.** *Proteome Sci.* 2006; **4**: 9.
PubMed Abstract | Publisher Full Text | Free Full Text
5. Tripet F, Touré YT, Dolo G, *et al.*: **Frequency of multiple inseminations in field-collected *Anopheles gambiae* females revealed by DNA analysis of transferred sperm.** *Am J Trop Med Hyg.* 2003; **68**(1): 1–5.
PubMed Abstract
6. Mueller JL, Ravi Ram K, McGraw LA, *et al.*: **Cross-species comparison of *Drosophila* male accessory gland protein genes.** *Genetics.* 2005; **171**(1): 131–143.
PubMed Abstract | Publisher Full Text | Free Full Text
7. Mpanya A, Hendrickx D, Vuna M, *et al.*: **Should I get screened for sleeping sickness? A qualitative study in Kasai province, Democratic Republic of Congo.** *PLoS Negl Trop Dis.* 2012; **6**(1): e1467.
PubMed Abstract | Publisher Full Text | Free Full Text
8. Grady SC, Messina JP, McCord PF: **Population vulnerability and disability in Kenya's tsetse fly habitats.** *PLoS Negl Trop Dis.* 2011; **5**(2): e957.
PubMed Abstract | Publisher Full Text | Free Full Text
9. Yuval B: **Mating systems of blood-feeding flies.** *Annu Rev Entomol.* 2006; **51**: 413–440.
PubMed Abstract | Publisher Full Text
10. Franco JR, Simarro PP, Diarra A, *et al.*: **Epidemiology of human African trypanosomiasis.** *Clin Epidemiol.* 2014; **6**: 257–275.
PubMed Abstract | Publisher Full Text | Free Full Text
11. Almeida FC, Desalle R: **Orthology, function and evolution of accessory gland proteins in the *Drosophila repleta* group.** *Genetics.* 2009; **181**(1): 235–245.
PubMed Abstract | Publisher Full Text | Free Full Text
12. Chen PS: **The accessory gland proteins in male *Drosophila*: structural, reproductive, and evolutionary aspects.** *Experientia.* 1996; **52**(6): 503–510.
PubMed Abstract | Publisher Full Text
13. Aguadé M: **Positive selection drives the evolution of the Acp29AB accessory gland protein in Drosophila.** *Genetics.* 1999; **152**(2): 543–551.
PubMed Abstract | Free Full Text
14. Davies SJ, Chapman T: **Identification of genes expressed in the accessory glands of male Mediterranean Fruit Flies (*Ceratitis capitata*).** *Insect Biochem Mol Biol.* 2006; **36**(11): 846–856.
PubMed Abstract | Publisher Full Text
15. Collins AM, Caperna TJ, Williams V, *et al.*: **Proteomic analyses of male contributions to honey bee sperm storage and mating.** *Insect Mol Biol.* 2006; **15**(5): 541–549.
PubMed Abstract | Publisher Full Text | Free Full Text
16. Begun DJ, Lindfors HA, Thompson ME, *et al.*: **Recently evolved genes identified from *Drosophila yakuba* and *D. erecta* accessory gland expressed sequence tags.** *Genetics.* 2006; **172**(3): 1675–1681.
PubMed Abstract | Publisher Full Text | Free Full Text
17. Li L, Stoeckert CJ Jr, Roos DS: **OrthoMCL: identification of ortholog groups for eukaryotic genomes.** *Genome Res.* 2003; **13**(9): 2178–2189.
PubMed Abstract | Publisher Full Text | Free Full Text
18. Kulohoma BW: **BMX: a tool for computing bacterial phyletic composition from orthologous maps.** *BMC Res Notes.* 2015; **8**: 51.
PubMed Abstract | Publisher Full Text | Free Full Text
19. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Res.* 2004; **32**(5): 1792–1797.
PubMed Abstract | Publisher Full Text | Free Full Text
20. Dereeper A, Guignon V, Blanc G, *et al.*: **Phylogeny.fr: robust phylogenetic analysis for the non-specialist.** *Nucleic Acids Res.* 2008; **36**(Web Server issue): W465–469.
PubMed Abstract | Publisher Full Text | Free Full Text
21. Korder B: **HIV Signature and Sequence Variation Analysis.** Netherlands: Kluwer Academic Publishers; 2000.
Reference Source
22. Krzywinski M, Schein J, Birol I, *et al.*: **Circos: an information aesthetic for comparative genomics.** *Genome Res.* 2009; **19**(9): 1639–1645.
PubMed Abstract | Publisher Full Text | Free Full Text
23. Wolfner MF: **The gifts that keep on giving: physiological functions and evolutionary dynamics of male seminal proteins in *Drosophila*.** *Heredity (Edinb).* 2002; **88**(2): 85–93.
PubMed Abstract | Publisher Full Text
24. Samakovlis C, Kylsten P, Kimbrell DA, *et al.*: **The andropin gene and its product, a male-specific antibacterial peptide in *Drosophila melanogaster*.** *EMBO J.* 1991; **10**(1): 163–169.
PubMed Abstract | Free Full Text
25. Yang G, Attardo GM, Lohs C, *et al.*: **Molecular characterization of two novel milk proteins in the tsetse fly (*Glossina morsitans morsitans*).** *Insect Mol Biol.* 2010; **19**(2): 253–262.
PubMed Abstract | Publisher Full Text | Free Full Text
26. Alves-Silva J, Ribeiro JM, Van Den Abbeele J, *et al.*: **An insight into the sialome of *Glossina morsitans morsitans*.** *BMC Genomics.* 2010; **11**: 213.
PubMed Abstract | Publisher Full Text | Free Full Text
27. Hughes GL, Ren X, Ramirez JL, *et al.*: ***Wolbachia* infections in *Anopheles gambiae* cells: transcriptomic characterization of a novel host-symbiont interaction.** *PLoS Pathog.* 2011; **7**(2): e1001296.
PubMed Abstract | Publisher Full Text | Free Full Text
28. Zügel U, Kaufmann SH: **Role of heat shock proteins in protection from and pathogenesis of infectious diseases.** *Clin Microbiol Rev.* 1999; **12**(1): 19–39.
PubMed Abstract | Free Full Text
29. Jain S, Rana V, Shrinet J, *et al.*: **Blood feeding and *Plasmodium* infection alters the miRNome of *Anopheles stephensi*.** *PLoS One.* 2014; **9**(5): e98402.
PubMed Abstract | Publisher Full Text | Free Full Text
30. Azevedo RV, Dias DB, Bretãs JA, *et al.*: **The transcriptome of *Lutzomyia longipalpis* (Diptera: Psychodidae) male reproductive organs.** *PLoS One.* 2012; **7**(4): e34495.
PubMed Abstract | Publisher Full Text | Free Full Text
31. Swanson WJ, Yang Z, Wolfner MF, *et al.*: **Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals.** *Proc Natl Acad Sci U S A.* 2001; **98**(5): 2509–2514.
PubMed Abstract | Publisher Full Text | Free Full Text
32. Ram KR, Wolfner MF: **Sustained post-mating response in *Drosophila melanogaster* requires multiple seminal fluid proteins.** *PLoS Genet.* 2007; **3**(12): e238.
PubMed Abstract | Publisher Full Text | Free Full Text
33. Strickler-Dinglasan PM, Guz N, Attardo G, *et al.*: **Molecular characterization of iron binding proteins from *Glossina morsitans morsitans* (Diptera: Glossinidae).** *Insect Biochem Mol Biol.* 2006; **36**(12): 921–933.
PubMed Abstract | Publisher Full Text | Free Full Text
34. Herndon LA, Wolfner MF: **A *Drosophila* seminal fluid protein, Acp26Aa, stimulates egg laying in females for 1 day after mating.** *Proc Natl Acad Sci U S A.* 1995; **92**(22): 10114–10118.
PubMed Abstract | Publisher Full Text | Free Full Text
35. Avila FW, Sirot LK, LaFlamme BA, *et al.*: **Insect seminal fluid proteins: identification and function.** *Annu Rev Entomol.* 2011; **56**: 21–40.
PubMed Abstract | Publisher Full Text | Free Full Text
36. Lung O, Tram U, Finnerty CM, *et al.*: **The *Drosophila melanogaster* seminal fluid protein Acp62F is a protease inhibitor that is toxic upon ectopic expression.** *Genetics.* 2002; **160**(1): 211–224.
PubMed Abstract | Free Full Text
37. Scolari F, Benoit JB, Michalkova V, *et al.*: **The Spermatophore in *Glossina morsitans morsitans*: Insights into Male Contributions to Reproduction.** *Sci Rep.* 2016; **6**: 20334.
PubMed Abstract | Publisher Full Text | Free Full Text

Open Peer Review

Current Referee Status: ✔ ✔

Version 2

Referee Report 06 December 2017

doi:10.21956/wellcomeopenres.14370.r28295

✔ **Kristipati Ravi Ram**
Embryotoxicology Laboratory, Environmental Toxicology Group, Council of Scientific & Industrial Research (CSIR)- Indian Institute of Toxicology Research, Lucknow, Uttar Pradesh, India

Authors have adequately addressed the concerns raised on the previous version.

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Referee Report 23 October 2017

doi:10.21956/wellcomeopenres.13476.r26501

? **Kristipati Ravi Ram**
Embryotoxicology Laboratory, Environmental Toxicology Group, Council of Scientific & Industrial Research (CSIR)- Indian Institute of Toxicology Research, Lucknow, Uttar Pradesh, India

The paper by Abry *et al*., is aimed at identification of accessory gland proteins (Acps) in five *Glossina* species through bioinformatic approaches. Given the importance of reproductive molecules in strategic designs of vector control, the reported data are of significant value. However, the study is quite preliminary, and hence it feels that authors have gone overboard with the interpretation of these preliminary results. Primarily, authors are required to be clear about the rationale of their study given that the testicular/MAG contributions towards spermatophore formation through transcriptomic and proteomic approaches in *G. morsitans* has already been reported earlier (Scolari *et al.* 2016, Scientific Reports). The inclusion of data on reciprocal hits would increase the reliability of predicted/putative Acps. Further, RT-PCR based analysis of the enrichment/expression of at least a couple of identified putative Acp orthologs in male accessory glands (MAGs) from any of the Glossina species would add excellent value to the data and support to the conclusions drawn. Alternatively, authors may compare their datasets with RNAseq/proteomic data of *G. morsitans* (Scolari *et al.*, 2016) to assess if the identified genes are enriched/expressed in MAGs or form the part of male contributions to the spermatophore.

Is the work clearly and accurately presented and does it cite the current literature?
Partly

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
I cannot comment. A qualified statistician is required.

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 20 Nov 2017

Benard Kulohoma,

We thank the reviewers for the comments. Our study is a preliminary *in silico* exploration of the MAG repertoire of *Glossina* species, and provides initial findings about genome annotations that identifies ACPs orthologs. Our aim was to establish whether there are male accessory gland (ACPs) orthologs across the Glossina species genomes, by comparing them to well characterised ACP orthologs from *Drosophila melanogaster* and *Anopheles gambiae*, given the importance of reproductive molecules in strategic designs of vector control.

- We have revised the manuscript to address the over-emphasis that ACPS have a play crucial role in the tsetse fly life reproductive cycle, and thus vector control. Lines 45, 73, 163, and 164.
- We have revised the manuscript to show that this is a preliminary study using the initial release of the Glossina genomes, which are publicly available, and future studies using RNASeq/transcriptome datasets of the male accessory glands and testis may identify rapidly evolving Glossina-specific ACPs, which perhaps do not bear a well-known protein domains, and could escape *in silico* investigation dependent on a comparative genomic approach alone. Lines 194 - 201.
- We have amended the manuscript to reflect that we identified possible signatures of selection, since a wider analysis including more closely related taxa was not performed. Lines 137, 140-143, and 165-166.

Competing Interests: None declared

Referee Report 04 September 2017

doi:10.21956/wellcomeopenres.13476.r25531

Emiliano Mancini
Department of Science, Roma Tre University, Rome, Italy

The paper by Abry *et al.* presents a comparative genomics analysis to identify male accessory gland proteins in *Glossina* species. The paper is an "*in silico*" exploration of the MAG repertoire of *Glossina*, and has the merit to open the road to further investigation on these important reproductive proteins and their functions. However I think that the authors should give the right relevance to their results, which are, in a sense, quite preliminary, but presented, in my opinion, with too much emphasis.

First of all the authors have to clarify the aim of the work, especially in the context of vector control. Although it is true that a deep understanding of ACPS in *Glossina* can be useful for limit fertility of *Glossina* species and thus control their spread (these proteins have important roles in female post-mating behaviour, such as single mating, storing and sustaining viability of sperms etc.), I think that ACPS do not necessarily have a realistic connection with the peculiar post-mating characteristic of the reproductive cycle of *Glossina* (e.g. absence of eggs deposition). Since this point is recurrently discussed and emphasized in the paper, I suggest the author either to better explain how the ACPS can have a role on the origin of this particular life cycle, or to eliminate this point which, otherwise, remains quite speculative.

It is important that authors stress the importance that their work should be completed in the future by RNAseq of male accessory glands and testis of *Glossina*. This because some rapidly evolving ACPS could escape *in silico* investigation, as they could be *Glossina*-specific and not corresponding to any ortholog, or simply because they do not bear a well-known protein domain. Species-specific proteins can not be obtained by a comparative genomic approach alone, so that RNA analysis is required to reveal (probably) the most important, peculiar ACPS in this important vector.

Finally, I would be very cautious in assigning positive selection to genes with dN/dS>1, since a wider analysis including more closely related taxa (and data on their intra-specific variability) should be performed. However, based on the dN/dS estimation here provided, I think that the authors should just argue about "possible signatures of positive selection" on these genes.

Is the work clearly and accurately presented and does it cite the current literature?
Partly

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 20 Nov 2017

Benard Kulohoma,

We thank the reviewers for the comments. Our study is a preliminary *in silico* exploration of the MAG repertoire of *Glossina* species, and provides initial findings about genome annotations that identifies ACPs orthologs. Our aim was to establish whether there are male accessory gland (ACPs) orthologs across the Glossina species genomes, by comparing them to well characterised ACP orthologs from *Drosophila melanogaster* and *Anopheles gambiae*, given the importance of reproductive molecules in strategic designs of vector control.

- We have revised the manuscript to address the over-emphasis that ACPS have a play crucial role in the tsetse fly life reproductive cycle, and thus vector control. Lines 45, 73, 163, and 164.
- We have revised the manuscript to show that this is a preliminary study using the initial release of the Glossina genomes, which are publicly available, and future studies using RNASeq/transcriptome datasets of the male accessory glands and testis may identify rapidly evolving Glossina-specific ACPs, which perhaps do not bear a well-known protein domains, and could escape *in silico* investigation dependent on a comparative genomic approach alone. Lines 194 - 201.
- We have amended the manuscript to reflect that we identified possible signatures of selection, since a wider analysis including more closely related taxa was not performed. Lines 137, 140-143, and 165-166.

Competing Interests: None declared