Insect Molecular Biology

Insect Molecular Biology (2017) **26**(1), 58–73 doi: 10.1111/imb.12271

Novel seminal fluid proteins in the seed beetle *Callosobruchus maculatus* identified by a proteomic and transcriptomic approach

H. Bayram, A. Sayadi, J. Goenaga*, E. Immonen and G. Arnqvist

Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden

Abstract

The seed beetle *Callosobruchus maculatus* is a significant agricultural pest and increasingly studied model of sexual conflict. Males possess genital spines that increase the transfer of seminal fluid proteins (SFPs) into the female body. As SFPs alter female behaviour and physiology, they are likely to modulate reproduction and sexual conflict in this species. Here, we identified SFPs using proteomics combined with a *de novo* transcriptome. A prior 2D-sodium dodecyl sulphate polyacrylamide gel electrophoresis analysis identified male accessory gland protein spots that were probably transferred to the female at mating. Proteomic analysis of these spots identified 98 proteins, a majority of which were also present within ejaculates collected from females. Standard annotation workflows revealed common functional groups for SFPs, including proteases and metabolic proteins. Transcriptomic analysis found 84 transcripts differentially expressed between the sexes. Notably, genes encoding 15 proteins were highly expressed in male abdomens and only negligibly expressed within females. Most of these sequences corresponded to 'unknown' proteins (nine of 15) and may represent rapidly evolving SFPs novel to seed beetles. Our combined analyses highlight 44 proteins for which there is strong evidence that they are SFPs. These results can inform further investigation, to better understand the molecular mechanisms of sexual conflict in seed beetles.

Keywords: evolution, reproduction, coleoptera, seminal fluid.

First published online 24 October 2016.

Correspondence: Helen Bayram, Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, Uppsala SE-752 36, Sweden. Tel.: + 46 18 471 2674; e-mail: helen.bayram@ebc.uu.se

*Present address: Aarhus Institute of Advanced Studies, Aarhus University, Aarhus, Denmark.

Introduction

Ejaculates are a complex mixture of sperm and seminal fluid (Perry *et al.*, 2013). Seminal fluid proteins (SFPs) are being increasingly studied owing to their effects on fertility and female physiology (Poiani, 2006). SFPs improve male fertility directly, maintaining sperm viability and motility (eg King *et al.*, 2011; Simmons & Beveridge, 2011; Smith & Stanfield, 2012). Furthermore, many SFPs influence female traits, such as egg production and remating rate (Mane *et al.*, 1983; Chapman, 2001) and are therefore likely targets of selection via sexual conflict (Arnqvist & Rowe, 2005; Sirot *et al.*, 2014). Sperm competition (Parker, 1970), the competition between sperm from different males to fertilize the same set of ova, is predicted to be a potent force driving both the evolution of SFPs (Dhole & Servedio, 2014) and the plasticity of their expression (Fedorka *et al.*, 2011; Ramm *et al.*, 2015). Indeed, there is now much evidence that SFPs are under positive Darwinian selection (eg Swanson *et al.*, 2001; Haerty *et al.*, 2007; Ramm *et al.*, 2008; for a review of reproductive protein evolution see Wilburn & Swanson, 2015).

The rapid evolution of SFPs can limit the techniques available for investigation in species without a described genome. The nature of many proteomic workflows – searching identified mass spectra against a database of predicted peptide sequences – requires a high level of genomic information. For many proteins, cross species matching may be sufficient for confident identifications (eg within mammalian sperm: Bayram *et al.*, 2016). However, the rapid evolution of SFPs results in unique or highly divergent sequences that may not be identified in this manner. Previous proteomic studies of SFPs have

thus often been limited to model species with extensive genomic information available [such as *Drosophila melanogaster* (Findlay *et al*., 2008), *Mus musculus* (Dean *et al*., 2011) and humans (Pilch & Mann, 2006)] or dedicated genome projects [eg *Apis mellifera* (Baer *et al*., 2009) and *Aedes aegypti* (Sirot *et al*., 2011)]. An alternative approach, using a *de novo* transcriptome to provide predicted protein sequences for protein identifications (Evans *et al*., 2012), allows a proteomic approach to be used to study the SFPs of species that lack a sequenced genome, but do have biologically interesting SFPs. Previously, this principle has been successfully applied using expressed sequence tag and 454 sequencing approaches to identify transcripts for proteomic identification of SFPs within crickets (Andrés *et al*., 2008; Marshall *et al*., 2009; Simmons *et al*., 2013), butterflies (Walters & Harrison, 2010) and abalone (Palmer *et al*., 2013). As technological advances continue to make the production of high-quality transcriptomic information faster and more affordable, proteomic methods for nonmodel organisms, relying on sequencing information from *de novo* transcriptomes, should become increasingly useful.

The seed beetle (*Callosobrucus maculatus*) is a widespread and economically very important agricultural pest and an emerging model organism in evolutionary biology. Males of this species have a suite of adaptations that improve their fertilization success during reproductive competition, such as sclerotized genital spines (Hotzy & Arnqvist, 2009; Hotzy *et al*., 2012). These spines increase the passage of seminal fluid substances into the body of mated females (Hotzy *et al*., 2012), suggesting an important role for the proteins within the seminal fluid of this species. Indeed, injecting females directly with SFPs alters female receptivity and male success in sperm competition (Yamane *et al*., 2008, 2015).

A recent study by Goenaga *et al*. (2015) used a 2D gel-based approach to identify 127 gel spots containing proteins that are both present in the male accessory glands and are transferred to the female at mating. Furthermore, comparisons of the accessory gland protein profile of 15 distinct populations revealed striking intraspecific variation and functional assays showed that relative protein abundance within several of these spots correlated with postmating reproductive phenotypes (such as male success in sperm competition and male ability to stimulate female egg laying). Here, we built upon this previous study; combining proteomic and transcriptomic analyses allowed us to identify SFPs from 55 of these transferred spots. We found 98 sequences in total, 85 of which have significant homology to known, annotated proteins. Comparisons between the sequences identified here and SFPs already characterized within insects show 21 of the sequences to be homologous. Analysis of transcript expression data revealed that 15 of these sequences are highly expressed in males and only negligibly present within females. Over half of these 15 sequences have no homology to annotated proteins. These are likely to be novel SFPs, possibly unique to seed beetles, worthy of further investigation.

Results and discussion

Protein identification

A previous analysis used 2D sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) of male accessory gland proteins, which were compared with gel spot profiles of mated and virgin females, to identify 127 spots that were transferred to the female at mating (Goenaga *et al*., 2015). Here, we chose 55 of those spots for protein identification. Out of these 55 spots, 34 were selected because they showed a strong interpopulation-level correlation with traits relating to reproductive success (Goenaga *et al*., 2015). The remaining 21 represented clear spots that were consistently found within all 15 populations tested (Fig. 1). Commonly, protein identifications are based upon predicted protein sequences taken from genomic information. Here, a recent and integrative *de novo* transcriptome for *C. maculatus* (Sayadi *et al*., 2016) was used to allow protein identification (Evans *et al*., 2012). Searching against the *de novo* transcriptome following matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) analyses gave significant protein hits, based on at least two unique peptides, for 23 of the 55 gel spot extractions. The results of the MALDI-TOF analysis for the remaining 32 gel spots were deemed inconclusive, as they either contained multiple proteins, which is common for 2D SDS-PAGE analysis (Campostrini *et al*., 2005), or contained a low protein concentration. These 32 spots were then analysed with higher resolution using liquid chromatography-tandem mass spectrometry (LC-MSMS), a more sensitive method that also allows the identification of multiple proteins from the same spot. After LC-MSMS analysis, at least one protein sequence was identified with significant confidence, based on a minimum of two unique peptide hits, in extractions from 30 out of the 32 gel spots. The two remaining gel spots gave no significant protein sequence hits.

As protein identifications were based on an in-house transcriptome (Sayadi *et al*., 2016) the open reading frame of each transcript hit was predicted using TransDecoder (http://transdecoder.github.io/; Haas *et al*., 2013) and checked manually before confirming the identified protein sequence. Following this confirmation, a total of 98 protein sequences was identified with confidence from the 55 gel spots analysed (Table 1). Identifying more proteins than spots was expected (Campostrini *et al*., 2005). Furthermore, several of the spots were

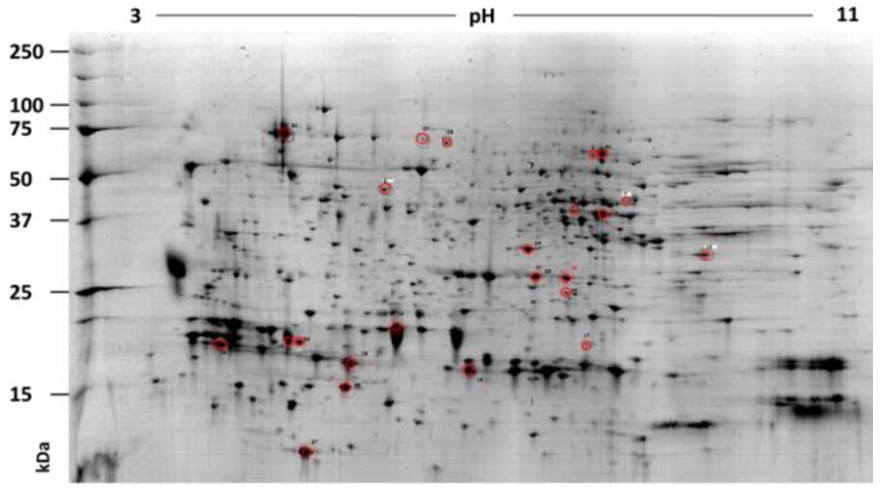

Figure 1. Representative 2D sodium dodecyl sulphate polyacrylamide gel electrophoresis gel of the male *Callosobruchus maculatus* accessory gland proteome. Representative spots collected for tryptic digestion and mass spectrometry analysis are circled. Twenty-three spots were collected from this particular gel. The remaining spots were collected from other, equivalent gels. The molecular mass is labelled on the left-hand side of the gel, and pH along the top. [Colour figure can be viewed at wileyonlinelibrary.com]

specifically chosen for analysis owing to their size and therefore the likelihood of obtaining good protein identifications. Of the 53 gel spots that yielded protein hits, 29 contained only one significantly identified protein sequence. There were four spots containing two proteins, and seven spots containing three. The remaining 13 gel spots contained more than three protein sequences (range four to 11) that were matched with significant confidence. Of the final 98 protein sequences, 27 were identified within multiple spots; these were either from spots on the same region of multiple gels, or adjoining spots on the same gel. The other 71 proteins were identified within single gel spots. We note that our protein identification from the 55 spots is probably conservative, considering that there is substantial intraspecific variation in SFPs (Goenaga *et al.*, 2015), which our in-house transcriptome may not fully capture.

Annotation and gene ontology

The 98 identified protein sequences were searched against the National Center for Biotechnology Information (NCBI) database of nonredundant protein sequences using Blast2Go, with e-values $< 1 \times 10^{-5}$. Homologous protein matches were found for 85 sequences, with a mean coverage of 96% (range: 65–100%) and a mean of 74% identities (range: 32–100%) (Table 1). Functional annotation of these 85 homologous proteins revealed many proteins that have metabolic biological processes (Fig. 2). This apparent overrepresentation may stem from the fact that metabolic proteins are relatively conserved and therefore more likely to provide significant sequence matches. Metabolic proteins secreted as SFPs have also been documented within other species (Collins *et al.*, 2004; Findlay *et al.*, 2008; Kelleher *et al.*, 2009).These proteins may, for example, assist with sperm longevity in storage. Some of these metabolic proteins may derive from the cells of the accessory glands and may not be true SFPs. The gel spots were selected based on their presence in male accessory glands and mated, but not virgin, females. Cellular metabolic proteins from the male accessory glands that are not present in virgin female samples, but do share a gel spot with a transferred SFP, or are produced by the mated female, will be seen within mated female samples and so considered within a transferred protein spot. We therefore have combined many analyses here to add confidence to our assumptions of which proteins are indeed SFPs (Table 1).

Additionally, many of the protein sequences were homologous to proteases and protease inhibitors, which are important components of seminal fluid in other species (Mueller *et al.*, 2004; LaFlamme & Wolfner, 2013). Male-derived protease cascades, which occur within the female reproductive tract in the presence of SFPs (Park & Wolfner, 1995), lead to active proteins able to stimulate ovulation and other postmating changes in the female (Heifetz *et al.*, 2005; LaFlamme *et al.*, 2012). Protease inhibitors may in part act to block catalytic enzymes produced by females, and are thus candidates for interaction with female-derived proteins and for sexual conflict (Lung *et al.*, 2002; Mueller *et al.*, 2008).

Similarity with known seminal fluid proteins

The protein sequences identified here in the seed beetle were compared against sequences previously identified by proteomic analysis of honeybee (*Apis mellifera*; Baer *et al.*, 2009) and *Drosophila melanogaster* (Findlay & Swanson, 2010) ejaculates. Additionally, predicted SFPs

Table 1. A summary of the 98 proteins identified here from the 55 gel spots

Spot #(s)	Transcript ID	In whole ejaculate	Predicted secreted†	Differentially expressed	MW (kDa)	Top homologous protein match	Species	Coverage (%)	Identity (%)
1	TR72098|c0_g2_i1				79.83	NADH-ubiquinone oxidoreductase 75 kDa subunit, mitochondrial	*Tribolium castaneum*	99	81
2	TR62718|c2_g1_i2		X	M	62.15	2-hydroxyacyl-CoA lyase 1	*Tribolium castaneum*	100	70
3	**TR3574|c4_g1_i3**	X	X	M*	60.96	**Serine protease inhibitor**	*Tribolium castaneum*	99	53
3	TR25349|c3_g4_i2	X			52.8	Protein hu-li tai shao	*Tribolium castaneum*	94	81
3	TR55622|c0_g1_i1		X		49.23	26S protease regulatory subunit 4	*Tribolium castaneum*	100	99
3	**TR4903|c1_g1_i1**	X	X		53.98	**Dihydrolipoyl dehydrogenase**	*Tribolium castaneum*	100	84
3	**TR8720|c0_g1_i1**	X	X	M	55.74	**Protein disulphide-isomerase A3**	*Tribolium castaneum*	100	77
3	**TR47303|c0_g3_i1**	X	X	M*	52.29	Carboxypeptidase Q	*Tribolium castaneum*	99	60
4	TR36188|c1_g1_i3		X		54.35	Succinate-semialdehyde dehydrogenase	*Tribolium castaneum*	96	75
4	TR44676|c0_g1_i1			M	50.77	Hydroxymethylglutaryl-CoA synthase 1	*Tribolium castaneum*	100	72
4	TR7946|c0_g2_i1		X		48.44	26S protease regulatory subunit 7	*Tribolium castaneum*	100	98
5	TR36757|c0_g1_i2			F	49.12	Elongation factor 1-gamma	*Tribolium castaneum*	100	77
5	TR5618|c0_g1_i1			F	44.96	DnaJ homologue subfamily A member 1	*Tribolium castaneum*	100	83
5	TR29747|c2_g6_i1			F	51.98	Elongation factor Tu, mitochondrial	*Tribolium castaneum*	99	84
6	**TR48652|c0_g2_i1**	X	X	M	52.89	Mitochondrial-processing peptidase subunit beta	*Tribolium castaneum*	100	87
8	**TR72129|c1_g1_i5**	X	X		47.36	Probable medium-chain specific acyl-CoA dehydrogenase, mitochondrial	*Tribolium castaneum*	96	68
9	TR27744|c0_g3_i1		X		36.73	N-acetyl-D-glucosamine kinase	*Tribolium castaneum*	95	55
9	**TR7050|c1_g1_i3**			M	36.08	**Regucalcin**	*Tribolium castaneum*	96	41
10	**TR3254|c0_g1_i2**	X	X	M	36.26	Aldo-keto reductase	*Dendroctonus ponderosae*	96	69
10	TR31205|c0_g1_i2	X			30.6	Hydroxyacyl-coenzyme A dehydrogenase, mitochondrial	*Tribolium castaneum*	100	74
10	**TR30158|c0_g2_i1**		X	M	36.28	Aldose reductase	*Tribolium castaneum*	100	79
11	TR2340|c0_g15_i1	X		MF	26.29	Lipid storage droplets surface-binding protein 1	*Tribolium castaneum*	87	55
12	TR44608|c1_g1_i1	X		M	23.44	Glutathione S-transferase	*Tenebrio molitor*	100	66
13	TR18113|c0_g2_i3			M	23.1	Protein DJ-1	*Tribolium castaneum*	99	67
13	**TR54559|c7_g2_i1**	X	X	M	9.17	Unknown			
14	TR72389|c4_g20_i2			F	19.32	Ragulator complex protein LAMTOR1	*Tribolium castaneum*	98	66
14	**TR3632|c1_g1_i5**	X	X		22.04	Peroxiredoxin 1	*Tribolium castaneum*	99	81
14	TR69877|c0_g2_i1				16.87	Eukaryotic translation initiation factor 1A, X-chromosomal	*Tribolium castaneum*	100	99
18	TR10785|c0_g1_i1		X		16.78	Actin-related protein 2/3 complex subunit 5	*Tribolium castaneum*	99	72
18	TR70813|c2_g1_i3	X		M	17.46	Troponin C, isoform 2	*Tribolium castaneum*	100	99
18	**TR33498|c0_g3_i1**	X	X	M*	19.18	Unknown			
19	**TR40631|c0_g1_i1**	X	X	M*	19.9	Unknown			
20	**TR33531|c0_g2_i1**	X	X	M*	11.78	Unknown			
23	TR74364|c0_g1_i2		X	F	25.83	MOB kinase activator-like 4	*Tribolium castaneum*	100	96
23	TR50022|c0_g2_i1			F+	28.54	Hydroxyacylglutathione hydrolase, mitochondrial	*Tribolium castaneum*	100	75
24	**TR28266|c2_g1_i2**	X	X		32.82	Unknown			
25	TR56935|c0_g1_i1			M	35.57	Ketimine reductase mu-crystallin	*Tribolium castaneum*	97	53
25	TR24091|c1_g1_i3	X		F+	39.55	Probable isocitrate dehydrogenase [NAD] subunit alpha, mitochondrial	*Tribolium castaneum*	99	90
26	TR33486|c0_g1_i1		X		13.89	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 5	*Tribolium castaneum*	100	77
28	TR47656|c0_g1_i2			F	57.29	Unknown			

Table 1 *Continued*

Spot #(s)	Transcript ID	In whole ejaculate	Predicted secreted†	Differentially expressed	MW (kDa)	Top homologous protein match	Species	Coverage (%)	Identity (%)
29	TR7191|c0_g1_i1				43.77	Short-chain specific acyl-CoA dehydrogenase, mitochondrial	*Tribolium castaneum*	98	84
29	TR36193|c3_g2_i5			M	45.36	NAD-dependent protein deacetylase sirtuin-2	*Tribolium castaneum*	91	67
29	TR37093|c0_g1_i1	X		M	50.47	Elongation factor 1-alpha	*Tribolium castaneum*	100	97
29	TR13504|c0_g1_i2		X	M	43.43	Cystathionine gamma-lyase	*Culex quinquefasciatus*	99	59
29	TR7212|c0_g1_i2				43.65	Pyruvate dehydrogenase E1 component subunit alpha, mitochondrial	*Tribolium castaneum*	100	74
29	**TR72449|c5_g3_i1**	X	X	M	44.49	**Serine protease inhibitor 3/4**	*Tribolium castaneum*	98	53
29	TR56146|c0_g1_i1		X		43.79	Short-chain specific acyl-CoA dehydrogenase, mitochondrial	*Tribolium castaneum*	99	72
31	TR21207|c0_g5_i5	X		M	67.98	NADP-dependent malic enzyme	*Tribolium castaneum*	99	70
40	TR38464|c0_g1_i1		X	M	66.08	Electron transfer flavoprotein-ubiquinone oxireductase, mitochondrial	*Tribolium castaneum*	100	83
41	**TR6343|c4_g5_i3**	X	X	M	49.51	**Chitinase-like protein Idgf4**	*Diaprepes abbreviatus*	100	71
44	**TR67329|c0_g1_i2**	X	X	F	13.85	**Profilin**	*Tribolium castaneum*	100	97
45	**TR10050|c2_g1_i1**	X	X	M	22.49	**Peptidyl-prolyl cis-trans isomerase B**	*Tribolium castaneum*	98	81
45	TR71966|c0_g3_i1			M	19.78	Adenine phosphoribosyltransferase	*Tribolium castaneum*	99	60
46	TR52534|c0_g1_i1				27.91	Proteasome subunit alpha type-7-1	*Tribolium castaneum*	93	96
46	TR52430|c8_g7_i1				27.26	Vesicle-associated membrane protein-associated protein B	*Tribolium castaneum*	96	57
46	**TR71852|c5_g32_i7**	X		M	36.93	**Voltage-dependent anion-selective channel-like**	*Tribolium castaneum*	100	72
46	TR44526|c0_g2_i8	X			28.27	14-3-3 protein zeta	*Tribolium castaneum*	100	96
46	TR6354|c0_g1_i1		X	F	32.05	Enoyl-CoA hydratase domain-containing protein 3, mitochondrial	*Tribolium castaneum*	91	62
46	TR60977|c0_g1_i1		X	F	27.47	Adenylate kinase	*Tribolium castaneum*	96	75
46	TR6860|c0_g1_i1		X	M	37.03	Probable uridine-cytidine kinase	*Tribolium castaneum*	86	75
49	TR17836|c0_g1_i1	X		M	34.85	Succinyl-CoA ligase subunit alpha, mitochondrial	*Tribolium castaneum*	100	89
49	**TR70797|c0_g2_i1**	X	X	M	51.11	Trifunctional enzyme subunit beta, mitochondrial	*Tribolium castaneum*	100	82
50	TR27857|c3_g5_i4		X	F	64.63	Hrp65 protein	*Tribolium castaneum*	75	64
51	TR841|c1_g1_i2				35	PIH1 domain-containing protein 1	*Tribolium castaneum*	98	58
51	TR75369|c0_g2_i1			F+	27.36	Bifunctional methylenetetrahydrofolate dehydrogenase/cyclohydrolase, mitochondrial isoform X3	*Tribolium castaneum*	96	78
51	TR8479|c0_g1_i2	X		M	35.83	Glyceraldehyde-3-phosphate dehydrogenase	*Colaphellus bowringi*	100	83
55	TR63577|c0_g1_i1	X		M	12.55	FABP-like protein	*Tribolium castaneum*	100	80
56	**TR73858|c0_g1_i1**	X	X	M*	18.87	Unknown			
56	**TR6332|c0_g1_i1**	X	X	F	17.55	**Epididymal secretory protein E1**	*Tribolium castaneum*	100	52
56	**TR10050|c2_g3_i1**	X		F	22.29	**Peptidyl-prolyl cis-trans isomerase**	*Tribolium castaneum*	100	82
56	**TR57145|c0_g1_i2**	X	X	F	18.74	**Nucleoside diphosphate kinase**	*Tribolium castaneum*	100	86
10, 14	**TR23829|c5_g10_i3**	X		MF	41.82	**Actin**	*Ornithodoros moubata*	100	99
10, 30, 38	**TR50156|c0_g4_i2**	X	X		72.88	**Heat shock protein 70b**	*Leptinotarsa decemlineata*	98	95
10, 38	**TR55001|c2_g2_i6**	X		F	71.45	**Heat shock protein 70a**	*Leptinotarsa decemlineata*	93	96
13, 14, 15, 16, 34	**TR32496|c0_g1_i1**	X	X	M*	21.9	Protein spaetzle	*Harpegnathos saltator*	73	37
14, 15	**TR54559|c7_g2_i3**	X	X	M*	18.72	Unknown			

Table 1 *Continued*

Spot #(s)	Transcript ID	In whole ejaculate	Predicted secreted†	Differentially expressed	MW (kDa)	Top homologous protein match	Species	Coverage (%)	Identity (%)
14, 15, 16	**TR6668|c0_g1_i1**	X	X	M*	18.06	**Kunitz-like protease inhibitor precursor**	*Ancylostoma caninum*	88	32
14, 18	**TR26246|c8_g1_i1**	X	X	M*	17.83	Unknown			
14, 34, 43	TR66839|c1_g1_i3	X		F	17.48	Eukaryotic translation initiation factor 5A	*Tribolium castaneum*	100	95
15, 20, 21, 34, 38, 44	**TR23829|c5_g10_i4**	X		M	41.88	**Actin-5C**	*Drosophila melanogaster*	100	100
14, 16	**TR54466|c2_g3_i4**		X	M*	19.07	Unknown			
17, 18	**TR54466|c2_g3_i1**	X	X	M*	19.22	Unknown			
18, 23	**TR32785|c0_g1_i1**	X	X	M*	27.52	**Trypsin-like serine protease inhibitor**	*Ceratosolen solmi marchali*	88	32
18, 35	**TR9596|c0_g1_i1**	X	X	M*	15.48	Unknown			
21, 22, 37	**TR57234|c6_g2_i3**	X	X		11.44	**Digestive cysteine protease intestain**	*Leptinotarsa decemlineata*	93	44
23, 33, 36	**TR57334|c3_g6_i4**	X	X	M	30.44	Unknown			
25, 29	**TR31224|c1_g1_i6**	X	X	M*	33.24	**Serine protease easter-like**	*Tribolium castaneum*	92	39
25, 29, 32, 51	TR23859|c2_g1_i1	X		M	39.55	Fructose-bisphosphate aldolase isoform X1	*Tribolium castaneum*	100	86
25, 29, 42	TR71505|c0_g3_i1		X	M	40.87	Acetyl-CoA acetyltransferase, cytosolic	*Tribolium castaneum*	99	63
27, 28	TR69749|c2_g2_i1	X		F	67.98	Lamin Dm0	*Tribolium castaneum*	100	74
45, 47	TR59712|c1_g2_i1			M	24.56	Glutathione S-transferase	*Tenebrio molitor*	94	61
46, 52	**TR55877|c0_g1_i7**	X	X	M	31.9	Enoyl-CoA hydratase, mitochondrial	*Tribolium castaneum*	100	74
48, 49	**TR33575|c0_g1_i1**	X	X	M	26.29	V-type proton ATPase subunit E	*Leptinotarsa decemlineata*	100	88
49, 51	TR28214|c0_g1_i3	X		F	33.03	RNA-binding protein squid	*Tribolium castaneum*	65	86
7, 51	TR49043|c0_g1_i1	X		M	46.46	Cytoplasmic NADP+ dependent isocitrate dehydrogenase	*Pogonistes gracilis*	98	87
9, 25, 29	**TR17057|c0_g1_i1**	X	X		33.91	**Serine protease easter-like**	*Cerapachys biroi*	97	39
9, 25, 51	TR18165|c0_g2_i5	X		MF	37.11	Probable transaldolase	*Tribolium castaneum*	89	80
9, 51	**TR12255|c7_g2_i7**	X	X		40.63	Glycerol-3-phosphate dehydrogenase [NAD(+)], cytoplasmic isoform X1	*Tribolium castaneum*	94	90

Transcript ID corresponds with the *de novo* transcriptome used for identification (Sayadi *et al.*, 2016).
The molecular weight was predicted using the Compute pI/Mw online tool (Gasteiger *et al.*, 2005).
The top BLAST protein matches are also reported here. Many proteins were also found within an analysis of four single ejaculates, collected from the female bursa copulatrix and analysed using liquid chromatography-tandem mass spectrometry (LC-MSMS).
Differential expression analysis was performed on all isoforms of transcripts that matched to the identified protein sequences.
Significant differences are indicated by letters: M, more highly expressed in males; F, more highly expressed in females. An * indicates the 15 sequences most highly expressed in males and only negligibly present in females. For proteins that had an isoform expressed in a significantly greater amount in females, and also an isoform with no significant difference in expression, there is a + symbol. Where proteins had isoforms with significant differences in both males and females, both letters are present.
†SignalP and secretomeP were used to predict whether protein sequences had signals of secretion (Bendtsen *et al.*, 2004; Petersen *et al.*, 2011).
Transcript ID in bold where they are confidently assumed to be seminal fluid proteins (SFPs). Homologous protein in bold if this is a known SFP in other species.
MW, molecular weight; NAD, nicotinamide adenine dinucleotide; NADH, reduced form of NAD; NADP, NAD phosphate; CoA, Coenzyme A; Tu, thermo unstable; LAMTOR1, late endosomal/lysosomal adaptor, MAPK and MTOR activator 1; FABP, fatty acid binding protein; Idgf4, imaginal disc growth factor 4; PIH1, protein interacting with Hsp90 1.

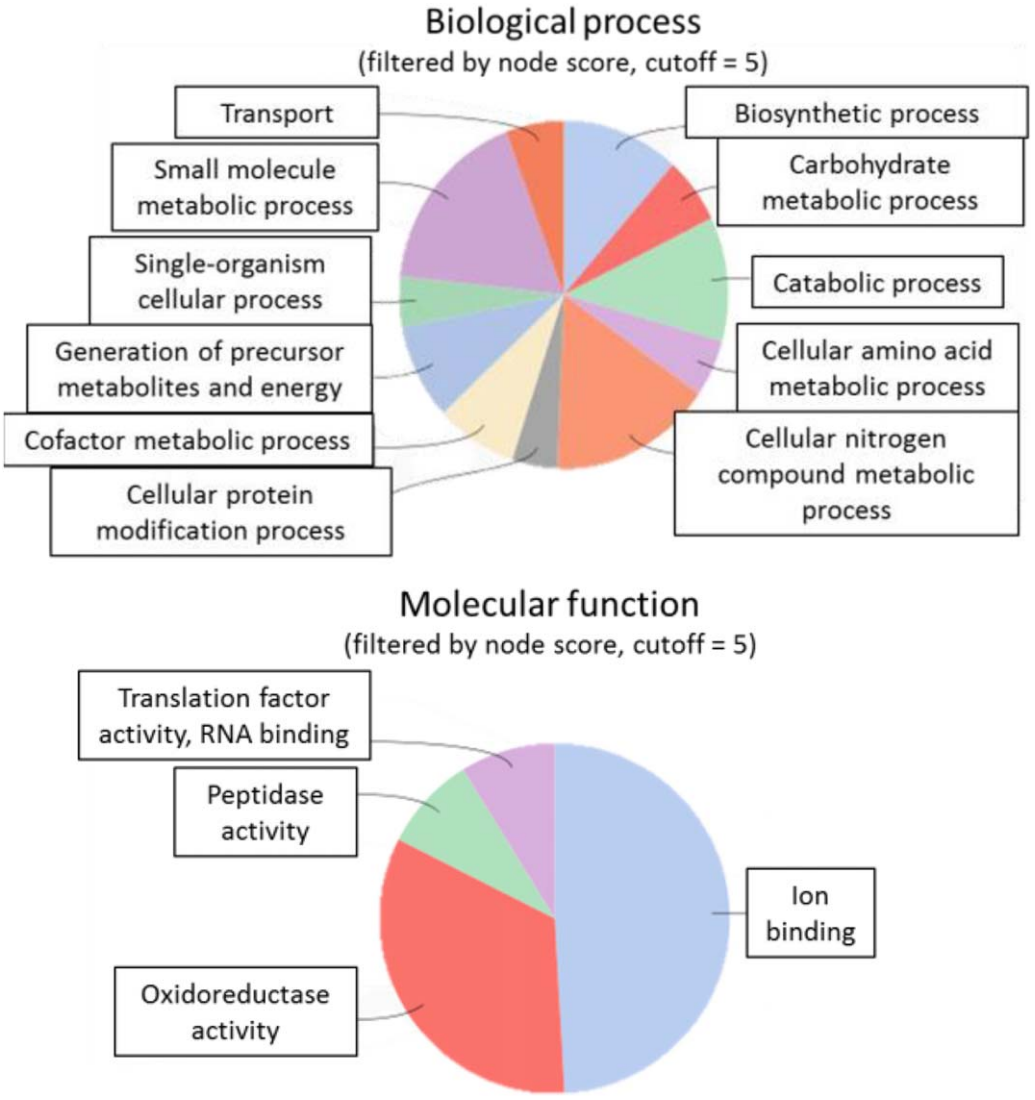

Figure 2. Functional annotation of the proteins homologous to protein sequences identified in the *Callosobruchus maculatus* accessory glands. Blast searching of the protein sequences identified from the *C. maculatus* accessory glands against all known protein sequences found 85 homologous proteins. The gene ontology information for biological processes and molecular functions of these 85 homologous proteins was collected within Blast2Go. [Colour figure can be viewed at wileyonlinelibrary.com]

of the red flour beetle (*Tribolium castaneum*), identified using proteomic (South *et al.*, 2011; Xu *et al.*, 2013) and microarray (Parthasarathy *et al.*, 2009) techniques, were collected for comparison. In total, 22 out of our 98 sequences matched to homologues previously identified as SFPs, with e-values $< 1 \times 10^{-6}$ and identities $\geq 30\%$ (Table 2). Almost all of these sequences (21 out of 22) were also found within the whole ejaculate analysis described below. Comparisons against red flour beetle seminal fluid protein sequences yielded 13 matches. There were also 12 homologues found within the *D. melanogaster* seminal fluid proteome and nine within the honeybee seminal fluid proteome. Of the 22 amino acid sequences with homologues in other seminal fluid proteomes, 11 were found in only one species, 10 within two species, and one – a serine protease – within all three species tested here.

Surprisingly few of the 98 proteins identified here were homologous to known SFPs, most likely in part because of the rapid evolution of SFPs (eg Swanson *et al.*, 2001) leading to interspecies divergence. Many of the proteins homologous to known SFPs were proteases or protease inhibitors. These are common components of the seminal fluid of many species (LaFlamme & Wolfner, 2013; Avila *et al.*, 2015). Protease matches tended to give lower sequence matching scores than the matches to the rest of the homologous proteins [identities 29%

Table 2. Proteins homologous to those previously identified in the seminal fluid of *Drosophila melanogaster*, *Tribolium castaneum* and/or *Apis mellifera* (Findlay *et al*., 2008; Baer *et al*., 2009; Parthasarathy *et al*., 2009; South *et al*., 2011; Xu *et al*., 2013)

		Homologous species match		
Transcript ID	Homologous protein name	*Tribolium castaneum*	*Apis mellifera*	*Drosophila melanogaster*
TR55001lc2_g2_i6	Heat shock protein 70a	**xr**	**xr**	
TR50156lc0_g4_i2	Heat shock protein 70b	**xr**	**x**	
TR6668lc0_g1_i1	Kunitz-like protease inhibitor precursor	**x**		
TR57145lc0_g1_i2	Nucleoside diphosphate kinase	**xr**		
TR10050lc2_g3_i1	Peptidyl-prolyl cis-trans isomerase	**x**		**x**
TR10050lc2_g1_i1	Peptidyl-prolyl cis-trans isomerase B	**xr**		**xr**
TR8720lc0_g1_i1	Protein disulphide-isomerase A3	**xr**		**x**
TR7050lc1_g1_i3	Regucalcin	**x**		
TR17057lc0_g1_i1	Serine protease easter-like	**x**	**x**	**x**
TR31224lc1_g1_i6	Serine protease easter-like	**x**		
TR3574lc4_g1_i3	Serine protease inhibitor	**x**		**x**
TR72449lc5_g3_i1	Serine protease inhibitor 3/4	**xr**		**x**
TR32785lc0_g1_i1	Trypsin-like serine protease inhibitor	**x**		**x**
TR23829lc5_g10_i3	Actin		**xr**	
TR23829lc5_g10_i4	Actin-5C		**xr**	
TR6343lc4_g5_i3	Chitinase-like protein Idgf4		**xr**	
TR57234lc6_g2_i3	Digestive cysteine protease intestain			**x**
TR4903lc1_g1_i1	Dihydrolipoyl dehydrogenase			**xr**
TR6332lc0_g1_i1	Epididymal secretory protein E1		**xr**	**xr**
TR3632lc1_g1_i5	Peroxiredoxin 1		**xr**	**xr**
TR67329lc0_g1_i2	Profilin		**xr**	
TR71852lc5_g32_i7	Voltage-dependent anion-selective channel-like			**xr**

Homologous matches were based on a tBlastn comparison of the 98 *Callosobruchus maculatus* protein sequences with genes collected using available data within the above references.
Proteins were considered homologous if they had e-values $< 1 \times 10^{-6}$ and identities $\geq 30\%$, marked with an 'x'. Additionally, a reciprocal BLAST was performed. Sequences for which the top hit for the known seminal fluid protein (SFP) from other species was the transcript for the corresponding putative SFP, indicating likely orthology, are marked with an 'r'. Idgf4, imaginal disc growth factor 4.

(23–35%) protease or inhibitor vs. 68% (32–99%) all other proteins]. Proteases present in both mammals and *D. melanogaster* contain conserved structures, despite sequence differences (Mueller *et al*., 2004). It is likely that within this protein class the functional motif is highly conserved, whereas the remaining sequence is subject to strong selection.

Determining whether homologues are likely to be orthologues, deriving from a common ancestor and diverging after a speciation event (Fitch, 2000), is important when trying to infer similar functions. Commonly, genes are considered orthologues across genomes if the best hit of their protein products corresponds to the best hit in the reciprocal analysis (Tatusov *et al*., 1997; Bork *et al*., 1998). A BLAST reciprocal best hit analysis was performed here to determine whether these matches are likely to be orthologues. Using the tblastn function in the desktop version of BLAST, the proteins that are known SFPs described in *A. mellifera*, *D. melanogaster* and *T. castaneum* were compared with the *C. maculatus* transcriptome. Following this analysis, six proteins from *T. castaneum*, seven proteins from *A. mellifera* and five proteins from *D. melanogaster* did indeed have the corresponding putative *C. maculatus* SFP transcript as the top match (Table 2).

Comparisons with whole ejaculate analysis

Spots from 2D SDS-PAGE gels of male accessory gland homogenates were deemed to contain candidate SFPs based upon a comparison of protein spots in the reproductive tracts of just-mated and virgin females; with those present in male accessory gland homogenates as well as in mated females, but not virgin females, being deemed candidate SFPs. To confirm that these proteins are indeed transferred to the female at mating, four individual male ejaculates were collected within the female reproductive tract (the bursa copulatrix) immediately after mating. These were analysed individually using LC-MSMS of the trypsin-digested peptides and over 400 proteins were identified in these samples. Of the 98 protein sequences identified from the gel spot analysis of male accessory glands, 60 were also identified within the entire ejaculates and reproductive bursa of immediately mated females (Table 1). This suggests that most of the 98 proteins are indeed ejaculated. The 38 proteins that were not identified may either be present in quantities too low to be detected using a whole ejaculate approach, or may be present within the accessory gland although not ejaculated. In addition, there is much intraspecific variation in the seminal fluid proteome in this

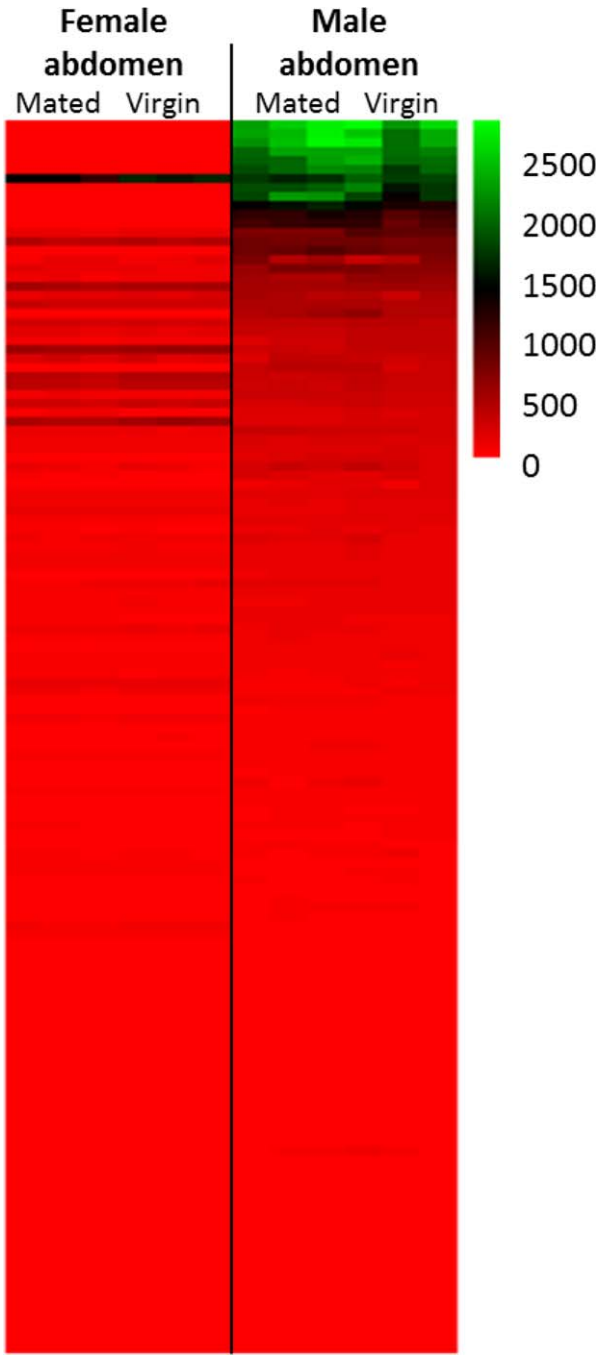

Figure 3. Heatmap of fragments per kilobase of transcript per million mapped reads values for the transcripts that matched to the 98 identified proteins. [Colour figure can be viewed at wileyonlinelibrary.com]

species (Goenaga *et al*., 2015). The whole ejaculate samples were taken from a single population out of those analysed with the 2D SDS-PAGE gel approach and some of the proteins present in the initial analysis may not be expressed within these individuals.

Over 400 proteins were identified by LC-MSMS analysis of single ejaculates within the female bursa. Many of these proteins are likely to derive from the female bursa copulatrix, and further analyses of these proteins are thus not included here. A stable isotope labelling approach would enable male and female proteins to be distinguished and all ejaculate proteins to be identified (Findlay *et al*., 2008; Dean *et al*., 2011; Sirot *et al*., 2011; Boes *et al*., 2014). As yet, this approach has not been possible in seed beetles.

To gain further confidence that we have identified transferred SFPs, we assessed signals of secretion within the identified protein sequences, using SignalP and SecretomeP (Bendtsen *et al*., 2004; Petersen *et al*., 2011). Of the 98 protein sequences, 54 showed signals of secretion. Of these, 37 were also identified within the whole ejaculate analysis.

Transcriptome expression data

Transcript expression data, for each of the 98 protein sequences identified here from the male accessory glands, were analysed from male and female abdominal samples ($N = 12$). Both virgin and mated samples were available, giving four biological sample categories (mated male abdomen, virgin male abdomen, mated female abdomen and virgin female abdomen). For each group there were three biological replicates, allowing statistical analysis using a negative binomial generalized linear model (GLM). The abdomens of both sexes of *C. maculatus* are to a very large extent composed of reproductive tissue. Genes that are highly expressed in males, but not expressed in females, are almost certainly male-specific reproductive proteins.

The 98 protein sequences were matched back to the transcriptome using BLAST, resulting in 137 transcript sequences with identical matching. The disparity between the number of protein and transcriptome sequences here is because of the presence of isoforms that are indistinct at a translated protein level. Comparing virgin male and female abdominal samples, 83 of the 137 transcripts analysed were significantly differentially expressed after a correction for multiple testing [using a false discovery rate (FDR) of 5%; Table S1]. Out of the 83 significantly differently expressed transcripts, 56 were expressed more highly in males and 27 in females. The results for mated individuals were very similar. In total, 84 transcripts were differentially expressed; 59 having higher expression in males, and 25 in females. There were five transcripts differentially expressed between male and female virgin abdomen samples that were not significantly differentially expressed in the mated samples, whereas six transcripts were differentially expressed in mated, but not virgin, abdomen samples. These 11 transcripts had lower expression differences than most of those tested here. As an example, using a cut-off q-value < 0.01 would exclude

23 transcripts, of which nine were only differentially expressed for either the virgin or mated samples.

The majority of transcripts were expressed both in female and male abdomens. This has previously been described when analysing both the proteome and transcriptome to identify SFPs (Boes *et al*., 2014), highlighting the importance of proteomic study as opposed to relying only upon differential transcript expression. Indeed, transcripts for 17 out of 98 protein sequences were more highly expressed within females than males. It is certainly possible these are proteins present in the cells of male accessory sex glands, and not SFPs. Indeed, most do not have known reproduction-specific functions but are cellular proteins, important for mitochondrial function, cell cycle progression or protein synthesis. However, for five of these 17 sequences, there is further evidence that they are indeed SFPs; including being known SFPs in other species. In these instances, these proteins are most likely expressed within both the male and female abdomens, and present within the ejaculate. For example, epididymal secretory protein is a common component of the ejaculate across many taxa (matching with *A. mellifera* and *D. melanogaster* SFPs here), and is also a cholesterol transporting protein that has more general activity elsewhere (Liou *et al*., 2006).

Of the 58 transcripts that had significantly higher expression in males, 15 were particularly highly expressed [fragments per kilobase of transcript per million mapped reads (FPKM) values over 200] and either not expressed at all in females, or expressed at negligible levels (Fig. 3). Considering the results of the homology searching for these sequences, nine of the 15 did not give significant BLAST hits. Within the 98 protein sequences identified, a total of 13 has no homologous proteins. That so many 'unknown' protein sequences are the most highly expressed male transcripts suggests that they are novel *C. maculatus* SFPs that are highly diverged from any previously identified SFPs in other species investigated. Indeed, protein sequence homology matching revealed that the most highly expressed transcript in both virgin and mated male abdomens matched, although with low confidence, to an 'ovulation-inducing factor' protein template (data not shown). As SFPs are commonly shown to evolve rapidly (eg Civetta & Singh, 1995; Swanson *et al*., 2001), owing to competition and conflict within and between the sexes, this suggests that SFPs in seed beetles may also undergo rapid evolution.

The remaining transcripts that were highly expressed in males but not at all expressed in females were homologous to five proteases and one protein, spaetzle, which works within the Toll immune pathway (Hoffmann *et al*., 1999). As discussed above, a variety of proteases are significant and typical components of the seminal fluid (LaFlamme & Wolfner, 2013). Proteins involved in the immune response are important within the ejaculate (eg Poiani, 2006; Dorus *et al*., 2012), both to protect sperm and to modulate the female immune response. The protein spaetzle, an important component of innate immunity pathways, has previously been identified as a protein transferred to females at mating in mosquitoes (Boes *et al*., 2014). Additionally, a small peptide that is induced within the spaetzle pathway is highly expressed within the oviducts of *D. melanogaster* in response to mating (Kapelnikov *et al*., 2008), suggesting a role for this pathway in female responses to mating.

All 98 proteins analysed here are male derived, as the samples were taken from whole male accessory sex gland homogenates. However, it is possible that, because many of the spots contain multiple proteins, not all are true SFPs. Some may instead be cellular proteins from the accessory sex glands. Combining the transcript expression information with the results from the whole ejaculate analysis, the secretome predictions and the homologous SFP results, allows confident designation of a number of the 98 protein sequences as SFPs. Under the requirements that a sequence has to be either present in the whole ejaculate and predicted to be secreted, or homologous to a known SFP, or highly expressed in males but negligibly in females, there are in total 44 protein sequences that can be considered true SFPs. In reality, most of these sequences fulfil many, or indeed all, of the above requirements. The remaining sequences may indeed be ejaculated proteins that were not identified within this whole ejaculate analysis; however, the evidence is currently not strong enough to confirm this.

Phenotypic correlations

The data presented here build upon a previous 2D-SDS PAGE analysis of male accessory sex glands (Goenaga *et al*., 2015). In this study, 2D gel images of male accessory gland homogenates were compared with those from virgin and just-mated females, revealing 127 spots that appeared to be transferred at mating. Using Progenesis SameSpots software (Nonlinear Dynamics, Newcastle, UK), these gel images were normalized to make the volume of spots (a measure of relative protein abundance) comparable quantitatively. Across 15 different populations of *C. maculatus*, the volume of gel spots correlated with reproductive phenotypes. Both defence (P1) and offence (P2) components of sperm competition success were measured using a standard sterile male technique (Boorman & Parker, 1976; Simmons, 2001). Here, females are mated to two males in succession, one of which is irradiated such that his sperm remain motile and able to fertilize eggs but carry lethal

mutations that render the eggs inviable, and the other male is focal and fertile. Here, P1 and P2 denote the proportion of offspring fertilized by the focal male when he is first or second to mate, respectively, with a given female in such a double mating experiment. Thus, for example, a protein spot for which volume correlated with P1 contains a protein/proteins that was statistically associated with males' relative fertilization success when mating as a first male. Additionally, protein spot size was correlated with fecundity when not mating in a competitive scenario; the number of offspring an individual male produces when mating to a standard virgin female. One aim of the present study was then to identify proteins within such spots.

As recently stressed by Sirot *et al*. (2014), the sexual selection paradigm of SFP evolution predicts both complexity and functional redundancy of SFPs. Because SFPs serve as agents that influence many different postmating aspects of female physiology and behaviour, any novel SFP that benefits males will be favoured by selection and we would thus predict the evolutionary accumulation of a diverse and at least partly redundant set of SFPs. Indeed, our previous study (Goenaga *et al*., 2015), investigating intraspecific associations between the size of 2D SDS-PAGE spots and reproductive phenotypes, found many spots to correlate with the same reproductive phenotypes. Our current findings, in combination with the results presented by Goenaga *et al*. (2015), are in line with these general predictions because many proteins appear to have similar functions in terms of their downstream effect on reproductive phenotypes. For example, we have now identified seven different SFPs that predict male sperm competition success and five different SFPs that predict female egg production after mating. Additionally, the 15 sequences that were found to be male specific in the transcriptomic analysis all correlated with reproductive phenotype, either female fecundity or male competitive fertilization success. These sequences occurred either within multiple spots, or within spots for which multiple proteins were present, and so firm conclusions about specific proteins cannot be drawn here. However, all evidence suggests that these proteins in particular are of significance to male reproductive success within *C. maculatus*.

Although 34 of the gel spots tested here were chosen because of their association with reproductive phenotype, the presence of multiple proteins within most of these spots limits the conclusions that can be made from these identifications. Excluding gel spots with multiple proteins, 12 identified sequences were from gel spots that correlated with a reproductive phenotype in a previous analysis (Goenaga *et al*., 2015; Table S2). Some caution must be taken when interpreting these results as correlation does not equal causation; however, they offer hypotheses for further investigation. For example, the size of the gel spot containing a protein homologous to the detoxification enzyme glutathione S-transferase was positively correlated with P1, or sperm competition success as the first male to mate. Glutathione S-transferase is well documented within mammalian sperm (eg Gupta, 2006) and has been identified as SFPs of both *Aedes aegypti* and *Aedes albopictus* (Sirot *et al*., 2011; Boes *et al*., 2014). This protein is known to be highly expressed in sperm storage organs of the honeybee (Collins *et al*., 2004), and is thought to offer sperm some protection against oxidative damage, which may explain the correlation with P1 fertilization success.

Three of the proteins identified in spots associated with fecundity or sperm competition success have no known homologous protein matches, and may be novel SFPs. The remaining single sequences within gel spots associated with a reproductive trait contained homologous proteins that have roles in mitochondrial energy production or lipid metabolism. There are both positive and negative associations with these proteins and fecundity and male success in sperm competition. The detailed causes for these correlations are difficult to untangle, but we note that it is perhaps not surprising that a cocktail of metabolic proteins within the seminal fluid affects male fertility. Metabolic proteins, particularly those involved in lipid metabolism, have been previously described in the seminal fluid of species within *Drosophila* genus (Findlay *et al*., 2008; Kelleher *et al*., 2009). These proteins may assist with sperm motility and storage within the female reproductive tract.

Conclusions

We have identified 98 candidate SFPs in *C. maculatus*, using proteomic analysis of 55 gel spots. These were predicted to be transferred to females based upon a 2D SDS-PAGE analysis. Most proteins identified here had functions that are in line with SFPs of other species. Analysis of whole ejaculates extracted from just-mated females further confirmed that many of these proteins are transferred at mating. Additionally, 54 of the 98 sequences show signals of secretion, a common requirement for ejaculated proteins. Combining this analysis with transcript expression data revealed that 84 transcripts were differentially expressed between the sexes, including 15 that were particularly highly expressed in males compared with females. The protein sequences encoded by all 15 of these transcripts had predicted signal sequences of secretion, and all bar one were found within the whole ejaculate analysis. Of these, nine had no significant homology to known proteins and are therefore considered novel SFPs. Further work is required to understand the precise mechanisms by which these proteins alter female responses to reproduction.

Experimental procedures

Origin of protein spots

This study follows a previous investigation of interpopulation diversification in the relative abundance of accessory gland proteins (Goenaga *et al.*, 2015), so we refer to this study for a more detailed description of sample collection and the identification of relevant spots. Briefly, these spots were picked from 2D SDS-PAGE gels, taken from accessory reproductive glands of males from 15 different focal populations of *C. maculatus*. Populations were reared under common garden conditions on black-eyed beans (*Vigna unguiculata*) in the laboratory at 29°C, 60% relative humidity (RH) and a 12 h light (L) : 12 h dark (D) light cycle. For each population, 60 pairs of accessory reproductive glands from 40–50 individual virgin males (0–1 day old) were dissected on ice and pooled into 60 μl Milli-Q water (Merk Milliporem Billerica, MA, USA) (four replicate samples per population). Samples were stored at −80°C prior to separation via 2D SDS-PAGE. Progenesis SameSpots software (Nonlinear Dynamics) was used to compare the 2D gels produced using the male accessory gland tissue with other 2D gels containing either virgin female reproductive bursae or just-mated female reproductive bursae. This allowed the location, on the 2D gels, of male-derived accessory gland proteins that are transferred to the female at mating to be deduced (Fig. 1). In total, 127 such spots were predicted to be transferred to the female and 55 of these were picked here for protein identification. These were either shown to correlate strongly with a postmating reproductive phenotype (26 spots), male aedeagus spine length (eight spots) or were large and consistently present within the accessory glands of all male populations (21 spots).

Protein identification using Mass Spectrometry

For proteomic analysis, the 55 isolated gel spots were destained twice with acetonitrile (ACN) and dried using a Speed Vac (Eppendorf, Hamburg, Germany) v at 30°C. Samples were incubated in 10 mM dithiothreitol (DTT) and then 50 mM iodoacetamide (IAA) to reduce and then alkylate the proteins. After rinsing the gel spots with ACN and drying in the Speed Vac, proteins were digested with trypsin (12.5 ng/μl in 25 mM ammonium bicarbonate, Promega, Madison, WI, USA) overnight at 37°C. The supernatant was collected and further peptides isolated from the gel spots by washing with 60% volume/volume % (v/v) ACN and 5% v/v formic acid (FA). The pooled extract was dried using the Speed Vac prior to MS analysis by MALDI-TOF. For MALDI-TOF analysis, dried samples were resuspended in solvent solution (30% ACN with 1% FA) and 1 μl of this, mixed 50:50 with a saturated α-cyano-4-hydroxycinnamic acid solution, added to a target plate. The sample and matrix were left to dry before MALDI-TOF analysis using an Ultraflex (Bruker, Billerica, MA, USA).

The acquired data (.RAW files) were searched against an in-house transcriptome data set (Sayadi *et al.*, 2016), produced using Trinity (Grabherr *et al.*, 2011) (see below for details). Searches were performed using the MASCOT search algorithm (Perkins et al., 1999) via Proteome Discoverer software (v. 1.4.0.288, Thermo Fisher Scientific, Bremen, Germany), allowing for carbamidomethyl (C), oxidation (M) and deamidation (N, Q) modifications, and two missed cleaves. The search parameters included: maximum 10 ppm and 0.02 Da error tolerance for the survey scan and MS/MS analysis. Proteins identified with at least two matching peptide sequences of 95% confidence per protein were analysed further, in both an automated and manual manner, to confirm the identified protein sequence. TransDecoder (https://transdecoder.github.io/; Haas *et al.*, 2013) was used to identify candidate protein-coding regions of the matched transcript sequences. These were manually compared with the peptide matches to ensure that all peptides assigned to a sequence were identified within a protein-coding region. The validity of each protein hit was further assessed by considering the quality of the peptide hits and the mass of the predicted protein in relation to the gel location. Samples that did not give clear protein identifications following MALDI-TOF analysis were also submitted for LC-MSMS analysis.

Dried peptides were resuspended in 0.1% FA and analysed with a QExactive Plus Orbitrap mass spectrometer (Thermo Fisher Scientific) equipped with a nano electrospray ion source. Peptides were upfront separated by reversed phase liquid chromatography using an EASY-nLC 1000 system (Thermo Fisher Scientific). A set-up of pre-column and analytical column was used. The pre-column was a 2-cm EASY-column (internal diameter (ID) 100 μm, 5 μm C18; Thermo Fisher Scientific) and the analytical column was a 10-cm EASY-column (ID 75 μm, 3 μm, C18; Thermo Fisher Scientific). Peptides were eluted with a 35-min linear gradient from 4 to 100% acetonitrile at 250 nl/min. The mass spectrometer was operated in positive ion mode, acquiring a survey mass spectrum with resolving power 70 000 and consecutive high collision dissociation fragmentation spectra of the 10 most abundant ions.

The raw data from the LC-MSMS was searched against the in-house transcriptome using MASCOT through Proteome Discoverer (Thermo Fisher Scientific). Again, up to two missed cleaves and modifications on carbamidomethyl (C) (fixed), and oxidation (M) and deamidation (N, Q) (variable) were allowed for. Only transcript matches with at least two high-quality peptide matches were considered for further analysis. Transcribed protein sequences were assessed, as for the MALDI-TOF data, to ensure that protein sequences were identified with confidence.

Whole ejaculate collections

To confirm that these proteins were indeed transferred to females at mating, entire ejaculates were collected from females immediately after mating, for proteomic analysis. Females ($N = 4$) from an isogenic reference population (South India SI4) were paired with males of the same population until mating occurred. Immediately after the pair separated (within seconds), the bursa copulatrix (containing the ejaculate) was dissected out from the female and placed into 5 μl lysis buffer (20 mM hydroxyethyl piperazineethanesulfonic acid (HEPES), 9 M urea, complete mini ethylenediaminetetraacetic acid-free protease inhibitor cocktail). A dissection needle was used to homogenize the ejaculate, which was briefly centrifuged before storing in the freezer prior to LC-MSMS analysis.

For proteomic analysis, samples were digested using trypsin in a similar manner to the gel spot samples. Proteins were reduced with DTT (final concentration 48 mM) and alkylated with IAA (final concentration 25 mM). Samples were diluted four

times with 50 mM ammonium bicarbonate prior to tryptic enzymatic digestion (enzyme : protein ratio 1:20, Promega) at 37°C overnight. The reaction was stopped by acidifying the sample with trifluoroacetic acid. Prior to the analysis by mass spectrometry, the peptides were purified by Pierce C18 Spin Columns (Thermo Fisher Scientific) and dried in a SpeedVac system. Peptides from ejaculate samples were analysed as described above for LC-MSMS of gel spots, with exception of the LC gradient, which lasted 90 min.

Homologous protein searching

Homologous protein matches and gene ontology information was gathered using Blast2Go (Conesa *et al*., 2005) by searching the protein sequences against all nonredundant protein sequences within the NCBI database using an e-value $< 1 \times 10^{-5}$. Proteins were considered homologous if they contained a minimum of 60% sequence coverage (mean 96%) and minimum 50% positive amino acid matches (mean 84%). Functional annotation was based upon these homologous protein matches. Graphical representations of these data were created using Blast2Go (Fig. 2). SignalP (4.1) and secretomeP (2.0) were used to predict whether the identified protein sequences were likely to be secreted (Bendtsen *et al*., 2004; Petersen *et al*., 2011).

The identified protein sequences were also compared with SFPs identified in the red flour beetle (*T. castaneum*; Parthasarathy *et al*., 2009; South *et al*., 2011; Xu *et al*., 2013) and to ejaculate proteins identified in honeybees (*A. mellifera*; Baer *et al*., 2009) and fruit flies (*D. melanogaster*; Findlay *et al*., 2008). The desktop version of Blast was used to create a database of the sequences collected from the literature and online resources (Camacho *et al*., 2009). The protein sequences identified within the gel spots were searched against this database using tBlastn. Proteins were considered homologous if they had e-values $< 1 \times 10^{-6}$ and identities $\geq 30\%$.

Transcriptome data

The in-house *de novo* transcriptome was prepared as described in Sayadi *et al*. (2016). Briefly, RNA was extracted from our isogenic *C. maculatus* reference population (SI4), reared on mung beans at 29°C, 60% RH and a 12 h L: 12 h D light cycle, at various developmental and physiological stages. To cover a range of expressed transcripts, 27 different libraries were prepared. These included samples from larvae and pupae, as well as replicated samples of male and female adults, both virgin and mated (thorax and abdomen separately). RNA was extracted from six individuals per sample, using a RNeasy Mini Kit (Qiagen, Valencia, CA, USA) following the manufacturer's protocol. The RNA sequencing libraries were prepared from 1 μg total RNA using the Illumina TruSeq stranded mRNA sample preparation kit (RS-200-9002DOC, Illumina, San Diego, California, USA). Library generation and sequencing were performed by the SNP&SEQ Technology Platform at Uppsala University. The resulting *de novo* transcriptome was used here for protein sequence identification.

Following protein identification, the transcript expression data were used to further investigate these sequences. Comparisons of the relevant transcripts, present in the abdomens of male and female, virgin and mated, adults, were performed. The 98 protein sequences identified from within the protein spots were searched against the transcriptome used to calculate the expression data using the tblastx function within the standalone Blast program. A total of 137 transcript IDs was then selected based on the best matches (e-values $< 1 \times 10^{-20}$; mean identities 99%, range: 84–100%). The discrepancy between the number of proteins and the number of transcripts is because of the presence of multiple isoforms. Of the 98 protein sequences, 69 matched a single isoform in the transcriptome. A further 19 protein sequences matched two isoforms, and eight matched three or more isoforms. Expression data were considered for all isoforms in order to be comprehensive.

Transcript expression data were analysed in R (v. 3.2.2) (R Development Core Team, 2011) using the packages edgeR v. 3.10.5 (Robinson *et al*., 2010) and limma v. 3.24.15 (Smyth, 2005; Ritchie *et al*., 2015). Data were normalized using a weighted trimmed mean of M-values (Robinson & Oshlack, 2010). A GLM likelihood ratio test, fitting a negative binomial GLM with Cox–Reid dispersion estimates, was used to investigate patterns of differential expression between the abdominal samples from males and females, both virgin and mated. The *P*-values were adjusted for multiple testing using the FDR (Benjamini & Hochberg, 1995). In order to visualize the differential expression, FPKM values were calculated for the whole transcriptome using the RSEM package v. 1.2.25 (Li & Dewey, 2011). The data for the relevant transcripts were isolated and a heatmap produced in R using the packages pheatmap (Kolde, 2015) and gplots (Warnes *et al*., 2015) (Fig. 3).

Acknowledgements

This work was supported by the Mass Spectrometry Based Proteomics Facility and the sequencing was performed by the SNP&SEQ Technology Platform at Uppsala University. These facilities are part of the National Genomics Infrastructure of Sweden and the Science for Life Laboratory. The SNP&SEQ Platform is supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation. Data storage was obtained and supported by Bioinformatics Infrastructure for Life Sciences. This work was financially supported by grants from the European Research Council (GENCON AdG-294333), the Swedish Research Council (621-2010-5266) and the Wallenberg Bioinformatics Long-term Support at Uppsala University.

References

Andrés, J.A., Maroja, L.S. and Harrison, R.G. (2008) Searching for candidate speciation genes using a proteomic approach: seminal proteins in field crickets. *Proc R Soc Lond B Biol Sci* **275**: 1975–1983.

Arnqvist, G. and Rowe, L. (2005) *Sexual Conflict*. Princeton University Press, Princeton, NJ.

Avila, F.W., Cohen, A.B., Ameerudeen, F.S., Duneau, D., Suresh, S., Mattei, A.L. *et al*. (2015) Retention of ejaculate by

Drosophila melanogaster females requires the male-derived mating plug protein PEBme. *Genetics* **200**: 1171.

Baer, B., Heazlewood, J.L., Taylor, N.L., Eubel, H. and Millar, A.H. (2009) The seminal fluid proteome of the honeybee *Apis mellifera*. *Proteomics* **9**: 2085–2097.

Bayram, H.L., Claydon, A.J., Brownridge, P.J., Hurst, J.L., Mileham, A., Stockley, P. *et al*. (2016) Cross-species proteomics in analysis of mammalian sperm proteins. *J Proteomics* **135**: 38–50.

Bendtsen, J.D., Jensen, L.J., Blom, N., Von Heijne, G. and Brunak, S. (2004) Feature-based prediction of non-classical and leaderless protein secretion. *Protein Eng Des Sel* **17**: 349–356.

Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Methodol* **57**: 289–300.

Boes, K.E., Ribeiro, J.M.C., Wong, A., Harrington, L.C., Wolfner, M.F. and Sirot, L.K. (2014) Identification and characterization of seminal fluid proteins in the Asian tiger mosquito, *Aedes albopictus*. *PLoS Negl Trop Dis* **8**: e2946.

Boorman, E. and Parker, G.A. (1976) Sperm (ejaculate) competition in *Drosophila melanogaster*, and the reproductive value of females to males in relation to female age and mating status. *Ecol Entomol* **1**: 145–155.

Bork, P., Dandekar, T., Diaz-Lazcoz, Y., Eisenhaber, F., Huynen, M. and Yuan, Y. (1998) Predicting function: from genes to genomes and back1. *J Mol Biol* **283**: 707–725.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. *et al*. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.

Campostrini, N., Areces, L.B., Rappsilber, J., Pietrogrande, M.C., Dondi, F., Pastorino, F. *et al*. (2005) Spot overlapping in two-dimensional maps: a serious problem ignored for much too long. *Proteomics* **5**: 2385–2395.

Chapman, T. (2001) Seminal fluid-mediated fitness traits in *Drosophila*. *Heredity* **87**: 511–521.

Civetta, A. and Singh, R.S. (1995) High divergence of reproductive tract proteins and their association with postzygotic reproductive isolation in *Drosophila melanogaster* and *Drosophila virilis* group species. *J Mol Evol* **41**: 1085–1095.

Collins, A.M., Williams, V. and Evans, J.D. (2004) Sperm storage and antioxidative enzyme expression in the honey bee, *Apis mellifera*. *Insect Mol Biol* **13**: 141–146.

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M. and Robles, M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**: 3674–3676.

Dean, M.D., Findlay, G.D., Hoopmann, M.R., Wu, C.C., MacCoss, M.J., Swanson, W.J. *et al*. (2011) Identification of ejaculated proteins in the house mouse (*Mus domesticus*) via isotopic labeling. *BMC Genomics* **12**: 306.

Dhole, S. and Servedio, M.R. (2014) Sperm competition and the evolution of seminal fluid composition. *Evolution* **68**: 3008–3019.

Dorus, S., Skerget, S. and Karr, T.L. (2012) Proteomic discovery of diverse immunity molecules in mammalian spermatozoa. *Syst Biol Reprod Med* **58**: 218–228.

Evans, V.C., Barker, G., Heesom, K.J., Fan, J., Bessant, C. and Matthews, D.A. (2012) De novo derivation of proteomes from transcriptomes for transcript and protein identification. *Nat Methods* **9**: 1207–1211.

Fedorka, K.M., Winterhalter, W.E. and Ware, B. (2011) Perceived sperm competition intensity influences seminal fluid protein production prior to courtship and mating. *Evolution* **65**: 584–590.

Findlay, G.D. and Swanson, W.J. (2010) Proteomics enhances evolutionary and functional analysis of reproductive proteins. *Bioessays* **32**: 26–36.

Findlay, G.D., Yi, X., MacCoss, M.J. and Swanson, W.J. (2008) Proteomics reveals novel *Drosophila* seminal fluid proteins transferred at mating. *PLoS Biol* **6**: e178.

Fitch, W.M. (2000) Homology: a personal view on some of the problems. *Trends Genet* **16**: 227–231.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D. *et al*. (2005) Protein identification and analysis tools on the ExPASy server. In *The Proteomics Protocols Handbook* (Walker, J., ed.), pp. 571–607. Humana Press, Totowa, NJ.

Goenaga, J., Yamane, T., Rönn, J. and Arnqvist, G. (2015) Within-species divergence in the seminal fluid proteome and its effect on male and female reproduction in a beetle. *BMC Evol Biol* **15**: 266.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I. *et al*. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotech* **29**: 644–652.

Gupta, G.S. (2006) *Proteomics of Spermatogenesis*. Springer Science & Business Media, New York.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J. *et al*. (2013) De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. *Nat Protoc* **8**: 1494–1512.

Haerty, W., Jagadeeshan, S., Kulathinal, R.J., Wong, A., Ram, K.R., Sirot, L.K. *et al*. (2007) Evolution in the fast lane: rapidly evolving sex-related genes in *Drosophila*. *Genetics* **177**: 1321–1335.

Heifetz, Y., Vandenberg, L.N., Cohn, H.I. and Wolfner, M.F. (2005) Two cleavage products of the *Drosophila* accessory gland protein ovulin can independently induce ovulation. *Proc Natl Acad Sci USA* **102**: 743–748.

Hoffmann, J.A., Kafatos, F.C., Janeway, C.A., and Ezekowitz, R.A. (1999) Phylogenetic perspectives in innate immunity. *Science* **284**: 1313–1318.

Hotzy, C. and Arnqvist, G. (2009) Sperm competition favors harmful males in seed beetles. *Curr Biol* **19**: 404–407.

Hotzy, C., Polak, M., Rönn, J.L. and Arnqvist, G. (2012) Phenotypic engineering unveils the function of genital morphology. *Curr Biol* **22**: 2258–2261.

Kapelnikov, A., Zelinger, E., Gottlieb, Y., Rhrissorrakrai, K., Gunsalus, K.C. and Heifetz, Y. (2008) Mating induces an immune response and developmental switch in the *Drosophila* oviduct. *Proc Natl Acad Sci USA* **105**: 13912–13917.

Kelleher, E.S., Watts, T.D., LaFlamme, B.A., Haynes, P.A. and Markow, T.A. (2009) Proteomic analysis of *Drosophila mojavensis* male accessory glands suggests novel classes of seminal fluid proteins. *Insect Biochem Mol Biol* **39**: 366–371.

King, M., Eubel, H., Millar, A.H. and Baer, B. (2011) Proteins within the seminal fluid are crucial to keep sperm viable in the honeybee *Apis mellifera*. *J Insect Physiol* **57**: 409–414.

Kolde, R. (2015) *pheatmap: Pretty Heatmaps*. R package version 1.0.8. https://CRAN.R-project.org/package=pheatmap

LaFlamme, B.A. and Wolfner, M.F. (2013) Identification and function of proteolysis regulators in seminal fluid. *Mol Reprod Dev* **80**: 80–101.

LaFlamme, B.A., Ravi Ram, K. and Wolfner, M.F. (2012) The *Drosophila melanogaster* seminal fluid protease "seminase" regulates proteolytic and post-mating reproductive processes. *PLoS Genet* **8**: e1002435.

Li, B. and Dewey, C.N. (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**: 323.

Liou, H.L., Dixit, S.S., Xu, S., Tint, G.S., Stock, A.M. and Lobel, P. (2006) NPC2, the protein deficient in Niemann-Pick C2 disease, consists of multiple glycoforms that bind a variety of sterols. *J Biol Chem* **281**: 36710–36723.

Lung, O., Tram, U., Finnerty, C.M., Eipper-Mains, M.A., Kalb, J.M. and Wolfner, M.F. (2002) The *Drosophila melanogaster* seminal fluid protein Acp62F is a protease inhibitor that is toxic upon ectopic expression. *Genetics* **160**: 211–224.

Mane, S.D., Tompkins, L. and Richmond, R.C. (1983) Male esterase 6 catalyzes the synthesis of a sex pheromone in *Drosophila melanogaster* females. *Science* **222**: 419–421.

Marshall, J.L., Huestis, D.L., Hiromasa, Y., Wheeler, S., Oppert, C., Marshall, S.A. *et al*. (2009) Identification, RNAi knockdown, and functional analysis of an ejaculate protein that mediates a postmating, prezygotic phenotype in a cricket. *PLoS One* **4**: e7537.

Mueller, J.L., Ripoll, D.R., Aquadro, C.F. and Wolfner, M.F. (2004) Comparative structural modeling and inference of conserved protein classes in *Drosophila* seminal fluid. *Proc Natl Acad Sci USA* **101**: 13542–13547.

Mueller, J.L., Linklater, J.R., Ram, K.R., Chapman, T. and Wolfner, M.R. (2008) Targeted gene deletion and phenotypic analysis of the *Drosophila melanogaster* seminal fluid protease inhibitor Acp62F. *Genetics* **178**: 1605–1614.

Palmer, M.R., McDowall, M.H., Stewart, L., Ouaddi, A., MacCoss, M.J. and Swanson, W.J. (2013) Mass spectrometry and next-generation sequencing reveal an abundant and rapidly evolving abalone sperm protein. *Mol Reprod Dev* **80**: 460–465.

Park, M. and Wolfner, M.F. (1995) Male and female cooperate in the prohormone-like processing of a *Drosophila melanogaster* seminal fluid protein. *Dev Biol* **171**: 694–702.

Parker, G.A. (1970) Sperm competition and its evolutionary consequences in the insects. *Biol Rev* **45**: 525–567.

Parthasarathy, R., Tan, A., Sun, Z., Chen, Z., Rankin, M. and Palli, S.R. (2009) Juvenile hormone regulation of male accessory gland activity in the red flour beetle, *Tribolium castaneum*. *Mech Dev* **126**: 563–579.

Perkins, D. N., Pappin, D. J. C., Creasy, D. M. and Cottrell, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, **20**: 3551–3567.

Perry, J.C., Sirot, L. and Wigby, S. (2013) The seminal symphony: how to compose an ejaculate. *Trends Ecol Evol* **28**: 414–422.

Petersen, T.N., Brunak, S., von Heijne, G. and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* **8**: 785–786.

Pilch, B. and Mann, M. (2006) Large-scale and high-confidence proteomic analysis of human seminal plasma. *Genome Biol* **7**: R40.

Poiani, A. (2006) Complexity of seminal fluid: a review. *Behav Ecol Sociobiol* **60**: 289–310.

R Development Core Team (2011) *R: a language and environment for statistical computing*. Vienna, Austria: the R foundation for Statistical Computing. http://www.R-project.org/

Ramm, S.A., Oliver, P.L., Ponting, C.P., Stockley, P. and Emes, R.D. (2008) Sexual selection and the adaptive evolution of mammalian ejaculate proteins. *Mol Biol Evol* **25**: 207–219.

Ramm, S.A., Edward, D.A., Claydon, A.J., Hammond, D.E., Brownridge, P., Hurst, J.L. *et al*. (2015) Sperm competition risk drives plasticity in seminal fluid composition. *BMC Biol* **13**: 87.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W. *et al*. (2015) Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**: e47.

Robinson, M.D. and Oshlack, A. (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* **11**: R25.

Robinson, M.D., McCarthy, D.J. and Smyth, G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.

Sayadi, A., Immonen, E., Bayram, H. and Arnqvist, G. (2016) The de novo transcriptome and its functional annotation in the seed beetle *Callosobruchus maculatus*. *PLoS ONE* **11**: e0158565.

Simmons, L.W. (2001) *Sperm Competition and Its Evolutionary Consequences in the Insects*. Princeton University Press, Princeton, NJ.

Simmons, L.W. and Beveridge, M. (2011) Seminal fluid affects sperm viability in a cricket. *PLoS One* **6**: e17975.

Simmons, L.W., Tan, Y.F. and Millar, A.H. (2013) Sperm and seminal fluid proteomes of the field cricket Teleogryllus oceanicus: identification of novel proteins transferred to females at mating. *Insect Mol Biol* **22**: 115–130.

Sirot, L.K., Hardstone, M.C., Helinski, M.E.H., Ribeiro, J.M.C., Kimura, M., Deewatthanawong, P. *et al*. (2011) Towards a semen proteome of the dengue vector mosquito: protein identification and potential functions. *PLoS Negl Trop Dis* **5**: e989.

Sirot, L.K., Wong, A., Chapman, T. and Wolfner, M.F. (2014) Sexual Conflict and Seminal Fluid Proteins: A Dynamic Landscape of Sexual Interactions. In *The Genetics and Biology of Sexual Conflict* (Rice, W.R., and Gavrilets, S., eds) pp. 49–72. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Smith, J.R. and Stanfield, G.M. (2012) A seminal fluid protease activates sperm motility in *C. elegans* males. *Worm* **1**: 151–154.

Smyth, G.K. (2005) Limma: linear models for microarray data. In *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* (Gentleman, R., Carey, V.J., Huber, W., Irizarry, R.A. and Dudoit, S. eds.), pp. 397–420. Springer, New York.

South, A., Sirot, L.K. and Lewis, S.M. (2011) Identification of predicted seminal fluid proteins in *Tribolium castaneum*. *Insect Mol Biol* **20**: 447–456.

Swanson, W.J., Clark, A.G., Waldrip-Dail, H.M., Wolfner, M.F. and Aquadro, C.F. (2001) Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. *Proc Natl Acad Sci USA* **98**: 7375–7379.

Tatusov, R.L., Koonin, E.V. and Lipman, D.J. (1997) A genomic perspective on protein families. *Science* **278**: 631–637.

Walters, J.R. and Harrison, R.G. (2010) Combined EST and proteomic analysis identifies rapidly evolving seminal fluid proteins in *Heliconius* butterflies. *Mol Biol Evol* **27**: 2000–2013.

Warnes, G.R., Bolker, B., Bonebakker, L., Gentleman, R., Liaw, W.H.A., Lumley, T. *et al*. (2015) *gplots: Various R Programming Tools for Plotting Data*. R package version 3.0.1. https://CRAN.R-project.org/package=gplots

Wilburn, D.B. and Swanson, W.J. (2015) From molecules to mating: rapid evolution and biochemical studies of reproductive proteins. *J Proteomics* **135**: 12–25.

Xu, J., Baulding, J. and Palli, S.R. (2013) Proteomics of *Tribolium castaneum* seminal fluid proteins: identification of an angiotensin-converting enzyme as a key player in regulation of reproduction. *J Proteomics* **78**: 83–93.

Yamane, T., Miyatake, T. and Kimura, Y. (2008) Female mating receptivity after injection of male-derived extracts in *Callosobruchus maculatus*. *J Insect Physiol* **54**: 1522–1527.

Yamane, T., Goenaga, J., Rönn, J.L. and Arnqvist, G. (2015) Male seminal fluid substances affect sperm competition success and female reproductive behavior in a seed beetle. *PLoS ONE* **10**: e0123770.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table S1. Fragments per kilobase of transcript per million mapped reads values for the transcripts that matched to the identified proteins. Data for all adult abdomen samples are presented here. Sex difference in transcript expression was calculated as described in the Experimental procedures. Positive values indicate a male-biased expression, and negative female biased. The given *P*-values were adjusted to account for multiple testing using the false discovery rate (FDR). Data for virgin and mated samples were analysed separately and the results are highly comparable.

Table S2. The reproductive phenotypes associated with gel spots containing a single protein. Top homologous protein matches are reported, with more information being present within Table 1. Trait values were collected through testing males of 15 *Callosobruchus maculatus* populations mated to a single female phenotype for fecundity (Fec, number of hatching offspring) and male sperm competition success, at multiple time points postmating. Both defence (P1) and offence (P2) aspects of sperm competition were measured using a standard sterile male technique. As described in Goenaga *et al*. (2015), the trait values were correlated against gel spot sizes taken from 2D gel images of male accessory gland proteins, using Progenesis SameSpots software (Nonlinear Dynamics). Spots with significant correlations were chosen for proteomic analysis. The results for the spots containing only one protein are presented here.