# Proteomics in Evolutionary Ecology

Running title: Evolutionary dynamics in the proteome

B. Baer & A. H. Millar

*Centre for Integrative Bee Research (CIBER) and ARC Centre of Excellence in Plant Energy Biology, Bayliss Building, The University of Western Australia, 6009 Crawley, Australia*

for correspondence
Phone ++61 8 6488 4495
Fax ++61 8 6488 1029
Email: boris.baer@uwa.edu.au

28

29 ## Abstract

30

31 Evolutionary ecologists are traditionally gene-focused, as genes propagate phenotypic
32 traits across generations and mutations and recombination in the DNA generate genetic
33 diversity required for evolutionary processes. As a consequence, the inheritance of
34 changed DNA provides a molecular explanation for the functional changes associated
35 with natural selection. A direct focus on proteins on the other hand, the actual molecular
36 agents responsible for the expression of a phenotypic trait, receives far less interest from
37 ecologists and evolutionary biologists. This is partially due to the central dogma of
38 molecular biology that appears to define proteins as the 'dead-end of molecular
39 information flow' as well as technical limitations in identifying and studying proteins and
40 their diversity in the field and in many of the more exotic genera often favored in
41 ecological studies. Here we provide an overview of a newly forming field of research that
42 we refer to as '*Evolutionary Proteomics*'. We point out that the origins of cellular
43 function are related to the properties of polypeptide and RNA and their interactions with
44 the environment, rather than DNA descent, and that the critical role of horizontal gene
45 transfer in evolution is more about coopting new proteins to impact cellular processes
46 than it is about modifying gene function. Furthermore, post-transcriptional and post-
47 translational processes generate a remarkable diversity of mature proteins from a single
48 gene, and the properties of these mature proteins can also influence inheritance through
49 genetic and perhaps epigenetic mechanisms. The influence of post-transcriptional
50 diversification on evolutionary processes could provide a novel mechanistic underpinning

for elements of rapid, directed evolutionary changes and adaptations as observed for a variety of evolutionary processes. Modern state-of the art technologies based on mass spectrometry are now available to identify and quantify peptides, proteins, protein modifications and protein interactions of interest with high accuracy and assess protein diversity and function. Therefore, proteomic technologies can be viewed as providing evolutionary biologist with exciting novel opportunities to understand very early events in functional variation on cellular molecular machinery that are acting as part of evolutionary processes.

Key words: Evolution, natural selection, sexual selection, peptide mass spectrometry, protein diversity, post-translational modification, protein-protein interaction.

## Introduction

Evolutionary theory as initially formulated by Charles Darwin [1] has become a foundation for biological sciences and ranks among mankind's most important scientific discoveries. The empirical support for evolutionary theory shows that traits under natural selection require two characteristics that make them evolvable: variation and inheritance. For evolutionary processes such as for example host-parasite / predator-prey coevolution, sexual selection or ecological adaptation to occur, phenotypic variation between individuals needs to be generated and maintained for a trait so that selection can differentially act upon them. Furthermore, traits need to be heritable so that individuals

with advantageous characteristics pass these onto the next generation and thereby change their frequency within a population, resulting in fundamental biological developments such as adaptation and speciation. The discovery of DNA by Watson and Crick [2] as the molecule responsible for biological information storage and inheritance offered biologists the possibility to develop and use a range of molecular tools such as PCR, sequencing and cloning to understand the implications of Darwinian thinking at the molecular level and across evolutionary timescales. Crick also formulated the central dogma for molecular biology [3] (Figure 1), where heritable information is coded as genes, typically DNA but sometimes RNA, from which proteins are produced via transcription and translation. The central dogma presents proteins as the endpoint of information flow where any changes are not translated back to RNA and DNA and thus proteins are typically not considered as drivers of evolutionary processes. As a consequence of this history, evolutionary biologists are predominantly gene-focused and the technical opportunities to aid their study of genes and genomes have developed at breathtaking speed up to the present day. The gene mutation paradigm as the key to evolution has dominated modern molecular biology, there have been prominent thinkers such as Woese [4-6] however who proposed RNA and translation to protein as central drivers of phylogenetic relationships in the tree of life. Woese also highlighted the role of horizontal gene transfer between prokaryotic cells (i.e. the swapping of DNA encoding for a whole new protein in bacteria and archea, Figure 1) as more critical to great swathes of evolution than point mutation of the organism's own genes [4, 5]. Horizontal gene transfer allows for rapid evolution to occur at the level of the ecosystem rather than the level of the organism and the introduction of an entirely new protein agent into a cellular

97 milieu and indeed into a protein network. The importance of horizontal gene transfer in
98 fungal evolution [7] and even in very recent examples of grains in pathogenicity of fungi
99 is well documented [8].
100
101 The full genome sequences of thousands of species are now available allowing
102 unprecedented base by base comparison of genes within and across families, genera and
103 kingdoms and increasingly more sophisticated methodologies are also available to
104 permanently or transiently manipulate gene sequence and expression to observed the
105 effects. However, while this genomics has generated solid empirical evidence for
106 evolutionary theory and provided detailed insights into evolutionary dynamics (e.g. [9]),
107 a range of more fundamental questions still remained unresolved. For example,
108 comparative genomics reveals that many genes often remain remarkably similar
109 throughout evolutionary history, providing, for example, only preliminary answers to the
110 question of why chimpanzees are chimpanzees and humans are humans based on DNA
111 sequence alone [10, 11]. It is widely acknowledged that regulation of expression of
112 genomes is the key differentiator between mammals but it remains unclear, how
113 differences in gene expression of an identical gene pool can generate the tremendous
114 phenotypic variation observed such as for example between humans and chimpanzees
115 [10, 11]. Furthermore, dependence of molecular evolution of DNA on random mutations
116 alone resulting in the eventual appearance of a gene with superior functionality [12-14]
117 would relegate evolution to depend predominantly on chance events acted on by selective
118 forces across generations. In our view while the focus on point mutation alone is
119 weakening, the current framework pursued by many researchers still provides an

unsatisfying and insufficient explanation for fast co-evolving traits such as for example those under sexual selection or host-parasite co-evolution, where heritable changes in phenotype can often become visible within a handful of generations [15-19].

The predominance of evolutionary studies still focuses on genes and genomes through measures of mutation rates and genotype frequency changes in populations. There is no current evolutionary framework or substantial research literature to understand the importance of the role of translated agents – the proteins and their function - as drivers of adaptation. This can be very simplistically illustrated by the co-occurrence of 'genome and evolution' and 'proteome and evolution' in PubMed: a close to 50-fold difference in co-occurrence exists. Woese [4] pointed out a similar dilemma for RNA biology a decade ago where the importance of studying the evolution of translation of RNA to protein did not fit within the molecular biology paradigm. As he pointed out in his seminal contribution, "*molecular biology has to bring evolution to the fore and integrate it fully – not hold it at arm's length*" [5]. Much has changed to resolve this as the explosion of data on numerous levels of RNA biology and the biological role of non-translated RNAs in influencing DNA [20] (Figure 1) has revealed a modern 'RNA world' in eukaryotes to mirror the ancestral RNA world at the time of archea and bacteria divergence [21, 22].

In a similar way, we contend in this review that proteins are crucial molecules to study directly when addressing the scientific questions typically investigated by evolutionary biologists for a variety of reasons. First they normally represent the functional units ("the agents") at the molecular level that are directly responsible for a phenotype seen on the

143 macroscopic scale. Secondly, most environmental factors, that are not direct mutagens,
144 act firstly on proteins and only secondarily on the genome. Thirdly, proteins are
145 responsible for determining transcriptional competency of significant portions of a given
146 genome by controlling eu/heterochromatin modulation and thus access of the
147 transcription machinery (Figure 1). This becomes increasingly more important given that
148 work over the last two decades has revealed that genes can produce a substantial variety
149 of proteins with fundamentally differed functions through both post-transcriptional
150 processing and post-translational modifications [23]. The idea of post-translational
151 marking as a driver of protein-protein interaction and of this marking as a 'protein
152 interaction code' has been proposed based on a range of examples in yeast [24]. Thirdly,
153 the state and function of a specific protein is influenced by both the proteome as a whole
154 and the influence of the prevailing environment. Fourthly, proteins are the agents for
155 epigenetic marking of genomes through histone modification and DNA methylation [25-
156 27] (Figure 1), and hence the proteome has the potential for trans-generational influence
157 both as the end product of the genome and also as epigenome modifying agents. Fifthly,
158 and as pointed out by several recent reviews [28-31], advances in protein biochemistry
159 now allow the assessment of the abundance, location, modification and function of
160 proteins, from isolated single proteins to complex mixtures from whole tissue extracts
161 [32, 33]. Proteomic technologies therefore provide capacity for the study of biological
162 mixtures of proteins and the use of a range of separation techniques from gel
163 electrophoresis to liquid chromatography coupled to different types of mass spectrometry
164 to analyze, quantify and identify differences in the proteome [28, 34, 35]. Separations in
165 gel, in liquid media and on solid surfaces provide physical arrays of proteins for

166 comparison of differences samples. Peptide mass spectrometry allows the pattern
167 matching based identification of peptides to track them back to all the specific genome
168 loci that encode them [36]. Increasing, the techniques of proteomics also allow the
169 assessment of smaller samples, faster and more accurately, and population level analysis
170 of individuals is already a reality [37]. Mass spectrometry can identify and quantify not
171 only the abundance of proteins, but also many modifications to proteins induced by
172 ecological stimuli and through genetic susceptibility to modification [32]. Changes in the
173 partnering and strength of protein-protein interactions will also soon be able to be
174 predicted, detected and quantified [38].
175
176 If we consequently consider protein synthesis and maturation as quantifiable mechanisms
177 to produce natural variation in gene products that selection can act upon and that can be
178 inherited, we need to integrate the analysis of proteins and their functions into a larger
179 framework for evolutionary biology using readily available molecular systems biology
180 approaches (see Table 1 for some examples). In our view this generates very exciting
181 opportunities for future research. There are a number of textbook examples where
182 proteins are involved in dynamics of evolutionary importance. For example trans-
183 generational movement of antibodies initiating immune competence from mother to child
184 [39], or Darwinian evolution of prion proteins in response to host competition [40]. But
185 what we will discuss in more detail here is that proteins more generally have an
186 underlying and fundamental role in phenotypic variation that is acted upon by evolution
187 across all species and considerable work is required to build and understand the
188 mechanisms that underpin it.

To build a strong framework of connection between proteomics and evolutionary theory we start with a general and understandable introduction into protein dynamics and protein networks and then consider the proteomic technologies and their possibilities and limitations for use by evolutionary biologists (Table1). To exemplify some of the theoretical considerations we then provide our own experience in attempting to bridge these worlds through work on reproductive processes and pathogen susceptibility in two model insects – the honeybee *Apis mellifera* and the leaf cutter ant *Atta colombica*. In the later paragraphs we point out the major questions and challenges that remain to be studied. Our main aim in writing this paper is to encourage biologists of all fields to consider recent advances in understanding of proteins and the broader field of proteomics for their future work on evolutionary questions.

## Proteins are more diverse than the genes that encode them

The primary structure of a protein is a string with each position occupied by one of 23 different amino acids. Each is decoded from defined sets of triplet bases containing one, two or three of the four bases of DNA. As macromolecules, proteins have a huge range of size, complexity and a much wider range of physical properties than the DNA that encodes them. Individual proteins can interact with substrates and products in catalysis, or with structural partner proteins through surface residue interactions. Different proteins can operate in aqueous or highly hydrophobic environments and in a temperature range of

over 120$^{o}$ C, from less than -20$^{o}$ C in arctic and alpine species to more than 100$^{o}$ C in species within volcanic vents [41, 42]. It is widely presumed by evolutionary biologists that proteins have little plasticity of function, no memory and very limited ability to adapt to changing molecular environments. However, protein biochemistry shows this is not really the case [43]. The functioning of proteins can generate alterations of their native conformation that can impact their future function. Proteins can also have their kinetic characteristics altered by external stimuli, thereby changing the way they interact with substrates and products. This can occur through covalent modification of amino acids by processes including phosphorylation, acetylation or glycosylation [24], or reversibly by allosteric activation or inhibition through the binding to proteins of small molecules other than substrates or products [44].

## Proteins are organized in networks and groups of networks

Complex phenotypic traits are not caused by a single protein but are usually the result of an organized co-occurrence of a set of proteins (Figure 1). Collectively such sets of proteins constitute functional networks with further aspects of flexibility. A protein network can have nodes and hubs, multiple inputs and outputs, and can operate at different states of flux depending on the starting conditions. One new enzyme added to the network can even reorganize the node-wiring diagram, bridge between pathways and thus impact the network flux more dramatically than its individual role as a single protein. Horizontal gene transfer is a means of introduction of new protein players into networks from other organisms in an ecosystem (Figure 1) [4, 6], and a raft of insights into the post-translational evolutionary processes regulating protein abundance have

recently been reviewed [45]. The complexity of such networks is further increased because proteins often have multiple functions and can therefore be found in multiple proteomic networks, providing links and possibilities for interactions between them.

The presence of complex proteomic networks underlying macroscopically observed phenomena has several potential consequences: The different subtasks of networks are at risk of interfering with each other, both at the level of the proteins themselves and at the level of the metabolic pools that interconnect them. For example, proteases involved in host-parasite interactions can damage other proteins both in the host and parasite [46, 47], while enzymes producing free radical species can cause oxidation of other proteins through their reactive products [48]. Consequently we might expect evolutionary trade-offs to be apparent within protein networks and the ways network components can evolve is likely to be more restricted for those components that fulfill core functions in each network. This means the evolutionary clock will tick at different speeds for the corresponding genes coding for these proteins within a given network, depending on the biological activity of proteins or subgroups of proteins and their degree of connectivity to functional networks. Interestingly the high frequency of selection of functional modifications in metabolic enzyme loci associated with the TCA cycle and glycolysis and their secondary roles in cells has already been highlighted [49].

## The potential for selection acting on proteomic networks

A problem we face is how best to describe the functioning as well as the evolvability of protein networks in a way that can illustrate the action of selective processes. Mathematicians use matrices to define the condition of an N-dimensional structure. By the same analogy we can describe a proteomic network as an N-dimensional matrix where each individual protein or pathway represents an additional dimension (vector) within the matrix and values within the matrix define individual network characteristics such as metabolic flux. The matrix is initially defined by genotype, but after its formation, it's phenotypic expression can be represented in several different ways and powerfully influenced the interaction of its components:

Firstly, a protein matrix can be seen as information processing machinery. Input variables (an input vector that consists of internal or environmental stimuli) are entered into the matrix and ultimately produce a response variable with a phenotypic expression on the cellular or individual level. The stimulus could be a protein or a metabolite (or several or a combination) that enters a biochemical network and finally generates a product that has biological activity; for example, a neurotransmitter to influence development, or a metabolic substrate or antioxidant to nurture or defend a cell or tissue.

Secondly, a protein matrix can be seen as an information storage device. In such a case the input variables can be seen as a matrix as well and the interaction of matrices produce a new but altered one. This change, dilution or modification will affect the possible output from what has been mentioned above. These modified functions can be "stored" within the matrix in a short time frame. For example proteins within networks could be

281 degraded by proteases or post-translationally modified or allosterically activated or
282 inhibited, and thereby biochemical pathways blocked or metabolic flows diverged that
283 would normally have been expected as physiological responses. Multiple modifications at
284 closely arranged sites can ultimately represent logic functions, building logic gates that
285 can impact the protein network function [32].
286

287 ## Evolutionary dynamics of protein network matrices

288

289 Our discussion about protein networks and selection acting on them as presented in the
290 last chapters also allows us to propose elements of a continuum, starting from instant
291 physiological effects influencing a trait under selection right through to the
292 transgenerational dynamics that evolutionary ecologists typically investigate in their
293 macroscopic trait studies. However, for biochemical credibility, such a pathway needs
294 definable and measureable components that underpin mechanisms. In our view, many of
295 the elements needed are already known and simply need to be placed within a connected
296 framework of evolutionary biology as outlined below.
297

298 A protein matrix is able to respond to its environment rapidly in seconds to minutes, such
299 as for example the detection of the presence of a pathogen. This produces a matrix output
300 that selection can operate on in minutes; for example, an oxygen radical is detoxified by
301 antioxidant machinery, or alternatively it damages a protein altering the flux through a
302 metabolic pathway, or a protein in phosphorylated changing its function or interaction
303 partners. The manipulation of the protein matrix then moves it into a new equilibrium or

304 phenotype with altered attributes over a period of hours; for example changes in
305 metabolic pools which can then feedback or feed forward on gene expression providing
306 new components to the matrix. Prolonged changes in gene expression will alter the
307 proteome, but can also influence both the mRNA pool and the smallRNA and miRNA
308 pools. Through mechanisms such as RNA-directed DNA methylation or alterations of
309 histone modification in regions of active transcription, over days to months, these altered
310 gene expression changes can modify epigenetic marks on the DNA leading to altered
311 epigenetic states of specific cells. Methylation patterning and corresponding changes in
312 histone binding and modification can in some cases be heritable, providing a
313 transgenerational flow of protein matrix attributes. There is also a further stage of this
314 process, as there is now clear evidence that methylated regions of DNA have higher
315 mutation rates [50] due to spontaneous deamination of methyl-cytosine and subsequent
316 mismatching during DNA duplication to replace a CG couple with an AT couple,
317 yielding a single nucleotide polymorphism. Hence changes to proteome matrices that can
318 feedback to gene expression have the potential to influence the evolution of genes that
319 influence them in a targeted fashion. Examples of the role of post-translational
320 modification processes [51] and of DNA methylation [52, 53] in evolutionary processes
321 have been reviewed.
322
323

## 324 Evolutionary biology from a protein perspective:

325 The ideas and hypotheses raised so far in this paper originate from the literature that has
326 combined protein biochemistry with evolutionary ecology. Such empirical cross

disciplinary work conducted over the last decade provides a number of interesting examples of how questions derived from evolutionary biology have been approached using a combination of discovery based proteomic analyses and the typical, hypotheses driven approaches used in evolutionary ecology (see Table 1 for more details and examples). For example, proteins have been shown to be phenotypically plastic, if they lack a defined stable secondary or tertiary structure [54]. Regions within proteins that show such plasticity play a key role in protein-protein interaction networks, which might provide functional advantages. Because of their higher capacity to rewire with other protein partners, they also seem to evolve rapidly [54]. Proteomics has also been used in phylogenetic studies to understand the evolution of traits of interest, for example key physiological traits such as protein translation throughout the tree of life [55]. Proteomic approaches revealed how individual proteins such as Hsp90 can impact developmental networks and thereby influence morphological phenotypes and their evolution [56]. An interesting example of protein based inheritance has recently been reported in the honeybee *Apis mellifera*, where the egg yolk protein vitellogenin can bind to a bacterial bee pathogen and is used as a carrier of microbial fragments to eggs, resulting in immune priming of offspring [57].

Sexual reproduction typically involves behavioral interactions by two or more individuals on the macroscopic level but numerous key processes are occurring at the level of proteins [58]. These interactions are typically characterized by the presence of strong selective forces acting on individual fitness and by fast evolutionary change. Males have to provide females with sufficient numbers of viable sperm but often compete against

each other within the females' sexual tract for access to eggs. Consequently, evolutionary ecologists had a long standing interest to untangle those traits that support the cooperative aspects of sexual reproduction from those responsible for the conflicts over paternity and understand their impact and male and female reproductive success. Proteomics has already been used in a wide range of different organisms to identify male reproductive proteins, which provided key insights into the make up of ejaculates [59-70] and triggered studies to unravel their effects on females and reproductive success [69, 71-75]. The identification of these proteins and a more detailed understanding about their function has enabled comparative studies to investigate the evolutionary history and phylogenetic relationships for some of these proteins [76-78].

In the next section we will further elaborate on such reproductive proteomics by summarizing our own findings, which offers us an opportunity to discuss our strategy and intent behind our scientific progress. Furthermore, commentary on the timeline of discovery and our unpublished and pilot data allows us to provide a more general overview of the benefits of proteomics for evolutionary research. Proteomics was without doubt the key tool we used over the last 10 years to gain major insights into highly complex biological processes and to begin to hone in on proximate mechanisms that underlie reproductive traits such as high quality sperm and long-term sperm storage. We started with the identification of lists of proteins in samples of interest. While this is often flagged as a limitation of proteomics, because it cannot provide causal relationships, it provided our protein landscape for the following years. As we illustrate below in detail, these parts lists guided consequently phenotypic studies providing not only answers to

373 our initial questions but a substantially broader and systemic insight into the molecular
374 function of social insect reproduction.
375

## 376 An example from our own research on insect reproduction

377

378 We studied sexual reproduction in Hymenopteran social insects, being the eusocial ants,
379 bees and wasps. Their societies are characterized by the presence of a single or very few
380 reproductive animals, typically referred to as queens, and a non-reproducing cast known
381 as workers [79-82]. Workers benefit from helping if they are related to the helped
382 individual, known as inclusive fitness benefits [83, 84]. Because helping incentive
383 increases with helper relatedness, social insects queens perform only a single round of
384 mate choice and sperm acquisition and never replenish sperm once they have started to
385 lay eggs. This has resulted in the evolution of spectacular reproductive adaptations in
386 species with large and long lived societies, where males produce exceptionally large
387 ejaculates of high quality [85], and queens store them for decades to sire millions of
388 offspring [86, 87]. Although these reproductive traits must have been key during social
389 evolution and will have contributed towards the remarkable ecological success of these
390 animals [88], the proximate mechanisms to achieve such high levels of fertility or their
391 evolution had remained completely unknown.
392 During mating males transfer ejaculates to females that consist of sperm and glandular
393 secretions known as seminal fluid, which has increasingly been recognized to contain key
394 molecules determining male and female reproductive success [89-92]. Seminal fluid can
395 also produce mating plugs or mating signs [81, 93-96], which influence the mating

behavior of bumblebee queens [97, 98], providing the first evidence that seminal fluid components are important in social insects.

With this background we started our collaborative work together by a discovery inspired proteomic characterization of seminal fluid in honeybees [99], revealing a first insight and basic parts map of its molecular landscape. This list of proteins proved to be of significant scientific value because functional analyses predicted biological functions and generated a wealth of hypothesis and ideas that consequently guided our further experimental work. We found that seminal fluid proteins consisted of three major groups of proteins, that (1) keep sperm alive (2) defend sperm and queens from pathogens and (3) are molecular agents of sexual conflict. Follow up experimental work confirmed that seminal fluid is indeed exceptionally potent in keeping sperm alive and that proteins are the key molecules responsible for this effect [100, 101]. The detection of antimicrobial proteins in the seminal fluid triggered a search for pathogens in honeybee ejaculates, and resulted in the identification of two widespread bee pathogens in ejaculates, *Nosema apis* and *Nosema ceranae* [102, 103]. Our finding of antifungal proteins in seminal fluid implied that males may be able to combat these infections in their ejaculates. Experimental follow up showed that seminal fluid is indeed remarkably efficient in killing *N. apis* spores (Peng et al., submitted). Moreover, our data revealed that the pathogen is killed in at least two distinctly different ways, implying that there is redundancy in the defense system of seminal fluid as well as some specificity, because the biologically active molecules show no antimicrobial activity against a series of non-pathogenic microbes. Both redundancy and specificity of honeybee antimicrobial proteins were novel findings that had not been reported previously for insects. Thirdly, we found

molecules that we predicted to be involved in sexual conflict. The presence of ejaculates from multiple males within a female's sexual tract can result in postcopulatory sexual selection, operating either as sperm competition [104] or cryptic female choice [105]. We hypothesized that the battlegrounds of these events are extracellular spaces dominated by secreted proteomes and the effective role and variability of these protein sets would be defining paternity success. We indeed found that seminal fluid proteins of polyandrous honeybees and leaf cutter ants are capable to kill sperm of rival males, known as sperm incapacitation [106].

This shows how our initial identification of proteins in the seminal fluid of honeybees generated a number of predictions about function, which were accurate because we were able to confirm the expected phenotypes through follow up experiments. The proteins identified were confirmed to be the biologically active molecules, and our functional analyses have already provided subsets of target proteins for further study [107]. Finally, our work indicated that seminal fluid proteins or protein networks interact with other proteomes, such as those of rival ejaculates, the queen or parasites, encouraging us to also identify these additional proteomes.

The honeybee sperm proteome revealed the presence of a very distinct subset of proteins [108], many of them being related to energy metabolism. Their high abundance in honeybee sperm implied that the survival of high quality sperm is closely associated with energy production. When we quantified the effect of some of these proteins on sperm metabolism we were able to confirm that these proteins are biologically active and are key for sperm survival. A second group of abundant proteins we detected in honeybee sperm are related to transcriptional or translational activities, which was surprising given

442 that sperm is often believed to be translationally and transcriptionally silent. Providing
443 sperm with radiolabeled amino acids confirmed that sperm indeed produce proteins at a
444 low rate [109].
445
446 As well as producing high quality sperm, social insects are also the 'world record
447 holders' for storing sperm [81, 86, 87]. To achieve this, queens provide sperm with
448 spermathecal fluid, a glandular secretion continuously added to sperm during storage.
449 The proteomic profiles of spermathecal fluid was distinct from that of seminal fluid [110]
450 indicating that sperm are able to survive in two very different biochemical environments.
451 The seminal fluid proteome forms a loosely connected network of proteins, consistent
452 with the expectation that these proteins are responsible for more individual tasks such as
453 keeping sperm alive, killing rival sperm and parasites or manipulating female physiology
454 [100, 101, 106, 110, 111]. The spermathecal proteome on the other hand keeps sperm
455 alive for years, and its high connectivity seems to provide a biochemical environment that
456 has been selected for maximized sperm survival [110]. We therefore expected the
457 proteome of sperm to adapt to these changes in their biochemical environment. We have
458 now confirmed this experimentally [107], and key enzymes with changed abundances
459 were as expected related to energy production. Consequently, we were able to provide
460 important molecular insights into the secrets of long-term sperm storage, which were
461 facilitated by the presence of a relatively small number of enzymes that maximize ATP
462 production and minimize oxidative stress.
463
464 Apart from the proteome differences between stored and ejaculated sperm we have also

found other proteomes to be remarkably plastic, for example between seminal fluid samples from males of different bee lineages [111] or between seminal fluid of infected and uninfected males (Grassl *et al* in preparation). These studies reveal that the proteomes of ejaculates provide a useful model system to understand proteomic networks and their functioning in an evolutionary framework, because the proteomes investigated are substantially smaller than those of entire cells, organs or organisms, and the strong selective forces or ecological stimuli that impact them result in fast changes in these secreted proteomes and functional changes in proteomic networks that can be measured in vitro. These studies merely mark a starting point and demonstrate the feasibility of fruitful collaborations between proteomics and evolutionary biology researchers – what we plan now are studies that quantify the phenotypic variation of proteomes and to quantify their hypothesized heritability. Furthermore, the identification of individual proteins or protein networks of interest now allow us to also conduct phylogenetic comparisons to understand the evolutionary history of proteins and the way underlying proteomic networks co-evolved with traits of interest, similar to what has been done in other species [77].

Our work on evolutionary proteomics revealed that proteomic investigations tend to take substantially longer to conduct and publish, as they generate larger datasets that require substantially more time to analyze than classical studies typically conducted in behavioral ecology. However, such detailed data mining resulted in the development of predictions and hypotheses and the accuracy of these predictions was found, in time, to be highly consistent with the mechanisms uncovered. Consequently, proteomics can also be used as a highly efficient and accurate tool for the development and formulation of testable

hypotheses.

## The future of evolutionary proteomics

Developing a framework for how proteins play an important role for evolutionary processes is critical for engagement of researchers with expertise in each area. Biochemical approaches to identify and study proteins and their abundance and functionality have been rapidly developing over the last decade and are now much more easily accessible for a wider range of scientists including evolutionary biologists [30, 31, 112]. Academic institutions as well as private companies offer collaborations and services to analyze protein samples of interest. With this technical feasibility, proteomes can be studied in an evolutionary framework that does not differ in any major way from any other phenotypic trait of interest studied by evolutionary biologists over the last 150 years. What is still needed are larger scale experimental studies quantifying natural variation in protein profiles within and between individuals of the same and different populations in order to understand how much of the theoretically achievable variation in proteomes is actually realized and selected for. Furthermore we need a better understanding of how the interactions between environmental factors and an individual's proteome operate, both in the short as well as in the long term. Such studies will provide crucial insights into an individual's opportunity to actively respond and adapt to changing environments and can test for the degree of heritability of modified proteomic networks characteristics. Finally, studies are needed that quantify the overall effect of proteins on evolutionary processes in order to understand whether they are only important for a

subset of processes such as sperm competition and immunity or whether they are in fact additional drivers of evolution more generally.

## Summary

The traditional split between biological sciences focused on either proximate or ultimate questions is starting to diminish. One of the reasons for this is that molecular insights into how life works have exponentially grown over the last decade, due in significant measure to spectacular technological breakthroughs that now allow the study of molecular dynamics in cells and entire organisms and even their extrapolation to wider habitats and ecosystems. Genomics has provided evolutionary biologists with new and exciting opportunities to understand and investigate evolutionary concepts. The rapidly evolving field of proteomics now needs researchers with evolutionary questions to link protein networks and their functioning to complex organismic characteristics.

## Acknowledgements

We were supported by the Australian Research Council (ARC) through a Queen Elizabeth II Fellowship and a Future Fellowship to BB, a Future Fellowship to AHM, an ARC Linkage Project to BB and the facilities of the ARC Centre of Excellence in Plant Energy Biology. We thank Barbara Baer-Imhoof, Koos Boomsma, Julia Grassl and Susanne den Boer for continuous discussions and feedback, which was important to develop and write this review.

535

536

537

538 ## References

539 [1] Darwin C. The origin of species by means of natural selection. London: John Murray;
540 1859.
541 [2] Watson JD, Crick FHC. Molecular structure of nucelic acids - a structure for
542 deoxyribose nucleic acid. Nature. 1953;171:737-8.
543 [3] Crick F. Central dogma of molecular biology. Nature. 1970;227:561-&.
544 [4] Woese CR. On the evolution of cells. Proc Natl Acad Sci U S A. 2002;99:8742-7.
545 [5] Woese CR. Translation: In retrospect and prospect. RNA. 2001;7:1055-67.
546 [6] Koonin EV. Carl Woese's vision of cellular evolution and the domains of life. RNA
547 Biol. 2014;11:197-204.
548 [7] Gladieux P, Ropars J, Badouin H, Branca A, Aguileta G, De Vienne DM, et al.
549 Fungal evolutionary genomics provides insight into the mechanisms of adaptive
550 divergence in eukaryotes. Mol Ecol. 2014;23:753-73.
551 [8] Oliver RP, Friesen TL, Faris JD, Solomon PS. *Stagonospora nodorum*: From
552 pathology to genomics and host resistance. Annu Rev Phytopathol. 2012;50:23-43.
553 [9] Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW, Hohenlohe PA, et al.
554 Genomics and the origin of species. Nat Rev Genet. 2014;15:176-92.
555 [10] Varki A, Nelson DL. Genomic comparisons of humans and chimpanzees. Annu Rev
556 Anthropol. 2007;36:191-209.
557 [11] Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B, et al.
558 Great ape genetic diversity and population history. Nature. 2013;499:471-5.
559 [12] Przeworski M, Coop G, Wall JD. The signature of positive selection on standing
560 genetic variation. Evolution. 2005;59:2312-23.
561 [13] Vitti JJ, Grossman SR, Sabeti PC. Detecting natural selection in genomic data. Annu
562 Rev Genet. 2013;47:97-120.
563 [14] Meyer RS, Purugganan MD. Evolution of crop species: genetics of domestication
564 and diversification. Nat Rev Genet. 2013;14:840-52.
565 [15] Ebert D. Experimental evolution of parasites. Science. 1998;282:1432-5.
566 [16] Firman RC, Garcia-Gonzalez F, Thyer E, Wheeler S, Yamin Z, Yuan M, et al.
567 Evolutionary change in testes tissue composition among experimental populations of
568 house mice. Evolution. 2015;69:848-55.
569 [17] Firman RC, Simmons LW. Experimental evolution of sperm competitiveness in a
570 mammal. BMC Evol Biol Bmc Evolutionary Biology. 2011;11.
571 [18] Snook RR, Robertson A, Crudgington HS, Ritchie MG. Experimental manipulation
572 of sexual selection and the evolution of courtship song in *Drosophila pseudoobscura*.
573 Behav Genet. 2005;35:245-55.
574 [19] Snook RR. Sexual selection: Conflict, kindness and chicanery. Curr Biol.
575 2001;11:R337-R41.

576 [20] Krishnan J, Mishra RK. Emerging trends of long non-coding RNAs in gene
577 activation. Febs J. 2014;281:34-45.
578 [21] Groen JN, Capraro D, Morris KV. The emerging role of pseudogene expressed non-
579 coding RNAs in cellular functions. Int J Biochem Cell Biol. 2014;54:350-5.
580 [22] Chen YN, Xiong XD. Long noncoding RNA and epigenetic regulation. Prog
581 Biochem Biophys. 2014;41:723-30.
582 [23] Black DL. Mechanisms of alternative pre-messenger RNA splicing. Annu Rev
583 Biochem. 2003;72:291-336.
584 [24] Winter DL, Erce MA, Wilkins MR. A web of possibilities: network-based discovery
585 of protein interaction codes. Journal of Proteome Research. 2014;13:5333-8.
586 [25] Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome
587 integrates intrinsic and environmental signals. Nature Genetics. 2003;33:245-54.
588 [26] Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and
589 beyond. Nat Rev Genet. 2012;13:484-92.
590 [27] Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C, Fouse SD, et al.
591 Conserved role of intragenic DNA methylation in regulating alternative promoters.
592 Nature. 2010;466:253-U131.
593 [28] Lesur A, Domon B. Advances in high-resolution accurate mass spectrometry
594 application to targeted proteomics. Proteomics. 2015;15:880-90.
595 [29] Boersema PJ, Kahraman A, Picotti P. Proteomics beyond large-scale protein
596 expression analysis. Curr Opin Biotechnol 2015;34:162-70.
597 [30] Karr TL. Application of proteomics to ecology and population biology. Heredity.
598 2008;100:200-6.
599 [31] Diz AP, Martinez-Fernandez M, Rolan-Alvarez E. Proteomics in evolutionary
600 ecology: linking the genotype with the phenotype. Mol Ecol. 2012;21:1060-80.
601 [32] Maarten Altelaar AF, Munoz J, Heck AJR. Next-generation proteomics: towards an
602 integrative view of proteome dynamics. Nature Rev Gen 2013;14:35-48.
603 [33] LaCava J, Molloy KR, Taylor MS, Domanski M, Chait BT, Rout MP. Affinity
604 proteomics to study endogenous protein complexes: pointers, pitfalls, preferences and
605 perspectives. Biotechniques. 2015;58:103-19.
606 [34] Christoforou AL, Lilley KS. Isobaric tagging approaches in quantitative proteomics:
607 the ups and downs. Anal Bioanal Chem. 2012;404:1029-37.
608 [35] Kline KG, Sussman MR. Protein quantitation using isotope-assisted mass
609 spectrometry. Annu Rev Biophys. 2010;39:291-308.
610 [36] Breker M, Schuldiner M. The emergence of proteome-wide technologies: systematic
611 analysis of proteins comes of age. Rev Mol Cell Biol. 2014;7:453-64.
612 [37] Nedelkov D, Kiernan UA, Niederkofler EE, Tubbs KA, Nelson RW. Population
613 proteomics: the concept, attributes, and potential for cancer biomarker research. Mol Cell
614 Proteomics. 2006;5:1811-8.
615 [38] Kastritis PL, Bonvin AM. On the binding affinity of macromolecular interactions:
616 daring to ask why proteins interact. J R Soc Interface. 2012;10:20120835.
617 [39] Carlier Y, Truyens C. Influence of maternal infection on offspring resistance
618 towards parasites. Parasitol Today. 1995;11:94-9.
619 [40] Li JL, Browning S, Mahal SP, Oelschlegel AM, Weissmann C. Darwinian evolution
620 of prions in cell culture. Science. 2010;327:869-72.

621 [41] De Maayer P, Anderson D, Cary C, Cowan DA. Some like it cold: understanding the
622 survival strategies of psychrophiles. EMBO Rep. 2014;15:508-17.
623 [42] Minic Z, Serre V, Herve G. Adaptation of organisms to extreme conditions of deep-
624 sea hydrothermal vents. C R Biol. 2006;329:527-40.
625 [43] Vuillon L, Lesieur C. From local to global changes in proteins: a network view. Curr
626 Opin Struct Biol. 2015;31:1-8.
627 [44] Kokkinidis M, Glykos NM, Fadouloglou VE. Protein flexibility and enzymatic
628 catalysis. Adv Protein Chem Struct Biol. 2012;87:181-218.
629 [45] Vogel C, Marcotte EM. Insights into the regulation of protein abundance from
630 proteomic and transcriptomic analyses. Nat Rev Genet. 2012;13:227-32.
631 [46] Amano A, Takeuchi H, Furuta N. Outer membrane vesicles function as offensive
632 weapons in host-parasite interactions. Microbes Infect 2010;12:791-8.
633 [47] Pejler G, Abrink M, Ringvall M, Wernersson S. Mast cell proteases. Adv Immunol.
634 2007;95:167-255.
635 [48] Gupta RK, Patel AK, Shah N, Choudhary AK, Jha UK, Yadav UC, et al. Oxidative
636 stress and antioxidants in disease and cancer: A review. Asian Pac J Cancer Prev.
637 2014;15:4405-9.
638 [49] Marden JH. Nature's inordinate fondness for metabolic enzymes: why metabolic
639 enzyme loci are so frequently targets of selection. Mol Ecol. 2013;22:5743-64.
640 [50] Cooper DN, Krawczak M. Cytosine methylation and the fate of CpG dinucleotides
641 in vertebrate genomes. Hum Genet. 1989;83:181-8.
642 [51] Beltrao P, Bork P, Krogan NJ, van Noort V. Evolution and functional cross-talk of
643 protein post-translational modifications. Mol Syst Biol. 2013;9.
644 [52] Carbone L, Harris RA, Vessere GM, Mootnick AR, Humphray S, Rogers J, et al.
645 Evolutionary breakpoints in the gibbon suggest association between Cytosine methylation
646 and karyotype evolution. PLoS Genet. 2009;5.
647 [53] Burggren WW, Crews D. Epigenetics in comparative biology: Why we should pay
648 attention. Integr Comp Biol. 2014;54:7-20.
649 [54] Mosca R, Pache RA, Aloy P. The role of structural disorder in the rewiring of
650 protein interactions through evolution. Mol Cell Proteomics. 2012;11.
651 [55] Mahlab S, Linial M. Speed controls in translating secretory proteins in Eukaryotes -
652 an evolutionary perspective. PLoS Comput Biol. 2014;10.
653 [56] Rutherford S, Knapp JR, Csermely P. Hsp90 and developmental networks.
654 AdvExpMedBiol. 2007;594:190-7.
655 [57] Salmela H, Amdam GV, Freitak D. Transfer of immunity from mother to offspring
656 is mediated via egg-yolk protein vitellogenin. PLOS PATHOGENS. 2015;11:e1005015.
657 [58] Birkhead TR, Hosken DJ, Pitnick S. Sperm Biology: An evolutionary perspective.
658 Amsterdam: Elsevier; 2009.
659 [59] Solomon B, Duncan MW. Proteomics of seminal fluid. In: Thongboonkerd V, editor.
660 Proteomics of human body fluids. Totowa, NJ: Humana Press Inc.; 2007. p. 467-93.
661 [60] Takemori N, Yamamoto MT. Proteome mapping of the *Drosophila melanogaster*
662 male reproductive system. Proteomics. 2009;9:2484-93.
663 [61] Oliva R, de Mateo S, Estanyol JM. Sperm cell proteomics. Proteomics. 2009;9:1004-
664 17.
665 [62] Miller I, Friedlein A, Tsangaris G, Maris A, Fountoulakis M, Gemeiner M. The
666 serum proteome of *Equus caballus*. Proteomics. 2004;4:3227-34.

667 [63] Kelly VC, Kuy S, Palmer DJ, Xu ZZ, Davis SR, Cooper GJ. Characterization of
668 bovine seminal plasma by proteomics. Proteomics. 2006;6:5826-33.
669 [64] Walker MJ, Rylett CM, Keen JN, Audsley N, Sajid M, Shirras AD, et al. Proteomic
670 identification of *Drosophila melanogaster* male accessory gland proteins, including a
671 pro-cathepsin and a soluble gamma-glutamyl transpeptidase. Proteome Sci. 2006;4:1-10.
672 [65] Tomar AK, Sooch BS, Singh S, Yadav S. Differential proteomics of human seminal
673 plasma: A potential target for searching male infertility marker proteins. Proteom Clin
674 ApplProteomics Clinical Applications. 2012;6:147-51.
675 [66] Fung KYC, Glode LM, Green S, Duncan MW. A comprehensive characterization of
676 the peptide and protein constituents of human seminal fluid. Prostate. 2004;61:171-81.
677 [67] Findlay GD, Yi XH, MacCoss MJ, Swanson WJ. Proteomics reveals novel
678 *Drosophila* seminal fluid proteins transferred at mating. PLOS Biology. 2008;6:1417-26.
679 [68] Sirot LK, Hardstone MC, Helinski MEH, Ribeiro JMC, Kimura M,
680 Deewatthanawong P, et al. Towards a semen proteome of the dengue vector mosquito:
681 Protein identification and potential functions. Plos Neglect Trop Dis. 2011;5.
682 [69] Georgiou SA, Sostaric E, Wong CH, Snijders APL, Wright PC, Moore HD, et al.
683 Gametes Alter the Oviductal Secretory Proteome Mol Cell Proteomics. 2005;4:1785-96.
684 [70] Wasbrough ER, Dorus S, Hester S, Howard-Murkin J, Lilley K, Wilkin E, et al. The
685 *Drosophila melanogaster* sperm proteome-II (DmSP-II). J Proteomics. 2010;73:2171-85.
686 [71] Wigby S, Sirot LK, Linklater JR, Buehner N, Calboli FCF, Bretman A, et al.
687 Seminal fluid protein allocation and male reproductive success. Curr Biol. 2009;19:751-
688 7.
689 [72] Kubli E. Sex-peptides: Seminal peptides of the *Drosophila* male. CMLS Cell Mol
690 Life Sci. 2003;60:1689-704.
691 [73] Yapici N, Kim YJ, Ribeiro C, Dickson BJ. A receptor that mediates the post-mating
692 switch in *Drosophila* reproductive behaviour. Nature. 2008;451:33-U1.
693 [74] Ram KR, Wolfner MF. Seminal influences: Drosophila Acps and the molecular
694 interplay between males and females during reproduction. Integr Comp Biol.
695 2007;47:427-45.
696 [75] Ram KR, Wolfner MF. Sustained post-mating response in *Drosophila melanogaster*
697 requires multiple seminal fluid proteins. PLoS Genet. 2007;3:2428-38.
698 [76] Dean MD, Clark NL, Findlay GD, Karn RC, Yi XH, Swanson WJ, et al. Proteomics
699 and comparative genomic investigations reveal heterogeneity in evolutionary rate of male
700 reproductive proteins in mice (*Mus domesticus*). Mol Biol Evol. 2009;26:1733-43.
701 [77] Ramm SA, McDonald L, Hurst JL, Beynon RJ, Stockley P. Comparative proteomics
702 reveals evidence for evolutionary diversification of rodent seminal fluid and its functional
703 significance in sperm competition. Mol Biol Evol. 2009;26:189-98.
704 [78] Marshall JL, Huestis DL, Garcia C, Hiromasa Y, Wheeler S, Noh S, et al.
705 Comparative proteomics uncovers the signature of natural selection acting on the
706 ejaculate proteomes of two cricket species isolated by postmating, prezygotic phenotypes.
707 Mol Biol Evol. 2011;28:423-35.
708 [79] Boomsma JJ, Baer B, Heinze J. The evolution of male traits in social insects. Annual
709 Review of Entomology. 2005;50:395-420.
710 [80] Baer B. Sexual selection in *Apis* bees. Apidologie. 2005;36:187-200.
711 [81] Baer B. The copulation biology of ants (Hymenoptera: Formicidae). Myrmecol
712 News. 2011;14:55-68.

713 [82] Baer B. Sexual Selection in Social Insects. In: Shuker D, Simmons LW, editors. The
714 Evolution of Insect Mating Systems. Oxford: Oxford University Press; 2014. p. 261-74.
715 [83] Hamilton WD. The genetical evolution of social behavior I & II. J Theor Biol.
716 1964;7:(I): 1-16; (II): 7-32.
717 [84] Trivers RL, Hare H. Hapoldiploidy and the evolution of the social insects. Science.
718 1976;191:249-63.
719 [85] Hunter FM, Birkhead TR. Sperm viability and sperm competition in insects. Curr
720 Biol. 2002;12:121-3.
721 [86] Weber NA. Gardening ants: The attines. Philadelphia: The American Philosophical
722 Society; 1972.
723 [87] Keller L, Genoud M. Extraordinary lifespans in ants: a test of evolutionary theories
724 of ageing. Nature. 1997;389:958-60.
725 [88] Baer B. Female Choice in Social Insects. In: Peretti A, Aisenberg A, editors. Cryptic
726 Female Choice in Arthropods: Patterns, Mechanisms and Prospects. Stuttgart: Springer;
727 2015. p. 461-77.
728 [89] Avila FW, Sirot LK, LaFlamme BA, Rubinstein CD, Wolfner MF. Insect seminal
729 fluid proteins: Identification and function. Annual Review of Entomology. 2011;56:21-
730 40.
731 [90] Chapman T. Seminal fluid-mediated fitness traits in *Drosophila*. Heredity.
732 2001;87:511-21.
733 [91] Chapman T. The soup in my fly: Evolution, form and function of seminal fluid
734 proteins. PLOS Biology. 2008;6:1379-82.
735 [92] Poiani A. Complexity of seminal fluid: a review. Behav Ecol Sociobiol.
736 2006;60:289-310.
737 [93] Polak M, Wolf Larry L, Starmer William T, Barker JSF. Function of the mating plug
738 in Drosophila hibisci Bock. Behav Ecol Sociobiol. 2001;49:196-205.
739 [94] Baer B, Maile R, Schmid-Hempel P, Morgan ED, Jones GR. Chemistry of a mating
740 plug in bumblebees. J Chem Ecol. 2000;26:1869-75.
741 [95] Mikheyev AS. Evidence for mating plugs in the fire ant *Solenopsis invicta*. Insectes
742 Sociaux. 2003;50:401-2.
743 [96] Mikheyev AS. Male accessory gland size and the evolutionary transition from single
744 to multiple mating in the fungus-gardening ants. Journal of Insect Science. 2004;4:1-5.
745 [97] Baer B, Morgan ED, Schmid-Hempel P. A non-specific fatty acid within the
746 bumblebee mating plug prevents females from remating. Proc Natl Acad Sci USA.
747 2001;98:3926-8.
748 [98] Sauter A, Brown MJF, Baer B, Schmid-Hempel P. Males of social insects can
749 prevent queens from multiple mating. Proc R Soc B. 2001;268:1449-54.
750 [99] Baer B, Heazlewood JL, Taylor NL, Eubel H, Millar AH. The seminal fluid
751 proteome of the honeybee *Apis mellifera*. Proteomics. 2009;9:2085-97.
752 [100] den Boer SPA, Boomsma JJ, Baer B. Honey bee males and queens use glandular
753 secretions to enhance sperm viability before and after storage. J Insect Physiol.
754 2009;55:538-43.
755 [101] King M, Eubel H, Millar AH, Baer B. Proteins within the seminal fluid are crucial
756 to keep sperm viable in the honeybee *Apis mellifera*. J Insect Physiol. 2011;57:409-14.
757 [102] Peng Y, Imhoof-Baer B, Millar AH, Baer B. Consequences of *Nosema apis*
758 infection for male honey bees and their fertility. Sci Rep. 2015;5:1-5.

[103] Roberts K, Evison S, Baer B, Hughes WOH. The cost of promiscuity: sexual transmission of Nosema microsporidian parasites in polyandrous honey bees. Sci Rep. 2015;5:1-7.
[104] Birkhead TR, Moller AP. Sperm competition and sexual selection. New York: Academic Press; 1998.
[105] Eberhard WG, Cordero C. Sexual conflict and female choice. Trends Ecol Evol. 2003;18:438-9.
[106] den Boer SPA, Baer B, Boomsma JJ. Seminal fluid mediates ejaculate competition in social insects. Science. 2010;327:1506-9.
[107] Poland V, Eubel H, King M, Solheim C, Millar AH, Baer B. Stored sperm differs from ejaculated sperm by proteome alterations associated with energy metabolism in the honeybee *Apis mellifera*. Mol Ecol. 2011;20:2643-54.
[108] Zareie R, Eubel H, Millar AH, Baer B. Long term survival of high quality sperm: Insights into the sperm proteome of the honeybee *Apis mellifera*. J Proteome Res. 2013;12:5180-8.
[109] Paynter E. Cellular energetics of long-term survival and storage in honeybee sperm. Perth: The University of Western Australia; 2015.
[110] Baer B, Eubel H, Taylor NL, O'Toole N, Millar AH. Insights into female sperm storage from the spermathecal fluid proteome of the honeybee *Apis mellifera*. Genome Biology. 2009;10:R 67.
[111] Baer B, Zareie R, Paynter E, Poland V, Millar AH. Seminal fluid proteins differ in abundance between genetic lineages of honeybees. J Proteomics. 2012;75:5646-53.
[112] Graves PR, Haystead TAJ. Molecular biologist's guide to proteomics. Microbiol Mol Biol Rev. 2002;66:39-63.
[113] Ram KR, Sirot LK, Wolfner MF. Predicted seminal astacin-like protease is required for processing of reproductive proteins in *Drosophila melanogaster*. Proc Natl Acad Sci U S A. 2006;103:18674-9.
[114] Wyatt TD. Proteins and peptides as pheromone signals and chemical signatures. Anim Behav. 2014;97:273-80.
[115] Georgiou AS, Snijders APL, Sostaric E, Aflatoonian R, Vazquez JL, Vazquez JM, et al. Modulation of the oviductal environment by gametes. J Proteome Res. 2007;6:4656-66.
[116] Wilburn DB, Bowen KE, Doty KA, Arumugam S, Lane AN, Feldhoff PW, et al. Structural insights into the evolution of a sexy protein: Novel topology and restricted backbone flexibility in a hypervariable pheromone from the Red-Legged Salamander, *Plethodon shermani*. PLoS One. 2014;9.

Table 1: A focus on proteins as drivers of ecological and evolutionary processes offer researchers from various fields novel opportunities to explain biological phenomena such as rapid evolutionary and adaptive changes, especially in cases where the traditional focus on DNA mutation followed by natural selection provide unsatisfying proximate explanations for observed phenotypes. Furthermore, molecular biologists focusing on the use of -omics approaches to understand variation in their datasets can use evolutionary and ecological explanations in hypothesis building for followup experiments..

| | Areas of Interest | Protein/Proteome Feature | Literature |
|---|---|---|---|
| **Evolution** (Proteome changes over generations) | Evolution of traits | Proteins modify genomic information flow through DNA modifications and structure | [50, 55, 56] |
| | Heritability | Proteins are all genome derived but can also act as non-genomic components of transgenerational inheritance | [25-27, 39, 57] |
| | Natural Selection | Proteins are the biologically active "agents" and proximate drivers of fitness | [40, 49, 100, 101, 108, 113] |
| | Natural Variation | Proteomes are substantially more diverse than the underlying genomes. | [23, 24] |
| | Speciation | Proteins can be drivers of reproductive isolation | [114] |
| **Ecology** (Proteome-Proteome and Proteome – Environment interactions | Adaptation | Proteins are drivers of ecological adaptation, and spread through introgression and horizontal gene transfer | [4, 5, 7, 43] |
| | Genotype x Environment Interactions | Protein networks are susceptible to environmental stimuli | [107] |

| | | | |
|---|---|---|---|
| | | and can transfer such information to influence the transciptome and the epigenome | |
| | Host-Parasite Interactions | Protein diversity can be a driver of virulence and immunity | [8, 40, 47] |
| | Phenotypic Plasticity | Protein networks and matrixes are highly variable and change due to genomic and ecological factors | [23, 49] [54, 115] |
| | Sexual Selection | Proteins can be molecular drivers of conflicts over paternity | [71-73, 116] |

806
807

808

# Figure Legends

## Figure 1

**The Central Dogma coupled to other regulatory steps and mechanisms of environment-dependent variation that influence the proteome.** The central blue box presents the primary flow of molecular information as found in all living organisms, through which DNA encodes for genes that are transcribed into RNA which in many cases are translated into proteins. The latter are principally responsible for the expression of a specific phenotype. Research over the past decades has now shown that this central protein building system is augmented by a range of more dynamic protein and proteome-modifying factors which are influenced by environmental factors. This presentation highlights the role of proteins and their variation as additional drivers of physiological processes and their evolution rather than simply as end points of molecular information flow.