Journal of Proteomics 135 (2016) 4–11

Contents lists available at ScienceDirect

Journal of Proteomics

journal homepage: www.elsevier.com/locate/jprot

Proteomics in evolutionary ecology

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

ARTICLE INFO

Article history:
Received 26 May 2015
Received in revised form 22 September 2015
Accepted 30 September 2015
Available online 17 October 2015

Keywords:
Evolution
Natural selection
Sexual selection
Peptide mass spectrometry
Protein diversity
Post-translational modification
Protein-protein interaction

ABSTRACT

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

1. 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] (Fig. 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

* Corresponding author.
E-mail address: boris.baer@uwa.edu.au (B. Baer).

http://dx.doi.org/10.1016/j.jprot.2015.09.031
1874-3919/

encoding for a whole new protein in bacteria and archea, Fig. 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 milieu and indeed into a protein network. The importance of horizontal gene transfer in fungal evolution [7] and even in very recent examples in pathogenicity of fungi is well documented [8].

The full genome sequences of thousands of species are now available allowing unprecedented base by base comparison of genes within and across families, genera and kingdoms and increasingly more sophisticated methodologies are also available to permanently or transiently manipulate gene sequence and expression to observed the effects. However, while this genomics has generated solid empirical evidence for evolutionary theory and provided detailed insights into evolutionary dynamics (e.g. [9]), a range of more fundamental questions still remained unresolved. For example, comparative genomics reveals that many genes often remain remarkably similar throughout evolutionary history, providing, for example, only preliminary answers to the question of why chimpanzees are chimpanzees and humans are humans based on DNA sequence alone [10,11]. It is widely acknowledged that regulation of expression of genomes is the key differentiator between mammals but it remains unclear, how differences in gene expression of an identical gene pool can generate the tremendous phenotypic variation observed such as for example between humans and chimpanzees [10,11]. Furthermore, dependence of molecular evolution of DNA on random mutations alone resulting in the eventual appearance of a gene with superior functionality [12–14] would relegate evolution to depend predominantly on chance events acted on by selective forces across generations. In our view while the focus on point mutation alone is 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] (Fig. 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

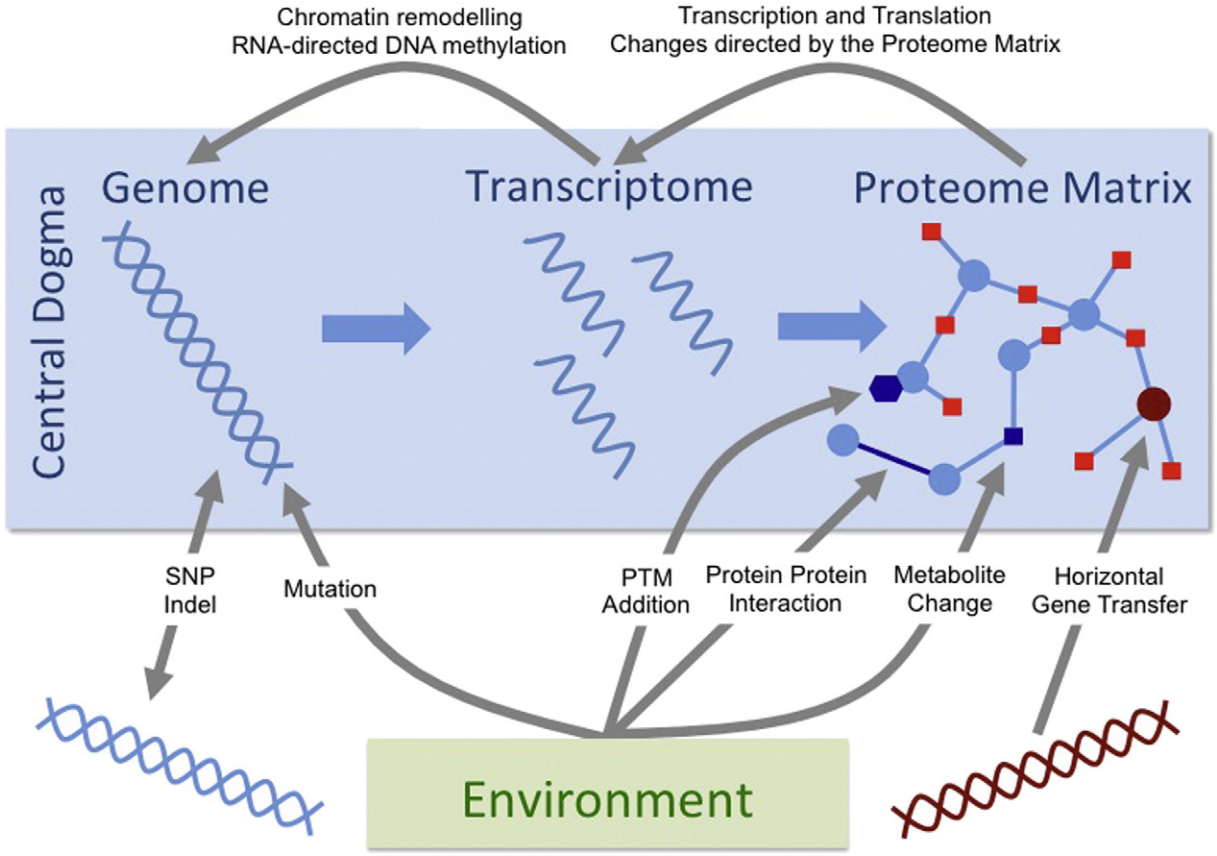

Fig. 1. The Central Dogma coupled to other regulatory steps and mechanisms of environment-dependent variation that influence the proteome. The central blue box presents the primary flow of molecular information as found in all living organisms, through which DNA encodes for genes that are transcribed 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.