Journal of Proteomics 135 (2016) 12–25

Contents lists available at ScienceDirect

Journal of Proteomics

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

Review

From molecules to mating: Rapid evolution and biochemical studies of reproductive proteins

Damien B. Wilburn *, Willie J. Swanson

Department of Genome Sciences, University of Washington, United States

ARTICLE INFO

Article history:
Received 16 March 2015
Received in revised form 9 June 2015
Accepted 10 June 2015
Available online 11 June 2015

Keywords:
Sexual selection
Reproduction
Pheromones
Fertilization
Evolution

ABSTRACT

Sexual reproduction and the exchange of genetic information are essential biological processes for species across all branches of the tree of life. Over the last four decades, biochemists have continued to identify many of the factors that facilitate reproduction, but the molecular mechanisms that mediate this process continue to elude us. However, a recurring observation in this research has been the rapid evolution of reproductive proteins. In animals, the competing interests of males and females often result in arms race dynamics between pairs of interacting proteins. This phenomenon has been observed in all stages of reproduction, including pheromones, seminal fluid components, and gamete recognition proteins. In this article, we review how the integration of evolutionary theory with biochemical experiments can be used to study interacting reproductive proteins. Examples are included from both model and non-model organisms, and recent studies are highlighted for their use of state-of-the-art genomic and proteomic techniques.

Significance: Despite decades of research, our understanding of the molecular mechanisms that mediate fertilization remain poorly characterized. To date, molecular evolutionary studies on both model and non-model organisms have provided some of the best inferences to elucidating the molecular underpinnings of animal reproduction. This review article details how biochemical and evolutionary experiments have jointly enhanced the field for 40 years, and how recent work using high-throughput genomic and proteomic techniques have shed additional insights into this crucial biological process.

Contents

1. Introduction

Sexual reproduction, while prevalent along every branch of the tree of life, remains a challenge for evolutionary biologists to explain [1]. Asexual reproduction offers the advantages of propagating twice the genetic material, lacks the costs associated with finding mates, and can more rapidly establish favorable epistatic effects [2]. Early mathematical models also supported asexual reproduction as the optimal reproductive strategy. However, in more realistic scenarios of dynamic ecosystems with changing environments and co-evolving symbiotes, frequent recombination is needed and natural selection favors sexual reproduction [3,4]. Given the breadth and diversity of sexually reproducing organisms, it is no surprise that various strategies have evolved to improve reproductive success. In animals, males often perform various courtship displays [5–7], deliver pheromones that affect female

* Corresponding author.
E-mail address: dwilburn@u.washington.edu (D.B. Wilburn).

http://dx.doi.org/10.1016/j.jprot.2015.06.007
1874-3919/

behavior and physiology [8–10], and regulate the contents of their ejaculate based on female quality [11,12]. Similarly, to improve mate fitness and quality, females must be able to discriminate between these cues for honest or dishonest signals of fitness [13]. Both the male and female characteristics involved can be modified by sexual selection. Under sexual reproduction, mates must be procured to provide complementary genetic material, much like a predator capturing prey for energy and nutrient acquisition. Just as predators and prey often evolve through arms race dynamics, the continual adaptation between elaborate male traits and female perception represents one of the most well characterized examples of rapid, exacerbated co-evolution [14].

The literature is rich with examples of co-evolving sexually selected traits [15]. For historical reasons, the majority of study systems have been visible characteristics such as body size, coloration, mating behaviors, and secondary sexual traits [16]. In recent decades, as molecular biology and biochemistry have advanced, research on sexually selected traits has broadened to include the study of reproductive proteins [14], which we broadly define as any polypeptide directly involved in reproduction. While all reproductive proteins may be subject to sexual selection, the most interesting examples are likely those that directly bind molecules derived from the other sex: examples include pheromones and their cognate receptors [17], interacting egg and sperm surface proteins [18], and seminal proteins that alter female physiology [19]. A recurring theme among reproductive proteins is rapid evolution. As sexual reproduction is an essential biological process for most animals, one might expect that the majority of the reproductive proteins would be under strong negative selection to maintain compatibility. However, the recurring pattern of rapidly evolving reproductive proteins has been observed in both vertebrates and invertebrates at several stages of reproduction [14]. Because selection is most likely to act on functionally important residues in a protein, signatures of positive Darwinian selection can often guide further investigation into their underlying biochemical mechanisms [20], with studies of reproductive proteins serving as exemplars of applying molecular evolutionary techniques to characterize protein function [21–25].

Near the turn of the century and following the completion of the human genome project, a surge of high throughput technologies emerged which have altered the size and scope of questions that biologists can now ask. Various next-generation sequencing (NGS) platforms permit *de novo* analysis of whole genomes and transcriptomes for both model and non-model organisms [26,27]. Likewise, advances in mass spectrometry (MS) now provide the opportunity to qualitatively and quantitatively characterize whole proteomes [3,28,29,30]. These techniques have additionally been adapted to a wide array of other specific "omic" applications (*e.g.*, metabolomics, phosphoproteomics, pharmacogenomics), but both NGS- and MS-based approaches are quickly becoming the standard for the initial characterization of any biological system [23]. Here we review the biochemical investigations of many reproductive proteins that span various levels of reproduction: pre-copulatory behavior (pheromones), copulation (seminal proteins), and fertilization (sperm/egg proteins). Recent studies in the field have employed NGS- and MS-based approaches, and we discuss how applying such "omics" techniques to reproductive systems may be further integrated with detailed mechanistic and theoretical evolutionary models.

2. Molecular evolution and models of sexual selection

The molecular evolution of any given trait is shaped by neutrality or some form of selection (balancing, directional, or disruptive) (Fig. 1A). Balancing selection reduces genetic diversity and stabilizes a trait at some optimum phenotype. Disruptive selection is the opposite of balancing selection and favors individuals with extreme phenotypes. Finally, directional selection shifts a trait towards a single extreme. Two suites of statistical tests which use either allele frequencies or nucleotide substitutions have been developed, and each tests for selection on relatively different time scales. In the first suite of analyses, assumptions are made concerning the rates at which specific mutations are accumulated and distributed among alleles, often within and between populations. These tests are particularly valuable for identifying recent selection following selective sweeps, but are also heavily influenced by population demographics and bottlenecks (for more thorough review, see [31]). The second set of analyses compares the frequency of nucleotide substitutions at codons within genes – usually between species – and describes trends on relatively longer time scales. In the absence of selection, most nucleotide substitutions (and amino acid substitutions) are free to accumulate at the basal mutation rate. The rate of synonymous substitutions (d_S) provides an estimate of this mutation rate, and under neutrality, non-synonymous substitutions (d_N) should similarly occur, yielding a ratio of $d_N/d_S \approx 1$. Because most non-synonymous substitutions alter the tertiary structure of a protein and negatively impact function, non-synonymous substitutions should occur more rarely ($d_N/d_S < 1$). Residues where non-synonymous substitutions are disfavored are described as under negative or purifying selection [20,32–34]. Unsurprisingly, the average d_N/d_S across the protein coding sequence for most genes is less than one, and in humans, the genome-wide average $d_N/d_S \sim 0.25$ [35]. The purging of deleterious mutations by purifying selection often results in stabilizing selection of a trait. However, under situations where rapid mutation may be adaptive, non-synonymous substitutions can accumulate more quickly than the mutation rate ($d_N/d_S > 1$) and the trait is described as under positive selection [32]. The forces leading to positive selection often generate directional selection, but disruptive selection is also possible when nearly any deviation from the mean is similarly favorable.

Since each residue differentially contributes to a given protein's structure and function, and it is likely that all three forces of selection are simultaneously acting on protein-coding genes to different degrees. For example, with serine proteases, most of the protein surface is covered in polar residues that are functionally neutral and highly interchangeable. Within the active site, purifying selection preserves the catalytic triad of serine, histidine, and aspartate that are critical for enzymatic activity, and mutations are rare save for cases of atypical function [36]. However, for select serine proteases that are involved in apoptosis, adaptive response to pathogens which cause cell death has promoted positive selection on active site residues that mediate inhibitor binding and substrate specificity [37]. While the identification of neutrally evolving sites and those under purifying selection can be advantageous for understanding protein function, both forces are common in maintaining protein function and do not necessarily reflect adaptation to specific stimuli. Hence, greater interest is often placed on sites under positive selection, and various statistical packages exist to compute d_N/d_S scores along phylogenetic trees for both whole genes and specific residues [20].

While positive selection and rapid evolution have been documented in a range of systems [33,38–40], they are practically hallmarks of interacting reproductive proteins [14]. Over the last few decades, various quantitative genetic models have been developed to address how elaborate male traits and female preferences may evolve. While qualitatively described by Fisher in the 1930s, Lande [41] was the first to formalize the theory using a genetic correlation matrix (Fig. 1B–D). To illustrate this, assume that females of a given species have a preference for some male ornament, such as large, colorful peacock tails. If there is some heritable component to both the male ornament and the female preference, female peacocks with strong preferences and males with bright tails should produce daughters and sons who carry both traits, leading to genetic correlation and linkage disequilibrium. This process may continue iteratively through generations, which can lead to greater genetic association, stronger preferences, and more pronounced ornaments as part of a model of runaway selection [16]. This has alternatively been dubbed "the sexy son" hypothesis [42]. However, various factors may restrain the characteristics from evolving indefinitely. Elaborate ornaments may be energetically costly, and the