Detecting Adaptation with Genome-Scale Molecular Evolutionary Analysis: An Educational Primer for Use with "RNA Interference Pathways Display High Rates of Adaptive Protein Evolution in Multiple Invertebrates"

Brian P. Lazzaro[1]
Cornell Institute of Host-Microbe Interactions and Disease, Department of Entomology, Cornell University, Ithaca, New York 14853
ORCID ID: 0000-0002-3881-0995 (B.P.L.)

ABSTRACT Hosts and pathogens impose coevolutionary pressure on each other as pathogens strive to establish themselves and hosts seek to suppress infection. RNA interference (RNAi) is a mechanism by which cells repress viruses and transposable elements, thereby serving as a form of immune defense. Previous studies have shown that antiviral RNAi genes evolve extraordinarily quickly in the fruit fly *Drosophila melanogaster*, suggesting that they may adaptively coevolve with viruses and transposable elements. An article by Palmer and colleagues extends this observation to nematodes and multiple insects. Their article can be combined with this Primer to demonstrate the use of comparative genomics and molecular evolutionary analyses in the measurement of natural selection.

KEYWORDS adaptation; antiviral; coevolution; host–pathogen; natural selection; RNAi; transposable element

Related article in *GENETICS*: Palmer, W. H., J. D. Hadfield, and D. J. Obbard, 2018 RNA-Interference pathways display high rates of adaptive protein evolution in multiple invertebrates. Genetics 208: 1585–1599.

TABLE OF CONTENTS

doi: https://doi.org/10.1534/genetics.118.301453
Manuscript received August 3, 2018; accepted for publication August 30, 2018
Available freely online through the author-supported open access option.
[1]Corresponding author: Cornell Institute of Host-Microbe Interactions and Disease, Department of Entomology, Cornell University, Ithaca, New York 14853. E-mail: bplazzaro@cornell.edu

IMMUNE response genes are frequently among the most rapidly evolving genes in the genome (Nielsen *et al.* 2005; Sackton *et al.* 2007; Waterhouse *et al.* 2007; Obbard *et al.* 2009). This is thought to indicate antagonistic coevolution between hosts and pathogens, where each reciprocally adapts to the other in a never-ending cycle of one-upmanship (Dawkins and Krebs 1979), as well as adaptation of the immune system to novel pathogens that invade host populations. Of course, not all components of the immune system will experience identical natural selective pressure. Some genes may evolve more quickly than others based on the details of their function, and on the nature of their interactions with pathogens and pathogen-derived molecules. Molecular evolutionary analyses can be used to identify the components of the immune response that experience the strongest natural selective pressures, and comparative analyses can reveal how universal these selective pressures are.

What Is RNA Interference?

RNA interference, or RNAi, describes a set of related mechanisms by which RNA molecules are targeted for silencing or degradation in cells (a short explanatory video can be viewed at https://www.youtube.com/watch?v=cK-OGB1_ELE). RNAi is triggered by the presence of double-stranded RNA, including the genomes of double-stranded RNA viruses and replicating single-stranded RNA viruses. Additionally, in eukaryotes, the expression of mature mRNAs can be post-transcriptionally regulated by short RNAs that bind the mRNA transcript to create small segments of double-stranded RNA (Ha and Kim 2014). These short RNAs are known as micro-RNAs, or miRNAs. RNAi can also be activated by short-interfering RNAs (siRNAs), which are produced in response to foreign RNA (*e.g.*, viruses) or to active transposable elements (Bronkhorst and van Rij 2014). Transposable element activity in the germline is additionally suppressed by RNAi mediated by piwi-interacting RNAs (piRNAs), which are expressed from graveyards of inactivated transposable elements stored in the genome (Thomson and Lin 2009; Czech and Hannon 2016). In insects, distinct pathways activate RNAi in response to miRNAs, siRNAs, and piRNAs. These pathways are hypothesized to be under different selective pressures: miRNAi pathways perform housekeeping functions in the cell and therefore may be expected to evolve largely under purifying selection; siRNAi and piRNAi pathways suppress pathogenic viruses and transposable elements, and therefore may evolve adaptively under host–pathogen conflict. Crucially, the homologous proteins in the various RNAi pathways have similar biochemical activities, but they operate in different contexts. Comparisons among them can therefore test whether patterns of adaptation are due to the biochemical functions of the proteins or to the contexts in which they operate. The power of such contrasts was elegantly demonstrated in a previous paper by Obbard *et al.* (2006), which showed that the siRNA pathway of *Drosophila melanogaster* contains some of the most rapidly and adaptively evolving genes in the genome, whereas the miRNA pathway of *D. melanogaster* evolves at a rate similar to that of other housekeeping processes.

Detecting Adaptation with DNA Sequence Data

Patterns of DNA sequence diversity within populations and between species can reveal both recent and ancient adaptation. Recurrent adaptation by the same protein over long evolutionary time results in a higher rate of amino acid change than would be expected in the absence of natural selection. This can be revealed by comparing rates of DNA and protein sequence divergence between species (Hughes and Nei 1988; McDonald and Kreitman 1991). Recent bouts of natural selection alter allele frequencies in the chromosomal regions that surround the selected genes (Smith and Haigh 1974, Kaplan *et al.* 1989). These impacts can be detected by sequencing alleles of the genes from multiple individuals sampled from a population (*e.g.*, Tajima 1989; Fu and Li 1993; Fay and Wu 2000). Sequence-based tests for adaptation are especially powerful when evaluated over the entire genome, thereby revealing genes whose evolution departs dramatically from genome norms (*e.g.*, Nielsen *et al.* 2005; Larracuente *et al.* 2008).

The concept of using the rate of amino acid divergence between species as a specific test for adaptive evolution was first developed for the Major Histocompatibility Complex (MHC) antigen-presenting genes of the vertebrate immune system (Hughes and Nei 1988). The premise was that the rate of divergence at synonymous codon sites (which do not change protein sequence when mutated) would provide a baseline rate of divergence between the species due to mutation and genetic drift without natural selection. Nonsynonymous mutations change the amino acid sequence of the protein, which is deleterious in most cases. Natural selection is expected to remove these negative mutations from the population, so they will not accumulate as differences between species. The rate of nonsynonymous divergence (d_N) is therefore generally predicted to be much lower than the rate of synonymous divergence (d_S), and indeed that is what is observed for the vast majority of genes in the genome of any organism. However, Hughes and Nei (1988) noted that the antigen-presenting domain of the MHC genes showed *higher* rates of nonsynonymous divergence (amino acid replacement) than synonymous divergence (silent genetic change), suggesting that amino acid substitutions in the antigen-presenting domain may be adaptively favored.

McDonald and Kreitman (1991) subsequently developed a more powerful test based on the same concept. The McDonald–Kreitman (MK) test uses a 2×2 contingency table to statistically compare the levels of synonymous and nonsynonymous divergence between a pair of species with the levels of synonymous and nonsynonymous polymorphism among individuals within one of the species (Figure 1). Assuming that synonymous mutations are not affected by natural selection, the number of synonymous polymorphisms within a species and the number of synonymous differences

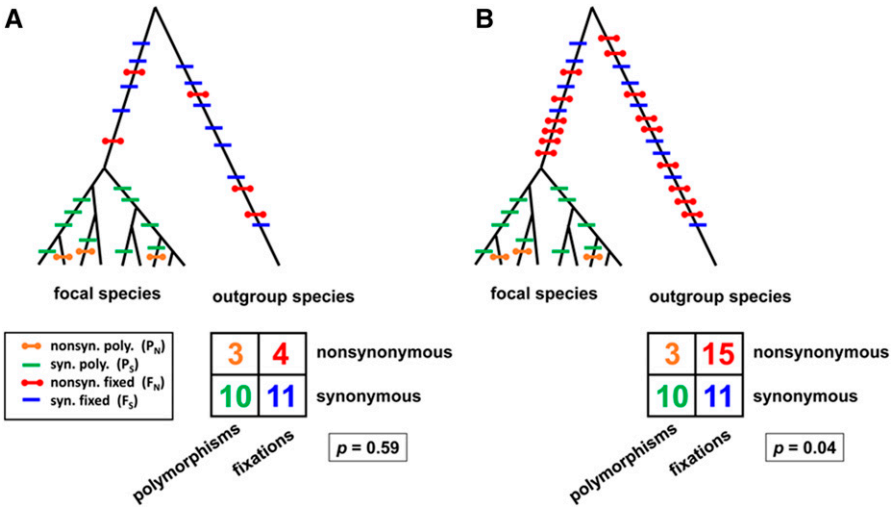

Figure 1 Schematic illustration of the conventional McDonald–Kreitman test. The test uses a 2×2 contingency table to compare the observed number of synonymous (syn.) polymorphisms (poly.) within species (P_S), nonsynonymous (nonsyn.) polymorphisms within species (P_N), synonymous differences fixed between species (F_S), and nonsynonymous differences fixed between species (F_N). A *P*-value is typically obtained with Fisher's exact test. (A) In the absence of adaptive evolution, the ratio P_N/P_S is similar to the ratio F_N/F_S, and is typically < 1 since purifying selection will eliminate the large fraction of nonsynonymous mutations that are deleterious. (B) When nonsynonymous mutations are adaptive, natural selection quickly drives them to fixation within a species. They are then observed as fixed differences between species, inflating the observed F_N and resulting in a significant McDonald–Kreitman test. The advanced tests used in Palmer *et al.* (2018) to screen RNA interference genes for ω_A and "selection effect" are fundamentally based on the McDonald–Kreitman concept.

between species should reflect the rates at which mutations arise in a population and drift to fixation. Some nonsynonymous mutations are inconsequential, but most are deleterious. A large faction of nonsynonymous mutations that arise should therefore be eliminated by purifying selection. A few may be observed as harmful polymorphisms within the species and some of the innocuous ones will drift to fixation between species, but the selective removal of deleterious mutations will generally result in the observed number of nonsynonymous polymorphisms and fixations being smaller than the number of synonymous polymorphisms and fixations. Importantly, in the absence of adaptive evolution, the ratio P_N/P_S should be similar to the ratio F_N/F_S because mutation, genetic drift, and purifying selection are the predominant forces both within and between species (Figure 1A). However, some nonsynonymous mutations may be adaptive. Natural selection should quickly drive these to fixation and they will increase the observed number of differences fixed between species (Figure 1B). Schlenke and Begun (2003) used the MK test to show that genes in the *D. melanogaster* and *D. simulans* immune systems tend to evolve adaptively. Palmer *et al.* (2018) screen for adaptive evolution in RNAi genes of several insect and nematode species using a more sophisticated version of the test, but the underlying premise is the same.

Crucially, the d_N/d_S and MK tests can only detect the accumulation of a large number of amino acid fixations over long evolutionary time. These tests are insensitive to rare adaptive events. However, recent adaptive fixations can be detected on an individual basis by virtue of their impact on surrounding genetic diversity. When strong natural selection drives an adaptive mutation to fixation in a species, the region of chromosome around the adaptive mutation "hitchhikes" to fixation along with it (Smith and Haigh 1974). All previously existing alleles are displaced by the adaptively favored chromosome, eliminating genetic diversity until it is slowly restored by new mutations (Figure 2). For an evolutionarily short time after it fixes, the adaptive mutation will therefore be surrounded by a window of reduced genetic diversity (the size of the window is dependent on the speed of fixation and the rate of meiotic recombination in the population). Various tests have been developed to measure the reduction in diversity and recovery of variation around such "selective sweeps" (*e.g.*, Tajima 1989; Fu and Li 1993; Fay and Wu 2000), based on the number of DNA polymorphisms in set of DNA sequences and the individual allele frequencies of each polymorphism [termed the "site frequency spectrum" (SFS)]. Palmer *et al.* (2018) apply an advanced test of the SFS to evaluate the relative likelihood of selective sweeps around RNAi genes.

Tests for adaptive evolution may be conducted in relation to an explicit null hypothesis. However, these null hypotheses often hinge on unrealistic assumptions, such as an assumption that there is no natural selection at all. Unrealistic assumptions can lead to inappropriate rejection of null hypotheses, as in cases where the demographic history of a species results in an SFS similar to that expected from adaptive evolution (Teshima *et al.* 2006). Therefore, it is often desirable to compare the values of test statistics obtained from focal study genes (*e.g.*, genes in RNAi pathways) to the complete set of genes in the genome. If the study genes exhibit a pattern of evolution that is different from that of the rest of the genome, the investigator may conclude that the study genes have evolved adaptively (*e.g.*, Nielsen *et al.* 2005). An alternative approach is to compare study genes to a similar-sized set of control genes that have arbitrary functions. This is the approach that Palmer *et al.* (2018) choose. These control genes can be matched to the study genes for factors like gene length, expression level, and chromosomal position, all of which may affect molecular evolutionary rates (*e.g.*, Larracuente *et al.* 2008), but

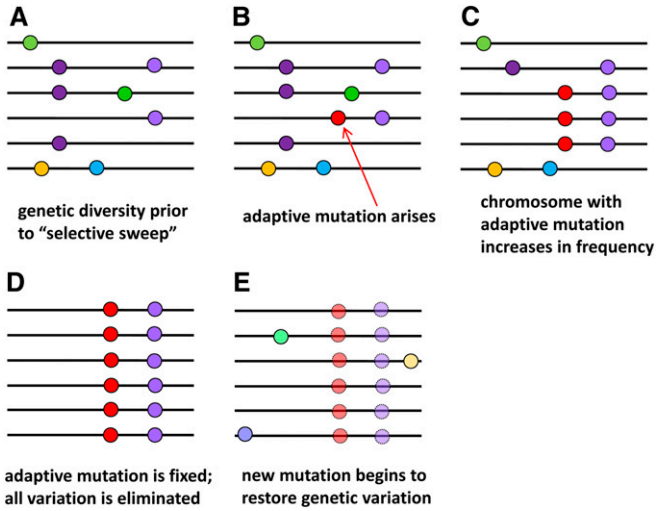

Figure 2 Schematic illustration of a selective sweep. Horizontal lines represent alleles present in a population. The circles represent mutations that distinguish the alleles. (A) Genetic diversity in the population before the selective sweep. (B) An adaptive mutation, indicated by the red circle, arises in the population. (C) The adaptive allele increases in frequency in the population, displacing other alleles. (D) When the adaptive allele reaches fixation, the chromosomal region surrounding the adaptive mutation is invariant in the population. The size of this region will depend on the speed with which the adaptive allele fixes and the rate at which meiotic recombination moves the adaptive mutation onto different genetic backgrounds (not shown). (E) Mutation will begin to restore genetic variation in the population, but these new polymorphisms will be rare in the population. Palmer *et al.* (2018) use a statistical test that scans the genome for segments of DNA that have low diversity and a site frequency spectrum that is skewed toward rare variants.

they will not have the biological function that is hypothesized to result in adaptive evolution.

The Value of Comparative Analyses and Public Databases

Comparisons across distantly related species can indicate whether a biological phenomenon is general, or whether it is specific to particular species or groups. In previous work, Obbard *et al.* (2006) have shown that siRNA genes of *D. melanogaster* evolve rapidly and adaptively compared to other genes in the genome, presumably because of conflict with viruses and transposable elements. Palmer *et al.* (2018) extend this analysis to include a new *D. melanogaster* data set, another *Drosophila* species (*D. pseudoobscura*), a mosquito (*Anopheles gambiae*), a honeybee (*Apis mellifera*), a moth and a butterfly (*Bombyx mandarina* and *Heliconius melpomene*), and two nematodes (*Pristionchus pacificus* and *Caenorhabditis briggsae*). They hypothesize that if conflict with viruses and/or transposable elements generally drives rapid evolution, they should see parallel adaptation in the siRNA genes of all of these species. The prediction is consistent with the findings from a recent paper by Enard *et al.* (2016), which showed that evolutionary escape from viruses can be a strong driver of adaptive evolution in broadly diverse mammalian proteins, including proteins that are not involved in RNAi.

The Palmer *et al.* (2018) study is possible because population genomic data sets (genome sequences from multiple individuals) have been generated for each of these species by other groups and the data are freely available in public archives. This project would have been impossible 5 years ago, when genome sequencing was more expensive and existing data sets were sparse. It would be impractically expensive and difficult even today if Palmer and colleagues would have had to generate the sequence data themselves. However, studies like that of Palmer *et al.* (2018) become possible with the falling cost of genome sequencing and research teams making the sequences they that generate freely available. In the age of genomics and open science, previously unimaginable projects are becoming inexpensive and feasible, and widespread data sharing allows secondary investigators to address questions that might never have been conceived by the primary data generators.

Unpacking the Study

The objective of this study is to test whether host–pathogen interactions drive adaptive evolution in RNAi genes. There are two levels of control in the study. The matched control genes that have no RNAi function demonstrate typical evolution of the genome, allowing the test of whether RNAi genes have elevated rates of adaptation. The housekeeping miRNAi genes serve as negative controls for the hypothesis that adaptation is driven by host–pathogen coevolution. If antiviral RNAi (viRNAi), siRNAi, and miRNAi genes all show more adaptive evolution than the functionally unrelated genome controls, then the adaptation must arise from some aspect of RNAi function, but may *not* be driven by host–pathogen conflict. However, if the viRNAi and siRNAi genes show higher rates of adaptation than the miRNAi genes and the background genome, then the investigators can conclude that host–parasite interactions probably drive adaptive evolution of the viRNAi and siRNAi systems. Obbard *et al.* (2009) previously came to this conclusion in a study of *D. melanogaster*, and the Palmer *et al.* (2018) study tests whether the phenomenon is generally consistent across multiple species of insects.

Curating the Data

Palmer *et al.* (2018) combine a variety of computational tools to identify RNAi and control genes in each species, retrieve the gene sequences from public databases, and ensure that the sequence data are of good quality for molecular evolutionary analysis. This rigor of this pipeline is very important in experimental practice; however, the details of how each computational tool works are not essential for most teaching contexts. An overview of what the investigators are trying to achieve and why they want to do it may be preferable to a nuts-and-bolts dissection of how exactly they did it.

The sequence data used in this study were generated in several different studies by different investigators using varied methods. The first tasks of Palmer *et al.* (2018) were therefore to retrieve the data from the public databases, identify the RNAi genes from each species, and assemble similar sets of control genes for all species. To identify the RNAi genes from each species, they used the already known and annotated gene sets from *D. melanogaster* and *C. elegans* as "query" sequences to find the most similar sequences in the genomes of the other insects and nematodes. If no good match to a query gene could be found in the genome being analyzed, the gene was classified as missing. If two or more equally good matches were found, the gene was classified as duplicated. The genomic control genes were the physically closest genes upstream or downstream of each RNAi gene that were roughly similar in length. Polymorphic alleles of all of these genes were then retrieved from the population genomic databases for each species. Gene regions were only included if they were covered by a minimum of five independent sequence reads (or two reads for *B. mandarina*, for which less sequencing had been performed) to maximize the probability that both alleles would be observed in a heterozygote.

Palmer *et al.* (2018) use molecular evolutionary tests that require comparisons to closely related species, termed "outgroups" (see Figure 1). For each of the main species analyzed, the authors retrieved homologous RNAi and control genes from the mostly closely related species for which a genome sequence was available. These outgroup species are typically $< 10\%$ diverged at the DNA sequence level. However, the phylogenetic structure of *Anopheles* mosquito species presented a complication. There is a constellation of species that are very closely related to *A. gambiae*, but these are so closely related (in some cases, they are hybridizing subspecies) that they cannot be used for analysis of sequence divergence. However, the next most closely related species are too divergent for optimal analysis (Obbard *et al.* 2007). Palmer *et al.* (2018) solve this problem by performing the *A. gambiae* analyses twice: once in comparison to the too-close *A. melas* and once in comparison to the too-divergent *A. christyi*. Both analyses give qualitatively similar results and they present only the findings from comparison to *A. melas* in the paper. The data for the outgroup nematode *C. nigoni* also presented a challenge. It is known that the sequence of this species is contaminated with DNA from a more distantly related species, *C. afra* (Thomas *et al.* 2015). To prevent this contamination from confounding the molecular evolutionary analyses, Palmer *et al.* (2018) exclude any sequence regions that show > 6 standard deviations higher divergence than the average between *C. nigoni* and *C. elegans*, assuming that these extraordinarily high-divergence regions reflect contaminating sequence from *C. afra*. The genomes of the outgroup bee *A. cerana* and butterfly *H. hecale* had been previously sequenced, but the raw sequence reads were not assembled into complete genomes. Therefore, Palmer *et al.* (2018) contrive a rapid pipeline to infer the gene sequences of these outgroup species from raw sequence data using established bioinformatic tools. They validated the accuracy of their pipeline by testing it with sequence data from *D. melanogaster* and *D. simulans*, two species whose complete genome sequences are well known.

Adaptive Divergence Between Species

Palmer *et al.* (2018) estimate the rate of adaptive protein evolution for each individual gene, as well as for genes pooled into various classifications to allow specific contrasts (*e.g.*, RNAi *vs.* control, viRNAi *vs.* miRNAi, etc.). They do this with two sophisticated extensions of the MK test illustrated above. In the first extension, they use a piece of software called DFE-α to estimate a parameter, ω_A, that can be interpreted as the proportion of amino acid differences between two species that were fixed by positive selection as opposed to genetic drift. ω_A is similar to d_N/d_S, with an emphasis on detecting the proportion of nonsynonymous substitutions that became fixed through adaptive evolution (as opposed to genetic drift). It is not necessary to understand the details of how this estimate is made [described in detail in Supplemental Material, Text S1 in Palmer *et al.* (2018)] to understand the biological conclusions of the Palmer *et al.* (2018) paper. It is sufficient to appreciate that the test is based on the MK concept, with a statistical correction that uses the whole-genome data set to model the demographic history of the species in question because demographic effects can impact evolutionary rates. To contrast the rates of adaptation between different classes of genes (*e.g.*, RNAi *vs.* control), Palmer *et al.* (2018) load the individual ω_A estimates from each gene into a novel analysis that they term a "multispecies meta-analysis," which allows them to estimate ω_A for entire pathways averaged across species. Thus, they can contrast whether the RNAi pathways differ from controls or from each other generically across species. These estimates have some uncertainty, and the posterior density in Figure 1B and Figure 2, B and C shows the likelihood that ω_A has a given value for each gene set. The highest point of the graphed distribution is the most likely value of ω_A and the width of the distribution indicates the confidence in that estimate.

The second extension of the MK test is the SnIPRE analysis [based on Eilertson *et al.* (2012)], which returns a value called the "selection effect" for each gene or gene set. In the SnIPRE analysis, mutations are again categorized into four classes: synonymous (silent) polymorphism within species, nonsynonymous (amino acid replacement) polymorphism within species, synonymous divergence fixed between species, or nonsynonymous divergence fixed between species (Figure 1). The data from multiple species pairs can be combined for simultaneous analysis, and the likelihood that a gene or gene set evolves at a high rate of adaptation is inferred from a statistical excess of nonsynonymous divergences. As with the DFE-α analysis, it is not necessary to understand the details of how the model operates to grasp the biological conclusions of the study. The essential point is to understand that a significantly positive "selection

effect" indicates a relative excess of amino acid divergence between species in a gene or gene set, and that this excess is interpreted as evidence for adaptive evolution.

Both the DFE-α and the SnIPRE approaches demonstrate higher rates of adaptive evolution in the RNAi genes than in the controls. This is shown, for example, as a significantly higher average ω_A for the RNAi genes in the DFE-α analysis ($\omega_A = 0.01$ in control genes and $\omega_A = 0.062$ in RNAi genes, $P < 0.001$). However, the DFE-α analysis also revealed a much higher *variance* in DFE-α among the RNAi genes, suggesting that different RNAi genes or pathways may evolve with different rates of adaptive evolution. Sure enough, a pathway-level analysis demonstrates that the housekeeping miRNA genes show no difference from the control genes in rate of adaptive amino acid divergence. However, the rate of adaptive divergence is much higher in the viRNA, piRNA, and siRNA classes. The SnIPRE analysis shows the same pattern. In all analyses, the viRNA genes show the strongest evidence for adaptive divergence between species, which is consistent with the hypothesis that interaction with viruses is a strong driver of adaptive protein evolution [see also Obbard *et al.* (2006) and Enard *et al.* (2016)]. The piRNA genes, which are responsible for suppressing transposable elements in the germline, also show strong signatures of adaptation that are most easily detected with the DFE-α analysis. Following Blumenstiel *et al.* (2016), Palmer *et al.* (2018) speculate that adaptation in piRNA genes may be driven by the invasion of new transposable elements into each species and by evolutionary fine-tuning of the response to existing transposable elements. The paper presents multiple additional analyses that parse the genes into different functional subclasses, and set up a variety of specific contrasts within and across species, but they all follow the same logic and support the same broad set of conclusions. A classroom exercise could be to examine each of the contrasts, and to have students identify the specific hypothesis being tested and interpret the results in each.

Recent Adaptation Within Species

Evolutionarily recent bouts of strong adaptation leave a signature of reduced diversity and distorted allele frequencies in the genome immediately surrounding adaptive mutations. The distribution of the individual allele frequencies of every polymorphic nucleotide in a genomic window is termed the site frequency spectrum. Palmer *et al.* (2018) use an algorithm called SweeD to scan for evidence of selective sweeps. The SweeD algorithm assesses the likelihood that a selective sweep has occurred in a given genomic interval based on the deviation of the SFS of that interval from what would be typically expected for the genome being analyzed. As with the DFE-α and SnIPRE analyses, it is not necessary for students to understand the details of the SweeD algorithm unless the course has an intensive focus on population genetic methods. For most classes, it will be sufficient to appreciate that SweeD identifies genomic regions that are likely to have experienced a recent selective sweep. Palmer *et al.* (2018) then contrast the abundance of apparent selective sweeps in genomic regions that include RNAi genes to those that contain control genes. The explicit hypothesis is that if RNAi genes exhibit adaptive evolution more often than the rest of the genome does, then there should be positive SweeD results at RNAi genes more often than at control genes. Conceptually analogous analyses are conducted to contrast other classifications of RNAi genes, such as viRNAi *vs.* miRNAi genes.

Palmer *et al.* (2018) find evidence of selective sweeps within 1 kb of a significantly larger proportion of RNAi genes than control genes in six of the eight species (limited genome annotation prevented analysis of the *P. pacificus* and *B. mandarina* genomes). However, all RNAi subpathways were equally likely to have experienced recent sweeps. Remarkably, two genes—*spn-E* (a piRNA gene) and *vig* (a siRNA gene)—showed evidence of a recent sweep in five of the six insect species. This degree of parallel adaptation indicates that these genes are frequently subject to strong selective pressure in diverse species and may be common targets of host–parasite coevolution. However, for the most part, different genes in each species show evidence of recent adaptive sweeps.

Connections to Genetics Concepts

Although first intuition might suggest that adaptive mutations will be obvious in complete genome sequences, in practice it is virtually impossible to determine from DNA sequence alone whether any particular mutation is (or was) favored by natural selection. The adaptive mutations themselves do not look any different from neutral mutations that become fixed between species by genetic drift. A challenge in evolutionary biology then becomes to infer which genes have experienced recent or recurrent adaptive evolution. Contemporary approaches for inferring historical adaptation from sequence data can be daunting for nonexperts, and the Methods sections of papers, such as that by Palmer *et al.* (2018), can seem impenetrable to the uninitiated. The key to teaching papers like this in an introductory classroom is to emphasize two fundamental concepts: (1) repeated episodes of protein adaptation will speed up the rate of divergence between species (Figure 1) and (2) rapid fixation of an adaptive mutation sweeps the surrounding chromosomal region to fixation as well (Figure 2). Modern methods employ sophisticated model testing to distinguish adaptive evolution from neutral evolution under complex demographic scenarios, which is a level of rigor that is essential for scientific practice. However, even the sophisticated models are based on these two fundamental concepts, and the more complex details can be set aside for more advanced classes.

The Palmer *et al.* (2018) article provides a compelling blend of fundamental molecular evolutionary concepts with cutting edge application. The overall biological interpretation is very accessible: host–pathogen coevolution leads to

adaptation in antiviral and antitransposable-element genes. The study can be used as a case example in the application of fundamental molecular evolutionary analyses in an introductory class, and the rigorous application of cutting-edge methodology can serve as a model for more advanced classes. The Palmer *et al.* (2018) article further serves to teach the value of genome-scale comparisons for drawing evolutionary inference, as all of their analyses gain power from the contrasts of multigene sets and pathways. Finally, the article serves as an endorsement of open science and data sharing. The paper relies on the independent and parallel observation of similar evolutionary patterns in different species to demonstrate the generality of host–pathogen coevolution in RNAi genes. This is only possible because the genome sequences are publicly available and freely shared. The Palmer *et al.* (2018) article can thus be used as a catalyst for classroom discussion of data ownership and sharing in global scientific research, which is especially pertinent in the postgenomic era of massive data generation.

Suggestions for Classroom Use

Instructors are encouraged to provide this Primer to students in conjunction with the article by Palmer *et al.* (2018). Students could be encouraged to focus on the Introduction and main figures of Palmer *et al.* (2018), supplemented with the background explanations provided above. Instructors may also want to employ a variant of the C.R.E.A.T.E. strategy of having students read and interpret several papers in sequence (Hoskins *et al.* 2007). A potential sequence could be to begin with the original d_N/d_S paper from Hughes and Nei (1988), then read the original article presenting the MK test (1991). These papers could be followed by Obbard *et al.* (2006), which showed that viRNAi genes are among the fastest evolving genes in the *D. melanogaster* genome, ultimately leading in to the present article by Palmer *et al.* (2018). The two initial papers are brief and accessible, and neatly present the logic behind the molecular evolutionary tests as originally conceived. The Obbard *et al.* (2006) paper provides an initial application of those methods to a small number of RNAi genes in a single species, and also presents an initial contrast of those genes to the evolutionary trajectory of the background genome. The paper provides a clear introduction to Palmer *et al.* (2018), which then addresses essentially the same question with more sophisticated methods in more species simultaneously. By reading these papers in series, students will come to appreciate the progression of both methodology and biological understanding over years of research.

Questions for Further Exploration

The following questions can be used to simulate discussion either in small groups or as a whole class. Students may also be assigned to prepare short answers to one or more questions in advance of a class meeting. Pondering these questions should promote deeper understanding of the Palmer *et al.* (2018) article and the concepts therein.

1. Palmer *et al.* (2018) do a parallel analysis of the RNAi genes of six species of insect and two species of nematode. What value is gained by doing this study in multiple species? Should it be expanded to an even larger number of species?
2. Two of the analyzed species yield negative estimates for ω_A in both control genes and RNAi genes. Which two species are those, and how is that observation interpreted?
3. Even in the two species for which ω_A is negative, the estimate is more negative in control genes than it is in RNAi genes. How is this observation interpreted?
4. Based on figure 2, which class of genes shows the strongest evidence of adaptive evolution?
5. What do the vertical gray bars in figure 4 indicate (note that there is a thin bar in the left panel in addition to the prominent bar in the right panel).
6. The authors observe a greater variance in estimated ω_A among RNAi genes than among control genes. What does this indicate about adaptation in RNAi pathways?
7. In figure 4, the authors show relatively little variance in estimated ω_A among genes within the viRNAi, piRNAi and siRNAi, and miRNAi pathways. What does this indicate about the effects of selection on these pathways?
8. The authors did not identify any genes that showed significant evidence of adaptive evolution in every species tested. However, there were strong parallels in the relative strengths of adaptive evolution on the different RNAi pathways across species, and seven individual genes were identified as showing strong indication of adaptive evolution in more than one-half of the species examined. Yet the species are distantly enough related that they should be infected by distinct viruses and transposable elements. Why would the same RNAi genes evolve adaptively in species that are infected by different parasites?
9. Demographic events like strong population bottlenecks followed by rapid population expansion can result in an SFS that is similar to the SFS expected after a selective sweep. Explain why this would be the case, using illustrations to support your argument.
10. Why are transposable elements damaging to host genomes? Why do host cells need to suppress the activity of transposable elements?
11. Some viruses have evolved the ability to inhibit RNAi mechanisms. Explain why viral inhibition of RNAi could result in adaptive evolution of the RNAi pathway.
12. The authors document adaptive evolution in viRNAi genes of multiple invertebrate hosts. Is this sufficient evidence to demonstrate coevolution? What other data might be desirable?

Literature Cited

Blumenstiel, J. P., A. A. Erwin, and L. W. Hemmer, 2016 What drives positive selection in the Drosophila piRNA machinery? The genomic autoimmunity hypothesis. Yale J. Biol. Med. 89: 499–512.

Bronkhorst, A. W., and R. P. van Rij, 2014 The long and short of antiviral defense: small RNA-based immunity in insects. Curr. Opin. Virol. 7: 19–28. https://doi.org/10.1016/j.coviro.2014.03.010

Czech, B., and G. J. Hannon, 2016 One loop to rule them all: the ping-pong cycle and piRNA-guided silencing. Trends Biochem. Sci. 41: 324–337. https://doi.org/10.1016/j.tibs.2015.12.008

Dawkins, R., and J. R. Krebs, 1979 Arms races between and within species. Proc. R. Soc. Lond. B Biol. Sci. 205: 489–511. https://doi.org/10.1098/rspb.1979.0081

Eilertson, K. E., J. G. Booth, and C. D. Bustamante, 2012 SnIPRE: selection inference using a Poisson random effects model. PLOS Comput. Biol. 8: e1002806. https://doi.org/10.1371/journal.pcbi.1002806

Enard, D., L. Cai, C. Gwennap, and D. A. Petrov, 2016 Viruses are a dominant driver of protein adaptation in mammals. Elife 5: e12469. https://doi.org/10.7554/eLife.12469

Fay, J. C., and C.-I. Wu, 2000 Hitchhiking under positive Darwinian selection. Genetics 155: 1405–1413.

Fu, Y.-X., and W.-H. Li, 1993 Statistical tests of neutrality of mutations. Genetics 133: 693–709.

Ha, M., and V. N. Kim, 2014 Regulation of microRNA biogenesis. Nat. Rev. Cell. Biol. 15: 509–524. https://doi.org/10.1038/nrm3838

Hoskins, S. G., L. M. Stevens, and R. H. Nehm, 2007 Selective use of the primary literature transforms the classroom into a virtual laboratory. Genetics 176: 1381–1389. https://doi.org/10.1534/genetics.107.071183

Hughes, A. L., and M. Nei, 1988 Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335: 167–170. https://doi.org/10.1038/335167a0

Kaplan, N. L., R. R. Hudson, and C. H. Langley, 1989 The "hitchhiking effect" revisited. Genetics 123: 887–899.

Larracuente, A. M., T. B. Sackton, A. J. Greenberg, A. Wong, N. D. Singh *et al.*, 2008 Evolution of protein-coding genes in *Drosophila*. Trends Genet. 24: 114–123. https://doi.org/10.1016/j.tig.2007.12.001

McDonald, J. H., and M. Kreitman, 1991 Adaptive protein evolution at the Adh locus in *Drosophila*. Nature 351: 652–654. https://doi.org/10.1038/351652a0

Nielsen, R., C. Bustamante, A. G. Clark, S. Glanowski, T. B. Sackton *et al.*, 2005 A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biol. 3: e170. https://doi.org/10.1371/journal.pbio.0030170

Obbard, D. J., F. M. Jiggins, D. L. Halligan, and T. J. Little, 2006 Natural selection drives extremely rapid evolution in antiviral RNAi genes. Curr. Biol. 16: 580–585. https://doi.org/10.1016/j.cub.2006.01.065

Obbard, D. J., Y.-M. Linton, F. M. Jiggins, G. Yan, and T. J. Little, 2007 Population genetics of Plasmodium resistance genes in *Anopheles gambiae*: no evidence for strong selection. Mol. Ecol. 16: 3497–3510. https://doi.org/10.1111/j.1365-294X.2007.03395.x

Obbard, D. J., J. J. Welch, K. W. Kim, and F. M. Jiggins, 2009 Quantifying adaptive evolution in the *Drosophila* immune system. PLoS Genet. 5: e1000698. https://doi.org/10.1371/journal.pgen.1000698

Palmer, W. H., J. D. Hadfield, and D. J. Obbard, 2018 RNA-Interference Pathways Display High Rates of Adaptive Protein Evolution in Multiple Invertebrates. Genetics 208: 1585–1599. https://doi.org/10.1534/genetics.117.300567

Sackton, T. B., B. P. Lazzaro, T. A. Schlenke, J. D. Evans, D. Hultmark *et al.*, 2007 Dynamic evolution of the innate immune system in *Drosophila*. Nat. Genet. 39: 1461–1468. https://doi.org/10.1038/ng.2007.60

Schlenke, T. A., and D. J. Begun, 2003 Natural selection drives *Drosophila* immune system evolution. Genetics 164: 1471–1480.

Smith, J. M., and J. Haigh, 1974 The hitch-hiking effect of a favourable gene. Genet. Res. 23: 23–25. https://doi.org/10.1017/S0016672300014634

Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123: 585–595.

Teshima, K. M., G. Coop, and M. Przeworski, 2006 How reliable are empirical genomic scans for selective sweeps? Genome Res. 16: 702–712. https://doi.org/10.1101/gr.5105206

Thomas, C. G., W. Wang, R. Jovelin, R. Ghosh, T. Lomasko *et al.*, 2015 Full-genome evolutionary histories of selfing, splitting, and selection in *Caenorhabditis*. Genome Res. 25: 667–678. https://doi.org/10.1101/gr.187237.114

Thomson, T., and H. Lin, 2009 The biogenesis and function of PIWI proteins and piRNAs: progress and prospect. Annu. Rev. Cell Dev. Biol. 25: 355–376. https://doi.org/10.1146/annurev.cellbio.24.110707.175327

Waterhouse, R. M., E. V. Kriventseva, S. Meister, Z. Xi, K. S. Alvarez *et al.*, 2007 Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes. Science 316: 1738–1743. https://doi.org/10.1126/science.1139862

Communicating editor: E. De Stasio