

HOW TO MEASURE NATURAL SELECTION

How to measure natural selection

John R. Stinchcombe^{*,1} , Joanna L. Kelley² and Jeffrey K. Conner³

¹Department of Ecology and Evolutionary Biology, Koffler Scientific Reserve at Joker's Hill, University of Toronto, Toronto, ON M5S 3B2, Canada; ²School of Biological Sciences, Washington State University, Pullman, WA 99164, USA; and ³Kellogg Biological Station, Department of Plant Biology, Michigan State University, Hickory Corners, MI 49060, USA

Introduction

Evolution by natural selection is a simple process. The only requirements—phenotypic and genetic variation in traits, and non-random survival and reproduction of individuals based on those traits—appear to be met in most species and populations. Despite its conceptual simplicity, a full understanding of the causes and consequences of evolution by natural selection is a tall order because it is both an ecological and a genetic process. The abiotic and biotic environment interact with phenotypic variation to cause selection; for changes in phenotypes to be transmitted across generations requires genetics. Given sufficient time and strong enough selection, the signatures of evolution by natural selection should be manifest in the distribution of traits and phenotypes across a landscape, and in patterns of genetic variation in the genome. Understanding how and why some individuals survive and reproduce better than others, the traits that allow them to do so, the genetic basis of those traits, and the signatures of past and present selection in patterns of variation in the genome remain at the top of the research agenda for evolutionary biology. Here we draw together a collection of seven papers that highlight new methodological and conceptual approaches to meeting this agenda.

Characterizing natural selection

Most of the current work attempting to measure natural selection within populations, that is, to describe the relationship between traits and relative fitness, draws on the seminal work of Robertson (1966), Price (1970), and Lande & Arnold (1983). These studies showed elegantly how selection can be represented as a covariance between a trait and relative fitness (Price 1970), how a response to selection requires a genetic covariance between breeding values for a trait and fitness (Robertson 1966; Price 1970), and how total selection can be partitioned into direct selection on a trait and indirect selection through correlated traits (Lande & Arnold 1983). Collectively, these papers spawned an enormous literature devoted to conceptual, methodological, statistical, and even philosophical approaches to estimating and partitioning the phenotypic

and genetic covariances between traits and some aspect of relative fitness. Three papers in this issue clarify unresolved aspects of this literature in meaningful and important ways.

First, Franklin & Morrissey (2017) consider the knotty issue of what to do when one can only obtain a proxy for fitness, such as body size, biomass, or growth rate. Their work builds on a classic framework developed by Arnold (1983). They find that using performance in lieu of relative fitness will only reflect true selection under limited conditions, and can lead to severe under-estimates of selection. Thomson & Hadfield (2017) consider what is in some ways the opposite issue: when one has multiple fitness components, which ones belong to whom? Should one count offspring survival as fitness components of their own (offspring) generation, or as part of the fitness of their parents? And how does one account for the fact that parental phenotypes might affect not only their own fecundity but also the survival of their offspring? They show that using offspring fitness components as part of parental fitness (“mixed fitness” in their terms) is common in studies of birds and mammals, but will only lead to correct estimates of selection and evolutionary change under very restrictive conditions. These papers not only offer quantitative and statistical guidance for investigators with similar data, but also have important implications for attempts to predict evolutionary responses and for interpreting the results of meta-analyses of selection (e.g., Kingsolver *et al.* 2012). An additional complication arises when the phenotype of an individual is itself affected by the (potentially related) individuals it interacts with. Hadfield & Thomson (2017) develop methods for understanding selection in these contexts in relation to Hamilton’s costs and benefits (Hamilton 1964). A major result of their study is a multivariate version of Hamilton’s rule: they derive conditions in which altruism can evolve in multi-trait contexts. Collectively, these three papers add to understanding natural selection by clarifying some of the key issues related to measuring selection in contemporary populations: what is fitness, whose fitness it is, whose phenotype it is, and how interactions with related individuals can shape the evolution of multiple traits. Importantly, all three papers contain firm recommendations, theory, and statistical methods to guide the field.

*Correspondence author. E-mail: John.Stinchcombe@utoronto.ca

Effects of natural selection on the genome

The next three papers in the issue tackle the genomic results of evolution by natural selection: namely, how we can detect natural selection from genomic data? The advent of high-throughput sequencing, combined with advances in molecular population genetics, offers the prospect of investigating the genomic signal of natural selection in unprecedented ways, in any population or species. These papers provide an overview of some of the key conceptual issues, such as the prevalence of hard or soft sweeps, recommendations for identifying and prioritizing putatively adaptive loci, and detecting selection on transposable elements.

Hermisson & Pennings (2017) review the basic theory and evidence from population genetics on hard or soft sweeps. Is most adaptive evolution due to the occurrence of new mutations that arise after the onset of selection and rapidly sweep to fixation? Or is does selection mainly act on standing variation, or recurrent mutation of beneficial alleles after the onset of selection? Hermisson & Pennings (2017) review the theoretical predictions and evidence to date. They find a high prevalence of soft selective sweeps in the data available to date, but note that future models, simulations, and interpretations will have to go beyond the soft-hard sweep dichotomy. The paper by Lotterhos *et al.* (2017) in this issue develops new methods of combining datasets to increase our power to detect selective sweeps.

An alternative to inferring the action of selection from patterns of nucleotide polymorphism is to identify loci that show unusual patterns of differentiation between populations. While this idea has a long history (Lewontin & Krakauer 1973), recent years have seen an explosion of methods and approaches for detecting so-called outlier loci. At their heart, many of these methods test for loci that show extreme patterns of variation or differentiation relative to either a neutral standard, or the genome-wide empirical distribution. The statistical and inference challenges of these models have recently begun to be explored (Lotterhos & Whitlock 2014). Lotterhos *et al.* (2017) provide clear methods for combining data across multiple outlier locus methods, either by combining the *P*-values from multiple tests or by combining the signal into a multivariate test for differentiation. Using a combination of simulations and analysis of empirical datasets, they provide much needed guidance for empiricists working in this area.

Villanueva-Cañas *et al.* (2017) approach the question of how selection acts on mobile genetic elements, such as transposable elements (TEs). They review a variety of approaches and data for detecting selection on TEs, many of which are familiar to those seeking to detect selection on “traditional” loci—analyses of polymorphism at linked sites, environmental associations, and functional assays—but which have unique challenges when applied to TEs. They highlight how improved sequencing techniques so-called high throughput that provide longer sequencing reads promise to fundamentally change our ability to detect selection on TEs.

These three papers highlight the diverse ways that evolutionary biologists and population geneticists have developed to

probe the effects of natural selection on the genome. They also highlight the unresolved aspects of this question, and point to new areas of investigation. As sequencing technologies and methods develop, we expect an exciting portrait of selective sweeps, genetic differentiation, and selection on genomic elements will emerge from multiple species and study systems.

Genetics and ecology of local adaptation

Given differential natural selection between populations, phenotypes and alleles affecting those phenotypes are expected to become differentiated between populations. How much do we know about the underlying loci or agents of selection? Wadgymar *et al.* (2017) review what is known about the loci underlying local adaptation, and find that rather than exhibiting trade-offs across environments (alleles favoured in one environment are disfavoured in another), most cases to date are of conditional neutrality, with alleles only affecting phenotypes in one setting. Rather shockingly, they also find that there have been remarkably few manipulative experiments performed to validate the proposed selective agent. They close by outlining how genome-wide association studies or pedigreed populations can be combined with experimental studies to validate both selective agents and the loci underlying adaptation.

Looking forward

T.H. Huxley's famous response to the idea of natural selection was “How extremely stupid not to have thought of that!” At the time, neither he nor Darwin could have foreseen the vast array of biological phenomena that could be studied and understood through the lens of natural selection. As the papers in this issue show, there are many open avenues for investigating natural selection and its consequences. Two themes seem apparent to us. The first is the tremendous quantitative and statistical sophistication that is being brought to measuring selection on phenotypes and genomes. We clearly have come a long way from using merely an intuitive match between features of organisms and their environment to detect selection and adaptation (cf. Antonovics 1976). Second, we see—and welcome—an increasing integration of the ecological and genetic aspects of studying natural selection. In the most general sense, the papers in this issue deal with how we can detect selection in a way that can be used to predict evolutionary responses, how selection affects the genome, and how selection and genetics underlie adaptive differentiation. We expect that these challenges will remain at the forefront of evolutionary biology for years to come.

References

- Antonovics, J. (1976) The input from population genetics: “the new ecological genetics”. *Systematic Botany*, **1**, 233–245.
- Arnold, S.J. (1983) Morphology, performance and fitness. *American Zoologist*, **23**, 347–361.
- Franklin, O.D. & Morrissey, M.B. (2017) Inference of selection gradients using performance measures as fitness proxies. *Methods in Ecology and Evolution*, **8**, 663–677.

- Hadfield, J.D. & Thomson, C.E. (2017) Interpreting selection when individuals interact. *Methods in Ecology and Evolution*, **8**, 688–699.
- Hamilton, W.D. (1964) The genetical evolution of social behaviour. II. *Journal of Theoretical Biology*, **7**, 17–52.
- Hermisson, J. & Pennings, P.S. (2017) Soft sweeps and beyond: understanding the patterns and probabilities of selection footprints under rapid adaptation. *Methods in Ecology and Evolution*, **8**, 700–716.
- Kingsolver, J.G., Diamond, S.E., Siepielski, A.M. & Carlson, S.M. (2012) Synthetic analyses of phenotypic selection in natural populations: lessons, limitations and future directions. *Evolutionary Ecology*, **26**, 1101–1118.
- Lande, R. & Arnold, S.J. (1983) The measurement of selection on correlated characters. *Evolution*, **37**, 1210–1226.
- Lewontin, R.C. & Krakauer, J. (1973) Distribution of gene frequency as a test of theory of selective neutrality of polymorphisms. *Genetics*, **74**, 175–195.
- Lotterhos, K.E., Card, D.C., Schaal, S.M., Wang, L., Collins, C. & Verity, R. (2017) Composite measures of selection can improve the signal-to-noise ratio in genome scans. *Methods in Ecology and Evolution*, **8**, 717–727.
- Lotterhos, K.E. & Whitlock, M.C. (2014) Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Molecular Ecology*, **23**, 2178–2192.
- Price, G.R. (1970) Selection and covariance. *Nature*, **227**, 520–521.
- Robertson, A. (1966) A mathematical model of the culling process in dairy cattle. *Animal Production*, **8**, 95–108.
- Thomson, C.E. & Hadfield, J.D. (2017) Measuring selection when parents and offspring interact. *Methods in Ecology and Evolution*, **8**, 678–687.
- Villanueva-Cañas, J.L., Rech, G.E., de Cara, M.A.R. & Gonzalez, J. (2017) Beyond SNPS: how to detect selection on transposable element insertions. *Methods in Ecology and Evolution*, **8**, 728–737.
- Wadgymar, S., Lowry, D.B., Gould, B.A., Byron, C.N., Mactavish, R. & Anderson, J.T. (2017) Identifying the targets and agents of selection: innovative methods to evaluate the processes that contribute to local adaptation. *Methods in Ecology and Evolution*, **8**, 738–749.

Received 9 May 2017; accepted 10 May 2017

Handling Editor: Robert Freckleton