Molecular Phylogenetics and Evolution xxx (2013) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Molecular Phylogenetics and Evolution

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

Review

Integrating phylogenetics, phylogeography and population genetics through genomes and evolutionary theory

Asher D. Cutter *

Department of Ecology & Evolutionary Biology, University of Toronto, 25 Willcocks St., Toronto, ON M5S 3B2, Canada

ARTICLE INFO

Article history:
Received 6 April 2013
Revised 6 June 2013
Accepted 12 June 2013
Available online xxxx

Keywords:
Phylogeography
Species trees
Speciation
Genome evolution
Coalescent theory

ABSTRACT

Evolutionary theory is primed to synthesize microevolutionary processes with macroevolutionary divergence by taking advantage of multilocus multispecies genomic data in the molecular evolutionary analysis of biodiversity. While coalescent theory bridges across timescales to facilitate this integration, it is important to appreciate the assumptions, caveats, and recent theoretical advances so as to most effectively exploit genomic analysis. Here I outline the connections between population processes and phylogeny, with special attention to how genomic features play into underlying predictions. I discuss empirical and theoretical complications, and solutions, relating to recombination and multifurcating genealogical processes, predictions about how genome structure affects gene tree heterogeneity, and practical choices in genome sequencing and analysis. I illustrate the conceptual implications and practical benefits of how genomic features generate predictable patterns of discordance of gene trees and species trees along genomes, for example, as a consequence of how regions of low recombination and sex linkage interact with natural selection and with the accumulation of reproductive incompatibilities in speciation. Moreover, treating population genetic parameters as characters to be mapped onto phylogenies offers a new way to understand the evolutionary drivers of diversity within and differentiation between populations. Despite a number of challenges conferred by genomic information, the melding of phylogenetics, phylogeography and population genetics into integrative molecular evolution is poised to improve our understanding of biodiversity at all levels.

Contents

---

* Fax: +1 416 978 5878.
E-mail address: asher.cutter@utoronto.ca

1055-7903/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ympev.2013.06.006

Please cite this article in press as: Cutter, A.D. Integrating phylogenetics, phylogeography and population genetics through genomes and evolutionary theory. Mol. Phylogenet. Evol. (2013), http://dx.doi.org/10.1016/j.ympev.2013.06.006

---

1. Introduction

Students of molecular phylogenetics, phylogeography and population genetics are broadly interested in understanding the same thing: evolutionary change of genomes and the organisms that host them. Where they differ is in emphasis, whether trying to draw inference about biodiversity from the interspecies genealogical histories of organisms, explicitly seeking to explain the demographic history of populations, or to characterize how natural selection gets recorded in DNA sequences – despite the evolutionary intercalation of phylogeny, demographic history, and natural selection. The generality of their common dependence on molecular sequence data provides a clear means of unifying phylogenetics, phylogeography and population genetics in a view of integrative molecular evolution. Recent and ongoing attempts have partially harmonized these subdisciplines (Edwards, 2009; Knowles, 2009). With renewed urgency, technical advances in DNA sequencing bring this integration to the fore (Carstens et al., 2012; McCormack et al., 2013): it is clear that 'phylogenomics' and 'population genomics' are the new standard for analysis. This is well exemplified by the impressive integrative work on our own species into how selection, demography and speciation history all interact throughout human and primate genomes (McVicker et al., 2009; Campbell and Tishkoff, 2010; The 1000 Genomes Project Consortium, 2012). The infiltration of genomics into these disciplines means that we can accelerate our understanding of biodiversity from micro- to macro-evolutionary scales. The genomic data itself provides the raw material to do so, exposing the predictable features of genome structure that must be accounted for, ripe for exploitation by researchers, provided that we more fully appreciate and incorporate the empirical and theoretical complexities that genomes help reveal.

My aim here is to raise attention to the key conceptual molecular evolutionary issues that bridge across phylogenetics, phylogeography and population genetics, with particular emphasis on recent progress in molecular population genetics and genomics that will prove valuable in the analysis and interpretation of phylogeography and phylogeny. In this review, I begin by introducing the special importance of gene tree heterogeneity and non-genealogical summaries to genome-scale studies (Section 2). I then describe how genome structure varies in predictable ways to affect gene tree heterogeneity, and its implications and potential to be exploited for understanding evolutionary history (Section 3). This is followed by empirical and theoretical caveats about genealogical coalescence, with emphasis on recent advances (Section 4), and about how to integrate evolutionary views of time across studies of differing depths of divergence (Section 5). I then outline practical genomic approaches to addressing problems in phylogenetics, phylogeography and population genetics (Section 6) before offering a view to ways in which integrative genomic analysis can span these subdisciplines (Section 7), with some cautions and areas of promise for future discoveries (Section 8).

2. Gene genealogies and gene summaries

The importance of considering many loci in phylogenetics, phylogeography and population genetics is now well-accepted. But there are different perspectives on how to integrate multi-locus data for inference. Gene trees provide a superb visual and quantitative way to consider the evolutionary process, but genome-scale data shines a bright light on some inherent, well-known challenges to this way of summarizing the evolution of populations. Non-genealogical summaries of molecular evolution provide complementary and alternative methods for some applications. In this section, to set the stage for more detailed issues, I outline some fundamental attributes of using gene trees and non-genealogical population genetic metrics for evaluating the many trajectories of evolution that get recorded across the genome.

2.1. Gene tree heterogeneity

The independent genealogical realizations of distinct loci provides a powerful way of determining the phylogeny of species relationships even when the genealogies of many or most loci do not reflect the true branching of speciation events in history (Edwards et al., 2007; Degnan and Rosenberg, 2009; Kubatko et al., 2009; Salichos and Rokas, 2013). In the face of speciation events clustered close in time or in the very recent past, different genes can differ in branch lengths or topology owing to heterogeneity in how or if allelic variation of a given locus in the species' common ancestor passes to the descendant lineages (Pamilo and Nei, 1988; Maddison, 1997), creating so-called 'anomalous gene trees' (Degnan and Rosenberg, 2006; Degnan and Rosenberg, 2009). The prevalence of gene tree heterogeneity depends on the time between speciation events relative to ancestral population sizes, with large ancestral population sizes and gene flow between incipient species near the time of ancestral nodes exacerbating such effects (Kubatko and Degnan, 2007; Eckert and Carstens, 2008).

Incorporating this heterogeneity in coalescence among loci can remove misleading inferences that could result from concatenation of loci and helps reveal the true 'species tree' in the face of both recent and ancient radiations (Knowles and Maddison, 2002; Edwards et al., 2007; Kubatko and Degnan, 2007; Edwards, 2009; Salichos and Rokas, 2013). This requires collecting data for many loci from multiple individuals in each species (or population), ideally, weighted by information content (Heled and Drummond, 2010; Lanier and Knowles, 2012; Salichos and Rokas, 2013). Just as coalescent theory has proved powerful in understanding evolution within populations (Hein et al., 2004; Wakeley, 2009), the 'multispecies coalescent' process has emerged as an important paradigm in understanding divergence between species (Degnan and Rosenberg, 2009; Edwards, 2009). This view simply makes explicit the idea that genealogies are a property of evolving populations that, owing to the stochasticity of genetic drift and recombination among their constituent individuals, can differ from locus to locus despite sharing an identical population history of speciation. This