BioOne COMPLETE

Next-generation sequencing and the expanding domain of phylogeography

Authors: Edwards, Scott V., Shultz, Allison J., and Campbell-Staton, Shane C.

Source: Folia Zoologica, 64(3) : 187-206

Published By: Institute of Vertebrate Biology, Czech Academy of Sciences

URL: https://doi.org/10.25225/fozo.v64.i3.a2.2015

BioOne Complete (complete.BioOne.org) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

Your use of this PDF, the BioOne Complete website, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at www.bioone.org/terms-of-use.

Usage of BioOne Complete content is strictly limited to personal, educational, and non - commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

BioOne sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.

Downloaded From: https://bioone.org/journals/Folia-Zoologica on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

Folia Zool. – 64 (3): 187–206 (2015)

Next-generation sequencing and the expanding domain of phylogeography

Scott V. EDWARDS*, Allison J. SHULTZ and Shane C. CAMPBELL-STATON

Department of Organismic and Evolutionary Biology, and Museum of Comparative Zoology, Harvard University, Cambridge, MA 02138, U.S.A.; e-mail: sedwards@fas.harvard.edu

Received 17 March 2015; Accepted 26 June 2015

Abstract. Phylogeography is experiencing a revolution brought on by next-generation sequencing methods. A historical survey of the phylogeographic literature suggests that phylogeography typically incorporates new questions, expanding on its classical domain, when new technologies offer novel or increased numbers of molecular markers. A variety of methods for subsampling genomic variation, including restriction site associated DNA sequencing (Rad-seq) and other next generation approaches, are proving exceptionally useful in helping to define major phylogeographic lineages within species as well as details of historical demography. Next-generation methods are also blurring the edges of phylogeography and related fields such as association mapping of loci under selection, and the emerging paradigm is one of simultaneously inferring both population history across geography and genomic targets of selection. However, recent examples, including some from our lab on *Anolis* lizards and songbirds, suggest that genome subsampling methods, while extremely powerful for the classical goals of phylogeography, may fail to allow phylogeography to fully achieve the goals of this new, expanded domain. Specifically, if genome-wide linkage disequilibrium is low, as is the case in many species with large population sizes, most genome subsampling methods will not sample densely enough to detect selected variants, or variants closely linked to them. We suggest that whole-genome resequencing methods will be essential for allowing phylogeographers to robustly identify loci involved in phenotypic divergence and speciation, while at the same time allowing free choice of molecular markers and further resolution of the demographic history of species.

Key words: Rad-seq, genotype-by-sequencing, bird, association mapping, natural selection, demographic history

Introduction

Like many fields in evolutionary biology, phylogeography has been consistently transformed by available technologies for assaying genetic variation. Indeed, the various approaches for measuring genetic variation have waxed and waned as available technologies come and go. A determinist viewpoint might suggest that major trends in marker use over the decades have been driven largely by available technologies, such as PCR and next-generation sequencing (reviewed in Brito & Edwards 2009). However, a more "free will" perspective might suggest that marker choice is driven also by the conceptual needs of the discipline and a rallying of the field behind core concepts that argue for one set of markers over another. What seems clear is that, as technologies available to phylogeographers have changed, the borders of the discipline – the types of questions and hypotheses that are tackled – also changed. In particular, phylogeography now appears to regularly pose questions that were traditionally the domain of its closely allied sister discipline, population genetics, and increasingly in the domain of finding the loci responsible for phenotypic variation in natural populations. These new questions themselves can also drive the shape of the field and the tools that phylogeographers adopt in order to answer them. In this essay we explore the history and future of phylogeography through this lens of changing technologies and questions. We suggest that the domains of phylogeography have expanded to include surveys of selection and covariation of genes and environment across the landscape as a result of the increasing ability to assay variation at large numbers of loci through next-generation sequencing. We suggest that trends in marker use imply both a degree of technological determinism as well as shifts in free choice of markers over time. Like other recent reviews, we foresee a day when a greatly expanded toolkit of markers and phylogeographic questions will be readily available through routine whole-genome resequencing of natural populations sampled across geographic space.

Molecular markers and the core concepts of phylogeography

In the 1990s, the polymerase chain reaction (PCR) was the primary driver, and this led to a proliferation

* *Corresponding Author*

Downloaded From: https://bioone.org/journals/Folia-Zoologica on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

of studies focusing on mitochondrial DNA in animals, and on chloroplast DNA in plants. The focus on organellar genomes was not necessarily prescribed by PCR, but the ease of amplification using PCR and the challenges of working routinely with nuclear genes perceived by phylogeographers made organelle genomes a practical focus. MtDNA and cpDNA were also attractive because they were polymorphic at the intraspecific level and experienced little if any recombination, making it straightforward to move from DNA sequence to gene tree without additional data manipulation. With this focus on organellar genomes came the oft-repeated caveats now familiar to every student of phylogeography (Edwards & Bensch 2009): that differences in gene flow between the sexes might cause organelle genomes to recover a biased history of a given species; that the smaller effective population size of organellar genomes might cause the genetic lineages to track population lineages more faithfully than the average marker, sometimes resulting in an overly simplistic view of population history; and, more recently, that natural selection on organelle genomes, particularly mtDNA, might yield estimates of genetic diversity, or spatial patterns in the distribution of that diversity, which do not reflect neutral processes or recent population history (Rand 2001, Excoffier & Ray 2008, Nabholz et al. 2008, 2009). Rather, lineage-specific mutation rates, which are hard to predict for a given species from first principles or life history parameters, appear to be the best predictor of variation in genetic diversity across species, at least in birds and mammals. Natural selection has also emerged as a key determinant of mitochondrial diversity within species, as well as of relationships between mtDNA haplotype or codon distributions and latitude or thermal environment, potentially compromising efforts to understand neutral diversity via neutral markers (Gerber et al. 2001, Ballard & Whitlock 2004, Ribeiro et al. 2011, Jobling 2012, Ballard & Pichaud 2014, Morales et al. 2015). Although genetic diversity within populations and species is not always a primary focus of phylogeographic studies, it is certainly a basic descriptor of population history and the forces governing it warrant our attention.

The first forays in to the nuclear genome in animal phylogeography came in the early 1990s in the form of PCR-amplified nuclear DNA sequences and via microsatellites. Diploid nuclear genes were typically amplified via PCR and sequenced directly (i.e. without cloning first), a practice that led to much hand-wringing over how to determine the phase of nuclear haplotypes comprising the PCR product from heterozygous individuals (Palumbi & Baker 1994, Hare & Palumbi 1999). The phase of nuclear alleles is important because only after determining phase is one able to coherently analyze alleles within populations or linked sites, even for PCR products of a few hundred base pairs. Even today, particularly when PCR-amplified nuclear genes are used in phylogenetics, the phase of nuclear alleles is often ignored, possibly because it is not deemed important when comparing highly divergent species. Recombination also had to be acknowledged and often this was accomplished by determining DNA tracts within which no detectable recombination was observed. Detection of recombination events was often accomplished through software focusing on phylogenetic discordances among sites within a sequence or by estimating linkage disequilibrium among sites within or between loci (Hudson & Kaplan 1988). Testing for and dealing with recombination, for example by retaining only sections of an alignment free of detectable recombination, was and continues to be important so as to ensure assumptions are not violated when building gene trees or when estimating population parameters that require either full linkage between sites within loci or complete independence of sites.

The specific ways in which sequence data were analyzed were very much constrained by the technical limits of PCR; typically data sets consisted of a few loci, each of a few hundreds base pairs, and an ecosystem of software emerged around this particular format. Discordances between nuclear and mitochondrial DNA came to the fore as researchers were able to directly observe them in phylogenetic and phylogeographic analyses (e.g. Godinho et al. 2008), and in particular in studies of hybrid zones. The effects of incomplete lineage sorting were also readily visible even in analyses employing few loci. PCR of nuclear genes was very much a brute-force operation, the number of loci being assayed directly proportional to the amount of effort and number of PCR experiments performed. At the zenith of the (PCR) nuclear age in phylogeography (Balakrishnan et al. 2010), typical studies included tens of loci, and there was mounting evidence that the uncertainty of estimates of demographic parameters decreased with increasing numbers of loci. In addition, population genetic theory suggested much the same: for example, the optimal sampling scheme for estimating genetic diversity within single populations is generally thought to maximize the number of loci at the expense

Downloaded From: https://bioone.org/journals/Folia-Zoologica on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

of the number of individuals (alleles) or length of loci (Nei & Roychoudhury 1974, Pluzhnikov & Donnelly 1996, Felsenstein 2006, Carling & Brumfield 2007). Surprisingly, the optimal sampling scheme for estimating genetic parameters from multiple populations partially linked by gene flow is still understudied. In the PCR era, given constraints on budgets and finite ability to sample individuals, there was a trade-off between the number of individuals or populations sampled and the number of loci assayed. We suggest that this trade-off is partially removed with the advent of next-generation sequencing.

Microsatellites and other simple sequence repeats also emerged in phylogeography in 1990s, following their discovery in the previous decade (Tautz & Renz 1984, Jeffreys et al. 1985). Although first employed extensively in the study of parentage in natural populations (Burke et al. 1989, Gyllensten et al. 1990), understandably these markers swept like wildfire through phylogeography. Indeed, despite the fact that microsatellites fail to capture critical components of the original spirit of phylogeography – in particular phylogeography's focus on phylogenetic lineages – they have historically been the most extensively used molecular marker in phylogeography. Their popularity is understandable because of their hypervariability – it is easy to be seduced by markers with such a large number of alleles and such potentially high resolving power. On the positive side, microsatellites have provided unquestionable insight into the demographic histories of literally thousands of species, and have helped expand phylogeography to incorporate and synergize with sister disciplines such as population genetics, landscape genetics, and even behavioral ecology. They also carry some information on the relationships among alleles – assuming a step-wise mutation model – and in principle have the ability to distinguish between recent and more ancient timescales of population divergence (e.g. F_{ST} *vs.* R_{ST} comparisons). On this logic, some authors have suggested they may be useful for estimating divergence times as well (Sun et al. 2009). On the down side, several authors have called for a reappraisal of the utility and neutrality assumptions of microsatellites and have questioned the high degree of enthusiasm for these markers in phylogeography (Brumfield et al. 2003, Morin et al. 2004, Zink & Barrowclough 2008, Edwards & Bensch 2009, Zink 2010). As put by Morin et al. (2004) "… the high information content (of microsatellites), a result of high mutation rates, comes at a price…". The challenges and deficiencies of microsatellites in phylogeography have been reviewed extensively elsewhere (Zink 2010, Albayrak et al. 2011, Perktaş et al. 2015), and include substantial homoplasy, making estimates of the number of mutations difficult; an inability to conduct robust phylogenetic analyses, and hence offering little continuity between phylogeography and phylogenetics; frequent null alleles; and difficulty comparing to sequence-based markers, including mtDNA. Less well-appreciated deficiencies of microsatellites include clear evidence that some simple sequence repeats are indeed functional – they are often involved in gene regulation in both microbes and eukaryotic genomes – and thus may be subject to selection. This last critique no doubt applies to other kinds of markers as well, including SNPs, but the frequent appeal to neutrality by users of microsatellites should be tempered by the increasing number of examples of functional roles for such markers (Liu et al. 2000, Metzgar et al. 2000, Sureshkumar et al. 2009, Tremblay et al. 2010, Grover & Sharma 2011, Gao et al. 2013).

There are no doubt still staunch defenders of microsatellites, and we do not mean to suggest that SNPs, sequence-based markers, or other alternatives to microsatellites are not above reproach. A major criticism of sequence based markers or SNPs in phylogeography has been the paucity of such markers and their low polymorphism. While these criticisms may have been valid in the PCR era. We suggest that they no longer apply meaningfully given the large number of SNPs now achievable with next-generation sequencing approaches. By contrast, although the number of microsatellite loci has been increasing in recent years, we do not know of efforts to assay variation targeted at microsatellites using next-generation approaches. Next-generation isolation of microsatellite loci, followed by PCR-based assays of variation, has been used with considerable success (Abdelkrim et al. 2009, Perry & Rowe 2011, Singham et al. 2012, Curto et al. 2013, Taguchi et al. 2013), but actually assaying variation and scaling up beyond PCR-based assays to our knowledge has only recently been developed for microsatellites in studies of phylogeography (but see Fordyce et al. 2011, Raposo et al. 2015, Suez et al. 2015). Although it is surely too early to tell, we suggest that this technical gap implies that the community does not place a high priority on scaling up for microsatellites, perhaps because it is thus far comfortable with the expanded power of SNPs in the next-generation sequencing era. Garrick et al. (2015) recently declared that "Compared to other classes of molecular markers, DNA sequence haplotypes and single nucleotide polymorphisms

Downloaded From: https://bioone.org/journals/Folia-Zoologica on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

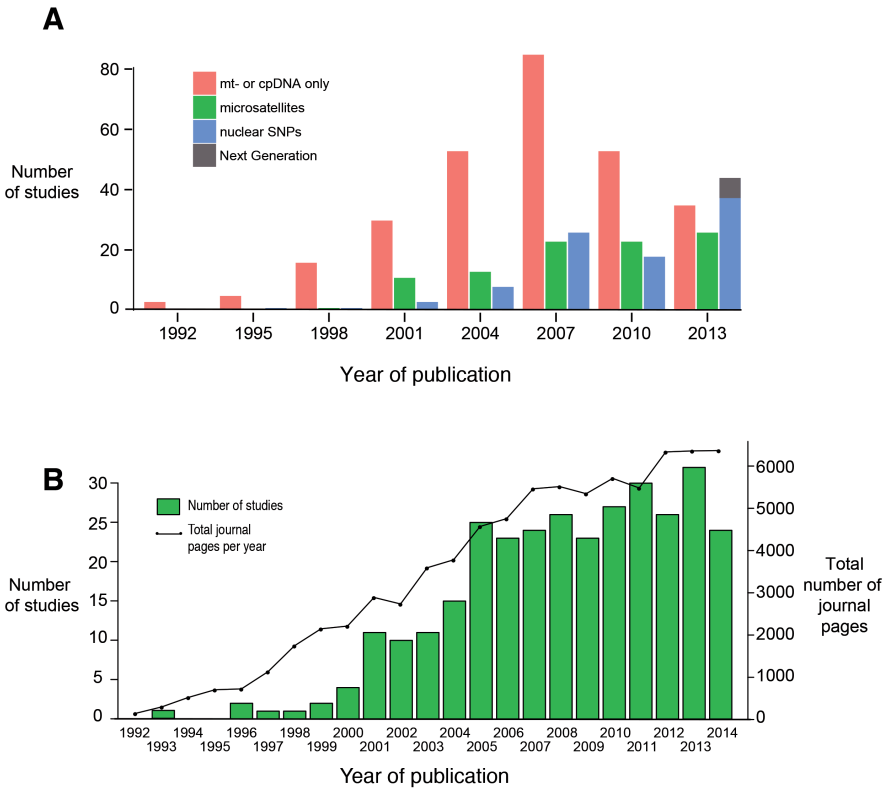


Fig. 1. Trends in the use of molecular markers over time from a survey of articles on phylogeography from the journal Molecular Ecology. (A) Plot of the number of articles from 1992 to 2013 using organelle DNA markers, microsatellites, nuclear sequence-based markers, coded according to the key at upper left. The category of nuclear sequence-based markers includes eight studies in 2013 using Rad-Seq, sequence capture, or another next-generation sequencing technique to genotype SNPs; the remainder of nuclear SNP studies used PCR approaches. Data on articles using nuclear SNPs and organelle markers are taken from Garrick et al. (2015) and are presented for every three years for easy comparison with that study. Overall, during the time period sampled there were 280 studies using only SNPs from mt- or cpDNA, 101 studies using nuclear SNPs, and 97 studies using nuclear microsatellites. There were 13 studies that included both nuclear SNPs and microsatellites, these were excluded from the plot. Sixty-three of the microsatellite studies and 79 of the nuclear SNP studies also included mt- or cpDNA SNPs and are included. (B) Trends in the total number of articles from Molecular Ecology using nuclear microsatellites sampled every year from 1992 to 2014. These 318 studies include those using other types of data such as mt- or cpDNA SNPs or nuclear SNPs. The number of pages in the journal *Molecular Ecology* per year is shown in a black line as a gauge on growth of the journal as a whole. The full list of articles using microsatellites can be found in the supplementary material.

(SNPs) should be more informative about historical events and processes … operating over timescales most relevant to the discipline (of phylogeography)" (Garrick et al. 2015). While we agree wholeheartedly with this statement, we suggest that much of the community might still favor microsatellites if given the choice, particularly in comparisons of closely related, endangered or very recently diverged populations. This preference, we suspect, is due in part because some labs may not yet have ready access to next-generation sequencing methods. But it might also be due to the perception that, due to their hypervariability, microsatellites have advantages over SNPs in many contexts, especially if they can be assayed in large numbers (Becquet et al. 2007, Kwong & Pemberton 2014).

We sought to determine whether changes in methodologies and markers used in phylogeography have been driven by choice or instead more by the availability of technologies adopted primarily for increasing the number of loci. We conducted a study parallel to that of Garrick et al. (2015) by reading abstracts for 397 papers that use microsatellites in phylogeography and were published in Molecular Ecology, a major outlet for phylogeographic research (Fig. 1, see legend for methods). Garrick et al. (2015)'s survey, which comprised 370 papers reporting on 508 single-species data sets, was interesting because it somewhat unexpectedly focused on SNPs, whereas our intuition was that, among nuclear loci, microsatellites were the main driver of phylogeography until recently. Our analysis confirms this suspicion: once the studies that only used mitochondrial DNA were pulled out from their analysis (a total of 280 studies, comprising 73.5 % of all "SNP" studies in our sample; see Fig. 1), the number of phylogeographic studies using microsats is comparable to, and sometimes exceeds, that using nuclear SNPs (Fig. 1A). Intriguingly,

Downloaded From: https://bioone.org/journals/Folia-Zoologica on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

phylogeographic studies employing only mtDNA do indeed seem to be declining since 2007, at least in the pages of Molecular Ecology. This decline could highlight a shift in preference within the field towards other types of genetic data or a shift in preference of journals against publishing studies that rely solely on mtDNA. Additionally, the year 2013 suggests a shift as studies employing nuclear SNPs begin to exceed those using microsatellites. This uptick does not seem to be driven entirely by next-generation sequencing, which only comprised eight studies in our sample, suggesting that SNPs may have risen in popularity independently of novel technologies and perhaps due to conceptual advances or available software. Examination of microsatellite studies in all years of our sample (Fig. 1B) suggest that, at best, this technology has leveled out in its popularity, particularly given the increasing number of pages in the journal over time. It will be interesting to see what the next five years brings in terms of the relative use of these various marker types in phylogeography. Given the pre-eminence of Molecular Ecology in the field of phylogeography, we suggest that the trends observed here may well reveal trends that will follow in time with the rest of the field.

Next-generation sequencing and the rise of sequence-based markers in phylogeography

Phylogeographers have appreciated for years that, despite their lower polymorphism, SNPs are much more common in the genome than microsatellites (Brumfield et al. 2003). Yet this point was almost moot because it was difficult if not impossible to take advantage of SNPs on a scale that would capitalize on their ubiquity. The advent of next-generation sequencing will likely increase the swing of the phylogeographic pendulum back in favor of SNPs and sequence-based markers once and for all. The adoption of partial genome survey methods such as Rad-seq will not only yield SNPs in sufficient numbers for phylogeography, but will prescribe the use of SNPs even more so than will whole-genome resequencing. Whole-genome phylogeographic studies are already the norm for model species such as humans (Altshuler et al. 2010, Reich et al. 2010, Hammer et al. 2011, Li & Durbin 2011, Stoneking & Krause 2011) and *Drosophila* (Yukilevich et al. 2010, Campo et al. 2013, Duchen et al. 2013, Reinhardt et al. 2014) and researchers will have many options for marker types once this phase is achieved. Until that time, by shifting the focus of phylogeography to sequence-based markers and SNPs, next-generation sequencing methods promise to stabilize and unify phylogeographic studies in many productive ways. To us they are a positive trend for phylogeography because of the many reasons that SNPs have previously been considered beneficial: they provide more natural comparisons to variation in organelle genomes and between studies, and, despite the challenges of recombination within loci, provide natural bridges to phylogenetic analysis.

Types and consequences of next-generation sequencing approaches in phylogeography

Aside from the use of next-generation sequencing approaches for isolating microsatellite loci, next-generation sequencing is making inroads into phylogeography in two main ways: through Rad-seq, which generates short (~100 bp) markers, typically with one or a few SNPs per locus (see Puritz et al. 2014 for a review of different Rad-seq methods); and through targeted capture approaches, which can be used to target already-defined sets of loci, such as exons or ultraconserved elements (UCEs) and their polymorphic flanking regions (Faircloth et al. 2012, Smith et al. 2014). Although transcriptome and amplicon sequencing have also both proven useful in phylogeography (Hedin et al. 2012, O'Neill et al. 2013), transcriptome sequencing will likely have less direct use in purely phylogeographic investigations (as opposed to the discovery of loci under selection; see below) because of its focus on loci that are relatively conserved but more likely under selection, and we predict that amplicon sequencing will ultimately prove less attractive to phylogeographers because of the labor involved and the smaller number of loci that can be assayed (but see McCormack & Faircloth 2013).

The emerging "core" approaches of targeted enrichment and Rad-seq promise to re-orient phylogeography towards sequence-based markers in different ways because of the types of data they each produce (Lemmon & Lemmon 2012, McCormack et al. 2012, 2013). Targeted enrichment approaches yield data that can be assembled into individual sequence-based markers spanning hundreds to potentially thousands of base pairs, resulting in haplotypes or consensus sequences within which there may be several to many SNPs that can in principle be subjected to phylogenetic analysis (Lemmon & Lemmon 2013). By contrast, Rad-seq typically yields loci that are too short to analyze using traditional phylogenetic methods; instead, researchers typically extract single or multiple SNPs from such Rad-loci and then analyze them as individual SNPs. In many ways the two approaches provide contrasting bridges to phylogenetics and

Downloaded From: https://bioone.org/journals/Folia-Zoologica on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

classical phylogeography, as well as pointing to complementary analytical approaches in the future. For example, because the loci yielded by targeted enrichment approaches to phylogeography can often be analyzed using standard phylogenetic methods for estimating gene trees, they provide a natural bridge to classical phylogeography. By contrast, although the SNPs generated by Rad-seq can be used to estimate phylogenetic relationships of populations or species ("species trees"), and indeed have been subjected to concatenation approaches in early examples (Emerson et al. 2010, Merz et al. 2013), currently these markers are used to bypass classical gene trees and instead estimate the species tree directly (Bryant et al. 2012, Chifman & Kubatko 2014, Rheindt et al. 2014). These two approaches can sometimes require different sets of analyses, and it may be that the toolkit for linked SNPs such as produced by targeted enrichment is still deeper than that available for analyzing SNPs.

Although both core methods will align phylogeography squarely on the use of SNPs, whether linked or unlinked in individual loci, these differences in continuity with classical "gene tree" phylogeography and analytical approaches are significant. Gene trees may be the lynchpin in this phylogeographic transition. Many have suggested that, despite their centrality to the origins of phylogeography (Avise et al. 1987), ultimately, gene trees are a nuisance parameter in phylogeography and, if anything, can be a distraction from the key levels of analysis and primary interests, which are populations and species, not genes. In this sense, Rad-seq may have the practical advantage of finally freeing the community conceptually from gene trees and haplotype networks, which are still a ubiquitous component of phylogeography. The ability and tendency to make and interpret gene trees has resulted in heated controversies in phylogeography, such as the conflicts between model-based approaches in phylogeography and more literal interpretations of gene trees, such promoted by nested-clade analysis (Nielsen & Beaumont 2009, Beaumont et al. 2010, Templeton 2010). It will be interesting to see how the analytical methods afforded by Rad-seq versus targeted enrichment influence the next ten years of phylogeography. It may be that the sheer number of loci generated by both methods moves the field forward in adopting the model-based approaches that are clearly appropriate for such data sets.

Power of next-generation genome subsampling methods for phylogeography

Genome subsampling methods such as Rad-seq and targeted capture re-sequencing, including UCE analysis, hold enormous promise for phylogeographic investigations of neutral processes such as population structure, species delimitation and historical demography. Phylogeography has traditionally emphasized sampling of individuals and populations over loci (Brito & Edwards 2009, Garrick et al. 2015). This bias is a natural and understandable outgrowth of one of the main motivations for phylogeography – to discover novel lineages of organismal biodiversity within species. However, it is now better appreciated that sampling robustly for loci as well as individuals is essential for increased precision of parameter estimates in phylogeography and for better accounting for stochastic variation among loci (Beerli & Felsenstein 1999, Edwards & Beerli 2000, Jennings & Edwards 2005, Felsenstein 2006, Carling & Brumfield 2007). The sheer number of loci revealed by methods such as Rad-seq, UCE analysis or targeted capture methods therefore comes as a welcomed boon over the relative dearth of loci captured in a typical PCR-based study. Considering one of the core goals of phylogeography is to describe the history of populations using genomic data, it should not be surprising that even a dozen loci (which is typical for the heyday of PCR-based phylogeography) are unlikely to capture the diverse signals of history across all chromosomes in a typical genome. In the case of Rad-seq, there is a serious issue involving the bias of the technique against older haplotypes that have experienced mutations in the restriction sites used to isolate DNA fragments – a bias that can compromise estimates of genetic variation (Arnold et al. 2013). Furthermore, the process of assembling a library from non-model species often involves grouping sequence reads by some similarity threshold. The choice of these thresholds is not necessarily a straightforward process, and highly divergent alleles may be inadvertently omitted prior to downstream analyses either because their differences exceed preset similarity thresholds, or they increase the proportion of missing data in the genotype by individual matrix (Huang & Knowles 2014, Harvey et al. 2015a). Still, for other next-gen subsampling approaches, the variation revealed by next-generation approaches – whether amplicon sequencing of ~100 loci or the tens of thousands of SNPs that a typical Rad-seq study reveals – is likely more than adequate for understanding the basic population history of most species.

Harvey et al. (2015b) recently compared the resolving power for phylogeography of data sets varying in size in terms of number and length of sequence-based

Downloaded From: https://bioone.org/journals/Folia-Zoologica on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

markers to estimate demographic parameters (effective population size, divergence time, migration rate) and species history in a Neotropical songbird with deep phylogeographic breaks. They found that increasing the number of loci up to 5000 provided increased resolution of the particular demographic histories that they studied, but that increasing the number of loci beyond this yielded minimal gains. Additionally, they found that increasing locus length past 500 bp did not yield additional improvements in resolution for the focal parameters of their study. This study therefore suggested that the numbers of loci revealed by genome subsampling methods such as Rad-seq are likely to be adequate for resolving population history on a variety of scales. Indeed, the initial round of empirical studies using data sets produced by Rad-seq or sequence capture, usually on the order of 2000-30000 SNPs or aligned markers appear quite satisfying in so far as they have revealed undiscovered phylogeographic lineages that significantly improve our understanding of the fit of genomic variation to environmental and topographic barriers to gene flow (e.g. Alcaide et al. 2014, Harvey & Brumfield 2015). We concur with Harvey et al. (2015b) that genome subsampling methods are likely to finally provide the appropriate level of genomic detail for the foreseeable future of phylogeography.

Natural selection and the expanding domain of phylogeography enabled by next-generation sequencing

Classically, phylogeography has focused on the neutral demographic history of species, a goal that has been facilitated by studies on organelle genomes as well as multilocus analyses of nuclear genes. However, the ability to scan genomes for thousands of loci at a time has helped expand the purview of phylogeography beyond neutral demographic histories to include the discovery of loci under selection. As we have seen, with the advent of diverse types of nuclear markers including SNPs, phylogeography has relaxed its original focus on gene trees; purists may even argue that the original definition of phylogeography included a substantial focus on mitochondrial gene trees, and that use of nuclear markers with their frequent recombination might constitute an expansion of the original definition of phylogeography (Avise et al. 1987). In the same way, the ability to examine variation on a large scale and to focus on, for example, variation in transcriptomes and exomes (e.g. Marra et al. 2014) where natural selection is likely to be more prevalent, has allowed phylogeography to embrace topics such as natural selection. This shift, although fascinating in its own right, was arguably not a part of Avise's original vision for the field (Avise et al. 1987). However, the ability to examine large numbers of loci, and to estimate robust distributions of alleles and allele frequencies across geography and the genome, immediately raises the possibility of phylogeography embracing studies of natural selection. Indeed, some of the most integrative studies in phylogeography thus far are those that combine robust geographic sampling and historical demographic inference with investigations of natural selection (Deagle et al. 2012, Jones et al. 2012a, b, Pearse et al. 2014, Schielzeth & Husby 2014, Wallberg et al. 2014). Have phylogeography and population genetics become synonymous? We suggest not. If there is any attribute that distinguishes phylogeography from population genetics it is robust geographic sampling of populations within a species; such sampling is arguably a hallmark of phylogeography, yet is often not required, or achieved, in even high quality studies of population genetics.

Lewontin's paradox and genetic variation within species

There is a long history of using population genetics to discover loci with a history of selection, beginning with Lewontin & Krakauer's (1973) observation that F_{ST} outliers could be useful in identifying loci under selection. The use of F_{ST} outliers has become common now, and, although there are caveats to the interpretation of high F_{ST} as an unambiguous signal of selection (Turner & Hahn 2010, Cruickshank & Hahn 2014), the ability now to study distributions of loci makes it a useful tool, especially when implemented with care and a consideration of the underlying demographic history (Johnston et al. 2014, Lotterhos & Whitlock 2014). Recent studies suggest that, in fact, avoiding natural selection entirely, even in the nuclear genome, may not be possible, whether studying whole-genome or transcriptome variation, because the imprint of selection may be much greater across the genome than originally envisioned by phylogeographers. For example, the overall level of diversity in the nuclear genome displayed in a species is often taken by phylogeographers to be a neutral indication of the historical effective population size, summarized for single populations by the equation $\pi = 4N\mu$. However, the small range of nuclear genetic diversities across species – usually considered to fall in a range of two orders of magnitude and often called "Lewontin's (1974) paradox" – has been a major

Downloaded From: https://bioone.org/journals/Folia-Zoologica on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

challenge for population geneticists and has profound implications for phylogeography as well. The paradox was a major impetus for the development of the nearly neutral theory, which placed emphasis on interactions between selection and drift and seemed to fit available data better than the strictly neutral theory (Ohta 1992, Ohta & Gillespie 1996). By postulating a nearly neutral zone in which the absolute value of the product Ns was substantially less than one, the nearly neutral theory – clearly a key departure from Kimura's strictly neutral theory (Kimura 1968, 1983) – was able to account in part for this paradox. Although there have been many estimates of the distribution of selection coefficients for key model species (Keightley & Eyre-Walker 2010), until recently there were few compelling data to really test these ideas across a wide range of species.

Two recent comparative studies of population genomics are relevant to Lewontin's paradox and have important implications for phylogeography. Corbett-Detig et al. (2015) conducted an exhaustive survey of genome-wide genetic variation in 40 eukaryotes and came to the conclusion that the small range of genetic diversities (π) among species could be explained in part by the greater ability of natural selection to reduce genetic variation in species with large population size. The implication of this paper is that for some species, particularly those with large populations, selection may depress the level of neutral variation at nearly every site in the genome, because selective sweeps are common and drift is relatively weak. This study has profound implications for phylogeographic studies of widespread species; remarkably, the frequent signals of natural selection that phylogeography has come to expect for mtDNA may also apply to the nuclear genome, particularly as it applies to overall diversity in species with large populations.

In another related study, Romiguier et al. (2014) recently surveyed population variation in transcriptomes of a variety of species across the tree of life. Like the study by Corbett-Detig et al. (2015) this study, although comprehensive, cannot be considered phylogeography because the main focus was not population history but the overall amount of variation. Whereas in some cases multiple populations per species were sampled and much of the genetic variation within each species may have been captured, in neither study was the general geographic sampling robust enough to be considered phylogeography. Romiguier et al. (2014) came to the startling conclusion that the amount of variation (π) in the transcriptome was best predicted by life history attributes and longevity, rather than by geographic range or other aspects of strictly neutral demography. Surprisingly their reasoning was largely based on a neutral argument: long-lived and other species with K-selected life history traits tend to be able to sustain smaller populations, and hence lower genetic diversities, than species with r-selected life history traits, because they live in more stable environments with fewer long-term perturbations. By contrast, r-selected species, which tend to have greater genetic diversity, possess the large populations that allow them to colonize and persist in unstable habitats. Although ecologists will likely find merit in this hypothesis, because it conceives of ecology and life history as the causal driving force behind population genetic variation, it is not very satisfying from a population genomics perspective. It is surprising that even negative selection on deleterious variation, traditionally considered a major force in regulating intraspecific genetic variation, especially in protein-coding regions, was only briefly mentioned as potentially important for explaining the positive correlation between life history and the ratio of nonsynonymous to synonymous nucleotide substitutions (d_n/d_s) within species. In this case, the smaller populations of long-lived K species result in higher d_n/d_s (driven largely by higher d_n) due to increased fixation of deleterious mutations in small populations, as envisioned by the nearly neutral theory (see also Weber et al. 2014). However, Romiguier et al. (2014) suggested that overall levels of synonymous substitution were largely driven by effective population size, which in turn was seen as a neutral consequence of life history variation. It will be important to verify the hypothesis of this work through genome-wide measurements of diversity as a part of detailed phylogeograhic analyses of diverse species.

Selection, recombination and hitchhiking in phylogeography

The above two studies provide an important contrast in how phylogeographers are beginning to think about genome-wide data. In particular, the issue of linkage disequilibrium (LD) and the potential for genome-wide hitchhiking has emerged as a key factor influencing patterns of variation in the era of whole-genome phylogeography. Aside from a few key emerging models for ecology and evolution, such as sticklebacks, honeybees and other groups (Jones et al. 2012a, Wallberg et al. 2014), the phylogeography of non-model species thus far has not dealt substantially with the effects of linkage on genomic variation across geographically sampled sets of populations.

Downloaded From: https://bioone.org/journals/Folia-Zoologica on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

In the PCR era, levels of LD were occasionally measured within or between the loci that could be assayed, often to confirm the independence of loci, but in general the data available to phylogeographers was not comprehensive enough to provide meaningful insight into the effects of hitchhiking on genome-wide variation. Levels of LD can be influenced by many factors, including neutral processes such as genetic drift and population bottlenecks; selective processes like selective sweeps and balancing selection; and genetic processes like rates of recombination and mutation (Slatkin 2008). The population recombination rate ($\rho = 4Nc$) has been measured in relatively few non-model species, whereas there are numerous estimates for populations of humans, mice or *Drosophila* (Smukowski & Noor 2011).

Population geneticists have been measuring LD for decades, and one can calculate the LD or r^2 value between any pair of markers, regardless of how densely the genome is sampled or whether the markers are on the same or different chromosomes. In the pre-genomic era, when LD values were calculated between loci in studies that only sparsely sampled the genome, the motivation was often to study the action of natural selection. But this was a very hit or miss endeavor: if the candidate genes between which LD was calculated were not involved in selective processes, the result was often underwhelming. By contrast, in the era of genomics, when the genome can be sampled much more densely, in principle one need not know the candidate genes under selection: LD can be exploited to discover loci that are the actual targets of selection without a priori suspicion that the loci measured are under selection themselves (Slatkin 2008). Correlated patterns variation across the genome due to hitchhiking have been used increasingly to produce a set of candidate genes responding to selection, without those candidates having been assayed directly. The candidate genes are usually physically close to or in the same linkage blocks as, the markers directly measured, and the assumption is that hitchhiking on the actual targets is causing departures from neutrality, such as high F_{ST}, in the assayed SNPs. This protocol, variably called selection mapping or association mapping, has resulted in the generation of hundreds of candidate genes in emerging model species that may be responding to environmental or other selective pressures (Hohenlohe et al. 2012). Increasingly studies are also taking advantage of the high LD created when two species or populations hybridize: chromosomal blocks deriving from each parental population can be used to identify genomic regions that underlie phenotypic traits in those populations in hybrids. This method, often called admixture mapping, has been used in humans extensively and increasingly in non-model species (Slate & Pemberton 2007, Pallares et al. 2014).

However, at the dawn of the whole genome era, such protocols can be challenging to implement, and can get stuck between a rock and a hard place (Fig. 2). On the one hand, methods such as Rad-seq, although delivering a large number of SNPs, still sample the genome sparsely, and can fall short of this goal, particularly in species where the population recombination rate is high. This failure to identify actual targets of selection through selection mapping arises because the genome-wide levels of LD can be so low as to cause the actual targets of selection to be effectively unlinked (in low or average LD with) the nearest neutral site whose variation is interrogated. The result will be little evidence for selection among those genomic SNPs that are assayed, and many targets of selection will be missed. Clades such as *Drosophila*, or many bird species, likely fall into this category, and may require whole-genome resequencing to more confidently identify targets of selection through hitchhiking (Backström et al. 2006, Ellegren et al. 2012, see Fig. 3; Backström et al. 2013). The advantage to such species with high recombination rates is that, when an outlier locus is detected, one can be sure that one is relatively close to the actual target of selection, although even whole genome resequencing studies in *Drosophila* have sometimes yielded mixed results, especially if selection is weak, very recent, or acting on standing variation. On the other hand, levels of genome-wide LD in a given species may be substantial because of recent population history, a history of domestication or an overall low population recombination rate. Canids are good examples of this pattern (see Fig. 3 and Boyko et al. 2010, Boyko 2011). In such cases, even sparse sampling of the genome will often uncover sites that appear to be under selection, having hitchhiked with the actual targets that could be megabases away. When LD is high, larger regions of the chromosome are dragged along by hitchhiking than when LD is small, and these large regions can sometimes capture hundreds of genes. But in this situation, the list of candidate genes will be so long that it becomes less than useful. Threespine sticklebacks (*Gasterosteus aculeatus*) may fall in this category: Rad-seq studies routinely identify F_{ST} outliers but the list of genes within linkage blocks can be long and often provide only a vague idea of actual targets of selection (Hohenlohe

Downloaded From: https://bioone.org/journals/Folia-Zoologica on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

et al. 2012). The timing and strength of the selection event can also be important for regulating the size of the hitchhiking chromosomal segments. Chromosome inversions will also protect blocks of the genome from recombination, causing hundreds of genes to remain in high LD and making identification of the actual targets of selection challenging or impossible without further methods development. Although genome subsampling methods such as Rad-seq will in general provide a coarser picture of hitchhiking loci, even whole genome re-sequencing will not be able to unambiguously identify the actual targets of selection if LD is high.

Rad-seq or other genome subsampling method

	Advantages	Disadvantages
Low LD species	SNPs exhibiting selection likely close to causal variants	Difficult to find SNPs linked to causal variants; many SNPs required
High LD species	Easier to find SNPs linked to causal variants	Long linkage blocks containing many candidate genes

Fig. 2. Advantages and disadvantages of studying species with low or high levels of linkage disequilibrium (LD) using genome subsampling methods such as Rad-seq to identify loci under selection. The matrix covers two sets of species: on the Y-axis, those with low or high levels of LD, as discussed in the main text. On the X-axis are listed the advantages or disadvantages of studying such species using the Rad-seq approach. In each cell is a description of common situations encountered in the search for loci under selection. See text for further discussion.

Examples: Rad-seq meets selection mapping in natural populations

Our thinking on the efficacy of Rad-seq to search for loci under selection issues has been influenced by recent results from our laboratory. We now use two case studies, from a lizard and a songbird, to illustrate the challenges of detecting selection and of identifying targets of selection with genome subsampling methods.

Adaptation and the evolution of cold tolerance in green anole lizards

The green anole lizard, *Anolis carolinensis*, is an ideal species to explore the molecular basis of climate-mediated local adaptation. The only anole native to the continental United States, this species occupies the highest latitudes of any of the nearly 400 species of the genus. The northern edge of the species' distribution is likely limited by winter temperatures (Williams 1969), but populations do not hibernate, as is common for most mid- and high- latitude reptiles. By retreating to sheltered sites and basking during sun exposure, northern populations are able to remain active and periodically feed in the winter months (Bishop & Echternacht 2004), despite regular ambient temperatures below freezing. Additionally, populations from different climates display significant differences in cold tolerance (Wilson & Echternacht 1987). The recent publication of the genome of this species (Alfoldi et al. 2011) provides a unique resource for understanding the molecular processes of evolutionary response to local environment and for identifying genes that may play a key role in physiological divergence between populations of a non-model species. Taking advantage of this opportunity, we used double-digest RAD sequencing (ddRad-seq, Peterson et al. 2012) to identify regions of the *A. carolinensis* genome associated with cold variation across the species' range. Using SphI and EcoRI restriction enzymes, we digested genomic DNA from 28 individuals representing six populations spanning the latitudinal extent of the natural range of *A. carolinensis*. We genotyped 20282 SNPs with a coverage of ≥ 10× for these individuals using the Stacks software package (Catchen et al. 2011, 2013). To search for regions of the genome associated with temperature variation across the species' range, we used georeferenced locality data for each population to retrieve estimates of the mean temperature of the coldest quarter of the year (BIO11) from the Worldclim database (Hijmans et al. 2005). We then used allele counts from the Rad-seq dataset to calculate Bayes factor associations and Pearson correlations using the Bayenv2 software package (Gunther & Coop 2013). Variant sites in the top 1 % of both Bayes factor associations and Pearson correlations were retained as candidate markers identifying regions of the genome that may be important for local adaptation to cold (Fig. 4). This analysis resulted in 72 candidate SNPs, all of which were noncoding: 67 % are located in intergenic regions, whereas 33 % map to introns.

Several genes in this dataset may be of interest for further study due to their close proximity to SNPs exhibiting geographic correlations with temperature and their potential involvement in oxygen regulation, which has been proposed as a major constraint for ectotherms under extreme temperature challenge (Portner et al. 2006, 2007). One of these variants is located 40.8 kb upstream from the first exon of Rho-associated protein kinase 2 (ROCK 2), whose signaling is important for regulation of pulmonary vasculature

Downloaded From: https://bioone.org/journals/Folia-Zoologica on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

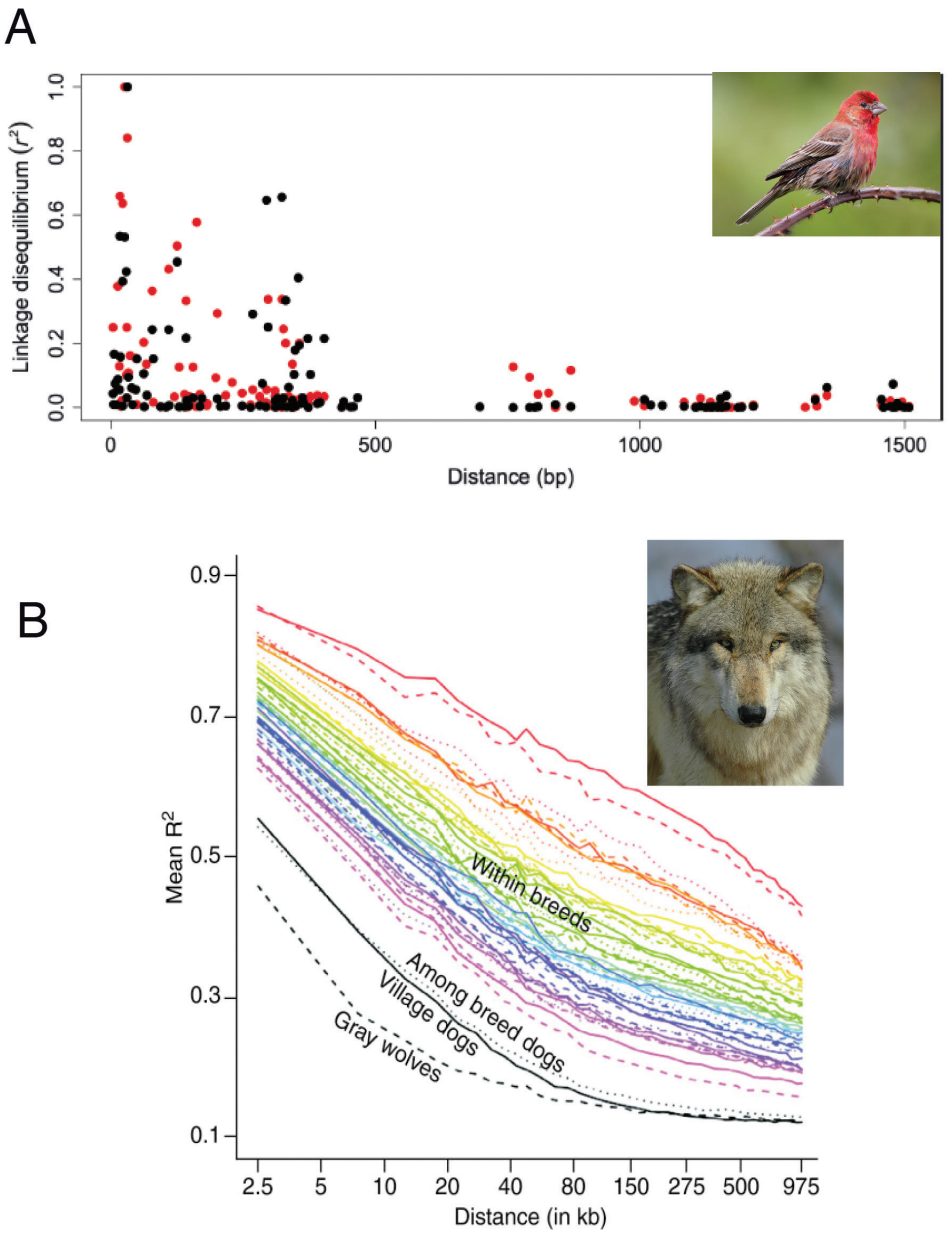


Fig. 3. Two examples of low- and high linkage disequilibrium species. A) Plot of pairwise r^2 (a measure of linkage disequilibrium) between SNPs across 1500 base pairs of the HSP90a gene in house finches, a common North American songbird. SNPs from an Arizona population are in black, those from an Alabama population in gray. This species is characterized by large population sizes, resulting in a high population recombination rate and low levels of LD. From Backström et al. (2013). B) Similar plot of pairwise r^2 and physical distance in kilobases in various dog breeds and wild populations of gray wolves (*Canis lupus*). Notice how high levels of LD extend hundreds of kilobases; those same levels of LD extend only a few hundred bases in the case of the house finches. Image from Boyko (2011); see also Boyko et al. (2010). Both images used under the Creative Commons 2 License https://creativecommons.org/licenses/by/2.0/.

(Riento & Ridley 2003, Noma et al. 2006, Seasholtz et al. 2006, Rankinen et al. 2008). Another SNP lies 1.46 kb upstream of the first exon of transcription factor 4 (TCF4), which is involved in regulation of breathing patterns (Zweier et al. 2007). Functional genomics studies are needed to better understand the potential role and importance of these and other physiological processes to temperature-mediated local adaptation within the green anole.

Temporal evolution of house finch populations before and after an epizootic

The house finch (*Haemorhous mexicanus*), one of the most common birds in both urban and rural environments in North America, is rapidly becoming a model system for avian study. It has been important in studies of rapid morphological adaptation, sexual selection, evolution of disease resistance, and invasion (Badyaev et al. 2012). The uniqueness of the house finch in studies of disease ecology is the result of its relationship with the pathogen

Downloaded From: https://bioone.org/journals/Folia-Zoologica on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

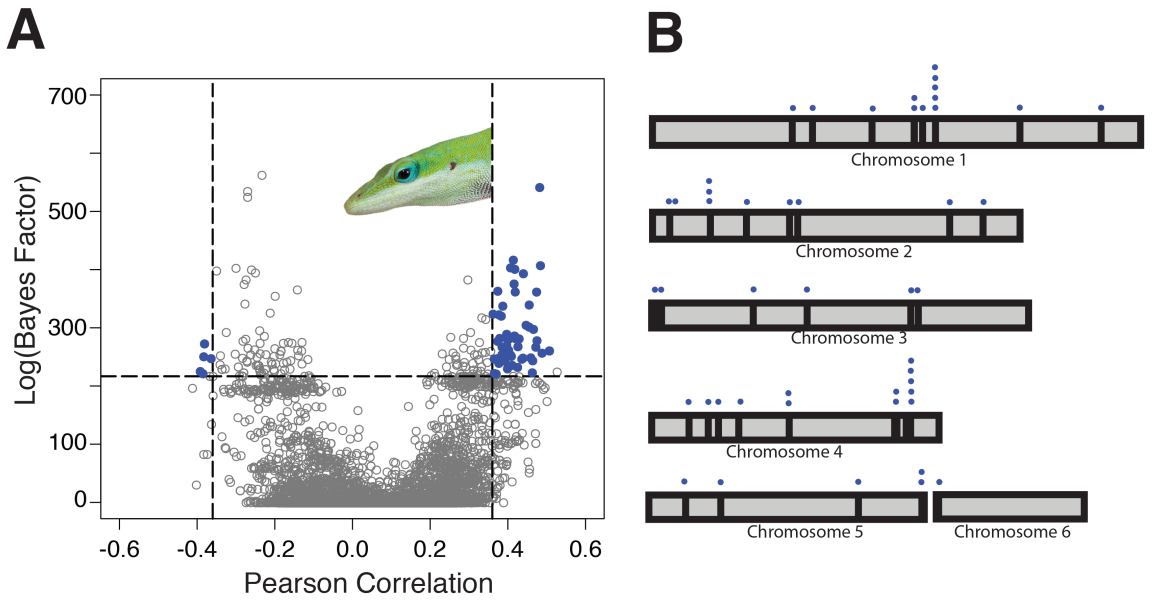


Fig. 4. Correlations of Rad-seq SNP variation and environmental variables across geographic space in the lizard *Anolis carolinensis*. A) Environmental associations across geographic space of each SNP identified by Rad-seq and the mean temperature of the coldest quarter of the year at each of six localities distributed across the species range, calculated in Bayenv2 (Gunther & Coop 2013). The horizontal and vertical dotted lines represent 99 % cutoffs for significance of Pearson correlation and Bayes Factor associations, respectively. Filled points indicate candidate SNPs falling within the top 1 % of both axes. B) The genomic position of each outlier SNP in panel A on the six annotated macrochromosomes of the *A. carolinensis* genome (Alfoldi et al. 2011).

Mycoplasma gallisepticum (MG). This poultry-associated bacterium was first documented in the house finch in 1994 in the Washington D.C. area (Ley et al. 1996, Hochachka & Dhondt 2000). MG infects the respiratory tract and causes severe conjunctivitis (Hochachka & Dhondt 2000), suppresses pathogen-specific components of the immune system (Bonneaud et al. 2011), and stimulates inflammatory responses (Gaunson et al. 2006, Mohammed et al. 2007, Adelman et al. 2013). The pathogen spread through the eastern population rapidly, and by 1998 had caused severe declines across the region, as high as 60 % in some areas (Dhondt et al. 1998). Infection experiments comparing gene expression responses of eastern individuals with 12 years of exposure and historically unexposed individuals suggested rapid evolution of gene expression, disease resistance (Bonneaud et al. 2011, Bonneaud et al. 2012) and disease tolerance (Adelman et al. 2013).

We collected a genome-wide SNP dataset using double-digest Rad-seq (Peterson et al. 2012) to identify regions of the genome with signatures of MG-mediated selection over time. As in the *Anolis* study, we digested the genome with SphI and EcoR1, and selected fragments from 276-324 base pairs long to recover homologous loci scattered randomly across the genome. In this preliminary study, we sampled only 11 individuals (22 chromosomes), five from pre-epizootic (1990) and six from post-epizootic (2003) populations from Alabama. We ran the Rad-seq libraries on a single lane of HiSeq Illumina 2500, generating a total of ~8 million paired-end reads (~4 million pairs), each 150 bp long. Using the Stacks pipeline (Catchen et al. 2011, 2013), we genotyped over 12000 SNPs from 2223 loci (Fig. 5). Of the 7260 comparisons of SNPs achieving our quality threshold between these time periods, we found 167 (2.3 %) significant shifts in allele frequency (Fisher's exact test p-value < 0.05) from 129 unique loci. Of these 129 loci, 68.2 % are in intergenic regions, 29.5 % are in introns, 2 loci fall within exons, and 1 within a 3′ UTR region. Although none of these loci retain significance with Bonferroni correction, we suspect that larger sample sizes of individuals from additional localities will improve detection. F_{ST} values in this collection of SNPs range from 0.208 to 1 (fixed differences). These regions with high F_{ST} are in or near genes with a variety of functions. One gene, PPP2R2C is involved in immune pathways in humans, and falls 13 kb away from a SNP with an F_{ST} of 1.

These two studies illustrate the promise but also the challenges of detecting selection and identifying candidate genes in vertebrate genomes using a genome subsampling method such as Rad-seq (Tiffin & Ross-Ibarra 2014). In both studies, most of the Rad-seq SNPs fell in noncoding regions whose relationship to

Downloaded From: https://bioone.org/journals/Folia-Zoologica on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

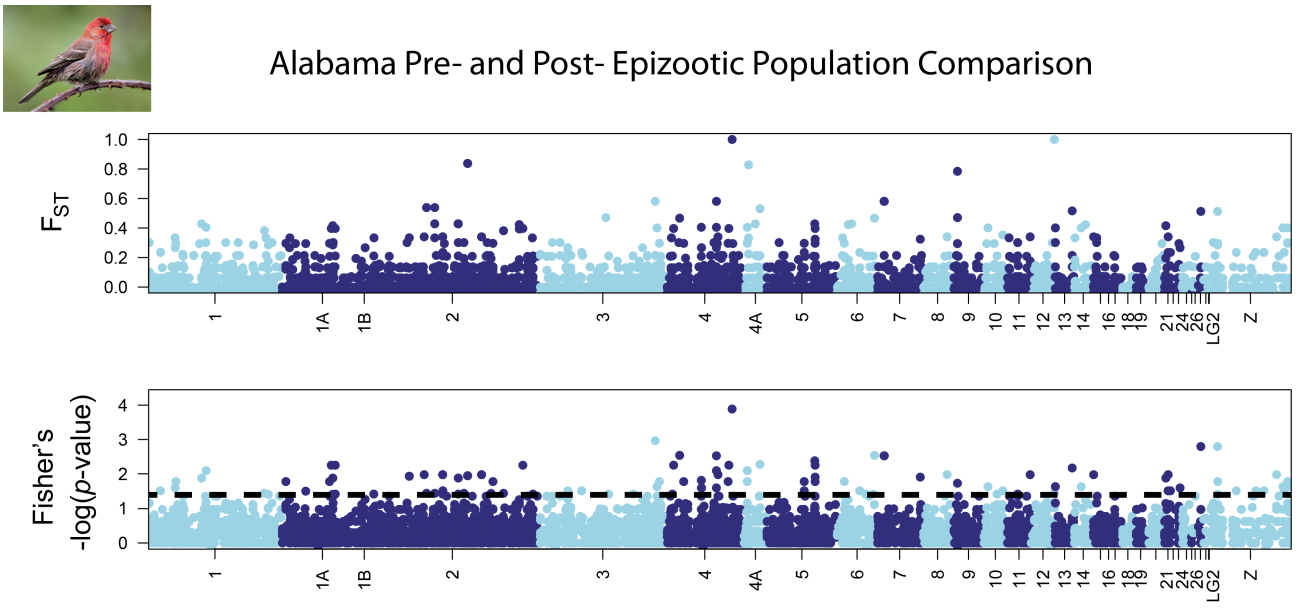


Fig. 5. Estimates of F_{ST} and associated Fisher's *p*-values for each SNP compared between pre- (1990) and post- *Mycoplasma* (2003) epizootic individuals sampled from Auburn, Alabama. The dotted line on the plot of *p*-values shows the uncorrected cutoff of $p = 0.05$. After Bonferroni correction, no SNPs achieve significant F_{ST} (see text). SNP positions are depicted as mapped to the zebra finch genome, assuming synteny between the house finch and zebra finch genomes. High values of F_{ST} are not always associated with high Fisher's *p*-values, usually due to incomplete data matrices and concomitant smaller sample sizes at those positions, a situation common in Rad-Seq data. Image of house finch from http://www.flickr.com/photos/11652987@N03/7315942062 and used under the Creative Commons 2 License https://creativecommons.org/licenses/by/2.0/.

nearby genes was unclear. In the example from *Anolis*, the candidate genes identified as being the closest to those SNPs that were correlated with environmental variables were often quite far away from the SNP used to tag them. In the example from house finches, the number of F_{ST} outliers in comparisons of pre- and post-epizootic populations was relatively small, perhaps because the observed level of LD in house finches is generally small, and certainly because of our small sample sizes. In the few avian species that have been studied with regard to recombination rate, rates on the autosomes are likely to be quite high, with levels of LD declining rapidly as one moves away from the focal SNP (Backström et al. 2006, Bullaughey et al. 2008, Janes et al. 2009, Li & Merila 2010, Ellegren 2014). We have found that LD in songbird populations, such as red-winged blackbirds (*Agelaius phoeniceus*) and house finches, falls off rapidly after a few hundred base pairs, a situation very reminiscent of populations of *Drosophila* (Edwards & Dillon 2004, Backström et al. 2013). In such species, LD often declines between SNPs less than 500 bp apart, which means that any SNP found to exhibit high differentiation or signatures of natural selection is unlikely to be useful in identifying candidate genes for a phenotypic trait even a few kb away. Thus it is unclear whether recent proposals to map QTL in natural populations using SNP-chips containing on the order of 10000 SNPs will be effective (Hagen et al. 2013).

Recent statistical models promise new power to estimate the distribution of effect sizes of loci underlying a phenotypic trait along whole chromosomes in natural populations. For example, using approximately 10000 SNPs, Santure et al. (2013) found that for most chromosomes in the genome of great tits studied at Whyndham Woods U.K., a model in which the effect of a given chromosome on continuous traits like wing length and clutch size was proportional to the length of that chromosome, and hence the number of genes on that chromosome. By extension, this result suggests that nearly every gene has a similar – and infinitesimally small – effect on variation in the focal trait. But it is also unclear whether the failure to reject such a null hypothesis is also due to the relatively meager sampling of the genome. Although 10000 SNPs seems like a large number, it is relatively small in terms of capturing variation in genomic blocks in high LD, especially for vertebrate genomes on the order of 1-3 Gb and in species with high population recombination rates (Edwards 2013).

Measuring hitchhiking is crucial enough to the expanding domain of phylogeography that we predict that, ultimately, the field will forsake genome subsampling approaches for phylogeographically

Downloaded From: https://bioone.org/journals/Folia-Zoologica on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

informed whole-genome resequencing studies. We are beginning to see the first glimpses of such studies in genomically unstudied species (e.g. Ojeda et al. 2014), and the results are as exciting as they are informative about the determinants of genomic variation and structure.

From phylogeography to genotype to phenotype

Loci whose variation has been influenced by natural selection are often also loci that underlie phenotypic traits. The search for loci underlying natural variation in phenotypic traits is a major thrust of modern evolutionary biology (Hoekstra et al. 2006, Hoekstra & Coyne 2007, Ellegren & Sheldon 2008, Rebeiz et al. 2009, Hiller et al. 2012, Jones et al. 2012b). Of the many methods for identifying such loci – including QTL and linkage mapping using pedigrees and crosses, or "systems genetics", in which multiple kinds of genomic, transcriptomic, and metabolomic data are integrated together (Feltus 2014) – association mapping is probably the method most closely allied to phylogeography. This alliance arises because association mapping uses population comparisons to find genomic loci that correlate with the distribution of a particular phenotype (Stinchcombe & Hoekstra 2008, reviewed in Kratochwil & Meyer 2015). Others have made a similar connection between landscape genetics and the search for loci underlying adaptive phenotypes (Jones et al. 2013). Association mapping has significant promise for closing the genotype-phenotype gap in non-model species and in many contexts has more statistical power than does mapping with pedigrees or controlled crosses (Schielzeth & Husby 2014). Indeed, the emerging trend is toward simultaneous inference of population history and identification of loci associated with phenotypic traits (e.g. Fumagalli et al. 2011, Linnen et al. 2013). Association mapping is likely to be most powerful in situations where most of the genomic variation is shared among a group of closely related populations or species, perhaps connected by high gene flow, but there exist marked phenotypic differences among those populations (Axelsson et al. 2013, Cullingham et al. 2014, Schielzeth & Husby 2014). In such situations, association mapping should reveal similar allele frequencies across nearly the entire genome, often due to shared standing variation, in both control and comparison populations, yet these populations should differ at loci whose variation correlates with the divergent phenotypes of interest. Precisely this situation has been found in those emerging cases where association mapping or candidate gene investigation has proved useful in non-model species, including several plants (Comeault et al. 2014, Cullingham et al. 2014, Johnston et al. 2014, Pearse et al. 2014, Roesti et al. 2014). Use of candidate genes in such situations can also be highly informative (Uy et al. 2009). Indeed, when populations have experienced a moderate history of divergence, such that the compared genomes differ at many sites due to neutral demographic divergence, it becomes essential to correct for such substructuring so as not to spuriously implicate loci showing strong allele frequency differences in the origin of phenotypic traits that between those populations (Pritchard et al. 2000, Patterson et al. 2006, Price et al. 2006, 2010).

Conclusions

This is an exciting time for phylogeography. We have a growing number of examples of studies in which employment of next-generation sequencing methods has yielded high resolution substructuring within species at a level of detail that far exceeds that formerly yielded by single locus mtDNA or microsatellite studies. It is now clear that a few thousand loci such as is typically yielded by methods such as Rad-seq may well be adequate to discover the major phylogeographic lineages within a species. Although the finer details of the history of species can always be clarified further with increasing genomic sampling (or individual sampling), many of these details are likely lost to historical reconstruction because of their age or their mild imprint on the pattern of genome-wide variation. While we expect whole-genome resequencing to become more common in phylogeography, it is unclear whether this method will be overkill if the focus is strictly on the core purview of phylogeography, namely reconstruction of the phylogenetic lineages and neutral demography within species. What is clear is that next-generation methods are causing a resurgence in the use of SNPs as opposed to microsatellites, enabling easier comparisons among loci and species and providing a uniform framework for comparative phylogeography (Hickerson et al. 2010, Andrew et al. 2013).

But next-generation sequencing has also broken down the conceptual edges of phylogeography, resulting in an expanded purview that has blurred the lines between core phylogeographic foci and other areas of related interest, such as identifying loci with a history of natural selection and underlying variation in adaptive traits. Studies combining historical reconstruction of phylogeographic history and a search for such adaptive loci are becoming more

Downloaded From: https://bioone.org/journals/Folia-Zoologica on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

common (Deagle et al. 2012, Jones et al. 2012a, b, Pearse et al. 2014, Wallberg et al. 2014), and the goal of identifying loci underlying adaptive traits is often as important as understanding the demographic history of a set of populations. This conceptually expanded phylogeography marks an important phase in the evolution of the field, and although always a glimmer in the eye of phylogeographers, has arguably been driven due to the recent arrival of high-throughput genomic approaches. As phylogeography expands its purview, it is becoming clear that genome subsampling methods such as Rad-seq, while extremely powerful for identifying phylogeographic clusters within species, may be inadequate for identifying loci that are targets of natural selection or linked to the true targets. Whole-genome resequencing will likely emerge as the standard tool, if not for traditional phylogeographic investigations, then certainly in the quest for loci underlying quantitative traits in natural populations and their history of divergence within species.

Acknowledgements

We thank Utku Perktaş for inviting us to contribute to this volume. We thank Tim Sackton, Fábio Raposo do Amaral, Ryan Garrick and Bryan Carstens for helpful discussion and sharing of data, and two anonymous reviewers and Ryan Garrick for helpful comments on the manuscript. We thank Emily Kay and Brant Peterson for suggestions on the Rad-Seq library preparation protocol, and Christian Daly and Jennifer Couget for help with Illumina sequencing. Work was supported by NSF grant DEB-IOS 0923088 to SVE and grants from the American Museum of Natural History, American Ornithologists Union, and Society for the Study of Evolution to AJS. Members of Jonathan Losos' lab provided helpful comments in the development of ideas and analytical approaches of the Anolis project. The Harvard Museum of Comparative Zoology Putnam Expedition Grant, Miyata Award and Robert A. Chapman Memorial Scholarship funded Anolis fieldwork. Anolis Rad-Seq work was supported by a National Science Foundation Doctoral Dissertation Improvement Grant (DDIG award # 1311484) to SCC and SVE.

Literature

Abdelkrim J., Robertson B., Stanton J.-A. & Gemmell N. 2009: Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. *BioTechniques 46: 185–192.*

Adelman J.S., Kirkpatrick L., Grodio J.L. & Hawley D.M. 2013: House finch populations differ in early inflammatory signaling and pathogen tolerance at the peak of *Mycoplasma gallisepticum* infection. *Am. Nat. 181: 674–689.*

Albayrak T., Gonzalez J., Drovetski S.V. & Wink M. 2012: Phylogeography and population structure of Kruper's nuthatch *Sitta krueperi* from Turkey based on microsatellites and mitochondrial DNA. *J. Ornithol. 153: 405–411.*

Alcaide M., Scordato E.S.C., Price T.D. & Irwin D.E. 2014: Genomic divergence in a ring species complex. *Nature 511: 83–85.*

Alfoldi J., Di Palma F., Grabherr M., Williams C., Kong L.S., Mauceli E., Russell P., Lowe C.B., Glor R.E., Jaffe J.D. et al. 2011: The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature 477: 587–591.*

Altshuler D., Durbin R.M., Abecasis G.R., Auton A., Brooks L.D., Gibbs R.A., Hurles M.E. & McVean G.A. et al. 2010: A map of human genome variation from population-scale sequencing. *Nature 467: 1061–1073.*

Andrew R.L., Bernatchez L., Bonin A., Buerkle C.A., Carstens B.C., Emerson B.C., Garant D., Giraud T., Kane N.C., Rogers S.M., Slate J., Smith H., Sork V.L., Stone G.N., Vines T.H., Waits L., Widmer A. & Rieseberg L.H. 2013: A road map for molecular ecology. *Mol. Ecol. 22: 2605–2626.*

Arnold B., Corbett-Detig R.B., Hartl D. & Bomblies K. 2013: RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Mol. Ecol. 22: 3179–3190.*

Avise J.C., Arnold J., Ball R.M., Bermingham E., Lamb T., Neigel J.E., Reeb C.A. & Saunders N.C. 1987: Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annu. Rev. Ecol. Syst. 18: 489–522.*

Axelsson E., Ratnakumar A., Arendt M.L., Maqbool K., Webster M.T., Perloski M., Liberg O., Arnemo J.M., Hedhammar A. & Lindblad-Toh K. 2013: The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature 495: 360–364.*

Backström N., Qvarnstrom A., Gustafsson L. & Ellegren H. 2006: Levels of linkage disequilibrium in a wild bird population. *Biol. Lett. 2: 435–438.*

Backström N., Shipilina D., Blom M.P.K. & Edwards S.V. 2013: Cis-regulatory sequence variation and association with *Mycoplasma* load in natural populations of the house finch (*Carpodacus mexicanus*). *Ecol. Evol. 3: 655–666.*

Badyaev A.V., Belloni V. & Hill G.E. 2012: House finch (*Carpodacus mexicanus*). In: Poole A. (ed.), The birds of North America. *Accessed 24 July 2015, Cornell Lab of Ornithology, Ithaca.*

Balakrishnan C.N., Lee J.Y. & Edwards S.V. 2010: Phylogeography and phylogenetics in the nuclear age. In: Grant P. & Grant R. (eds.), Searching for the causes of evolution: from field observations to mechanisms. *Princeton University Press, Princeton, New Jersey: 65–88.*

Ballard J.W.O. & Pichaud N. 2014: Mitochondrial DNA: more than an evolutionary bystander. *Funct. Ecol. 28: 218–231.*

Ballard J.W.O. & Whitlock M.C. 2004: The incomplete natural history of mitochondria. *Mol. Ecol. 13: 729–744.*

Beaumont M.A., Nielsen R., Robert C., Hey J., Gaggiotti O., Knowles L., Estoup A., Panchal M., Corander J., Hickerson M., Sisson S.A., Fagundes N., Chikhi L., Beerli P., Vitalis R., Cornuet J.M., Huelsenbeck J., Foll M., Yang Z.H., Rousset F., Balding D. & Excoffier L. 2010: In defence of model-based inference in phylogeography REPLY. *Mol. Ecol. 19: 436–446.*

Becquet C., Patterson N., Stone A.C., Przeworski M. & Reich D. 2007: Genetic structure of chimpanzee populations. *PLoS Genet. 3: e66.*

Beerli P. & Felsenstein J. 1999: Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics 152: 763–773.*

Downloaded From: https://bioone.org/journals/Folia-Zoologica on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

Bishop D.C. & Echternacht A.C. 2004: Emergence behavior and movements of winter-aggre gated green anoles (*Anolis carolinensis*) and the thermal characteristics of their crevices in Tennessee. *Herpetologica 60: 168–177.*
Bonneaud C., Balenger S.L., Russell A.F., Zhang J., Hill G.E. & Edwards S.V. 2011: Rapid evolution of disease resistance is accompanied by functional changes in gene expression in a wild bird. *Proc. Natl. Acad. Sci. U. S. A. 108: 7866–7871.*
Bonneaud C., Balenger S.L., Zhang J., Edwards S.V. & Hill G.E. 2012: Innate immunity and the evolution of resistance to an emerging infectious disease in a wild bird. *Mol. Ecol. 21: 2628–2639.*
Boyko A.R. 2011: The domestic dog: man's best friend in the genomic era. *Genome Biol. 12: 216.*
Boyko A.R., Quignon P., Li L., Schoenebeck J.J., Degenhardt J.D., Lohmueller K.E., Zhao K.Y., Brisbin A., Parker H.G., vonHoldt B.M., Cargill M., Auton A., Reynolds A., Elkahloun A.G., Castelhano M., Mosher D.S., Sutter N.B., Johnson G.S., Novembre J., Hubisz M.J., Siepel A., Wayne R.K., Bustamante C.D. & Ostrander E.A. 2010: A simple genetic architecture underlies morphological variation in dogs. *PLoS Biol. 8: e10000451.*
Brito P. & Edwards S.V. 2009: Multilocus phylogeography and phylogenetics using sequence-based markers. *Genetica 135: 439–455.*
Brumfield R., Nickerson D.A., Beerli P. & Edwards S.V. 2003: The utility of single nucleotide polymorphisms in inferences of population history. *Trends Ecol. Evol. 18: 249–256.*
Bryant D., Bouckaert R., Felsenstein J., Rosenberg N.A. & RoyChoudhury A. 2012: Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Mol. Biol. Evol. 29: 1917–1932.*
Bullaughey K., Przeworski M. & Coop G. 2008: No effect of recombination on the efficacy of natural selection in primates. *Genome Res. 18: 544–554.*
Burke T., Davies N.B., Bruford M.W. & Hatchwell B.J. 1989: Parental care and mating-behaviour of polyandrous dunnocks *Prunella modularis* related to paternity by DNA fingerprinting. *Nature 338: 249–251.*
Campo D., Lehmann K., Fjeldsted C., Souaiaia T., Kao J. & Nuzhdin S.V. 2013: Whole-genome sequencing of two North American *Drosophila melanogaster* populations reveals genetic differentiation and positive selection. *Mol. Ecol. 22: 5084–5097.*
Carling M.D. & Brumfield R.T. 2007: Gene sampling strategies for multi-locus population estimates of genetic diversity (theta). *PLoS ONE 2: e160.*
Catchen J.M., Amores A., Hohenlohe P., Cresko W. & Postlethwait J.H. 2011: Stacks: building and genotyping loci de novo from short-read sequences. *G3 1: 171–182.*
Catchen J., Hohenlohe P.A., Bassham S., Amores A. & Cresko W.A. 2013: Stacks: an analysis tool set for population genomics. *Mol. Ecol. 22: 3124–3140.*
Chifman J. & Kubatko L. 2014: Quartet inference from SNP data under the coalescent model. *Bioinformatics 30: 3317–3324.*
Comeault A.A., Soria-Carrasco V., Gompert Z., Farkas T.E., Buerkle C.A., Parchman T.L. & Nosil P. 2014: Genome-wide association mapping of phenotypic traits subject to a range of intensities of natural selection in *Timema cristinae*. *Am. Nat. 183: 711–727.*
Corbett-Detig R.B., Hartl D.L. & Sackton T.B. 2015: Natural selection constrains neutral diversity across a wide range of species. *PLoS Biol. 13: e1002112.*
Cruickshank T.E. & Hahn M.W. 2014: Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol. Ecol. 23: 3133–3157.*
Cullingham C.I., Cooke J.E.K. & Coltman D.W. 2014: Cross-species outlier detection reveals different evolutionary pressures between sister species. *New Phytol. 204: 215–229.*
Curto M.A., Tembrock L.R., Puppo P., Nogueira M., Simmons M.P. & Meimberg H. 2013: Evaluation of microsatellites of *Catha edulis* (qat; Celastraceae) identified using pyrosequencing. *Biochem. Syst. Ecol. 49: 1–9.*
Deagle B.E., Jones F.C., Chan Y.G.F., Absher D.M., Kingsley D.M. & Reimchen T.E. 2012: Population genomics of parallel phenotypic evolution in stickleback across stream-lake ecological transitions. *Proc. R. Soc. Lond. B 279: 1277–1286.*
Dhondt A.A., Tessaglia D.L. & Slothower R.L. 1998: Epidemic mycoplasmal conjunctivitis in house finches from eastern North America. *J. Wildlife Dis. 34: 265–280.*
Duchen P., Živković D., Hutter S., Stephan W. & Laurent S. 2013: Demographic inference reveals African and European admixture in the North American *Drosophila melanogaster* population. *Genetics 193: 291–301.*
Edwards S.V. 2013: Next-generation QTL mapping: crowdsourcing SNPs, without pedigrees. *Mol. Ecol. 22: 3885–3887.*
Edwards S.V. & Beerli P. 2000: Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution 54: 1839–1854.*
Edwards S.V. & Bensch S. 2009: Looking forwards or looking backwards in avian phylogeography? A comment on Zink and Barrowclough 2008. *Mol. Ecol. 18: 2930–2933.*
Edwards S.V. & Dillon M. 2004: Hitchhiking and recombination in birds: evidence from Mhc-linked and unlinked loci in red-winged blackbirds (*Agelaius phoeniceus)*. *Genet. Res. 84: 175–192.*
Ellegren H. 2014: Genome sequencing and population genomics in non-model organisms. *Trends Ecol. Evol. 29: 51–63.*
Ellegren H. & Sheldon B.C. 2008: Genetic basis of fitness differences in natural populations. *Nature 452: 169–175.*
Ellegren H., Smeds L., Burri R., Olason P.I., Backström N., Kawakami T., Kunstner A., Makinen H., Nadachowska-Brzyska K., Qvarnstrom A., Uebbing S. & Wolf J.B.W. 2012: The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature 491: 756–760.*
Emerson K.J., Merz C.R., Catchen J.M., Hohenlohe P.A., Cresko W.A., Bradshaw W.E. & Holzapfel C.M. 2010: Resolving postglacial phylogeography using high-throughput sequencing. *Proc. Natl. Acad. Sci. U. S. A. 107: 16196–16200.*
Excoffier L. & Ray N. 2008: Surfing during population expansions promotes genetic revolutions and structuration. *Trends Ecol. Evol. 23: 347–351.*
Faircloth B.C., McCormack J.E., Crawford N.G., Harvey M.G., Brumfield R.T. & Glenn T.C. 2012: Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Syst. Biol. 61: 717–726.*

Downloaded From: https://bioone.org/journals/Folia-Zoologica on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

Felsenstein J. 2006: Accuracy of coalescent likelihood estimates: do we need more sites, more sequences, or more loci? *Mol. Biol. Evol. 23: 691–700.*
Feltus F.A. 2014: Systems genetics: a paradigm to improve discovery of candidate genes and mechanisms underlying complex traits. *Plant Sci. 223: 45–48.*
Fordyce S.L., Ávila-Arcos M.C., Rockenbauer E., Børsting C., Frank-Hansen R., Petersen F.T., Willerslev E., Hansen A.J., Morling N. & Gilbert M.T.P. 2011: High-throughput sequencing of core STR loci for forensic genetic investigations using the Roche Genome Sequencer FLX platform. *BioTechniques 51: 127–133.*
Fumagalli M., Sironi M., Pozzoli U., Ferrer-Admetlla A., Ferrer-Admettla A., Pattini L. & Nielsen R. 2011: Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. *PLoS Genet. 7: e1002355.*
Gao C.H., Ren X.D., Mason A.S., Li J.N., Wang W., Xiao M.L. & Fu D.H. 2013: Revisiting an important component of plant genomes: microsatellites. *Funct. Plant Biol. 40: 645–661.*
Garrick R.C., Bonatelli I.A.S., Hyseni C., Morales A., Pelletier T.A., Perez M.F., Rice E., Satler J.D., Symula R.E., Thomé M.T.C. & Carstens B.C. 2015: The evolution of phylogeographic datasets. *Mol. Ecol. 24: 1164–1171.*
Gaunson J.E., Philip C.J., Whithear K.G. & Browning G.F. 2006: The cellular immune response in the tracheal mucosa to *Mycoplasma gallisepticum* in vaccinated and unvaccinated chickens in the acute and chronic stages of disease. *Vaccine 24: 2627–2633.*
Gerber A.S., Loggins R., Kumar S. & Dowling T.E. 2001: Does nonneutral evolution shape observed patterns of DNA variation in animal mitochondrial genomes? *Annu. Rev. Genet. 35: 539–566.*
Godinho R., Crespo E. & Ferrand N. 2008: The limits of mtDNA phylogeography: complex patterns of population history in a highly structured Iberian lizard are only revealed by the use of nuclear markers. *Mol. Ecol. 17: 4670–4683.*
Grover A. & Sharma P.C. 2011: Is spatial occurrence of microsatellites in the genome a determinant of their function and dynamics contributing to genome evolution? *Curr. Sci. 100: 859–869.*
Gunther T. & Coop G. 2013: Robust identification of local adaptation from allele frequencies. *Genetics 195: 205–220.*
Gyllensten U.B., Jakobsson S. & Temrin H. 1990: No evidence for illegitimate young in monogamous and polygynous warblers. *Nature 343: 168–170.*
Hagen I.J., Billing A.M., Ronning B., Pedersen S.A., Parn H., Slate J. & Jensen H. 2013: The easy road to genome-wide medium density SNP screening in a non-model species: development and application of a 10K SNP-chip for the house sparrow (*Passer domesticus*). *Mol. Ecol. Resour. 13: 429–439.*
Hammer M., Woerner A., Mendez F., Watkins J. & Wall J. 2011: Genetic evidence for archaic admixture in Africa. *Proc. Natl. Acad. Sci. U. S. A. 108: 15123–15128.*
Hare M.P. & Palumbi S.R. 1999: The accuracy of heterozygous base calling from diploid sequence and resolution of haplotypes using allele-specific sequencing. *Mol. Ecol. 8: 1750–1752.*
Harvey M.G. & Brumfield R.T. 2015: Genomic variation in a widespread Neotropical bird (*Xenops minutus*) reveals divergence, population expansion, and gene flow. *Mol. Phylogenet. Evol. 83: 305–316.*
Harvey M.G., Judy C.D., Seeholzer G.F., Maley J.M., Graves G.R. & Brumfield R.T. 2015a: Similarity thresholds used in DNA sequence assembly from short reads can reduce the comparability of population histories across species. *PeerJ 3: e895.*
Harvey M.G., Smith B.T., Glenn T.C., Faircloth B.C. & Brumfield R.T. 2015b: Sequence capture versus restriction site associated DNA sequencing for phylogeography. *arXiv: 1312.6439.*
Hedin M., Starrett J., Akhter S., Schonhofer A.L. & Shultz J.W. 2012: Phylogenomic resolution of paleozoic divergences in harvestmen (Arachnida, Opiliones) via analysis of next-generation transcriptome data. *PLoS ONE 7: e42888.*
Hickerson M.J., Carstens B.C., Cavender-Bares J., Crandall K.A., Graham C.H., Johnson J.B., Rissler L., Victoriano P.F. & Yoder A.D. 2010: Phylogeography's past, present, and future: 10 years after Avise, 2000. *Mol. Phylogenet. Evol. 54: 291–301.*
Hijmans R.J., Cameron S.E., Parra J.L., Jones P.G. & Jarvis A. 2005: Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol. 25: 1965–1978.*
Hiller M., Schaar B.T., Indjeian V.B., Kingsley D.M., Hagey L.R. & Bejerano G. 2012: A "forward genomics" approach links genotype to phenotype using independent phenotypic losses among related species. *Cell Rep. 2: 817–823.*
Hochachka W.M. & Dhondt A.A. 2000: Density-dependent decline of host abundance resulting from a new infectious disease. *Proc. Natl. Acad. Sci. U. S. A. 97: 5303–5306.*
Hoekstra H.E. & Coyne J.A. 2007: The locus of evolution: evo devo and the genetics of adaptation. *Evolution 61: 995–1016.*
Hoekstra H.E., Hirschmann R.J., Bundey R.A., Insel P.A. & Crossland J.P. 2006: A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science 313: 101–104.*
Hohenlohe P.A., Bassham S., Currey M. & Cresko W.A. 2012: Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. *Philos. Trans. R. Soc. Lond. B 367: 395–408.*
Huang H. & Knowles L.L. 2014: Unforeseen consequences of excluding missing data from next-generation sequences: simulation study of RAD sequences. *Syst. Biol. 63: 263–271.*
Hudson R.R. & Kaplan N.L. 1988: The coalescent process in models with selection and recombination. *Genetics 120: 831–840.*
Janes D.E., Ezaz T., Graves J.A.M. & Edwards S.V. 2009: Recombination and nucleotide diversity in the sex chromosomal pseudoautosomal region of the emu, *Dromaius novaehollandiae*. *J. Hered. 100: 125–136.*
Jeffreys A.J., Wilson V. & Thein S.L. 1985: Hypervariable minisatellite regions in human DNA. *Nature 314: 67–73.*
Jennings W.B. & Edwards S.V. 2005: Speciational history of Australian grass finches (*Poephila*) inferred from 30 gene trees. *Evolution 59: 2033–2047.*
Jobling M.A. 2012: The impact of recent events on human genetic diversity. *Philos. Trans. R. Soc. Lond. B 367: 793–799.*

Downloaded From: https://bioone.org/journals/Folia-Zoologica on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

Johnston S.E., Orell P., Pritchard V.L., Kent M.P., Lien S., Niemela E., Erkinaro J. & Primmer C.R. 2014: Genome-wide SNP analysis reveals a genetic basis for sea-age variation in a wild population of Atlantic salmon (*Salmo salar*). *Mol. Ecol. 23: 3452–3468.*

Jones F.C., Chan Y.F., Schmutz J., Grimwood J., Brady S.D., Southwick A.M., Absher D.M., Myers R.M., Reimchen T.E., Deagle B.E., Schluter D. & Kingsley D.M. 2012a: A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. *Curr. Biol. 22: 83–90.*

Jones F.C., Grabherr M.G., Chan Y.F., Russell P., Mauceli E., Johnson J., Swofford R., Pirun M., Zody M.C., White S., Birney E., Searle S., Schmutz J., Grimwood J., Dickson M.C., Myers R.M., Miller C.T., Summers B.R., Knecht A.K., Brady S.D., Zhang H.L., Pollen A.A., Howes T., Amemiya C., Lander E.S., Di Palma F., Lindblad-Toh K., Kingsley D.M., Platf B.I.G.S. & Team W.G.A. 2012b: The genomic basis of adaptive evolution in threespine sticklebacks. *Nature 484: 55–61.*

Jones M.R., Forester B.R., Teufel A.I., Adams R.V., Anstett D.N., Goodrich B.A., Landguth E.L., Joost S. & Manel S. 2013: Integrating landscape genomics and spatially explicit approaches to detect loci under seletion in clinal populatoins. *Evolution 67: 3455–3468.*

Keightley P.D. & Eyre-Walker A. 2010: What can we learn about the distribution of fitness effects of new mutations from DNA sequence data? *Philos. Trans. R. Soc. Lond. B 365: 1187–1193.*

Kimura M. 1968. Evolutionary rate at the molecular level. *Nature 217: 624–626.*

Kimura M. 1983: The neutral theory of molecular evolution. *Cambridge University Press, Cambridge.*

Kratochwil C.F. & Meyer A. 2015: Closing the genotype-phenotype gap: emerging technologies for evolutionary genetics in ecological model vertebrate systems. *BioEssays 37: 213–226.*

Kwong M. & Pemberton T.J. 2014: Sequence differences at orthologous microsatellites inflate estimates of human-chimpanzee differentiation. *BMC Genomics 15: 990.*

Lemmon A.R. & Lemmon E.M. 2012: High-throughput identification of informative nuclear loci for shallow-scale phylogenetics and phylogeography. *Syst. Biol. 61: 745–761.*

Lemmon E.M. & Lemmon A.R. 2013: High-throughput genomic data in systematics and phylogenetics. *Annu. Rev. Ecol. Evol. Syst. 44: 99–121.*

Lewontin R. 1974: The genetic basis of evolutionary change. *Columbia University Press, New York.*

Lewontin R.C. & Krakauer J. 1973: Distribution of gene frequency as a test of theory of selective neutrality of polymorphisms. *Genetics 74: 175–195.*

Ley D.H., Berkhoff J.E. & McLaren J.M. 1996: *Mycoplasma gallisepticum* isolated from house finches (*Carpodacus mexicanus*) with conjunctivitis. *Avian Dis. 40: 480–483.*

Li H. & Durbin R. 2011: Inference of human population history from individual whole-genome sequences. *Nature 475: 493–496.*

Li M.H. & Merila J. 2010: Sex-specific population structure, natural selection, and linkage disequilibrium in a wild bird population as revealed by genome-wide microsatellite analyses. *BMC Evol. Biol. 10: 66.*

Linnen C.R., Poh Y.P., Peterson B.K., Barrett R.D.H., Larson J.G., Jensen J.D. & Hoekstra H.E. 2013: Adaptive evolution of multiple traits through multiple mutations at a single gene. *Science 339: 1312–1316.*

Liu T., Wahlberg S., Burek E., Lindblom P., Rubio C. & Lindblom A. 2000: Microsatellite instability as a predictor of a mutation in a DNA mismatch repair gene in familial colorectal cancer. *Gene. Chromosome. Canc. 27: 17–25.*

Lotterhos K.E. & Whitlock M.C. 2014: Evaluation of demographic history and neutral parameterization on the performance of F-ST outlier tests. *Mol. Ecol. 23: 2178–2192.*

Marra N.J., Romero A. & DeWoody J.A. 2014: Natural selection and the genetic basis of osmoregulation in heteromyid rodents as revealed by RNA-seq. *Mol. Ecol. 23: 2699–2711.*

McCormack J.E. & Faircloth B.C. 2013: Next-generation phylogenetics takes root. *Mol. Ecol. 22: 19–21.*

McCormack J.E., Hird S.M., Zellmer A.J., Carstens B.C. & Brumfield R.T. 2013: Applications of next-generation sequencing to phylogeography and phylogenetics. *Mol. Phylogenet. Evol. 66: 526–538.*

McCormack J.E., Maley J.M., Hird S.M., Derryberry E.P., Graves G.R. & Brumfield R.T. 2012: Next-generation sequencing reveals phylogeographic structure and a species tree for recent bird divergences. *Mol. Phylogenet. Evol. 62: 397–406.*

Merz C., Catchen J.M., Hanson-Smith V., Emerson K.J., Bradshaw W.E. & Holzapfel C.M. 2013: Replicate phylogenies and post-glacial range expansion of the pitcher-plant mosquito, *Wyeomyia smithii*, in North America. *PLoS ONE 8: e72262.*

Metzgar D., Bytof J. & Wills C. 2000: Selection against frameshift mutations limits microsatellite expansion in coding DNA. *Genome Res. 10: 72–80.*

Mohammed J., Frasca S., Cecchini K., Rood D., Nyaoke A.C., Geary S.J. & Silbart L.K. 2007: Chemokine and cytokine gene expression profiles in chickens inoculated with *Mycoplasma gallisepticum* strains Rlow or GT5. *Vaccine 25: 8611–8621.*

Morales H.E., Pavlova A., Joseph L. & Sunnucks P. 2015: Positive and purifying selection in mitochondrial genomes of a bird with mitonuclear discordance. *Mol. Ecol. 24: 2820–2837.*

Morin P.A., Luikart G., Wayne R.K. & Grp S.N.P.W. 2004: SNPs in ecology, evolution and conservation. *Trends Ecol. Evol. 19: 208–216.*

Nabholz B., Glemin S. & Galtier N. 2009: The erratic mitochondrial clock: variations of mutation rate, not population size, affect mtDNA diversity across birds and mammals. *BMC Evol. Biol. 9: 54.*

Nabholz B., Mauffrey J.F., Bazin E., Galtier N. & Glemin S. 2008: Determination of mitochondrial genetic diversity in mammals. *Genetics 178: 351–361.*

Nei M. & Roychoudhury A.K. 1974: Sampling variances of heterozygosity and genetic distance. *Genetics 76: 379–390.*

Nielsen R. & Beaumont M.A. 2009: Statistical inferences in phylogeography. *Mol. Ecol. 18: 1034–1047.*

Noma K., Oyama N. & Liao J.K. 2006: Physiological role of ROCKs in the cardiovascular system. *Am. J. Physiol.-Cell Physiol. 290: C661–C668.*

Downloaded From: https://bioone.org/journals/Folia-Zoologica on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

O'Neill E.M., Schwartz R., Bullock C.T., Williams J.S., Shaffer H.B., Aguilar-Miguel X., Parra-Olea G. & Weisrock D.W. 2013: Parallel tagged amplicon sequencing reveals major lineages and phylogenetic structure in the North American tiger salamander (*Ambystoma tigrinum*) species complex. *Mol. Ecol. 22: 111–129.*

Ohta T. 1992: The nearly neutral theory of molecular evolution. *Annu. Rev. Ecol. Syst. 23: 263–286.*

Ohta T. & Gillespie J.H. 1996: Development of neutral and nearly neutral heories. *Theor. Popul. Biol. 49: 128–142.*

Ojeda D.I., Dhillon B., Tsui C.K.M. & Hamelin R.C. 2014: Single-nucleotide polymorphism discovery in *Leptographium longiclavatum*, a mountain pine beetle-associated symbiotic fungus, using whole-genome resequencing. *Mol. Ecol. Resour. 14: 401–410.*

Pallares L.F., Harr B., Turner L.M. & Tautz D. 2014: Use of a natural hybrid zone for genome-wide association mapping of craniofacial traits in the house mouse. *Mol. Ecol. 23: 5756–5770.*

Palumbi S.R. & Baker C.S. 1994: Contrasting population-structure from nuclear intron sequences and mtDNA of humpback whales. *Mol. Biol. Evol. 11: 426–435.*

Patterson N., Price A.L. & Reich D. 2006: Population structure and eigenanalysis. *PLoS Genet. 2: e190.*

Pearse D.E., Miller M.R., Abadia-Cardoso A. & Garza J.C. 2014: Rapid parallel evolution of standing variation in a single, complex, genomic region is associated with life history in steelhead/rainbow trout. *Proc. R. Soc. Lond. B 281: 20140012.*

Perktaş U., Gür H., Sağlam İ.K. & Quintero E. 2015: Climate-driven range shifts and demographic events over the history of Kruper's nuthatch *Sitta krueperi*. *Bird Study 62: 14–28.*

Perry J.C. & Rowe L. 2011: Rapid microsatellite development for water striders by next-generation sequencing. *J. Hered. 102: 125–129.*

Peterson B.K., Weber J.N., Kay E.H., Fisher H.S. & Hoekstra H.E. 2012: Double digest RADseq: an inexpensive method for De Novo SNP discovery and genotyping in model and non-model species. *PLoS ONE 7: e37135.*

Pluzhnikov A. & Donnelly P. 1996: Optimal sequencing strategies for surveying molecular genetic diversity. *Genetics 144: 1247–1262.*

Portner H.O., Bennett A.F., Bozinovic F., Clarke A., Lardies M.A., Lucassen M., Pelster B., Schiemer F. & Stillman J.H. 2006: Trade-offs in thermal adaptation: the need for a molecular to ecological integration. *Physiol. Biochem. Zool. 79: 295–313.*

Portner H.O., Peck L. & Somero G. 2007: Thermal limits and adaptation in marine Antarctic ectotherms: an integrative view. *Philos. Trans. R. Soc. Lond. B 362: 2233–2258.*

Price A.L., Patterson N.J., Plenge R.M., Weinblatt M.E., Shadick N.A. & Reich D. 2006: Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet. 38: 904–909.*

Price A.L., Zaitlen N.A., Reich D. & Patterson N. 2010: New approaches to population stratification in genome-wide association studies. *Nat. Rev. Genet. 11: 459–463.*

Pritchard J.K., Stephens M. & Donnelly P. 2000: Inference of population structure using multilocus genotype data. *Genetics 155: 945–959.*

Puritz J.B., Matz M.V., Toonen R.J., Weber J.N., Bolnick D.I. & Bird C.E. 2014: Demystifying the RAD fad. *Mol. Ecol. 23: 5937–5942.*

Rand D.M. 2001: The units of selection on mitochondrial DNA. *Annu. Rev. Ecol. Syst. 32: 415–448.*

Rankinen T., Church T., Rice T., Markward N., Blair S.N. & Bouchard C. 2008: A major haplotype block at the Rho-associated kinase 2 locus is associated with a lower risk of hypertension in a recessive manner: the HYPGENE study. *Hypertension Res. 31: 1651–1657.*

Raposo do Amaral F., Neves L.G., Resende M.F.R., Mobili F., Miyaki C.Y., Pellegrino K.C.M. & Biondo C. 2015: Ultraconserved elements sequencing as a low-cost source of complete mitochondrial genomes and microsatellite markers in non-model amniotes. *PLoS ONE 10: e0138446.*

Rebeiz M., Pool J.E., Kassner V.A., Aquadro C.F. & Carroll S.B. 2009: Stepwise modification of a modular enhancer underlies adaptation in a *Drosophila* population. *Science 326: 1663–1667.*

Reich D., Green R.E., Kircher M., Krause J., Patterson N., Durand E.Y., Viola B., Briggs A.W., Stenzel U., Johnson P.L.F., Maricic T., Good J.M., Marques-Bonet T., Alkan C., Fu Q., Mallick S., Li H., Meyer M., Eichler E.E., Stoneking M., Richards M., Talamo S., Shunkov M.V., Derevianko A.P., Hublin J.-J., Kelso J., Slatkin M. & Pääbo S. 2010: Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature 468: 1053–1060.*

Reinhardt J.A., Kolaczkowski B., Jones C.D., Begun D.J. & Kern A.D. 2014: Parallel geographic variation in *Drosophila melanogaster*. *Genetics 197: 361–373.*

Rheindt F.E., Fujita M.K., Wilton P.R. & Edwards S.V. 2014: Introgression and phenotypic assimilation in *Zimmerius* flycatchers (Tyrannidae): population genetic and phylogenetic inferences from genome-wide SNPs. *Syst. Biol. 63: 134–152.*

Ribeiro A.M., Lloyd P. & Bowie R.C.K. 2011: A tight balance between natural selection and gene flow in a Southern African arid-zone endemic bird. *Evolution 65: 3499–3514.*

Riento K. & Ridley A.J. 2003: Rocks: multifunctional kinases in cell behaviour. *Nat. Rev. Mol. Cell Biol. 4: 446–456.*

Roesti M., Gavrilets S., Hendry A.P., Salzburger W. & Berner D. 2014: The genomic signature of parallel adaptation from shared genetic variation. *Mol. Ecol. 23: 3944–3956.*

Romiguier J., Gayral P., Ballenghien M., Bernard A., Cahais V., Chenuil A., Chiari Y., Dernat R., Duret L., Faivre N., Loire E., Lourenco J.M., Nabholz B., Roux C., Tsagkogeorga G., Weber A.A.T., Weinert L.A., Belkhir K., Bierne N., Glemin S. & Galtier N. 2014: Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature 515: 261–263.*

Santure A.W., De Cauwer I., Robinson M.R., Poissant J., Sheldon B.C. & Slate J. 2013: Genomic dissection of variation in clutch size and egg mass in a wild great tit (*Parus major*) population. *Mol. Ecol. 22: 3949–3962.*

Schielzeth H. & Husby A. 2014: Challenges and prospects in genome-wide quantitative trait loci mapping of standing genetic variation in natural populations. *Ann. N. Y. Acad. Sci. 1320: 35–57.*

Seasholtz T.M., Wessel J., Rao F.W., Rana B.K., Khandrika S., Kennedy B.P., Lillie E.O., Ziegler M.G., Smith D.W., Schork N.J., Brown J.H. & O'Connor D.T. 2006: Rho kinase polymorphism influences blood pressure and systemic vascular resistance in human twins – role of heredity. *Hypertension 47: 937–947.*

Downloaded From: https://bioone.org/journals/Folia-Zoologica on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

Singham G.V., Vargo E.L., Booth W., Othman A.S. & Lee C.Y. 2012: Polymorphic microsatellite loci from an indigenous Asian fungus-growing termite, *Macrotermes gilvus* (Blattodea: Termitidae) and cross amplification in related taxa. *Environ. Entomol. 41: 426–431.*

Slate J. & Pemberton J.M. 2007: Admixture and patterns of linkage disequilibrium in a free-living vertebrate population. *J. Evol. Biol. 20: 1415–1427.*

Slatkin M. 2008: Linkage disequilibrium – understanding the evolutionary past and mapping the medical future. *Nat. Rev. Genet. 9: 477–485.*

Smith B.T., Harvey M.G., Faircloth B.C., Glenn T.C. & Brumfield R.T. 2014: Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. *Syst. Biol. 63: 83–95.*

Smukowski C.S. & Noor M.A. 2011: Recombination rate variation in closely related species. *Heredity 107: 496–508.*

Stinchcombe J.R. & Hoekstra H.E. 2008: Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity 100: 158–170.*

Stoneking M. & Krause J. 2011: Learning about human population history from ancient and modern genomes. *Nat. Rev. Genet. 12: 603–614.*

Suez M., Behdenna A., Brouillet S., Graça P., Higuet D. & Achaz G. 2015: MicNeSs: genotyping microsatellite loci from a collection of (NGS) reads. *Mol. Ecol. Resour.: doi:10.1111/1755-0998.12467.*

Sun J.X., Mullikin J.C., Patterson N. & Reich D.E. 2009: Microsatellites are molecular clocks that support accurate inferences about history. *Mol. Biol. Evol. 26: 1017–1027.*

Sureshkumar S., Todesco M., Schneeberger K., Harilal R., Balasubramanian S. & Weigel D. 2009: A genetic defect caused by a triplet repeat expansion in *Arabidopsis thaliana*. *Science 323: 1060–1063.*

Taguchi M., Shigenobu Y., Ohkubo M., Yanagimoto T., Sugaya T., Nakamura Y., Saitoh K. & Yokawa K. 2013: Characterization of 12 polymorphic microsatellite DNA loci in the blue shark, *Prionace glauca*, isolated by next generation sequencing approach. *Conserv. Genet. Resour. 5: 117–119.*

Tautz D. & Renz M. 1984: Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Res. 12: 4127–4138.*

Templeton A.R. 2010: Coalescent-based, maximum likelihood inference in phylogeography. *Mol. Ecol. 19: 431–435.*

Tiffin P. & Ross-Ibarra J. 2014: Advances and limits of using population genetics to understand local adaptation. *Trends Ecol. Evol. 29: 673–680.*

Tremblay D.C., Alexander G., Moseley S. & Chadwick B.P. 2010: Expression, tandem repeat copy number variation and stability of four macrosatellite arrays in the human genome. *BMC Genomics 11: 632.*

Turner T.L. & Hahn M.W. 2010: Genomic islands of speciation or genomic islands and speciation? *Mol. Ecol. 19: 848–850.*

Uy J.A.C., Moyle R.G., Filardi C.E. & Cheviron Z.A. 2009: Difference in plumage color used in species recognition between incipient species is linked to a single amino acid substitution in the Melanocortin-1 receptor. *Am. Nat. 174: 244–254.*

Wallberg A., Han F., Wellhagen G., Dahle B., Kawata M., Haddad N., Simoes Z.L.P., Allsopp M.H., Kandemir İ., De la Rua P., Pirk C.W. & Webster M.T. 2014: A worldwide survey of genome sequence variation provides insight into the evolutionary history of the honeybee *Apis mellifera*. *Nat. Genet. 46: 1081–1088.*

Weber C., Nabholz B., Romiguier J. & Ellegren H. 2014: Kr/Kc but not dN/dS correlates positively with body mass in birds, raising implications for inferring lineage-specific selection. *Genome Biol. 15: 542.*

Williams E.E. 1969: Ecology of colonization as seen in zoogeography of anoline lizards on small islands. *Q. Rev. Biol. 44: 345–389.*

Wilson M.A. & Echternacht A.C. 1987: Geographic variation in the critical thermal minimum of the anole, *Anolis carolinensis* (Sauria, Iguanidae), along a latitudinal gradient. *Comp. Biochem. Physiol. A 87: 757–760.*

Yukilevich R., Turner T.L., Aoki F., Nuzhdin S.V. & True J.R. 2010: Patterns and processes of genome-wide divergence between North American and African *Drosophila melanogaster*. *Genetics 186: 219–239.*

Zink R.M. 2010: Drawbacks with the use of microsatellites in phylogeography: the song sparrow *Melospiza melodia* as a case study. *J. Avian Biol. 41: 1–7.*

Zink R.M. & Barrowclough G.F. 2008: Mitochondrial DNA under siege in avian phylogeography. *Mol. Ecol. 17: 2107–2121.*

Zweier C., Peippo M.M., Hoyer J., Sousa S., Bottani A., Clayton-Smith J., Reardon W., Saraiva J., Cabral A., Gohring I., Devriendt K., de Ravel T., Bijlsma E.K., Hennekam R.C.M., Orrico A., Cohen M., Dreweke A., Reis A., Nurnberg P. & Rauch A. 2007: Haploinsufficiency of TCF4 causes syndromal mental retardation with intermittent hyperventilation (Pitt-Hopkins syndrome). *Am. J. Hum. Genet. 80: 994–1001.*

Supplementary online materials

Table S1. The full list of articles using microsatellites
(Excel file; URL: http://www.ivb.cz/folia/download/edwards_et_al._table_s1_supplementary_material.xlsx).

Downloaded From: https://bioone.org/journals/Folia-Zoologica on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use