DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY
Office for Scholarly Communication

Next-generation sequencing and the expanding domain of phylogeography

Citation
Edwards, S. V, A. Shultz, and S. C. Campbell-Staton. 2015. Next-generation sequencing and the expanding domain of phylogeography. Folia Zoologica 64 (3): 187–206.

Published Version
http://www.ivb.cz/folia-zoologica-archive.html?vol=64&no=3

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:27716528

Terms of Use
This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

In press, *Folia Zoologica*
Special Issue on Phylogeography

Next-generation sequencing and the expanding domain of phylogeography

Scott V. Edwards[1*], Allison J. Shultz[1] and Shane C. Campbell-Staton[1]

[1]Department of Organismic and Evolutionary Biology, and Museum of Comparative Zoology, Harvard University, Cambridge, MA 02138 USA

*Corresponding author:

Scott Edwards

Department of Organismic and Evolutionary Biology

Harvard University

Cambridge, MA 02138, USA

Email: sedwards@fas.harvard.edu

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 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.

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 and Edwards 2009). However, a more 'free will' perspective might suggest that marker choice is also 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 change. 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

23 phylogeographic questions will be readily available through routine whole-genome resequencing
24 of geographically sampled populations.

25 *Molecular markers and the core concepts of phylogeography*

26 In the 1990s, the polymerase chain reaction (PCR) was the primary driver, and this led to
27 a proliferation of studies focusing on mitochondrial DNA in animals, and on chloroplast DNA in
28 plants. The focus on organellar genomes was not necessarily prescribed by PCR, but the ease of
29 amplification using PCR and the challenges of working routinely with nuclear genes perceived
30 by phylogeographers made organelle genomes a practical focus. MtDNA and cpDNA were also
31 attractive because they were polymorphic at the intraspecific level and experienced little if any
32 recombination, making it straightforward to move from DNA sequence to gene tree without
33 additional data manipulation. With this focus on organellar genomes came the oft-repeated
34 caveats now familiar to every student of phylogeography (Edwards and Bensch 2009): that
35 differences in gene flow between the sexes might cause organelle genomes to recover a biased
36 history of a given species; that the smaller effective population size of organellar genomes might
37 cause the genetic lineages to track population lineages more faithfully than the average marker,
38 sometimes resulting in an overly simplistic view of population history; and, more recently, that
39 natural selection on organelle genomes, particularly mtDNA, might yield estimates of genetic
40 diversity, or spatial patterns in the distribution of that diversity, which do not reflect neutral
41 processes or recent population history (Rand 2001; Excoffier and Ray 2008; Nabholz et al. 2008,
42 2009). Rather, lineage-specific mutation rates, which are hard to predict for a given species from
43 first principles or life history parameters, appear to be the best predictor of variation in genetic
44 diversity across species, at least in birds and mammals. Natural selection has also emerged as a
45 key determinant of mitochondrial diversity within species, as well as of relationships between

46 mtDNA haplotype or codon distributions and latitude or thermal environment, potentially
47 compromising efforts to understand neutral diversity via neutral markers (Gerber et al. 2001;
48 Ballard and Whitlock 2004; Ribeiro et al. 2011; Jobling 2012; Ballard and Pichaud 2014;
49 Morales et al. 2015). Although genetic diversity within populations and species is not always a
50 primary focus of phylogeographic studies, it is certainly a basic descriptor of population history
51 and the forces governing it warrant our attention.

52 The first forays in to the nuclear genome in animal phylogeography came in the early
53 1990s in the form of PCR-amplified nuclear DNA sequences and via microsatellites. Diploid
54 nuclear genes were typically amplified via PCR and sequenced directly (i.e., without cloning
55 first), a practice that led to much hand-wringing over how to determine the phase of nuclear
56 haplotypes comprising the PCR product from heterozygous individuals (Palumbi and Baker
57 1994; Hare and Palumbi 1999). The phase of nuclear alleles is important because only after
58 determining phase is one able to coherently analyze alleles within populations or linked sites,
59 even for PCR products of a few hundred base pairs. Even today, particularly when PCR-
60 amplified nuclear genes are used in phylogenetics, the phase of nuclear alleles is often ignored,
61 possibly because it is not deemed important when comparing highly divergent species.
62 Recombination also had to be acknowledged and often this was accomplished by determining
63 DNA tracts within which no detectable recombination was observed. Detection of
64 recombination events was often accomplished through software focusing on phylogenetic
65 discordances among sites within a sequence or by estimating linkage disequilibrium among sites
66 within or between loci (Hudson and Kaplan 1988). Testing for and dealing with recombination,
67 for example by retaining only sections of an alignment free of detectable recombination, was and
68 continues to be important so as to ensure assumptions are not violated when building gene trees

69 or when estimating population parameters that require either full linkage between sites within
70 loci or complete independence of sites.

71 The specific ways in which sequence data were analyzed were very much constrained by
72 the technical limits of PCR; typically data sets consisted of a few loci, each of a few hundreds
73 base pairs, and an ecosystem of software emerged around this particular format. Discordances
74 between nuclear and mitochondrial DNA came to the fore as researchers were able to directly
75 observe them in phylogenetic and phylogeographic analyses (e.g. Godinho et al. 2008), and in
76 particular in studies of hybrid zones. The effects of incomplete lineage sorting were also readily
77 visible even in analyses employing few loci. PCR of nuclear genes was very much a brute-force
78 operation, the number of loci being assayed directly proportional to the amount of effort and
79 number of PCR experiments performed. At the zenith of the (PCR) nuclear age in
80 phylogeography (Balakrishnan et al. 2010), typical studies included tens of loci, and there was
81 mounting evidence that the uncertainty of estimates of demographic parameters decreased with
82 increasing numbers of loci. In addition, population genetic theory suggested much the same: for
83 example, the optimal sampling scheme for estimating genetic diversity within single populations
84 is generally thought to maximize the number of loci at the expense of the number of individuals
85 (alleles) or length of loci (Nei and Roychoudhury 1974; Pluzhnikov and Donnelly 1996;
86 Felsenstein 2006; Carling and Brumfield 2007). Surprisingly, the optimal sampling scheme for
87 estimating genetic parameters from multiple populations partially linked by gene flow is still
88 understudied. In the PCR era, given constraints on budgets and finite ability to sample
89 individuals, there was a trade-off between the number of individuals or populations sampled and
90 the number of loci assayed. We suggest that this trade-off is partially removed with the advent
91 of next generation sequencing.

Microsatellites and other simple sequence repeats also emerged in phylogeography in 1990s, following their discovery in the previous decade (Tautz and 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. Fst vs. Rst 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 and Barrowclough 2008; Edwards and Bensch 2009; Zink 2010; Albayrak et al. 2012; Perktas et al. 2015). 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

115 (Zink 2010), and include substantial homoplasy, making estimates of the number of mutations
116 difficult; an inability to conduct robust phylogenetic analyses, and hence offering little continuity
117 between phylogeography and phylogenetics; frequent null alleles; and difficulty comparing to
118 sequence-based markers, including mtDNA. Less well-appreciated deficiencies of
119 microsatellites include clear evidence that some simple sequence repeats are indeed functional –
120 they are often involved in gene regulation in both microbes and eukaryotic genomes - and thus
121 may be subject to selection. This last critique no doubt applies to other kinds of markers as well,
122 including SNPs, but the frequent appeal to neutrality by users of microsatellites should be
123 tempered by the increasing number of examples of functional roles for such markers (Liu et al.
124 2000; Metzgar et al. 2000; Sureshkumar et al. 2009; Tremblay et al. 2010; Grover and Sharma
125 2011; Gao et al. 2013).

126 There are no doubt still staunch defenders of microsatellites, and we do not mean to
127 suggest that SNPs, sequence-based markers, or other alternatives to microsatellites are not above
128 reproach. A major criticism of sequence based markers or SNPs in phylogeography has been the
129 paucity of such markers and their low polymorphism. While these criticisms may have been
130 valid in the PCR era, we suggest that they no longer apply meaningfully given the large number
131 of SNPs now achievable with next-generation sequencing approaches. By contrast, although the
132 number of microsatellite loci has been increasing in recent years, we do not know of efforts to
133 assay variation targeted at microsatellites using next-generation approaches. Next-generation
134 isolation of microsatellite loci, followed by PCR-based assays of variation, has been used with
135 considerable success (Abdelkrim et al. 2009; Perry and Rowe 2011; Singham et al. 2012; Curto
136 et al. 2013; Taguchi et al. 2013), but actually assaying variation and scaling up beyond PCR-
137 based assays to our knowledge has not occurred yet for microsatellites in studies of

138 phylogeography (but see Fordyce et al. (2011) for other applications). Although it is surely too
139 early to tell, we suggest that this technical gap implies that the community does not place a high
140 priority on scaling up for microsatellites, perhaps because it is thus far comfortable with the
141 expanded power of SNPs in the next-generation sequencing era. Garrick et al. (2015) recently
142 declared that "Compared to other classes of molecular markers, DNA sequence haplotypes and
143 single nucleotide polymorphisms (SNPs) should be more informative about historical events and
144 processes … operating over timescales most relevant to the discipline [of phylogeography]"
145 (Garrick et al. 2015). While we agree wholeheartedly with this statement, we suggest that much
146 of the community might still favor microsatellites if given the choice, particularly in comparisons
147 of closely related, endangered or very recently diverged populations. This preference, we
148 suspect, is due in part because some labs may not yet have ready access to next-generation
149 sequencing methods. But it might also be due to the perception that, due to their hypervariability,
150 microsatellites have advantages over SNPs in many contexts, especially if they can be assayed in
151 large numbers (Becquet et al. 2007; Kwong and Pemberton 2014).
152 We sought to determine whether changes in methodologies and markers used in
153 phylogeography have been driven by choice or instead more by the availability of technologies
154 adopted primarily for increasing the number of loci. We conducted a study parallel to that of
155 Garrick et al (2015) by reading abstracts for 397 papers that use microsatellites in
156 phylogeography and were published in *Molecular Ecology*, a major outlet for phylogeographic
157 research (Fig. 1, see legend for methods). Garrick et al. (2015)'s survey, which comprised 370
158 papers reporting on 508 single-species data sets, was interesting because it somewhat
159 unexpectedly focused on SNPs, whereas our intuition was that, among nuclear loci,
160 microsatellites were the main driver of phylogeography until recently. Our analysis confirms

161 this suspicion: once the studies that only used mitochondrial DNA were pulled out from their
162 analysis (a total of 280 studies, comprising 73.5 % of all 'SNP' studies in our sample; see Fig. 1),
163 the number of phylogeographic studies using microsats is comparable to, and sometimes exceeds,
164 that using nuclear SNPs (Fig. 1A). Intriguingly, phylogeographic studies employing only
165 mtDNA do indeed seem to be declining since 2007, at least in the pages of *Molecular Ecology*.
166 This decline could highlight a shift in preference within the field towards other types of genetic
167 data or a shift in preference of journals against publishing studies that rely solely on mtDNA.
168 Additionally, the year 2013 suggests a shift as studies employing nuclear SNPs begin to exceed
169 those using microsatellites. This uptick does not seem to be driven entirely by next-generation
170 sequencing, which only comprised 8 studies in our sample, suggesting that SNPs may have risen
171 in popularity independently of novel technologies and perhaps due to conceptual advances or
172 available software. Examination of microsatellite studies in all years of our sample (Fig. 1B)
173 suggest that, at best, this technology has leveled out in its popularity, particularly given the
174 increasing number of pages in the journal over time. It will be interesting to see what the next
175 five years brings in terms of the relative use of these various marker types in phylogeogeography.
176 Given the pre-eminence of *Molecular Ecology* in the field of phylogeography, we suggest that
177 the trends observed here may well reveal trends that will follow in time with the rest of the field.

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

179 Phylogeographers have appreciated for years that, despite their lower polymorphism,
180 SNPs are much more common in the genome than microsatellites (Brumfield et al. 2003). Yet
181 this point was almost moot because it was difficult if not impossible to take advantage of SNPs
182 on a scale that would capitalize on their ubiquity. The advent of next-generation sequencing will
183 likely increase the swing of the phylogeographic pendulum back in favor of SNPs and sequence-

184 based markers once and for all. The adoption of partial genome survey methods such as Rad-seq
185 will not only yield SNPs in sufficient numbers for phylogeography, but will prescribe the use of
186 SNPs even more so than will whole-genome resequencing. Whole-genome phylogeographic
187 studies are already the norm for model species such as humans (Consortium et al. 2010; Reich et
188 al. 2010; Hammer et al. 2011; Li and Durbin 2011; Stoneking and Krause 2011) and *Drosophila*
189 (Yukilevich et al. 2010; Campo et al. 2013; Duchen et al. 2013; Reinhardt et al. 2014) and
190 researchers will have many options for marker types once this phase is achieved. Until that time,
191 by shifting the focus of phylogeography to sequence-based markers and SNPs, next-generation
192 sequencing methods promise to stabilize and unify phylogeographic studies in many productive
193 ways. To us they are a positive trend for phylogeography because of the many reasons that SNPs
194 have previously been considered beneficial: they provide more natural comparisons to variation
195 in organelle genomes and between studies, and, despite the challenges of recombination within
196 loci, provide natural bridges to phylogenetic analysis.
197 *Types and consequences of next-generation sequencing approaches in phylogeography*:
198 Aside from the use of next-generation sequencing approaches for isolating microsatellite loci,
199 next-generation sequencing is making inroads into phylogeography in two main ways: through
200 Rad-seq, which generates short (~100 bp) markers, typically with one or a few SNPs per locus
201 (see Puritz et al. 2014 for a review of different Rad-seq methods); and through targeted capture
202 approaches, which can be used to target already-defined sets of loci, such as exons or
203 ultraconserved elements (UCEs) and their polymorphic flanking regions (Faircloth et al. 2012;
204 Smith et al. 2014). Although transcriptome and amplicon sequencing have also both proven
205 useful in phylogeography (Hedin et al. 2012; O'Neill et al. 2013), transcriptome sequencing will
206 likely have less direct use in purely phylogeographic investigations (as opposed to the discovery

207 of loci under selection; see below) because of its focus on loci that are relatively conserved but
208 more likely under selection, and we predict that amplicon sequencing will ultimately prove less
209 attractive to phylogeographers because of the labor involved and the smaller number of loci that
210 can be assayed (but see Mccormack and Faircloth 2013).

211 The emerging 'core' approaches of targeted enrichment and Rad-seq promise to re-orient
212 phylogeography towards sequence-based markers in different ways because of the types of data
213 they each produce (Lemmon and Lemmon 2012; McCormack et al. 2012; McCormack et al.
214 2013). Targeted enrichment approaches yield data that can be assembled into individual
215 sequence-based markers spanning hundreds to potentially thousands of base pairs, resulting in
216 haplotypes or consensus sequences within which there may be several to many SNPs that can in
217 principle be subjected to phylogenetic analysis (Lemmon and Lemmon 2013). By contrast, Rad-
218 seq typically yields loci that are too short to analyze using traditional phylogenetic methods;
219 instead, researchers typically extract single or multiple SNPs from such Rad-loci and then
220 analyze them as individual SNPs. In many ways the two approaches provide contrasting bridges
221 to phylogenetics and classical phylogeography, as well as pointing to complementary analytical
222 approaches in the future. For example, because the loci yielded by targeted enrichment
223 approaches to phylogeography can often be analyzed using standard phylogenetic methods for
224 estimating gene trees, they provide a natural bridge to classical phylogeography. By contrast,
225 although the SNPs generated by Rad-seq can be used to estimate phylogenetic relationships of
226 populations or species ('species trees'), and indeed have been subjected to concatenation
227 approaches in early examples (Emerson et al. 2010; Merz et al. 2013), currently these markers
228 are used to bypass classical gene trees and instead estimate the species tree directly (Bryant et al.
229 2012; Rheindt et al. 2014; Chifman and Kubatko 2014). These two approaches can sometimes

230 require different sets of analyses, and it may be that the toolkit for linked SNPs such as produced
231 by targeted enrichment is still deeper than that available for analyzing SNPs.
232 Although both core methods will align phylogeography squarely on the use of SNPs,
233 whether linked or unlinked in individual loci, these differences in continuity with classical 'gene
234 tree' phylogeography and analytical approaches are significant. Gene trees may be the lynchpin
235 in this phylogeographic transition. Many have suggested that, despite their centrality to the
236 origins of phylogeography (Avise et al. 1987), ultimately, gene trees are a nuisance parameter in
237 phylogeography and, if anything, can be a distraction from the key levels of analysis and primary
238 interests, which are populations and species, not genes. In this sense, Rad-seq may have the
239 practical advantage of finally freeing the community conceptually from gene trees and haplotype
240 networks, which are still a ubiquitous component of phylogeography. The ability and tendency
241 to make and interpret gene trees has resulted in heated controversies in phylogeography, such as
242 the conflicts between model-based approaches in phylogeography and more literal interpretations
243 of gene trees, such promoted by nested-clade analysis (Nielsen and Beaumont 2009; Beaumont
244 et al. 2010; Templeton 2010). It will be interesting to see how the analytical methods afforded
245 by Rad-seq versus targeted enrichment influence the next ten years of phylogeography. It may
246 be that the sheer number of loci generated by both methods moves the field forward in adopting
247 the model-based approaches that are clearly appropriate for such data sets.
248 *Power of next generation genome subsampling methods for phylogeography*: Genome
249 subsampling methods such as Rad-seq and targeted capture re-sequencing, including UCE
250 analysis, hold enormous promise for phylogeographic investigations of neutral processes such as
251 population structure, species delimitation and historical demography. Phylogeography has
252 traditionally emphasized sampling of individuals and populations over loci (Brito and Edwards

253 2009; Garrick et al. 2015). This bias is a natural and understandable outgrowth of one of the
254 main motivations for phylogeography – to discover novel lineages of organismal biodiversity
255 within species. However, it is now better appreciated that sampling robustly for loci as well as
256 individuals is essential for increased precision of parameter estimates in phylogeography and for
257 better accounting for stochastic variation among loci (Beerli and Felsenstein 1999; Edwards and
258 Beerli 2000; Felsenstein 2001; Jennings and Edwards 2005; Carling and Brumfield 2007). The
259 sheer number of loci revealed by methods such as Rad-seq, UCE analysis or targeted capture
260 methods therefore comes as a welcomed boon over the relative dearth of loci captured in a
261 typical PCR-based study. Considering one of the core goals of phylogeography is to describe the
262 history of populations using genomic data, it should not be surprising that even a dozen loci
263 (which is typical for the heyday of PCR-based phylogeography) are unlikely to capture the
264 diverse signals of history across all chromosomes in a typical genome. In the case of Rad-seq,
265 there is a serious issue involving the bias of the technique against older haplotypes that have
266 experienced mutations in the restriction sites used to isolate DNA fragments – a bias that can
267 compromise estimates of genetic variation (Arnold et al. 2013). Furthermore, the process of
268 assembling a library from non-model species often involves grouping sequence reads by some
269 similarity threshold. The choice of these thresholds is not necessarily a straightforward process,
270 and highly divergent alleles may be inadvertently omitted prior to downstream analyses either
271 because their differences exceed preset similarity thresholds, or they increase the proportion of
272 missing data in the genotype by individual matrix (Harvey et al. 2015a; Huang and Knowles, in
273 press). Still, for other next-gen subsampling approaches, the variation revealed by next-
274 generation approaches – whether amplicon sequencing of ~100 loci or the tens of thousands of

275 SNPs that a typical Rad-seq study reveals – is likely more than adequate for understanding the
276 basic population history of most species.

277 Harvey et al. (2015b) recently compared the resolving power for phylogeography of data
278 sets varying in size in terms of number and length of sequence-based markers to estimate
279 demographic parameters (effective population size, divergence time, migration rate) and species
280 history in a Neotropical songbird with deep phylogeographic breaks. They found that increasing
281 the number of loci up to 5000 provided increased resolution of the particular demographic
282 histories that they studied, but that increasing the number of loci beyond this yielded minimal
283 gains. Additionally, they found that increasing locus length past 500 bp did not yield additional
284 improvements in resolution for the focal parameters of their study. This study therefore
285 suggested that the numbers of loci revealed by genome subsampling methods such as Rad-seq
286 are likely to be adequate for resolving population history on a variety of scales. Indeed, the
287 initial round of empirical studies using data sets produced by Rad-seq or sequence capture,
288 usually on the order of 2000-30,000 SNPs or aligned markers appear quite satisfying in so far as
289 they have revealed undiscovered phylogeographic lineages that significantly improve our
290 understanding of the fit of genomic variation to environmental and topographic barriers to gene
291 flow (e.g., Alcaide et al. 2014; Harvey and Brumfield 2015). We concur with Harvey et al.
292 (2015b) that genome subsampling methods are likely to finally provide the appropriate level of
293 genomic detail for the foreseeable future of phylogeography.

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

296 Classically, phylogeography has focused on the neutral demographic history of species, a

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

320 *Lewontin's paradox and genetic variation within species*: There is a long history of using
321 population genetics to discover loci with a history of selection, beginning with Lewontin and
322 Krakauer's (1973) observation that F_{st} outliers could be useful in identifying loci under selection.
323 The use of F_{st} outliers has become common now, and, although there are caveats to the
324 interpretation of high F_{st} as an unambiguous signal of selection (Turner and Hahn 2010;
325 Cruickshank and Hahn 2014), the ability now to study distributions of loci makes it a useful tool,
326 especially when implemented with care and a consideration of the underlying demographic
327 history (Johnston et al. 2014; Lotterhos and Whitlock 2014). Recent studies suggest that, in fact,
328 avoiding natural selection entirely, even in the nuclear genome, may not be possible, whether
329 studying whole-genome or transcriptome variation, because the imprint of selection may be
330 much greater across the genome than originally envisioned by phylogeographers. For example,
331 the overall level of diversity in the nuclear genome displayed in a species is often taken by
332 phylogeographers to be a neutral indication of the historical effective population size,
333 summarized for single populations by the equation $\pi = 4N\mu$. However, the small range of
334 nuclear genetic diversities across species – usually considered to fall in a range of two orders of
335 magnitude and often called "Lewontin's (1974) paradox" – has been a major challenge for
336 population geneticists and has profound implications for phylogeography as well. The paradox
337 was a major impetus for the development of the nearly neutral theory, which placed emphasis on
338 interactions between selection and drift and seemed to fit available data better than the strictly
339 neutral theory (Ohta 1992; Ohta and Gillespie 1996). By postulating a nearly neutral zone in
340 which the absolute value of the product Ns was substantially less than 1, the nearly neutral
341 theory – clearly a key departure from Kimura's strictly neutral theory (Kimura 1968; Kimura
342 1983) - was able to account in part for this paradox. Although there have been many estimates

343 of the distribution of selection coefficients for key model species (Keightley and Eyre-Walker
344 2010), until recently there were few compelling data to really test these ideas across a wide range
345 of species.

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

358 In another related study, Romiguier et al (2014) recently surveyed population variation in
359 transcriptomes of a variety of species across the tree of life. Like the study by Corbett-Detig et al.
360 (2015) this study, although comprehensive, cannot be considered phylogeography because the
361 main focus was not population history but the overall amount of variation. Whereas in some
362 cases multiple populations per species were sampled and much of the genetic variation within
363 each species may have been captured, in neither study was the general geographic sampling
364 robust enough to be considered phylogeography. Romiguier et al. (2014) came to the startling
365 conclusion that the amount of variation (π) in the transcriptome was best predicted by life history

366 attributes and longevity, rather than by geographic range or other aspects of strictly neutral
367 demography. Surprisingly their reasoning was largely based on a neutral argument: long-lived
368 and other species with K-selected life history traits tend to be able to sustain smaller populations,
369 and hence lower genetic diversities, than species with r-selected life history traits, because they
370 live in more stable environments with fewer long-term perturbations. By contrast, r-selected
371 species, which tend to have greater genetic diversity, possess the large populations that allow
372 them to colonize and persist in unstable habitats. Although ecologists will likely find merit in this
373 hypothesis, because it conceives of ecology and life history as the causal driving force behind
374 population genetic variation, it is not very satisfying from a population genomics perspective. It
375 is surprising that even negative selection on deleterious variation, traditionally considered a
376 major force in regulating intraspecific genetic variation, especially in protein-coding regions,
377 was only briefly mentioned as potentially important for explaining the positive correlation
378 between life history and the ratio of nonsynonymous to synonymous nucleotide substitutions
379 (d_n/d_s) within species. In this case, the smaller populations of long-lived K species result in
380 higher d_n/d_s (driven largely by higher d_n) due to increased fixation of deleterious mutations in
381 small populations, as envisioned by the nearly neutral theory (see also Weber et al. 2014).
382 However, Romiguier et al (2014) suggested that overall levels of synonymous substitution were
383 largely driven by effective population size, which in turn was seen as a neutral consequence of
384 life history variation. It will be important to verify the hypothesis of this work through genome-
385 wide measurements of diversity as a part of detailed phylogeograhic analyses of diverse species.
386 *Selection, recombination and hitchhiking in phylogeography:* The above two studies
387 provide an important contrast in how phylogeographers are beginning to think about genome-
388 wide data. In particular, the issue of linkage disequilibrium (LD) and the potential for genome-

389 wide hitchhiking has emerged as a key factor influencing patterns of variation in the era of
390 whole-genome phylogeography. Aside from a few key emerging models for ecology and
391 evolution, such as sticklebacks, honeybees and other groups (Jones et al. 2012a; Wallberg et al.
392 2014), the phylogeography of non-model species thus far has not dealt substantially with the
393 effects of linkage on genomic variation across geographically sampled sets of populations. In
394 the PCR era, levels of LD were occasionally measured within or between the loci that could be
395 assayed, often to confirm the independence of loci, but in general the data available to
396 phylogeographers was not comprehensive enough to provide meaningful insight into the effects
397 of hitchhiking on genome-wide variation. Levels of LD can be influenced by many factors,
398 including neutral processes such as genetic drift and population bottlenecks; selective processes
399 like selective sweeps and balancing selection; and genetic processes like rates of recombination
400 and mutation (Slatkin 2008). The population recombination rate ($\rho = 4Nc$) has been measured in
401 relatively few non-model species, whereas there are numerous estimates for populations of
402 humans, mice or *Drosophila* (Smukowski and Noor 2011).

403 Population geneticists have been measuring LD for decades, and one can calculate the
404 LD or r^2 value between any pair of markers, regardless of how densely the genome is sampled or
405 whether the markers are on the same or different chromosomes. In the pre-genomic era, when
406 LD values were calculated between loci in studies that only sparsely sampled the genome, the
407 motivation was often to study the action of natural selection. But this was a very hit or miss
408 endeavor: if the candidate genes between which LD was calculated were not involved in
409 selective processes, the result was often underwhelming. By contrast, in the era of genomics,
410 when the genome can be sampled much more densely, in principle one need not know the
411 candidate genes under selection: LD can be exploited to discover loci that are the *actual targets*

412 *of selection* without a priori suspicion that the loci measured are under selection themselves
413 (Slatkin 2008). Correlated patterns variation across the genome due to hitchhiking have been
414 used increasingly to produce a set of candidate genes responding to selection, without those
415 candidates having been assayed directly. The candidate genes are usually usually physically
416 close to or in the same linkage blocks as, the markers directly measured, and the assumption is
417 that hitchhiking on the actual targets is causing departures from neutrality, such as high F_{st}, in the
418 assayed SNPs. This protocol, variably called selection mapping or association mapping, has
419 resulted in the generation of hundreds of candidate genes in emerging model species that may be
420 responding to environmental or other selective pressures (Hohenlohe et al. 2012). Increasingly
421 studies are also taking advantage of the high LD created when two species or populations
422 hybridize: chromosomal blocks deriving from each parental population can be used to identify
423 genomic regions that underlie phenotypic traits in those populations in hybrids. This method,
424 often called admixture mapping, has been used in humans extensively and increasingly in non-
425 model species (Slate and Pemberton 2007; Pallares et al. 2014)

426 However, at the dawn of the whole genome era, such protocols can be challenging to
427 implement, and can get stuck between a rock and a hard place (Fig. 2). On the one hand,
428 methods such as Rad-seq, although delivering a large number of SNPs, still sample the genome
429 sparsely, and can fall short of this goal, particularly in species where the population
430 recombination rate is high. This failure to identify actual targets of selection through selection
431 mapping arises because the genome-wide levels of LD can be so low as to cause the actual
432 targets of selection to be effectively unlinked (in low or average LD with) the nearest neutral site
433 whose variation is interrogated. The result will be little evidence for selection among those
434 genomic SNPs that are assayed, and many targets of selection will be missed. Clades such as

435 *Drosophila*, or many bird species, likely fall into this category, and may require whole-genome
436 resequencing to more confidently identify targets of selection through hitchhiking (Backstrom et
437 al. 2006; Ellegren et al. 2012, see Fig. 3; Backstrom et al. 2013). The advantage to such species
438 with high recombination rates is that, when an outlier locus is detected, one can be sure that one
439 is relatively close to the actual target of selection, although even whole genome resequencing
440 studies in *Drosophila* have sometimes yielded mixed results, especially if selection is weak, very
441 recent, or acting on standing variation. On the other hand, levels of genome-wide LD in a given
442 species may be substantial because of recent population history, a history of domestication or an
443 overall low population recombination rate. Canids are good examples of this pattern (see Fig. 3
444 and Boyko et al. 2010; Boyko 2011). In such cases, even sparse sampling of the genome will
445 often uncover sites that appear to be under selection, having hitchhiked with the actual targets
446 that could be megabases away. When LD is high, larger regions of the chromosome are dragged
447 along by hitchhiking than when LD is small, and these large regions can sometimes capture
448 hundreds of genes. But in this situation, the list of candidate genes will be so long that it becomes
449 less than useful. Threespine sticklebacks (*Gasterosteus aculeatus*) may fall in this category:
450 Rad-seq studies routinely identify F_{st} outliers but the list of genes within linkage blocks can be
451 long and often provide only a vague idea of actual targets of selection (Hohenlohe et al. 2012).
452 The timing and strength of the selection event can also be important for regulating the size of the
453 hitchhiking chromosomal segments. Chromosome inversions will also protect blocks of the
454 genome from recombination, causing hundreds of genes to remain in high LD and making
455 identification of the actual targets of selection challenging or impossible without further methods
456 development. Although genome subsampling methods such as Rad-seq will in general provide a

457 coarser picture of hitchhiking loci, even whole genome re-sequencing will not be able to
458 unambiguously identify the actual targets of selection if LD is high.

459 **Examples: Rad-seq meets selection mapping in natural populations**

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

464 *Adaptation and the evolution of cold tolerance in Green Anole lizards*: The green anole
465 lizard, *Anolis carolinensis*, is an ideal species to explore the molecular basis of climate-mediated
466 local adaptation. The only anole native to the continental United States, this species occupies the
467 highest latitudes of any of the nearly 400 species of the genus. The northern edge of the species'
468 distribution is likely limited by winter temperatures (Williams 1969), but populations do not
469 hibernate, as is common for most mid- and high- latitude reptiles. By retreating to sheltered sites
470 and basking during sun exposure, northern populations are able to remain active and periodically
471 feed in the winter months (Bishop and Echternacht 2004), despite regular ambient temperatures
472 below freezing. Additionally, populations from different climates display significant differences
473 in cold tolerance (Wilson and Echternacht 1987). The recent publication of the genome of this
474 species (Alfoldi et al. 2011) provides a unique resource for understanding the molecular
475 processes of evolutionary response to local environment and for identifying genes that may play
476 a key role in physiological divergence between populations of a non-model species. Taking
477 advantage of this opportunity, we used double digest RAD sequencing (ddRad-seq; Peterson et al.

478 2012) to identify regions of the *A. carolinensis* genome associated with cold variation across the
479 species' range.
480 Using SphI and EcoRI restriction enzymes, we digested genomic DNA from 28
481 individuals representing 6 populations spanning the latitudinal extent of the natural range of *A.*
482 *carolinensis*. We genotyped 20,282 SNPs with a coverage of ≥ 10X for these individuals using
483 the Stacks software package (Catchen et al. 2011; Catchen et al. 2013). To search for regions of
484 the genome associated with temperature variation across the species' range, we used
485 georeferenced locality data for each population to retrieve estimates of the mean temperature of
486 the coldest quarter of the year (BIO11) from the Worldclim database (Hijmans et al. 2005). We
487 then used allele counts from the RADseq dataset to calculate Bayes factor associations and
488 Pearson correlations using the Bayenv2 software package (Gunther and Coop 2013). Variant
489 sites in the top 1% of both Bayes factor associations and Pearson correlations were retained as
490 candidate markers identifying regions of the genome that may be important for local adaptation
491 to cold (Fig. 4). This analysis resulted in 72 candidate SNPs, all of which were noncoding: 67%
492 are located in intergenic regions, whereas 33% map to introns.
493 Several genes in this dataset may be of interest for further study due to their close
494 proximity to SNPs exhibiting geographic correlations with temperature and their potential
495 involvement in oxygen regulation, which has been proposed as a major constraint for ectotherms
496 under extreme temperature challenge (Portner et al. 2006; Portner et al. 2007). One of these
497 variants is located 40.8kb upstream from the first exon of Rho-associated protein kinase 2
498 (ROCK 2), whose signaling is important for regulation of pulmonary vasculature (Riento and
499 Ridley 2003; Noma et al. 2006; Seasholtz et al. 2006; Rankinen et al. 2008). Another SNP lies
500 1.46 kb upstream of the first exon of transcription factor 4 (TCF4), which is involved in

501 regulation of breathing patterns (Zweier et al. 2007). Functional genomics studies are needed to
502 better understand the potential role and importance of these and other physiological processes to
503 temperature-mediated local adaptation within the green anole.
504 *Temporal evolution of House Finch populations before and after an epizootic*: The
505 House Finch (*Haemorhous mexicanus*), one of the most common birds in both urban and rural
506 environments in North America, is rapidly becoming a model system for avian study. It has been
507 important in studies of rapid morphological adaptation, sexual selection, evolution of disease
508 resistance, and invasion (Badyaev et al. 2012). The uniqueness of the House Finch in studies of
509 disease ecology is the result of its relationship with the pathogen *Mycoplasma gallisepticum*
510 (MG). This poultry-associated bacterium was first documented in the House Finch in 1994 in the
511 Washington DC area (Ley et al. 1996; Hochachka and Dhondt 2000a). MG infects the
512 respiratory tract and causes severe conjunctivitis (Hochachka and Dhondt 2000b), suppresses
513 pathogen-specific components of the immune system (Bonneaud et al. 2011), and stimulates
514 inflammatory responses (Gaunson et al. 2006; Mohammed et al. 2007; Adelman et al. 2013). The
515 pathogen spread through the eastern population rapidly, and by 1998 had caused severe declines
516 across the region, as high as 60% in some areas (Dhondt et al. 1998). Infection experiments
517 comparing gene expression responses of eastern individuals with 12 years of exposure and
518 historically unexposed individuals suggested rapid evolution of gene expression, disease
519 resistance (Bonneaud et al. 2011; Bonneaud et al. 2012) and disease tolerance (Adelman et al.
520 2013).
521 We collected a genome-wide SNP dataset using double-digest RADseq (Peterson et al.
522 2012) to identify regions of the genome with signatures of MG-mediated selection over time. As
523 in the *Anolis* study, we digested the genome with SphI and EcoR1, and selected fragments from

524 276-324 base pairs long to recover homologous loci scattered randomly across the genome. In
525 this preliminary study, we sampled only 11 individuals (22 chromosomes), 5 from pre-epizootic
526 (1990) and 6 from post-epizootic (2003) populations from Alabama. We ran the Rad-seq
527 libraries on a single lane of HiSeq Illumina 2500, generating a total of ~8 million paired-end
528 reads (~4 million pairs), each 150 bp long. Using the Stacks pipeline (Catchen et al. 2011;
529 Catchen et al. 2013), we genotyped over 12,000 SNPs from 2223 loci (Fig. 5). Of the 7260
530 comparisons of SNPs achieving our quality threshold between these time periods, we found 167
531 (2.3%) significant shifts in allele frequency (Fisher's exact test p-value < 0.05) from 129 unique
532 loci. Of these 129 loci, 68.2% are in intergenic regions, 29.5% are in introns, 2 loci fall within
533 exons, and 1 within a 3' UTR region. Although none of these loci retain significance with
534 Bonferroni correction, we suspect that larger sample sizes of individuals from additional
535 localities will improve detection. F_{st} values in this collection of SNPs range from 0.208 to 1
536 (fixed differences). These regions with high F_{st} are in or near genes with a variety of functions.
537 One gene, PPP2R2C is involved in immune pathways in humans, and falls 13 kb away from a
538 SNP with an F_{st} of 1.

539 These two studies illustrate the promise but also the challenges of detecting selection and
540 identifying candidate genes in vertebrate genomes using a genome subsampling method such as
541 Rad-seq (Tiffin and Ross-Ibarra 2014). In both studies, most of the Rad-seq SNPs fell in
542 noncoding regions whose relationship to nearby genes was unclear. In the example from *Anolis*,
543 the candidate genes identified as being the closest to those SNPs that were correlated with
544 environmental variables were often quite far away from the SNP used to tag them. In the
545 example from House Finches, the number of F_{st} outliers in comparisons of pre- and post-
546 epizootic populations was relatively small, perhaps because the observed level of LD in House

547 Finches is generally small, and certainly because of our small sample sizes. In the few avian
548 species that have been studied with regard to recombination rate, rates on the autosomes are
549 likely to be quite high, with levels of LD declining rapidly as one moves away from the focal
550 SNP (Backstrom et al. 2006; Bullaughey et al. 2008; Janes et al. 2009; Li and Merila 2010;
551 Ellegren 2014). We have found that LD in songbird populations, such as Red-winged Blackbirds
552 (*Agelaius phoeniceus*) and House Finches, falls off rapidly after a few hundred base pairs, a
553 situation very reminiscent of populations of *Drosophila* (Edwards and Dillon 2004; Backstrom et
554 al. 2013). In such species, LD often declines between SNPs less than 500 bp apart, which means
555 that any SNP found to exhibit high differentiation or signatures of natural selection is unlikely to
556 be useful in identifying candidate genes for a phenotypic trait even a few kb away. Thus it is
557 unclear whether recent proposals to map QTL in natural populations using SNP-chips containing
558 on the order of 10,000 SNPs will be effective (Hagen et al. 2013).
559 Recent statistical models promise new power to estimate the distribution of effect sizes of
560 loci underlying a phenotypic trait along whole chromosomes in natural populations. For
561 example, using approximately 10,000 SNPs, Santure et al. (2013) found that for most
562 chromosomes in the genome of Great Tits studied at Whyndham Woods UK, a model in which
563 the effect of a given chromosome on continuous traits like wing length and clutch size was
564 proportional to the length of that chromosome, and hence the number of genes on that
565 chromosome. By extension, this result suggests that nearly every gene has a similar – and
566 infinitesimally small – effect on variation in the focal trait. But it is also unclear whether the
567 failure to reject such a null hypothesis is also due to the relatively meager sampling of the
568 genome. Although 10,000 SNPs seems like a large number, it is relatively small in terms of

569 capturing variation in genomic blocks in high LD, especially for vertebrate genomes on the order
570 of 1-3 Gb and in species with high population recombination rates (Edwards 2013).

571 Measuring hitchhiking is crucial enough to the expanding domain of phylogeography that
572 we predict that, ultimately, the field will forsake genome subsampling approaches for
573 phylogeographically informed whole-genome resequencing studies. We are beginning to see the
574 first glimpses of such studies in genomically unstudied species (e.g., Ojeda et al. 2014), and the
575 results are as exciting as they are informative about the determinants of genomic variation and
576 structure.

577 *From phylogeography to genotype to phenotype*: Loci whose variation has been
578 influenced by natural selection are often also loci that underlie phenotypic traits. The search for
579 loci underlying natural variation in phenotypic traits is a major thrust of modern evolutionary
580 biology (Hoekstra et al. 2006; Hoekstra and Coyne 2007; Ellegren and Sheldon 2008; Rebeiz et
581 al. 2009; Hiller et al. 2012; Jones et al. 2012b). Of the many methods for identifying such loci –
582 including QTL and linkage mapping using pedigrees and crosses, or "systems genetics", in
583 which multiple kinds of genomic, transcriptomic, and metabolomic data are integrated together
584 (Feltus 2014) – association mapping is probably the method most closely allied to
585 phylogeography. This alliance arises because association mapping uses population comparisons
586 to find genomic loci that correlate with the distribution of a particular phenotype (Stinchcombe
587 and Hoekstra 2008; reviewed in Kratochwil and Meyer 2015). Others have made a similar
588 connection between landscape genetics and the search for loci underlying adaptive phenotypes
589 (Jones et al. 2013). Association mapping has significant promise for closing the genotype-
590 phenotype gap in non-model species and in many contexts has more statistical power than does
591 mapping with pedigrees or controlled crosses (Schielzeth and Husby 2014). Indeed, the

592 emerging trend is toward simultaneous inference of population history and identification of loci
593 associated with phenotypic traits (e.g., Fumigalli et al. 2011; Linnen et al. 2013). Association
594 mapping is likely to be most powerful in situations where most of the genomic variation is
595 shared among a group of closely related populations or species, perhaps connected by high gene
596 flow, but there exist marked phenotypic differences among those populations (Axelsson et al.
597 2013; Cullingham et al. 2014; Schielzeth and Husby 2014). In such situations, association
598 mapping should reveal similar allele frequencies across nearly the entire genome, often due to
599 shared standing variation, in both control and comparison populations, yet these populations
600 should differ at loci whose variation correlates with the divergent phenotypes of interest.
601 Precisely this situation has been found in those emerging cases where association mapping or
602 candidate gene investigation has proved useful in non-model species, including several plants
603 (Comeault et al. 2014; Cullingham et al. 2014; Johnston et al. 2014; Pearse et al. 2014; Roesti et
604 al. 2014). Use of candidate genes in such situations can also be highly informative (Uy et al.
605 2009). Indeed, when populations have experienced a moderate history of divergence, such that
606 the compared genomes differ at many sites due to neutral demographic divergence, it becomes
607 essential to correct for such substructuring so as not to spuriously implicate loci showing strong
608 allele frequency differences in the origin of phenotypic traits that between those populations
609 (Pritchard et al. 2000; Patterson et al. 2006; Price et al. 2006; Price et al. 2010).

610 **Conclusions**

611 This is an exciting time for phylogeography. We have a growing number of examples of
612 studies in which employment of next-generation sequencing methods has yielded high resolution
613 substructuring within species at a level of detail that far exceeds that formerly yielded by single
614 locus mtDNA or microsatellite studies. It is now clear that a few thousand loci such as is

615 typically yielded by methods such as Rad-seq may well be adequate to discover the major
616 phylogeographic lineages within a species. Although the finer details of the history of species
617 can always be clarified further with increasing genomic sampling (or individual sampling), many
618 of these details are likely lost to historical reconstruction because of their age or their mild
619 imprint on the pattern of genome-wide variation. While we expect whole-genome resequencing
620 to become more common in phylogeography, it is unclear whether this method will be overkill if
621 the focus is strictly on the core purview of phylogeography, namely reconstruction of the
622 phylogenetic lineages and neutral demography within species. What is clear is that next-
623 generation methods are causing a resurgence in the use of SNPs as opposed to microsatellites,
624 enabling easier comparisons among loci and species and providing a uniform framework for
625 comparative phylogeography (Hickerson et al. 2010; Andrew et al. 2013).
626 But next-generation sequencing has also broken down the conceptual edges of
627 phylogeography, resulting in an expanded purview that has blurred the lines between core
628 phylogeographic foci and other areas of related interest, such as identifying loci with a history of
629 natural selection and underlying variation in adaptive traits. Studies combining historical
630 reconstruction of phylogeographic history and a search for such adaptive loci are becoming more
631 common (Deagle et al. 2012; Jones et al. 2012a; Jones et al. 2012b; Pearse et al. 2014; Wallberg
632 et al. 2014), and the goal of identifying loci underlying adaptive traits is often as important as
633 understanding the demographic history of a set of populations. This conceptually expanded
634 phylogeography marks an important phase in the evolution of the field, and although always a
635 glimmer in the eye of phylogeographers, has arguably been driven due to the recent arrival of
636 high-throughput genomic approaches. As phylogeography expands its purview, it is becoming
637 clear that genome subsampling methods such as Rad-seq, while extremely powerful for

638 identifying phylogeographic clusters within species, may be inadequate for identifying loci that
639 are targets of natural selection or linked to the true targets. Whole-genome resequencing will
640 likely emerge as the standard tool, if not for traditional phylogeographic investigations, then
641 certainly in the quest for loci underlying quantitative traits in natural populations and their
642 history of divergence within species.
643
644 **Acknowledgements**: We thank Utku Perktas for inviting us to contribute to this volume.
645 We thank Tim Sackton, Ryan Garrick and Bryan Carstens for helpful discussion and sharing of
646 data, and two anonymous reviewers and Ryan Garrick for helpful comments on the manuscript.
647

648

649 **References**

650 Abdelkrim, J., B. Robertson, J.-A. Stanton, and N. Gemmell. 2009. Fast, cost-effective
651 development of species-specific microsatellite markers by genomic sequencing.
652 BioTechniques 46:185–192.
653 Adelman, J. S., L. Kirkpatrick, J. L. Grodio, and D. M. Hawley. 2013. House Finch populations
654 differ in early inflammatory signaling and pathogen tolerance at the peak of Mycoplasma
655 gallisepticum infection. The American Naturalist 181:674-689.
656 Albayrak, T., J. Gonzalez, S. V. Drovetski, and M. Wink. 2012. Phylogeography and population
657 structure of Kruper's Nuthatch Sitta krueperi from Turkey based on microsatellites and
658 mitochondrial DNA. J Ornithol 153:405-411.
659 Alcaide, M., E. S. C. Scordato, T. D. Price, and D. E. Irwin. 2014. Genomic divergence in a ring
660 species complex. Nature 511:83-85.
661 Andrew, R. L., L. Bernatchez, A. Bonin, C. A. Buerkle, B. C. Carstens, B. C. Emerson, D.
662 Garant, T. Giraud, N. C. Kane, S. M. Rogers, J. Slate, H. Smith, V. L. Sork, G. N. Stone,
663 T. H. Vines, L. Waits, A. Widmer, and L. H. Rieseberg. 2013. A road map for molecular
664 ecology. Mol Ecol 22:2605-2626.
665 Arnold, B., R. B. Corbett-Detig, D. Hartl, and K. Bomblies. 2013. RADseq underestimates
666 diversity and introduces genealogical biases due to nonrandom haplotype sampling. Mol
667 Ecol 22:3179-3190.
668 Avise, J. C., J. Arnold, R. M. Ball, E. Bermingham, T. Lamb, J. E. Neigel, C. A. Reeb, and N. C.
669 Saunders. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between
670 population genetics and systematics. Annu Rev Ecol Syst 18:489-522.
671 Axelsson, E., A. Ratnakumar, M. L. Arendt, K. Maqbool, M. T. Webster, M. Perloski, O. Liberg,
672 J. M. Arnemo, A. Hedhammar, and K. Lindblad-Toh. 2013. The genomic signature of
673 dog domestication reveals adaptation to a starch-rich diet. Nature 495:360-364.
674 Backstrom, N., A. Qvarnstrom, L. Gustafsson, and H. Ellegren. 2006. Levels of linkage
675 disequilibrium in a wild bird population. Biol Lett 2:435-438.
676 Backstrom, N., D. Shipilina, M. P. K. Blom, and S. V. Edwards. 2013. Cis-regulatory sequence
677 variation and association with Mycoplasma load in natural populations of the house finch
678 (Carpodacus mexicanus). Ecology and Evolution 3:655-666.
679 Badyaev, A. V., V. Belloni, and G. E. Hill. 2012. House Finch (Haemorhous mexicanus)*in* A.
680 Poole, ed. Ithaca, Cornell Lab of Ornithology.
681 Balakrishnan, C. N., L. J. Y., and S. V. Edwards. 2010. Phylogeography and phylogenetics in the
682 nuclear age, Pages 65-88 *in* P. Grant, and R. Grant, eds. Searching for the Causes of
683 Evolution: From Field Observations to Mechanisms. Princeton, NJ, Princeton University
684 Press.
685 Ballard, J. W. O., and N. Pichaud. 2014. Mitochondrial DNA: more than an evolutionary
686 bystander. Funct Ecol 28:218-231.
687 Ballard, J. W. O., and M. C. Whitlock. 2004. The incomplete natural history of mitochondria.
688 Mol Ecol 13:729-744.
689 Beaumont, M. A., R. Nielsen, C. Robert, J. Hey, O. Gaggiotti, L. Knowles, A. Estoup, M.
690 Panchal, J. Corander, M. Hickerson, S. A. Sisson, N. Fagundes, L. Chikhi, P. Beerli, R.
691 Vitalis, J. M. Cornuet, J. Huelsenbeck, M. Foll, Z. H. Yang, F. Rousset, D. Balding, and

692 L. Excoffier. 2010. In defence of model-based inference in phylogeography REPLY. Mol
693 Ecol 19:436-446.
694 Becquet, C., N. Patterson, A. C. Stone, M. Przeworski, and D. Reich. 2007. Genetic structure of
695 chimpanzee populations. PLoS Genet 3:e66.
696 Beerli, P., and J. Felsenstein. 1999. Maximum-likelihood estimation of migration rates and
697 effective population numbers in two populations using a coalescent approach. Genetics
698 152:763-773.
699 Bonneaud, C., S. L. Balenger, A. F. Russell, J. Zhang, G. E. Hill, and S. V. Edwards. 2011.
700 Rapid evolution of disease resistance is accompanied by functional changes in gene
701 expression in a wild bird. Proceedings of the National Academy of Sciences 108:7866-
702 7871.
703 Bonneaud, C., S. L. Balenger, J. Zhang, S. V. Edwards, and G. E. Hill. 2012. Innate immunity
704 and the evolution of resistance to an emerging infectious disease in a wild bird. Mol Ecol
705 21:2628-2639.
706 Boyko, A. R. 2011. The domestic dog: man's best friend in the genomic era. Genome Biol
707 12:216.
708 Boyko, A. R., P. Quignon, L. Li, J. J. Schoenebeck, J. D. Degenhardt, K. E. Lohmueller, K. Y.
709 Zhao, A. Brisbin, H. G. Parker, B. M. vonHoldt, M. Cargill, A. Auton, A. Reynolds, A. G.
710 Elkahloun, M. Castelhano, D. S. Mosher, N. B. Sutter, G. S. Johnson, J. Novembre, M. J.
711 Hubisz, A. Siepel, R. K. Wayne, C. D. Bustamante, and E. A. Ostrander. 2010. A Simple
712 Genetic Architecture Underlies Morphological Variation in Dogs. Plos Biol 8:e10000451.
713 Brito, P., and S. V. Edwards. 2009. Multilocus phylogeography and phylogenetics using
714 sequence-based markers. Genetica 135:439-455.
715 Brumfield, R., D. A. Nickerson, P. Beerli, and S. V. Edwards. 2003. The utility of single
716 nucleotide polymorphisms in inferences of population history. Trends in Ecology and
717 Evolution 18:249-256.
718 Bryant, D., R. Bouckaert, J. Felsenstein, N. A. Rosenberg, and A. RoyChoudhury. 2012.
719 Inferring Species Trees Directly from Biallelic Genetic Markers: Bypassing Gene Trees
720 in a Full Coalescent Analysis. Mol Biol Evol 29:1917-1932.
721 Bullaughey, K., M. Przeworski, and G. Coop. 2008. No effect of recombination on the efficacy
722 of natural selection in primates. Genome Res 18:544-554.
723 Burke, T., N. B. Davies, M. W. Bruford, and B. J. Hatchwell. 1989. Parental Care and Mating-
724 Behavior of Polyandrous Dunnocks Prunella-Modularis Related to Paternity by DNA
725 Fingerprinting. Nature 338:249-251.
726 Campo, D., K. Lehmann, C. Fjeldsted, T. Souaiaia, J. Kao, and S. V. Nuzhdin. 2013. Whole-
727 genome sequencing of two North American Drosophila melanogasterpopulations reveals
728 genetic differentiation and positive selection. Mol Ecol 22:5084-5097.
729 Carling, M. D., and R. T. Brumfield. 2007. Gene sampling strategies for multi-locus population
730 estimates of genetic diversity (theta). Plos One 2:e160.
731 Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko. 2013. Stacks: an
732 analysis tool set for population genomics. Mol Ecol 22:3124-3140.
733 Catchen, J. M., A. Amores, P. Hohenlohe, W. Cresko, and J. H. Postlethwait. 2011. Stacks:
734 building and genotyping Loci de novo from short-read sequences. G3 1:171-182.
735 Chifman, J., and L. Kubatko. 2014. Quartet inference from SNP data under the coalescent model.
736 Bioinformatics 30:3317-3324.

737 Comeault, A. A., V. Soria-Carrasco, Z. Gompert, T. E. Farkas, C. A. Buerkle, T. L. Parchman,
738 and P. Nosil. 2014. Genome-Wide Association Mapping of Phenotypic Traits Subject to
739 a Range of Intensities of Natural Selection in Timema cristinae*. Am Nat 183:711-727.
740 Consortium, G. P., G. R. Abecasis, D. Altshuler, A. Auton, L. D. Brooks, R. M. Durbin, R. A.
741 Gibbs, M. E. Hurles, and G. A. McVean. 2010. A map of human genome variation from
742 population-scale sequencing. Nature 467:1061-1073.
743 Corbett-Detig, R. B., D. L. Hartl, and T. B. Sackton. 2015. Natural selection constrains neutral
744 diversity across a wide range of species. Plos Biol 13:e1002112.
745 Cruickshank, T. E., and M. W. Hahn. 2014. Reanalysis suggests that genomic islands of
746 speciation are due to reduced diversity, not reduced gene flow. Mol Ecol 23:3133-3157.
747 Cullingham, C. I., J. E. K. Cooke, and D. W. Coltman. 2014. Cross-species outlier detection
748 reveals different evolutionary pressures between sister species. New Phytologist 204:215-
749 229.
750 Curto, M. A., L. R. Tembrock, P. Puppo, M. Nogueira, M. P. Simmons, and H. Meimberg. 2013.
751 Evaluation of microsatellites of Catha edulis (qat; Celastraceae) identified using
752 pyrosequencing. Biochem Syst Ecol 49:1-9.
753 Deagle, B. E., F. C. Jones, Y. G. F. Chan, D. M. Absher, D. M. Kingsley, and T. E. Reimchen.
754 2012. Population genomics of parallel phenotypic evolution in stickleback across stream-
755 lake ecological transitions. P R Soc B 279:1277-1286.
756 Dhondt, A. A., D. L. Tessaglia, and R. L. Slothower. 1998. Epidemic mycoplasmal conjunctivitis
757 in house finches from eastern North America. J. Wildlife Dis. 34:265-280.
758 Duchen, P., D. Živković, S. Hutter, W. Stephan, and S. Laurent. 2013. Demographic inference
759 reveals African and European admixture in the North American Drosophila melanogaster
760 population. Genetics 193:291-301.
761 Edwards, S., and S. Bensch. 2009. Looking forwards or looking backwards in avian
762 phylogeography? A comment on Zink and Barrowclough 2008. Mol Ecol 18:2930-2933.
763 Edwards, S. V. 2013. Next-generation QTL mapping: crowdsourcing SNPs, without pedigrees.
764 Mol Ecol 22:3885-3887.
765 Edwards, S. V., and P. Beerli. 2000. Perspective: Gene divergence, population divergence, and
766 the variance in coalescence time in phylogeographic studies. Evolution 54:1839-1854.
767 Edwards, S. V., and M. Dillon. 2004. Hitchhiking and recombination in birds: evidence from
768 Mhc-linked and unlinked loci in red-winged blackbirds (*Agelaius phoeniceus)*. Genet Res
769 84:175-192.
770 Ellegren, H. 2014. Genome sequencing and population genomics in non-model organisms.
771 Trends Ecol Evol 29:51-63.
772 Ellegren, H., and B. C. Sheldon. 2008. Genetic basis of fitness differences in natural populations.
773 Nature 452:169-175.
774 Ellegren, H., L. Smeds, R. Burri, P. I. Olason, N. Backstrom, T. Kawakami, A. Kunstner, H.
775 Makinen, K. Nadachowska-Brzyska, A. Qvarnstrom, S. Uebbing, and J. B. W. Wolf.
776 2012. The genomic landscape of species divergence in Ficedula flycatchers. Nature
777 491:756-760.
778 Emerson, K. J., Merz, C. R., Catchen, J. M., Hohenlohe, P. a, Cresko, W. a, Bradshaw, W. E., &
779 Holzapfel, C. M. (2010). Resolving postglacial phylogeography using high-throughput
780 sequencing. Proceedings of the National Academy of Sciences of the United States of
781 America, 107(37), 16196–16200. doi:10.1073/pnas.1006538107

782 Excoffier, L., and N. Ray. 2008. Surfing during population expansions promotes genetic
783 revolutions and structuration. TREE 23:347-351.
784 Faircloth, B. C., J. E. Mccormack, N. G. Crawford, M. G. Harvey, R. T. Brumfield, and T. C.
785 Glenn. 2012. Ultraconserved elements anchor thousands of genetic markers spanning
786 multiple evolutionary timescales. Syst Biol 61:717–726.
787 Felsenstein, J. 2001. Accuracy of coalescent likelihood estimates: do we need more sites, more
788 sequences, or more loci? Mol Biol Evol 23:691-700.
789 Felsenstein, J. 2006. Accuracy of coalescent likelihood estimates: do we need more sites, more
790 sequences, or more loci? Mol Biol Evol 23:691-700.
791 Feltus, F. A. 2014. Systems genetics: a paradigm to improve discovery of candidate genes and
792 mechanisms underlying complex traits. Plant science : an international journal of
793 experimental plant biology 223:45-48.
794 Fordyce, S. L., M. C. Ávila-Arcos, E. Rockenbauer, C. Børsting, R. Frank-Hansen, F. T.
795 Petersen, E. Willerslev, A. J. Hansen, N. Morling, and M. T. P. Gilbert. 2011. High-
796 throughput sequencing of core STR loci for forensic genetic investigations using the
797 Roche Genome Sequencer FLX platform. BioTech. 51:127–133.
798 Fumagalli, M., Sironi, M., Pozzoli, U., Ferrer-Admetlla, A., Ferrer-Admettla, A., Pattini, L., &
799 Nielsen, R. 2011. Signatures of environmental genetic adaptation pinpoint pathogens as
800 the main selective pressure through human evolution. PLoS Genetics, 7(11), e1002355.
801 doi:10.1371/journal.pgen.1002355
802 Gao, C. H., X. D. Ren, A. S. Mason, J. N. Li, W. Wang, M. L. Xiao, and D. H. Fu. 2013.
803 Revisiting an important component of plant genomes: microsatellites. Functional Plant
804 Biology 40:645-661.
805 Garrick, R. C., I. A. S. Bonatelli, C. Hyseni, A. Morales, T. A. Pelletier, M. F. Perez, E. Rice, J.
806 D. Satler, R. E. Symula, M. T. C. Thomé, and B. C. Carstens. 2015. The evolution of
807 phylogeographic datasets. Mol. Ecol. 24:1164-1171.
808 Gaunson, J. E., C. J. Philip, K. G. Whithear, and G. F. Browning. 2006. The cellular immune
809 response in the tracheal mucosa to Mycoplasma gallisepticum in vaccinated and
810 unvaccinated chickens in the acute and chronic stages of disease. Vaccine 24:2627-2633.
811 Gerber, A. S., R. Loggins, S. Kumar, and T. E. Dowling. 2001. Does nonneutral evolution shape
812 observed patterns of DNA variation in animal mitochondrial genomes? Annual Review
813 of Genetics 35:539-566.
814 Godinho, R., E. Crespo, and N. Ferrand. 2008. The limits of mtDNA phylogeography: complex
815 patterns of population history in a highly structured Iberian lizard are only revealed by
816 the use of nuclear markers. Mol Ecol 17:4670-4683.
817 Grover, A., and P. C. Sharma. 2011. Is spatial occurrence of microsatellites in the genome a
818 determinant of their function and dynamics contributing to genome evolution? Curr Sci
819 India 100:859-869.
820 Gunther, T., and G. Coop. 2013. Robust Identification of Local Adaptation from Allele
821 Frequencies. Genetics 195:205-220.
822 Gyllensten, U. B., S. Jakobsson, and H. Temrin. 1990. No Evidence for Illegitimate Young in
823 Monogamous and Polygynous Warblers. Nature 343:168-170.
824 Hagen, I. J., A. M. Billing, B. Ronning, S. A. Pedersen, H. Parn, J. Slate, and H. Jensen. 2013.
825 The easy road to genome-wide medium density SNP screening in a non-model species:
826 development and application of a 10K SNP-chip for the house sparrow (Passer
827 domesticus). Molecular Ecology Resources 13:429-439.

828 Hammer, M., A. Woerner, F. Mendez, J. Watkins, and J. Wall. 2011. Genetic evidence for
829 archaic admixture in Africa. Proceedings of the National Academy of Sciences
830 108:15123-15128.
831 Hare, M. P., and S. R. Palumbi. 1999. The accuracy of heterozygous base calling from diploid
832 sequence and resolution of haplotypes using allele-specific sequencing. Mol Ecol 8:1750-
833 1752.
834 Harvey, M. G., and R. T. Brumfield. 2015. Genomic variation in a widespread Neotropical bird
835 (Xenops minutus) reveals divergence, population expansion, and gene flow. Mol
836 Phylogenet Evol 83:305-316.
837 Harvey, M. G., C. D. Judy, G. F. Seeholzer, J. M. Maley, G. R. Graves, and R. T. Brumfield.
838 2015a. Similarity thresholds used in DNA sequence assembly from short reads can
839 reduce the comparability of population histories across species. PeerJ 3:e895.
840 Harvey, M. G., B. T. Smith, T. C. Glenn, B. C. Faircloth, and R. T. Brumfield. 2015b. Sequence
841 Capture Versus Restriction Site Associated DNA Sequencing for Phylogeography.
842 arXiv:1312.6439.
843 Hedin, M., J. Starrett, S. Akhter, A. L. Schonhofer, and J. W. Shultz. 2012. Phylogenomic
844 resolution of paleozoic divergences in harvestmen (Arachnida, Opiliones) via analysis of
845 next-generation transcriptome data. Plos One 7:e42888.
846 Hickerson, M. J., B. C. Carstens, J. Cavender-Bares, K. A. Crandall, C. H. Graham, J. B.
847 Johnson, L. Rissler, P. F. Victoriano, and A. D. Yoder. 2010. Phylogeography's past,
848 present, and future: 10 years after Avise, 2000. Mol Phylogenet Evol 54:291-301.
849 Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution
850 interpolated climate surfaces for global land areas. International Journal of Climatology
851 25:1965-1978.
852 Hiller, M., B. T. Schaar, V. B. Indjeian, D. M. Kingsley, L. R. Hagey, and G. Bejerano. 2012. A
853 "forward genomics" approach links genotype to phenotype using independent phenotypic
854 losses among related species. Cell reports 2:817-823.
855 Hochachka, W. M., and A. A. Dhondt. 2000a. Density-dependent decline of host abundance
856 resulting from a new infectious disease. Proc Natl Acad Sci U S A 97:5303-5306.
857 Hochachka, W. M., and A. A. Dhondt. 2000b. Density-dependent decline of host abundance
858 resulting from a new infectious disease. Proceedings of the National Academy of
859 Sciences 97:5303-5306.
860 Hoekstra, H. E., and J. A. Coyne. 2007. The locus of evolution: Evo devo and the genetics of
861 adaptation. Evolution 61:995-1016.
862 Hoekstra, H. E., R. J. Hirschmann, R. A. Bundey, P. A. Insel, and J. P. Crossland. 2006. A single
863 amino acid mutation contributes to adaptive beach mouse color pattern. Science 313:101-
864 104.
865 Hohenlohe, P. A., S. Bassham, M. Currey, and W. A. Cresko. 2012, Extensive linkage
866 disequilibrium and parallel adaptive divergence across threespine stickleback genomes.
867 Phil. Trans. Roy. Soc. B 367:395-408.
868 Huang, H., and L.L. Knowles. In press. Unforseen consequences of excluding missing data from
869 next-generation sequences: simulation study of RAD sequences. Syst. Biol.
870 Hudson, R. R., and N. L. Kaplan. 1988. The coalescent process in models with selection and
871 recombination. Genetics 120:831-840.

872 Janes, D. E., T. Ezaz, J. A. M. Graves, and S. V. Edwards. 2009. Recombination and nucleotide
873 diversity in the sex chromosomal pseudoautosomal region of the Emu, *Dromaius*
874 *novaehollandiae*. J Hered 100:125-136.
875 Jeffreys, A. J., V. Wilson, and S. L. Thein. 1985. Hypervariable Minisatellite Regions in Human
876 DNA. Nature 314:67-73.
877 Jennings, W. B., and S. V. Edwards. 2005. Speciational history of Australian grass finches
878 (*Poephila*) inferred from 30 gene trees. Evolution 59:2033-2047.
879 Jobling, M. A. 2012. The impact of recent events on human genetic diversity. Philos T R Soc B
880 367:793-799.
881 Johnston, S. E., P. Orell, V. L. Pritchard, M. P. Kent, S. Lien, E. Niemela, J. Erkinaro, and C. R.
882 Primmer. 2014. Genome-wide SNP analysis reveals a genetic basis for sea-age variation
883 in a wild population of Atlantic salmon (Salmo salar). Mol Ecol 23:3452-3468.
884 Jones, F. C., Y. F. Chan, J. Schmutz, J. Grimwood, S. D. Brady, A. M. Southwick, D. M. Absher,
885 R. M. Myers, T. E. Reimchen, B. E. Deagle, D. Schluter, and D. M. Kingsley. 2012a. A
886 Genome-wide SNP Genotyping Array Reveals Patterns of Global and Repeated Species-
887 Pair Divergence in Sticklebacks. Curr Biol 22:83-90.
888 Jones, F. C., M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli, J. Johnson, R. Swofford, M.
889 Pirun, M. C. Zody, S. White, E. Birney, S. Searle, J. Schmutz, J. Grimwood, M. C.
890 Dickson, R. M. Myers, C. T. Miller, B. R. Summers, A. K. Knecht, S. D. Brady, H. L.
891 Zhang, A. A. Pollen, T. Howes, C. Amemiya, E. S. Lander, F. Di Palma, K. Lindblad-
892 Toh, D. M. Kingsley, B. I. G. S. Platf, and W. G. A. Team. 2012b. The genomic basis of
893 adaptive evolution in threespine sticklebacks. Nature 484:55-61.
894 Jones, M. R., B. R. Forester, A. I. Teufel, R. V. Adams, D. N. Anstett, B. A. Goodrich, E. L.
895 Landguth, S. Joost, and S. Manel. 2013. Integrating landscape genomics nd spatially
896 explicit approaches to detect loci under seletion in clinal populatoins. Evolution 67:3455-
897 3468.
898 Keightley, P. D., and A. Eyre-Walker. 2010. What can we learn about the distribution of fitness
899 effects of new mutations from DNA sequence data? Philosophical transactions of the
900 Royal Society of London. Series B, Biological sciences 365:1187-1193.
901 Kimura, M. 1968. Evolutionary rate at the molecular level. Nature 217:624-626.
902 Kimura, M. 1983, The Neutral Theory of Molecular Evolution. Cambridge, Cambridge
903 University Press.
904 Kratochwil, C. F., and A. Meyer. 2015. Closing the genotype-phenotype gap: Emerging
905 technologies for evolutionary genetics in ecological model vertebrate systems. Bioessays
906 37:213-226.
907 Kwong, M., and T. J. Pemberton. 2014. Sequence differences at orthologous microsatellites
908 inflate estimates of human-chimpanzee differentiation. Bmc Genomics 15:990.
909 Lemmon, A. R., and E. M. Lemmon. 2012. High-Throughput Identification of Informative
910 Nuclear Loci for Shallow-Scale Phylogenetics and Phylogeography. Syst Biol 61:745-
911 761.
912 Lemmon, E. M., and A. R. Lemmon. 2013. High-Throughput Genomic Data in Systematics and
913 Phylogenetics. Annual Review of Ecology, Evolution, and Systematics 44:99-121.
914 Lewontin, R. 1974, The Genetic Basis of Evolutionary Change. New York, Columbia University
915 Press.
916 Lewontin, R. C., and J. Krakauer. 1973. Distribution of gene frequency as a test of theory of
917 selective neutrality of polymorphisms. Genetics 74:175-195.

918 Ley, D. H., J. E. Berkhoff, and J. M. McLaren. 1996. Mycoplasma gallisepticum isolated from
919 house finches (Carpodacus mexicanus) with conjunctivitis. Avian Diseases. 40:480-483.
920 Li, H., and R. Durbin. 2011. Inference of human population history from individual whole-
921 genome sequences. Nature 475:493-496.
922 Li, M. H., and J. Merila. 2010. Sex-specific population structure, natural selection, and linkage
923 disequilibrium in a wild bird population as revealed by genome-wide microsatellite
924 analyses. BMC Evol Biol 10:66.
925 Linnen, C. R., Y. P. Poh, B. K. Peterson, R. D. H. Barrett, J. G. Larson, J. D. Jensen, and H. E. Hoekstra.
926 2013. Adaptive Evolution of Multiple Traits Through Multiple Mutations at a Single Gene.
927 Science 339:1312-1316.
928 Liu, T., S. Wahlberg, E. Burek, P. Lindblom, C. Rubio, and A. Lindblom. 2000. Microsatellite
929 instability as a predictor of a mutation in a DNA mismatch repair gene in familial
930 colorectal cancer. Genes Chromosomes & Cancer 27:17-25.
931 Lotterhos, K. E., and M. C. Whitlock. 2014. Evaluation of demographic history and neutral
932 parameterization on the performance of F-ST outlier tests. Mol Ecol 23:2178-2192.
933 Marra, N. J., A. Romero, and J. A. DeWoody. 2014. Natural selection and the genetic basis of
934 osmoregulation in heteromyid rodents as revealed by RNA-seq. Mol Ecol 23:2699-2711.
935 Mccormack, J. E., and B. C. Faircloth. 2013. Next-generation phylogenetics takes root. Mol Ecol
936 22:19-21.
937 McCormack, J. E., S. M. Hird, A. J. Zellmer, B. C. Carstens, and R. T. Brumfield. 2013.
938 Applications of next-generation sequencing to phylogeography and phylogenetics. Mol
939 Phylogenet Evol 66:526-538.
940 McCormack, J. E., J. M. Maley, S. M. Hird, E. P. Derryberry, G. R. Graves, and R. T. Brumfield.
941 2012. Next-generation sequencing reveals phylogeographic structure and a species tree
942 for recent bird divergences. Mol Phylogenet Evol 62:397-406.
943 Merz, C., Catchen, J. M., Hanson-Smith, V., Emerson, K. J., Bradshaw, W. E., & Holzapfel, C.
944 M. (2013). Replicate phylogenies and post-glacial range expansion of the pitcher-plant
945 mosquito, Wyeomyia smithii, in North America. PloS One, 8(9), e72262.
946 doi:10.1371/journal.pone.0072262
947 Metzgar, D., J. Bytof, and C. Wills. 2000. Selection against frameshift mutations limits
948 microsatellite expansion in coding DNA. Genome Res 10:72-80.
949 Mohammed, J., J. Frasca, Salvatore, K. Cecchini, D. Rood, A. C. Nyaoke, S. J. Geary, and L. K.
950 Silbart. 2007. Chemokine and cytokine gene expression profiles in chickens inoculated
951 with Mycoplasma gallisepticum strains Rlow or GT5. Vaccine 25:8611-8621.
952 Morales, H.E., A. Pavlova, L. Joseph, and P. Sunnucks. 2015. Positive and purifying selection in
953 mitochondrial genomes of a bird with mitonuclear discordance. Mol. Ecol. in press.
954 Morin, P. A., G. Luikart, R. K. Wayne, and S. N. P. W. Grp. 2004. SNPs in ecology, evolution
955 and conservation. Trends Ecol Evol 19:208-216.
956 Nabholz, B., S. Glemin, and N. Galtier. 2009. The erratic mitochondrial clock: variations of
957 mutation rate, not population size, affect mtDNA diversity across birds and mammals.
958 BMC Evol Biol 9:54.
959 Nabholz, B., J. F. Mauffrey, E. Bazin, N. Galtier, and S. Glemin. 2008. Determination of
960 mitochondrial genetic diversity in mammals. Genetics 178:351-361.
961 Nei, M., and A.K. Roychoudhury. 1974. Sampling variances of heterozygosity and genetic
962 distance. Genetics 76:379-390.

963 Nielsen, R., and M. A. Beaumont. 2009. Statistical inferences in phylogeography. Mol Ecol
964 18:1034-1047.
965 Noma, K., N. Oyama, and J. K. Liao. 2006. Physiological role of ROCKs in the cardiovascular
966 system. American Journal of Physiology-Cell Physiology 290:C661-C668.
967 O'Neill, E. M., R. Schwartz, C. T. Bullock, J. S. Williams, H. B. Shaffer, X. Aguilar-Miguel, G.
968 Parra-Olea, and D. W. Weisrock. 2013. Parallel tagged amplicon sequencing reveals
969 major lineages and phylogenetic structure in the North American tiger salamander
970 (Ambystoma tigrinum) species complex. Mol Ecol 22:111-129.
971 Ohta, T. 1992. The nearly neutral theory of molecular evolution. Annu Rev Ecol Syst 23:263-
972 286.
973 Ohta, T., and J. H. Gillespie. 1996. Development of Neutral and Nearly Neutral Theories. Theor
974 Popul Biol 49:128-142.
975 Ojeda, D. I., B. Dhillon, C. K. M. Tsui, and R. C. Hamelin. 2014. Single-nucleotide
976 polymorphism discovery in Leptographium longiclavatum, a mountain pine beetle-
977 associated symbiotic fungus, using whole-genome resequencing. Molecular Ecology
978 Resources 14:401-410.
979 Pallares, L. F., B. Harr, L. M. Turner, and D. Tautz. 2014. Use of a natural hybrid zone for
980 genomewide association mapping of craniofacial traits in the house mouse. Mol Ecol
981 23:5756-5770.
982 Palumbi, S. R., and C. S. Baker. 1994. Contrasting Population-Structure from Nuclear Intron
983 Sequences and Mtdna of Humpback Whales. Mol Biol Evol 11:426-435.
984 Patterson, N., A. L. Price, and D. Reich. 2006. Population structure and eigenanalysis. PLoS
985 Genet Volume 2:e190.
986 Pearse, D. E., M. R. Miller, A. Abadia-Cardoso, and J. C. Garza. 2014. Rapid parallel evolution
987 of standing variation in a single, complex, genomic region is associated with life history
988 in steelhead/rainbow trout. P R Soc B 281:20140012.
989 Perktaş, U., Gür, H., Sağlam, İ. K., & Quintero, E. 2015. Climate-driven range shifts and
990 demographic events over the history of Kruper's Nuthatch Sitta krueperi. Bird Study, 62:
991 14–28. doi:10.1080/00063657.2014.977220
992 Perry, J. C., and L. Rowe. 2011. Rapid Microsatellite Development for Water Striders by Next-
993 Generation Sequencing. J Hered 102:125-129.
994 Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher, and H. E. Hoekstra. 2012. Double Digest
995 RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model
996 and Non-Model Species. Plos One 7:e37135.
997 Pluzhnikov, A., and P. Donnelly. 1996. Optimal sequencing strategies for surveying molecular
998 genetic diversity. Genetics 144:1247–1262.
999 Portner, H. O., A. F. Bennett, F. Bozinovic, A. Clarke, M. A. Lardies, M. Lucassen, B. Pelster, F.
1000 Schiemer, and J. H. Stillman. 2006. Trade-offs in thermal adaptation: The need for a
1001 molecular to ecological integration. Physiological and Biochemical Zoology 79:295-313.
1002 Portner, H. O., L. Peck, and G. Somero. 2007. Thermal limits and adaptation in marine Antarctic
1003 ectotherms: an integrative view. Philos T R Soc B 362:2233-2258.
1004 Price, A. L., N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, and D. Reich. 2006.
1005 Principal components analysis corrects for stratification in genome-wide association
1006 studies. Nat Genet 38:904-909.
1007 Price, A. L., N. A. Zaitlen, D. Reich, and N. Patterson. 2010. New approaches to population
1008 stratification in genome-wide association studies. Nat Rev Genet 11:459-463.

1009 Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using
1010 multilocus genotype data. Genetics 155:945-959.
1011 Puritz, J. B., M. V. Matz, R. J. Toonen, J. N. Weber, D. I. Bolnick, and C. E. Bird. 2014.
1012 Demystifying the RAD fad. Mol Ecol 23:5937-5942.
1013 Rand, D.M. 2001. The units of selection on mitochondrial DNA. Annu. Rev. Ecol. Syst. 32:415-
1014 448.
1015 Rankinen, T., T. Church, T. Rice, N. Markward, S. N. Blair, and C. Bouchard. 2008. A Major
1016 Haplotype Block at the Rho-Associated Kinase 2 Locus Is Associated with a Lower Risk
1017 of Hypertension in a Recessive Manner: The HYPGENE Study. Hypertension Research
1018 31:1651-1657.
1019 Rebeiz, M., J. E. Pool, V. A. Kassner, C. F. Aquadro, and S. B. Carroll. 2009. Stepwise
1020 Modification of a Modular Enhancer Underlies Adaptation in a Drosophila Population.
1021 Science 326:1663-1667.
1022 Reich, D., R. E. Green, M. Kircher, J. Krause, N. Patterson, E. Y. Durand, B. Viola, A. W.
1023 Briggs, U. Stenzel, P. L. F. Johnson, T. Maricic, J. M. Good, T. Marques-Bonet, C. Alkan,
1024 Q. Fu, S. Mallick, H. Li, M. Meyer, E. E. Eichler, M. Stoneking, M. Richards, S. Talamo,
1025 M. V. Shunkov, A. P. Derevianko, J.-J. Hublin, J. Kelso, M. Slatkin, and S. Pääbo. 2010.
1026 Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature
1027 468:1053-1060.
1028 Reinhardt, J. A., B. Kolaczkowski, C. D. Jones, D. J. Begun, and A. D. Kern. 2014. Parallel
1029 geographic variation in Drosophila melanogaster. Genetics 197:361-373.
1030 Rheindt, F. E., Fujita, M. K., Wilton, P. R., & Edwards, S. V. (2014). Introgression and
1031 phenotypic assimilation in zimmerius flycatchers (Tyrannidae): Population genetic and
1032 phylogenetic inferences from genome-wide SNPs. Systematic Biology, 63(2), 134–152.
1033 doi:10.1093/sysbio/syt070
1034 Ribeiro, A. M., P. Lloyd, and R. C. K. Bowie. 2011. A tight balance between natural selection
1035 and gene flow in a Southern African arid-zone endemic bird. Evolution 65:3499-3514.
1036 Riento, K., and A. J. Ridley. 2003. Rocks: Multifunctional kinases in cell behaviour. Nature
1037 Reviews Molecular Cell Biology 4:446-456.
1038 Roesti, M., S. Gavrilets, A. P. Hendry, W. Salzburger, and D. Berner. 2014. The genomic
1039 signature of parallel adaptation from shared genetic variation. Mol Ecol 23:3944-3956.
1040 Romiguier, J., P. Gayral, M. Ballenghien, A. Bernard, V. Cahais, A. Chenuil, Y. Chiari, R.
1041 Dernat, L. Duret, N. Faivre, E. Loire, J. M. Lourenco, B. Nabholz, C. Roux, G.
1042 Tsagkogeorga, A. A. T. Weber, L. A. Weinert, K. Belkhir, N. Bierne, S. Glemin, and N.
1043 Galtier. 2014. Comparative population genomics in animals uncovers the determinants of
1044 genetic diversity. Nature 515:261-U243.
1045 Santure, A. W., I. De Cauwer, M. R. Robinson, J. Poissant, B. C. Sheldon, and J. Slate. 2013.
1046 Genomic dissection of variation in clutch size and egg mass in a wild great tit (Parus
1047 major) population. Mol Ecol 22:3949-3962.
1048 Schielzeth, H., and A. Husby. 2014. Challenges and prospects in genome-wide quantitative trait
1049 loci mapping of standing genetic variation in natural populations. Year in Evolutionary
1050 Biology 1320:35-57.
1051 Seasholtz, T. M., J. Wessel, F. W. Rao, B. K. Rana, S. Khandrika, B. P. Kennedy, E. O. Lillie, M.
1052 G. Ziegler, D. W. Smith, N. J. Schork, J. H. Brown, and D. T. O'Connor. 2006. Rho
1053 kinase polymorphism influences blood pressure and systemic vascular resistance in
1054 human twins - Role of heredity. Hypertension 47:937-947.

1055 Singham, G. V., E. L. Vargo, W. Booth, A. S. Othman, and C. Y. Lee. 2012. Polymorphic
1056 Microsatellite Loci From an Indigenous Asian Fungus-Growing Termite, Macrotermes
1057 gilvus (Blattodea: Termitidae) and Cross Amplification in Related Taxa. Environ
1058 Entomol 41:426-431.
1059 Slate, J., and J. M. Pemberton. 2007. Admixture and patterns of linkage disequilibrium in a free-
1060 living vertebrate population. J Evol Biol 20:1415-1427.
1061 Slatkin, M. 2008. Linkage disequilibrium--understanding the evolutionary past and mapping the
1062 medical future. Nat Rev Genet 9:477-485.
1063 Smith, B. T., M. G. Harvey, B. C. Faircloth, T. C. Glenn, and R. T. Brumfield. 2014. Target
1064 capture and massively parallel sequencing of ultraconserved elements for comparative
1065 studies at shallow evolutionary time scales. Syst Biol 63:83–95.
1066 Smukowski, C. S., and M. A. Noor. 2011. Recombination rate variation in closely related species.
1067 Heredity (Edinb) 107:496-508.
1068 Stinchcombe, J. R., and H. E. Hoekstra. 2008. Combining population genomics and quantitative
1069 genetics: finding the genes underlying ecologically important traits. Heredity 100:158-
1070 170.
1071 Stoneking, M., and J. Krause. 2011. Learning about human population history from ancient and
1072 modern genomes. Nat Rev Genet 12:603-614.
1073 Sun, J. X., J. C. Mullikin, N. Patterson, and D. E. Reich. 2009. Microsatellites are molecular
1074 clocks that support accurate inferences about history. Mol. Biol. Evol. 26:1017–1027.
1075 Sureshkumar, S., M. Todesco, K. Schneeberger, R. Harilal, S. Balasubramanian, and D. Weigel.
1076 2009. A Genetic Defect Caused by a Triplet Repeat Expansion in *Arabidopsis thaliana*.
1077 Science 323:1060-1063 .
1078 Taguchi, M., Y. Shigenobu, M. Ohkubo, T. Yanagimoto, T. Sugaya, Y. Nakamura, K. Saitoh,
1079 and K. Yokawa. 2013. Characterization of 12 polymorphic microsatellite DNA loci in the
1080 blue shark, Prionace glauca, isolated by next generation sequencing approach.
1081 Conservation Genetics Resources 5:117-119.
1082 Tautz, D., and M. Renz. 1984. Simple Sequences Are Ubiquitous Repetitive Components of
1083 Eukaryotic Genomes. Nucleic Acids Res 12:4127-4138.
1084 Templeton, A. R. 2010. Coalescent-based, maximum likelihood inference in phylogeography.
1085 Mol Ecol 19:431-435.
1086 Tiffin, P., & Ross-Ibarra, J. 2014. Advances and limits of using population genetics to
1087 understand local adaptation. Trends in Ecology & Evolution, 29(12), 673–680.
1088 doi:10.1016/j.tree.2014.10.004
1089 Tremblay, D. C., G. Alexander, S. Moseley, and B. P. Chadwick. 2010. Expression, tandem
1090 repeat copy number variation and stability of four macrosatellite arrays in the human
1091 genome. Bmc Genomics 11:632.
1092 Turner, T. L., and M. W. Hahn. 2010. Genomic islands of speciation or genomic islands and
1093 speciation? Mol Ecol 19:848-850.
1094 Uy, J. A. C., R. G. Moyle, C. E. Filardi, and Z. A. Cheviron. 2009. Difference in Plumage Color
1095 Used in Species Recognition between Incipient Species Is Linked to a Single Amino
1096 Acid Substitution in the Melanocortin-1 Receptor. Am Nat 174:244-254.
1097 Wallberg, A., F. Han, G. Wellhagen, B. Dahle, M. Kawata, N. Haddad, Z. L. P. Simoes, M. H.
1098 Allsopp, I. Kandemir, P. De la Rua, C. W. Pirk, and M. T. Webster. 2014. A worldwide
1099 survey of genome sequence variation provides insight into the evolutionary history of the
1100 honeybee Apis mellifera. Nat Genet 46:1081-1088.

1101 Weber, C., B. Nabholz, J. Romiguier, and H. Ellegren. 2014. Kr/Kc but not dN/dS correlates positively
1102 with body mass in birds, raising implications for inferring lineage-specific selection. Genome
1103 Biol 15:542.
1104 Yukilevich, R., T. L. Turner, F. Aoki, S. V. Nuzhdin, and J. R. True. 2010. Patterns and
1105 Processes of Genome-Wide Divergence Between North American and African
1106 Drosophila melanogaster. Genetics 186:219-239.
1107 Zink, R. M. 2010. Drawbacks with the use of microsatellites in phylogeography: the song
1108 sparrow Melospiza melodia as a case study. J Avian Biol 41:1-7.
1109 Zink, R. M., and G. F. Barrowclough. 2008. Mitochondrial DNA under siege in avian
1110 phylogeography. Mol Ecol 17:2107-2121.
1111 Zweier, C., M. M. Peippo, J. Hoyer, S. Sousa, A. Bottani, J. Clayton-Smith, W. Reardon, J.
1112 Saraiva, A. Cabral, I. Gohring, K. Devriendt, T. de Ravel, E. K. Bijlsma, R. C. M.
1113 Hennekam, A. Orrico, M. Cohen, A. Dreweke, A. Reis, P. Nurnberg, and A. Rauch. 2007.
1114 Haploinsufficiency of TCF4 causes syndromal mental retardation with intermittent
1115 hyperventilation (Pitt-Hopkins syndrome). American Journal of Human Genetics 80:994-
1116 1001.

1117

1118

1119 **Figure Legends**

1120

1121 Figure 1. Trends in the use of molecular markers over time from a survey of articles on
1122 phylogeography from the journal *Molecular Ecology*. (A) Plot of the number of articles from
1123 1992 to 2013 using organelle DNA markers, microsatellites, nuclear sequence-based markers,
1124 color coded according to the key at upper left. The category of nuclear sequence-based markers
1125 includes 8 studies in 2013 using RadSeq, sequence capture, or another next-generation
1126 sequencing technique to genotype SNPs; the remainder of nuclear SNP studies used PCR
1127 approaches. Data on articles using nuclear SNPs and organelle markers are taken from Garrick
1128 et al. (2015) and are presented for every three years for easy comparison with that
1129 study. Overall, during the time period sampled there were 280 studies using only SNPs from
1130 mt- or cpDNA, 101 studies using nuclear SNPs, and 97 studies using nuclear
1131 microsatellites. There were 13 studies that included both nuclear SNPs and microsatellites; these
1132 were excluded from the plot. Sixty-three of the microsatellite studies and 79 of the nuclear SNP
1133 studies also included mt- or cpDNA SNPs and are included. (B) Trends in the total number of
1134 articles from *Molecular Ecology* using nuclear microsatellites sampled every year from 1992 to
1135 2014. These 318 studies include those using other types of data such as mt- or cpDNA SNPs or
1136 nuclear SNPs. The number of pages in the journal *Molecular Ecology* per year is shown in a
1137 black line as a guage on growth of the journal as a whole. The full list of articles using
1138 microsatellites can be found in the supplementary material.

1139 Figure 2: Advantages and disadvantages of studying species with low or high levels of
1140 linkage disequilibrium (LD) using genome subsampling methods such as Rad-seq to identify loci
1141 under selection. The matrix covers two sets of species: on the Y-axis, those with low or high

1142 levels of LD, as discussed in the main text. On the X-axis are listed the advantages or
1143 disadvantages of studying such species using the Rad-seq approach. In each cell is a description
1144 of common situations encountered in the search for loci under selection. See text for further
1145 discussion.

1146 Figure 3: Two examples of low- and high linkage disequilibrium species. A) Plot of
1147 pairwise r^2 (a measure of linkage disequilibrium) between SNPs across 1500 base pairs of the
1148 HSP90a gene in House Finches, a common North American songbird. SNPs from an Arizona
1149 population are in black, those from an Alabama population in red. This species is characterized
1150 by large population sizes, resulting in a high population recombination rate and low levels of LD.
1151 From Backström et al. (2013). B) Similar plot of pairwise r^2 and physical distance in kilobases
1152 in various dog breeds and wild populations of gray wolves (*Canis lupus*). Notice how high levels
1153 of LD extend hundreds of kilobases; those same levels of LD extend only a few hundred bases in
1154 the case of the House Finches. Image from Boyko (2011); see also Boyko et al. (2010). Both
1155 images used under the Creative Commons 2 License
1156 https://creativecommons.org/licenses/by/2.0/.

1157 Figure 4. Correlations of Rad-seq SNP variation and environmental variables across
1158 geographic space in the lizard *Anolis carolinensis*. A) Environmental associations across
1159 geographic space of each SNP identified by RADseq and the mean temperature of the coldest
1160 quarter of the year at each of 6 localities distributed across the species range, calculated in
1161 Bayenv2 (Gunther and Coop 2013). The horizontal and vertical dotted lines represent 99%
1162 cutoffs for significance of Pearson correlation and Bayes Factor associations, respectively. Filled
1163 points indicate candidate SNPs falling within the top 1% of both axes. B) The genomic position

1164 of each outlier SNP in panel A on the 6 annotated macrochromosomes of the *A. carolinensis*

1165 genome (Alfoldi et al. 2011).

1166 Figure 5. Estimates of F_{st} and associated Fisher's p-values for each SNP compared

1167 between pre- (1990) and post- *Mycoplasma* (2003) epizootic individuals sampled from Auburn,

1168 Alabama. The dotted line on the plot of p-values shows the uncorrected cutoff of p =

1169 0.05. After Bonferroni correction, no SNPs achieve significant F_{st} (see text). SNP positions are

1170 depicted as mapped to the Zebra Finch genome, assuming synteny between the House Finch and

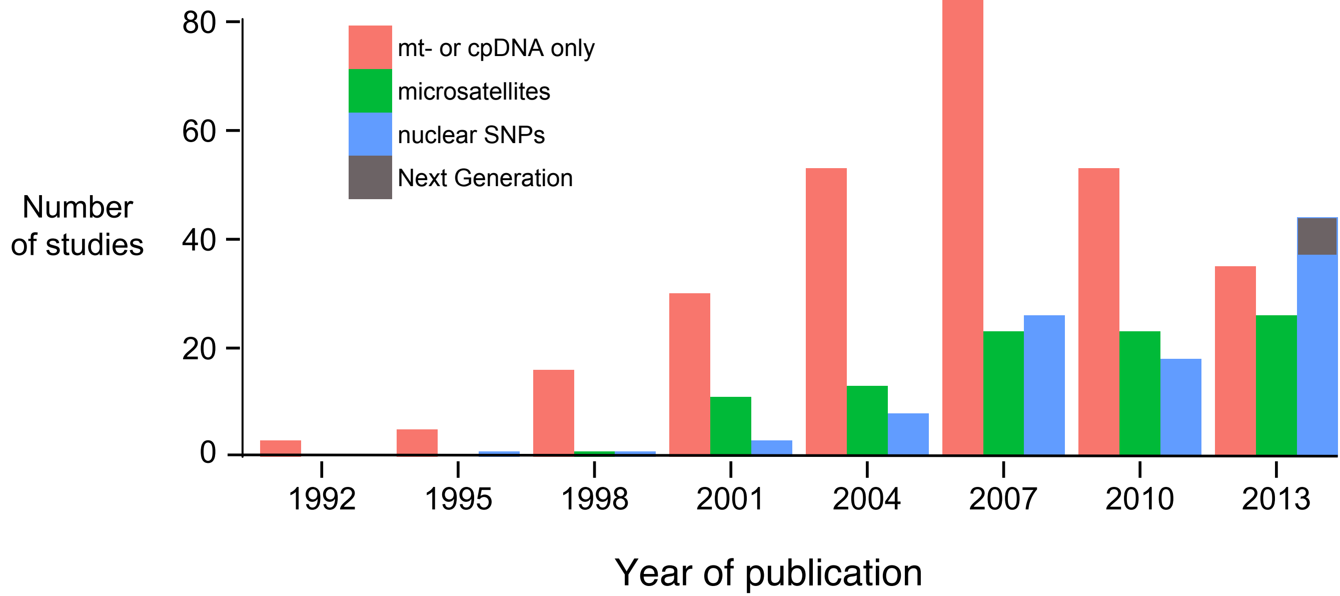
1171 Zebra Finch genomes. High values of F_{st} are not always associated with high Fisher's p-values,

1172 usually due to incomplete data matrices and concomitant smaller sample sizes at those positions,

1173 a situation common in RadSeq data. Image of House Finch from

1174 http://www.flickr.com/photos/11652987@N03/7315942062 and used under the Creative

1175 Commons 2 License https://creativecommons.org/licenses/by/2.0/.



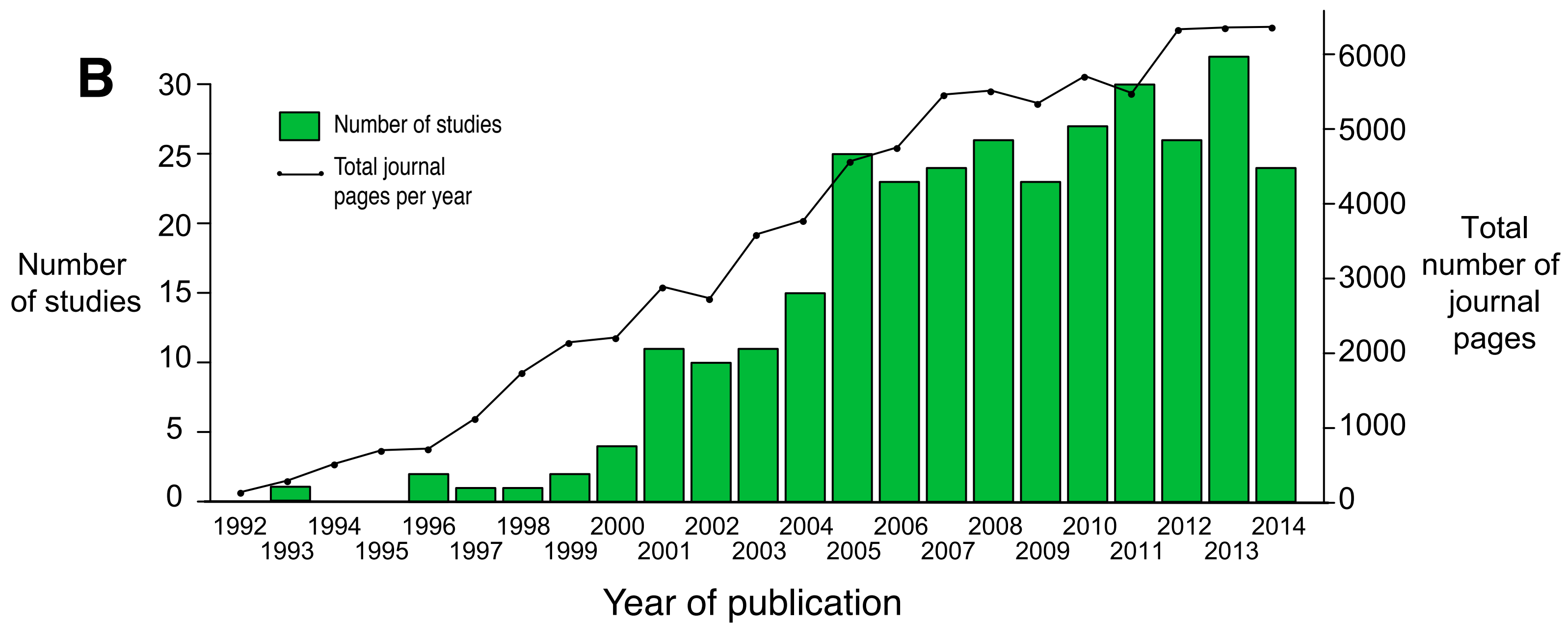


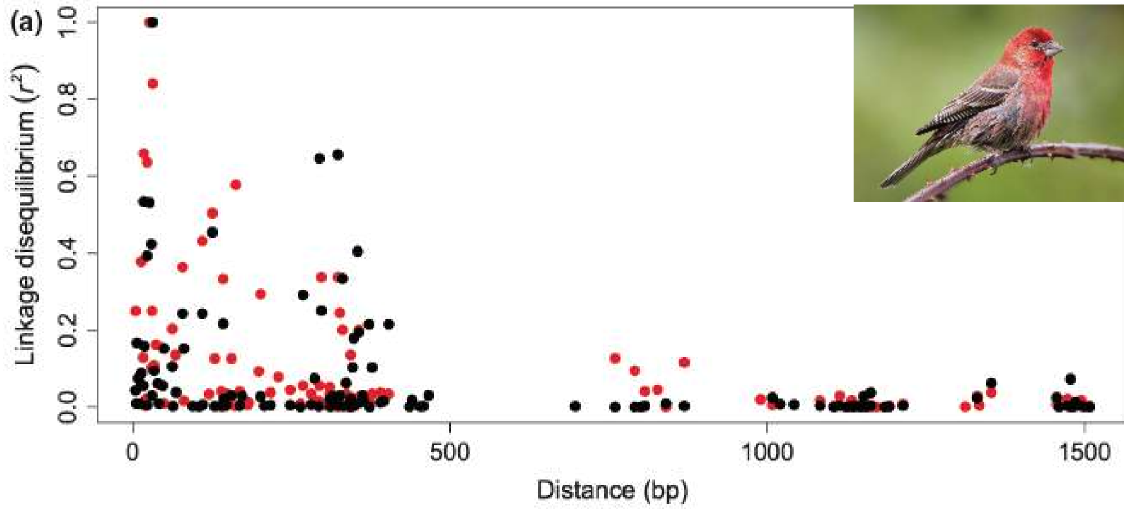
Figure 1

Rad-seq or other genome subsamping 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

Figure 2

A



B

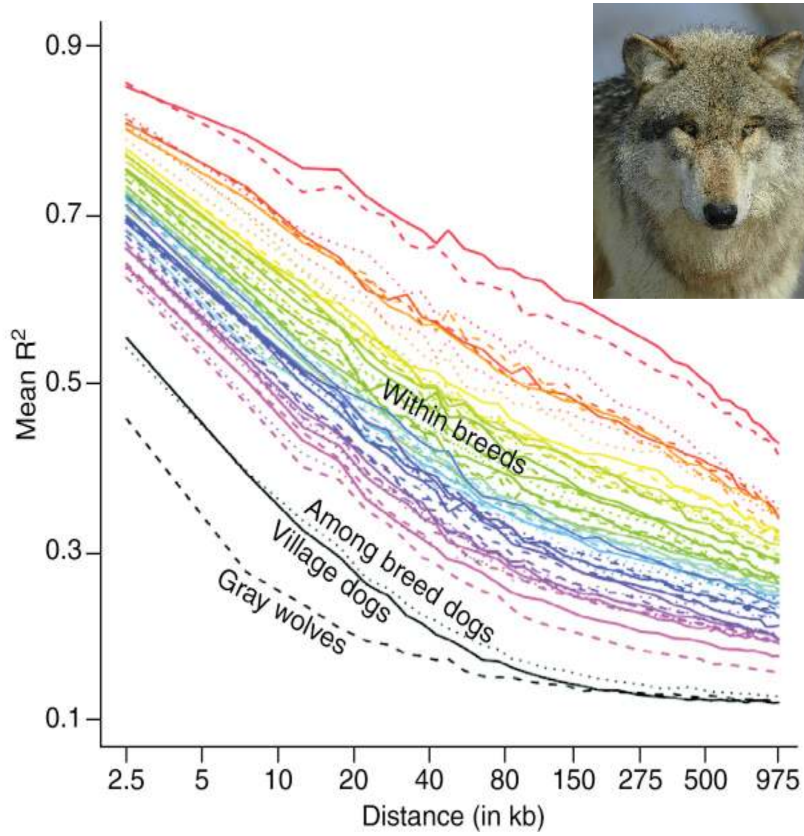
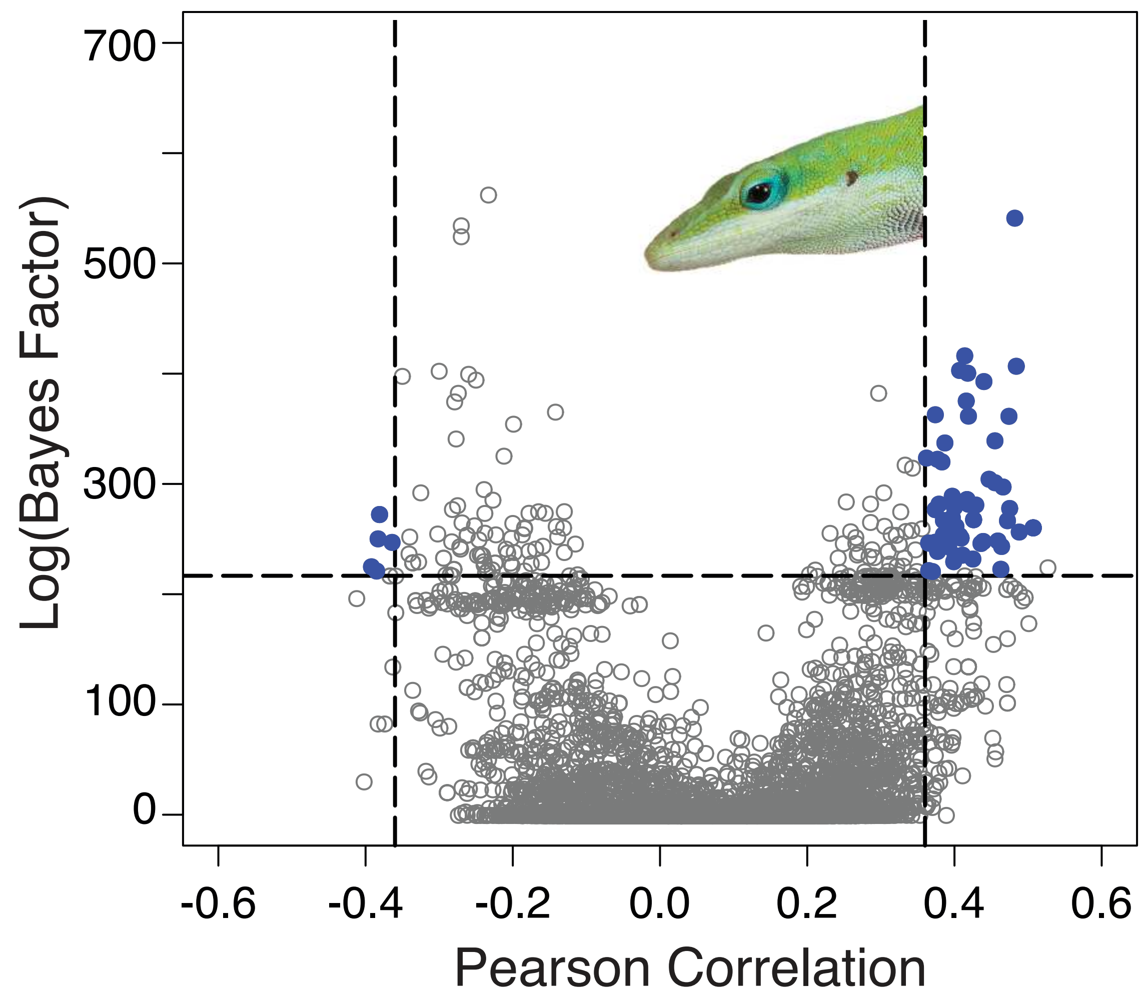


Figure 3



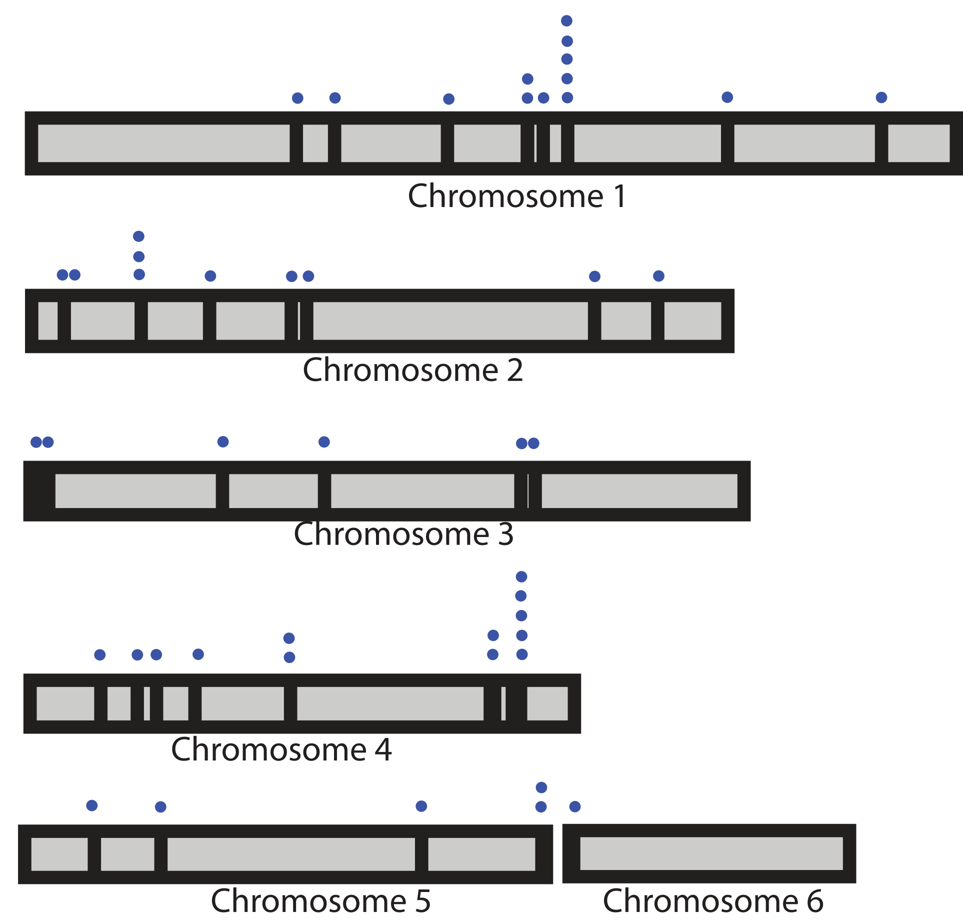
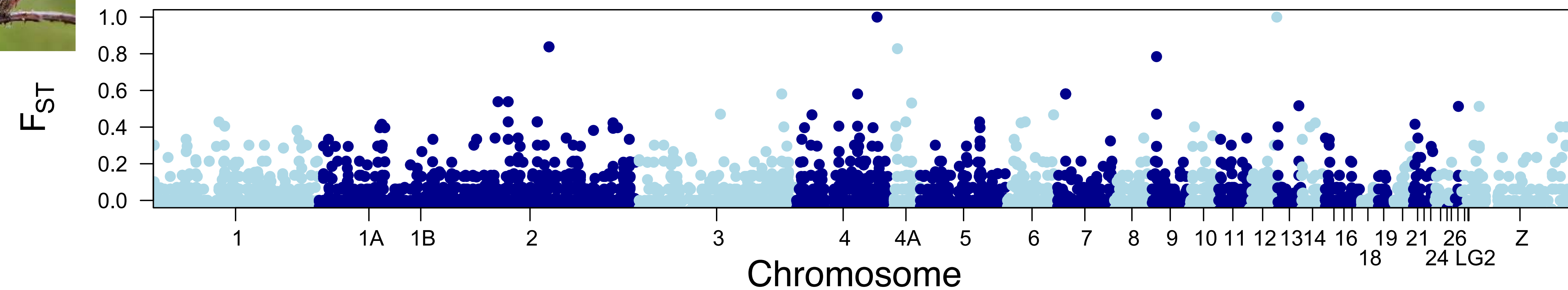


Figure 4



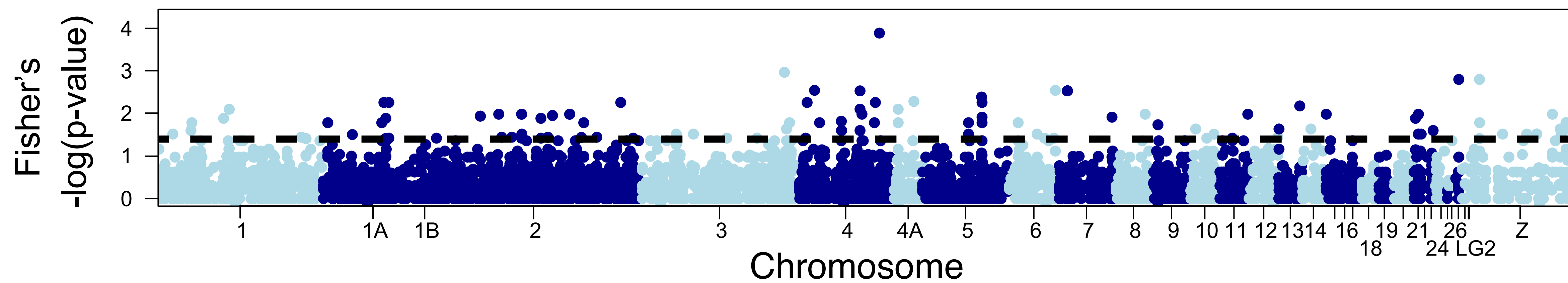


Figure 5