NEWS AND VIEWS

PERSPECTIVE

Widespread yet heterogeneous genomic divergence

PATRIK NOSIL* and JEFF L. FEDER†

Department of Animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield S10 2TN, UK, †Department of Biology, Notre Dame University, South Bend, IN 11111, USA

Genetic differentiation during adaptive divergence and speciation is heterogeneous among genomic regions. Some regions can be highly differentiated between populations, for example, because they harbour genes under divergent selection or those causing reproductive isolation and thus are resistant to gene flow. Other regions might be homogenized by gene flow and thus weakly differentiated. Debates persist about the number of differentiated regions expected under divergence with gene flow, and their causes, size, and genomic distribution. In this issue of *Molecular Ecology*, a study of freshwater stickleback used next-generation sequencing to shed novel insight into these issues (Roesti *et al.* 2012). Many genomic regions distributed across the genome were strongly differentiated, indicating divergence with gene flow can involve a greater number of loci than often thought. Nonetheless, differentiation of some regions, such as those near the centre of chromosomes where recombination is reduced, was strongly accentuated over others. Thus, divergence was widespread yet highly heterogeneous across the genome. Moreover, different population pairs varied in patterns of differentiation, illustrating how genomic divergence builds up across stages of the speciation process. The study demonstrates how variation in different evolutionary processes, such as selection and recombination rate, can combine to result in similar genomic patterns. Future work could focus on teasing apart the contributions of different processes for causing differentiation, a task facilitated by experimental manipulations.

Next-generation sequencing (NGS) is empowering a comparative genomics approach to studying adaptation and

Correspondence: Patrik Nosil, Fax: 0114 2220002;
E-mail: patrik.nosil@colorado.edu

speciation. The goal here is to not just identify individual speciation genes contributing to reproductive isolation (Coyne & Orr 2004), but to determine how these genes are arrayed through the genome and the role this genome structure plays in speciation (Nosil & Feder 2012). Genome structure is particularly important for facilitating or impeding divergence in the face of gene flow. NGS is enabling extensive genome scans to identify 'outlier loci' displaying exceptional differentiation between populations (Fig 1). These outliers can denote regions affected by divergent natural selection ('selection' hereafter), whereas, in contrast, the remainder of the genome (presumably neutral) is homogenized by gene flow (Barton & Bengtsson 1986; Charlesworth *et al.* 1997; Feder & Nosil 2010). Such patterns of heterogeneous genomic divergence have now been documented in many taxa (Nosil *et al.* 2009). However, debate remains about the causes and meaning of these patterns.

There is a continuum of possibilities for genomic divergence. At one end of the spectrum, a few regions under strong divergent selection may initially drive speciation, with subsequent 'divergence hitchhiking' around these regions leading to a few large clusters of pronounced, physically linked genetic differentiation (Via & West 2008). At the other end of the continuum, speciation is because of selection on numerous regions across the genome, with such multifarious selection potentially reducing average genome-wide gene flow to the extent that divergence occurs across much of the genome (via a process recently termed 'genome hitchhiking' Feder *et al.* 2012; Feder & Nosil 2010). It is critical to note that intermediate scenarios are possible (and likely) and that under the latter scenario of genome hitchhiking, genomic divergence is still expected to be highly heterogeneous. There might be some general uplifting of baseline neutral differentiation across the genome caused by lowered overall gene flow, but regions subject to strong selection or low recombination will still be exceptionally differentiated (and divergence surrounding them might drop off rapidly).

In this issue of *Molecular Ecology*, Roesti *et al.* (2012) describe such a pattern of both widespread and heterogeneous genomic divergence based on NGS of thousands of RAD-tag SNP markers between freshwater lake and stream forms of the three-spined stickleback, *Gasterosteus aculeatus* (Figs 2 and 3). We focus here on four main findings and their significance for furthering our understanding of speciation genomics.

First and most generally, they found that numerous genomic regions are highly differentiated between lake and stream forms. This result suggests that speciation, even with gene flow and in early stages, may involve more
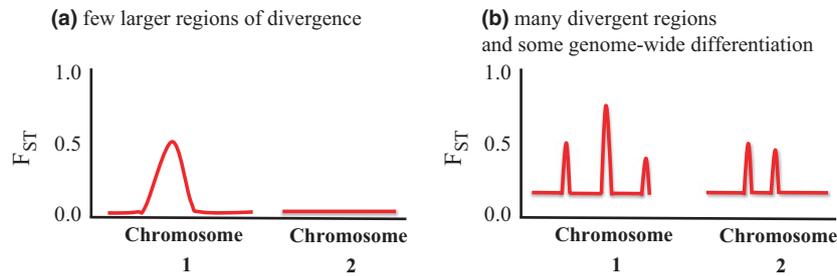
(a) few larger regions of divergence

(b) many divergent regions and some genome-wide differentiation

Fig. 1. Two ends of the spectrum of genomic differentiation with gene flow (i.e., many intermediate scenarios are possible). (a) Genomic divergence in a few large and isolated regions, with homogenization of the remainder of the genome by gene flow. Such a pattern of scattered 'genomic islands' of divergence is consistent with divergence hitchhiking potentially contributing greatly to speciation-with-gene-flow. (b) Genomic divergence in many regions consistent with multifarious selection on many traits and genes, establishing conditions conducive for genome hitchhiking. Note that under divergence of many regions, some general baseline differentiation across the genome (i.e., even of neutral regions) might occur, but differentiation can nonetheless still be highly heterogeneous across the genome. See text for details.

Fig. 2. Representative adult male stickleback from Robert's lake (bottom) and from its outlet stream (top). Fish from streams and lakes differ for many ecologically relevant traits including gill raker number and length, body depth, and body size, as measured by Roesti *et al.* (2012). Photo credit Daniel Berner.

genomic regions than often thought (e.g., see Michel *et al.* 2010). The finding is consistent with recent theory (Feder & Nosil 2010; Feder *et al.* 2012) implying: (i) that widespread genomic divergence may require divergent selection on many loci; and (ii) that these loci may be distributed widely across the genome rather than being clustered in a few isolated 'islands of speciation'. Other recent studies employing NGS or SNP-chip technologies have reported similar patterns of widespread genomic divergence (Lawniczak *et al.* 2010; McGaugh & Noor 2012), including studies of different stickleback systems (Deagle *et al.* 2012; Hohenlohe *et al.* 2010). Thus, it appears that a combination of recent technological and theoretical advances is challenging the view that speciation typically involves change in just a few genes of large effect. When ecological speciation occurs, it may often involve multifarious selection on many traits and change in numerous genes, several of which have large effect, even in its early stages. In turn, this might allow for widespread genomic divergence facilitated by genome hitchhiking.

Second, the authors report greater divergence in the centre of chromosomes, relative to the periphery, coupled with evidence that recombination is reduced in the centre of chromosomes. Past studies have reported accentuated divergence in regions of low recombination, such as within chromosomal inversions (McGaugh & Noor 2012; Michel *et al.* 2010; Strasburg *et al.* 2009) or proximity to centromeres (Nachman & Payseur 2012). The results of Roesti *et al.* (2012) are consistent with this past work, but differ in that accentuated divergence was not associated with known inversions or proximity to centromeres. Their findings indicate that attention to broad-scale recombination rate variation across the genome warrants attention in future studies. For example, the results indicate that anonymous genome scans should be cautious about interpreting divergent selection as the (only) factor contributing to accentuated genetic divergence, and further illustrate how the search for causal associations between phenotype and genotype can be complicated by low recombination. Importantly, as in the past studies (e.g., McGaugh & Noor 2012; Michel *et al.* 2010; Nachman & Payseur 2012; Strasburg *et al.* 2009), although divergence in Roesti *et al.* (2012) was accentuated in regions of low recombination, differentiation of other regions was also observed.

Third, they report patterns suggesting accentuated differentiation at divergent selected loci is not always expected. The ectodysplasin gene (eda) affects lateral plate number in stickleback. Plate number among populations generally varies according to ecological conditions (Colosimo *et al.* 2005; Reimchen 1995). Although Roesti *et al.* (2012) report accentuated divergence at regions flanking eda between lake and stream stickleback, they found that the eda locus itself was not strongly differentiated. This result concurs with recent work noting how selection can sometimes leave somewhat nonintuitive patterns in the genome (Bierne *et al.* 2011) and urges caution when interpreting patterns of genomic divergence.

Fourth, Roesti *et al.* (2012) report a positive relationship between the degree of morphological differentiation for ecologically relevant traits between lake and stream

Fig. 3. Typical habitat used by stickleback residing in lakes and streams (left: Boot Lake; right: outlet stream of Robert's Lake). Apart from differences in abiotic features, these habitats differ strikingly in prey types available to stickleback; zooplankton is a major prey resource in lakes, whereas benthic prey is the major resource in streams. Differences in fish phenotypes equate with these abiotic and ecological features of the two habitats. Photo credit Daniel Berner.

populations and the degree of genomic divergence. Specifically, more phenotypically divergent populations had greater levels of baseline genetic differentiation and more heterogeneous patterns of genomic divergence (e.g., greater discrepancy between divergence in the centre versus periphery of chromosomes). They also report a large degree of nonparallelism in genomic divergence among population pairs from different watersheds, a finding that compliments past work in other stickleback systems where parallelism has been reported for some regions, but not for others (Colosimo *et al.* 2005; Deagle *et al.* 2012; Hohenlohe *et al.* 2010). Although nonparallelism at the genetic level means that different gene regions could be responsible for the parallel adaptive phenotypic responses to lake and stream habitats seen across watersheds, other possibilities include the following: (i) different patterns of linkage disequilibrium for neutral markers surrounding target genes in different populations; and (ii) geographically variable sources of selection in addition to those between lake vs. streams may be affecting patterns of genomic divergence.

Collectively, the results of Roesti *et al.* (2012) illustrate how genomic divergence can differentially build up at differing points in the speciation process. At the initiation of speciation, very high gene flow may strongly preclude genomic divergence, except perhaps for the most strongly selected loci. This may correspond to the situation reported by Roesti *et al.* (2012) at Misty Lake, where fish show minimal morphological and genetic differentiation and migration is presumed high. In comparison, later in the speciation process when more loci differentiate, widespread divergence may occur via genome hitchhiking (Feder *et al.* 2012; Nosil & Feder 2012). This may correspond to the Joe and Boot's systems. Further genomic studies across the speciation continuum are required to better characterize these stages of speciation. However, a surprising result of current studies, including that of Roesti *et al.* (2012), is that taxa appear to reach relatively widespread

genomic differentiation comparatively rapidly following colonization of novel habitats.

Future directions

Future work could benefit from a number of methodological and conceptual advances. For example, models that incorporate genotype uncertainty, rather than assuming genotypes are known, are required to better deal with the uneven coverage among individuals and gene regions that is inherent in NGS data (Gompert & Buerkle 2011; Gompert *et al.* 2012). Future work could also tackle the fact that different evolutionary processes can generate similar patterns in the genome. For example, both strong selection and low recombination rate can potentially facilitate exceptional differentiation for specific gene regions (Feder & Nosil 2010). The study by Roesti *et al.* (2012) shows how progress can be made by characterizing the contribution recombination rate makes to genomic divergence. Further development of better metrics for quantifying genomic divergence is required to help push the field from vague metaphorical descriptions towards more quantitative treatments of patterns and ultimately inference of underlying processes.

Along these lines, observational genome scans, even those using many loci, can be biased towards underestimating the number of divergently selected loci, because weakly selected loci (or those in regions of high recombination) will be only moderately differentiated and thus deemed part of the neutrally evolving genome. Experimental studies, as recently seen in microbes (Barrick *et al.* 2009) and flies (Burke *et al.* 2010; Michel *et al.* 2010), might more directly test the number of genomic regions subject to selection. Ultimately, detailed studies of natural populations will be coupled with field experiments to yield a comprehensive picture of the causes and consequences of genomic divergence and its role in fostering adaptation and speciation.

References

Barrick JE, Yu DS, Yoon SH *et al.* (2009) Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature*, **461**, U1243–U1274.

Barton N, Bengtsson BO (1986) The barrier to genetic exchange between hybridizing populations. *Heredity*, **57**, 357–376.

Bierne N, Welch J, Loire E, Bonhomme F, David P (2011) The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology*, **20**, 2044–2072.

Burke MK, Dunham JP, Shahrestani P *et al.* (2010) Genome-wide analysis of a long-term evolution experiment with Drosophila. *Nature*, **467**, 587–590.

Charlesworth B, Nordborg M, Charlesworth D (1997) The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetical Research*, **70**, 155–174.

Colosimo PF, Hosemann KE, Balabhadra S *et al.* (2005) Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science*, **307**, 1928–1933.

Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates, Inc., Sunderland, Minnesota, USA.

Deagle BE, Jones FC, Chan YF *et al.* (2012) Population genomics of parallel phenotypic evolution in stickleback across stream-lake ecological transitions. *Proceedings of the Royal Society B-Biological Sciences*, doi:10.1098/rspb.2011.1552.

Feder JL, Nosil P (2010) The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution*, **64**, 1729–1747.

Feder JL, Geiji R, Yeaman S, Nosil P (2012) Establishment of new mutations under divergence and genome hitchhiking. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **367**, 461–474.

Gompert Z, Buerkle CA (2011) A hierarchical Bayesian model for next-generation population genomics. *Genetics*, **187**, 903–917.

Gompert Z, Lucas LK, Nice CC *et al.* (2012) Genomic regions with a history of divergent selection affect fitness of hybrids between two butterfly species. *Evolution*, doi: 10.1111/j.1558-5646.2012.01587.x.

Hohenlohe PA, Bassham S, Etter PD *et al.* (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *Plos Genetics*, **6**, e1000862, doi:10.1371/journal.pgen.1000862.

Lawniczak MKN, Emrich SJ, Holloway AK *et al.* (2010) Widespread divergence between incipient *Anopheles gambiae* species revealed by whole genome sequences. *Science*, **330**, 512–514.

McGaugh SE, Noor MAF (2012) Genomic impacts of chromosomal inversions in parapatric *Drosophila* species. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **367**, 422–429.

Michel AP, Sim S, Powell THQ *et al.* (2010) Widespread genomic divergence during sympatric speciation. *Proceedings of the National Academy of Sciences*, **107**, 9724–9729.

Nachman MW, Payseur BA (2012) Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **367**, 409–421.

Nosil P, Feder JL (2012) Genomic divergence during speciation: causes and consequences. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **367**, 332–342.

Nosil P, Funk DJ, Ortíz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, **18**, 375–402.

Reimchen TE (1995) Predator-induced cyclical changes in lateral plate frequencies of Gasterosteus. *Behaviour*, **132**, 1079–1094.

Roesti M, Hendry AP, Salzburger W, Berner D (2012) Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. *Molecular Ecology*, **21**, 2852–2862.

Strasburg JL, Scotti-Saintagne C, Scotti I, Lai Z, Rieseberg LH (2009) Genomic patterns of adaptive divergence between chromosomally differentiated sunflower species. *Molecular Biology and Evolution*, **26**, 1341–1355.

Via S, West J (2008) The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology*, **17**, 4334–4345.