MOLECULAR ECOLOGY

Molecular Ecology (2012) doi: 10.1111/j.1365-294X.2012.05509.x

FROM THE COVER

Genome divergence during evolutionary diversification as revealed in replicate lake–stream stickleback population pairs

MARIUS ROESTI,* ANDREW P. HENDRY,† WALTER SALZBURGER* and DANIEL BERNER*

Zoological Institute, University of Basel, Vesalgasse 1, CH-4051 Basel, Switzerland, †Department of Biology and Redpath Museum, McGill University, 859 Sherbrooke St W., Montreal, QC, Canada H3A 2K6

Abstract

Evolutionary diversification is often initiated by adaptive divergence between populations occupying ecologically distinct environments while still exchanging genes. The genetic foundations of this divergence process are largely unknown and are here explored through genome scans in multiple independent lake–stream population pairs of threespine stickleback. We find that across the pairs, overall genomic divergence is associated with the magnitude of divergence in phenotypes known to be under divergent selection. Along this same axis of increasing diversification, genomic divergence becomes increasingly biased towards the centre of chromosomes as opposed to the peripheries. We explain this pattern by within-chromosome variation in the physical extent of hitchhiking, as recombination is greatly reduced in chromosome centres. Correcting for this effect suggests that a great number of genes distributed widely across the genome are involved in the divergence into lake vs. stream habitats. Analyzing additional allopatric population pairs, however, reveals that strong divergence in some genomic regions has been driven by selection unrelated to lake–stream ecology. Our study highlights a major contribution of large-scale variation in recombination rate to generating heterogeneous genomic divergence and indicates that elucidating the genetic basis of adaptive divergence might be more challenging than currently recognized.

Keywords: F_{ST} outlier, *Gasterosteus aculeatus*, gene flow, next generation sequencing, population genomics, RAD, speciation

Received 29 November 2011; revision received 20 January 2012; accepted 24 January 2012

Introduction

Speciation often begins with the adaptive divergence of populations into selectively different ecological environments despite the presence of initially high gene flow (Endler 1977; Schilthuizen 2000; Coyne & Orr 2004; Gavrilets 2004; Via 2009; Sobel *et al.* 2010). The molecular underpinnings of this process remain poorly understood (Wu 2001; Via 2009; Nosil & Schluter 2011). One fundamental unresolved question is how genetic differentiation that builds up between diverging populations is distributed across the genome. Adaptive divergence between populations certainly implies that selection is strong enough to overcome the homogenizing effect of gene flow at ecologically relevant loci (hereafter 'QTLs') (Wu 2001; Nosil *et al.* 2009; Via 2009). But how many QTLs are involved, how are they arranged across the genome, and how does their divergence influence selectively neutral parts of the genome? Opinions differ widely. At one extreme, some studies argue that, because of hitchhiking, divergence at a few QTLs of major effect can protect large genomic regions from gene flow between selective environments (Turner *et al.* 2005; Via & West 2008; Via 2009). Within these regions, divergence between environments will be elevated relative to the rest of the genome, and additional QTLs can become recruited for further adaptive divergence. At the other extreme, adaptive divergence might involve numerous QTLs of relatively small effect, in which case

Correspondence: Daniel Berner, Fax: +41 (0)61 267 0301;
E-mail: daniel.berner@unibas.ch

the hitchhiking of neutral regions along with selected QTLs is predicted to be greatly restricted physically (Barton & Bengtsson 1986; Feder & Nosil 2010). In this case, genetic divergence will either be highly localized or will build up homogeneously throughout the entire genome if reproductive barriers associated with adaptive divergence restrict gene flow effectively enough.

Evaluating the generality of these extreme (as well as intermediate) views on genomic divergence during speciation with gene flow is currently precluded by the scarcity of empirical evidence. The most powerful empirical solutions to this problem are expected to emerge from studies providing high-resolution genome-wide data from multiple replicate population pairs in the initial stages of ecological divergence. These 'species in waiting' are particularly informative because the genomic footprints of selection will not yet have been obscured by evolutionary processes acting *after* the completion of reproductive isolation (Coyne & Orr 2004; Via 2009; Nosil & Schluter 2011). Moreover, the incorporation of multiple population pairs differing in their magnitude of divergence allows an explicit examination of how genomic divergence builds up, as opposed to providing only a single temporal snapshot. Our study adopts this approach by combining the power of high-throughput sequencing technology with the availability of an emerging model for studying divergence with gene flow: replicate lake and stream populations of threespine stickleback fish (*Gasterosteus aculeatus*).

Threespine stickleback inhabit contiguous lake and stream habitats in many watersheds that were colonized independently by marine ancestors following the last glacial retreat (Reimchen *et al.* 1985; Lavin & McPhail 1993; Thompson *et al.* 1997; Hendry & Taylor 2004; Berner *et al.* 2009, 2010; Deagle *et al.* 2011). Different lake and stream population pairs typically exhibit similar directions of phenotypic divergence in a number of traits as a response to similar divergent selection (Reimchen *et al.* 1985; Lavin & McPhail 1993; Hendry & Taylor 2004; Berner *et al.* 2009; Kaeuffer *et al.* 2011; Deagle *et al.* 2011). This adaptive divergence likely represents the initial stage of speciation because it frequently coincides with the emergence of at least partial reproductive isolation. (Although it does not necessarily imply that divergence will ever become complete.) In particular, strong shifts in neutral marker allele frequencies occur across lake–stream transitions of just a few hundred metres, even in the absence of physical dispersal barriers (Berner *et al.* 2009). Here, we examine four evolutionarily independent lake and outlet stream stickleback population pairs ('systems') from Vancouver Island, Canada. These systems differ in their magnitude of divergence because of differences in the strength of divergent selection, the time for divergence and/or differences in available genetic variation (Hendry & Taylor 2004; Moore *et al.* 2007; Berner *et al.* 2009; Kaeuffer *et al.* 2011). Importantly, this variation among systems allows us to investigate genomic patterns along a gradient of divergence (Nosil & Schluter 2011). We here present genome scans for all four lake–stream systems based on thousands of markers obtained through Illumina sequencing of restriction site-associated DNA.

Material and methods

Study populations and phenotypic analysis

Our study builds on stickleback sampled from one lake and one outlet stream site in the Boot, Joe's, Misty, and Robert's watersheds on Vancouver Island, British Columbia, Canada (sites Boot 'L' and 'S2', Joe's 'L' and 'S2', Misty 'L' and 'S6', and Robert's 'L' and 'S2' in Berner *et al.* 2009). The population pair in each of these systems derives from independent postglacial colonization by marine ancestors (Hendry & Taylor 2004; Berner *et al.* 2009). Absolute barriers to dispersal between lakes and streams are absent in all systems, providing the opportunity for gene flow between the habitats. Details on sampling methods and the populations are provided in Berner *et al.* (2009). This analysis is based on 27 individuals per site (216 in total).

For phenotypic traits, we quantified gill raker number and length, and landmark-based body size and relative body depth, as described in Berner *et al.* (2008, 2011). These traits are known to show strong genetically based divergence between lake and stream populations (Lavin & McPhail 1993; Sharpe *et al.* 2008; Berner *et al.* 2011). We here combined these data into a single multivariate summary metric of within-system phenotypic divergence by mean-scaling each trait and then calculating the Euclidean distance between the lake and the stream sample (univariate patterns are shown in Appendix S1, Supporting information).

Marker generation and quantification of population divergence

To obtain genetic markers, we first prepared libraries of individually barcoded, restriction site-associated DNA (RAD; Baird *et al.* 2008) by largely following the protocol in Hohenlohe *et al.* (2010). Each of the 12 total libraries combined RAD from 18 individuals and was single-end sequenced with 76 cycles in a separate lane on an Illumina genome analyzer IIx. The resulting reads (NCBI short read archive accession number SRP007695) were sorted individually by barcode and then aligned to the reference stickleback genome (Ensembl database version 63.1, assembly Broad S1) by using Novoalign

v2.07.06 (http://novocraft.com). We tolerated an equivalent of approximately six high-quality mismatches or gaps and enforced unique alignment, thereby excluding data from repeated elements. Alignments were BAM-converted using Samtools v0.1.11 (Li *et al.* 2009).

For each individual and RAD locus, we then determined the consensus diploid genotype if ten or more replicate reads were available or a haploid consensus genotype if replication was below ten. This threshold was chosen because we identified heterozygote diploids for variable nucleotide positions by a binomial test with insufficient power at low replication. This test involved calculating the binomial likelihood of the observed variant frequency distribution under the null hypothesis of heterozygosity (i.e. assuming a probability of 0.5 for both variants) and accepted heterozygosity if the likelihood was >0.01. Consensus genotyping was quality aware in that bases with a >0.01 error probability were ignored.

To identify single nucleotide polymorphisms (including a small fraction of microindels, hereafter simply subsumed under 'SNPs'), we pooled the individual consensus genotypes from both habitats within a system at each RAD locus. If a locus was represented by at least 27 consensus genotypes from each habitat (i.e. each individual contributed at least one haplotype on average), we screened every nucleotide position of the locus for variants. Otherwise, the locus was ignored because the quantification of population differentiation was considered unreliable.

Before detected SNPs could be used as genetic markers for analysis, we had to eliminate those lacking the potential to adequately capture the signatures of drift and selection because of a low minor allele frequency. We did so by discarding SNPs with a minor allele frequency of <0.25 (justification and details given in Appendix S2, Supporting information). This filter also effectively eliminated sequencing errors and PCR artefacts from the data but reduced the number of polymorphic RAD loci substantially (e.g. from 12 495 to 4127 in the Boot system). Summary statistics on library size, read coverage, alignment success and marker numbers are provided in Appendix S3 (Supporting information). The remaining (informative) SNPs were then used to calculate F_{ST} based on haplotype diversity (Nei & Tajima 1981, equation 7). For loci harbouring multiple SNPs, we retained for analysis only the one yielding the highest F_{ST} value. However, working with F_{ST} averaged over all SNPs at a given RAD locus, or drawing a single SNP at random, produced similar results supporting identical conclusions in all analyses. Furthermore, using as an alternative divergence metric the chi-square ratio calculated from allele frequencies within a population pair also produced consistent results throughout, highlighting the robustness of our F_{ST}-based strategy.

Differentiation and recombination rate within chromosomes

Genome-wide F_{ST} patterns suggested a systematic bias of lake–stream divergence towards the chromosome centres (hereafter called 'chromosome centre-biased divergence', CCBD; see Results). To formally quantify this observation, we divided each chromosome physically into its 'centre' (inner 50% of a chromosome's sequence) and its 'peripheries' (outer 25% on each side). We then subtracted mean F_{ST} of all markers in the periphery from mean F_{ST} of the markers in the centre and calculated the mean and 95% confidence interval for this CCBD metric within each lake–stream system by using all chromosomes as data points ($N = 21$).

To explore whether CCBD was associated with recombination rate, we extracted information on genetic (linkage) distance (in cM) and physical distance (in mb) for the SNPs and microsatellite markers underlying the stickleback linkage maps presented in Albert *et al.* (2008) and Greenwood *et al.* (2011). The ratio of genetic by physical distance for neighbouring markers then provided an estimate of the average recombination rate for that marker interval (Appendix S4, Supporting information). In addition, we used information on the physical location of the centromere on each chromosome (Urton *et al.* 2011) to evaluate whether heterogeneity in divergence and recombination rate along chromosomes was related to centromere position.

Sliding window analysis screening for outlier regions

The magnitude of population divergence at a given locus proved dependent on chromosome position at a large physical scale (CCBD). Screening for localized regions of high divergence (F_{ST} 'outliers'), potentially indicating hitchhiking along with QTLs under divergent selection (Storz 2005; Nielsen 2005), thus first required an adjustment of F_{ST} values to account for CCBD (see Discussion). To do so, we subjected system- and chromosome-specific F_{ST} data to locally weighted scatterplot smoothing ('LOESS', a nonparametric regression) with chromosome position as predictor. (The polynomial degree was zero in all analyses; hence, LOESS produced a moving average). We used a relatively high bandwidth (0.3) to capture only the coarse heterogeneity in divergence within a chromosome. We then calculated 'residual divergence' at each marker as the difference between the raw and the fitted F_{ST} values. Because CCBD increased with overall divergence (see below),

this procedure had a large effect in the Boot system but a relatively minor effect in the other systems.

To explore the number and physical arrangement of outlier regions, *residual* divergence within each system was subjected to sliding window analysis using LOESS with a narrow bandwidth (0.03) facilitating visualization while adequately conserving small-scale divergence heterogeneity along each chromosome. We excluded the Misty system from this analysis because we suspected a low signal to noise ratio in this barely differentiated lake–stream pair. Outlier significance thresholds were determined empirically based on a resampling strategy (Appendix S5, Supporting information). In addition to the 'parapatric' lake–stream comparisons within each system, we also performed 'allopatric' comparisons between populations of the same habitat type (i.e. lake–lake and stream–stream population pairings). Parapatric vs. allopatric comparisons then allowed us to compare patterns of genomic divergence across different ecological settings.

All analyses except for sequence alignment were performed in R (R Development Core Team 2010), making use of the R-Bioconductor packages ShortRead (Morgan *et al.* 2009), Rsamtools, and Biostrings.

Results

We found dramatic differences among systems in the magnitude of overall baseline genomic divergence (Fig. 1), and this paralleled the magnitude of differences among systems in phenotypic divergence (see red and blue bars in Fig. 2). In particular, approximately 0.6% of the markers in the most divergent system (Boot) reached fixation of alternative variants between the habitats. By contrast, no locus reached appreciable divergence in the Misty system. Furthermore, heterogeneity in divergence along the genome increased with increasing baseline divergence (Fig. 1).

As noted earlier, a striking pattern towards higher F_{ST} values in the chromosome centres than in the chromosome peripheries was evident, particularly in the Boot system. A metric based on the difference in mean F_{ST} between markers from the centre and from the peripheries of each chromosome confirmed this pattern (Fig. 2), which we call 'chromosome centre-biased divergence' (CCBD). CCBD averaged across chromosomes within systems was related to the overall magnitude of phenotypic and baseline genetic lake–stream divergence in those systems: that is, CCBD was absent in the undifferentiated Misty system but was very

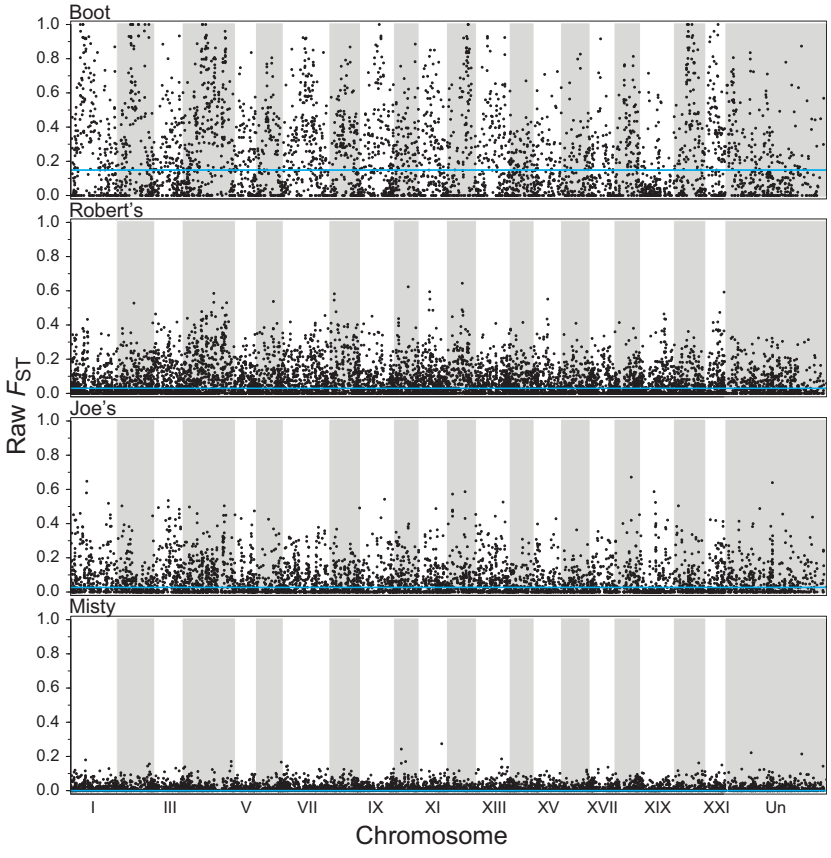

Fig. 1 Genome-wide divergence in four independent population pairs (systems) of lake and stream stickleback. The dots show F_{ST} values for each marker on each chromosome; the chromosomes are separated by white and grey background shading. ['Un' is the artificial chromosome consisting of concatenated unanchored scaffolds. Also, chromosome XIX was corrected for misassembly (Ross & Peichel 2008) in all analyses.] Total marker coverage per system ranges between 4127 and 8417 (Appendix S3, Supporting information). The blue horizontal line represents baseline divergence defined as genome-wide median F_{ST} (Misty: 0; Joe's: 0.027; Robert's: 0.030; Boot: 0.149). Moving from the bottom (Misty) to the top (Boot), note increasing magnitudes of baseline divergence, and increasing heterogeneity in divergence across the genome.

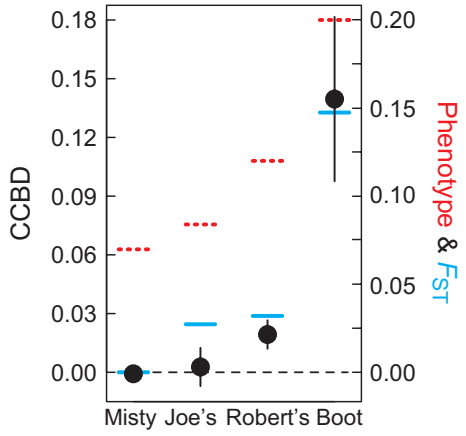

Fig. 2 The emergence of chromosome centre-biased divergence (CCBD). CCBD is expressed as the difference between the chromosome centre (inner 50% of the sequence) and the chromosome peripheries (outer 50%) in the magnitude of differentiation (F_{ST}) between the lake and stream habitat within each system. Dots and error bars are means and 95% confidence intervals across the 21 chromosomes. Positive values indicate relatively greater divergence in the centre of the chromosomes as opposed to their peripheries. CCBD emerges when divergence becomes substantial, as quantified by phenotypic divergence and genome-wide median F_{ST} (dashed red and solid blue horizontal bars, both referring to the right axis). Phenotypic divergence integrates four ecologically important and genetically based morphological traits (Appendix S1, Supporting information).

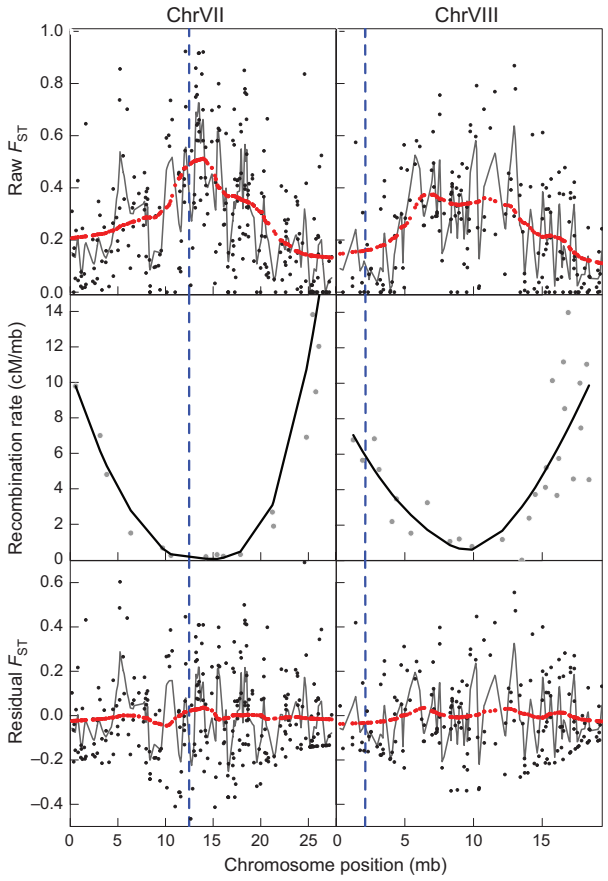

Fig. 3 *Top panels*. Lake–stream divergence in the Boot system along chromosomes seven and eight. Black dots give the raw F_{ST} value at each marker, the grey line connects F_{ST} values predicted by a fine-scale smoother (LOESS, bandwidth = 0.03), and red dots represent F_{ST} values predicted by a coarse smoother (bandwidth = 0.3). These data illustrate that population divergence is greater in the chromosome centres than in their peripheries (CCBD). Divergence profiles for all other chromosomes in the Boot system are presented in Appendix S6a (Supporting information). *Middle panels*. Recombination rates for marker intervals along the same two chromosomes show that recombination is dramatically lower in the chromosome centres relative to the peripheries. Note that heterogeneity along the chromosomes in both F_{ST} and recombination rate is independent from the position of the centromere, indicated by the dashed blue vertical line (chromosome seven is metacentric, whereas chromosome eight is telocentric; Urton *et al.* 2011). Recombination rates and centromere positions for the other chromosomes are presented in Appendix S4 (Supporting information). *Bottom panels*. Separating locus-specific signatures of selection from CCBD. The plotting conventions are as in the top row, except that the underlying data points are residual F_{ST} obtained by subtracting the values predicted by the coarse smoother (red dots) in the top panels from the raw F_{ST} values (black dots) in the top panels.

strong in the Boot system exhibiting greatest progress in divergence (the top row in Fig. 3 shows a fine-scale illustration of CCBD for two exemplary chromosomes in the Boot system; patterns on all chromosomes for that system are presented in supporting online Appendix S6a, Supporting information).

All chromosomes for which enough data were available exhibited a valley of reduced recombination around their centre (Fig. 3, middle row; Appendix S4, Supporting information). The variation in recombination rate was often dramatic, with a 10 fold or higher reduction in the centre of some chromosomes relative to their peripheries (see e.g. chromosome VII, Fig. 3). Both CCBD and physical variation in recombination rate were unrelated to the position of the centromere (Fig. 3; Appendix S4, Supporting information).

After adjusting raw F_{ST} values for CCBD ('residual F_{ST}'; Fig. 3, bottom), our sliding window analyses found outlier regions in relatively high numbers throughout the genome in all three systems (Fig. 4; Misty excluded owing to the overall lack of differentiation). A qualitative comparison indicated that outlier regions were relatively inconsistent across the systems. For instance, we found no peak exceeding the $P < 0.01$ threshold in all three systems. Significant outliers were also observed in allopatric population comparisons, with an exemplar allopatric comparison shown in Fig. 4 (bottom). Interestingly, some outlier regions

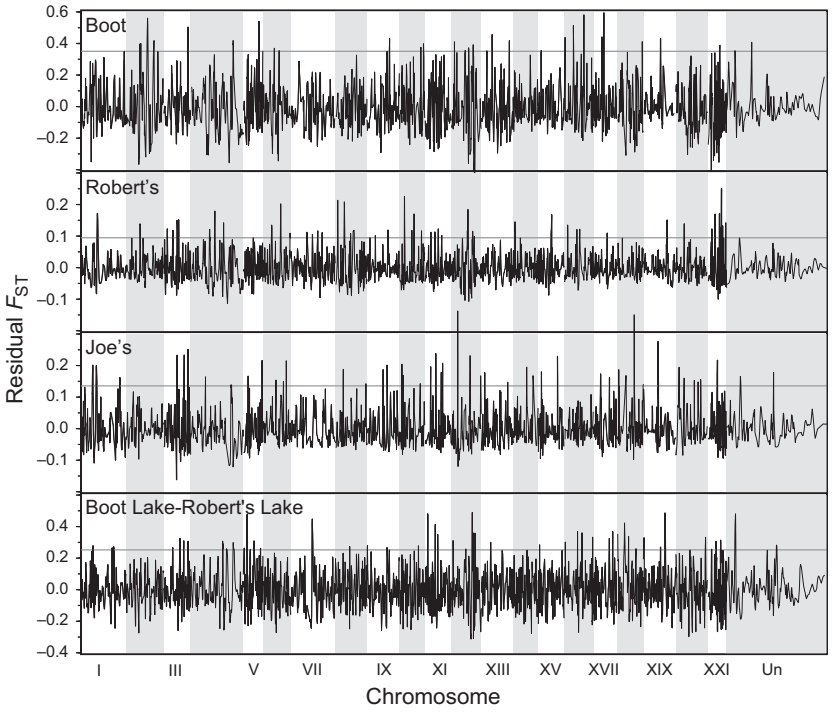

Fig. 4 Sliding window analyses visualizing genome-wide divergence. The top three panels show lake–stream comparisons within the three divergent systems (the Misty system was excluded, because divergence was minimal). The bottom panel shows an exemplary allopatric population comparison (Boot Lake-Robert's Lake; 8735 markers, median F_{ST} = 0.266). The divergence profiles are based on residual (CCBD-corrected) F_{ST} (see Fig. 3). Grey horizontal lines indicate $P < 0.01$ significance thresholds for outlier regions determined by a resampling approach (Appendix S5, Supporting information). Note that the relatively weakly divergent Robert's and Joe's systems are plotted on a two-fold finer scale than the Boot and the allopatric comparisons.

suggested by the parapatric lake–stream comparisons also emerged in allopatric comparisons. A particularly clear case involves the two high-divergence peaks flanking the low-divergence *Ectodysplasin* (*Eda*) locus at a distance of 1–2 mb (Fig. 5).

Discussion

We used stickleback population pairs from lake and stream habitats in multiple independent watersheds to characterize how genomes diverge when populations diversify in the face of gene flow. One major finding is that striking differences are evident among systems in the overall magnitude of lake–stream genomic divergence (Fig. 1) and that these differences match those previously documented for phenotypes and microsatellites (Berner *et al.* 2009). In particular, while baseline divergence is substantial and a number of markers have reached fixation for alternative variants in the Boot system, divergence is weaker in the Robert's and Joe's systems, and negligible in the Misty system. Given that gene flow is known to be very high from Misty Lake into the Misty outlet stream, despite evidence for strong divergent selection (Hendry *et al.* 2002; Moore *et al.* 2007; Berner *et al.* 2009), our analysis here provides a robust demonstration of genome-wide constraints on adaptive divergence as a result of homogenizing gene flow. That is, gene flow in the Misty system overwhelms divergence even for the loci likely subject to the strongest divergent selection.

Chromosome centre-biased divergence (CCBD)

Another major finding is that increasing phenotypic and genetic divergence leads to relatively stronger divergence in chromosome centres than towards their peripheries (CCBD), which contributes to increasing heterogeneity (or variance) in divergence across the genome. The most straightforward explanation for CCBD is the coincidence of adaptive divergence at multiple QTLs and reduced recombination rate in the chromosome centres. The reason is that, for a given magnitude of divergence at a QTL, associated hitchhiking will extend deeper into the neutral neighbourhood if the QTL is located in a genomic region where the recombination rate is relatively reduced (Barton & Bengtsson 1986; Kaplan *et al.* 1989; Charlesworth *et al.* 1997; Feder & Nosil 2010). Moreover, CCBD will be particularly pronounced if a chromosome harbours *multiple* QTLs under divergent selection, because the hitchhiking effect of the QTLs will tend to cumulate more strongly in the centre than in the periphery. Our analysis of recombination rate based on stickleback linkage maps is consistent with this hypothesized mechanism: stickleback chromosomes consistently display reduced recombination in their centres relative to their peripheries (see also Hohenlohe *et al.* 2012). Similar within-chromosome variation in recombination rate has recently been reported from several genetic model organisms [*C. elegans*: Rockman & Kruglyak (2009); zebrafish: Bradley *et al.* (2011); mice, rats, humans: Jensen-Seaman *et al.*

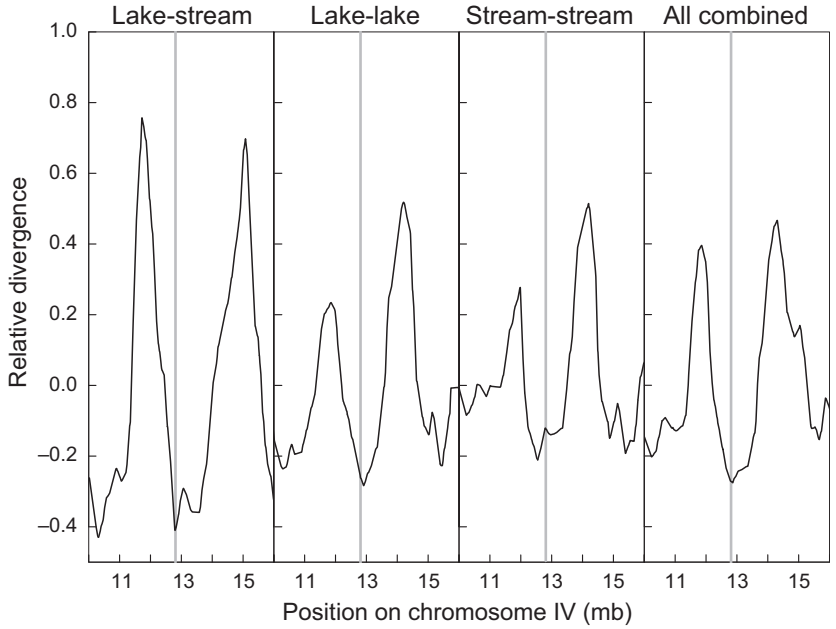

Fig. 5 Divergence profiles along a 6 mb segment of chromosome IV centred at the *Ectodysplasin* (*Eda*) locus (grey vertical line), which is the key genetic factor in lateral plate reduction (Colosimo *et al*. 2005). Profiles are shown for parapatric (lake-stream) and allopatric (lake–lake, stream–stream) population comparisons, and for an analysis combining all these data. The parapatric class averages divergence data from all four lake–stream systems, while each allopatric class averages data from all six possible population pairings. The combined analysis thus integrates data from 16 total population comparisons, yielding a total of 261 markers and an average inter-marker spacing of 23 kb within the displayed chromosome segment. Data averaging was made possible by expressing divergence in each population comparison in units of baseline divergence ('relative divergence'; obtained by scaling residual divergence by genome-wide mean F_{ST}). Note that *Eda* consistently displays below-baseline divergence but is flanked by a high-divergence peak on each side.

(2004); Borodin *et al.* (2008); Chowdhury *et al.* (2009)]. The emergence of CCBD during adaptive divergence might thus be a common phenomenon.

While the reason for within-chromosome variation in recombination rate remains unclear—it appears unrelated to the position of the centromere—the phenomenon has at least two important implications. First, the associated CCBD challenges the conceptual dichotomy between divergence beginning with the emergence of a few large and isolated differentiated regions associated with large-effect QTLs (Via & West 2008; Via 2009), vs. more homogeneous genome-wide divergence associated with numerous QTLs of minor effect (Feder & Nosil 2010). That is, given reduced recombination in chromosome centres, even minor-effect QTLs might drive strong marker divergence over large genomic regions when they happen to co-localize in chromosome centres, whereas large-effect QTLs might not generate much hitchhiking when located in the highly recombining peripheries. Our study thus highlights a key role of variable recombination rate in generating heterogeneous genomic divergence during evolutionary diversification and indicates that the prevailing focus on pericentric regions and inversions (Butlin 2005; Kirkpatrick & Barton 2006; Hoffmann & Rieseberg 2008; Feder & Nosil 2009; Noor & Bennett 2009) misses important variation in recombination rate at a much larger physical scale.

A second implication of within-chromosome variation in recombination rate and CCBD is methodological. Because hitchhiking is expected to be more extensive in chromosome centres, the probability of a particular marker detecting the signature of a locus under selection is relatively higher in the chromosome centres. In addition, genomic regions under divergent selection in nonmodel organisms are often identified by anonymous genome scans that do not map markers to a reference genome or a linkage map (e.g. Beaumont & Balding 2004; Foll & Gaggiotti 2008; Excoffier *et al.* 2009). These approaches assume that locus-specific F_{ST} values can be evaluated against a *genome-wide* baseline. CCBD undermines this assumption and hence leads to a systematic bias towards identifying outliers at markers located near chromosome centres. That is, anonymous genome scans cannot separate localized signatures of hitchhiking associated with *specific* selected QTLs from diffuse, large-scale heterogeneity in divergence along chromosomes driven by *multiple* selected QTLs and large-scale reduced recombination. Our strategy to address this problem was to express divergence at each marker as the deviation of the raw F_{ST} value from the F_{ST} value

predicted by a coarse smoothing function capturing CCBD (yielding 'residual divergence'; Fig. 3). We do not claim that this *ad hoc* empirical standardization is optimal. Until more sophisticated methods are developed, however, localized signatures of selection in systems exhibiting CCBD are certainly better inferred based on residual F_{ST} than on raw F_{ST}.

Outlier analysis

Our sliding window analyses based on residual (CCBD-adjusted) divergence suggested the presence of dozens of outliers spread throughout the genome within each system. Many loci thus appear to contribute to adaptive divergence, as has also recently been inferred for *Anopheles* mosquitoes (Lawniczak *et al.* 2010), and marine vs. freshwater stickleback (Hohenlohe *et al.* 2010). This finding contradicts the idea that during the early stages of speciation, divergence builds up in only a few genomic hotspots associated with major QTLs (Via & West 2008; Via 2009). Our inference of numerous selected QTLs is also consistent with the observation of CCBD; if only a few loci were targeted by selection, strong and consistent CCBD would not be expected to emerge.

We also found that divergence profiles were rather inconsistent among our systems, making it difficult to identify genetic regions of *general* importance to lake–stream stickleback divergence. A similar conclusion was reached in a recent lower-resolution genome scan using several lake–stream stickleback populations from another region of British Columbia (Deagle *et al.* 2011). Possible explanations include differences in the nature of divergent lake–stream selection among the systems (Berner *et al.* 2008, 2009; Kaeuffer *et al.* 2011), or the possibility that responses to similar divergent selection involve different QTLs in the different systems (Arendt & Reznick 2008; Kaeuffer *et al.* 2011). The latter would not be surprising, as many traits involved in adaptive divergence between lakes and streams are likely polygenic (Peichel *et al.* 2001; Albert *et al.* 2008; Greenwood *et al.* 2011).

It is also possible, however, that the inconsistency in outliers among systems reflects a fundamental limitation of genome scans. Drawing on theory (Slatkin & Wiehe 1998; Barton 2000; Bierne 2010), we predict that the recurrent fixation of an *unconditionally* favourable QTL allele from the standing genetic variation will generate peaks of high population divergence in neutral regions flanking the QTL on both sides, while the QTL itself will remain undifferentiated. The reason is that different copies of the favourable allele will share their immediate neutral neighbourhood, while potentially being associated with different neutral variants further away from the QTL. The pattern we found at the *Eda* locus across parapatric *and* allopatric population comparisons (Fig. 5) is consistent with this scenario. *Eda* is the key genetic factor underlying adaptive lateral plate reduction following freshwater colonization by stickleback, and all of our (low-plated) populations are likely fixed for the same derived *Eda* allele available in the ancestral standing variation of the colonizing marine fish (Colosimo *et al.* 2005). The twin peaks flanking *Eda* therefore reflect hitchhiking with a single unconditionally favourable allele (i.e. an allele favoured in *both* lakes and streams) rather than two separate signatures of divergent lake–stream selection. Interestingly, Deagle *et al.* (2011) inferred a locus presumably influenced by divergent lake–stream selection almost exactly at the tip of the left peak flanking *Eda* in our analysis (at 12 mb; see the first marker in their Table 2). Similarly, Jones *et al.* (2012) interpreted two outlier regions flanking *Eda* (at 11.4 mb and 15.7 mb; see their Fig. 3) as indicating loci involved in the divergence of sympatric benthic–limnetic stickleback. In the light of our findings, these interpretations need to be revised.

Overall, the conclusion that lake–stream divergence involves numerous QTLs is probably robust. However, the above-mentioned considerations (Slatkin & Wiehe 1998; Bierne 2010) and results highlight that regions of high divergence identified in (replicate) genome scans are not necessarily related to *divergent* selection mediated by the causal factor of interest (here lake–stream ecology). Allele frequency shifts at QTLs driven by *any* type of selection within a local population can generate outliers in linked markers between populations (Charlesworth *et al.* 1997; Charlesworth 1998).

Conclusions

Our genome scan comparisons of multiple lake–stream stickleback population pairs have shown that increasing phenotypic divergence coincides with increasing overall genomic divergence, and with increasing large-scale heterogeneity in divergence across the genome. Heterogeneous divergence is strongly driven by within-chromosome variation in recombination rate, a phenomenon that might be common and hence requires conceptual integration in speciation genetics. Large-scale heterogeneous divergence also represents an unappreciated methodological challenge to genome scans in search for selected loci. Our study further suggests that lake–stream divergence involves shifts at numerous QTLs throughout the genome but also cautions that inferring the selective context underlying regions of high divergence is less straightforward than generally recognized.

Acknowledgements

We thank A.-C. Grandchamp, J.-S. Moore and K. Hudson for aiding field work; P. Etter and W. Cresko for sharing their

expertise in RAD library preparation; B. Aeschbach and N. Boileau for facilitating wet laboratory work; I. Nissen for Illumina sequencing at the Quantitative Genomics Facility, D-BSSE, ETH Zürich; L. Zimmermann for IT support; C. Hercus for modifying Novoalign; M. Morgan and H. Pagès for coding suggestions; D. Ebert, C. Peichel, H. Hoekstra, H. Gante and three referees for valuable comments and suggestions on data analysis; J. Urton for details on centromere positions; M. Hansen and T. Vines for efficient manuscript handling. APH was funded by the Natural Sciences and Engineering Research Council (NSERC) of Canada, WS by the European Research Council (ERC, Starting Grant 'INTERGENADAPT'), the Swiss National Science Foundation and the University of Basel, and DB by the Swiss National Science Foundation (PBBSA-111216 and Ambizione PZ00P3_126391/1).

References

Albert AYK, Sawaya S, Vines TH *et al.* (2008) The genetics of adaptive shape shift in stickleback: pleiotropy and effect size. *Evolution*, **62**, 76–85.

Arendt J, Reznick D (2008) Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends in Ecology and Evolution*, **23**, 26–32.

Baird NA, Etter PD, Atwood TS *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, **3**, e3376.

Barton NH (2000) Genetic hitchhiking. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, **355**, 1553–1562.

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

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**, 969–980.

Berner D, Adams DC, Grandchamp A-C, Hendry AP (2008) Natural selection drives patterns of lake-stream divergence in stickleback foraging morphology. *Journal of Evolutionary Biology*, **21**, 1653–1665.

Berner D, Grandchamp A-C, Hendry AP (2009) Variable progress toward ecological speciation in parapatry: stickleback across eight lake-stream transitions. *Evolution*, **63**, 1740–1753.

Berner D, Roesti M, Hendry AP, Salzburger W (2010) Constraints on speciation suggested by comparing lake-stream stickleback divergence across two continents. *Molecular Ecology*, **19**, 4963–4978.

Berner D, Kaeuffer R, Grandchamp A-C *et al.* (2011) Quantitative genetic inheritance of morphological divergence in a lake-stream stickleback ecotype pair: implications for reproductive isolation. *Journal of Evolutionary Biology*, **24**, 1975–1983.

Bierne N (2010) The distinctive footprints of local hitchhiking in a varied environment and global hitchhiking in a subdivided population. *Evolution*, **64**, 3254–3272.

Borodin PM, Karamysheva TV, Belonogova NM *et al.* (2008) Recombination map of the common shrew, *Sorex araneus* (eulipotyphla, mammalia). *Genetics*, **178**, 621–632.

Bradley KM, Breyer JP, Melville DB *et al.* (2011) An SNP-based linkage map for zebrafish reveals sex determination loci. *G3: Genes, Genomes, Genetics*, **1**, 3–9.

Butlin RK (2005) Recombination and speciation. *Molecular Ecology*, **14**, 2621–2635.

Charlesworth B (1998) Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution*, **15**, 538–543.

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.

Chowdhury R, Bois PRJ, Feingold E, Sherman SL, Cheung VG (2009) Genetic analysis of variation in human meiotic recombination. *PLoS Genetics*, **5**, e1000648.

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, Sunderland, Massachusetts.

Deagle BE, Jones FC, Chan YF *et al.* (2011) Population genomics of parallel phenotypic evolution in stickleback across stream-lake ecological transitions. *Proceedings of the Royal Society of London Series B Biological Sciences*, in press.

Endler JA (1977) *Geographic Variation, Speciation, and Clines*. Princeton University, Princeton, New Jersey.

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity*, **103**, 285–298.

Feder JL, Nosil P (2009) Chromosomal inversions and species differences: when are genes affecting adaptive divergence and reproductive isolation expected to reside within inversions? *Evolution*, **63**, 3061–3075.

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

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, **180**, 977–993.

Gavrilets S (2004) *Fitness Landscapes and the Origin of Species*. Princeton University, Princeton, New Jersey.

Greenwood AK, Jones FC, Chan YF *et al.* (2011) The genetic basis of divergent pigment patterns in juvenile threespine sticklebacks. *Heredity*, **107**, 155–166.

Hendry AP, Taylor EB (2004) How much of the variation in adaptive divergence can be explained by gene flow? An evaluation using lake-stream stickleback pairs *Evolution*, **58**, 2319–2331.

Hendry AP, Taylor EB, McPhail JD (2002) Adaptive divergence and the balance between selection and gene flow: lake and stream stickleback in the Misty system. *Evolution*, **56**, 1199–1216.

Hoffmann AA, Rieseberg LH (2008) Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? *Annual Review of Ecology and Systematics*, **39**, 21–42.

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.

Hohenlohe PA, Bassham S, Currey M, Cresko WA (2012) Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. *Philosophical Transactions of the Royal Society B, Biological Sciences*, **367**, 395–408.

Jensen-Seaman MI, Furey TS, Payseur BA *et al.* (2004) Comparative recombination rates in the rat, mouse, and human genomes. *Genome Research*, **14**, 528–538.

Jones FC, Chan YF, Schmutz J *et al.* (2012) A genome-wide SNP genotyping array reveals patterns of global and repeated Species-pair divergence in sticklebacks. *Current Biology*, **22**, 83–90.

Kaeuffer R, Peichel C, Bolnick DI, Hendry AP (2011) Parallel and nonparallel aspects of ecological, phenotypic, and genetic divergence across replicate population pairs of lake and stream stickleback. *Evolution*, **66**, 402–418.

Kaplan NL, Hudson RR, Langley CH (1989) The hitchhiking effect revisited. *Genetics*, **123**, 887–899.

Kirkpatrick M, Barton N (2006) Chromosome inversions, local adaptation and speciation. *Genetics*, **173**, 419–434.

Lavin PA, McPhail JD (1993) Parapatric lake and stream sticklebacks on northern Vancouver Island: disjunct distribution or parallel evolution? *Canadian Journal of Zoology*, **71**, 11–17.

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.

Li H, Handsaker B, Wysoker A *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

Moore JS, Gow JL, Taylor EB, Hendry AP (2007) Quantifying the constraining influence of gene flow on adaptive divergence in the lake-stream threespine stickleback system. *Evolution*, **61**, 2015–2026.

Morgan M, Anders S, Lawrence M *et al.* (2009) ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data. *Bioinformatics*, **25**, 2607–2608.

Nei M, Tajima F (1981) DNA polymorphism detectable by restriction endonucleases. *Genetics*, **97**, 145–163.

Nielsen R (2005) Molecular signatures of natural selection. *Annual Review of Genetics*, **39**, 197–218.

Noor MAF, Bennett SM (2009) Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species *Heredity*, **103**, 439–444.

Nosil P, Schluter D (2011) The genes underlying the process of speciation. *Trends in Ecology and Evolution*, **26**, 160–167.

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

Peichel CL, Nereng KS, Oghi KA *et al.* (2001) The genetic architecture of divergence between threespine stickleback species. *Nature*, **414**, 901–905.

R Development Core Team(2010) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reimchen TE, Stinson EM, Nelson JS (1985) Multivariate differentiation of parapatric and allopatric populations of threespine stickleback in the Sangan River watershed, Queen Charlotte Islands. *Canadian Journal of Zoology*, **63**, 2944–2951.

Rockman MV, Kruglyak L (2009) Recombinational landscape and population genomics of *Caenorhabditis elegans*. *PLoS Genetics*, **5**, e1000419.

Ross JA, Peichel CL (2008) Molecular cytogenetic evidence of rearrangements on the Y chromosome of the threespine stickleback fish. *Genetics*, **179**, 2173–2182.

Schilthuizen M (2000) Dualism and conflicts in understanding speciation. *BioEssays*, **22**, 1134–1141.

Sharpe DMT, Räsänen K, Berner D, Hendry AP (2008) Genetic and environmental contributions to the morphology of lake and stream stickleback: implications for gene flow and reproductive isolation. *Evolutionary Ecology Research*, **10**, 849–866.

Slatkin M, Wiehe T (1998) Genetic hitch-hiking in a subdivided population. *Genetical Research*, **71**, 155–160.

Sobel JM, Chen GF, Watt LR, Schemske DW (2010) The biology of speciation. *Evolution*, **64**, 295–315.

Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology*, **14**, 671–688.

Thompson CE, Taylor EB, McPhail JD (1997) Parallel evolution of lake-stream pairs of threespine sticklebacks (*Gasterosteus*) inferred from mitochondrial DNA variation. *Evolution*, **51**, 1955–1965.

Turner TL, Hahn MW, Nuzhdin SV (2005) Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology*, **3**, e285.

Urton JR, McCann SM, Peichel CL (2011) Karyotype differentiation between two stickleback species (Gasterosteidae). *Cytogenetic and Genome Research*, **135**, 150–159.

Via S (2009) Natural selection in action during speciation. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 9939–9946.

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

Wu CI (2001) Genes and speciation. *Journal of Evolutionary Biology*, **14**, 889–891.

M.R. is a PhD student in the Salzburger laboratory and has general interest in the processes underlying biological diversification. He is currently investigating diversification in lake–stream stickleback from population genomic and ecological angles. A.H. investigates factors that influence the evolution of biological diversity, including natural selection, gene flow, adaptation and reproductive isolation. He conducts research in a number of study systems, including Darwin's finches (Galápagos Islands), guppies (Trinidad and Tobago) and stickleback (British Columbia). W.S. is Assistant Professor at the Zoological Institute of the University of Basel. The research of his team focuses on the genetic basis of adaptation, evolutionary innovation and animal diversification. The laboratory's homepage at http://www.evolution.unibas.ch/salzburger/ provides further details on the group's (research) activities. D.B. is interested in adaptation and uses lake and stream stickleback populations for his empirical research.

Data accessibility

DNA sequences: NCBI short read archive accession number SRP007695. Morphological data for Berner *et al.* (2010): Dryad digital repository (doi:10.5061/dryad.1960).

Supporting information

Appendix S1 Univariate divergence in four morphological traits within each lake–stream system.

Appendix S2 Strategy adopted to eliminate uninformative polymorphisms from the marker data sets.

Appendix S3 Summary statistics on library size, read coverage, alignment success and marker numbers.

Appendix S4 Recombination rate along stickleback chromosomes.

Appendix S5 Resampling approach used to determine significance thresholds for F_{ST} outlier regions.

Appendix S6 Genome-wide sliding window profile of lake–stream divergence within the Boot system (both raw F_{ST} and residual F_{ST}).

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.