Received: 28 April 2017 | Revised: 15 September 2017 | Accepted: 18 September 2017

DOI: 10.1111/mec.14373

ORIGINAL ARTICLE

WILEY MOLECULAR ECOLOGY

Genomics of adaptive divergence with chromosome-scale heterogeneity in crossover rate

Daniel Berner[1] | Marius Roesti[1,2]

[1]Zoological Institute, University of Basel, Basel, Switzerland

[2]Department of Zoology, University of British Columbia, Vancouver, BC, Canada

Correspondence
Marius Roesti, Department of Zoology, University of British Columbia, Vancouver, BC, Canada.
Email: roesti@zoology.ubc.ca

Abstract

Genetic differentiation between divergent populations is often greater in chromosome centres than peripheries. Commonly overlooked, this broadscale differentiation pattern is sometimes ascribed to heterogeneity in crossover rate and hence linked selection within chromosomes, but the underlying mechanisms remain incompletely understood. A literature survey across 46 organisms reveals that most eukaryotes indeed exhibit a reduced crossover rate in chromosome centres relative to the peripheries. Using simulations of populations diverging into ecologically different habitats through sorting of standing genetic variation, we demonstrate that such chromosome-scale heterogeneity in crossover rate, combined with polygenic divergent selection, causes stronger hitchhiking and especially barriers to gene flow across chromosome centres. Without requiring selection on new mutations, this rapidly leads to elevated population differentiation in the low-crossover centres relative to the high-crossover peripheries of chromosomes ("Chromosome Centre-Biased Differentiation", CCBD). Using simulated and empirical data, we then show that strong CCBD between populations can provide evidence of polygenic adaptive divergence with a phase of gene flow. We further demonstrate that chromosome-scale heterogeneity in crossover rate impacts analyses beyond that of population differentiation, including the inference of phylogenies and parallel adaptive evolution among populations, the detection of genetic loci under selection, and the interpretation of the strength of selection on genomic regions. Overall, our results call for a greater appreciation of chromosome-scale heterogeneity in crossover rate in evolutionary genomics.

KEYWORDS

barrier to gene flow, hitchhiking, linked selection, population differentiation, recombination, standing genetic variation

1 | INTRODUCTION

During meiosis—the characteristic cell division of sexually reproducing eukaryotes—homologous chromosomes commonly exchange genetic material through crossover. Crossover can, on the one hand, promote adaptation by bringing together beneficial alleles located on the same chromosome but in different genetic backgrounds, or by releasing beneficial alleles from deleterious genetic backgrounds (Fisher, 1930; Hill & Robertson, 1966; Muller, 1932). On the other hand, crossover can constrain adaptation, most obviously so in a population experiencing gene flow from another population residing in a selectively different habitat. Here, crossover can uncouple unfavourable foreign alleles and link them to locally favourable alleles within a population, making their elimination by natural selection less efficient than in the absence of crossover, in which case the foreign alleles would remain coupled as a single block of large

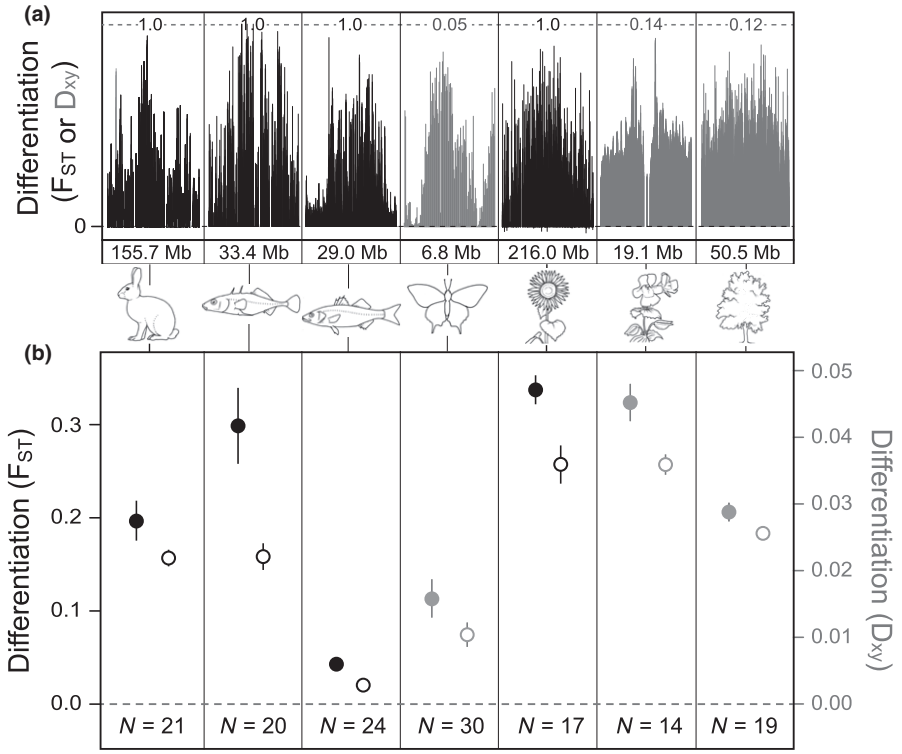

FIGURE 1 Chromosome Centre-Biased Differentiation (CCBD) in natural systems. (a) Genetic differentiation along a single representative chromosome between divergent populations or closely related species of rabbits (Carneiro et al., 2014), threespine stickleback (Roesti et al., 2012a), seabass (Tine et al., 2014), swallowtails (Li et al., 2015), sunflowers (Renaut et al., 2013), monkeyflowers (Puzey et al., 2017) and cottonwood trees (Christe et al., 2017). Differentiation is quantified by either F_{ST} (black; *Y*-axis scaled from 0 to 1) or D_{xy} (grey; *Y*-axis scaled individually). The length of the chromosome displayed is indicated on the bottom (note the variation among the taxa). (b) In the same seven organismal systems, CCBD is a general, genomewide feature of divergence, as indicated by greater average genetic differentiation in chromosome centres (filled circles) than peripheries (open circles). Differentiation quantified by F_{ST} and D_{xy} is shown in black (left *Y*-axis) and grey (right *Y*-axis), respectively. Shown are the grand means across all chromosomes, with vertical bars indicating parametric 95% CIs. The total number of chromosomes (known sex chromosomes excluded) within each genome is indicated on the bottom

deleterious effect (Barton, 1983; Kirkpatrick & Barton, 2006; Yeaman & Whitlock, 2011). This perspective of crossover as a constraint to adaptive divergence is supported by population genomic insights from chromosomal inversions. Inversions are structural genomic rearrangements suppressing crossover in relatively localized chromosome regions (Sturtevant & Beadle, 1936) and can thereby lock together co-adapted alleles into single loci of exceptionally strong selective effect (Feder & Nosil, 2009; Kirkpatrick & Barton, 2006; Ortiz-Barrientos, Reiland, Hey, & Noor, 2002; Rieseberg, 2001). Indeed, inversions often emerge as particularly highly differentiated chromosome regions in genome scans between populations diverged with recent or ongoing gene flow (e.g., Joron et al., 2011; Kirubakaran et al., 2016; Küpper et al., 2016; Lee et al., 2017; Poelstra et al., 2014; Puzey, Willis, & Kelly, 2017; Roesti, Kueng, Moser, & Berner, 2015; Wang et al., 2013).

In this article, however, we aim to demonstrate that a genetic pattern at a larger physical scale than most inversions—that of entire chromosomes—also supports the idea of crossover as a constraint to adaptive divergence. The phenomenon in question is stronger genetic differentiation between divergent populations in the centre of chromosomes than in the chromosome peripheries. We will hereafter refer to this pattern as CCBD—"Chromosome Centre-Biased Differentiation" (Roesti, Hendry, Salzburger, & Berner, 2012a). Indeed, genomewide scans for population differentiation based on well-assembled reference genomes are beginning to suggest that CCBD is common and taxonomically widespread: evidence of CCBD in seven organismal systems is presented in Figure 1.

A potential explanation for CCBD is offered by the population genetic theory on "linked selection" stimulated by early reports of a positive association between crossover rate and levels of genetic diversity within a genome (Begun & Aquadro, 1992; Nachman, Bauer, Crowell, & Aquadro, 1998). In genome regions with a relatively low crossover rate, loci are more tightly linked and selection on a new mutation will cause hitchhiking (Kaplan, Hudson, & Langley, 1989; Maynard Smith & Haigh, 1974) across a physically more extensive chromosome section than in high-crossover regions. Consequently, the continuous selective spread of beneficial mutations (genetic draft model; Gillespie, 2000) and the elimination of recurrent deleterious mutations (background selection model; Charlesworth, Morgan, & Charlesworth, 1993; Hudson & Kaplan, 1995) will cause a relatively reduced effective population size and thus stronger drift in regions with a low crossover rate. The result is reduced

genetic diversity within populations and elevated differentiation among populations in genomic regions with a low crossover rate.

Applying this theory to entire chromosomes, the taxonomically widespread observation of CCBD (Figure 1) leads to the prediction that the rate of crossover should commonly be reduced in chromosome centres relative to the peripheries. We evaluated this idea by performing a survey of the genetic mapping literature, yielding information on the physical distribution of crossover along chromosomes in 46 eukaryotic organisms, including animals, plants and fungi (presented as Table S1). This survey makes clear that a broadscale deficiency in crossover across the centre of chromosomes relative to their peripheries—a pattern often appearing unrelated to the location or presence of centromeres (see Table S1)—is taxonomically so consistent (43 of 46 species) that it may reflect a general mechanistic feature of sexual reproduction.

Nevertheless, the above linked selection theory based on recurrent hitchhiking along with beneficial or deleterious de novo mutations does not provide a universally valid explanation for CCBD, because even populations known to have diverged for a few thousand generations only—and hence from standing (i.e., pre-existing) genetic variation (Barrett & Schluter, 2008; Messer & Petrov, 2013)—can exhibit characteristic broadscale heterogeneity in differentiation along chromosomes (e.g., Hohenlohe et al., 2010; Kolaczkowski, Kern, Holloway, & Begun, 2011; Lawniczak et al., 2010; Neafsey et al., 2010; Roesti et al., 2012a). An additional explanation for CCBD can be derived, however, from the theory on the barrier to gene flow around a locus under divergent selection (Aeschbacher, Selby, Willis, & Coop, 2017; Barton, 1979; Barton & Bengtsson, 1986; Feder & Nosil, 2010; Gavrilets & Cruzan, 1998; Roesti, Gavrilets, Hendry, Salzburger, & Berner, 2014): if chromosome centres display a relative deficiency of crossover and if divergent selection between ecologically different populations acts on multiple loci spread across chromosomes, gene flow between the populations should be relatively reduced and population differentiation elevated in chromosome centres (Nachman & Payseur, 2012; Roesti, Moser, & Berner, 2013; Roesti et al., 2012a). This is because in chromosome centres with reduced crossover rate, maladaptive foreign alleles should remain relatively strongly linked as a genomic block of large deleterious effect, which should facilitate their selective elimination.

CCBD in young organismal systems under divergent selection may thus represent a cumulative signature of multiple geneflow barriers of variable physical extent along chromosomes. This view, however, is complicated by the possibility that genetic hitchhiking caused by the selective sorting of standing variation may contribute to heterogeneous population differentiation as well. Formal theory resolving this ambiguity is lacking. The first goal of this study is therefore to use multilocus simulations to investigate under which conditions CCBD arises when populations start diverging, and in particular, to understand the relative role of hitchhiking versus geneflow barriers as underlying drivers. We then expand our focus beyond population differentiation to demonstrate—using both simulated and empirical data—how broadscale heterogeneity in crossover rate along chromosomes impacts other population genetic metrics and inferential approaches, thus highlighting a greatly under-appreciated conceptual and methodological challenge in evolutionary genomics.

2 | MATERIALS AND METHODS

2.1 | Empirical examples of CCBD and crossover literature survey

To illustrate that CCBD is a taxonomically widespread feature of broadscale genome divergence, we acquired genomewide differentiation data (F_{ST} or D_{xy}) mapped to chromosome-scale reference genomes from various organismal systems: rabbits (*Oryctolagus cuniculus cuniculus* versus *O. c. algirus*; Carneiro et al., 2014), threespine stickleback (*Gasterosteus aculeatus*, lake versus stream; Roesti et al., 2012a), European seabass (*Dicentrarchus labrax*, Atlantic versus Mediterranean; Tine et al., 2014), swallowtails (*Papilio machaon* versus *P. xuthus*; Li et al., 2015), sunflowers (*Helianthus annuus* versus *H. petiolaris*; Renaut et al., 2013), monkeyflowers (*Mimulus guttatus* ecotypes; Puzey et al., 2017) and cottonwood trees (*Populus alba* versus *P. tremula*; Christe et al., 2017). In six of these seven taxon pairs (all except swallowtails), levels of gene flow have been investigated explicitly, in all cases yielding strong evidence of substantial recent genetic exchange. For each system, we visualized the F_{ST} or D_{xy} profile for a single representative chromosome (i.e., chromosome 3, 4, 1A, 2, 4, 2 and 1). To quantify the magnitude of CCBD across *all* chromosomes within each taxon pair, we divided each chromosome into a central (inner 50%) and two peripheral (outer 25% to each side) physical segments and calculated mean differentiation and the associated parametric 95% confidence interval (hereafter "CI") for each segment type across all chromosomes (sex chromosomes were excluded). This pragmatic delimitation of chromosome "centre" and "periphery," resembling the broadscale heterogeneity in crossover rate along chromosomes in diverse organisms (e.g., Anderson et al., 2003; Backström et al., 2010; Bekele, Wieckhorst, Friedt, & Snowdon, 2013; Bhakta, Jones, & Vallejos, 2015; Roesti et al., 2013; Tine et al., 2014), is employed throughout the theoretical and empirical parts of this study.

To explore how consistently the crossover rate is reduced in chromosome centres relative to peripheries across eukaryotic genomes, we performed a literature search for studies presenting genomewide estimates of crossover rate in conjunction with a well-assembled physical map of the genome (i.e., a chromosome-level assembly) (closing date of the literature search: 31 December 2016). For each study meeting these criteria, we then evaluated whether the crossover rate was reduced in chromosome centres relative to the peripheries. A study was rated as positive evidence if it included an explicit qualitative statement by its authors that the crossover rate was generally greater towards the chromosome peripheries than in the centre. Studies lacking such an interpretation were inspected for graphical representations of the crossover landscape (typically genetic marker distance or intermarker crossover rates plotted against physical chromosome position) and were rated as positive evidence if the

focal pattern was unambiguous visually and relatively consistent among chromosomes (studies presenting data from a single chromosome only were ignored). If none of these two criteria were met, a study qualified as negative evidence of a reduced crossover rate in chromosome centres. Overall, this survey included 55 studies from 46 total species (26 animals, 18 plants and 2 fungi).

2.2 | Standard simulation model for adaptive divergence with heterogeneous crossover rate

To scrutinize the relationship between broadscale heterogeneity in crossover rate along chromosomes (i.e., lower crossover rate in chromosome centres than peripheries; hereafter simply referred to as "heterogeneous crossover rate") and CCBD, we implemented individual-based forward simulations. In the base model (hereafter called "standard" model), two populations derived from a common ancestor diverge into ecologically different habitats (H_0 and H_1) in the face of gene flow. Each population consists of a constant number N of haploid hermaphrodites, each represented by a single chromosome. The chromosome harbours 100 evenly distributed genetic loci, of which a fraction—the "selected loci" (*SL*)—are under divergent natural selection between the habitats (the alleles 0 and 1 are favoured in H_0 and H_1, respectively). The other loci, hereafter *NL* for "neutral loci," are selectively neutral, biallelic (0, 1) SNPs. At the onset of each simulation, alleles at the *NL* are drawn at random with an expected frequency of 0.5, thus minimizing the likelihood of loci becoming monomorphic across both populations, and hence uninformative, during the simulations. However, drawing starting frequencies for the alleles from the uniform or the exponential distribution produced similar results supporting identical conclusions (Fig. S1a,b; Table S2 presents an overview of all checks performed to validate our standard simulation model). Alleles at the *SL* are drawn at random with an expected frequency f for the allele 1. The two populations then evolve under divergent selection over g generations. In every generation, the reproductive contribution of an individual to the subsequent offspring generation is a stochastic function of its genotype (Berner & Thibert-Plante, 2015). Specifically, we first calculate absolute fitness as an individual's deviation from the genotypic optimum in the focal habitat. This deviation is given by the count of the locally unfavourable alleles across all *SL* on the chromosome multiplied by the per-locus selection coefficient s. The effect of the loci is thus additive. Given that genetic variation in complex traits, such as fitness, is well described by additive contributions from loci (Hill, Goddard, & Visscher, 2008; Lynch & Walsh, 1998), we consider this choice adequate. However, simulations using a multiplicative fitness function produced similar results supporting identical conclusions (Table S2; Fig. S1c). To then calculate the relative fitness of an individual, its absolute fitness is scaled by the sum of the absolute fitness across all individuals within the focal population. Finally, $N/2$ mating pairs are formed by randomly drawing two individuals with replacement with a probability proportional to their relative fitness, and each pair contributes two offspring to the subsequent generation.

During reproduction, crossover between the two parental chromosomes occurs with a fixed probability of 0.5. Translated to a diploid organism (our modelled system is haploid), this crossover rate per chromosome corresponds to one obligate crossover per meiosis (or a genetic map length of 50 cM), a biologically reasonable assumption (e.g., Backström et al., 2010; Bhakta et al., 2015; Borodin et al., 2008; Fledel-Alon et al., 2009; Roesti et al., 2013). However, simulations with a twofold higher or lower crossover rate produced qualitatively similar results supporting the same conclusions (Table S2; Fig. S2). Moreover, inspecting a limited set of parameter combinations by modelling diploid individuals also produced very similar results (Table S2; Fig. S1d), thus justifying exploring the full-parameter range using the simpler haploid model. Based on the findings of our literature survey above, our default simulations assume a heterogeneous crossover rate along the chromosome. The crossover rate is biased towards the chromosome's peripheries by assuming that a fraction *cbias* of the crossover events occurs in the peripheries and 1-*cbias* in the centre. After the production of the offspring cohort within each habitat, the genotypes of all individuals at all loci are recorded for both populations. Finally, bidirectional migration between the habitats occurs at a rate m, a scheme corresponding to juvenile dispersal, and selection begins anew. Our model ignores de novo mutation as a source of genetic variation because our main interest is in the early stages of population divergence well known to be fuelled by standing genetic variation (e.g., Domingues et al., 2012; Jones et al., 2012; Roesti et al., 2014; Tennessen & Akey, 2011; reviewed in Barrett & Schluter, 2008; Messer & Petrov, 2013).

2.3 | Parameterization of the standard model

The above standard model of adaptive divergence with gene flow and heterogeneous crossover rate was parametrized with biologically plausible default settings and then explored by changing each parameter separately while holding all else constant at default. The default parameter settings were as follows: population size $N = 10{,}000$, number of selected loci $SL = 16$, skew in the initial frequency of the selected alleles $f = 0.5$, generations of evolution $g = 1{,}000$, strength of divergent selection $s = 0.01$, periphery-bias in crossover rate $cbias = 0.9$ and migration rate $m = 0.01$. An overview of the full-parameter ranges explored is provided as Table S3. The selection strength s was generally locus-specific. However, holding the cumulative selection strength across all *SL* constant at $s = 0.16$ by rescaling the per-locus selection coefficient when changing the number of *SL* produced qualitatively similar results (Table S2; Fig. S3). Unless specified otherwise, results in this study were generated with the standard model using default parameter settings (the corresponding R-code is provided as Appendix S2) and represent averages across 100 simulation replicates.

To confirm that our results were not contingent on modelling a single chromosome per individual only, we scaled up our standard model such that each individual comprised three independently segregating chromosomes. We here thus tracked the

evolution of three times as many loci. This multi-chromosome model confirmed that our single-chromosome findings were robust (Table S2; Fig. S4).

2.4 | Modifications of the standard model

In what follows, we describe the rationale and methodological details of three major modifications of the above standard simulation model that were used to explore specific aspects of interest. Unless stated otherwise, these models were run with default parameter settings. An overview of all models is provided as Table S4.

1. "Restricted geneflow" model: This model allowed us to distinguish between the *independent* contributions of hitchhiking and the geneflow barrier to CCBD—a major objective of our study. We modified the standard model such that when populations exchanged migrants, only the alleles at the *SL* of a migrant moved into the recipient population. All alleles at the *NL*, however, remained in the source population, but became associated with the *SL* alleles of a migrant from the foreign population. The migration of *SL* alleles occurred by fully preserving their physical coupling (Fig. S5a provides a visual outline of this model). Hence, the idea of this *restricted geneflow* model was to generate continuous immigration of blocks of unfavourable foreign alleles at the *SL*, like in the standard model, thereby providing the opportunity for sustained hitchhiking at linked *NL* caused by the selective elimination of *SL* alleles within a population. Because migration at the *NL* was suppressed, however, geneflow barriers could not operate as drivers of CCBD at the *NL*. In the *restricted geneflow* model, hitchhiking was therefore the only driving force of CCBD at the *NL*. To infer the unique contribution of the geneflow barrier to CCBD, the magnitude of CCBD (see below for how CCBD was quantified) in the *restricted geneflow* model was subtracted from the magnitude of CCBD in the standard model (where both hitchhiking and geneflow barriers operated jointly in driving CCBD). That the *restricted geneflow* model performed properly was confirmed by the observation that divergence at the *SL* was exactly like in the standard model (compare Fig. S5b to Figure 2a).
2. "Secondary contact" model: To examine the effect of heterogeneity in crossover rate in a situation with migration after a period of adaptive divergence in physical isolation—that is, in allopatry—we implemented a *secondary contact* model. Here, each population initially had the *SL* completely fixed for the locally favourable allele (i.e., complete local adaptation). Evolution then occurred as in the standard model. The standard and the *secondary contact* model provide two extreme situations, with starting allele frequencies at the *SL* either fully balanced in both populations, or completely opposed. The outcome of any short phase of evolution in isolation, leading to incomplete adaptive divergence, is expected to lie between these extremes. To confirm, we ran a further secondary contact simulation in which population divergence first occurred in allopatry for 1,000 generations, followed by 2,000 generations of divergence with gene flow.
3. "Different SL" model: In conjunction with the standard model, this model served to compare patterns of differentiation between population pairs having diverged independently with a completely different set of loci under divergent selection (i.e., nonparallel adaptive evolution). The *different SL* model was identical to the standard model except that all *SL* were shifted systematically by one position along the chromosome relative to the standard model (visualized in Fig. S6). Translated to a real chromosome, this physical shift is substantial. On a 20-Mb chromosome, for instance, the distance between the corresponding *SL* in the two models would amount to >200 kb. The relationship between the standard and the *different SL* model can be interpreted as two flavours of nonparallel evolution: either as population pairs diverging under similar divergent selection but exploiting adaptive variation at completely nonoverlapping sets of genetic loci, or as population pairs diverging along distinct ecological axes altogether.

2.5 | How a heterogeneous crossover rate influences population differentiation along chromosomes

In a first set of analyses, data generated with the above models were used to characterize patterns of population differentiation along the chromosome. Throughout the theoretical part, differentiation was expressed as the absolute difference in the frequency of the allele 1 between two populations, using data from all individuals from both diverging populations. To ensure the robustness of this differentiation metric, we confirmed that quantifying differentiation as F_{ST} and as D_{xy} supported identical conclusions (Table S2; Fig. S7). Population differentiation along the chromosome was visualized either as temporal snapshot at all loci after the default time of 1,000 generations of evolution or by expressing the magnitude of CCBD in a single index—mean differentiation in the centre minus mean differentiation in the peripheries—and plotting this index as time series over 2,000 generations. To exclude potential artificial results arising from averaging across multiple simulation runs and to appreciate variation among single simulation replicates, we further explored data from a small number of randomly selected, individual simulation replicates produced with default parameter settings (Table S2; Fig. S8).

In empirical studies, CCBD is often mirrored by a negative genomewide correlation between genetic differentiation and crossover rate (e.g., Aeschbacher et al., 2017; Burri et al., 2015; Renaut et al., 2013; Roesti et al., 2013; Samuk et al., 2017; Tine et al., 2014). We thus explored this correlation with our standard model, focusing separately on the *SL* and *NL*. For the *SL*, we estimated the crossover rate around each focal *SL* as the mean number of crossovers occurring between the *SL* and its neighbouring *NL* on each side, based on 300,000 crossover breakpoint locations observed in a single simulation replicate. The crossover rate thus estimated was then related to the magnitude of population differentiation at each

SL. For the *NL*, we split the chromosome in nonoverlapping windows of five loci, determined for each window the mean number of crossover breakpoints falling in an interval between adjacent loci, and related this crossover rate estimate to the average population differentiation across the *NL* within each window.

2.6 | Importance of gene flow in driving CCBD

Adaptive divergence is thought to often begin in the face of some gene flow. We were thus particularly interested in investigating how gene flow influences CCBD, and whether the magnitude of CCBD could serve as an indicator of population divergence having occurred in the face of genetic exchange. This specific investigation required comparing CCBD among different geneflow scenarios. To make such comparisons powerful, it was necessary to adjust CCBD for the level of overall differentiation across the entire chromosome, because the presence or absence of gene flow affects CCBD and overall differentiation simultaneously. Several adjustment approaches were explored, but subtracting chromosome-wide mean differentiation from our standard CCBD index (i.e., mean differentiation in the centre minus mean differentiation in the periphery) yielded the most sensitive statistic. Other approaches, however, including division instead of subtraction, yielded qualitatively similar results. Also, we here considered the *NL* only, but including the *SL* led to the same conclusions. "CCBD adjusted for overall differentiation" was thus calculated for four distinct types of pairwise population comparisons differing in the mode of gene flow: (i) parapatric populations diverging with gene flow between each other according to our standard model; (ii) populations diverging in complete isolation, that is, allopatry (standard model with *m* set to 0); (iii) populations from different standard simulation replicates adapted to different habitats (corresponding to allopatric comparisons of populations having diverged independently in the face of gene flow with nonfocal populations); and (iv) populations from different simulations adapted to the same habitat type (Figure 5a illustrates these different scenarios).

2.7 | Chromosome-scale heterogeneity in selection targets as a potential confounding factor

We examined to what degree CCBD could be influenced by broadscale heterogeneity in the distribution of selection targets along chromosomes. The motivation was that if organisms consistently display a higher density of selection targets in chromosome centres relative to the peripheries, this could provide an alternative explanation for CCBD: even with a uniform crossover rate along a chromosome, target-enriched central regions would be affected by selection more strongly than target-poor chromosome peripheries (Aeschbacher et al., 2017). To evaluate this possibility, we used two complementary approaches. First, we investigated for 22 organisms with well-assembled and annotated genomes whether chromosome centres are generally richer in potential targets of selection than chromosome peripheries. Second, we performed a theoretical investigation in which we skewed the distribution of the *SL* to either the centre or the periphery of the chromosome (Table S4), and then tested how this influences CCBD, considering both uniform and heterogeneous crossover rates along the chromosome. Full details of these analyses are given in Methods S1.

2.8 | How a heterogeneous crossover rate impacts differentiation-based genomic analyses

Empirical studies commonly characterize population differentiation by summarizing the genomewide distribution of locus-specific differentiation estimates in a histogram. We analysed how this distribution is affected by heterogeneous (i.e., periphery-biased; *cbias* = 0.9) and uniform (*cbias* = 0.5) crossover rate along the chromosome by plotting the frequency distribution of differentiation values from 30 replicate simulations. For clarity, these analyses focused on the *NL* only, although considering both the *SL* and *NL* supported the same conclusions.

Next, we tested whether a heterogeneous crossover rate can drive differentiation patterns consistent with parallel genomic divergence even when divergent selection is nonparallel. For this, we let multiple pairs of populations diverge with completely different sets of loci under divergent selection using the standard and the *different SL* models (Table S4; Fig. S6). Each model was run in 20 replicate simulations with both uniform and heterogeneous crossover rate. We then asked whether the presence of a heterogeneous crossover rate increases the probability of inferring shared high-differentiation loci (i.e., "outliers") between independent population comparisons relative to a uniform crossover rate. For this, we formed all 400 possible pairwise combinations of population comparisons generated with the two different models, and for each combination, we determined the number of outliers shared between the two comparisons. This analysis considered only those loci neutral in *both* model types ($N = 68$, Fig. S6; including the *SL* produced qualitatively similar results). Outliers were defined as the seven loci exhibiting the strongest differentiation within a population comparison (hence roughly the top 10%). In a complementary analysis using the same data, we examined how heterogeneity in crossover rate influences the chromosome-wide correlation in differentiation values between two population comparisons with different loci under selection. For this, we calculated the correlation among locus-specific differentiation values for all standard versus *different SL* population comparison pairings under both uniform and heterogeneous crossover rate, again considering only the loci neutral in both models. We then visualized the frequency distribution of the resulting correlation coefficients separately for each crossover scheme.

To identify putative loci (and genes) under selection between populations, empirical studies often employ automated scans for high-differentiation outliers that do not require mapping markers to a reference genome. We examined how chromosome-scale heterogeneity in crossover rate can influence such outlier searches by analysing allele frequency data generated with the standard model and

both uniform and heterogeneous crossover rate in BAYESCAN2.1 (Foll & Gaggiotti, 2008). For both crossover schemes, we analysed four replicate population comparisons using the software's default settings. These analyses supported qualitatively similar conclusions irrespective of whether all loci or just the *NL* were considered (we present the former).

2.9 | Consequences of a heterogeneous crossover rate beyond population differentiation

We next expanded our focus by asking whether broadscale heterogeneity in crossover rate along a chromosome affects commonly used population genetic analyses and metrics beyond simple population differentiation. These investigations used data generated with the standard model run with default parameter settings and considered the *NL* only (using all loci consistently produced similar conclusions).

First, we tested for the influence of heterogeneous crossover rate on within-population genetic diversity along the chromosome. We thus averaged the minor allele frequency at each locus within each population generated in the 100 replicate simulations and visualized the resulting measure of genetic diversity along the chromosome.

Second, we characterized within-population haplotype structure along the chromosome, reflecting the antagonistic influence of selection and crossover on the integrity of chromosome segments. For this, we used genotype data from all individuals from a single simulated population to visualize the weighted average of "extended haplotype homozygosity" across both alleles (EHHS; Sabeti et al., 2002) at one focal locus from the chromosome centre (position 50) and two peripheral loci (positions 8 and 92) (patterns of EHHS were highly consistent across simulation replicates, justifying the presentation of data from a single one). To obtain point estimates of haplotype structure along the chromosome, we additionally integrated locus-specific haplotype homozygosity values, yielding the iES statistic (Tang, Thornton, & Stoneking, 2007). iES values from both populations generated in 25 replicate simulations were then averaged and plotted against their chromosome position (ignoring the three most peripheral loci on either side to avoid edge artefacts). Both EHHS and iES were computed with the R package REHH (Gautier & Vitalis, 2012).

Third, we quantified linkage disequilibrium (LD) between loci along the chromosome. LD was quantified as pairwise R^2 using MCLD (Zaykin, Pudovkin, & Weir, 2008) and visualized using the R package LDHEATMAP (Shin, Blay, McNeney, & Graham, 2006). Patterns of LD were highly consistent across individual simulation replicates and populations so that we present the analysis of a single population only.

Fourth, we examined the influence of a heterogeneous crossover rate on phylogenetic population separation. We sampled 50 individuals from each of the two diverging populations and constructed a phylogenetic tree separately for the loci in the chromosome centre and the ones in the chromosome peripheries. The difference between these trees was highly consistent across simulations and phylogenetic algorithms (i.e., neighbour-joining, maximum likelihood, and maximum parsimony as implemented in the R package PHANGORN; Schliep, 2011); hence, a neighbour-joining tree from a single replicate was visualized with FIGTREE version 1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/). To express the phylogenetic separation between populations quantitatively, we use the R package GENEALOGICALSORTING (http://www.genealogicalsorting.org) to calculate the genealogical sorting index (gsi; Cummings, Neel, & Shaw, 2008) separately for central versus peripheral loci. We here used neighbour-joining trees based on 500 random individuals from each population, averaged the two population-specific gsi values and visualized the frequency distribution of these averages across all 100 simulation replicates.

Finally, we investigated how a heterogeneous crossover rate influences the genetic structure among populations. We here used STRUCTURE (Pritchard, Stephens, & Donnelly, 2000) to analyse the clustering of 50 random individuals sampled from each of the two simulated populations, considering loci from the chromosome centre and from the periphery separately. The software was run in five replicates, each with a burnin and MCMC run length of 50,000 and with two predefined clusters ($K = 2$). Data from 10 replicate simulations were analysed in this way, but the outcome was highly consistent so that only a single replicate is presented.

2.10 | Empirical validation of theoretical results

To confirm key patterns emerging from our theoretical investigation empirically, we re-analysed genomewide sequence data from three young (i.e., postglacial, less than 10,000 generations old) population pairs of threespine stickleback diverging independently into selectively different lake and stream habitats within the Boot, Robert's and Joe's drainages on Vancouver Island (BC, Canada) (Berner, Grandchamp, & Hendry, 2009; Roesti et al., 2012a). Within each drainage, the lake and stream population are in direct contact and have diverged under extensive gene flow (Berner et al., 2009; Roesti et al., 2012a). The data were generated by RAD sequencing (Baird et al., 2008) using the Sbf1 restriction enzyme and 27 individuals per population. Details on RAD library preparation, short read processing, consensus genotyping at RAD loci and the extraction of SNPs from these loci is described in detail in Roesti et al. (2012a) and Roesti, Salzburger, and Berner (2012b). For all empirical analyses below, we provide a brief methodological overview only and refer to the Methods S2 for full detail.

A first set of analyses used genomewide SNPs from the Boot lake–stream stickleback population pair. To evaluate the idea that a heterogeneous crossover rate drives CCBD, we tested for an association between the magnitude of periphery-bias in crossover rate and the magnitude of CCBD across stickleback chromosomes, quantifying the former based on a previous characterization of the crossover landscape in threespine stickleback (Roesti et al., 2013) We also tested whether heterogeneity in the distribution of potential selection targets, as opposed to a heterogeneous crossover rate, can explain variation in the magnitude of CCBD among stickleback chromosomes. We next visualized the F_{ST} frequency distribution for all SNPs from chromosome centres versus peripheries to test for a chromosome centre bias in high-differentiation values within the stickleback genome. The same data were also used to a search for putative loci under selection with BAYESCAN (Foll & Gaggiotti, 2008). We

also constructed independent phylogenetic trees for central versus peripheral SNPs and evaluated whether the extent of phylogenetic separation between the two divergent populations was, both in absolute and in relative terms, different between the two trees. Finally, we tested for a difference in within-population linkage disequilibrium between alleles in chromosome centres versus peripheries.

Our simulation-based finding that geneflow barriers are essential in driving substantial CCBD stimulated a second set of empirical analyses, using genomewide SNP data from all three lake–stream stickleback population pairs. We here explored whether CCBD, adjusted for overall (mean) differentiation as described above, is more extreme in population comparisons when divergence occurred in the presence of gene flow (parapatry) than in its absence (allopatry). This analysis, using both F_{ST} and D_{XY} as measures of population differentiation, again considered distinct types of pairwise population comparisons differing in the mode of gene flow, as described above for simulated data (see Figure 5a for a conceptual outline).

Unless specified otherwise, all simulations, analyses and graphing were performed using the *R* language (R Development Core Team, 2015).

3 | RESULTS AND DISCUSSION

3.1 | CCBD and the underlying mechanisms

Heterogeneity in crossover rate across the genome and selection on new mutations is widely recognized to jointly shape patterns of genetic diversity within and between species over long timescales (Charlesworth et al., 1993; Cutter & Payseur, 2013; Gillespie, 2000; Hudson & Kaplan, 1995; Nachman, 2002). By contrast, the consequences of a heterogeneous crossover rate to rapid population differentiation remain relatively poorly explored. Considering a taxonomically widespread trend in the distribution of crossover—a relatively reduced crossover rate in the chromosome centre relative to the peripheries (Table S1)—our simulations of polygenic adaptive divergence from standing genetic variation identify Chromosome Centre-Biased Differentiation (CCBD), a broadscale differentiation pattern commonly observed in nature (Figure 1). Our models further allow us to assess the relative importance of two distinct genetic mechanisms—hitchhiking versus the geneflow barrier—in establishing and maintaining CCBD.

At the selected loci (*SL*), CCBD establishes rapidly as the populations adapt to their local habitats, but stabilizes in magnitude as migration–selection balance is attained (Figure 2a, top panel; the outcome of individual simulation replicates is presented in Fig. S8). At the neutral loci (*NL*), by contrast, CCBD increases even after the migration–selection balance at the *SL* is reached (Figure 2a, bottom panel). At both the *SL* and *NL*, divergence thus generates a negative correlation between the local magnitude of genetic differentiation and the rate of crossover across the chromosome (Figure 3), an association often detected in empirical genome scans of population differentiation (e.g., Aeschbacher et al., 2017; Burri et al., 2015; Renaut et al., 2013; Roesti et al., 2013; Samuk et al., 2017; Tine et al., 2014). The sustained increase in CCBD at the *NL* through time suggests the

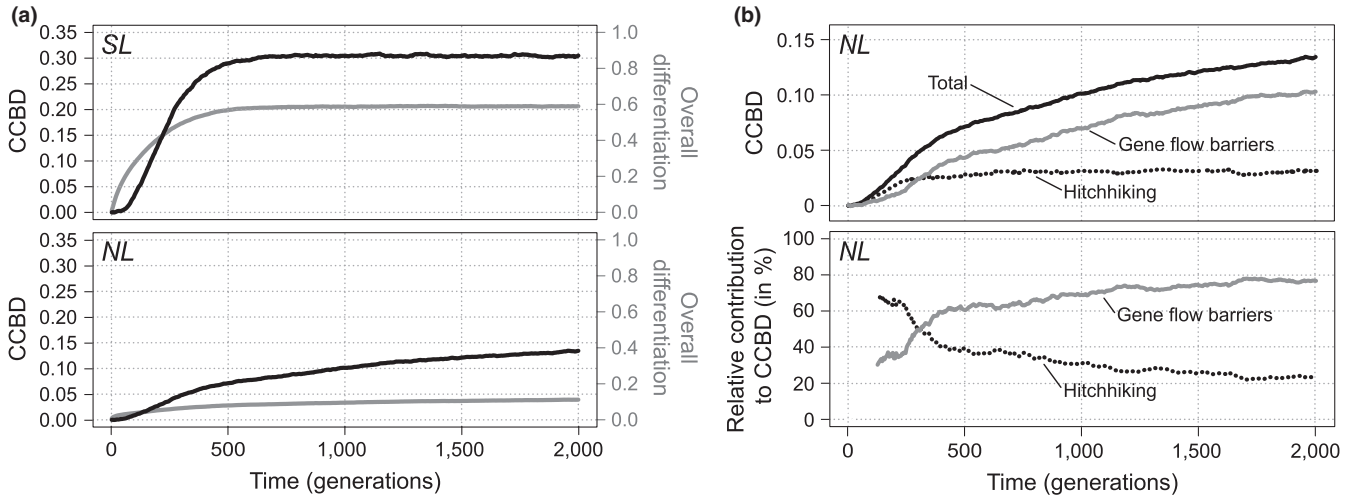

FIGURE 2 Emergence of CCBD during adaptive divergence. (a) Magnitude of CCBD and chromosome-wide mean (overall) differentiation at the selected (*SL*; top panel) and neutral (*NL*; bottom panel) loci in our standard model of divergence with gene flow (run at default: size of each population N = 10,000; 100 total loci, of which *SL* = 16 are under divergent selection with a selection coefficient s = 0.01; crossover bias towards the peripheries *cbias* = 0.9; migration rate m = 0.01). CCBD (black curves) is quantified as mean magnitude of population differentiation in the chromosome centre minus mean differentiation in the two peripheries combined. Overall differentiation (grey curves, right Y-axes) represents the mean differentiation between the populations across either the *SL* or *NL* on the entire chromosome. (b) The *restricted geneflow* model, run at default as in (a), reveals that hitchhiking (black dotted curve) contributes substantially to total CCBD at the *NL* (black solid curve; corresponding to CCBD in the bottom panel of (a), but only in the beginning of population divergence. After this phase, CCBD is largely driven by heterogeneity in the strength of geneflow barriers across the chromosome (grey solid curve; obtained by subtracting the hitchhiking effect seen in the *restricted geneflow* model from overall CCBD observed in the standard model). The top and bottom panels of (b) show the contributions of hitchhiking versus the barrier to gene flow to total CCBD as absolute values and as relative proportions (the latter ignoring data from the first 125 generations, during which the proportions were too noisy)

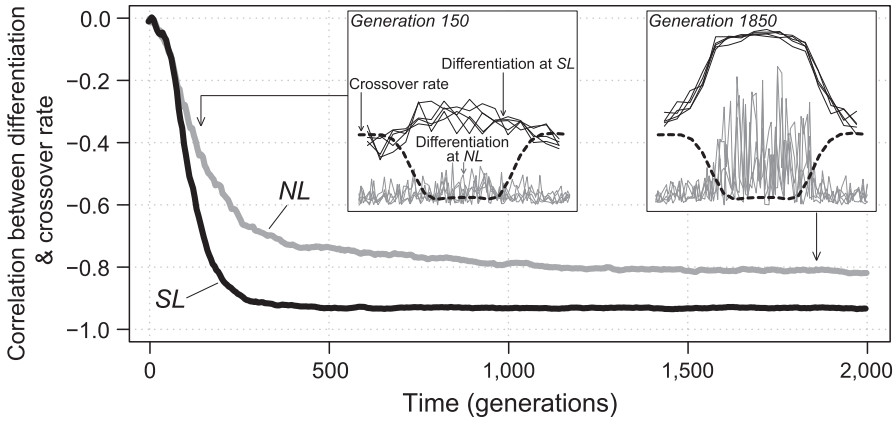

FIGURE 3 CCBD mirrored by the association between genetic differentiation and crossover rate. The emergence of CCBD goes hand in hand with the establishment of a chromosome-wide negative correlation between genetic population differentiation and crossover rate, here shown separately for the *SL* and *NL*. The insert panels visualize the magnitude of these variables at two time points during divergence for five exemplary simulation replicates. For clarity, the crossover rate was smoothed for plotting, although the correlations in the main graphic are based on precise crossover counts. The data were generated using the standard simulation model with default parameter settings

influence of geneflow barriers arising as a consequence of adaptive divergence at the *SL*. The reason is that once the alleles at the *SL* reach migration–selection balance, recurrent allele frequency shifts at these loci due to ongoing maladaptive gene flow are relatively minor, so that hitchhiking is no longer expected to be a major driver of population differentiation at the *NL*. To evaluate this idea directly, we modified our standard simulation model such that gene flow between the populations still occurred at the *SL* (with all alleles at the *SL* migrating together as coupled haplotypes like in the standard model; visualized in Fig. S5), but was completely suppressed at the *NL*. This *restricted geneflow* model, allowing selection to influence differentiation at the *NL* exclusively via hitchhiking, confirmed an independent contribution of hitchhiking to CCBD (Figure 2b). This contribution, however, was marginal beyond the initial phase of divergence characterized by rapid allele frequency shifts within the diversifying populations.

Taken together, this first set of analyses demonstrates that CCBD arises from two distinct mechanisms operating jointly: hitchhiking and the geneflow barrier. Hitchhiking, however, is important primarily before migration–selection balance at the *SL* is established, so that in the course of population divergence, the geneflow barrier rapidly becomes the main driver of CCBD. The ultimate consequence of heterogeneity in the strength of both hitchhiking and the geneflow barrier is variation in effective population size—and hence in the rate of evolution by drift—along a chromosome. Specifically, geneflow barriers make genetic exchange among populations more difficult in the chromosome centre than in the peripheries as soon as adaptive divergence at the *SL* is substantial. The reason is that in a chromosome's centre, where the crossover rate is low, multiple maladaptive alleles from the foreign population and their associated *NL* remain coupled as large cosegregating blocks (i.e., haplotypes). These haplotypes are selected against relatively effectively. In the peripheries, by contrast, frequent crossover tends to break down maladaptive foreign alleles and their associated neutral neighbourhood into shorter haplotypes, which have a relatively small negative selective effect and can thus persist more easily in the foreign habitat. The consequence is crossover rate-mediated heterogeneity in effective gene flow—and thus population size—along the chromosome: in the peripheries, the two diverging populations remain connected more strongly through genetic exchange than in the chromosome centre, thus slowing differentiation by drift in the peripheries relative to the centre. We highlight that the emergence of CCBD through such geneflow barriers between connected populations differs fundamentally from previous explanations for the association between the crossover rate and genetic differentiation based on de novo mutation and hitchhiking (Charlesworth et al., 1993; Gillespie, 2000; Hudson & Kaplan, 1995). We in no way challenge that the latter process can account for elevated population differentiation in low-crossover regions over longer timescales (Burri et al., 2015; Charlesworth, Nordborg, & Charlesworth, 1997; Cruickshank & Hahn, 2014; Noor & Bennett, 2009; Rockman, Skrovanek, & Kruglyak, 2010). However, in young populations diverging from standing variation into selectively different habitats in the face of genetic exchange, CCBD should primarily reflect heterogeneity in the strength of geneflow barriers along chromosomes. Even over long time spans these barriers may contribute to CCBD, provided that some hybridization between diverging populations persists.

3.2 | Biological variables influencing CCBD

Having developed a mechanistic understanding of the emergence and maintenance of CCBD, we next examined how CCBD is influenced by key biological variables. Consistent with the above conclusion that selection-mediated heterogeneity in drift along a chromosome is the major driver of CCBD, we observe that CCBD at the *NL* is most pronounced at an absolute scale when the diverging populations are small (Figure 4a). This is because in very large populations, drift within a genome region is weak irrespective

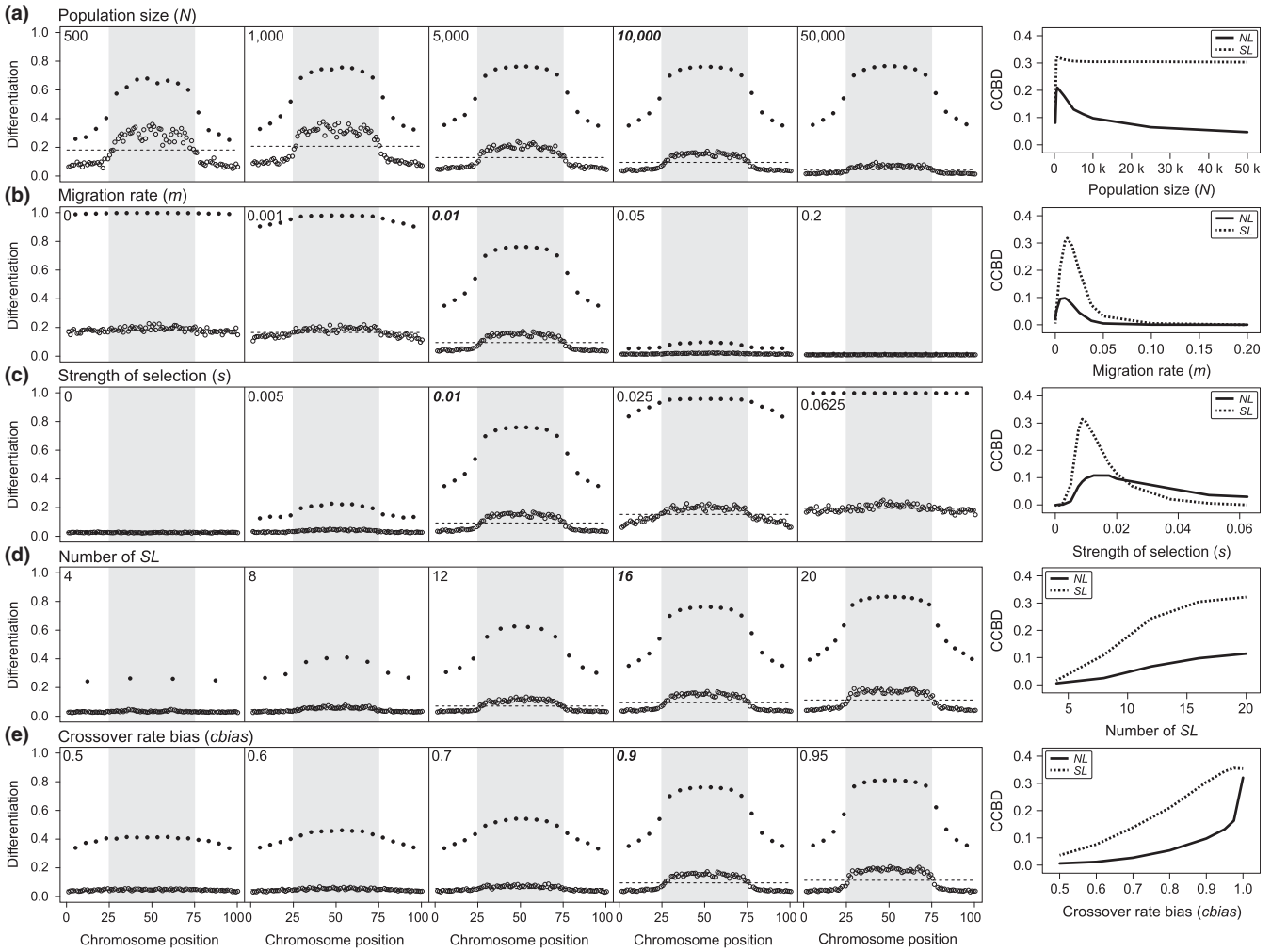

FIGURE 4 Influence of biological variables on CCBD. The variables include (a) the size of each population, (b) the reciprocal migration rate between the populations, (c) the strength of divergent selection on each selected locus (*SL*), (d) the number of *SL* along the chromosome and (e) how strongly the crossover rate is biased towards the chromosome peripheries. Each variable was modified individually while holding all other variables constant at their default value (the focal values are indicated within each subpanel, with default values in bold italics). The magnitude of population differentiation at the *SL* (filled circles) and *NL* (open circles) is shown for five representative parameter values. Mean differentiation across *all* loci on the chromosome is shown as dashed horizontal line. The column on the right summarizes CCBD (quantified as in Figure 2) separately for the *SL* and *NL* across the full parameter range explored (Table S3)

of whether or not the region is influenced by a selective barrier to gene flow. The rates of stochastic population differentiation in the chromosome centre and in the peripheries thus converge as population size increases. At the *SL*, by contrast, CCBD is little affected by population size because differentiation along the chromosome reflects a relatively deterministic balance between migration and selection.

We next addressed the role of gene flow in CCBD. When the migration rate is very low relative to the overall strength of divergent selection, geneflow barriers are inconsequential because the opportunity for gene flow to impede differentiation by drift is consistently low across the entire chromosome (Figure 4b). Hence, only weak CCBD emerges, driven by hitchhiking during the early phase of divergence. This outcome is most evident when divergence occurs in complete isolation—that is, in allopatry (Figure 4b where $m = 0$; Fig. S9a). Interestingly, minimal CCBD driven by hitchhiking can also be observed in allopatric comparisons between ecologically different or similar populations when each of them has diverged separately in the face of gene flow from another, nonfocal population (Fig. S9b,c). At the other extreme, a very high migration rate relative to the overall strength of selection also impedes CCBD because the *SL* cannot reach sufficient differentiation at migration–selection balance to produce an effective barrier to gene flow (Figure 4b). Analogously, CCBD arises only at intermediate strengths of divergent selection on the *SL* (Figure 4c). Both very weak selection (or the absence of any selection, a scenario presented with temporal resolution in Fig. S10a) and strong selection relative to migration preclude the emergence of CCBD because either substantial divergence at the *SL* and the associated geneflow barriers fail to build up, or because gene flow is impeded effectively across the entire chromosome (Barton & Bengtsson, 1986; Feder & Nosil, 2010).

We further find that when the number of *SL* is low, substantial CCBD does not arise (Figure 4d). This is because with a constant selection coefficient across *SL*, the variance in fitness between short versus long haplotypes from the foreign habitat is low when a chromosome holds only a few *SL*, leading to relatively homogeneous gene flow along the entire chromosome. Also, the crossover rate between a few distant *SL* is necessarily high, thus allowing alleles at these loci to segregate almost independently from each other. (Qualitatively similar results were obtained when varying the number of *SL* while holding their cumulative strength of selection constant; see Fig. S3a.) Moreover, the difference in crossover rate between the chromosome centre and the peripheries is a key determinant of CCBD: at the *SL*, the magnitude of CCBD increases roughly linearly with the strength of periphery bias in crossover rate, while at the *NL*, CCBD increases dramatically towards the upper end of the parameter range (Figure 4e). Interestingly, subtle CCBD (particularly at the *SL*) can arise even in the absence of any crossover rate heterogeneity along the chromosome (see also Fig. S4 for similar evidence from multi-chromosome simulations). This "edge effect" arises during the establishment of migration–selection balance (Fig. S10b) and hence reflects a slight hitchhiking advantage of long central chromosome segments encompassing many favourable alleles over short peripheral segments harbouring few favourable alleles only. (This is analogous to the relatively weaker reduction in genetic diversity towards chromosome tips observed within a population experiencing background selection with uniform crossover rate [Nordborg, Charlesworth, & Charlesworth, 1996].) The edge effect is therefore a specific consequence of the linear morphology of chromosomes. Furthermore, over a broad parameter range, we find no material consequences of imbalance in the starting frequency of alleles at the *SL*; that is, making locally favourable alleles initially frequent in one population but rare in the other (Fig. S3b). This is consistent with the conclusion that hitchhiking is of minor importance to CCBD; the pattern is primarily shaped by persistent geneflow barriers established once the alleles at the *SL* are at migration–selection balance.

Finally, we tested whether heterogeneity in the distribution of selection targets across chromosomes could be an alternative explanation for CCBD. Even with a uniform crossover rate, chromosome centres could be affected by selection more strongly if they were enriched in selection targets relative to chromosome peripheries, thus driving CCBD analogously to heterogeneity in crossover rate. Our empirical evaluation of the distribution of potential selection targets in 22 organisms, however, made clear that chromosome-scale heterogeneity in the distribution of potential selection targets is either absent (animals, fungi; Fig. S11a, top), or occurs in the direction opposite to the prediction of a higher density of targets in the chromosome centres than the peripheries (plants; Fig. S11a, bottom). In addition, simulations revealed that even if substantial broadscale heterogeneity in selection target density in the predicted direction existed, this would drive only weak heterogeneity in population differentiation (Fig. S11b). Combined, these analyses rule out a heterogeneous distribution of selection targets along chromosomes as a general cause for CCBD.

In summary, our results demonstrate that substantial CCBD represents a characteristic outcome of polygenic divergent selection in the face of gene flow and heterogeneous crossover rate along chromosomes. Combined, these main factors cause a stronger barrier to gene flow in the chromosome centre than in the peripheries, whereas the absence of any one of these factors precludes the emergence of strong CCBD. Subtle CCBD might still emerge in the absence of gene flow, or weaker yet when divergence with gene flow occurs in the absence of heterogeneity in crossover rate, driven by hitchhiking alone. Clearly, however, the extent of CCBD relative to overall differentiation remains very weak in *any* type of allopatric population comparison (compare Fig. S9 with Figure 2a). After accounting for overall population differentiation, substantial CCBD can thus provide a qualitative indication of adaptive divergence with gene flow (Figures 5b, S12). The observation of substantial CCBD alone, however, does not allow inferring that population divergence was *initiated* in the face of gene flow: in our secondary contact model with populations diverging with gene flow after an initial phase of evolution in allopatry, CCBD at both the *SL* and *NL* converges to the pattern seen under primary divergence as soon as migration–selection balance at the *SL* is established (Fig. S10c,d; see also Cruickshank & Hahn, 2014). The secondary contact scenario further suggests that CCBD could not only be driven by loci under divergent *ecological* selection: environment-independent genetic incompatibilities accumulated in isolation could act as barrier loci too, provided that they still permit some hybridization following secondary contact (see also Bierne, Welch, Loire, Bonhomme, & David, 2011).

3.3 | Analytical consequences of chromosome-scale heterogeneity in crossover rate

Motivated by the recognition that broadscale heterogeneity in crossover rate along chromosomes can profoundly influence patterns of differentiation between populations, we next assessed whether and how this may affect current differentiation-based concepts and inferential procedures in population genomics. We first focused on the distribution of locus-specific differentiation values (e.g., F_{ST} values) obtained by a genomic comparison of two populations under divergent selection. The shape of this distribution has been proposed to indicate whether or not populations have diverged in the face of gene flow (Feder, Egan, & Nosil, 2012): with gene flow, the distribution is expected to be strongly right-tailed because substantial differentiation is possible only at a relatively small number of loci under divergent selection, while most loci are selectively neutral and thus homogenized between the populations. Conversely, in the absence of gene flow, divergence at all loci in the genome is unconstrained, so that the distribution of differentiation values is predicted to be less skewed. Because the influence of other biological variables on the shape of this distribution has not been explored, we compared the distribution of locus-specific allele frequency shifts in our simulations with and without heterogeneity in crossover rate along the

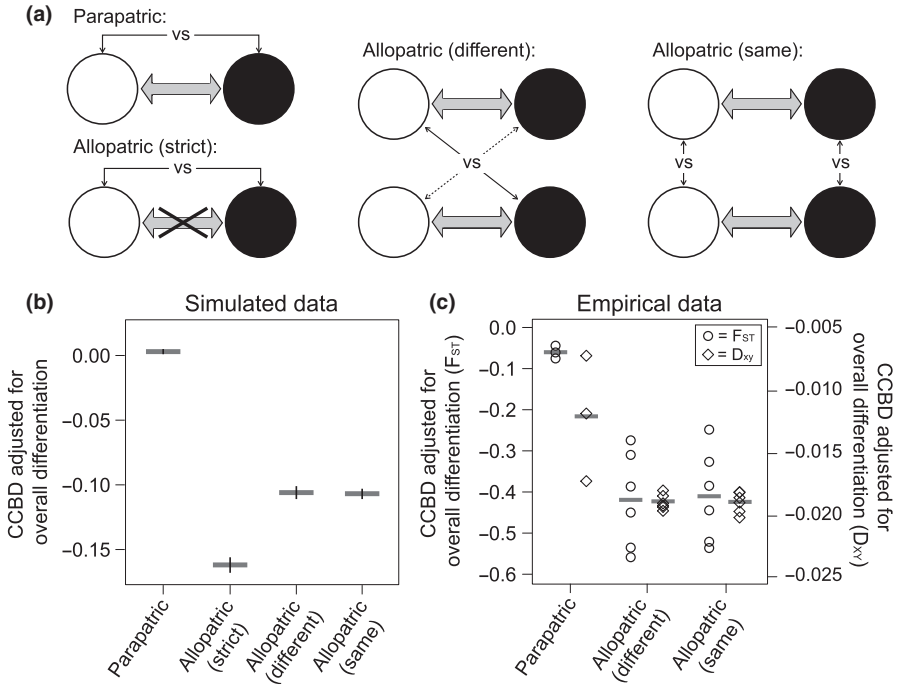

FIGURE 5 Magnitude of CCBD adjusted for overall differentiation in different geographic and ecological population comparisons. (a) Schematic of the different population comparisons considered, with white and black circles representing populations in selectively distinct habitats. Large grey arrows indicate migration (gene flow) between two populations, and thin black arrows connect the "focal" populations being compared genetically. The scenarios include the following: (i) Standard divergence with gene flow ("Parapatric"). (ii) Divergence in complete isolation ["Allopatric (strict)"], that is, none of the focal populations exchanges migrants with any other population. (iii) The focal populations diverge by exchanging migrants with a separate, nonfocal population occupying a selectively different habitat ["Allopatric, (different)"]. These latter two comparisons are considered nonparallel because they involve populations adapted to different habitat types. (iv) The focal allopatric populations evolve by exchanging migrants with a nonfocal population, but occupy selectively similar habitats ["Allopatric (same)"]; hence, these comparisons are parallel. Adjusted by the magnitude of overall differentiation between populations, CCBD is more pronounced in comparisons of populations having diverged with gene flow between each other than in allopatry: (b) For simulated data, mean-adjusted CCBD is expressed as the difference in population differentiation between the centre and the periphery minus overall differentiation across the entire chromosome (shown are mean values and their parametric 95% CIs over 100 replicate simulations). (c) Empirical results are based on genomewide SNPs from three independent lake–stream stickleback population pairs (Vancouver Island, Canada). Mean-adjusted CCBD was here expressed analogously to (b) as the centre-periphery difference in differentiation—quantified by both F_{ST} and D_{xy}—minus chromosome-wide mean differentiation, averaged across all 20 autosomes for each population comparison. The comparisons include the three pairs of geographically adjoining lake–stream populations within each watershed ("Parapatric"), and all possible pairwise combinations between populations from different watersheds occupying either a different ["Allopatric (different)"] or the same ["Allopatric (same)"] habitat. Shown are the raw data points and their means as grey bars

chromosome. This revealed that for a given migration rate, a heterogeneous crossover rate causes a strongly right-tailed distribution—and elevated overall differentiation—because the associated CCBD implies that relatively extreme differentiation values are possible in the chromosome centre (Figure 6a). It follows that beside a heterogeneous crossover rate, any variable demonstrated above to affect the magnitude of CCBD (Figure 4)—including gene flow—has the potential to influence the shape of the distribution of differentiation values. Because robust information about the number of loci under divergent selection, the strength of selection, or population sizes is usually lacking in empirical systems, drawing conclusions about levels of gene flow or stages of speciation based on F_{ST} distributions (Feder et al., 2012; Seehausen et al., 2014) appears problematic, unless well-replicated or controlled comparative study designs are employed.

Our second investigation of analytical challenges associated with a heterogeneous crossover rate concerned the inference of parallel (or "convergent"; Arendt & Reznick, 2008) evolution based on the comparison of replicate genome scans from multiple independent population pairs believed to have diverged along similar ecological axes. In this case, a positive correlation in the magnitude of genomewide differentiation values and in particular the detection of *shared* high-differentiation outliers between the pairs is commonly taken as evidence of parallel adaptive evolution (e.g., Fraser, Kunstner, Reznick, Dreyer, & Weigel, 2015; Gagnaire, Pavey, Normandeau, & Bernatchez, 2013; Westram et al., 2014). The underlying assumption is that genomic regions displaying exceptionally high habitat-related differentiation in independent population comparisons reflect repeated, deterministic (parallel) selection on the same loci. To examine whether this reasoning is valid in the face of heterogeneity in crossover rate,

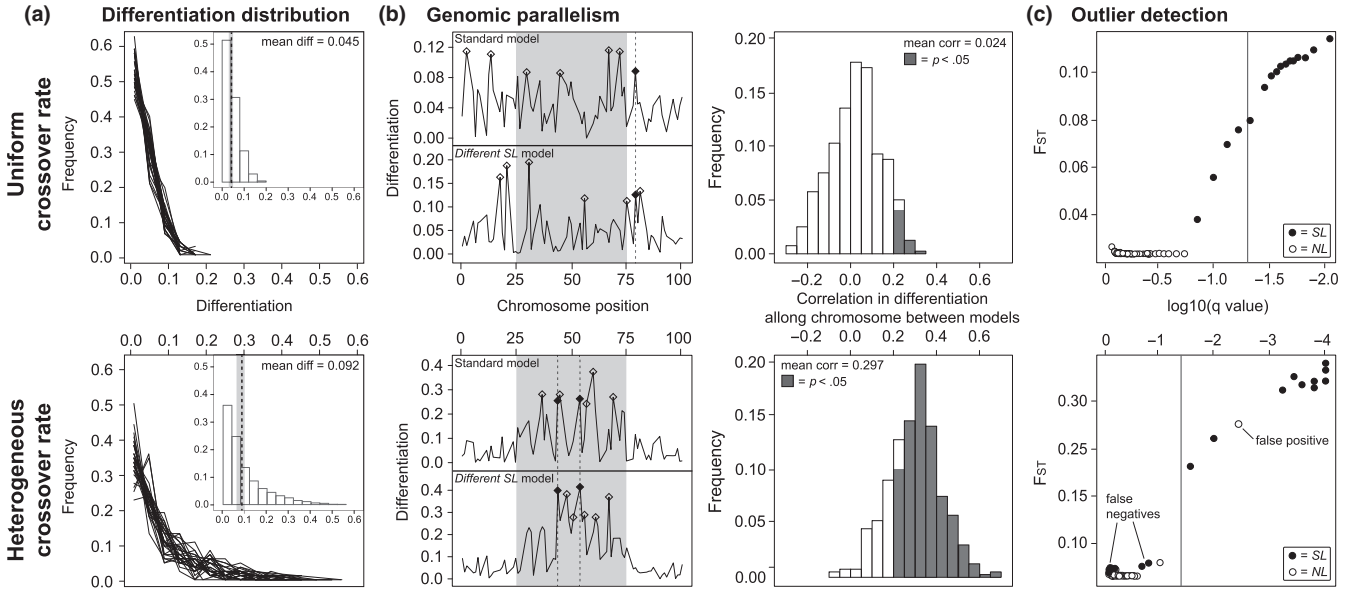

FIGURE 6 Heterogeneity in crossover rate affects differentiation-based genomic analyses and interpretations. For all panels in this figure, the upper row shows data from populations diverging with a uniform crossover rate along the chromosome (standard model with *cbias* = 0.5), while the data presented in the lower row were generated with a heterogeneous (periphery-biased) crossover rate (*cbias* = 0.9), thus leading to CCBD. (a) CCBD results in a relatively right-tailed distribution of genetic differentiation values across selectively neutral loci in pairwise population comparisons (the black curves show data from 30 individual simulations). CCBD further causes average population differentiation across all *NL* to be relatively elevated, as seen in the insert histograms displaying data pooled across 100 simulations. The grey vertical bars in the insert panels show the range of chromosome-wide mean differentiation, and the black dashed lines indicate the grand mean. (b) Association in locus-specific differentiation between population pairs evolved under the standard versus the *different SL* model. These two models have completely nonoverlapping sets of loci under selection, hence population pairs evolve along distinct selective axes (Fig. S6). The left column of panel (b) shows differentiation profiles along the chromosome for a single representative population pair for each model type (considering only the 68 loci neutral in both models). Diamonds indicate high-differentiation outliers within each population comparison, and filled diamonds connected by dashed lines indicate outliers shared between the two population pairs. The right column of panel (b) shows the distribution of correlation coefficients across 400 pairwise combinations of population pairs evolved under nonparallel selection (i.e., each coefficient quantifies the correlation between differentiation values from a single run with the standard versus and the *different SL* model). Positive correlation coefficients exhibiting $p < .05$ are shaded in grey, and the mean across all 400 coefficients is given within each panel. (c) Genome scans for selection outliers with and without CCBD. Using data from a single representative simulation, the magnitude of differentiation (F_{ST}) at the *SL* and *NL* is plotted against the strength of evidence of selection (*X*-axis; note that the scale of this axis differs between the upper and lower panel). The vertical line denotes the significance threshold at a false discovery rate of 0.05

we ran simulations in which two population pairs evolved independently with completely nonoverlapping sets of loci under divergent selection (i.e., the standard versus the *different SL* model). We thus explicitly simulated nonparallel adaptive evolution. These analyses revealed that the probability of a *NL* to emerge as a shared outlier between both population comparisons was roughly twice as high when the crossover rate was heterogeneous, as opposed to uniform, along the chromosome (mean number of shared outliers: 1.51 versus 0.80, $p = .0001$). Also, with a heterogeneous crossover rate, outliers occurred mainly in the low-crossover chromosome centre (Figure 6b, left column). The obvious reason is that loci in chromosome centres tended to reach elevated divergence (i.e., CCBD) in both population comparisons irrespective of the precise targets of selection. Similarly, with a uniform crossover rate along the chromosome, the correlation in locus-specific differentiation strength between the population pairs evolving under nonparallel divergent selection peaked at zero, and associated *p*-values below .05 occurred at the type 1 error frequency expected in the absence of any relationship (0.058) (Figure 6b, right column, top). With a heterogeneous crossover rate, however, the independent population pairs generally showed a positive correlation in differentiation values, with most coefficients exhibiting $p < .05$ (Figure 6b, right column, bottom). That is, even when the population pairs had completely different loci targeted by selection along the chromosome, CCBD produced a positive correlation in the magnitude of genomewide differentiation values between the pairs. Interpreting such a positive correlation—or the overlap in differentiation outliers in independent genome scans—as evidence of parallel adaptive evolution or a shared "genomic architecture of adaptation" can thus be misleading: inferring determinism in selection across the genome requires the characterization of the actual targets of selection with high marker resolution and accounting for broadscale heterogeneity in baseline differentiation along chromosomes (Roesti et al., 2012a).

A common endeavour with genomic data from populations under divergent selection is to identify how many and which loci (and associated genes) are involved in adaptive divergence. In a next step, we thus asked whether a heterogeneous crossover rate can bias this type of

analysis when using commonly used outlier detection software applicable to organisms lacking a well-assembled reference genome. When using such software, it is generally assumed that loci under selection can be isolated statistically from a background of selectively neutral differentiation across the genome. We thus screened our simulated data sets for selection outlier loci using BAYESCAN (Foll & Gaggiotti, 2008). An observation qualitatively consistent across data sets was that with a uniform crossover rate along the chromosome, those loci producing the strongest evidence of divergent selection were, as expected, the loci truly under selection (*SL*) (Figure 6c, top). With a heterogeneous crossover rate, however, *SL* located in the high-crossover chromosome peripheries were systematically missed as selection outliers (false negatives), while *NL* located in the chromosome centre were sometimes identified as targets of selection (false positives; Figure 6c, bottom). Moreover, with a heterogeneous crossover rate, all outliers reached relatively stronger evidence of selection (compare the scales of the X-axes between the top and bottom panel in Figure 6c). This confirms that in the presence of CCBD, as driven by broadscale heterogeneity in crossover rate along chromosomes, anonymous (i.e., reference genome-disabled) outlier detection methods are generally unreliable: the distinction between selected and neutral loci is blurred because the magnitude of differentiation around a locus depends on the crossover rate in the locus' neighbourhood and hence its position within the genome.

3.4 | Consequences of heterogeneity in crossover rate beyond population differentiation

Having studied the consequences of broadscale heterogeneity in crossover rate along chromosomes to population differentiation, we expanded our focus on its implications in other analytical contexts. First, we confirmed that divergence with gene flow in the presence of a heterogeneous crossover rate reduces—by modifying effective population size and hence drift along a chromosome—genetic diversity in a chromosome's centre relative to its peripheries *within* a population (Figure 7a). The chromosome centre also exhibited a lower rate of haplotype decay than the peripheries (Figure 7b), and correspondingly, elevated linkage disequilibrium (LD) within populations (Figure 7c). Analysing time series revealed that heterogeneity in LD along the chromosome emerges rapidly while an adapting population experiences strong allele frequency shifts at the *SL*, but is little affected by the migration context (Fig. S13). The differences in haplotype structure and LD between central and peripheral chromosome regions in our models are thus driven mainly by hitchhiking. This again highlights that hitchhiking and the geneflow barrier go hand in hand; the relative importance of each mechanism is dependent on the stage of divergence (i.e., ongoing allele frequency changes versus migration–selection balance at the *SL*) and the genetic statistic of interest. These analyses also illustrate why genomic approaches to detecting selection based on extended haplotype structure and elevated LD are most powerful in populations displaying recent or ongoing allele frequency shifts (Oleksyk, Smith, & O'Brien, 2010; Sabeti et al., 2006).

Finally, we examined the impact of a heterogeneous crossover rate on common measures of population structure. For this, we constructed phylogenetic trees separately for the *NL* from the chromosome centre and from the periphery and found deeper genealogical separation in trees inferred from central loci (Figure 7d). This is analogous to empirical observations of higher phylogenetic resolution in regions of low crossover rate in relatively ancient clades (Hobolth, Dutheil, Hawks, Schierup, & Mailund, 2011; Pease & Hahn, 2013; Prüfer et al., 2012), a pattern previously ascribed to sustained linked selection caused by new mutations. Similarly, a standard algorithm for identifying distinct genetic clusters among individuals (STRUCTURE, Pritchard et al., 2000) detected stronger population structure based on the loci in the chromosome centre than those in the peripheries (Figure 7e). Together, these analyses highlight the profound impact of the interaction between selection and heterogeneity in crossover rate. Conclusions from many population genetic and genomic analyses profit from considering the broadscale crossover landscape in which the underlying markers are embedded.

3.5 | Empirical validation of theoretical findings in lake–stream stickleback fish

To illustrate the empirical relevance of our theoretical results, we examined whether exemplary patterns emerging from our simulations can be recovered in an organismal system showing recent population divergence in the face of gene flow: parapatric populations of threespine stickleback fish diverging into adjoining, ecologically different lake and stream habitats under polygenic divergent selection (Berner et al., 2009; Deagle et al., 2012; Feulner et al., 2015; Marques et al., 2016; Ravinet, Prodohl, & Harrod, 2013; Roesti et al., 2012a, 2015; Stuart et al., 2017). This organism displays a consistently reduced crossover rate in chromosome centres relative to the peripheries (Roesti et al., 2013). We inspected genomewide differentiation data in such a lake–stream population pair and found that using chromosomes as data points, the magnitude of CCBD is strongly related to the magnitude of heterogeneity in crossover rate (Figure 8a). By contrast, we found no systematic bias in the distribution of potential selection targets across stickleback chromosomes and accordingly no association between this distribution and CCBD (Fig. S14). Hence, broadscale heterogeneity in the density of genes or in the length of transcripts does not seem to be responsible for CCBD.

We next compared the frequency distributions of F_{ST} values between chromosome centres and peripheries. As expected when CCBD is present, SNPs with strong differentiation were mainly located in chromosome centres (Figure 8b) (see also Figure 1a and Roesti et al., 2012a). For instance, among the 73 SNPs displaying $F_{ST} > 0.9$, 67 (92%) were from chromosome centres (binomial test for homogeneous distribution, $p < .001$). The same was true for 73 (61%) of the 119 SNPs proposed as loci under selection by BAYESCAN ($p = .008$). When ignoring heterogeneity in crossover rate and thus in baseline differentiation along chromosomes, these findings might stimulate the misleading conclusion that loci in chromosome centres tend to have larger selective effect sizes (i.e.,

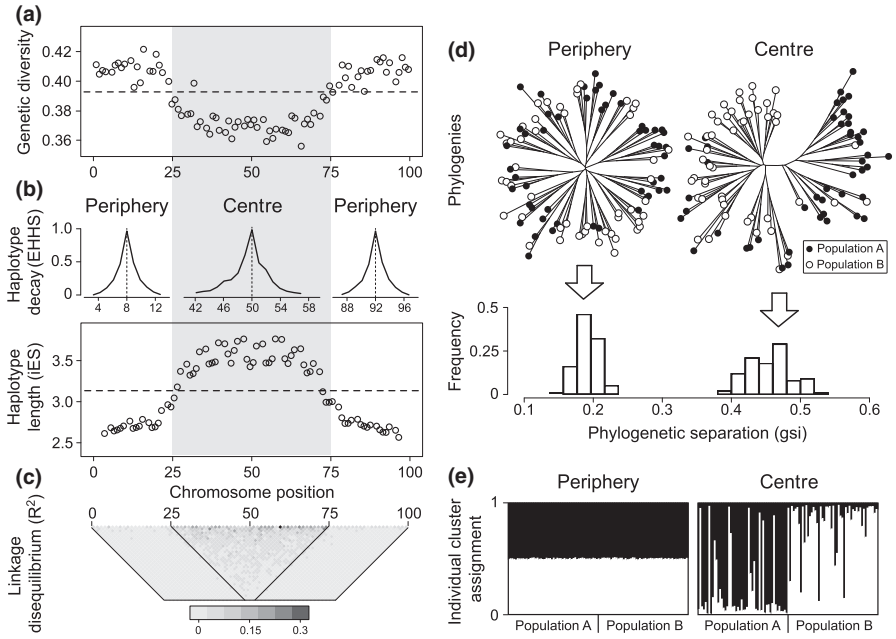

FIGURE 7 Consequences of heterogeneity in crossover rate beyond CCBD. (a) Divergence with gene flow and a relatively reduced crossover rate in the chromosome centre leads to lower genetic diversity in the chromosome centre (shaded in grey) than in the peripheries within a population. The dots are means over both populations and 100 simulation replicates, and the dashed horizontal line indicates the mean across the entire chromosome. (b) The rate of haplotype decay within a population is lower around loci in the chromosome centre than around peripheral loci, as shown for two exemplary loci from the chromosome periphery (positions 8 and 92) and a locus from the chromosome centre (position 50). The lower panel displays a locus-specific haplotype length metric along the entire chromosome (dots are means across both populations from 25 replicates). (c) As a consequence of longer haplotype tracts, loci in the chromosome centre also exhibit extended linkage disequilibrium within a population. (d) With a heterogeneous crossover rate, loci from the chromosome centre exhibit stronger phylogenetic separation between populations than loci from the periphery. The upper panel shows neighbour-joining trees based on subsamples from both populations generated in a single simulation. The lower panel describes the strength of phylogenetic separation expressed by the genealogical sorting index (gsi) based on central versus peripheral loci across 100 replicate simulations. (e) Diverging populations also appear genetically more distinct when applying the STRUCTURE clustering algorithm to central versus peripheral loci. Individual assignment probabilities to each of two genetic clusters (black and white) are visualized by vertical bars. Data are from a single simulation replicate. With default parameters, loci from the centre indicate two distinct populations whereas peripheral loci indicate a single homogeneous cluster. (a–e) are all based on the standard model and consider the *NL* only

selection coefficients) than peripheral ones (see also Roesti et al., 2012a) or that the physical arrangement of loci important to adaptation has been shaped directly through selection (i.e., clustered genomic architecture). Both may perhaps be true in some natural systems, but analytical approaches confounded by heterogeneity in crossover rate cannot provide convincing evidence. An analogous caveat has been raised for the mapping of quantitative trait loci (QTL mapping; Noor, Cunningham, & Larkin, 2001; Noor & Bennett, 2009; Rockman, 2012).

Matching the predictions from our simulations (Figure 7d), a phylogenetic tree based on SNPs from chromosome centres revealed deeper genealogical sorting than a tree based on peripheral polymorphisms (Figure 8c), a pattern also detected in taxa exhibiting much more ancient divergence (Hobolth et al., 2011). Within populations, SNPs from the chromosome centres also exhibited stronger LD (Figure 8d) and reduced genetic diversity (see Fig. 4b in Roesti et al., 2013) in accordance with our simulation results (Figure 7a,c).

Finally, we included SNP data from two additional, evolutionarily independent lake–stream stickleback population pairs to evaluate our theoretical expectation that strong CCBD provides a qualitative indicator of divergence with gene flow (Figures 5b and S12). For this, we contrasted the magnitude of CCBD in the three total parapatric lake–stream population comparisons to the magnitude of CCBD in all pairwise allopatric comparisons. We here separately considered comparisons of allopatric populations residing in the same habitat type (i.e., comparisons between allopatric lake populations or between allopatric stream populations) and in different habitats (between allopatric lake and stream populations) and adjusted CCBD by the overall (mean) differentiation following our theoretical investigation (see Figures 5b and S12). The resulting metric was dramatically higher in parapatric than in allopatric comparisons (Figure 5c; see also Samuk et al., 2017), as predicted by our simulations (Figure 5b). This pattern held irrespective of the type of habitat contrast used for the allopatric comparisons, and irrespective of the divergence metric considered (Figure 5a,c). When multiple population comparisons are available, the strength of CCBD relative to overall genomic differentiation can thus indicate whether adaptive divergence was influenced by gene flow.

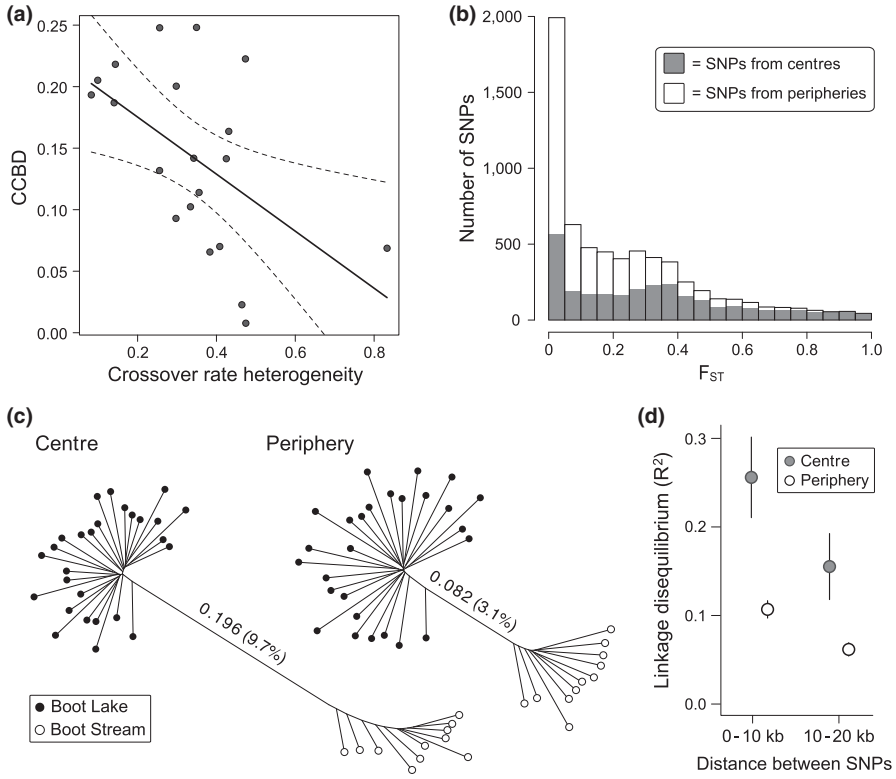

FIGURE 8 Consequences of chromosome-scale heterogeneity in crossover rate in a natural system. (a) In an exemplary lake and stream population pair of stickleback fish diverging in parapatry, the magnitude of CCBD is related to the magnitude of heterogeneity in crossover rate when using chromosomes as data points (Spearman's rank correlation = −0.635, p = .003). Crossover rate heterogeneity was calculated by dividing mean crossover rate in a chromosome's centre by the mean crossover rate in its peripheries; hence, lower values along the X-axis represent stronger chromosome periphery-biased crossover rate. Dots indicate the 20 autosomes, and dashed lines depict the 95% confidence limits around the linear regression. (b) The F_{ST} frequency distribution reveals more strongly differentiated SNPs in the chromosome centres (mean F_{ST} = 0.32; number of SNPs = 2,833) than in the peripheries (mean F_{ST} = 0.16; number of SNPs = 3,673). (c) A neighbour-joining tree based on SNPs from chromosome centres indicates a deeper phylogenetic separation between the two populations than a tree based on SNPs from the peripheries of chromosomes. The numbers associated with each tree specify the length of the major branch separating the two populations, both absolute and relative (in %) to the entire tree length. (d) SNPs from chromosome centres show greater linkage disequilibrium than the peripheral ones within populations. This holds both for relatively contiguous (0–10 kb apart) and more distant (10–20 kb) SNPs. Error bars are parametric 95% CIs across mean values from all autosomes

4 | CONCLUSION

In summary, our simulations have shown that heterogeneity in crossover rate along chromosomes, in interaction with polygenic divergent selection and migration, generates heterogeneity in the strength of hitchhiking and particularly in the efficacy of geneflow barriers along a chromosome. A major outcome is "Chromosome Centre-Biased Differentiation" (CCBD) between populations, a broadscale genomic signature highlighting the constraining effect of crossover during adaptive divergence. CCBD can emerge rapidly through the selective sorting of standing genetic variation—new mutations are not required. Although empirical support for CCBD from genomewide scans of population differentiation is yet limited to relatively few organismal systems with well-developed genomic resources (Figure 1), CCBD may prove to be a biologically widespread pattern of genome divergence among populations under ecological diversification. The reason is that a reduced crossover rate in chromosome centres is taxonomically widespread (Table S1) and that adaptive divergence commonly has a highly polygenic basis and occurs in the face of some gene flow (e.g., Bergland, Behrman, O'Brien, Schmidt, & Petrov, 2014; Carneiro et al., 2014; Elgvin et al., 2017; Foote et al., 2016; Fournier-Level et al., 2011; Lamichhaney et al., 2015; Lawniczak et al., 2010; Martin et al., 2013; Puzey et al., 2017; Roesti et al., 2015; Soria-Carrasco et al., 2014; Tine et al., 2014). The key ingredients for CCBD are thus generally given in nature, and testing for this chromosome-scale differentiation pattern should become a routine step in genomic analyses. When CCBD is present, we should be cautious when inferring population structure and phylogenetic relationships, when attempting to estimate the strength of divergent selection on specific loci, when interpreting genomic clustering of signatures of selection, and when using replicate genome scans to infer deterministic natural selection from shared patterns of differentiation or diversity.

ACKNOWLEDGEMENTS

We are grateful to the following people for sharing empirical differentiation data: Pierre-Alexandre Gagnaire, François Bonhomme (*Dicentrarchus*); Joshua Puzey, John Kelly (*Mimulus*); Sébastien Renaut, Loren Rieseberg (*Helianthus*); Wei Zhang, Marcus Kronforst (*Papilio*). Xavier Thibert-Plante made a valuable suggestion on the fitness scheme in the model. We also thank Jeff Feder and nine anonymous reviewers for suggestions to improve this study and Lotti Roesti for visualizing a genuine Berner Roesti (Fig. S15). Both authors received financial support from the Swiss National Science Foundation (SNF) and the University of Basel. MR received additional financial support through a Janggen-Pöhn fellowship.

CONFLICT OF INTEREST

The authors declare no competing financial interests.

DATA ACCESSIBILITY

Tables S1-S4, Figures S1–S15 and Methods S1 & S2 can be found online as Appendix 1. The R-code for the standard simulation model is provided as Appendix S2. Stickleback RAD sequences can be obtained from NCBI (SRA) under the Accession No SRP007695.

AUTHOR CONTRIBUTIONS

D.B. designed the study, conceived and implemented the simulation tools, contributed to data analysis, interpreted results, wrote the first manuscript version and the final manuscript. M.R. initiated and designed the study, carried out the simulations, implemented and conducted the analyses of simulated and empirical data, visualized and interpreted results and wrote the final manuscript.

ORCID

Marius Roesti http://orcid.org/0000-0002-7408-4804

REFERENCES

Aeschbacher, S., Selby, J. P., Willis, J. H., & Coop, G. (2017). Population-genomic inference of the strength and timing of selection against gene flow. *Proceedings of the National Academy of Sciences of the United States of America*, *114*, 7061–7066.

Anderson, L. K., Doyle, G. G., Brigham, B., Carter, J., Hooker, K. D., Lai, A., … Stack, S. M. (2003). High-resolution crossover maps for each bivalent of *Zea mays* using recombination nodules. *Genetics*, *165*, 849–865.

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

Backström, N., Forstmeier, W., Schielzeth, H., Mellenius, H., Nam, K., Bolund, E., … Ellegren, H. (2010). The recombination landscape of the zebra finch *Taeniopygia guttata* genome. *Genome Research*, *20*, 485–495.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., … Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, *3*, e3376.

Barrett, R. D. H., & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, *23*, 38–44.

Barton, N. H. (1979). Gene flow past a cline. *Heredity*, *43*, 333–339.

Barton, N. H. (1983). Multilocus clines. *Evolution*, *37*, 454–471.

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

Begun, D. J., & Aquadro, C. F. (1992). Levels of naturally occurring DNA polymorphism correlate with recombination rates in *Drosophila melanogaster*. *Nature*, *356*, 519–520.

Bekele, W. A., Wieckhorst, S., Friedt, W., & Snowdon, R. J. (2013). High-throughput genomics in sorghum: From whole-genome resequencing to a SNP screening array. *Plant Biotechnology Journal*, *11*, 1112–1125.

Bergland, A. O., Behrman, E. L., O'Brien, K. R., Schmidt, P. S., & Petrov, D. A. (2014). Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in Drosophila. *PLoS Genetics*, *10*, e1004775.

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

Berner, D., & Thibert-Plante, X. (2015). How mechanisms of habitat preference evolve and promote divergence with gene flow. *Journal of Evolutionary Biology*, *28*, 1641–1655.

Bhakta, M. S., Jones, V. A., & Vallejos, C. E. (2015). Punctuated distribution of recombination hotspots and demarcation of pericentromeric regions in *Phaseolus vulgaris* L. *PLoS ONE*, *10*, e0116822.

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.

Borodin, P. M., Karamysheva, T. V., Belonogova, N. M., Torgasheva, A. A., Rubtsov, N. B., & Searle, J. B. (2008). Recombination map of the common shrew, *Sorex araneus* (eulipotyphla, mammalia). *Genetics*, *178*, 621–632.

Burri, R., Nater, A., Kawakami, T., Mugal, C. F., Olason, P. I., Smeds, L., … Ellegren, H. (2015). Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of Ficedula flycatchers. *Genome Research*, *25*, 1656–1665.

Carneiro, M., Albert, F. W., Afonso, S., Pereira, R. J., Burbano, H., Campos, R., … Ferrand, N. (2014). The genomic architecture of population divergence between subspecies of the European rabbit. *PLoS Genetics*, *10*, e1003519.

Charlesworth, B., Morgan, M. T., & Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics*, *134*, 1289–1303.

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.

Christe, C., Stölting, K. N., Paris, M., Fraisse, C., Bierne, N., & Lexer, C. (2017). Adaptive evolution and segregating load contribute to the genomic landscape of divergence in two tree species connected by episodic gene flow. *Molecular Ecology*, *26*, 59–76.

Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, *23*, 3133–3157.

Cummings, M. P., Neel, M. C., & Shaw, K. L. (2008). A genealogical approach to quantifying lineage divergence. *Evolution*, *62*, 2411–2422.

Cutter, A. D., & Payseur, B. A. (2013). Genomic signatures of selection at linked sites: Unifying the disparity among species. *Nature Reviews Genetics*, *14*, 262–274.

Deagle, B. E., Jones, F. C., Chan, Y. F., Absher, D. M., Kingsley, D. M., & Reimchen, T. E. (2012). Population genomics of parallel phenotypic

evolution in stickleback across stream-lake ecological transitions. *Proceedings of the Royal Society B*, *279*, 1277–1286.

Domingues, V. S., Poh, Y. P., Peterson, B. K., Pennings, P. S., Jensen, J. D., & Hoekstra, H. E. (2012). Evidence of adaptation from ancestral variation in young populations of beach mice. *Evolution*, *66*, 3209–3223.

Elgvin, T. O., Trier, C. N., Tørresen, O. K., Hagen, I. J., Lien, S., Nederbragt, A. J., … Sætre, G.-P. (2017). The genomic mosaicism of hybrid speciation. *Science Advances*, *3*, e1602996.

Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation-with-gene-flow. *Trends in Genetics*, *28*, 342–350.

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

Feulner, P. G., Chain, F. J., Panchal, M., Huang, Y., Eizaguirre, C., Kalbe, M., … Milinski, M. (2015). Genomics of divergence along a continuum of parapatric population differentiation. *PLoS Genetics*, *11*, e1004966.

Fisher, R. A. (1930). In *The genetical theory of natural selection*. Oxford, UK: Oxford University Press.

Fledel-Alon, A., Wilson, D. J., Broman, K. W., Wen, X., Ober, C., Coop, G., & Przeworski, M. (2009). Broad-scale recombination patterns underlying proper disjunction in humans. *PLoS Genetics*, *5*, e1000658.

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.

Foote, A. D., Vijay, N., Avila-Arcos, M. C., Baird, R. W., Durban, J. W., Fumagalli, M., … Wolf, J. B. (2016). Genome-culture coevolution promotes rapid divergence of killer whale ecotypes. *Nature Communications*, *7*, 11693.

Fournier-Level, A., Korte, A., Cooper, M.D., Nordborg, M., Schmitt, J., & Wilczek, A.M. (2011). A map of local adaptation in *Arabidopsis thaliana*. *Science*, *334*, 86–89.

Fraser, B. A., Kunstner, A., Reznick, D. N., Dreyer, C., & Weigel, D. (2015). Population genomics of natural and experimental populations of guppies *(Poecilia reticulata)*. *Molecular Ecology*, *24*, 389–408.

Gagnaire, P.-A., Pavey, S. A., Normandeau, E., & Bernatchez, L. (2013). The genetic architecture of reproductive isolation during speciation-with-gene-flow in lake whitefish species pairs assessed by rad sequencing. *Evolution*, *67*, 2483–2497.

Gautier, M., & Vitalis, R. (2012). *rehh:* An R package to detect footprints of selection in genome-wise SNP data from haplotype structure. *Bioinformatics*, *28*, 1176–1177.

Gavrilets, S., & Cruzan, M. B. (1998). Neutral gene flow across single locus clines. *Evolution*, *52*, 1277–1284.

Gillespie, J. H. (2000). Genetic drift in an infinite population: The pseudo-hitchhiking model. *Genetics*, *155*, 909–919.

Hill, W. G., Goddard, M. E., & Visscher, P. M. (2008). Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genetics*, *4*, e1000008.

Hill, W. G., & Robertson, A. (1966). Effect of linkage on limits to artificial selection. *Genetical Research*, *8*, 269–294.

Hobolth, A., Dutheil, J. Y., Hawks, J., Schierup, M. H., & Mailund, T. (2011). Incomplete lineage sorting patterns among human, chimpanzee, and orangutan suggest recent orangutan speciation and widespread selection. *Genome Research*, *21*, 349–356.

Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., & Cresko, W. A. (2010). Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, *6*, e1000862.

Hudson, R. R., & Kaplan, N. L. (1995). Deleterious background selection with recombination. *Genetics*, *141*, 1605–1617.

Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., … Kingsley, D. M. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, *484*, 55–61.

Joron, M., Frezal, L., Jones, R. T., Chamberlain, N. L., Lee, S. F., Haag, C. R., … Richard, H. (2011). Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature*, *477*, 203–206.

Kaplan, N. L., Hudson, R. R., & Langley, C. H. (1989). The hitchhiking effect revisited. *Genetics*, *123*, 887–899.

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

Kirubakaran, T. G., Grove, H., Kent, M. P., Sandve, S. R., Baranski, M., Nome, T., … Andersen, O. (2016). Two adjacent inversions maintain genomic differentiation between migratory and stationary ecotypes of Atlantic cod. *Molecular Ecology*, *25*, 2130–2143.

Kolaczkowski, B., Kern, A. D., Holloway, A. K., & Begun, D. J. (2011). Genomic differentiation between temperate and tropical Australian populations of *Drosophila melanogaster*. *Genetics*, *187*, 245–260.

Küpper, C., Stocks, M., Risse, J. E., Dos Remedios, N., Farrell, L. L., McRae, S. B., … Burke, T. (2016). A supergene determines highly divergent male reproductive morphs in the ruff. *Nature Genetics*, *48*, 79–83.

Lamichhaney, S., Berglund, J., Almen, M. S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., … Andersson, L. (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, *518*, 371–375.

Lawniczak, M. K. N., Emrich, S. J., Holloway, A. K., Regier, A. P., Olson, M., White, B., … Besansky, N. J. (2010). Widespread divergence between incipient *Anopheles gambiae* species revealed by whole genome sequences. *Science*, *330*, 512–514.

Lee, C.-R., Wang, B., Mojica, J. P., Mandáková, T., Prasad, K. V. S. K., Goicoechea, J. L., … Mitchell-Olds, T. (2017). Young inversion with multiple linked QTLs under selection in a hybrid zone. *Nature Ecology & Evolution*, *1*, 0119.

Li, X., Fan, D., Zhang, W., Liu, G., Zhang, L., Zhao, L., & … Wang, W. (2015). Outbred genome sequencing and CRISPR/Cas9 gene editing in butterflies. *Nature Communications*, *6*, 8212.

Lynch, M., & Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Sunderland: Sinauer Associates.

Marques, D. A., Lucek, K., Meier, J. I., et al. (2016). Genomics of rapid incipient speciation in sympatric threespine stickleback. *PLoS Genetics*, *12*, e1005887.

Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., … Jiggins, C. D. (2013). Genomewide evidence for speciation with gene flow in Heliconius butterflies. *Genome Research*, *23*, 1817–1828.

Maynard Smith, J., & Haigh, J. (1974). Hitch-hiking effect of a favorable gene. *Genetical Research*, *23*, 23–35.

Messer, P. W., & Petrov, D. A. (2013). Population genomics of rapid adaptation by soft selective sweeps. *Trends in Ecology & Evolution*, *28*, 659–669.

Muller, H. J. (1932). Some genetic aspects of sex. *The American Naturalist*, *66*, 118–138.

Nachman, M. W. (2002). Variation in recombination rate across the genome: Evidence and implications. *Current Opinion in Genetics & Development*, *12*, 657–663.

Nachman, M. W., Bauer, V. L., Crowell, S. L., & Aquadro, C. F. (1998). DNA variability and recombination rates at X-linked loci in humans. *Genetics*, *150*, 1133–1141.

Nachman, M. W., & Payseur, B. A. (2012). Recombination rate variation and speciation: Theoretical predictions and empirical results from rabbits and mice. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences*, *367*, 409–421.

Neafsey, D. E., Lawniczak, M. K. N., Park, D. J., Redmond, S. N., Coulibaly, M. B., Traore, S. F., … Muskavitch, M. A. T. (2010). SNP genotyping defines complex gene-flow boundaries among African malaria vector mosquitoes. *Science*, *330*, 514–517.

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

Noor, M. A. F., Cunningham, A. L., & Larkin, J. C. (2001). Consequences of recombination rate variation on quantitative trait locus mapping studies: Simulations based on the *Drosophila melanogaster* genome. *Genetics*, *159*, 581–588.

Nordborg, M., Charlesworth, B., & Charlesworth, D. (1996). The effect of recombination on background selection. *Genetical Research*, *67*, 159–174.

Oleksyk, T. K., Smith, M. W., & O'Brien, S. J. (2010). Genomewide scans for footprints of natural selection. *Philosophical Transactions of the Royal Society B*, *365*, 185–205.

Ortiz-Barrientos, D., Reiland, J., Hey, J., & Noor, M. A. F. (2002). Recombination and the divergence of hybridizing species. *Genetica*, *116*, 167–178.

Pease, J. B., & Hahn, M. W. (2013). More accurate phylogenies inferred from low-recombination regions in the presence of incomplete lineage sorting. *Evolution*, *67*, 2376–2384.

Poelstra, J. W., Vijay, N., Bossu, C. M., Lantz, H., Ryll, B., Muller, I., … Wolf, J. B. (2014). The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science*, *344*, 1410–1414.

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

Prüfer, K., Munch, K., Hellmann, I., Akagi, K., Miller, J. R., Walenz, B., … Paabo, S. (2012). The bonobo genome compared with the chimpanzee and human genomes. *Nature*, *486*, 527–531.

Puzey, J. R., Willis, J. H., & Kelly, J. K. (2017). Population structure and local selection yield high genomic variation in Mimulus guttatus. *Molecular Ecology*, *26*, 519–535.

R Development Core Team. (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ravinet, M., Prodohl, P. A., & Harrod, C. (2013). Parallel and nonparallel ecological, morphological and genetic divergence in lake-stream stickleback from a single catchment. *Journal of Evolutionary Biology*, *26*, 186–204.

Renaut, S., Grassa, C. J., Yeaman, S., Moyers, B. T., Lai, Z., Bowers, J. E., … Rieseberg, L. H. (2013). Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nature Communications*, *4*, 1827.

Rieseberg, L. H. (2001). Chromosomal rearrangements and speciation. *Trends in Ecology & Evolution*, *16*, 351–358.

Rockman, M. V. (2012). The QTN program and the alleles that matter for evolution: All that's gold does not glitter. *Evolution*, *66*, 1–17.

Rockman, M. V., Skrovanek, S. S., & Kruglyak, L. (2010). Selection at linked sites shapes heritable phenotypic variation in *C. elegans*. *Science*, *330*, 372–376.

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

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

Roesti, M., Kueng, B., Moser, D., & Berner, D. (2015). The genomics of ecological vicariance in threespine stickleback fish. *Nature Communications*, *6*, 8767.

Roesti, M., Moser, D., & Berner, D. (2013). Recombination in the threespine stickleback genome–patterns and consequences. *Molecular Ecology*, *22*, 3014–3027.

Roesti, M., Salzburger, W., & Berner, D. (2012b). Uninformative polymorphisms bias genome scans for signatures of selection. *BMC Evolutionary Biology*, *12*, 94.

Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z. P., Richter, D. J., Schaffner, S. F., … Lander, E. S. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature*, *419*, 832–837.

Sabeti, P. C., Schaffner, S. F., Fry, B., Lohmueller, J., Varilly, P., Shamovsky, O., … Lander, E. S. (2006). Positive natural selection in the human lineage. *Science*, *312*, 1614–1620.

Samuk, K., Owens, G. L., Delmore, K. E., Miller, S. E., Rennison, D. J., & Schluter, D. (2017). Gene flow and selection interact to promote adaptive divergence in regions of low recombination. *Molecular Ecology*, *26*, 4378–4390.

Schliep, K. P. (2011). *phangorn:* Phylogenetic analysis in R. *Bioinformatics*, *27*, 592–593.

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., … Widmer, A. (2014). Genomics and the origin of species. *Nature Reviews Genetics*, *15*, 176–192.

Shin, J.-H., Blay, S., McNeney, B., & Graham, J. (2006). *LDheatmap:* An R function for graphical display of pairwise linkage disequilibrium between single nucleotide frequencies. *Journal of Statistical Software*, *16*, 1–8.

Soria-Carrasco, V., Gompert, Z., Comeault, A. A., Farkas, T. E., Parchman, T. L., Johnston, J. S., … Nosil, P. (2014). Stick Insect genomes reveal natural selection's role in parallel speciation. *Science*, *344*, 738–742.

Stuart, Y. E., Veen, T., Weber, J. N., Hanson, D., Ravinet, M., Lohman, B. K., … Bolnick, D. I. (2017). Contrasting effects of environment and genetics generate a continuum of parallel evolution. *Nature Ecology & Evolution*, *1*, 0158.

Sturtevant, A. H., & Beadle, G. W. (1936). The relations of inversions in the X chromosome of Drosophila melanogaster to crossing over and disjunction. *Genetics*, *21*, 554–604.

Tang, K., Thornton, K. R., & Stoneking, M. (2007). A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS Biology*, *5*, e171.

Tennessen, J. A., & Akey, J. M. (2011). Parallel adaptive divergence among geographically diverse human populations. *PLoS Genetics*, *7*, e1002127.

Tine, M., Kuhl, H., Gagnaire, P. A., Louro, B., Desmarais, E., Martins, R. S., … Reinhardt, R. (2014). European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nature Communications*, *5*, 5770.

Wang, J., Wurm, Y., Nipitwattanaphon, M., Riba-Grognuz, O., Huang, Y. C., Shoemaker, D., & Keller, L. (2013). A Y-like social chromosome causes alternative colony organization in fire ants. *Nature*, *493*, 664–668.

Westram, A. M., Galindo, J., Alm Rosenblad, M., Grahame, J. W., Panova, M., & Butlin, R. K. (2014). Do the same genes underlie parallel phenotypic divergence in different *Littorina saxatilis* populations? *Molecular Ecology*, *23*, 4603–4616.

Yeaman, S., & Whitlock, M. C. (2011). The genetic architecture of adaptation under migration-selection balance. *Evolution*, *65*, 1897–1911.

Zaykin, D. V., Pudovkin, A., & Weir, B. S. (2008). Correlation-based inference for linkage disequilibrium with multiple alleles. *Genetics*, *180*, 533–545.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Berner D, Roesti M. Genomics of adaptive divergence with chromosome-scale heterogeneity in crossover rate. *Mol Ecol*. 2017;26:6351–6369. https://doi.org/10.1111/mec.14373