Supporting information

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

Authors

Daniel Berner[1] and Marius Roesti[1,2,*]

Affiliation

[1] Zoological Institute, University of Basel, Vesalgasse 1, 4051 Basel, Switzerland

[2] Department of Zoology, University of British Columbia, 6270 University Boulevard, Vancouver, BC V6T1Z4, Canada

* corresponding author (roesti@zoology.ubc.ca)

Content

Tables S1-S4

Figures S1-S15

Methods S1 & S2

References

Table S1. Organisms exhibiting a reduced crossover rate in the center of chromosomes as compared to the peripheral chromosome domains (upper part of the table), and organisms not displaying this broad-scale crossover pattern (lower part of the table), sorted alphabetically by species names within each organismal kingdom. The list is based on a literature survey of 46 total species for which both well-assembled chromosomes (physical positions) and crossover rate information along these chromosomes were available. Most species listed as providing evidence of relatively reduced crossover rate in chromosome centers were rated so because the authors of the corresponding studies themselves explicitly inferred a general broad-scale reduction in crossover rate in chromosome centers relative to the peripheries (E for 'explicit' in the Evidence column). In some studies, the pattern was not discussed by the authors but was unambiguous when inspecting graphics presenting the association between physical chromosome position and crossover rate (V for 'visual' in the Evidence column). To help readers appraise our categorization, the Reference(s) column specifies in parentheses the key graphical item in the corresponding publication used as basis for our rating.

Seven exemplary species in which broad-scale heterogeneity in crossover rate along chromosomes is clearly unrelated to the location of the centromere position are marked by an astersik (*). We note that in some taxa like nematodes (including *Caenorhabditis*) or Lepidoptera (including *Heliconius*), the chromosomes are holocentric (i.e., they lack a localized centromere) so that broad-scale heterogeneity in crossover rate must be unrelated to the centromere location. In most organisms, however, the relationship between crossover rate and centromere location cannot be evaluated, either because the chromosomes are (close to) metacentric (i.e., centromere is located in the physical center of a chromosome), or because the centromere positions are unknown.

Our survey reveals clearly that a reduced crossover rate in chromosome centers is a general biological phenomenon, although three exceptions emerge.

Scientific name	**Colloquial designation**	**Kingdom (Subgroup)**	**Evidence**	**Reference(s)**
Organisms with reduced crossover rate in chromosome centers relative to the peripheries:				
Aedes aegypti	Yellow fever mosquito	Animal (insect)	E	Juneja et al. 2014 (Fig. 5)
Bos primigenius	Cattle	Animal (vertebrate)	E	Sandor et al. 2012 (Suppl. Fig. 2d)
Caenorhabditis briggsae	Worm	Animal (nematode)	E	Ross et al. 2011 (Fig. 2)
Caenorhabditis elegans	Worm	Animal (nematode)	E	Rockman & Kruglyak 2009 (Fig. 1)
Canis lupus familiaris	Dog	Animal (vertebrate)	E	Wong et al. 2010 (Fig. 1)
Danio rerio	Zebrafish	Animal (vertebrate)	E	Bradley et al. 2011 (Fig. 2)

Daphnia magna	Waterflea	Animal (crustacean)	E	Dukic et al. 2016 (Fig. 4)
Dicentrarchus labrax	Seabass	Animal (vertebrate)	E	Tine et al. 2014 (Fig. 4c)
Drosophila melanogaster	Fruit fly	Animal (insect)	V	Mackay et al. 2012 (Fig. 2a)
*Felis catus**	Domestic cat	Animal (vertebrate)	E	Li et al. 2016 (Fig. 2)
Ficedula albicollis	Collared flycatcher	Animal (vertebrate)	E	Kawakami et al. 2014 (Figs. 5, 6 & 8)
Gallus gallus	Chicken	Animal (vertebrate)	E	Groenen et al. 2009 (Fig. 2)
*Gasterosteus aculeatus**	Threespine stickleback	Animal (vertebrate)	E	Roesti et al. 2013, Glazer et al. 2015; (e.g., Fig. 1 in Roesti et al.)
Heliconius melpomene	Passion-vine butterfly	Animal (insect)	E	Davey et al. 2016; Simon Martin, personal communication
*Homo sapiens**	Human	Animal (vertebrate)	E	Kong et al. 2002, Jensen-Seaman et al. 2004, Tapper et al. 2005, Chowdhury et al. 2009, Auton et al. 2012; (e.g., Fig. 1a in Auton et al.)
Ictalurus punctatus	Channel catfish	Animal (vertebrate)	V	Liu et al. 2016 (Suppl. Fig. 2a-e)
Mus musculus	House mouse	Animal (vertebrate)	E	Brunschwig et al. 2012 (Fig. 1b)
Nasonia vitripennis, N. giraulti	Parasitic wasp	Animal (insect)	E	Niehuis et al. 2010 (Fig. 1)
Pan troglodytes verus	Chimpanzee	Animal (vertebrate)	E	Auton et al. 2012 (Fig. 1a)
Pungitius pungitius	Ninespine stickleback	Animal (vertebrate)	V	Rastas et al. 2015 (Fig. 3a,b)
Rattus norvegicus	Brown rat	Animal (vertebrate)	E	Jensen-Seaman et al. 2004 (Fig. 1)
Salmo salar	Atlantic salmon	Animal (vertebrate)	E	Tsai et al. 2016 (Fig. 2)
Sorex araneus	Common shrew	Animal (vertebrate)	E	Borodin et al. 2008 (Fig. 2 & 3)
*Sus scrofa**	Wild boar	Animal (vertebrate)	E	Tortereau et al. 2012 (Fig. 1 & 3)
Taeniopygia guttata	Zebra finch	Animal (vertebrate)	E	Backström et al. 2010 (Fig. 1 & 2)
Saccharomyces cerevisiae	Yeast	Fungus	E	Barton et al. 2008 (Fig. 4)
Zymoseptoria tritici	-	Fungus	E	Croll et al. 2015 (Fig. 2 & Suppl. Fig. 5)
Beta vulgaris	Sugar beet	Plant	E	Dohm et al. 2012 (Fig. 3)

Brachypodium distachyon	Purple false brome	Plant	E	Huo et al. 2011 (Fig. 2)
Brassica napus	Rapeseed	Plant	V	Wang et al. 2015 (Fig. 3)
Citrullus lanatus	Watermelon	Plant	E	Ren et al. 2012 (Fig. 4)
*Cucumis melo**	Muskmelon	Plant	E	Argyris et al. 2015 (Fig. 2)
*Glycine max**	Soybean	Plant	E	Schmutz et al. 2010 (Suppl. Fig. 1)
Helianthus annuus	Common sunflower	Plant	V	Renaut et al. 2013 (Suppl. Fig. 5)
Juglans regia	Walnut	Plant	E	Luo et al. 2015 (Fig. 1)
Miscanthus sinensis	Silvergrass	Plant	V	Swaminathan et al. 2012 (Fig. 4a)
Oryza sativa	Rice	Plant	V	Tian et al. 2009 (Fig. 2)
Phaseolus vulgaris	Common bean	Plant	E	Bhakta et al. 2015 (Fig. 4)
Solanum lycopersicum	Tomato	Plant	E	Tomato Genome Consortium 2012 (Suppl. Fig. 1)
Solanum tuberosum	Potato	Plant	V	Endelman et al. 2016 (Suppl. Fig. 4)
Sorghum bicolor	Sorghum	Plant	V	Bekele et al. 2013 (Fig. 7)
Triticum aestivum	Common wheat	Plant	E	Akhunov et al. 2003, Gardner et al. 2016 (e.g., Fig. 5 in Gardner et al.)
*Zea mays**	Maize	Plant	E	Anderson et al. 2003, Bauer et al. 2013, Li et al. 2015 (e.g., Fig. 2a in Li et al.)
Organisms without reduced crossover rate in chromosome centers relative to the peripheries:				
Apis mellifera	Honey bee	Animal (insect)	V	Solignac et al. 2007, Liu et al. 2015 (e.g., Fig. 4 in Solignac et al.)
Arabidopsis thaliana	Thale cress	Plant	V	Signer et al. 2005, Salomé et al. 2012, Choi et al. 2013 (e.g., Suppl. Figs. 9-11 in Singer et al.)
Medicago truncatula	Barrelclover	Plant	E	Paape et al. 2012 (Fig. 1)

Table S2. Overview of analyses performed to validate assumptions underlying the standard simulation model.

Model element	Check	Result	Visualization
Starting allele frequencies	The starting allele frequencies at all genetic loci are determined by sampling randomly from an uniform or from an exponential distribution (standard model: the starting allele frequency is 0.5 at all genetic loci)	Modifications of the starting allele frequencies at the genetic loci do not alter the conclusions drawn from the base model	Fig. S1a,b
Fitness function	Individual fitness is a multiplicative function of the allelic states across loci (standard model: additive fitness)	Whether fitness is calculated additively or multiplicatively does not impact conclusions	Fig. S1c
Reproductive scheme	Within each generation of evolution, each population forms *N* mating pairs producing one offspring each (standard model: *N*/2 mating pairs producing two offspring each)	Results are quantitatively very similar between the two reproductive schemes	details not presented
Ploidy of individuals	Individuals are modeled as diploids (standard model: haploid individuals)	Ploidy does not change simulation results and conclusions	Fig. S1d
Crossover rate	Overall (i.e., average) crossover rate is two fold higher or lower than in the standard model	Differentiation, including CCBD, becomes slightly weaker with increasing crossover rate. Importantly, however, the overall crossover rate proves to have no qualitative effect on the general pattern of differentiation across the chromosome.	Fig. S2
Selection coefficient per locus	Cumulative selection strength across all selected loci is held constant, and hence when changing the number of loci under selection (*SL*), the per-locus selection coefficient changes (standard model: the selection coefficient is defined at the level of the locus, hence the cumulative selection strength varies when changing *SL*)	Ploidy does not change simulation results and conclusions	Fig. S3a
Chromosomal representation of individuals	Individuals are represented by three chromosomes each (standard model: a single chromosome)	Conclusions are not contingent on the number of chromosomes	Fig. S4
Differentiation metric	F_{ST} and D_{xy} used as alternative metrics to quantify population differentiation (standard model: simple allele frequency difference)	Results and conclusions are robust to the specific differentiation metric applied	Fig. S7
Simulation averaging	Interpretation of individual simulation replicates (standard presentation is averages across 100 simulation replicates)	Despite noise, the general features identified with the standard model can be observed in single simulation replicates	Fig. S8

Table S3. Full parameter ranges explored with the standard simulation model of multilocus divergence with gene flow. Default values are indicated in bold.

Parameter	Description	Parameter range considered
N	Number of haploid individuals per population	250, 500, 750, 1,000, 2,500, 5,000, 7,500, **10,000**, 25,000, 50,000
SL	Number of selected loci	4, 8, 12, **16**, 20
s	Selection coefficient for each locus under selection	0, 0.00125, 0.0025, 0.005, 0.0075, **0.01**, 0.0125, 0.0175, 0.02, 0.025, 0.0375, 0.05, 0.0625
m	Migration rate (proportion of each population moving to the other population in the beginning of each life cycle)	0, 0.001, 0.005, **0.01**, 0.0125, 0.0175, 0.025, 0.0375, 0.05, 0.1, 0.2
g	Generations of evolution	10, 20, 50, 100, 200, 500, **1,000**, 2,000, 5,000
f	Starting frequency of the derived allele at the selected loci	0.0005, 0.001, 0.005, 0.01, 0.025, 0.05, 0.1, 0.2, 0.3, 0.4, **0.5**
cbias	Proportion of crossover occurring in the chromosome periphery	0.5, 0.6, 0.7, 0.8, **0.9**, 0.95, 0.975, 1

Table S4. Overview of the standard simulation model and all its modifications.

Name	Description	Visualization
Standard model	Base model for two populations adapting under polygenic divergent selection to different environments in the presence of gene flow.	Basis of all analysis and figures unless specified otherwise
Restricted gene flow model	The selected loci (*SL*), but not the neutral loci (*NL*), are exchanged as entire haplotypes during migration between the diverging populations (see Fig. S5). This allows investigating the effect of hitchhiking *independently* from the effect of the gene flow barrier in driving population differentiation at the *NL.* Subtracting the magnitude of differentiation at the *NL* observed in the *restricted gene flow* model from the magnitude of differentiation in the standard model therefore isolates the characteristic contribution of the gene flow barrier to population differentiation.	Fig. 2b; conceptual visualization: Fig. S5
S*econdary contact* model	Two secondary contact scenarios were considered: (i) Both diverging populations have all *SL* fixed for their locally favorable allele in the beginning of the simulations, mimicking a situation in which two populations have reached complete local adaptation during a phase without any gene flow (allopatry). Evolution otherwise occurs as in the standard model. (ii) The population first diverge in allopatry (m = 0) for 1,000 generations, followed by 2,000 generations of divergence with gene flow (during which differentiation was tracked).	Fig. S10c,d
Different SL model	All *SL* are shifted by one position along the chromosome relative to the standard model (see Fig. S6). The loci under selection are thus not the same in the two models. Comparing differentiation within a population pair diverged under this *different SL* model to differentiation within a population pair diverged under standard model conditions allows investigating how CCBD can result in patterns consistent with parallel genomic divergence, despite the absence parallelism in divergent selection between population pairs.	Fig. 6b; conceptual visualization: Fig. S6
Selection target density model	The density of *SL* is skewed to either the center or the peripheries of the chromosome. That is, the model assumes either 10 *SL* in the periphery and 6 *SL* in the center, or vice versa. Data obtained from this model allow investigating whether a difference in the density of selection targets between the chromosome center versus periphery could explain the emergence of strong CCBD.	Fig. S11b

Figure S1

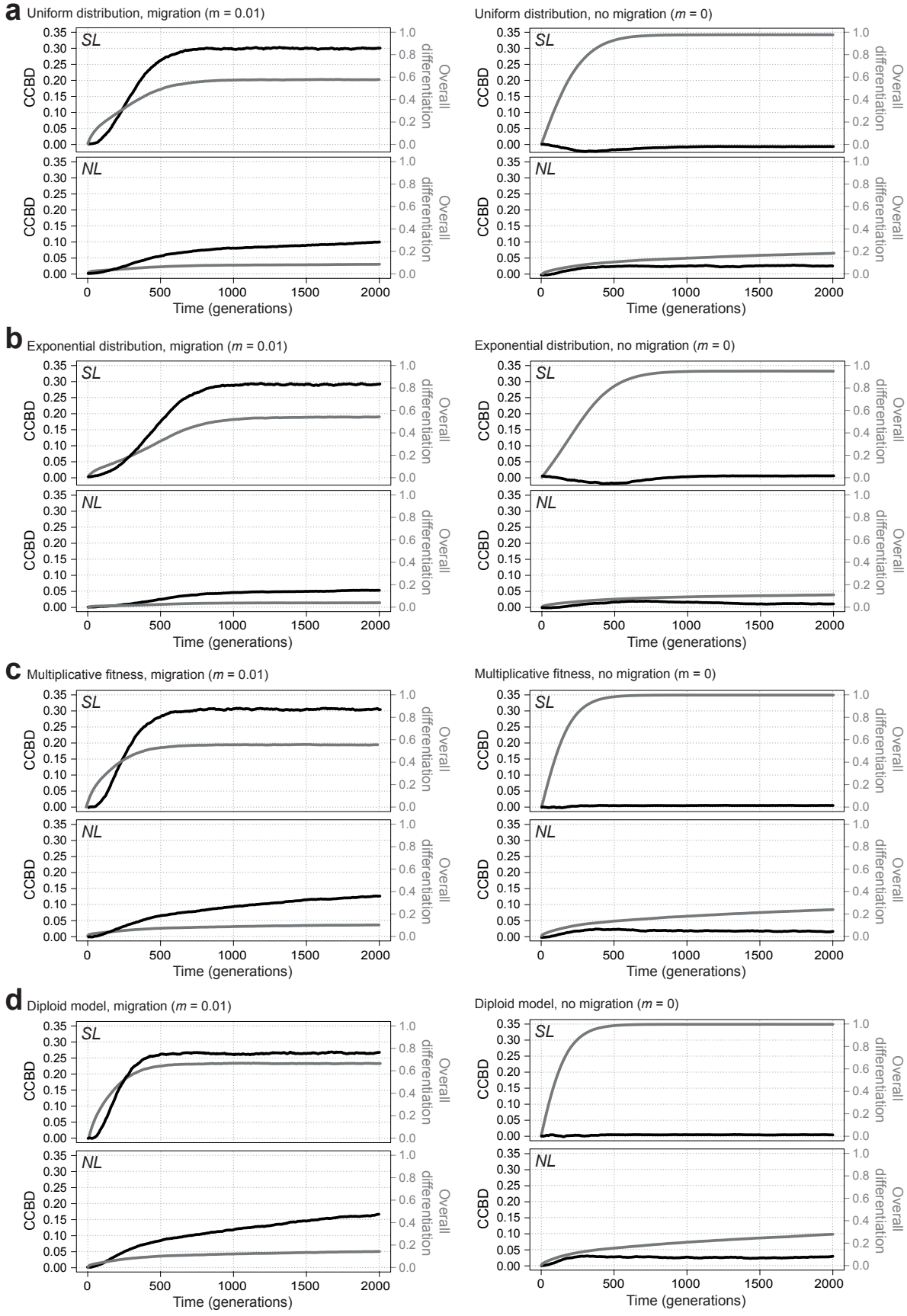

a Uniform distribution, migration (m = 0.01)
Uniform distribution, no migration (m = 0)
b Exponential distribution, migration (m = 0.01)
Exponential distribution, no migration (m = 0)
c Multiplicative fitness, migration (m = 0.01)
Multiplicative fitness, no migration (m = 0)
d Diploid model, migration (m = 0.01)
Diploid model, no migration (m = 0)
SL
NL
CCBD
Overall differentiation
Time (generations)

Figure S1. Complementary simulations to validate the robustness of our default model assumptions. Results obtained from simulations where the starting allele frequencies for all loci were determined by sampling randomly from **(a)** the uniform and **(b)** the exponential distribution were comparable to the results from our default simulation models (where the starting allele frequencies were 0.5) (see also Fig. S10b). **(c)** To confirm that our results were not dependent on the fitness definition, we performed simulations in which individual fitness was a multiplicative function of the number of locally unfavorable alleles. That is, fitness was specified as $(1\text{-}s)^n$, where n corresponds to the number of locally unfavorable alleles along the chromosome. **(d)** To confirm the robustness of the results generated with our default haploid simulation model, we implemented an analogous model with diploid individuals. The total number of chromosomes in the population here coincided with the one in the haploid model (i.e., N), but the chromosomes were contained in pairs within hermaphrodite, sexually reproducing individuals. Accordingly, individual fitness and hence the probability to reproduce was determined by the number of locally unfavorable alleles at the loci under selection across both chromosomes (i.e., $2*SL$ selected loci). Individuals selected to form a reproductive pair ($N/4$ pairs in total) produced four haploid gametes analogously to typical meiosis in natural system with one crossover: two gametes exhibited crossover between the chromosomes while two remained non-recombined. All reproductive pairs then produced two diploid offspring, each combining a gamete selected at random from each parent. The figure's plotting conventions are identical to the ones in Fig. 2a. Curves show means across 100 replicate simulation runs, both for standard model settings and for population divergence in allopatry (i.e., $m = 0$).

Figure S2

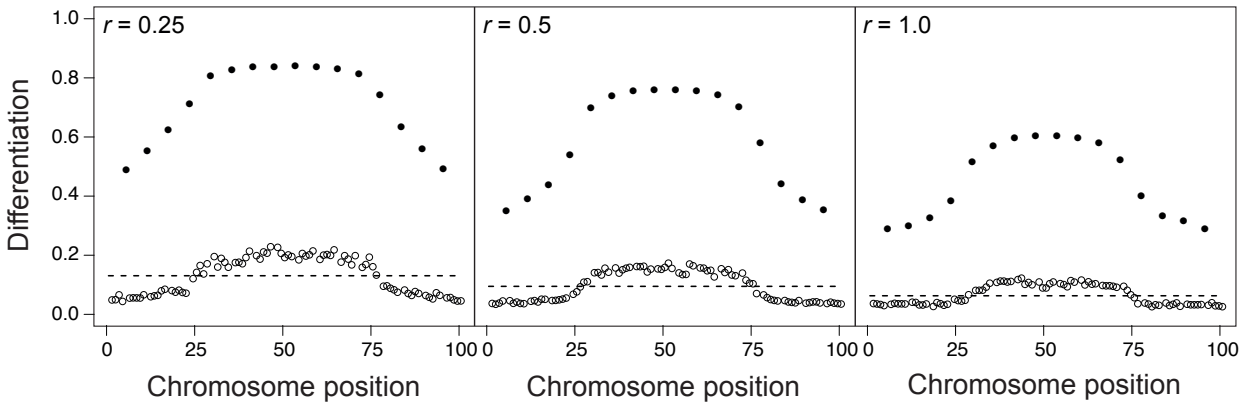


Figure S2. Influence of variation in the overall crossover rate *(r)* on CCBD. Simulations run with a twofold lower (left panel) and higher (right panel) overall crossover rate than that chosen as our biologically realistic default (central panel) demonstrate that CCBD and overall differentiation become slightly weaker with increasing crossover rate. Importantly, however, the overall crossover rate proves to have no qualitative effect on the general pattern of differentiation across the chromosome. Plotting conventions are as in Fig. 4.

Figure S3

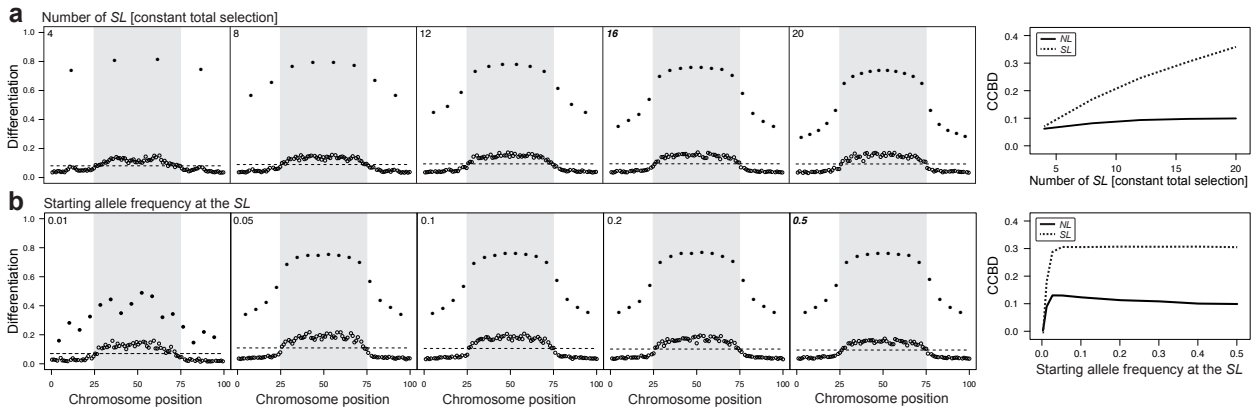


Figure S3. Influence of variation in the number of *SL* with constant total selection strength, and of the initial allele frequency of the two alleles at the *SL*, on CCBD. (**a**) The magnitude of CCBD at the *SL* and *NL* increases with the number of loci under selection. Contrary to Fig. 4d, the cumulative strength of selection across all *SL* was here held constant, so that the per-locus selection coefficient decreased proportionally with increasing locus number. (**b**) As long as the minor allele frequency at the *SL* in the beginning of the simulations is greater than a few percent, the imbalance between the two alleles has only a trivial influence on CCBD. Across this broad parameter range, the magnitude of CCBD declines subtly as the allele frequency approaches perfect balance (i.e., $f = 0.5$). The reason is that in the population initially more strongly maladapted, the favorable alleles drive more extensive initial hitchhiking when they rise from a low starting allele frequency. However, if the minor allele frequency at the *SL* is highly imbalanced (*f* below a few percent), the outcome of evolution is qualitatively different. In this range, locally favorable alleles are so rare in one of the habitats they will generally segregate in isolation from other favorable alleles; that is, they will rarely recombine into a haplotype substantially fitter than immigrant haplotypes. In this population, beneficial alleles at the *SL* thus become prone to swamping by gene flow. Consequently, the magnitude of adaptive divergence between the populations declines as *f* approaches zero – the population whose beneficial alleles are initially very rare will remain relatively maladapted, thus impeding the emergence of gene flow barriers and CCBD. All plotting conventions are as in Fig. 4.

Figure S4

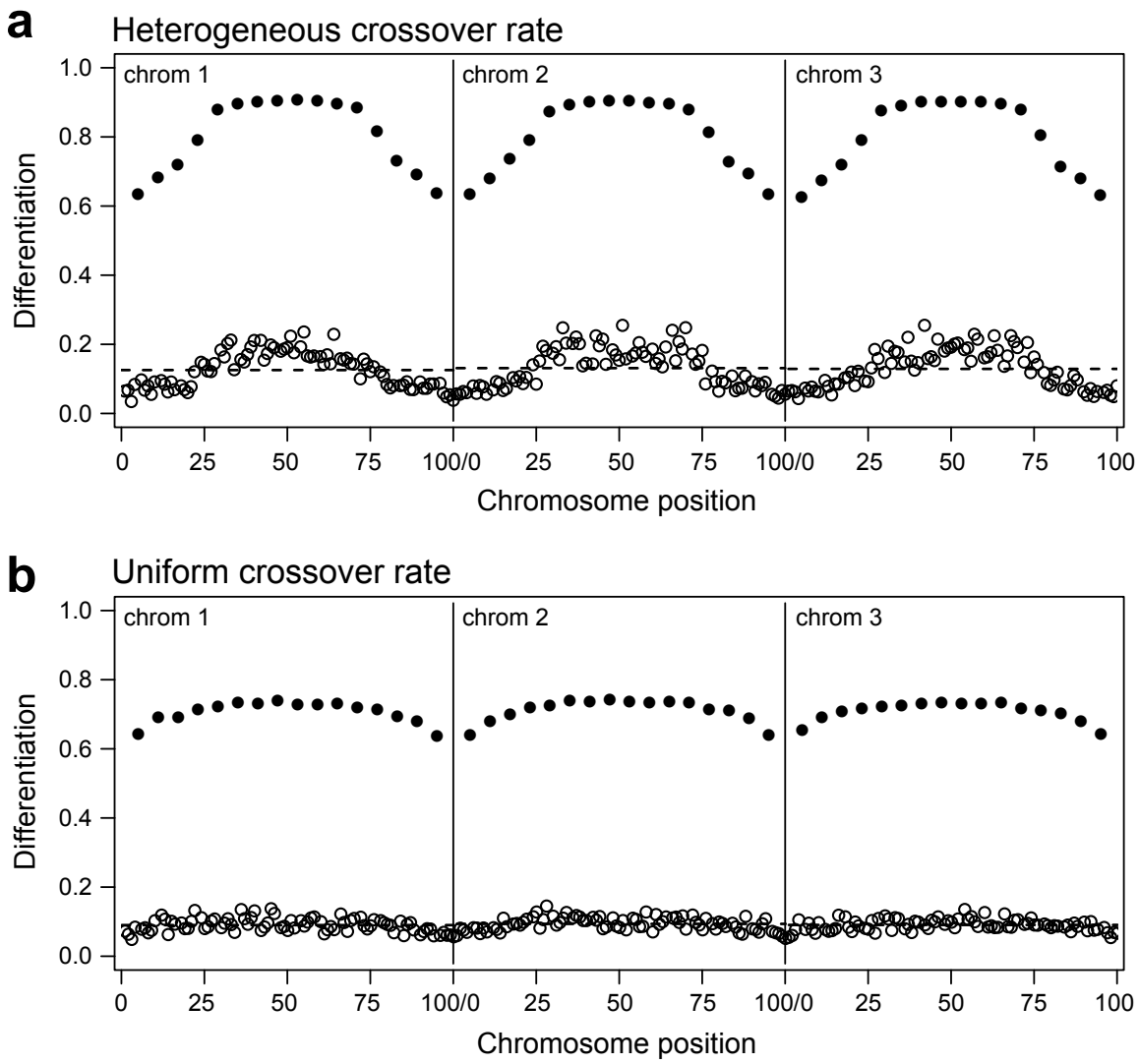


Figure S4. Extension of the standard simulation model to multiple chromosomes. To confirm the robustness of the analyses based on our standard single-chromosome model, additional simulations were performed in which each individual consisted of three independently segregating chromosomes, each containing loci under divergent selection. (**a**) With default parameter settings (hence 48 total *SL*, given as filled circles), strong CCBD emerges, as in the standard model (Fig. 4). The model in (**b**) differs from **a** in that the crossover rate is uniform across the chromosomes (i.e., the parameter *cbias* is set to 0.5, see Fig. 4e), a condition still slightly promoting hitchhiking in the chromosome peripheries (see Fig. S10b). Plotting conventions are as in Fig. 4.

Figure S5

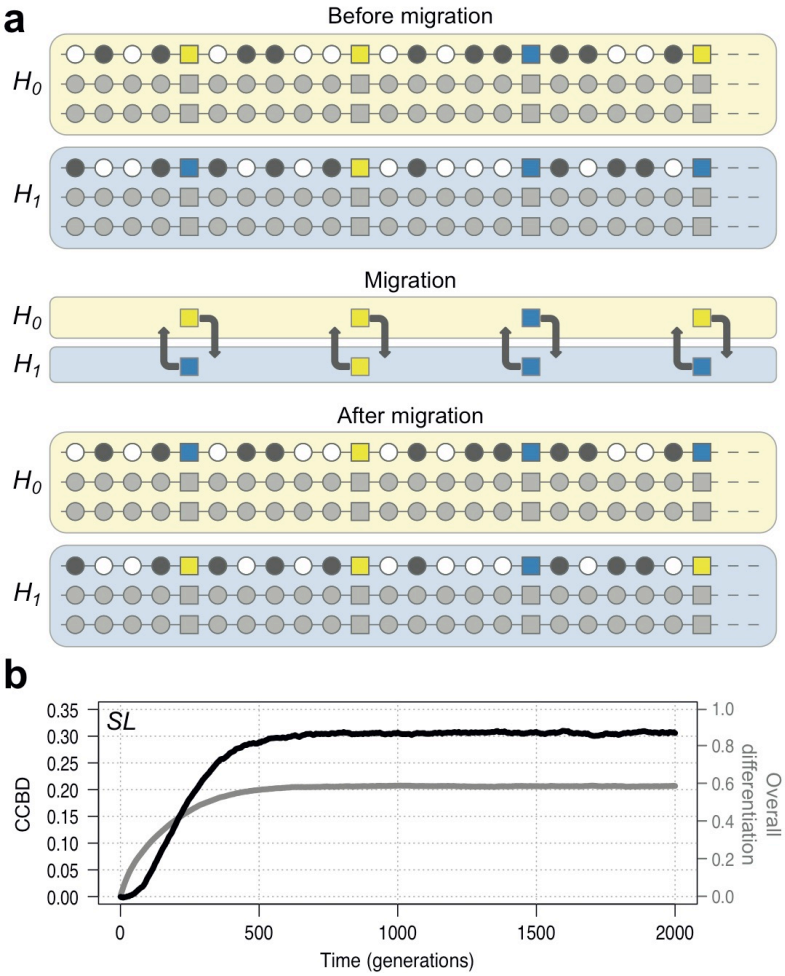


Figure S5. Details on the *restricted gene flow* model used to identify the contribution of hitchhiking – irrespective of the gene flow barrier – to CCBD. (**a**) Schematic of the migration scheme. First, individuals from each of the two selectively different habitats (H_0 and H_1) are drawn at random for migration to the opposite habitat. In this graphic, individuals are represented as single chromosomes, with neutral loci (*NL*, displaying black or white alleles) shown as circles and loci under divergent selection (*SL*) shown as squares (yellow and blue alleles favorable in H_0 and H_1, respectively). For clarity, the schematic shows only the first 23 loci of each chromosome. During the migration phase, pairs of migrants from the two habitats mutually exchange their alleles only at the *SL*. These alleles migrate as intact haplotypes, while the recipient neutral backgrounds also maintain their haplotype structure. The *SL* thus evolve exactly as in the standard model of divergence with gene flow [compare (**b**) to the left panel in Fig. 2a]. Since the alleles at the *NL* are not exchanged between the populations (i.e., gene flow at the *NL* is zero), however, selection can influence divergence at the *NL* only through hitchhiking caused by selection on the *SL*.The barrier to gene flow thus cannot operate as a driver of CCBD at the *NL*. The contribution of the gene flow barrier to CCBD at the *NL* can be then identified by calculating the differences in the magnitude of CCBD between the standard model (where both hitchhiking and the gene flow barrier operate) and the restricted gene flow model (where only hitchhiking operates).

Figure S6

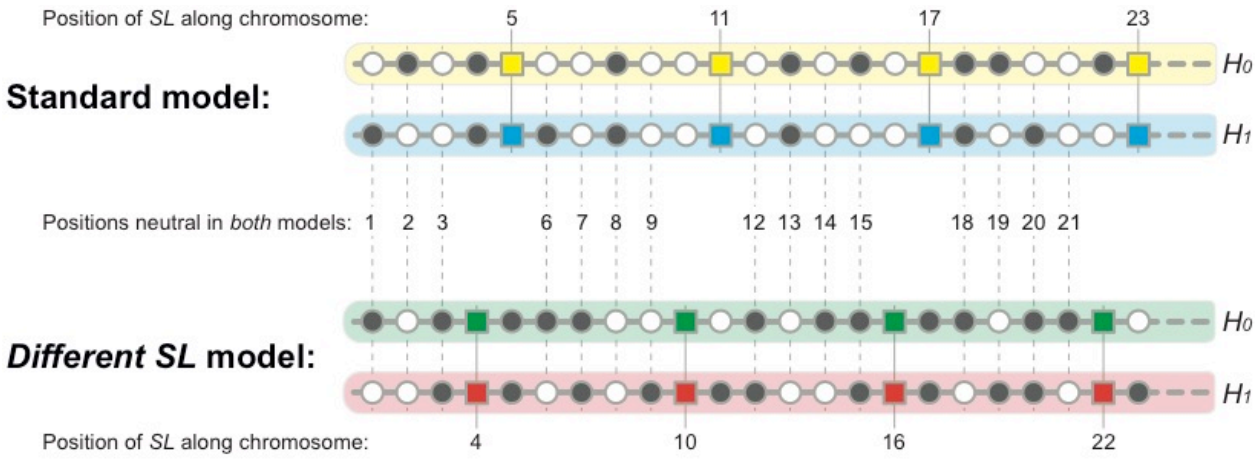


Figure S6. Schematic of the *different SL* model versus the standard simulation model. Neutral loci are shown as circles and display different alleles in black and white. Loci under divergent selection (*SL*) between the two populations residing in ecologically different habitats (H_0 and H_1; indicated by different background colors) are shown as squares. For clarity, only one individual per population and only the first 23 loci of each chromosome are shown. The only difference between the two models is that the *SL* are shifted by one physical position along the chromosome. Hence, the population pairs in the two models diverge along completely different (non-parallel) selective axes. Loci neutral in *both* models (indicated by dashed lines; total N = 68) were used to screen population pairs from the two different models for (non-)shared patterns in chromosome-wide differentiation values and high-differentiation 'outliers'.

Figure S7

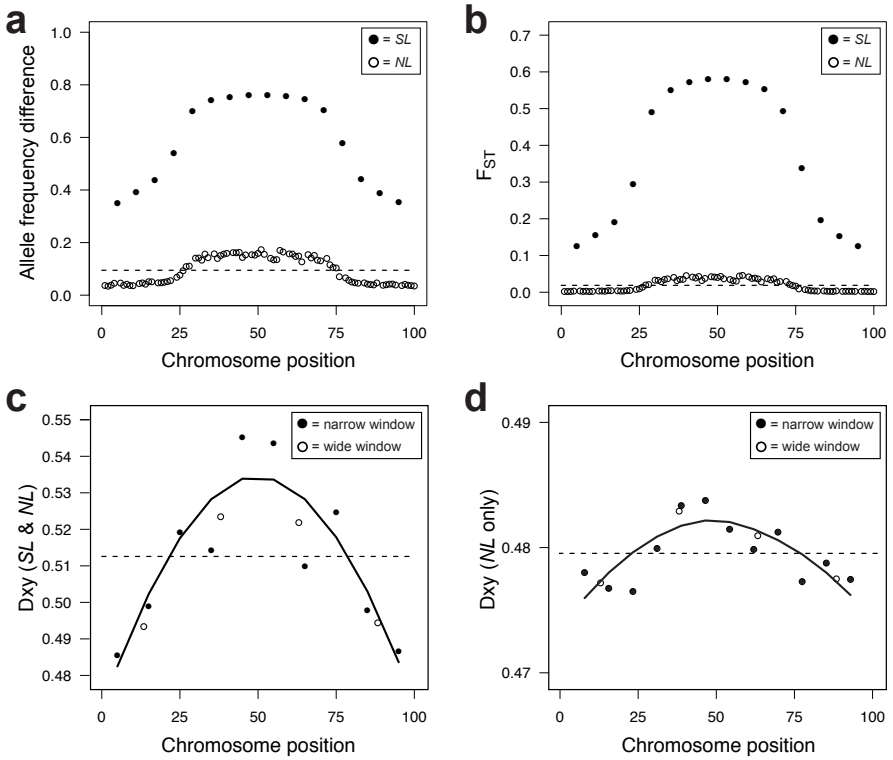


Figure S7. CCBD captured by different metrics for genetic population divergence (allele frequency difference, F_{ST}, and D_{xy}). To verify the robustness and generality of the standard metric used to quantify population differentiation (and thus CCBD) in our study, we compared multiple alternative statistics of population differentiation calculated from the same simulated data sets. (**a**) Our standard approach used throughout the paper was to quantify differentiation as the absolute difference between the populations in the frequency of the allele 1. In (**b**), allele frequency data from the loci were used to calculate differentiation as F_{ST} based on haplotype diversity (equation 7 in Nei & Tajima 1981; see e.g., Roesti et al. 2012a, 2015). The plots in the bottom row show differentiation expressed as D_{xy} (Nei & Li 1979), representing the mean number of nucleotide differences between pairs of sequences from two different populations. This metric is generally calculated based on allelic information from a stretch of DNA and not for single base positions. We thus analogously computed D_{xy} for non-overlapping windows along the chromosome. For (**c**), we considered both the *SL* and *NL* within the sliding windows, each spanning either 10 loci (i.e., 'high resolution'; raw and smoothed values shown as black dots and black line) or 25 loci (i.e., 'coarse resolution'; raw values shown as white dots). For (**d**), we considered the *NL* only and used a sliding window size of 7 (high resolution) and 21 (coarse resolution) loci. To calculate D_{xy}, we formed 100 pairs of chromosomes selected at random without replacement from each population, whereas **a** and **b** are based on the complete data from each population. Note that D_{xy} in **c** and **d** is high compared to values commonly reported in empirical studies, because all loci on our simulated chromosome were initially polymorphic within each population. Throughout the graphic, data points represent averages across 100 replicate simulations generated with the standard simulation model and default parameter settings.

Figure S8

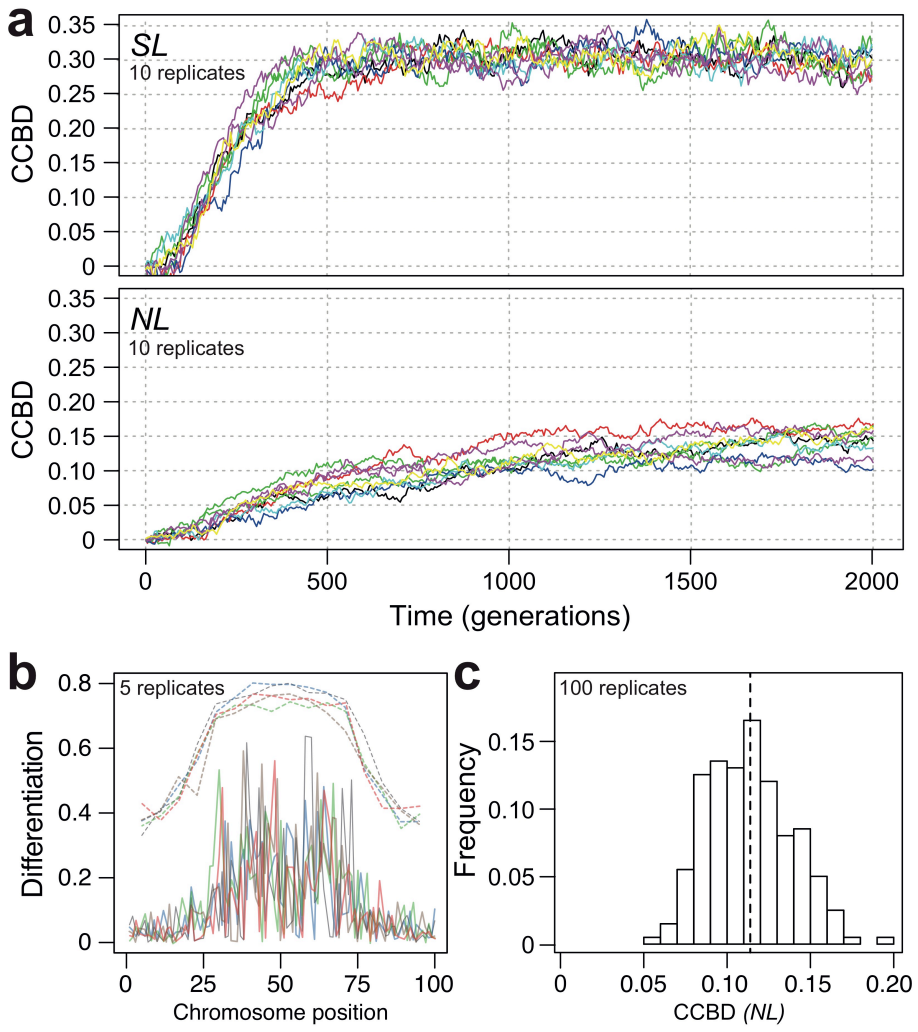


Figure S8. CCBD in individual simulation replicates. To illustrate variation among individual simulation runs, (**a**) visualizes the emergence of CCBD over time at the selected *(SL)* and neutral *(NL)* loci for ten randomly selected replicates, and (**b**) shows differentiation along the chromosome at the two types of loci (*SL* and *NL* represented by dashed and solid lines) for five random replicates. The simulations were run with default parameter settings, and the plotting conventions are analogous to Fig. 2a and Fig. 4. (**c**) Histogram showing the distribution of the magnitude of CCBD at the neutral loci across 100 replicate simulations (the vertical dashed line indicates the mean). Note that bias in differentiation toward the chromosome center is observed at both the *SL* and *NL* in all individual simulation runs.

Figure S9

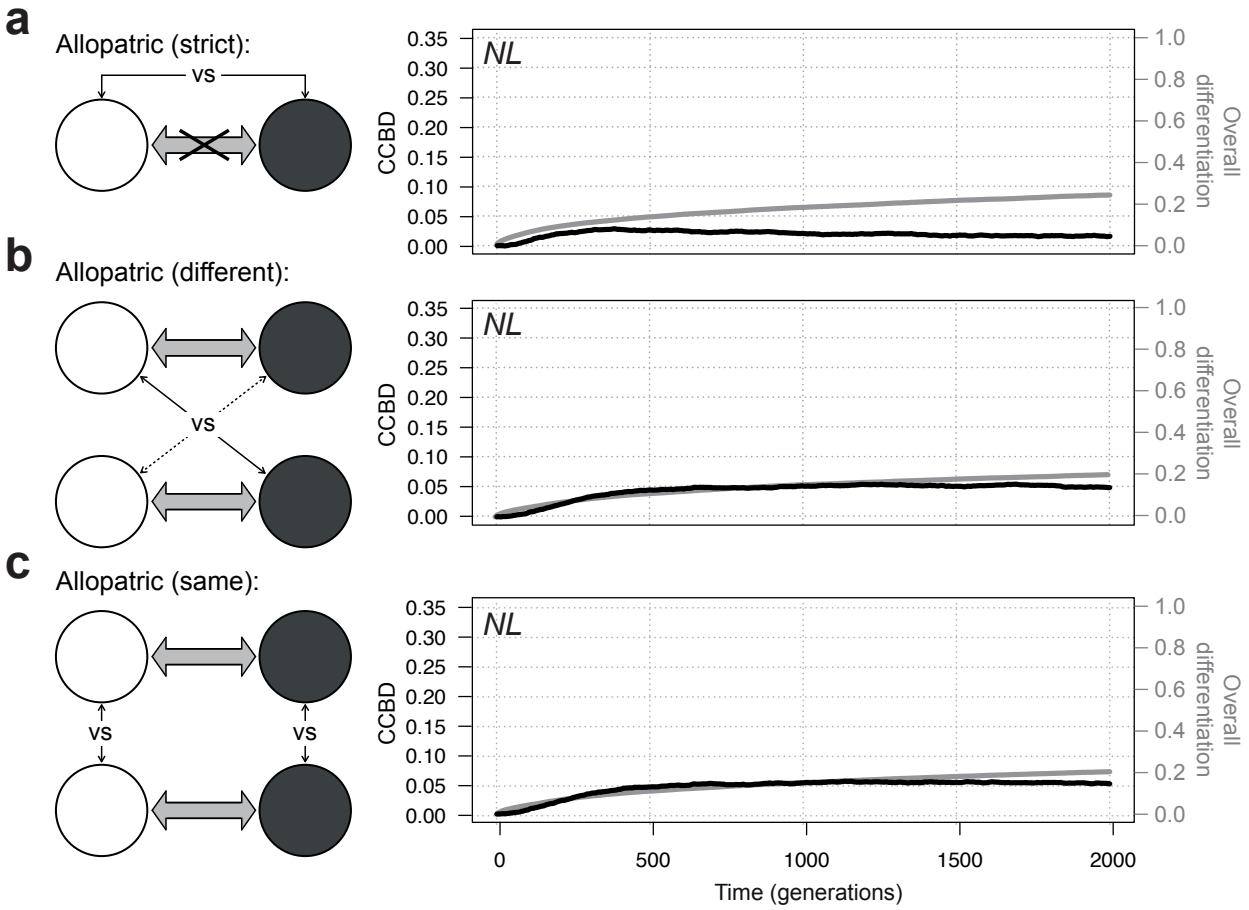


Figure S9. Magnitude of CCBD in different types of allopatric population comparisons. The schematics in the left column recapitulate the migration regimes and the type of comparisons among populations (for further details see Fig. 5a). The graphs in the right column show the corresponding magnitude of CCBD and population differentiation at the *NL* over time. The data were generated using the standard model with default parameter settings, and the plotting conventions for the right column follow those of Fig. 2a. Note that in all three scenarios, CCBD is weak relative to the magnitude of overall differentiation across the chromosome (compare to Fig. 2a, bottom), and arises from hitchhiking while the populations establish migration-selection balance at the *SL*. Hence, even when gene flow barriers may drive increasingly strong CCBD over time within pairs of populations diverging with migration (Fig. 2a), allopatric comparisons involving such populations will not reveal this barrier effect.

Figure S10

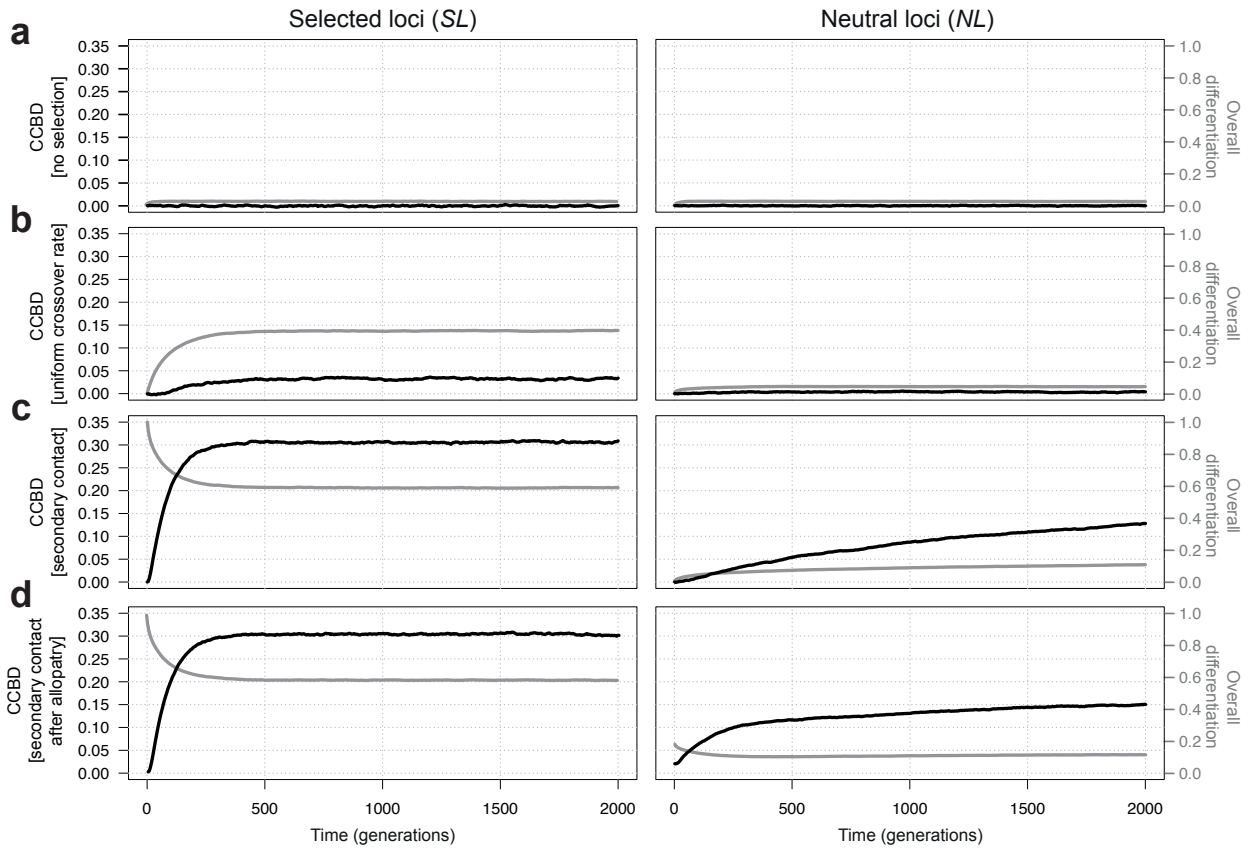


Figure S10. The influence of the complete absence of selection, a uniform crossover rate, and secondary contact on the magnitude of CCBD and overall population differentiation. (**a**) Simulations with the standard model in which divergent selection is absent (s = 0) make clear that CCBD is a fundamental signature of adaptation to ecologically distinct habitats. (**b**) Even when the crossover rate is completely uniform along the chromosome (b = 0.5), subtle CCBD can arise, an effect detectable mainly at the *SL*. Because the magnitude of CCBD does not materially increase beyond the initial phase during which migration-selection balance is established, the pattern must be driven by hitchhiking. This was confirmed by simulations based on the restricted gene flow model with b = 0.5; details not presented. (**c**) Secondary contact scenario. Here, divergence with gene flow starts with both populations being perfectly adapted to their home habitats (i.e., fixed for opposed alleles at the *SL*), corresponding to migration after adaptive divergence in complete isolation. The result differs from the outcome of the standard primary divergence model (Fig. 2a) only in the early phase of evolution during which migration-selection balance is established. Beyond this phase, the results of the two models converge. (**d**) A very similar result was obtained when we modeled an initial phase of 1,000 generations of allopatric population divergence (i.e., m = 0), followed by 2,000 generations of gene flow (m = 0.01) during which we tracked CCBD and overall (mean) differentiation. In comparison to **c**, however, slight CCBD (and mean differentiation) caused by hitchhiking during the allopatric phase is noticeable at the beginning of divergence with gene flow (secondary contact) already. All panels are plotted according to Fig. 2a.

Figure S11

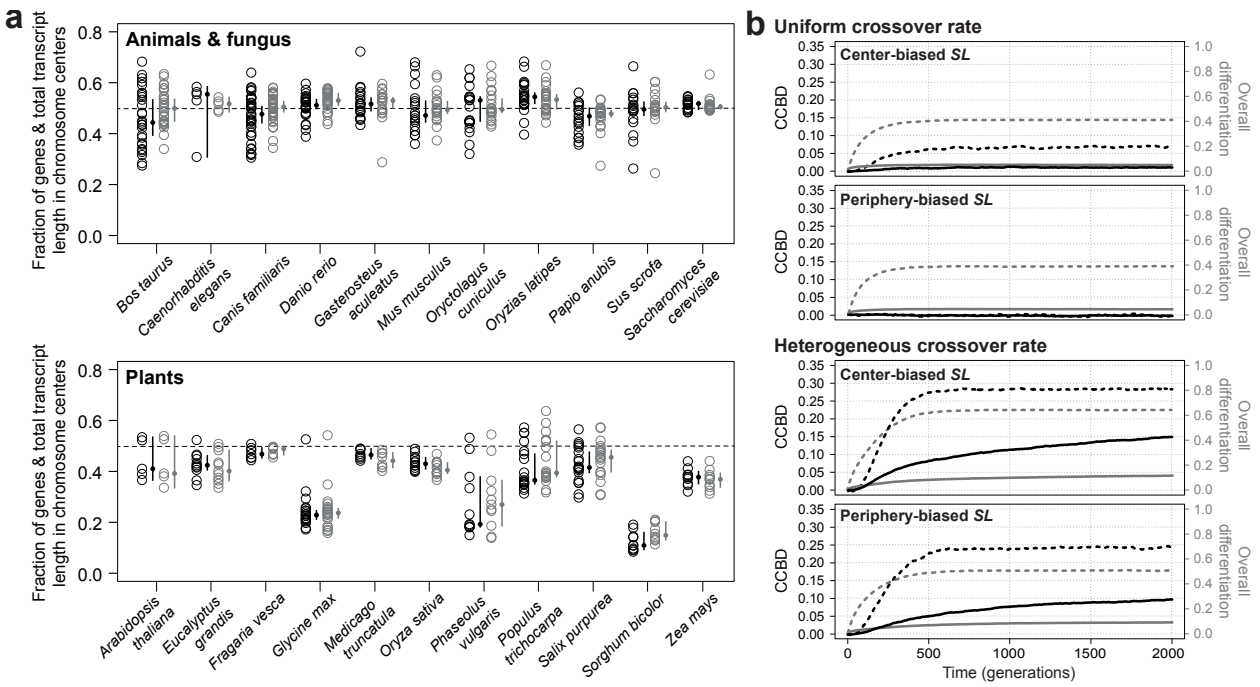


Figure S11. Distribution of selection targets along chromosomes and its contribution to CCBD. (**a**) Chromosome-scale distribution of two proxies of potential selection targets – the density of genes and the total transcript length – within the genome of 22 organisms (ten animals, one fungus, eleven plants). For both variables, the distributions were characterized by dividing the values from the center of each chromosome by the corresponding value from the entire chromosome. A value of 0.5 indicates the absence of any bias (dashed gray line). For each organism, the chromosome-specific bias in gene density (black circles) and in total transcript length (gray circles) is given, along with the median and 95% bootstrap CI across all autosomes to the right of them. In contrast to animals and the fungus, plants exhibit a systematic bias in both the density of genes (grand mean of all plant species: 0.352, 95% bootstrap CI: 0.228-0.430) and total transcript length (grand mean: 0.364, CI: 0.279-0.442) across chromosomes. This bias, however, is opposite to what is expected if a heterogeneous distribution of selection targets along chromosomes was instrumental for CCBD, because chromosome peripheries – and not chromosome centers – exhibit an increased density of potential selection targets. (**b**) Simulations exploring how a bias in the physical distribution of *SL* influences CCBD (black curves) and overall (mean) differentiation (gray curves), shown for *SL* (dashed) and *NL* (solid). The density of *SL* was either biased toward the chromosome's center (center-biased *SL*) or peripheries (periphery-biased *SL*). To assess the consequences of a biased distribution of *SL* independently from heterogeneity in crossover rate, simulations were run with both heterogeneous (*cbias* = 0.9) and uniform (*cbias* = 0.5) crossover rate across the chromosome. Although a substantial bias in the distribution of the *SL* was modeled (i.e., the high-density chromosome region harbored 67% more *SL* than the low-density region), the consequences on population differentiation remain relatively trivial, emphasizing that CCBD must primarily reflect heterogeneity in the rate of crossover.

Figure S12

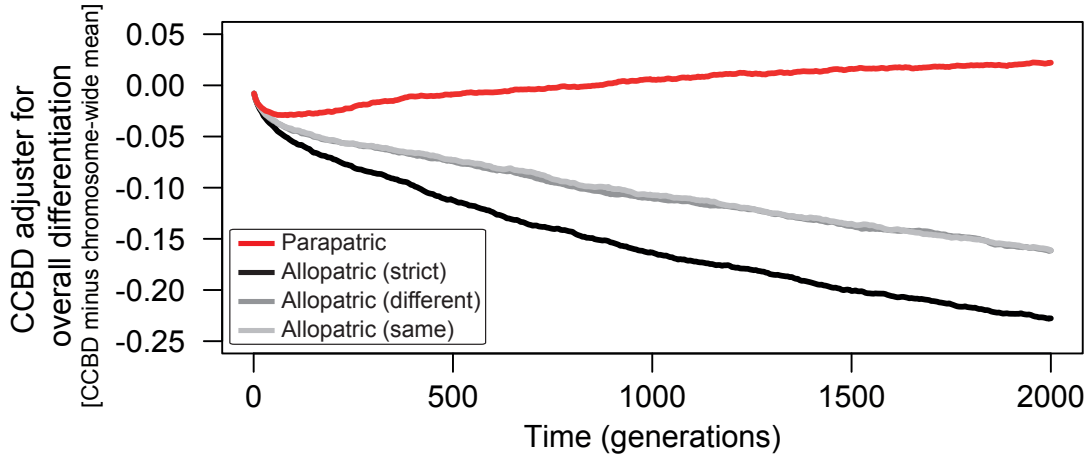


Figure S12. CCBD adjusted for overall (mean) differentiation in different population contrasts. Our simulations of adaptive population divergence with (Fig. 2) versus without (Fig. S9) gene flow indicated that the magnitude of CCBD in relation to overall differentiation should be indicative whether or not divergence has occurred in the face of gene flow. We thus calculated 'mean-adjusted CCBD'. For this, we subtracted from the magnitude of CCBD (i.e., mean differentiation in the center minus mean differentiation in the periphery) the overall (mean) differentiation across all loci. We considered four different scenarios: (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, non-focal population occupying a selectively different habitat ['Allopatric, (different)']. These scenarios are illustrated in Fig. 5a and Fig. S9. CCBD adjusted for overall differentiation is dramatically higher between populations diverging with than without gene flow.

Figure S13

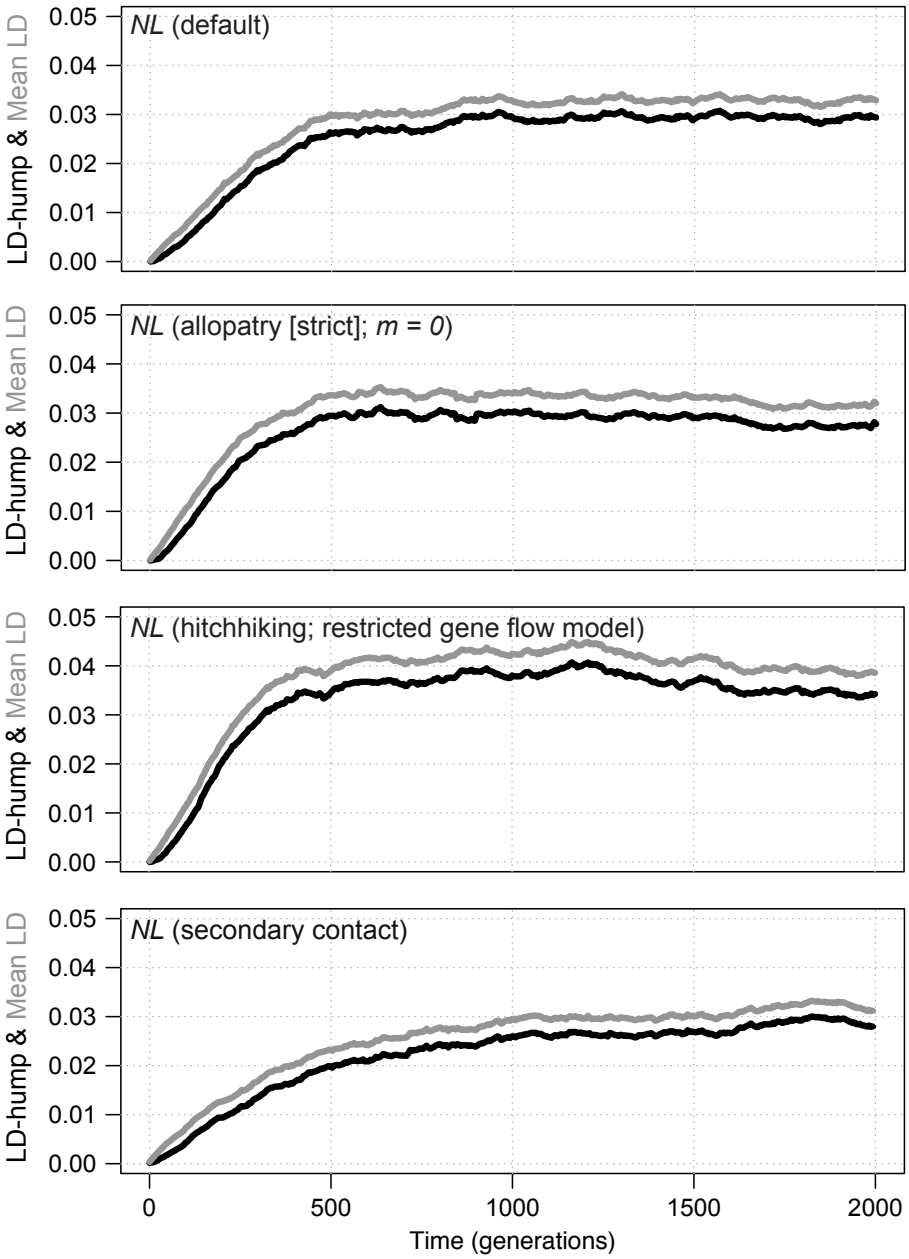


Figure S13. Establishment of linkage disequilibrium along the chromosome over time. Average magnitude of LD across the chromosome (gray line), and the difference in LD between the chromosome center and the peripheries ('LD-hump'; black curve), during evolution in the standard, the allopatry (strict; see Fig. 5a), the restricted gene flow, and the secondary contact models. LD is expressed as mean pairwise R^2 within a population between a focal locus and its two flanking loci on either side, considering only the *NL*. The plots are based on 25 replicate simulations, integrating data from both populations per replicate. Note that in all models (hence consistent across different migration contexts), patterns of LD consistently emerge mostly during the initial phase of substantial allele frequency changes at the selected loci (Fig. 2a), indicating that LD is caused primarily by hitchhiking.

Figure S14

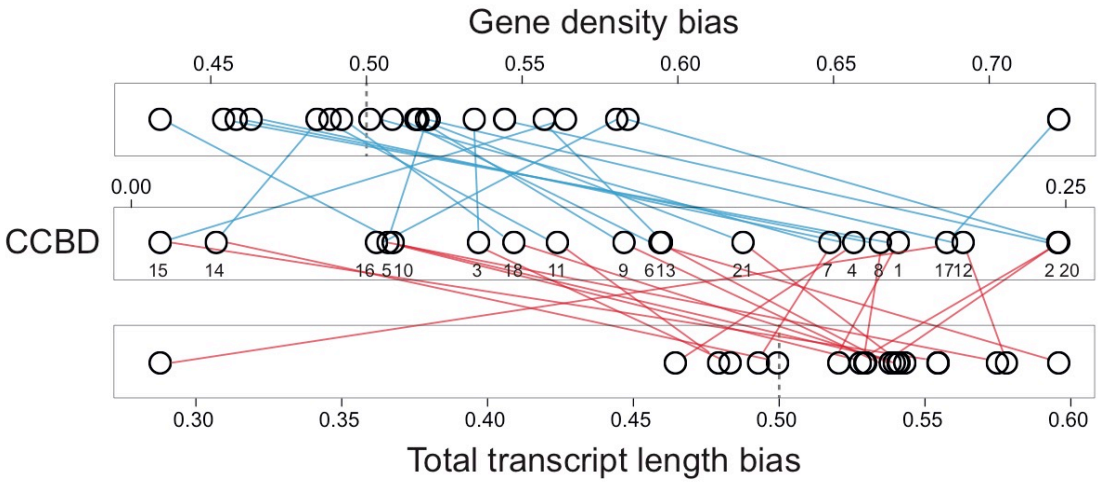
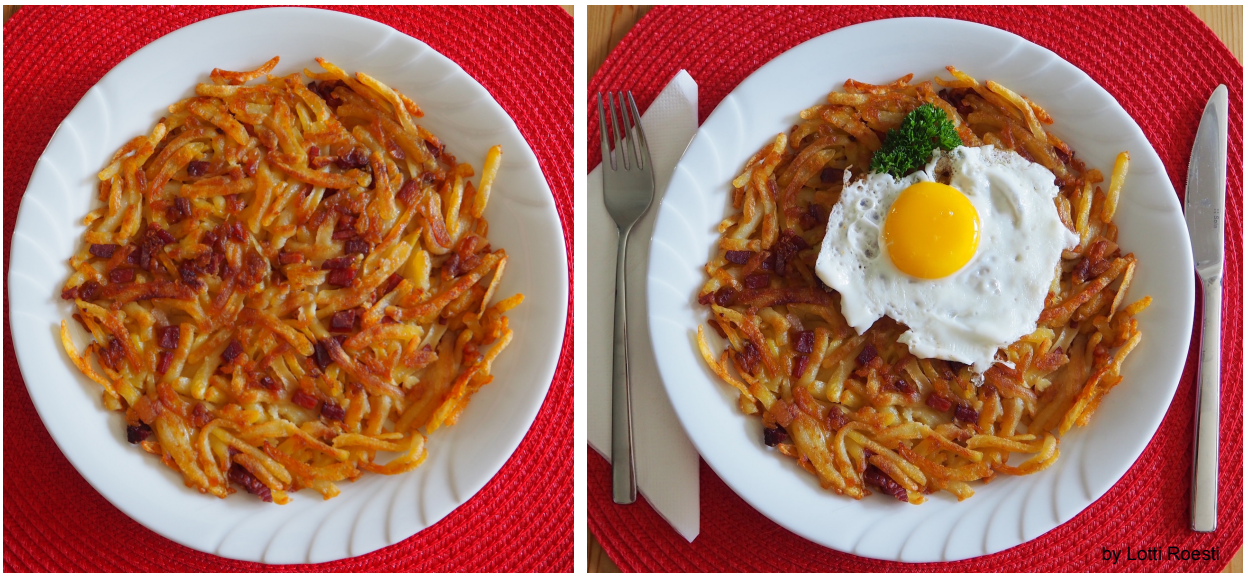


Figure S14. Bias in the distribution of potential selection targets across chromosomes does not explain magnitude of CCBD in the stickleback genome. Gene density bias reflects the number of genes in the center divided by the total gene number of a chromosome for all 20 stickleback autosomes. Accordingly, total transcript length bias was calculated by dividing the cumulative length of all transcripts in an autosome's center by its total transcript length. The dashed gray vertical lines represent the expected value (0.5) in the absence of any bias between a chromosome's center versus its peripheries. The small numbers in the central panel identify the chromosomes. We find that the extent of CCBD, here calculated for two parapatric lake and stream stickleback populations (Boot drainage; see Fig. 8), neither correlates with the magnitude in biased gene density (Spearman rank correlation coefficient = 0.044, P = 0.856) nor biased transcript length (Spearman rank correlation coefficient = -0.203, P = 0.394). This finding opposes the strong association found between the bias in the rate of crossover and CCBD (see Fig. 8a).

Figure S15



Recipe for
Berner Roesti – a Swiss national dish

Ingredients (4 servings):

- 800 g potatoes (waxy), boil not too soft the day before
- 120 g bacon, diced
- 1 teaspoon salt
- 3 tablespoons butter (lard or vegetable oil works too)

1. Grate the potatoes coarsely, mix with bacon and salt. Let rest for some minutes
2. Melt butter in a frying pan, add the potato mix from above and stir well to mix the fat and the potatoes
3. Fry at medium heat by turning repeatedly (if you feel confident, you can flip the Berner Roesti in the air) until a crispy, golden crust has formed. The final Berner Roesti can be cut almost like a cake. Enjoy!

Methods S1

We examined to what degree CCBD could be influenced by broad-scale heterogeneity in the distribution of selection targets along chromosomes. The motivation was that if organisms consistently displayed a higher density of selection targets in chromosome centers relative to the peripheries, this could provide an alternative explanation for CCBD: even with a uniform crossover rate along a chromosome, central regions relatively enriched in targets of selection would be affected by selection more strongly than target-poor chromosome peripheries. To evaluate this possibility, we used two complementary analyses.

First, we assessed whether chromosomes of natural organisms generally display a higher density of genes or a greater total transcript length – two proxies for the density of selection targets – in chromosome centers than peripheries. For this, we haphazardly selected 22 organisms with well-assembled and annotated genomes from two genome browsers: ten animal and one fungus species from ENSEMBL (BioMart) and eleven plant species from Phytozome v11.0. For these species, we retrieved all unique genes and transcripts for each chromosome and expressed bias in the density of genes and in the total transcript length as the ratio of the number of genes and the total length of all transcripts in the center (i.e., central 50% of each chromosome) relative to the entire chromosome. We then calculated the median of both metrics across all autosomes (sex-chromosomes were excluded), and the respective 95% bootstrap CI by resampling the autosome-specific values 10,000 times (all non-parametric CI estimation in this study followed this scheme).

Second, we performed a theoretical investigation by implementing a '*selection target density* model' (Table S4), another modification of the standard model in which we increased the density of *SL* by 67% in the chromosome center relative to the peripheries, or vice versa (i.e., the high-density chromosome region harbored ten and the low-density region only six *SL*). The *selection target density* model was run with the usual default parameter values. However, to appreciate the effect of heterogeneous *SL* density independently from heterogeneity in crossover rate, we applied this model with both heterogeneous and uniform crossover rate along the chromosome.

Methods S2

We validated key patterns emerging from our theoretical investigation empirically. For this, we used genome-wide sequence data available for three lake-stream stickleback population pairs (Roesti et al. 2012a). These pairs evolved postglacially, and thus in less than 10,000 generations, in independent drainages on Vancouver Island (BC, Canada; the Boot, Robert's and Joe's drainages). Within each pair, the two populations are in direct contact and diverge into selectively different lake and stream habitats 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,b).

A first set of analyses focused on SNP data from the Boot lake-stream population pair. Polymorphisms were filtered for a pooled minor allele frequency of at least 0.2 to ensure high marker information content (Roesti et al. 2012b), and by considering only a single marker per RAD locus (Roesti et al. 2012a). The remaining 6,506 genome-wide SNPs were then used to calculate F_{ST} between the lake and the stream population.

First, we were interested whether the magnitude of CCBD between these parapatric lake-stream stickleback reflect heterogeneity in crossover rate across chromosomes, or whether CCBD could be explained by heterogeneity in the density of available selection targets, again quantified as both gene density and transcript length. We thus calculated CCBD between Boot lake and stream stickleback following our standard rationale by subtracting mean differentiation (based on F_{ST}) in the peripheries from mean differentiation in the center for each chromosome. To assess chromosome-scale bias in crossover rate, we re-used previously published crossover rate estimates for intervals between adjacent SNP-markers (N = 1,872) across the stickleback genome (Roesti et al. 2013). To calculate average crossover rate in non-overlapping 1 Mb windows across the genome, each SNP-marker interval contributed proportionally to its physical overlap with a genomic window to the calculation of the mean crossover rate within that window. We then expressed crossover bias by dividing the mean crossover rate across all central windows by the mean of all peripheral windows within each chromosome. Bias of potential selection targets along each chromosome was quantified by following the general protocol also applied to the other organisms (see above). To allow for comparison among the different data sets, all data were matched to the improved assembly (Roesti et al. 2013) of the stickleback reference genome (Jones et al. 2012). We finally used correlation tests to quantify the association between crossover bias and CCBD, and between bias in the density of potential selection targets and CCBD across all autosomes. Because one chromosome showed extremely strong heterogeneity in the density of selection targets, we report the results using the Spearman's rank correlation coefficient as test statistic, although using Pearson's *r* led to the same conclusions. We evaluated the magnitude of each test statistic against its empirical random distribution established by permuting the data 10,000 times (all P-values in this paper are two-tailed and were generated using permutation strategies; Manly 2007). For visualization, we plotted crossover bias against CCBD for each autosome, and added the linear regression with its 95% confidence limits.

Second, we visualized the F_{ST} frequency distribution between Boot lake and stream stickleback, and indicated for each frequency bin the proportion of F_{ST} values from the chromosome center and the peripheries. The same data were also subjected to a search for putative loci under selection with *BayeScan* (Foll & Gaggiotti 2008; default settings were used).

Third, to investigate differences in phylogenetic structure inferred from polymorphisms from the chromosomes' centers versus peripheries, we first excluded from the raw SNP data set generated for Boot stickleback all individuals with < 75% diploid genotypes, and SNPs with > 25% missing data and/or a pooled minor allele frequency of < 0.05. The remaining SNPs were further filtered by accepting only a single polymorphism per RAD locus, and then split according to their chromosomal location (center versus periphery; N = 4,853 and 5,840 SNPs). The peripheral markers were sub-sampled to 4,853 to match the number of central ones, and for each SNP group, a separate neighbor-joining phylogeny was generated with *phangorn* (maximum likelihood methods with different models of sequence evolution always yielded comparable results). *FigTree* v1.4.2 was used to visualize the trees, and to calculate the absolute length of the branch separating the lake and stream population and the length of this branch relative to total tree size. These metrics here reflected the depth of the genealogical population separation more meaningfully than the gsi: due to the substantial overall genetic divergence of this lake-stream pair, the latter index was already at its maximum for both the central and peripheral tree.

Fourth, we estimated LD in chromosome centers and peripheries. For that, the SNP data from the Boot lake-stream pair were filtered as for the phylogenies. For each of the two SNP groups, R^2 was calculated for all pairwise marker combinations on each of the 20 autosomes within each population using the *R* package *LDcorSV*. Chromosome-specific mean values of LD were then combined to obtain a grand mean estimate (and its parametric 95% CI) for chromosome centers and peripheries across the genome. This procedure was carried out separately for SNPs 0-10 kb and 10-20 kb apart. The results were highly consistent across both populations; hence, we visualized the results for the lake population only.

In a second set of analyses, we tested whether mean-adjusted CCBD is more extreme between divergent populations evolved in the presence of gene flow (parapatry), as opposed to between populations evolved in isolation (allopatry). For this, we used the genome-wide RAD stickleback data from all three replicate drainages. To calculate F_{ST}, we filtered SNPs as described for the calculation F_{ST} of Boot lake-stream pair above. As an alternative differentiation measure, we calculated nucleotide divergence (D_{xy}) considering all pairwise consensus sequence comparisons per RAD locus between two populations. At least 20 haploid consensus sequences per population had to be present for a RAD locus to stay in this analysis. D_{xy} was then calculated using entire RAD reads (64 bp), separately for each locus. Genome-wide F_{ST} and D_{xy} values were then used to calculate mean-adjusted CCBD. This involved calculating, for every chromosome, the difference in the mean between central and peripheral differentiation values, and subtracting from this difference the chromosome-wide mean differentiation across all markers. Using all autosomes as data points, these chromosome-specific values were then averaged to a single index for each population comparison. In analogy to our theoretical analyses, these comparisons included the three parapatric lake-stream pairs within each drainage, all six allopatric combinations of populations occupying different habitats in different watersheds (i.e., Boot lake versus

Robert's stream), and all six allopatric combinations of populations residing in the same habitat (i.e., lake-lake, stream-stream; see Fig. 5a). The average number of F_{ST} and D_{xy} values in these pairwise comparisons ranged from 5,131 to 9,718 (mean = 8,289) and from 29,622 to 69,798 (mean = 37,720), respectively.

References

Akhunov ED, Goodyear AW, Geng S*, et al.* (2003) The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. *Genome Res* **13**, 753–763.

Anderson LK, Doyle GG, Brigham B*, et al.* (2003) High-resolution crossover maps for each bivalent of *Zea mays* using recombination nodules. *Genetics* **165**, 849–865.

Argyris JM, Ruiz-Herrera A, Madriz-Masis P*, et al.* (2015) Use of targeted SNP selection for an improved anchoring of the melon *(Cucumis melo L.)* scaffold genome assembly. *BMC Genomics* **16**, 4.

Auton A, Fledel-Alon A, Pfeifer S*, et al.* (2012) A fine-scale chimpanzee genetic map from population sequencing. *Science* **336**, 193–198.

Backström N, Forstmeier W, Schielzeth H*, et al.* (2010) The recombination landscape of the zebra finch *Taeniopygia guttata* genome. *Genome Res* **20**, 485–495.

Barton AB, Pekosz MR, Kurvathi RS, Kaback DB (2008) Meiotic recombination at the ends of chromosomes in *Saccharomyces cerevisiae*. *Genetics* **179**, 1221–1235.

Bauer E, Falque M, Walter H*, et al.* (2013) Intraspecific variation of recombination rate in maize. *Genome Biology* **14**, 1–17.

Bekele WA, Wieckhorst S, Friedt W, Snowdon RJ (2013) High-throughput genomics in sorghum: from whole-genome resequencing to a SNP screening array. *Plant Biotechnol J* **11**, 1112–1125.

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

Bhakta MS, Jones VA, Vallejos CE (2015) Punctuated distribution of recombination hotspots and demarcation of pericentromeric regions in *Phaseolus vulgaris* L. *PLoS One* **10**, e0116822.

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

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

Brunschwig H, Levi L, Ben-David E*, et al.* (2012) Fine-scale maps of recombination rates and hotspots in the mouse genome. *Genetics* **191**, 757–764.

Choi K, Zhao X, Kelly KA*, et al.* (2013) Arabidopsis meiotic crossover hot spots overlap with H2A.Z nucleosomes at gene promoters. *Nat Genet* **45**, 1327–1336.

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

Croll D, Lendenmann MH, Stewart E, McDonald BA (2015) The Impact of recombination hotspots on genome evolution of a fungal plant pathogen. *Genetics* **201**, 1213–1228.

Davey JW, Chouteau M, Barker SL*, et al.* (2016) Major improvements to the *Heliconius melpomene* genome assembly used to confirm 10 chromosome fusion events in 6 million years of butterfly evolution. *G3* **6**, 695–708.

Dohm JC, Lange C, Holtgrawe D*, et al.* (2012) Palaeohexaploid ancestry for Caryophyllales inferred from extensive gene-based physical and genetic mapping of the sugar beet genome *(Beta vulgaris). Plant J* **70**, 528–540.

Dukic M, Berner D, Roesti M, Haag CR, Ebert D (2016) A high-density genetic map reveals variation in

recombination rate across the genome of *Daphnia magna*. *BMC Genet* **17**, 137.

Endelman JB, Jansky SH (2016) Genetic mapping with an inbred line-derived F2 population in potato. *Theor Appl Genet* **129**, 935–943.

Fledel-Alon A, Wilson DJ, Broman KW*, et al.* (2009) Broad-scale recombination patterns underlying proper disjunction in humans. *PLoS Genet* **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.

Gardner KA, Wittern LM, Mackay IJ (2016) A highly recombined, high-density, eight-founder wheat MAGIC map reveals extensive segregation distortion and genomic locations of introgression segments. *Plant Biotechnol J* **14**, 1406–1417.

Glazer AM, Killingbeck EE, Mitros T, Rokhsar DS, Miller CT (2015) Genome assembly improvement and mapping convergently evolved skeletal traits in sticklebacks with genotyping-by-sequencing. *G3* **5**, 1463.

Groenen MA, Wahlberg P, Foglio M*, et al.* (2009) A high-density SNP-based linkage map of the chicken genome reveals sequence features correlated with recombination rate. *Genome Res* **19**, 510–519.

Huo N, Garvin DF, You FM*, et al.* (2011) Comparison of a high-density genetic linkage map to genome features in the model grass *Brachypodium distachyon*. *Theor Appl Genet* **123**, 455–464.

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

Jones FC, Grabherr MG, Chan YF*, et al.* (2012) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55–61.

Juneja P, Osei-Poku J, Ho YS*, et al.* (2014) Assembly of the genome of the disease vector *Aedes aegypti* onto a genetic linkage map allows mapping of genes affecting disease transmission. *PLoS Negl Trop Dis* **8**, e2652.

Kawakami T, Smeds L, Backström N*, et al.* (2014) A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. *Mol Ecol* **23**, 4035–4058.

Kong A, Gudbjartsson DF, Sainz J*, et al.* (2002) A high-resolution recombination map of the human genome. *Nat Genet* **31**, 241–247.

Li G, Hillier LW, Grahn RA, *et al.* (2016) A high-resolution SNP array-based linkage map anchors a new domestic cat draft genome assembly and provides detailed patterns of recombination. G3 **6**, 1607–1616.

Li X, Li L, Yan J (2015) Dissecting meiotic recombination based on tetrad analysis by single-microspore sequencing in maize. *Nat Commun* **6**, 6648.

Liu Z, Hillier LW, Grahn RA, *et al.* (2016) The channel catfish genome sequence provides insights into the evolution of scale formation in teleosts. *Nat Commun* **7**, 11757.

Liu H, Zhang X, Huang J*, et al.* (2015) Causes and consequences of crossing-over evidenced via a high-resolution recombinational landscape of the honey bee. *Genome Biol* **16**, 15.

Mackay TF, Richards S, Stone EA*, et al.* (2012) The *Drosophila melanogaster* Genetic Reference Panel. *Nature* **482**, 173–178.

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci USA* **76**, 5269–5273.

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

Niehuis O, Gibson JD, Rosenberg MS*, et al.* (2010) Recombination and its impact on the genome of the haplodiploid parasitoid wasp *Nasonia*. *PLoS One* **5**, e8597.

Paape T, Zhou P, Branca A*, et al.* (2012) Fine-scale population recombination rates, hotspots, and correlates of recombination in the *Medicago truncatula* genome. *Genome Biol Evol* **4**, 726–737.

Rastas P, Calboli FCF, Guo B, Shikano T, Merilä J (2015) Construction of ultra-dense linkage maps with Lep-MAP2: stickleback F2 recombinant crosses as an example. *Genome Biol Evol* **8**, 78–93.

Ren Y, Zhao H, Kou Q*, et al.* (2012) A High Resolution Genetic Map Anchoring Scaffolds of the Sequenced Watermelon Genome. *PLoS One* **7**, e29453.

Renaut S, Grassa CJ, Yeaman S*, et al.* (2013) Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nat Commun* **4**, 1827.

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

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

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

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

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

Ross JA, Koboldt DC, Staisch JE*, et al.* (2011) *Caenorhabditis briggsae* recombinant inbred line genotypes reveal inter-strain incompatibility and the evolution of recombination. *PLoS Genet* **7**, e1002174.

Salomé PA, Bomblies K, Fitz J*, et al.* (2012) The recombination landscape in *Arabidopsis thaliana* F2 populations. *Heredity* **108**, 447–455.

Sandor C, Li W, Coppieters W*, et al.* (2012) Genetic variants in REC8, RNF212, and PRDM9 influence male recombination in cattle. *PLoS Genet* **8**, e1002854.

Schmutz J, Cannon SB, Schlueter J*, et al.* (2010) Genome sequence of the palaeopolyploid soybean. *Nature* **463**, 178–183.

Singer T, Fan Y, Chang H-S*, et al.* (2005) A high-resolution map of Arabidopsis recombinant inbred lines by whole-genome array hybridization. *PLoS Genet* **2**, e144.

Solignac M, Mougel F, Vautrin D, Monnerot M, Cornuet JM (2007) A third-generation microsatellite-based linkage map of the honey bee, *Apis mellifera,* and its comparison with the sequence-based physical map. *Genome Biol* **8**, R66.

Swaminathan K, Chae WB, Mitros T*, et al.* (2012) A framework genetic map for *Miscanthus sinensis* from RNAseq-based markers shows recent tetraploidy. *BMC Genomics* **13**, 1–17.

Tapper W, Collins A, Gibson J*, et al.* (2005) A map of the human genome in linkage disequilibrium units. *Proc Natl Acad Sci USA* **102**, 11835–11839.

Tian Z, Rizzon C, Du J*, et al.* (2009) Do genetic recombination and gene density shape the pattern of

DNA elimination in rice long terminal repeat retrotransposons? *Genome Res* **19**, 2221–2230.

Tine M, Kuhl H, Gagnaire PA*, et al.* (2014) European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nat Commun* **5**, 5770.

Tomato Genome C (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* **485**, 635–641.

Tortereau F, Servin B, Frantz L*, et al.* (2012) A high density recombination map of the pig reveals a correlation between sex-specific recombination and GC content. *BMC Genomics* **13**, 586.

Tsai HY, Robledo D, Lowe NR*, et al.* (2016) Construction and annotation of a high density SNP linkage map of the atlantic salmon *(Salmo salar)* genome. *G3* **6**, 2173–2179.

Wang X, Yu K, Li H*, et al.* (2015) High-Density SNP Map Construction and QTL Identification for the Apetalous Character in *Brassica napus* L. *Front Plant Sci* **6**, 1164.

Wong AK, Ruhe AL, Dumont BL*, et al.* (2010) A comprehensive linkage map of the dog genome. *Genetics* 184, 595–436.