Selection and gene flow shape genomic islands that control floral guides

Hugo Tavares1,a, Annabel Whibley2, David L. Field2,c, Desmond Bradley3, Matthew Couchman3, Lucy Copsey3, Joane Elleouet4, Monique Burrous5, Christophe Andalo6, Miaomiao Li7,e,9, Qun Li7,f, Yongbiao Xuee,f,g,h, Alexandra B. Rebocho5, Nicolas H. Bartonb,2, and Enrico Coen1,2

1Department of Cell and Developmental Biology, John Innes Centre, NR4 7UH Norwich NR4 7UH, United Kingdom; 2Institute of Science and Technology Austria, 3400 Klosterneuburg, Austria; 3Department of Botany and Biodiversity Research, Faculty of Life Sciences, University of Vienna, A-1030 Vienna, Austria; 4Laboratoire Évolution et Diversité Biologique, UMR 5174 CNRS–Université Paul Sabatier, 31062 Toulouse Cédex 9, France; 5State Key Laboratory of Molecular Developmental Biology, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, 100101 Beijing, China; 6National Center for Plant Gene Research, Chinese Academy of Sciences, 100101 Beijing, China 7School of Life Sciences, University of Chinese Academy of Sciences, 100190 Beijing, China; and 8Department of Genetics, Chinese Academy of Sciences, 100101 Beijing, China

Edited by Nils Chr. Stenseth, University of Oslo, Oslo, Norway, and approved September 12, 2018 (received for review February 6, 2018)

Genomes of closely-related species or populations often display localized regions of enhanced relative sequence divergence, termed genomic islands. It has been proposed that these islands arise through selective sweeps and/or barriers to gene flow. Here, we genetically dissect a genomic island that controls flower color pattern differences between two subspecies of Antirrhinum majus, A.m.striatum and A.m.pseudomajus, and relate it to clinal variation across a natural hybrid zone. We show that selective sweeps likely raised relative divergence at two tightly-linked MYB-like transcription factors, leading to distinct flower patterns in the two subspecies. The two patterns provide alternate floral guides and create a strong barrier to gene flow where populations come into contact. This barrier affects the selected flower color genes and tightly-linked loci, but does not extend outside of this domain, allowing gene flow to lower relative divergence for the rest of the chromosome. Thus, both selective sweeps and barriers to gene flow play a role in shaping genomic islands: sweeps cause elevation in relative divergence, while heterogeneous gene flow flattens the surrounding “sea,” making the island of divergence stand out. By showing how selective sweeps establish alternative adaptive phenotypes that lead to barriers to gene flow, our study sheds light on possible mechanisms leading to reproductive isolation and speciation.

Genome scans of closely-related species or populations have revealed “genomic islands” as peaks of high relative sequence divergence (Fₛ) that stand out against a lower “sea” of divergence (1–5). The causes of genomic islands remain unclear, but they have been suggested to contain key loci involved in local adaption and/or reproductive isolation (6). However, their significance for speciation with or without gene flow between populations is a matter of debate (6–9). One hypothesis is that gene flow is unimpeded across most of the genome, reducing between-population diversity, except for loci under divergent selection and loci in close physical linkage to selected loci (8). Another hypothesis is that genomic islands reflect selective sweeps, where specific alleles are raised relative to divergence for the rest of the chromosome. Thus, both selective sweeps and barriers to gene flow play a role in shaping genomic islands: sweeps cause elevation in relative divergence, while heterogeneous gene flow flattens the surrounding “sea,” making the island of divergence stand out. By showing how selective sweeps establish alternative adaptive phenotypes that lead to barriers to gene flow, our study sheds light on possible mechanisms leading to reproductive isolation and speciation.

Significance

Populations often show “islands of divergence” in the genome. Analysis of divergence between subspecies of Antirrhinum that differ in flower color patterns shows that sharp peaks in relative divergence occur at two causal loci. The island is shaped by a combination of gene flow and multiple selective sweeps, showing how divergence and barriers between populations can arise and be maintained.

Author contributions: H.T., A.W., D.L.F., N.H.B., and E.C. designed research; H.T., A.W., D.L.F., D.B., M.C., L.C., J.E., and A.B.R. performed research; H.T., A.W., D.L.F., D.B., M.C., L.C., M.B., C.A., M.L., Q.L., Y.X., and N.H.B. contributed new reagents/analytic tools; H.T., A.W., D.L.F., D.B., M.C., M.L., N.H.B., and E.C. analyzed data; and H.T., A.W., N.H.B., and E.C. wrote the paper. The authors declare no conflict of interest.

This article is an open access article distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

Data deposition: The genomic sequence data reported in this paper are available at European Nucleotide Archive (ENA), https://www.ebi.ac.uk/ena (accession no. ENA PRJEB28287), and the RNAseq data have been deposited in the Gene Expression Omnibus (GEO) database, https://www.ncbi.nlm.nih.gov/geo (accession no. GSE118621).

1Present address: Sainsbury Laboratory, University of Cambridge, Cambridge, United Kingdom.

2To whom correspondence may be addressed. Email: Nick.Barton@list.ac.uk or enrico.coen@jic.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1801832115/-/DCSupplemental.

Published online October 8, 2018.
EL is tightly linked to ROS but has not been previously isolated (11, 14). Selection at ROS has been inferred from analysis of a hybrid zone between *A. m. striatum* and *A. m. pseudomajus*: both magenta pigmentation and ROS allele frequencies show sharp clines, ∼1 km wide, whereas markers >5 CM from ROS show more uniform allele frequency distributions (11).

Flower color differences between *A. m. striatum* and *A. m. pseudomajus* are unlikely to be maintained by adaptation to local conditions, as there are no clear differences in environment or pollinators across the hybrid zone (16). Rather, hybrids and recombinants may be selected against because their flower patterns are less effective as signposts for bee entry than the parental patterns (12, 17). In these cases (27), selection at ROS is likely due to fixation of one or more favorable mutations (selective sweeps). The right *F*ₚ peak, ~150 kb downstream of ROS, is also primarily due to a decrease in πₑ (lower green points, Fig. 3A). πₑ is reduced in both populations, for both the left and right peaks, indicating at least four sweeps (i.e., two for each of the two populations). By contrast, the middle peak does not have low πₑ but, rather, relatively high πₑ (light blue points, Fig. 3A). The middle peak is absent or reduced in some population comparisons (detailed below), suggesting that selective sweeps were not involved in generating it. The above results thus indicate that only the left and right *F*₂ peaks arose through selective sweeps.

**Mapping the Causal Loci.** To determine whether the regions subject to selective sweeps had phenotypic effects, we introgressed *ros EL* from *A. m. striatum* into *A. m. majus* (*ROS el*) and genotyped F2 populations. Recombinants were backcrossed or self-pollinated to determine their homozygous phenotypes (Fig. 4B–F). Regions causing the ROS phenotype mapped to the left *F*₂ peak, while the EL phenotype mapped to the middle and/or right *F*₂ peaks. The limits of the *ROS* and *EL* regions were further refined by crossing plants heterozygous for *ros EL* (from *A. m. striatum*) and *ROS el* (from *A. m. pseudomajus* or *A. m. majus*) to a *ros El/El* line. Screening 10,261 progeny yielded 26 *ROS EL* recombinants, mapping *EL* to an interval of ~50 kb (Fig. 4G), below the right *F*₂ peak. The map distance between *ROS* and *EL* was 0.5 CM, corresponding to ~3 CM/Mbp, which is of the
same order as the genome-wide average of 1.8 cM/Mbp. No phe- notypic effect mapped to the middle $F_a$ peak.

To determine whether the flower color phenotypes reflect variation in gene expression levels, we performed RNAseq on flower buds from homozygous progeny of individuals used in the genetic mapping experiments. Two of fifteen genes detected in the ROS-EL region showed highly significant expression differences (Fig. 4F, $q < 0.001$; SI Appendix, Fig. S6). One transcript derived from ROS1 and was about 10 times more abundant for samples with a dominant ROS allele compared with those with recessive $ros$, consistent with ROS conferring strong magenta. The second differential transcript encoded a MYB-like transcription factor with 57% protein identity to ROS1 (SI Appendix, Fig. S6). Two $EL$-MYB was expressed about threefold more in samples with a dominant $EL$ allele compared with those with recessive $el$, consistent with it being a repressor of magenta pigmentation (SI Appendix, Fig. S6C). These results indicate that $EL$ encodes a MYB-like transcription factor and show that at least some of the differences in gene activity are transcriptional. The $EL$-MYB gene maps to the rightmost $F_a$ peak (Fig. 4A). Two other transcripts showed differences in expression between el and $EL$ genotypes (genes 5 and 14, Fig. 4F, $q < 0.01$, $q < 0.05$, respectively) but showed a much weaker correlation with genotype than the $EL$-MYB gene (SI Appendix, Fig. S6 B and C).

We also analyzed recombinants, termed $ROS1^*$, with breakpoints just downstream of the $ROS1$ gene (Fig. 4H). $ROS1^*$ is expressed at a similar level to $A.m.pseudomajus$ $ROS1$, although it carries the $ROS1$ coding and upstream region of $A.m.pseudomajus$ (SI Appendix, Fig. S6C). Thus, variation in $ROS1$ transcript levels largely maps to a downstream enhancer. The paler flowers of $A.m.pseudomajus$ $ROS1^*$ compared with $A.m.pseudomajus$ $ROS1$ (Fig. 1E vs. Fig. 1H) suggests that variation in the coding region also contributes to the phenotype. Taken together with the observation of low $\pi_w$ for only the left and right $F_a$ peaks, these findings suggest that selective sweeps at $ROS$ and $EL$ caused these $F_a$ peaks.

**Gene Flow Lowers $F_a$ Outside the ROS/EL Region.** Sequence pools for populations of $A.m.pseudomajus$ and $A.m.striatum$ away from the center of the hybrid zone ($\sim 20$ km apart instead of $\sim 2.5$ km) showed a higher median $F_a$ ($0.048 \pm 0.0008$ compared with $0.040 \pm 0.0004$) and more variable profile for chromosome 6 than for nearby populations (Figs. 2H and I, 3B, and 5). By contrast, $F_a$ values at $ROS$, $EL$, and the intervening region were similar to
Comparison of within- and between-population divergence in the ROS/EL region. Relationship between \(F_{st}\) and \(p_{st}\) for pools sampled either side of the hybrid zone, separated by \(\sim 2.5\) km (A, YP1 and MP2, corresponding to Fig. 2 B and C) or \(\sim 20\) km (B, YP4 and MP11, corresponding to Fig. 2 H and I). Summarized in 10-kb windows, with a color gradient indicating the respective \(F_{st}\) (light colors, low; dark colors, high). The left, middle, and right \(F_{st}\) peaks indicated in Fig. 2C are shown as red, light blue, and green points, respectively. The dark blue points indicate windows between those \(F_{st}\) peaks. Other windows from around the ROS region are shown in gray.

those for the nearby populations (Figs. 2 H and I and 5). More remote populations showed a further increase in \(F_{st}\) for chromosome 6, with some comparisons yielding numerous \(F_{st}\) peaks, so that those at ROS and EL no longer stood out (Figs. 2 J and K and 5 and SI Appendix, Fig. S3 A and D and Table S9). Such a pattern of “isolation by distance” is often seen and indicates that gene flow reduces local divergence. In contrast, \(F_{st}\) is elevated across the whole ROS/EL region (Fig. 5), as expected from a strong barrier to gene flow generated by selection on ROS and EL (28). The statistical significance of these patterns is considered in SI Appendix, Supplementary Text S1.3.

A barrier to gene flow is also expected to cause sharp clines at any loci within it, regardless of whether they are selected. Indeed, we observe sharp clines at all divergent SNPs within or near the genomic islands, including those that lie outside ROS or EL (Fig. 2L and SI Appendix, Supplementary Text S2 and Fig. S7). Of the \(\sim 6 \times 10^5\) biallelic SNPs on chromosome 6, 115 showed frequency differences greater than 0.8 between the outer pools (\(\sim 20\) km apart). One hundred and one of these differential SNPs were within \(\sim 0.5\) Mbp ROS/EL region (Fig. 2M and SI Appendix, Fig. S3C), 14 of which were within the ROS and EL \(F_{st}\) peaks. 74 were between these peaks, and 13 were in flanking regions. Comparing SNP allele frequencies in the pools showed that the 14 differential SNPs within the ROS and EL \(F_{st}\) peaks, together with most of the 74 SNPs from the intervening region, exhibited clines centered at the hybrid zone (Fig. 2 N and O), confirmed and further refined by individual genotyping (Fig. 2L and SI Appendix, Fig. S7). The remaining differential SNPs, including 14 that were distributed sparsely along the chromosome (Fig. 2M), mainly showed a frequency change over a geographic region where the population density is low (Fig. 2P and SI Appendix, Fig. S7C). The change in frequency for these SNPs likely reflects fluctuations caused by the reduced gene flow created by the population density gap.

These findings support the hypothesis of a selective barrier at the ROS/EL region. The yellow flower patterning gene SULF exhibits steep SNP clines centered at the same geographically location as ROS-EL clines (12), supporting the idea that selection on flower color is the basis of the barrier.

Based on the 0.5-cM distance between ROS and EL, recombinants should be generated at hybrid zones, at a rate of 0.5% per heterozygote. Genotyping 2,393 individuals at the hybrid zone, using haplotype-specific markers in ROSI and EL, identified 201 recombinant haplotypes, which reached \(\sim 10\%\) frequency at the center of the hybrid zone (Fig. 4 J and K). Genotyping and test-crossing of progeny grown from 27 recombinants confirmed that most gave the expected phenotypes (SI Appendix, Supplementary Text S3).

Assuming a neutral model with no selection against recombinants, we estimated a lower bound of \(\sim 85\) generations for the age of this hybrid zone (SI Appendix, Supplementary Text S4). If the hybrid zone is older than this, then selection must have acted to eliminate recombinants. A
Simulations of gene flow and selective sweeps. Combined effects of $\pi$ and $F_{st}$ indicated by pale red (Supplementary Text S1.3) and (red, green) sweep through the separate populations, $G$ indicated by green. Note that $EL$ has decreased due to $C$ outside $EL$ peaks. By $EL$, $\pi$ and $F$ and $= 0.2$ would be generated, with contact being made $N$ observed in pools YP1, MP2, 2.5 km $H$ and is estimated as roughly 8.3 to be readily alleles becoming fixed in (Supplementary Text S1.6) Taken together, the clines, genetic analysis, transcrip-tion, divergence between populations at different geographic peaks. This argument illustrates the value of having two linked loci (pink). For each boxplot: the horizontal waistline indicates the median, the point indicates the mean, the length of the box indicates the interquartile range, and the whiskers extend to the data minima and maxima. For each genomic region, three $A.m.striatum$/$A.m.pseudomajus$ comparisons are shown, separated by 2.5 km (YP1 and MP2), 20 km (YP4 and MP11), or 100 km (ML-CN). Distributions are based on values calculated for 10-kb windows, 1-kb step size. Windows overlying ROS and $EL$: midpoints 530-575 kb and 707-720 kb on ROS scaffold. Windows between ROS and $EL$: midpoints 576-706 kb on ROS scaffold.

Fig. 5. Relative divergence between populations at different geographic locations. Notched boxplots of $F_{st}$, for three genomic regions: chromosome 6 (gray, from position >35Mb excluding the ROS/EL region), interval between ROS and $EL$ (blue), and the ROS and $EL$ loci (pink). For each boxplot: the horizontal waistline indicates the median, the point indicates the mean, the length of the waist indicates the 95% confidence interval of the median, the box indicates the interquartile range, and the whiskers extend to the data minima and maxima. For each genomic region, three $A.m.striatum$/$A.m.pseudomajus$ comparisons are shown, separated by 2.5 km (YP1 and MP2), 20 km (YP4 and MP11), or 100 km (ML-CN). Distributions are based on values calculated for 10-kb windows, 1-kb step size. Windows overlying ROS and $EL$: midpoints 530-575 kb and 707-720 kb on ROS scaffold. Windows between ROS and $EL$: midpoints 576-706 kb on ROS scaffold.

The observation that flower color variation under selection derives from two closely-linked loci ($ROS$ and $EL$) seems to lend support to the idea that divergent loci tend to cluster because linkage hinders swamping of locally adapted alleles (5, 29). However, other pigment loci under selection (e.g., $SULF$) are unlinked to $ROS$ and $EL$, showing that tight linkage is not essential. Moreover, $ROS$ and $EL$ are both MYB-like transcription factors and so may be clustered due to gene duplication. Thus, clustering may not be due to selection for linkage (SI Appendix, Supplementary Text S1.6).

Role of Selective Sweeps and Barriers to Gene Flow in Generating Genomic Islands. Taken together, the clines, genetic analysis, transcriptional differences, and analysis of $F_{st}$ peaks indicate that the $ROS/EL$ genomic island and its surround have been shaped by two processes: (i) historic selective sweeps that led to different $ROS$ and $EL$ alleles becoming fixed in $A.m.pseudomajus$ and $A.m.striatum$ populations and (ii) selection against hybrid genotypes generated note attached to a herbarium specimen of $A.m.pseudomajus$ from 1928 (London Natural History Museum) describes extensive color variation under selection in $EL$.

Barriers to gene flow observed at $ROS/EL$ raise the ques-tion of whether this alone could be responsible for the $F_{st}$ peaks. According to this view, the drop in $F_{st}$ in the intervening region between the peaks would be due to gene flow. However, selection at two linked loci ($ROS$ and $EL$) generates a strong barrier to gene flow throughout the intervening region because two recombination events are required to transfer a neutral allele onto the opposite genetic background (SI Appendix, Supplementary Text S5 and Figs. S15 and S16). A barrier of this form would therefore not be expected to generate two separate sharp peaks in $F_{st}$ as is observed. Thus, the barrier to gene flow alone cannot be responsible for the two sharp $F_{st}$ peaks. This argument illustrates the value of having two linked loci for distinguishing hypotheses. A further advantage of having two linked loci is that it allows a region of elevated $F_{st}$ to be readily picked out because the barrier extends over 0.5 cM and >200 kb. Single selected loci would generate a barrier over a narrow region, which would be harder to detect.

Fig. 6. Simulations of gene flow and selective sweeps. Combined effects of a barrier to gene flow and selective sweeps on $F_{st}$ (Left) and on $\pi$ and $F_{is}$ (Right). (A and F) A homogeneous population is split by a geographic barrier. (B and G) Alleles at $ROS$ and $EL$ (red, green) sweep through the separate populations, reducing diversity, $\pi_{na}$, generating peaks in $F_{st}$. (C and H) Further sweeps occur at $ROS$ and $EL$, strengthening the $F_{st}$ peaks. By $t = 0.2 N_e$ generations, divergence has increased genome-wide, with $F_{st} = 0.05$. At this time, the divergent populations meet and exchange genes everywhere except between $ROS$ and $EL$. (D and I) By time 0.5 $N_e$, $F_{st}$ outside $ROS/EL$ has decreased due to mixing (Left, black), but has increased between $ROS$ and $EL$ (Left, blue). Although in this scenario, population contact was established at 0.2 $N_e$, similar final profiles for $F_{st}$ and $\pi_{na}$ would be generated, with contact being made earlier or later than this. (E and J) The $\pi_{na}$, $\pi_{na}$ observed in pools YP1, MP2, 2.5 km apart, with the maximum $F_{st}$ observed at $ROS$ indicated by pale red (E) or red (J), and at $EL$ indicated by green. Note that $N_e$ is estimated as roughly $8.3 \times 10^4$ (SI Appendix, Supplementary Text S1.3). For further details, see SI Appendix, Supplementary Text S1.5.
where *A. m. pseudomajus* and *A. m. striatum* populations meet, creating a local barrier to gene flow (28). We performed simulations to explore scenarios consistent with the data and modes of selection.

To provide constraints on simulations, we first estimated the age of the selective sweeps. Based on the residual diversity within the sharp peaks at *ROS* and *EL*, we estimated the date of the most recent sweeps to be ~90,000 generations ago (SI Appendix, Supplementary Text S1); this is an upper bound, since “soft sweeps” might not have eliminated all diversity. We also estimated the age of the barrier to gene flow. As detailed in SI Appendix, Supplementary Text S1, the time required for *Fₚ* in the *ROS/EL* interval to accumulate to the observed value of 0.125 is *T* = 0.5 *Nₑ ~ 45,000* generations (where *Nₑ* = effective population size). Thus, both analyses suggest that selection and a barrier to gene flow were established roughly *Nₑ ~ 10⁶* generations ago.

We assume that a homogeneous ancestral population is first split by a geographic barrier, allowing sweeps to occur independently in each population (Fig. 6 A and F, for simplicity assuming an initial *Fₚ* = 0.0). Geographic separation is a simple way of ensuring that alleles swept in one population do not sweep into the other, although other scenarios such as environmental heterogeneity are possible; the sequence data are also compatible with divergence in primary contact. Sweeps at *ROS* and *EL* (red, green in Fig. 6 B and C) reduce diversity, *πₜ* generating peaks in *Fₚ*. These sweeps presumably reflect the selective advantage of a change in flower color, compared with the ancestral phenotype in each population. Given that both populations underwent sweeps, the ancestral flower phenotype would have been different from both of the current phenotypes in *A. m. pseudomajus* or *A. m. striatum*. Further sweeps at *ROS* and *EL* strengthen the *Fₚ* peaks (Fig. 6 C and H). Unlike the simulations, in real populations, it is possible that global and/or local sweeps occur at many other genetic loci and spatial locations, in addition to *ROS* and *EL*, creating a more rugged *Fₚ* profile across the genome.

After a period of time (*0.2 Nₑ* generations in the simulation shown in Fig. 6), the divergent populations came into contact. Gene flow leads to a lowering of *Fₚ* from the chromosome-wide average, except at loci where a barrier has been established. We propose that a barrier to gene flow occurs for only a subset of swept loci: those for which epistatic interactions or frequency dependence maintain divergence. *ROS* and *EL* represent one such case, as their interactions, together with loci controlling yellow lead, to alternative floral guides. Other loci that underwent sweeps, but led to no incompatibility (presumably the majority of sweeps) would undergo gene flow, with the allele conferring higher overall fitness going to fixation in both populations. By time 0.5 *Nₑ*, *Fₚ* outside *ROS/EL* has decreased due to gene flow (gray), but has further increased between *ROS* and *EL* (blue) because of the local barrier to gene flow (Fig. 6 D and J). The resulting *Fₚ*, *πₜ*, and *πₑ* values are comparable to those observed (compare Fig. 6 D and I with Fig. 6 E and J).

According to the above scenario, selective sweeps led to fixation of different alleles in each population, and selection maintains a local barrier to gene flow. Multiple changes in alleles are involved, a reasonable assumption given these events occurred over a period of ~10⁶ generations, extending over glacial periods, during which populations and the environment were in a state of flux. Our analysis indicates that both selective sweeps and barriers to gene flow combine to shape genomic islands of differentiation. The barrier to gene flow at *ROS/EL* is insufficient to prevent exchange for much of the genome. However, if the barrier were more severe and applied to additional loci, it could prevent gene flow more completely, leading to speciation. The mechanisms that created the genomic islands may therefore represent partial steps toward reproductive isolation and speciation.

**Materials and Methods**

Full details of plant material, DNA extraction, genome sequence analysis, population genomics, genotyping, SNP analysis for geographic, and RNAseq analysis are given in SI Appendix, Materials and Methods. Details on inference from pairwise diversity and divergence, geographiccline analysis, and genotypic screens are given in SI Appendix. Genomic sequence datasets are available at European Nucleotide Archive (ENA) with accession numbers PRJEB22827, and RNAseq datasets are deposited in National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) with accession number GSE118621. Associated scripts are provided at linked public data repositories as detailed in SI Appendix, Materials and Methods, and further information on the hybrid zone is available at [www.anttspec.org](http://www.anttspec.org).

**Acknowledgments.** Many thanks to Christophe Thébaud for sharing his finding of the herbarium specimen referenced in the text. This work was supported by Biotechnology and Biological Sciences Research Council Grants BBS/E/J/000P9773 and BB/G0093251 (to E.C.), ERC Grant 201252 (to N.H.B.), and a PhD scholarship (to H.T.) from the Portuguese Foundation for Science and Technology (FCT), through the Human Potential Operating Programme (POPH) of the National Strategic Reference Framework (QREN), within the European Social Fund (Scholarship SFRH/BD/60892/2009). This research was supported in part by the Norwich Bioscience Institute Computing Infrastructure for Science (CiS) group.

1. Hohenlohe PA, et al. (2010) Population genomics of parallel adaptation in threespine sticklebacks using ultradiploid tags. *Proc Natl Acad Sci USA* 107:2267–2272.
2. Martin SH, et al. (2013) Genome-wide evidence for speciation with gene flow in Heliconius butterflies. *Genome Res* 23:1817–1828.
3. Poelstra JW, et al. (2014) The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science* 344:1410–1414.
4. Clarkson CS, et al. (2014) Adaptive introgression between Anopheles sibling species eliminates a major genomic island but not reproductive isolation. *Nat Commun* 5:4248.
5. Ellegren H, et al. (2012) The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature* 491:756–760.
6. Pennisi E (2014) Disputed islands. *Science* 345:611–613.
7. Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol Ecol* 23:3133–3157.
8. Ma T, et al. (2018) Recent genomic polymorphisms and divergence hitchhiking contribute to genomic islands of divergence within a poplar species complex. *Proc Natl Acad Sci USA* 115:E236–E243.
9. Wolz JB, Ellegren H (2017) Making sense of genomic islands of differentiation in light of speciation. *Nat Rev Genet* 17:87–100.
10. 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. *Genet Res* 70:155–174.
11. Whibley AC, et al. (2006) Evolutionary pathways underlying flower color variation in *Helianthus annuus*. *Genetics* 173:1461–1476.
12. Bradley D, et al. (2017) Evolution of flower color pattern through selection on regulatory small RNAs. *Science* 358:925–928.
13. Zwickl J, Michaelis P, Schierwater B (1992) Untersuchungen an dem Genus Antirrhinum. *Z Indukt Abstamm Vererbungsl* 28. Barton N, Jones KN, Baum DA (2010) Sympatric sister species of Californian Antirrhinum and their transiently specialized pollinators. *Am Mid Nat* 164:337–347.
14. Counterman BA, et al. (2010) Genomic hotspots for adaptation: The population genetics of Mullerian mimicry in Heliconius erato. *PLoS Genet* 6:e1000976.
15. Mallet J, Barton NH (1989) Strong natural selection in a warning-color hybrid zone. *Evolution* 43:421–431.
16. Khimoun A, et al. (2012) Ecology predicts parapatic distributions in two closely-related Antirrhinum subspecies. *Evol Biol* 37:51–64.
17. Shang Y, et al. (2011) The molecular basis for venation patterning of pigmentation and its effect on pollinator attraction in flowers of *Antirrhinum*. *New Phytol* 189:602–615.
18. Geggel RL, Lavery TM (2005) Flower constancy in bumblebees: A test of the trait variability hypothesis. *Anim Behav* 69:939–949.
19. Smithson A, Macnair MR (1996) Frequency-dependent selection by pollinators: Mechanisms and consequences with regard to behaviour of bumblebees Bombus terrestris (L.) (Hymenoptera: Apidae). *J Evol Biol* 9:571–588.
20. Oyama RK, Jones KN, Baum DA (2010) Sympatric sister species of Californian *Antirrhinum* and their transiently specialized pollinators. *Am Mid Nat* 164:337–347.
21. Joron M, Mallet JL (1998) Diversity in mimicry: Paradox or paradigm? *Trends Ecol Evol* 13:254–261.
22. Jiggins CD (2017) The Ecology and Evolution of Heliconius Butterflies (Oxford Univ Press, Oxford).
23. Nadeau NJ, et al. (2012) Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philos Trans R Soc Lond B Biol Sci* 367:343–353.
24. Schlötterer C, Tobler R, Kofler R, Nolte V (2014) Sequencing pools of individuals-Mining genome-wide polymorphism data without big funding. *Nat Rev Genet* 15:749–763.
25. Payne SA, Riehberg DL (2016) A genomic perspective on hybridization and speciation. *Mol Ecol* 25:2317–2330.
26. Barton N, Bengtsson BO (1986) The barrier to genetic exchange between hybridizing populations. *Hereditas (Edins)* 57:357–376.
27. Yeaman S, Aeschbacher S, Bürger R (2016) The evolution of genomic islands by increased establishment probability of linked alleles. *Mol Ecol* 25:2542–2558.