

occurs. Thus, although cavity-circulation models driven by regional ocean circulation models coupled to global climate models might yield differing spatiotemporal variations in melt, they should produce patterns of retreat similar to those we have simulated, but with tighter constraints on the timing.

An important feature of our numerical simulations is that they reveal a strong sensitivity to mechanical and/or rheological weakening of the margins, which can accelerate the rate of collapse by decades to centuries. Thus, future models will require careful treatment of shear margins to accurately project sea-level rise. Our simulations also assume that there is no retreat of the ice-shelf front. Full or partial ice-shelf collapse should produce more rapid retreat than we have simulated. In addition, we have not modeled ocean-driven melt that extends immediately upstream of the grounding line, which could also accelerate retreat (32).

Our simulations provide strong evidence that the process of marine ice-sheet destabilization is already under way on Thwaites Glacier, largely in response to high subshelf melt rates. Although losses are likely to be relatively modest over the next century (<0.25 mm/year of sle), rapid collapse (>1 mm/year of sle) will ensue once the grounding line reaches the basin's deeper regions, which could occur within centuries. Such rapid collapse would probably spill over to adjacent catchments, undermining much of West Antarctica (18). Similar behavior also may be under way on neighboring Pine Island Glacier (12, 33). Unless CDW recedes sufficiently to reduce melt well below present levels, it is dif-

ficult to foresee a stabilization of the Thwaites system, even with plausible increases in surface accumulation. Although our simple melt parameterization suggests that a full-scale collapse of this sector may be inevitable, it leaves large uncertainty in the timing. Thus, ice-sheet models fully coupled to ocean/climate models are required to reduce the uncertainty in the chronology of a collapse. Nonetheless, the similarity between our highest melt rates and present observations suggests that collapse may be closer to a few centuries than to a millennium.

#### References and Notes

- H. D. Pritchard, R. J. Arthern, D. G. Vaughan, L. A. Edwards, *Nature* **461**, 971–975 (2009).
- A. Shepherd, *Geophys. Res. Lett.* **29**, 1364 (2002).
- E. Rignot, *Geophys. Res. Lett.* **35**, L12505 (2008).
- A. Shepherd et al., *Science* **338**, 1183–1189 (2012).
- S. S. Jacobs, A. Jenkins, C. F. Giulivi, P. Dutrieux, *Nat. Geosci.* **4**, 519–523 (2011).
- M. Thoma, A. Jenkins, D. Holland, S. Jacobs, *Geophys. Res. Lett.* **35**, L18602 (2008).
- E. Rignot, S. S. Jacobs, J. Mouginot, B. Scheuchl, *Science* **341**, 266–270 (2013).
- M. A. Deponter et al., *Nature* **502**, 89–92 (2013).
- H. D. Pritchard et al., *Nature* **484**, 502–505 (2012).
- A. Shepherd, D. Wingham, E. Rignot, *Geophys. Res. Lett.* **31**, L23402 (2004).
- A. Payne, A. Vieli, A. Shepherd, D. Wingham, E. Rignot, *Geophys. Res. Lett.* **31**, L23401 (2004).
- I. Joughin, B. E. Smith, D. M. Holland, *Geophys. Res. Lett.* **37**, L20502 (2010).
- I. Joughin, R. B. Alley, D. M. Holland, *Science* **338**, 1172–1176 (2012).
- O. Gagliardini, G. Durand, T. Zwinger, R. C. A. Hindmarsh, E. Le Meur, *Geophys. Res. Lett.* **37**, L14501 (2010).
- B. Medley et al., *Cryosphere Discuss.* **8**, 953–998 (2014).
- J. Mouginot, E. Rignot, B. Scheuchl, *Geophys. Res. Lett.* **41**, 1576–1584 (2014).

17. T. J. Hughes, *J. Glaciol.* **27**, 518–525 (1981).

18. J. Holt et al., *Geophys. Res. Lett.* **33**, L09502 (2006).

19. C. Schoof, *J. Geophys. Res. Earth Surf.* **112**, F03S28 (2007).

- J. Weertman, *J. Glaciol.* **13**, 3–11 (1974).
- J. H. Mercer, *Nature* **271**, 321–325 (1978).
- D. R. MacAyeal, *J. Geophys. Res. Solid Earth* **94**, 4071–4087 (1989).
- Model details are provided in the supplementary materials.
- I. Joughin et al., *J. Glaciol.* **55**, 245–257 (2009).
- E. Rignot, J. Mouginot, B. Scheuchl, *Geophys. Res. Lett.* **38**, L10504 (2011).
- E. Rignot, S. S. Jacobs, *Science* **296**, 2020–2023 (2002).
- C. Genthon, G. Krinner, H. Castebrunet, *Ann. Glaciol.* **50**, 55–60 (2009).
- E. Larour, E. Rignot, I. Joughin, D. Aubry, *Geophys. Res. Lett.* **32**, L05503 (2005).
- C. P. Borstad, E. Rignot, J. Mouginot, M. P. Schodlok, *Cryosphere Discuss.* **7**, 3567–3610 (2013).
- K. J. Tinto, R. E. Bell, *Geophys. Res. Lett.* **38**, L20503 (2011).
- S. S. R. Jamieson et al., *Nat. Geosci.* **5**, 799–802 (2012).
- B. R. Parizek et al., *J. Geophys. Res.* **118**, 638–655 (2013).
- L. Favier et al., *Nat. Clim. Change* **4**, 117–121 (2014).

**Acknowledgments:** Comments by M. Maki, D. Shapero, and four anonymous reviewers improved the manuscript. NSF supported I.J. and B.M.'s contribution (grants ANT-0424589 and ANT-0631973), and B.E.S. was supported by NASA (grant NNX09AE47G). We acknowledge NASA's Operation IceBridge, which provided much of the data used to constrain the model. The data used to constrain the model are archived at the University of Washington Library (<http://dx.doi.org/10.6069/H5CC0XMK>).

#### Supplementary Materials

[www.sciencemag.org/content/344/6185/735/suppl/DC1](http://www.sciencemag.org/content/344/6185/735/suppl/DC1)

Materials and Methods

Supplementary Text

Figs. S1 and S2

References (34–37)

27 November 2013; accepted 21 April 2014  
10.1126/science.1249055

lel speciation) (10), but the genome-wide consequences of this process are unclear (7, 8, 11). Although some genomic regions will likely diverge repeatedly during parallel speciation, many might show idiosyncratic patterns because of contingencies such as the order in which mutations arise (7, 8, 10, 11).

Even if repeated divergence occurs for some genomic regions (7, 12, 13), the underlying causes of this parallelism often remain speculative because it is difficult to disentangle the

<sup>1</sup>Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK. <sup>2</sup>Department of Biology, Utah State University, Logan, UT 84322, USA. <sup>3</sup>Department of Biology, University of Nevada, Reno, NV 89557, USA.

<sup>4</sup>Department of Entomology, Texas A&M University, College Station, TX 77843, USA. <sup>5</sup>Department of Botany, University of Wyoming, Laramie, WY 82071, USA. <sup>6</sup>Department of Biology, Notre Dame University, South Bend, IN 46556, USA.

<sup>7</sup>J. F. Blumenbach Institute of Zoology and Anthropology, University of Göttingen, 37073 Göttingen, Germany. <sup>8</sup>Department of Ecology and Evolution, University of Lausanne, Lausanne CH-1015, Switzerland. <sup>9</sup>Department of Ecology and Evolutionary Biology, Rice University, Houston, TX 77005, USA. <sup>10</sup>Department of Biological Sciences, Simon Fraser University, Burnaby, BC V5A 1S6, Canada.

\*These authors contributed equally to this work.

†Corresponding author. E-mail: p.nosil@sheffield.ac.uk

## Stick Insect Genomes Reveal Natural Selection's Role in Parallel Speciation

Víctor Soria-Carrasco,<sup>1\*</sup> Zachariah Gompert,<sup>2\*</sup> Aaron A. Comeault,<sup>1</sup> Timothy E. Farkas,<sup>1</sup> Thomas L. Parchman,<sup>3</sup> J. Spencer Johnston,<sup>4</sup> C. Alex Buerkle,<sup>5</sup> Jeffrey L. Feder,<sup>6</sup> Jens Bast,<sup>7</sup> Tanja Schwander,<sup>8</sup> Scott P. Egan,<sup>9</sup> Bernard J. Crespi,<sup>10</sup> Patrik Nosil<sup>1†</sup>

Natural selection can drive the repeated evolution of reproductive isolation, but the genomic basis of parallel speciation remains poorly understood. We analyzed whole-genome divergence between replicate pairs of stick insect populations that are adapted to different host plants and undergoing parallel speciation. We found thousands of modest-sized genomic regions of accentuated divergence between populations, most of which are unique to individual population pairs. We also detected parallel genomic divergence across population pairs involving an excess of coding genes with specific molecular functions. Regions of parallel genomic divergence in nature exhibited exceptional allele frequency changes between hosts in a field transplant experiment. The results advance understanding of biological diversification by providing convergent observational and experimental evidence for selection's role in driving repeatable genomic divergence.

Whether evolution is predictable and repeatable is difficult to test yet central to our understanding of biological diversification (1–6). Instances of repeated, parallel evolution in response to similar environmental pressures provide evidence of evolution by natu-

ral selection and can involve repeated divergence at specific genes (7–9). Indeed, parallel evolution of individual phenotypic traits has been estimated to involve the same genomic regions 30 to 50% of the time (8). Parallel evolution can also result in replicate species formation (i.e., paral-

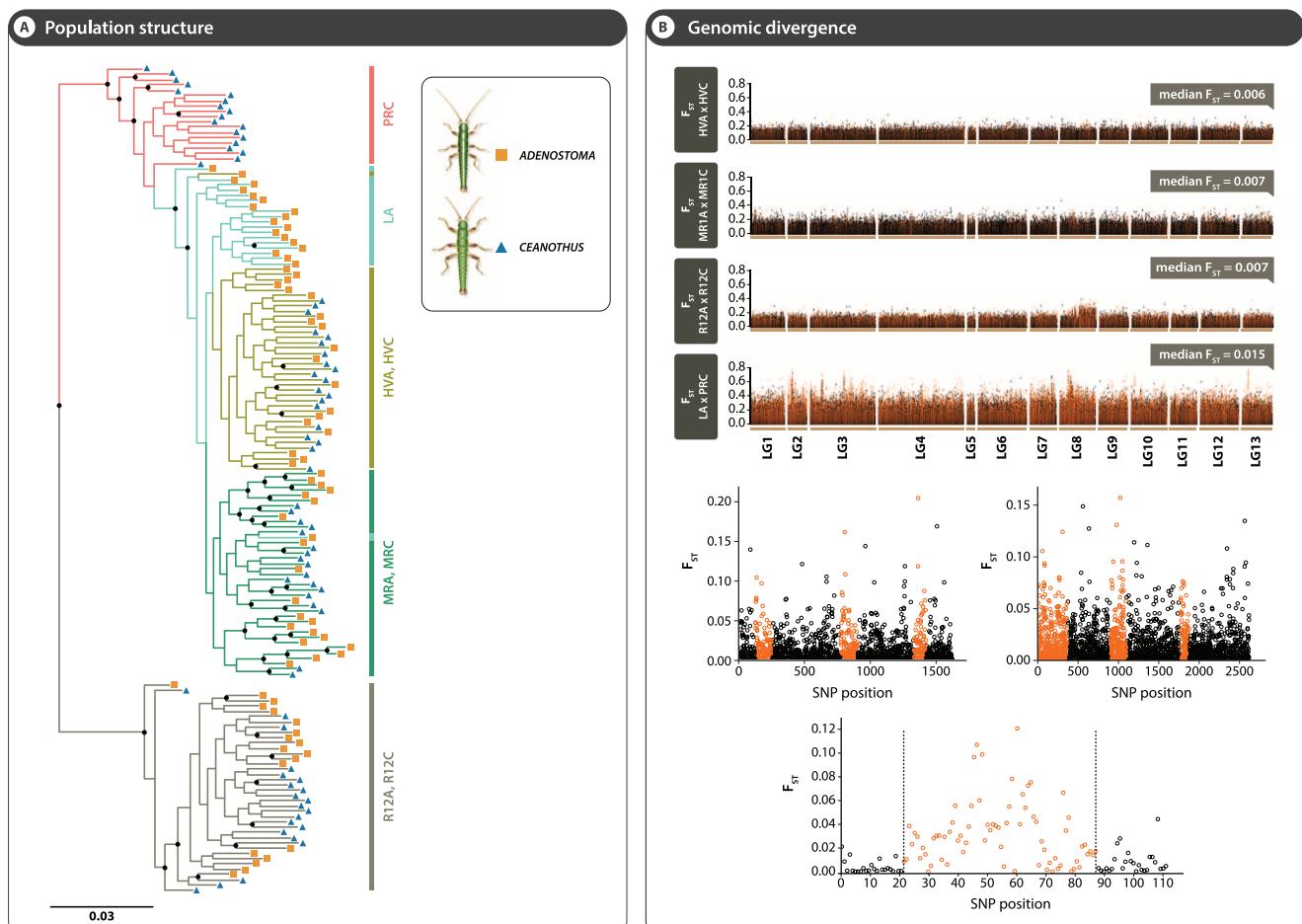
contributions of the many factors that can be involved, such as selection, mutation, recombination, gene flow, population size, and the genetic architecture of traits (2, 11, 14). We chose to study these questions with a wingless, herbivorous stick insect (*Timema cristinae*) endemic to California that has repeatedly evolved ecotypes adapted to different host plant species: *Adenostoma fasciculatum* and *Ceanothus spinosus* (Fig. 1). Populations of this species exhibit high gene flow between adjacent populations on different hosts, little or no gene flow between more distant geographic localities, and repeated evolution of partial reproductive isolation between different ecotype pairs (15–17). Consistent with the main prediction of parallel speciation, pairs of populations in different localities on the same host exhibit weaker reproductive isolation than do different ecotypes, independent from genetic distance. Thus, the

ecotypes of *T. cristinae* are in the early stages of parallel ecological speciation and have evolved increased sexual isolation through reinforcement.

We annotated the reference genome for this species (18) [assembly length of 1,027,063,217 base pairs (bp), 50% of assembly in scaffolds at least 312,000 bp long (N50)] and used sequence data from genetic crosses to assemble it into linkage groups (19). We then resequenced 160 additional whole genomes from replicate ecotype pairs and compared patterns of genomic divergence to allele frequency changes in populations experimentally transplanted to different host plants. These data allowed us to analyze the effects of adaptation on genomic divergence, experimentally test the hypothesis that accentuated and parallel genetic divergence between natural populations is driven by parallel divergent selection, rigorously quantify divergence in both coding and noncoding regions, rule out undetected

genetic differences because of low coverage across the genome, and quantify the repeatability of genetic divergence for individual loci and the genome as a whole.

Our data indicate that parallel speciation can involve both repeatable and idiosyncratic divergence at the genetic level and that the repeated component involves many regions affected by divergent selection. We examined four replicate population pairs on *Ceanothus* versus *Adenostoma* ( $n = 8$  populations) to characterize genomic divergence between ecotypes ( $n = 19$  to 21 individuals per population). Three of these pairs were directly adjacent to one another, and the fourth was separated by 6.4 km (Table 1 and fig. S1). We discovered many rare genetic variants (e.g., 2,526,553 single-nucleotide variants with minor allele frequency, MAF, <1%) and retained 4,391,556 single-nucleotide polymorphisms (SNPs) with MAF >1% that mapped to linkage



**Fig. 1. Population structure and whole-genome divergence between host-plant ecotypes of *T. cristinae*.** (A) A maximum likelihood tree from whole-genome data from individuals from *T. cristinae* populations (rooted with other species of *Timema* not shown). Black circles depict nodes with >80% bootstrap support (100 replicates). Populations LA, HVA, MRA, and R12A were collected from *Adenostoma*. Populations PRC, HVC, MRC, and R12C were collected from *Ceanothus*. The geographic locations of the populations are depicted in fig. S1. (B) Patterns of genetic divergence ( $F_{ST}$ ) along the genome. LG indicates linkage group. Colors denote whether SNPs occur

in moderately (black) or highly (red/orange) divergent genetic regions, delimited with a HMM approach (weakly differentiated regions in blue are too sparse to see). Results are shown at three different scales. (Top)  $F_{ST}$  estimates at all ~4.4 million SNPs for each of the four population pairs. (Middle) Genomic divergence for two scaffolds (scaffolds 1271 on LG7 and 2230 on LG6 from HVA × HVC). (Bottom) Magnified view of genomic divergence for one parallel HMM divergence region that harbored a SNP showing parallel allele frequency divergence between hosts upon transplantation (scaffold 556 on LG4 from R12A × R12C; see also Fig. 3). [Credit: courtesy of R. Ribas.]

**Table 1. Characteristics of the four pairs of natural populations of *T. cristinae* on different host-plant species.**  $N_e m$  values are from (16).

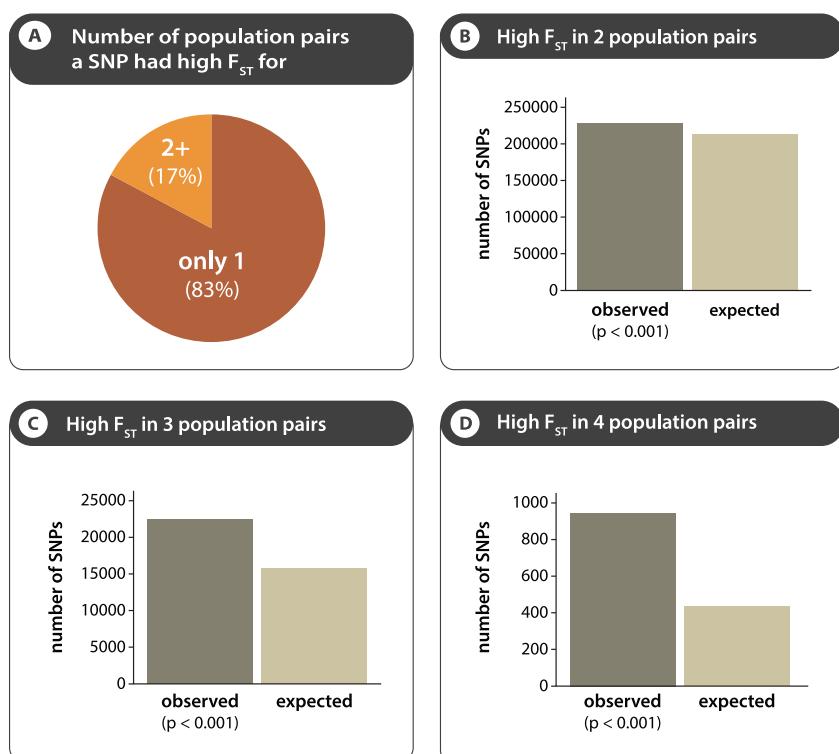
Population Pair	HVA × HVC	MR1A × MR1C	R12A × R12C	LA × PRC
<i>Population characteristics</i>				
Geography	adjacent	adjacent	adjacent	separated
Gene flow ( $N_e m$ )	28	37	93	11
No. individuals sequenced (per individual population)	20, 20	20, 20	21, 21	19, 19
<i>Genome characteristics from all SNPs</i>				
Mean MAF	17%	20%	19%	20%
Mean $F_{ST}$	0.013	0.015	0.015	0.031
Median $F_{ST}$	0.006	0.007	0.007	0.015
Range of $F_{ST}$	0.000–0.358	0.000–0.461	0.000–0.398	0.000–0.8000
90th quantile	0.035	0.041	0.041	0.083
95th quantile	0.050	0.058	0.056	0.116
<i>Characteristics of HMM divergence regions</i>				
Number	7041	3800	10,628	18,135
% of all SNPs in HMM divergence regions	13%	8%	19%	30%
Mean (median) size in bp	10,460 (4729)	6276 (2278)	10,580 (4830)	7255 (2411)
SD of size in bp	15,514	9942	15,462	12,106
Mean $F_{ST}$	0.022	0.028	0.039	0.051

groups for further analysis (across all SNPs, mean coverage per individual per SNP is 5.0).

Genomic divergence between ecotype pairs varied geographically. A phylogenomic tree grouped *T. cristinae* individuals by geographic locality and rejected grouping by host [ $P < 0.00001$ , approximately unbiased test (Fig. 1 and fig. S2)] (19). This finding was supported by principal components (PC) analysis [variance partitioning of PC1 is 93.7% by geography and 0.2% by host (fig. S1)]. The average genome-wide fixation index ( $F_{ST}$ ) for the geographically separated pair was twice that of the adjacent pairs, with all three adjacent pairs exhibiting similarly low average  $F_{ST}$  [median ~0.007; mean ~0.015 (Fig. 1 and Table 1)]. No fixed differences were observed between ecotype pairs. These levels of population differentiation are lower than those found in many studies that analyzed the genomic consequences of divergence (12, 20, 21), emphasizing that we are examining the early stages of speciation.

Patterns of divergence between ecotype pairs also varied across the genome, with  $F_{ST}$  for the most differentiated SNPs being about 60 times greater than the median value (Table 1 and fig. S3). A hidden Markov model (HMM) (22) classified the majority of the genome as moderately divergent, but we observed numerous small- to modest-sized regions of high divergence (hereafter referred to as HMM divergence regions) distributed across all linkage groups. Across population pairs, HMM divergence regions varied in number from 3800 to 18,135, comprising 8 to 30% of the genome with mean size from 6276 to 10,580 bp (standard deviation 9942 to 15,514) (Fig. 1 and fig. S4).

We tested whether divergence between replicate population pairs frequently involved the same genomic regions. To increase precision, we focused these analyses on SNPs. We identified SNPs that were highly divergent between each population pair (those above the 90th empirical quantile of the  $F_{ST}$  distribution; divergent SNPs

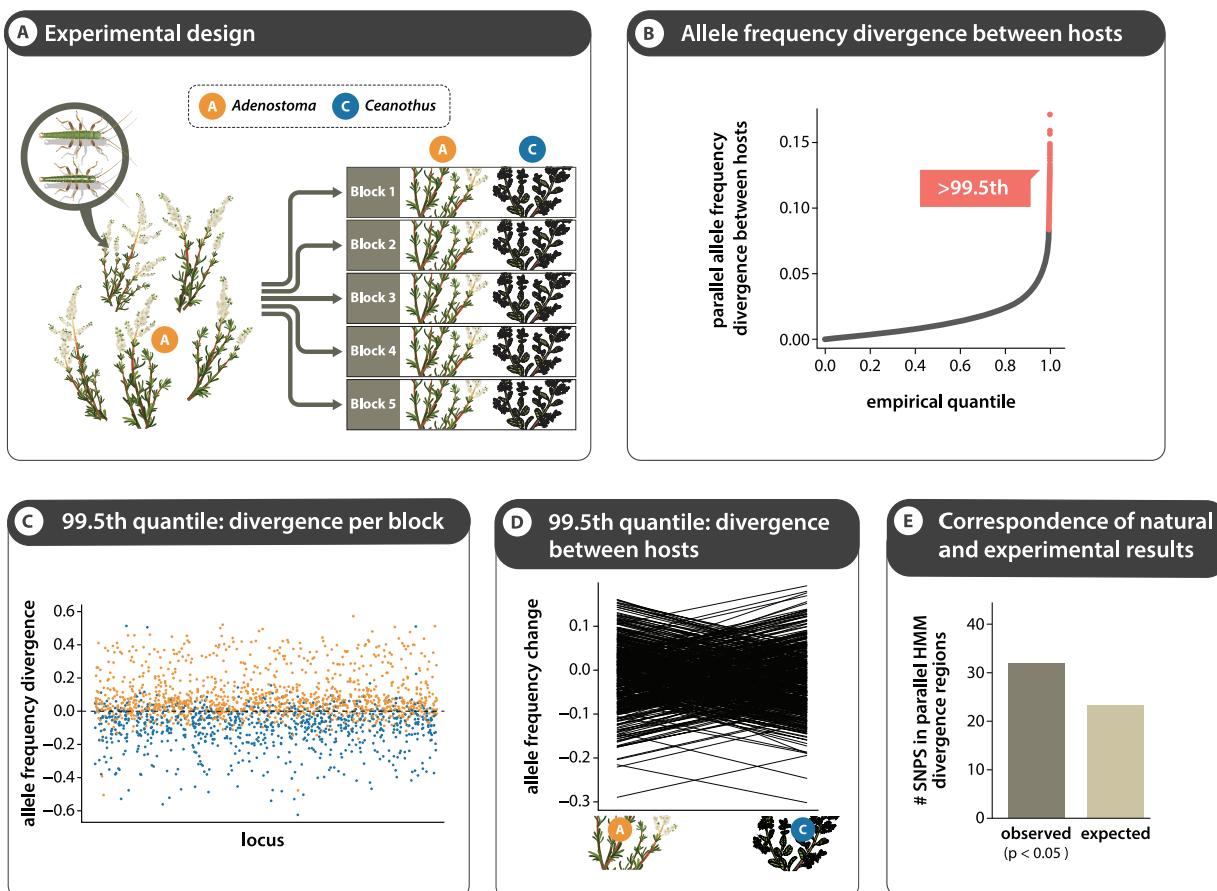


**Fig. 2. Quantification of parallel and nonparallel divergence across population pairs.** (A) The majority of divergent SNPs (those above the 90th quantile of the empirical  $F_{ST}$  distribution) were so only in a single comparison, indicative of largely nonparallel genetic divergence across the genome (pie chart; see table S1 for full results, including those from the 99th quantile). Nonetheless, a significant enrichment of parallelism was observed (bar plots, all  $P < 0.0001$ , permutation tests of observed versus expected numbers of parallel divergent SNPs). (B) High  $F_{ST}$  in two population pairs. (C) High  $F_{ST}$  in three population pairs. (D) High  $F_{ST}$  in four population pairs. [Credit: courtesy of R. Ribas.]

hereafter). The majority of divergent SNPs (83%) were divergent only in a single population pair. Thus, accentuated divergence was largely nonparallel, supporting repeated and largely independent evolutionary divergence of ecotype pairs in different localities. Nonparallel divergence could

reflect different selective pressures acting on ecotypes in different localities, geographically variable trait genetic architecture, independent genetic drift, or a combination of these factors.

Nonetheless, 17% of divergent SNPs were divergent in two or more population pairs. This



**Fig. 3. Allele frequency changes across the genome in a field transplant experiment.** (A) The experimental design, where individuals from population LA were transplanted to a nearby location of five paired blocks that contained a single plant individual of each host species. (B) An empirical quantile plot of the mean divergence in allele frequencies between hosts for all SNPs. (C) SNPs showing the largest parallel divergence between hosts

(>99.5th quantile) for individual blocks (orange, greater change on *Adenostoma*; blue, greater change on *Ceanothus*). (D) SNPs showing the largest parallel divergence between hosts (>99.5th quantile) averaged across blocks. (E) SNPs showing parallel divergence between hosts in the experiment were in parallel HMM divergence regions between natural population pairs more often than expected by chance ( $P < 0.05$ , randomization test). [Credit: courtesy of R. Ribas.]

degree of parallelism was greater than expected by chance (Fig. 2) (all  $P < 0.001$ , permutation test) and most pronounced for SNPs that were highly divergent in all population pairs [i.e., twofold greater than null expectations; congruent results were obtained for divergent SNPs above the 99th quantile (table S1)]. Last, these SNPs that were divergent in multiple population pairs were in HMM divergence regions that occurred between multiple population pairs more often than expected by chance [all  $P < 0.001$ , randomization test (table S2 and fig. S5)], demonstrating congruence between SNP-based and HMM results. The observed parallelism is consistent with repeatable genetic divergence by natural selection.

Even parallel genetic divergence can be affected by factors other than divergent selection, such as background selection and reduced recombination (2, 11). We thus tested whether HMM divergence regions occurring in multiple population pairs (i.e., parallel HMM divergence regions) harbor SNPs exhibiting divergent allele-frequency changes between multiple experimental population pairs transplanted to different plant species

in the field. In a paired blocks design, we transplanted *T. cristinae* from population LA onto nearby *Ceanothus* and *Adenostoma* plants that were devoid of *T. cristinae* [ $n = 2000$  individuals and 5 blocks (Figs. 1 and 3 and fig. S6)]. There was little or no dispersal after transplantation (19). Thirty-one individuals from LA were not transplanted and preserved as representative “ancestors” of the experimental populations [genomic data from these individuals provide a good representation of genetic variation in LA, particularly given low linkage disequilibrium (16) in this population (fig. S3)]. We collected the first-generation descendants of the transplanted insects 1 year later (these insects feature a single generation per year with a long egg diapause through summer, fall, and winter).

By sequencing these descendants ( $n = 418$  F<sub>1</sub> individuals) and their ancestors ( $n = 31$ ), we observed variable allele frequency changes between hosts across the genome, with most SNPs exhibiting weak to moderate divergence between hosts across a generation (Fig. 3). Numerous SNPs exhibited larger allele frequency changes

between hosts that were remarkably consistent across experimental replicates (i.e., blocks). Thirty-two of the 213 SNPs showing the greatest parallel and divergent allele frequency changes between hosts in the experiment (i.e., those above the 99.5th empirical quantile of such changes) were in parallel HMM divergence regions inferred for natural population pairs, which was more than the 23 expected by chance ( $P < 0.05$ , randomization test). These 32 SNPs were widely distributed across the genome (e.g., found on 32 different scaffolds and 12 of 13 linkage groups). These results are consistent with the hypothesis that parallel HMM divergence regions in natural populations are enriched for loci affected by host-related divergent selection. Our analyses focused on genomic responses to selection, and thus the role of recombination and correlations among loci in such responses warrants future work.

We examined the function of genomic regions exhibiting parallel divergence. We identified the 0.01% of SNPs with the greatest evidence of parallel divergence on the basis of

a metric that combines information on  $F_{ST}$  levels and the direction of allele frequency differences between the four natural population pairs on *Adenostoma* and *Ceanothus* [hereafter referred to as parallel divergence SNPs,  $n = 439$  SNPs; these results are robust to different metrics of parallelism and quantile cutoffs (table S3)]. These parallel divergence SNPs exhibited a 1.5-fold enrichment for being in coding regions of genes compared with expectations from all SNPs (13 and 8%, respectively,  $P < 0.001$ , randomization test), consistent with a role for coding changes even early in speciation.

Gene ontology (GO) analyses revealed that parallel divergence SNPs reside in regions containing genes involved in a range of putative biological processes, including metabolism and signal transduction, and exhibiting various molecular functions [we used GO annotations from the nearest gene on the same scaffold for each parallel divergence SNP (table S4)]. Metal ion binding and calcium ion binding were the two statistically overrepresented molecular functions in regions harboring the parallel divergence SNPs (seven- and threefold enrichment relative to all SNPs,  $P < 0.00001$  and  $P = 0.01$ , respectively, randomization test). Specifically, almost half of the parallel divergence SNPs with molecular function GO annotations (20 of 41 = 48.8%) were in regions harboring genes implicated in metal ion binding, compared with only 6.6% of all SNPs. Consistent with these results from SNPs, the 32 parallel HMM divergence regions associated with parallel divergence in the experiment contained genes related to binding of metals (e.g., iron) and calcium (table S5). Local adaptation to balance the nutritional versus toxic effects of metals has been described in other insects (23), and metals are known to affect traits that differ between the *T. cristinae* ecotypes, such as pigmentation and mandible morphology (24, 25).

Last, we examined the function of SNPs that were divergent between only a single population pair (i.e., nonparallel divergence SNPs,  $n = 440$ ) (19). These SNPs were in regions significantly enriched for genes exhibiting several functions, including one that involved metal binding [i.e., zinc ion binding,  $P < 0.05$ , randomization test (table S6)]. This finding raises the possibility that nonparallel genetic divergence is sometimes the result of adaptive divergence between host ecotypes and that it includes cases of parallelism at the functional level, although further work is required to test these hypotheses.

Our data show that early stages of parallel speciation in *T. cristinae* involve mostly nonparallel genetic divergence between ecotypes. However, numerous regions of the genome have diverged in parallel, and these include more coding genes than expected by chance and harbor loci showing experimentally induced allele frequency changes between hosts. These results indicate that divergent selection plays a role in repeated genomic divergence between ecotypes. Furthermore,

our results suggest that, although repeated evolutionary scenarios (i.e., replaying the tape of life) would likely result in idiosyncratic outcomes, there may be a repeatable component driven by selection that can be detected, even at the genome-wide level and during the complex process of speciation.

#### References and Notes

- D. L. Stern, *Nat. Rev. Genet.* **14**, 751–764 (2013).
- R. D. H. Barrett, H. E. Hoekstra, *Nat. Rev. Genet.* **12**, 767–780 (2011).
- M. K. Burke, *Proc. R. Soc. London Ser. B* **279**, 5029–5038 (2012).
- D. M. Weinreich, N. F. Delaney, M. A. Depristo, D. L. Hartl, *Science* **312**, 111–114 (2006).
- J. R. Meyer *et al.*, *Science* **335**, 428–432 (2012).
- J. B. Losos, *Lizards in an Evolutionary Tree: Ecology and Adaptive Radiation of Anoles* (Univ. California Press, Berkeley, CA, 2009).
- F. C. Jones *et al.*, *Nature* **484**, 55–61 (2012).
- G. L. Conte, M. E. Arnegard, C. L. Peichel, D. Schlüter, *Proc. R. Soc. London Ser. B* **279**, 5039–5047 (2012).
- Heliconius Genome Consortium, *Nature* **487**, 94–98 (2012).
- D. Schlüter, L. M. Nagel, *Am. Nat.* **146**, 292–301 (1995).
- J. L. Feder, S. P. Egan, P. Nosil, *Trends Genet.* **28**, 342–350 (2012).
- H. Ellegren *et al.*, *Nature* **491**, 756–760 (2012).
- S. H. Martin *et al.*, *Genome Res.* **23**, 1817–1828 (2013).
- H. Ellegren, *Trends Ecol. Evol.* **29**, 51–63 (2014).
- P. Nosil, *Am. Nat.* **169**, 151–162 (2007).
- P. Nosil *et al.*, *Proc. R. Soc. London Ser. B* **279**, 5058–5065 (2012).
- P. Nosil, B. J. Crespi, C. P. Sandoval, *Nature* **417**, 440–443 (2002).
- Z. Gompert *et al.*, *Ecol. Lett.* **17**, 369–379 (2014).
- Materials and methods are available as supplementary materials on *Science* Online.

- M. K. N. Lawniczak *et al.*, *Science* **330**, 512–514 (2010).
- S. Renaud *et al.*, *Nature Commun.* **4**, 1827 (2013).
- T. Hofer, M. Foll, L. Excoffier, *BMC Genomics* **13**, 107 (2012).
- T. van Ooik, M. J. Rantala, *Ann. Zool. Fenn.* **47**, 215–222 (2010).
- J. F. V. Vincent, *Structural Biomaterials* (Princeton Univ. Press, Princeton, NJ, 1990).
- N. T. Dittmer, M. R. Kanost, *Insect Biochem. Mol. Biol.* **40**, 179–188 (2010).

**Acknowledgments:** The work was funded by the European Research Council (grant R/129639) and Utah State University start-up funds. We thank J. Slate, A. Beckerman, D. Schlüter, and two anonymous reviewers for constructive feedback and the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics (funded by Wellcome Trust grant reference 090532/Z/09/Z and Medical Research Council Hub grant G0900747 91070) for generation of the sequencing data. O. Simakov and A. Kapusta provided scripts for transposable elements annotation. The data reported in this paper are tabulated in the supplementary materials and archived in the following databases. The raw sequencing reads are in the NCBI short read archive (BioProject ID: PRJNA243533), and the whole genome assembly and annotation at <http://nosil-lab.group.shef.ac.uk/resources>. Additional data and computer source code have been deposited in the Dryad repository, datadryad.org, and are also available from the authors upon request. The authors declare no conflicts of interest.

#### Supplementary Materials

- [www.sciencemag.org/content/344/6185/738/suppl/DC1](http://www.sciencemag.org/content/344/6185/738/suppl/DC1)  
Materials and Methods  
Supplementary Text  
Figs. S1 to S6  
Tables S1 to S6  
References (26–80)

12 February 2014; accepted 18 April 2014  
10.1126/science.1252136

## NMR Spectroscopy of Native and in Vitro Tissues Implicates PolyADP Ribose in Biomineralization

W. Ying Chow,<sup>1</sup> Rakesh Rajan,<sup>2</sup> Karin H. Muller,<sup>3</sup> David G. Reid,<sup>1</sup> Jeremy N. Skepper,<sup>3</sup> Wai Ching Wong,<sup>1</sup> Roger A. Brooks,<sup>2</sup> Maggie Green,<sup>4</sup> Dominique Bihann,<sup>5</sup> Richard W. Farndale,<sup>5</sup> David A. Slatter,<sup>5</sup> Catherine M. Shanahan,<sup>6</sup> Melinda J. Duer<sup>1\*</sup>

Nuclear magnetic resonance (NMR) spectroscopy is useful to determine molecular structure in tissues grown in vitro only if their fidelity, relative to native tissue, can be established. Here, we use multidimensional NMR spectra of animal and in vitro model tissues as fingerprints of their respective molecular structures, allowing us to compare the intact tissues at atomic length scales. To obtain spectra from animal tissues, we developed a heavy mouse enriched by about 20% in the NMR-active isotopes carbon-13 and nitrogen-15. The resulting spectra allowed us to refine an in vitro model of developing bone and to probe its detailed structure. The identification of an unexpected molecule, poly(adenosine diphosphate ribose), that may be implicated in calcification of the bone matrix, illustrates the analytical power of this approach.

Understanding the atomic-level structure of the extracellular matrix (ECM) of tissues is a prerequisite for detailed understanding of both biological and mechanical functions of those tissues. However, while nanoscopic and longer structural length scales can be studied by various forms of microscopy (1–3), there is a

paucity of techniques able to probe atomic-level structures in such complex and heterogeneous materials as the ECM. Multidimensional nuclear magnetic resonance (NMR) spectroscopy is capable in principle of yielding structural information (4–6), even in such a difficult situation, but only if <sup>13</sup>C and <sup>15</sup>N labeling of the molecular