

Mapping the Epigenetic Basis of Complex Traits

Sandra Cortijo,^{1*†} René Wardenaar,^{2*} Maria Colomé-Tatché,^{2*‡} Arthur Gilly,^{3§} Mathilde Etcheverry,¹ Karine Labadie,³ Erwann Caillieux,¹ Frédéric Hospital,⁴ Jean-Marc Aury,³ Patrick Wincker,^{3,5,6} François Roudier,¹ Ritsert C. Jansen,² Vincent Colot,^{1||} Frank Johannes^{2||}

Quantifying the impact of heritable epigenetic variation on complex traits is an emerging challenge in population genetics. Here, we analyze a population of isogenic *Arabidopsis* lines that segregate experimentally induced DNA methylation changes at hundreds of regions across the genome. We demonstrate that several of these differentially methylated regions (DMRs) act as bona fide epigenetic quantitative trait loci (QTL^{epi}), accounting for 60 to 90% of the heritability for two complex traits, flowering time and primary root length. These QTL^{epi} are reproducible and can be subjected to artificial selection. Many of the experimentally induced DMRs are also variable in natural populations of this species and may thus provide an epigenetic basis for Darwinian evolution independently of DNA sequence changes.

Methylation of cytosines is an epigenetic mark involved in the silencing of transposable elements (TEs) and genes (1). Despite its functional conservation across many species (2, 3), intraspecific surveys have revealed widespread variation in DNA methylation patterns within populations (4–6). Estimates in the model plant *Arabidopsis thaliana* indicate that heritable changes in the methylation status of clusters of cytosines, which could be function-

ally more relevant than individual cytosines (7), arise spontaneously at rates similar to that of DNA sequence mutations (8, 9). A key challenge in population genetics is to show that epigenetic variants exist independently of cis- or trans-acting DNA sequence changes, are stably transmitted over many sexual generations, and are associated with heritable phenotypic variation (10). Addressing this challenge using natural populations continues to pose major technical difficulties.

To overcome these difficulties, we established in *Arabidopsis* a population of so-called epigenetic recombinant inbred lines (epiRILs) that have almost identical DNA sequences but segregate many differences in DNA methylation (11). To derive this population, a plant homozygous for the recessive *ddm1-2* mutation was first crossed with a near-isogenic wild-type (WT) individual. The *ddm1-2* mutation leads to a loss of DNA methylation and silencing over transposable elements (TEs) mainly, with potential consequences

¹Institut de Biologie de l'École Normale Supérieure, Centre National de la Recherche Scientifique (CNRS), UMR 8197, Institut National de la Santé et de la Recherche Médicale (INSERM) U 1024, Paris F-75005, France. ²Groningen Bioinformatics Centre, GBB, University of Groningen, 9747 AG Groningen, Netherlands. ³Commissariat à l'Énergie Atomique (CEA), Institut de Génétique (IG), Genoscope, 2 Rue Gaston Crémieux, 91057 Evry, France. ⁴Institut National de la Recherche Agronomique (INRA), UMR 1313, Génétique Animale et Biologie Intégrative, Domaine de Vilvert, 78352 Jouy en Josas, France. ⁵CNRS, UMR 8030, CP5706, Evry, France. ⁶Université d'Evry, CP5706, Evry, France.

*These authors contributed equally to this work.

†Present address: Sainsbury Laboratory, University of Cambridge, Cambridge, CB2 1LR, UK.

‡Present address: European Research Institute for the Biology of Ageing, University of Groningen, UMCG, A. Deusinglaan 1, NL-9713 AV Groningen, Netherlands.

§Present address: Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK.

||Corresponding author. E-mail: colot@biologie.ens.fr (V.C.); f.johannes@rug.nl (F.J.)

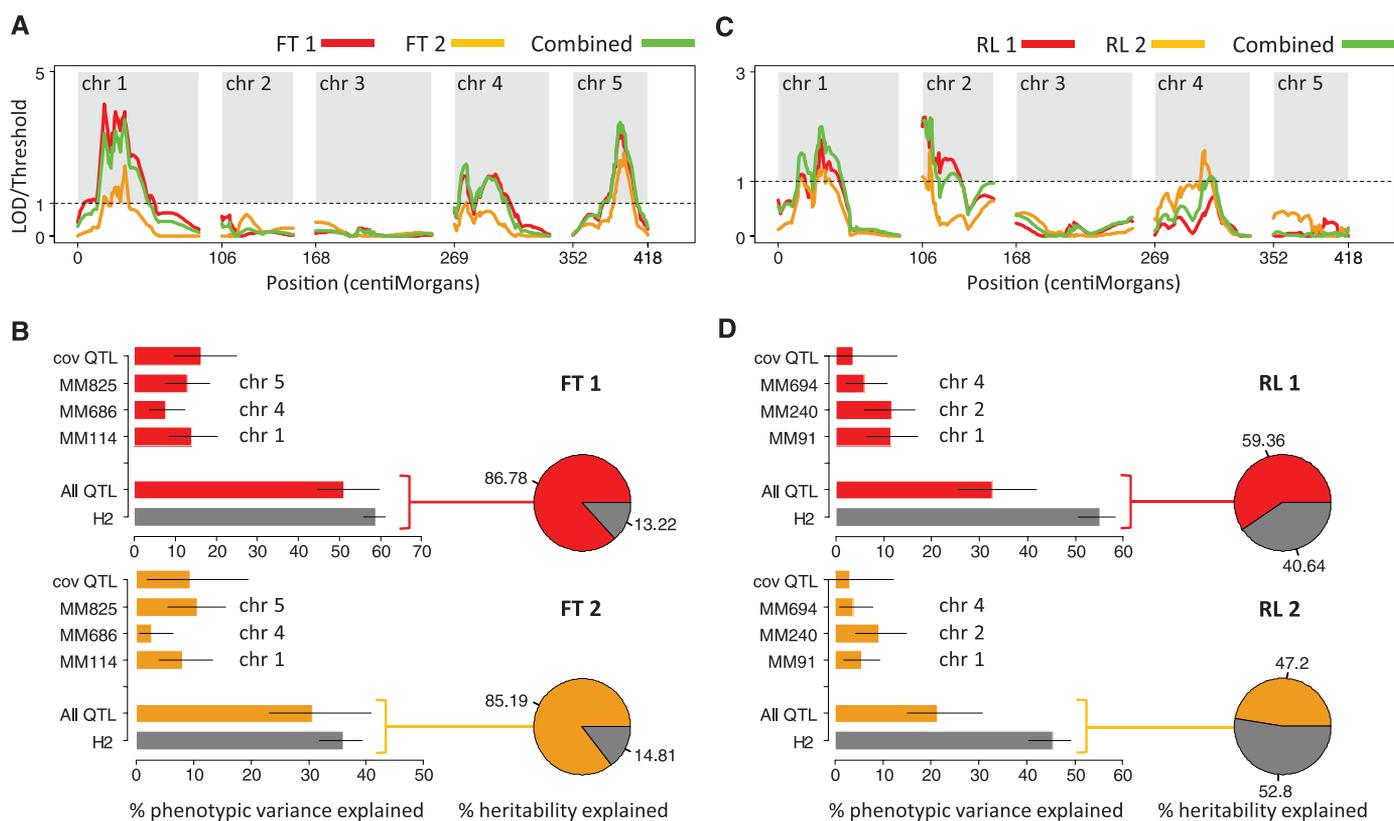


Fig. 1. Interval mapping results. (A) QTL mapping profiles for two independent flowering time measurements (FT1 and FT2), as well as their average (combined). FT1 was measured in the greenhouse and FT2 in a field experiment. (B) Percentages of phenotypic variance and of broad-sense heritability (H^2) explained

by the peak QTL markers (MM#). Error bars, ± 1 standard error of the estimate. (C) QTL mapping profiles for two independent primary root length measurements (RL1 and RL2), as well as their average (combined). Both RL1 and RL2 were measured in a climate-controlled growth chamber. (D) Same as in (B), but for RL.

on the expression of neighboring genes (12), but transposition events are relatively rare (13). Importantly, some of the DNA methylation and expression changes induced by *ddm1-2* are inherited independently of the mutation (14, 15). A single F1 *DDM1/ddm1-2* individual was therefore backcrossed to the WT parental line, and after selection of F2 progeny homozygous for the *DDM1* allele, the epiRILs ($N > 500$) were selfed for six generations (fig. S1).

Phenotypic analysis revealed significant broad-sense heritability in the epiRILs, with estimates ranging from about 0.05 to 0.4 (11, 16, 17). Theoretical predictions indicate that these heritability values are consistent with a small number of parentally derived quantitative trait loci (QTL) (17, 18). We hypothesized that these QTL are caused by stably inherited DNA methylation changes originating from the *ddm1-2* mutant founder parent. Indeed, a survey of the DNA methylomes of a selected set of 123 epiRILs identified hundreds of parental differentially methylated regions (DMRs) showing Mendelian segregation patterns (19).

Using an informative subset of 126 of these DMRs as physical markers, we were able to derive a genetic map covering 81.9% of the total genome (19).

Here, we used this map in conjunction with classical linkage analysis to search for epigenetic quantitative trait loci (QTL^{epi}) underlying complex traits in the epiRIL population.

We interval-mapped (20) two highly heritable (and weakly correlated) complex traits, flowering time (FT1) and primary root length (RL1) (Fig. 1, fig. S2, and table S1). The FT1 phenotype was obtained in a greenhouse experiment (11), whereas RL1 was measured in a climate-controlled growth chamber (21). Linkage analysis detected highly significant QTL for FT1 on chromosome 1 (chr 1) [40.59 cM; logarithm of the odds ratio for linkage (LOD) = 8.72], chr 4 (30 cM; LOD = 4.43), and chr 5 (41.73 cM; LOD = 8.53) (Fig. 1A and table S2). For these three QTL, the plants that inherited the WT epigenotype at the peak marker flowered significantly later than those with the *ddm1-2* inherited epigenotype (fig. S3A). The combined additive effects of these QTL explained 86.78%

of the broad-sense heritability for the trait and 51.14% of the total phenotypic variance (Fig. 1B and table S3). All three QTL were confirmed using independent flowering time data (FT2) collected in a field experiment (17) (Fig. 1, A and B, and table S3), which indicates that these QTL are robust across environmental settings. For RL1, we detected significant QTL on chr 1 (38 cM; LOD = 4.9), chr 2 (6.47 cM; LOD = 5.28), and chr 4 (50 cM; LOD = 2.65) (Fig. 1C and table S2). The WT-inherited epigenotype at the peak QTL markers was associated with longer primary roots compared with the *ddm1-2* inherited epigenotype (fig. S3B). The combined additive effect of these QTL explained 59.36% of the estimated broad-sense heritability and 32.69% of the total phenotypic variance (Fig. 1D and table S4). Again, all three QTL were confirmed with data from a replicate phenotyping experiment (RL2) (Fig. 1, C and D, and table S4).

Our linkage mapping results indicate that the broad-sense heritability in the epiRILs is mainly due to causal variants originating from the pa-

A

Trait	QTL		Inserted sequence				Number of epiRILs shared			Phenotypic effect (<i>p</i> -value)	
	Chr	Start	Stop	Name	Start	Stop	Seq	PCR	Total	TE	QTL marker
FT1	1	15908858	17261949	ATENSPM3	16836944	16837131	5 / 52	2 / 27	7 / 79	0.0076	5.20E-07
RL1	1	13617271	17523612	ATENSPM3	16836944	16837131	5 / 52	2 / 27	7 / 79	0.023	0.00012
RL1	1	13617271	17523612	ATENSPM3 / HELITRON	17407431	17410302	2 / 52	0 / 27	2 / 79	0.75	0.00012
RL1	4	8906583	11824662	ATCOPIA93	9649457	9651158	13 / 52	9 / 27	22 / 79	0.073	0.020
RL1	4	8906583	11824662	ATCOPIA78	10736583	10740627	18 / 52	8 / 27	26 / 79	0.99	0.020

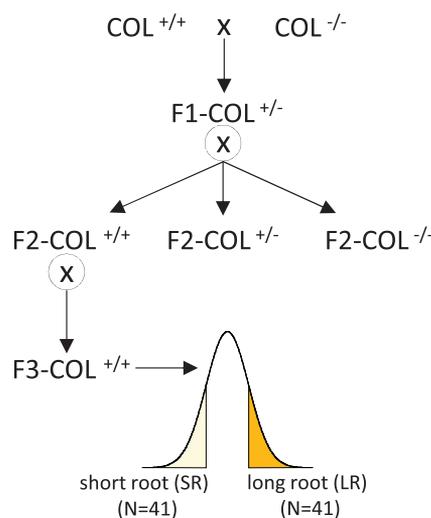
B

Cross between two new WT (COL^{+/+}) and *ddm1-2* (COL^{-/-}) founder plants.

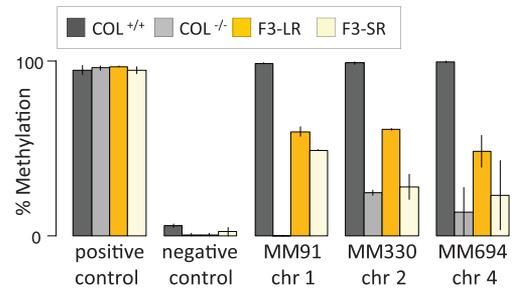
Selfing of a F1 plant epi-heterozygote for all peak QTL markers on chr 1, 2, and 4.

Selfing of a single F2 with wildtype genotype at the *DDM1* locus.

Phenotyping and selective epigenotyping of short and long root F3 plants.



C



D

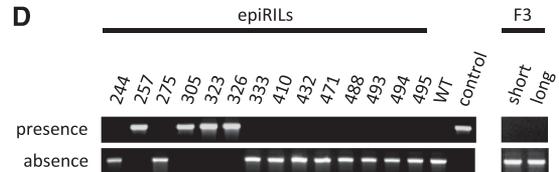


Fig. 2. Ruling out *ddm1-2*-derived TE insertions as a cause for the epiRIL QTL. (A) Resequencing of 52 epiRILs and targeted PCR of an additional 27 epiRILs detected four shared insertions in the RL and FT QTL intervals (coordinates are according to TAIR10). Phenotypic analysis testing for the effect of TEs and peak QTL markers [*P* values from multiple regression models (table S8)]. (B) Validation of the RL QTL by selective epigenotyping of 82 short- and long-root F3 progeny obtained using the crossing scheme shown. (C) Quantitative PCR analysis of McrBC-digested DNA of tail-selected samples was used to determine

DNA methylation levels at the peak markers MM91, MM330, and MM694 on chr 1, 2, and 4, respectively (right panel). We used marker MM330 on chr 2 instead of the peak marker MM240. These markers are in tight linkage, but MM330 was easier to assay by PCR. Error bars, ± 1 SEM. (D) Example of the presence or absence of the most common insertion (ATCOPIA 93) in a sample of epiRILs and the pools of short- and long-root F3 individuals. (Top) Results of PCR with primer pairs designed to amplify one end of the element and its flanking sequence. (Bottom) Results of PCR with primer pairs designed to amplify the WT sequence. Positive control: epiRIL 55.

rental generation and not from later generations of inbreeding. To examine the possibility that these parentally derived causal variants are transposable element (TE) insertions that occurred in the *ddm1-2* parental line rather than DMRs, we resequenced a representative sample of 52 of the 123 epiRILs (tables S5 and S6). Our analysis revealed, in addition to several nonshared TE insertions, a total of four shared TE insertions in the RL and FT QTL confidence intervals (Fig. 2A, fig. S4, and table S6): two shared TE insertions in the chr 1 interval and two in the chr 4 interval. We were able to confirm the shared TE insertions using targeted polymerase chain reaction (PCR) assays in an additional 27 epiRILs (Fig. 2, A and D, and tables S5 to S7).

However, further analysis revealed that shared TE insertions were not consistently inherited from

the *ddm1-2* parent (fig. S4 and table S6). Thus, we found that the ATCOPIA78 insertion on chr 4 lies in an interval of WT origin and is present in about one-third of the epiRILs, which indicates that this insertion occurred in the F1 individual rather than in the *ddm1-2* parent. Similarly, the two ATENSPM3 insertions on chr 1 lie in intervals of *ddm1-2* origin but are absent in some epiRILs with this epigenotype, suggesting that they either occurred in the parental *ddm1-2* line, with excisions in some epiRILs, or else arose in the F1 individual. Finally, the ATCOPIA93 insertion on chr 4 lies in an interval that is of *ddm1-2* origin in some epiRILs and of WT origin in others, reflecting a highly dynamic inheritance pattern. Attempts to associate these shared TE insertions with phenotypes revealed a significant association for ATENSPM3 (chr 1) with root length and flowering

time (FT, $P = 0.0076$; RL, $P = 0.023$) (Fig. 2A and table S8A) and a borderline significant effect for the ATCOPIA93 (chr 4) insertion on primary root length ($P = 0.073$). However, phenotypic effects were much weaker than those of the peak QTL markers (Fig. 2A and table S8, B and C), which implies that these shared TE insertions are unlikely causal.

To support this conclusion, we crossed a new pair of WT and *ddm1-2* founder plants (Fig. 2B) (21) and selfed the F1 to select a single *DDMI/DDMI* F2 individual that was “epi-heterozygote” for all three QTL peak markers on chr 1, 2, and 4 (21). After an additional selfing, we selected F3 progeny from the long and short extremes of the RL distribution. DNA methylation analysis confirmed the association of short and long primary roots with the *ddm1-2*-like and WT methylation states at the

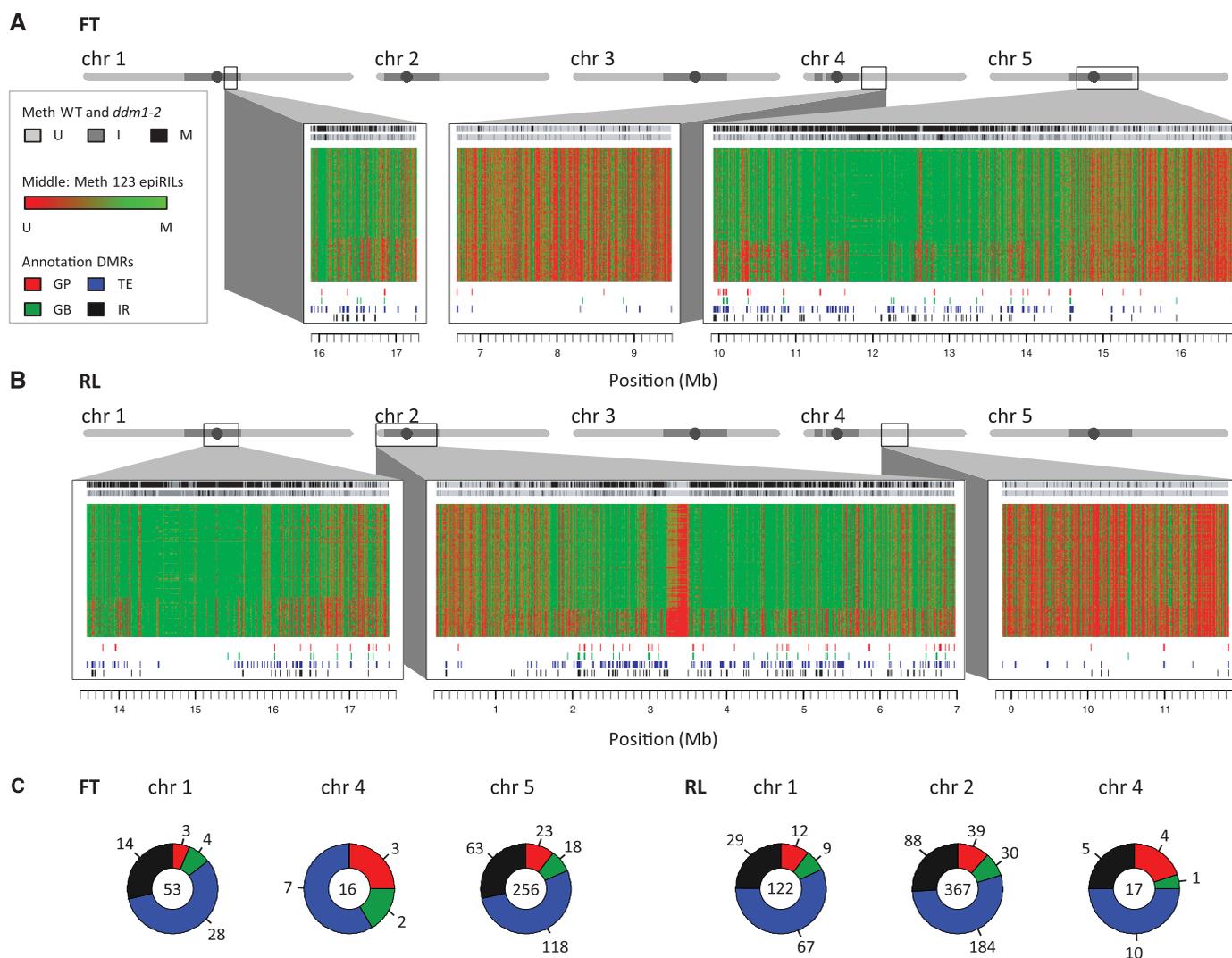


Fig. 3. DNA methylation profiles of candidate DMRs in the epiRIL QTL intervals. (A and B) Location and annotation of candidate DMRs detected for FT (A) and RL (B). The top part of the rectangles shows the DNA methylation profile of the WT and *ddm1-2* parents, respectively (U, unmethylated; I, intermediate DNA methylation; M, high-level DNA methylation). The DNA methylation profiles of the epiRILs are indicated below and are ordered according to the epigenotype of the peak marker [from WT (top) to *ddm1-2* (bottom)]. The bottom part of

the rectangles shows the annotations that overlap with the DMRs (GP, gene promoters; GB, gene bodies; TE, transposable element sequences; IR, intergenic regions). A schematic representation of each chromosome is plotted above each rectangle. (C) Number of candidate DMRs detected for each QTL interval (values inside circles) and the number of unique annotations (values outside circles) with which they overlap. DMRs can overlap multiple annotations (tables S12 and S13). Colors and abbreviations are as in (A).

peak QTL marker, respectively (Fig. 2C) (21). Furthermore, the tail-selected F3 individuals contained none of the shared TE insertions identified in the epiRIL population (Fig. 2D and table S6). We thus conclude that the epiRIL QTL are most likely caused by the heritable (*ddm1-2* induced) loss of DNA methylation in the QTL intervals.

Next, we searched for putative causal DMRs in the RL and FT QTL intervals. We analyzed the methylomes (~165–base pair resolution) of the 123 epiRILs and their founders (19) and required candidate DMRs to be in approximate linkage disequilibrium with the peak QTL marker and displaying clear differences in DNA methylation states and expression levels between the WT and the *ddm1-2* founder lines (figs. S5 to S8). Our search revealed 325 candidate DMRs within the FT QTL intervals (chr1, 53; chr4, 16; chr5, 256), which mapped to 44 unique genes (including promoter regions), 153 annotated TE sequences, and 77 intergenic regions (Fig. 3, A and C, and tables S9, S10, and S12). For the RL QTL intervals, we detected 506 candidate DMRs (chr 1, 122; chr2, 367; chr4, 17), mapping to 71 unique genes (including promoter regions), 261 annotated TE sequences, and 122 intergenic regions (Fig. 3, B and C, and tables S9, S11, and S13). Further analysis of these candidate DMRs did not identify any obvious flowering time and root length genes (tables S10 and S11), which could be consistent with the lower amplitude of phenotypic variation observed among the epiRILs than among highly contrasted accessions (16). However, we cannot rule out that the candidate DMRs are in LD with causal DMRs that could not be called with our method. Ultimately, fine-mapping approaches and targeted manipulation

of selected DMRs will be required to identify causal regions.

Our analysis of the epiRILs demonstrates that induced DMRs can be stably inherited independently of DNA sequence changes and function as epigenetic quantitative trait loci (QTL^{epi}). Phenotypically, the detected QTL^{epi} have all the necessary properties to become targets of natural or artificial selection. Taking advantage of the single-nucleotide resolution methylomes of 138 natural accessions (4) (table S14), we could show that about 30% of the heritable DMRs identified in the epiRIL population overlap with naturally occurring DMRs among these accessions (figs. S9 and S10). Therefore, these epiRIL DMRs may have been historical targets of epimutations in the wild, either through trans-induced *ddm1*-like mutation events or else through still unknown mechanisms. This finding indicates in turn that DMRs could also act as QTL^{epi} in natural populations and thus constitute a measurable component of the so-called “missing heritability.” This possibility may have deep implications on how we delineate and interpret the heritable basis of complex traits.

References and Notes

1. J. A. Law, S. E. Jacobsen, *Nat. Rev. Genet.* **11**, 204–220 (2010).
2. S. Feng *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 8689–8694 (2010).
3. A. Zemach, I. E. McDaniel, P. Silva, D. Zilberman, *Science* **328**, 916–919 (2010).
4. R. J. Schmitz *et al.*, *Nature* **495**, 193–198 (2013).
5. S. R. Eichten *et al.*, *Plant Cell* **25**, 2783–2797 (2013).
6. H. Heyn *et al.*, *Genome Res.* **23**, 1363–1372 (2013).
7. D. Weigel, V. Colot, *Genome Biol.* **13**, 249 (2012).
8. C. Becker *et al.*, *Nature* **480**, 245–249 (2011).
9. R. J. Schmitz *et al.*, *Science* **334**, 369–373 (2011).
10. F. Johannes, V. Colot, R. C. Jansen, *Nat. Rev. Genet.* **9**, 883–890 (2008).

11. F. Johannes *et al.*, *PLOS Genet.* **5**, e1000530 (2009).
12. Z. Lippman *et al.*, *Nature* **430**, 471–476 (2004).
13. S. Tsukahara *et al.*, *Nature* **461**, 423–426 (2009).
14. T. Kakutani, K. Munakata, E. J. Richards, H. Hirochika, *Genetics* **151**, 831–838 (1999).
15. F. K. Teixeira *et al.*, *Science* **323**, 1600–1604 (2009).
16. V. Latzel, Y. Zhang, K. Karlsson Moritz, M. Fischer, O. Bossdorf, *Ann. Bot.* **110**, 1423–1428 (2012).
17. F. Roux *et al.*, *Genetics* **188**, 1015–1017 (2011).
18. F. Johannes, M. Colomé-Tatché, *Genetics* **188**, 215–227 (2011).
19. M. Colomé-Tatché *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16240–16245 (2012).
20. E. S. Lander, D. Botstein, *Genetics* **121**, 185–199 (1989).
21. Materials and methods and supplementary text are available as supporting material on Science Online.

Acknowledgments: We thank O. Bossdorf, C. Richards, T. Day, and K. Verhoeven for their input on an earlier version of this report. This work was supported by grants from the Netherlands Organization for Scientific Research (to F.J., R.C.J., R.W., and M.C.-T.); a University of Groningen Rosalind Franklin Fellowship to M.C.-T.; the Agence Nationale de la Recherche (ANR-09-BLAN-0237 EPIMOBILE to V.C. and P.W.; ANR-06-GPLA-010 TAG, Investissements d’Avenir ANR-10-LABX-54 MEMO LIFE, and ANR-11-IDEX-0001-02 PSL* Research University to V.C.); and the European Union (EpiGeneSys FP7 Network of Excellence number 257082, to V.C.). S.C. and M.E. were supported by Ph.D. studentships from the Ministère de l’Enseignement Supérieur et de la Recherche, with additional support from the Fondation pour la Recherche Médicale (to S.C.). Sequence reads for the 52 epiRILs and parental lines are deposited at the European Bioinformatics Institute under accession number ERP004507.

Supplementary Material

www.sciencemag.org/content/343/6175/1145/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S11
Tables S1 to S14
References (22–28)

5 November 2013; accepted 22 January 2014
Published online 6 February 2014;
10.1126/science.1248127

A Single Gene Affects Both Ecological Divergence and Mate Choice in *Drosophila*

Henry Chung,¹ David W. Loehlin,¹ Héloïse D. Dufour,¹ Kathy Vaccarro,¹ Jocelyn G. Millar,² Sean B. Carroll^{1*}

Evolutionary changes in traits involved in both ecological divergence and mate choice may produce reproductive isolation and speciation. However, there are few examples of such dual traits, and the genetic and molecular bases of their evolution have not been identified. We show that methyl-branched cuticular hydrocarbons (mbCHCs) are a dual trait that affects both desiccation resistance and mate choice in *Drosophila serrata*. We identify a fatty acid synthase *mFAS* (*CG3524*) responsible for mbCHC production in *Drosophila* and find that expression of *mFAS* is undetectable in oenocytes (cells that produce CHCs) of a closely related, desiccation-sensitive species, *D. birchii*, due in part to multiple changes in cis-regulatory sequences of *mFAS*. We suggest that ecologically influenced changes in the production of mbCHCs have contributed to reproductive isolation between the two species.

The evolution of traits with dual roles in ecological divergence and mate choice could cause populations to become reproductively isolated through local adaptation (1, 2).

However, the direct contribution of ecological divergence to speciation is uncertain, in part because relatively few systems have been investigated experimentally, and the genes affecting

dual traits have not been identified. Insect cuticular hydrocarbons (CHCs) are potential dual traits (3). CHCs seal the cuticle, protecting against desiccating environments (4), and can also act as pheromones that influence mate choice and mating success (5). Whereas specific CHC molecules act as pheromones (6, 7), no specific CHC or class of CHCs has been demonstrated to be involved directly in desiccation resistance.

One class of CHCs, methyl-branched CHCs (mbCHCs), which have melting points above ambient temperature, are hypothesized to help maintain a barrier against evaporative water loss (8). *D. serrata*, a habitat generalist found outside of and on the fringes of the rainforest on the east coast of Australia, is relatively desiccation resistant and produces relatively large amounts of mbCHCs (29% of all CHCs) (Fig. 1A). In contrast, its close relative *D. birchii*, a habitat specialist found exclusively in the humid rainforest, is ex-

¹Howard Hughes Medical Institute and Laboratory of Molecular Biology, University of Wisconsin, Madison, WI 53706, USA. ²Department of Entomology, University of California, Riverside, CA 92521, USA.

*Corresponding author. E-mail: sbcarroll@wisc.edu