Spring 2025 – Epigenetics and Systems Biology
Discussion Session (Epigenetics and Evolutionary Biology)
Michael K. Skinner – Biol 476/576
Week 16 (April 24)

Epigenetics and Evolutionary Biology

Primary Papers

1. Luo, et al. (2020) Cell Reports 33:108306. (PMID: 33113358)
2. Aagaard, et al. (2022) Mol Ecol. (22):5765-5783. (PMID 36112081)
3. McNew, et al. (2017) BMC Evolution 17(1):183. (PMID: 28835203)

Discussion

Student 17 – Ref #1 above

- What was the experimental design and model system?
- What epigenetic process and gene network effects were observed?
- Does this provide evidence for environmental induction of epigenetic alterations in a gene network for evolutionary adaptation?

Student 24 – Ref #2 above

- What is the model system, phenotypic change, and environmental factor?
- What epigenetic change was observed?
- How did the environment, epigenetics and genetics integrate?

Student 25 – Ref #3 above

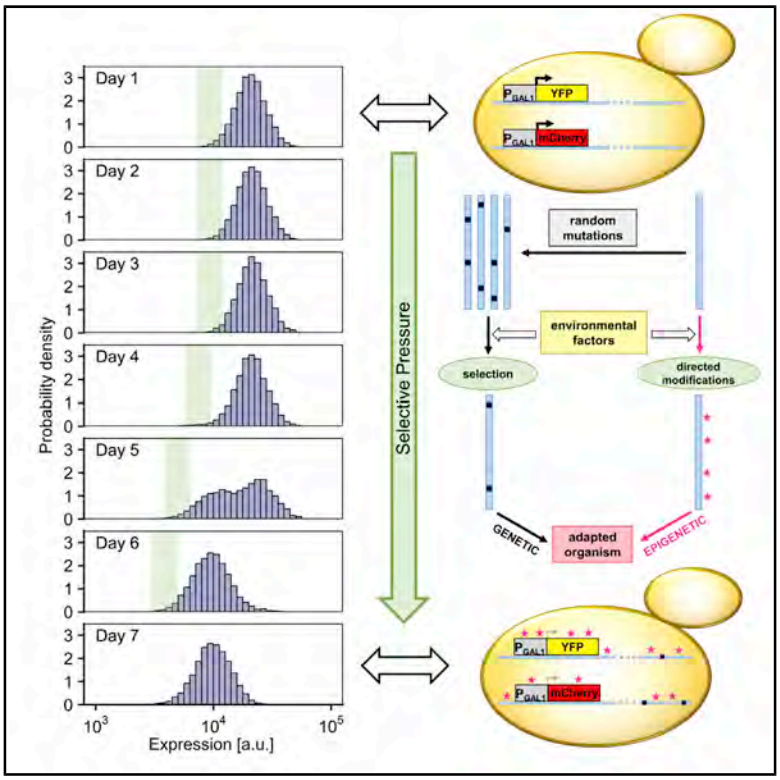
- What was the experimental design and approach?
- What molecular alterations were observed in what cell types?
- What molecular mechanism can promote rapid evolutionary events?

Cell Reports

Article

Epigenetic Mechanisms Contribute to Evolutionary Adaptation of Gene Network Activity under Environmental Selection

Graphical Abstract



Highlights

- Gated sorting and regrowth of low-expression cells causes sustained expression decrease
- Chromatin environment affects expression changes in response to selective pressure
- Genetic causes alone cannot explain the expression reduction under selection
- Inheritance of epigenetic factors contributes to the observed phenotypic change

Authors

Xinyue Luo, Ruijie Song, David F. Moreno, Hong-Yeoul Ryu, Mark Hochstrasser, Murat Acar

Correspondence

murat.acar@yale.edu

In Brief

Luo et al. demonstrate how epigenetic mechanisms contribute to the evolution of gene network activity. Subjecting yeast cells to repeated environmental selection based on the activity of the galactose network, they observe sustained changes in reporter expression level. They characterize the epigenetic and genetic factors contributing to the observed phenotypes.

Luo et al., 2020, Cell Reports *33*, 108306
October 27, 2020 © 2020 The Author(s).
https://doi.org/10.1016/j.celrep.2020.108306

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Epigenetic Mechanisms Contribute to Evolutionary Adaptation of Gene Network Activity under Environmental Selection

Xinyue Luo,[1,2,6] Ruijie Song,[2,3,6] David F. Moreno,[1,2] Hong-Yeoul Ryu,[4,7] Mark Hochstrasser,[1,4] and Murat Acar[1,2,3,5,8,*]
[1]Department of Molecular Cellular and Developmental Biology, Yale University, 219 Prospect Street, New Haven, CT 06511, USA
[2]Systems Biology Institute, Yale University, 850 West Campus Drive, West Haven, CT 06516, USA
[3]Interdepartmental Program in Computational Biology and Bioinformatics, Yale University, 300 George Street, Suite 501, New Haven, CT 06511, USA
[4]Department of Molecular Biophysics and Biochemistry, Yale University, 266 Whitney Avenue, New Haven, CT 06520, USA
[5]Department of Physics, Yale University, 217 Prospect Street, New Haven, CT 06511, USA
[6]These authors contributed equally
[7]Present address: School of Life Sciences, BK21 Plus KNU Creative BioResearch Group, College of National Sciences, Kyungpook National University, Daegu 41566, Republic of Korea
[8]Lead Contact
*Correspondence: murat.acar@yale.edu
https://doi.org/10.1016/j.celrep.2020.108306

SUMMARY

How evolution can be facilitated by epigenetic mechanisms has received refreshed attention recently. To explore the role epigenetic inheritance plays in evolution, we subject isogenic wild-type yeast cells expressing P_{GAL1}-YFP (yellow fluorescent protein) to selection by daily sorting based on reporter expression. We observe expression-level reductions in multiple replicates sorted for the lowest expression that persist for several days, even after lifting the selection pressure. Reduced expression is due to factors in the galactose (GAL) network rather than global factors. Results using a constitutively active GAL network are in overall agreement with findings with the wild-type network. We find that the local chromatin environment of the reporter has a significant effect on the observed phenotype. Genome sequencing, chromatin immunoprecipitation (ChIP)-qPCR, and sporulation analysis provide further insights into the epigenetic and genetic contributors to the expression changes observed. Our work provides a comprehensive example of the role played by epigenetic mechanisms on gene network evolution.

INTRODUCTION

Since Charles Darwin published *On the Origin of Species* in 1859, the concept of evolution by natural selection has occupied a prominent place in modern biology (Darwin, 1859). Darwin himself, of course, had no knowledge of the molecular details of this process: DNA would not be established as the genetic material for another 85 years (Avery et al., 1944). The neo-Darwinian evolution theory combines modern knowledge of genetics and molecular biology with Darwin's thinking (Olson-Manning et al., 2012), but classic neo-Darwinian evolution theory is focused on genetics as the primary molecular mechanism and has substantial difficulties with the fact that beneficial genetic mutations occur at an extremely low rate (Day and Bonduriansky, 2011; Jablonka and Raz, 2009; Kuzawa and Thayer, 2011; Nei and Nozawa, 2011), to the point where some evolutionary biologists have called for a rethinking of the entire evolutionary theory (Laland et al., 2014).

The concept of inheritance of acquired characteristics is frequently attributed to Lamarck (Skinner, 2015), though perhaps inaccurately (Burkhardt, 2013). Nonetheless, the so-called neo-Lamarckian theory, grounded on epigenetic mechanisms, has received increased attention recently (Burggren, 2014; Day and Bonduriansky, 2011; Skinner et al., 2015). A key postulate of the neo-Lamarckian theory is that environment directly alters phenotype generationally (Figure 1A) (Skinner, 2015). In this context, epigenetic mechanisms can be the mediator for the environment to directly alter phenotypic variation and its subsequent inheritance (Skinner, 2015).

Evolutionary consequences of epigenetic inheritance have been studied in recent years, showing how epigenetic control of gene expression affects adaptation (Bódi et al., 2017; Bonduriansky and Day, 2009; Bonduriansky et al., 2012; Halfmann et al., 2012; Klironomos et al., 2013; Kronholm and Collins, 2016; Stajic et al., 2019). Nongenetic inheritance can be mediated in several ways, such as by the inheritance of epigenetic states, cytoplasmic factors, and nutrients (Bonduriansky and Day, 2009). Nongenetic inheritance and its evolutionary implications have been conceptualized in a general framework, showing that by decoupling phenotypic change from the genotype, nongenetic inheritance could circumvent the limitations of genetic inheritance (Bonduriansky and Day, 2009). Nongenetic and genetic inheritance mechanisms

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

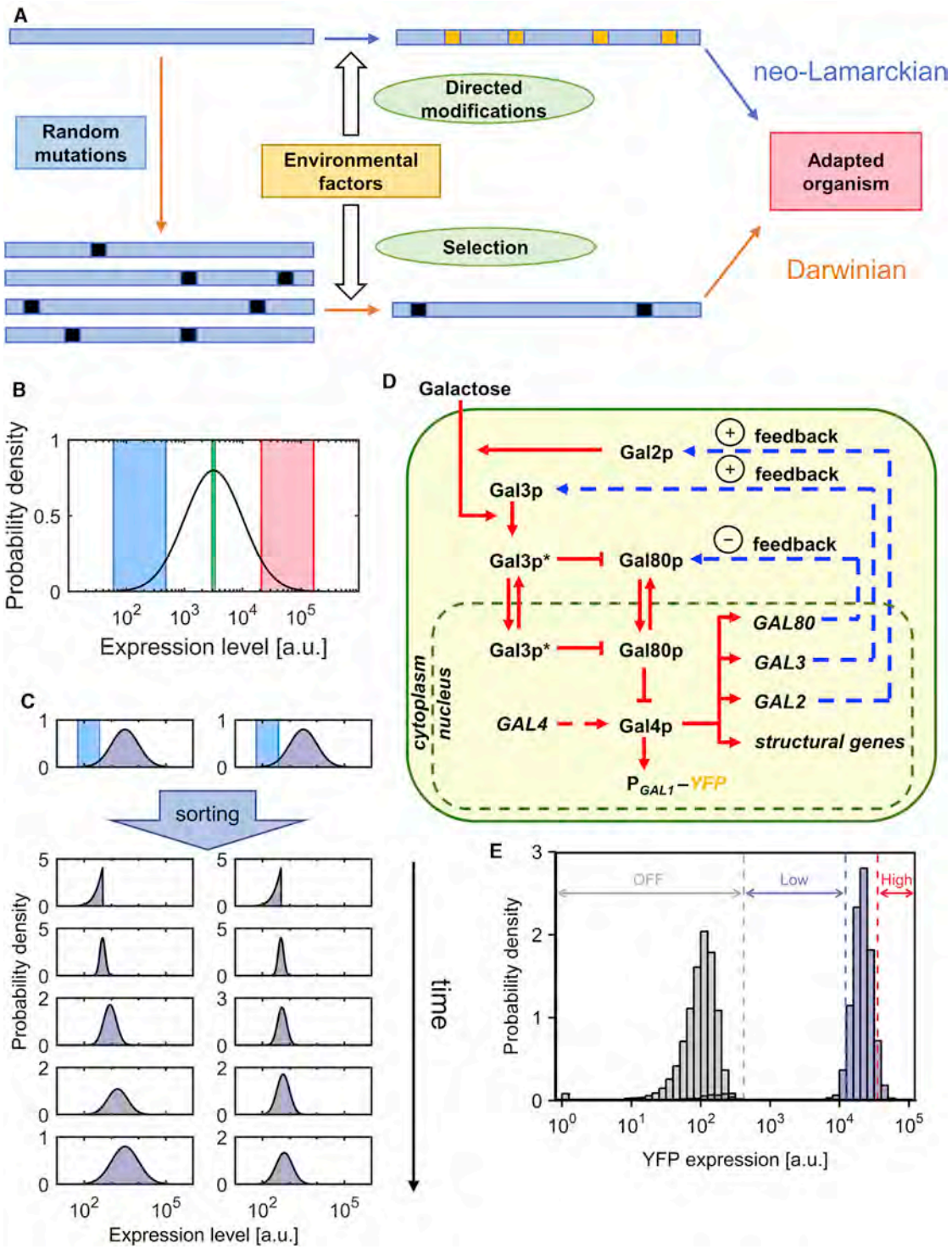


Figure 1. Evolutionary Models, Study Design, and Model Network

(A) Illustration of Darwinian (orange arrows) and neo-Lamarckian (blue arrows) models of evolution. Blue rectangles represent DNA. Black and yellow squares represent genetic mutations and epigenetic modifications, respectively.

(B) Illustration of the three possible sorting gates used: lowest 5% (blue), middle 5% (green), and highest 5% (red).

(C) Illustration of two possible outcomes of the sorting experiment. After being sorted for the lowest-expressing cells, the initial sharp distribution of expression levels will gradually relax, but it may relax either back to the same distribution as the original (left) or to a different distribution with a lower mean (right).

(legend continued on next page)

are not mutually exclusive. For example, theoretical predictions suggested that nongenetic inheritance could increase the rate of both phenotypic and genetic change (Bonduriansky et al., 2012). Also, theoretical and computational work showed how the interplay of heritable epigenetic changes with genetic changes could affect adaptive evolution (Klironomos et al., 2013) and how the effect of epigenetic mutations on adaptive walks depended on their stability and fitness effects relative to genetic mutations (Kronholm and Collins, 2016).

In addition to this theoretical work, experimental studies further focused on the evolutionary consequences of nongenetic heterogeneity and inheritance across generations (Acar et al., 2005, 2008; Bódi et al., 2017; Chatterjee and Acar, 2018; Halfmann et al., 2012; Huang, 2009; Peng et al., 2015; Stajic et al., 2019; Tyedmers et al., 2008; Xue and Acar, 2018a, 2018b). For example, yeast prion proteins can act as epigenetic elements of inheritance (Halfmann et al., 2012), and it has been hypothesized that the yeast prion [*PSI*$^{+}$] provides a mechanism to increase survival in fluctuating environments (Tyedmers et al., 2008); it has also been shown that prions are a common mechanism for phenotypic inheritance in wild strains of *Saccharomyces* (Halfmann et al., 2012). In another experimental study focusing on heterogeneity, it has been shown that phenotypic heterogeneity facilitates adaptive evolution, with the heterogeneity being an evolving trait when populations are under chronic selection pressure (Bódi et al., 2017). As the final example, when tuning low, intermediate, and high levels of heritable silencing of a reporter under selection by insertion within silent subtelomeric yeast chromatin, epigenetic gene silencing has been found to alter the mechanisms and rate of evolutionary adaptation (Stajic et al., 2019).

The concepts of epigenetic inheritance and memory are tightly linked and often used interchangeably to refer to non-DNA-based inheritance (Bonduriansky and Day, 2009). While epigenetic inheritance refers to the passage of certain epigenetic marks to the offspring (Lacal and Ventura, 2018), epigenetic memory is defined as the process of establishing and maintaining a heritable transcriptional state (Acar et al., 2005; Kaufmann et al., 2007; Kundu et al., 2007). Work from the van Oudenaarden group (Acar et al., 2005) described transcriptional memory in the yeast galactose (GAL) network by showing that yeast cells "remember" whether they were previously exposed to high or low concentrations of galactose. Using an engineered GAL network (Acar et al., 2005) where single yeast cells switch between "on" and "off" states of the network, another study (Kaufmann et al., 2007) from the same group measured inheritance of the dynamic gene-expression state and found that several generations after cells have separated, many closely related cell pairs switched with high degrees of synchrony. Providing mechanistic insights into the mediation of epigenetic memory in the GAL network, one study (Kundu et al., 2007) showed that the rate of transcriptional induction of *GAL1* was regulated by the prior expression state; the epigenetic state was inherited by daughter cells, and the SWI/SNF chromatin remodeling enzyme was essential for *GAL1* epigenetic memory. Another study (Brickner et al., 2007) demonstrated that the yeast *GAL1* gene is recruited to the nuclear periphery upon transcriptional activation, and it remains at the periphery for generations after it is repressed, with localization at the periphery serving as a form of memory of recent transcriptional activation. Finally, Tzamarias and colleagues (Zacharioudakis et al., 2007) further showed that the residual activity of the *GAL1*-encoded galactokinase preserves memory in progeny cells by rapidly turning on the Gal4p activator upon cells' re-exposure to galactose.

Despite these studies, a comprehensive example of the role played by neo-Lamarckian epigenetic mechanisms on evolution in the context of a gene network has been lacking. Here, we directly explore the role epigenetic inheritance plays in short-timescale microevolution. We subjected yeast cells to repeated environmental selection based on the expression level of a fluorescent protein reporting on the activity of the canonical GAL network (Figure 1D) (Acar et al., 2005, 2008, 2010; Elison et al., 2018; Luo et al., 2018) over a period of 7 days. We observed reductions in expression level in multiple replicates sorted for the lowest expression that persisted even after the selection pressure was lifted. Using whole-genome sequencing (WGS), chromatin immunoprecipitation (ChIP)-qPCR experiments, and sporulation analysis, we characterized the epigenetic and genetic factors contributing to the persistent expression-level reductions observed.

RESULTS

Applying Environmental Selection on WT GAL Network Activity

To explore the role epigenetic inheritance (Bintu et al., 2016; Bird, 2002; Kouzarides, 2007; Li and Zhang, 2014; Zhou et al., 2011) may play in short-timescale microevolution, we designed an experiment in which a population of isogenic wild-type (WT) yeast cells expressing the yellow fluorescent protein (YFP) under the *GAL1* promoter is subjected to repeated environmental selection in the form of daily sorting based on the expression level of the reporter as measured by flow cytometry. During a 7-day period and corresponding to approximately 101 generations with a 100-min doubling time, the cells were sorted daily based on YFP expression level, and only cells whose expression level is within a particular range (either the lowest 5%, the middle 5%, or the highest 5%; Figure 1E) were selected and allowed to grow further in the same environment (Figures 1B and 1C).

Immediately after sorting, the expression-level distribution of the selected cells is extremely narrow, and it gradually relaxes over time, either to the original distribution if the sorting procedure had no lasting effect on the expression level or to a different distribution with different statistical properties (Figure 1C). By monitoring the expression-level distributions over the 7-day sorting period, one can discern if the sorting intervention had any impact on the expression of the reporter. To determine if any change in the reporter expression observed was transient or lasting, immediately after the 7-day sorting period, the population was grown 3 additional days free from selection pressure

(D) Gene network architecture of the WT yeast GAL network whose activity is reported using a P_{GAL1}-YFP reporter.
(E) Sorting gate determination for WT cells. The off-peak position is determined using expression data measured from uninduced cells (gray bars); the sorting gates in the induced samples (purple bars) are then selected considering only "on" cells.

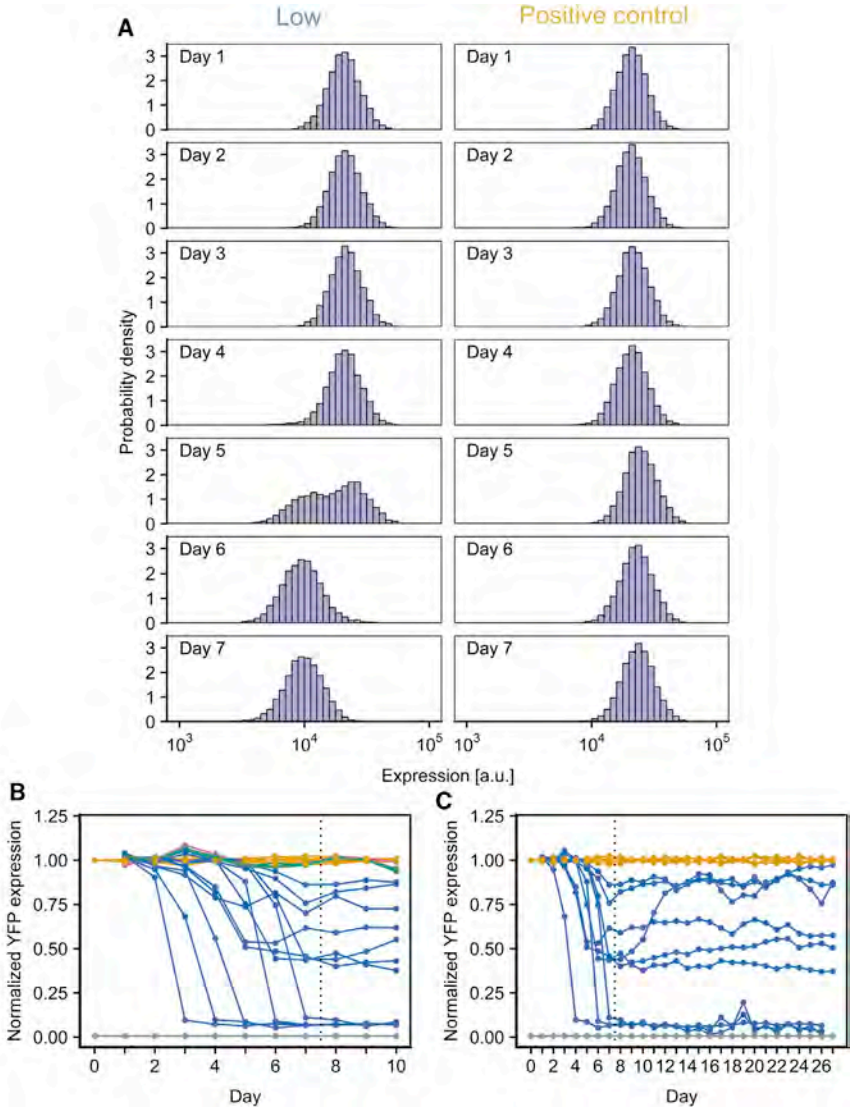


Figure 2. Applying Environmental Selection on WT GAL Network Activity

(A) Sample YFP expression distributions for one of the WT replicates sorted for the lowest expression (left) compared to the positive control (right) over the first 7 days of the experiment.

(B) Normalized YFP expression levels from the YFP-sorting experiment in WT cells. Pink indicates samples sorted for the highest expression, green indicates samples sorted for the median expression, blue indicates samples sorted for the lowest expression, orange indicates positive control, and gray indicates negative control. The dashed line indicates the time at which expression-level-based sorting is terminated. All expression levels are normalized to the corresponding positive control.

(C) Normalized YFP expression levels from the extended YFP-sorting experiment in WT cells. Blue indicates samples sorted for the lowest expression, orange indicates positive control, and gray indicates negative control. The dashed line indicates the time at which expression-level-based sorting is terminated. All expression levels are normalized to the corresponding positive control.

to see if the expression-level distribution reverted back to the original after the selection pressure was lifted.

We found no significant expression-level changes in cells sorted for the middle or highest expression levels (Figure 2B). It is unsurprising that cells already at the middle expression levels retained their character. Given the already high expression levels from the *GAL1* promoter, the lack of change in the cells sorted for the highest expression level may be simply because there is little room for it to increase further. On the other hand, all 12 biological replicates sorted for the lowest expression levels displayed marked reduction in expression (Figures 2A and 2B) to varying degrees that persisted during the 3-day selection-free growth period. Nine replicates were grown (free of selection) for a further 16 days (approximately 230 generations), and 8 of the 9 retained the expression-level reduction (Figure 2C).

To better understand the possible causes driving the observed expression-level reduction, we introduced a second fluorescent reporter protein, mCherry, into the cells. We constructed two strains in which mCherry was either driven by the *TEF1* promoter or the *GAL1* promoter and performed the same YFP-sorting experiment on these strains. We found no change in the expression level of mCherry in the P_{TEF1}-mCherry strain (Figure S1) but significant reduction in the mCherry expression level in the P_{GAL1}-mCherry strain (Figure 3). Further, the level and timing of mCherry expression-level reduction in the P_{GAL1}-mCherry strain was synchronized with that of YFP (Figure 3C). We therefore conclude that the observed reduction in expression was likely due to some factor specific to the GAL network rather than a global factor that can be expected to also affect the expression from P_{TEF1}-mCherry.

Dissecting System Behavior in the Constitutively Active GAL Network

The natural GAL network contains a number of interacting regulators forming feedback loops (Acar et al., 2010; Peng et al., 2016). This complicates the interpretation of the results. For example, the synchronized reduction in double-reporter expression could be due to the dynamics of the epigenetic regulation of the *GAL1* promoter activity, or it could be due to the upstream regulatory elements in the WT GAL network. To eliminate such complication, we deleted the *GAL80* gene—which codes for a repressor through which other GAL network regulators exert their effects—effectively converting the P_{GAL1} promoter into a constitutive promoter dependent only on the transcription factor Gal4p (Figure 4A) and repeated the YFP-sorting experiments in

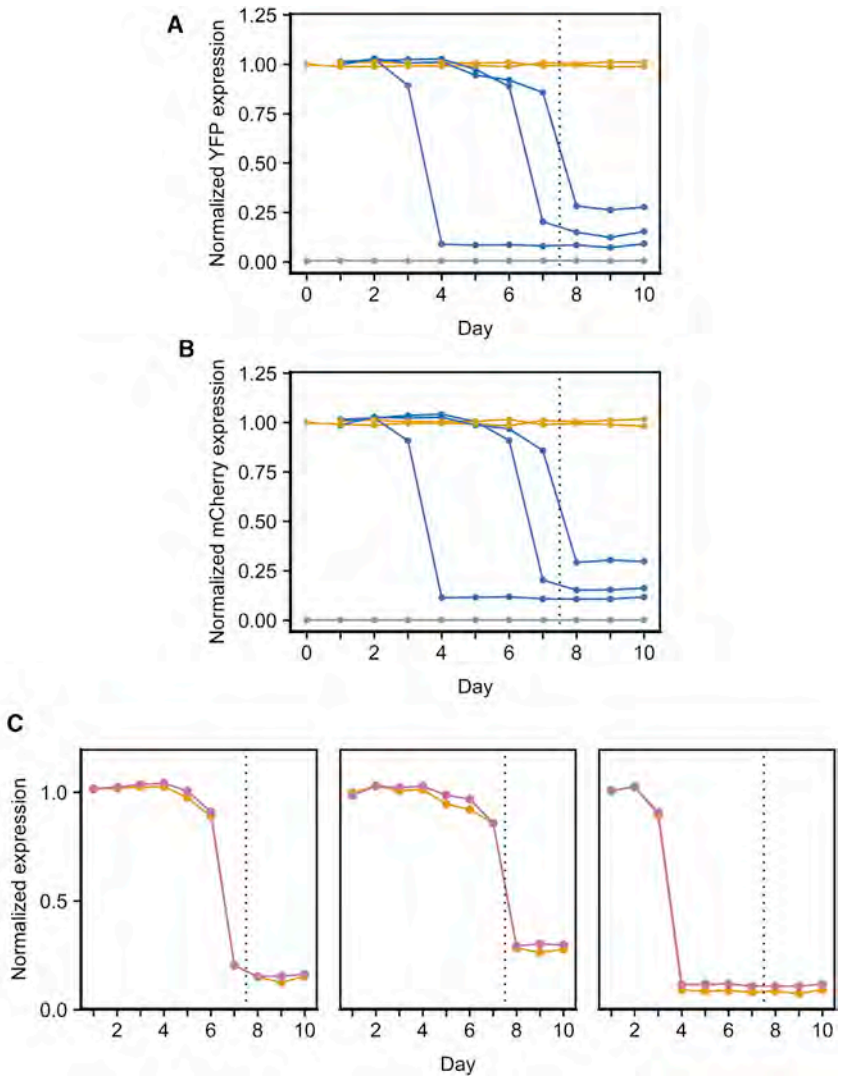


Figure 3. Expression Dynamics from the YFP-Sorting Experiment in WT Cells Containing P_{GAL1}-YFP and P_{GAL1}-mCherry

(A and B) Normalized YFP (A) and mCherry (B) expression levels from the YFP-sorting experiment in WT cells containing P_{GAL1}-YFP and P_{GAL1}-mCherry. Blue indicates samples sorted for the lowest expression, orange indicates positive control, and gray indicates negative control. The dashed line indicates the time at which expression-level-based sorting is terminated. All expression levels are normalized to the positive control.

(C) Comparison of YFP (orange) and mCherry (pink) expression-level trajectories of the three samples sorted for the lowest expression in (A) and (B). Each subpanel represents one sample. The dashed line indicates the time at which expression-level-based sorting is terminated. All expression levels are normalized to the positive control.

See also Figure S1.

this strain. A total of nine biological replicates were used, two of which were found to have accumulated mutations in *GAL4* or the reporter during the course of the experiment and were excluded from further analysis (Data S1).

We found that most biological replicates in which the cells were sorted for the lowest-YFP-expressing cells continued to display a downward shift in mean YFP expression level, although the extent of the shift varied from replicate to replicate, and in one replicate there was no significant change (Figures 4B and 4C, blue curves); five replicates were further grown free of selection for 8 additional days (approximately 115 generations), and all retained their reduced expression levels during this selection-free period (Figures 4E and 4F). Cells sorted for the middle or highest YFP expression did not display any significant changes in either YFP or mCherry expression (Figures 4B and 4C, green and pink curves), just like WT cells. Measuring the doubling times of the cells belonging to colonies isolated from the biological replicate that displayed the largest downward shift in YFP expression (~70%) showed a small increase in doubling times for five colonies isolated from the 7th-day culture, compared to the selection-free positive control (Figure S2A).

We also observed a downward shift in mCherry expression level despite sorting cells in the YFP channel. In all but one of the biological replicates, the YFP and mCherry expression levels were in agreement, but in one replicate, the level of expression-level reduction was significantly different (Figure 4D), though the timing of reduction was similar. This suggests that at least two underlying mechanisms are in play. One mechanism affects the *GAL1* promoter activity generally, driving the synchronized behavior seen both here and previously (Figure 3). But another mechanism, apparently specific to the P_{GAL1}-YFP reporter, must exist that drives the divergence in expression-level reduction between the two reporters, as seen in the last biological replicate.

Measuring Noise Dynamics under Environmental Selection

To see how the selection pressure potentially influences the expression heterogeneity, we next examined the level of noise in P_{GAL1}-YFP and P_{GAL1}-mCherry expression when *gal80Δ* cells were sorted throughout the 7-day period. We discerned no change in the level of P_{GAL1}-YFP noise (coefficient of variation [CV]) in the positive control (no gating) sample or when cells were sorted for the middle or highest expression (Figure S3A).

However, we observed an increase in noise in several samples when cells were sorted for the lowest expression (Figure S3A, blue). Like the change in the expression level itself, this noise increase was stable when the cells were grown selection-free for an additional 10 days after the 7-day sorting period (Figures S3C and S3D, blue). Moreover, in two samples, an accompanying

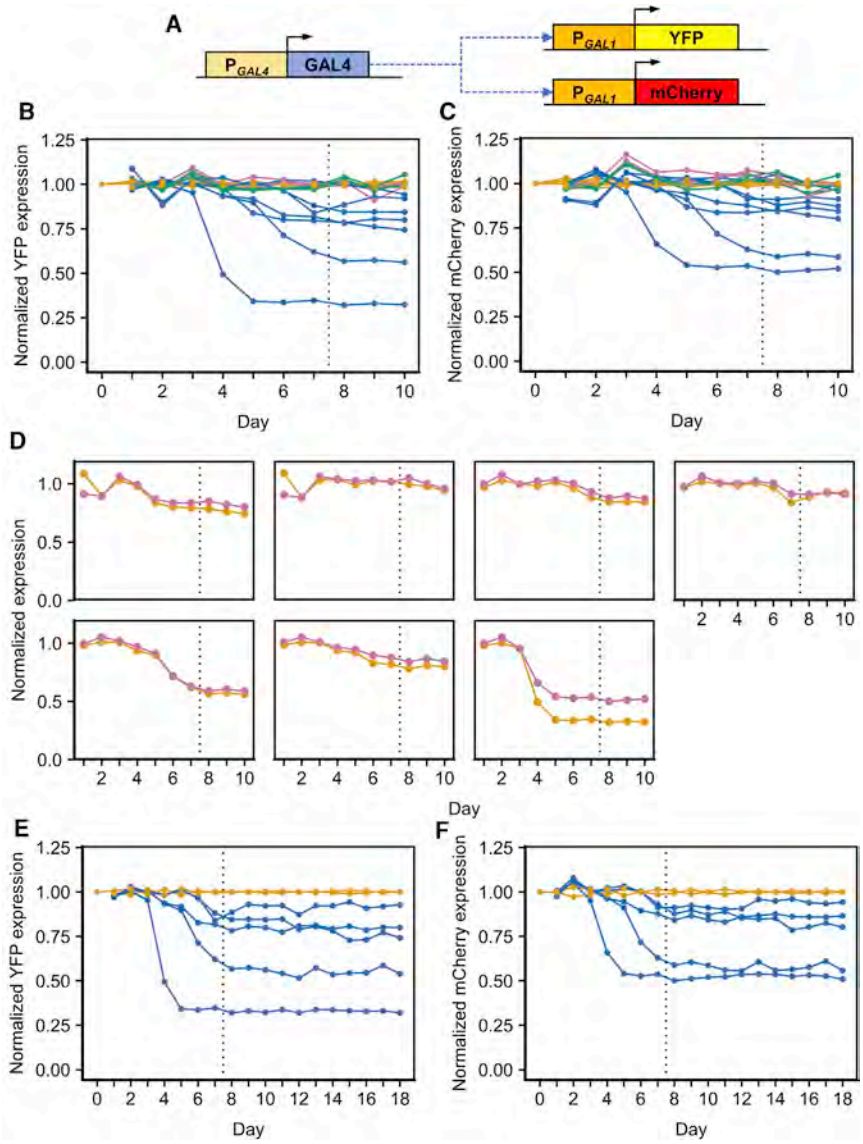


Figure 4. Dissecting System Behavior in the Constitutively Active GAL Network

(A) Gene network architecture of the *gal80Δ* strain.

(B and C) Normalized YFP (B) and mCherry (C) expression levels from the YFP-sorting experiment in *gal80Δ* cells. Pink indicates samples sorted for the highest expression, green indicates samples sorted for the median expression, blue indicates samples sorted for the lowest expression, and orange indicates positive control. The dashed line indicates the time at which expression-level-based sorting is terminated. All expression levels are normalized to the positive control.

(D) Comparison of YFP (orange) and mCherry (pink) expression-level trajectories of the nine samples sorted for the lowest YFP expression in *gal80Δ* cells. Each subpanel represents one sample. The dashed line indicates the time at which expression-level-based sorting is terminated. All expression levels are normalized to the positive control.

(E and F) Normalized YFP (E) and mCherry (F) expression levels from the extended YFP-sorting experiment in *gal80Δ* cells. Blue indicates samples sorted for the lowest expression, and orange indicates positive control. The dashed line indicates the time at which expression-level-based sorting is terminated. All expression levels are normalized to the positive control.

See also Figures S2–S5 and Data S1, S2, S3, and S4.

increase in P_{GAL1}-mCherry noise can be seen when cells were sorted for the lowest P_{GAL1}-YFP expression (Figures S3A–S3D). Under the extreme selection pressure applied during the sorting process, it is unsurprising that higher levels of noise in the protein whose expression is under selection (YFP) may prove evolutionarily beneficial (by increasing the number of cells having reduced expression levels and hence selected during the sorting process). On the other hand, it is likely not advantageous to have higher noise in the expression level of mCherry—or, as a proxy, the structural genes of the *GAL* network, which are responsible for metabolizing the GAL taken from the static environment. Diverging from the optimal level of *GAL* network expression in the environment carries a fitness cost (and higher noise means that more cells are diverging from the optimal level), which could explain why the noise level does not display the same degree of synchronization behavior as the expression-level reduction.

The Effect of the Local Chromatin Environment on Observed Results

To understand the potential influence of genomic loci on the level of expression-level reduction, we performed the same sorting experiment but based on the expression level of P_{GAL1}-mCherry (which was integrated into the *ura3* locus) rather than P_{GAL1}-YFP (which was in the *ho* locus). A total of nine biological replicates were used, two of which were found to have accumulated mutations in *GAL4* or the reporter during the experiment and were excluded from further analysis. Measuring the resulting *GAL1* promoter activity levels during the 7-day sorting period as before, we found expression-level reduction to be significantly more difficult to achieve (Figures 5A–5C), if not impossible, compared to sorting when the reporter cassette is integrated into the *ho* locus, with only one biological replicate out of seven displaying a significant reduction in expression compared to the positive control. We similarly quantified the level of noise in P_{GAL1}-YFP and P_{GAL1}-mCherry expression during these mCherry-sorting experiments (Figures 5D and 5E). We did not detect substantial and persistent changes in noise levels of the kind we had seen previously (Figures S3A and S3B).

Together, these observations suggest that the genomic locus at which the cassette is integrated, and hence the local chromatin structure and epigenetic markers, has a significant effect

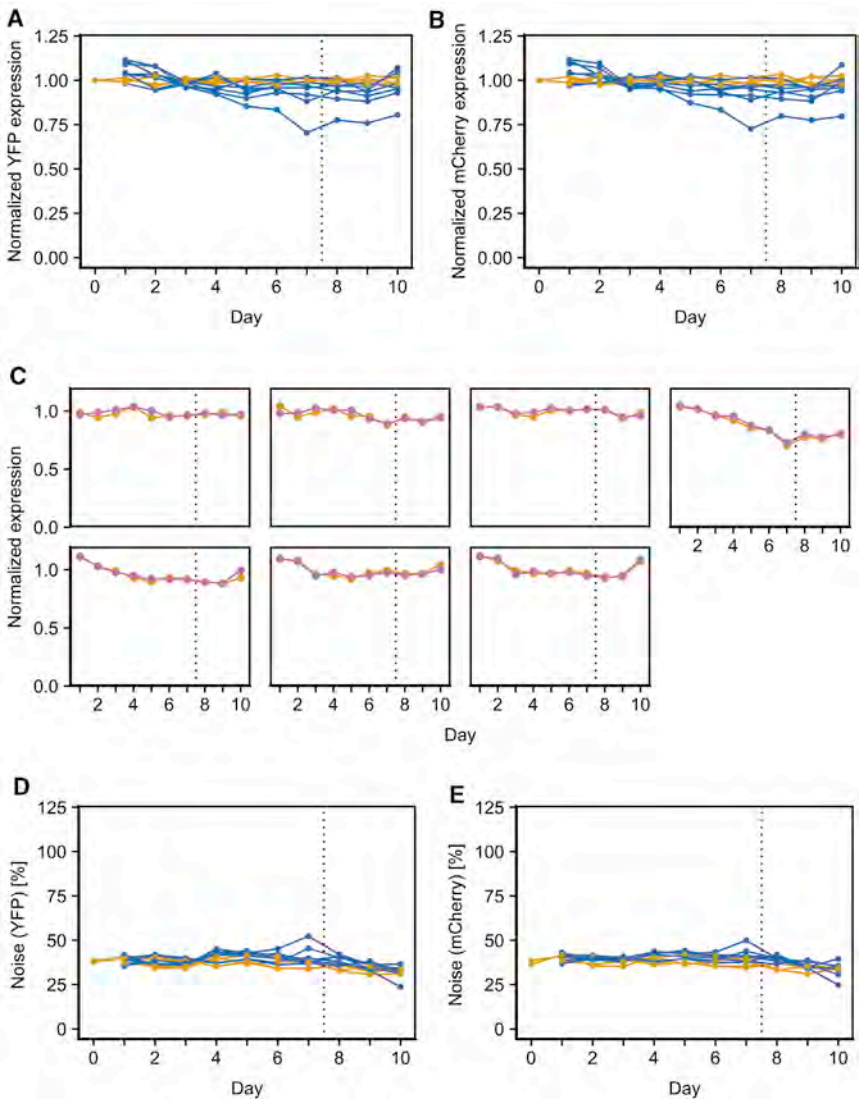


Figure 5. The Effect of the Local Chromatin Environment on Observed Results

(A and B) Normalized YFP (A) and mCherry (B) expression levels from the mCherry-sorting experiment in *gal80Δ* cells. Blue indicates samples sorted for the lowest expression, and orange indicates positive control. The dashed line indicates the time at which expression-level-based sorting is terminated. All expression levels are normalized to the positive control.

(C) Comparison of YFP (orange) and mCherry (pink) expression-level trajectories of the nine samples sorted for the lowest mCherry expression in *gal80Δ* cells. Each subpanel represents one sample. The dashed line indicates the time at which expression-level-based sorting is terminated. All expression levels are normalized to the positive control.

(D and E) Noise in YFP (D) and mCherry (E) expression levels from the mCherry-sorting experiment in *gal80Δ* cells. Blue indicates samples sorted for the lowest expression, and orange indicates positive control. The dashed line indicates the time at which expression-level-based sorting is terminated.

on the phenotype we observed. The *ho* locus appears to be significantly more susceptible to experiencing expression-level reduction in response to the selection pressure we applied compared to the *ura3* locus. Especially given that the genetic mutation rate appears to be approximately constant between the two experiments (in both cases, two out of nine biological replicates were found to have accumulated mutations in the relevant genes), the diverging outcomes strongly suggests that a non-genetic mechanism is involved in suppressing the YFP expression at the *ho* locus.

As noted, sequencing detected no mutations in *GAL4* or the two reporter cassettes (totaling approximately 8 kbp) in the biological replicates under consideration. In addition, we did not detect any substantial fitness changes in the sorted populations passed from one day to the next; if anything, the sorted populations divide slightly slower than unsorted cells (Figure S2). We therefore hypothesized that the observed downward shifts in expression level were due to epigenetic changes in the transcription factor gene *GAL4* and/or in the reporter promoters: the accumulated epigenetic changes "lock" the chromatin into a closed state and are enriched by the daily sorting process. Given the experimental observations, such locks were then necessarily strong enough to persist through hundreds of generations of selection-free growth.

WGS to Explore Genetic Causes of Observed Phenotypes

Next, we performed WGS to evaluate any contributions from global genetic factors on the observed reduction in YFP expression levels. For this, we focused on two biological replicates—the FL6 and FL9 populations in the *gal80Δ* background (STAR Methods)—from which we had seen significant reduction in P_{GAL1}-YFP and P_{GAL1}-mCherry expression on Day7, compared to Day0 expression levels, after gated sorting in the YFP channel. Our local sequencing of the Day7 FL6 and FL9 populations in the reporter cassettes and the *GAL4* region did not identify any mutations.

We isolated five single colonies from the FL6 and FL9 populations on Day7 and randomly selected two single colonies from each of the two groups for performing WGS on them. As controls, we also included in these WGS characterizations two randomly selected single colonies isolated from the Day0 population, as well as two randomly selected single colonies isolated from the positive control group on Day7. Results obtained from the sequencing of each isolated colony were compared to those obtained from the sequencing of the Day0 colonies (Data S2, S3, and S4). While mutations in intergenic promoter regions may also

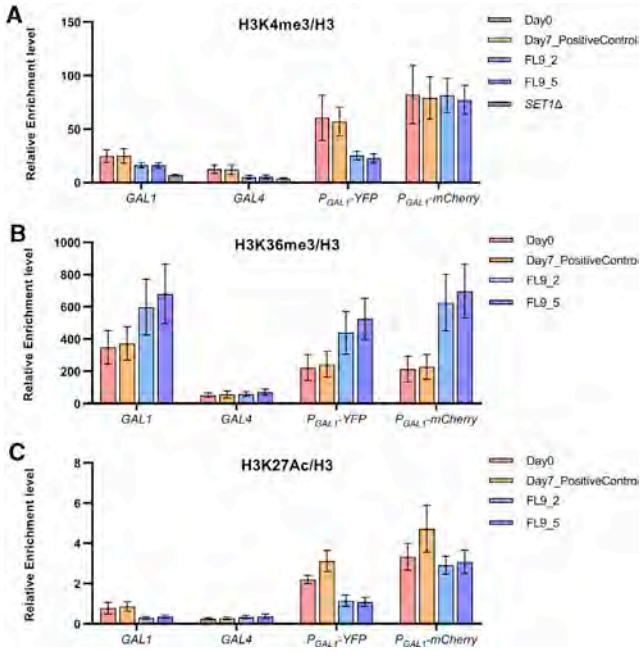


Figure 6. Epigenetic Modification Levels Quantified by ChIP-qPCR

Three types of epigenetic marks—H3K4me3 (A), H3K36me3 (B), and H3K27Ac (C)—were characterized at four genetic loci—*GAL1*, *GAL4*, P_{GAL1}-YFP, and P_{GAL1}-mCherry—in four isogenic populations: Day0, Day7 positive control, FL9_2, and FL9_5. An additional strain, *SET1*Δ, was included for H3K4me3 (A) as a technical control. Error bars indicate SEM (n = 4).

See also Figures S6 and S7 and Tables S1 and S2.

have phenotypic consequences, for the sake of interpretability, we focused on the SNPs (single-nucleotide polymorphisms) causing amino acid alterations in open reading frames (ORFs). We then selected the common ORF mutations found in both single colonies isolated from the FL6 and FL9 populations. Five common mutations were identified for the FL6 colonies (in *FEN2*, *GPM2*, *IRA2*, *NUP133*, and *RPN4*), while the FL9 colonies shared three mutations (in *APL1*, *BDS1*, and *SRB8*).

To see the isolated effects of these mutations on the P_{GAL1}-YFP and P_{GAL1}-mCherry expression levels, we attempted to clone them singly and in combination into a single colony isolated from the unevolved Day0 population using CRISPR. For the FL6 group mutations, we successfully cloned the mutations in the *FEN2*, *IRA2*, and *NUP133* genes one at a time and combinatorically, but the mutations in the *GPM2* and *RPN4* genes could not be cloned because of challenges associated with CRISPR. Measuring the P_{GAL1}-YFP and P_{GAL1}-mCherry expression levels in each constructed strain, we did not see any major changes in expression levels caused by the mutations in *FEN2*, *IRA2*, or *NUP133* genes relative to the Day0 isogenic background without these mutations (Figure S4A). Regarding the mutations in the *GPM2* and *RPN4* genes that could not be cloned, *GPM2* is a nonfunctional homolog of *GPM1* phosphoglycerate mutase, and *RPN4* codes for a transcription factor that stimulates expression of proteasome genes. Despite the potential relevance of *RPN4* for the phenotypes we observed, the degradation experiments we performed for the colonies isolated from the FL6 population did not show differential degradation dynamics for YFP or mCherry (Figures S5A and S5B). Nevertheless, we cannot fully exclude the possibility that the mutations in *GPM2* and/or *RPN4* might exert effects on the phenotypes we observed if they were cloned into the Day0 unevolved, single-colony-derived population.

For the mutations on the *APL1*, *BDS1*, and *SRB8* genes of the FL9 group, on the other hand, we combinatorically constructed all eight strains carrying these mutations one at a time and together. Measuring the P_{GAL1}-YFP and P_{GAL1}-mCherry expression levels in each constructed strain, we saw consistent changes in P_{GAL1}-YFP and P_{GAL1}-mCherry expression levels, relative to the Day0 isogenic background without the mutations (Figure S4B). The degradation experiments we performed for the colonies isolated from the FL9 population did not show differential degradation dynamics for YFP or mCherry (Figures S5C and S5D). The mutation in *SRB8*, coding for a subunit of the RNA polymerase II (RNA Pol II) mediator complex, led to 75% and 50% reductions in the P_{GAL1}-YFP and P_{GAL1}-mCherry expression levels, respectively, across all strains carrying the mutation. Since these expression reductions in levels of YFP and mCherry are very similar to the ones observed at the end of the 7-day sorting period, the mutation in the *SRB8* gene can account for the phenotypic changes observed in one of the biological replicates (sorting group FL9). However, the differential effects of the *SRB8* mutation on YFP and mCherry expression levels indicate that integration-locus-specific epigenetic factors still play a role on the main phenotype of gene expression reduction under environmental selection.

Measuring Acetylation and Methylation Levels on System Components

To further investigate the effect of epigenetic factors on the difference in expression-level decrease between the two reporters as a result of the *SRB8* mutation, we examined the chromatin modification levels at the *GAL4*, *GAL1*, P_{GAL1}-YFP, and P_{GAL1}-mCherry loci in the two WGS-characterized isogenic colonies of the FL9 group (FL9_2 and FL9_5), as well as isogenic colonies isolated from the Day0 population and Day7 positive control population. For this, we tested for three different types of histone modifications via ChIP-qPCR: trimethylation of histone H3 lysine 4 (H3K4me3), which positively correlates with transcriptional activity; trimethylation of histone H3 lysine 36 (H3K36me3), which represses transcription and is known to be associated with HDAC (Histone DeACetylase) recruitment; and acetylation of histone H3 lysine 27 (H3K27ac), which is associated with active transcription.

As expected based on the YFP and mCherry expression levels, ChIP-qPCR results from the Day0 and Day7 positive control colonies showed higher overall H3K4me3 and H3K27Ac levels but lower H3K36me3 levels, compared to the results observed from the FL9_2 and FL9_5 colonies (Figures 6A–6C; Table S1). More specifically, at the endogenous *GAL1* locus, we saw reductions in H3K4me3 and H3K27Ac levels and an increase in H3K36me3 level in the FL9_2 and FL9_5 colonies, compared to the Day7 positive control colony, suggesting that the local transcriptional activities at the *GAL1* locus in FL9_2 and FL9_5 are lower than in Day0 and Day7 positive control. This is consistent with other observations: in FL9_2 and FL9_5,

the YFP protein level was reduced significantly (approximately 70% reduction compared to the Day7 positive control), and there was a significant decrease in the mRNA level of YFP as quantified by qRT-PCR (Figures S6 and S7). We also observed a comparable trend of change in the epigenetic marks at the P_{GAL1}-YFP reporter, indicating that the three types of chromatin modifications tested are similar on these two genetic elements and that both loci are likely governed by the same epigenetic modification machineries that act on the *GAL1* promoter.

We did not see a clear difference between the FL9 colonies and the Day0 and Day7 control colonies at the *GAL4* locus with respect to the H3K36me3 and H3K27Ac modifications, although there seemed to be some reduction of the H3K4me3 mark in FL9_2 and FL9_5. We interpret this as a consequence of a lack of local chromatin-repressing machinery, considering that *GAL4* is expressed constitutively.

Interestingly, while we observed similar trends in H3K36me3 and H3K27Ac modifications at P_{GAL1}-mCherry compared to *GAL1* and P_{GAL1}-YFP, we saw no differentiating trend in the H3K4me3 modification among the tested colonies at P_{GAL1}-mCherry. This divergence suggests a difference in the local chromatin dynamics between the two reporters. It is possible that the locus where P_{GAL1}-mCherry is placed, *URA3* on chromosome V, has a distinct local chromatin regulatory mechanism that interferes with *GAL1*-promoter-specific regulation. The lack of the H3K4me3 mark at P_{GAL1}-mCherry relative to P_{GAL1}-YFP suggests that transcriptional activity at the former is higher than the latter. Indeed, while YFP protein level in FL9_2 and FL9_5 on Day7 was reduced by 70% relative to Day0, mCherry protein level was reduced by only 50%. Therefore, the difference in expression reduction between the two reporters is associable with the difference in the local chromatin modification levels between the two loci where they are located. Together, these results solidify the role of epigenetic modifications on the expression levels of the two reporters.

Sporulation-Based Assessment of Genetic versus Epigenetic Contributions on the Observed Phenotypes

To rule out the possibility that the observed phenotypes of the evolved strains could be explained on purely genetic grounds, we crossed our evolved strains (FL6_2 and FL9_2) with the equivalent of our unevolved WT strain (Day0) of opposite mating type. As a control, we crossed two unevolved WT strains of opposite mating type. After mating, sporulation, and tetrad dissection, we measured the YFP and mCherry expression levels displayed by the progeny of each cross after growing the cells in the same media conditions as used during our evolution experiments. As expected, all offspring of the WT-to-WT cross showed expression levels very similar to the parental strain (Figure 7A).

The FL6_2-to-WT cross generated offspring that was very heterogeneous in expression, contrary to what would be expected from plausible Mendelian genetic mechanisms (Figure 7B). Moreover, the lack of offspring clustering around the parental FL6_2 strain suggested that the mutations on *GPM2* and *RPN4* were not relevant to the observed phenotypes; were they relevant, Mendelian genetics would predict that half (or a quarter) of the offspring spores would carry the mutations on one (or both) of those genes and display a similar phenotype to that of the parental strain, but the fraction of the spores displaying a similar phenotype to that of the parental strain was actually much lower. Surprisingly, a substantial fraction of the spores displayed reporter expression levels higher than the WT parental strain's expression, especially for the YFP reporter. While there was some correlation (Pearson's r = 0.697) between the YFP and mCherry expression levels, we saw that some spores displayed YFP levels higher than the WT but mCherry levels closer to the level displayed by the evolved FL6_2's low mCherry expression level. Meiosis and sporulation are complex cellular programs involving the creation and repair of double-strand breaks on the DNA in certain recombination hotspots (Kolodkin et al., 1986; Lichten and Goldman, 1995; Smith and Nicolas, 1998), and it is known that epigenetic state influences the meiotic recombination hotspots (Brachet et al., 2012). Therefore, a potential explanation for these expression levels is that epigenetic changes carried by the evolved parental strain might have caused unusual meiotic recombination events in the offspring.

The offspring generated from the FL9_2-to-WT cross displayed three clearly distinct clusters of YFP and mCherry expression: one coincided with the parental WT strain's reporter expressions, one coincided with the parental FL9_2 strain's expressions, and the third cluster displayed a WT-like YFP expression and a FL9_2-like mCherry expression (Figure 7C). The reporter expression composition of the third cluster would not be expected based on a purely genetic inheritance pattern, as all genetic components controlling the network are the same for both reporters. Since the only difference between the two genetic constructs (P_{GAL1}-YFP and P_{GAL1}-mCherry) is their chromosomal integration site, we attribute the differential gene expression to the differential impact of the epigenetic marks between the YFP and the mCherry loci, as we described in the previous section.

To explain these three expression clusters displayed by the offspring of the FL9_2-to-WT cross, we propose a model in which the mutation in the RNA Pol II mediator subunit Srb8 contributes to the reporter's downregulation, but epigenetic marks at the YFP reporter locus make its expression level independent of which *SRB8* allele the cell is carrying (Figure 7D), potentially through an epigenetically facilitated compensation mechanism maintaining the overall progression of RNA Pol II irrespective of the Srb8 subunit activity. The inheritance pattern we observed from the offspring of the FL9_2-to-WT crossing supports the presence of epigenetic modifications leading to WT-like YFP expression in one of the three expression clusters. More specifically, the parental FL9_2 strain, which bears both the *SRB8* mutated allele and differential epigenetic marks at the YFP insertion locus (Figures 6A and 6C), displays low YFP and mCherry expression (Figure 7C); the unevolved strain bearing the mutated *SRB8* allele in a WT-like chromatin environment also displays low YFP and mCherry expression (Figure S4B). Therefore, the epigenetic mark we are proposing to explain the offspring's third cluster should favor WT-like YFP expression, as only this inheritance model would generate a 3:1 inheritance pattern for the YFP expression and a 2:2 pattern for the mCherry expression, which matches our observations (Figures 7E and 7F).

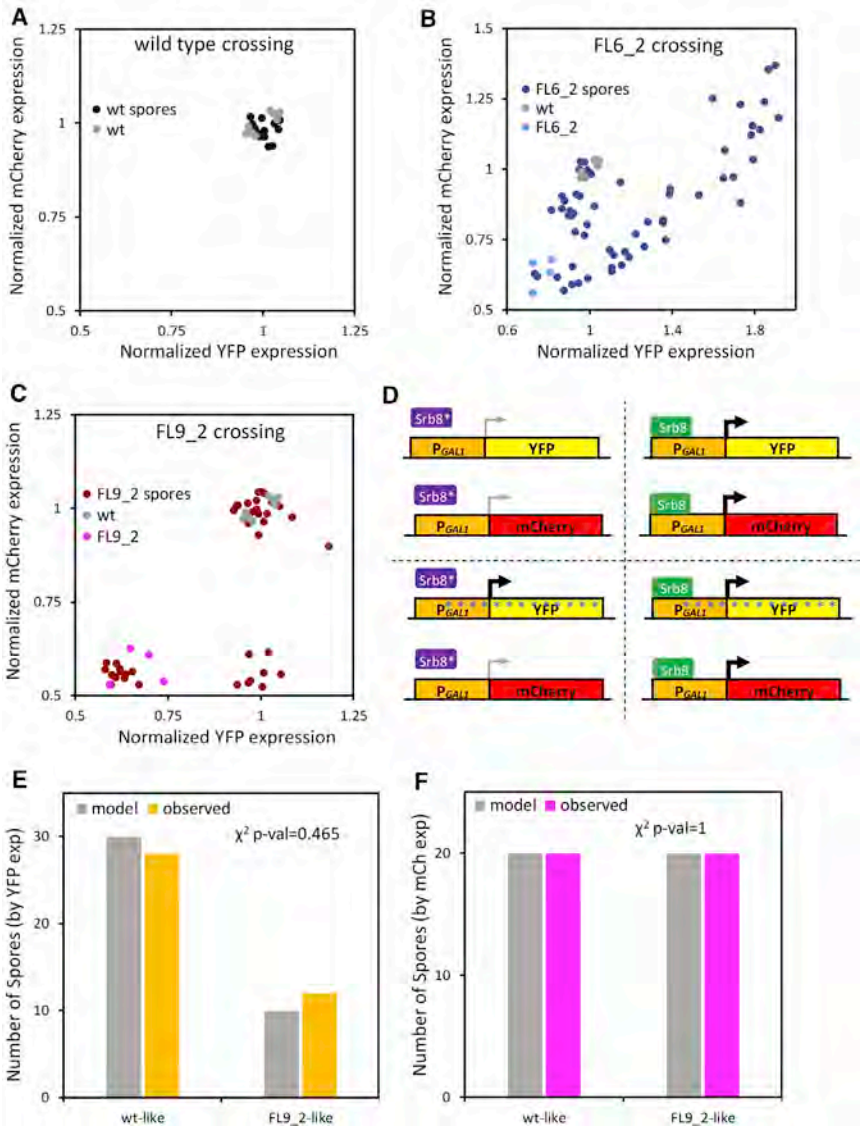


Figure 7. Sporulation-Based Assessment of Genetic versus Epigenetic Contributions on the Observed Phenotypes

(A–C) Scatterplot of the mean measured YFP and mCherry expression displayed by the spores coming from the WT-to-WT crossing (A), the FL6_2-to-WT crossing (B), and the FL9_2-to-WT crossing (C). The expression levels displayed by the parental strains are also shown.

(D) Proposed model explaining the expression distributions of the offspring obtained from the FL9_2-to-WT crossing. The purple Srb8* represents the mutated version identified by WGS, while the green Srb8 represents the WT version. The asterisks along the YFP locus represent the proposed epigenetic mark modulating reporter expression independently from the *SRB8* allele inherited. The color and width of the arrows indicate the strength of the gene expression (thin and gray, FL9_2-like low expression; thick and black, WT-like expression).

(E and F) Bar plots showing the number of spores displaying a given phenotype (WT-like or FL9_2-like), expected by the model and observed experimentally, for the YFP (E) and mCherry (F) expression phenotypes.

Overall, these results indicate that the mechanism behind the observed expression reduction during our evolution experiments cannot be explained only by genetic causes. Epigenetic modifications must be contributing to the differential expression pattern exhibited by the two reporters within the same cell and to the overall evolved phenotype emerging at the end of the evolution process.

DISCUSSION

In this report, we explored the role epigenetic inheritance plays in short-timescale microevolution. We observed reductions in expression level in multiple replicates sorted for the lowest expression that persisted for hundreds of generations, long after the selection pressure was lifted. The amount of decrease in expression level was locus specific, implicating the involvement of local chromatin environment in the process. Performing WGS characterizations on isogenic colonies obtained from two populations, we found that one case of the persistent expression-level reduction was due to genetic factors, while experiments performed for the other case did not indicate a genetic contributor. Measuring the level of chromatin modification marks on system components supported the conclusion that epigenetic regulation differences between integration loci could explain differential YFP and mCherry expression under the same promoter. Finally, results from mating and sporulation experiments provided evidence for the involvement of non-genetic inheritance mechanisms as contributors to the differential expression pattern exhibited by the two reporters in the same cell.

For the replicate that is guided mainly genetically, given that a single mutation in the *SRB8* gene is sufficient to reproduce the decreased YFP and mCherry protein levels measured on Day7, a plausible explanation for this could be that the impaired mRNA synthesis machinery led to a loss in mRNA production and subsequently in protein production in the cell. Since Srb8 is involved in global RNA synthesis, one would expect a global reduction in mRNA levels in the cell as well. We indeed saw a significant reduction in the mRNA level of the housekeeping *ACT1* gene (Figure S7A). Moreover, we found that the isogenic colonies isolated from the FL9 group had significantly increased doubling time compared to both Day0 (~3% increase) and Day7 positive controls (~4% increase), revealing a reduced fitness level that is potentially attributable to inefficient mRNA

synthesis. Despite the genetic contributions to the observed phenotypes, we note that crossing and sporulation analysis showed that the inheritance pattern of this strain's traits was not explainable by a solely genetic mechanism.

Our study involves observation of YFP and mCherry expression dynamics with reporter constructs integrated in different genomic loci. We note that genomic loci differ not only in their local chromatin environment, but also in their sequence context (e.g., presence or absence of certain enhancers), which may be a contributor to expression-level differences between the reporters.

Genetic and epigenetic mechanisms do not have to be mutually exclusive. In response to a particular environmental condition, both kinds of mechanisms can play roles and complement each other. Epigenetic mechanisms generally operate at a shorter timescale than genetic mechanisms, allowing a faster response to changing environmental conditions (Bonduriansky et al., 2012; Burggren, 2016). On the other hand, genetic mechanisms operate at a longer timescale but also produce a more permanent response.

To test whether short-term epigenetic inheritance interacts with genetic change, a recent study (Stajic et al., 2019) used an experimental evolution setup in yeast by tuning low, intermediate, and high levels of heritable silencing of a *URA3* reporter under selection. The authors showed that heritable gene expression through epigenetic chromatin states contributed to adaptive evolution; however, their results and interpretations were not free from mutational effects. More specifically, heritable silencing drove population size expansion and rapid epigenetic adaptation, eventually leading to genetic assimilation of the silent phenotype by mutations. Also, at intermediate or low levels of heritable silencing, the study showed that populations evolved more rapidly by accumulation of adaptive mutations.

Natural environments are usually not fully static, but they fluctuate over time. Memory of previous expression levels, from whatever source, can function as a double-edged sword in fluctuating environments. On one hand, having some memory of the optimal expression level in the current environment confers a fitness advantage in the present (Brickner et al., 2007; Zacharioudakis et al., 2007). On the other hand, locking the expression at a particular level would prevent the cell from responding to environmental changes. Thus, fully persistent memory of expression level would be expected to be detrimental in a fluctuating external environment (Acar et al., 2008; Bódi et al., 2017).

Of the various kinds of heritable factors, genetic mutations are certainly among the most persistent; the cell is full of mechanisms aiming to ensure that genetic materials are faithfully passed from one generation to the next, and moreover, reverting a genetic mutation naturally is even more difficult given the randomness of the mutagenesis process and the rarity of gain-of-function mutations. However, epigenetic mechanisms of reducing gene expression are likely easier to revert when the environment demands it (Klironomos et al., 2013); after all, chromatin is routinely remodeled during the cell cycle (Deniz et al., 2016; Raynaud et al., 2014). Thus, compared to genetic mechanisms, the wide variety of epigenetic mechanisms of regulating gene expression levels are much more easily tuned to environmental demands. While some epigenetic changes may disappear within a few generations (Kundu et al., 2007), others can be highly persistent (Catania et al., 2020). Thus, the epigenetic toolset allows the cell to strike a balance between memorizing gene expression states and being plastic to external environmental changes.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - *Saccharomyces* cerevisiae with the W303 genetic background
- METHOD DETAILS
  - Construction of yeast strains and plasmids
  - Flow cytometry data acquisition and sorting
  - mRNA transcript levels determination by qRT-PCR
  - Local sequencing of key GAL network components
  - Whole Genome Sequencing (WGS) sample preparation
  - Measuring doubling-times of cell populations
  - SNP introduction with CRISPR-Cas9
  - Degradation dynamics of fluorescent proteins
  - ChIP-qPCR experiments
  - Mating, sporulation and tetrad dissection
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Flow cytometry data analysis
  - Whole Genome Sequencing data analysis
  - χ^2 test for checking inheritance models for sporulation outcomes

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2020.108306.

ACKNOWLEDGMENTS

We thank J. Gendron, E. Karatekin, L. Mirny, G.P. Wagner, and Acar Lab members for comments and feedback on various stages of this work. X.L. acknowledges support through the China Scholarship Council-Yale World Scholars Program. M.H. acknowledges funding from the US National Institutes of Health (GM136325). M.A. acknowledges funding from the US National Institutes of Health (1DP2AG050461-01 and 1U54CA209992-01).

AUTHOR CONTRIBUTIONS

X.L., R.S., and M.A. designed the experiments and analyses, interpreted the data and results, and designed and prepared the manuscript. X.L. constructed the strains, performed the experiments, collected the data, and contributed to data analysis. R.S. analyzed the data and performed statistical analyses. X.L., H.-Y.R., and M.H. performed the ChIP-qPCR experiments and helped interpret their results. D.F.M. performed the mating and sporulation experiments, including strain construction; flow-cytometry measurements and analyses; and interpreting, plotting, and writing the results with M.A. D.F.M. also

contributed to the revision of the manuscript. M.H. dissected the tetrads. M.A. conceived and supervised the project. All authors read and approved the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 22, 2019
Revised: August 17, 2019
Accepted: October 2, 2020
Published: October 27, 2020

REFERENCES

Acar, M., Becskei, A., and van Oudenaarden, A. (2005). Enhancement of cellular memory by reducing stochastic transitions. Nature *435*, 228–232.

Acar, M., Mettetal, J.T., and van Oudenaarden, A. (2008). Stochastic switching as a survival strategy in fluctuating environments. Nat. Genet. *40*, 471–475.

Acar, M., Pando, B.F., Arnold, F.H., Elowitz, M.B., and van Oudenaarden, A. (2010). A general mechanism for network-dosage compensation in gene circuits. Science *329*, 1656–1660.

Ahn, S.H., Kim, M., and Buratowski, S. (2004). Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3′ end processing. Mol. Cell *13*, 67–76.

Avery, O.T., Macleod, C.M., and McCarty, M. (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types: Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type iii. J. Exp. Med. *79*, 137–158.

Barbieri, E.M., Muir, P., Akhuetie-Oni, B.O., Yellman, C.M., and Isaacs, F.J. (2017). Precise Editing at DNA Replication Forks Enables Multiplex Genome Engineering in Eukaryotes. Cell *171*, 1453–1467.e13.

Bintu, L., Yong, J., Antebi, Y.E., McCue, K., Kazuki, Y., Uno, N., Oshimura, M., and Elowitz, M.B. (2016). Dynamics of epigenetic regulation at the single-cell level. Science *351*, 720–724.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes Dev. *16*, 6–21.

Bódi, Z., Farkas, Z., Nevozhay, D., Kalapis, D., Lázár, V., Csörgő, B., Nyerges, Á., Szamecz, B., Fekete, G., Papp, B., et al. (2017). Phenotypic heterogeneity promotes adaptive evolution. PLoS Biol. *15*, 1–26.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

Bonduriansky, R., and Day, T. (2009). Nongenetic Inheritance and Its Evolutionary Implications. Annu. Rev. Ecol. Evol. Syst. *40*, 103–125.

Bonduriansky, R., Crean, A.J., and Day, T. (2012). The implications of nongenetic inheritance for evolution in changing environments. Evol. Appl. *5*, 192–201.

Brachet, E., Sommermeyer, V., and Borde, V. (2012). Interplay between modifications of chromatin and meiotic recombination hotspots. Biol. Cell *104*, 51–69.

Brickner, D.G., Cajigas, I., Fondufe-Mittendorf, Y., Ahmed, S., Lee, P.C., Widom, J., and Brickner, J.H. (2007). H2A.Z-mediated localization of genes at the nuclear periphery confers epigenetic memory of previous transcriptional state. PLoS Biol. *5*, e81.

Burggren, W.W. (2014). Epigenetics as a source of variation in comparative animal physiology - or - Lamarck is lookin' pretty good these days. J. Exp. Biol. *217*, 682–689.

Burggren, W. (2016). Epigenetic Inheritance and Its Role in Evolutionary Biology: Re-Evaluation and New Perspectives. Biology (Basel) *5*, 24.

Burkhardt, R.W., Jr. (2013). Lamarck, evolution, and the inheritance of acquired characters. Genetics *194*, 793–805.

Catania, S., Dumesic, P.A., Pimentel, H., Nasif, A., Stoddard, C.I., Burke, J.E., Diedrich, J.K., Cook, S., Shea, T., Geinger, E., et al. (2020). Evolutionary Persistence of DNA Methylation for Millions of Years after Ancient Loss of a De Novo Methyltransferase. Cell *180*, 263–277.e20.

Chatterjee, M., and Acar, M. (2018). Heritable stress response dynamics revealed by single-cell genealogy. Sci. Adv. *4*, e1701775.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al.; 1000 Genomes Project Analysis Group (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156–2158.

Darwin, C. (1859). On the origin of species by means of natural selection, or, the preservation of favoured races in the struggle for life (J. Murray).

Day, T., and Bonduriansky, R. (2011). A unified approach to the evolutionary consequences of genetic and nongenetic inheritance. Am. Nat. *178*, E18–E36.

Deniz, Ö., Flores, O., Aldea, M., Soler-López, M., and Orozco, M. (2016). Nucleosome architecture throughout the cell cycle. Sci. Rep. *6*, 19729.

Elison, G.L., Xue, Y., Song, R., and Acar, M. (2018). Insights into Bidirectional Gene Expression Control Using the Canonical GAL1/GAL10 Promoter. Cell Rep. *25*, 737–748.e4.

Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Véronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., André, B., et al. (2002). Functional profiling of the Saccharomyces cerevisiae genome. Nature *418*, 387–391.

Griffiths, A.J.F., Miller, J.H., Suzuki, D.T., Lewontin, R.C., and Gelbart, W.M. (2000). An Introduction to Genetic Analysis, Seventh Edition (W. H. Freeman).

Hahne, F., LeMeur, N., Brinkman, R.R., Ellis, B., Haaland, P., Sarkar, D., Spidlen, J., Strain, E., and Gentleman, R. (2009). flowCore: a Bioconductor package for high throughput flow cytometry. BMC Bioinformatics *10*, 106.

Halfmann, R., Jarosz, D.F., Jones, S.K., Chang, A., Lancaster, A.K., and Lindquist, S. (2012). Prions are a common mechanism for phenotypic inheritance in wild yeasts. Nature *482*, 363–368.

Huang, S. (2009). Non-genetic heterogeneity of cells in development: more than just noise. Development *136*, 3853–3862.

Jablonka, E., and Raz, G. (2009). Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. Q. Rev. Biol. *84*, 131–176.

Janke, C., Magiera, M.M., Rathfelder, N., Taxis, C., Reber, S., Maekawa, H., Moreno-Borchart, A., Doenges, G., Schwob, E., Schiebel, E., and Knop, M. (2004). A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. Yeast *21*, 947–962.

Kaiser, C., Michaelis, S., and Mitchell, A. (1994). Methods in Yeast Genetics: A Cold Spring Harbor Laboratory Course Manual (Cold Spring Harbor Laboratory Press).

Kaufmann, B.B., Yang, Q., Mettetal, J.T., and van Oudenaarden, A. (2007). Heritable stochastic switching revealed by single-cell genealogy. PLoS Biol. *5*, e239.

Klironomos, F.D., Berg, J., and Collins, S. (2013). How epigenetic mutations can affect genetic evolution: model and mechanism. BioEssays *35*, 571–578.

Kolodkin, A.L., Klar, A.J.S., and Stahl, F.W. (1986). Double-strand breaks can initiate meiotic recombination in S. cerevisiae. Cell *46*, 733–740.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693–705.

Kronholm, I., and Collins, S. (2016). Epigenetic mutations can both help and hinder adaptive evolution. Mol. Ecol. *25*, 1856–1868.

Kundu, S., Horn, P.J., and Peterson, C.L. (2007). SWI/SNF is required for transcriptional memory at the yeast GAL gene cluster. Genes Dev. *21*, 997–1004.

Kuzawa, C.W., and Thayer, Z.M. (2011). Timescales of human adaptation: the role of epigenetic processes. Epigenomics *3*, 221–234.

Lacal, I., and Ventura, R. (2018). Epigenetic Inheritance: Concepts, Mechanisms and Perspectives. Front. Mol. Neurosci. *11*, 292.

Laland, K., Uller, T., Feldman, M., Sterelny, K., Müller, G.B., Moczek, A., Jablonka, E., Odling-Smee, J., Wray, G.A., Hoekstra, H.E., et al. (2014). Does evolutionary theory need a rethink? Nature *514*, 161–164.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv, arXiv:1303.3997. https://arxiv.org/abs/1303.3997.

Li, E., and Zhang, Y. (2014). DNA methylation in mammals. Cold Spring Harb. Perspect. Biol. *6*, a019133.

Lichten, M., and Goldman, A.S.H. (1995). Meiotic recombination hotspots. Annu. Rev. Genet. *29*, 423–444.

Luo, X., Song, R., and Acar, M. (2018). Multi-component gene network design as a survival strategy in diverse environments. BMC Syst. Biol. *12*, 85.

McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. Genome Biol. *17*, 122.

Nei, M., and Nozawa, M. (2011). Roles of mutation and selection in speciation: from Hugo de Vries to the modern genomic era. Genome Biol. Evol. *3*, 812–829.

Olson-Manning, C.F., Wagner, M.R., and Mitchell-Olds, T. (2012). Adaptive evolution: evaluating empirical support for theoretical predictions. Nat. Rev. Genet. *13*, 867–877.

Peng, W., Liu, P., Xue, Y., and Acar, M. (2015). Evolution of gene network activity by tuning the strength of negative-feedback regulation. Nat. Commun. *6*, 6226.

Peng, W., Song, R., and Acar, M. (2016). Noise reduction facilitated by dosage compensation in gene networks. Nat. Commun. *7*, 12959.

Raynaud, C., Mallory, A.C., Latrasse, D., Jégu, T., Bruggeman, Q., Delarue, M., Bergounioux, C., and Benhamed, M. (2014). Chromatin meets the cell cycle. J. Exp. Bot. *65*, 2677–2689.

Ryu, H.-Y., and Ahn, S. (2014). Yeast histone H3 lysine 4 demethylase Jhd2 regulates mitotic rDNA condensation. BMC Biol. *12*, 75.

Sikorski, R.S., and Hieter, P. (1989). A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. Genetics *122*, 19–27.

Skinner, M.K. (2015). Environmental epigenetics and a unified theory of the molecular aspects of evolution: A neo-Lamarckian concept that facilitates neo-Darwinian evolution. Genome Biol. Evol. *7*, 1296–1302.

Skinner, M.K., Guerrero-Bosagna, C., and Haque, M.M. (2015). Environmentally induced epigenetic transgenerational inheritance of sperm epimutations promote genetic mutations. Epigenetics *10*, 762–771.

Smith, K.N., and Nicolas, A. (1998). Recombination at work for meiosis. Curr. Opin. Genet. Dev. *8*, 200–211.

Stajic, D., Perfeito, L., and Jansen, L.E.T. (2019). Epigenetic gene silencing alters the mechanisms and rate of evolutionary adaptation. Nat. Ecol. Evol. *3*, 491–498.

Tyedmers, J., Madariaga, M.L., and Lindquist, S. (2008). Prion switching in response to environmental stress. PLoS Biol. *6*, e294.

Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. Curr. Protoc. Bioinformatics *43*, 11.10.1–11.10.33.

Wach, A., Brachat, A., Pöhlmann, R., and Philippsen, P. (1994). New heterologous modules for classical or PCR-based gene disruptions in Saccharomyces cerevisiae. Yeast *10*, 1793–1808.

Xue, Y., and Acar, M. (2018a). Mechanisms for the epigenetic inheritance of stress response in single cells. Curr. Genet. *64*, 1221–1228.

Xue, Y., and Acar, M. (2018b). Live-Cell Imaging of Chromatin Condensation Dynamics by CRISPR. iScience *4*, 216–235.

Zacharioudakis, I., Gligoris, T., and Tzamarias, D. (2007). A yeast catabolic enzyme controls transcriptional memory. Curr. Biol. *17*, 2041–2046.

Zhou, V.W., Goren, A., and Bernstein, B.E. (2011). Charting histone modifications and the functional organization of mammalian genomes. Nat. Rev. Genet. *12*, 7–18.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
α-H3K4me3	Abcam	Cat# ab8580; RRID:AB_306649
α-H3K36me3	Abcam	Cat# ab9050; RRID:AB_306966
α-H3Ac	Millipore	Cat# 07-360; RRID:AB_310550
α-H3	Abcam	Cat# ab1791; RRID:AB_302613
Experimental Models: Organisms/Strains		
S. cerevisiae: Strain background: W303 MATα	Our Lab Stocks	MA0001
S. cerevisiae: Strain background: W303 MATa	Our Lab Stocks	MA0002
S. cerevisiae: Strain background: W303 MATα ho::HIS5-P_{GAL1}-YFP	Our Lab Stocks	WP35
S. cerevisiae: Strain background: W303 MATα, ho::HIS5-P_{GAL1}-YFP, ura3::URA3-P_{GAL1}-mCherry	This paper	WP35URAPg1mC
S. cerevisiae: Strain background: W303 MATα, ho::HIS5-P_{GAL1}-YFP, ura3::URA3-P_{TEF1}-mCherry	This paper	WP35URAPtefmC
S. cerevisiae: Strain background: W303 MATα, ho::HIS5-P_{GAL1}-YFP, ura3::URA3-P_{GAL1}-mCherry, gal80Δ::NatNT2	This paper	XLUYmCgΔ80
S. cerevisiae: Strain background: W303 MATa, ho::HIS5-P_{GAL1}-YFP, ura3::URA3-P_{GAL1}-mCherry, gal80Δ::KanMX4	This paper	DMY375
Recombinant DNA		
Plasmid: pRS306	(Sikorski and Hieter, 1989)	N/A
Plasmid: pRS305	(Sikorski and Hieter, 1989)	N/A
Plasmid: pYM17	(Janke et al., 2004)	N/A
Plasmid: pFA6-kanMX4	(Wach et al., 1994)	N/A
Plasmid: HIS5-P_{GAL1}-YFP	Our Lab Stocks	N/A
Plasmid: URA3-P_{GAL1}-mCherry	This paper	N/A
Plasmid: URA3-P_{TEF1}-mCherry	This paper	N/A
Plasmid: pRS314-CAS9	Our Lab Stocks	N/A
Chemicals, Peptides, and Recombinant Proteins		
Gibson Assembly® Master Mix	New England BioLabs	E2611S
Restriction Enzyme: BstBI	New England BioLabs	R0519S
iTaqTM Universal SYBR® Green Supermix	Bio-Rad	1725120
Protein G-Sepharose	GE Healthcare	17-0618-01
Pronase	Roche	11 459 643 001
High Capacity RNA-to-cDNA kit	Applied Biosystems	4388950
YeaStar Genomic DNA Kit	ZYMO Research	D2002
Cycloheximide	Sigma-Aldrich	C7698
cOmplete protease inhibitor cocktail	Roche	11697498001
PMSF	AmericanBio	AB01620
β-glucuronidase	Sigma-Aldrich	G7017

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Oligonucleotides		
Primers for qRT-PCR and ChIP-qPCR, see Table S2	This paper	N/A
Deposited Data		
GenBank: SAMN11440943	GenBank	N/A
GenBank: SAMN11440944	GenBank	N/A
GenBank: SAMN11440945	GenBank	N/A
GenBank: SAMN11440946	GenBank	N/A
GenBank: SAMN11440947	GenBank	N/A
GenBank: SAMN11440948	GenBank	N/A
GenBank: SAMN11440949	GenBank	N/A
GenBank: SAMN11440950	GenBank	N/A
Software and Algorithms		
NEBuilder® Assembly Tool	New England BioLabs	N/A
BD FACSuite	BD Biosciences	N/A
R	www.R-project.org	N/A
Bioconductor flowCore package	(Hahne et al., 2009)	N/A
Trimmomatic	(Bolger et al., 2014)	N/A
BWA-MEM	(Li, 2013)	N/A
Picard's MarkDuplicates	https://github.com/broadinstitute/picard	N/A
GATK tools	(Van der Auwera et al., 2013)	N/A
VCFtools	(Danecek et al., 2011)	N/A
VEP tool	(McLaren et al., 2016)	N/A
Other		
BD FACS-Aria	Becton Dickinson	N/A
Cary 60 UV-Vis Spectrometer	Agilent Technologies	N/A
Zeiss Tetrad "Advanced Yeast Dissection Microscope"	Carl Zeiss	N/A
Illumina HiSeq4000	Illumina	N/A

RESOURCE AVAILABILITY

Lead Contact

Requests for further information and for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Murat Acar (murat.acar@yale.edu).

Materials Availability

Yeast strains and plasmids used in this study are described in the Key Resources Table and will be made available upon request from the Lead Contact, Murat Acar (murat.acar@yale.edu).

Data and Code Availability

The accession numbers for the sequencing data from WGS runs reported in this paper are GenBank: SAMN11440943, GenBank: SAMN11440944, GenBank: SAMN11440945, GenBank: SAMN11440946, GenBank: SAMN11440947, GenBank: SAMN11440948, GenBank: SAMN11440949, GenBank: SAMN11440950. These numbers are also listed in the Key Resources Table.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Saccharomyces cerevisiae with the W303 genetic background

All yeast *Saccharomyces* cerevisiae strains constructed are based on the haploid W303 strain background. Complete genotypic descriptions of all strains can be found in the Key Resources Table.

All cultures were grown in synthetic minimal media with histidine dropout and appropriate supplements of other amino-acids. Culture growths were performed in a 30°C shaker (225rpm) in a volume of 1mL. After 48hrs of growth on histidine-dropout minimal media plates containing 2% glucose, strains were grown in liquid minimal media for 22hr ("overnight") in the presence of 0.1% mannose as a non-inducing sole carbon source. This was followed by a 72hr induction period in liquid minimal media containing 0.1% mannose and 0.2% galactose as carbon sources.

METHOD DETAILS

Construction of yeast strains and plasmids

Strains used to study the GAL network were built on WP35 which is a haploid wild-type strain carrying single copy of the P_{GAL1}-*YFP* reporter in the *ho* locus. The double reporter strains carrying a second reporter (P_{GAL1}-*mCherry-tCYC1* or P_{TEF1}-*mCherry-tCYC1*) were constructed with the following steps. First, plasmids carrying P_{GAL1}-*mCherry-tCYC1* or P_{TEF1}-*mCherry-tCYC1* on the pRS306 backbone (Sikorski and Hieter, 1989) were constructed using the Gibson Assembly® Master Mix and NEBuilder® Assembly Tool (New England BioLabs). The resulting plasmids were then linearized within the *URA3* gene at BstBI cut site and transformed into WP35 using the standard lithium acetate (LiOAc) transformation technique. qPCR was performed to select colonies carrying single copy of the second reporter. To obtain the strain XLUYmCgΔ80, the P_{AgTEF}-*natNT2-tADH1* cassette from pYM17 (Euroscarf, Janke et al., 2004) was integrated into the double reporter strain carrying P_{GAL1}-*YFP-tCYC1* and P_{GAL1}-*mCherry-tCYC1* to replace the *GAL80* gene by using 60bps homology regions immediately before and after *GAL80*.

For the mating and sporulation experiments, a MATa counterpart for XLUYmCgΔ80 strain was constructed, starting with the MA0002 strain. First, a single copy of *HIS5*-P_{GAL1}-*YFP-tCYC1* was inserted into the *ho* locus by using 60bp homology regions around the *ho* locus. Then, the P_{GAL1}-*mCherry-tCYC1* plasmid was linearized by BstBI and inserted in single-copy into the *URA3* locus. Finally, the *GAL80* ORF was deleted by amplifying the P_{AgTEF}-*KanR-tAgTEF1* cassette from the *pFA6-kanMX4* plasmid (Wach et al., 1994) with 60bp homology regions immediately before and after the *GAL80* ORF.

Flow cytometry data acquisition and sorting

After the induction period described in the "Experimental Model and Subject Details" section above, the expression distribution of ~50,000 cells were measured by flow cytometry (FACS-Aria; Becton Dickinson) at flow rates between 4 to 8 (flow rate scale of 1-11 corresponds to approximately 10-80 μL/min), and the cell-sorting period was initiated. During the 7-day-long sorting period, populations underwent expression-based sorting once a day, followed by the selection-free growth period lasting from 3 to 20 days during which the entire expression distribution was passed from one day to the next instead of gated-sorting. Every 24hrs during the 7-day-long gated-sorting period, individual cells were sorted into fresh media of the same type by applying fluorescence intensity based gates rendering the highest, middle and lowest 4.8%–5.2% of the total cell population (referred to as HIGH, MID, LOW). In "forward" sorting groups, the gates were selected based on YFP fluorescence, while in the "reverse" sorting groups they were selected based on mCherry fluorescence. 450 individual cells were sorted for the HIGH groups, and 600 cells were sorted for all other groups. To minimize potential variations in the size and/or morphology of the sorted cells, the gating process also involved applying a narrow FSC-SSC (ForwardSCatter-SideSCatter) range corresponding to the densest ~20% of the total cell population. Grown cultures taken from Day0, Day7, the last day of the selection-free period, as well as certain other days throughout each sorting period were frozen on the same day for further analysis. Starting from the overnight growth period and until the end of the gated and selection-free sorting periods, cell densities were kept low (between OD_{600} 0.2 and 0.3) to prevent nutrient depletion.

mRNA transcript levels determination by qRT-PCR

Selected cell populations of the strain XLUYmCgΔ80 frozen after sorting on Day0 and Day7 were recovered as single isogenic colonies from glycerol stocks streaked on 2% glucose minimal media plates with histidine dropout. Colonies were then grown overnight in liquid minimal media containing 0.1% mannose as the sole carbon source, and induced in minimal media containing 0.1% mannose and 0.2% galactose for 48hrs in 50mL volume, reaching a final OD_{600} of less than 0.2. Fluorescence levels of the induced cells were recorded by flow cytometry right before harvesting for total RNA. cDNA was prepared by using the High Capacity RNA-to-cDNA kit from Applied Biosystems. The resulting cDNA was then used in qPCR reactions to quantify mRNA levels of genes of interest. For qPCR, we used the iTaq™ Universal SYBR® Green Supermix from Bio-Rad and targeted 4 genes with 2 sets of primers for each: ACT1 (primer pair EPIACT1-2F and EPIACT1-2R were used as the endogenous control for ΔC_T calculation), YFP, mCherry, GAL1, and GAL4. The relative transcription levels for samples within the same sorting experimental group were calculated with the Day0 population's transcript levels used as the control. The qPCR primers used are listed on Table S2. The amplicons were between 158 and 161bps long.

Local sequencing of key GAL network components

To see whether or not mutations were accumulated on the genetic components relevant to the GAL network activity, frozen cell populations from Day0 and Day7 sorting groups of the strain XLUYmCgΔ80 (the groups that showed at least 20% decrease in YFP or mCherry expression compared to corresponding Day0 expression) were recovered from glycerol stocks streaked on 2% glucose minimal media plates with histidine dropout. Populations were then grown overnight in liquid minimal media containing 0.1%

mannose as the sole carbon source, and induced in minimal media containing 0.1% mannose and 0.2% galactose for 48hrs in 1mL volume, reaching a final OD_{600} of less than 0.3. After the induction period, expression measurements were performed by flow cytometry and also genomic DNA contents were extracted from the induced populations to sequence key genetic components of the GAL network. All re-induced Day7 populations exhibited similar expression levels relative to expression levels of their corresponding Day0 populations; in other words, freezing and re-induction after the actual sorting process did not alter the relative expression levels in these populations. For sequencing, we selected the LOW ("L") sorting groups of the strain XLUYmCgΔ80 which showed over 20% decrease in reporter expression; the local sequencing was performed for the P_{GAL1}-YFP, P_{GAL1}-mCherry, and P_{GAL4}-GAL4 constructs from the beginning of the promoters to the end of the terminators. With "F or R" indicating "Forward or Reverse" sorting based on YFP or mCherry, these Day7 sequenced groups of sorted populations were named as FL1, FL2, FL6, FL8, FL9, RL2, RL5, and RL6; the numbers indicate the identity of the biological replicate from the 7-day-long sorting experiment. Sequencing was first performed on the population level for these sorting groups; genomic DNA was prepared from the entire sorted populations from Day7, and no apparent mutation was identified. Then, 5 randomly-selected single colonies were isolated from each population for isogenic expression characterization, and sequencing was performed on select single colonies which had similar expression profile as the corresponding original population. No mutation on the P_{GAL1}-YFP, P_{GAL1}-mCherry, P_{GAL4}-GAL4 constructs was found in colonies isolated from the FL1, FL6, FL9, RL2, RL5, whose full sequences are given in Data S1; however, mutations/changes were found in colonies isolated from FL2, FL8 and RL6 (Data S1).

Whole Genome Sequencing (WGS) sample preparation

Out of the 5 single colonies isolated from the FL6 and FL9 populations on Day7 (for which local-sequenced for the P_{GAL1}-YFP, P_{GAL1}-mCherry, P_{GAL4}-GAL4 constructs did not identify any mutations), we randomly selected 2 single colonies from each of the FL6 and FL9 groups for performing whole genome sequencing on them. As controls, we also included in these WGS characterizations 2 randomly-selected single colonies isolated from the Day0 population, as well as 2 randomly selected single colony isolated from the positive control group on Day7.

Cells from each single colony were recovered from glycerol stock on 2% glucose minimal media plates with histidine dropout, and grown in 10mL YPD liquid media until the cell-density (OD_{600}) reached ~1. The YeaStar Genomic DNA Kit (ZYMO Research) was used for genomic DNA extraction. The process was repeated 2-3 times until 1-5 μg of purified DNA ($OD_{260/280}$ between 1.8 and 2) concentrated in 50 μL of TE buffer was acquired for each sample. The purified DNA were pooled and sequenced at the Yale Center for Genome Analysis with Illumina HiSeq4000 (paired-end, 150bp) targeting 200X coverage.

Measuring doubling-times of cell populations

Five isogenic colonies isolated from each of the Day7 FL6 and FL9 populations, as well as one isogenic colony isolated from each of the Day0 and Day7 positive controls, were recovered from glycerol stocks and streaked on 2% glucose minimal media plates with histidine dropout. Colonies were then grown overnight in liquid minimal media containing 0.1% mannose as the sole carbon source, and induced in minimal media containing 0.1% mannose and 0.2% galactose for 48hrs in 1mL volume, reaching a final OD_{600} of less than 0.3. Following the induction period, cultures were continuously grown in the same media conditions, and the growth rate analyses were performed based on the dilution rates and the OD_{600} values measured at 6 different time points across the next 52-55hrs. At each time point, all cultures were diluted to maintain OD_{600} below 0.55 to keep growth at log-phase and to prevent nutrition depletion. The average log-phase doubling-time $t_{doubling}$ was calculated (Figures S2A and S2B) using the following formula:

$$t_{doubling} = t_{duration} \Big/ \log_2\left(\frac{D_{end}}{D_{start}} * \prod_{k=1}^{N} d_k\right)$$

$t_{duration}$: duration between the start and end of the continuous culture growth
D_{end}: OD_{600} at the end of continuous culture growth
D_{start}: OD_{600} at the start of continuous culture growth
N: total number of dilutions (here N = 6)
d_k: dilution rate at time point k

SNP introduction with CRISPR-Cas9

Select mutations identified from WGS were cloned into a single colony from Day0 to see if each or all of them could result in the phenotypic changes observed in the sorting experiment. To choose candidate mutations, we first selected mutations within ORFs that cause changes in their corresponding amino acids. We then selected the common mutations found in both single colonies isolated from a group: 5 common mutations for FL6 (*FEN2*, *GPM2*, *IRA2*, *NUP133, RPN4*), and 3 common mutations for FL9 (*APL1*, *BDS1, SRB8*). To introduce these mutations back into a single colony from Day0, a centromeric plasmid with backbone pRS314 carrying a Cas9 cassette was first transformed into a Day0 single colony. For each mutation, the resulting strain was transformed with

donor DNA carrying the mutation together with a plasmid with backbone pRS305 carrying guide RNA cassette targeting the mutation site. The final strains were sequenced locally to verify the intended genetic alterations.

Degradation dynamics of fluorescent proteins

Five isogenic colonies isolated from each of the Day7 FL6 and FL9 populations, as well as one isogenic colony isolated from each of the Day0 and Day7 positive control, were recovered from glycerol stocks and streaked on 2% glucose minimal media plates with histidine dropout. Colonies were then grown overnight in liquid minimal media containing 0.1% mannose as the sole carbon source, and induced in minimal media containing 0.1% mannose and 0.2% galactose for 48hrs, reaching a final OD_{600} of ~0.2 in a total volume of 5mL for each sample. Cycloheximide (Sigma-Aldrich, C7698) was then added to the cultures at the final concentration of 10 μg/mL. Samples were taken from the cycloheximide-treated cultures for fluorescence measurements with flow cytometry (FACS-Aria, Becton Dickinson) at the following time points while the cultures continued being incubated in a 30°C shaker: 0hr (right before adding cycloheximide), 15min (right after adding cycloheximide), 1.5hr, 4hr, 6hr, 21hr and 28.5hr.

During the flow cytometry measurements, due to the cycloheximide treatment, changes were observed in the position of the total cell population based on the FSC-SSC readings. Therefore, a large FSC-SSC gate covering the densest 40%–80% of the total population was applied.

ChIP-qPCR experiments

From the single colonies selected for WGS, we selected the two colonies from FL9 together with one single colony from each of the Day0 and Day7 positive control group for histone modification characterization with ChIP-qPCR. We tested three types of histone modifications at GAL1, GAL4, YFP and mCherry open reading frames: H3K4me3, H3K36me3 and H3Ac. A yeast strain with *set1Δ* background was used as negative control for the H3K4me3 group. Cells from each single colony were recovered from glycerol stock on 2% glucose minimal media plates with histidine dropout, grown overnight in minimal media containing 0.1% mannose as the sole carbon source, and induced in minimal media containing 0.1% mannose and 0.2% galactose for 48hrs. We used 300mL of culture with OD_{600} ~0.4 to initiate ChIP.

ChIP experiments were performed as described previously (Ahn et al., 2004; Ryu and Ahn, 2014). Briefly, formaldehyde was added to a final concentration of 1% for 20 min. Cross-linking was quenched by addition of glycine to 240 mM. Cells were collected by centrifugation, washed in TBS twice, and then lysed with glass beads in FA lysis buffer {50 mM HEPES-KOH at pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na deoxycholate, 0.1% SDS, cOmplete protease inhibitor cocktail (Roche, 11697498001), 1 mM PMSF (AmericanBio, AB01620)}. Sheared chromatin by sonication was incubated with Protein G-Sepharose (GE Healthcare, 17-0618-01) bound with anti-H3K4me3 (Abcam, ab8580), anti-H3K36me3 (Abcam, ab9050), anti-H3Ac (Millipore, 07-360), or anti-H3 (Abcam, ab1791). Following washings, eluted chromatin fragments were treated with pronase (Roche, 11 459 643 001), and DNA was purified by phenol/chloroform extraction. qPCR assays were performed using 1:8 diluted DNA template, and then the results for methylated or acetylated H3 were normalized to total histone H3 signals and the internal control (a fragment amplified from an untranscribed region on ChrIV).

Forward and reverse primer sequences used for ChIP-qPCR are listed on Table S2. EPIYFP pair targets 169-327bp from 5′ of YFP ORF; EPImC-1 pair targets 177-338bp from 5′ of mCherry ORF; EPIGal1-1 pair targets 319-476bp from 5′ of GAL1 ORF; EPIGal4-1 pair targets 2225-2385bp from 5′ of GAL4 ORF; IntIV pair is the endogenous control and targets an intergenic region on chromosome IV.

Mating, sporulation and tetrad dissection

The desired strains were grown in YPD plates overnight. Then, a small amount of the fresh patch of cells was mixed with its mating counterpart in a fresh YPAD plate (YPD supplemented with 20mg/L Ade) and incubated for 4h at 30°C. Zygote formation was checked by microscopy and a portion of the mating patch was transferred to YPD+Nat+G418 plate, which selected for diploid cells. Single diploid colonies were transferred to GNA pre-sporulation plates (Giaever et al., 2002) (5% glucose, 3% nutrient broth, 1% yeast extract, 2% agar) and grown for a day at 30°C, and then transferred to sporulation plates (Kaiser et al., 1994) (1% potassium acetate, 0.1% yeast extract, 0.05% dextrose, 2% agar) where they were incubated at room temperature for 4-5 days. Tetrad dissection was performed on a YPD plate after degrading the ascus wall with β-glucuronidase (Sigma-Aldrich, G7017), and then incubated at 30°C for 2 days. The spores coming from each tetrad were spotted on YPD+G418 and YPD+Nat plates to check that they were displaying a proper segregation pattern of the markers, which qualified them for further analysis.

QUANTIFICATION AND STATISTICAL ANALYSIS

Flow cytometry data analysis

Each sample of flow cytometry (FACSAria, Becton Dickinson) data were analyzed in R using the Bioconductor flowCore software package (Hahne et al., 2009). The FSC-SSC gate was chosen to cover the densest portion of the total population and eliminate individuals with unusual morphologies, such as dying cells and cell debris; the same gate was used for all samples gathered during a single experiment. Each FACS sample had on average ~7500 cells after gating. Log-amplified fluorescence measurements for the

gated cells were converted to linear scale for analysis. When needed for the wild-type strain, a threshold for ON state was selected based on fluorescence measurements from uninduced cells and applied uniformly to all relevant samples.

The raw expression level of each strain on each day during the multi-day experiment is measured by averaging the single-cell reporter fluorescence as measured by flow cytometry. To control for the effect of day-to-day variations, the raw expression levels were normalized using the average expression level of the same reporter in the positive control samples measured on the same day.

Whole Genome Sequencing data analysis

The sequencing reads were first trimmed using Trimmomatic (Bolger et al., 2014) with the following settings: "LEADING:3 TRAILING:3 SLIDINGWINDOW:2:30" (Barbieri et al., 2017). The filtered reads from each sample were independently aligned to the current version of S288C reference genome from Ensembl using BWA-MEM (Li, 2013) and converted to BAM format using SAMtools. Picard's MarkDuplicates (https://github.com/broadinstitute/picard) was used to mark duplicates in the resulting BAM files. We then realigned the reads with the GATK tools (Van der Auwera et al., 2013) RealignerTargetCreator and IndelRealigner, and variants were called using GATK's HaplotypeCaller with "-ploidy 1." SNPs and Indels were extracted from the resulting file and filtered manually using GATK's SelectVariants and VariantFiltration. For SNP, we used the following parameters: -filterExpression "QD<2.0 || FS>50.0 || MQ<50.0 || SOR>3.0 || MQRankSum<-12.5 || ReadPosRankSum<-8.0." For Indel, we used these parameters: -filterExpression "QD<2.0 || FS>200.0 || InbreedingCoeff<-0.8 || SOR>10.0 || ReadPosRankSum<-20.0." Finally, we used VCFtools (Danecek et al., 2011) to identify variants in FL6, FL9 and Day7-positive-control samples relative to Day0, which were then annotated with Ensembl's VEP tool (McLaren et al., 2016).

χ^2 test for checking inheritance models for sporulation outcomes

The proposed inheritance model was tested as described (Griffiths et al., 2000). Briefly, the χ^2 test statistic was calculated as $\sum_{i=1}^{n} \frac{(E_i - O_i)^2}{E_i}$, where n is the number of classes for a phenotype, E_i is the expected number of individuals under that class according to the model, and O_i is the experimentally-observed number of individuals classified under that class. This test statistic was compared with the χ^2 distribution with n-1 degrees of freedom to obtain the likelihood of the experimental observation if the assumed model was true (p value).

Check for updates

Received: 10 February 2022 | Revised: 1 September 2022 | Accepted: 6 September 2022

DOI: 10.1111/mec.16696

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Adapting to climate with limited genetic diversity: Nucleotide, DNA methylation and microbiome variation among populations of the social spider *Stegodyphus dumicola*

Anne Aagaard[1] | Shenglin Liu[1] | Tom Tregenza[2] | Marie Braad Lund[3] | Andreas Schramm[3] | Koen J. F. Verhoeven[4] | Jesper Bechsgaard[1] | Trine Bilde[1]

[1]Section for Genetics, Ecology & Evolution, Department of Biology, Aarhus University, Aarhus C, Denmark

[2]Centre for Ecology & Conservation, School of Biosciences, University of Exeter, Penryn Campus, UK

[3]Section for Microbiology, Department of Biology, Aarhus University, Aarhus C, Denmark

[4]Terrestrial Ecology Department, Netherlands Institute of Ecology (NIOO-KNAW), Wageningen, The Netherlands

Correspondence
Anne Aagaard, Section for Genetics, Ecology & Evolution, Department of Biology, Aarhus University, Aarhus C, Denmark.
Email: anneaagaard@bio.au.dk

Funding information
Natur og Univers, Det Frie Forskningsråd, Grant/Award Number: 6108-00565; Novo Nordisk Foundation Interdisciplinary Synergy Grant, Grant/Award Number: NNF16OC0021110

Handling Editor: Victoria L. Sork

Abstract

Understanding the role of genetic and nongenetic variants in modulating phenotypes is central to our knowledge of adaptive responses to local conditions and environmental change, particularly in species with such low population genetic diversity that it is likely to limit their evolutionary potential. A first step towards uncovering the molecular mechanisms underlying population-specific responses to the environment is to carry out environmental association studies. We associated climatic variation with genetic, epigenetic and microbiome variation in populations of a social spider with extremely low standing genetic diversity. We identified genetic variants that are associated strongly with environmental variation, particularly with average temperature, a pattern consistent with local adaptation. Variation in DNA methylation in many genes was strongly correlated with a wide set of climate parameters, thereby revealing a different pattern of associations than that of genetic variants, which show strong correlations to a more restricted range of climate parameters. DNA methylation levels were largely independent of *cis*-genetic variation and of overall genetic population structure, suggesting that DNA methylation can work as an independent mechanism. Microbiome composition also correlated with environmental variation, but most strong associations were with precipitation-related climatic factors. Our results suggest a role for both genetic and nongenetic mechanisms in shaping phenotypic responses to local environments.

KEYWORDS
adaptation, DNA methylation, low evolutionary potential, microbiome, phenotypic plasticity, social spiders

1 | INTRODUCTION

The environment fluctuates at a range of timescales and in space across species ranges. If environmental changes occur over periods that are many multiples of species generation times, or if there are restrictions on gene flow between locations, organisms can evolve naturally selected adaptations to this variation (Charlesworth et al., 2017). Additionally, and even in the absence of local

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2022 The Authors. *Molecular Ecology* published by John Wiley & Sons Ltd.

1365294x, 2022, 22, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.16696, Wiley Online Library on [03/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

adaptation, organisms may be able to cope with environmental variation through the capacity of a single genotype to express a range of phenotypes. This phenotypic plasticity gives organisms the potential to respond to environmental variation at short spatial or temporal scales (Fox et al., 2019; Ghalambor et al., 2007, 2015; Gonzalo-Turpin & Hazard, 2009). Phenotypes that are modulated plastically can be expressed very briefly, for example in behavioural reactions, or the rapid production of heat shock proteins (Dahlgaard et al., 1998; Gong et al., 2012), or they can last over multiple generations (Hanson et al., 2017; Nätt et al., 2012), depending on the type of plastic response and its underlying mechanism.

An important question, not least in the context of rapid global change, is whether and how fast adaptive responses can enable organisms to cope with environmental fluctuations and variability (Fox et al., 2019). In particular, species with low population genetic diversity have limited capacity for genetic adaptations when challenged by environmental change (Ørsted et al., 2019; Sgrò et al., 2011; Willi et al., 2006). This raises the question of whether this evolutionary constraint can be compensated for by nongenetic mechanisms with the potential to shape phenotypes (Donelson et al., 2019; Lande, 2009; Sgrò et al., 2016). Advancing our understanding of phenotypic responses shaped by nongenetic mechanisms is important, as they may play a key role in modulating adaptive responses to environmental change in species with low genetic diversity.

Local phenotypic responses can occur via mechanisms other than genetic adaptations, including epigenetic marks (e.g., histone modifications and DNA methylation) that may regulate gene function, and may be mitotically and/or meiotically heritable (Cavalli, 2006; Heckwolf et al., 2019; Henikoff et al., 2004; Holliday, 1987; Wu & Morris, 2001). Epigenetic changes in, for example, DNA methylation profiles can alter the phenotype of the individual (Cubas et al., 1999; Heckwolf et al., 2019; Jablonka, 2017). The various functions of DNA methylation as an epigenetic feature are only partially understood, but a role in relation to phenotypic change, such as by regulation of gene function, has been suggested in several species (Gatzmann et al., 2018; Keller et al., 2016; Liu, Aagaard, et al., 2019; Sarda et al., 2012; Varriale, 2014; Xu et al., 2021). Within invertebrates, methylation is enriched in gene bodies, but the function of this pattern remains unclear (Duncan et al., 2022). The highly structured distribution of DNA methylation across the genome suggests a functional role, and various hypotheses have been proposed, such as regulating gene expression, either directly (*cis*) or by modifying histone acetylation (*trans*), alternative splicing and stabilizing gene expression (Choi et al., 2020; Duncan et al., 2022; Gatzmann et al., 2018; Kvist et al., 2018; Lev Maor et al., 2015; Liu, Ma, et al., 2019; Neri et al., 2017; Xu et al., 2021). Several studies have demonstrated that DNA methylation profiles can change as a function of environmental stressors in common garden experiments, in species such as corals, sticklebacks, cockroaches and dandelions (Dimond & Roberts, 2020; Metzger & Schulte, 2017; Peña et al., 2021; Verhoeven et al., 2010). This is consistent with the idea that variation in phenotypes may be mediated by environmentally induced changes in DNA methylation profiles, which may facilitate the ability of populations to cope with changes in local climatic conditions on a shorter timescale than that of adaptive genetic changes. Furthermore, studies have shown that epigenetic changes can be heritable and may persist across generations (Harney et al., 2022; Nätt et al., 2012; Riddle & Richards, 2002; Sutherland et al., 2000). Studies on how DNA methylation variation is structured across geographical locations and combining this with variation in environmental factors such as, for example, temperature and precipitation, can inform and substantiate hypotheses on the role of DNA methylation in generating locally advantageous phenotypes. Environmental association studies have revealed strong relationships between epigenetic variants and climatic or environmental parameters (Fischer et al., 2013; Gugger et al., 2016; Rico et al., 2014; Verhoeven et al., 2010). However, it is often unclear if such relationships reflect epigenetic variants, or geographical variation in genetic control over epigenetic variants (Dubin et al., 2015).

It is also possible that local responses can be mediated by host–symbiont interactions. All organisms engage in interactions with microbes, and the microbiome represents a source of variation. Symbiotic interactions have huge potential to modulate host phenotype. Indeed, there is ample evidence to suggest that symbiotic interactions with the bacterial microbiome can shape numerous physiological, reproductive and behavioural functions of the host (Bang et al., 2018; Dunbar et al., 2007; Moran et al., 2019; Mueller et al., 2020). Responses in host phenotype mediated by changes in microbiome composition may contribute to improved performance of individuals in their local environment (Henry et al., 2019; Mueller et al., 2020), through new and potentially locally beneficial functions such as improved nutrition, energy production, temperature resistance or pathogen protection (Lynch & Hsiao, 2019; McFall-Ngai et al., 2013; Raza et al., 2020). Adjustments of the microbiome that provide beneficial local adjustments in host phenotype will naturally depend on the specific context, and can vary from changes in overall microbiome composition, to presence/absence and abundance of specific microbes or strains of microbes, and/or changes in strain composition of specific microbial species (Rennison et al., 2019; Rudman et al., 2019; Shigenobu & Wilson, 2011; Wernegreen, 2012). In pea aphids, populations harbour different strains of the obligate symbiont *Buchnera*, which differ in the expression of a heat-shock gene caused by a deletion in the promoter sequence. Aphid populations harbouring low-expression *Buchnera* perform better in colder environments, while populations harbouring high-expression *Buchnera* perform better in warmer environments (Dunbar et al., 2007). In reef-corals it was recently shown that their symbiont composition is shaped by environmental temperature and potentially mediates adaptive host phenotypes (Herrera et al., 2021). Association studies between microbiome composition and environmental variation are, however, relatively scarce (Busck et al., 2020; Suzuki et al., 2019; Walters et al., 2020), and only a few studies have revealed differences in host phenotypes as a function of the environmental context and its microbiome (Walters et al., 2020).

Social spiders of the genus *Stegodyphus* harbour very low species-wide genetic diversity, and *S. dumicola* is known to have one of the lowest genetic diversities recorded in any animal species

1365294x, 2022, 22, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.16696, Wiley Online Library on [03/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(Leffler et al., 2012; Settepani et al., 2017). It has been suggested that the lack of genetic diversity in this species may reduce its evolutionary potential (Settepani et al., 2017), but nevertheless they persist across broad climatic gradients in southern Africa, spanning several climate zones (Majer et al., 2015; Ngaira, 2007). We therefore hypothesize that responses to local environmental factors caused by nongenetic variants, such as DNA methylation and microbiome composition, may facilitate adaptive responses to local conditions. Although some level of heritable variation conferring local adaptation may be present (e.g., in response to temperature challenges; Malmos et al., 2021), the high level of genetic similarity of populations provides an excellent opportunity to evaluate the role of epigenetic and microbiome variation in population differentiation.

A first step towards revealing the molecular mechanisms underlying population-specific responses can be taken through environmental association studies (Morgan et al., 2018; Rellstab et al., 2015; Thomas, 2010; Ungerer et al., 2008). If there are population-specific evolutionary adaptations to climate, we expect to see associations between environmental parameters and genetic variants. Similarly, associations between epigenetic and microbial variants and environmental parameters are predicted if these features, either as induced or as inherited variants, have a role in phenotypic responses to local environmental factors. We use an environmental association approach in which we examine the relationship between genetic, epigenetic and bacterial microbiome diversity with a set of climatic parameters in populations of social spiders. Our aim is to characterize the mechanisms that may govern phenotypic responses of the social spider *S. dumicola* to different climatic variables within their natural habitats. Given low levels of genetic variation in the *S. dumicola* system (populations and species-wide), we hypothesize that DNA methylation and microbiome composition contribute to *S. dumicola* population differentiation, and are associated with environmental and climatic variation across populations.

2 | MATERIALS AND METHODS

2.1 | Study species and sampling

Stegodyphus dumicola is one of three independently evolved social spider species from the genus *Stegodyphus* (Settepani et al., 2016). *S. dumicola* live in family groups in nests of hundreds to thousands of individuals that live their entire life in and around their natal nest, resulting in extremely high levels of inbreeding (Lubin & Bilde, 2007). Additionally, sex ratios are highly female-biased and only a small proportion of females participate in reproduction. Genetic drift is consequently a strong evolutionary force in this species (Settepani et al., 2014, 2017). *S. dumicola* is distributed in the southern part of Africa (Majer et al., 2015), across a range of climatic conditions (Figure 1).

We sampled one female from each of 15 *S. dumicola* nests in each of six different populations, 90 females (from 90 different nests) in total, during December 2015. Five populations are located on a north–south gradient in Namibia and one population is located in South Africa (Figure 1; Table S1). Individual spiders were cut in half and placed directly in DNA extraction buffer (ATL buffer, DNeasy Blood & Tissue; Qiagen) in the field and transported to the laboratory at Aarhus University at ambient temperature. Cutting the spiders in half ensures proper penetration of buffer into the samples. One spider from each nest (i.e., a total of 90 spiders) was used for sequencing.

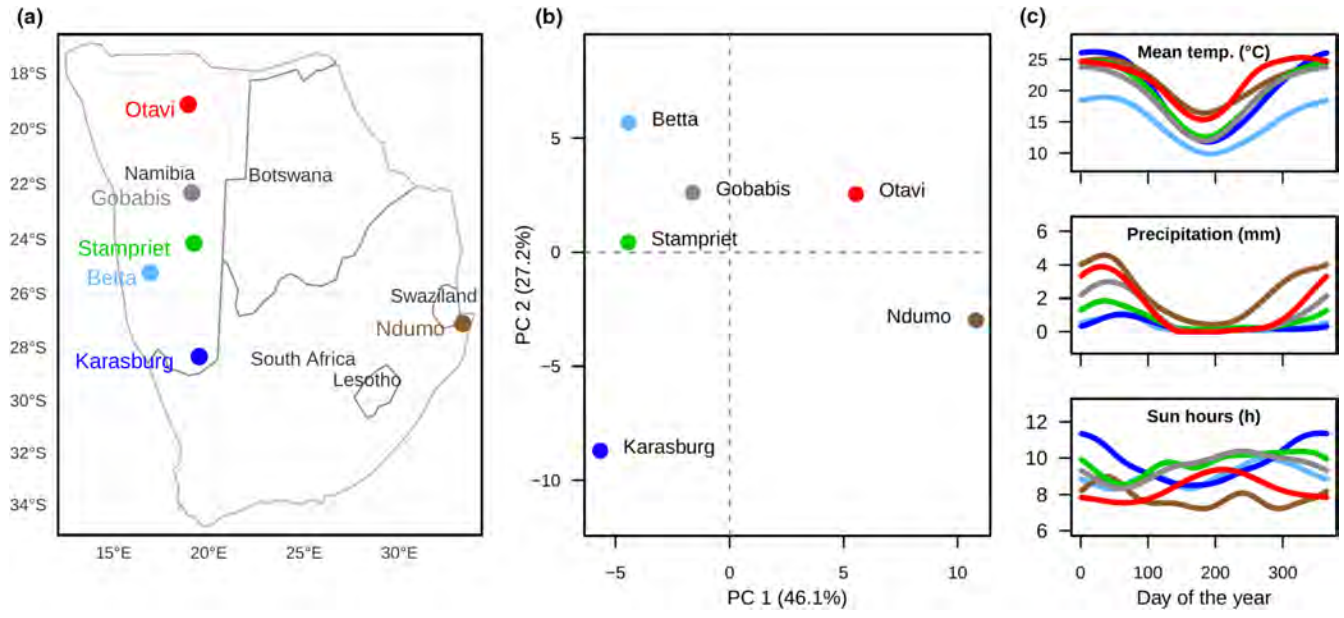


FIGURE 1 (a) Map of southern Africa showing the locations of social spider populations. (b) Climatic separation of the geographical locations on the two main environmental axes from the PCA (see details in Figures S1–S3 and Table S2). (c) Yearly variation in three climatic variables; top: Mean temperature, Centre: Precipitation, bottom: Daily hours of direct sun.

1365294x, 2022, 22, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.16696, Wiley Online Library on [03/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.2 | DNA extraction, sequencing and quantitative PCR

DNA was extracted from all samples using the DNeasy Blood and Tissue kit from Qiagen following the animal tissue protocol. Prior to extraction, samples were homogenized using a pellet pestle. For each round of DNA extraction, one extraction blank (i.e., no sample was added to the tube) was included. The resulting DNA extracts were used for either (i) whole-genome (WG) resequencing, (ii) whole-genome bisulphite (WGB) resequencing or (iii) bacterial 16S rRNA gene amplicon sequencing and quantitative polymerase chain reaction (qPCR).

For WG and WGB sequencing, we pooled the DNA from each population in equimolar ratios before construction of WG and WGB libraries, and sequenced on a HiSeq2500 platform. While individuals may not be equally represented in pooled sequencing, it enables us to find population-specific differences, while still being cost efficient. After first WG sequencing, another set of libraries were made from the same DNA samples to obtain high enough coverage, and a total of 12 WG libraries were sequenced. For bacterial 16S rRNA gene amplicon sequencing, the primers Bac341F and Bac 805R (Herlemann et al., 2011) were used to amplify the V3–V4 region and libraries were prepared according to Illumina's 16S Metagenomic Sequencing Library Preparation guide. Paired-end (2×301 bp) sequencing was done on a MiSeq desktop sequencer (Illumina). Both DNA extraction blanks and PCR negatives were included for amplicon sequencing. Samples were run in two independent sequencing runs.

Quantitative polymerase chain reaction was used to estimate the number of bacteria in individual spiders as described previously (Busck et al., 2022). To compensate for differences in spider body size, we normalized the bacterial 16S rRNA gene copy number to a spider gene copy number (gene 5F, Settepani et al., 2016), and this ratio is referred to as bacterial load (number of bacterial 16S rRNA gene copies/number of spider gene copies). Highly similar coverage of the gene 5F relative to the entire genome indicates a single copy in all populations. All qPCRs were run in triplicate. For details see Busck et al. (2022).

2.3 | Whole genome mapping and variant calling

Whole-genome sequencing of the 12 libraries resulted in 622 Gb of raw data (paired-end reads, each of 150 bp and insert size of 300 or 500 bp). WGB sequencing of the six libraries resulted in 274 Gb of raw data (paired-end reads with each read 100 bp and insert size of 169–225 bp). The raw data were filtered using TRIM GALORE version 0.4.1 by allowing "--trim1." After the filtering, 264 Gb remained.

We mapped the WG resequencing reads to the *S. dumicola* genome (Liu, Aagaard, et al., 2019) using BWA (version 0.7.15) "aln" (Li & Durbin, 2009) allowing a maximum of two mismatches and converted them to bam files using SAMTOOLS (version 1.2) (Li et al., 2009). We extracted single nucleotide polymorphisms (SNPs) using SAMTOOLS mpileup with minimum mapping quality of 20 (Li, 2011) and the POPOOLATION2 (version 1.201) script snp-frequency-diff.pl (--min-count 1 --min-coverage 1 --max-coverage 50,50,50,50,50,40) (Kofler, Pandey, & Schlötterer, 2011). We estimated nucleotide diversities (π) for each population using a Variance-sliding.pl script (--window-size 10,000 --step-size 10,000 --min-count 5 --min-coverage 10 --max-coverage 400 --min-qual 20 --pool-size 15) from POPOOLATION2 (Kofler, Orozco-terWengel, et al., 2011), after converting bam files to pileup files using the mpileup function in SAMTOOLS (Li et al., 2009). To obtain a single estimate of genetic diversity for all samples, we downsampled population bam files to the same size and merged them using the view and merge functions in SAMTOOLS (Li et al., 2009) before converting to pileup format and Variance-sliding.pl. To construct the phylogenetic relationship among the studied populations, WG resequencing data from all individuals from each location were mapped to the *S. dumicola* genome (Liu, Aagaard, et al., 2019) using BWA (version 0.7.15) "aln" (Li & Durbin, 2009) allowing a maximum of two mismatches and converted to location-specific bam files using SAMTOOLS (version 1.2) (Li et al., 2009). We called variants into vcf files using BCFTOOLS version 1.5 ("mpileup" without indel calling [-I] and "call") (Li, 2011). We extracted coding positions using SAMTOOLS "faidx" (Li et al., 2009), and we called consensus sequences using BCFTOOLS version 1.5 "consensus" (Danecek & McCarthy, 2017). We joined consensus sequences into a single concatenated sequence per location and aligned them. We reconstructed a neighbour-joining phylogeny using MEGA X (Kumar et al., 2018). In total, 1000 bootstraps were used to add support to the topology. Genewise F_{ST} estimates were calculated using POPOOLATION2 scripts Create-genewise-sync.pl and fst-sliding.pl (--min-count 3 --min-coverage 20 --max-coverage 100 --pool-size 30 --min-covered-fraction 0.0 --window-size 1,000,000 --step-size 1,000,000).

We mapped WGB sequencing reads with BISMARK (version 0.19.9) (Krueger & Andrews, 2011) using --bowtie1. Methylation status of all C sites was called using Bismark_methylation_extractor and coverage was extracted using the bismark2bedgraph script. DNA methylations were filtered to only include sites with a depth above 10 and below 30, and proportions of C sites methylated in CpG, CHG and CHH (where H = A, T, or C) context were calculated. The methylation level for each cytosine in CpG context was determined as the ratio of reads indicating methylation over the total number of reads for that position, a level referred to as site methylation level (SML) (Schultz et al., 2012). DNA methylations within gene bodies were extracted using the genome annotation and BEDTOOLS INTERSECT version 2.29.2 (Quinlan & Hall, 2010), and the weighted methylation level (WML) of all CpG sites in each gene separately was estimated (mean of proportions of mapped reads being methylated in all CpG sites) (Schultz et al., 2012). Nei's F_{ST} (Nei & Kumar, 2000) was calculated for each gene and between each population pair.

2.4 | 16S rRNA gene amplicon analysis

We obtained 16S rRNA amplicon sequences from individuals from 78 nests (between 11 and 15 per population). From four nests (Otavi)

1365294x, 2022, 22, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.16696, Wiley Online Library on [03/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

two individuals were sequenced per nest, the duplicate individuals were merged in PHYLOSEQ (McMurdie & Holmes, 2013) prior to calculating relative abundances of the amplicon sequence variants (ASVs). qPCR data were obtained from 69 of the individuals (between nine and 14 per population). A sample summary is given in Table S1.

CUTADAPT (Martin, 2011) was used for barcode and primer removal and sequence quality trimming. Using R version 3.6.1 (R Core Team, 2019) the two independent sequencing runs were processed separately using DADA2 version 1.12.1 (Callahan et al., 2016) for quality filtering, denoising and merging of paired-end reads. Filtering was set to maxEE = (2, 2), trncQ = 2 and truncLen = 280 and 200 bp for forward and reverse reads, respectively, in order to identify ASVs. Data from the two sequencing runs were merged prior to chimera finding and classification using DADA2 and Silva small subunit (SSU) reference database nr132 (Quast et al., 2013). ASVs were filtered to a minimum length of 400 bp, and nonbacterial ASVs, chloroplasts and mitochondrial ASVs were excluded. Samples with fewer than 8000 reads were removed from further analysis.

Using the R package PHYLOSEQ version 1.28.0 (McMurdie & Holmes, 2013), all samples were subsampled to the same read depth of 8227 reads (the smallest sample size, seed = 42). Both relative ASV abundances and absolute ASV abundances (based on bacterial load from qPCR analyses) were estimated. ASVs were filtered to only contain ASVs with a prevalence above 25% in at least one population. This retained 57 ASVs for absolute abundance, and 60 ASVs for relative abundance. Bray–Curtis dissimilarity matrices were obtained using the vegdist function in VEGAN version 2.5-6 (Oksanen et al., 2019). Bray–Curtis dissimilarities were calculated for all ASVs across all nests between and within populations, as well as for single ASVs and genera across each population using population-wise abundance means.

2.5 | Environmental variables

Thirty years (1961–1990) of climate data were downloaded using the application NEW_LOCCLIM_1.10 (Grieser et al., 2006), which interpolates climate station measurements (FAOCLIM database) to the input GPS positions from the six populations, and outputs daily climate estimations of selected variables. Three to seven nest GPS points (Table S1) were used to represent each population to create a mean climate estimate for each of the populations. Downloaded climatic variables included 30-year mean daily estimates of mean, maximum and minimum temperature (°C), precipitation (mm), potential evapotranspiration (mm), sun fraction (%), day length (hr), sun hours (hr), water vapour pressure (hPa) and wind speed ($km\,h^{-1}$). Shepard's Interpolation method was used for temperature data, while a thin-plate-spline was used for the remaining variables. For each of these variables, monthly and yearly mean, maximum, minimum and variation was calculated, along with the number of days where temperatures exceeded or went below set thresholds. Longitude, latitude and altitude were also included. The means of all monthly estimates were calculated, to obtain a monthly based yearly mean. All in all, 99 climate variable factors were calculated for the six populations.

To reduce the number of environmental variables, a principal component analysis (PCA) was run on 96 out of 99 variables (those solely containing zeros were excluded), using scaled and centred prcomp in R. A summary can be seen in Figures S1–S3 and Table S2. Based on these analyses, the first five PC axes were applied as composite environmental variables explaining a substantial amount of variation of the initial 96 environmental variables. The distance between populations in PC axes was calculated using dist() in the R stats package. Because previous research has indicated that temperature and precipitation are particularly important drivers for local phenotypic responses in arthropod species (Gefen et al., 2015; Malmos et al., 2021; Toolson, 1982) we selected 51 aspects of temperature and precipitation (see *x*-axis Figure 6) (many of which may be correlated), and directly calculated population distances using dist() from the R stats package. For an explanation on how these parameters were calculated, see Table S3.

2.6 | Environmental association analyses

Genetic (F_{ST}), DNA methylation (F_{ST}) and microbiome (Bray–Curtis dissimilarity) distances among the six populations were correlated to the distance in environmental PC axes through a set of analyses. (i) Partial Mantel tests were run for each gene separately to test for correlations between environmental axes and genetic divergence (F_{ST}) based on gene-body SNPs corrected for neutral population structure based on all SNPs (overall F_{ST}), and between environmental axes and DNA methylation divergence based on gene-body WML (gene-wise F_{ST}), using two different corrections: assumed neutral population structure based on all SNPs (overall F_{ST}); and *cis*-genetic variation based on the SNP variation in the given gene (gene-wise F_{ST}). (ii) Partial Mantel tests were run to test for correlations between microbiome ASVs and environmental axes, while correcting for neutral population structure based on all SNPs (overall F_{ST}). (iii) Multiple regressions were run on distance matrices (MRM function from the ECODIST R package; Goslee & Urban, 2007) featuring F_{ST} distance measures for gene-body SNPs and WML, and Bray–Curtis dissimilarity of microbiome variables as a function of the distance in each environmental axis. Genes with no variation between populations were removed before analyses (all $F_{ST} = 0$). For gene-body WML corrected for *cis* genetic variation, associations using simple Mantel tests were used in cases with no SNPs within the given gene. To assess temperature and precipitation associations explicitly, we ran the above-mentioned Mantel and Partial Mantel tests, exchanging distance in PC axes with distances in 51 individual temperature and precipitation aspects. *p*-Values are not particularly informative in this type of analysis where sample sizes can be extremely large and numerous closely related correlations are run. To identify biologically significant relationships we compared two distributions: (i) a distribution of actual correlation coefficients stemming from the above-mentioned partial Mantel analyses, and (ii) an expected distribution of the same correlations as in (i), but where the environmental axis or climate parameter were permuted (hereafter

1365294x, 2022, 22, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.16696, Wiley Online Library on [03/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

termed the null distribution). To discern which climatic parameter or environmental axes showed stronger correlations than expected based on the null distribution, we took the number of genes (both nucleotide and methylation) in the actual distribution exceeding the 99.99th percentile of the null distribution. This arbitrary threshold represents a conservative approach to identifying correlations that may represent adaptive variants in response to climate, and that are unlikely to occur only by chance. For the microbiome, a similar approach was used, but because we only analyse 61 ASVs, the 99.99th percentile is not meaningful. Instead, we used the highest correlation coefficient from the null distribution as a threshold and included those that exceeded the highest correlation coefficient of the null distribution.

Gene ontology (GO) enrichment analyses were run for genes strongly correlated (above 99.99% threshold) to every environmental axis for gene-wise SNPs corrected for population structure, WML corrected for population structure and *cis*-genetic variation. Gene functional annotation was performed using EGGNOG orthology data and EGGNOG MAPPER (emapper-2.1.9) using DIAMOND SEARCH version 0.9.21 (Buchfink et al., 2021; Cantalapiedra et al., 2021; Huerta-Cepas et al., 2019), while ontology enrichment analysis was done using the R packages GOSTATS version 2.52.0 and GSEABASE version 1.58.0. (Falcon & Gentleman, 2007; Morgan et al., 2022).

For data handling and analyses run in R, we used the following main packages: USEDIST version 0.4.0 (Bittinger, 2020) and DPLYR version 1.0.6 (Wickham et al., 2021), while graphics were performed using base R (R Core Team, 2019), VENNDIAGRAM version 1.6.20 (Chen, 2018) and TMAP version 2.3.2 (Tennekes, 2018).

3 | RESULTS

3.1 | Whole genome sequencing and bisulphite sequencing

The WG sequencing mapping rates of the six spider populations varied between 71.9% and 80.8% when mapped to the reference genome, with coverage depth ranging from 21 to 29 (Table S4). The total number of SNPs called in each population varied from 655,000 to 1,119,000 (Table S4). For WGB sequencing (WGBS) λDNA was used as a control for a bisulphite conversion rate, and 99% of the unmethylated cytosines were converted. WGBS mapping rates of the six populations varied between 40.5% and 46.6%. The total number of sites that were at least partially methylated (>0 reads suggesting methylation) in each population varied from 3,441,377 to 6,218,880, while the number of sites that were methylated across all reads in each population varied from 398,114 to 580,440, about 9% on average (Table S4). Most methylations were found in CpG sequence context, and of all cytosines in CpG context, between 9.4% and 11.3% were at least partially methylated in the six populations. Methylations in CHG and CHH sequence context comprised less than 1% in all populations. Mapping and SNP/methylation statistics are summarized in Table S4.

3.2 | Bacterial microbiome

On average 30,321 quality filtered reads were obtained from 82 samples and the minimum and maximum number of reads were 8227 and 49,237, respectively (Table S1). A total of 3378 bacterial ASVs were identified, but the 10 most abundant ASVs accounted for more than 80% of all reads (Table S1, ASV table). The bacterial load (calculated as the number of bacterial 16S rRNA gene copies divided by the number of spider gene copies) was determined in 73 out of the 82 samples used for amplicon sequencing (Table S1, sample list). The average bacterial load was 5.7, ranging from 0.02 to 61.6, and the load did not differ significantly between populations (ANOVA, $p = .0825$, Figure S4).

The spider microbiome was dominated by *Mycoplasma*, *Diplorickettsia*, *Borrelia* and *Weeksellaceae* (Figure S5), corroborating a previous study in *Stegodyphus dumicola* (Busck et al., 2020, 2022). Bray–Curtis dissimilarity between individuals from different nests differed between populations, with the highest dissimilarity index found in Otavi (Figure S6). Overall, individuals from the same population had more similar microbiome composition compared with individuals from different populations. Significant differences in Bray–Curtis dissimilarities between populations were not driven by any single population (Figure S6). We recovered similar results whether analyses were based on relative or absolute abundances (Figure S6).

3.3 | Population phylogeny and population genetic diversity

A phylogenetic reconstruction of populations is shown in Figure 2. Phylogenetic relationships among populations cannot be predicted directly from geographical locations; for example, we show that Betta is phylogenetically closest to Otavi, while geographically closest to Stampriet. Genome-wide nucleotide diversity varied from 0.00021 in Betta to 0.00071 in Gobabis, while nucleotide diversity for all samples pooled was 0.00091.

3.4 | Population divergences—Genetic, DNA methylation and microbiome

Pairwise molecular distances among populations, estimated as F_{ST} values, when averaging over all genes, were between 0.04 and 0.15 based on SNPs, and between 0.004 and 0.01 for WML. Pairwise microbiome ASV distances among populations, estimated as Bray–Curtis dissimilarities, ranged from 0.21 to 0.87 for relative abundances and 0.19 to 0.80 for absolute abundances (Figure S7c,d). We found significant isolation-by-distance when considering nucleotide variation distance ($r = .75$, $p = .048$), and this correlation was predominantly driven by the Ndumo population, which is a long way to the east of the Namibian populations (Figure 3). No isolation-by-distance was found when analysing WML and microbiome distances

1365294x, 2022, 22, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.16696, Wiley Online Library on [03/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

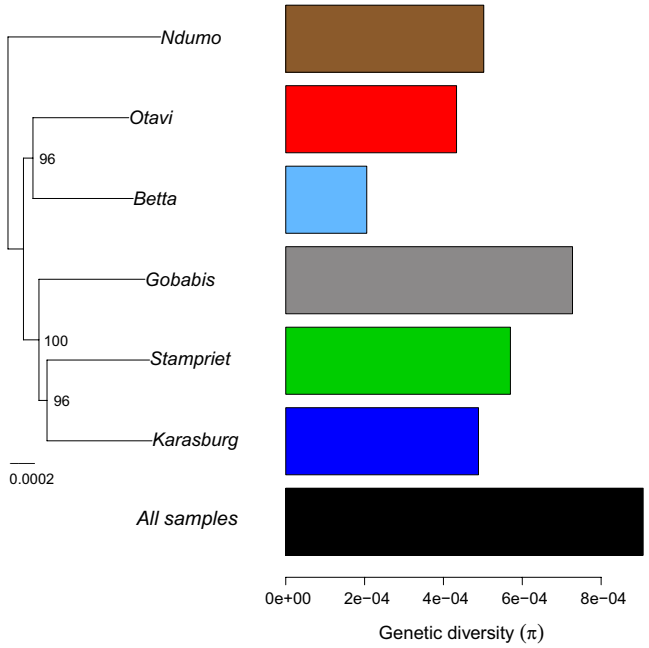


FIGURE 2 Phylogenetic relationships among social spider populations (location names), bars show nucleotide diversities (π) for populations, and the black bar shows genetic diversity across all populations joined. Bootstrap values above 60% are shown.

(Figure 3). The distribution of F_{ST} values estimated per gene across the entire genome (Figure S7a,b) revealed that population differentiation based on DNA methylation (gene-body WML) is generally lower than population differentiation based on genetic variants (SNPs). However, for DNA methylation data, there is a long tail (Figure S7b), indicating that some genes are strongly differentiated among populations.

3.5 | Environmental parameters

The six populations differ substantially in local climate, for example in mean and seasonal fluctuations in temperature and precipitation (see Figures 1 and S8). An overview of all investigated climate variables and their population patterns is provided in Figure S8. PCA on the environmental factors resulted in five axes each explaining a substantial amount of variance in the data (Figures S2 and S3). The first three axes explained 92% of the total variance. The populations are relatively well separated on PC1 and PC2 (Figure 1), with Karasburg and Ndumo being more different compared to the other populations (see Figure S9 for plots of the remaining PC axes). The 20 main loadings driving each PC axis were extracted (Table S2), and this revealed that PC1 is driven mainly by precipitation, minimum temperature, sunshine fraction and sunshine hours, PC2 is mainly driven by maximum temperature and a mixture of other variables, while PC3 is driven mainly by wind, day length and mean temperature. PC4 is driven by temperature, potential evapotranspiration and water vapour pressure, while PC5 does not reveal any clear patterns regarding climate variables (see Table S2 for details).

3.6 | Environmental association analyses

When averaging across all loci or symbionts, we found no isolation by environment (Wang & Bradburd, 2014) across any of the PC axes (Figure 4 and Figure S10). The lack of an overall isolation by environment makes it possible to identify individual variants potentially involved in responses to local environments, and we subsequently analysed each gene/symbiont separately.

For most correlations between variation in climate axes and variation in genetic, DNA methylation and microbiome features, the histograms of correlation coefficients for the null distributions appear normal (Figure 5a, grey distributions), while the distribution of actual correlation coefficients are right skewed in most cases (Figure 5a, coloured distributions; Figure S11). The peak of the observed correlation coefficient distribution mainly falls within the negative correlation coefficients, an observation that is difficult to interpret as climate-related responses, while the right-hand tail represents the most strongly positively correlated genes/microbiome features, which represent candidates for local adaptive variants. In the right-hand tail, we generally see an excess of genes/microbiome features compared to the null distribution (Figures 5a and S11).

When correcting for neutral population structure, partial Mantel correlations between the distance in climatic axes and the genetic distance among populations showed substantially more genes correlating strongly than expected based on the null distribution obtained by permuting the environmental axes, especially PC2 and PC4 (Figure 5a,b). For DNA methylation, partial Mantel analyses revealed more genes strongly correlating to PC axes than expected based on the null distribution (most clearly PC3 and PC4, Figure 5a,c). The same overall pattern was revealed, regardless of whether we corrected for population structure (Figure 5c, dark bars) or *cis* nucleotide variation (light bars). GO term enrichment analysis revealed that various broad categories of biological processes and molecular functions were enriched in genes strongly correlated to environment (Table S5), but none of them were clearly related to climate. We note that 2413 genes could not be functionally annotated, a common issue for nonmodel organisms. To further investigate whether the genes that show strong correlations to the climate axes were shared between nucleotide variants and DNA methylation variants, we used Venn diagrams (Figure 5e).

A very low number of genes showed a strong correlation with both DNA methylation variants and nucleotide variants (Figure 5e and Table S6, overlap). This is in contrast to the large number of genes co-occurring in both DNA methylation variants corrected for population structure and *cis* genetic variants within each gene (between 60% and 95%, Figure 5d and Table S6, overlap). The large overlap among gene-wise methylation variants corrected for population structure or corrected for *cis* genetic structure indicates that variation in DNA methylation is not a function of *cis*-nucleotide variation. This suggests that DNA methylation is either a function of *trans*-nucleotide variation or arises independently of nucleotide variation.

Correlation analyses of microbiome Bray–Curtis dissimilarity across all ASVs and genera and divergence in climate axes revealed

1365294x, 2022, 22, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.16696, Wiley Online Library on [03/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

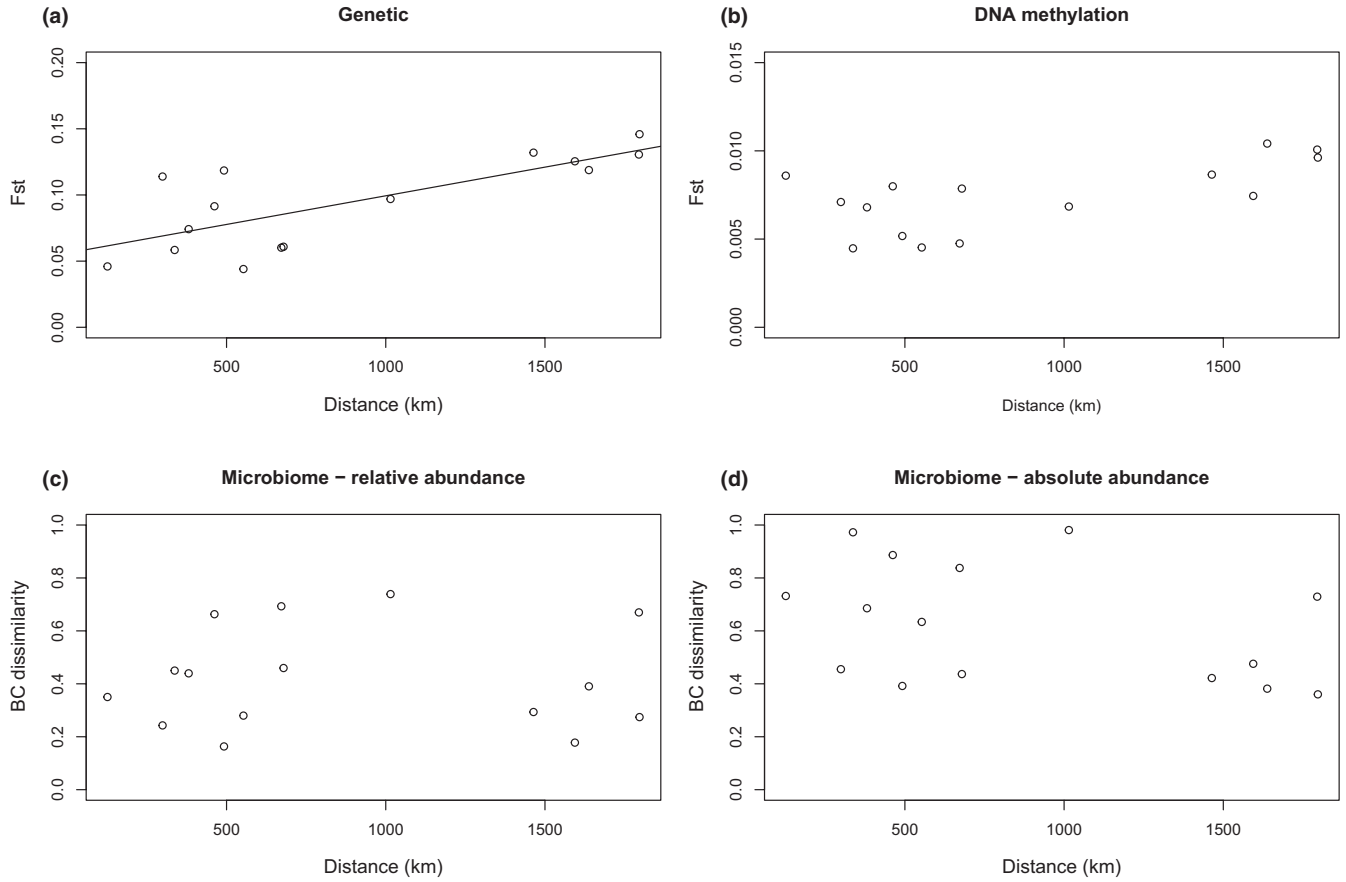


FIGURE 3 Isolation-by-distance plots of (a) genetic divergence (F_{ST}), (b) DNA methylation divergence (F_{ST}), and (c,d) microbiome divergence (Bray–Curtis [BC] dissimilarity); BC dissimilarities were estimated as a function of both (c) relative and (d) absolute abundance. Isolation-by-distance was only significant when considering genetic divergence, but this was driven by the Ndumo population, and no isolation-by-distance was observed within Namibia.

more ASVs/genera correlating strongly to climate axes than expected based on the null distribution (Figure 5a,d). Examples of the strongest correlations between climate and gene-wise SNP, WML and microbiome ASVs are shown in Figure S12, while distribution plots of correlation coefficients are presented in Figures 5a and S11.

To verify the results based on Mantel tests, we also analysed the associations using multiple regression tests on distance matrices (Castellano & Balletto, 2002; Guillot & Rousset, 2013; Legendre, 2000; Legendre et al., 2015; Raufaste & Rousset, 2001). These analyses yielded results very similar to the Mantel tests (Figure S14), suggesting that the revealed patterns are robust.

Genetic variation at individual genes correlated most closely especially with specific mean temperature parameters, as well as yearly minimum precipitation (Figure 6 top, blue bars). DNA methylation at individual genes often correlated with parameters related to minimum temperature as well as yearly minimum precipitation (Figure 6 top, orange bars). Both genetic and methylation variation within genes seem to correlate strongly with specific aspects of maximum temperature in a large number of genes (Figure 6 top). For the microbiome presented as ASVs or genera, many specific aspects of both mean temperature and precipitation were found to correlate more strongly with microbiome than the strongest correlation from the null distribution (Figure 6 bottom). Distribution plots of correlation coefficients for the correlations between genetic, DNA methylation, microbiome variation, and temperature and precipitation parameters can be seen in Figure S15.

Heatmaps of the microbiome data (Figures S17 and S18) show that significant and very strong correlations are not generally driven by one or a few ASVs. Most ASVs correlate strongly and/or significantly with few climatic parameters. An exception is *Enhydrobactor* (ASV 27, absolute abundance Figure S17a), which correlates with multiple climate parameters (Figure S17). A clear clustering of ASVs is evident around the precipitation parameters, mainly driven by *Mycoplasma* (ASV 4) and *Proteus* (ASV 26) (relative abundance, Figure S17b), but many ASVs contribute to the cluster. When sorting the ASVs according to abundance, however, no clear clustering was seen (Figure S18a,b).

4 | DISCUSSION

Populations of the social spider species Stegodyphus *dumicola* inhabit wide climatic gradients across southern Africa, raising the question of how they respond to variation in local conditions in the face of extremely low species-wide genetic diversity (Settepani et al., 2017). We investigated sources of variation that potentially

1365294x, 2022, 22, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.16696, Wiley Online Library on [03/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

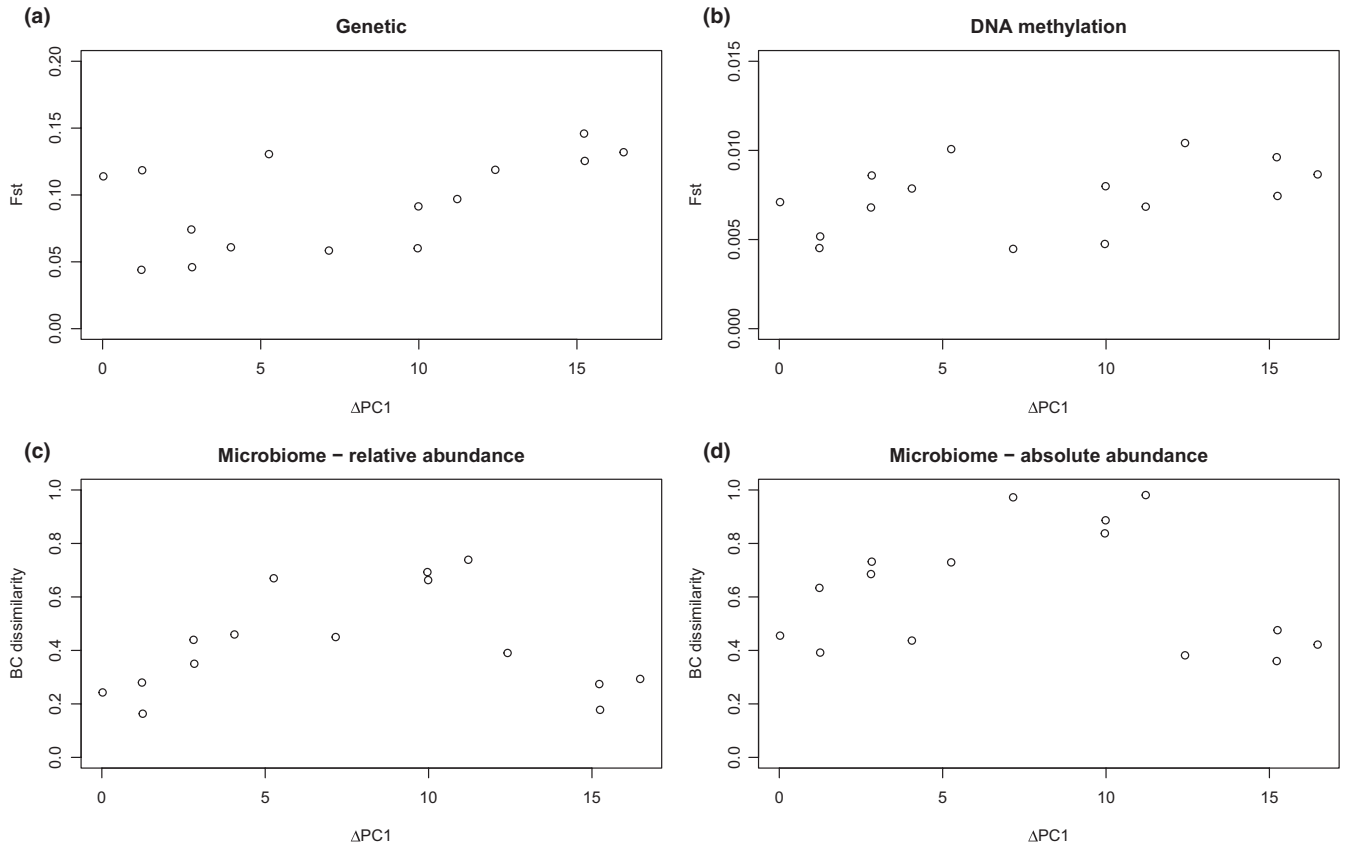


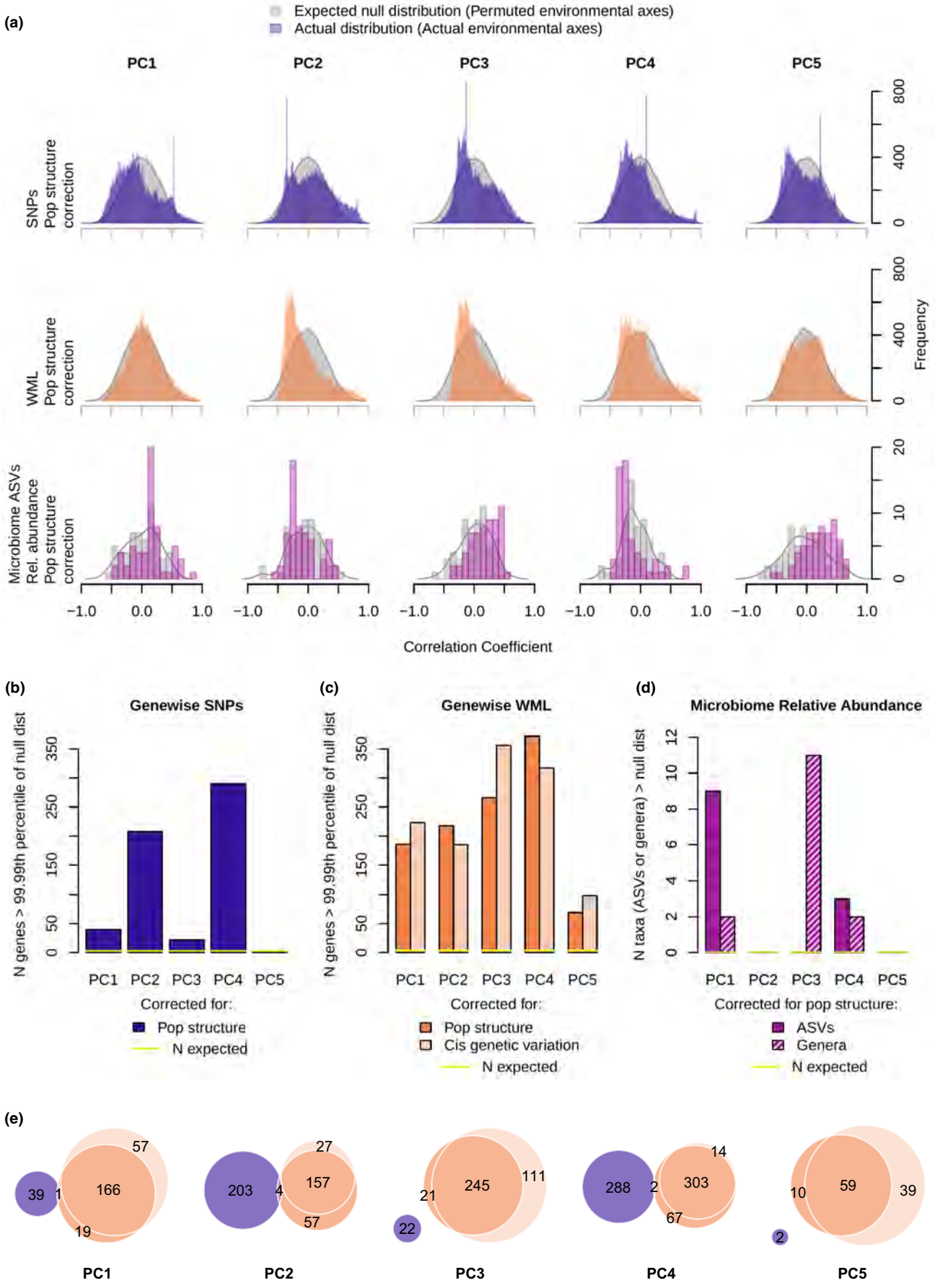
FIGURE 4 Isolation-by-environment plots after averaging (a) genetic, (b) DNA methylation and (c,d) microbiome divergence across all loci and symbionts. The environmental divergence presented here is distance on PC axis 1. Isolation-by-environment plots with PC axes 2–5 are shown in Figure S10.

mediate local responses, by correlating environmental variation with genetic, epigenetic and microbiome variation. Here we highlight three main conclusions that we discuss in more detail below. (i) In *S. dumicola*, despite low species-wide genetic diversity, we find genetic variants associated with environmental variation, consistent with patterns of local adaptation to environmental conditions, particularly in relation to mean temperatures. (ii) DNA methylation variation is associated with environmental variation, as expected if there is an epigenetic role in response to local climatic conditions. DNA methylations show different environmental association patterns than those of genetic variants, and show strong associations across temperature aspects. (iii) The bacterial microbiome correlates with environmental variation; most notably we detect the strongest associations with mean temperature and humidity-related climatic factors. These results suggest that both genetic adaptation and responses mediated by nongenetic mechanisms might contribute to population differentiation in *S. dumicola*.

4.1 | Population genetics

Population genetic structure was characterized by weak but significant isolation-by-distance, mainly driven by the South African Ndumo population, which is distant from the Namibian populations (>1500 km) (Figure 3a). When assessing only the five Namibian populations, it is clear that geographical and genetic distances do not match, suggesting that populations do not differentiate due to geographical distance. The genomic differentiation among populations (F_{ST} estimates of 9.4%) could be the result of neutral evolution, especially as a result of recurrent extinction and colonization events and genetic drift, or the differentiation could be caused by selection. Despite large census population sizes of *S. dumicola* (Settepani et al., 2017), we estimated extremely low genome-wide population-specific genetic diversities (on average $\pi = 0.00048$) (Figure 2), corroborating similar findings of a RAD sequencing study (Settepani et al., 2017). In small populations characterized by inbreeding and lack of gene flow, such as seen in the social spiders, we expect to detect strong population genetic structure caused by lineage divergence. However, high population extinction/colonization rates can act to homogenize genetic structure (Settepani et al., 2016, 2017), and indeed, species-wide genetic diversity was very low (across populations: $\pi = 0.00091$). This genetic pattern is expected to impede the efficacy of selection and local adaptation (Jensen & Bachtrog, 2011; Settepani et al., 2017), since the probability of segregating variants that are advantageous in a changing environment is lower when genetic diversity is low (Barrett & Schluter, 2008; Lande & Shannon, 1996; Rousselle et al., 2020). Nevertheless, we identified associations between genetic and

1365294x, 2022, 22, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.16696, Wiley Online Library on [03/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License



1365294x, 2022, 22, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.16696, Wiley Online Library on [03/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

FIGURE 5 Correlation results when running partial Mantel tests correlating gene-wise genetic distance, gene-wise methylation distance and distance in microbiome ASVs/genera with environmental distances. (a) Distribution plots of correlation coefficients for the expected null distribution (grey), made by permuting the environmental axes, and the actual distribution (coloured), made with the actual environmental axes. For the distribution plots of the remaining data sets see Figure S11. (b,c) Number of genes exceeding the 99.99th percentile of the null distribution. (b) Gene-wise nucleotide variation (blue). (c) Gene-wise DNA methylation variation corrected for population structure (dark orange) and *cis* genetic variation (light orange). (d) Number of microbiome features with a correlation coefficient exceeding the highest correlation coefficient in the null distribution. The microbiome is represented as relative abundance of ASVs (purple) and genera (hashed). Absolute abundance is given in Figure S13. The horizontal yellow lines indicate the theoretically expected number of genes/microbiome features (0.01 percent of correlations). (e) Venn diagrams showing the number of genes co-occurring between nucleotide variation (blue), DNA methylation corrected for population structure (dark orange) and DNA methylation corrected for *cis* genetic variation (light orange) (Table S6).

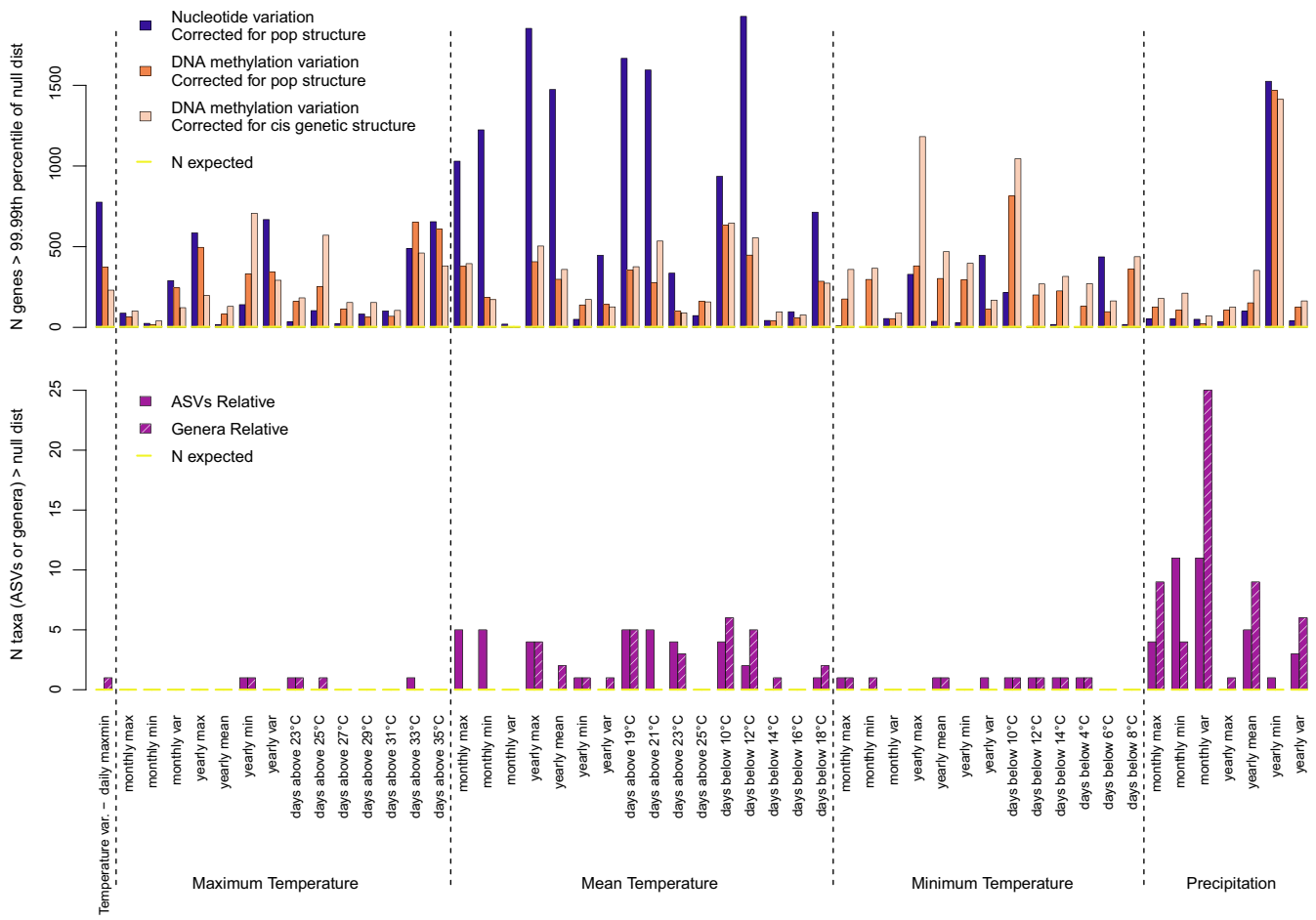


FIGURE 6 Correlations, presented as the number of genes or microbiome components correlating very strongly with distance in specific temperature and precipitation parameters. Upper graph: Number of genes correlating more strongly than the 99.99th percentile of the null distribution. Blue bars: Gene-wise nucleotide variation, orange bars: DNA Methylation variation, dark orange bars: Corrected for population structure, light orange bars: Corrected for *cis* genetic variation. Bottom graph: Number of microbiome features represented as relative abundance correlating more strongly with environmental parameters than the highest correlation coefficient of the null distribution. Purple bars: Single ASVs, hashed bars: Genera. A similar representation of results for absolute abundance is presented in Figure S16.

environmental variants (Figure 5a), consistent with the existence of adaptive genetic diversity. In support of this scenario, the distribution of genetic differentiation among populations of individual genes is relatively even (Figure S7a), contrasting with the expected left-skewed bell shape for neutral loci (Schwartz et al., 2007), suggesting that selection may have affected individual genes differently.

4.2 | Association patterns between climate variables and nucleotide variation

We identified hundreds of strong associations between nucleotide variants averaged across genes and climate axes (Figure 5b), contradicting our hypothesis that populations with strong drift and low efficacy of selection harbour low amounts of adaptive genetic variation.

1365294x, 2022, 22, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.16696, Wiley Online Library on [03/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Focusing on associations between different aspects of temperature and precipitation and nucleotide variation within genes suggests that aspects of mean temperature (Figure 6) are the strongest drivers of local adaptation to climate. In populations with high genetic drift, strong selection is required to maintain adaptive variation and local adaptation. Exposure to high temperatures is known to exert selection for phenotypic responses to avoid heat stress, leading to local adaptation in temperature responses in natural populations of arthropods (Sørensen & Loeschcke, 2002; Tregenza et al., 2021; Williams et al., 2012). In addition, our study species, *S. dumicola*, shows population-specific variation in behavioural and physiological responses to high temperatures (Barton, 2011; Malmos et al., 2021; Sandfeld et al., 2022), substantiating a role for genetic variation in temperature adaptation. However, relationships between genes that are differentiated between climatically divergent populations and specific phenotypic adaptations have yet to be established.

4.3 | DNA methylation variation

We found relatively high CpG methylation in *S. dumicola* (about 10%) (Table S4) as compared with most invertebrates (see overview by Bewick et al., 2017; de Mendoza et al., 2020). This finding corroborates previous findings in social spiders (Liu, Aagaard, et al., 2019). With no indication of genome-wide isolation-by-environment (Figure 4), the overall pattern of DNA methylation is seemingly not shaped by climate. However, analyses of DNA methylation of single genes separately revealed a long tail of genes that show strong differentiation between populations (Figure S7b). It is possible that methylation in these genes is responsible for aspects of phenotype that relate to temperature tolerance (Agwunobi et al., 2021), for example by potentially regulating gene expression or being involved in alternative splicing (Flores et al., 2013; Lev Maor et al., 2015; Liu, Aagaard, et al., 2019; Xu et al., 2021). The average level of DNA methylation divergence among populations was considerably lower than that of nucleotide divergence (Figure S7a,b, on average $F_{ST(WML)} = 0.007$). However, because these F_{ST} estimates are based on different types of data and analyses, they are not easily comparable (Coates et al., 2009). One plausible reason for lower divergence in DNA methylation compared with nucleotide divergence is that a large proportion of DNA methylation may be constrained in its variation, for example due to roles in development (Gao et al., 2012), differences between tissues or individual differences (Marshall et al., 2019). This implies that only a subset of methylation variants may differentiate among populations. DNA methylations can be induced by environmental variance as shown in fish, mammals, plants, birds and invertebrates (Bossdorf et al., 2008; Cropley et al., 2012; Harney et al., 2022; Heckwolf et al., 2019; Nätt et al., 2012), and are proposed to have the potential to modulate phenotypes in a plastic manner (Duncan et al., 2022; Flores et al., 2012; Gatzmann et al., 2018; Marshall et al., 2019). Furthermore, inducible DNA methylations can be transmitted across generations, which exposes them to evolutionary forces at least on short timescales (Cubas et al., 1999; Nätt et al., 2012; Riddle & Richards, 2002; Sutherland et al., 2000).

4.4 | Association patterns between climate variables and DNA methylation variation

Currently, we do not know whether DNA methylations can mediate local responses to ecological factors in spiders. However, we identified an excess of strong associations between variants in DNA methylation and divergence across climate axes (Figure 5c). Some genes differed substantially in their patterns of methylation between populations, a pattern consistent with a role in mediating responses to local climate differences. Methylations mediating responses to the local climate have been found in other species; for example, gene body methylation of a coral changed in response to transplantation, and corals with methylation patterns more similar to the local specimens had higher fitness (Dixon et al., 2018). Methylation patterns are influenced by environmental factors such as salinity in *Daphnia* (Asselman et al., 2015), and both salinity and temperature in the ascidian *Ciona* (Hawes et al., 2018). In a cockroach, *Diploptera punctata*, methylation patterns in heat shock protein 70 respond to temperature, potentially providing a fast-response mechanism to regulate expression of heat shock proteins (Peña et al., 2021). We found that associations between DNA methylation level and climatic variables were largely independent of *cis*-genetic variation and of overall genetic population structure (Figure 5e, overlap between dark/light orange). This indicates that DNA methylation is not solely a function of the local DNA sequence itself, but we cannot rule out that DNA methylation is regulated by *trans*-acting loci or influenced by SNPs further upstream of the gene region. Very few genes (seven in total) showed evidence of strong associations between both nucleotide and DNA methylation variants and climate axes (Figure 5e, overlap blue/orange), suggesting that nucleotide variants and DNA methylation to a large extent are independent. These few genes present interesting candidates to investigate functional relationships in more detail, for example by using experimental molecular methods combined with analyses of gene expression associated with phenotypic changes. Speculative explanations for the pattern in these genes include: (i) locally adapted genes with DNA methylation variants fine-tuning the local response; (ii) genetically based differences in gene expression cause differences in DNA methylation patterns (Secco et al., 2015); and (iii) a plastic gene that has become locally adapted (plasticity first hypothesis, Perry et al., 2018).

More genes showed a strong association to climate in their DNA methylation than nucleotide variants (Figure 5b,c). This may be surprising considering the lower number of genes that show significant differentiation among populations in DNA methylation compared to nucleotides (Figure S7a,b). Drift and limited gene flow act to increase differentiation of genetic variants, which may not be the case for DNA methylation variants if they are plastically induced. This could lead to more genes that are divergent with respect to nucleotide variation than to DNA methylation. However, if DNA methylations are generally transmitted across generations, drift will also lead to differentiation

1365294x, 2022, 22, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.16696, Wiley Online Library on [03/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

of DNA methylation among populations. Despite the lower number of genes diverging in DNA methylation, many more genes are both strongly and significantly associated with climate in DNA methylation. This suggests that genes that are differentially methylated among populations are important in mediating local responses to climate.

The scrutiny of selected temperature and precipitation variables and their associations to DNA methylation variation within genes suggests that DNA methylation may be particularly involved in responses to differences in minimum temperature, but many genes associate strongly across environmental parameters (Figure 6, orange bars). Associations between DNA methylation and temperature have previously been identified. For example, DNA methylation level in 43 RAD loci was highly associated with maximum temperature in the oak species *Quercus lobata* in California (Gugger et al., 2016). Methylation level in Hsp70 responds to heat in the mollusc *Biomphalaria glabratahe* (Ittiprasert et al., 2015). However, our results suggest that minimum temperature may be a driver for differential DNA methylation, potentially by responding to low temperature through fast plastic responses. This was found in an alpine Brassicaceae that respond to chilling with an alteration of DNA methylation, suggesting that methylations mediate fast responses to cold stress (Song et al., 2015). In the goldenrod gall moth, low temperatures cause an increase in expression of DNA methyltransferases (Williamson et al., 2021). A study on ticks showed the importance of DNA methyltransferases in regulating the cold response (Agwunobi et al., 2021); they knocked out the DNA methyltransferases and ticks subsequently exposed to sublethal temperatures died.

Our results also indicate that local responses to precipitation may be mediated primarily by DNA methylation variation (Figure 6). In plants it has been shown that epigenetic signals may guide development of stomatal cells in response to relative humidity in the environment (Tricker et al., 2012), and in humans associations between blood cell methylation patterns and ambient relative humidity were identified, which furthermore interacts with temperature (Bind et al., 2014).

4.5 | Bacterial microbiome composition

Analyses of *S. dumicola* microbiome composition revealed no pattern of isolation by distance. Within the same location, we recovered substantial variation in microbiome composition between nests, with only little additional variation found between different geographical locations (Figure S6). This finding corroborates previous microbiome studies of *S. dumicola* populations across Southern Africa (Busck et al., 2020, 2022). The actual symbionts identified in the *S. dumicola* microbiome also overlap substantially. We know from preliminary data that bacteria are not vertically but rather socially transmitted among nest mates in *S. dumicola* (our unpublished data). Within nests, the microbiome composition does not change much across generations, suggesting relatively high transmission fidelity within nests (Busck et al., 2022), but nests within a population often carry substantially different microbiome compositions (Busck et al., 2020, 2022).

4.6 | Association patterns between climate variables and microbiome variation

Correlation analyses revealed an excess of strong associations between microbiome composition and the majority of environmental axes (Figure 5d). In some cases, the microbiome composition or presence of certain strains was found to associate with the ambient environment of the host, revealing a host phenotype better fitted to a particular environment (Dunbar et al., 2007; Herrera et al., 2021). Such changes in microbiome composition can be caused by mutualistic relationships with the host, or differential survival within hosts across a climate gradient. Several aspects of precipitation associated more than expected with relative abundance of bacterial symbionts (Figure 6). A previous study showed an association between high precipitation and microbiome composition (Busck et al., 2022), and together these studies indicate that the microbiome composition of *S. dumicola* is shaped by aspects of humidity. This association is driven by 10 ASVs from different taxonomic groups that correlate strongly and/or significantly with aspects of precipitation (Figure S17). An effect of precipitation-related variables has also been found in mosquitoes, where the gut microbiome changes along a landscape–moisture gradient (Medeiros et al., 2021), and exposure to altered humidity has been shown to change the microbiome in mice (Yin et al., 2022). However, our association study cannot discern whether symbiont abundance is shaped directly by humidity irrespective of the host or indirectly by the host as a response to humidity. Further steps are required to disentangle these processes and investigate a potential functional relationship with the host. In contrast to the patterns recovered for genetic and epigenetic variation, we detected a more scattered pattern of strong associations between microbiome composition and aspects of mean temperature (Figure 5d), a pattern driven primarily by five ASVs from different genera (Figure S17). A relationship between temperature and microbiome was found in other species, for example for the insect *Wolbachia* in relation to mean temperature (Woodhams et al., 2020), or for the *Drosophila* gut microbiome (Mazzucco & Schlötterer, 2021; Walters et al., 2020).

We find that the most abundant symbionts are not necessarily those showing the strongest correlations with the environment (Figure S18), indicating that strict abundance filters on microbiome data may remove functionally important symbionts. Under the assumption of a mutualistic relationship between host and symbiont, this implies that symbionts that govern host phenotypic responses may be found among the less abundant symbionts in the microbiome community. An example of a low-abundance taxon that can contribute valuable functions for the host is found in the human gut: Christensenellaceae are associated with health and longevity despite mostly being present with a relative abundance well below 0.1% (Kong et al., 2016; Waters & Ley, 2019). In addition to abundance and presence/absence, strain variation may be important in a mutualistic relationship between host and symbiont. Strain variation

1365294x, 2022, 22, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.16696, Wiley Online Library on [03/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

caused by indel polymorphism was shown to be important in providing the host with different functions in the aphid symbiont *Buchnera* (Dunbar et al., 2007).

4.7 | Concluding remarks

We aimed to take a step towards understanding varied sources of variation that allow a species to occupy a range of habitats, by examining associations between environmental variation and genetic, epigenetic and microbiome variation. We identified gene-wise genetic variants that are associated strongly with environmental variation, particularly in mean temperature, a result which is consistent with local genetic adaptation. DNA methylations show different environmental association patterns compared with genetic variants, by having strong correlations to all climate axes and across aspects of temperature and precipitation. This pattern follows the expectation of an epigenetic role in responses to local climatic conditions. The microbiome also correlated with environmental variation, but also showed an independent pattern of association with most strong associations being with mean temperature and humidity-related climatic factors. We hypothesize that nongenetic sources of variation underlying adaptive responses to environmental change may be important in species with low standing variation. The next steps would be to assess functional relationships and determine whether molecular variants associated with phenotypic change are inducible and/or transmitted across generations. Common garden studies designed to substantiate links between the environment, phenotypic change and underlying molecular mechanism may be useful for establishing specific functional relationships, while causal relationships may require molecular experiments and analyses of gene expression connected with phenotypic change.

AUTHOR CONTRIBUTIONS

AA, TB, JB and AS designed the research and obtained funding; AA, SL, JB and MBL performed research and data analysis; KJFV, TB, TT, AS and MBL aided by suggesting new analysis methods; all authors wrote the manuscript, with substantial contributions from AA, JB, TB and TT.

ACKNOWLEDGEMENTS

We thank Virginia Settepani, Anja Junghanns and Tharina Bird for collecting spiders from the field. We also owe thanks to Marie Rosenstand Hansen, Susanne Nielsen, Britta Poulsen and Lykke Beinta Bjærge Bamdali for spider care and technical assistance in the laboratory, and Mette Marie Busck for establishing the microbiome data analysis pipeline. We are indebted to GenomeDK and Aarhus University for providing computational resources for the computing of data (GenomeDK cluster, genome.au.dk). The study was supported the Danish Council for Independent Research (grant no. DFF - 6108-00565) and by a Novo Nordisk Foundation Interdisciplinary Synergy Grant (NNF16OC0021110).

Permits for exporting the spiders for research purposes were obtained from the Ministry of Environment and Tourism, Windhoek, Namibia, permit no. 2110/2016, Ezemvelo KZN Wildlife, Pietermaritzburg, KwaZulu-Natal, South Africa, permit no. OP 5074/2014 and Department of Environment and Nature Conservation, Kimberly, South Africa, permit no. FAUNA 1450/2015.

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

DATA AVAILABILITY STATEMENT

All data have been uploaded to GenBank: Bioproject PRJNA754001. Previously published genetic data are available at GenBank: Bioproject PRJNA510316. A document with information about which files are available at which bioproject has been uploaded as Supporting Information for this article.

ORCID

Anne Aagaard https://orcid.org/0000-0001-7044-5177
Tom Tregenza https://orcid.org/0000-0003-4182-2222
Marie Braad Lund https://orcid.org/0000-0003-1543-1009
Andreas Schramm https://orcid.org/0000-0002-7614-9616
Koen J. F. Verhoeven https://orcid.org/0000-0003-3002-4102
Jesper Bechsgaard https://orcid.org/0000-0003-3273-0174
Trine Bilde https://orcid.org/0000-0002-0341-161X

REFERENCES

Agwunobi, D. O., Zhang, M., Shi, X., Zhang, S., Zhang, M., Wang, T., Masoudi, A., Yu, Z., & Liu, J. (2021). DNA methyltransferases contribute to cold tolerance in ticks *Dermacentor silvarum* and *Haemaphysalis longicornis* (Acari: Ixodidae). *Frontiers in Veterinary Science*, *8*, 1–17. https://doi.org/10.3389/fvets.2021.726731

Asselman, J., de Coninck, D. I., Vandegehuchte, M. B., Jansen, M., Decaestecker, E., de Meester, L., vanden Bussche, J., Vanhaecke, L., Janssen, C. R., & de Schamphelaere, K. A. (2015). Global cytosine methylation in *Daphnia magna* depends on genotype, environment, and their interaction. *Environmental Toxicology and Chemistry*, *34*(5), 1056–1061. https://doi.org/10.1002/ETC.2887

Bang, C., Dagan, T., Deines, P., Dubilier, N., Duschl, W. J., Fraune, S., Hentschel, U., Hirt, H., Hülter, N., Lachnit, T., Picazo, D., Pita, L., Pogoreutz, C., Rädecker, N., Saad, M. M., Schmitz, R. A., Schulenburg, H., Voolstra, C. R., Weiland-Bräuer, N., … Bosch, T. C. G. (2018). Metaorganisms in extreme environments: Do microbes play a role in organismal adaptation? *Zoology*, *127*, 1–19. https://doi.org/10.1016/j.zool.2018.02.004

Barrett, R. D. H., & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, *23*(1), 38–44. https://doi.org/10.1016/j.tree.2007.09.008

Barton, B. T. (2011). Local adaptation to temperature conserves top-down control in a grassland food web. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1721), 3102–3107. https://doi.org/10.1098/rspb.2011.0030

Bewick, A. J., Vogel, K. J., Moore, A. J., & Schmitz, R. J. (2017). Evolution of DNA methylation across insects. *Molecular Biology and Evolution*, *34*(3), 654–665. https://doi.org/10.1093/molbev/msw264

1365294x, 2022, 22, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.16696, Wiley Online Library on [03/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Bind, M.-A., Zanobetti, A., Gasparrini, A., Peters, A., Coull, B., Baccarelli, A., Tarantini, L., Koutrakis, P., Vokonas, P., Schwartz, J. (2014). Effects of temperature and relative humidity on DNA methylation. *Epidemiology (Cambridge, Mass.)*, *25*(4), 561. https://doi.org/10.1097/EDE.0000000000000120

Bittinger, K. (2020). *usedist: Distance matrix utilities*. Retrieved from https://cran.r-project.org/package=usedist

Bossdorf, O., Richards, C. L., & Pigliucci, M. (2008). Epigenetics for ecologists. *Ecology Letters*, *11*(2), 106–115. https://doi.org/10.1111/j.1461-0248.2007.01130.x

Buchfink, B., Reuter, K., & Drost, H. G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods*, *18*(4), 366–368. https://doi.org/10.1038/s41592-021-01101-x

Busck, M. M., Lund, M. B., Bird, T. L., Bechsgaard, J. S., Bilde, T., & Schramm, A. (2022). Temporal and spatial microbiome dynamics across natural populations of the social spider *Stegodyphus dumicola*. *FEMS Microbiology Ecology*, *98*, 1–11. https://doi.org/10.1093/femsec/fiac015

Busck, M. M., Settepani, V., Bechsgaard, J., Lund, M. B., Bilde, T., & Schramm, A. (2020). Microbiomes and specific symbionts of social spiders: Compositional patterns in host species, populations, and nests. *Frontiers in Microbiology*, 1–14. https://doi.org/10.3389/FMICB.2020.01845

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583. https://doi.org/10.1038/nmeth.3869

Cantalapiedra, C. P., Herɲandez-Plaza, A., Letunic, I., Bork, P., & Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional annotation, Orthology assignments, and domain prediction at the metagenomic scale. *Molecular Biology and Evolution*, *38*(12), 5825–5829. https://doi.org/10.1093/MOLBEV/MSAB293

Castellano, S., & Balletto, E. (2002). Is the partial mantel test inadequate? *Evolution*, *56*(9), 1871–1873. https://doi.org/10.1111/j.0014-3820.2002.tb00203.x

Cavalli, G. (2006). Chromatin and epigenetics in development: Blending cellular memory with cell fate plasticity. *Development*, *133*(11), 2089–2094. https://doi.org/10.1242/dev.02402

Charlesworth, D., Barton, N. H., & Charlesworth, B. (2017). The sources of adaptive variation. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1855), 1–12. https://doi.org/10.1098/rspb.2016.2864

Chen, H. (2018). *VennDiagram: Generate High-Resolution Venn and Euler Plots*. Retrieved from https://cran.r-project.org/package=VennDiagram%0A

Choi, J., Lyons, D. B., Kim, M. Y., Moore, J. D., & Zilberman, D. (2020). DNA methylation and histone H1 jointly repress transposable elements and aberrant intragenic transcripts. *Molecular Cell*, *77*(2), 310–323.e7. https://doi.org/10.1016/J.MOLCEL.2019.10.011

Coates, B. S., Sumerford, D. V., Miller, N. J., Kim, K. S., Sappington, T. W., Siegfried, B. D., & Lewis, L. C. (2009). Comparative performance of single nucleotide polymorphism and microsatellite markers for population genetic analysis. *Journal of Heredity*, *100*(5), 556–564. https://doi.org/10.1093/jhered/esp028

Cropley, J. E., Dang, T. H. Y., Martin, D. I. K., & Suter, C. M. (2012). The penetrance of an epigenetic trait in mice is progressively yet reversibly increased by selection and environment. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1737), 2347–2353. https://doi.org/10.1098/rspb.2011.2646

Cubas, P., Vincent, C., & Coen, E. (1999). An epigenetic mutation responsible for natural variation in floral symmetry. *Nature*, *401*(6749), 157–161. https://doi.org/10.1038/43657

Dahlgaard, J., Loeschcke, V., Michalak, P., & Justesen, J. (1998). Induced thermotolerance and associated expression of the heat-shock protein Hsp70 in adult *Drosophila melanogaster*. *Functional Ecology*, *12*(5), 786–793. https://doi.org/10.1046/j.1365-2435.1998.00246.x

Danecek, P., & McCarthy, S. A. (2017). BCFtools/csq: Haplotype-aware variant consequences. *Bioinformatics*, *33*(13), 2037–2039. https://doi.org/10.1093/bioinformatics/btx100

de Mendoza, A., Lister, R., & Bogdanovic, O. (2020). Evolution of DNA methylome diversity in eukaryotes. *Journal of Molecular Biology*, *432*(6), 1687–1705. https://doi.org/10.1016/j.jmb.2019.11.003

Dimond, J. L., & Roberts, S. B. (2020). Convergence of DNA methylation profiles of the reef coral *Porites astreoides* in a novel environment. *Frontiers in Marine Science*, *6*, 792. https://doi.org/10.3389/fmars.2019.00792

Dixon, G., Liao, Y., Bay, L. K., & Matz, M. V. (2018). Role of gene body methylation in acclimatization and adaptation in a basal metazoan. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(52), 13342–13346. https://doi.org/10.1073/pnas.1813749115

Donelson, J. M., Sunday, J. M., Figueira, W. F., Gaitán-Espitia, J. D., Hobday, A. J., Johnson, C. R., … Munday, P. L. (2019). Understanding interactions between plasticity, adaptation and range shifts in response to marine environmental change. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, *374*(1768), 20180186. https://doi.org/10.1098/rstb.2018.0186

Dubin, M. J., Zhang, P., Meng, D., Remigereau, M. S., Osborne, E. J., Paolo Casale, F., Drewe, P., Kahles, A., Jean, G., Vilhjálmsson, B., Jagoda, J., Irez, S., Voronin, V., Song, Q., Long, Q., Rätsch, G., Stegle, O., Clark, R. M., & Nordborg, M. (2015). DNA methylation in *Arabidopsis* has a genetic basis and shows evidence of local adaptation. *eLife*, *4*, 1–23. https://doi.org/10.7554/eLife.05255

Dunbar, H. E., Wilson, A. C. C., Ferguson, N. R., & Moran, N. A. (2007). Aphid thermal tolerance is governed by a point mutation in bacterial symbionts. *PLoS Biology*, *5*(5), 1006–1015. https://doi.org/10.1371/journal.pbio.0050096

Duncan, E. J., Cunningham, C. B., & Dearden, P. K. (2022). Phenotypic plasticity: What has DNA methylation got to do with it? *13*(2), 110. Retrieved from https://www.mdpi.com/2075-4450/13/2/110/htm

Falcon, S., & Gentleman, R. (2007). Using GOstats to test gene lists for GO term association. *Bioinformatics*, *23*(2), 257–258. https://doi.org/10.1093/BIOINFORMATICS/BTL567

Fischer, M. C., Rellstab, C., Tedder, A., Zoller, S., Gugerli, F., Shimizu, K. K., … Widmer, A. (2013). Population genomic footprints of selection and associations with climate in natural populations of *Arabidopsis halleri* from the Alps. *Molecular Ecology*, *22*(22), 5594–5607. https://doi.org/10.1111/mec.12521

Flores, K. B., Wolschin, F., & Amdam, G. V. (2013). The role of methylation of DNA in environmental adaptation. *Integrative and Comparative Biology*, *53*(2), 359–372. https://doi.org/10.1093/icb/ict019

Flores, K. B., Wolschin, F., Corneveaux, J. J., Allen, A. N., Huentelman, M. J., & Amdam, G. V. (2012). Genome-wide association between DNA methylation and alternative splicing in an invertebrate. *BMC Genomics*, *13*(1), 1–9. https://doi.org/10.1186/1471-2164-13-480

Fox, R. J., Donelson, J. M., Schunter, C., Ravasi, T., & Gaitán-Espitia, J. D. (2019). Beyond buying time: The role of plasticity in phenotypic adaptation to rapid environmental change. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, *374*(1768), 1–9. https://doi.org/10.1098/rstb.2018.0174

Gao, F., Liu, X., Wu, X. P., Wang, X. L., Gong, D., Lu, H., Xia, Y., Song, Y., Wang, J., Du, J., Liu, S., Han, X., Tang, Y., Yang, H., Jin, Q., Zhang, X., & Liu, M. (2012). Differential DNA methylation in discrete developmental stages of the parasitic nematode *trichinella spiralis*. *Genome Biology*, *13*(10), R100. https://doi.org/10.1186/gb-2012-13-10-r100

Gatzmann, F., Falckenhayn, C., Gutekunst, J., Hanna, K., Raddatz, G., Carneiro, V. C., & Lyko, F. (2018). The methylome of the marbled crayfish links gene body methylation to stable expression of poorly accessible genes. *Epigenetics & Chromatin*, *11*(1), 1–12. https://doi.org/10.1186/s13072-018-0229-6

1365294x, 2022, 22, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.16696, Wiley Online Library on [03/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Gefen, E., Talal, S., Brendzel, O., Dror, A., & Fishman, A. (2015). Variation in quantity and composition of cuticular hydrocarbons in the scorpion *Buthus occitanus* (Buthidae) in response to acute exposure to desiccation stress. *Comparative Biochemistry and Physiology. Part A*, *182*, 58–63. https://doi.org/10.1016/J.CBPA.2014.12.004

Ghalambor, C. K., Hoke, K. L., Ruell, E. W., Fischer, E. K., Reznick, D. N., & Hughes, K. A. (2015). Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature. *Nature*, *525*(7569), 372–375. https://doi.org/10.1038/nature15256

Ghalambor, C. K., McKay, J. K., Carroll, S. P., & Reznick, D. N. (2007). Adaptive versus nonadaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Functional Ecology*, *21*(3), 394–407. https://doi.org/10.1111/j.1365-2435.2007.01283.x

Gong, T. W., Fairfield, D. A., Fullarton, L., Dolan, D. F., Altschuler, R. A., Kohrman, D. C., & Lomax, M. I. (2012). Induction of heat shock proteins by hyperthermia and noise overstimulation in Hsf1−/− mice. *JARO – Journal of the Association for Research in Otolaryngology*, *13*(1), 29–37. https://doi.org/10.1007/s10162-011-0289-9

Gonzalo-Turpin, H., & Hazard, L. (2009). Local adaptation occurs along altitudinal gradient despite the existence of gene flow in the alpine plant species *Festuca eskia*. *Journal of Ecology*, *97*(4), 742–751. https://doi.org/10.1111/j.1365-2745.2009.01509.x

Goslee, S. C., & Urban, D. L. (2007). The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, *22*(7), 1–19. https://doi.org/10.18637/jss.v022.i07

Grieser, J., Gommes, R., & Bernardi, M. (2006). New LocClim – the local climate estimator of FAO. *Geophysical Research Abstracts*, *8*(1), 8305.

Gugger, P. F., Fitz-Gibbon, S., Pellegrini, M., & Sork, V. L. (2016). Species-wide patterns of DNA methylation variation in *Quercus lobata* and their association with climate gradients. *Molecular Ecology*, *25*(8), 1665–1680. https://doi.org/10.1111/mec.13563

Guillot, G., & Rousset, F. (2013). Dismantling the mantel tests. *Methods in Ecology and Evolution*, *4*(4), 336–344. https://doi.org/10.1111/2041-210x.12018

Hanson, D., Hu, J., Hendry, A. P., & Barrett, R. D. H. (2017). Heritable gene expression differences between lake and stream stickleback include both parallel and antiparallel components. *Heredity*, *119*(5), 339–348. https://doi.org/10.1038/hdy.2017.50

Harney, E., Paterson, S., Collin, H., Chan, B. H. K., Bennett, D., & Plaistow, S. J. (2022). Pollution induces epigenetic effects that are stably transmitted across multiple generations. *Evolution Letters*, *6*(2), 118–135. https://doi.org/10.1002/EVL3.273

Hawes, N. A., Tremblay, L. A., Pochon, X., Dunphy, B., Fidler, A. E., & Smith, K. F. (2018). Effects of temperature and salinity stress on DNA methylation in a highly invasive marine invertebrate, the colonial ascidian *Didemnum vexillum*. *PeerJ*, *2018*(6), 1–18. https://doi.org/10.7717/PEERJ.5003/SUPP-4

Heckwolf, M. J., Meyer, B. S., Häsler, R., Höppner, M. P., Eizaguirre, C., & Reusch, T. B. H. (2019). Two different epigenetic pathways detected in wild three-spined sticklebacks are involved in salinity adaptation. BioRxiv, 649574. doi: https://doi.org/10.1101/649574

Henikoff, S., Furuyama, T., & Ahmad, K. (2004). Histone variants, nucleosome assembly and epigenetic inheritance. *Trends in Genetics*, *20*(7), 320–326. https://doi.org/10.1016/j.tig.2004.05.004

Henry, L. P., Bruijning, M., Forsberg, S. K. G., & Ayroles, J. F. (2019). Can the microbiome influence host evolutionary trajectories? BioRxiv, 700237. doi: https://doi.org/10.1101/700237

Herlemann, D. P. R., Labrenz, M., Jürgens, K., Bertilsson, S., Waniek, J. J., & Andersson, A. F. (2011). Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME Journal*, *5*(10), 1571–1579. https://doi.org/10.1038/ismej.2011.41

Herrera, M., Klein, S. G., Campana, S., Chen, J. E., Prasanna, A., Duarte, C. M., & Aranda, M. (2021). Temperature transcends partner specificity in the symbiosis establishment of a cnidarian. *ISME Journal*, *15*(1), 141–153. https://doi.org/10.1038/s41396-020-00768-y

Holliday, R. (1987). The inheritance of epigenetic defects. *Science, New Series*, *238*(4824), 163–170.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., Mende, D. R., Letunic, I., Rattei, T., Jensen, L. J., von Mering, C., & Bork, P. (2019). eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, *47*(D1), D309–D314. https://doi.org/10.1093/NAR/GKY1085

Ittiprasert, W., Miller, A., Knight, M., Tucker, M., & Hsieh, M. H. (2015). Evaluation of cytosine DNA methylation of the *Biomphalaria glabratahe* at shock protein 70 locus after biological and physiological stresses. *Journal of Parasitology and Vector Biology*, *7*(10), 182–193. https://doi.org/10.5897/JPVB2015.0207

Jablonka, E. (2017). The evolutionary implications of epigenetic inheritance. *Interface Focus*, *7*(5), 1–46. https://doi.org/10.1098/rsfs.2016.0135

Jensen, J. D., & Bachtrog, D. (2011). Characterizing the influence of effective population size on the rate of adaptation: Gillespie's Darwin domain. *Genome Biology and Evolution*, *3*(1), 687–701. https://doi.org/10.1093/gbe/evr063

Keller, T. E., Han, P., & Yi, S. V. (2016). Evolutionary transition of promoter and gene body DNA methylation across invertebrate-vertebrate boundary. *Molecular Biology and Evolution*, *33*(4), 1019–1028. https://doi.org/10.1093/molbev/msv345

Kofler, R., Orozco-terWengel, P., de Maio, N., Pandey, R. V., Nolte, V., Futschik, A., Kosiol, C., & Schlötterer, C. (2011). Popoolation: A toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS One*, *6*(1), e15925. https://doi.org/10.1371/journal.pone.0015925

Kofler, R., Pandey, R. V., & Schlötterer, C. (2011). PoPoolation2: Identifying differentiation between populations using sequencing of pooled DNA samples (Pool-seq). *Bioinformatics*, *27*(24), 3435–3436. https://doi.org/10.1093/bioinformatics/btr589

Kong, F., Hua, Y., Zeng, B., Ning, R., Li, Y., & Zhao, J. (2016). Gut microbiota signatures of longevity. *Current Biology*, *26*(18), R832–R833. https://doi.org/10.1016/j.cub.2016.08.015

Krueger, F., & Andrews, S. R. (2011). Bismark: A flexible aligner and methylation caller for bisulfite-seq applications. *Bioinformatics*, *27*(11), 1571–1572. https://doi.org/10.1093/bioinformatics/btr167

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549. https://doi.org/10.1093/molbev/msy096

Kvist, J., Gonçalves Athanàsio, C., Shams Solari, O., Brown, J. B., Colbourne, J. K., Pfrender, M. E., & Mirbahai, L. (2018). Pattern of DNA methylation in *daphnia*: Evolutionary perspective. *Genome Biology and Evolution*, *10*(8), 1988–2007. https://doi.org/10.1093/gbe/evy155

Lande, R. (2009). Adaptation to an extraordinary environment by evolution of phenotypic plasticity and genetic assimilation. *Journal of Evolutionary Biology*, *22*(7), 1435–1446. https://doi.org/10.1111/j.1420-9101.2009.01754.x

Lande, R., & Shannon, S. (1996). The role of genetic variation in adaptation and population persistence in a changing environment. *Evolution*, *50*(1), 434–437. https://doi.org/10.1111/j.1558-5646.1996.tb04504.x

Leffler, E. M., Bullaughey, K., Matute, D. R., Meyer, W. K., Ségurel, L., Venkat, A., Andolfatto, P., & Przeworski, M. (2012). Revisiting an old Riddle: What determines genetic diversity levels within species? *PLoS Biology*, *10*(9), e1001388. https://doi.org/10.1371/journal.pbio.1001388

Legendre, P. (2000). Comparison of permutation methods for the partial correlation and partial mantel tests. *Journal of Statistical Computation*

1365294x, 2022, 22, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.16696, Wiley Online Library on [03/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

and Simulation, *67*(1), 37–73. https://doi.org/10.1080/00949650008812035

Legendre, P., Fortin, M. J., & Borcard, D. (2015). Should the mantel test be used in spatial analysis? *Methods in Ecology and Evolution*, *6*(11), 1239–1247. https://doi.org/10.1111/2041-210X.12425

Lev Maor, G., Yearim, A., & Ast, G. (2015). The alternative role of DNA methylation in splicing regulation. *Trends in Genetics*, *31*(5), 274–280. https://doi.org/10.1016/J.TIG.2015.03.002

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*(21), 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Liu, S., Aagaard, A., Bechsgaard, J., & Bilde, T. (2019). DNA methylation patterns in the social spider, Stegodyphus dumicola. *Genes*, *10*(2), 1–17. https://doi.org/10.3390/genes10020137

Liu, Y., Ma, S., Chang, J., Zhang, T., Chen, X., Liang, Y., & Xia, Q. (2019). Programmable targeted epigenetic editing using CRISPR system in *Bombyx mori*. *Insect Biochemistry and Molecular Biology*, *110*, 105–111. https://doi.org/10.1016/J.IBMB.2019.04.013

Lubin, Y., & Bilde, T. (2007). The evolution of sociality in spiders. *Advances in the Study of Behavior*, *37*, 83–145. https://doi.org/10.1016/S0065-3454(07)37003-4

Lynch, J. B., & Hsiao, E. Y. (2019). Microbiomes as sources of emergent host phenotypes. *Science*, *365*(6460), 1405–1409. https://doi.org/10.1126/science.aay0240

Majer, M., Svenning, J. C., & Bilde, T. (2015). Habitat productivity predicts the global distribution of social spiders. *Frontiers in Ecology and Evolution*, *3*(101), 1–10. https://doi.org/10.3389/fevo.2015.00101

Malmos, K. G., Lüdeking, A. H., Vosegaard, T., Aagaard, A., Bechsgaard, J., Sørensen, J. G., & Bilde, T. (2021). Behavioural and physiological responses to thermal stress in a social spider. *Functional Ecology*, *35*, 2728–2742. https://doi.org/10.1111/1365-2435.13921

Marshall, H., Lonsdale, Z. N., & Mallon, E. B. (2019). Methylation and gene expression differences between reproductive and sterile bumblebee workers. *Evolution Letters*, *3*(5), 485–499. https://doi.org/10.1002/evl3.129

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10. https://doi.org/10.14806/ej.17.1.200

Mazzucco, R., & Schlötterer, C. (2021). Long-term gut microbiome dynamics in *Drosophila melanogaster* reveal environment-specific associations between bacterial taxa at the family level. *Proceedings of the Royal Society B*, *288*(1965), 20212193. https://doi.org/10.1098/RSPB.2021.2193

McFall-Ngai, M., Hadfield, M. G., Bosch, T. C. G., Carey, H. V., Domazet-Lošo, T., Douglas, A. E., Dubilier, N., Eberl, G., Fukami, T., Gilbert, S. F., Hentschel, U., King, N., Kjelleberg, S., Knoll, A. H., Kremer, N., Mazmanian, S. K., Metcalf, J. L., Nealson, K., Pierce, N. E., … Wernegreen, J. J. (2013). Animals in a bacterial world, a new imperative for the life sciences. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(9), 3229–3236. https://doi.org/10.1073/pnas.1218525110

McMurdie, P. J., & Holmes, S. (2013). Phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*, *8*(4), 1–11. https://doi.org/10.1371/journal.pone.0061217

Medeiros, M. C. I., Seabourn, P. S., Rollins, R. L., & Yoneishi, N. M. (2021). Mosquito microbiome diversity varies along a landscape-scale moisture gradient. *Microbial Ecology*, *1*, 1–8. https://doi.org/10.1007/s00248-021-01865-x

Metzger, D. C. H., & Schulte, P. M. (2017). Persistent and plastic effects of temperature on dna methylation across the genome of threespine stickleback (*gasterosteus aculeatus*). *Proceedings of the Royal Society B: Biological Sciences*, *284*(1864), 1–7. https://doi.org/10.1098/rspb.2017.1667

Moran, N. A., Ochman, H., & Hammer, T. J. (2019). Evolutionary and ecological consequences of gut microbial communities. *Annual Review of Ecology, Evolution, and Systematics*, *50*, 451–475. https://doi.org/10.1146/annurev-ecolsys-110617-062453

Morgan, M., Falcon, S., & Gentleman, R. (2022). GSEABase: Gene set enrichment data structures and methods.

Morgan, T. J., Herman, M. A., Johnson, L. C., Olson, B. J. C. S., & Ungerer, M. C. (2018). Ecological genomics: Genes in ecology and ecology in genes, *61*(4), v–vii. https://doi.org/10.1139/Gen-2018-0022

Mueller, E. A., Wisnoski, N. I., Peralta, A. L., & Lennon, J. T. (2020). Microbial rescue effects: How microbiomes can save hosts from extinction. *Functional Ecology*, *34*(10), 2055–2064. https://doi.org/10.1111/1365-2435.13493

Nätt, D., Rubin, C. J., Wright, D., Johnsson, M., Beltéky, J., Andersson, L., & Jensen, P. (2012). Heritable genome-wide variation of gene expression and promoter methylation between wild and domesticated chickens. *BMC Genomics*, *13*(1), 1–12. https://doi.org/10.1186/1471-2164-13-59

Nei, M., & Kumar, S. (2000). *Molecular evolution and phylogenetics (first edit)*. Oxford University Press, USA.

Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., Maldotti, M., Anselmi, F., & Oliviero, S. (2017). Intragenic DNA methylation prevents spurious transcription initiation. *Nature*, *543*(7643), 72–77. https://doi.org/10.1038/nature21373

Ngaira, K. J. W. (2007). Impact of climate change on agriculture in Africa by 2030. *Scientific Research and Essays*, *2*(7), 238–243. https://doi.org/10.5897/SRE.9000567

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., … Wagner, H. (2019). Vegan: Community ecology pack. Retrieved from https://cran.r-project.org/package=vegan

Ørsted, M., Hoffmann, A. A., Sverrisdóttir, E., Nielsen, K. L., & Kristensen, T. N. (2019). Genomic variation predicts adaptive evolutionary responses better than population bottleneck history. *PLoS Genetics*, *15*(6), 1–18. https://doi.org/10.1371/journal.pgen.1008205

Peña, M. V., Piskobulu, V., Murgatroyd, C., & Hager, R. (2021). DNA methylation patterns respond to thermal stress in the viviparous cockroach *Diploptera punctata*. *Epigenetics*, *16*(3), 313–326. https://doi.org/10.1080/15592294.2020.1795603

Perry, B. W., Schield, D. R., & Castoe, T. A. (2018). Evolution: Plasticity versus selection, or plasticity and selection? *Current Biology*, *28*(18), R1104–R1106. https://doi.org/10.1016/J.CUB.2018.07.050

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, *41*(D1), 590–596. https://doi.org/10.1093/nar/gks1219

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

R Core Team. (2019). *R: A language and environment for computing (R Foundation for Statistical Computing, ed.)*. R Foundation for Statistical Computing Retrieved from https://www.r-project.org/

Raufaste, N., & Rousset, F. (2001). Are partial Mantel test adequate? *Evolution*, *55*(8), 1703–1705. https://doi.org/10.1554/0014-3820(2001)055[1703:APMTA]2.0.CO;2

1365294x, 2022, 22, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.16696, Wiley Online Library on [03/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Raza, M. F., Wang, Y., Cai, Z., Bai, S., Yao, Z., Awan, U. A., … Zhang, H. (2020). Gut microbiota promotes host resistance to low-temperature stress by stimulating its arginine and proline metabolism pathway in adult *Bactrocera dorsalis*. *PLoS Pathogens*, *16*(4), e1008441. https://doi.org/10.1371/journal.ppat.1008441

Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., & Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, *24*(17), 4348–4370. https://doi.org/10.1111/mec.13322

Rennison, D. J., Rudman, S. M., & Schluter, D. (2019). Parallel changes in gut microbiome composition and function during colonization, local adaptation and ecological speciation. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1916), 1–9. https://doi.org/10.1098/rspb.2019.1911

Rico, L., Ogaya, R., Barbeta, A., & Peñuelas, J. (2014). Changes in DNA methylation fingerprint of *Quercus ilex* trees in response to experimental field drought simulating projected climate change. *Plant Biology*, *16*(2), 419–427. https://doi.org/10.1111/plb.12049

Riddle, N. C., & Richards, E. J. (2002). The control of natural variation in cytosine methylation in *Arabidopsis*. *Genetics*, *162*(1), 355–363. https://doi.org/10.1093/genetics/162.1.355

Rousselle, M., Simion, P., Tilak, M. K., Figuet, E., Nabholz, B., & Galtier, N. (2020). Is adaptation limited by mutation? A timescale-dependent effect of genetic diversity on the adaptive substitution rate in animals. *PLoS Genetics*, *16*(4), 1–24. https://doi.org/10.1371/journal.pgen.1008668

Rudman, S. M., Greenblum, S., Hughes, R. C., Rajpurohit, S., Kiratli, O., Lowder, D. B., … Schmidt, P. (2019). Microbiome composition shapes rapid genomic adaptation of *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(40), 20025–20032. https://doi.org/10.1073/pnas.1907787116

Sandfeld, T., Malmos, K. G., Nielsen, C. B., Lund, M. B., Aagaard, A., Bechsgaard, J., … Schramm, A. (2022). Metabolite profiling of the social spider *Stegodyphus dumicola* along a climate gradient. *Frontiers in Ecology and Evolution*, *10*, 1–11. https://doi.org/10.3389/fevo.2022.841490

Sarda, S., Zeng, J., Hunt, B. G., & Yi, S. V. (2012). The evolution of invertebrate gene body methylation. *Molecular Biology and Evolution*, *29*(8), 1907–1916. https://doi.org/10.1093/molbev/mss062

Schultz, M. D., Schmitz, R. J., & Ecker, J. R. (2012). 'Leveling' the playing field for analyses of single-base resolution DNA methylomes. *Trends in Genetics*, *28*(12), 583–585. https://doi.org/10.1016/j.tig.2012.10.012

Schwartz, M. K., Luikart, G., & Waples, R. S. (2007). Genetic monitoring as a promising tool for conservation and management. *Trends in Ecology & Evolution*, *22*(1), 25–33. https://doi.org/10.1016/j.tree.2006.08.009

Secco, D., Wang, C., Shou, H., Schultz, M. D., Chiarenza, S., Nussaume, L., … Lister, R. (2015). Stress induced gene expression drives transient DNA methylation changes at adjacent repetitive elements. *eLife*, *4*, 1–26. https://doi.org/10.7554/ELIFE.09343.001

Settepani, V., Bechsgaard, J., & Bilde, T. (2014). Low genetic diversity and strong but shallow population differentiation suggests genetic homogenization by metapopulation dynamics in a social spider. *Journal of Evolutionary Biology*, *27*(12), 2850–2855. https://doi.org/10.1111/jeb.12520

Settepani, V., Bechsgaard, J., & Bilde, T. (2016). Phylogenetic analysis suggests that sociality is associated with reduced effectiveness of selection. *Ecology and Evolution*, *6*(2), 469–477. https://doi.org/10.1002/ece3.1886

Settepani, V., Schou, M. F., Greve, M., Grinsted, L., Bechsgaard, J., & Bilde, T. (2017). Evolution of sociality in spiders leads to depleted genomic diversity at both population and species levels. *Molecular Ecology*, *26*(16), 4197–4210. https://doi.org/10.1111/mec.14196

Sgrò, C. M., Lowe, A. J., & Hoffmann, A. A. (2011). Building evolutionary resilience for conserving biodiversity under climate change. *Evolutionary Applications*, *4*(2), 326–337. https://doi.org/10.1111/j.1752-4571.2010.00157.x

Sgrò, C. M., Terblanche, J. S., & Hoffmann, A. A. (2016). What can plasticity contribute to insect responses to climate change? *Annual Review of Entomology*, *61*, 433–451. https://doi.org/10.1146/annurev-ento-010715-023859

Shigenobu, S., & Wilson, A. C. C. (2011). Genomic revelations of a mutualism: The pea aphid and its obligate bacterial symbiont. *Cellular and Molecular Life Sciences*, *68*(8), 1297–1309. https://doi.org/10.1007/s00018-011-0645-2

Song, Y., Liu, L., Feng, Y., Wei, Y., Yue, X., He, W., … An, L. (2015). Chilling- and freezing- induced alterations in cytosine methylation and its association with the cold tolerance of an alpine Subnival plant, *Chorispora bungeana*. *PLoS One*, *10*(8), e0135485. https://doi.org/10.1371/JOURNAL.PONE.0135485

Sørensen, J. G., & Loeschcke, V. (2002). Natural adaptation to environmental stress via physiological clock-regulation of stress resistance in *drosophila*. *Ecology Letters*, *5*(1), 16–19. https://doi.org/10.1046/J.1461-0248.2002.00296.X

Sutherland, H. G. E., Kearns, M., Morgan, H. D., Headley, A. P., Morris, C., Martin, D. I. K., & Whitelaw, E. (2000). Reactivation of heritably silenced gene expression in mice. *Mammalian Genome*, *11*(5), 347–355. https://doi.org/10.1007/s003350010066

Suzuki, T. A., Martins, F. M., & Nachman, M. W. (2019). Altitudinal variation of the gut microbiota in wild house mice. *Molecular Ecology*, *28*(9), 2378. https://doi.org/10.1111/MEC.14905

Tennekes, M. (2018). Tmap: Thematic maps in R. *Journal of Statistical Software*, *84*(1), 1–39. https://doi.org/10.18637/JSS.V084.I06

Thomas, D. (2010). Gene–environment-wide association studies: Emerging approaches. *Nature Reviews Genetics*, *11*(4), 259–272. https://doi.org/10.1038/nrg2764

Toolson, E. C. (1982). Effects of rearing temperature on cuticle permeability and epicuticular lipid composition in *Drosophila pseudoobscura*. *Journal of Experimental Zoology*, *222*(3), 249–253. https://doi.org/10.1002/JEZ.1402220307

Tregenza, T., Rodríguez-Muñoz, R., Boonekamp, J. J., Hopwood, P. E., Sørensen, J. G., Bechsgaard, J., Settepani, V., Hegde, V., Waldie, C., May, E., Peters, C., Pennington, Z., Leone, P., Munk, E. M., Greenrod, S. T. E., Gosling, J., Coles, H., Gruffydd, R., Capria, L., … Bilde, T. (2021). Evidence for genetic isolation and local adaptation in the field cricket *Gryllus campestris*. *Journal of Evolutionary Biology*, *34*(10), 1624–1636. https://doi.org/10.1111/JEB.13911

Tricker, P. J., Gibbings, J. G., Rodríguez López, C. M., Hadley, P., & Wilkinson, M. J. (2012). Low relative humidity triggers RNA-directed de novo DNA methylation and suppression of genes controlling stomatal development. *Journal of Experimental Botany*, *63*(10), 3799–3813. https://doi.org/10.1093/JXB/ERS076

Ungerer, M. C., Johnson, L. C., & Herman, M. A. (2008). Ecological genomics: Understanding gene and genome function in the natural environment. *Heredity*, *100*(2), 178–183. https://doi.org/10.1038/sj.hdy.6800992

Varriale, A. (2014). DNA methylation, epigenetics, and evolution in vertebrates: Facts and challenges. *International Journal of Evolutionary Biology*, *2014*, 1–7. https://doi.org/10.1155/2014/475981

Verhoeven, K. J. F., Jansen, J. J., van Dijk, P. J., & Biere, A. (2010). Stress-induced DNA methylation changes and their heritability in asexual dandelions. *New Phytologist*, *185*(4), 1108–1118. https://doi.org/10.1111/j.1469-8137.2009.03121.x

Walters, A. W., Hughes, R. C., Call, T. B., Walker, C. J., Wilcox, H., Petersen, S. C., Rudman, S. M., Newell, P. D., Douglas, A. E., Schmidt, P. S., & Chaston, J. M. (2020). The microbiota influences the *Drosophila melanogaster* life history strategy. *Molecular Ecology*, *29*(3), 639–653. https://doi.org/10.1111/MEC.15344

1365294x, 2022, 22, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.16696, Wiley Online Library on [03/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Wang, I. J., & Bradburd, G. S. (2014). Isolation by environment. *Molecular Ecology*, *23*(23), 5649–5662. https://doi.org/10.1111/MEC.12938

Waters, J. L., & Ley, R. E. (2019). The human gut bacteria Christensenellaceae are widespread, heritable, and associated with health. *BMC Biology*, *17*(1), 1–11. https://doi.org/10.1186/s12915-019-0699-4

Wernegreen, J. J. (2012). Mutualism meltdown in insects: Bacteria constrain thermal adaptation. *Current Opinion in Microbiology*, *15*(3), 255–262. https://doi.org/10.1016/j.mib.2012.02.001

Wickham, H., François, R., Henry, L., & Müller, K. (2021). Dplyr: A grammar of data manipulation. Retrieved from https://cran.r-project.org/package=dplyr

Willi, Y., Van Buskirk, J., & Hoffmann, A. A. (2006). Limits to the adaptive potential of small populations. *Annual Review of Ecology, Evolution, and Systematics*, *37*, 433–458. https://doi.org/10.1146/annurev.ecolsys.37.091305.110145

Williams, B. R., Van Heerwaarden, B., Dowling, D. K., & Sgrò, C. M. (2012). A multivariate test of evolutionary constraints for thermal tolerance in *Drosophila melanogaster*. *Journal of Evolutionary Biology*, *25*(7), 1415–1426. https://doi.org/10.1111/j.1420-9101.2012.02536.x

Williamson, S. M., Ingelson-Filpula, W. A., Hadj-Moussa, H., & Storey, K. B. (2021). Epigenetic underpinnings of freeze avoidance in the goldenrod gall moth, *Epiblema scudderiana*. *Journal of Insect Physiology*, *134*, 104298. https://doi.org/10.1016/J.JINSPHYS.2021.104298

Woodhams, D. C., Bletz, M. C., Becker, C. G., Bender, H. A., Buitrago-Rosas, D., Diebboll, H., Huynh, R., Kearns, P. J., Kueneman, J., Kurosawa, E., LaBumbard, B., Lyons, C., McNally, K., Schliep, K., Shankar, N., Tokash-Peters, A. G., Vences, M., & Whetstone, R. (2020). Host-associated microbiomes are predicted by immune system complexity and climate. *Genome Biology*, *21*(1), 1–20. https://doi.org/10.1186/S13059-019-1908-8

Wu, C. T., & Morris, J. R. (2001). Genes, genetics, and epigenetics: A correspondence. *Science*, *293*(5532), 1103–1105. https://doi.org/10.1126/SCIENCE.293.5532.1103

Xu, G., Lyu, H., Yi, Y., Peng, Y., Feng, Q., Song, Q., … Zheng, S. (2021). Intragenic DNA methylation regulates insect gene expression and reproduction through the MBD/Tip60 complex. *IScience*, *24*(2), 102040. https://doi.org/10.1016/J.ISCI.2021.102040

Yin, H., Zhong, Y., Wang, H., Hu, J., Xia, S., Xiao, Y., … Xie, M. (2022). Short-term exposure to high relative humidity increases blood urea and influences colonic urea-nitrogen metabolism by altering the gut microbiota. *Journal of Advanced Research*. https://doi.org/10.1016/J.JARE.2021.03.004

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Aagaard, A., Liu, S., Tregenza, T., Braad Lund, M., Schramm, A., Verhoeven, K. J. F., Bechsgaard, J., & Bilde, T. (2022). Adapting to climate with limited genetic diversity: Nucleotide, DNA methylation and microbiome variation among populations of the social spider *Stegodyphus dumicola*. *Molecular Ecology*, *31*, 5765–5783. https://doi.org/10.1111/mec.16696

1365294x, 2022, 22, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.16696, Wiley Online Library on [03/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

McNew et al. BMC Evolutionary Biology (2017) 17:183
DOI 10.1186/s12862-017-1025-9

BMC Evolutionary Biology

RESEARCH ARTICLE Open Access

Epigenetic variation between urban and rural populations of Darwin's finches

Sabrina M. McNew[1], Daniel Beck[2], Ingrid Sadler-Riggleman[2], Sarah A. Knutie[1], Jennifer A. H. Koop[1], Dale H. Clayton[1] and Michael K. Skinner[2*]

Abstract

Background: The molecular basis of evolutionary change is assumed to be genetic variation. However, growing evidence suggests that epigenetic mechanisms, such as DNA methylation, may also be involved in rapid adaptation to new environments. An important first step in evaluating this hypothesis is to test for the presence of epigenetic variation between natural populations living under different environmental conditions.

Results: In the current study we explored variation between populations of Darwin's finches, which comprise one of the best-studied examples of adaptive radiation. We tested for morphological, genetic, and epigenetic differences between adjacent "urban" and "rural" populations of each of two species of ground finches, *Geospiza fortis* and *G. fuliginosa,* on Santa Cruz Island in the Galápagos. Using data collected from more than 1000 birds, we found significant morphological differences between populations of *G. fortis*, but not *G. fuliginosa*. We did not find large size copy number variation (CNV) genetic differences between populations of either species. However, other genetic variants were not investigated. In contrast, we did find dramatic epigenetic differences between the urban and rural populations of both species, based on DNA methylation analysis. We explored genomic features and gene associations of the differentially DNA methylated regions (DMR), as well as their possible functional significance.

Conclusions: In summary, our study documents local population epigenetic variation within each of two species of Darwin's finches.

Keywords: Epigenetics, *Geospiza*, Copy number variation, Galápagos Islands, DNA methylation

Background

Studies of the molecular basis of evolutionary change have focused almost exclusively on genetic mechanisms. However, recent work suggests that heritable modifications to gene expression and function, independent of changes to DNA sequence, may also be involved in the evolution of phenotypes [1–3]. One of the most common of these epigenetic mechanisms is DNA methylation, i.e. the chemical attachment of methyl groups (CH_3) to nucleotides (usually a cytosine followed by a guanine- "CpG") [4]. Methylation can be induced by the environment and affect gene expression and phenotypic traits without changing the DNA sequence itself [5–8]. Importantly, some patterns of methylation are heritable, meaning they have the potential to evolve [9–14]. Indeed, because DNA methylation modifications (epimutations) are more common than genetic mutations [15], they may play a role in the rapid adaptation of individuals to new or variable environments [16].

Environmentally-induced epimutations may be a component of the adaptive radiation of closely related species to new environments [17]. For example, Skinner et al. [18] showed that epigenetic variation is significantly correlated with phylogenetic distance among five closely related species of Darwin's finches in the Galápagos Islands. Although the adaptive significance of this epigenetic variation is unknown, some of the variants are associated with genes related to beak morphology, cell signaling, and melanogensis. The results of this study suggest that epigenetic changes accumulate over macroevolutionary time and further suggest that epigenetic changes may contribute to the evolution of adaptive phenotypes.

* Correspondence: skinner@wsu.edu
[2]Center for Reproductive Biology, School of Biological Sciences, Washington State University, Pullman, WA 99164-4236, USA
Full list of author information is available at the end of the article

BioMed Central

© The Author(s). 2017 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

Epigenetic variation also occurs among populations within single species [15, 19–24]. Some population epigenetic studies report correlations between methylation patterns and environmental factors, suggesting that differences in methylation are involved in local adaptation to different environments [21, 22, 24]. For example, in a study of populations of two salt marsh specialist plants living along a salinity gradient, Foust et al. [24] found that ground salinity is more closely correlated with epigenetic variation than genetic variation.

The purpose of our study was to investigate epigenetic variation between populations of each of two species of Darwin's finches: the medium ground finch *(Geospiza fortis)* and the small ground finch *(G. fuliginosa)* (Fig. 1). Darwin's finches are a closely related group of about 16 species endemic to the Galápagos Islands [25–28]. Long-term studies show rapid phenotypic changes in populations of finches in response to environmental pressures, including competition [26]. The molecular basis of these phenotypic changes is poorly known. Although recent genomic studies have identified alleles in several putative genes associated with beak size and shape [28–30], most genetic markers show little differentiation among populations or species [28, 30–34].

Epigenetic variation may contribute to the phenotypic diversity of Darwin's finch populations that cannot be detected through genomic studies. As an initial test of this hypothesis, we compared components of the morphology, genetics, and epigentics in populations of finches living at El Garrapatero, a relatively undisturbed locality, to populations living near Puerto Ayora (Academy Bay), the largest town in the Galápagos Islands (Fig. 1). Hereafter, we refer to these as the "rural" and "urban" sites, respectively. The two sites, which are only 10 km apart, are both arid, lowland scrub habitat along the south and south-eastern coast of the island. Vegetative cover, based on remote sensing spectroradiometric indicies, is slightly higher at the urban site; however, cover is highly correlated between the two sites year-round (Additional file 1: Figure S1). Despite the overall ecological similarity of the sites, anthropogenic disturbance at the urban site has increased dramatically over the past fifty years [35]. Observational studies suggest urbanization has effects on finch behavior and diet: birds in the urban population incorporate novel, human foods into their diets, whereas finches in the rural popuation do not [36]. To further explore potential impacts of urbanization of Puerto Ayora on ground finches, we tested for morphological, genetic, and

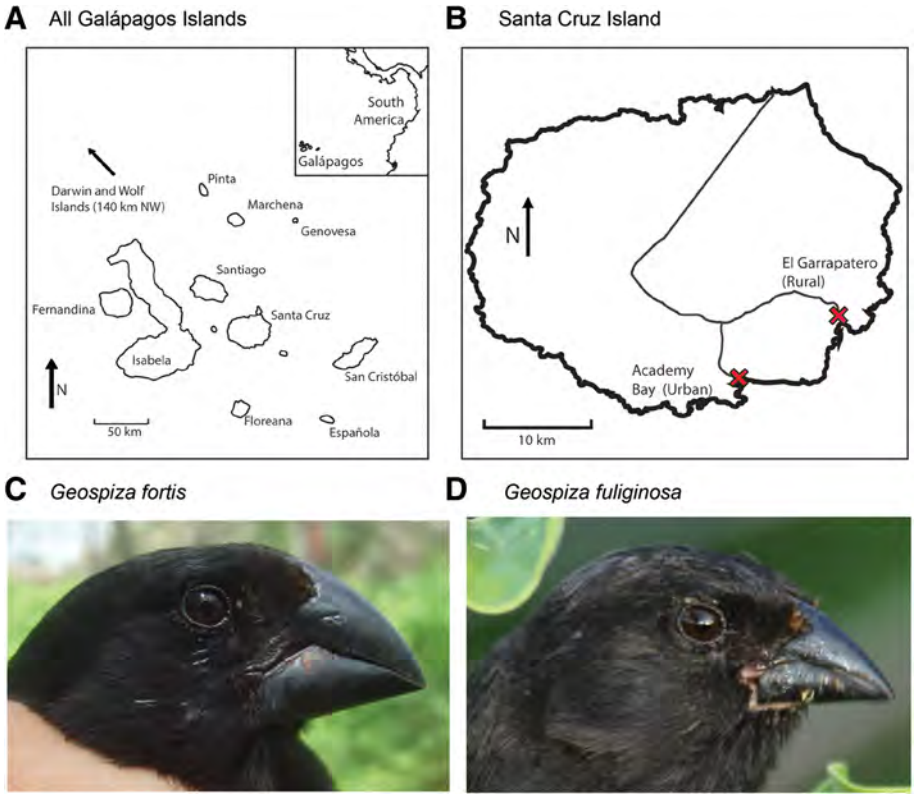


Fig. 1 Study sites and species. **a** The Galápagos Archipelago. **b** Santa Cruz Island; Roads are indicated by *narrow grey lines* and study sites by *red* Xs. **c** *Geospiza fortis*; photo by J.A.H.K. **d** *Geospiza fuliginosa*; photo by S.A.K. Maps in (**a**) and (**b**) are modified from © 2016 Google

epigenetic differences between urban and rural populations in each of two species of finches.

Methods

Study sites and species

We studied each of two populations of *G. fortis* and *G. fuliginosa* living in urban and rural environments (urban: Academy Bay; 0° 44′ 21.3″ S, 90° 18 ‘06.3″ W; rural: El Garrapatero; 0° 41′ 15.7″ S, 90° 13′ 18.3″ W). The two localities, which are separated by about 10 km, are both in the arid coastal zone of Santa Cruz Island (Fig. 1). *Geospiza fortis* and *G. fuliginosa* are among the most abundant species of finches at these study sites. There appears to be little movement of finches between populations. Over the course of a decade-long banding study (2002–2012), during which more than 3700 finches were captured- and more than 300 individuals recaptured- only one bird (a female *G. fortis*) was shown to have moved between the two sites (J. Raeymaekers pers. comm.).

Field work and sample collection

Finches were captured at the two study sites January–April 2008–2016. The birds were mist-netted and banded with individually numbered Monel bands in order to track individuals. They were aged and sexed using size and plumage characteristics [37]. Morphological measurements were taken from each individual including beak depth, beak width, beak length, tarsus length, wing chord, and body mass, following Grant and Grant (2014) [26], with the exception that wing chord was measured unflattened. Principle components were calculated from untransformed data for the three body measurements (mass, wing chord, and tarsus) and for the three beak measurements (length, width, and depth) to provide aggregate measures of body size and beak size and shape [38]. We evaluated morphological differences between urban and rural sites using linear mixed effects models (LMM), with site as a fixed effect, and year as a random effect to control for variation among years and investigators. Separate models were run for each morphological measurement, as well as body size (PC1 body) and beak size and shape (PC1 beak and PC2 beak). *P*-values were adjusted with a Bonferroni correction for multiple tests. Morphological analyses were run in the program RStudio using R version 3.2.1 with the packages pwr, plotrix, lme4, and lmerTest [39–42].

Blood and sperm samples for epigenetic and genetic analyses were collected from a subset of birds captured January–April 2009–2013 at the two study sites. Blood samples (<90 μl) were taken from finches via brachial venipuncture. The samples were stored on wet ice in the field and, within six hours of collection, erythrocytes were purified by centrifugation. Sperm samples (~5 μl) were taken from a subset of males. The sperm samples were obtained by gentle squeezing of the cloacal protuberance of reproductively active males. Blood erythrocytes and sperm samples were stored in a − 20 °C freezer in the Galápagos. Following each field season, they were transferred to a − 80 °C freezer in the USA for long-term storage. All field procedures were approved by the University of Utah Institutional Animal Care and Use Committee (protocols #07–08004, #10–07003 and #13–06010) and by the Galápagos National Park.

Genomic DNA preparation

Genomic DNA from finch red blood cells (erythrocytes) was prepared using the Qiagen DNAeasy Blood and Tissue Kit (Qiagen, Valencia CA). The manufacturer's instructions for nucleated blood samples were followed, but in the final DNA elution step H_2O was used instead of the buffer provided in the kit. Genomic DNA from finch sperm was prepared as follows: collected sperm suspension was adjusted to 100 μl with 1 x Phosphate Buffered Saline (PBS) then 820 μl DNA extraction buffer (50 mM Tris pH 8, 10 mM EDTA pH 8, 0.5% SDS) and 80 μl 0.1 M dithiothreitol (DTT) were added and the sample was incubated at 65 °C for 15 min. Next, 80 μl Proteinase K (20 mg/ml) were added and the sample was incubated on a rotator at 55 °C for 2 h. After incubation, 300 μl of protein precipitation solution (Promega, A795A) were added, then the sample was mixed and incubated on ice for 15 min, then spun at 4 °C at 13,000 rpm for 20 min. The supernatant was transferred to a fresh tube, then precipitated over night with the same volume of 100% isopropanol and 2 μl glycoblue at −20 °C. The sample was then centrifuged and the pellet washed with 75% ethanol, then air-dried and re-suspended in 100 μl H_2O. DNA concentration was measured using a Nanodrop Spectrophotometer (Thermo Fisher).

CNV-Seq protocol

To test for genetic differences between the urban and rural populations we sequenced DNA extracted from red blood cells (erythrocytes) and compared genetic copy number variation (CNV) [18]. CNV, defined as the changes in the number of repeat element copies of more than >1 kb of DNA, is increasingly recognized as one of the most common and functionally important markers of genetic variation [43]. The basic copy number variation (CNV) was determined through genomic sequencing of the same samples used for epigenetic analysis. Read numbers at specific loci were compared genome wide to identify CNV [18]. Erythrocyte DNA pools were generated by combining equal amounts of extracted DNA from five individuals. Each pool contained a total of 2 μg of genomic DNA. Three pools of five individuals each were created per species, per site.

Pooling samples for genomic analysis provides an accurate and cost-effective way of comparing populations [44]. Pooling decreases power, compared to sequencing individual samples. Although minor differences in copy number between populations may be missed [45], large differences between groups should be detected.

The pools were diluted to 130 μl with 1 x TE buffer and sonicated in a Covaris M220 with the manufacturer's preset program to create fragments with a peak at 300 bp. Aliquots of the pools were run on a 1.5% agarose gel to confirm fragmentation. The NEBNext DNA Library Kit for Illumina was used to create libraries for each pool, with each pool receiving a separate index primer. The libraries were sent to the University of Nevada, Reno Genomics Core for NGS on the Illumina HiSeq 2500 using a paired end PE50 application. All 6 pooled sequencing libraries for each species were run in one sequencing lane to generate approximately 30 million reads per pool. The read depth across the genome was then assessed to identify CNV and statistically assessed with a Bayesian analysis. The genome-wide paired end read depth was approximately 2× with the CNV read depth being a total of 300 to 6000 reads per CNV detected.

Methylated DNA Immunoprecipitation (MeDIP)

Following Skinner et al. [18], we used erythrocytes as a purified somatic cell type to compare differentially methylated regions (DMRs) between populations of each of the two species. For a subset of birds, we also compared DMR of germ line cells (sperm). DMRs between urban and rural populations were identified by the methylated DNA immunoprecipitation (MeDIP) of genomic DNA. MeDIP is an enrichment-based technique that uses an antibody to preferentially precipitate methylated regions of the genome that are then sequenced [46]. DMRs are identified by comparing coverage between groups of interest. MeDIP is a cost-effective way to evaluate genomic CpG methylation, and provides highly concordant results to other sequencing-based DNA methylation methods, such as bisulfite sequencing [47]. Because MeDIP surveys methylation genome-wide, it can be used to identify genomic characteristics associated with methylation. For instance, studies have found relationships between CpG density, methylation, and effects on gene transcription [6].

For analysis of erythrocytes, genomic DNA was extracted from the same individuals as used in the CNV pools. Each erythrocyte pool included five individuals and contained a total of 6 μg of genomic DNA. Sperm pools included two individuals and contained a total of 1.8 μg of genomic DNA. Three pools were generated per species, per site (for a total of $n = 6$ individuals per species, per site for sperm and $n = 15$ individuals per species per site for erythrocytes to consider biological variation of the pools and analysis). All pools were diluted to 150 μl with 1× Tris-EDTA (TE, 10 mM Tris, 1 mM EDTA) and sonicated with a probe sonicator using 5 × 20 pulses at 20% amplitude. Fragment size (200–800 bp) was verified on 1.5% agarose gel. Sonicated DNA was diluted to 400 μl with 1xTE and heated to 95 °C for 10 min, then shocked in ice water for 10 min. Next, 100 μl of 5 x immunoprecipitation (IP) buffer (50 mM Sodium Phosphate pH 7, 700 mM NaCl, 0.25% Triton X-100) and 5 μg of 5-mC monoclonal antibody (Diagenode, C15200006–500) were added and the sample was incubated on a rotator at 4 °C over night. The next day Protein A/G Agarose Beads from Santa Cruz Biotechnology, Santa Cruz CA, were prewashed with 1xPBS/0.1% BSA and re-suspended in 1 x IP buffer. Eighty μl of the bead slurry were added to each sample and incubated at 4 °C for 2 h on a rotator. The bead-DNA-antibody complex was washed 3 times with 1 x IP buffer by centrifuging at 6000 rpm for 2 min and resuspending in 1 x IP buffer. After the last wash the bead-complex was re-suspended in 250 μl of digestion buffer (50 mM Tris pH 8, 10 mM EDTA pH 8, 0.5% SDS) with 3.5 μl Proteinase K (20 mg/ml) per sample and incubated on a rotator at 55 °C for 2 h. After incubation, DNA was extracted with the same volume of Phenol-Chloroform-Isoamylcohol, then with the same volume of chloroform. To the supernatant from chloroform extraction, 2 μl glycoblue, 20 μl 5 M sodium chloride and 500 μl 100% cold ethanol were added. DNA was precipitated at −20 °C over night, then spun for 20 min at 13,000 rpm at 4 °C, washed with 75% ethanol, and air-dried. The dry pellet was re-suspended in 20 μl H_2O and concentration measured in Qubit using a Qubit ssDNA Assay Kit (Life Technologies, Carlsbad, CA).

MeDIP-Seq protocol

The next step for DMR identification involved sequencing the MeDIP DNA to identify differential methylation at specific genomic loci by assessing read numbers for the different samples. The MeDIP pools were used to create sequencing libraries for next generation sequencing (NGS) at the University of Nevada, Reno Genomics Core Laboratory using the NEBNext® Ultra™ RNA Library Prep Kit for Illumina®, starting at step 1.4 of the manufacturer's protocol to generate double stranded DNA. After this step the manufacturer's protocol was followed. Each pool received a separate index primer. NGS was performed at the same laboratory using the Illumina HiSeq 2500 with a paired end PE50 application, with a read size of approximately 50 bp and approximately 100 million reads per pool. Two separate sequencing libraries, one rural and one urban, were run in each lane. The read depth for identified differential DNA methylated regions (DMRs) ranged from approximately 100 to >1000 total reads per DMR.

Bioinformatics

Basic read quality was verified using summaries produced by the FastQC program [48]. The reads for each sample for both CNV and DMR analyses were mapped to the zebra finch (*Taenopygia guttata*) genome using Bowtie2 [49] with default parameter options. The mapped read files were converted to sorted BAM files using SAMtools [50]. The cn.MOPS R package [51] was used to identify potential CNV. The cn.mops default information gain thresholds were used for this analysis. The cn.MOPS analysis detects CNVs by modeling read depth across all samples. The model predicts copy number for a given window based on observed read counts. The model uses a Bayesian framework to determine whether copy number for a give window differs significantly from 2. The length of the CNV is determined by comparing copy number of adjacent windows on the genome and joining those with the same copy number into one segment. A CNV call occurs when copy number for a given genomic segment varies from that of other samples. CNV detection with cn.MOPS is robust to low-coverage sequencing data (0.18–0.46 for 75 bp reads) and performs well when comparing 6 or more samples [51]. The window size used by the cn.MOPS analysis was chosen dynamically for each chromosome based on the read coverage. The chromosomes' window size ranged from approximately 25 kb to 60 kb. Only CNV that occurred in either all urban or all rural pools were compared. Although some individual pools had higher numbers of CNV, only CNV that occur red among all the pools were included in the analysis. The CNV are identified using the difference between the posterior and prior distributions from the Bayesian analysis to estimate information gain.

To identify DMR, the reference genome was broken into 100 bp windows. The MEDIPS R package [52] was used to calculate differential coverage between the urban and rural localities. The edgeR *p*-value [53] was used to determine the relative difference between the two localities for each genomic window. Windows with an edgeR *p*-value less than 10^{-3} were considered DMR. The DMR edges were extended until no genomic window with an edgeR *p*-value less than 0.1 remained within 1000 bp of the DMR. The DMR that included at least two windows with an edgeR *p*-value $<10^{-3}$ ("multiple-window DMR") were then selected for further analysis. Because no fully assembled or annotated genome exists for any Darwin's finch species, we aligned DMR with the zebra finch genome. CpG density and gene associations were then calculated for the DMR, based on alignment with the reference genome. Though we previously found high (>98%) homology between Darwin's finch and zebra finch genomes using tiling arrays [18], some differences were expected. Thus, associations of DMR with genes are likely to be under-estimates. To validate the epigenetics and gene associations, a similar analysis was also done with the draft *G. fortis* genome [54]. All the DMR sequence and genomic data obtained in the current study have been deposited in the NCBI public GEO database (GEO # GSE87825).

DMR clusters were identified with an in-house R script (www.skinner.wsu.edu under genomic data) using a 2 Mb sliding window with 50 kb intervals. DMR were annotated using the biomaRt R package [55] to access the Ensembl database [56]. The genes that overlapped with DMR were then input into the KEGG pathway search [57, 58] to identify associated pathways. A 10 kb flanking sequence was added to each DMR to consider potential localization in promoter regions of the gene as previously described [18, 59]. The DMR associated genes were manually sorted into gene classification groups by consulting information provided by the DAVID, Panther, and Uniprot databases incorporated into an internal curated database (www.skinner.wsu.edu under

Table 1 Mean (± 1SE) values for morphological characteristics of *G. fortis* and *G. fuliginosa* at rural vs. urban sites.

	G. fortis		*G. fuliginosa*	
Morphological	Rural	Urban	Rural	Urban
Character	*N* = 560	*N* = 245	*N* = 171	*N* = 121
Beak depth	11.48 ± 0.06	11.98 ± 0.09**	7.40 ± 0.04	7.42 ± 0.06
Beak width	9.89 ± 0.04	10.24 ± 0.07**	6.8 ± 0.03	6.82 ± 0.04
Beak length	11.71 ± 0.04	12.02 ± 0.07***	8.56 ± 0.04	8.46 ± 0.09
Tarsus length	21.00 ± 0.06	21.15 ± 0.09	18.83 ± 0.11	18.67 ± 0.09
Wing chord	69.3 ± 0.19	70.4 ± 0.29**	61.26 ± 0.31	61.1 ± 0.30
Body mass	21.23 ± 0.13	22.2 ± 0.23*	13.87 ± 0.15	13.76 ± 0.14
PC1 Body	−0.13 ± 0.06	0.29 ± 0.09***	0.07 ± 0.09	−0.10 ± 0.10
PC1 Beak	−0.17 ± 0.07	0.40 ± 0.11***	0.01 ± 0.09	−0.01 ± 0.15
PC2 Beak	−0.01 ± 0.02	0.02 ± 0.03	0.07 ± 0.04	−0.09 ± 0.10

Statistically significant differences between populations at $P < 0.01$, 0.001, and <0.0001 are indicated by *, ** and ***, respectively

Table 2 Differentially methylated regions (DMR) between urban and rural populations based on different cell types

Species/Cell Type	Number of Windows*					Sum of Multiple (≥2) Window DMR**
	1	2	3	4	5	
G. fortis Erythrocytes	2742	125	4	0	0	129
G. fortis Sperm	1160	97	9	3	1	110
G. fuliginosa Erythrocytes	4339	314	9	1	0	324
G. fuliginosa Sperm	1765	133	6	0	0	139

Only DMR that were significant at $P < 0.001$ are included
*DMR detected in one window alone were considered "single-window" variants (Fig. 2)
**DMR detected in two or more adjacent windows were considered "multiple-window" variants and used in subsequent analyses (Figs. 2, 3, 4, 5 and 6)

genomic data). To assess that the DMR were not false positives due to random biological variation within populations, a pairwise comparison analysis (individual pool comparison) on the genomic sequence data within the individual urban or rural sites and cell populations was performed [60].

Results

Morphology

We used 1097 birds captured between 2008 and 2016 for morphological analyses. We controlled for slight variation among years in traits by including year as a random effect in all analyses. At both sites, *G. fortis* was significantly larger than *G. fuliginosa* in all morphological traits (linear mixed-effects models $P < 0.0001$). Within species, urban *G. fortis* was significantly larger than rural *G. fortis* for all direct morphological measurements, except tarsus length (Table 1). Composite measures of *G. fortis* body and beak size (PC1 body and PC1 beak) also differed between the two sites; however, there was no difference in beak shape (PC2 beak). In contrast to *G. fortis*, *G. fuliginosa* did not differ significantly between the urban and rural populations in any of the morphological measurements or composite measures (Table 1). Because we captured more *G. fortis* than *G. fuliginosa* we did a power analysis for *G. fuliginosa*, using the effect size of morphological differences found in the *G. fortis* populations (smallest effect size = 0.256 (wing chord); largest effect size = 0.358 (beak depth)). Power for comparisons of *G. fuliginosa* appeared adequate for detecting similar effect sizes (0.69–0.91).

Copy number variation (CNV)

Mean read depth genome-wide for pools used in CNV analysis varied between 1.08× and 1.30× (overall mean = 1.22×). The total read depth of individual variants ranged from 300 to 6000. We identified unique CNV in three of six *G. fortis* pools and five of six *G. fuliginosa* pools. The total number of variants per pool ranged from 1 to 20. However, no variants were exclusive to all urban or all rural pools for either *G. fortis* or *G. fuliginosa* (Additional file 2: Fig. S2).

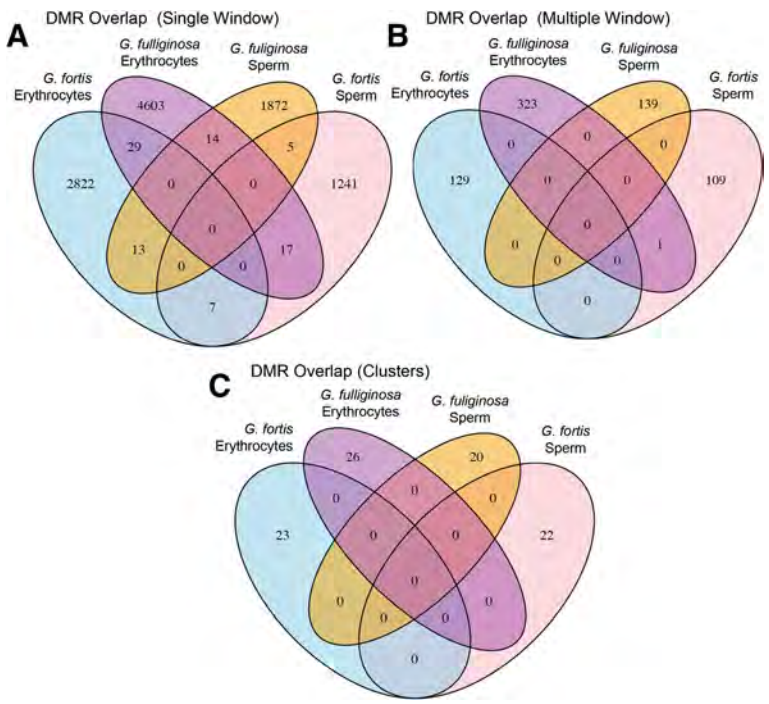


Fig. 2 DMR overlap between species and cell types. Each value is the number of differentially methylated regions between the urban and rural populations. Overlapping colors in the figure show the number of DMR that are shared between the two species or the two cell types. DMRs detected within **a** single 100 bp windows, **b** 2–5 adjacent 100 bp windows, **c** 2–7 Mb regions

Therefore, while there was variation within populations in copy number at various loci in both *G. fortis* and *G. fuliginosa* (e.g., FB2 & 12), there were no fixed differences between the urban and rural populations for either species. It is unclear why certain pools had more variants than others; however variation was consistent among chromosomes.

To control for underestimation of CNV differences due to reads that did not align to the zebra finch genome, we performed a similar analysis aligning reads to the un-assembled *Geospiza fortis* genome [54]. The average proportion of reads aligned to the *G. fortis* genome was higher (two-fold). However, we still did not find any differences in CNV between the urban and rural populations for either species of Darwin's finch. A limitation of this CNV analysis is that only large variants (>24 Kbp) can be detected reliably; smaller variants (<10 Kbp or less) may have escaped detection.

Differential DNA methylation regions (DMRs)

DMRs were found between populations for both cell types and both species (Table 2). We report the number of DMRs at *p*-value cut-offs ranging from <0.01 to <1e-05 in Additional file 3: Table S1; Additional file 4: Table S2 and Additional file 5: Table S3. The analyses reported below are restricted to DMRs significant at a level of $P < 0.001$. We evaluated differences on three "regional" scales (Fig. 2): single 100 bp window DMRs, multiple window DMRs, and "DMR clusters", i.e. statistically over-represented DMR clusters of 3–10 DMRs spanning 2–7 Mb [18] (Additional file 6: Table S4A-D). We focus on multiple-window DMRs (Additional file 4: Table S2 and Additional file 5: Table S3), i.e. DMRs detected independently in adjacent windows, because they further reduce the likelihood of false positives and provide a set of highly reproducible DMRs [18]. Multiple-window DMRs were used in the analysis of the genomic features of DMRs reported below.

There was little overlap between species or cell types in the regions that were differentially methylated between urban and rural populations (Fig. 2). A small proportion of single window DMRs (Fig. 2A) was shared between species and/or cell types. However, there were virtually no shared multiple-window DMRs (Fig. 2B) or clusters of DMRs (Fig. 2C) between species and/or cell types.

For both species and cell types, multiple-window DMRs usually were detected in only two multiple 100 bp windows; however, a limited number (<10% of total DMRs) were found in 3–5 multiple windows (Table 2). Based on extension of edges of multiple-window DMRs (extension

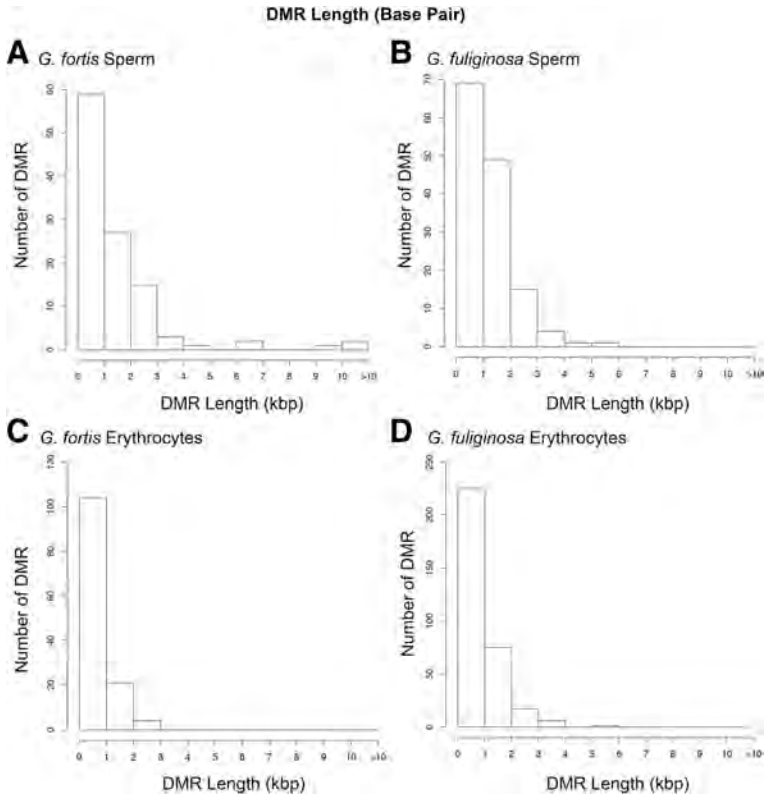


Fig. 3 DMR length (kb) in **a** *G. fortis* sperm. **b** *G. fuliginosa* sperm. **c** *G. fortis* erythrocytes. **d** *G. fuliginosa* erythrocytes. Only multiple-window DMR significant at a *p*-value threshold of $<10^{-3}$ are included

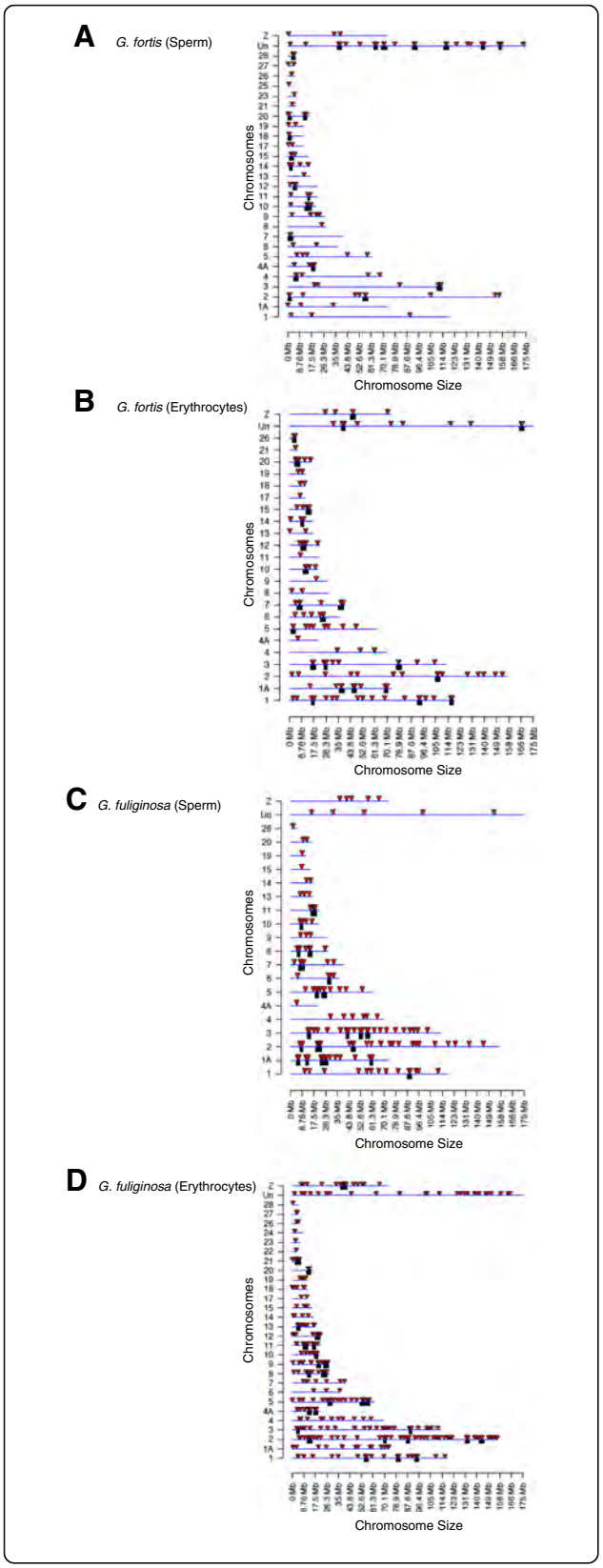


Fig. 4 Chromosomal locations of DMR identified in *Geospiza fortis* sperm **a** and erythrocytes (**b**) and *G. fuliginosa* sperm (**c**) and erythrocytes (**d**). Locations are based on alignment to the zebra finch (*Taeniopygia guttata*) genome. *Red arrowheads* indicate DMR and *black boxes* indicate DMR clusters. Only multiple-window DMR significant at a *p*-value threshold of $<10^{-3}$ are shown

of adjacent 100 bp windows with $p < 0.1$; see Methods) we estimated that most DMRs were 500–1000 bp in length (Fig. 3). Many of the DMRs in this study were clustered together, consistent with previous studies showing that DMRs are not evenly distributed across the genome [59]. Based on alignment to the zebra finch genome, we plotted the chromosomal locations of multiple-window DMRs and DMR clusters (Fig. 4). DMRs were present on all chromosomes in both sperm and erythrocytes of both species; however, the chromosomal locations of DMRs differed between the cell types and species.

We evaluated the location of DMRs with respect to nucleotide composition. CpG density was highest in DMRs of *G. fortis* sperm cells (Fig. 5A). DMRs in *G. fortis* erythrocytes and both cell types of *G. fuliginosa* were most often found in lower density CpG regions of the genome (<1 CpG site/100 bp; Fig. 5B-D). We estimated that the DMRs typically had approximately 10 CpG sites clustered within 1 kb regions.

We identified potential genes associated with DMRs through alignment with the zebra finch reference genome. DMRs within 10 kb of a gene (such that the promoter is included) have the potential to influence the gene's expression and/or pathways associated with that gene [59]. Different categories of genes were methylated in the two cell types and species (Fig. 6, specific genes listed in Additional file 7: Table S5). The most common gene categories associated with DMRs were metabolism, cell signaling and transcription (Fig. 6). Gene categories associated with DMRs differed significantly between the two species (Chi-square test, $p = 0.039$) and marginally between the two cell types (Chi-square test; $p = 0.078$). Pathway analysis (KEGG) showed DMRs associated with several genes (GALNT14, SGMS1, ENO2, PLCH2) in metabolic pathways of *G. fortis* sperm. DMRs were associated with different genes (GCLC, PRIM2, ALD1A3, AK4, ACACA) in metabolic pathways of *G. fuliginosa* sperm. *Geospiza fortis* erythrocyte DMRs were associated with genes (CACNA1H, FGF8, MRAS, RAP1A) in the MAPK signaling pathway. *Geospiza fuliginosa* erythrocyte DMRs were not associated with any particular pathway.

When the DMR data sets for both species and cell types were compared, KEGG pathways with the most DMR-associated genes were metabolic pathways, and MAPK and TGFß/BMP signaling pathways. Metabolic pathways included glycolysis, in which genes involved with pyruvate and acetate production were associated with

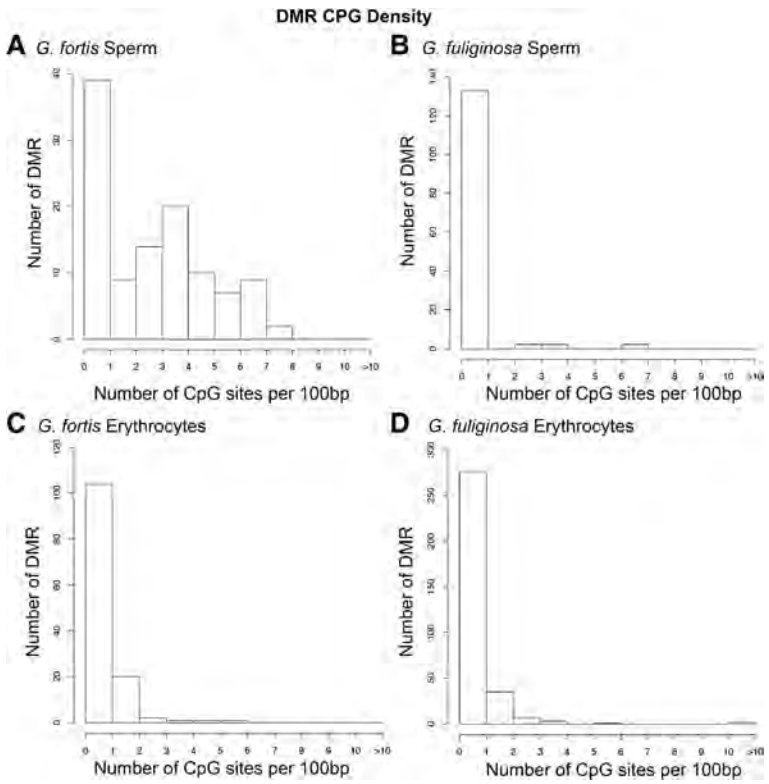


Fig. 5 The CpG density of DMR in *Geospiza fortis* sperm (**a**), *G. fuliginosa* sperm (**b**), *G. fortis* erythrocytes (**c**) and *G. fuliginosa* erythrocytes (**d**). Only multiple-window DMR significant at a *p*-value threshold of $<10^{-3}$ are included

DMRs in both finch species (Additional file 8: Figure S3 and Additional file 9: Figure S4). Other metabolic pathways associated with DMRs included genes involved in purine metabolism and glycosylation (Additional file 7: Tables S5). Signaling pathways were also associated with DMRs in both species and cell types. Three genes in the TGFß/BMP pathway were associated with DMRs between *G. fuliginosa* populations (erythrocytes and sperm combined): BMP5, BMP7 and FST (Fig. 7). MAPK, a common pathway for many regulatory processes, such as cell growth, contained a high number of DMR-associated genes in both finch species (Additional file 8: Figure S3 and Additional file 9: Figure S4).

Genomic correlates of our DMR and CNV data were analyzed using the well-annotated zebra finch genome. In addition, our sequencing data were also compared to the *G. fortis* shotgun sequence database [54]. In contrast to the zebra finch genome, the *G. fortis* genome is neither assembled, nor annotated, meaning that limited data analysis is possible. The pooled individual sample read number was approximately 100 million reads for both genome analyses. The overall read alignment rate was 47–48% for the zebra finch analysis and 70–75% for the *G. fortis* genome analysis. Although previous analysis using tiling arrays suggested a 98% similarity in tiling array hybridization of the genome [18], the next generation sequencing analysis shows that more differences exist, likely in non-coding regions. The zebra finch genome analysis revealed twice the number of DMRs compared to the *G. fortis* genome analysis. This was largely due to the incomplete nature of the *G. fortis* genome. Nevertheless, analysis with both the zebra finch and *G. fortis* genomes identified epigenetic alterations between the rural and urban sites. To test whether methylation variation between sites was greater than within sites we conducted a pairwise comparison analysis (comparison of individual pools) within each species and rural or urban populations for specific cell types. We identified a number of DMRs between individual pools, which suggests that there is epigenetic variation within the study populations. However, few DMRs were found in multiple pools from the same population. Moreover, almost none of these DMRs were also found between urban and rural populations (Additional file 10: Figure S5). Thus, the DMRs identified between urban and rural populations are not an artifact of sampling within-population variation.

Discussion

Darwin's finches are well known for their phenotypic variability and evolution in response to changing environmental conditions [26]. In addition to genetic variation,

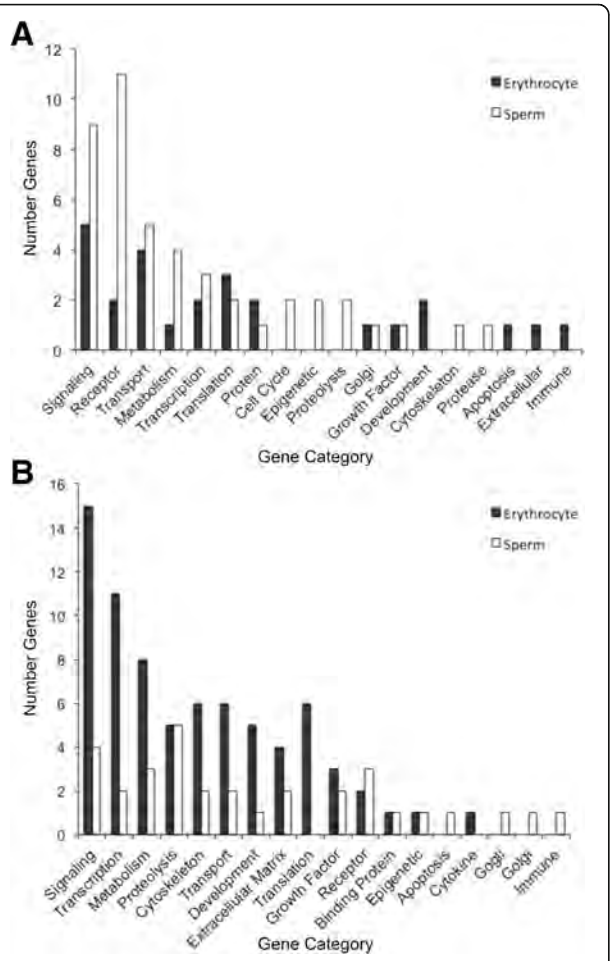


Fig. 6 Gene categories associated with DMR detected in (**a**) *G. fortis* and (**b**) *G. fuliginosa*. Only multiple-window DMR significant at a *p*-value threshold of $<10^{-3}$ are included

epigenetic variation - such as differential DNA methylation - may exist between natural populations living under different environmental conditions. The goal of this paper was to test for morphological, genetic, and epigenetic differences between urban and rural populations within each of two species of Darwin's finches. We found that *G. fortis* individuals at the urban site (Academy Bay) were larger than those at the rural site (El Garrapatero). In contrast, *G. fuliginosa* individuals did not differ morphologically between the sites. We did not find genetic differentiation between populations of either species based on CNV comparisons. However, we did find epigenetic (DMR) differences between urban and rural populations of both species of finches.

We found urban *G. fortis* were larger in nearly all morphological measurements compared to rural *G. fortis* (Table 1), which may be due to increased food availability at the urban site. Previous work suggests that urbanization around Academy Bay has relaxed selection on finch beak size [35, 36]. Urbanization is associated with a shift in the distribution of beak size in *G. fortis*: beak size is strongly bimodal at the rural site, whereas bimodality has decreased at the urban site concurrently with human population growth [35]. Both studies propose that increased food availability at the urban site has altered the selective landscape for *G. fortis* [35, 36]. Beak size is highly heritable in *Geospiza* finches; e.g. mid-parent vs. mid-offspring values estimate heritability of beak depth in *G. fortis* to be 0.74 [61].

In contrast, *G. fuliginosa* showed no morphological differentiation between sites (Table 1). *Geospiza fortis* is phenotypically more variable than *G. fuliginosa* on Santa Cruz Island [61]. As a result, *G. fortis* may have undergone more rapid local adaptation than *G. fuliginosa*. Although *G. fuliginosa* and *G. fortis* have overlapping dietary niches, they do show some degree of specialization [27]. It is possible that urbanization has had a greater selective effect on *G. fortis* than *G. fuliginosa*. Alternatively, morphological differences in *G. fortis* may be driven by hybridization between *G. fortis* and the slightly larger *G. magnirostris*. Hybridization between *G. fortis* and *G. magnirostris* has been documented on Santa Cruz [62]. While we have no information on relative rates of hybridization at our study sites, *G. magnirostris* is more abundant at the urban site than the rural site (4.56% of urban birds captured, compared to 1.86% of rural birds captured; unpublished data 2008–2016).

Despite differences in morphology between populations of *G. fortis*, we found no genetic differences between the urban and rural populations, based on the CNV comparisons made. Because CNV sequence coverage was limited, we may have overlooked small CNV, but larger CNV should have been detected between the two populations. CNV is a sensitive index of genetic differentiation between populations; indeed, some studies have found that CNV accounts for more genetic variation than SNPs [63–65]. Recent work has also linked CNV to rapid evolution in pepper moths [66] and primates [67].

Our study is first to explore genetic variation between populations of Darwin's finches using large–scale genomic features (CNV). Like our study, previous studies using smaller-scale genomic markers (microsatellites, nuclear introns, and mitochondrial DNA) detected little or no genetic structure within populations of either *G. fortis* or *G. fuliginosa* [31, 34, 68]. Two recent studies of genomic variation among Darwin's finches using SNPs did identify variable sites associated with variation in beak morphology [29, 30]. However, most of the genes associated with beak morphology in the two studies were different. These inconsistent results suggest that other forms of variation, such as large scale CNVs, could underlie phenotypic differences. However, our

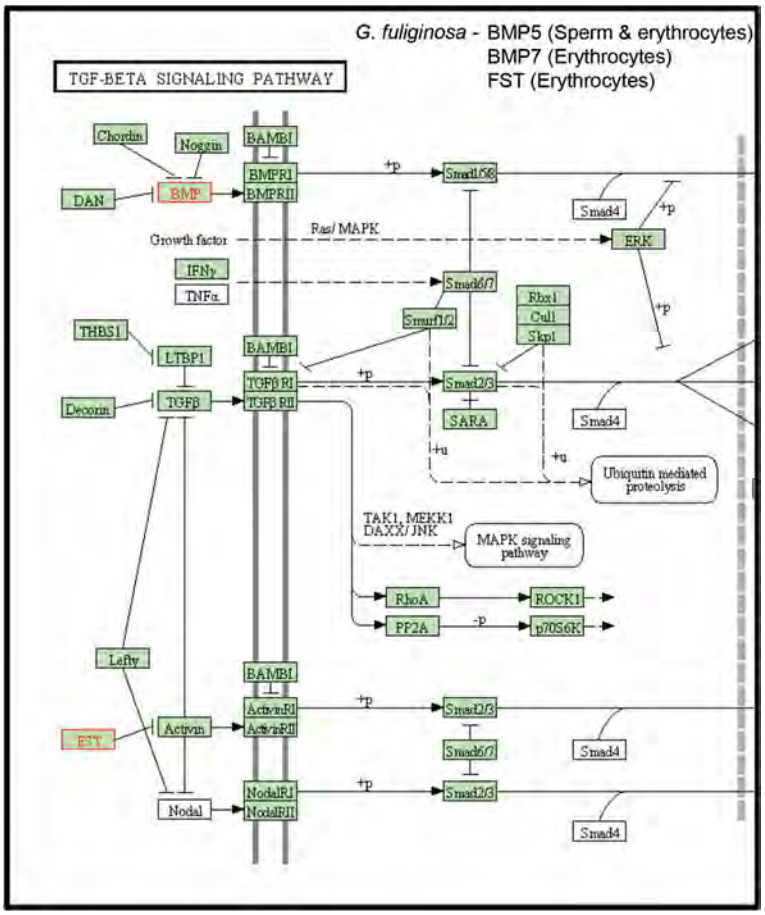


Fig. 7 TGFB/BMP pathway. Genes associated with DMR are listed and outlined in *red* in the pathway

results show that negligible large size CNV changes exist between the rural and urban populations of *G. fortis* or *G. fuliginosa*.

In contrast to our genetic results, we found a large number of epigenetic differences between urban and rural populations in both species of finches and both cell types (Fig. 2). Although DMRs were found in both species, few of the same genomic regions were differentially methylated in *G. fortis* and *G. fuliginosa*. These data suggest that methylation patterns are species-specific, even when comparing closely related species. This may mean that *G. fortis* and *G. fuliginosa* are responding to environmental changes at the urban site in different ways. The lack of overlap in DMRs between the two species may reflect differences in their diets [27]. As discussed above, dietary differences may also have contributed to the morphological differences between urban and rural populations of *G. fortis*.

Although DMRs were also found in both cell types, few of the same genomic regions were differentially methylated in sperm and erythrocytes. Because methylation is involved with cell differentiation [6, 69], some lack of similarity in erythrocyte and sperm DMR is expected. The differences between the genomic regions that were differentially methylated in sperm and erythrocytes may provide clues as to the functional significance of the epimutations. DMRs in somatic cells, such as erythrocytes, potentially reflect effects of the environment on physiology of the birds. DMRs in germ cells, such as sperm, are more likely to be transgenerationally inherited and contribute to evolution. Recent studies show that heritability of methylation variants can be high, but that this varies among loci [12]. However, without following multiple generations of individuals with known ancestry, we cannot determine which of the DMRs in our study are heritable. It is possible that many of the DMRs we detected were plastic responses to the environment. Analysis of Darwin's finches with known pedigrees - from long-term studies of banded populations - may be a way in which to distinguish heritable from non-heritable epimutations in the future.

While locations of DMRs varied between species and cell types (Fig. 4), they had genomic features in common. DMRs were typically 500–1000 bp in length (Fig. 3) and

many were clustered in 2–7 Mb regions. Most DMRs were in areas of low CpG density known as "CpG deserts" (Fig. 5). Many studies of DNA methylation have focused on the gene-silencing effects of methylation in high-density "CpG islands" near transcriptional start sites [6]. However, DMRs in other genomic regions, such as CpG deserts, can have other important effects on gene regulation and expression [6, 70]. Methylation of cytosines increases the rate of cytosine to thymine transitions [71]. Thus, over time, methylation can cause CpG-poor regions in the genome to accumulate. The persistence of conserved clusters of methylated CpG sites within CpG deserts suggests that these regions are likely conserved and under purifying selection [70]. Thus, these types of DMRs may have a functional role in regulating gene expression and could be subject to selection.

Many of the DMRs we detected were associated with metabolic and signaling genes (Fig. 6). Previous work has suggested that novel food sources at the urban site are changing the diet of finches [27]. While we did not quantify phenotypic traits related to metabolism, it is possible that DMRs associated with metabolic genes are associated with other physiological differences between the urban and rural populations.

We also found DMRs associated with genes in the bone morphogenic protein (BMP) / transferring growth factor beta (TGFß) pathway (Fig. 7). Expression of Bmp4 is related to beak shape in *Geospiza* finches [72]; however, it is unknown what factors regulate gene expression at this locus. We previously found that this pathway was differentially methylated among species of Darwin's finches [18]. These data suggest that DNA methylation may play a role in regulating expression of genes in this pathway and therefore may influence finch morphology.

Our study compared just two populations - one rural and one urban – and therefore we cannot be certain that urbanization is the key environmental change influencing finch morphology and/or epigenetics in our study. Moreover, it is possible that differences between the two populations are the result of epigenetic drift, rather than differential selection. Some dispersal of *G. fortis* between the urban and rural populations has been documented through mark-recapture studies; but it is not very common (J. Raeymaekers pers. comm.). Low levels of gene flow between populations would preclude divergence of the rural and urban populations due to drift. However, much more work is needed to understand the basis of epigenetic variation and its relationship to phenotypic variation in populations of Darwin's finches.

Conclusions

We found epigenetic differences between adjacent populations of each of two species of Darwin's finches. We do not know which of the DMRs are responses to environmental differences between the urban and rural sites, versus the result of random epigenetic drift. However, as the environmental differences between our sites are recent (<60 years) any methylation changes associated with urbanization have spread quickly. As in other recent studies [19, 20, 22], the functional relationship between environmental and epigenetic variation is not well understood. Nevertheless, these results are consistent with a potential role of epigenetic variation in rapid adaptation to changing environments. Future studies are needed to determine what effects DMR have on phenotypes, and to what extent these methylation patterns may play a role in evolution.

Additional files

Additional file 1: Figure S1. Comparison of vegetative cover at the rural site (El Garrapatero) versus urban site (Puerto Ayora, Academy Bay) over the course of the study. Cover was dervied from Normalized Difference Vegetative Index (NDVI) values generated from satellite imagery (ORNL DAAC. 2008. MODIS Collection 5 Land Products Global Subsetting and Visualization Tool. ORNL DAAC, Oak Ridge, Tennessee, USA. Accessed May 08, 2017 http://dx.doi.org/10.3334/ORNLDAAC/1241). Values range from 0-1 with 1 reprensenting the highest vegetation cover. (PDF 850 kb)

Additional file 2: Figure S2. Copy number variation (CNV) between the rural and urban populations. (A) CNV analysis summary for the *G. fortis* erythrocytes showing read depth and alignment, and CNV numbers per pool with chromosomes containing CNV indicated, and no overlap between rural and urban pools indicated. (B) CNV analysis summary for the *G. fuliginosa* erythrocytes with Read Mapping Summary, overall CNV per pool and chromosome, and no overlapping CNV identified. (PDF 20 kb)

Additional file 3: Table S1. The number of DMR detected at single window and multiple window scales at increasing levels of significance. (PDF 61 kb)

Additional file 4: Table S2. Description of multiple-window DMR detected in *G. fortis* sperm **(A)** and erythrocytes **(B)**. Description includes DMR name, chromosome number, DMR start site, length in base pair (bp), number of multiple sites, minimum p-value, CpG number per sequence length, CpG density (CpG number / 100 bp) and DMR gene association. "NA" indicates DMR associated with a gene that did not align to the zebra finch reference genome. (PDF 126 kb)

Additional file 5: Table S3. Description of multiple-window DMR detected in *G. fuliginosa* sperm **(A)** and erythrocytes **(B)**. Description includes DMR name, chromosome number, DMR start site, length in base pair (bp), number of multiple sites, minimum p-value, CpG number per sequence length, CpG density (CpG number / 100 bp) and DMR gene association. "NA" indicates DMR associated with a gene that did not align to the zebra finch reference genome. (PDF 154 kb)

Additional file 6: Table S4. Description of DMR clusters detected in *G. fortis* sperm **(A)** and erythrocytes **(B)** and *G. fuliginosa* sperm **(C)** and erythrocytes **(D)**. Description includes DMR in cluster, chromosome number, cluster start site, cluster stop site, length in bp, and minimum *p*-value. (PDF 103 kb)

Additional file 7: Table S5. Gene associations with DMR detected in *G. fortis* sperm **(A)** and erythrocytes **(B)** and *G. fuliginosa* sperm **(C)** and erythrocytes **(D)**. Description includes DMR name, gene symbol, entrez gene identification, chromosome number, start position site, ensemble gene identification number, gene description and gene classification category. (PDF 225 kb)

Additional file 8: Figure S3. MAPK signaling pathway. Genes associated with DMR are listed and outlined in red in the pathway. (PDF 109 kb)

Additional file 9: Figure S4. Glycolysis metabolism pathway. Genes associated with DMR are listed and outlined in red in the pathway. (PDF 66 kb)

Additional file 10: Figure S5. DMRs identified in pairwise comparison of pools within populations: **(A)** *G. Fuliginosa* RBC urban analysis, **(B)** *G. fuliginosa*-RBC rural analysis, **(C)** *G. fortis* RBC urban analysis, and **(D)** *G. fortis* rural analysis. Numbers indicate DMRs between urban (U) or rural (R) individual pools (1-3). "Full analysis" are DMRs identified between urban and rural pools. DMRs identified in the full analysis were found independently of within-site variation. (PDF 98 kb)

Abbreviations

BMP: bone morphogenic protein; CNV: copy number variation; DDT: dithiothreitol; DMR: differentially DNA methylated region; IP: immunoprecipitation; LMM: linear mixed effects models; NGS: next generation sequencing; PBS: Phosphate Buffered Saline; TGFß: transferring growth factor beta

Acknowledgments

We acknowledge the advice and critical review of Dr. Eric Nilsson (WSU). The following people contributed to sample collection: Céline Le Bohec, Sarah Bush, Roger Clayton, Miriam Clayton, Oliver Tiselma, Elena Arriero, Andrew Bartlow, Daniela Vargas, Emily DiBlasi, Jordan Herman, Pricilia Espina, Kiyoko Gotanda, Sofia Carvajal, Joost Raeymakers and Janaí Yepez. We thank Ms. Jayleana Barton for molecular technical assistance and Ms. Heather Johnson for assistance in preparation of the manuscript. The research was supported by a Templeton grant to MKS, National Science Foundation grants DEB-0816877 and DEB-1342600 to DHC, and an NSF Graduate Research Fellowship to SMM. Dr. Sarah A. Knutie's present address: Department of Ecology and Evolutionary Biology, Storrs, CT 06269-3043, USA. Dr. Jennifer A. H. Koop's present address: Biology Department, University of Massachusetts-Dartmouth, Dartmouth, MA 02747-2300, USA.

Funding

The research was supported by a Templeton grant to MKS, a National Science Foundation grant to DHC, and an NSF Graduate Research Fellowship to SMM.

Availability of data and materials

All the DMR sequence and genomic data obtained in the current study have been deposited in the NCBI public GEO database (GEO # GSE87825).

Author contributions

DHC and MKS designed the study; SMM, SAK and JAHK collected the samples, DB and ISR analyzed the genomic data, SMM, DHC and MKS analyzed the data and wrote the manuscript with help from the other authors. All authors read and approved the final manuscript.

Ethics approval and consent to participate

All field procedures were approved by the University of Utah Institutional Animal Care and Use Committee (protocols #07–08004, #10–07003 and #13–06010) and by the Galápagos National Park.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

[1]Department of Biology, University of Utah, Salt Lake City, UT 84112-0840, USA. [2]Center for Reproductive Biology, School of Biological Sciences, Washington State University, Pullman, WA 99164-4236, USA.

Received: 26 January 2017 Accepted: 26 July 2017
Published online: 24 August 2017

References

1. Bossdorf O, Richards CL, Pigliucci M. Epigenetics for ecologists. Ecol Lett. 2008;11(2):106–15.
2. Day T, Bonduriansky R. A unified approach to the evolutionary consequences of genetic and nongenetic inheritance. Am Nat. 2011;178(2):E18–36.
3. Robertson M, Richards C. Non-genetic inheritance in evolutionary theory - the importance of plant studies. Non-Genetic Inherit. 2015;2:3–11.
4. Angers B, Castonguay E, Massicotte R. Environmentally induced phenotypes and DNA methylation: how to deal with unpredictable conditions until the next generation and after. Mol Ecol. 2010;19(7):1283–95.
5. Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet. 2003; 33(Suppl):245–54.
6. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet. 2012;13(7):484–92.
7. Duncan EJ, Gluckman PD, Dearden PK. Epigenetics, plasticity, and evolution: how do we link epigenetic change to phenotype? J Exp Zool B Mol Dev Evol. 2014;322(4):208–20.
8. Jirtle RL, Skinner MK. Environmental epigenomics and disease susceptibility. Nat Rev Genet. 2007;8(4):253–62.
9. Crews D, Gore AC, Hsu TS, Dangleben NL, Spinetta M, Schallert T, Anway MD, Skinner MK. Transgenerational epigenetic imprints on mate preference. Proc Natl Acad Sci U S A. 2007;104(14):5942–6.
10. Richards CL, Bossdorf O, Pigliucci M. What role does heritable epigenetic variation play in phenotypic evolution? Bioscience. 2010;60:232–7.
11. Latzel V, Zhang Y, Karlsson Moritz K, Fischer M, Bossdorf O. Epigenetic variation in plant responses to defence hormones. Ann Bot. 2012;110(7):1423–8.
12. Janowitz Koch I, Clark MM, Thompson MJ, Deere-Machemer KA, Wang J, Duarte L, Gnanadesikan GE, McCoy EL, Rubbi L, Stahler DR, et al. The concerted impact of domestication and transposon insertions on methylation patterns between dogs and grey wolves. Mol Ecol. 2016;25(8):1838–55.
13. Verhoeven KJ, vonHoldt BM, Sork VL. Epigenetics in ecology and evolution: what we know and what we need to know. Mol Ecol. 2016;25(8):1631–8.
14. Skinner MK, Manikkam M, Guerrero-Bosagna C. Epigenetic transgenerational actions of environmental factors in disease etiology. Trends Endocrinol Metab. 2010;21(4):214–22.
15. Richards CL, Schrey AW, Pigliucci M. Invasion of diverse habitats by few Japanese knotweed genotypes is correlated with epigenetic differentiation. Ecol Lett. 2012;15(9):1016–25.
16. Liu QA. The impact of climate change on plant epigenomes. Trends Genet. 2013;29(9):503–5.
17. Flatscher R, Frajman B, Schonswetter P, Paun O. Environmental heterogeneity and phenotypic divergence: can heritable epigenetic variation aid speciation? Genet Res Int. 2012;2012:698421.
18. Skinner MK, Guerrero-Bosagna C, Haque MM, Nilsson EE, Koop JAH, Knutie SA, Clayton DH. Epigenetics and the evolution of Darwin's Finches. Genome Biology & Evolution. 2014;6(8):1972–89.
19. Lira-Medeiros CF, Parisod C, Fernandes RA, Mata CS, Cardoso MA, Ferreira PC. Epigenetic variation in mangrove plants occurring in contrasting natural environment. PLoS One. 2010;5(4):e10326.
20. Liu S, Sun K, Jiang T, Ho JP, Liu B, Feng J. Natural epigenetic variation in the female great roundleaf bat (Hipposideros Armiger) populations. Mol Gen Genomics. 2012;287(8):643–50.
21. Gugger PF, Fitz-Gibbon S, PellEgrini M, Sork VL. Species-wide patterns of DNA methylation variation in Quercus Lobata and their association with climate gradients. Mol Ecol. 2016;25(8):1665–80.
22. Lea AJ, Altmann J, Alberts SC, Tung J. Resource base influences genome-wide DNA methylation levels in wild baboons (Papio Cynocephalus). Mol Ecol. 2016;25(8):1681–96.
23. Zhao Y, Tang JW, Yang Z, Cao YB, Ren JL, Ben-Abu Y, Li K, Chen XQ, Du JZ, Nevo E. Adaptive methylation regulation of p53 pathway in sympatric speciation of blind mole rats, Spalax. Proc Natl Acad Sci U S A. 2016;113(8):2146–51.
24. Foust CM, Preite V, Schrey AW, Alvarez M, Robertson MH, Verhoeven KJ, Richards CL. Genetic and epigenetic differences associated with environmental gradients in replicate populations of two salt marsh perennials. Mol Ecol. 2016;25(8):1639–52.

25. Podos J. Correlated evolution of morphology and vocal signal structure in Darwin's finches. Nature. 2001;409(6817):185–8.
26. Grant PR, Grant BR. 40 years of evolution. Princeton, NJ: Princeton University Press; 2014.
27. De Leon LF, Podos J, Gardezi T, Herrel A, Hendry AP. Darwin's finches and their diet niches: the sympatric coexistence of imperfect generalists. J Evol Biol. 2014;27(6):1093–104.
28. Lamichhaney S, Berglund J, Almen MS, Maqbool K, Grabherr M, Martinez-Barrio A, Promerova M, Rubin CJ, Wang C, Zamani N, et al. Evolution of Darwin's finches and their beaks revealed by genome sequencing. Nature. 2015;518(7539):371–5.
29. Lamichhaney S, Han F, Berglund J, Wang C, Almen MS, Webster MT, Grant BR, Grant PR, Andersson L. A beak size locus in Darwin's finches facilitated character displacement during a drought. Science. 2016;352(6284):470–4.
30. Chaves JA, Cooper EA, Hendry AP, Podos J, De Leon LF, Raeymaekers JA, MacMillan WO, Uy JA. Genomic variation at the tips of the adaptive radiation of Darwin's finches. Mol Ecol. 2016;25(21):5282–95.
31. de Leon LF, Bermingham E, Podos J, Hendry AP. Divergence with gene flow as facilitated by ecological differences: within-island variation in Darwin's finches. Philos Trans R Soc Lond Ser B Biol Sci. 2010;365(1543):1041–52.
32. Sato A, O'HUigin C, Figueroa F, Grant PR, Grant BR, Tichy H, Klein J. Phylogeny of Darwin's finches as revealed by mtDNA sequences. Proc Natl Acad Sci U S A. 1999;96(9):5101–6.
33. Petren K, Grant B, Grant P. A phylogeny of Darwin's finches based on microsatellite DNA length variation. Proc R Soc Lond B. 1999;266:321–9.
34. Farrington HL, Lawson LP, Clark CM, Petren K. The evolutionary history of Darwin's finches: speciation, gene flow, and introgression in a fragmented landscape. Evolution. 2014;68(10):2932–44.
35. Hendry AP, Grant PR, Rosemary Grant B, Ford HA, Brewer MJ, Podos J. Possible human impacts on adaptive radiation: beak size bimodality in Darwin's finches. Proc Biol Sciences / R Soc. 2006;273(1596):1887–94.
36. De Leon LF, Raeymaekers JA, Bermingham E, Podos J, Herrel A, Hendry AP. Exploring possible human influences on the evolution of Darwin's finches. Evolution. 2011;65(8):2258–72.
37. Grant P. Ecology and evolution of Darwin's finches. Princeton, NJ: Princeton University Press; 1986.
38. Grant PR, Grant BR. Evolution of character displacement in Darwin's finches. Science. 2006;313(5784):224–6.
39. Kuznetsova A, Brockhoff P, Christensen R: lmerTest: Tests in Linear Mixed Effects Models. R package version 2.0–33. 2016.
40. Bates D, Maechler M, Bolker B, Walker S. lme4: linear mixed-effects models using Eigen and S4. J Stat Softw. 2015;67:1–48.
41. Lemon J. Plotrix: a package in the red light district of R. R-News. 2006;6:8–12.
42. Champely S: pwr: Basic functions for power analysis. . R package version 12–1 2017.
43. Freeman JL, Perry GH, Feuk L, Redon R, McCarroll SA, Altshuler DM, Aburatani H, Jones KW, Tyler-Smith C, Hurles ME, et al. Copy number variation: new insights in genome diversity. Genome Res. 2006;16(8):949–61.
44. Schlotterer C, Tobler R, Kofler R, Nolte V. Sequencing pools of individuals - mining genome-wide polymorphism data without big funding. Nat Rev Genet. 2014;15(11):749–63.
45. Zhang W, Carriquiry A, Nettleton D, Dekkers JC. Pooling mRNA in microarray experiments and its effect on power. Bioinformatics. 2007;23(10):1217–24.
46. Taiwo O, Wilson GA, Morris T, Seisenberger S, Reik W, Pearce D, Beck S, Butcher LM. Methylome analysis using MeDIP-seq with low DNA concentrations. Nat Protoc. 2012;7(4):617–36.
47. Harris RA, Wang T, Coarfa C, Nagarajan RP, Hong C, Downey SL, Johnson BE, Fouse SD, Delaney A, Zhao Y, et al. Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. Nat Biotechnol. 2010;28(10):1097–105.
48. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.
49. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. Nat Methods. 2012;9(4):357–9.
50. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. Genome project data processing S: the sequence alignment/map format and SAMtools. Bioinformatics. 2009;25(16):2078–9.
51. Klambauer G, Schwarzbauer K, Mayr A, Clevert DA, Mitterecker A, Bodenhofer U, Hochreiter S. cn.MOPS: mixture of Poissons for discovering copy number variations in next-generation sequencing data with a low false discovery rate. Nucleic Acids Res. 2012;40(9):e69.
52. Lienhard M, Grimm C, Morkel M, Herwig R, Chavez L. MEDIPS: genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments. Bioinformatics. 2014;30(2):284–6.
53. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139–40.
54. Li B, Li H, Parker P, Wang J. The genome of Darwin's finch (Geospiza Fortis). GigaDB. 2012;
55. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/bioconductor package biomaRt. Nat Protoc. 2009;4(8):1184–91.
56. Cunningham F, Amode MR, Barrell D, Beal K, Billis K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fitzgerald S, et al. Ensembl 2015. Nucleic Acids Res. 2015;43(Database issue):D662–9.
57. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27–30.
58. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. 2014;42(Database issue):D199–205.
59. Skinner MK, Manikkam M, Haque MM, Zhang B, Savenkova M. Epigenetic Transgenerational inheritance of somatic Transcriptomes and epigenetic control regions. Genome Biol. 2012;13(10):R91.
60. Shnorhavorian M, Schwartz SM, Stansfeld B, Sadler-Riggleman I, Beck D, Skinner MK. Differential DNA Methylation Regions in Adult Human Sperm Following Adolescent Chemotherapy: Potential for Epigenetic Inheritance. PloS One. 2017;12(2):e0170085.
61. Grant P, Grant R. How and why species multiply: the radiation of Darwin's finches Princeton. NJ: Princeton University Press; 2008.
62. Huber SK, De Leon LF, Hendry AP, Bermingham E, Podos J. Reproductive isolation of sympatric morphs in a population of Darwin's finches. Proceedings Biological sciences / The Royal Society. 2007;274(1619):1709–14.
63. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, et al. Global variation in copy number in the human genome. Nature. 2006;444(7118):444–54.
64. Beckmann JS, Estivill X, Antonarakis SE. Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. Nat Rev Genet. 2007;8(8):639–46.
65. McCarroll SA, Kuruvilla FG, Korn JM, Cawley S, Nemesh J, Wysoker A, Shapero MH, de Bakker PI, Maller JB, Kirby A, et al. Integrated detection and population-genetic analysis of SNPs and copy number variation. Nat Genet. 2008;40(10):1166–74.
66. Van't Hof AE, Campagne P, Rigden DJ, Yung CJ, Lingley J, Quail MA, Hall N, Darby AC, Saccheri IJ. The industrial melanism mutation in British peppered moths is a transposable element. Nature. 2016;534(7605):102–5.
67. Niu AL, Wang YQ, Zhang H, Liao CH, Wang JK, Zhang R, Che J, Su B. Rapid evolution and copy number variation of primate RHOXF2, an X-linked homeobox gene involved in male reproduction and possibly brain function. BMC Evol Biol. 2011;11:298.
68. Petren K, Grant PR, Grant BR, Keller LF. Comparative landscape genetics and the adaptive radiation of Darwin's finches: the role of peripheral isolation. Mol Ecol. 2005;14(10):2943–57.
69. Bestor TH. The DNA methyltransferases of mammals. Hum Mol Genet. 2000;9(16):2395–402.
70. Skinner MK, Guerrero-Bosagna C. Role of CpG deserts in the epigenetic Transgenerational inheritance of differential DNA methylation regions. BMC Genomics. 2014;15(1):692.
71. Cooper DN, Youssoufian H. The CpG dinucleotide and human genetic disease. Hum Genet. 1988;78(2):151–5.
72. Abzhanov A, Protas M, Grant BR, Grant PR, Tabin CJ. Bmp4 and morphological variation of beaks in Darwin's finches. Science. 2004;305(5689):1462–5.