Ecological divergence of wild strawberry DNA methylation patterns at distinct spatial scales

Group Members: Colin Schuller, Mina Shumaly, Sarah Phelps, and Supantha Dey

Differentially methylated cytosines in Fragaria vesca

De Kort, Hanne, KU Leuven, https://orcid.org/0000-0003-2516-0134

hanne.dekort@kuleuven.be

Published Jan 13, 2021 on Dryad. https://doi.org/10.5061/dryad.zs7h44j6r

Cite this dataset

De Kort, Hanne (2021). Differentially methylated cytosines in Fragaria vesca [Dataset]. Dryad. https://doi.org/10.5061/dryad.zs7h44j6r

Abstract

Epigenetic change is considered relatively unstable and short-lived, questioning its contribution to long-term adaptive potential. However, epigenetic modifications can accumulate in the presence of environmental stress, resulting in beneficial epigenetic memories where environments are challenging. Diverging epigenetic memories have been observed across large spatial scales, and can persist through multiple generations even in the absence of the causative environmental stressor. It is unknown, however, to what extent epigenetic variation contributes to fine-scale population structure and evolution. We compared DNA methylation patterns between a steep, altitudinal gradient (<2 km) and a wide spatial gradient (>500 km) using whole genome bisulfite sequencing data from 30 *Fragaria vesca* plants germinated and grown in controlled conditions. To assess the stability of spatial epigenetic variation in the presence of an environmental stressor, we applied acute drought stress to part of the plants and quantified drought-induced changes in

Data files

Jan 13, 2021 version files 8.95 MB

Download full dataset

Related works

Primary article

https://doi.org/10.1111/mec.15689

Share

Metrics

105 views

Received: 13 January 2020 | Revised: 1 October 2020 | Accepted: 2 October 2020

DOI: 10.1111/mec.15689

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Ecological divergence of wild strawberry DNA methylation patterns at distinct spatial scales

Hanne De Kort[1] | Bart Panis[2] | Dieter Deforce[3] | Filip Van Nieuwerburgh[3] | Ollivier Honnay[1]

[1]Plant Conservation and Population Biology, University of Leuven, Leuven, Belgium

[2]Bioversity International, K.U. Leuven, Leuven, Belgium

[3]Laboratory of Pharmaceutical Biotechnology, Ghent University, Ghent, Belgium

Correspondence
Hanne De Kort, Plant Conservation and Population Biology, University of Leuven, Leuven, Belgium.
Email: hanne.dekort@kuleuven.be

Funding information
Fonds Wetenschappelijk Onderzoek, Grant/Award Number: 12P6517N

Abstract

Epigenetic change is considered relatively unstable and short-lived, raising questions of its contribution to long-term adaptive potential. However, epigenetic modifications can accumulate in the presence of environmental stress, resulting in beneficial epigenetic memories where environments are challenging. Diverging epigenetic memories have been observed across large spatial scales, and can persist through multiple generations. It is unknown, however, to what extent epigenetic variation contributes to fine-scale population structure and evolution. We compared DNA methylation patterns between a steep, altitudinal gradient (<2 km) and a wide spatial gradient (>500 km) using whole genome bisulphite sequencing data from 30 *Fragaria vesca* plants germinated and grown in controlled conditions. To assess the stability of

Background on Biological Question

- Previous data showed that strawberry plants adapted to stressful high-altitude conditions grew slower and had fewer flowers than their counterparts adapted to lower altitudes. Illustrating adaptive divergence across altitudinal gradients especially since genetic diversity of this species is fairly limited across its range.

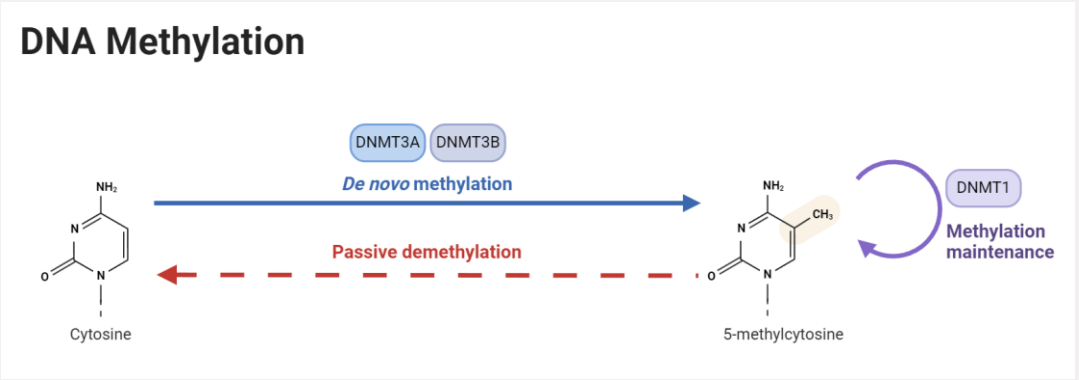
They hypothesize that due to the limited genetic diversity of this species, in combination with the pronounced altitude phenotype divergence, there are substantial adaptive epigenetic modifications along increased levels of stress.

1) Do epigenetic memories diverge with altitude, and does DNA methylation change with altitude?

2) Are these fine-scale genome wide methylation patterns comparable across a large geographical range

3) Are altitudinal DMCs enriched for and specific GO terms?

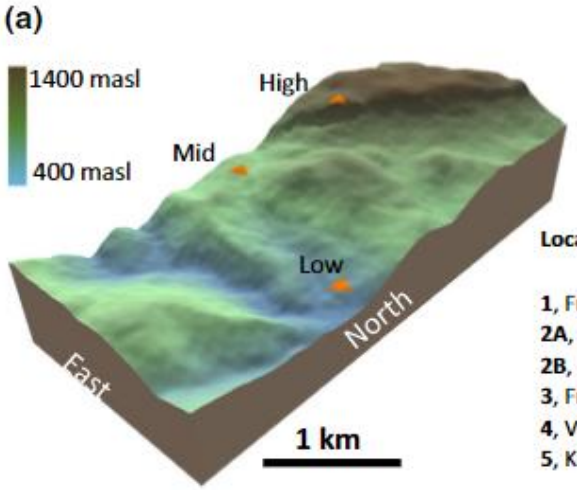
4) Does acute drought induce any detectable changes?



Sample Collection

- Seeds were collected from five plants at (a) three nearby locations in the French Pyrenees, (b) one location in the French Vosges and (c) one location in Poland.
- One seedling per mother plant was randomly selected from every location -> growth chamber
- To compare the magnitude of inherited epigenetic memories to intragenerational epigenetic change acquired through acute drought stress:

Second seedling per mother plant was raised and was subjected to reduced soil moisture levels.



(b)

Location (N)	Altitude	Space	Latitude	Longitude
1, French Pyrenees (5)	High		42.93007	1.151337
2A, French Pyrenees (5)	Mid(Dry)		42.927851	1.165021
2B, French Pyrenees (5)	Mid(Moist)		42.927851	1.165021
3, French Pyrenees (5)	Low	West	42.942821	1.170258
4, Vosges (eastern France) (5)		Central	48.197752	5.884375
5, Karkonosze (Poland) (4)		East	50.883056	15.559722

Material and Methods

- Whole genome bisulphite sequencing and DMC (differentially methylated cytosines) calling (Patterns across conditions or samples)
- Methylation profiling of altitude, space and drought
- DMC enrichment analysis (Process or feature overrepresented in DMC?)

Chr_Position	Type	Context	strand	pvalue	qvalue	meth.diff	|meth.dif
Fvb1_00655	Spatial	CHG	F	6.31E-06	0.0045	25.92593	25.92593
Fvb1_04977	Spatial	CG	F	5.81E-05	0.013121	72.72727	72.72727
Fvb1_10010587	Spatial	CG	F	3.39E-05	0.008824	54.97475	54.97475
Fvb1_10010703	Spatial	CG	R	5.06E-06	0.001874	69.14045	69.14045
Fvb1_10021815	Spatial	CG	R	0.000374	0.046431	89.6779	89.6779
Fvb1_10021941	Spatial	CG	R	0.000235	0.034295	44.6586	44.6586
Fvb1_10036672	Spatial	CG	F	7.73E-05	0.016057	92.78543	92.78543
Fvb1_10037813	Spatial	CG	F	0.00016	0.026552	83.81262	83.81262

...

...

Final Goal:

- Studying the ecological divergence in DNA methylation patterns of wild strawberries.
- How these patterns change in response to various environmental gradients (~ Adaptation and Evolution).

Raw data (30 final samples)

We regenerated all the plots!

Documentation of Figure 2a

DATA MANIPULATION

Create 3 data frames that correspond to the 3 sampling types

```
alt_df = strawberry[strawberry['Type'] == 'Altitudinal']
```

Create 3 subcategories for each context

```
alt_gc = alt_df[alt_df['Context'] == 'CG']  # all 3 contexts
```

Covert the subcategory counts for each major category to log10

```
altitide_counts = 492, 134, 73
```

```
log_altitude = [np.log10(n) for n in altitide_counts]
```

PLOT BAR CHART

Set categories, context, and values

Create an index to group Contexts within the same Type together on the x-axis

```
index = np.arange(len(categories))
```

Set context colors to match original figure

Loop to plot everything

```
for i, category in enumerate(context): plt.bar (index
+ i * bar_width, values[:, i], bar_width,
label=category,  color=context_colors[category])
```

Define axes and chart labels

Figure 2a

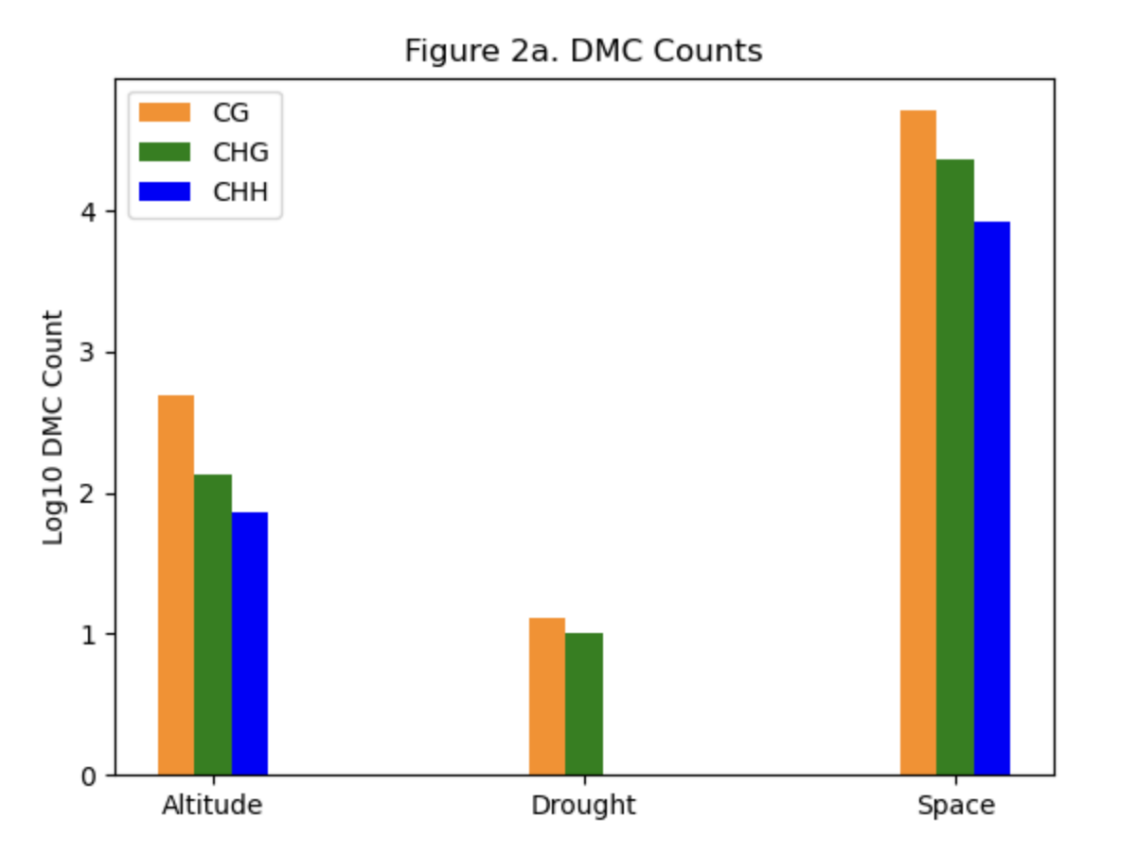


Figure generated from data

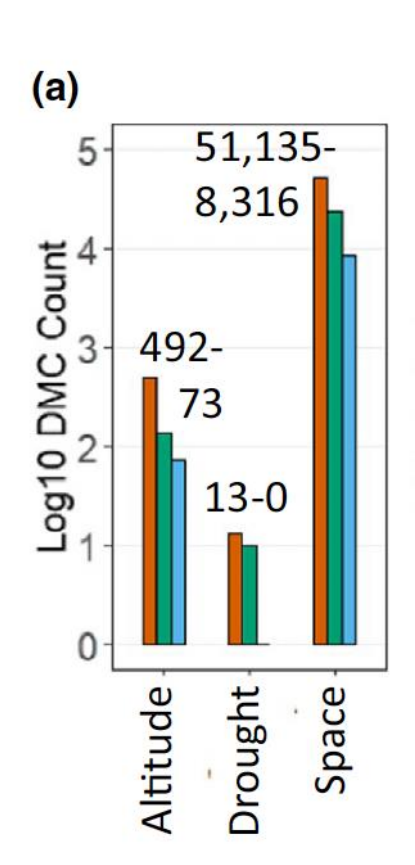


Figure from original publication

Documentation of Figure 2b

Set hue and order to match the original figure

```
hue_order = ['CG', 'CHG', 'CHH']

order = ['Altitudinal', 'Drought', 'Spatial'] # order the type variables to match the original figure
```

Plot the box plot

```
sns.catplot(x='Type', y='absolute.meth', hue='Context', kind='box', data=strawberry, order=order,

        hue_order=hue_order, height=5, aspect=1, palette= context_colors, legend_out=False)
```

Change plot labels to match original figure

Figure 2b

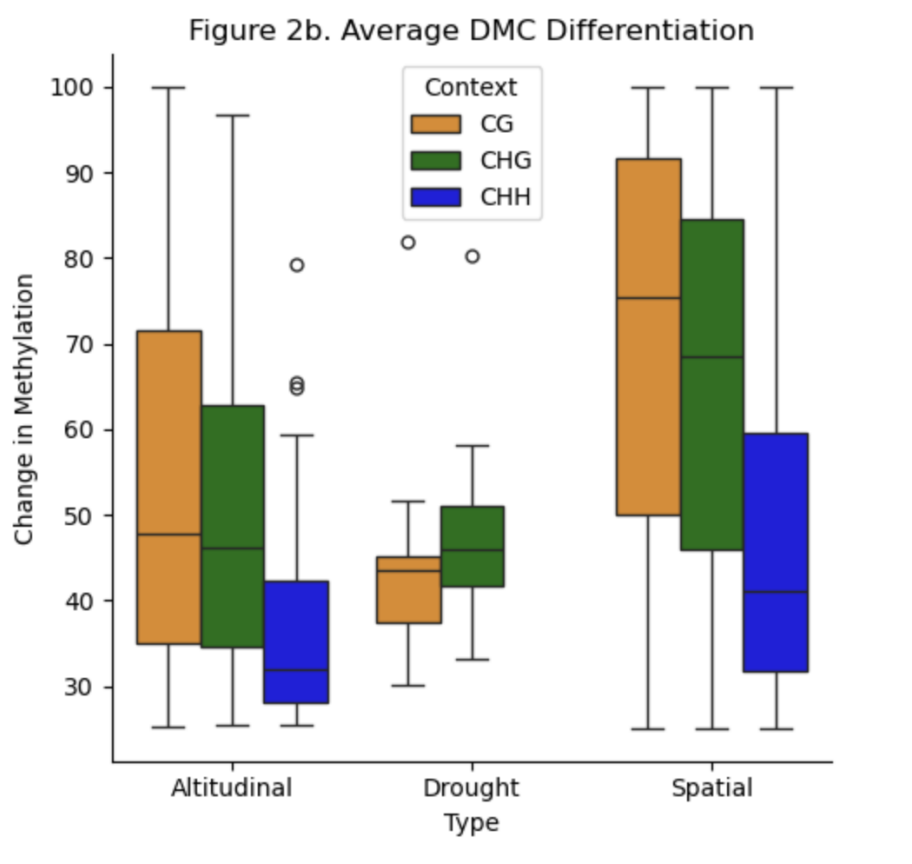


Figure generated from data

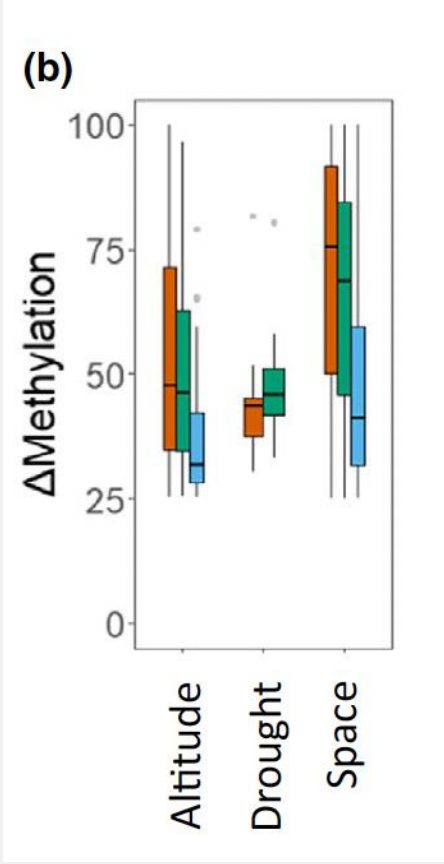
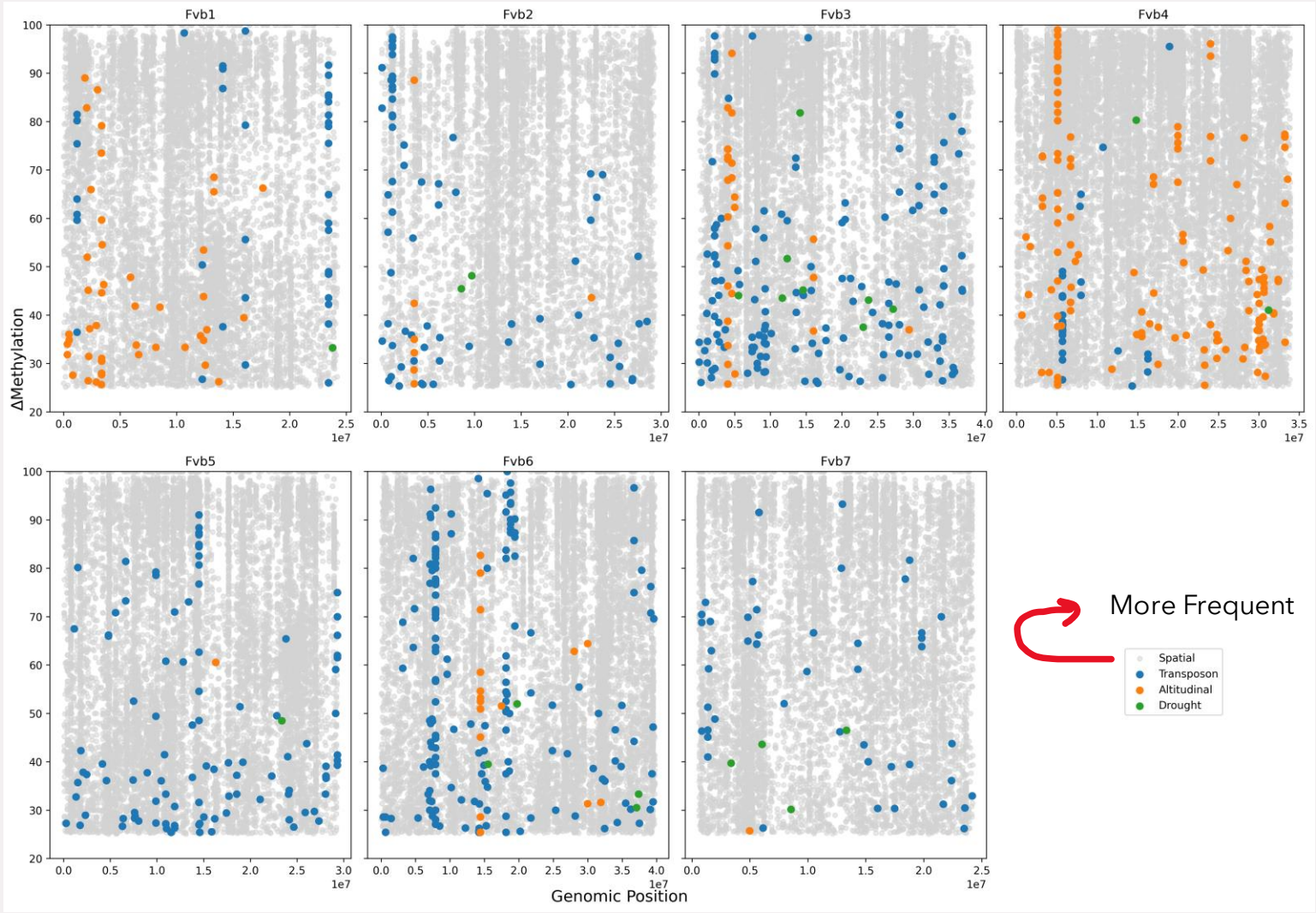


Figure from original publication

Figure 3

Genome-wide differentiation in methylation for the DMCs in all sequence contexts



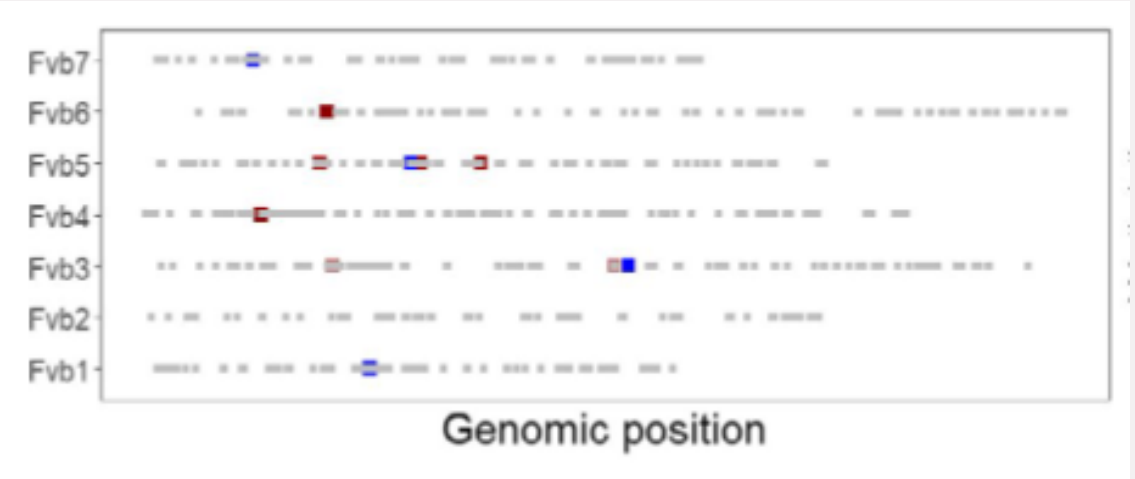
Chr_Position	|meth.diff|	Type
Fvb1_006554	25.92592593	Spatial
Fvb1_10730084	33.33333333	Altitudinal
Fvb3_12498700	54.16666667	Spatial
Fvb4_3121472	41.05263158	Drought

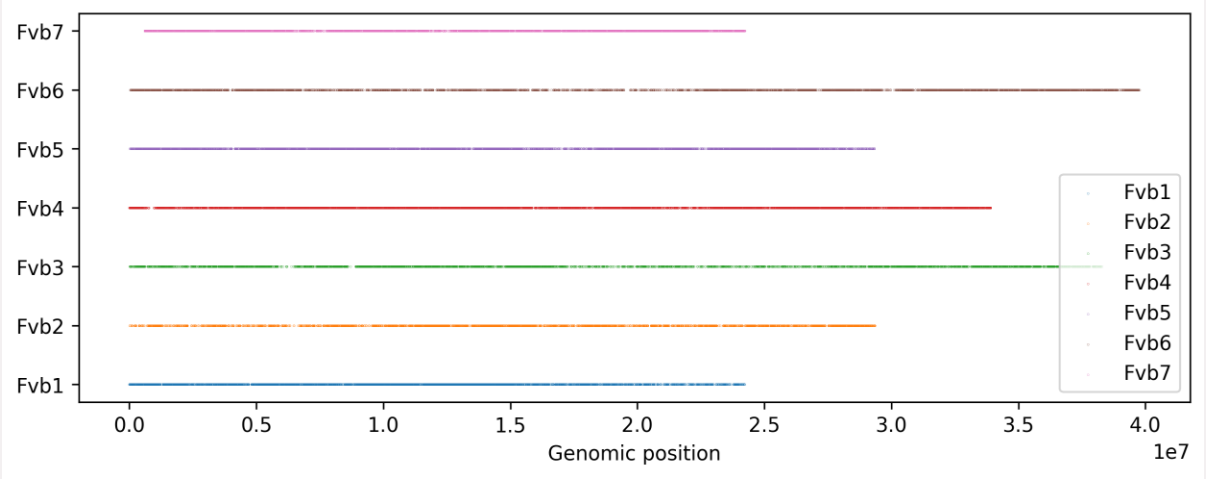
- Comparative analysis, showing that methylation changes vary notably across different contexts and environmental gradients in addition to how these factors influence genome-wide epigenetic patterns.
- Detected DMCs along the spatial gradient, the altitudinal gradient and between the soil moisture treatments. These DMCs often clustered together in genomic islands of differential methylation.
- Compared the methylation levels at single cytosines between samples.

Figure 4a (Clusters of DMCs)

- The distribution of DMC clusters along the genome under different environmental gradients, highlighting regions with dense epigenetic activity which could be crucial for adaptation.

Clustering behavior of DMCs





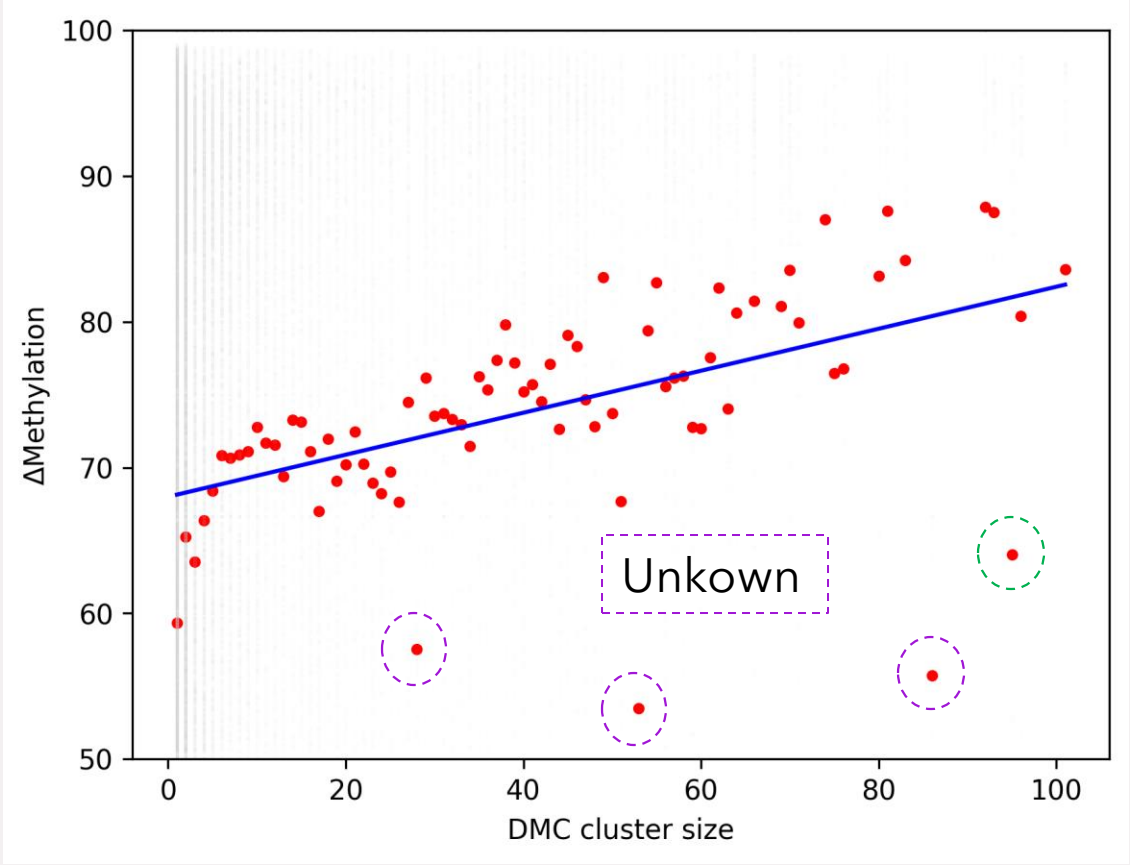
- Spatial
- Altitudinal
- Spatial clusters with exceptionally low DMC differentiation)

Data availability!

Figure 4b

Chr_Position	|meth.diff|	Clusters50_500	ClusterSize
Fvb1_164086	77.52188809	71	7
Fvb1_164092	96.07843137	71	7
Fvb1_164094	82.59905232	71	7

- Size of these clusters (number of DMCs in a cluster) vs their degree of methylation differentiation, indicating how cluster size might correlate with responsiveness to environmental changes.
- Spatial organization of methylation changes and their potential functional significance, suggesting that some regions of the genome are more susceptible or resistant to epigenetic changes based on environmental influences.



histone acetylation required for transcription (FvH4_1g16810)

Figure 5

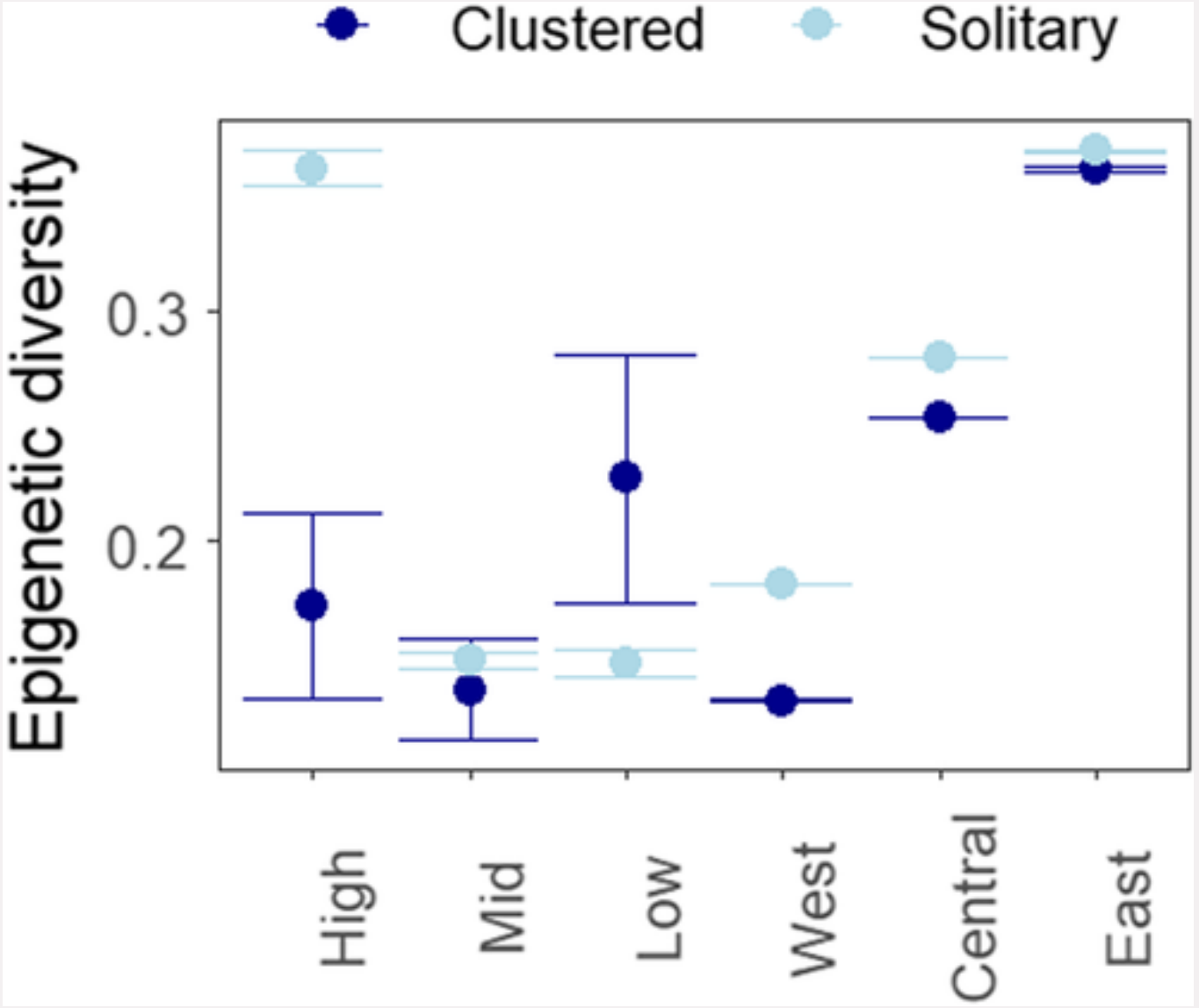
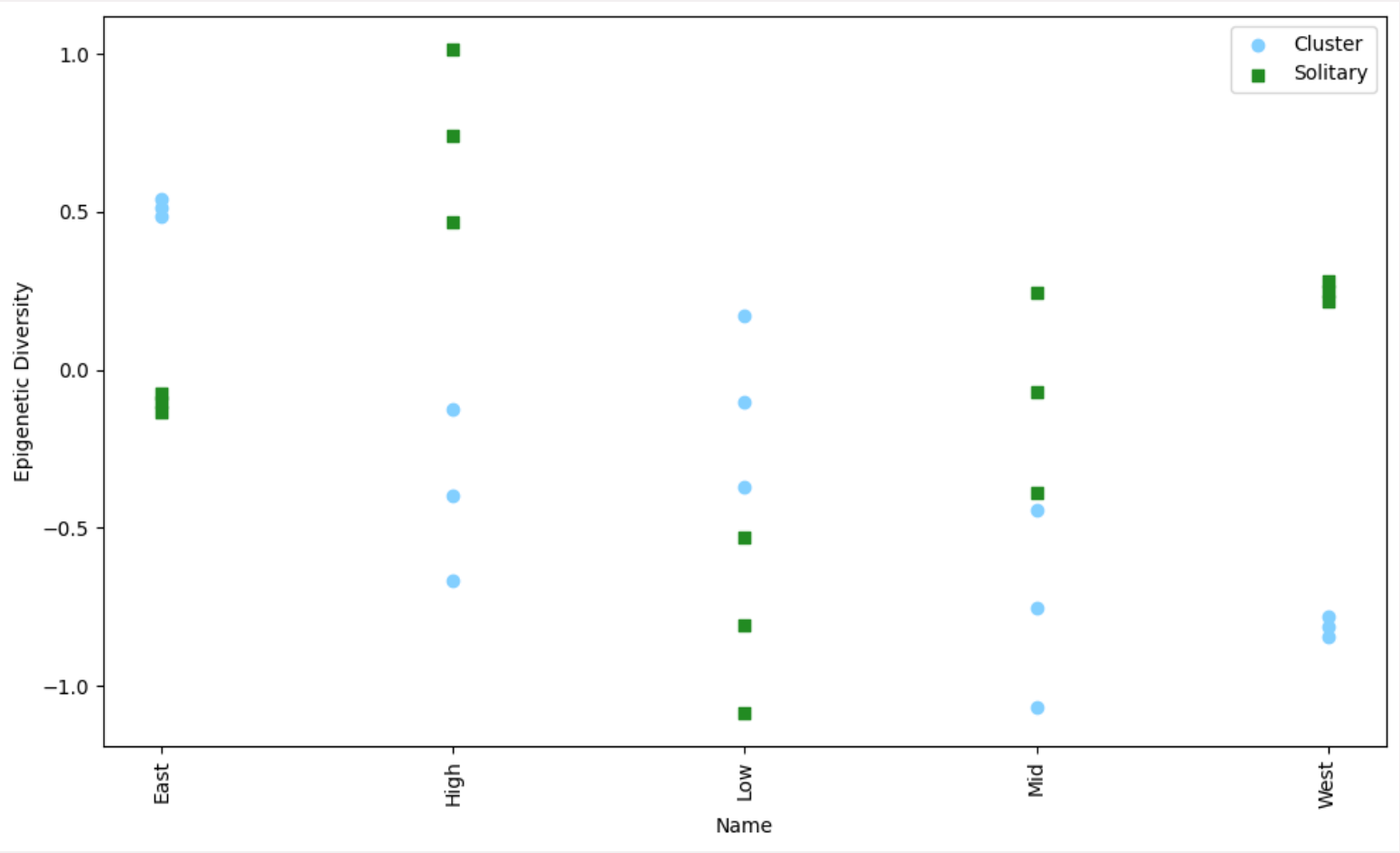

Clustered
Solitary
Epigenetic diversity
0.3
0.2
High
Mid
Low
West
Central
East

Figure 5 Continued



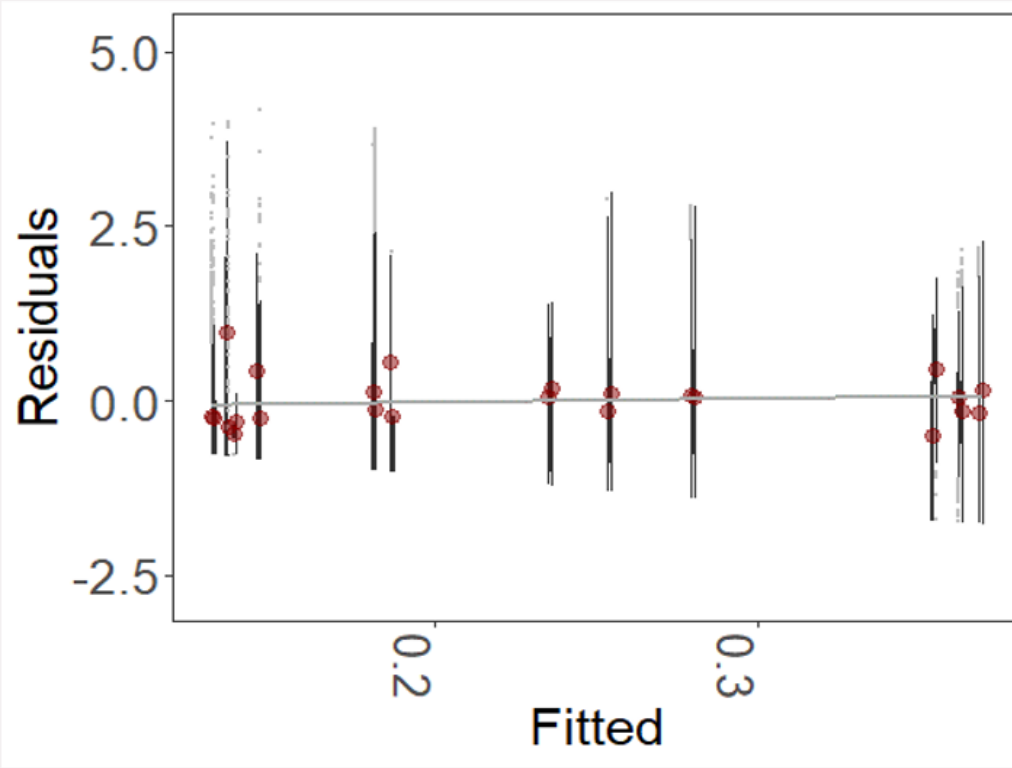


Figure 5 Continued

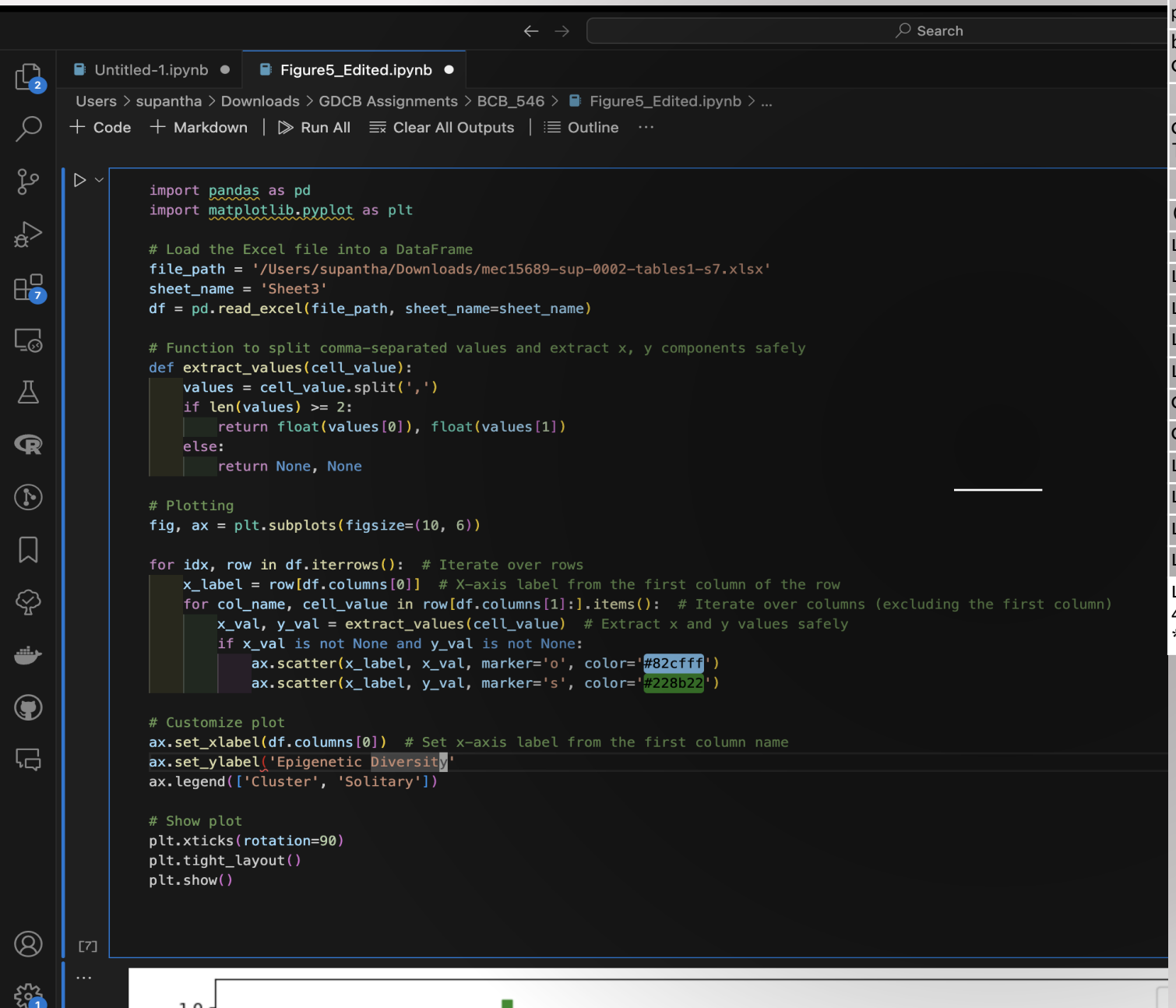


Table S4. Output of betaregression model used to test whether epigenetic diversity varies between populations and beteensolitary and clustered DMCs while controling for sequence context.

`betareg(formula = SQRT_ED ~ Location + Clustering + Context + Location * Clustering, data = Dataset)`

Coefficients (mean model with logit link):

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-1.074526	0.020806	-51.644	< 2e-16	***
LocationEast	0.511876	0.028329	18.069	< 2e-16	***
LocationHigh	-0.395972	0.269209	-1.471	0.14133	
LocationLow	-0.099675	0.271809	-0.367	0.71384	
LocationMid	-0.755079	0.310637	-2.431	0.01507	*
LocationWest	-0.812787	0.030747	-26.434	< 2e-16	***
ClusteringSolitary	0.131258	0.021140	6.209	5.33e-10	***
ContextNonCG	-0.004981	0.005451	-0.914	0.36080	
LocationEast:ClusteringSolitary	-0.102977	0.029033	-3.547	0.00039	***
LocationHigh:ClusteringSolitary	0.741669	0.272457	2.722	0.00649	**
LocationLow:ClusteringSolitary	-0.807768	0.276135	-2.925	0.00344	**
LocationMid:ClusteringSolitary	-0.071562	0.315866	-0.227	0.82077	
LocationWest:ClusteringSolitary	0.249938	0.031472	7.942	1.99e-15	***

		Upper Limit	Lower Limit
East	**0.511876,-0.102977**	0.540205,-0.073944	0.483547,-0.13201
High	-0.395972,0.741669	-0.126763,1.014126	-0.665181,0.469212
Low	-0.099675,-0.807768	0.172134,-0.531633	-0.371484,-1.083903
Mid	-0.755079,-0.071562	-0.444442,0.244304	-1.065716,-0.387428
West	-0.812787,0.249938	-0.78204,0.28141	-0.843534,0.218466

- Solitary DMCs showed more Epigenetic Diversity than the Clustered DMCs, which agrees with the hypothesis the authors proposed.
- Figure 5 might have been generated before applying the regression model and from raw data. Hence, we couldn't properly reproduce the figure, as Table S4 data doesn't correspond to the figure.

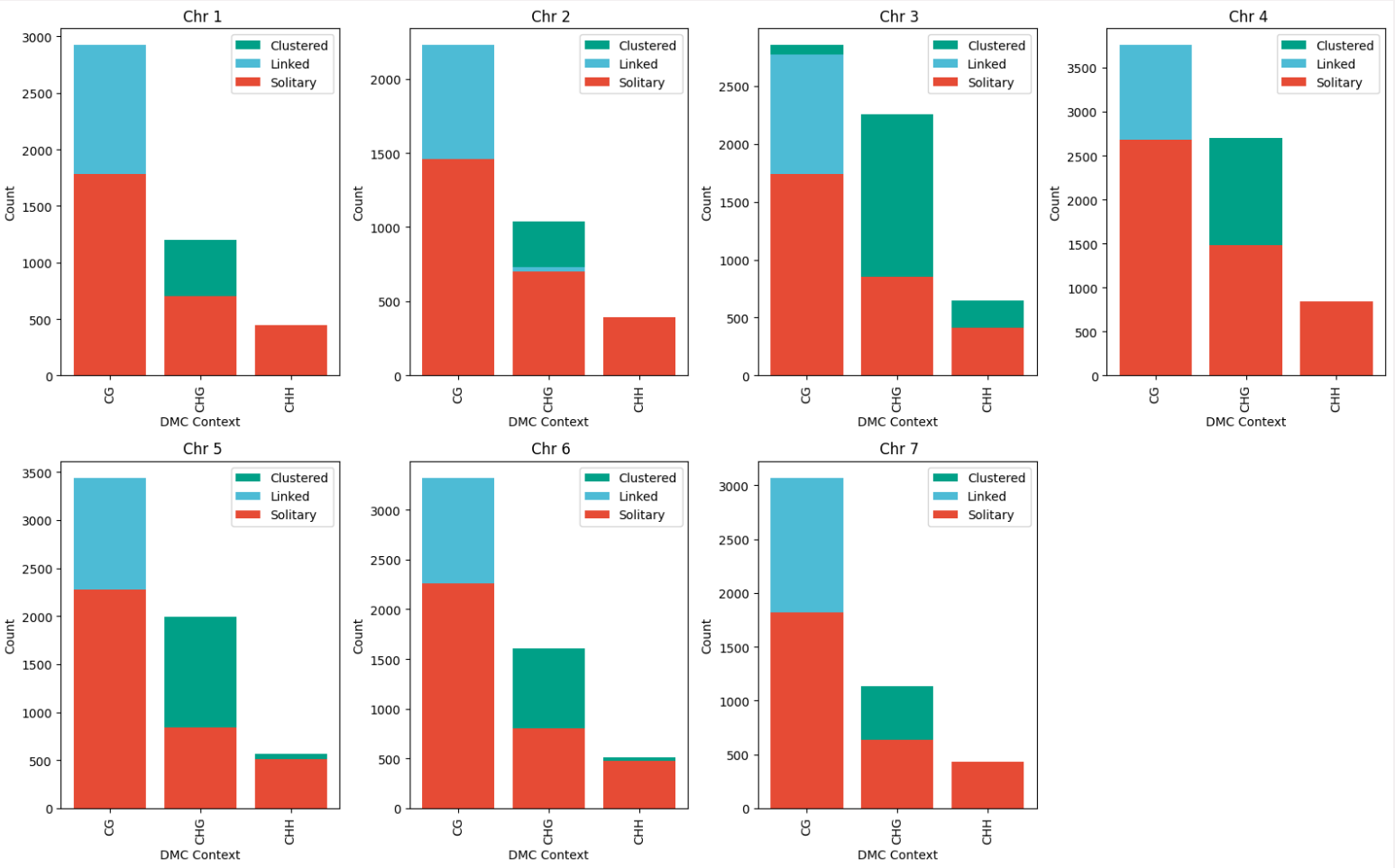
Extra Figure

M19

	A	B	C	D	E	F	G	H
1	Chr	Context	Clustering					
2	1	CHG	Solitary					
3	1	CG	Linked					
4	1	CG	Solitary					
5	1	CG	Clustered					
6	1	CG	Solitary					
7	1	CG	Solitary					
8	1	CG	Solitary					
9	1	CG	Solitary					
10	1	CHH	Solitary					
11	1	CG	Solitary					
12	1	CHH	Solitary					
13	1	CHG	Solitary					
14	1	CHH	Solitary					
15	1	CHG	Linked					
16	1	CG	Solitary					
17	1	CG	Solitary					
18	1	CHG	Solitary					
19	1	CG	Solitary					
20	1	CG	Solitary					
21	1	CG	Solitary					
22	1	CHG	Solitary					
23	1	CHH	Solitary					
24	1	CHG	Solitary					
25	1	CHG	Linked					
26	1	CHG	Solitary					
27	1	CG	Solitary					
28	1	CG	Solitary					
29	1	CG	Solitary					
30	1	CG	Solitary					
31	1	CHG	Solitary					
32	1	CG	Linked					
33	1	CG	Clustered					
34	1	CHG	Solitary					
35	1	CG	Solitary					
36	1	CG	Solitary					
37	1	CHH	Solitary					
38	1	CG	Clustered					
39	1	CG	Solitary					
40	1	CHH	Solitary					
41	1	CG	Clustered					
42	1	CHG	Solitary					
43	1	CG	Solitary					
44	1	CG	Solitary					
45	1	CG	Linked					
46	1	CG	Linked					
47	1	CG	Solitary					
48	1	CHG	Solitary					
49	1	CG	Solitary					
50	1	CG	Solitary					
51	1	CG	Solitary					
52	1	CG	Solitary					
53	1	CHH	Solitary					
54	1	CHH	Solitary					

DMCs | Clustering | Clustering stats | ED stats | Sheet3

Ready | Accessibility: Investigate



Code snippet for Extra Figure

- #Necessary modules loading
- import pandas as pd
- import matplotlib.pyplot as plt
- # We load the paper's Excel file into a DataFrame
- file_path = "mec15689-sup-0002-tables1-s7.xlsx"
- sheet_name = "Sheet2"
- df = pd.read_excel(file_path, sheet_name=sheet_name)

```
# Excluding the rows with NaN values in 'Context' or 'Clustering' columns
df_filtered = df.dropna(subset=['Context', 'Clustering'])

df_filtered = df_filtered[~df_filtered['Chr'].astype(str).str.startswith('0*')] # For some reaosns, there were names starting with 0*, instead of 1 to 7. We excluded them.

# Getting 1 to 7 chr from column.
unique_values = df_filtered['Chr'].unique()

# My favorite colors
custom_palette = ['#00A087', '#4DBBD5', '#E64B35']

# Plotting
plt.figure(figsize=(16, 10))

# Cycling through the colors in the palette as needed
color_index = 0

for i, value in enumerate(unique_values, 1):
    plt.subplot(2, 4, i)
```

Problems and Possibilities

- Some of the tables couldn't be translated to the associated figures.
- The authors discussed about Gene Ontology for a good portion of the paper. However, there wasn't a figure that could show us about different enriched GO terms. There was one plot in the supplementary table section.
  They generated this with shinyGO.
  For GO figures, online tools seems to be the best option.

Figure 6.

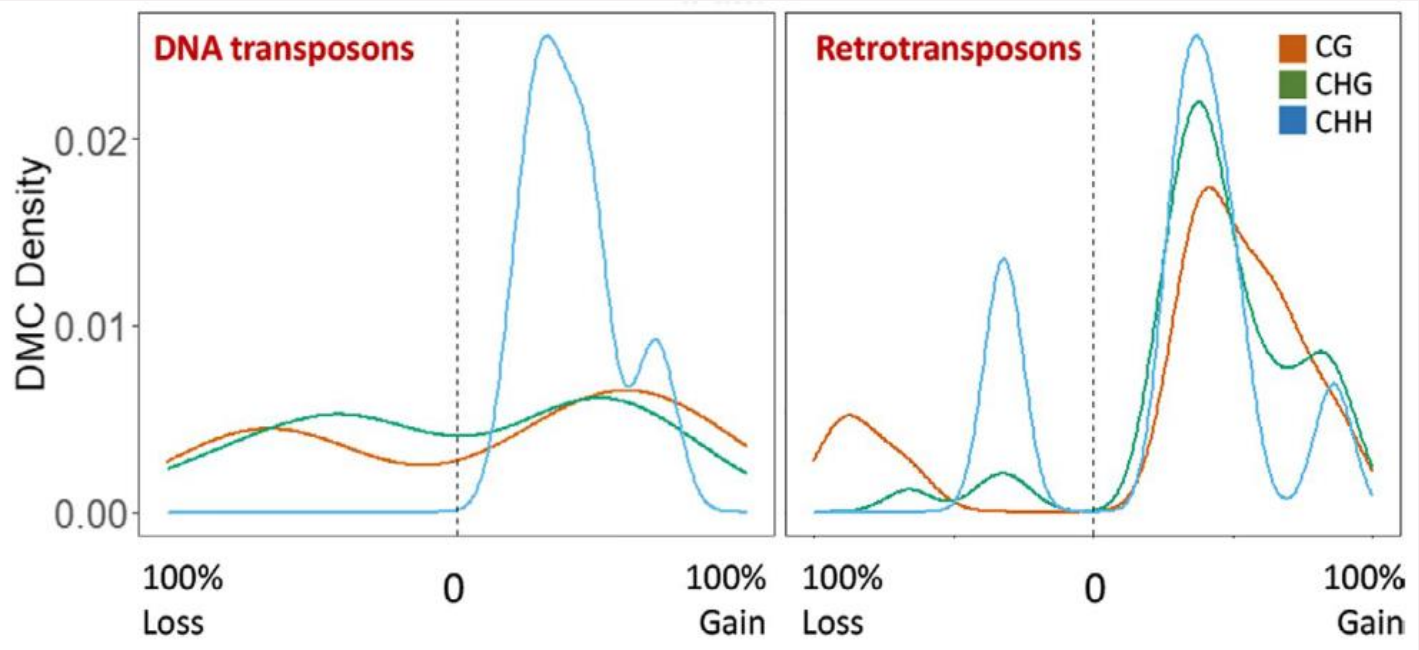
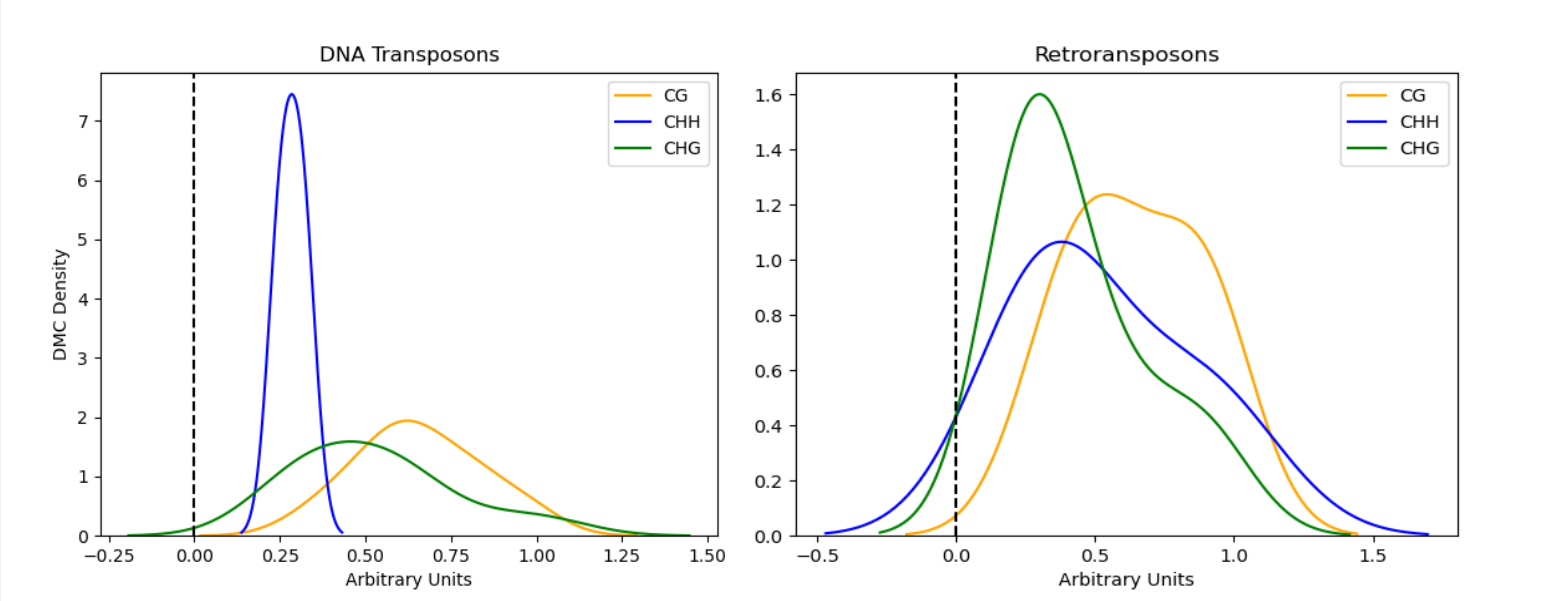


Figure 6. DMC density plot showing meth



- Total of 30 TE's were found to contain one or more DMC.
- All DMC TE's were found at the large spacial scale.
- CG DMCs were only found in retrotransposons
- Not shown: negligible DMCs between acute soil level dryness

Problems with figure 6:

- Many crucial datapoints are unavailable through their filtered dataset leading to...
  - Possible conclusion differences.
  - Missing "loss-of" data.
  - *Scaling differences.*

Overall challenges:

- *Raw dataset is over 300Gb between 60 different sequencing files (30 different biological samples)*
- *Programing language was not described directly.*
  - FASTQ Screen
  - BISMARK for mapping
  - Methylkit in R
  - Filtering steps
  - Clustering with Bumphunter
  - Betareg for effect size
  - OMICSBOX for GO

Documentation

- Programming language: Python 3
- Article's Publicly available dataset was used with minimum modifications.
- Clear instructions, codes and libraries are documented

in our GitHub.

GitHub Repository

Paper Citation:

De Kort, H., Panis, B., Deforce, D., Van Nieuwerburgh, F., & Honnay, O. (2020). Ecological divergence of wild strawberry DNA methylation patterns at distinct spatial scales. In Molecular Ecology (Vol. 29, Issue 24, pp. 4871–4881). Wiley. https://doi.org/10.1111/mec.15689.

Paper Conclusions

- Both small and large spatial scales showed evidence of holding epigenetic memory and methylation differentiation
    - *Distinct to large: demethylation of transposable elements*
- Short-term stress (drought) did not impact patterns of epigenetic marks
- What population the strawberry plants originated from had a larger effect than short-term environmental stress in shaping epigenetic marks