Searching for selection:

mix oil, air, water, a bit of seasonality, and genotypic variance.

Karl Fetter

Burke Lab/Plant Biology

April 8th, 2021

Searching for selection...

- Allele frequencies evolve under neutral or non-neutral processes.

- Allele frequency variation is linked to phenotypic variation by the central dogma.

- Tests for selection can be based on a sample of genetic loci and phenotypes.
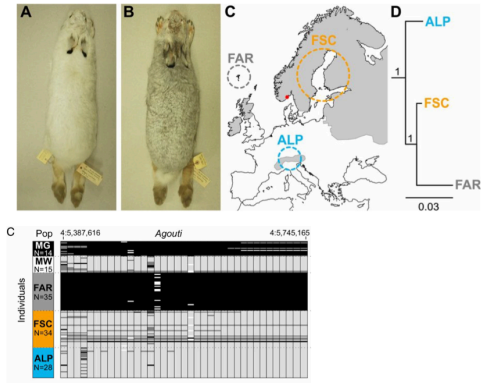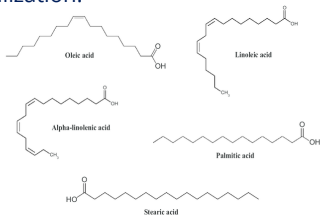
Abiotic selection

Iwona Giska et al. "Introgression drives repeated evolution of winter coat color polymorphism in hares". In: *Proceedings of the National Academy of Sciences* 116.48 (2019), pp. 24150–24156
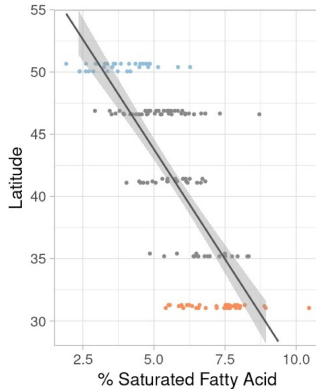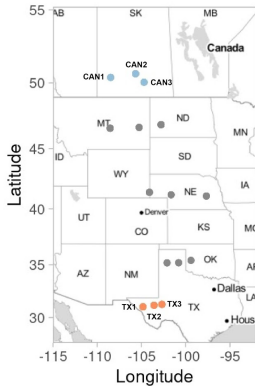
Cold stress & fatty acids

- Cold temperatures can be challenging for plants.
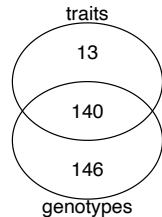- Fatty acids are important in membrane stabilization.
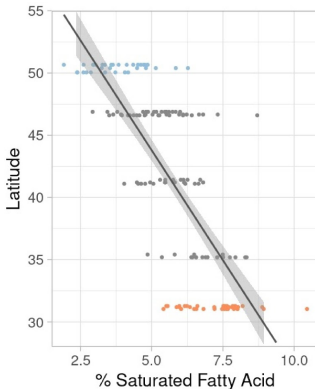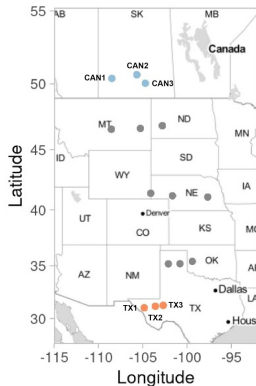

	Yag	Cetinel
% PUFA	72	77
% Germination	25	75

Table 1: Kaymak, "Potential effect of seed fatty acid profile of pepper (Capsicum annuum L.) cultivars on germination at various temperatures."

Initial Observation

Initial Observation

Questions & Tests

- Is the % saturated fatty acid content measured from oils under spatially-explicit selection?

- - $Q_{ST} - F_{ST}$

- Is there evidence of an abiotic driver of selection?

- - Multiple regression of climatic data to % saturated fatty acids.

Inference of selection - Q_{ST}-F_{ST}

- Q_{ST} is an analogue to F_{ST}. A quantitative trait with an additive genetic basis is expected to equal F_{ST} in the absence of selection.

$$Q_{ST} = \left(\frac{\sigma^2_{A_b}}{\sigma^2_{A_b} + \sigma^2_{A_w}} \right)$$

- Q_{ST} is the amount of phenotypic variance between populations relative to the total phenotypic variance.

- Q_{ST} is a spatially explicit statistic and appropriate for landscape genetic studies.

Three outcomes are possible:
- $Q_{ST} > F_{ST}$: directional selection
- $Q_{ST} = F_{ST}$: drift (neutral)
- $Q_{ST} < F_{ST}$: balancing selection

Qst model specification (Animal Model)

$$y_{ijk} \sim 1 + B_i + P_j + I_k + \epsilon_{ijk}$$

Y, the mean saturated fatty acid value for each family
at the ith block, j^{th} population, k^{th} genotype,
B_i, blocking effects in greenhouse,
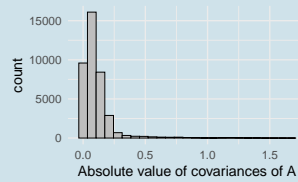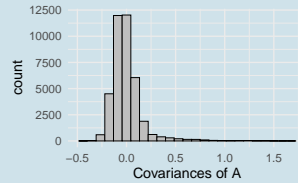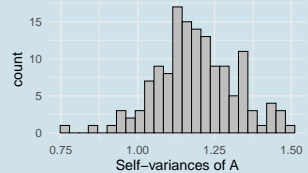P_j, genetic population,
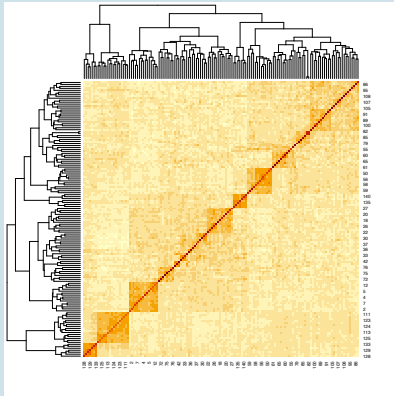I_k, a random effect of genotype,
ϵ_{ijk}, error term.

$$I \sim N(0, A)$$

A $= \sigma_u^2 * G$
σ_u^2 is the variance component expalined by individual effects (estimated),
and G is the additive genetic relationship matrix estimated from genetic loci.

Genomic relatedness matrix (G).
Estimated with `AGHmatrix`.

Mixed model results

```
> qst_mod
 Family: gaussian
  Links: mu = identity; sigma = identity
Formula: sat ~ 1 + block + deme + (1 | gr(ind_code, cov = G))
   Data: dat1 (Number of observations: 140)
Samples: 4 chains, each with iter = 6000; warmup = 3000; thin = 1;
         total post-warmup samples = 12000

Group-Level Effects:
~ind_code (Number of levels: 140)
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)     0.78      0.15     0.51     1.08 1.00     2413     4468

Population-Level Effects:
          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept     4.60      0.29     4.02     5.18 1.00     9648     8920
block         0.10      0.10    -0.09     0.28 1.00    11657     9012
demesouth     1.79      0.41     0.97     2.59 1.00     6827     7465

Family Specific Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma     0.96      0.10     0.78     1.16 1.00     2552     4025

Samples were drawn using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

UNIVERSITY OF
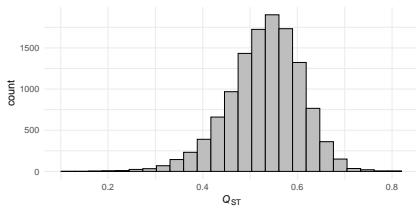**GEORGIA**

Qst Estimates

$$Q_{ST} = \left(\frac{\sigma_{A_b}^2}{\sigma_{A_b}^2 + \sigma_{A_w}^2} \right)$$

```
out <- qst_mod %>%
  brms::posterior_samples(.vars=2:4, .funs=funs(.^2))

# Calculate Qst from the variance components.
qst = out$b_demesouth /
  (out$b_demesouth + 2 * out$sd_ind_code__Intercept)

# Print the median & interval estimates
tidybayes::median_qi(qst)
y         ymin      ymax   .width .point .interval
0.5371864 0.3607508 0.6673257   0.95  median
```
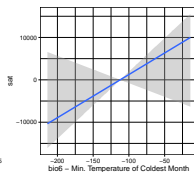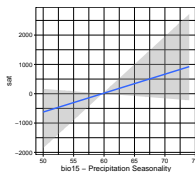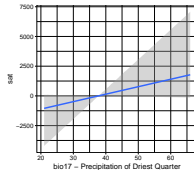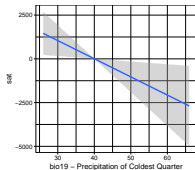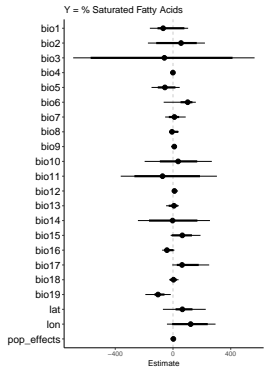

- $Q_{ST} - F_{ST} = 0.53 - 0.21 = 0.32$

Mixed effects model

$$y \sim E + XY + P_j + \epsilon_j$$

Conclusions

- $Q_{ST} - F_{ST}$ test indicates directional selection.
- This test needs refinement and confidence intervals put around it. We can use bootstrapping, Whitlock & Guillame (2009) method, ...

- The precipitation of the coldest quarter (bio19) is the only environmental variable significant at 95 % CI.
- Precipitation seasonality (bio15) and minimum temperature of the coldest month (bio6) are significant at the 66 % CI.
- These results suggest temperature and precipitation alone aren't sufficient to explain the directional selection.