

Searching for selection:

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

Karl Fetter

Burke Lab/Plant Biology

April 22nd, 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.

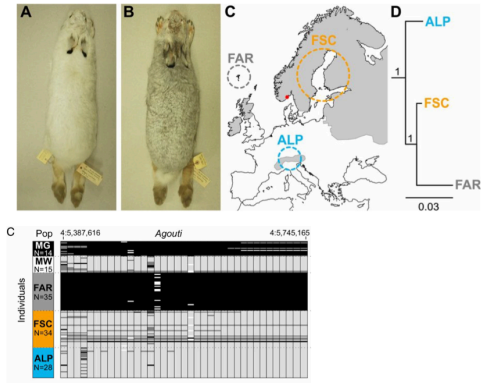


Abiotic selection

Iwona Giska et al.

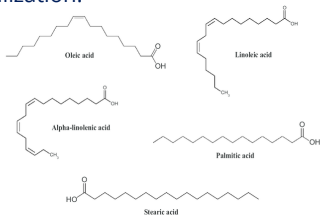
“Introgression drives repeated evolution of winter coat color polymorphism in hares”.

In: *Proceedings of the National Academy of Sciences* 116.48 (2019), pp. 24150–24156



Cold stress & fatty acids

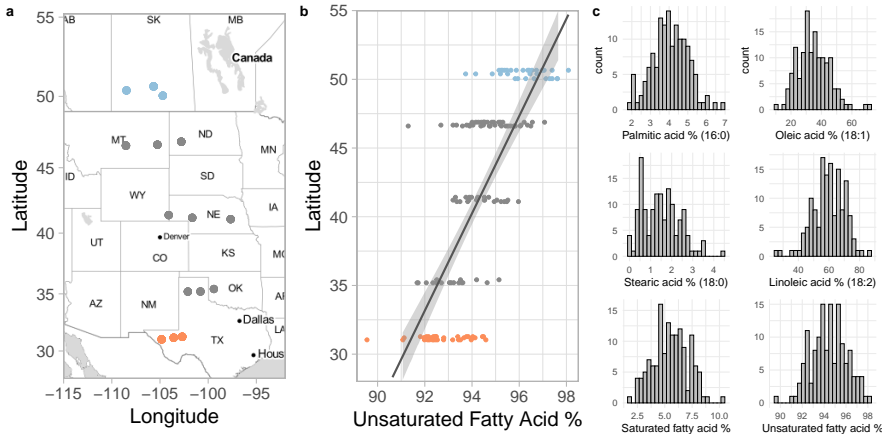
- Cold temperatures can be challenging for plants.
- Fatty acids are important in membrane stabilization.



	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



Questions & Tests

- Is the unsaturated fatty acid content measured in seed oil under spatially-explicit selection?

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

- Is there evidence of an abiotic driver of selection?

- Multiple regression model

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.

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

- 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)

$$y_{ijk} \sim 1 + B_j + P_k + I_l + \epsilon_{ijk}$$

y_i is the mean saturated fatty acid value for each family
at the j^{th} block, k^{th} population, l^{th} genotype,

B_j is the blocking effects in greenhouse,

P_k is the genetic population,

I_l is a random effect of genotype,

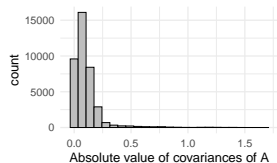
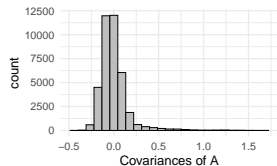
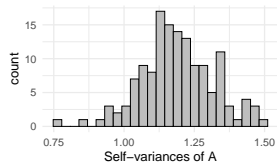
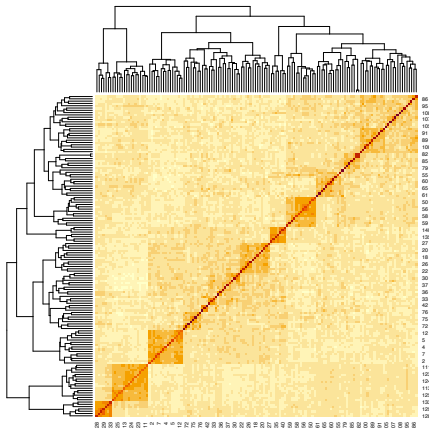
ϵ_{ijkl} is the error term.

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

$$A = \sigma_u^2 * G$$

σ_u^2 is the variance component explained 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).

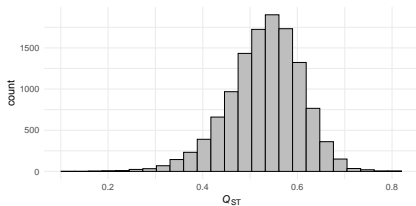
Qst Estimates

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

```
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
```



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

Multiple regression model

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

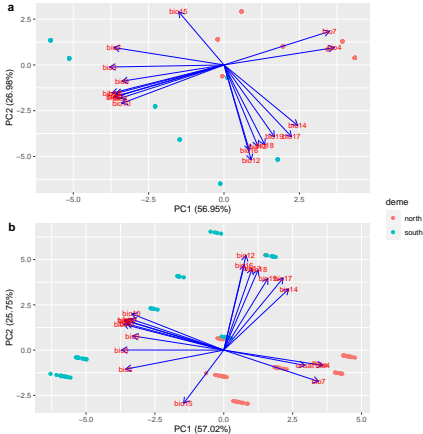
y_i is the mean unsaturated fatty acid value for each family,

P_j is the genetic population,

ϵ_{ik} is the error term.

```
mod02 <- brm(sat ~ bio1 + bio2 + bio3 + bio4 +  
             bio5 + bio6 + bio7 + bio8 + bio9 +  
             bio10 + bio11 + bio12 + bio13 + bio14 +  
             bio15 + bio16 + bio17 + bio18 + bio19 +  
             lat + lon +  
             (1|pop_code),  
             data = dat,  
             family = "gaussian",  
             chains = 10L, seed = 2398, iter = 5000) %>%  
add_criterion("loo", relou = TRUE)
```

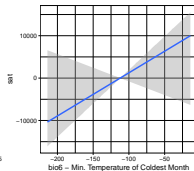
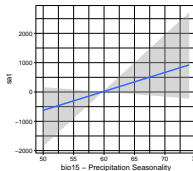
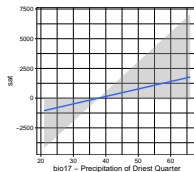
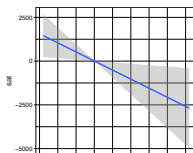
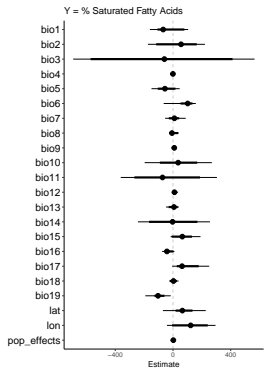
Climate Variables



- BIO1 = Annual Mean Temperature
- BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp))
- BIO3 = Isothermality (BIO2/BIO7) ($\times 100$)
- BIO4 = Temperature Seasonality (standard deviation $\times 100$)
- BIO5 = Max Temperature of Warmest Month
- BIO6 = Min Temperature of Coldest Month
- BIO7 = Temperature Annual Range (BIO5-BIO6)
- BIO8 = Mean Temperature of Wettest Quarter
- BIO9 = Mean Temperature of Driest Quarter
- BIO10 = Mean Temperature of Warmest Quarter
- BIO11 = Mean Temperature of Coldest Quarter
- BIO12 = Annual Precipitation
- BIO13 = Precipitation of Wettest Month
- BIO14 = Precipitation of Driest Month
- BIO15 = Precipitation Seasonality (Coefficient of Variation)
- BIO16 = Precipitation of Wettest Quarter
- BIO17 = Precipitation of Driest Quarter
- BIO18 = Precipitation of Warmest Quarter
- BIO19 = Precipitation of Coldest Quarter

Multiple regression results

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



Conclusions

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