An R package for heritability calculations for plant breeding trials

Fonti Kar, Emi Tanaka

ANU AAGI

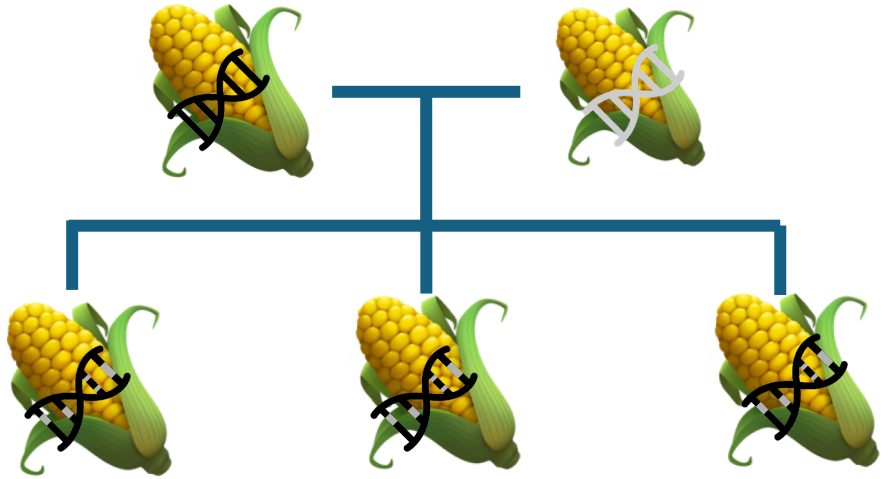
hi. 👋

hi. 👋

Selective plant breeding

Traits must:

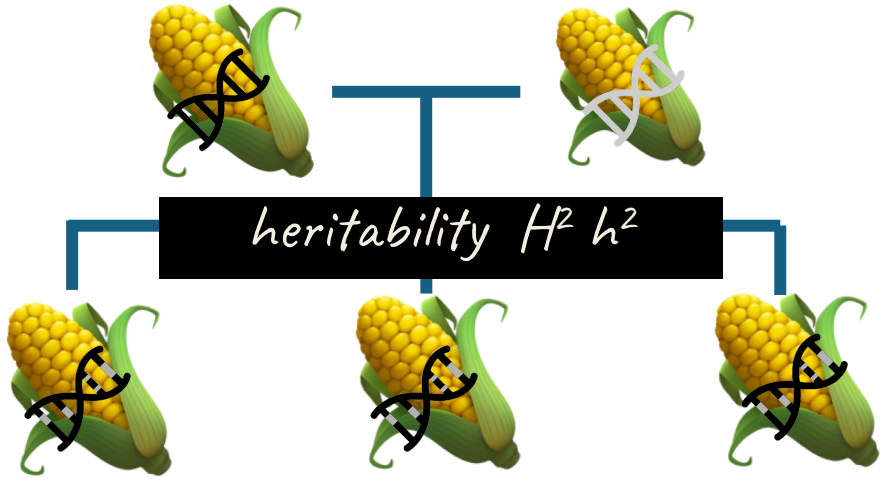
- show variability
- have a genetic basis



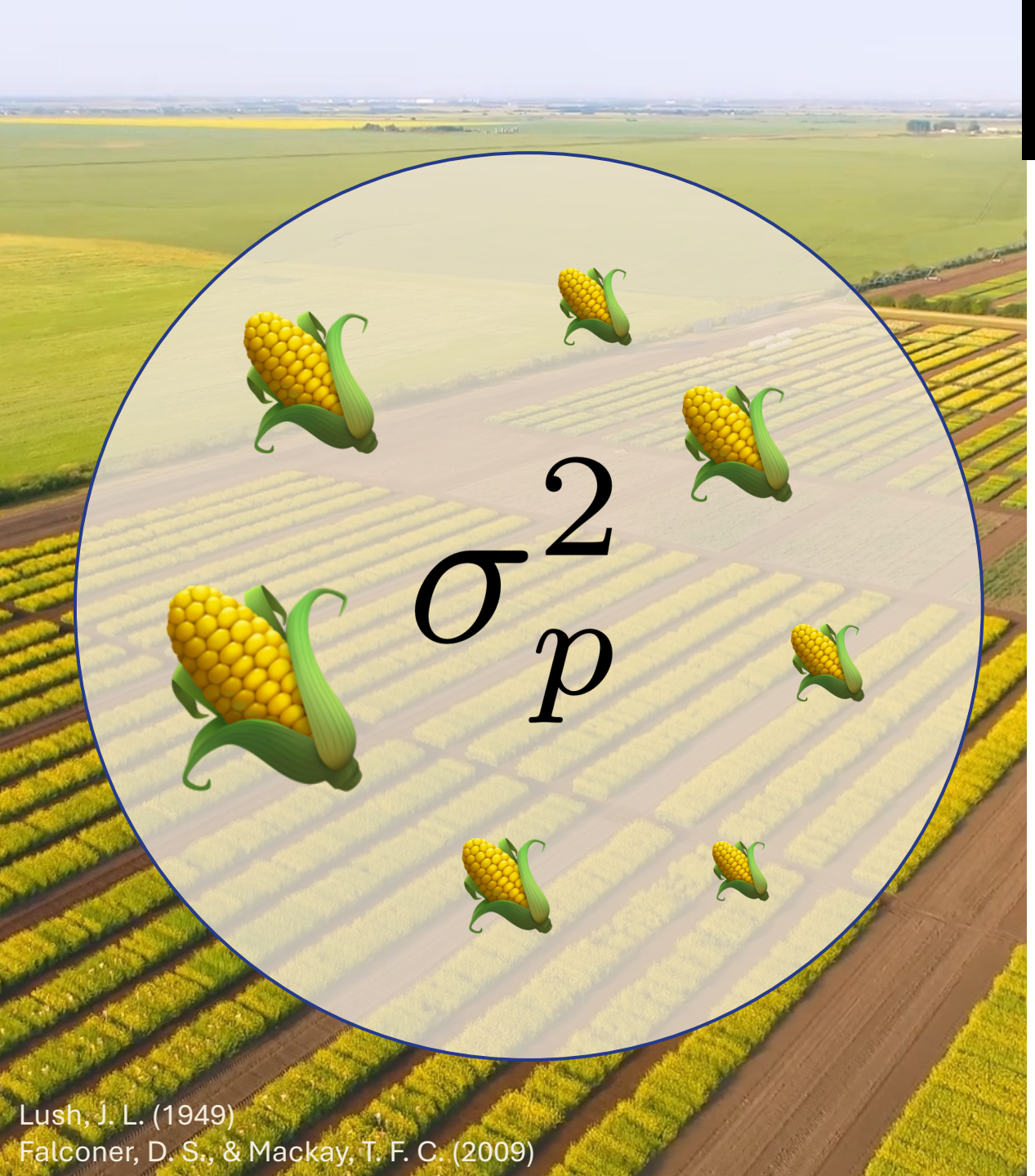
Selective plant breeding

Traits must:

- show variability
- have a genetic basis



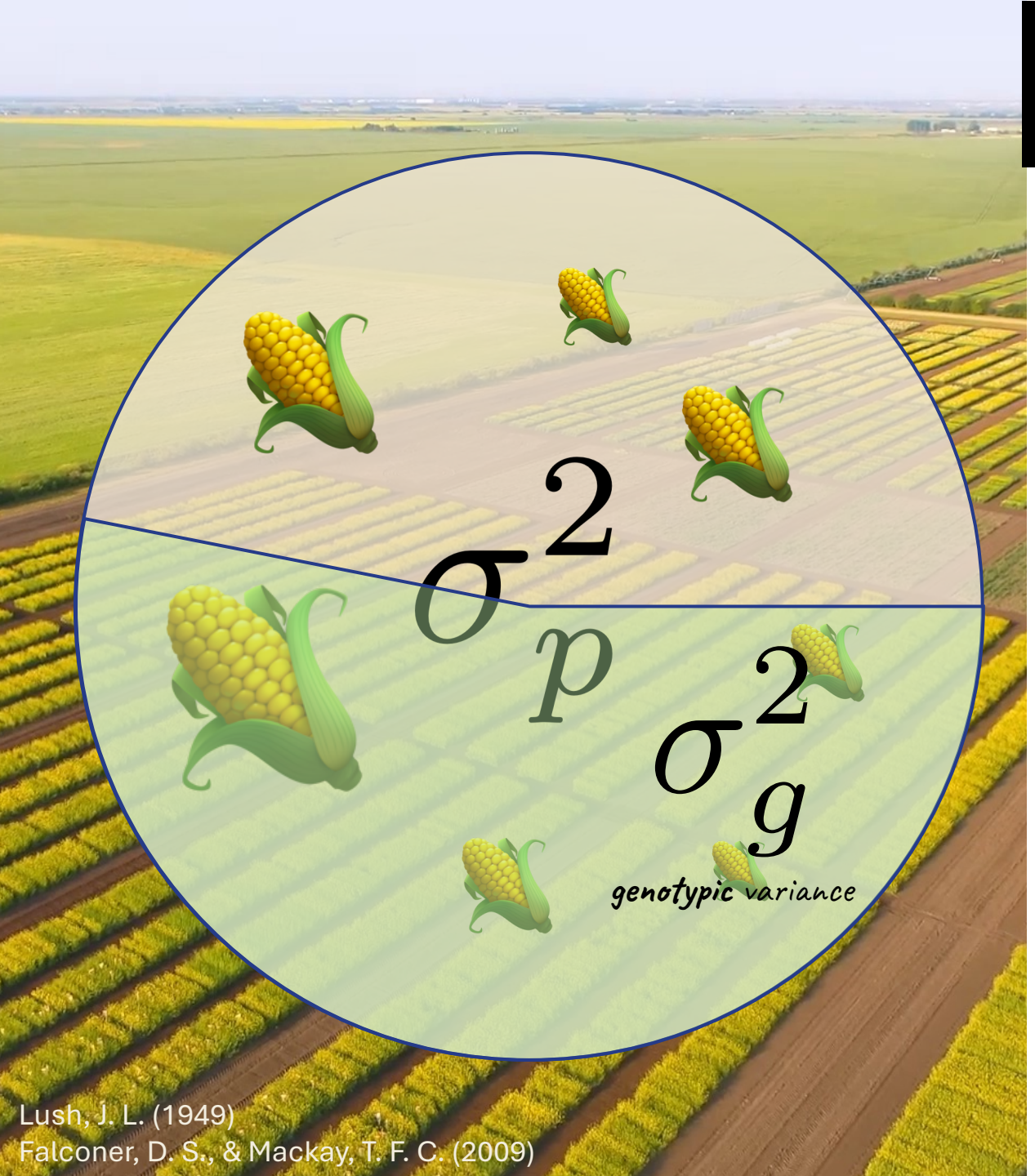
Total phenotypic variation



- Environmental 🐛🌡️🌦️
- Genetic 🧬
- Peripheral 🚜📏🪏

Lush, J. L. (1949)
Falconer, D. S., & Mackay, T. F. C. (2009)

Proportion of variation explained by genetic differences

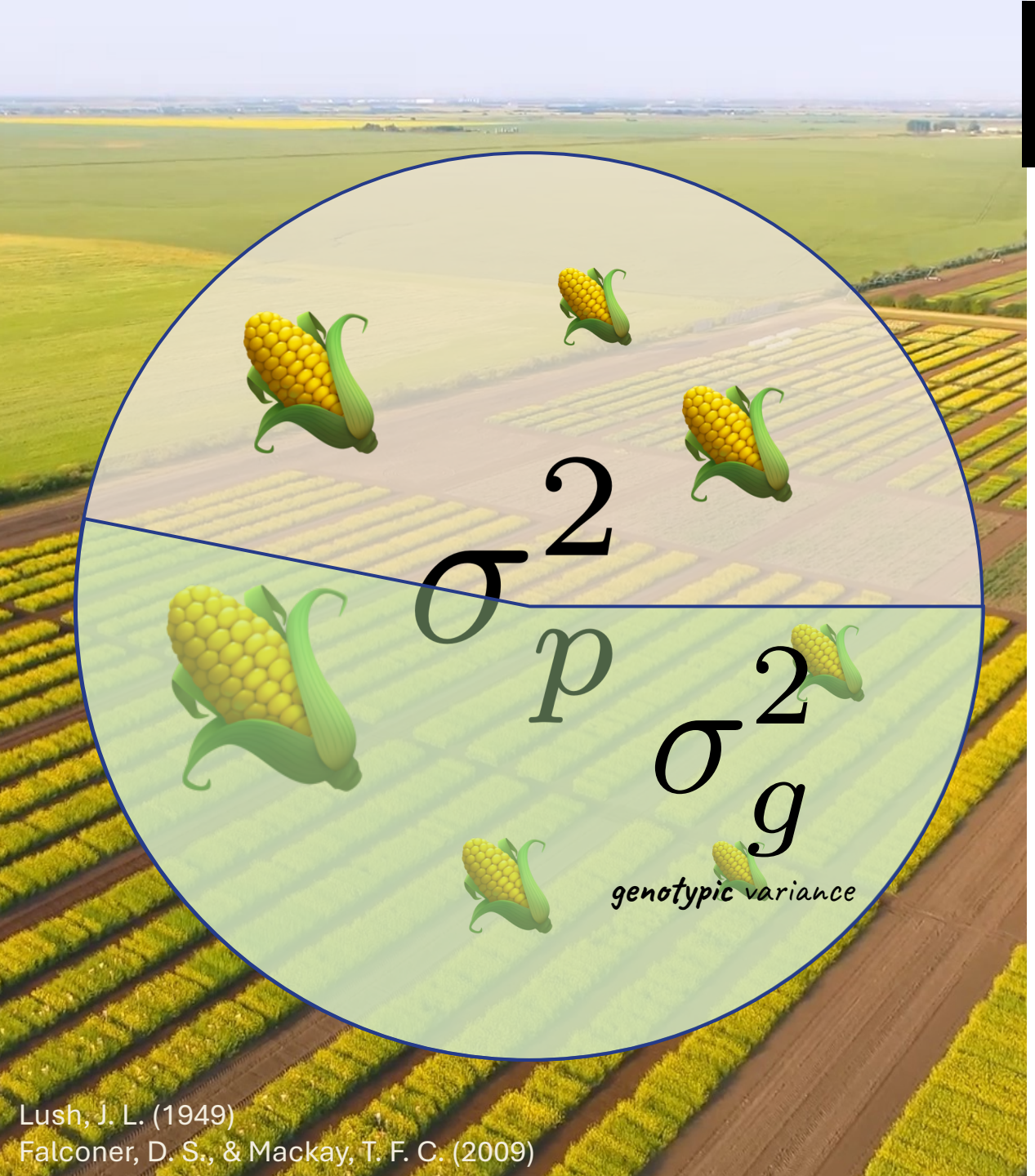


$$H^2 = \frac{\sigma^2_g}{\sigma^2_p}$$

broad-sense heritability

Lush, J. L. (1949)
Falconer, D. S., & Mackay, T. F. C. (2009)

Proportion of variation explained by genetic differences



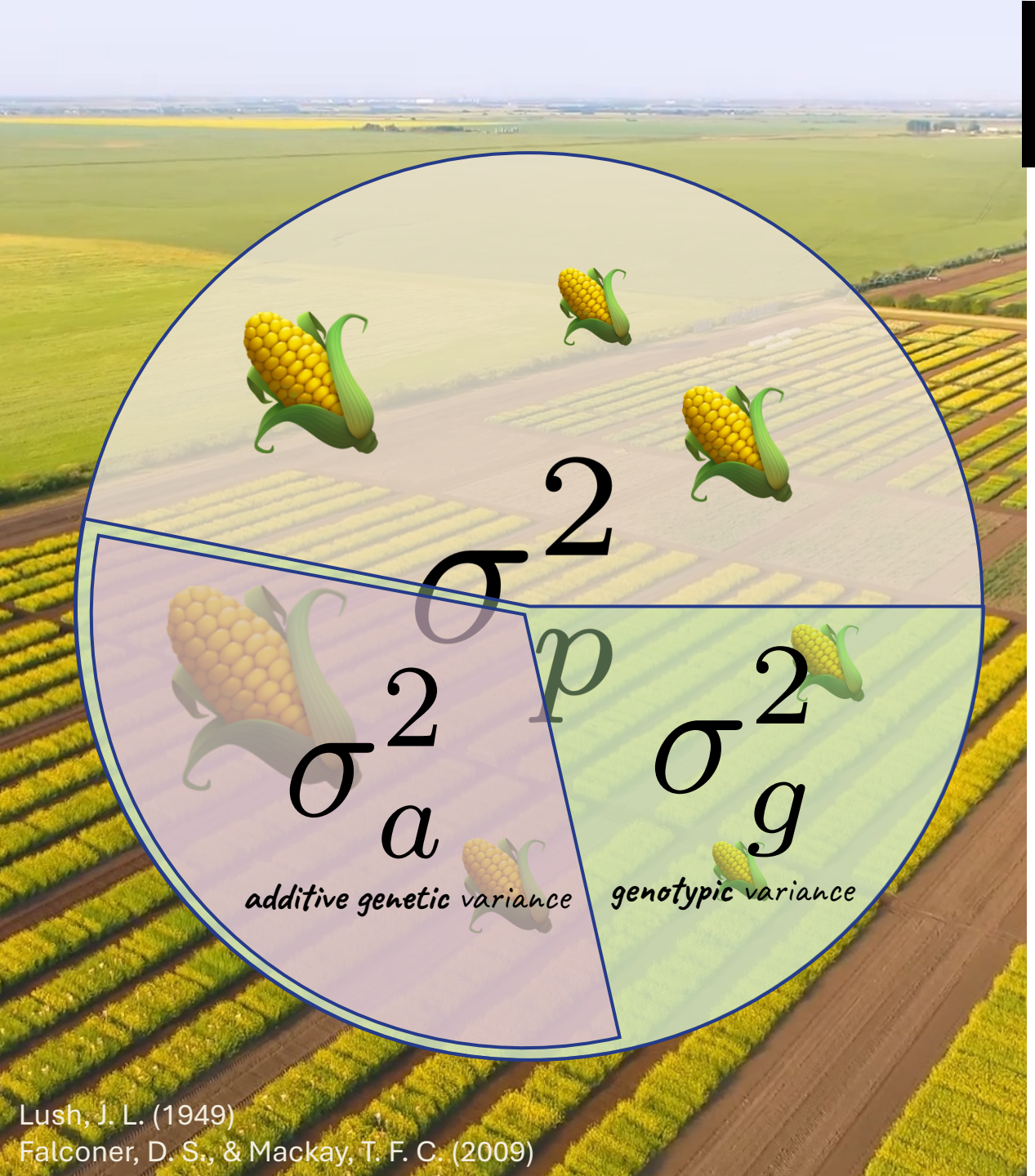
$$H^2 = \frac{\sigma^2_g}{\sigma^2_p}$$

broad-sense heritability

- additive
- epistatic
- dominance

Lush, J. L. (1949)
Falconer, D. S., & Mackay, T. F. C. (2009)

Proportion of variation explained by genetic differences



$$H^2 = \frac{\sigma^2_g}{\sigma^2_p}$$

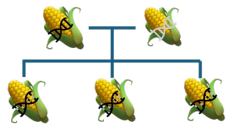
broad-sense heritability

- additive
- epistatic
- dominance

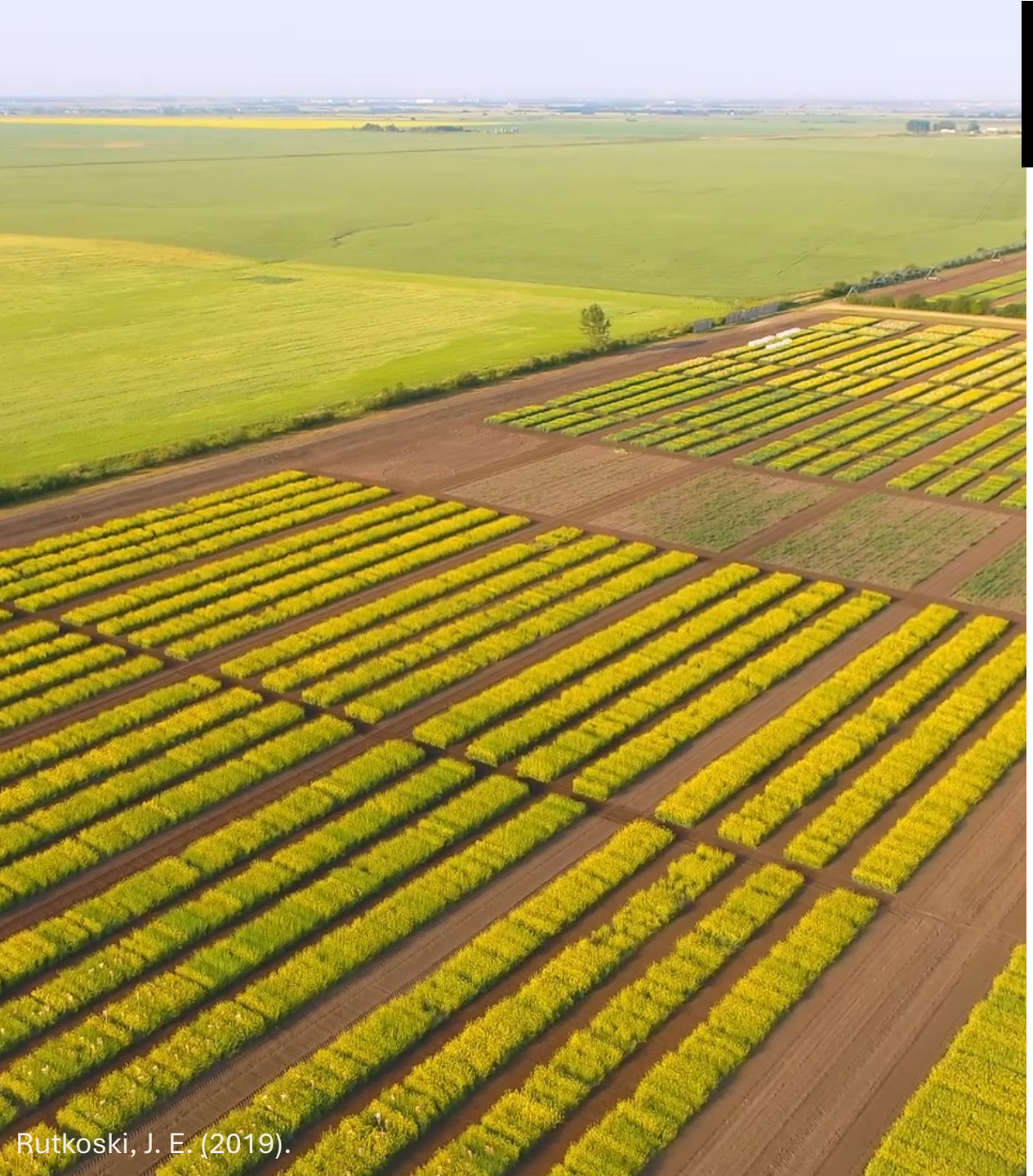
$$h^2 = \frac{\sigma^2_a}{\sigma^2_p}$$

narrow-sense heritability

"stuff that is passed on reliably"



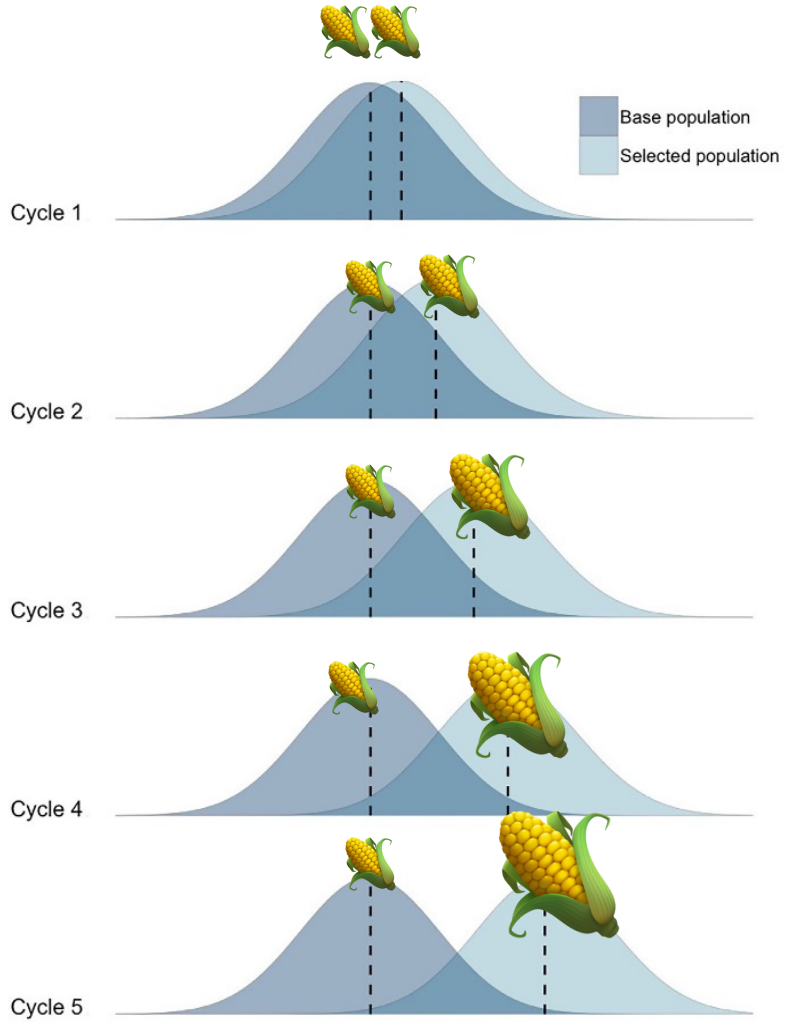
Lush, J. L. (1949)
Falconer, D. S., & Mackay, T. F. C. (2009)



Rutkoski, J. E. (2019).

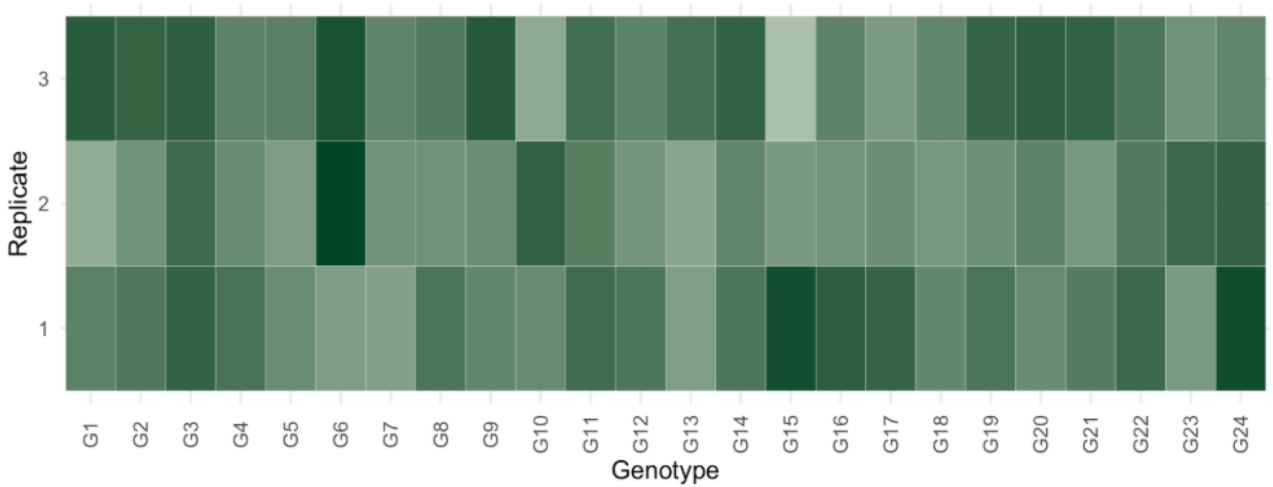
Genetic gain $\propto H^2$

- Genetic gain = improvement in average phenotype

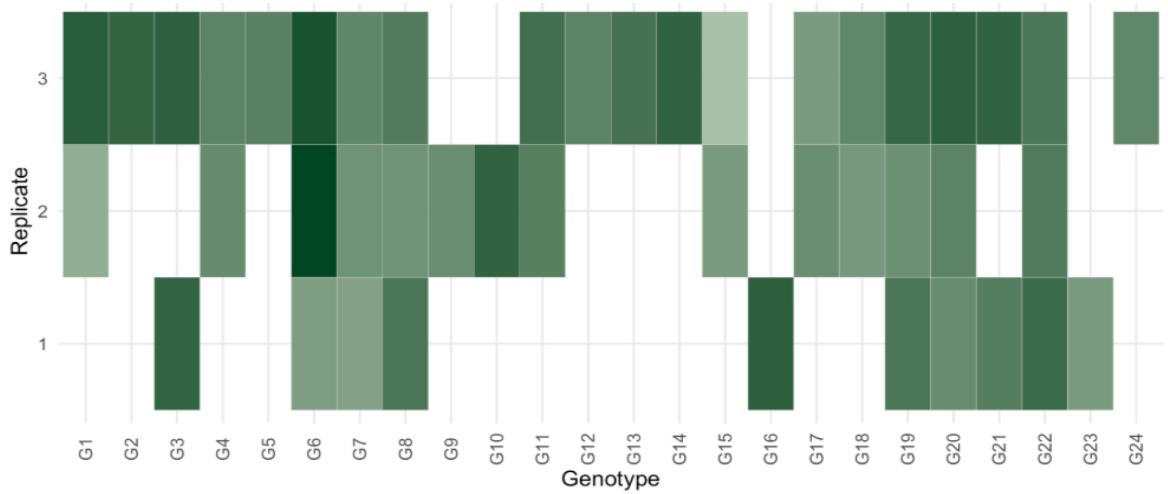


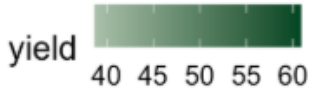
Field experiments are complex and messy

balanced 💪



unbalanced 😤





- Classical H^2 was developed for balanced designs
- Generally, in unbalanced scenarios H^2 is underestimated

Competing estimates for heritability 💡

$$H^2_{Standard} = \frac{\sigma^2_g}{\sigma^2_g + \frac{1}{n_g}\sum^{i=1}_{n_g} \sigma^2_p / n_{gi}} \qquad H^2_{Cullis} = 1 - \frac{PEV^{BLUP}_{\overline{\Delta}..}}{2\sigma^2_g}$$

$$H^2_{Piepho} = \frac{\sigma^2_g}{\sigma^2_g + \overline{PEV^{BLUE}_g}/2} \qquad H^2_{\Delta_{ij}} = 1 - \frac{\overline{PEV^{\text{BLUP}}_{ij}}}{2\sigma^2_g}$$

$$H^2_{Oakey} = \frac{\sum^{n_g}_{i=n_z+1} \lambda_i}{\sum^{\lambda_i \neq 0}_{n_g}}$$

Competing estimates for heritability 💡

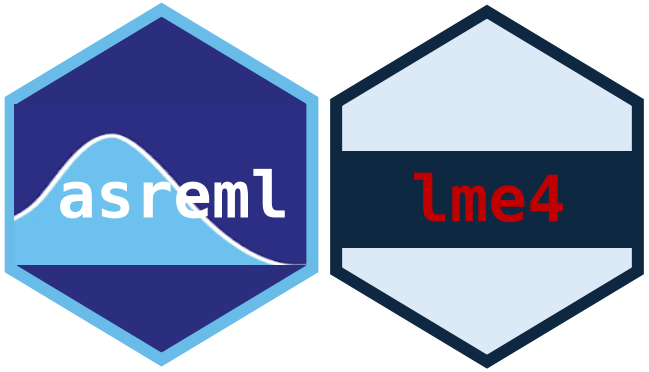
$$H^2_{Standard} = \frac{\sigma^2_g}{\sigma^2_g + \frac{1}{n_g}\sum^{i=1}_{n_g} \sigma^2_p / n_{gi}}$$

$$H^2_{Cullis} = 1 - \frac{PEV^{BLUP}_{\overline{\Delta}..}}{2\sigma^2_g}$$

Under balanced settings, all are approximately equal

$$H^2_{Oakey} = \frac{\sum^{n_g}_{i=n_z+1} \lambda_i}{\sum^{\lambda_i \neq 0}_{n_g}}$$

Mixed model magic ✨



$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_g\boldsymbol{u}_g + \mathbf{Z}_p\boldsymbol{u}_p + \boldsymbol{e}$$

e.g. yield

fixed effects

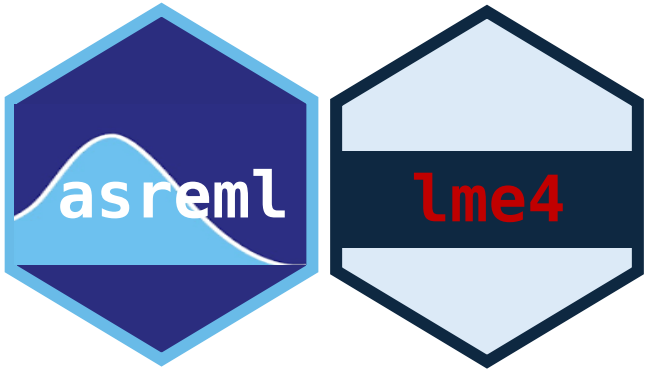
random genotype effect

other random effects

$$\begin{bmatrix} \boldsymbol{u}_g \\ \boldsymbol{u}_p \\ \boldsymbol{e} \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{G}_g & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R} \end{bmatrix} \right)$$

assumes random effects and errors are not correlated

Mixed model magic ✨



$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_g\boldsymbol{u}_g + \mathbf{Z}_p\boldsymbol{u}_p + \boldsymbol{e}$$

e.g. yield — fixed effects — random genotype effect — other random effects

Best Linear Unbiased Estimator

BLUE

$\hat{\boldsymbol{\beta}}$

Best Linear Unbiased Predictor

BLUP

for genotypes

$\tilde{\boldsymbol{u}}_g$

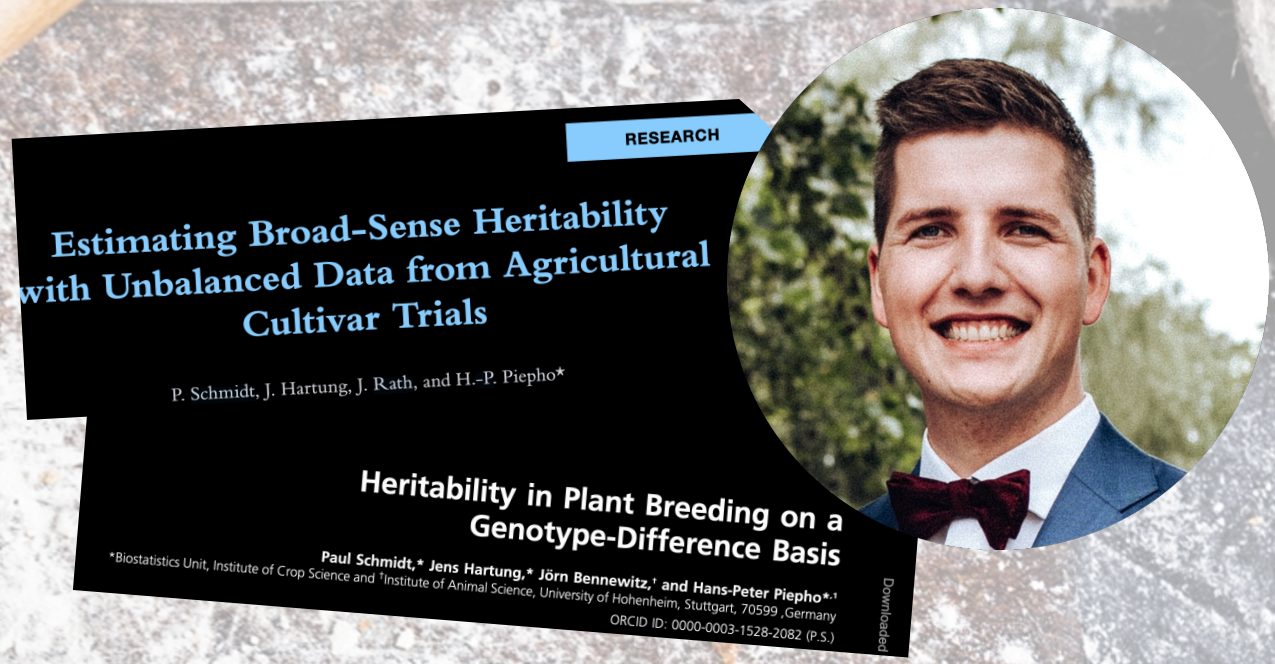
prediction error variances

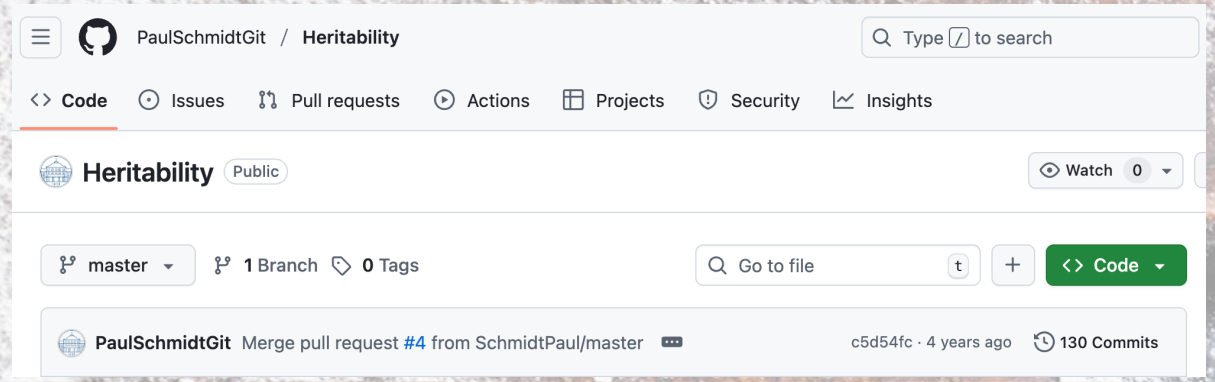
$$\text{PEV}_g^{\text{BLUE}} = var(\tilde{\boldsymbol{u}}_{\boldsymbol{g}} - \boldsymbol{u}_{\boldsymbol{g}})$$

$$\text{PEV}_g^{\text{BLUP}} = var(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$$

Motivations

1. Code exists but needs TLC
2. Variance components are hard to extract from model objects
3. Unclear how we are arriving on final values





https://github.com/PaulSchmidtGit/Heritability

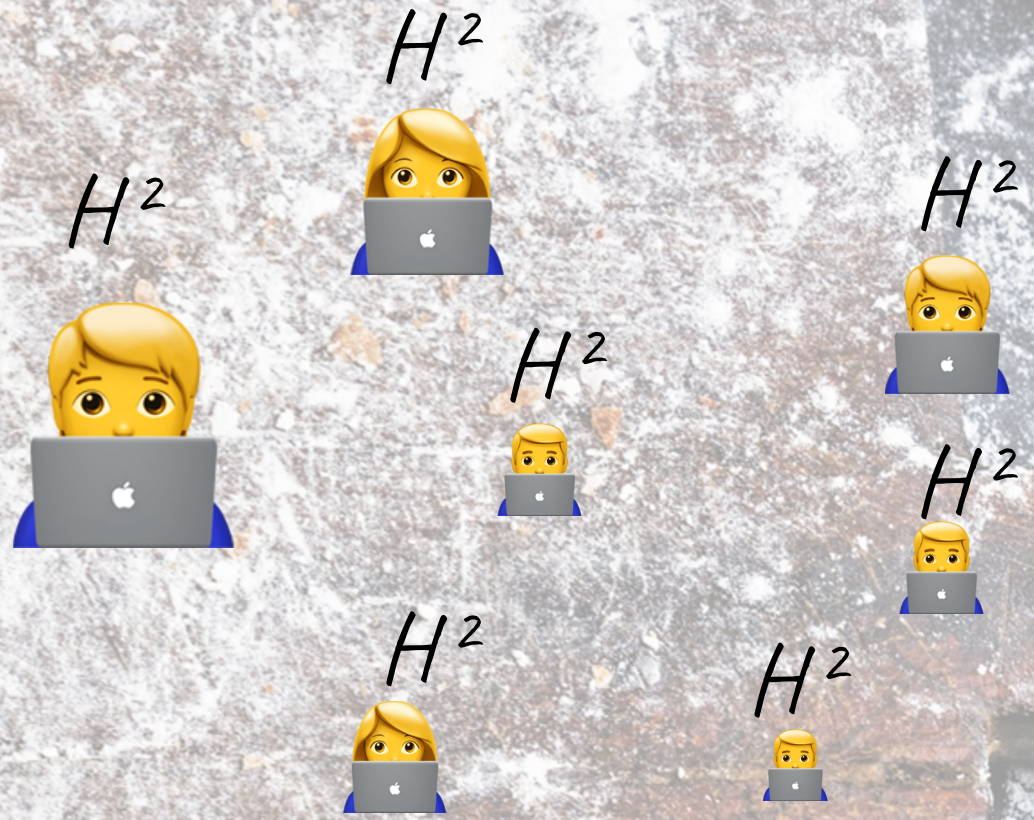
Motivations

1. Code exists but needs TLC
2. Variance components are hard to extract from model objects
3. Unclear how we are arriving on final values

```r
PEV_from_lme4 <- function(model) {
  vc <- lme4::VarCorr(model)
  ngrps <- lme4::ngrps(model)
  # Note the index and kronecker order needs to be followed careful downstream
  Glist <- lapply(names(vc), function(agrp) {
    Matrix::kronecker(vc[[agrp]], diag(ngrps[[agrp]]))
  })
  G <- do.call(Matrix::bdiag, Glist)

  n <- nrow(model@frame)
  R <- diag(n) * stats::sigma(model)^2

  X <- as.matrix(lme4::getME(model, "X"))
  Z <- as.matrix(lme4::getME(model, "Z"))

  C11 <- t(X) %*% solve(R) %*% X
  C12 <- t(X) %*% solve(R) %*% Z
  C21 <- t(Z) %*% solve(R) %*% X
  C22 <- t(Z) %*% solve(R) %*% Z + solve(G)

  C <- rbind(
    cbind(C11, C12),
    cbind(C21, C22)
  )
  solve(C)
}
```

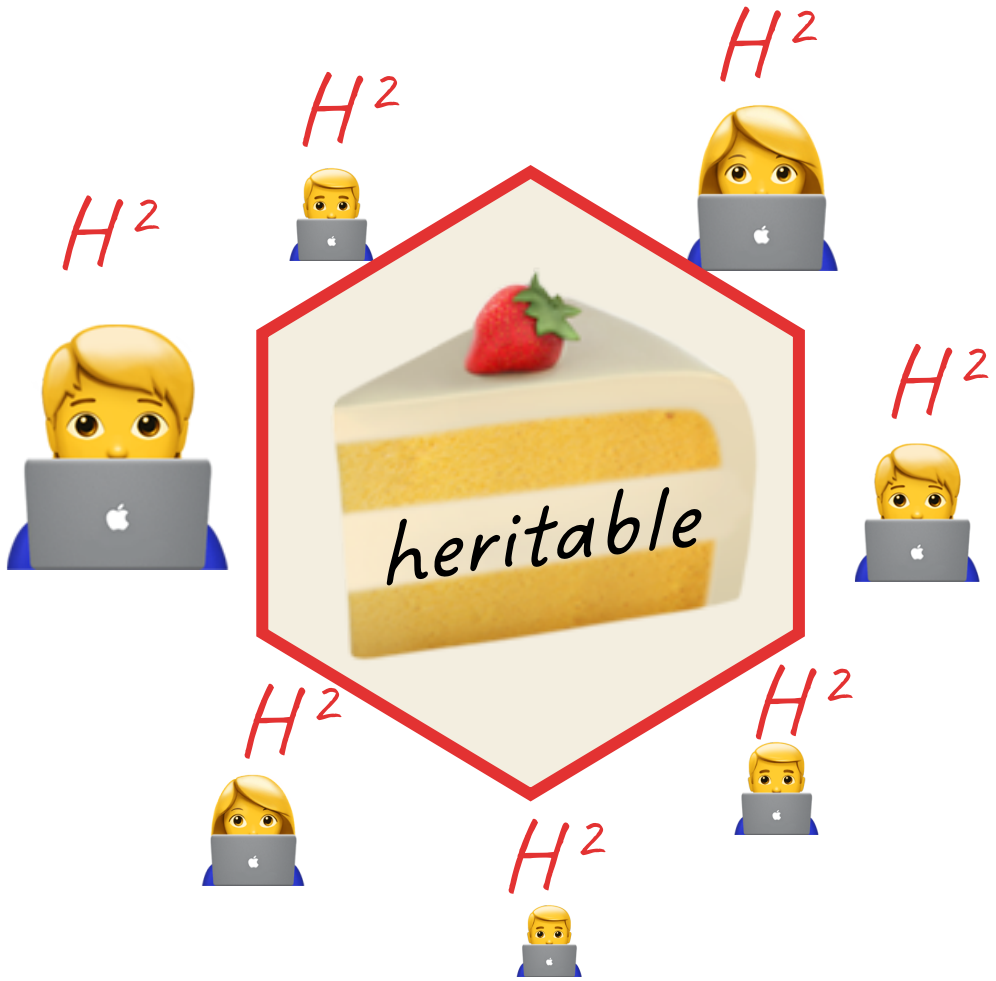
Motivations

1. Code exists but needs TLC
2. Variance components are hard to extract from model objects
3. Unclear how we are arriving on final values 🎂



Our solution

- One-stop shop for 5 methods
- Transparent + reproducible
- Supports models from `lme4` + `asreml`
- Methods are validated



```r
# devtools::install_github("anu-aagi/heritable")
library(heritable)
```

- Works for **single field environment** only (no G x E)

your genotype variable

which heritability method you want

Broad-sense

```r
H2(model, target = "gen", method = c("Standard", "Cullis", ...))
```

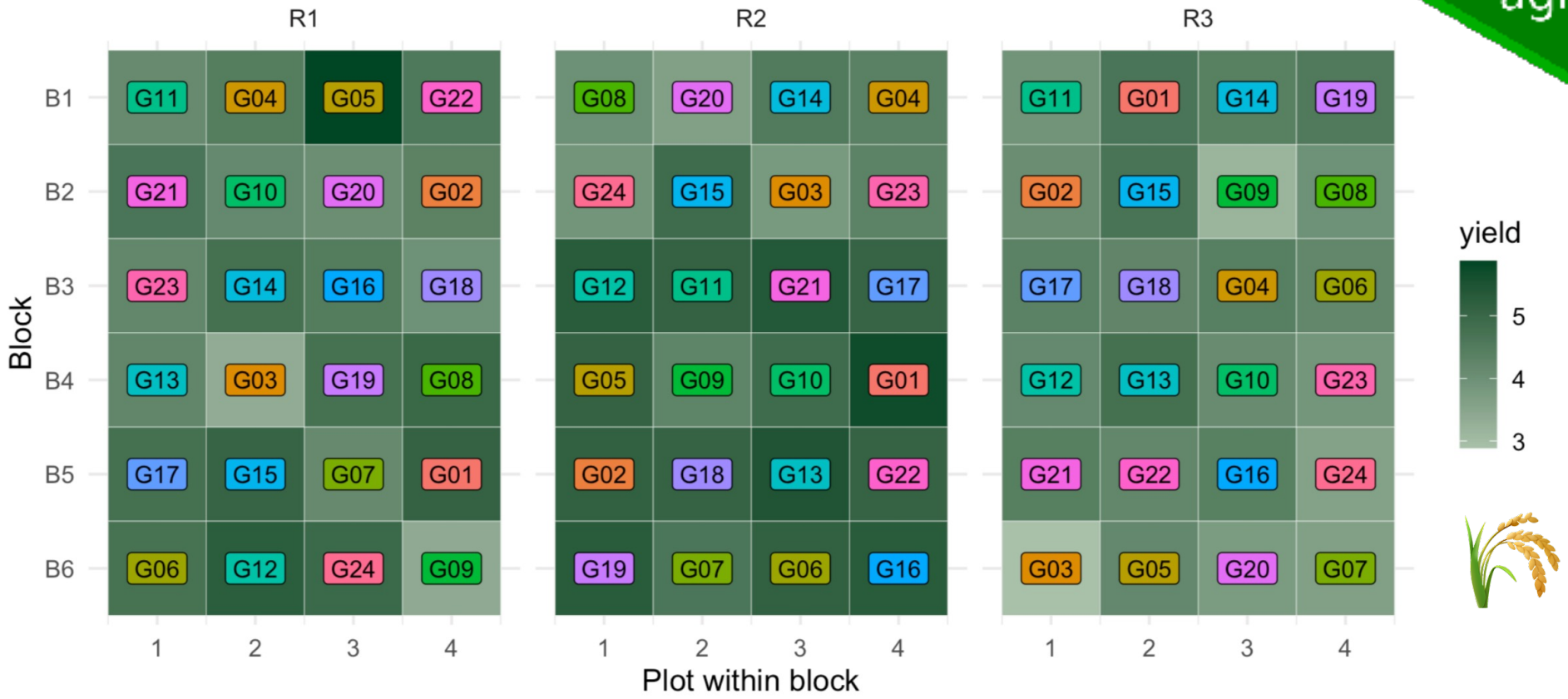
- Broad-sense assumes $\mathbf{G}_g = \sigma_g^2 \mathbf{I}_{n_g}$

Narrow-sense

```r
h2(model, target = "gen", method = c("Standard", "Cullis", ...))
```

Demo : a simple, incomplete case study

```
1 library(agridat)
2 john.alpha
```



- Genotypes not fully replicated across all blocks

The same model with different software

asreml

```
1 library(asreml)
2 fit_asreml <- asreml(yield ~ rep,
3                      random =~ gen + rep:block,
4                      data = agridat::john.alpha)
```

lme4

```
1 library(lme4)
2 fit_lme4 <- lmer(yield ~ rep + (1|gen) + (1|rep:block),
3                  data = agridat::john.alpha)
```

- `heritable` supports both `asreml` and `lme4` model objects

Standard

```
 1 library(heritable)
 2 H2_Standard(fit_asreml, target = "gen")
```

```
[1] 0.8400648
```

```
 1 H2_Standard(fit_lme4, target = "gen")
```

```
[1] 0.8400678
```

$$H^2_{Standard} = \frac{\sigma^2_g}{\sigma^2_g + \frac{1}{n_g}\sum^{i=1}_{n_g} + \sigma^2_p / n_{gi}}$$

number of genotype replicates

accounts of unbalanced-ness

number of genotypes

Lush, J. L. (1949)
Falconer, D. S., & Mackay, T. F. C. (2009)

Cullis'

```
 2 H2_Cullis(fit_asreml, target = "gen")
[1] 0.8090841
 1 H2_Cullis(fit_lme4, target = "gen")
[1] 0.8091338
```

- 👍 Recommended by CGIAR Excellence in Breeding Program

$$H^2_{Cullis} = 1 - \frac{\overline{PEV^{BLUP}_{\Delta..}}}{2\sigma^2_g}$$

mean of the variance difference of two BLUPS

Cullis, B. R., Smith, A. B., & Coombes, N. E. (2006)
Covarrubias-Pazaran (2019)

Differences between genotypes

```
1 H2_Delta_pairwise(fit_asreml, target = "gen", type = "BLUP")
```

	G01	G02	G03	G04	G05	G06	G07
G01	NA	0.8037763	0.8048406	0.8048406	0.8176590	0.8048318	0.8176594
G02	0.8037763	NA	0.8048406	0.8046516	0.8048318	0.8030883	0.8029864
G03	0.8048406	0.8048406	NA	0.8037763	0.8159703	0.8039318	0.8174866
G04	0.8048406	0.8046516	0.8037763	NA	0.8167590	0.8159703	0.8047864
G05	0.8176590	0.8048318	0.8159703	0.8167590	NA	0.8039406	0.8168145

- 👍 Accounts for genotypic covariances

$$H^2_{\Delta ij} = 1 - \frac{PEV^{BLUP}_{\Delta ij}}{2\sigma^2_g}$$

Schmidt, P., Hartung, J., Bennewitz, J., & Piepho, H.-P. (2019).

Genotype-specific

```
 1 H2_Delta_by_genotype(fit_asreml, target = "gen", type = "BLUP")
```

```
$G01
        H2D_i
G01 0.8090555

$G02
        H2D_i
G02 0.8090555
```

- Average of the **pairwise differences** for genotype *i*

$$H^2_{\Delta i.} = 1 - \frac{PEV^{BLUP}_{\Delta i.}}{2\sigma^2_g}$$

Schmidt, P., Hartung, J., Bennewitz, J., & Piepho, H.-P. (2019).

Overall H2 from differences

```
 1 H2_Delta(fit_asreml, target = "gen")
[1] 0.8090841
 1 H2_Delta(fit_lme4, target = "gen")
[1] 0.8091338
```

- Average **across all pairwise combinations*** *Equivalent to Cullis

$$H^2_{\Delta..} = 1 - \frac{PEV^{BLUP}_{\Delta..}}{2\sigma^2_g}$$

Schmidt, P., Hartung, J., Bennewitz, J., & Piepho, H.-P. (2019).

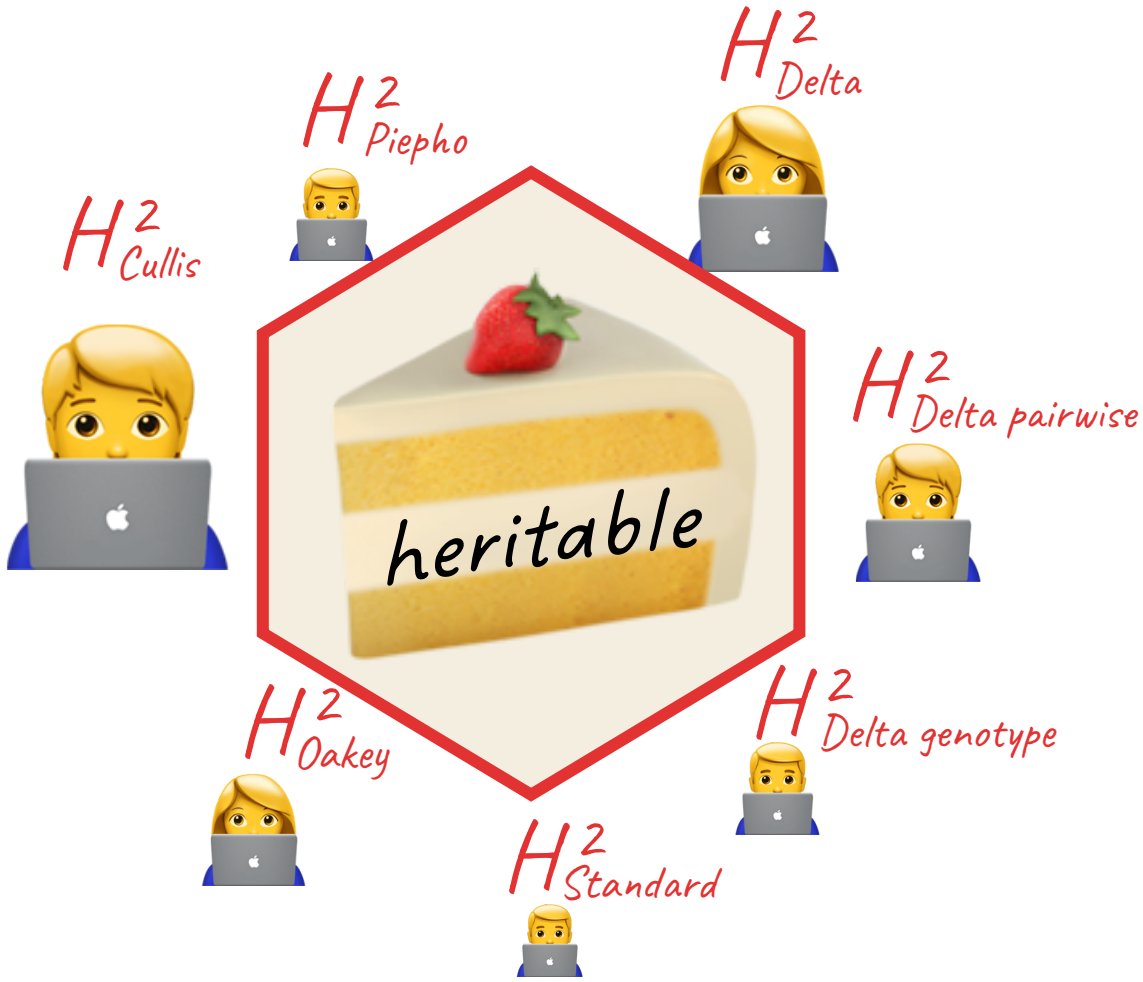
Underlying computation engine matters

```
 1 library(tidyverse)
 2 tibble(model = list(fit_lme4, fit_asreml)) |>
 3   mutate(H2 = map(model, ~H2(.x, target = "gen"))) |>
 4   unnest_wider(H2)
```

```
# A tibble: 2 × 6
  model     Cullis Oakey Piepho Delta Standard
  <list>     <dbl> <dbl>  <dbl> <dbl>    <dbl>
1 <lmerMod>  0.809 0.809  0.797 0.809    0.840
2 <asreml>   0.809 0.809  0.803 0.809    0.840
```

Differences in H2 / h2

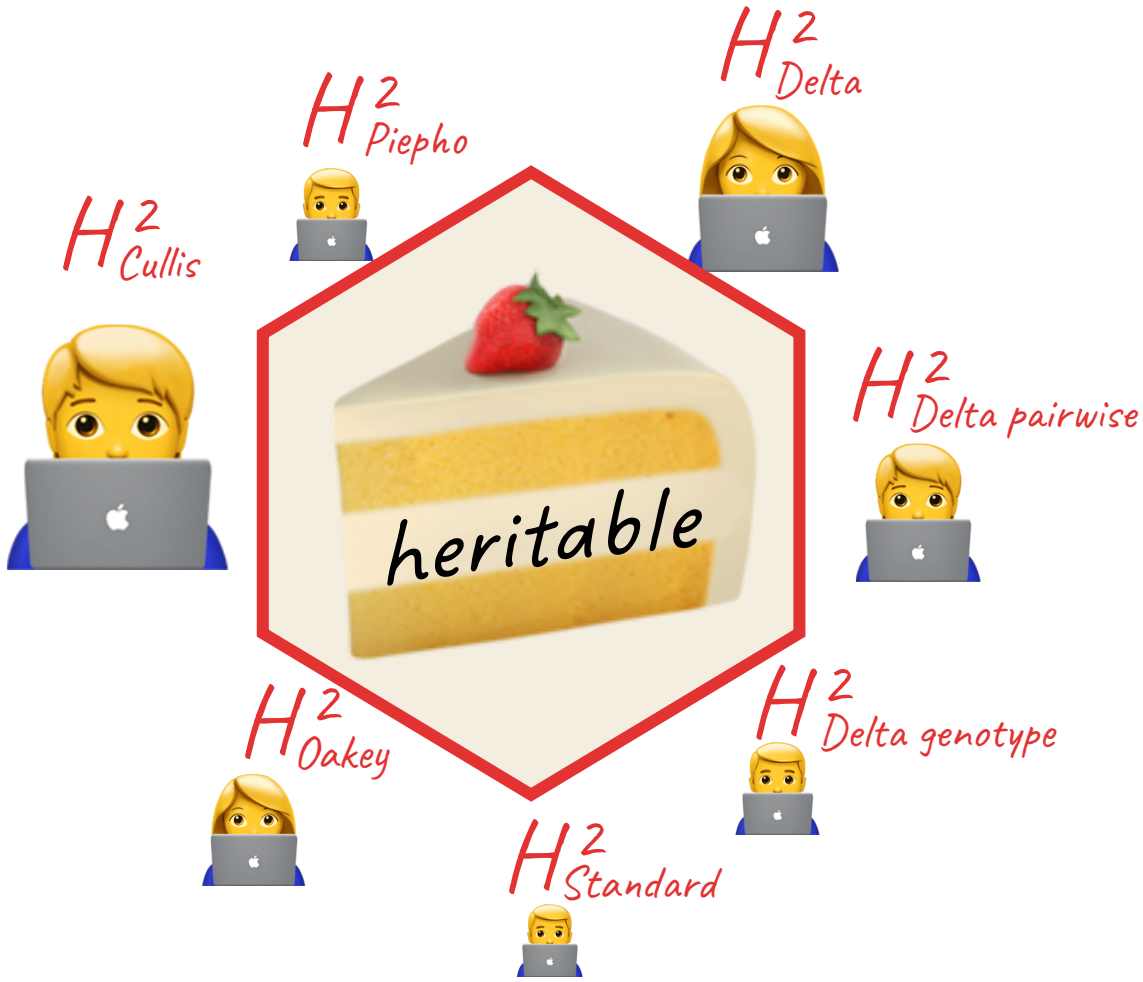
- Biology
- Statistical choice
- Computation engine



Standardise our workflow
Improve our reporting!

What's next

- Multiple environments
- Confidence intervals
- Simulation study

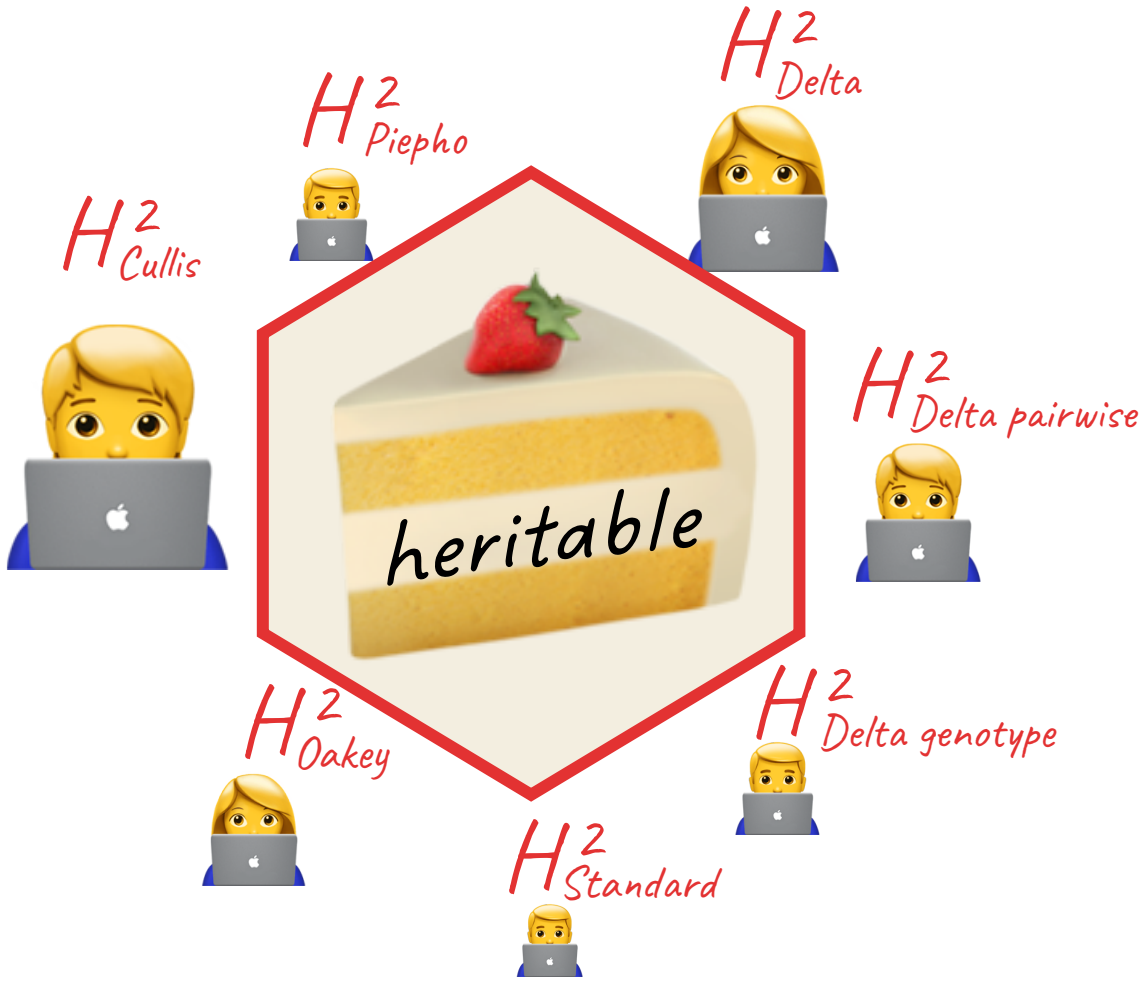


https://github.com/anu-aagi/heritable

lifecycle experimental

Thank you folks!

- One-stop shop for 5 methods
- Transparent + reproducible
- Supports models from `lme4` + `asreml`
- Methods are validated



https://github.com/anu-aagi/heritable

lifecycle experimental

Appendix*

- PEV = C_{11}

$$\underbrace{\begin{bmatrix} \mathbf{X}^\top\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}^\top\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}^\top\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}^\top\mathbf{R}^{-1}\mathbf{Z}+\mathbf{G}^{-1} \end{bmatrix}}_{\mathbf{C}} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \tilde{\boldsymbol{b}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^\top\mathbf{R}^{-1}\boldsymbol{y} \\ \mathbf{Z}^\top\mathbf{R}^{-1}\boldsymbol{y} \end{bmatrix}$$

Variance of prediction errors

$$\operatorname{var}\left(\begin{bmatrix} \hat{\boldsymbol{\beta}}-\boldsymbol{\beta} \\ \tilde{\boldsymbol{b}}-\boldsymbol{b} \end{bmatrix}\right) = \mathbf{C}^{-1} = \begin{bmatrix} (\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X})^{-1} & -(\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{Z}\mathbf{G} \\ -\mathbf{G}\mathbf{Z}^\top\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X})^{-1} & \mathbf{G}-\mathbf{G}\mathbf{Z}^\top\mathbf{P}\mathbf{Z}\mathbf{G} \end{bmatrix}$$

*details I am still learning and brushed over

Piepho's

```
 1 H2_Piepho(fit_asreml, target = "gen")
[1] 0.802976
 1 H2_Piepho(fit_lme4, target = "gen")
[1] 0.7966375
```

- Requires ad-hoc fitting genotype as a fixed effect
- 👎 Covariance of genotypes are not accounted for

$$H^2_{Piepho} = \frac{\sigma^2_g}{\sigma^2_g + \overline{PEV_g^{BLUE}}/2}$$

Piepho, H.-P., & Möhring, J. (2007).

Oakey's

```
 2 H2_Oakey(fit_asreml, target = "gen")
[1] 0.8090841
 1 H2_Oakey(fit_lme4, target = "gen")
[1] 0.8091338
```

- Eigen value method after maximisation of variance components
- 👍 Account of genetic covariances 👷

$$H^2_{Oakey} = \frac{\sum_{i=n_z+1}^{n_g} \lambda_i}{n_g - n_z}$$

Oakey, H., Verbyla, A., Pitchford, W., Cullis, B., & Kuchel, H. (2006).

p-rep
20 iterations
Fixed residuals