**Table S1:** Estimates of direct and indirect selection via educational attainment based on separate structural equation models for each of 13 anthropometric traits, corresponding to Fig. 4 in the main text. An individual’s rural/urban origin, number of siblings, and parental socioeconomic position (SEP) are included as exogenous covariates. It is assumed that each trait and covariate can influence lifetime reproductive success either directly or indirectly, as mediated by educational attainment. Robust standard errors were calculated using weighted least squares with mean- and variance-adjusted test statistics (option ‘WLSMV’ in the ‘lavaan’ package).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Direct selection | SE | *p*-value | Indirect selection | SE | *p­*-value | Sample size |
| Height | **−0.018** | **0.008** | **0.03** | **−0.007** | **0.001** | **<10-3** | 6451 |
| Weight | **−0.024** | **0.008** | **0.001** | **−0.004** | **0.001** | **<10-3** | 6182 |
| BMI | **−0.026** | **0.008** | **0.0006** | −0.001 | 0.001 | 0.15 | 6181 |
| Face width | −0.012 | 0.008 | 0.12 | **−0.003** | **0.001** | **<10-3** | 6449 |
| Jaw width | **−0.026** | **0.010** | **0.008** | **−0.003** | **0.001** | **0.003** | 4547 |
| Face roundness (FWHR) | 0.015 | 0.008 | 0.06 | −0.0004 | 0.0008 | 0.61 | 6439 |
| Cranial volume | −0.010 | 0.008 | 0.21 | **−0.007** | **0.001** | **<10-3** | 6449 |
| Shoulder width | −0.010 | 0.008 | 0.18 | **−0.002** | **0.0009** | **0.003** | 6449 |
| Hip width | **−0.022** | **0.008** | **0.005** | **−0.005** | **0.001** | **<10-3** | 6437 |
| Shoulder/hip ratio | 0.015 | 0.008 | 0.055 | **0.003** | **0.0009** | **0.001** | 6442 |
| Breast dev. rate | −0.015 | 0.010 | 0.13 | 0.00 | 0.001 | 0.68 | 4028 |
| Hand grip strength | 0.000 | 0.009 | 0.96 | **−0.005** | **0.001** | **<10-4** | 5155 |
| Lung capacity | −0.003 | 0.009 | 0.74 | **−0.005** | **0.001** | **<10-4** | 4922 |

**Table S2:** Estimates of direct and indirect selection via educational attainment based on a single structural equation model involving 7 anthropometric traits (*n* = 3555), corresponding to Fig. 5 in the main text. An individual’s rural/urban origin, number of siblings, and parental socioeconomic position (SEP) are included as exogenous covariates. It is assumed that each trait and covariate can influence lifetime reproductive success either directly or indirectly, as mediated by educational attainment. Robust standard errors were calculated using weighted least squares with mean- and variance-adjusted test statistics (option ‘WLSMV’ in the ‘lavaan’ package).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Direct selection | SE | *p*-value | Indirect selection | SE | *p*-value |
| Height | −0.014 | 0.012 | 0.26 | **−0.004** | **0.001** | **0.003** |
| BMI | **−0.023** | **0.010** | **0.02** | 0.001 | 0.001 | 0.64 |
| Face roundness (lFWHR) | 0.016 | 0.010 | 0.11 | **−**0.001 | 0.001 | 0.41 |
| Shoulder/hip ratio | 0.019 | 0.010 | 0.052 | **0.002** | **0.001** | **0.045** |
| Breast dev. rate | 0.001 | 0.011 | 0.95 | 0.002 | 0.001 | 0.14 |
| Hand grip strength | 0.012 | 0.011 | 0.24 | **−0.003** | **0.001** | **0.03** |
| Lung capacity | 0.013 | 0.011 | 0.25 | **−0.03** | **0.001** | **0.02** |

**Figure S1:** Path diagram for structural equation models of the effects of traits on lifetime reproductive success, including indirect effects via education and the age at first birth (AFB). The model allows for a direct effect of traits on fitness (red), as well as three types of indirect effect, acting via the following pathways:

1. via education only (black)
2. via AFB only(blue)
3. via education followed by AFB (yellow)

Single-headed arrows represent potential causal effects, whereas double-headed arrows represent covariances that are not analysed causally. An individual’s rural/urban origin, number of siblings, and parental socioeconomic position (SEP) are included as covariates. Note that this model excludes women with zero lifetime reproductive success.

**Table S3:** Estimates of direct and indirect selection via educational attainment and the age at first birth (AFB), based on the structural equation model summarized in Fig. S1 (*n* = 3256). An individual’s AFB was standardized as the absolute deviation (in years) from the mean age at first birth for each year cohort. Note that this model excludes women with zero lifetime reproductive success (299 women were excluded for this reason). Robust standard errors were calculated using weighted least squares with mean- and variance-adjusted test statistics (option ‘WLSMV’ in the ‘lavaan’ package).As expected, AFB has a strong and significant negative effect on fitness, even after controlling for measured traits and other covariates (the estimated path coefficient from AFB to LRS is , ). Selection via AFB accounts for a modest to substantial proportion of total selection, depending on the trait.

trait(s)

education

AFB

covariates

LRS

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Pathway | Selection | SE | *p*-value |
| Height | Direct | −0.0054 | 0.011 | 0.62 |
| Via education only | −0.0005 | 0.0006 | 0.45 |
| Via AFB only | −0.0034 | 0.0035 | 0.33 |
| **Via education then AFB** | **−0.0026** | **0.0009** | **0.004** |
| BMI | Direct | −0.0123 | 0.0085 | 0.15 |
| Via education only | 0.0000 | 0.0002 | 0.91 |
| Via AFB only | 0.0011 | 0.0031 | 0.73 |
| Via education then AFB | 0.0001 | 0.0008 | 0.91 |
| Face roundness (lFWHR) | Direct | 0.0103 | 0.0090 | 0.25 |
| Via education only | −0.0001 | 0.0002 | 0.66 |
| Via AFB only | 0.0046 | 0.0028 | 0.10 |
| Via education then AFB | 0.0004 | 0.0007 | 0.60 |
| Shoulder/hip ratio | Direct | 0.0060 | 0.0091 | 0.51 |
| Via education only | 0.0001 | 0.0003 | 0.47 |
| Via AFB only | 0.0051 | 0.0026 | 0.053 |
| Via education then AFB | 0.0012 | 0.0007 | 0.10 |
| Breast dev. rate | Direct | −0.0013 | 0.0100 | 0.90 |
| Via education only | 0.0001 | 0.0002 | 0.53 |
| Via AFB only | 0.0020 | 0.0031 | 0.52 |
| Via education then AFB | 0.0008 | 0.0008 | 0.34 |
| Hand grip strength | Direct | 0.0025 | 0.0096 | 0.79 |
| Via education only | −0.0003 | 0.0004 | 0.47 |
| Via AFB only | 0.0011 | 0.0030 | 0.72 |
| **Via education then AFB** | **−0.0015** | **0.0008** | **0.04** |
| Lung capacity | Direct | 0.0062 | 0.0103 | 0.55 |
| Via education only | −0.0004 | 0.0005 | 0.45 |
| Via AFB only | 0.0009 | 0.0031 | 0.79 |
| **Via education then AFB** | **−0.0021** | **0.0008** | **0.01** |

**Table S4:** Standardized selection differentials for 13 anthropometric traits, after excluding all but one girl from each group of known relatives (cf. main analysis in Table 1 of the main text). Results shown in bold remain significant after Holm-Bonferroni correction for multiple comparisons.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Total selection (*d*) | *p*-value | Directional selection (*s*) | *p*-value | Non-directional selection (*dN*) | *p*-value | Sample size |
| Height | **0.037** | **< 10-4** | **−0.037** | **< 10-4** | 0.000 | 0.43 | 9454 |
| Weight | **0.035** | **< 10-4** | **−0.035** | **< 10-4** | 0.000 | 0.42 | 9073 |
| BMI | **0.025** | **< 10-3** | **−0.025** | **< 10-3** | 0.000 | 0.29 | 9072 |
| Face width | 0.014 | 0.03 | −0.014 | 0.03 | 0.0002 | 0.25 | 9449 |
| Jaw width | **0.034** | **< 10-4** | **−0.034** | **< 10-4** | 0.000 | 0.91 | 6353 |
| Face roundness (lFWHR) | 0.017 | 0.009 | 0.015 | 0.02 | 0.002 | 0.11 | 9432 |
| Cranial volume | **0.017** | **0.007** | **−0.017** | **0.007** | 0.000 | 0.72 | 9451 |
| Shoulder width | 0.012 | 0.06 | −0.012 | 0.051 | 0.000 | 0.29 | 9448 |
| Hip width | **0.030** | **< 10-4** | **−0.030** | **< 10-4** | 0.000 | 0.89 | 9428 |
| Shoulder/hip ratio | **0.021** | **< 10-3** | **0.021** | **< 10-3** | 0.000 | 0.65 | 9421 |
| Breast dev. rate | 0.016 | 0.04 | −0.016 | 0.04 | 0.000 | 0.60 | 6201 |
| Hand grip strength | 0.004 | 0.70 | 0.002 | 0.81 | 0.002 | 0.14 | 7748 |
| Lung capacity | 0.011 | 0.15 | −0.011 | 0.13 | 0.000 | 0.46 | 7314 |

**Table S5:** Standardized linear and quadratic selection gradients for 13 anthropometric traits, after excluding all but one girl from each group of known relatives (cf. main analysis in Table 2 of the main text).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Linear selection gradient ) | SE | *p*-value | Quadratic selection gradients , | SE | *p*-value | Sample size |
| Height | −0.015 | 0.008 | 0.08 | −0.014  −0.021 | 0.009  0.011 | 0.11  0.052 | 6042 |
| Weight | **−0.024** | **0.008** | **0.003** | −0.017  −0.016 | 0.009  0.008 | 0.056  0.07 | 5802 |
| BMI | **−0.027** | **0.008** | **<10-3** | −0.017  **−0.023** | 0.009  **0.009** | 0.07  **0.007** | 5801 |
| Face width | −0.003 | 0.008 | 0.73 | −0.001  −0.019 | 0.008  0.012 | 0.92  0.11 | 6040 |
| Jaw width | **−0.021** | **0.010** | **0.03** | **−0.021**  −0.007 | **0.010**  0.014 | **0.04**  0.60 | 4240 |
| Face roundness (lFWHR) | **0.023** | **0.008** | **0.006** | **0.025**  **−0.037** | **0.008**  **0.012** | **0.003**  **0.001** | 6030 |
| Cranial volume | −0.004 | 0.008 | 0.65 | −0.004  0.006 | 0.008  0.011 | 0.60  0.58 | 6040 |
| Shoulder width | −0.012 | 0.008 | 0.14 | −0.013  −0.018 | 0.008  0.009 | 0.11  0.06 | 6041 |
| Hip width | **−0.023** | **0.008** | **0.005** | **−0.021**  −0.014 | **0.008**  0.011 | **0.01**  0.22 | 6028 |
| Shoulder/hip ratio | 0.014 | 0.008 | 0.09 | 0.014  0.001 | 0.008  0.009 | 0.09  0.92 | 6026 |
| Breast dev. rate | −0.014 | 0.010 | 0.15 | −0.014  0.012 | 0.010  0.012 | 0.17  0.32 | 3783 |
| Hand grip strength | 0.005 | 0.009 | 0.56 | 0.006  −0.005 | 0.009  0.011 | 0.51  0.67 | 4822 |
| Lung capacity | −0.001 | 0.009 | 0.95 | −0.001  0.004 | 0.010  0.011 | 0.88  0.69 | 4600 |

**Table S6:** Estimates of direct and indirect selection via educational attainment based on separate structural equation models for each of 13 anthropometric traits, after excluding all but one girl from each group of known relatives (cf. main analysis in Table S1).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Direct selection | SE | *p*-value | Indirect selection | SE | *p­*-value | Sample size |
| Height | −0.012 | 0.009 | 0.16 | **−0.009** | **0.002** | **<10-5** | 6044 |
| Weight | **−0.022** | **0.009** | **0.01** | **−0.006** | **0.001** | **<10-4** | 5804 |
| BMI | **−0.027** | **0.009** | **0.002** | −0.002 | 0.001 | 0.12 | 5803 |
| Face width | −0.002 | 0.008 | 0.82 | **−0.006** | **0.001** | **<10-4** | 6042 |
| Jaw width | **−0.021** | **0.010** | **0.04** | **−0.005** | **0.001** | **0.001** | 4242 |
| Face roundness (FWHR) | **0.022** | **0.008** | **0.009** | −0.001 | 0.001 | 0.33 | 6032 |
| Cranial volume | −0.001 | 0.008 | 0.92 | **−0.011** | **0.002** | **<10-5** | 6042 |
| Shoulder width | −0.010 | 0.009 | 0.22 | **−0.004** | **0.001** | **0.003** | 6043 |
| Hip width | **−0.021** | **0.008** | **0.01** | **−0.007** | **0.001** | **<10-5** | 6030 |
| Shoulder/hip ratio | 0.013 | 0.008 | 0.11 | **0.004** | **0.001** | **0.001** | 6028 |
| Breast dev. rate | −0.014 | 0.009 | 0.15 | 0.001 | 0.001 | 0.48 | 3785 |
| Hand grip strength | 0.007 | 0.009 | 0.43 | **−0.008** | **0.002** | **<10-5** | 4824 |
| Lung capacity | 0.001 | 0.010 | 0.88 | **−0.007** | **0.002** | **<10-4** | 4602 |

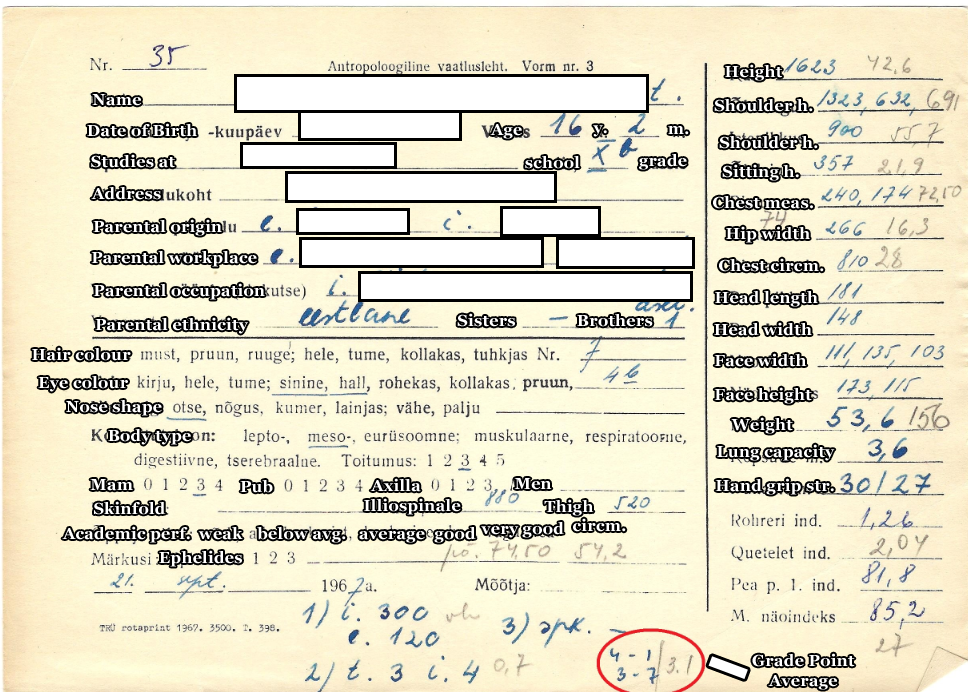
**Table S7:** Estimates of direct and indirect selection via educational attainment based on a single structural equation model involving 7 anthropometric traits (*n* = 3555), after excluding all but one girl from each group of known relatives (cf. main analysis in Table S2).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Direct selection | SE | *p*-value | Indirect selection | SE | *p*-value |
| Height | −0.008 | 0.013 | 0.50 | **−0.005** | **0.002** | **0.005** |
| BMI | **−0.025** | **0.012** | **0.03** | 0.001 | 0.001 | 0.70 |
| Face roundness (lFWHR) | **0.025** | **0.011** | **0.02** | **−**0.001 | 0.001 | 0.43 |
| Shoulder/hip ratio | 0.018 | 0.011 | 0.09 | 0.003 | 0.001 | 0.051 |
| Breast dev. rate | −0.001 | 0.012 | 0.94 | 0.002 | 0.002 | 0.29 |
| Hand grip strength | 0.015 | 0.011 | 0.17 | **−0.003** | **0.002** | **0.03** |
| Lung capacity | 0.016 | 0.012 | 0.18 | **−0.03** | **0.002** | **0.03** |

**Fig. S2:** Individual fitness curves for 13 anthropometric traits, showing the estimated relationships between standardized residuals of trait values and expected relative lifetime reproductive success (mean SEM), after excluding all but one girl from each group of known relatives (cf. main analysis in Fig. 3 of the main text).

****

**Fig. S3.** Example of Juhan Aul’s datasheet for recording anthropometric data.



# This R script relates to the paper:

# Natural selection on anthropometric traits of Estonian girls.

# It requires the dataset 'EstonianGirlsData.csv'.

# We quantified selection on 13 anthropometric traits in a sample of 4000 - 10000 Estonian girls.

# The traits are as follows (see main text for details):

height

weight

BMI # body mass index

face\_width

jaw\_width

face\_roundness # (lFWHR, lower face width/height ratio)

cranial\_vol # cranial volume

shoulder\_width

hip\_width

shoulder\_to\_hip # ratio of shoulder width/hip width

breast\_dev # breast development rate

HGS # hand-grip strength

lung\_capacity # vital lung capacity

# Note that the dataset does not contain raw trait values, but rather standardized residuals of raw trait values

# on smooth functions of age and birth date, which were obtained using generalized additive models

# (see main text for details).

# The dataset also contains the following biosocial covariates:

education # educational attainment, coded as 1 = primary (<=8 years of schooling), 2 = secondary, 3 = tertiary

rural # coded as 1 = rural origin, 0 = urban origin

n\_siblings # number of siblings in the individual's family

SEP # parental socioeconomic position, coded as 1 = unskilled manual workers, 2 = skilled manual workers, 3 = non-manual works (highest value of mother or father was taken)

## Main code

# This code requires the packages 'mgcv' and 'lavaan'

library(mgcv)

library(lavaan)

EstonianGirlsData$education <- factor(EstonianGirlsData$education,ordered=TRUE,levels=c(1,2,3))

EstonianGirlsData$SEP <- factor(EstonianGirlsData$SEP,ordered=TRUE,levels=c(1,2,3))

## Section 1

# calculating overall selection (Table 1)

# BEFORE you run this code, please load the functions at the end of this script (search for ### to find them)

# example here for trait 'height' (change as necessary)

z<-EstonianGirlsData$height

w<-EstonianGirlsData$rel\_LRS

# remove all individuals where either the trait value or relative LRS has value 'NA'

combined<-cbind(z,w)

combined<-combined[which(is.na(combined[,1])==FALSE & is.na(combined[,2])==FALSE),]

z<-combined[,1]

w<-combined[,2]

# this function returns values of:

# the distributional selection differential, DSD (which measures total selection)

# the linear selection differential, s (which measures directional selection)

# the non-directional component of the DSD (which measures non-directional selection)

# the values of the DSD and dN are based on smoothed estimates of expected relative LRS (see main text)

fitted.DSD(z,w)

# returns p-values for the above statistics based on exact permutation tests

# depending on your machine this step may take a while

DSD.permutation.test(z,w,bootlength=10000)

## Section 2

# Lande-Arnold linear and quadratic selection gradients (Table 2)

# linear model to calculate linear selection gradient 'beta' on the trait 'height' (change trait as necessary)

# with education, rural/urban environment, number of siblings as parental SEP as covariates

linear.model <- lm(rel\_LRS ~ height + education + rural + n\_siblings + SEP,

data=EstonianGirlsData)

summary(linear.model)

# quadratic model to calculate quadratic selection gradients 'beta\_Q' and 'gamma' on the trait 'height' (change trait as necessary)

# with education, rural/urban environment, number of siblings as parental SEP as covariates

# note that 'gamma' is twice the regression coefficient for the squared term in this model (Stinchcombe et al. 2008)

quadratic.model <- lm(rel\_LRS ~ height + I(height^2) + education + rural + n\_siblings + SEP,

data=EstonianGirlsData)

summary(quadratic.model)

## Section 3

# individual fitness curves (Fig. 3)

# generalized additive model (GAM) including a smoothed function of the trait 'height' (change trait as necessary)

# with education, rural/urban environment, number of siblings as parental SEP as covariates

gam.model <- gam(rel\_LRS ~ s(height) + education + rural + n\_siblings + SEP,

family = gaussian,

data=EstonianGirlsData)

# plot the individual fitness curve (mean +/- SEM)

plot.gam(gam.model,rug=FALSE,shade=TRUE,seWithMean=TRUE,xlim=c(-3.5,3.5),shift=1,xlab="trait",ylab="fitness curve")

# quantitative summary of the model

summary(gam.model)

## Section 4

# structural equation models (Fig. 4)

# the first model includes only one anthropometric trait

# example here is for the trait 'height' (replace as necessary)

# the variables 'direct' and 'indirect' represent direct selection via the path height -> rel\_LRS

# and indirect selection via the path rel\_LRS -> education -> rel\_LRS

SEM.model <- 'rel\_LRS ~ beta31\*height + beta32\*education + rural + n\_siblings + SEP2dummy + SEP3dummy

education ~ beta21\*height + rural + n\_siblings + SEP2dummy + SEP3dummy

direct := beta31

indirect := beta21\*beta32'

# fit the SEM model

SEM.fit <- sem(SEM.model,

data=EstonianGirlsData,

estimator="WLSMV")

#summary of the SEM model

summary(SEM.fit, nd=5)

# the next model includes 7 anthropometric traits simultaneously

AFB.SEM.model <- 'rel\_LRS ~ betaW1\*height + betaW2\*BMI + betaW5\*face\_roundness + betaW6\*cranial\_vol + betaW9\*shoulder\_to\_hip + betaW10\*breast\_dev + betaW11\*HGS + betaW12\*lung\_capacity+ betaWE\*education + rural + n\_siblings + SEP2dummy + SEP3dummy

education ~ betaE1\*height + betaE2\*BMI + betaE5\*face\_roundness + betaE6\*cranial\_vol + betaE9\*shoulder\_to\_hip + betaE10\*breast\_dev + betaE11\*HGS + betaE12\*lung\_capacity + rural + n\_siblings + SEP2dummy + SEP3dummy

direct\_height := betaW1

indirect\_height := betaWE\*betaE1

direct\_BMI := betaW2

indirect\_BMI := betaWE\*betaE2

direct\_fround := betaW5

indirect\_fround := betaWE\*betaE5

direct\_cran := betaW6

indirect\_cran := betaWE\*betaE6

direct\_SHratio:= betaW9

indirect\_SHratio := betaWE\*betaE9

direct\_breast:= betaW10

indirect\_breast := betaWE\*betaE10

direct\_HGS := betaW11

indirect\_HGS := betaWE\*betaE11

direct\_lung := betaW12

indirect\_lung := betaWE\*betaE12'

AFB.fit <- sem(AFB.SEM.model,

data=EstonianGirlsDataAFB,

estimator="WLSMV")

#summary of the SEM model

summary(AFB.fit, nd=5)

# lastly, this model includes 7 anthropometric traits simultaneously,

# as well as the age at first birth as a variable mediating the fitness

# effects of traits and educations

AFB.SEM.model <- 'rel\_LRS ~ betaW1\*height + betaW2\*BMI + betaW5\*face\_roundness + betaW6\*cranial\_vol + betaW9\*shoulder\_to\_hip + betaW10\*breast\_dev + betaW11\*HGS + betaW12\*lung\_capacity+ betaWE\*education + betaWA\*AFB + rural + n\_siblings + SEP2dummy + SEP3dummy

AFB ~ betaAE\*education + betaA1\*height + betaA2\*BMI + betaA5\*face\_roundness + betaA6\*cranial\_vol + betaA9\*shoulder\_to\_hip + betaA10\*breast\_dev + betaA11\*HGS + betaA12\*lung\_capacity + rural + n\_siblings + SEP2dummy + SEP3dummy

education ~ betaE1\*height + betaE2\*BMI + betaE5\*face\_roundness + betaE6\*cranial\_vol + betaE9\*shoulder\_to\_hip + betaE10\*breast\_dev + betaE11\*HGS + betaE12\*lung\_capacity + rural + n\_siblings + SEP2dummy + SEP3dummy

height\_direct := betaW1

height\_E := betaWE\*betaE1

height\_A := betaWA\*betaA1

height\_EA := betaWA\*betaAE\*betaE1

BMI\_direct := betaW2

BMI\_E := betaWE\*betaE2

BMI\_A := betaWA\*betaA2

BMI\_EA := betaWA\*betaAE\*betaE2

fround\_direct := betaW5

fround\_E := betaWE\*betaE5

fround\_A := betaWA\*betaA5

fround\_EA := betaWA\*betaAE\*betaE5

SH\_direct := betaW9

SH\_E := betaWE\*betaE9

SH\_A := betaWA\*betaA9

SH\_EA := betaWA\*betaAE\*betaE9

breast\_direct := betaW10

breast\_E := betaWE\*betaE10

breast\_A := betaWA\*betaA10

breast\_EA := betaWA\*betaAE\*betaE10

HGS\_direct := betaW11

HGS\_E := betaWE\*betaE11

HGS\_A := betaWA\*betaA11

HGS\_EA := betaWA\*betaAE\*betaE11

lung\_direct := betaW12

lung\_E := betaWE\*betaE12

lung\_A := betaWA\*betaA12

lung\_EA := betaWA\*betaAE\*betaE12'

AFB.fit <- sem(AFB.SEM.model,

data=EstonianGirlsDataAFB,

estimator="WLSMV")

#summary of the SEM model

summary(AFB.fit, nd=5)

### Section 5

# functions needed to calculate the distributional selection differential

# run all code below this line before starting

DSD<-function(Z,W,standard=FALSE)

{ # Extract dimensions from Z and standardize traits if required

if(is.vector(Z)){

n <- length(Z)

ntraits <- 1

if(standard){Z <- standardize(Z)}

}

else{

n <- dim(Z)[1]

ntraits <- dim(Z)[2]

if(standard){Z<-apply(Z,2,standardize)}

}

# Set up a data frame DSD.vector for the results, with row names given by rows

rows<-if (is.null(colnames(Z))) {paste("z", 1:ntraits, sep = "")} else {colnames(Z)}

DSD.vector<-data.frame(DSD=vector(length=ntraits),row.names=rows)

# Calculate the DSD for each trait Zi

for (i in c(1:ntraits)){

if(is.vector(Z)) {Zi <- Z} else {Zi<-Z[,i]}

# Order trait values from smallest to largest, and order fitness correspondingly

or<-order(Zi)

Zi<-Zi[or]

Wi<-W[or]

# Calculate relative fitness

w<-Wi/mean(Wi)

# Calculate the DSD for Zi; the formula here is equivalent to that in Henshaw & Zemel (2017), except with a divisor of (n-1) instead of n to reduce finite-sample bias.

cumulant<-cumsum(1-w)[-n]

DSD.vector[i,]<-(1/(n-1))\*sum(diff(Zi)\*abs(cumulant))

}

DSD.vector

}

fitted.DSD <-function(Z,W,standard=FALSE){

wgam <- gam(W~s(Z), na=na.exclude,family=gaussian)

Wfitted<-abs(fitted(wgam))

# Extract dimensions from Z and standardize traits if required

if(is.vector(Z)){

ntraits <- 1

if(standard){Z <- standardize(Z)}

}

else{

ntraits <- dim(Z)[2]

if(standard){Z<-apply(Z,2,standardize)}

}

# Save values of the DSD

DSDtemp=DSD(Z,Wfitted)$DSD

# Row names for the results

rows<-if (is.null(colnames(Z))) {paste("z", 1:ntraits, sep = "")} else {colnames(Z)}

# Return a data frame with values for d, dD, and dN

data.frame(DSD=DSDtemp,s=cov(Z,W/mean(W)),dN=DSDtemp-abs(cov(Z,Wfitted/mean(Wfitted))),row.names=rows)

}

DSD.permutation.test<-function(Z,W,standard=FALSE,bootlength=10000)

{

# Extract dimensions from Z and standardize traits if required

if(is.vector(Z)){

ntraits <- 1

if(standard){Z <- standardize(Z)}

}

else{

ntraits <- dim(Z)[2]

if(standard){Z<-apply(Z,2,standardize)}

}

# Row names for the results

rows<-if (is.null(colnames(Z))) {paste("z", 1:ntraits, sep = "")} else {colnames(Z)}

# Set up a data frame for the results

results<-data.frame(DSD=vector(length=ntraits),

pvalue.DSD=vector(length=ntraits),

s=vector(length=ntraits),

pvalue.s=vector(length=ntraits),

dN=vector(length=ntraits),

pvalue.dN=vector(length=ntraits),

row.names=rows)

# Save the observed values of d, dD, dN and s to the results data frame

DSD.components.temp=fitted.DSD(Z,W,standard)

results$DSD<-DSD.components.temp$DSD

results$dN<-DSD.components.temp$dN

results$s<-DSD.components.temp$s

#Set up tables that will later contain values of d, dD, dN and s corresponding to each permutation

boottable.DSD <- matrix(nrow = bootlength, ncol = ntraits)

boottable.dN <- matrix(nrow = bootlength, ncol = ntraits)

boottable.s <- matrix(nrow = bootlength, ncol = ntraits)

# For each permutation, run this loop

for(i in 1:bootlength){ if(i%%10==0){print(i)}

# Permute fitness values

Wtemp<-sample(W)

# Save values of d, dD, dN and s for this permutation to the boottables

DSD.components.temp=fitted.DSD(Z,Wtemp,standard)

boottable.DSD[i,]<-DSD.components.temp$DSD

boottable.dN[i,]<-DSD.components.temp$dN

boottable.s[i,]<-DSD.components.temp$s

}

# Calculate p-values for d, dD, dN and s for each trait as the proportion of boottable values that exceed the observed value (for s, absolute values are used)

for (i in 1:ntraits){

results$pvalue.DSD[i] <- length(boottable.DSD[,i][boottable.DSD[,i]>results$DSD[i]])/bootlength

results$pvalue.dN[i] <- length(boottable.dN[,i][boottable.dN[,i]>results$dN[i]])/bootlength

results$pvalue.s[i] <- length(boottable.s[,i][abs(boottable.s[,i])>abs(results$s[i])])/bootlength

}

results

}