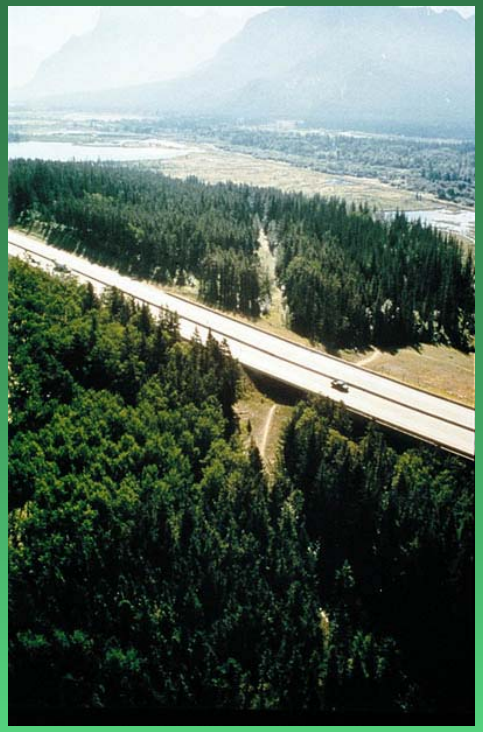
Using genetic differentiation to quantify inter-population movement: A simulation modelling approach

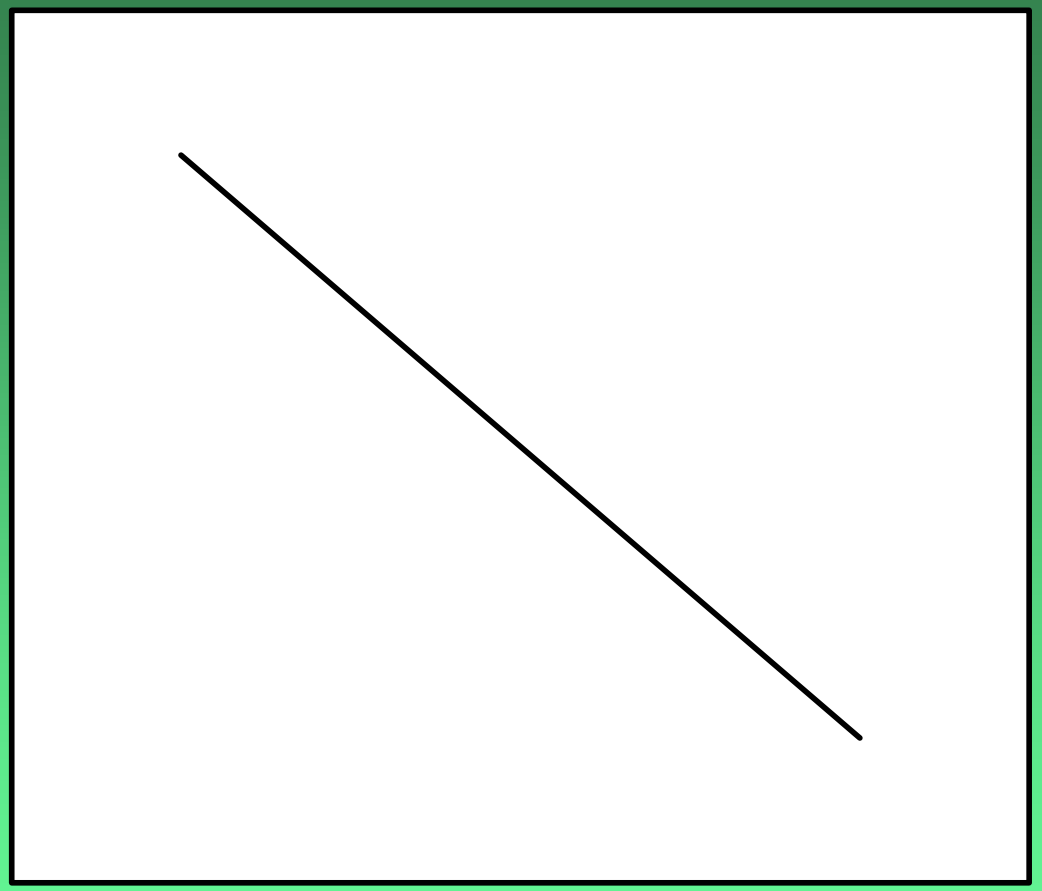
Julie Brennan
Lenore Fahrig
& Lutz Tischendorf
Carleton University
Ottawa, Canada



Overview:

- Genetic differentiation and movement
- Model structure and function
- Results of simulations
- Implications

Genetic differentiation



Inter-population movement

BUT…

- Models are based on assumptions often either violated or not testable
- inadequate sample sizes are available to assess genetic structure
- a more appropriate method could have been used

Assumptions

- genetic equilibrium
- demographic equilibrium
- no mutation
- gene flow=migration
- neutral to selection
- migration random wrt alleles
- no recombination

Genetic measure: F_{st} (Wright 1943; Nei 1973)

$$F_{st} = \frac{H_t - H_s}{H_t} \approx \frac{1}{4Nm + 1}$$

H_s: <u>Expected</u> within-subpopulation heterozygocity

H_t: <u>Expected</u> heterozygocity in a panmictic population

Genetic measure: F_{st}

$$F_{st} = \frac{H_t - H_s}{H_t} \approx \frac{1}{4Nm + 1}$$

Nm: number of migrants per generation

Genetic measure: G_{st} (Nei 1977)

$$G_{st} = \boxed{\frac{H_t - H_s}{H_t}} = \frac{D_{st}}{H_t} \approx \frac{1}{4Nm + 1}$$

H_s and H_t similar to those for F_{st} but:

- are corrected for observed heterozygocity
- use harmonic mean of population size

Genetic measure: G_{st}

$$G_{st} = \frac{H_t - H_s}{H_t} = \frac{D_{st}}{H_t} \approx \frac{1}{4Nm + 1}$$

D_{st}: heterozygocity among samples

dependent on number of samples

Genetic measure: G'_{st} (Nei & Chesser 1983)

$$G'_{st} = \frac{H'_t - H_s}{H'_t} = \frac{D'_{st}}{H'_t} \approx \frac{1}{4Nm+1}$$

Genetic measure: G'_{st}

$$G'_{st} = \frac{H'_t - H_s}{H'_t} = \boxed{\frac{D'_{st}}{H'_t}} \approx \frac{1}{4Nm+1}$$

D'_{st}: corrected for number of samples

$$D'_{st} = \frac{sample\#}{sample\#-1} * D_{st}$$

Genetic measure: G'_{st}

$$G'_{st} = \frac{H'_t - H_s}{H'_t} = \frac{D'_{st}}{H'_t} \approx \frac{1}{4Nm+1}$$

D'_{st}: corrected for number of samples

$$D'_{st} = \frac{sample\#}{sample\#-1} * D_{st}$$

and

$$H'_t = H_s + D'_{st}$$

Genetic measure: Θ (Weir & Cockerham 1984)

$$\theta = \frac{\sigma_a}{\sigma_a + \sigma_b + \sigma_w} \approx \frac{1}{4Nm + 1}$$

Genetic measure: Θ

$$\theta = \frac{\sigma_a}{\sigma_a + \sigma_b + \sigma_w} \approx \frac{1}{4Nm + 1}$$

Corrects for sample size and number of samples

Adjusts when samples sizes differ

Equal to G'_{st} when sample sizes even

Genetic measure: R_{st} (Slatkin 1995)

$$R_{ST} = \frac{\overline{S} - S_W}{\overline{S}} \approx \frac{1}{1 + 4Nm\left(\frac{d_s}{d_s - 1}\right)}$$

Genetic measure: R_{st}

$$R_{ST} = \boxed{\frac{\overline{S} - S_W}{\overline{S}}} \approx \frac{1}{1 + 4Nm\left(\frac{d_s}{d_s - 1}\right)}$$

Based on the variance in the number of repeats in microsatellite DNA; mutation present

Does not adjust for sample size or heterozygocity

Genetic measure: R_{st}

$$R_{ST} = \frac{\overline{S} - S_W}{\overline{S}} \approx \frac{1}{1 + 4Nm\left(\frac{d_s}{d_s - 1}\right)}$$

Nm adjusted for number of sampled populations: d_s

Genetic measure: Private alleles (Slatkin 1985)

- **Based on the frequencies of alleles found only in 1 population (p)**
- **Log of p is linearly related to log of Nm:**

$$\ln(p) = -0.505 \ln(Nm) + 2.44$$

Genetic measure: Private alleles

Rearranged, with adjustment for sample size:

$$Nm = e^{\left(\ln(p)+\left(\frac{2.44}{-5.05}\right)\right)} * \boxed{\frac{25}{n}}$$

Genetic measure: Genetic assignment test (Waser & Strobeck 1998)

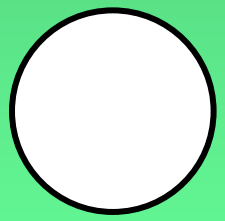
Sampled individual = migrant if:

sampled population likelihood

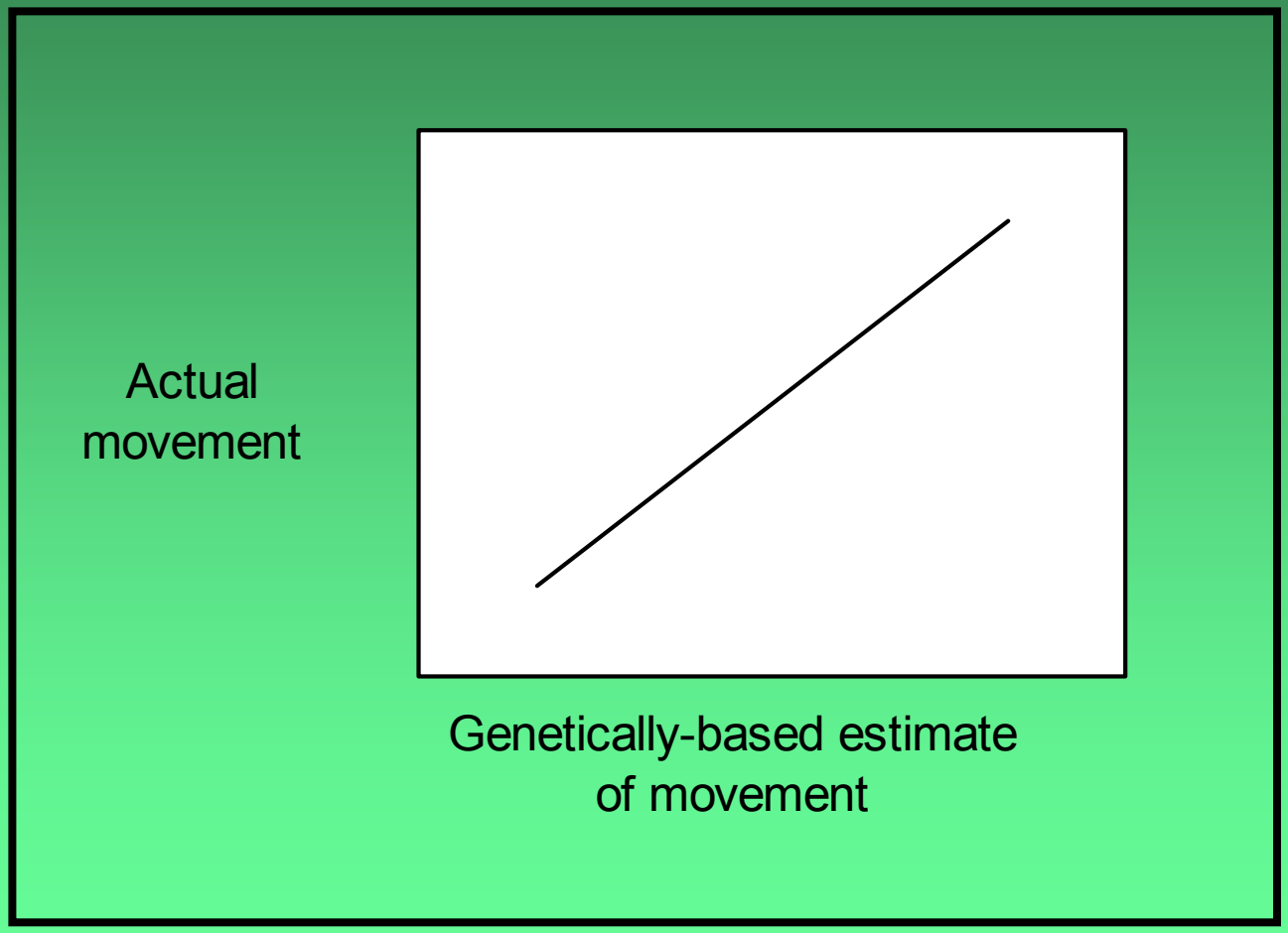
< another population likelihood

How good is an estimate?

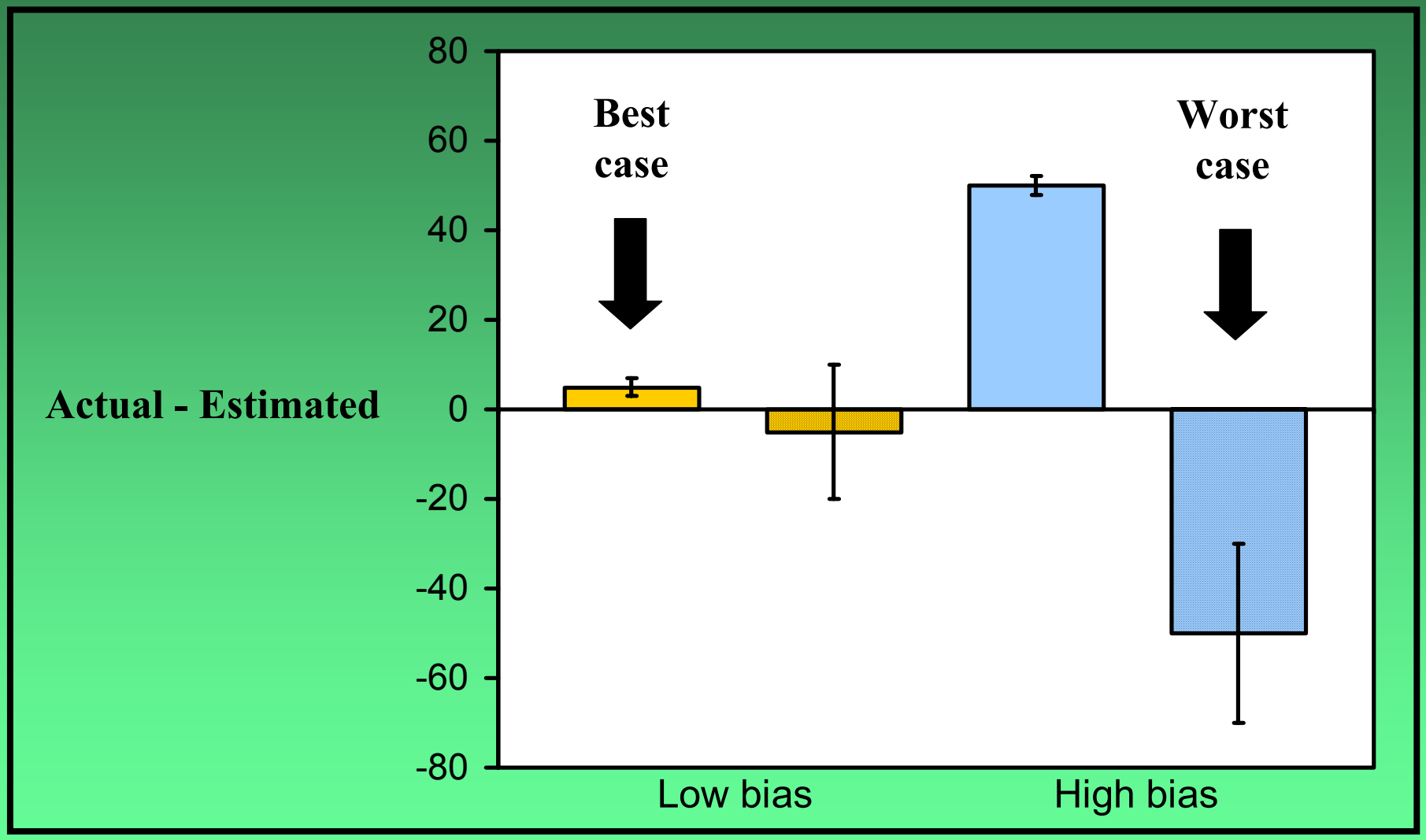
- Does it track actual movement?
- Is it unbiased?
- Is it consistent?

POPULATIONS	GENETIC MEASURES		BASED ESTIMATE OF MOVEMENT (estimated)
	F_{ST} G_{ST} G'_{ST} Θ	→	$Nm = \frac{1}{4 * gen.measure} - 0.25$
Actual Movement (reality) 	R_{ST}	→	$Nm = \frac{d_s - 1}{4 d_s} * \left(\frac{1}{R_{ST}} - 1 \right)$
	Private alleles	→	$Nm = e^{\left(\ln(p) + \left(\frac{2.44}{-5.05} \right) \right)} * \frac{25}{\bar{n}}$
	Genetic assignment test	→	*Number of misassigned individuals*

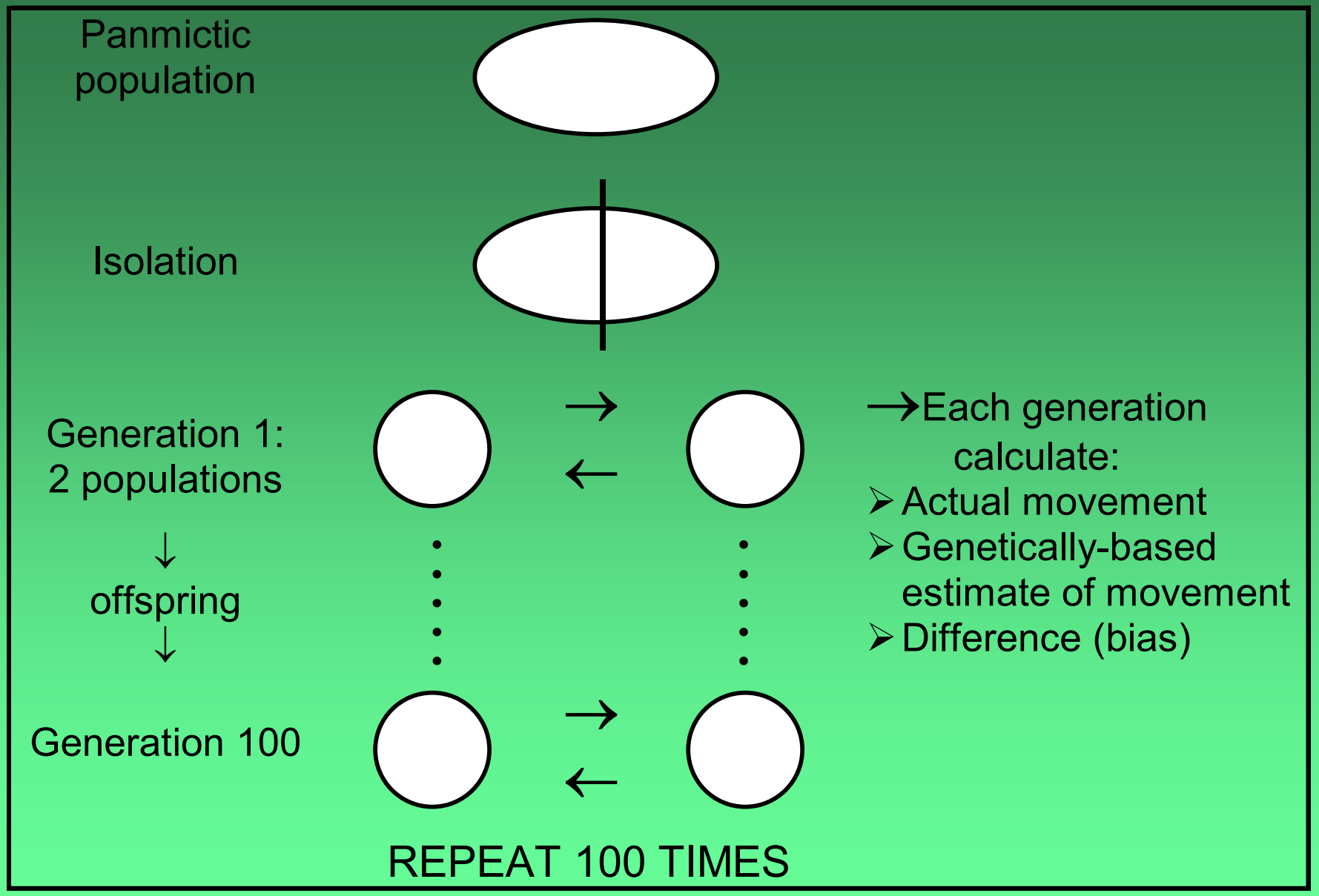
Does it track actual movement?



Is it unbiased and consistent?



The Simulations


Panmictic population
Isolation
Generation 1: 2 populations
offspring
Generation 100
→Each generation calculate:
➢ Actual movement
➢ Genetically-based estimate of movement
➢ Difference (bias)
REPEAT 100 TIMES

Method used

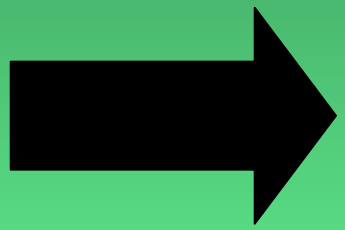
Demography

Generations since isolation

Sample size

Number of loci

Number of alleles



Movement amount
Bias
Consistency

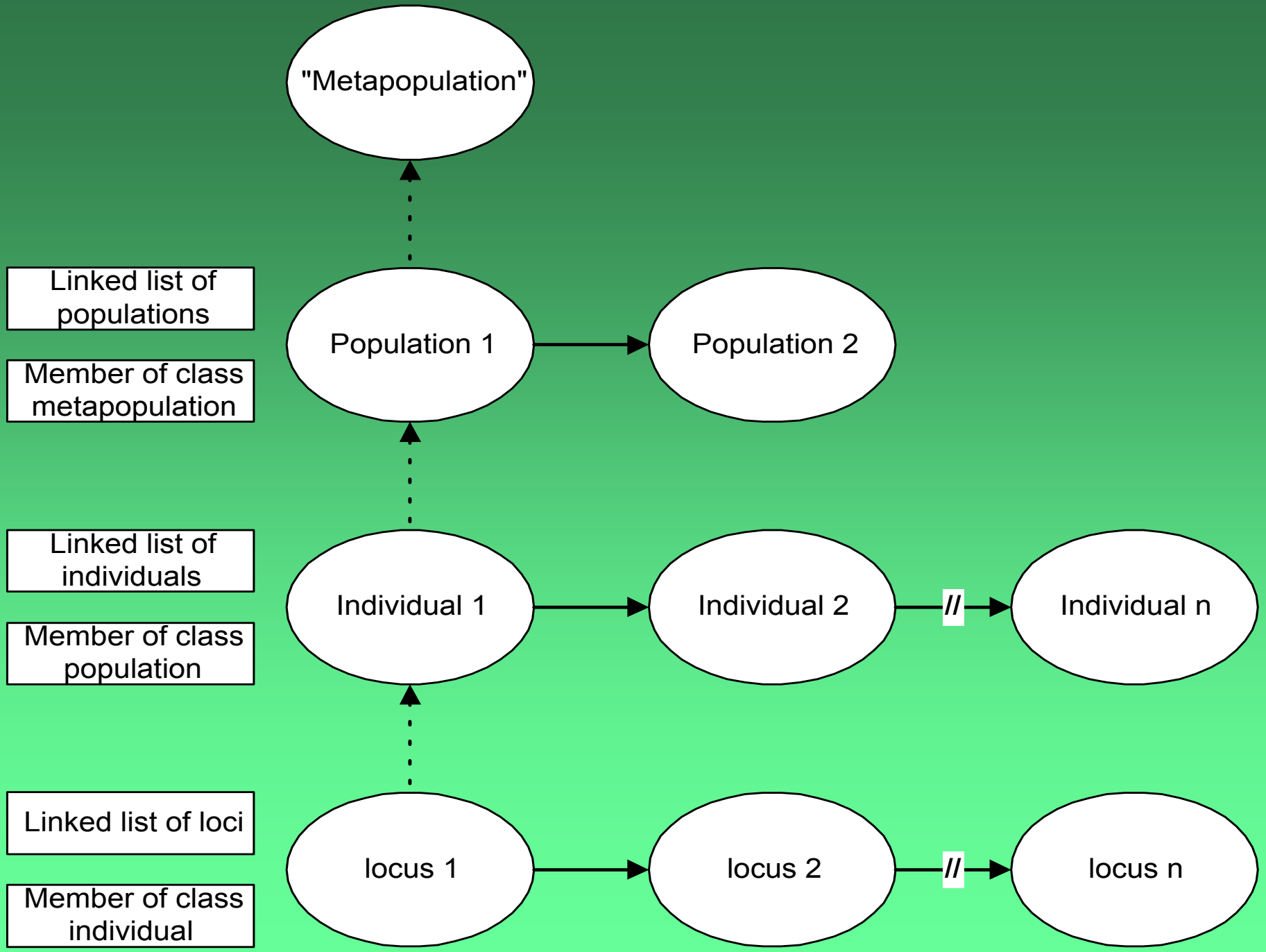
Simulation model assumptions

- 2 populations of equal initial size
- equal sample sizes
- birth rate = death rate
- closed system

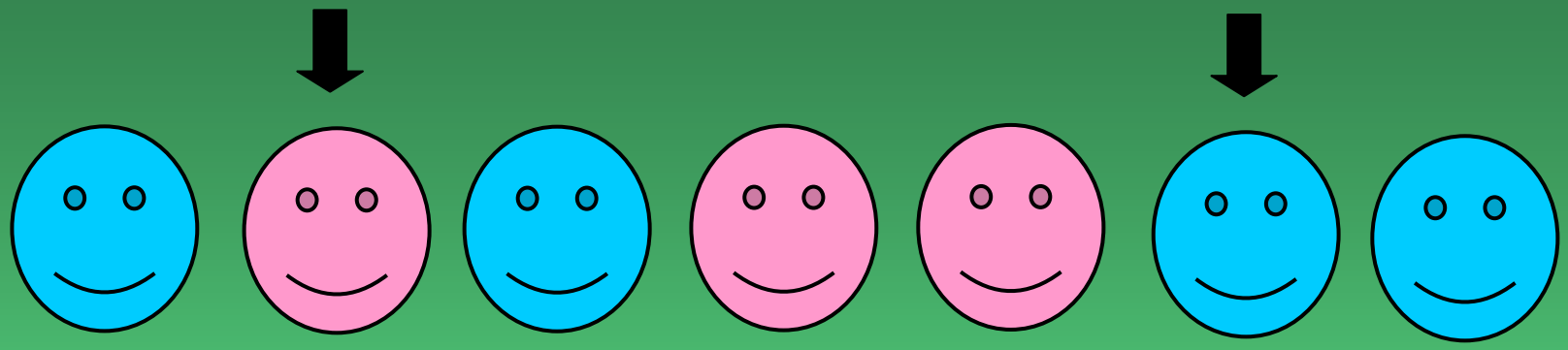
- symmetrical movement
- no movement mortality
- movement not spatially explicit

- polygynous mating
- equal probability of mating of residents and migrants

The model

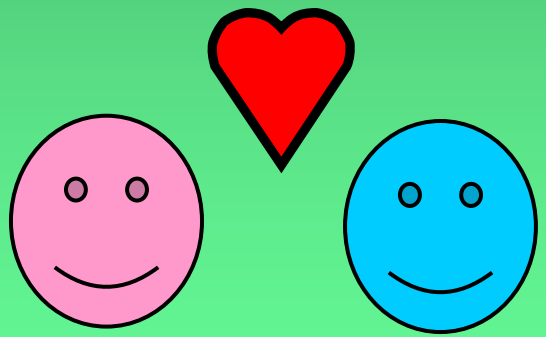

"Metapopulation"
Linked list of populations
Population 1
Population 2
Member of class metapopulation
Linked list of individuals
Individual 1
Individual 2
//
Individual n
Member of class population
Linked list of loci
locus 1
locus 2
//
locus n
Member of class individual

REPRODUCTION

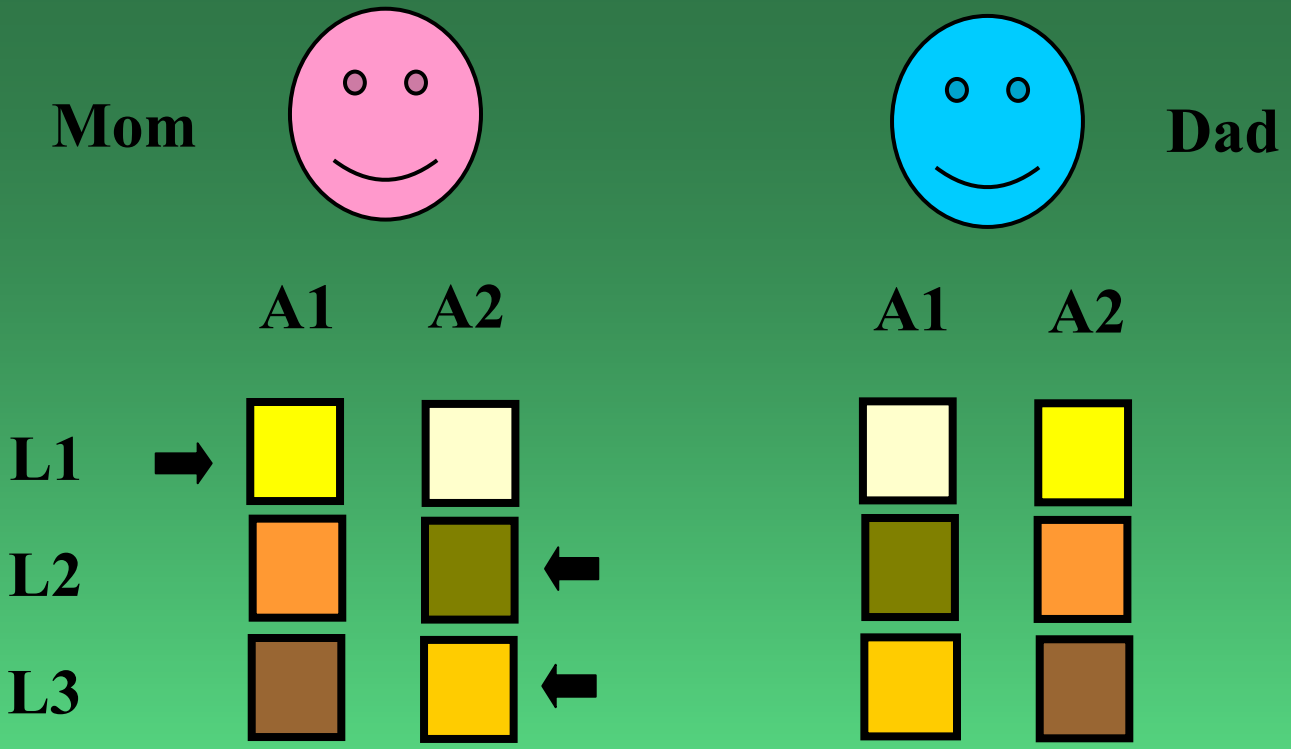


(FOR EACH POPULATION)

REPRODUCTION



(FOR EACH POPULATION)


Mom
Dad
A1
A2
A1
A2
L1
L2
L3

Mom

Dad

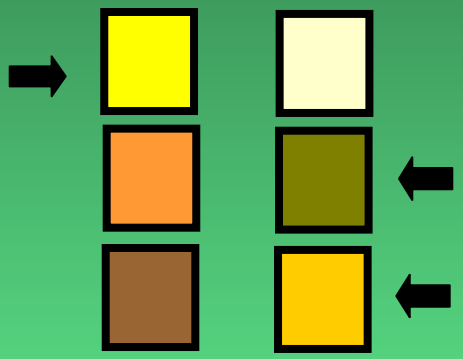
A1 A2

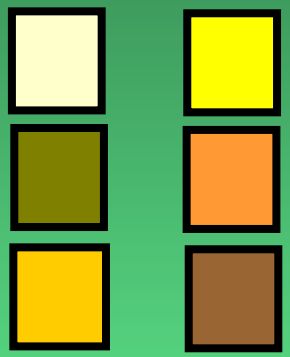
A1 A2

L1

L2

L3



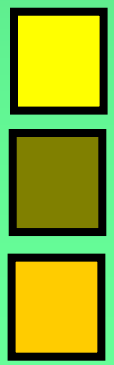


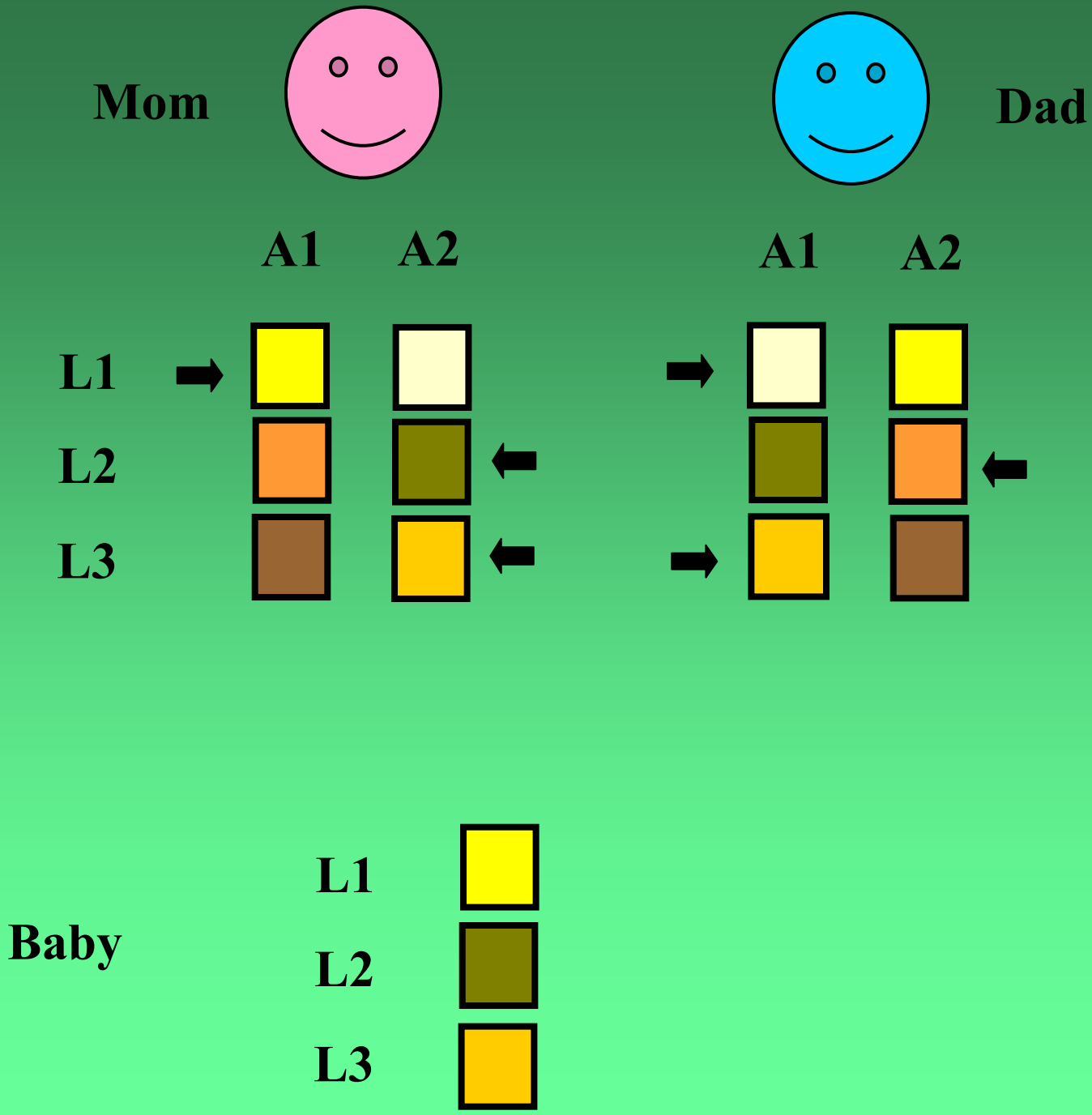
Baby

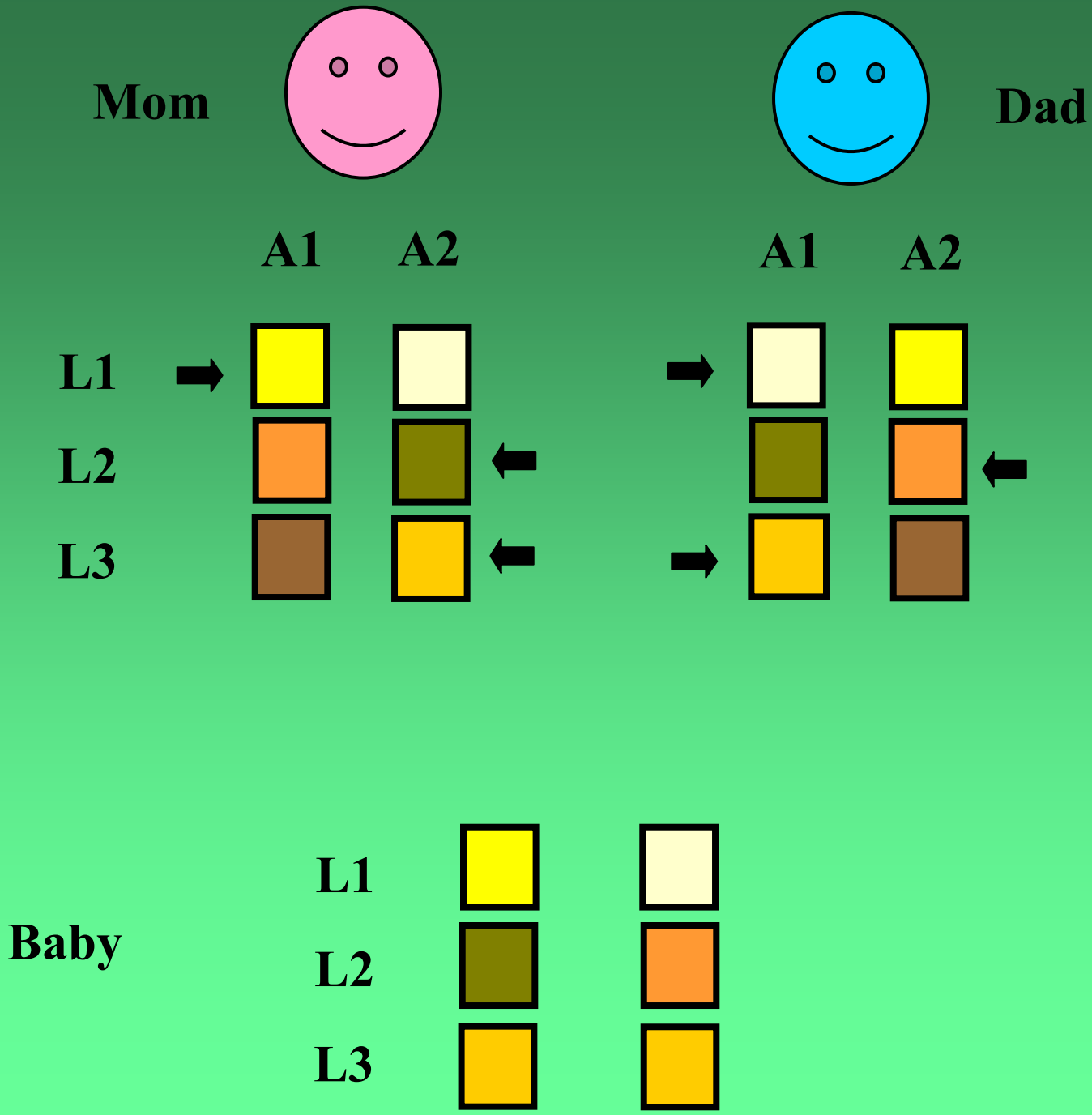
L1

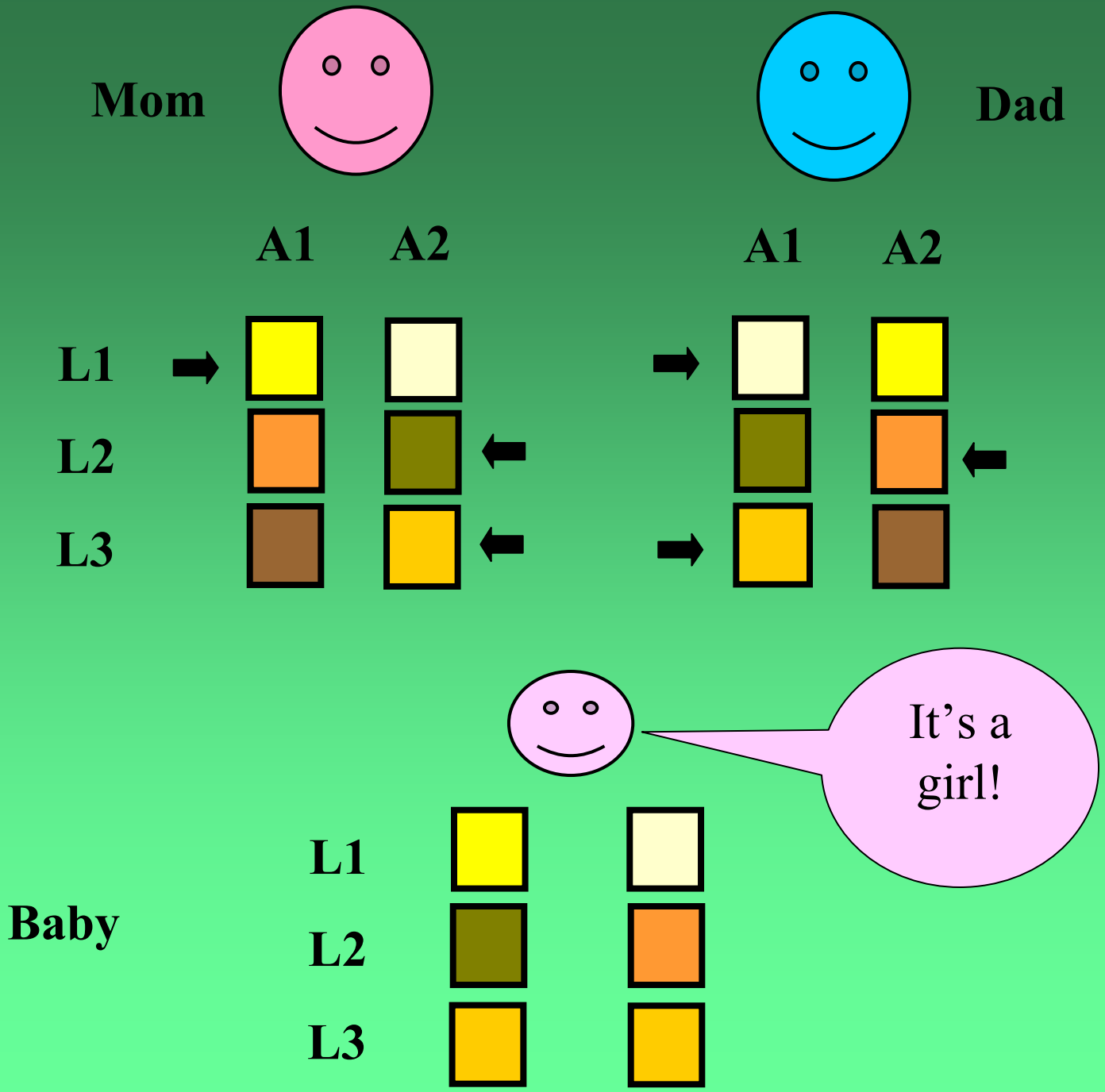
L2

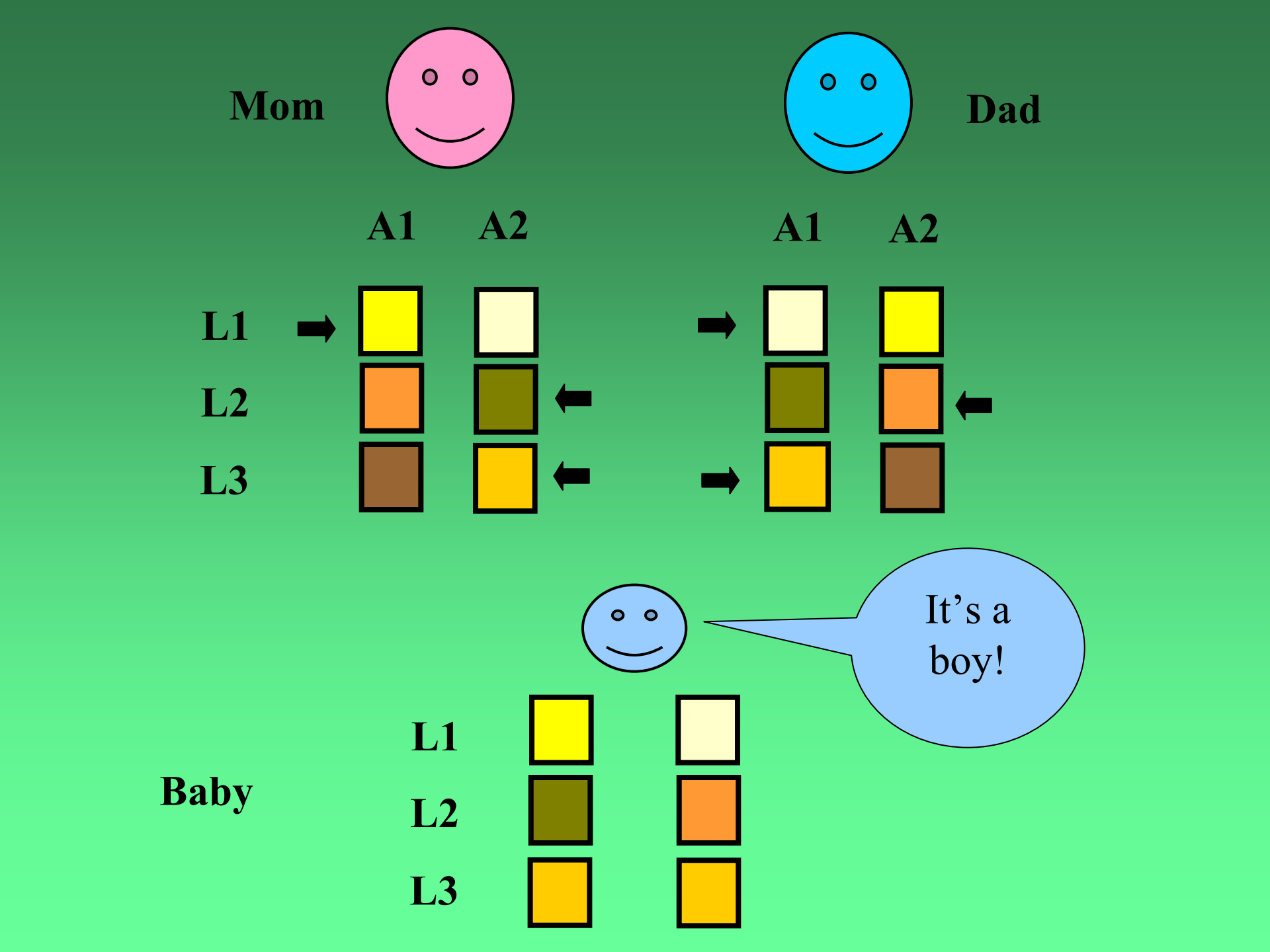
L3



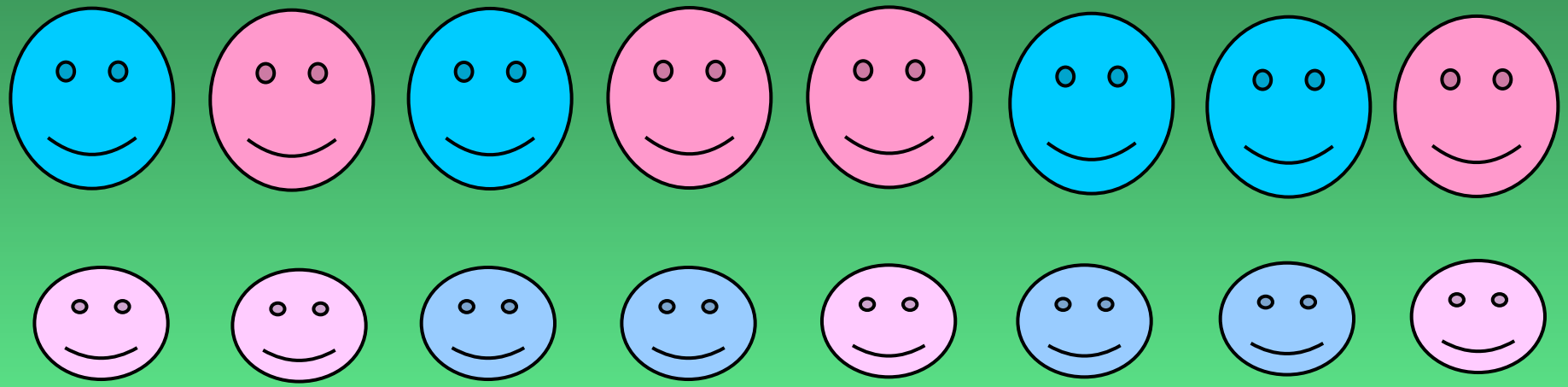

Mom
Dad
A1
A2
A1
A2
L1
L2
L3
Baby
L1
L2
L3


Mom
Dad
A1
A2
A1
A2
L1
L2
L3
Baby
L1
L2
L3


Mom
Dad
A1
A2
A1
A2
L1
L2
L3
It's a girl!
Baby
L1
L2
L3

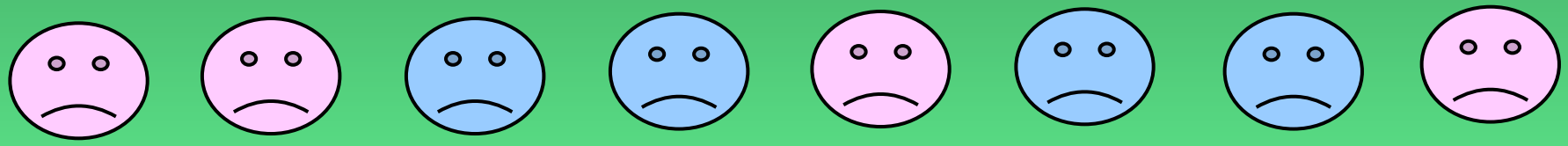

Mom
Dad
A1
A2
A1
A2
L1
L2
L3
It's a boy!
Baby
L1
L2
L3

MORTALITY



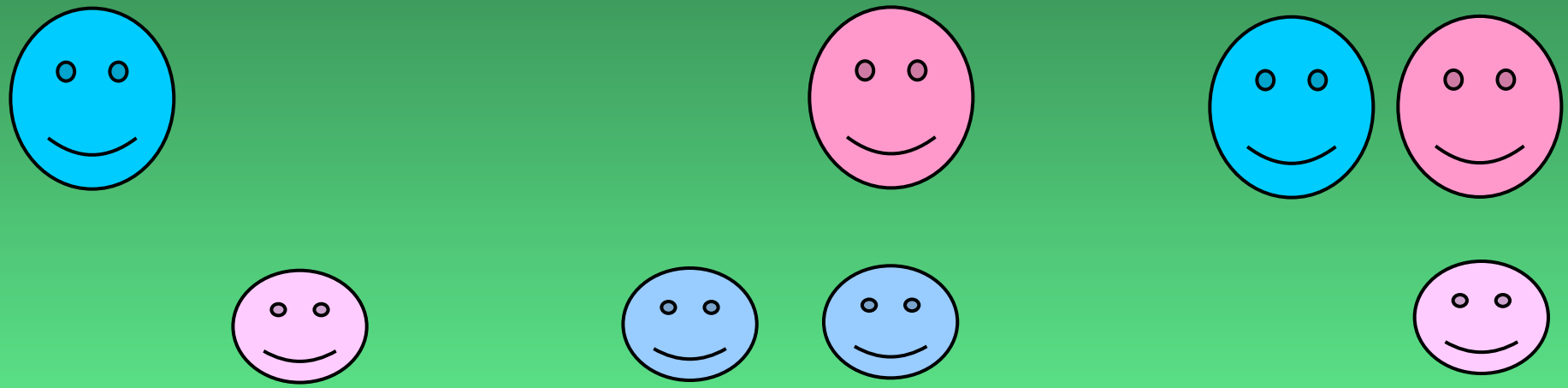
MORTALITY

Non-overlapping generations

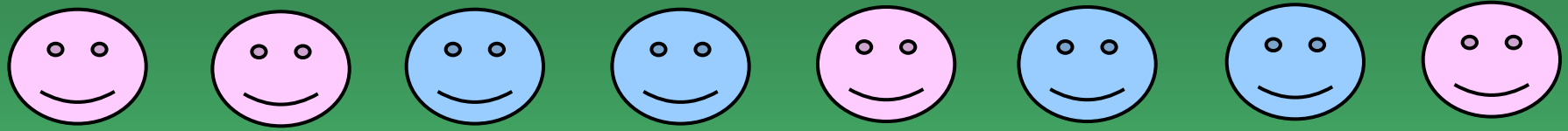


MORTALITY

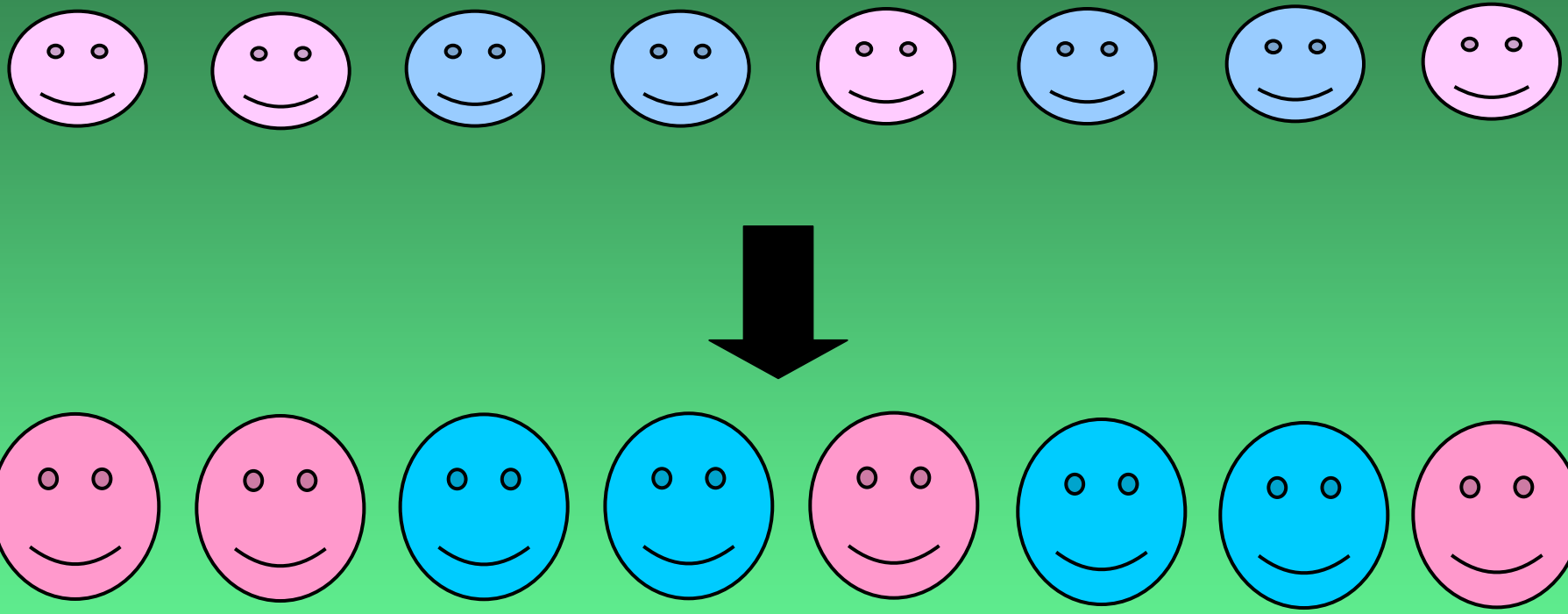
overlapping generations



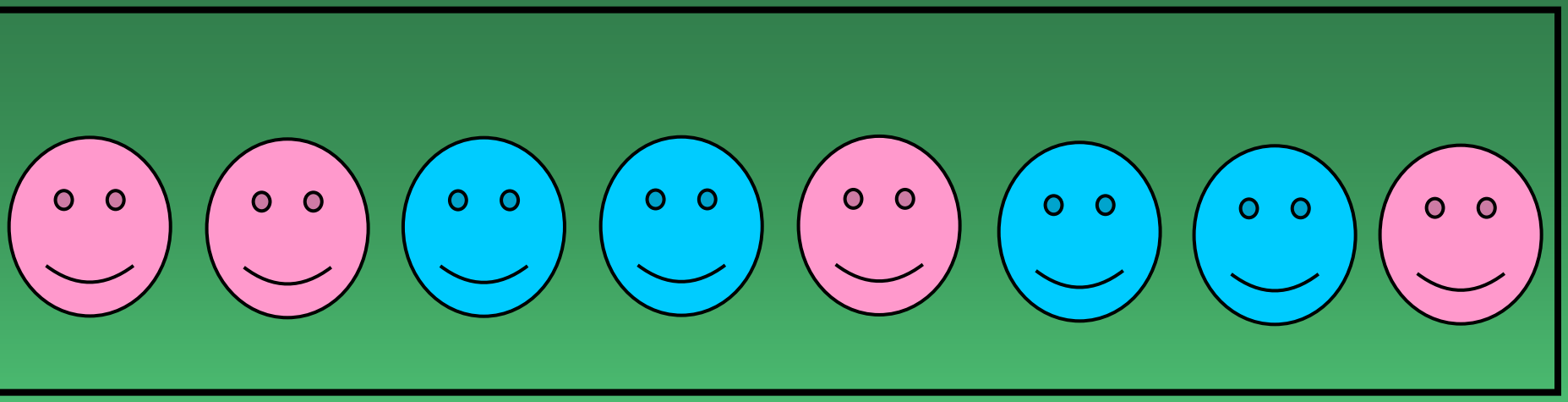
AGING



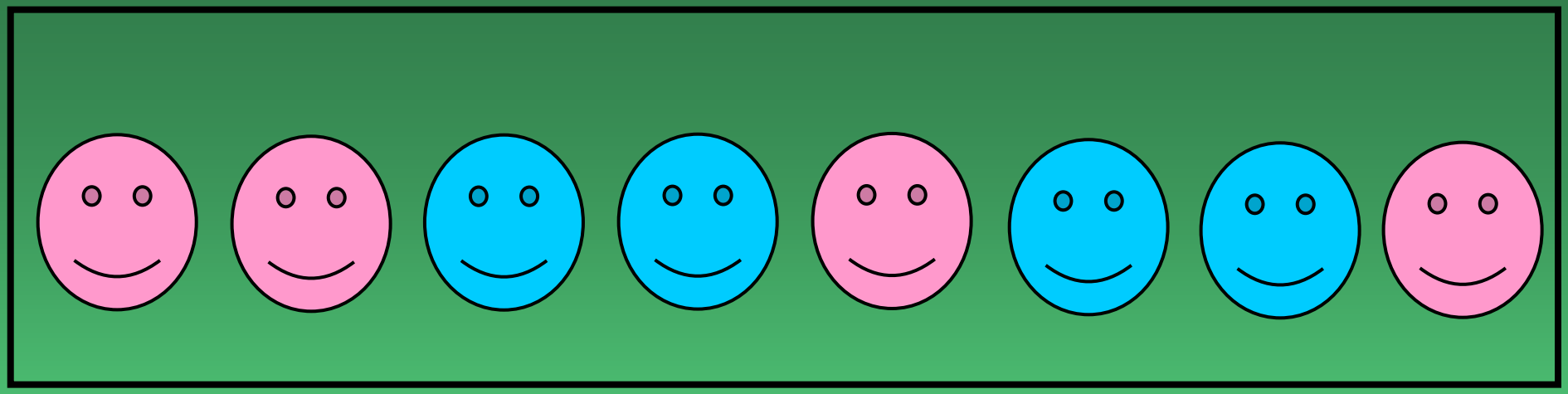
AGING

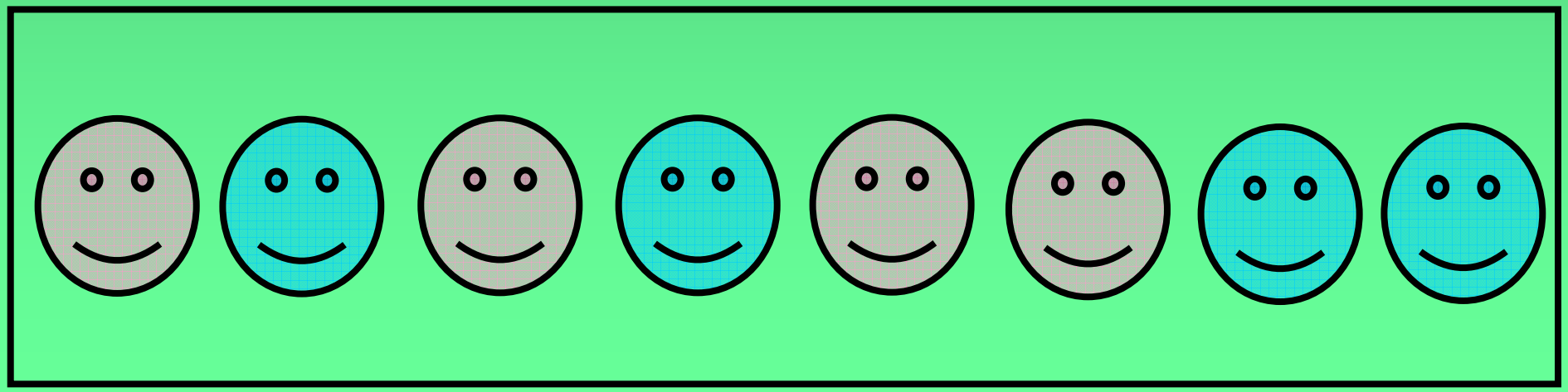


DISPERSAL

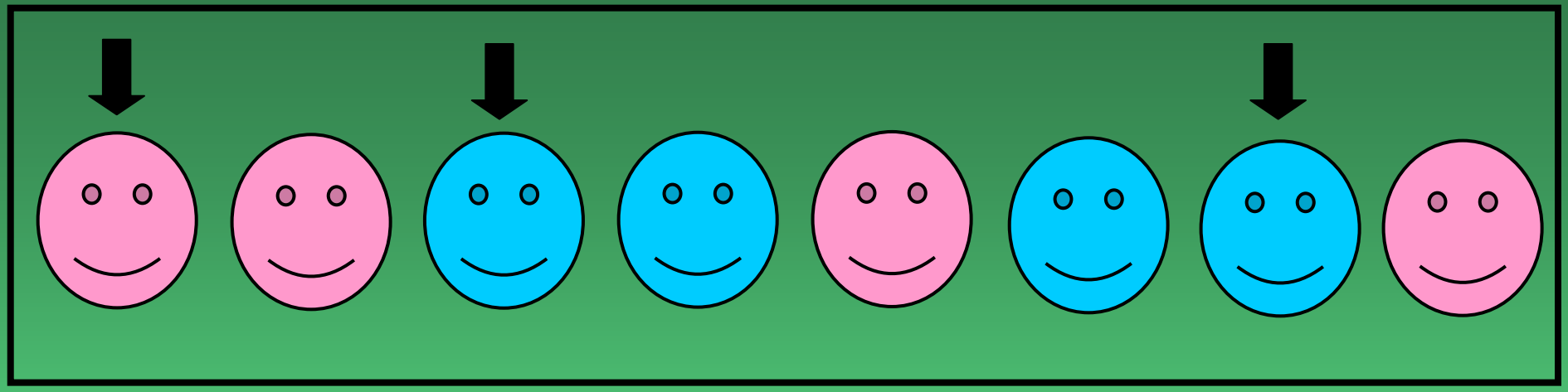


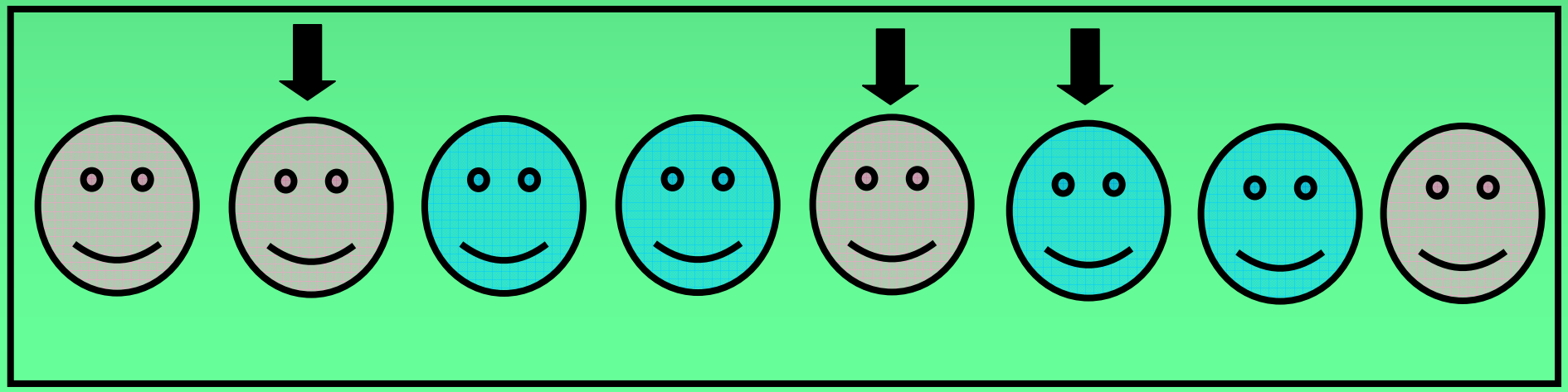
DISPERSAL



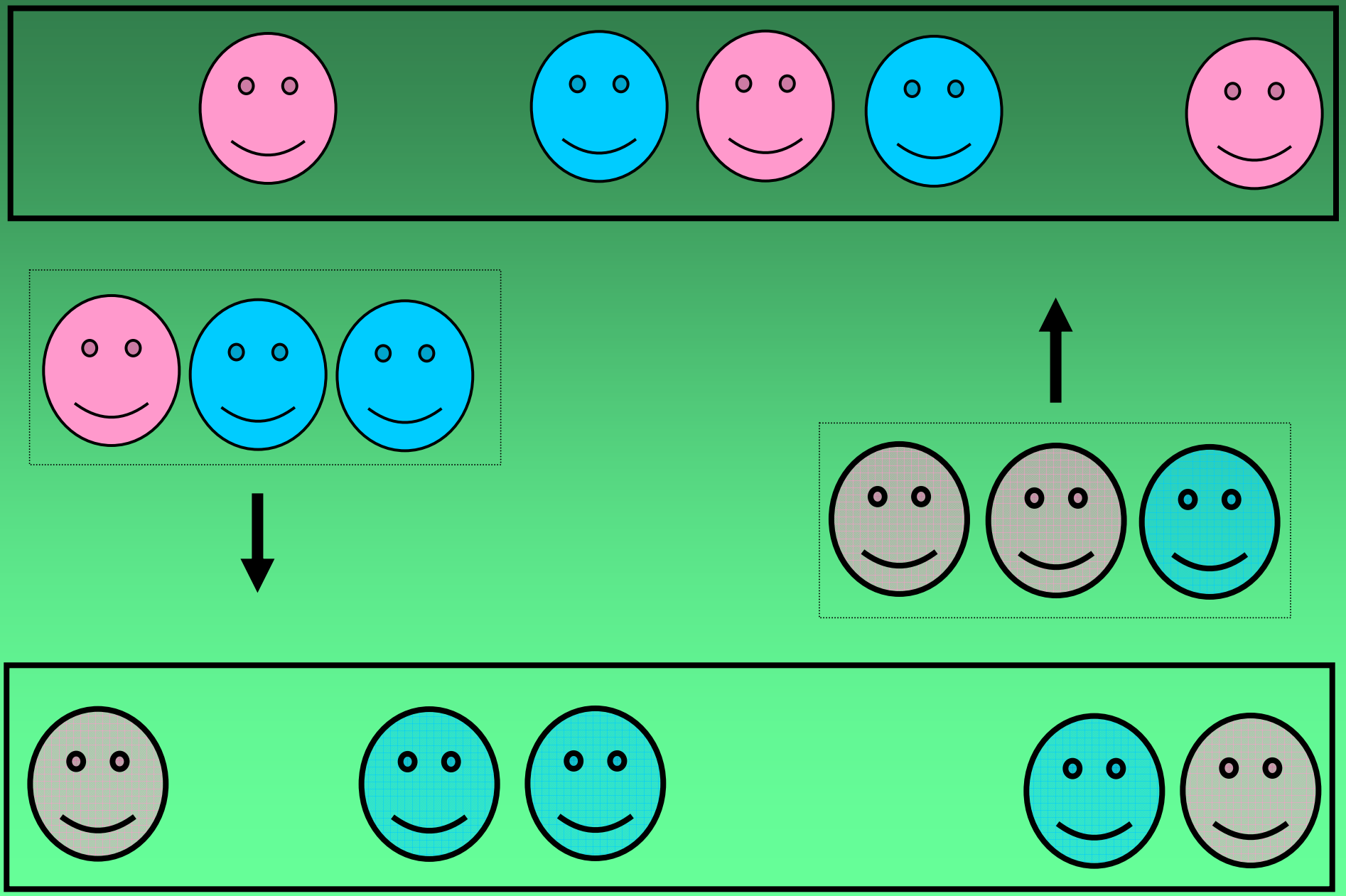


DISPERSAL

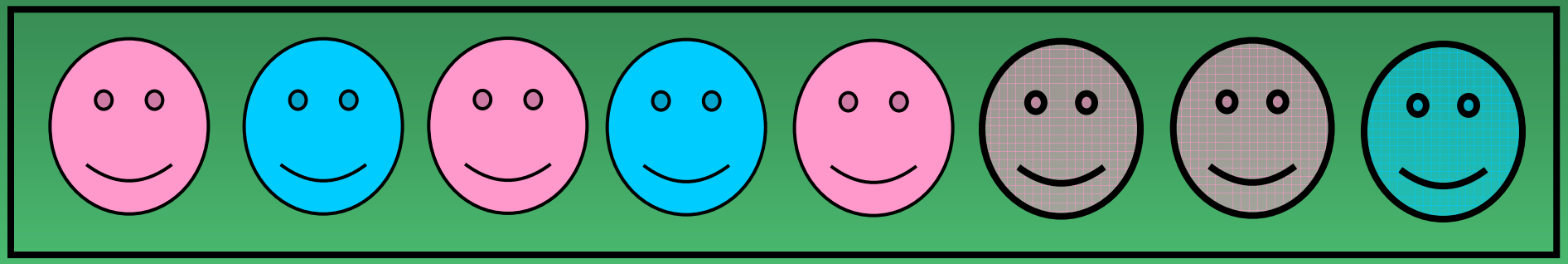


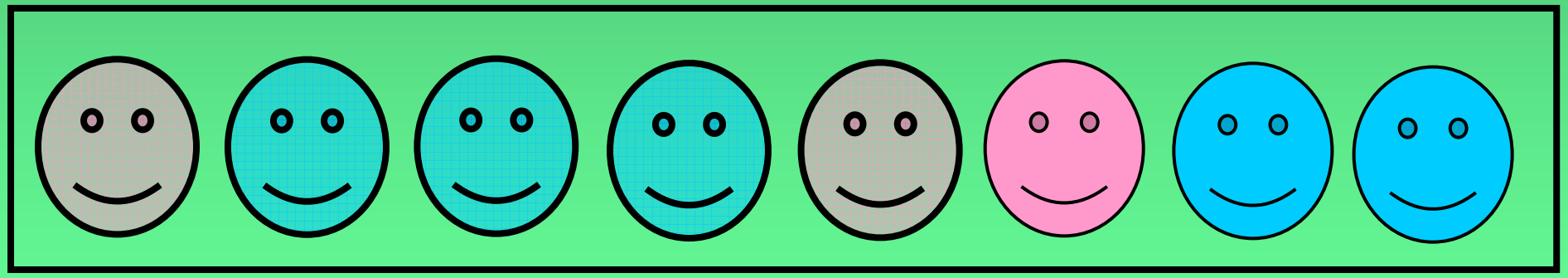


DISPERSAL



DISPERSAL





Calculate:

Actual movement

Estimated movement based on:

F_{ST}; G_{ST}; G'_{ST}; Θ;

R_{ST}; Private allele; GAT

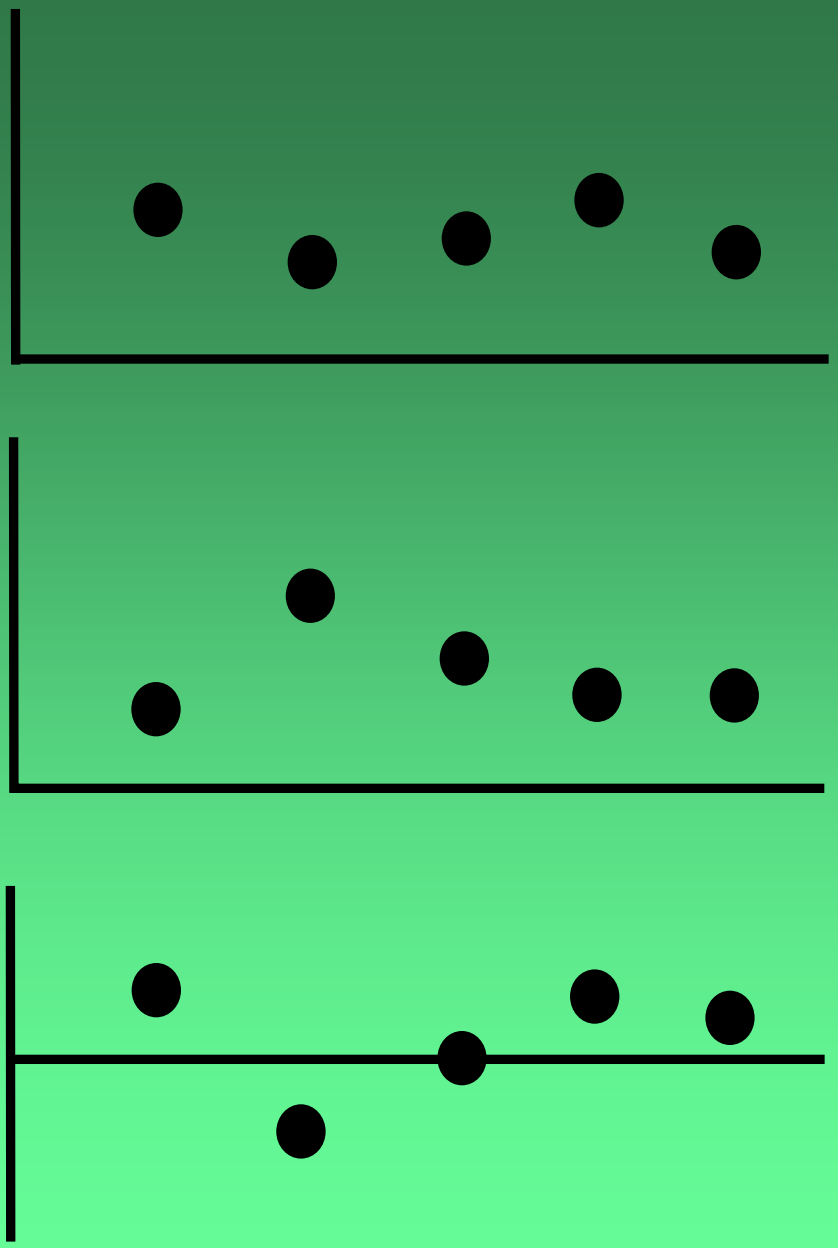
Difference (bias)

Output

Actual movement

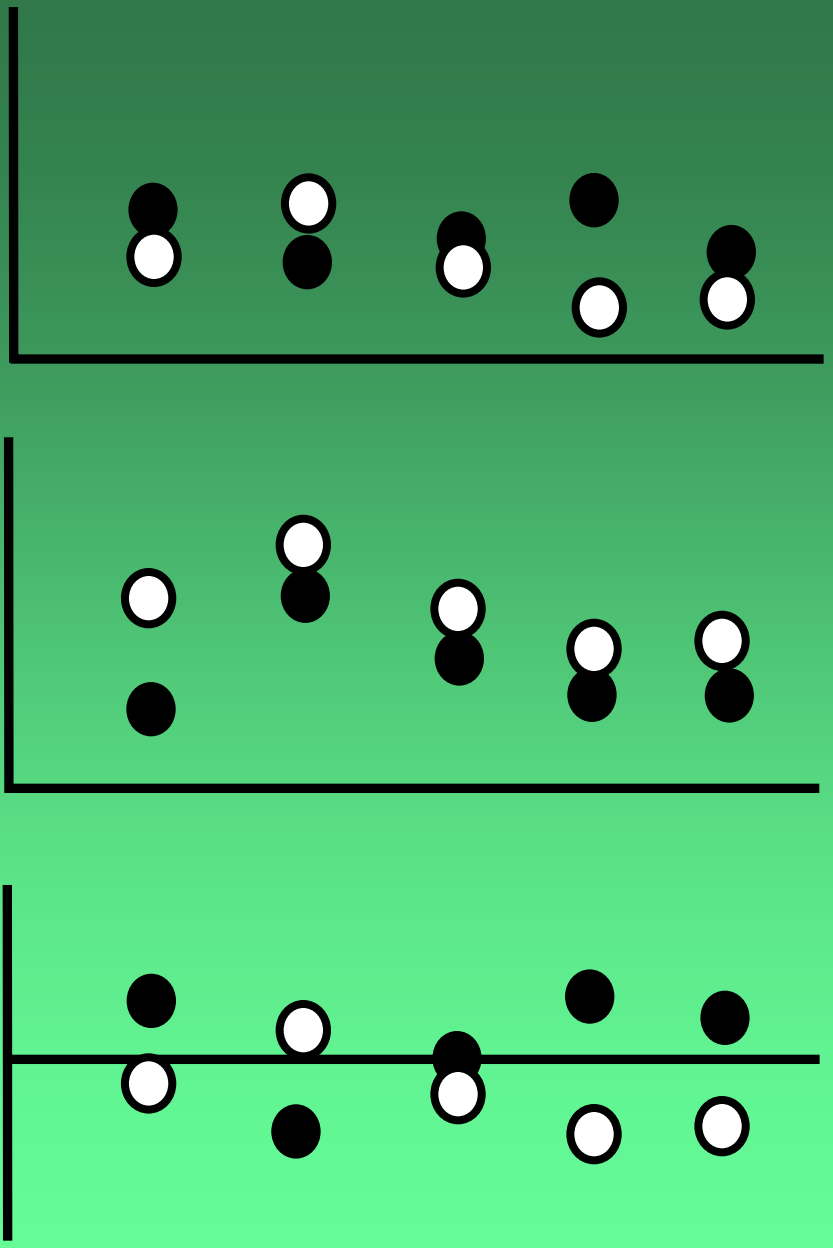
Estimated movement

Actual-estimated (bias)

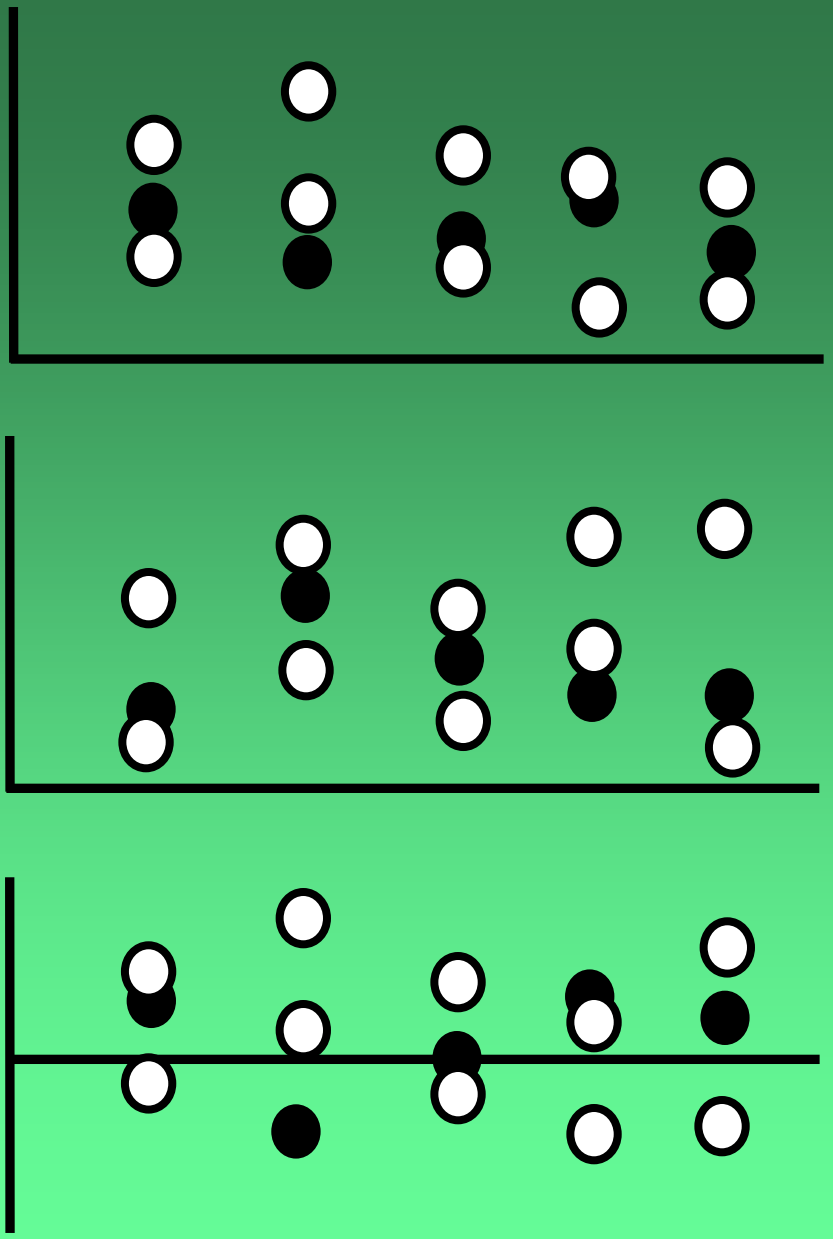


AFTER 1 REPETITION

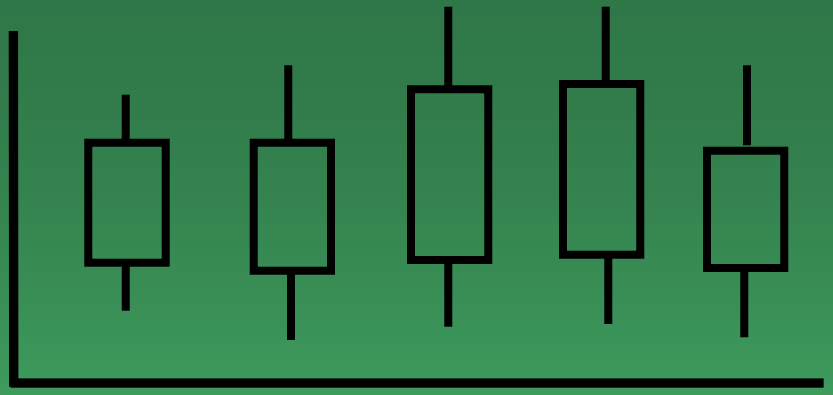
Generations since isolation



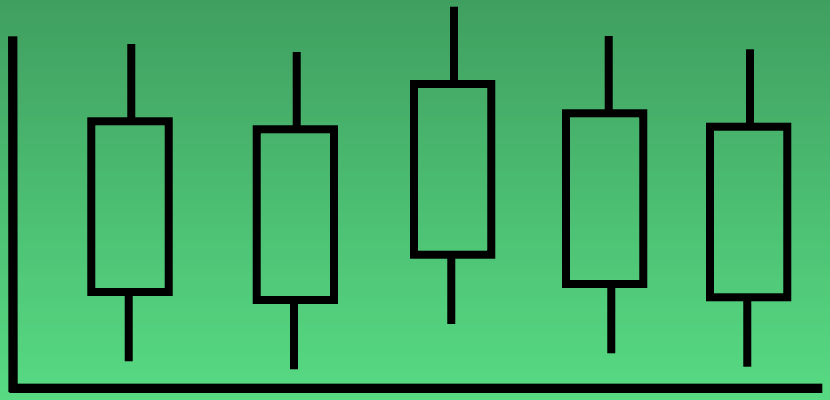
AFTER 2 REPETITIONS


Actual movement
Estimated movement
Actual-estimated (bias)
AFTER 3 REPETITIONS
Generations since isolation

Actual movement

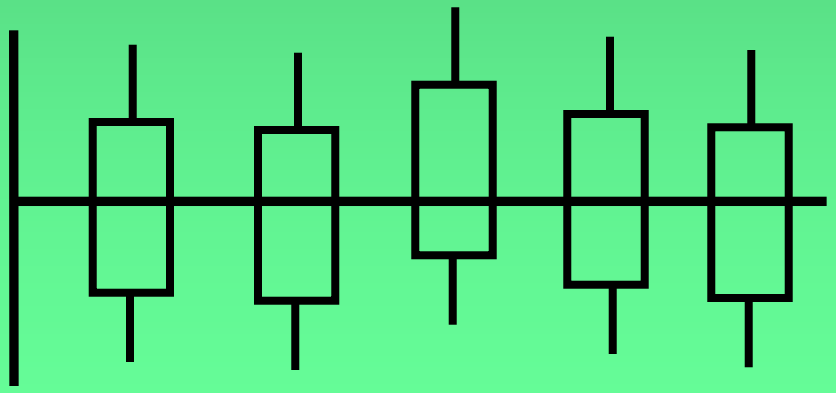


Estimated movement



AFTER <u>100</u> REPETITIONS

Actual-estimated (bias)



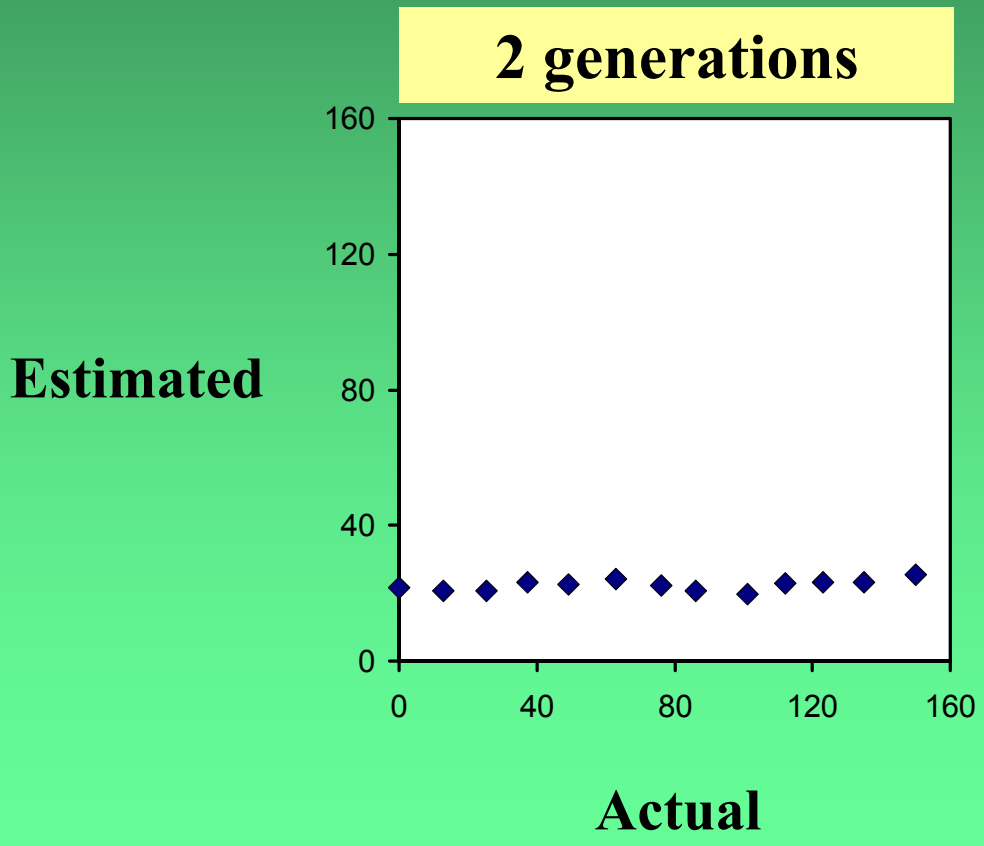
Generations since isolation

Subset of parameters considered:

- Initial population size: 500
- Sample sizes: 20, 120
- Generations since isolation: 2, 50
- Number of loci: 15
- Number of alleles: 2
- Movement probability: 0 - 0.30
- Generation overlap? none

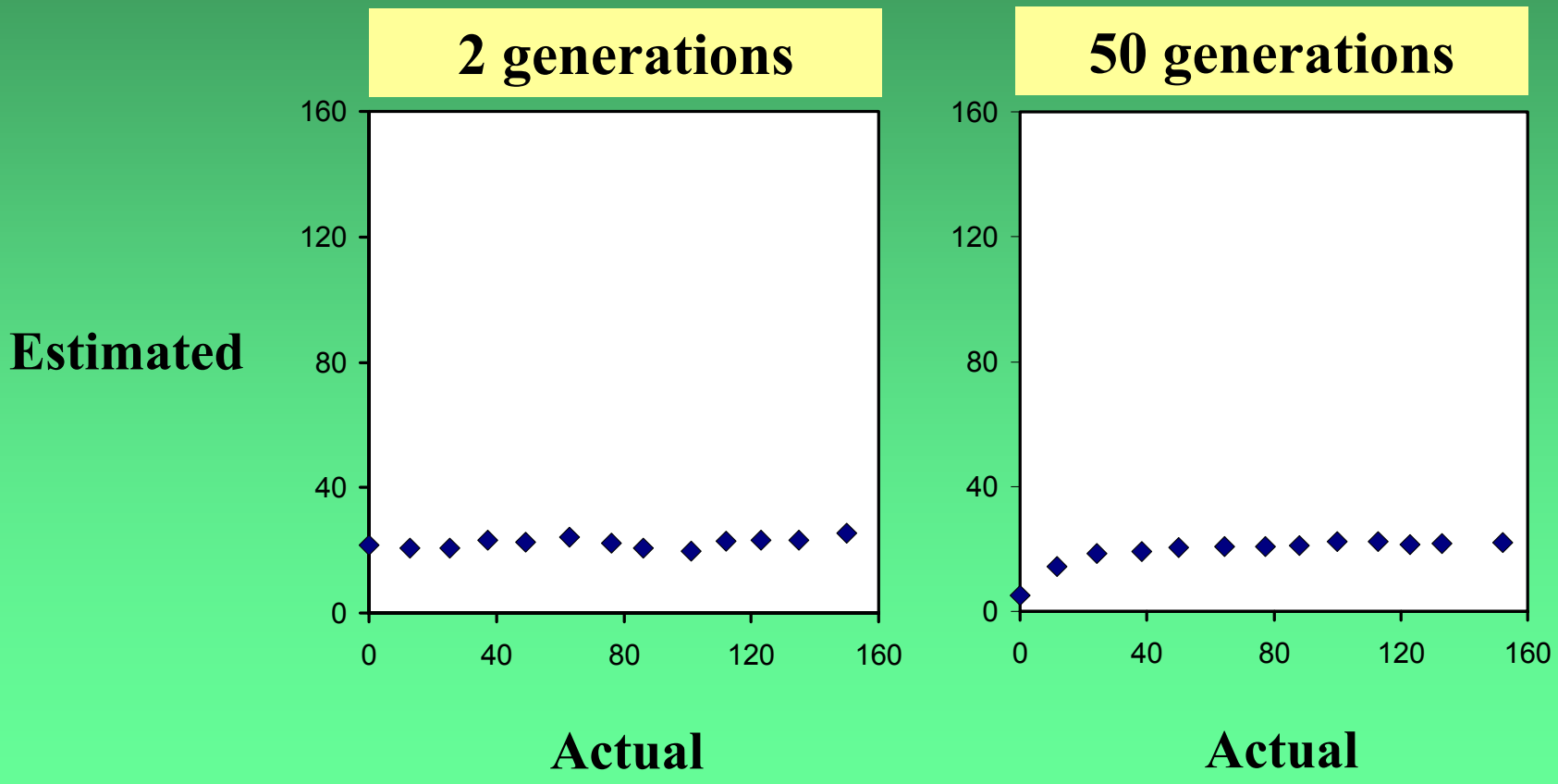
F_{st} based estimate of movemnt vs actual movement:

- sample size = 20



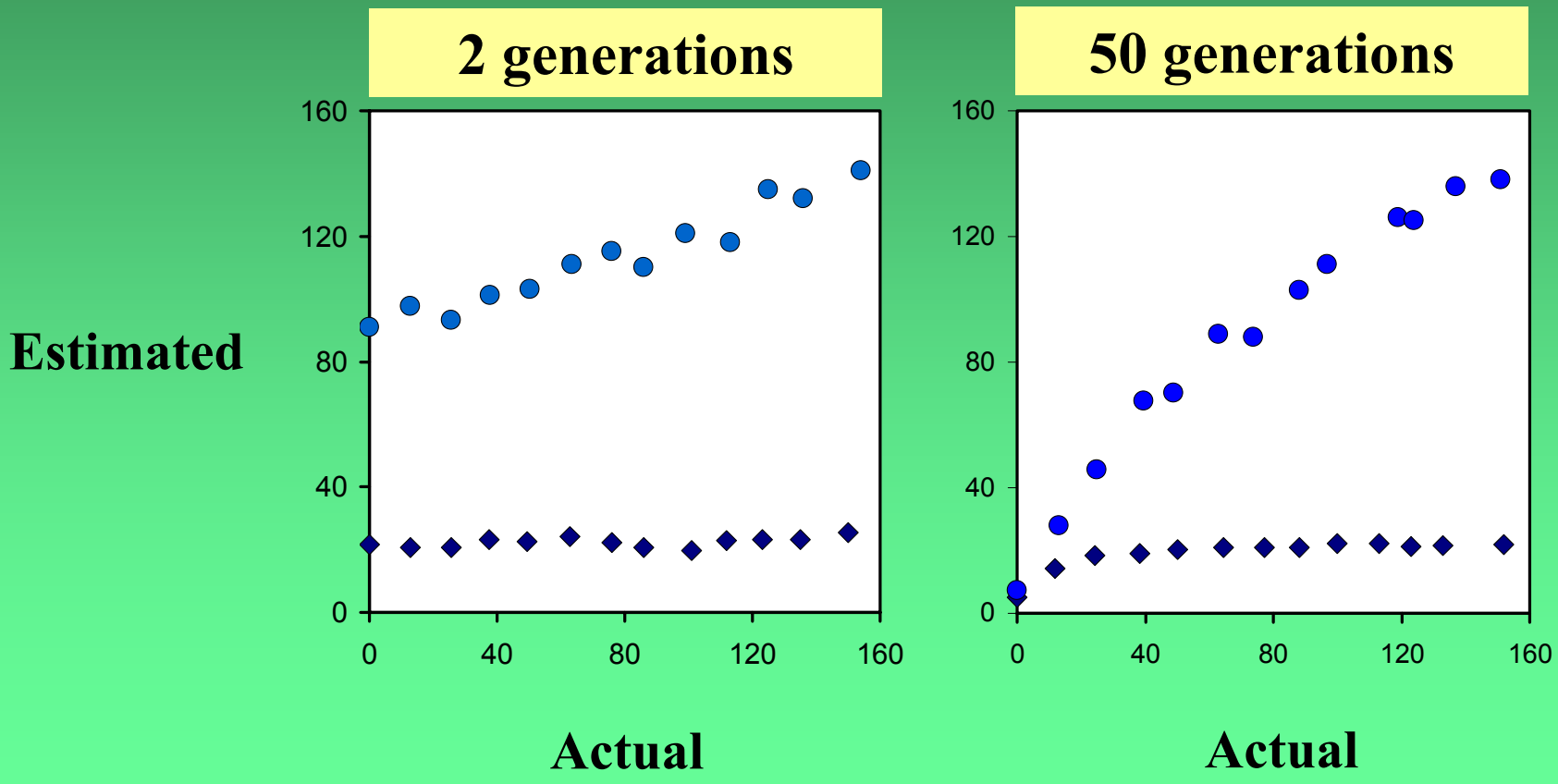
F_{st} based estimate of movemnt vs actual movement:

- sample size = 20



F_{st} based estimate of movemnt vs actual movement:

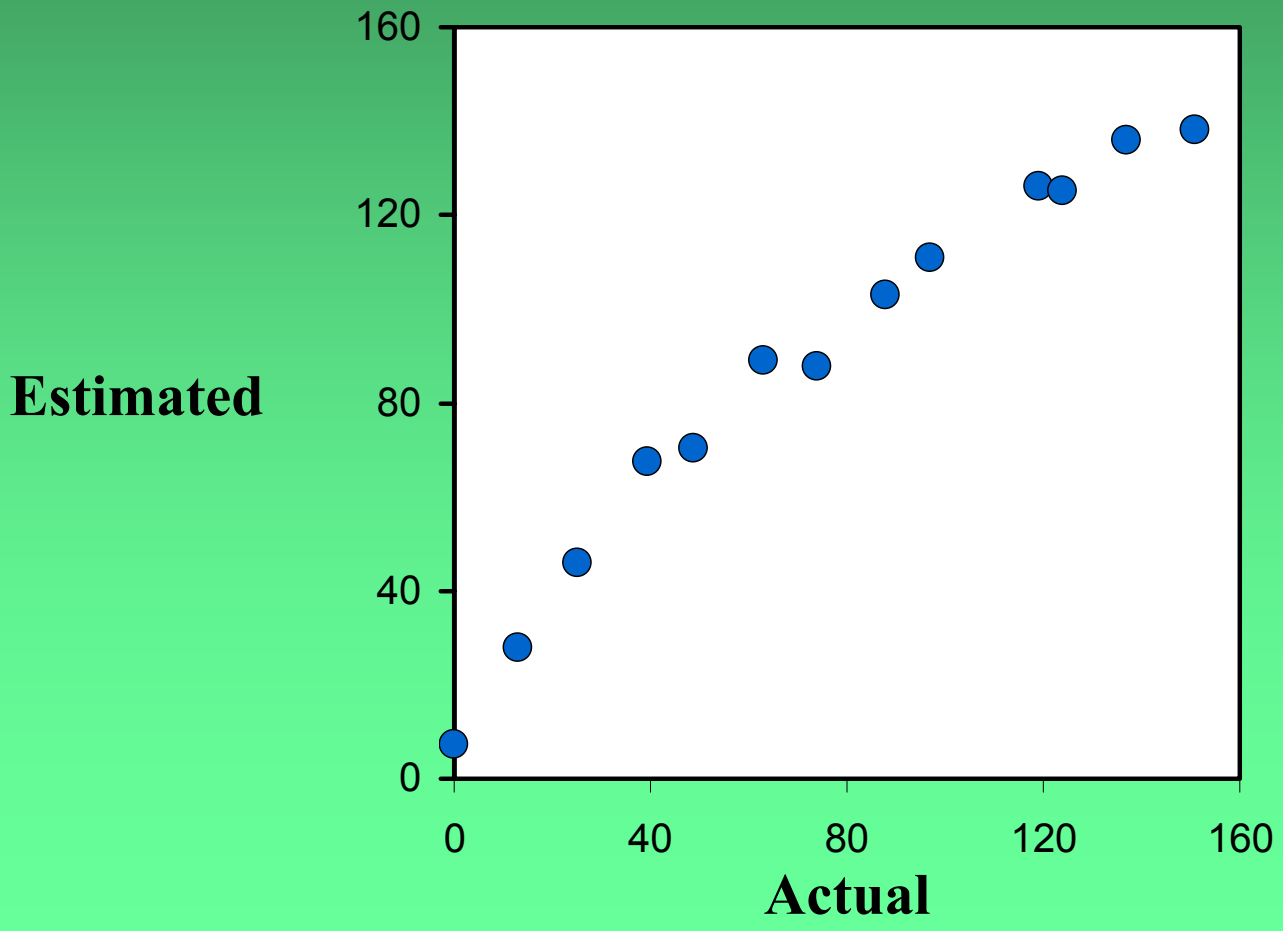
- sample size = 20
- sample size = 120



Genetically-based estimates:

- **sample size = 120**
- **50 generations since isolation**

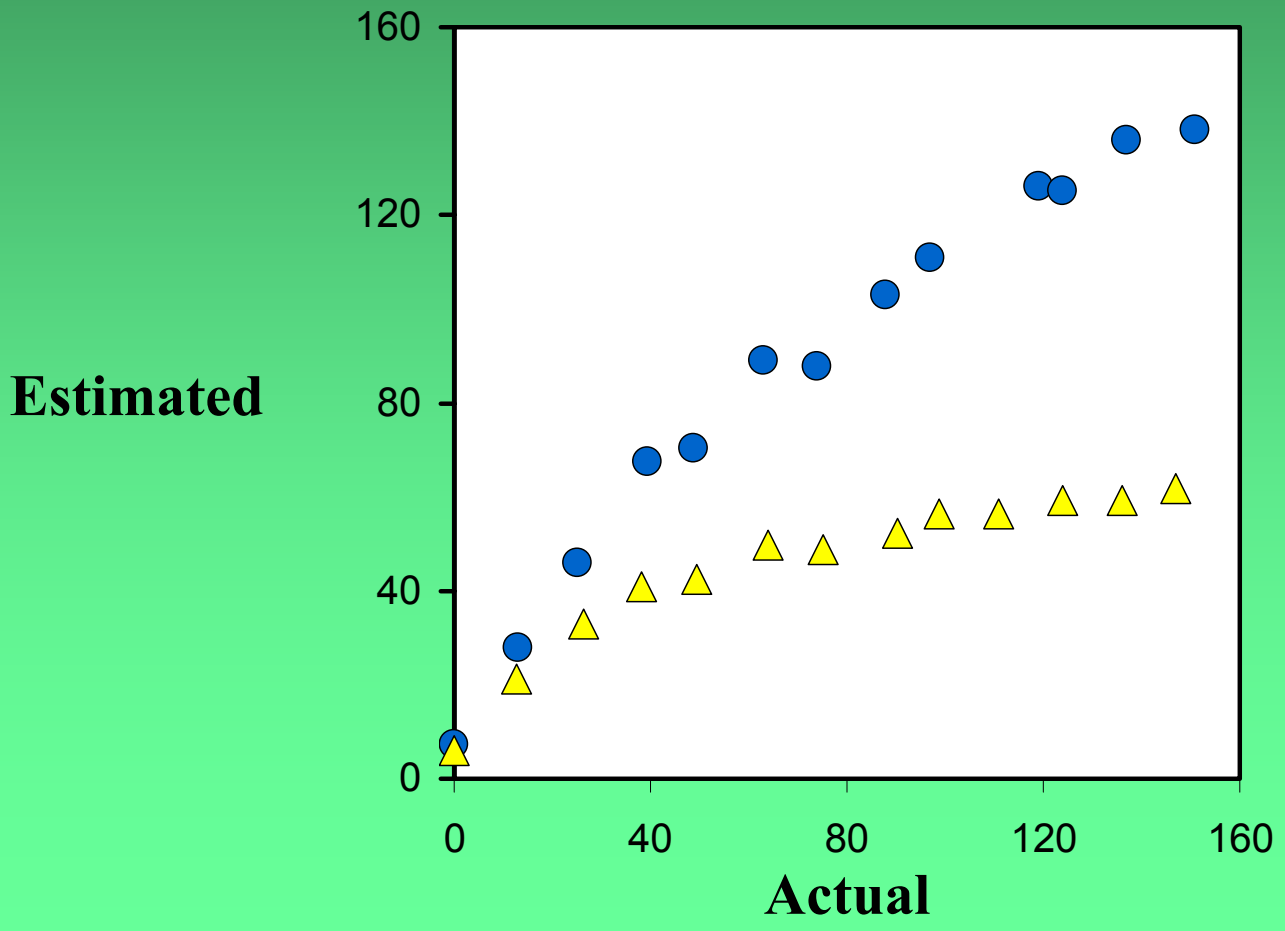
F_{st}



Genetically-based estimates:

- **sample size = 120**
- **50 generations since isolation**

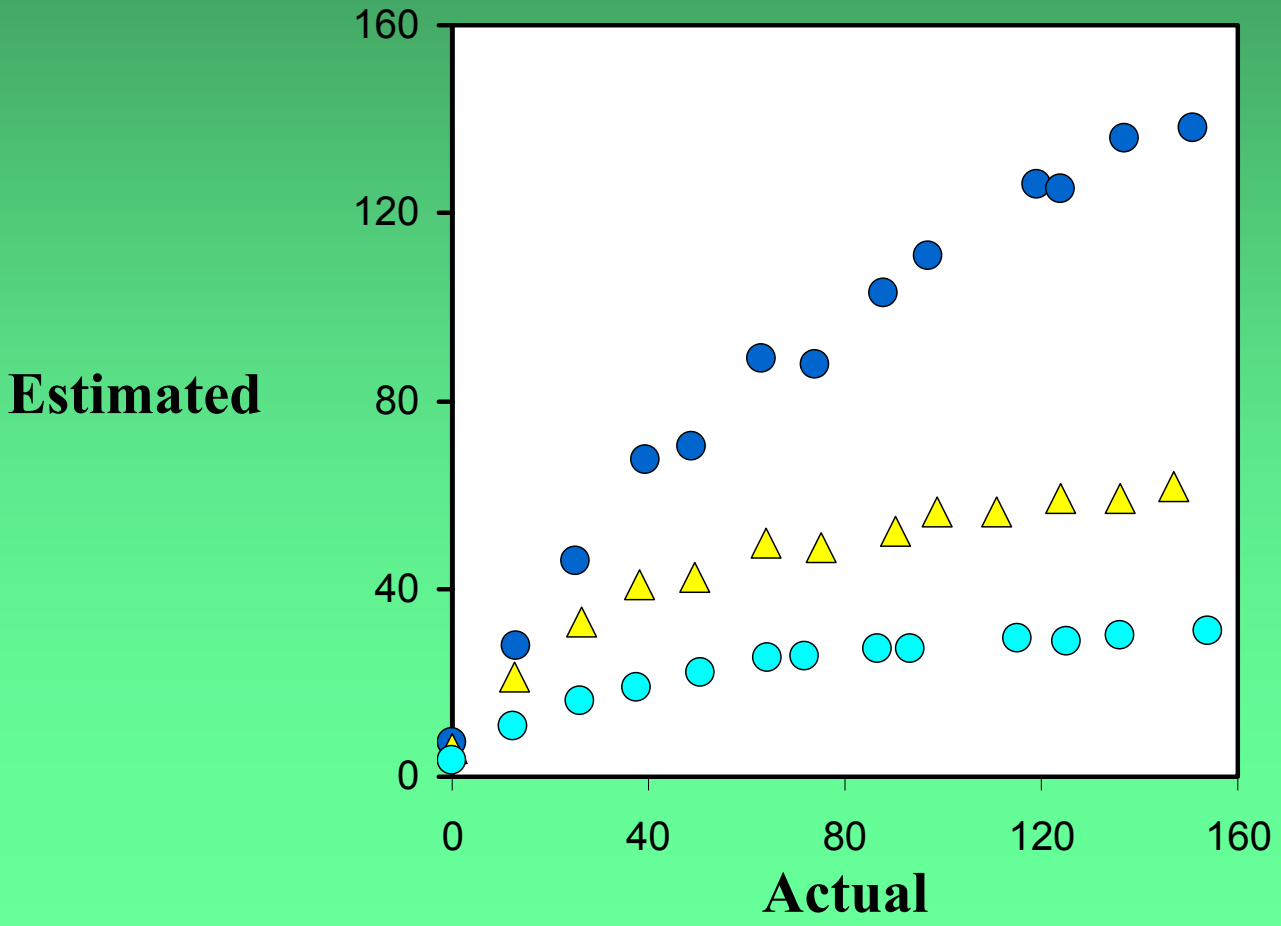
F_{st}; G_{st}



Genetically-based estimates:

- **sample size = 120**
- **50 generations since isolation**

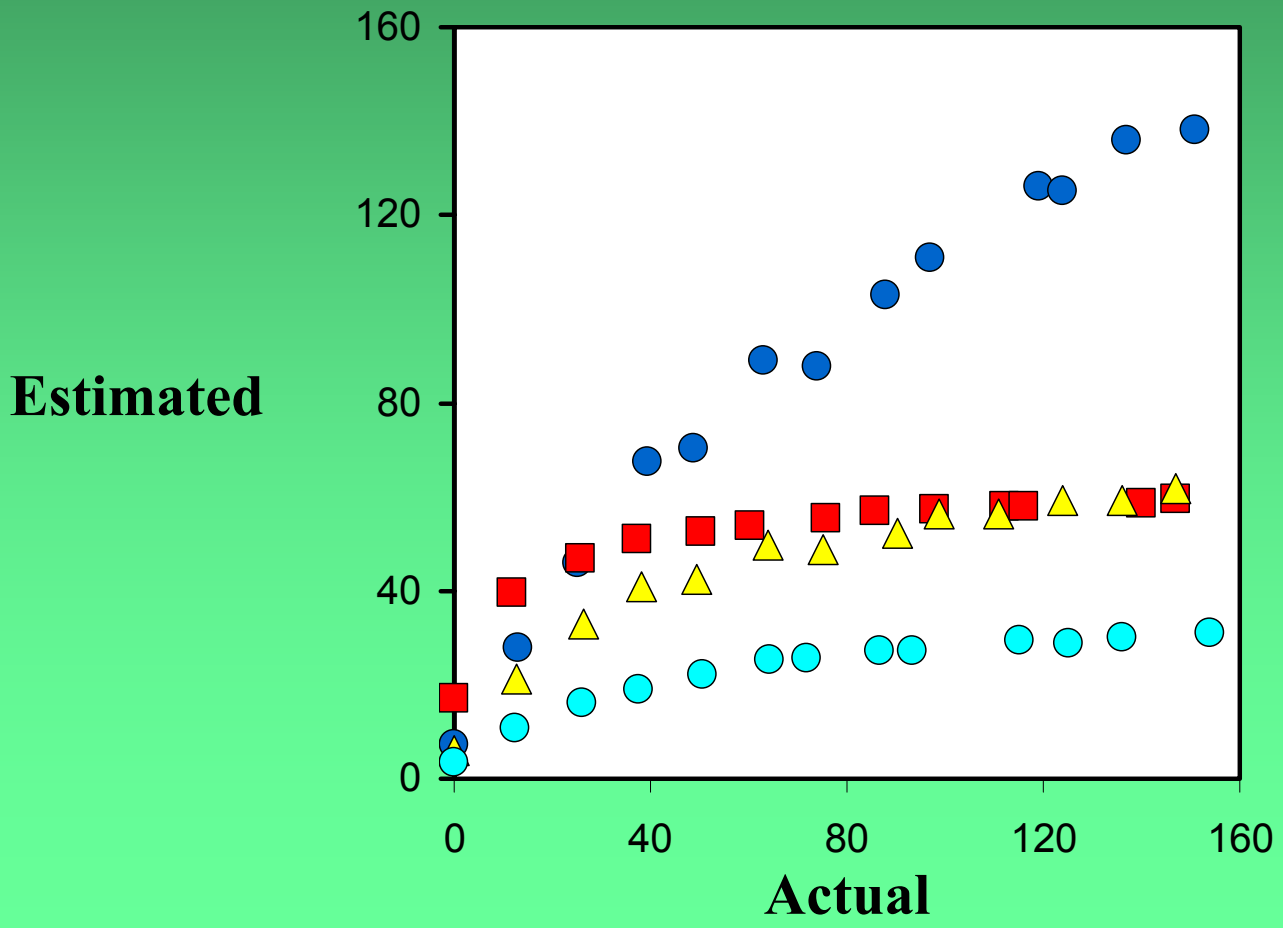
F_{st}; G_{st} ; G'_{st}



Genetically-based estimates:

- **sample size = 120**
- **50 generations since isolation**

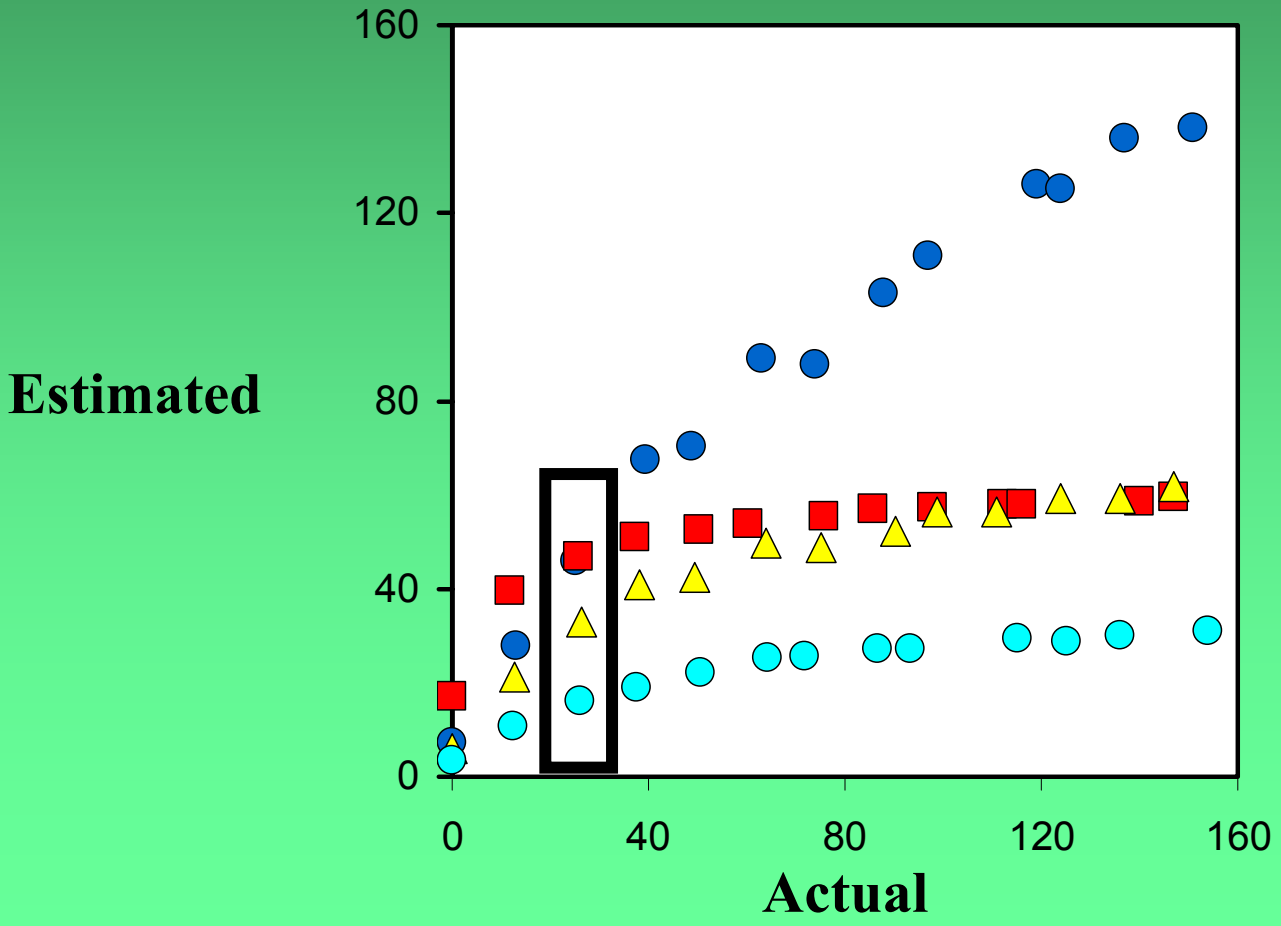
F_{st}; G_{st} ; G'_{st}; GAT



Genetically-based estimates:

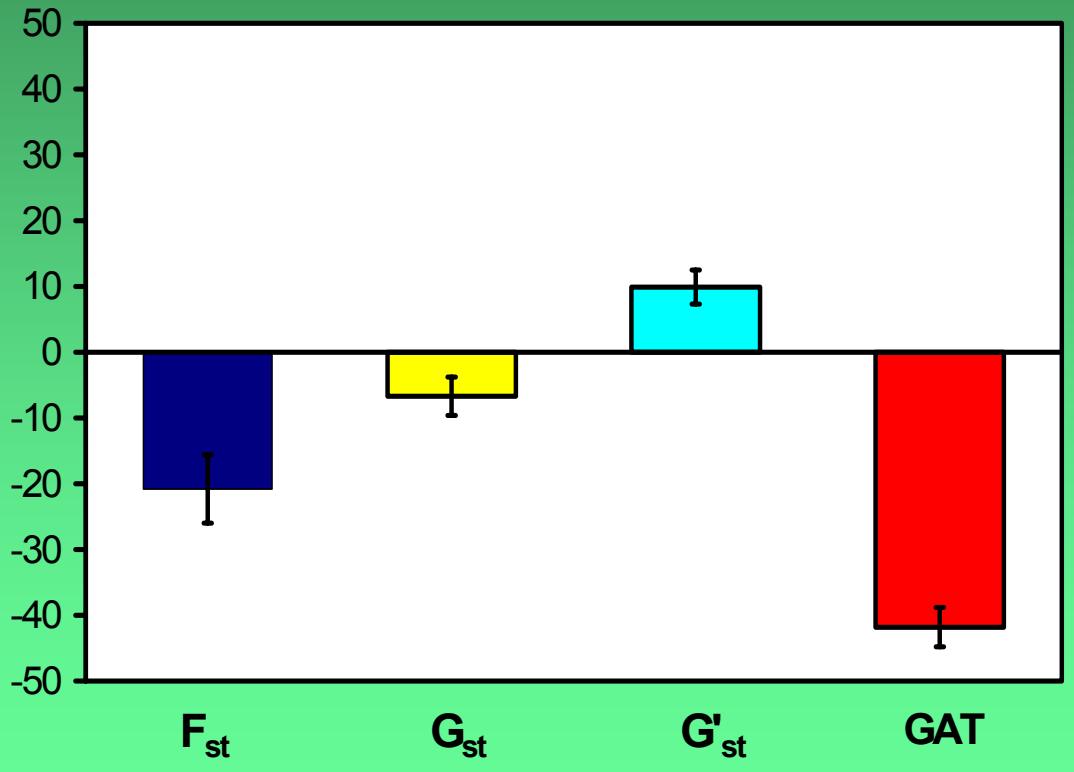
- **sample size = 120**
- **50 generations since isolation**

F_{st}; G_{st} ; G'_{st}; GAT



Unbiased and consistent?

Actual-Estimated





Considerations:

- **number of generations since isolation**
- **population size**
- **sample size**
- **demography**
- **number of loci and alleles**
- **method used to analyze the genetic data**

Conclusion

- Genetic differentiation can measure movement qualitatively, sometimes
- Quantify movement within a predictable margin of error, given the knowledge of conditions
- Potential to build a “road map” to the best way to estimate movement in the field

Acknowledgements

- Landscape Ecology Lab at Carleton University
- Drs. M. Forbes and C.S. Findlay
- NSERC and FCAR scholarships to J. Brennan
- NSERC research grant to L. Fahrig
- DFG postdoctoral scholarship to L. Tischendorf

Genetic measure: Θ

$$\theta = \frac{\sigma_a}{\sigma_a + \sigma_b + \sigma_w} \approx \frac{1}{4Nm + 1}$$

$$\sigma_a = \frac{\bar{n}}{n_c}\left\{ s^2 - \frac{1}{\bar{n}-1}\left[\bar{p}\left(1-\bar{p}\right) - \frac{r-1}{r}s^2 - \frac{1}{4}\bar{h} \right] \right\}$$

$$\sigma_b = \frac{\bar{n}}{\bar{n}-1}\left[\bar{p}\left(1-\bar{p}\right) - \frac{r-1}{r}s^2 - \frac{2\bar{n}-1}{4\bar{n}}\bar{h} \right]$$

$$\sigma_w = \frac{1}{2}\bar{h}$$