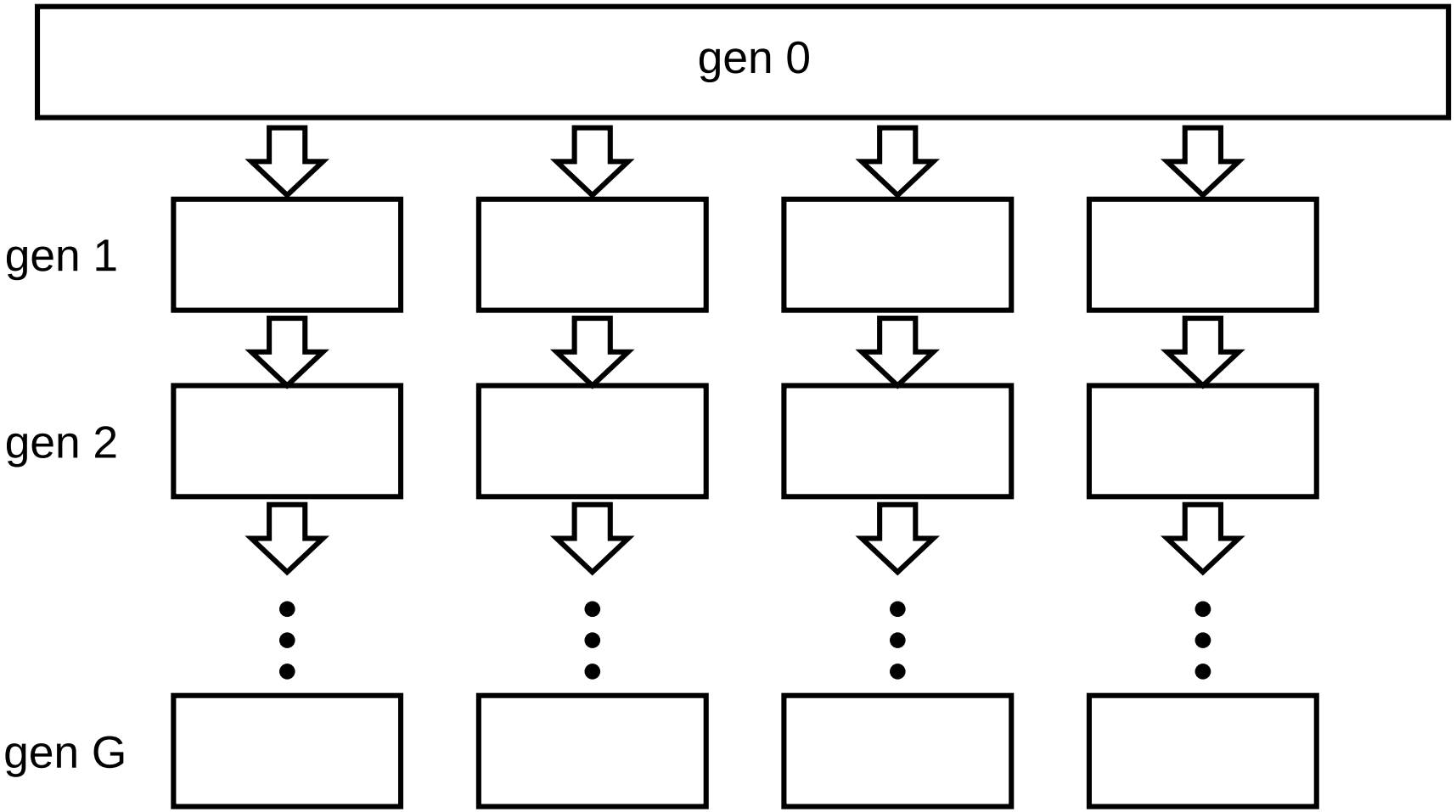
Fragmented Populations

Fragmented populations

- Large natural population with gene flow across the population is fragmented into a number of smaller populations.
  - Habitat destruction; creation of man-made barriers; alteration of landscapes.

- What are the consequences?

Fragmented populations



Fragmented populations

- The smaller subpopulations are more susceptible to the forces of drift …
  - drift acts more quickly within smaller sub-populations than it did in the larger original population.

- Variation will be lost in subpopulations.
- Sub-populations will start to diverge.

Fragmentation exercise

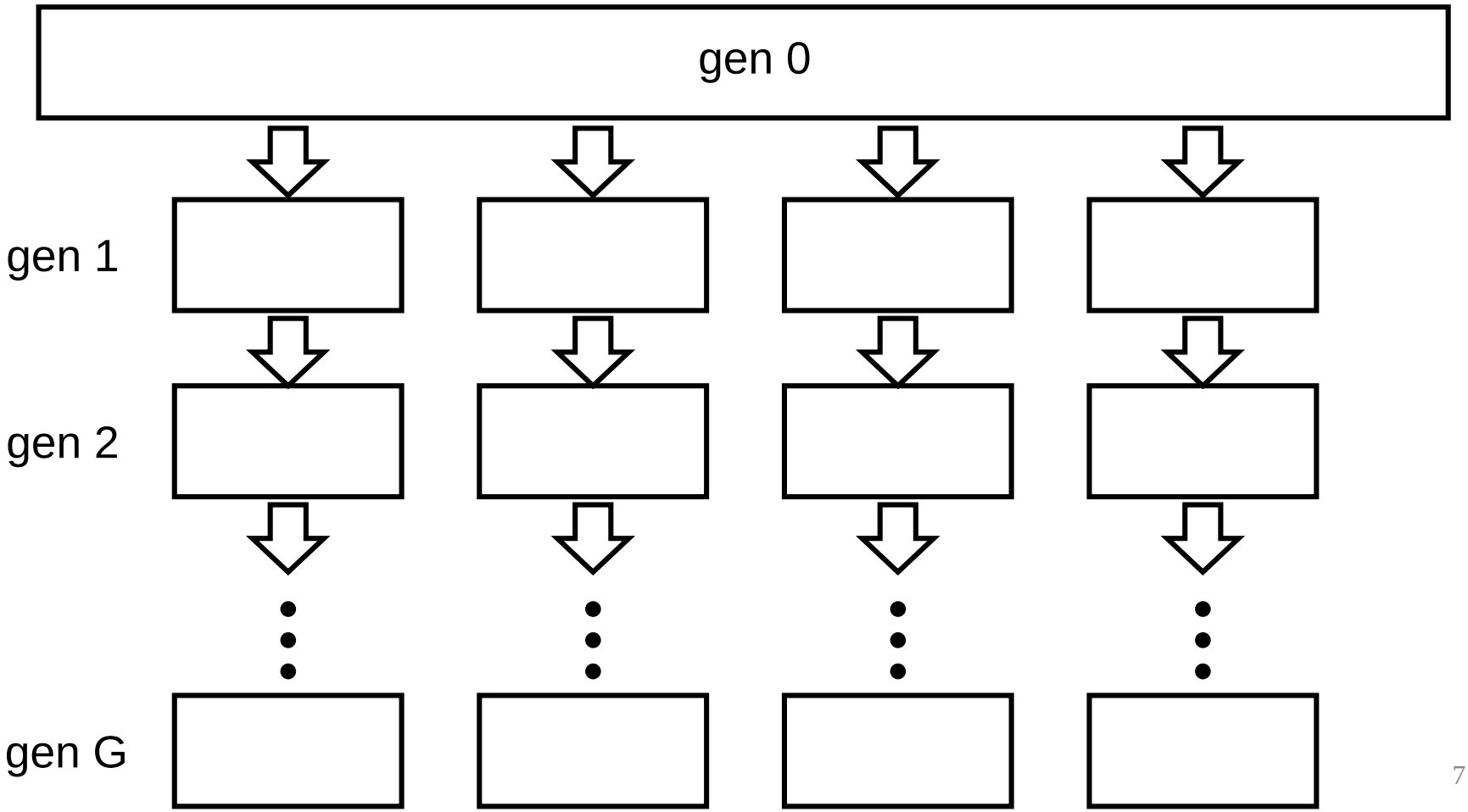
- Can we measure the amount of divergence between subpopulations?

- Run the inbreeding tool a few (say 5-6) times.
- Consider each simulation to be one subpopulation from a historically larger, now fragmented population.

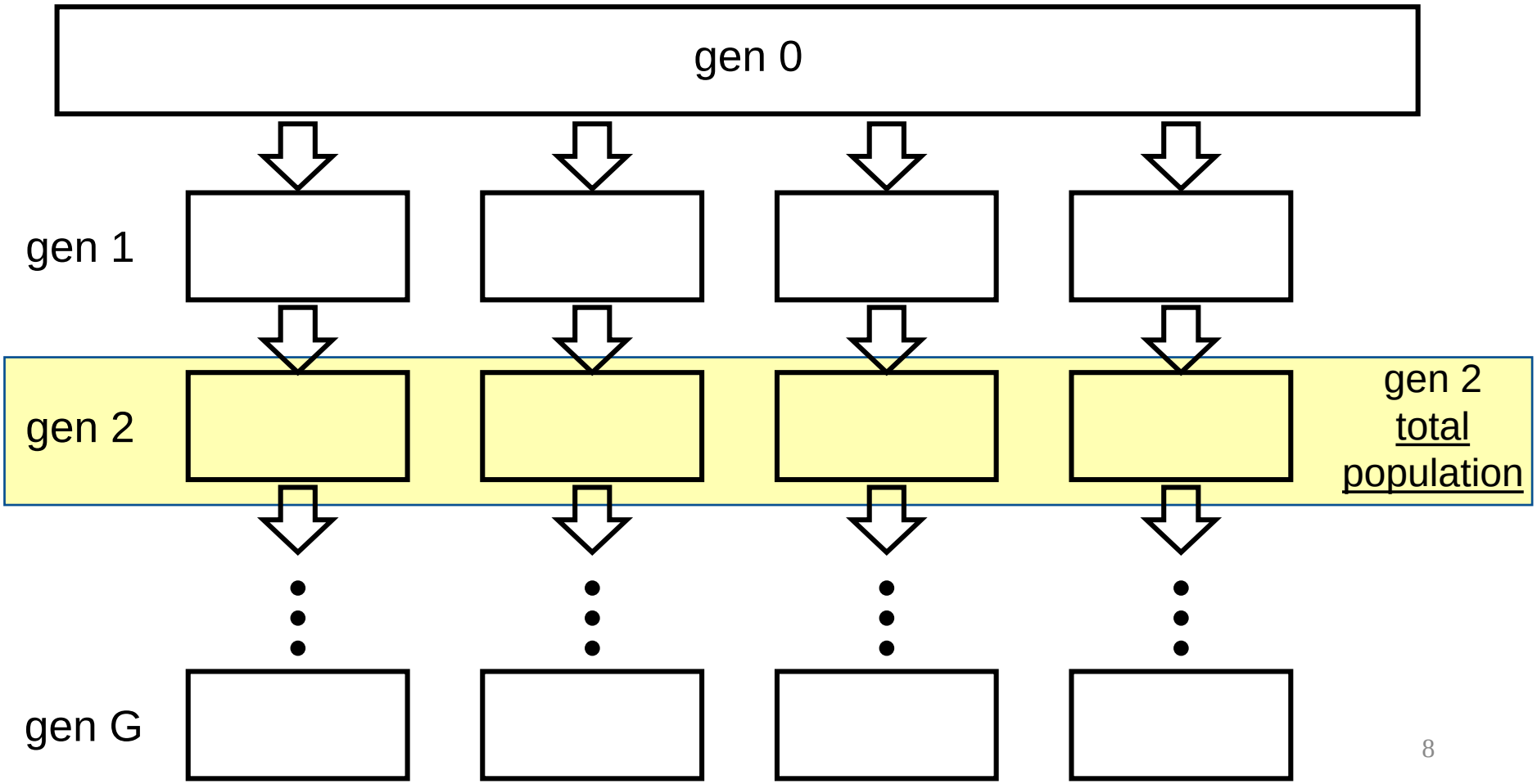
Exercise

- Extreme case: no migration between subpopulations.
- Assume random mating within subpopulations
  - can relax this assumption later.
- Consider the conglomeration of all the subpopulations to be the total population.

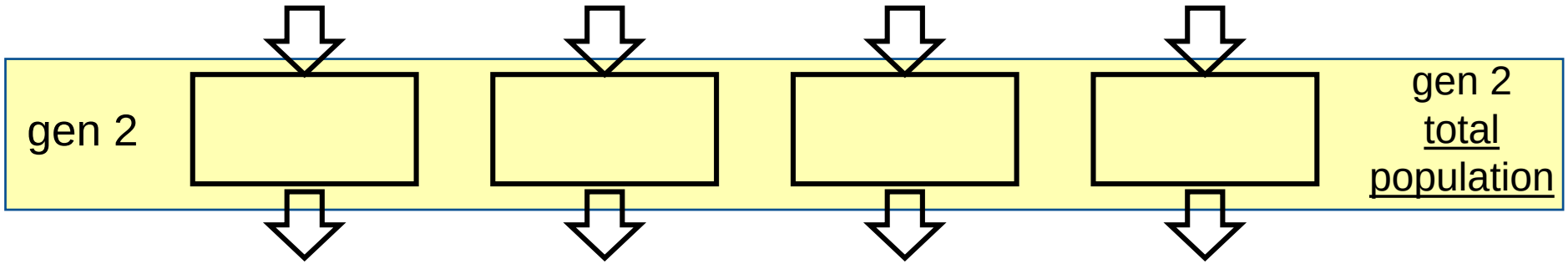
- Consider the conglomeration of all the subpopulations to be the total population.



- Consider the conglomeration of all the subpopulations to be the total population.



- Consider the conglomeration of all the subpopulations to be the total population.
- These individuals don't cross-breed: you simply collect individuals and call the collection the "total population."



(Each one on its own is a subpopulation)

Fragmentation exercise

- Run the inbreeding tool a few (say 5-6) times.
- For each run, pay attention to the amount of genetic variation you see within subpopulations in the first few generations versus the last few generations:
  - how many different alleles there are and the proportion of heterozygous genotypes.

Fragmentation exercise

- For the conglomerate total population, in which generations (first few or final few) does substantial genetic variation exist within subpopulations?

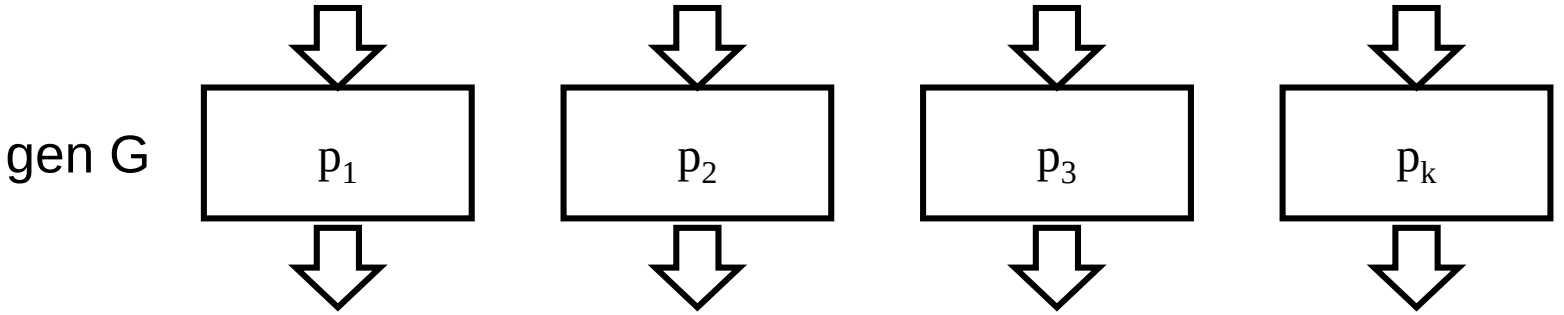
- For which generations does genetic variation exist mainly across the total population (rather than appearing within subpopulations)?

Wright's F statistics

- F_{ST} is the most commonly used.
- Measure of divergence between fragmented subpopulations.
  - Expected to be between 0 and 1.
  - Larger values indicate higher divergence of subpopulations.
- When subpopulations are highly diverged, most of the genetic variation exists at the level of the total population
  - not within subpopulations.

Interpreting F_{ST}

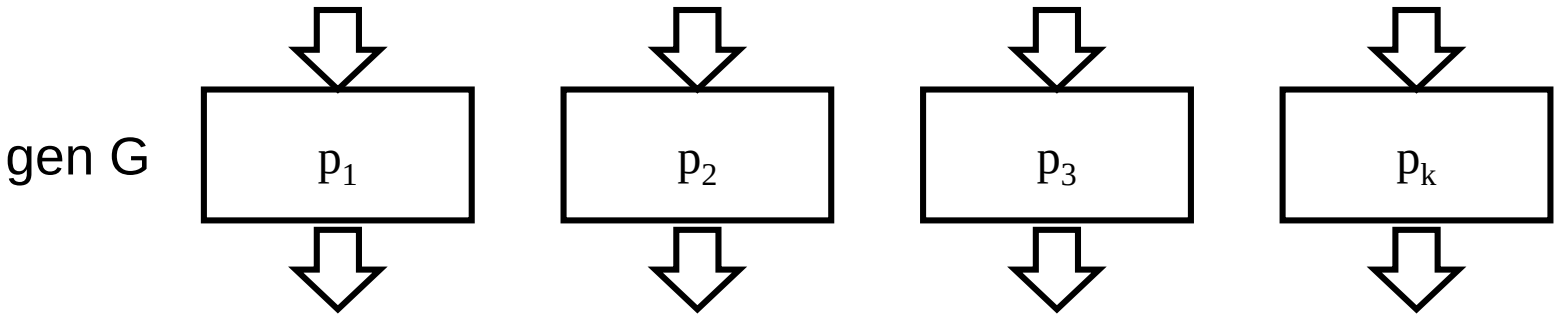
- Measure of variances of allele frequencies within subpopulations …



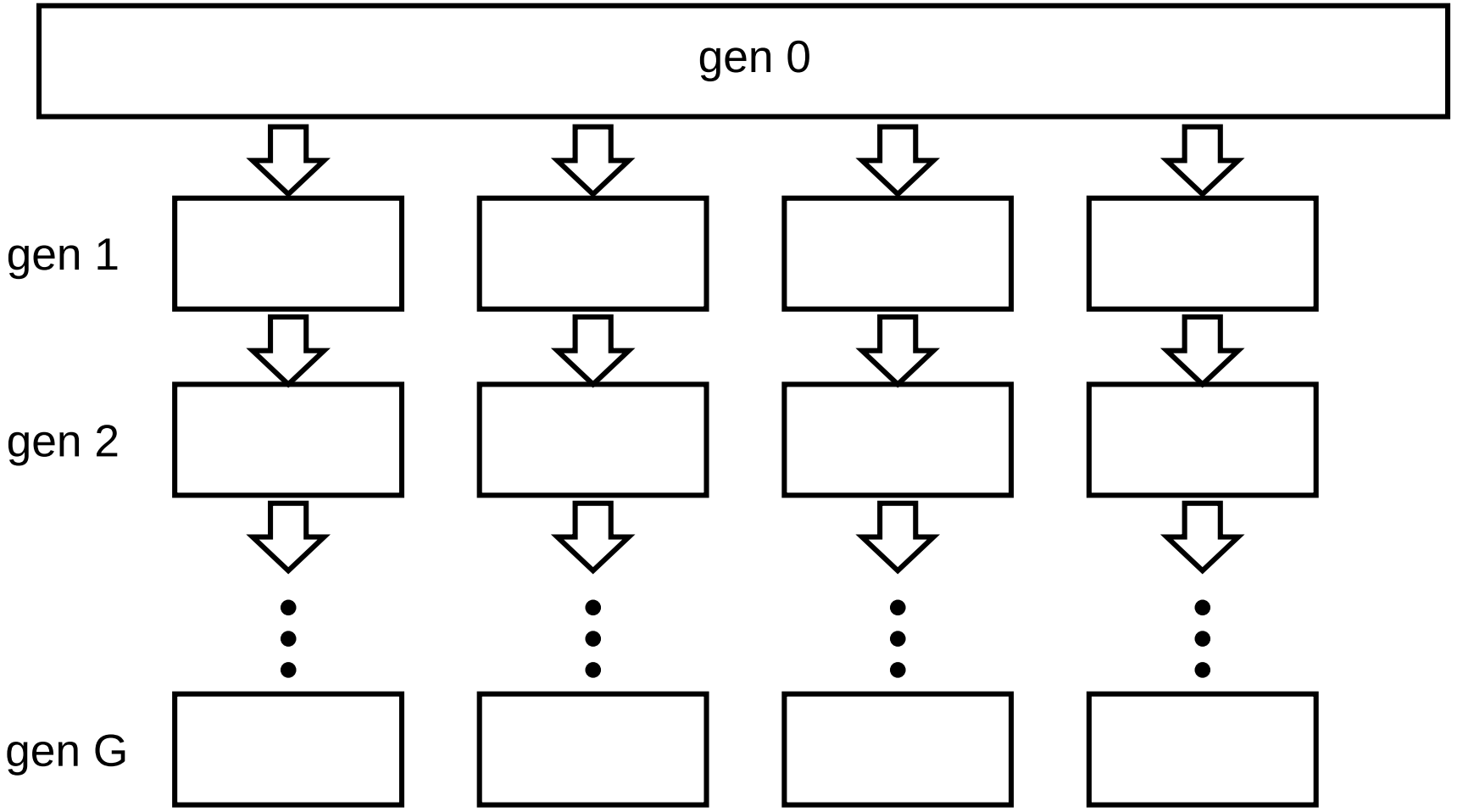
- For k subpopulations, have k allele frequencies (for a given allele), p_1, …, p_k.
- If the allele frequencies are very different between subpopulations, F_{ST} is large.
  - populations have diverged substantially.

Interpreting F_{ST}

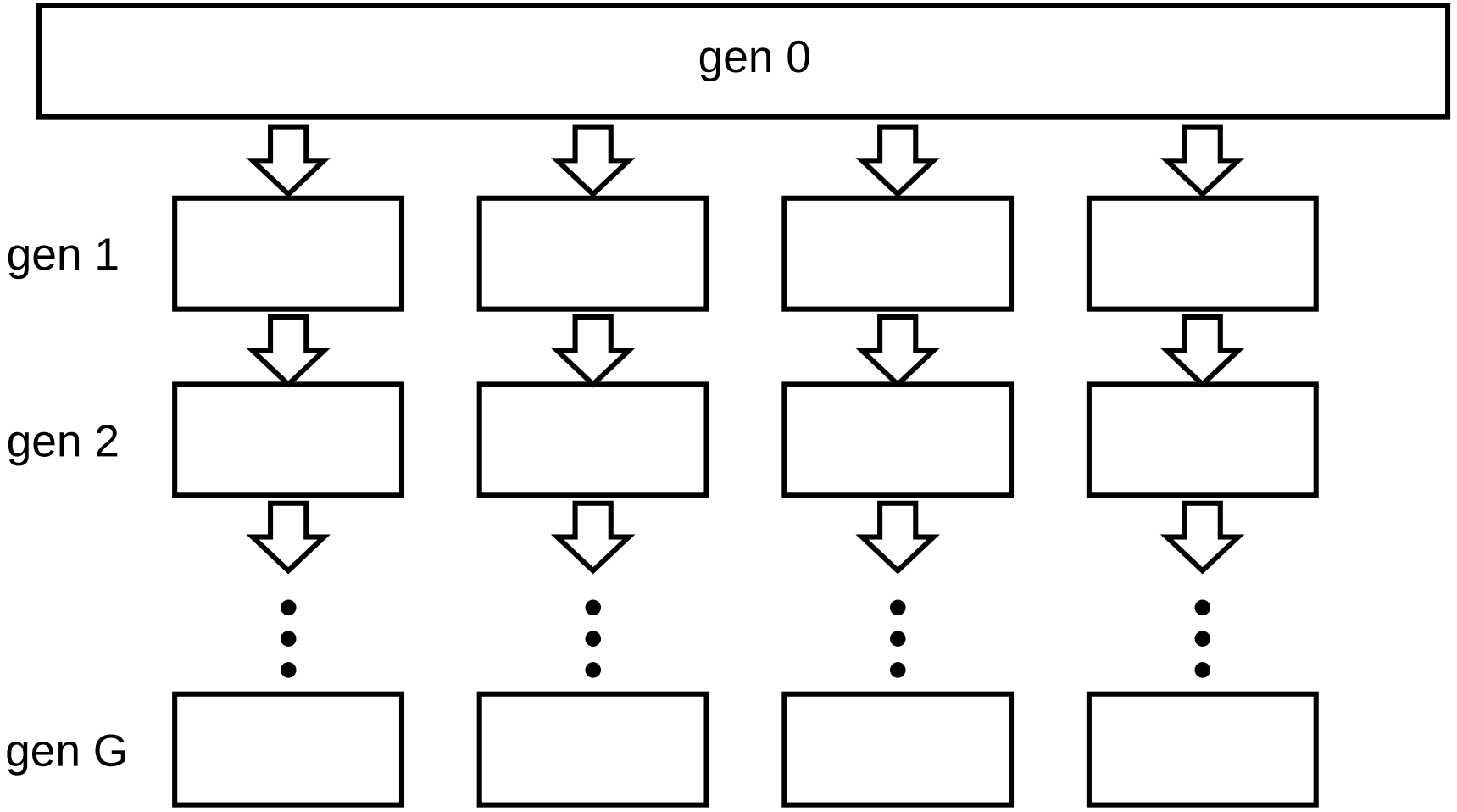
- Measure of variances of allele frequencies within subpopulations …



- For k subpopulations, have k allele frequencies (for a given allele), p_1, …, p_k.
- If the allele frequencies are very similar between subpopulations, F_{ST} is small.
  - populations have not diverged substantially.



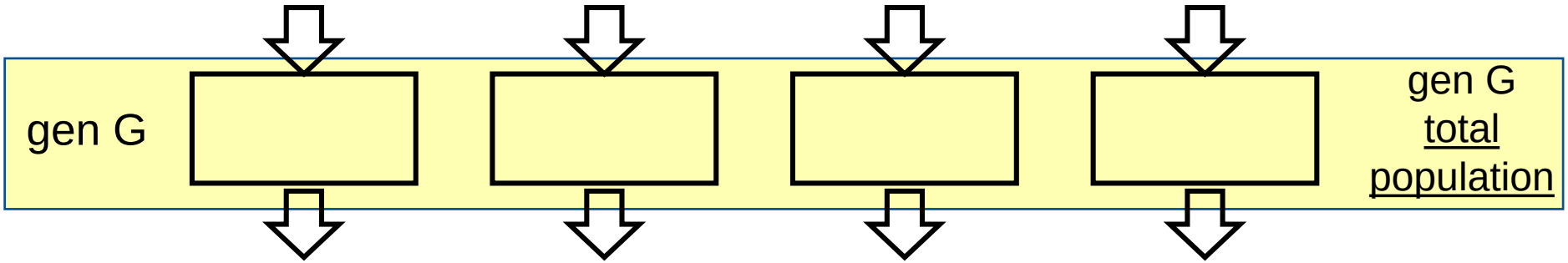
Early in the process (before much time has passed after fragmentation), the subpopulations' genetic composition will be similar to each other. Allele frequencies will be similar (variance will be low).



Later in the process, genetic variation will start to be lost within the individual subpopulations. Which alleles become rarer in each subpopulation is random. Variation in p gets larger.

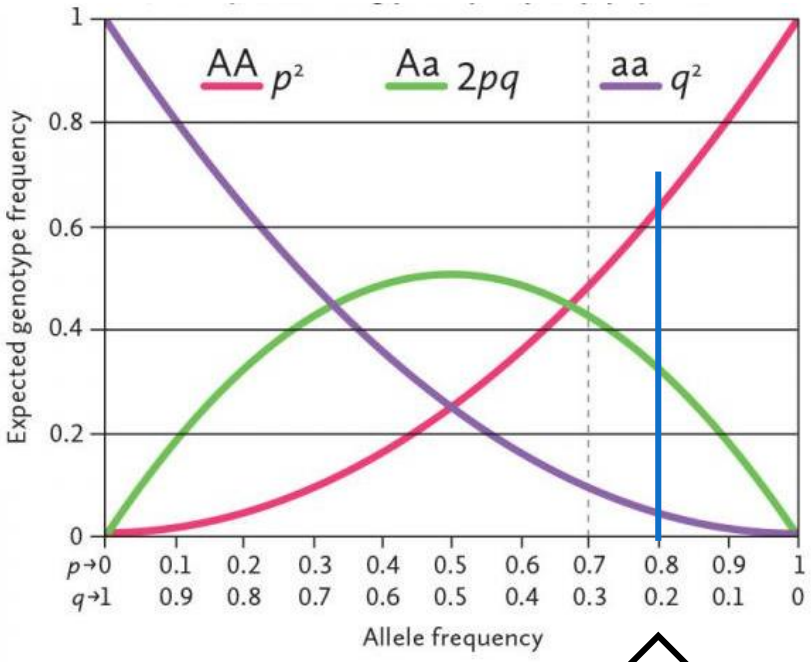
Another way to interpret F_{ST}

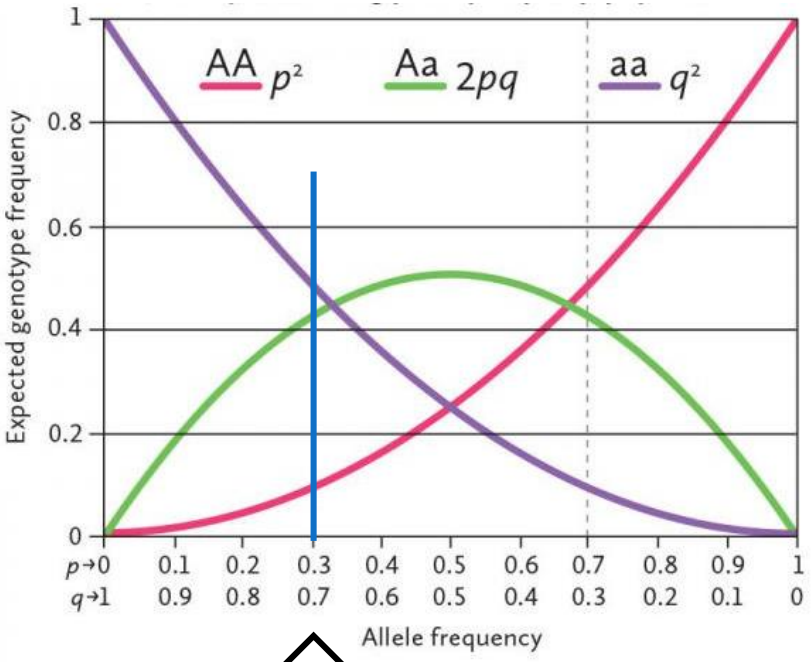
- If there is random mating (or close to it) in the subpopulations, we expect to find H-W genotype frequencies within the subpopulations.



- But, alleles are being fixed and lost within these subpopulations, so allele frequencies are going to their extremes.

H-W genotype frequencies





p in one sub-pop might be here …

… and in another one p might be here.

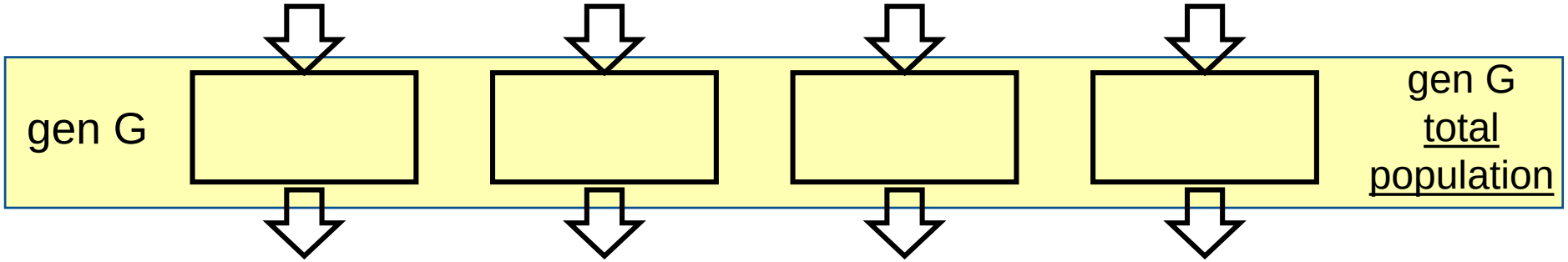
A is common in this sub-pop, so **AA** is also common.

A is rarer in this sub-pop, so **AA** is also rarer.

(p was highly similar in both sub-populations at the time of fragmentation)

F_{ST}

- What does the heterozygosity/ homozygosity of the <u>total</u> population look like?



Heterozygosity in the total pop

- Consider a locus with two alleles, A and a.
  - Let's assume a copy of A and a copy of a cannot be IBD.
- So, heterozygous genotypes cannot contain alleles that are IBD.
- What's the frequency of the Aa genotype in the total population?

Heterozygosity in the total pop

- P_{Aa} = Pr (chose two alleles from the population that are not IBD) x Pr (one of them is an A allele) x Pr (one of them is an a allele)

- $P_{Aa} = (1 - F)\ 2p_{tot}q_{tot}$

  F = Fixation index = Pr (two randomly chosen alleles from a population are IBD)

Homozygous frequencies

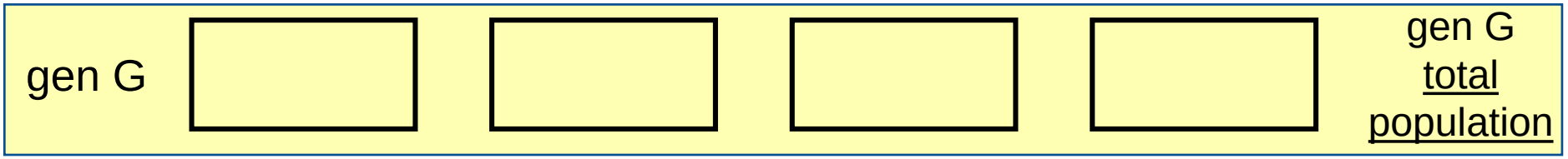
- P_{AA} = Pr (chose two alleles not IBD) x
  Pr (both of them are A) +
  Pr (chose two alleles that are IBD)
  Pr (they are an A allele)

- $P_{AA} = (1 - F)\, p^2_{tot} + F\, p_{tot}$

- $P_{AA} = p^2_{tot} + F\, p_{tot} q_{tot}$

In the total population

- $P_{AA} = p^2_{tot} + F\ p_{tot}q_{tot}$
- $P_{Aa} = 2p_{tot}q_{tot} - F\ 2p_{tot}q_{tot}$
- $P_{aa} = q^2_{tot} + F\ p_{tot}q_{tot}$



In the total population

- $P_{AA} = p^2_{tot} + F\ p_{tot}q_{tot}$
- $P_{Aa} = 2p_{tot}q_{tot} - F\ 2p_{tot}q_{tot}$
- $P_{aa} = q^2_{tot} + F\ p_{tot}q_{tot}$

H-W genotype frequencies

In the total population

- $P_{AA} = p^2_{tot}$ $+ F\ p_{tot}q_{tot}$
- $P_{Aa} = 2p_{tot}q_{tot}$ $- F\ 2p_{tot}q_{tot}$
- $P_{aa} = q^2_{tot}$ $+ F\ p_{tot}q_{tot}$

degree of departure from H-W genotype frequencies

Interpreting F_{ST}

- When considering a fragmented population, the Fixation index for the <u>total</u> population *is* Wright's F_{ST}
- Using the heterozygosity of the total population:

$$P_{Aa} = 2pq - F_{ST}\,2pq$$

$$F_{ST} = \frac{2pq - P_{Aa}}{2pq}$$

(all frequencies are for the total population, this just makes it easier to read)

Interpreting F_{ST}

- $F_{ST} = \dfrac{2pq - P_{Aa}}{2pq}$

(all frequencies are for the total population)

- 2pq is the heterozygosity expected if there is random mating across the total population (no fragmentation).
- P_{Aa} is the actual heterozygosity of the total population.

Wright's F statistics

- Subscripts indicate the level of the population structure from which alleles are being drawn.
- F_{ST} (the most commonly used)
  - "ST" – IBD of alleles w/in Subpopulations with respect to the Total population.
- F_{IT}
  - "IT" – alleles w/in Individuals wrt the Total population.
- F_{IS}
  - "IS" – alleles w/in Individuals wrt the Subpopulation.

F_{IT}

- This is an inbreeding coefficient.
- F_{IT} = Pr (two alleles within an individual are IBD wrt the total population)
  - Sample an individual from the total population, examine the two alleles that individual carries at a locus.
  - Relevant allele frequencies are p_{tot}, q_{tot}.
- If there is random mating within subpopulations, $F_{IT} = F_{ST}$.
  - (under random mating it is equivalent to sample two alleles from a subpopulation or two alleles within an individual in a subpop)

F_{IS}

- This is an inbreeding coefficient.
- F_{IS} = Pr (two alleles within an individual are IBD wrt the subpopulation)
  - Sample an individual from one subpopulation, examine the two alleles that individual carries at a locus.
  - Relevant allele freqs are $p_{subpop(i)}$, $q_{subpop(i)}$.
- If there is random mating within subpopulations, $F_{IS} = 0$.
  - IBD is only considered wrt the previous generation, not further back in time.

Other measures

- G_{ST}
  - Extension of Wrights's F Statistics theory to multiple alleles
    - e.g. microsatellites

- Φ_{ST}
  - Equation structurally similar to F_{ST}, but using nucleotide diversity measures in place of heterozygosity measures.

Hawaiian Petrel

Heredity (Edinb). 2012 Jul; 109(1): 19–28.
Published online 2012 Mar 21. doi: 10.1038/hdy.2012.7

Table 2

Population differentiation of historic and modern Hawaiian petrels based on mitochondrial and nuclear intron data sets

	Hawaii	*Maui*	*Lanai*	*Molokai*	*Kauai*
Hawaii	—	0.092[a]	0.060	NA	0.064[a]
Maui	0.068[a]	—	0.095[a]	NA	−0.030
Lanai	0.405[a]	0.543[a]	—	NA	0.145[a]
Molokai	0.226[a]	0.404[a]	0.037	—	NA
Kauai	0.511[a]	0.574[a]	0.633[a]	0.424[a]	—

Abbreviation: NA, not available.

Pairwise F_{ST} values for the *Cytochrome b* gene are below the diagonal, whereas those for a data set of sequences from three nuclear introns are above.

[a]Indicates the estimate is significantly different from zero after correction for multiple tests.

The Condor 99:812–815

GENETIC DIVERSITY AND DIVERGENCE OF ENDANGERED GALÁPAGOS AND HAWAIIAN PETREL POPULATIONS[1]

ROBERT A. BROWNE, DAVID J. ANDERSON[2] AND JEFFREY N. HOUSER
Department of Biology, Wake Forest University, Winston-Salem, NC 27109

FELIPE CRUZ
Servicio Parque Nacional Galápagos, Isla Santa Cruz, Galápagos, Ecuador

KEVIN J. GLASGOW
Department of Biology, Wake Forest University, Winston-Salem, NC 27109

CATHLEEN NATIVIDAD HODGES
Haleakala National Park, P.O. Box 369, Makawao, Maui, HI 96768

GREG MASSEY
Olinda Endangered Species Propagation Facility, 535 Olinda Rd., Makawao, HI 96768

- Based on a small number of markers, the authors estimated F_{ST} between Galapagos and Hawaiian petrels to be 1. Their conclusion is that these should be treated as different species of petrels.

Sierra Nevada Red Fox

- Native range being infiltrated by exotic populations.

Sierra Nevada Red Fox

Perrine, et al.

Conserv Genet (2007) 8:1083–1095

Table 3 Pairwise F_{ST} and Φ_{ST} estimates among three historic (pre-1950) and four modern (post-1950) California red fox populations

	Historic SN	Historic CS	Historic SV	Modern CS	Modern SV	Modern BA	Modern SC
Historic SN[a]	–	–0.08	0.51*	0.00	0.45*	0.36*	0.27*
Historic CS	–0.10	–	0.51	0.22	0.41*	0.18	0.11
Historic SV	0.42*	0.42	–	0.75*	–0.06	0.40*	0.32*
Modern CS	0.06	0.22	0.73*	–	0.54*	0.31*	0.21
Modern SV	0.54*	0.60*	–0.06	0.80*	–	0.40*	0.33*
Modern BA	0.44*	0.39*	0.40*	0.65*	0.51*	–	0.09
Modern SC	0.38*	0.31*	0.33*	0.56*	0.44*	0.18	–

Below diagonal measures are based solely on haplotype frequencies (F_{ST}); above diagonal estimates incorporate pairwise differences in sequence divergence (Φ_{ST})

[a] SN = Sierra Nevada, CS = Cascades, SV = Sacramento Valley, BA = San Francisco Bay Area, SC = Southern California

* significant at $\alpha = 0.05$ using sequential Bonferroni correction for multiple tests (Rice 1989)

- Native and exotic populations appear to remain distinct, and native populations are not highly diverged from one another.