The Family of F Coefficients

The Inbreeding coefficient, *f*

The Math: Derived from the coefficient of consanguinity(kinship): Pr(that 2 randomly chosen alleles are Identical By Descent (IBD))

The Concept: The probability that an INDIVIDUAL has two alleles at a locus that are IBD

$$f = \sum_{i=1}^{p} \left(\frac{1}{2}\right)^{n_i} (1 + f_{ai})$$

Inbreeding and heterozygosity

Some really ugly math reduces & tells us that inbreeding reduces heterozygosity:

$$H_f = H_r(1 - f)$$

Which rearranges to:

$$f = \frac{H_r - H_f}{H_r}$$

Which looks a lot like testing deviation from HW…

Inbreeding and heterozygosity

Thus there are two interpretations of *f*, the *inbreeding coefficient*

1. *f* is the probability that two alleles within an individual are IBD, and
2. *f* is the proportionate reduction in heterozygosity of an inbred population relative to the reference population

The Family of F coefficients

Sewell Wright, 1951

We can use this deviation from heterozygosity to measure population subdivision:

-Within Subpopulations (F_{IS})

-Due to Subdivision Alone (F_{ST})

-Within a total population (F_{IT})

Within Subpops: F_{IS}

Assuming that all populations are the same size, the average deviation in heterozygosities within subpops is:

$$F_{IS} = \frac{H_S - H_I}{H_S}$$

Factors affecting within population: drift, selection, inbreeding, assortative mating; can vary from -1 to +1

Due to subdivision: F_{ST}

Assuming all pops the same size, the avg. deviation in heterozygosities due to subdivision alone is:

$$F_{ST} = \left(\frac{H_T - H_S}{H_T} \right)$$

Factors affecting among subpops: drift, selection, Wahlund Effect; can vary from 0 to +1

Total population: F_{IT}

Assuming all pops the same size, average deviation in total population:

$$F_{IT} = \frac{H_T - H_I}{H_T}$$

Factors affecting at level of total pop: drift, selection, inbreeding, Wahlund; can vary from -1 to +1

The family of F coefficients…

These terms are NOT additive:

$$\left(1-F_{IT}\right)=\left(1-F_{IS}\right)\left(1-F_{ST}\right)$$

F_{ST} - dev. due to subdivision

- -This is the most commonly used of the 3
- -Used to elucidate population structure
- -Higher value indicates more division/less gene flow
- -Several variations exist:

F_{ST} - alternate versions

Nei's (1973,1977) G_{ST} developed for cases of multiple loci. His *Coefficient of Gene Diversity* looks like F_{ST}, but takes into account expected gene diversity based on multiple alleles. G_{ST} and F_{ST} are often used interchangeably in the literature.

F_{ST} - alternate versions

Hedrick (1999, 2005) pointed out some problems with Nei's G_{ST}- namely that the range may not vary from 0 to 1(as F_{ST} does) and can be small even if the subpops have non-overlapping alleles

He proposed a standardization:

Hedrick, 2005, Evolution 59(8) that is worth reading if you work with msats or mtDNA

F_{ST} - alternate versions

Slatkin (1995) proposed R_{ST} as a statistic to estimate demographic parameters from microsatellite loci.

-assumes generalized stepwise mutation model at msat loci w/in same species

-using R_{ST} and F_{ST} together may shed light on time scales of populations of interest

F_{ST} - alternate versions

Weir and Cockerham (1984) proposed Θ for *pairs* of alleles between individuals within populations.

-Depends on population size and history

-Doesn't depend on sampling scheme: numbers of alleles/locus, sample size per population, number of populations sampled

F_{ST} - alternate versions

Weir and Cockerham (1984) proposed Θ for *pairs* of alleles between individuals within populations.

-Depends on population size and history

-Doesn't depend on sampling scheme: numbers of alleles/locus, sample size per population, number of populations sampled

For anyone keeping score, this paper has been cited 3460X

So why can't we just use F_{ST} all the time?

" Under certain conditions, inappropriate for gene flow estimation & can lead to incorrect or absurd conclusions" (Neigel 2002)

-When gene flow is weak/N_e is large- small amounts of gene flow can overcome weak drift and give an F_{ST} near 0

-When F_{ST} is small, the variance of its estimator generates a large variance in estimator of N_m

So why can't we just use F_{ST} all the time?

If drift and gene flow are both weak…

-Can calculate precise equilibrium F_{ST}, but observed value may differ due to time to reach equilibrium

-This is particularly applicable to large, historically connected populations that are now isolated

So why can't we just use F_{ST} all the time?

If drift and gene flow are both weak…

-Effects of other forces (e.g. mutation and selection) may be significant. (esp. mutation)

-When N_e is small (<1000) and rate of mutation is lower than migration, insensitive to mutation rate

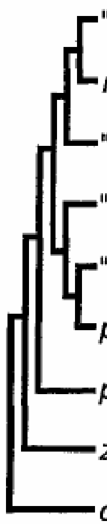
-Large populations, markers with high mutation rates (e.g. mtDNA, msats), effect of mutation can be high

So why can't we just use F_{ST} all the time?

-F_{ST} estimates are based on cumulative effects

-b/c of this averaging, not good estimates between specific pairs w/in networks or for estimating instaneous rates of gene flow

F_{ST} in action: *Rhagoletis* radiation
Berlocher 2000



phylogeny	pops analyzed	$\bar{H}$	F_{ST}	clines	unique high frequency allozymes	unique morphology ♀	unique morphology ♂
"plum fly"	2	0.206	?	?	*Fum*^{158}	yellow foreleg [1]	
mendax	26	0.154	0.015	no	*Fum*^{158}	yellow foreleg [1]	
"sparkleberry fly"	12	0.115	0.018	no	*Amy*^{143}	—	surstylus[2]
"fl. dogwood fly"	21	0.134	0.084	weak	—	—	—
"mayhaw fly"	1	0.179	?	?	—	—	—
pomonella S.	22	0.160	0.148	strong	—	—	—
pomonella N.		0.223			*Aat-1*^{100}, *Dia-2*^{100}	—	—
zephyria	2	0.115	?	?	*Had*^{111}	—	surstylus[3]
cornivora	1	0.183	?	?	*Had*^{133},8 more	ovipositor [4]	surstylus[5]

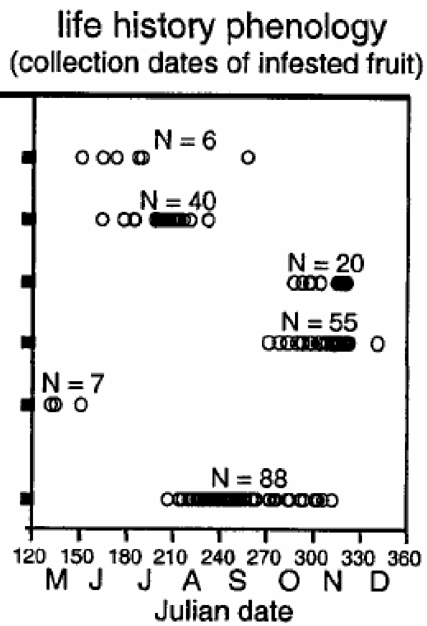


FIG. 4. Summary of population genetic, morphological data, and eclosion phenology for the *pomonella* complex. The tree is from Figure 3c. H-values are means over all samples of 17-locus values, with *Rhagoletis pomonella* values being the mean for the five southernmost and five northernmost sites in Berlocher and McPheron (1996). See text for references to genetic data and life-history phenology. Morphological characters that distinguish taxa from the rest of the group are: 1, approximately 75% of *R. mendax* and plum flies have the back of femur I yellow instead of black (Bush 1966; Jenkins 1996; S. H. Berlocher unpubl. data); 2, surstyli appear more widely spaced than in remainder of group (Bush 1966; Payne and Berlocher 1995a); 3, surstyli distinctly wider ventrally than dorsally instead of parallel (Bush 1966; Westcott 1982); 4, teeth in proximal part of oviscape with single point instead of multiple points (Jenkins 1996); 5, dorsal edge of surstylus with sharp top instead of smooth (Jenkins 1996).