SERIK SAGITOV, Chalmers Tekniska Högskola, April 14, 2008

## 2. Mutation, migration and selection

2.1 mutation
2.2 migration
2.3 selection
2.4 directional diploid selection
2.5 overdominance
2.6 other types of selection
2.7 mutation-selection balance

| Evolutionary forces interfering in HWE<br>mutation and migration increase genetic variation<br>selection and RGD decrease genetic variation |
|---|

### 2.1 Mutation

One locus two alleles model: $A$ wildtype, $a$ mutant
forward mutation rate $\mu$ per generation: $A \to a$
backward mutation rate $\nu$: $a \to A$

Typically $\mu$ is about
$10^{-4} - 10^{-6}$ mutations per gene per generation

| $p_t$ = population frequency of allele $A$ in generation $t$ |
|---|

### Irreversible mutation

If $\nu = 0$, then pure loss of alleles $A$ each generation
$p_t = p_{t-1}(1-\mu) = p_0(1-\mu)^t \approx p_0 e^{-\mu t}$
Fig 4.1 p153 (5.1 p165): $p_t \to 0$ under mutation pressure

| $\Delta p = -p\mu$ incremental frequency $\Delta p = p_t - p_{t-1}$ |
|---|

### Half-life of an allele

The number of generations halving the wildtype allele frequency
$t_{0.5} = \frac{\ln 2}{\mu} = \frac{0.693}{\mu}$ solves equation $p_t = 0.5 \cdot p_0$

If $\mu = 10^{-4}$, the half-life is $t_{0.5} = 6,930$ generations
if $\mu = 10^{-6}$, the half-life is $t_{0.5} = 693,000$ generations

**Ex 1: mutation rate estimation**

Fig 4.3 p155 (5.3 p167): infection resistance gene in *E.coli*
if the cumulative mutation rate $\mu t$ is small, then
$p_t \approx p_0(1 - \mu t)$; if moreover $p_0 \approx 1$, then $q_t \approx q_0 + \mu t$

**Ex 2: transposon deletion**

D. mauritania, a site with transposon *mariner* insertion
spontaneous deletion at rate $\mu = 0.01$: $A \to a$

If, $D_0 = 1$, find $t$ needed to reach $R_t = 0.05$:
assuming random mating $q_t = \sqrt{R_t} = 0.224$
linear approximation $q_t = 0.01 \cdot t$ gives $t = 23$
exact formula $q_t = 1 - (0.99)^t$ gives $t = 26$ generations

**Reversible mutation**

If $\mu > 0$ and $\nu > 0$, then the allele $A$ loss and gain interplay:
$p_t = p_{t-1}(1 - \mu) + q_{t-1}\nu = p_{t-1}(1 - \mu - \nu) + \nu$

$$\boxed{\Delta p = -p(\mu + \nu) + \nu}$$

Equlibrium frequency $\hat{p} = \frac{\nu}{\mu+\nu}$ solves $\Delta p = 0$
$p_t = \hat{p} + (p_0 - \hat{p})(1 - \mu - \nu)^t$

Fig 4.4 p157 (5.4 p169): $\mu = 10^{-4}$, $\nu = 10^{-5}$, $\hat{p} = 0.091$

**Ex 3: intrachromosomal recombination**

Salmonella bacteria switch between two forms of flagella at
high rates: $\mu = 8.6 \cdot 10^{-4}$, $\nu = 4.7 \cdot 10^{-3}$

Observed results for two Salmonella cultures

| $t$ | 0 | 30 | 700 |
|---|---|---|---|
| $p_t$ | 0 | 0.16 | 0.85 |

| $t$ | 0 | 388 | 700 |
|---|---|---|---|
| $p_t$ | 1 | 0.88 | 0.86 |

Expected frequencies

1: $p_t = 0.845(1 - (0.994)^t)$, $p_{30} = 0.13$, $p_{700} = 0.83$

2: $p_t = 0.845 + 0.155(0.994)^t$, $p_{388} = 0.86$, $p_{700} = 0.85$

Expected equlibrium frequency $\hat{p} = 0.845$

## 2.2 Migration

| Immigration rate $m$ into a subpopulation = the subpopulation proportion quota for new immigrants arriving each generation |
|---|

If $m = 0.05$, then 5% of the subpopulation individuals have immigrated during the last generation period

### One-way migration

Fig 6.16 p 295 (5.14 p190): mainland to island migration

mainland frequencies are fixed $p^*$, $q^*$

Island frequencies change

$p_t = (1 - m)p_{t-1} + mp^*$

$=$ {non-imm. with $A$} + {immigrants with $A$}

$$\boxed{\Delta p = -pm + mp^*}$$

Convergence to the mainland frequency

$p_t = p^* + (1 - m)^t(p_0 - p^*)$ so that $p_t \to p^*$

### Ex 4: migration rate estimation

White (Georgia) = mainland, blacks (Georgia) = island

| gene | $M$ | $S$ | $Fy^a$ | $Jk^a$ | $Js^a$ | $\beta^s$ |
|---|---|---|---|---|---|---|
| blacks (W.Africa) | .474 | .172 | .000 | .693 | .117 | .090 |
| blacks (Georgia) | .484 | .157 | .045 | .743 | .123 | .043 |
| whites (Georgia) | .507 | .279 | .422 | .536 | .002 | .000 |
| $\hat{m}$ per generation | .035 | -.013 | .011 | -.028 | -.005 | .071 |

MN data: $t = 10$, $p_0 = 0.474$, $p_t = 0.484$

$p^* = 0.507$, $(1-m)^{10} = \frac{0.507-0.484}{0.507-0.474}$, $m = 0.035$

Variation in $\hat{m}$ is mostly due to

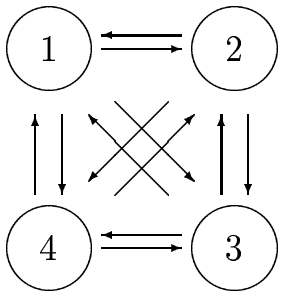
uncertainty of the origin of black Americans and

variablility of gene frequencies across West Africa

Most reliable is Duffy blood groups

since allele $Fy^a$ is absent in W. Africa

**Island model of migration**




$n = 4$

$p_t$ = allele $A$ frequency in a certain subpopulation

$\bar{p}$ = allele $A$ frequency in the metatpopulation constant over time

The same dynamics as with mainland to island migration

$p_t = (1-m')p_{t-1} + m'\bar{p}$, where $m' = m \cdot \frac{n}{n-1}$

Gene flow eliminates differences among subpopulations

$p_t = \bar{p} + (1-m')^t(p_0 - \bar{p})$ so that $p_t \to \bar{p}$

Fig 6.19 p299 (5.16 p194)

evolution similar to reversible mutation

difference in rates: $m \gg \mu$

## 2.3 Selection

### Haploid selection

Absolute fitnesses $W_A$, $W_a$

= offspring numbers for two bacteria strains $A$ and $a$

Fig 5.1 p201 (6.1 p213): two potential growth rates

Carrying capacity of the habitat is limited

focus on the allele competition within a population

relative fitnesses $w_A : w_a = W_A : W_a$

$\boxed{w_A = 1,\ w_a = 1 - s,\ \text{haploid selection coefficient } s}$

Two potential growth rates

$X_t = X_{t-1}W_A$ number of alleles $A$ in generation $t$

$Y_t = Y_{t-1}W_a$ number of alleles $a$ in generation $t$

Allele frequencies

$p_t = \frac{X_t}{X_t+Y_t}$, $q_t = \frac{Y_t}{X_t+Y_t}$ odds ratio $\frac{p_t}{q_t} = \frac{X_t}{Y_t}$

$\frac{p_t}{q_t} = \frac{p_{t-1}}{q_{t-1}}(1-s)^{-1} = \frac{p_0}{q_0}(1-s)^{-t}$

$p_t = \frac{p_0}{p_0+q_0(1-s)^t}$

$\boxed{\text{Haploid selection } \Delta p = spq, \text{ if } s \approx 0}$

Fixation of the favored allele

$p_t \to 1$ if $s > 0$ and $p_t \to 0$ if $s < 0$ as $t \to \infty$

Estimate $s$ using linear regression

$\ln(\frac{p_t}{q_t}) = \ln(\frac{p_0}{q_0}) - t\ln(1-s)$ or

$\ln(\frac{p_t}{q_t}) = \ln(\frac{p_0}{q_0}) + st$, if $s \approx 0$

**Ex 5: selection coefficient estimation**

Enzyme 6PGD in E.coli:

involved in metabolism of gluconate not ribose

Two alleles $RM43A$ $(p)$ and $RM77C$ $(q)$ code for

two allozymes of 6PGD in natural populations

| growth medium | $p_0$ | $p_{35}$ | $\ln(\frac{p_{35}}{q_{35}}) - \ln(\frac{p_0}{q_0})$ | $\hat{s}$ |
|---|---|---|---|---|
| gluconate | 0.455 | 0.898 | 2.36 | 0.065 |
| ribose (control) | 0.594 | 0.587 | $-0.029$ | $-0.0008$ |

Allele $RM43A$ is better fit to the pure gluconate media

control result is not significant

## Diploid selection

| genotype | $AA$ | $Aa$ | $aa$ |
|---|---|---|---|
| relative fitness | $w_{AA}$ | $w_{Aa}$ | $w_{aa}$ |

Three types of the diploid selection

directional selection:

$w_{AA} > w_{Aa} > w_{aa}$ or $w_{AA} < w_{Aa} < w_{aa}$

overdominance: $w_{Aa} > w_{AA}$ and $w_{Aa} > w_{aa}$

stabilizing selection against homozygotes

underdominance: $w_{Aa} < w_{AA}$ and $w_{Aa} < w_{aa}$

disruptional selection against heterozygotes

Biological components of human fitness

survival to maturity, mating success, and fertility

Two stage life history model

adults $\rightarrow$ random mating $\rightarrow$ newborns $\rightarrow$ adults

fitness is proportional to P(Survival to maturity)

Genotype frequencies

in adults $(D, H, R)$ and newborns $(d,\ h,\ r)$

From newborns to adults survival to maturity

$D : H : R = dw_{AA} : hw_{Aa} : rw_{aa}$

From adults to next generation newborns

random mating $d_{\text{next}} = p^2,\ h_{\text{next}} = 2pq,\ r_{\text{next}} = q^2$

Two relations combined

$D_{\text{next}} = p^2\frac{w_{AA}}{\bar{w}},\ H_{\text{next}} = 2pq\frac{w_{Aa}}{\bar{w}},\ R_{\text{next}} = q^2\frac{w_{aa}}{\bar{w}}$

Average fitness $\bar{w} = p^2w_{AA} + 2pqw_{Aa} + q^2w_{aa}$

is close to one if selection is weak

$$\Delta p = \frac{pq}{\bar{w}}(p(w_{AA} - w_{Aa}) + q(w_{Aa} - w_{aa}))$$

## 2.4 Directional diploid selection

Directional selection favoring allele $A$

$w_{AA} = 1$, $w_{Aa} = 1 - hs$, $w_{aa} = 1 - s$

diploid selection coefficient $s > 0$

Degree of dominance or heterozygous effect $h$

$h = 0$: harmful allele $a$ is recessive in fitness

$h = 1$: harmful allele $a$ is dominant in fitness

$0 < h < 1$: incomplete dominance

$h = 0.5$: additive selection

$$\boxed{\Delta p = spq(ph + q(1-h)), \text{ if } s \approx 0}$$

### Allele fixation dynamics

Directional selection eventually fixes the favored allele

Fig. 5.3 p213 (6.3 p225): three different fixation curves

$h = 0$, $\Delta p = spq^2$ $\qquad \boxed{\ln(\frac{p_t}{q_t}) + \frac{1}{q_t} = \ln(\frac{p_0}{q_0}) + \frac{1}{q_0} + st}$

$h = 0.5$, $\Delta p = \frac{s}{2}pq$ $\qquad \boxed{\ln(\frac{p_t}{q_t}) = \ln(\frac{p_0}{q_0}) + \frac{st}{2}}$

$h = 1$, $\Delta p = sp^2q$ $\qquad \boxed{\ln(\frac{p_t}{q_t}) - \frac{1}{p_t} = \ln(\frac{p_0}{q_0}) - \frac{1}{p_0} + st}$

Rare diseases: natural selection eliminates

dominant diseases more effectively than recessive ones

Additive selection is similar to haploid selection

with haploid selection coefficient $\frac{s}{2}$

### Ex 6: industrial melanism

melanic allele $A$, wildtype allele $a$

Fig 2.4 p65 (3.4 p86): if no lichens, then $A$ is a favored dominant

melanic moth frequency: 1% in 1848, 95% in 1898

Selection coefficient estimation (1 generation = 1 year)

$p_0^2 + 2p_0q_0 \approx 0.01$, $p_0 \approx 0.005$ $\quad$ $\ln(\frac{p_0}{q_0}) + \frac{1}{q_0} \approx -4.29$

$1 - q_{50}^2 \approx 0.95$, $p_{50} \approx 0.776$ $\quad$ $\ln(\frac{p_{50}}{q_{50}}) + \frac{1}{q_{50}} \approx 5.72$

Solve equation

$5.72 = -4.29 + 50s$ to find $s = 0.20$

**Ex 7: pesticide resistance**

US in 1940's 7% crops lost to insects

new environment: use of chemical pesticides

US in 1985 13% crops lost to insects

natural selection: 400 pest species evolved resistance

If $p_0$ and $p_t$ are small, then

$h = 0$: $\ln(p_t) + 1 = \ln(p_0) + 1 + st$ and $t = \frac{1}{s}\ln(\frac{p_t}{p_0})$

$h = 0.5$: $\ln(p_t) = \ln(p_0) + \frac{st}{2}$ and $t = \frac{2}{s}\ln(\frac{p_t}{p_0})$

| $s = 0.5$ | $p_t/p_0 = 10^2$ | $p_t/p_0 = 10^4$ | $p_t/p_0 = 10^7$ |
|---|---|---|---|
| $h = 0$ | $t = 9.2$ | $t = 18.4$ | $t = 32$ |
| $h = 0.5$ | $t = 18.4$ | $t = 36.8$ | $t = 64$ |

## 2.5 Overdominance

Overdominance favors heterozygotes

$w_{Aa} = 1$, $w_{AA} = 1 - s_1$, $w_{aa} = 1 - s_2$

$\Delta p = \frac{pq}{\bar{w}}(qs_2 - ps_1)$, $\bar{w} = 1 - p^2s_1 - q^2s_2$

Equlibrium frequencies $\hat{p} = \frac{s_2}{s_1+s_2}$, $\hat{q} = \frac{s_1}{s_1+s_2}$

Fig 5.4 p217 (6.4 p229): equilibrium frequency $\hat{p}$ maximizes $\bar{w}$

if $p = \hat{p}$, then $\frac{d\bar{w}}{dp} = 2\hat{q}s_2 - 2\hat{p}s_1 = 0$, $\bar{w}_{\max} = 1 - \frac{s_1s_2}{s_1+s_2}$

Segregational load $L = 1 - \bar{w}_{\max} = \frac{s_1s_2}{s_1+s_2}$

Due to gene segregation $\bar{w}$ is always less than
theoretical maximal genotype fitness $w_{Aa} = 1$
Equlibrium genotype frequencies
$D = \frac{1-s_1}{\bar{w}} \cdot \hat{p}^2$, $H = \frac{2}{\bar{w}} \cdot \hat{p}\hat{q}$, $R = \frac{1-s_2}{\bar{w}} \cdot \hat{q}^2$
excess of heterozygotes $F = 1 - \frac{1}{\bar{w}} = -\frac{s_1 s_2}{s_1 + s_2 - s_1 s_2}$

**Ex 8: sickle-cell anemia**

Fig 5.5 p219 (6.5 p231): regions in Africa with
incidences of malaria and sickle-cell anemia
Gene coding for $\beta$ chain of hemoglobin
$A$ = normal allele, $S$ = anemia allele
Relative fitnesses in Africa regions with malaria
$w_{AS} = 1$, $w_{AA} = 0.9$ (malaria), $w_{SS} = 0.2$ (anemia)
Selection coefficients and equilibrium frequencies
$s_1 = 0.1$, $s_2 = 0.8$, $\hat{p} = \frac{8}{9} = 0.89$, $\hat{q} = \frac{1}{9} = 0.11$
$\hat{q} = 0.11$ is close to the average $q$ across West Africa
considerable variation in $q$ among local populations

**Ex 9: multiple alleles selection**

West Africa: $S$ is prevalent, it is found a rare allele $C$

| genotype | $AA$ | $AS$ | $SS$ | $AC$ | $SC$ | $CC$ |
|---|---|---|---|---|---|---|
| premalarial environment | 1.0 | 1.0 | 0.2 | 1.0 | 0.7 | 1.0 |
| malarial environment $w$ | 0.9 | 1.0 | 0.2 | 0.9 | 0.7 | 1.3 |

Historical stable equilibrium $A : S = 8 : 1$, $\bar{w} = 0.911$
genotype ratio for new allele $AC : SC : CC = 8 : 1 : 0$
marginal $w_C = 0.9 \cdot \frac{8}{9} + 0.7 \cdot \frac{1}{9} = 0.878$ less than $\bar{w}$
fixation of $C$ starts from $p_C = 0.073$

**Ex 10: rat control in WWII**

Warfarin environment: a blood anticoagulant
two alleles $S$ = normal, $R$ = resistant to warfarin

Relative fitnesses in two environments
no warfarin: $w_{SS} = 1$, $w_{SR} = 0.77$, $w_{RR} = 0.46$
warfarin: $w_{SS} = 0.68$, $w_{SR} = 1$, $w_{RR} = 0.37$

Equlibrium warfarin frequency $\hat{q} = \frac{0.32}{0.32+0.63} = 0.34$
after stopping with warfarin how many generations it takes to go from $q_0 = \hat{q}$ down to $q_t = 0.01$?

Additive selection case
$1 : 0.77 : 0.46 \approx 1 : 0.75 : 0.50$ implying $s = h = 0.5$
$\ln(\frac{0.99}{0.01}) = \ln(\frac{0.66}{0.34}) + \frac{t}{4}$, $t = 16$ generations

## 2.6 Other types of selection

### Underdominance

Heterozygote inferiority: $w_{AA} > w_{Aa}$ and $w_{aa} > w_{Aa}$
$w_{Aa} = 1$, $w_{AA} = 1 + s_1$, $w_{aa} = 1 + s_2$

$$\boxed{\Delta p = \frac{pq}{\bar{w}}(ps_1 - qs_2),\ \bar{w} = 1 + p^2 s_1 + q^2 s_2}$$

Fig 5.7 p223 (6.7 p235): three equilibria
two locally stable equilibria $\hat{p} = 0$, $\hat{p} = 1$
one unstable equilibrium $\hat{p} = \frac{s_2}{s_1+s_2}$

**Ex 11: disruptive selection**

North American lacewings (insects): green or brown
two extreme colors provide camouflage in two different niches, but intermediate color offers no protection

This type of selection maintains population diversity
it might even cause one species to evolve into two

**Frequency-dependent selection**

Genotype fitness decreases with its frequency

$w_{AA} = 1 - c \cdot p^2$, $w_{Aa} = 1 - 2c \cdot p \cdot q$, $w_{aa} = 1 - c \cdot q^2$

$c$ is a positive constant of proportionality

$$\boxed{\Delta p = \frac{c}{\bar{w}} pq(q-p)(p^2 - pq + q^2)}$$

Stable equilibrium $\hat{p} = \hat{q} = 0.5$

despite heterozygote inferiority in the equilibrium state

$w_{AA} = w_{aa} = 1 - \frac{c}{4}$, $w_{Aa} = 1 - \frac{c}{2}$

## 2.7 Mutation-selection balance

Directional selection favoring allele $A$

increases $p$ $\qquad \Delta p = spq[ph + q(1-h)]$

Irreversible harmful recurrent mutation of rate $\mu$

decreases $p$ $\qquad \Delta p = -p\mu$

Combined effect $\qquad \Delta p = spq[ph + q(1-h)] - p\mu$

$$\boxed{\text{Equilibrium equation:} \quad pqh + q^2(1-h) = \frac{\mu}{s}}$$

Equlibrium frequencies of the harmful allele

$$\boxed{\hat{q} = \sqrt{\frac{\mu}{s}}, \text{ if } h = 0} \qquad \boxed{\hat{q} = \frac{\mu}{hs}, \text{ if } 0 < h \leq 1, \, p \approx 1}$$

Typically $\mu = 10^{-5}$ to $10^{-6}$ while $s = 10^{-1}$ to $10^{-2}$

$h = 1$ (dominant disease): $\hat{q} = 10^{-3}$ to $10^{-5}$

$h = 0$ (recessive disease): $\hat{q} = 3 \cdot 10^{-2}$ to $3 \cdot 10^{-3}$

To illustrate let $\mu = 5 \cdot 10^{-6}$ and $s = 1$

$\hat{q} = \sqrt{\mu} = 0.0022$ if $h = 0$ should be compared with

$\hat{q} = \frac{\mu}{h} = 0.0002$ if $h = 0.025$

**Ex 12: Huntington disease**

Severe inherited dominant disorder: degeneration of the neuromuscular system after age 35

Michigan sample frequency $\hat{q} = 5 \cdot 10^{-5}$

$w_{AA} = 1$, $w_{Aa} = 0.81$ due to late onset, $w_{aa} = 0$

Estimation of mutation rates in humans

$\hat{\mu} = \hat{q}hs \approx 10^{-5}$ mutations per gene per generation

**Ex 13: cystic fibrosis**

Two possible explanations of the polymorphism $\hat{q} = 0.02$

mutation-selection balance $w_{AA} = w_{Aa} = 1$, $w_{aa} = 0$

overdominance $w_{AA} < w_{Aa} = 1$, $w_{aa} = 0$

Mutation-selection balance $s = 1$, $h = 0$

$\mu = \hat{q}^2 = 0.0004$ unrealistically high mutation rate

Overdominance $\hat{q} = \frac{s_1}{1+s_1}$, $s_1 = 0.02$

heterozygotes are resistant against typhoid fever

2% advantage in heterozygous fitness

**Mutation load**

Mutation load = reduction in average fitness caused by recurrent harmful mutation

$$\boxed{\text{Mutation load } L = 1 - \bar{w} \text{ for } 0 \leq h \leq 1,\ s > 0}$$

Haldane-Muller principle

1. if $h = 0$, then $L = \mu$ is independent of $s$
2. if $h > 0$, then $L = 2\mu$ is independent of $s$ and $h$

The effect of deleterious mutation on the mean population fitness depends only on mutation rate and not on severity of mutations

Milder mutations are present at higher frequency whereas more severe mutations have lower frequency

**Segregation distortion**

Non-Mendelian segregation

heterozygotes $Aa$ produce a skewed ratio $k : (1-k)$ of $A$-gametes and $a$-gametes

Segregation distortion without selection

leads to fixation of allele $A$ if $k > 0.5$

Assuming random mating

$p_t = p_{t-1}^2 + k2p_{t-1}q_{t-1}$, $q_t = q_{t-1}^2 + (1-k)2p_{t-1}q_{t-1}$

$\Delta p = pq(2k-1)$ like add. selection with $\frac{s}{2} = 2k-1$

**Ex 14: segregation distortion chromosome**

SD chromosome in *D.melanogaster*

$A$ = SD chromosome, $a$ = wildtype chromosome

segregation ratio $= 0.75 : 0.25$

Selection against SD chromosome

$AA$ is lethal, $w_{AA} = 0$, $w_{Aa} = w_{aa} = 1$

Combined effect of segregation distortion and selection

$p_t = 1.5p_{t-1}q_{t-1}\frac{1}{\bar{w}}$

$q_t = q_{t-1}^2\frac{1}{\bar{w}} + 0.5p_{t-1}q_{t-1}\frac{1}{\bar{w}}$

$\bar{w} = 2pq + q^2 = q(1+p)$

Incremental and equilibrium frequencies

$\Delta p = \frac{p(0.5-p)}{1+p}$ and $\hat{p} = 0.5$