



معاونت علمی و فناوری
شبکه دانش کشاورزی
سلسله برنامه‌های ویدیو کنفرانس انتقال دانش به‌روز در گستره ملی بخش کشاورزی

عنوان:

روش‌های شناسایی نشانه‌های انتخاب و کاربرد آن در علوم دامی

سخنران:

محمود امیری رودبار

عضو هیأت علمی (استادیار) مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی صفی آباد دزفول

۶ شهریور ۱۴۰۱ - ساعت: ۸:۳۰-۹:۱۵

تنوع، سازگاری و اهلی سازی

- توزیع گونه ها در جهان ← عوامل محیطی، تغذیه ای و بیماری متنوع ← سازگاری
← تنوع

- داروین و والاس (۱۸۵۸): فراوانی صفات مفید (باعث افزایش شانس زنده ماندن و تولید مثل می شوند) با افزایش زمان بیشتر می شود.

- دانشمندان به دنبال یافتن شواهد تکاملی و سازگاری

- اهلی سازی شامل فرآیندهای تکاملی مانند رانش ژنی و همخونی، انتخاب طبیعی،
انتخاب مصنوعی



شواهد نشانه های انتخاب (Signature of selection)

- اصلاحگران دام و گیاه به عنوان اولین مشاهده گره
- اولین صفت شناسایی شده تحت انتخاب در انسان مربوط به بیماری های سلول خونی در نواحی با شیوع بالای مالاریا (Allison AC. 1954).
- با روش های ژنتیک تکاملی تعداد زیادی صفات مرتبط با سازگاری شناسایی گردید، در مرحله بعد جایگاه های ژنتیکی مرتبط معرفی گردید (forward genetics approach)
- تکنولوژی ژنومیک با کمک پویش کل ژنومی باعث کمک به یافتن ژن های کاندید برای فرآیندهای سازگاری و شایستگی شده
- شناسایی ژن های کاندید برای انتخاب موجب (i) اثبات تکامل و (ii) مفهومی بیولوژیکی از تنوع
- آللهای شناسایی شده در انسان با (i) مقاومت به بیماری های عفونی، و یا (ii) بیماری های ژنتیکی

شناسایی نشانه های انتخاب بر پایه:

- مشاهده ساده:

- صفات مرتبط با شایستگی بیشتر با احتمال بیشتری به نتاج منتقل شده و در نتیجه باعث افزایش فراوانی در جمعیت طی زمان می شود

- مشاهده ژومیک:

- انتخاب موجب انتقال غیر تصادفی آلل مطلوب به دلیل اثر فنوتیپی به نسل بعد خواهد شد

حالت های انتخاب

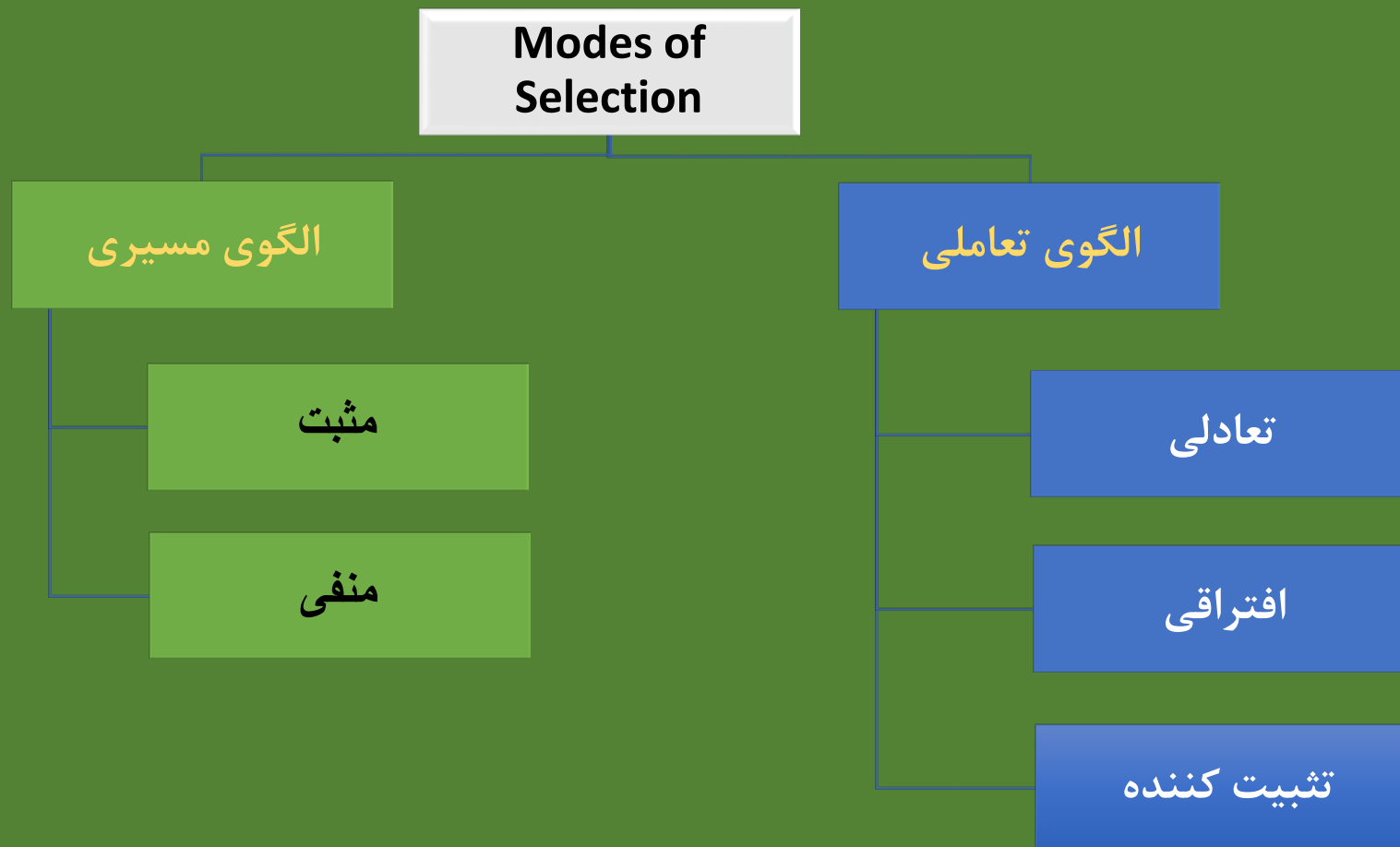
• **الگوی مسیری (Directional manner):** یک آلل مفید بوده و بصورت مثبت انتخاب شده و یا نا مطلوب بوده و به صورت منفی انتخاب می گردد

✓ جهش های تصادفی بیشتر از نوع نامطلوب (**deleterious**) هستند. جهش های نامطلوب در حال حذف یک شکلی از انتخاب منفی بوده که به عنوان انتخاب زمینه (**background selection**) عنوان می شود

• **الگوی تعاملی (Interactional manner):** بروز صفت هدف تحت تاثیر چندین آلل در یک جایگاه

- i. انتخاب تعادلی (**balancing**)، نگهداری چند آلل ← فراغلبه یا انتخاب وابسته به فراوانی
- ii. انتخاب افتراقی (**diversifying**)، نگهداری صفات متضاد ← صفات در حد وسط نا مطلوب
- iii. انتخاب تثبیت (**stabilizing**) کننده، مطلوبیت مقادیر فنوتیپی متوسط ← انتخاب تعادلی یا انتخاب مثبت برای آلل هایی که بروز صفات در حد وسط ایجاد می نمایند

حالت های انتخاب



کدام حالت انتخاب مهم است؟ و چرا؟

• بیشتر مطالعات بر روی معرفی روش های ژنومیک برای شناسایی **انتخاب مثبت** متمرکز بودند

✓ **از منظر کاربردی:** انتخاب منفی در نواحی محافظت شده مشاهده شده، و انتخاب تعادلی بسیار نامحسوس است، اما انتخاب مثبت ردپاهای بسیار محسوس روی سطح ژنوم ایجاد می کند

✓ **از منظر تئوری:** انتخاب مثبت به عنوان مکانیسم اولیه سازگاری شناخته می شود

روش های معمول برای شناسایی نواحی تحت انتخاب

- روش های کلان تکاملی (**Macroevolutionary**): بیشتر به دنبال انتخاب بین گونه ها بوده

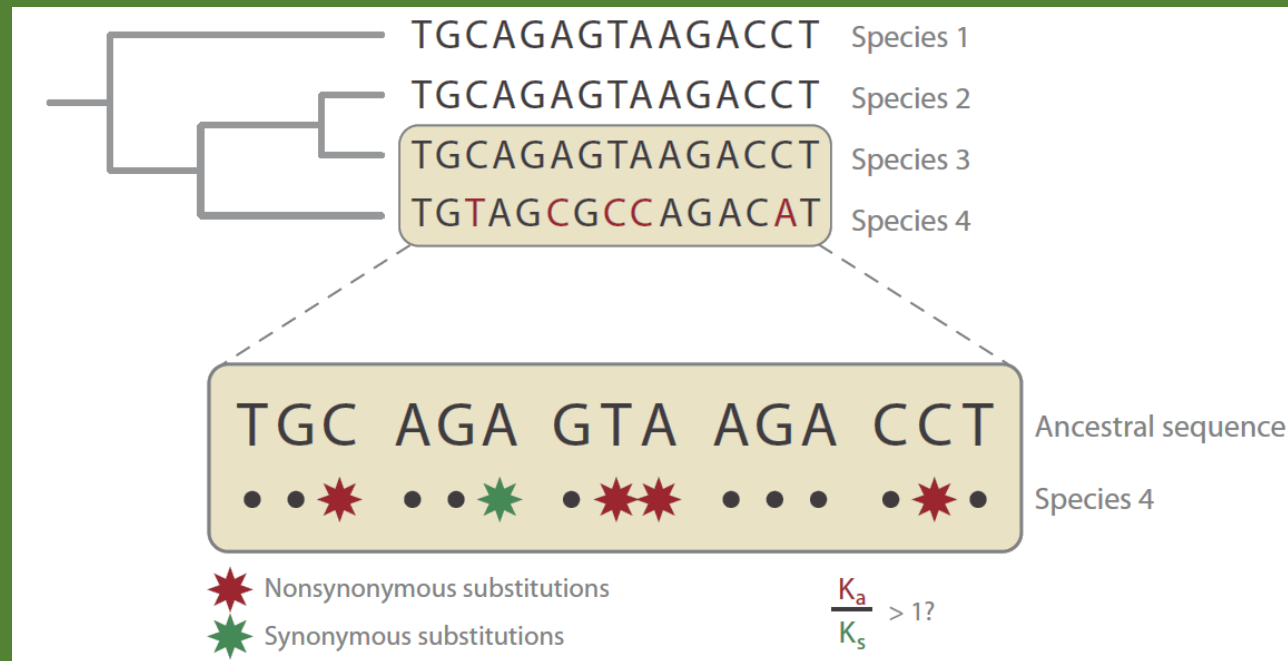
- روش های ریز تکاملی (**Microevolutionary**): به دنبال ردپای انتخاب در درون گونه ها بوده

روش های کلان تکاملی

	Approach	Intuition	Representative tests
Methods for macroevolution	Gene-based methods	Synonymous substitutions are (assumed to be) selectively neutral. Thus, they tell us about the background rate of evolution. If the rate of nonsynonymous substitution differs significantly, it is suggestive of selection.	K_a/K_s (also referred to as d_N/d_S or ω)
			McDonald-Kreitman test (MKT)
	Other rate-based methods	Levels of polymorphism and divergence should be correlated (because both are primarily functions of the mutation rate) unless selection causes one to exceed the other.	Hudson-Kreitman-Aguadé (HKA) test MKT
Regions that undergo accelerated change in one lineage but are conserved in related lineages are probable candidates for selection.			Identification of accelerated regions

Ka/Ks: روش های بر پایه ژن

• نسبت نرخ جهش های نامترادف (nonsynonymous) به مترادف (synonymous)



$$\begin{cases} \frac{K_a}{K_s} > 1 \Rightarrow \text{Positive selection} \\ \frac{K_a}{K_s} < 1 \Rightarrow \text{Negative selection} \end{cases}$$

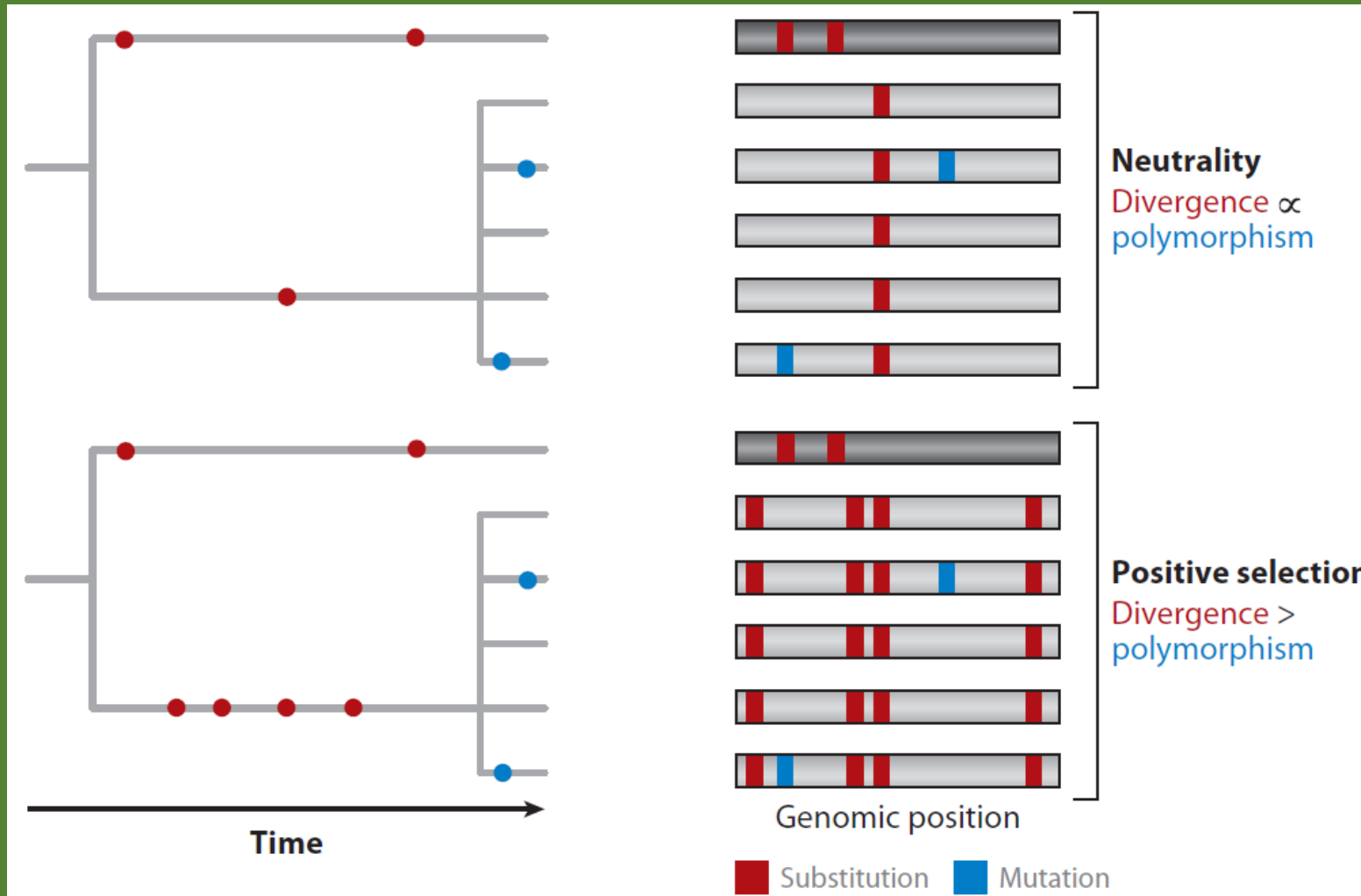
روش های بر پایه ژن: MK

$$\text{McDonald-Kreitman test (MKT)} = \frac{\frac{K_a}{K_s}_{\text{between}}}{\frac{K_a}{K_s}_{\text{within}}} \Rightarrow \begin{cases} \text{MKT} > 1 \Rightarrow \text{Positive selection} \\ \text{MKT} < 1 \Rightarrow \text{Weak negative selection} \\ \text{or balancing selection} \end{cases}$$

	<u>Within species</u>	<u>Between species</u>
Synonymous	a	b
Non-Synonymous	c	d

$$X^2 = \frac{n(ad-bc)^2}{[(a+b)(a+c)(b+d)(c+d)]}$$

روش های برپایه نرخ سایر عوامل: HKA



- استفاده از اطلاعات تنوع (diversity) و واگرایی (divergence)

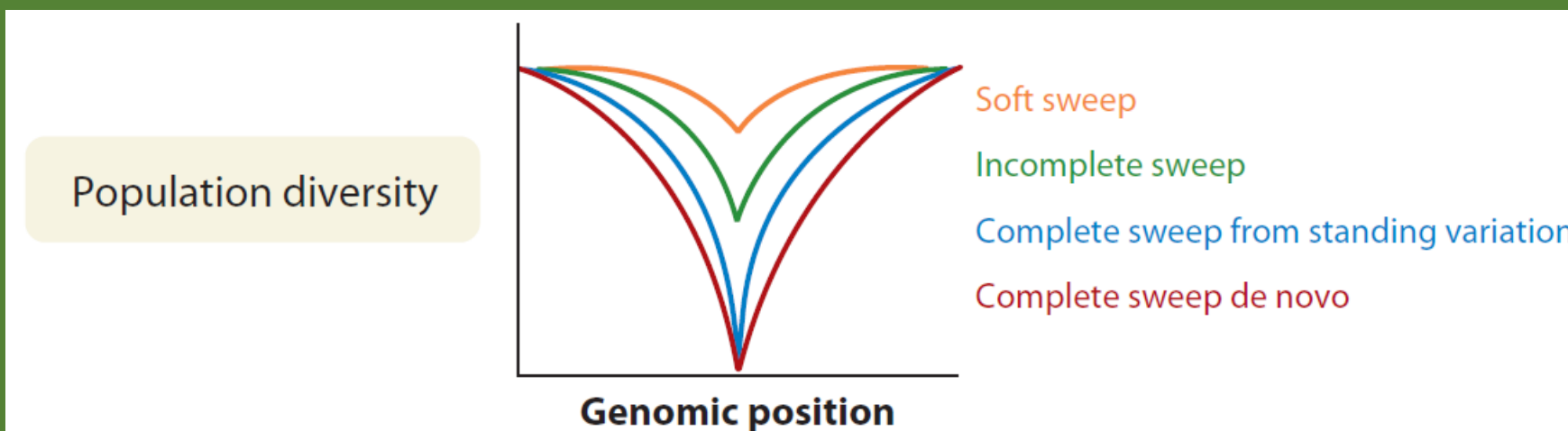
روش های برپایه نرخ سایر عوامل: حفظ شده (conserved) یا تسریع شده (accelerated)

- مقایسه اطلاعات ژنومیک برای شناسایی نواحی که بین گونه ها بسیار محافظت شده هستند، اما یک افزایش محسوس در نرخ جهش در یک گونه مشاهده می شود (مانند مقایسه ژن **HAR1F** در شامپانزه و انسان)

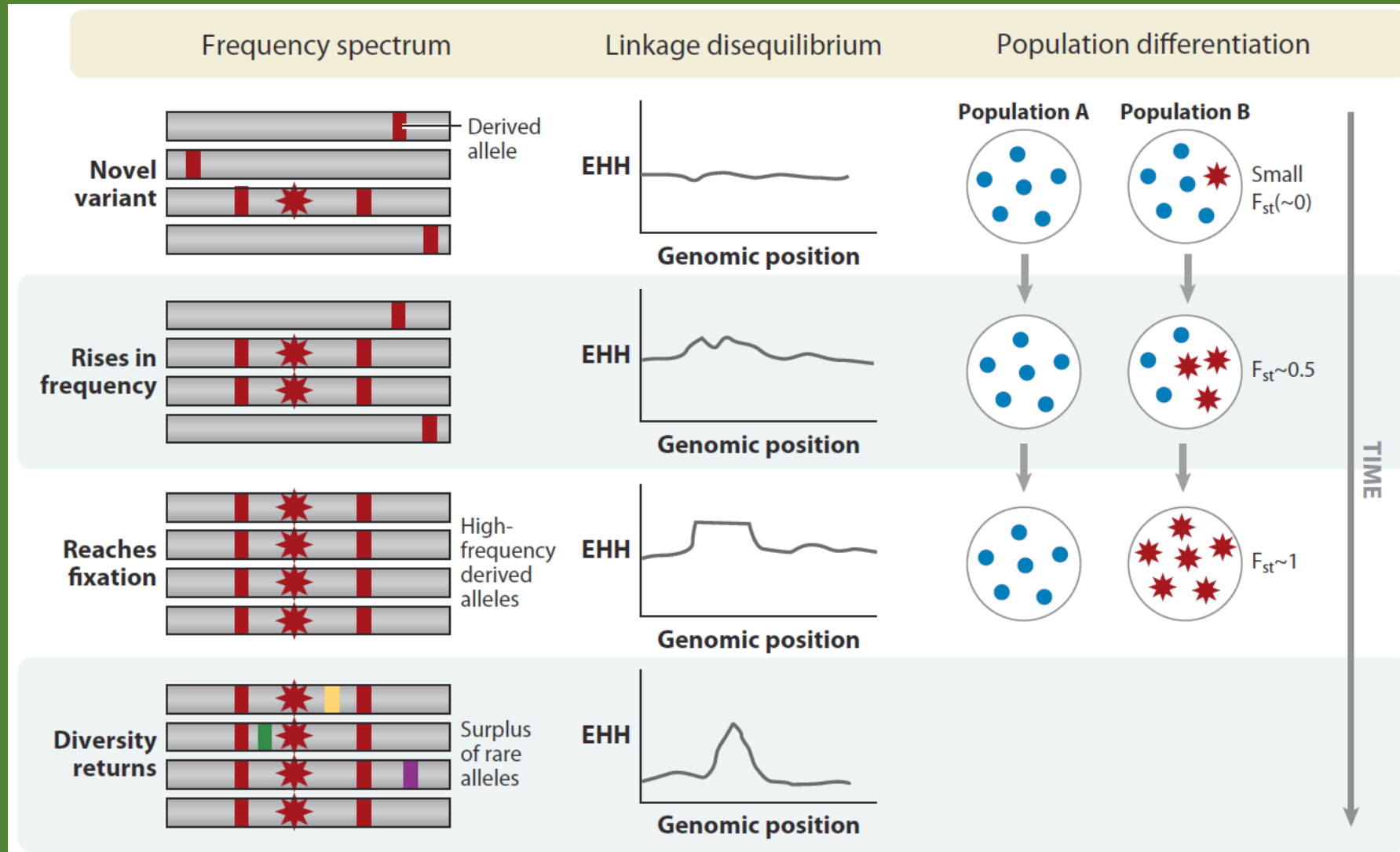


شناسایی نواحی تحت انتخاب ریز تکاملی

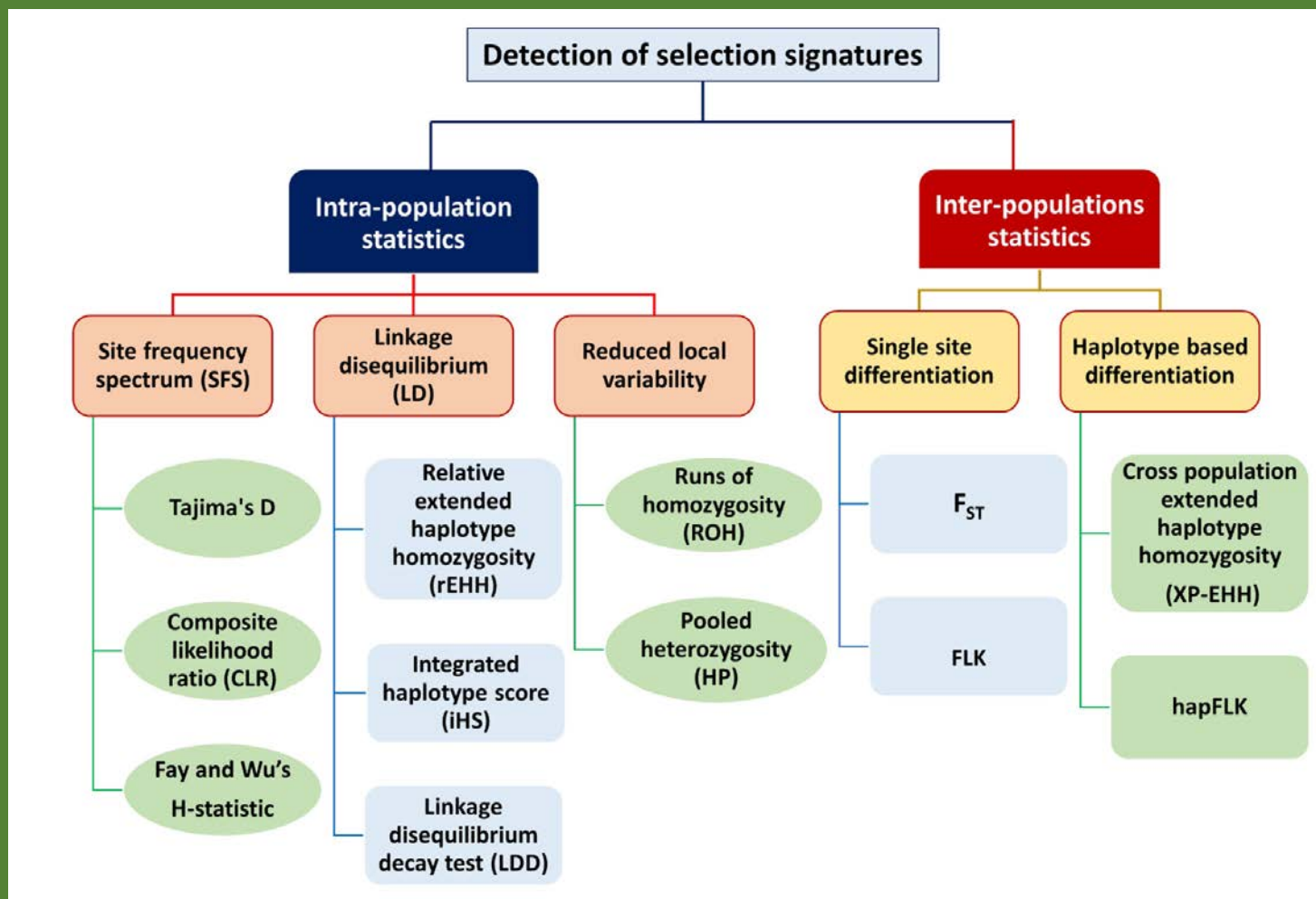
- انتخاب مثبت باعث جاروی آлл مفید به فراوانی بالاتر یا تثبیت (**fixation**) در درون یک جمعیت خواهد شد
- افزایش فراوانی آلل مطلوب به همراه واریانت های کناری (روی یک هاپلوتایپ) موجب کاهش تنوع ژنتیکی در سطح جمعیتی خواهد شد



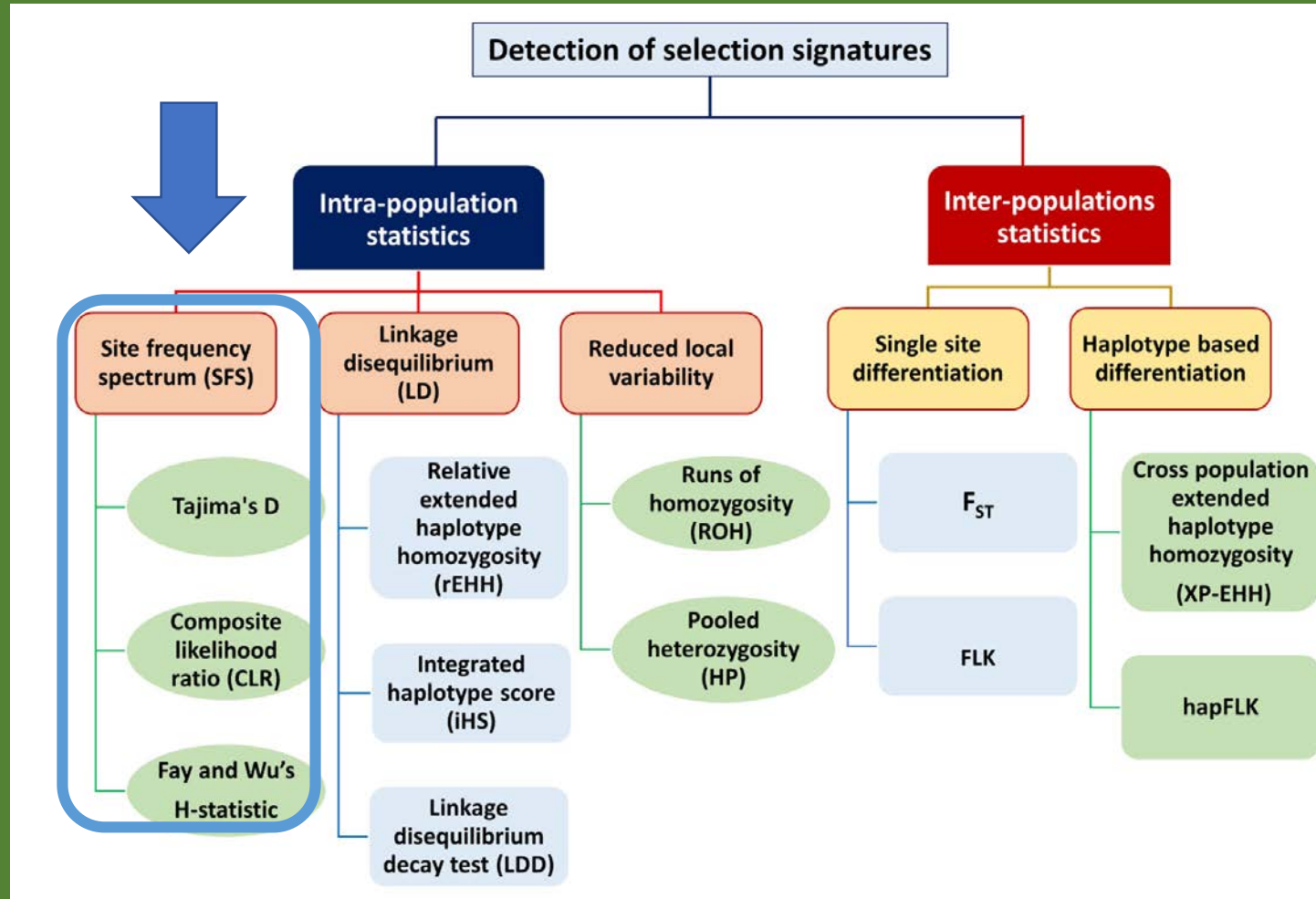
روش های شناسایی ریز تکامل: براساس آلی



روش های شناسایی ریز تکامل: براساس جمعیتی



طیف فراوانی تک جایگاهی (SFS)



SFS: Tajima's D

Tajima 1989: تفاوت جفتی میانگین (π) با تعداد کل چندشکلی های در حال تفرق (S)

$$D = \frac{d}{\sqrt{\hat{V}(d)}} = \frac{\hat{k} - \frac{S}{a_1}}{\sqrt{[e_1 S + e_2 S(S-1)]}}$$

where

$e_1 = \frac{c_1}{a_1}$	$e_2 = \frac{c_2}{a_1^2 + a_2}$
$c_1 = b_1 - \frac{1}{a_1}$	$c_2 = b_2 - \frac{n+2}{a_1 n} + \frac{a_2}{a_1^2}$
$b_1 = \frac{n+1}{3(n-1)}$	$b_2 = \frac{2(n^2 + n + 3)}{9n(n-1)}$
$a_1 = \sum_{i=1}^{n-1} \frac{1}{i}$	$a_2 = \sum_{i=1}^{n-1} \frac{1}{i^2}$

D = 0 \Rightarrow under neutrality

D < 0 \Rightarrow positive selection

D > 0 \Rightarrow balancing selection

SFS: Fay and Wu's H statistic

- آماره H (Fay and Wu, 2000): بر اساس فراوانی آللهای اجدادی و جایگزین با فرض دانستن آلل اجدادی می باشد

Tajima's D

$$\hat{\theta}_W = \left(\sum_{i=1}^{n-1} \frac{1}{i} \right)^{-1} \sum_{i=1}^{n-1} S_i,$$

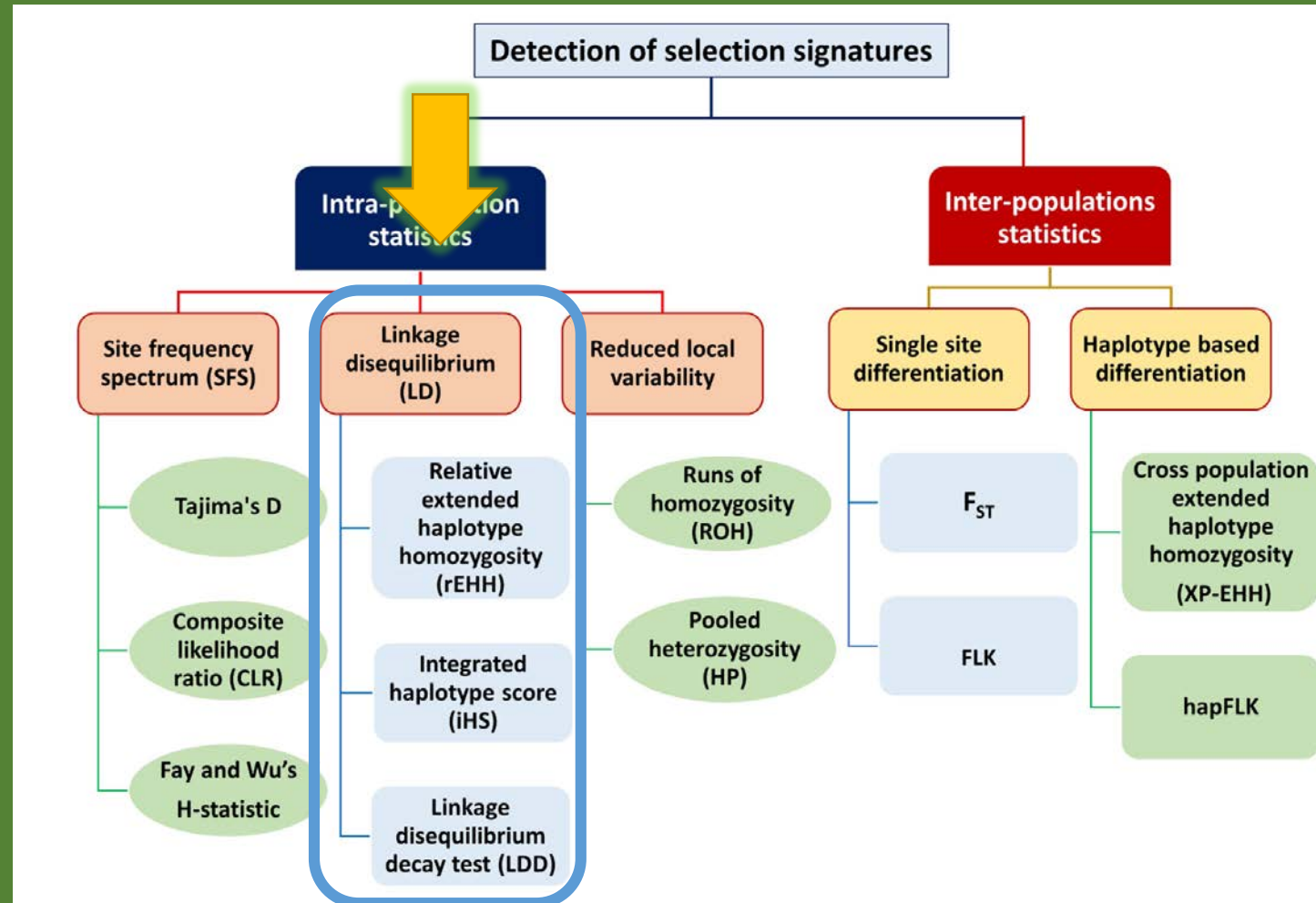
$$\hat{\theta}_\pi = \sum_{i=1}^{n-1} \frac{2S_i i (n - i)}{n(n - 1)}$$

H-statistic

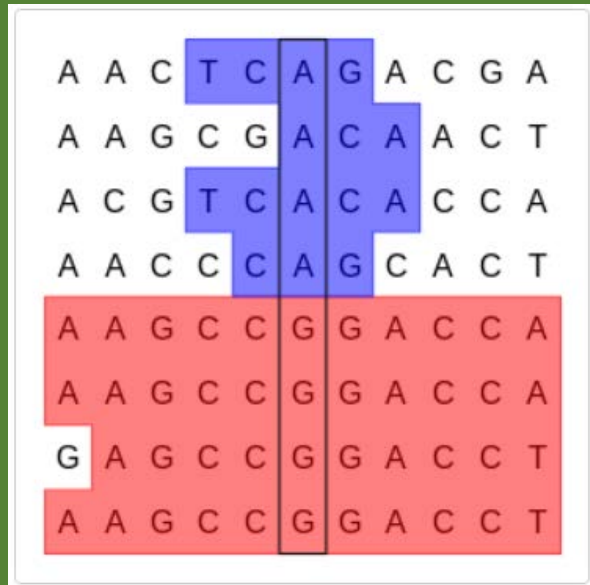
$$\hat{\theta}_H = \sum_{i=1}^{n-1} \frac{2S_i i^2}{n(n - 1)}$$

✓ این روش بر روی شناسایی نواحی تحت انتخاب جدید متمرکز بوده و برای آلل های غیر اجدادی با فراوانی میانه تا زیاد مناسب بوده و مکمل روش Tajima's D می باشد

روش های مبتنی بر اساس عدم تعادل پیوستگی



نسبت هموزیگوسیتی هاپلوتایپ گسترش یافته (rEHH)



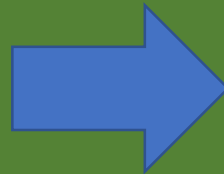
	rs1	rs2	rs3	rs4	rs5	rs6	rs7	rs8	rs9	rs10	rs11
HG1_1	0	0	1	1	0	0	0	0	0	1	1
HG1_2	0	0	0	0	1	0	1	0	1	0	0
HG2_1	0	1	0	1	0	0	1	0	0	0	1
HG2_2	0	0	1	0	0	0	0	1	1	0	0
HG3_1	0	0	0	0	0	1	0	0	0	0	1
HG3_2	0	0	0	0	0	1	0	0	0	0	1
HG4_1	1	0	0	0	0	1	0	0	0	0	0
HG4_2	0	0	0	0	0	1	0	0	0	0	0



$$EHH_{rs6,rs7}^a = \frac{1}{n_a(n_a - 1)} \sum_{k=1}^{K_{rs6,rs7}^a} n_k(n_k - 1) = \frac{1}{4 \cdot 3} (2 \cdot 1 + 2 \cdot 1) = \frac{1}{3} .$$

نسبت هموزیگوسیتی هاپلوتایپ گسترش یافته (rEHH)

A	A	C	T	C	A	G	A	C	G	A
A	A	G	C	G	A	C	A	A	C	T
A	C	G	T	C	A	C	A	C	C	A
A	A	C	C	C	A	G	C	A	C	T
A	A	G	C	C	G	G	A	C	C	A
A	A	G	C	C	G	G	A	C	C	A
G	A	G	C	C	G	G	A	C	C	T
A	A	G	C	C	G	G	A	C	C	T

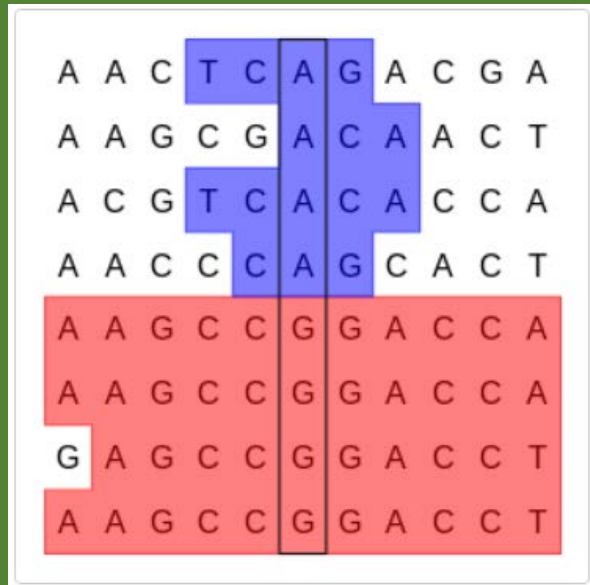


	rs1	rs2	rs3	rs4	rs5	rs6	rs7	rs8	rs9	rs10	rs11
HG1_1	0	0	1	1	0	0	0	0	0	1	1
HG1_2	0	0	0	0	1	0	1	0	1	0	0
HG2_1	0	1	0	1	0	0	1	0	0	0	1
HG2_2	0	0	1	0	0	0	0	1	1	0	0
HG3_1	0	0	0	0	0	1	0	0	0	0	1
HG3_2	0	0	0	0	0	1	0	0	0	0	1
HG4_1	1	0	0	0	0	1	0	0	0	0	0
HG4_2	0	0	0	0	0	1	0	0	0	0	0



$$EHH_{rs6,rs8}^a = \frac{1}{n_a(n_a - 1)} \sum_{k=1}^{K_{rs6,rs8}^a} n_k(n_k - 1) = \frac{1}{4 \cdot 3} (1 \cdot 0 + 2 \cdot 1 + 1 \cdot 0) = \frac{1}{6}$$

نسبت هموزیگوسیتی هاپلوتایپ گسترش یافته (rEHH)

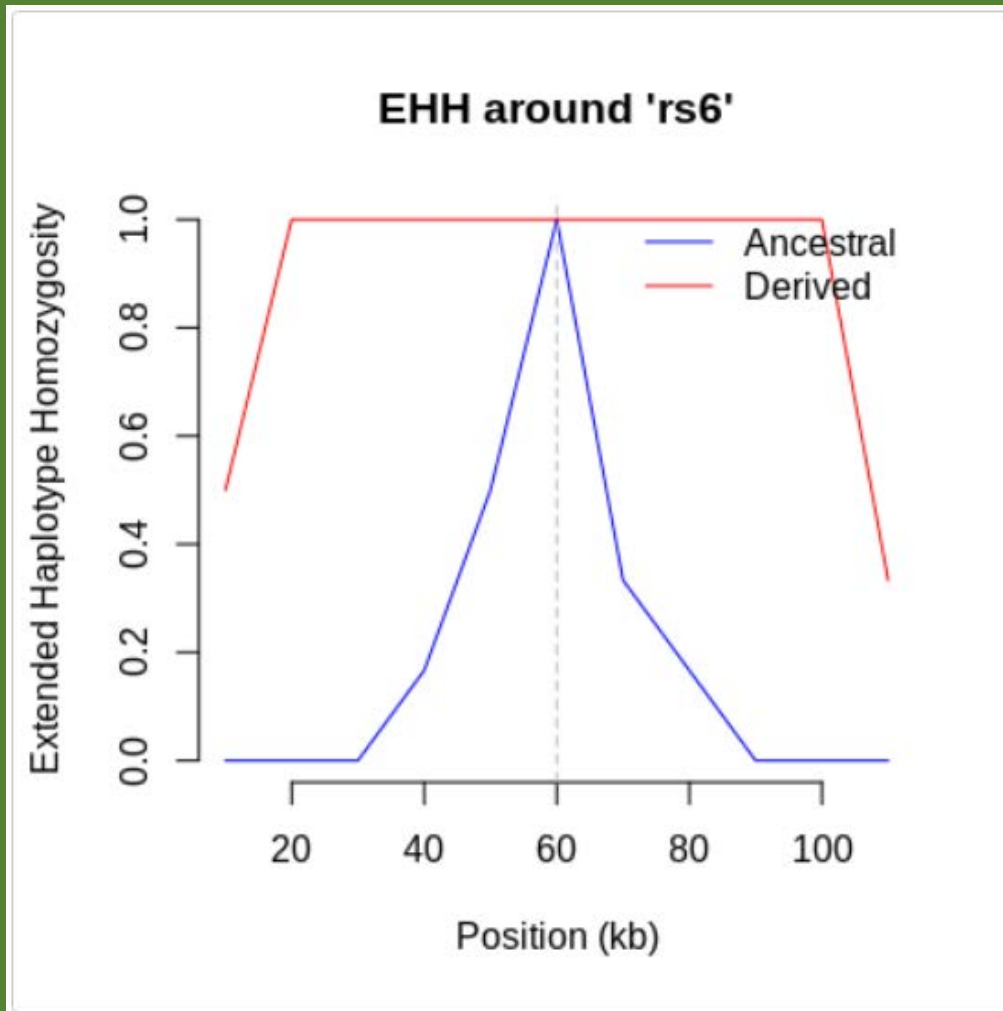


	rs1	rs2	rs3	rs4	rs5	rs6	rs7	rs8	rs9	rs10	rs11
HG1_1	0	0	1	1	0	0	0	0	0	1	1
HG1_2	0	0	0	0	1	0	1	0	1	0	0
HG2_1	0	1	0	1	0	0	1	0	0	0	1
HG2_2	0	0	1	0	0	0	0	1	1	0	0
HG3_1	0	0	0	0	0	1	0	0	0	0	1
HG3_2	0	0	0	0	0	1	0	0	0	0	1
HG4_1	1	0	0	0	0	1	0	0	0	0	0
HG4_2	0	0	0	0	0	1	0	0	0	0	0



$$EHH_{rs6,rs7}^d = EHH_{rs6,rs8}^d = \frac{1}{4 \cdot 3} \sum_{k=1}^1 4 \cdot 3 = 1$$

نسبت هموزیگوسیتی هاپلوتاایپ گسترش یافته (rEHH)



```
An object of class "ehh"
[[1]]
[1] "rs6"

[[2]]
FREQ_A FREQ_D
  0.5   0.5

[[3]]
      POSITION      EHH_A      EHH_D NHAPLO_A NHAPLO_D
rs1      10000 0.0000000 0.5000000         0         4
rs2      20000 0.0000000 1.0000000         0         4
rs3      30000 0.0000000 1.0000000         4         4
rs4      40000 0.1666667 1.0000000         4         4
rs5      50000 0.5000000 1.0000000         4         4
rs6      60000 1.0000000 1.0000000         4         4
rs7      70000 0.3333333 1.0000000         4         4
rs8      80000 0.1666667 1.0000000         4         4
rs9      90000 0.0000000 1.0000000         4         4
rs10     100000 0.0000000 1.0000000         0         4
rs11     110000 0.0000000 0.3333333         0         4

[[4]]
  IHH_A  IHH_D
18816.67    NA
```


انتگرال نمره هاپلوتایپی (iHS)

$$\text{uniHS}(s) = \ln \left(\frac{i\text{HH}^A(s)}{i\text{HH}^D(s)} \right)$$

$$i\text{HS}(s) = \frac{\text{uniHS}(s) - \text{mean}(\text{uniHS}|p_s)}{\text{sd}(\text{uniHS}|p_s)}$$

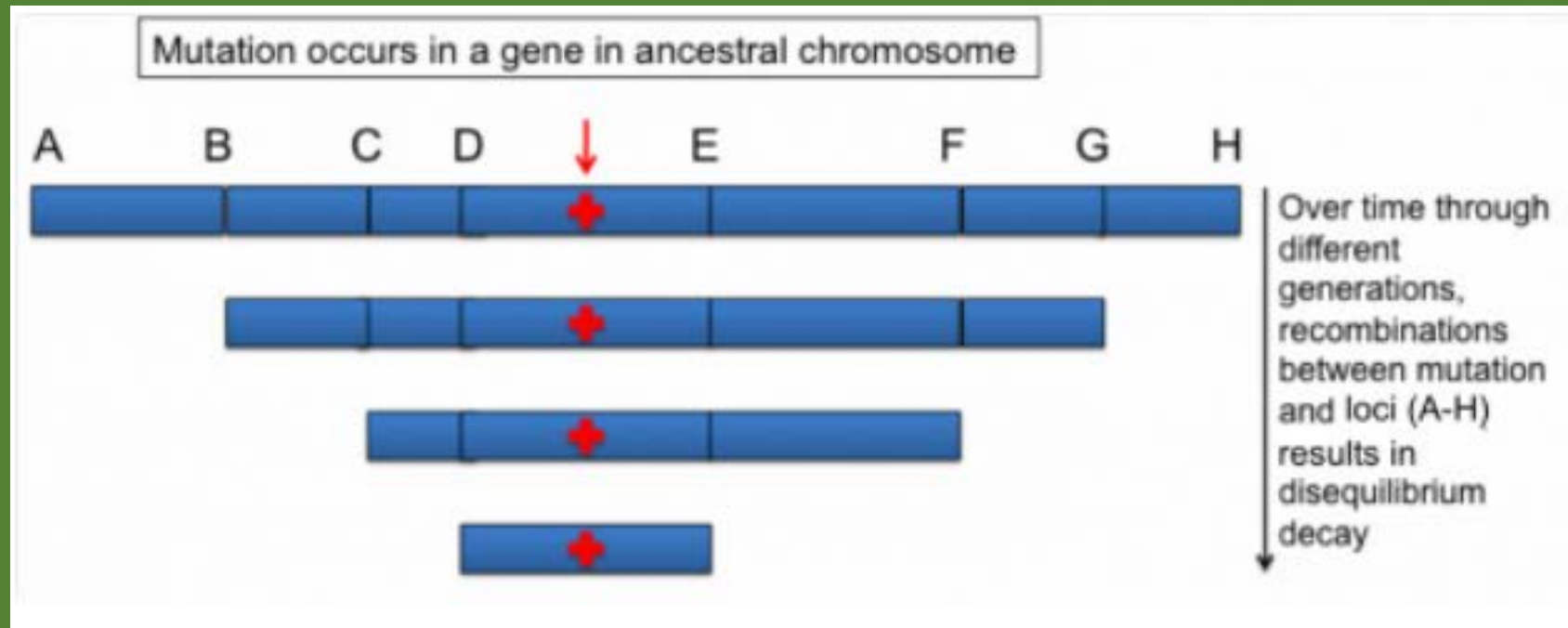
$i\text{HS} < -2 \Rightarrow$ Selection on derived allele

$i\text{HS} < 2$ and $> -2 \Rightarrow$ No selection

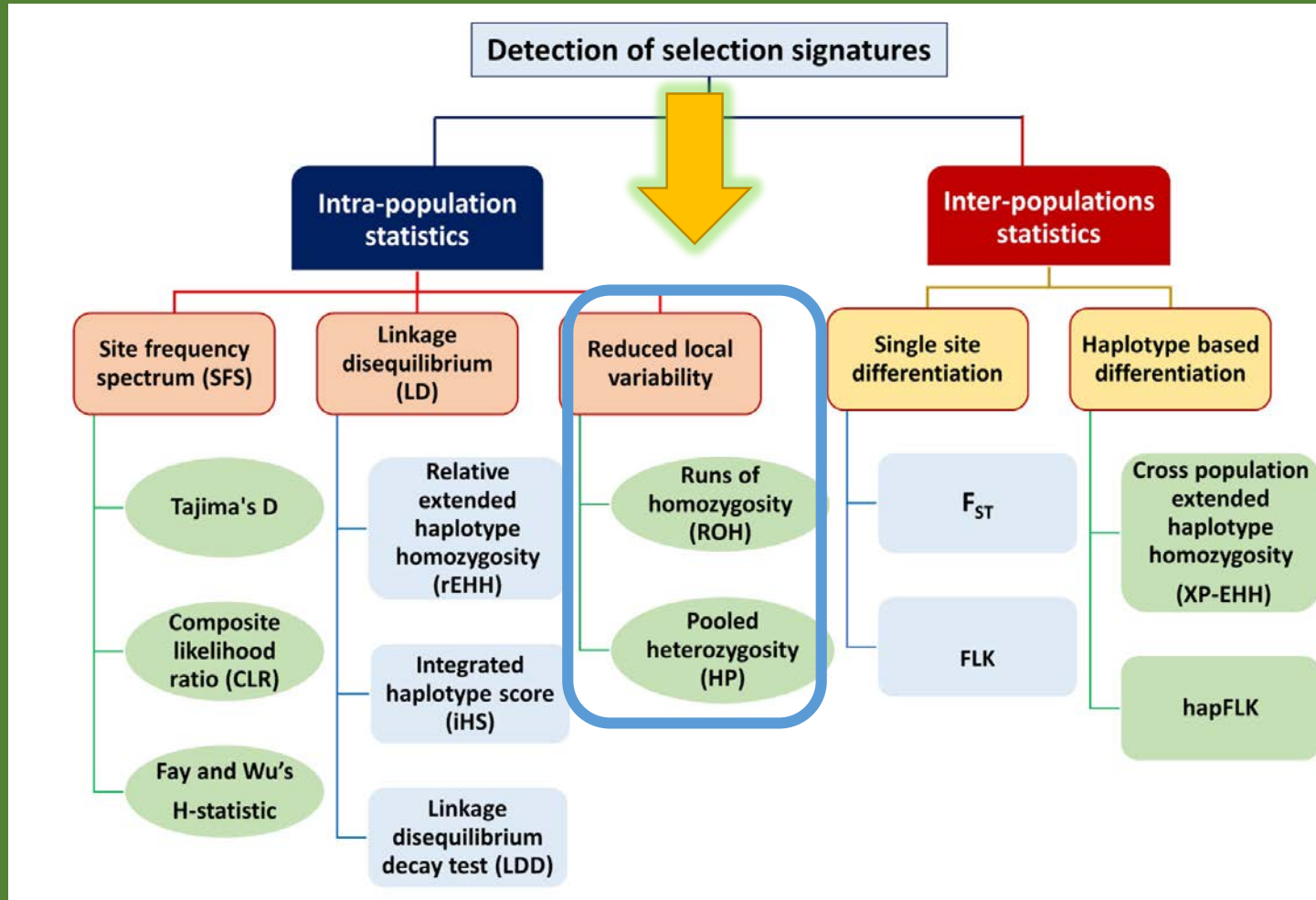
$i\text{HS} > 2 \Rightarrow$ Selection on ancestral allele

کاهش در عدم تعادل پیوستگی (LDD)

- برای این روش از آماره r^2 استفاده می شود

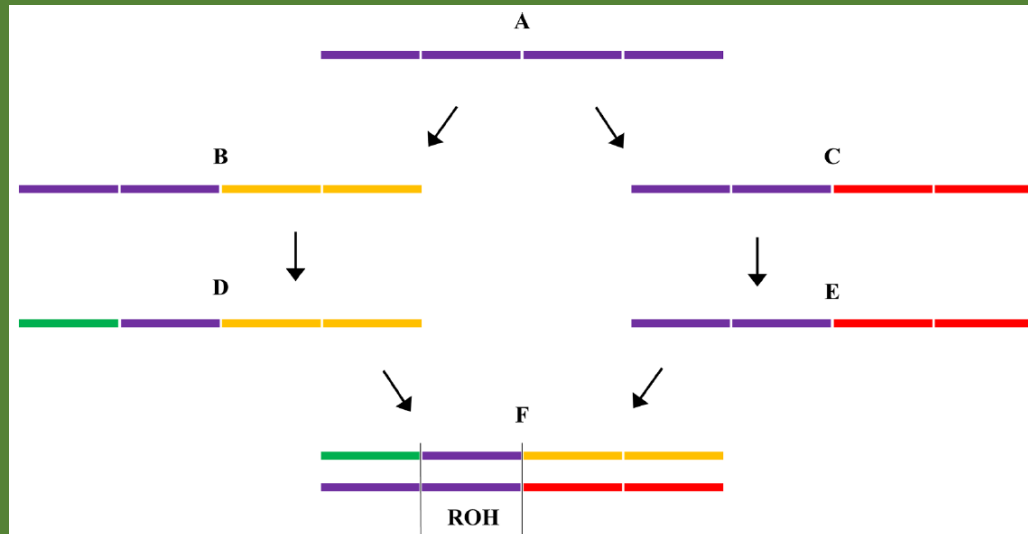


تنوع محلی کاهش یافته



ROH

- ROH به عنوان قطعات هموزیگوت پیوسته با طول زیاد در ژنوم بوده، که فرض می‌شوند از یک جد مشترک به ارث رسیده اند (McQuillan et al., 2008)
- افزایش طول ROH برابر با افزایش احتمال وجود نشانه انتخاب



□ کاربردهای این آماره:

✓ محاسبه همخونی ژنومیک

✓ مطالعه ساختار جمعیت

✓ بررسی تاریخچه جمعیت

هتروزیگوسیتی ادغام شده (H_p)

- براساس شمارش آلی محاسبه می شود (Rubin et al., 2010).

□ مقادیر H_p به فرم استاندارد تبدیل شده تا هم سیگنال ها به راحتی مشخص شده و هم با مطالعات قبلی قابل مقایسه باشد

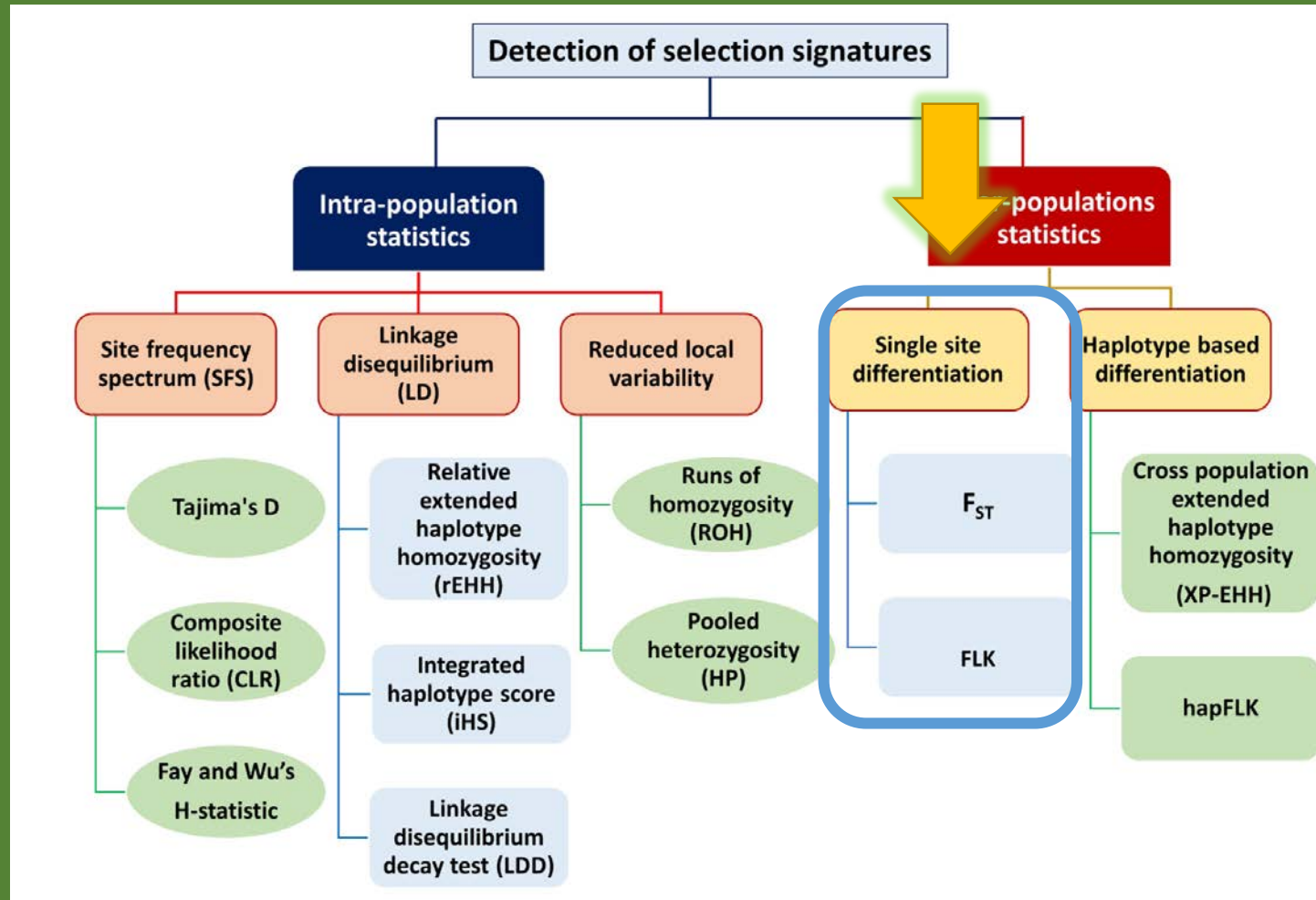
تعداد خوانش آل با بیشترین فراوانی

پنجره با 1 جایگاه

$$H_p = \frac{2 \sum_{i=1}^l n_i \sum_{i=1}^l (N_i - n_i)}{\left(\sum_{i=1}^l n_i + \sum_{i=1}^l (N_i - n_i) \right)^2}$$

تعداد کل خوانش در جایگاه i

تمایز جمعیت به صورت تک جایگاهی



شاخص تثبیت (F_{ST})

- آماره F_{ST} بر پایه تفاوت های مشاهده شده در فراوانی آللها در بین جمعیت ها محاسبه می شود.

	Genotype		
	<i>AA</i>	<i>Aa</i>	<i>aa</i>
Population 1	125	250	125
Population 2	50	30	20
Population 3	100	500	400

شاخص تثبیت (F_{ST})

- آماره F_{ST} بر پایه تفاوت های مشاهده شده در فراوانی آللها در 1 ن جمعیت ها محاسبه می شود.

	Genotype		
	<i>AA</i>	<i>Aa</i>	<i>aa</i>
Population 1	125	250	125
Population 2	50	30	20
Population 3	100	500	400

$$p_1 \text{ (frequency of allele } A \text{ in Pop. 1)} = (2 \cdot 125 + 250) / 1,000 = 0.50$$

$$q_1 = 0.5$$

$$p_2 = (2 \cdot 50 + 30) / 200 = 0.65$$

$$q_2 = 0.35$$

$$p_3 = (2 \cdot 100 + 500) / 2,000 = 0.35$$

$$q_3 = 0.65$$

$$\bar{p} \text{ (frequency of allele } A \text{ over all populations)}$$

$$(2 \cdot 125 + 250 + 2 \cdot 50 + 30 + 2 \cdot 100 + 500) / (1,000 + 200 + 2,000) = 1,330 / 3,200 = 0.4156$$

or

$$(0.5 \cdot 1,000 + 0.65 \cdot 200 + 0.35 \cdot 2,000) / 3,200$$

$$\bar{q} \text{ (frequency of allele } a \text{ over all populations)}$$

$$(2 \cdot 125 + 250 + 2 \cdot 20 + 30 + 2 \cdot 400 + 500) / (1,000 + 200 + 2,000) = 1,870 / 3,200 = 0.5844$$

شاخص تثبیت (F_{ST})

- آماره F_{ST} بر پایه تفاوت های مشاهده شده در فراوانی آللها در **1** جمعیت ها محاسبه می شود.

	Genotype		
	<i>AA</i>	<i>Aa</i>	<i>aa</i>
Population 1	125	250	125
Population 2	50	30	20
Population 3	100	500	400

$$p_1 \text{ (frequency of allele } A \text{ in Pop. 1)} = (2*125 + 250)/1,000 = 0.50$$

$$q_1 = 0.5$$

$$p_2 = (2*50 + 30)/200 = 0.65$$

$$q_2 = 0.35$$

$$p_3 = (2*100 + 500)/2,000 = 0.35$$

$$q_3 = 0.65$$

$$\bar{p} \text{ (frequency of allele } A \text{ over all populations)}$$

$$(2*125 + 250 + 2*50 + 30 + 2*100 + 500)/(1,000 + 200 + 2,000) = 1,330 / 3,200 = 0.4156$$

or

$$(0.5*1,000 + 0.65*200 + 0.35*2,000) / 3,200$$

$$\bar{q} \text{ (frequency of allele } a \text{ over all populations)}$$

$$(2*125 + 250 + 2*20 + 30 + 2*400 + 500) / (1,000 + 200 + 2,000) = 1,870 / 3,200 = 0.5844$$

2

Observed heterozygosity (here we are counting heterozygous **genotypes**):

$$H_{obs1} = 250/500 = 0.5$$

$$H_{obs2} = 30/100 = 0.3$$

$$H_{obs3} = 500/1000 = 0.5$$

Expected heterozygosity ($2pq$):

$$H_{exp1} = 2*0.5*0.5 = 0.5 \quad [\text{Observed} = \text{Expected}]$$

$$H_{exp2} = 2*0.65*0.35 = 0.46 \quad [\text{Observed} < \text{Expected}]$$

$$H_{exp3} = 2*0.35*0.65 = 0.46 \quad [\text{Observed} > \text{Expected}]$$

شاخص تثبیت (F_{ST})

- آماره F_{ST} بر پایه تفاوت های مشاهده شده در فراوانی آللها در 1 جمعیت ها محاسبه می شود.

	Genotype		
	<i>AA</i>	<i>Aa</i>	<i>aa</i>
Population 1	125	250	125
Population 2	50	30	20
Population 3	100	500	400

$$p_1 \text{ (frequency of allele } A \text{ in Pop. 1)} = (2 \cdot 125 + 250) / 1,000 = 0.50$$

$$q_1 = 0.5$$

$$p_2 = (2 \cdot 50 + 30) / 200 = 0.65$$

$$q_2 = 0.35$$

$$p_3 = (2 \cdot 100 + 500) / 2,000 = 0.35$$

$$q_3 = 0.65$$

$$p\text{-bar (frequency of allele } A \text{ over all populations)}$$

$$(2 \cdot 125 + 250 + 2 \cdot 50 + 30 + 2 \cdot 100 + 500) / (1,000 + 200 + 2,000) = 1,330 / 3,200 = 0.4156$$

or

$$(0.5 \cdot 1,000 + 0.65 \cdot 200 + 0.35 \cdot 2,000) / 3,200$$

$$q\text{-bar (frequency of allele } a \text{ over all populations)}$$

$$(2 \cdot 125 + 250 + 2 \cdot 20 + 30 + 2 \cdot 400 + 500) / (1,000 + 200 + 2,000) = 1,870 / 3,200 = 0.5844$$

3

2

Observed heterozygosity (here we are counting heterozygous **genotypes**):

$$H_{\text{obs1}} = 250 / 500 = 0.5$$

$$H_{\text{obs2}} = 30 / 100 = 0.3$$

$$H_{\text{obs3}} = 500 / 1000 = 0.5$$

Expected heterozygosity ($2pq$):

$$H_{\text{exp1}} = 2 \cdot 0.5 \cdot 0.5 = 0.5 \quad [\text{Observed} = \text{Expected}]$$

$$H_{\text{exp2}} = 2 \cdot 0.65 \cdot 0.35 = 0.46 \quad [\text{Observed} < \text{Expected}]$$

$$H_{\text{exp3}} = 2 \cdot 0.35 \cdot 0.65 = 0.46 \quad [\text{Observed} > \text{Expected}]$$

Heterozygosity indices (over individuals, subpopulations and total population)

H_S based on **expected** heterozygosities in populations

$$[= (H_{\text{exp1}} \cdot N_1 + H_{\text{exp2}} \cdot N_2 + H_{\text{exp3}} \cdot N_3) / N_{\text{TOTAL}}] = (0.5 \cdot 500 + 0.46 \cdot 100 + 0.46 \cdot 1000) / 1,600 = 0.4691$$

H_T based on **expected** heterozygosities overall:

$$[= 2 \cdot p\text{-bar} \cdot q\text{-bar}] = 2 \cdot 0.4156 \cdot 0.5844 = 0.4858$$

شاخص تثبیت (F_{ST})

- آماره F_{ST} بر پایه تفاوت های مشاهده شده در فراوانی آللها در بین جمعیت ها محاسبه می شود.

4

$$F_{ST} [= (H_T - H_S)/H_T] = (0.4858 - 0.4691) / 0.4858 = 0.0344$$

FLK (Bonhomme et al., 2010)

$$T_{\text{LK}} = \frac{n-1}{\bar{F}_{\text{ST}}} F_{\text{ST}}$$

$$F_{\text{ST}} = \frac{s_p^2}{\bar{p}(1-\bar{p})} = \frac{((1/(n-1)) \sum_{i=1}^n (p_i - \bar{p})^2)}{\bar{p}(1-\bar{p})}$$

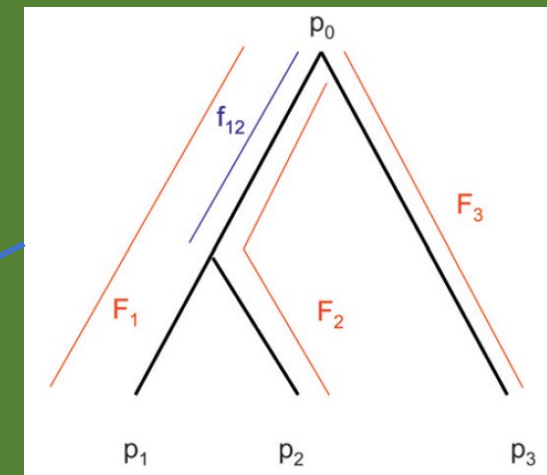
$$T_{F\text{-LK}} = (\mathbf{p} - \hat{p}_0 \mathbf{1}_n)' \mathbb{V}(\mathbf{p})^{-1} (\mathbf{p} - \hat{p}_0 \mathbf{1}_n) = \frac{Q}{\hat{p}_0(1-\hat{p}_0)}$$

$\mathbf{p} = (p_1, \dots, p_j, \dots, p_n)$ be the n-vector of allelic frequencies of an allele in the n populations

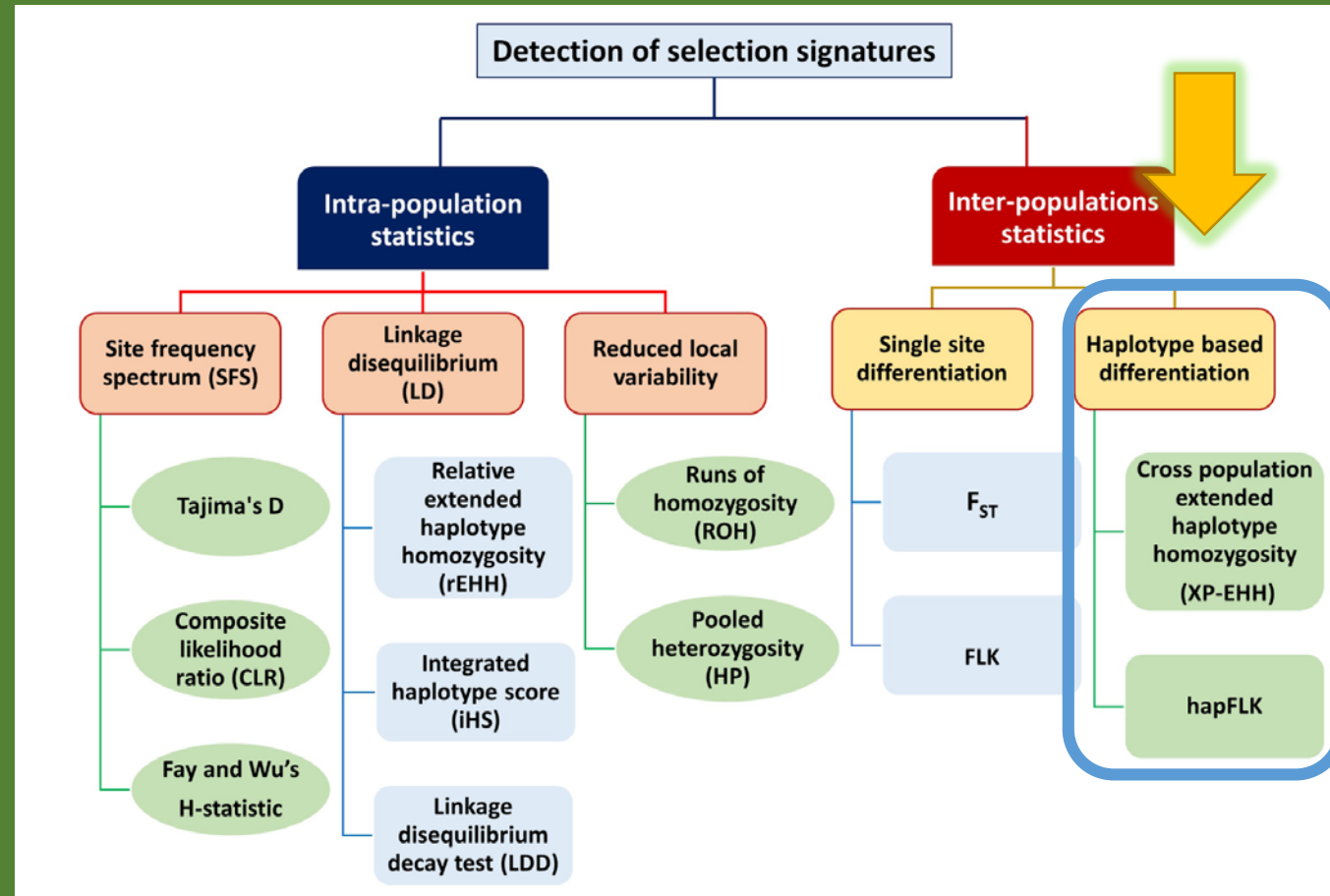
$$\hat{p}_0 = \frac{\mathbf{1}_n' \mathcal{F}^{-1} \mathbf{p}}{\mathbf{1}_n' \mathcal{F}^{-1} \mathbf{1}_n}$$

founder allele frequency

The kinship matrix from neighbor-joining (NJ) tree



روش های مبتنی بر هاپلوتایپ



XP-EHH (Sabeti et al., 2007) and Rsb (Tang et al. 2007)

	rs1	rs2	rs3	rs4	rs5	rs6	rs7	rs8	rs9	rs10	rs11
HG1_1	0	0	1	1	0	0	0	0	0	1	1
HG1_2	0	0	0	0	1	0	1	0	1	0	0
HG2_1	0	1	0	1	0	0	1	0	0	0	1
HG2_2	0	0	1	0	0	0	0	1	1	0	0
HG3_1	0	0	0	0	0	1	0	0	0	0	1
HG3_2	0	0	0	0	0	1	0	0	0	0	1
HG4_1	1	0	0	0	0	1	0	0	0	0	0
HG4_2	0	0	0	0	0	1	0	0	0	0	0



$$EHHS_{rs6,rs6} = \frac{1}{n_s(n_s - 1)} \left(\sum_{k=1}^{K_{rs6,rs6}} n_k(n_k - 1) \right) = \frac{1}{8 \cdot 7} (4 \cdot 3 + 4 \cdot 3) = \frac{3}{7}$$

XP-EHH (Sabeti et al., 2007) and Rsb (Tang et al. 2007)

	rs1	rs2	rs3	rs4	rs5	rs6	rs7	rs8	rs9	rs10	rs11
HG1_1	0	0	1	1	0	0	0	0	0	1	1
HG1_2	0	0	0	0	1	0	1	0	1	0	0
HG2_1	0	1	0	1	0	0	1	0	0	0	1
HG2_2	0	0	1	0	0	0	0	1	1	0	0
HG3_1	0	0	0	0	0	1	0	0	0	0	1
HG3_2	0	0	0	0	0	1	0	0	0	0	1
HG4_1	1	0	0	0	0	1	0	0	0	0	0
HG4_2	0	0	0	0	0	1	0	0	0	0	0



$$\text{EHHS}_{rs6,rs7} = \frac{1}{n_s(n_s - 1)} \left(\sum_{k=1}^{K_{rs6,rs7}} n_k(n_k - 1) \right) = \frac{1}{8 \cdot 7} (2 \cdot 1 + 2 \cdot 1 + 4 \cdot 3) = \frac{2}{7}$$

hapFLK (Fariello et al. 2013)

- مشابه روش FLK بوده ولی از اطلاعات هاپلوتایپی استفاده می شود
- در این روش هم برای اندازه جمعیت موثر تصحیح صورت میگیرد
- برآورد ماتریکس کینشپ (F) مشابه FLK می باشد ولی بجای محاسبه از طریق فراوانی آلی از فراوانی هاپلوتایپی استفاده می کند

نرم افزارهای مورد استفاده برای شناسایی نشانه های انتخاب (Saravanan et al., 2020)

Program/tool	Latest version ¹	Methods applied	Platform ²	GUI/CUI ³	License ⁴	Source link ¹
Arlequin	3.5.2.2 (2015)	Tajima's D	W	GUI	F	http://cmpg.unibe.ch/software/arlequin35/
BayeScan	2.1 (2012)	F _{ST} (Bayesian approach)	L/M	CUI/GUI	F	http://cmpg.unibe.ch/software/BayeScan/index.html
cgaTOH	(2013)	ROH	W/L/M	CUI	F	http://www.cs.kent.edu/~zhao/TOH/
DetectRUNS R-package	0.9.6 (2019)	ROH	W/L/M	CUI	F	https://CRAN.R-project.org/package=detectRUNS
DnaSP	6.12.03 (2019)	Tajima's D, Fay and Wu's H, Fu and Li's test, MKT	W/L	CUI/GUI	F	http://www.ub.edu/dnasp
Hapbin	(2015)	EHH, iHS, XP-EHH	W/L/M	CUI	F	https://github.com/evotools/hapbin
HapFLK	1.4 (2017)	HapFLK, FLK	W/L/M	CUI	F	https://forge-dga.jouy.inra.fr/projects/hapflk/files
HGDP Selection Browser	(2009)	iHS, XP-EHH, F _{ST} , CLR	W/L/M	CUI	F	http://hgdp.uchicago.edu/Software/
HierFstat R-package	0.04-22 (2015)	F _{ST}	W/L/M	CUI	F	https://CRAN.R-project.org/package=hierfstat
OmegaPlus	2.2.2 (2013)	LD based "ω"	L	CUI	F	http://www.exelixis-lab.org/software.html
Pcadapt R-package	4.3.1 (2020)	PCAdapt	W/L/M	CUI	F	https://CRAN.R-project.org/package=pcadapt
PLINK	1.90 (2015)	ROH, F _{ST}	W/L/M	CUI	F	http://zzz.bwh.harvard.edu/plink/
PoPoolation	1.2.2 (2011)	Tajima's Pi, Watterson's Theta and Tajima's D	W/L/M	CUI	F	https://sourceforge.net/projects/popoolation/
PopGenome	2.7.5 (2020)	Tajima's D, F _{ST} (Bayesian approach)	W/L/M	CUI	F	https://CRAN.R-project.org/package=PopGenome
Rehh R-package	3.1.0 (2020)	EHH, iHS, Rsb, XP-EHH	W/L/M	CUI	F	https://CRAN.R-project.org/package=rehh
Selscan	1.2.0 (2017)	EHH, iHS, XP-EHH, nSL	W/L/M	CUI	F	https://github.com/szpiech/selscan
SNP Variation Suite (SVS)	8.4.0 (2015)	F _{ST}	W/L/M	CUI/GUI	L	https://www.goldenhelix.com/forms/SNP_Variation
SweeD	3.2.1 (2013)	CLR	L	CUI	OS	https://cme.h-its.org/exelixis/web/software/sweeD/index.html
Sweep	1.1 (2010)	LRH, EHH	W/L	GUI	F	https://software.broadinstitute.org/mpg/sweep/
Sweepfinder2	1.0 (2016)	CLR	L	CUI	F	http://degiorgiogrupo.fau.edu/sf2.html
TRES	1.0 (2015)	Wright's F _{ST}	W/L/M	CUI/GUI	F	http://mlkd.csd.auth.gr/bio/tres
VariScan	2.0.3 (2012)	Tajima's D, Fu and Li's tests	W/L/M	CUI/GUI	F	http://www.ub.edu/softevol/variscan
VCFtools	0.1.13 (2015)	Tajima's D, F _{ST}	L/M	CUI	F	https://github.com/vcftools/vcftools
XP-CLR	1.1.2 (2019)	XP-CLR	L	CUI	F	https://reich.hms.harvard.edu/software
ZANARDI	(2015)	ROH	L/M	CUI	OS	https://github.com/bioinformatics-ptp/Zanardi

روش های ترکیبی

ترکیب نمرات آزمون برای چندین سایت
در یک منطقه مجاور

**Composite likelihood
ratio (CLR)**

**Cross-population composite
likelihood ratio (XP-CLR)**

ترکیب چندین تست مستقل در یک
سایت

DH test

**Composite of
multiple signals
(CMS)**

**de-correlated composite of
multiple signals (DCMS)**

نشانه های انتخاب در گاو

□ تعداد محدودی مارکر، کروموزوم خاص و STR

- ناحیه BTA6 جزو نواحی ژنومی شناسایی شده به عنوان کاندید انتخاب بوده که در گاو قرمز نروژی مطالعه گردید (Hayes et al. 2008).
- EHH ← آنگوس و هلشتاین ← BTA14 (Marques et al. 2008) ۳ ناحیه در BTA19 و ۳ ناحیه در BTA29 (Prasad et al. 2008)

□ آرایه SNP، و توالی یابی نسل بعد

- از سال ۲۰۰۹ تا ۲۰۱۵

نشانه های انتخاب در گوسفند و بز

- اولین دام های اهلی شده حدود ۱۰۵۰۰ سال پیش (Alberto et al., 2018)
- اهلی سازی و انتخاب های با واسطه انسان در نقاط مختلف جهان همراه با سازگاری با محیط های متفاوت

hapFLK و Fst، ROH ✓

نشانه های انتخاب در گونه های دیگر

□ شناسایی نشانه های انتخاب در گونه های اسب، گاومیش، شتر، جوجه، خوک و یاک

✓ Fst، ROH و XP-EHH ← اسب خونگرم سوئدی ← نواحی مرتبط با عملکرد ورزشی
(Ablondi et al. 2019)

✓ Fst ← گاومیش خوزستانی و آذری ← تولید شیر، رشد، عملکرد متابولیکی و تغییرات مرتبط با اهلی شدن مانند نمو سیستم عصبی و ایمنی (Mokhber et al., 2018)

✓ GBS ← Hp و Fst

کاربردهای مطالعات نشانه های انتخاب

- فهم بهتر از فرآیند اهلی شدن، نمو نژاد، ساختار جمعیتی، تکامل سازگاری، شناسایی ریشه اصلی تنوع فنوتیپی و نتایج فرآیندهای انتخاب
- همچنین این روش ها می توانند مکملی برای مطالعات پویش کلی ژنومی (GWAS) و انتخاب ژنومی بوده، و برای پرورش و برنامه های حفاظتی مورد استفاده قرار گیرند.
- بررسی تنوع ژنتیکی تنها میزان تفاوت در بین نژادها را مطالعه کرده ولی مطالعات نشانه های انتخاب پا را فراتر گذاشته و دلیل اصلی برای این تمایزات را نشان می دهند

کاربردهای مطالعات نشانه های انتخاب

• در مباحث علوم دامی، یافتن نواحی تحت انتخاب از منظر شناسایی ساختار ژنتیکی دام و تعیین مسیر ژنی برای تنوع صفات مهم اقتصادی بسیار مهم می باشند

□ برای مثال:

✓ شناسایی نواحی از ژنوم که موجب درک بهتری از تنوع صفات مهم اقتصادی و سازگاری با اقلیم سرد در گاو شده (Ghoreishifar et al. 2020)

✓ شناسایی نواحی کاندید برای انتخاب در گوسفندان موجب یافتن ژن ها مرتبط با سازگاری محلی و یا تنوع نژادها شده (Eydivandi et al. 2020; Eydivandi et al. 2021)

کلام آخر

- شناسایی ژن های منجر به همبستگی های ژنتیکی
- ژن های شناسایی شده در بررسی نشانه های انتخاب می توانند موجب پیدا کردن و اثبات جهش های علی در نواحی شناسایی شده در مطالعات QTL شده و ژن های مرتبط با صفات اقتصادی مشخص شوند
- همچنین این نتایج می توانند ژن ها و مسیرهای ژنی مرتبط با صفات را شناسایی کنند، ولی این ژن ها در نژادهای مختلف به صورت متفاوتی تحت انتخاب بوده باشد

با تشکر از توجه شما!!!!



معاونت علمی و فناوری
شبکه دانش کشاورزی
سلسله برنامه‌های ویدیو کنفرانس انتقال دانش به‌روز در گستره ملی بخش کشاورزی

عنوان:

روش‌های شناسایی نشانه‌های انتخاب و کاربرد آن در علوم دامی

سخنران:

محمود امیری رودبار

عضو هیأت علمی (استادیار) مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی صفی آباد دزفول

۶ شهریور ۱۴۰۱ - ساعت: ۸:۳۰-۹:۱۵