

Site Frequency Spectrum (SFS)

André Minderico
64517

O que é SFS?

Sequências de DNA Hipotéticas

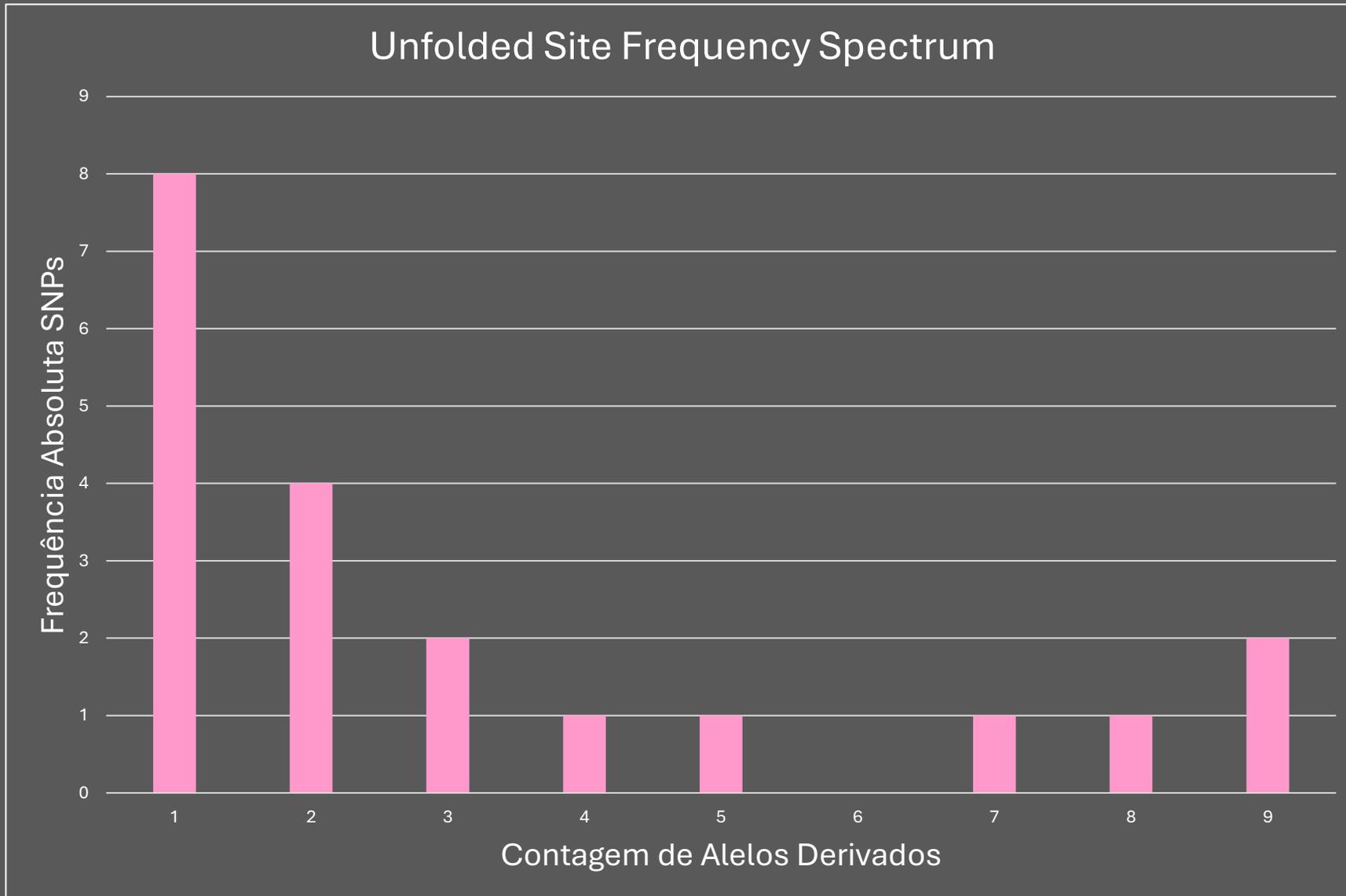
	SNP 1	SNP 2	SNP 3	SNP 4	SNP 5	SNP 6	SNP 7	SNP 8	SNP 9	SNP 10	SNP 11	SNP 12	SNP 13	SNP 14	SNP 15	SNP 16	SNP 17	SNP 18	SNP 19	SNP 20
Outgroup	C	C	A	A	G	A	T	C	T	C	C	T	T	G	A	G	T	G	G	G
Sample 1	C	C	A	A	G	C	T	C	C	T	C	T	T	G	A	G	G	G	A	G
Sample 2	C	T	A	A	G	C	A	C	C	C	C	T	G	G	A	G	G	G	A	C
Sample 3	G	T	A	T	A	C	T	C	C	C	T	C	T	G	A	G	T	T	G	G
Sample 4	C	T	A	A	G	C	T	C	T	C	C	T	T	G	A	G	G	G	A	G
Sample 5	G	T	A	A	G	A	T	C	C	T	C	T	G	G	A	G	G	G	A	G
Sample 6	C	T	A	A	G	A	T	C	C	C	C	T	T	C	T	G	T	G	A	G
Sample 7	C	T	G	A	G	C	T	G	C	C	T	T	T	C	A	G	T	G	A	G
Sample 8	C	T	A	A	A	C	T	C	C	C	C	T	T	G	A	G	T	T	A	G
Sample 9	C	T	A	A	A	C	T	C	C	C	T	T	T	G	A	T	T	G	G	G
Sample 10	G	T	A	A	G	C	T	C	C	C	T	T	T	G	A	G	G	G	G	G

Derived	3	9	1	1	3	8	1	1	9	2	4	1	2	2	1	1	5	2	7	1
Minor	3	1	1	1	3	2	1	1	1	2	4	1	2	2	1	1	5	2	3	1

Booker, T. R., Jackson, B. C., & Keightley, P. D. (2017). Detecting positive selection in the genome. *BMC biology*, 15(1), 98.

<https://doi.org/10.1186/s12915-017-0434-y>

Distribuição das Frequências Alélicas



Sequências de DNA Hipotéticas

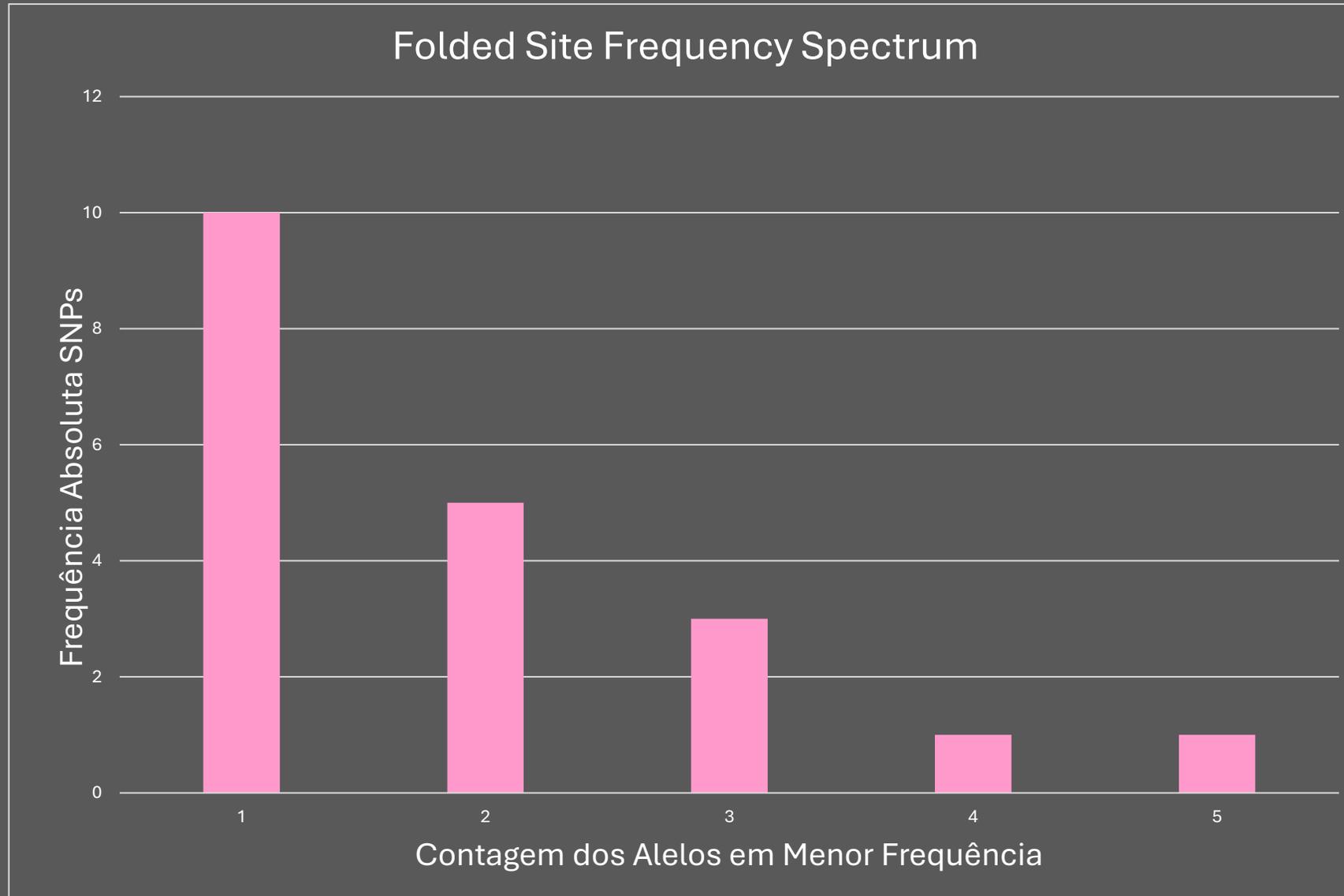
	SNP 1	SNP 2	SNP 3	SNP 4	SNP 5	SNP 6	SNP 7	SNP 8	SNP 9	SNP 10	SNP 11	SNP 12	SNP 13	SNP 14	SNP 15	SNP 16	SNP 17	SNP 18	SNP 19	SNP 20
Outgroup	C	C	A	A	G	A	T	C	T	C	C	T	T	G	A	G	T	G	G	G
Sample 1	C	C	A	A	G	C	T	C	C	T	C	T	T	G	A	G	G	G	A	G
Sample 2	C	T	A	A	G	C	A	C	C	C	C	T	G	G	A	G	G	G	A	C
Sample 3	G	T	A	T	A	C	T	C	C	C	T	C	T	G	A	G	T	T	G	G
Sample 4	C	T	A	A	G	C	T	C	T	C	C	T	T	G	A	G	G	G	A	G
Sample 5	G	T	A	A	G	A	T	C	C	T	C	T	G	G	A	G	G	G	A	G
Sample 6	C	T	A	A	G	A	T	C	C	C	C	T	T	C	T	G	T	G	A	G
Sample 7	C	T	G	A	G	C	T	G	C	C	T	T	T	C	A	G	T	G	A	G
Sample 8	C	T	A	A	A	C	T	C	C	C	C	T	T	G	A	G	T	T	A	G
Sample 9	C	T	A	A	A	C	T	C	C	C	T	T	T	G	A	T	T	G	G	G
Sample 10	G	T	A	A	G	C	T	C	C	C	T	T	T	G	A	G	G	G	G	G

Derived	3	9	1	1	3	8	1	1	9	2	4	1	2	2	1	1	5	2	7	1
Minor	3	1	1	1	3	2	1	1	1	2	4	1	2	2	1	1	5	2	3	1

Booker, T. R., Jackson, B. C., & Keightley, P. D. (2017). Detecting positive selection in the genome. *BMC biology*, 15(1), 98.

<https://doi.org/10.1186/s12915-017-0434-y>

Distribuição das Frequências Alélicas



Site Frequency Spectrum

Também chamado **Allele Frequency Spectrum**, é a distribuição das frequências alélicas num determinado número de posições no genoma, onde houver polimorfismos, de uma população.

Watterson Estimator

- Método para descrever a **diversidade genética** numa população
- É, também, uma medida da taxa de mutações neutras numa população

$$\theta = 4N_e\mu$$

Pressupostos:

- n indivíduos haploides
- Nucleótidos infinitos
- $n \ll N_e$

Watterson Estimator

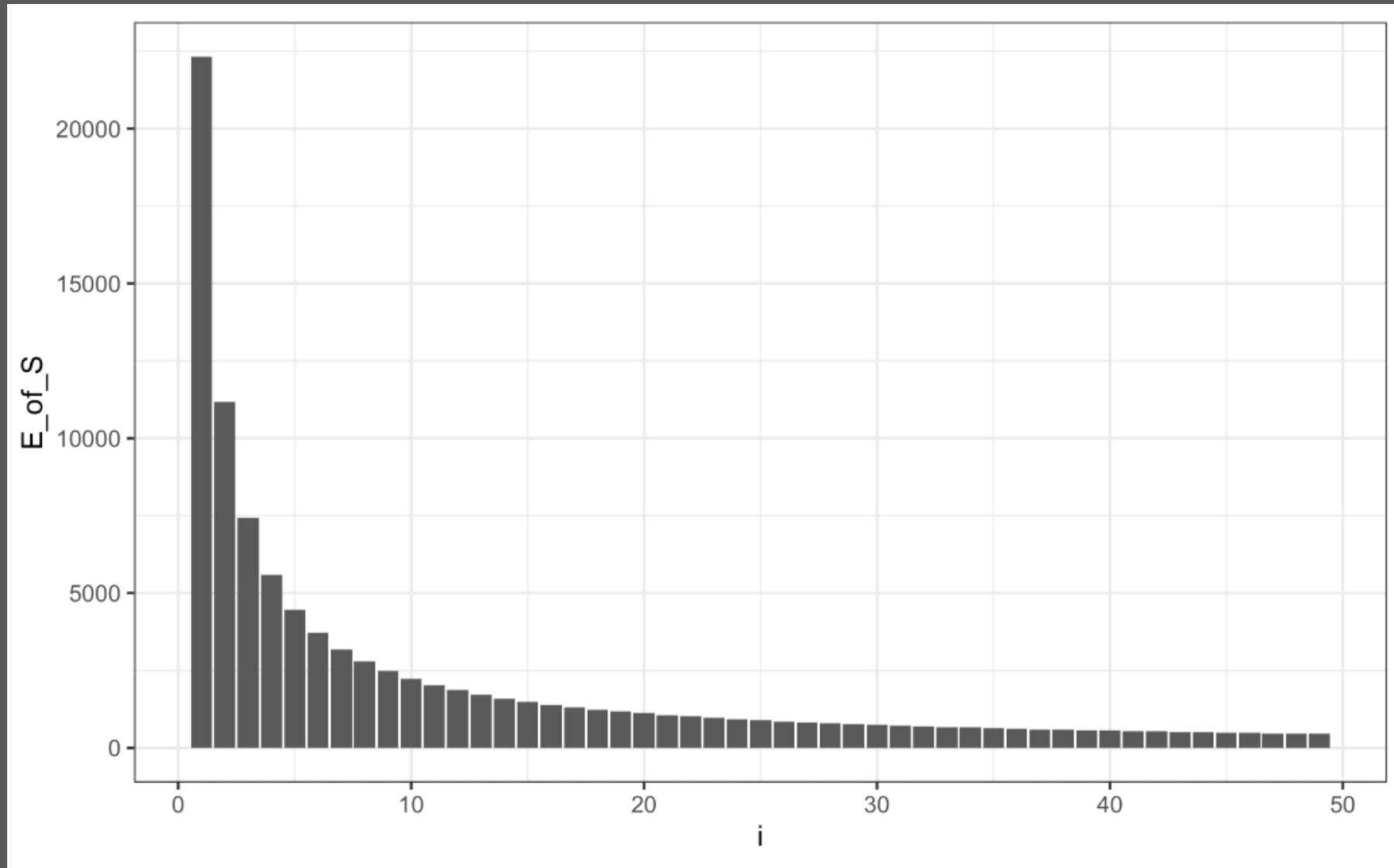
segregation sites

$$\hat{\theta}_w = \frac{K}{a_n}$$

$$a_n = \sum_{i=1}^{n-1} \frac{1}{i}$$

Série harmónica para n-1 individuos

Distribuição Esperada para θ_w



<https://eriquande.github.io/coallescent-hands-on/004-one-dimensional-SFS.nb.html>

Como medir?

Watterson Estimator

Tajima's D

Fu and Li's test

Fay and Wu's H

Zeng, Fu, Shi and Wu's test

Tajima's D

Pairwise Difference

$\binom{n}{2}$

Watterson's θ_w

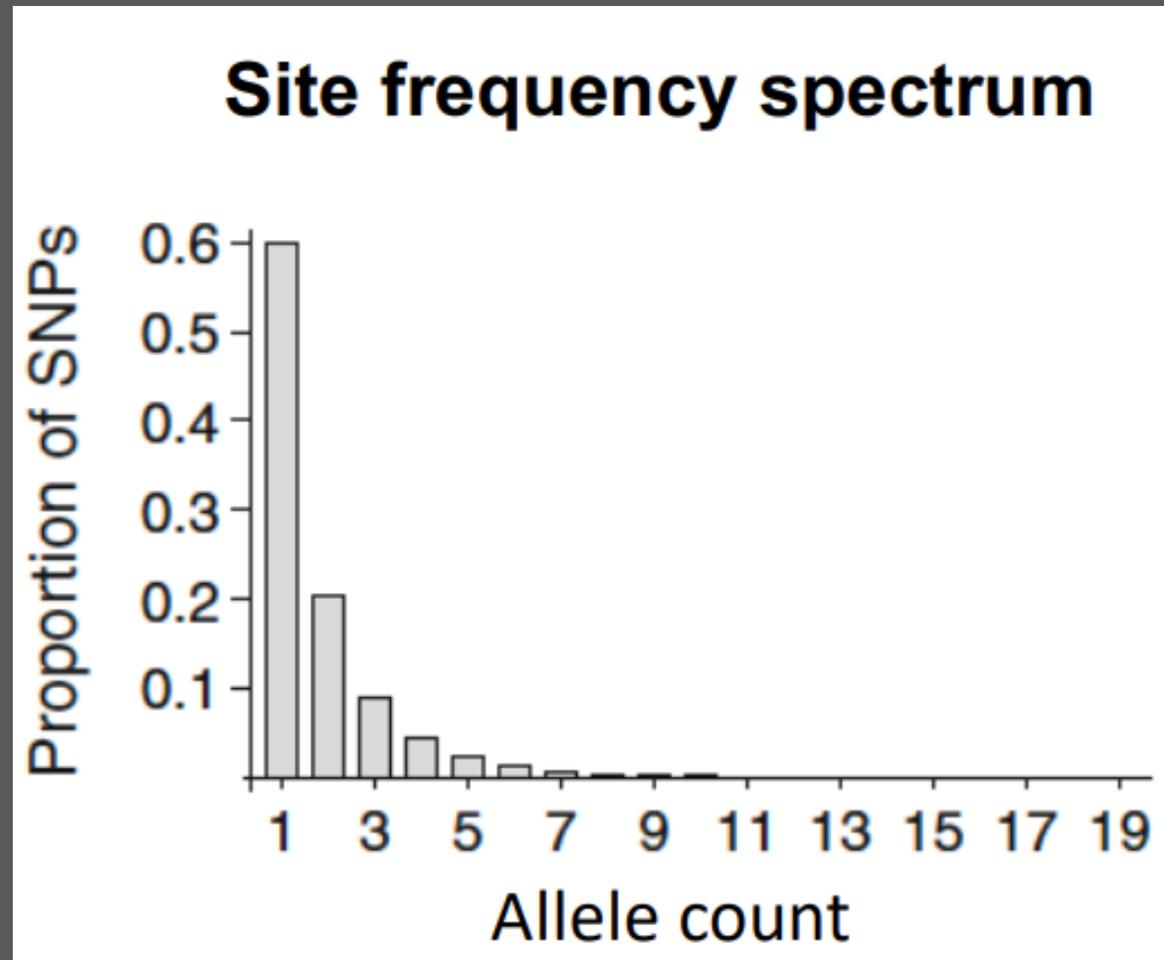
$$D = \frac{d}{\sqrt{\hat{V}(d)}} = \frac{\hat{k} - \frac{S}{a_1}}{\sqrt{[e_1 S + e_2 S(S - 1)]}}$$

Desvio Padrão da diferença
(no numerador)

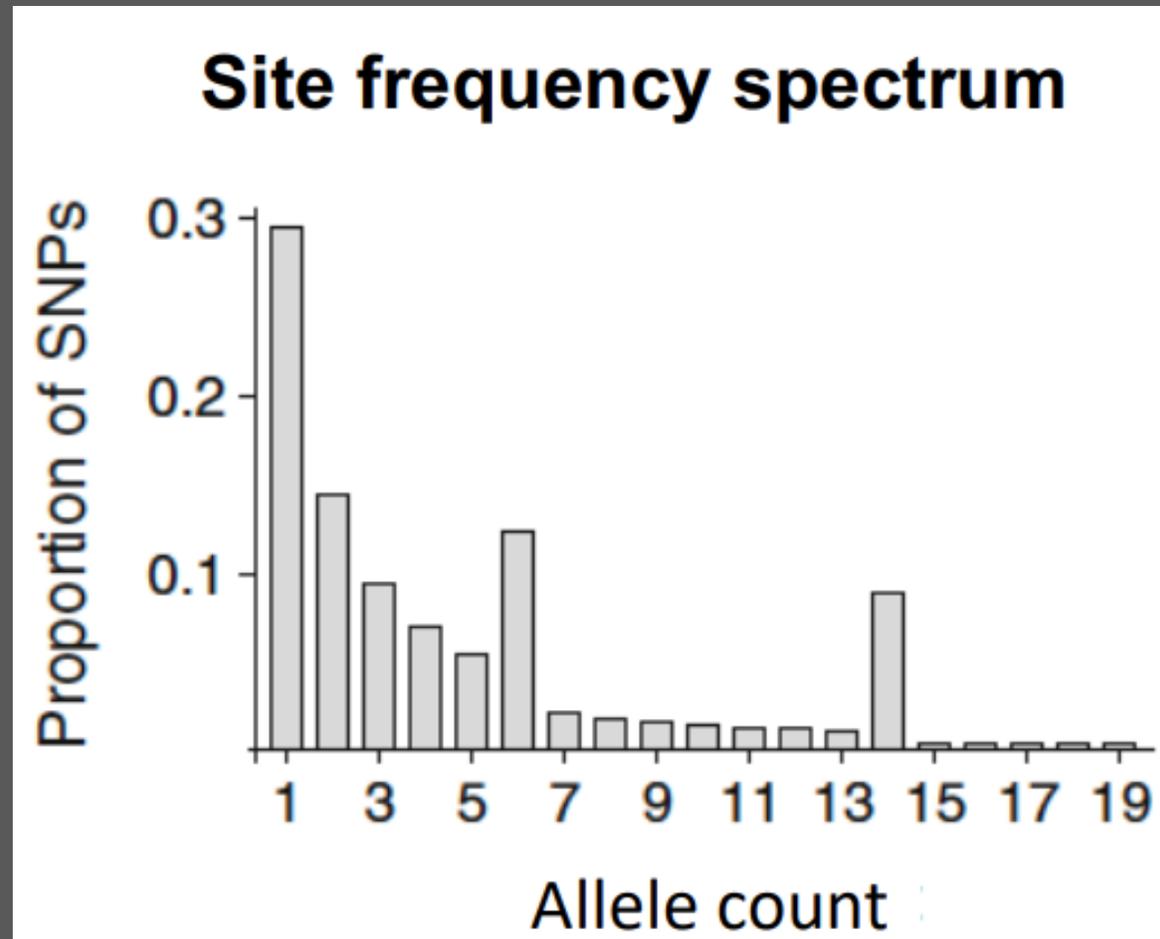
Variação da SFS à “Norma”

Todos os processos que afetem a frequência alélica (seletivos, migratórios, deriva genética, estrutura populacional, etc.), inevitavelmente alteram a distribuição esperada da Site Frequency Spectrum.

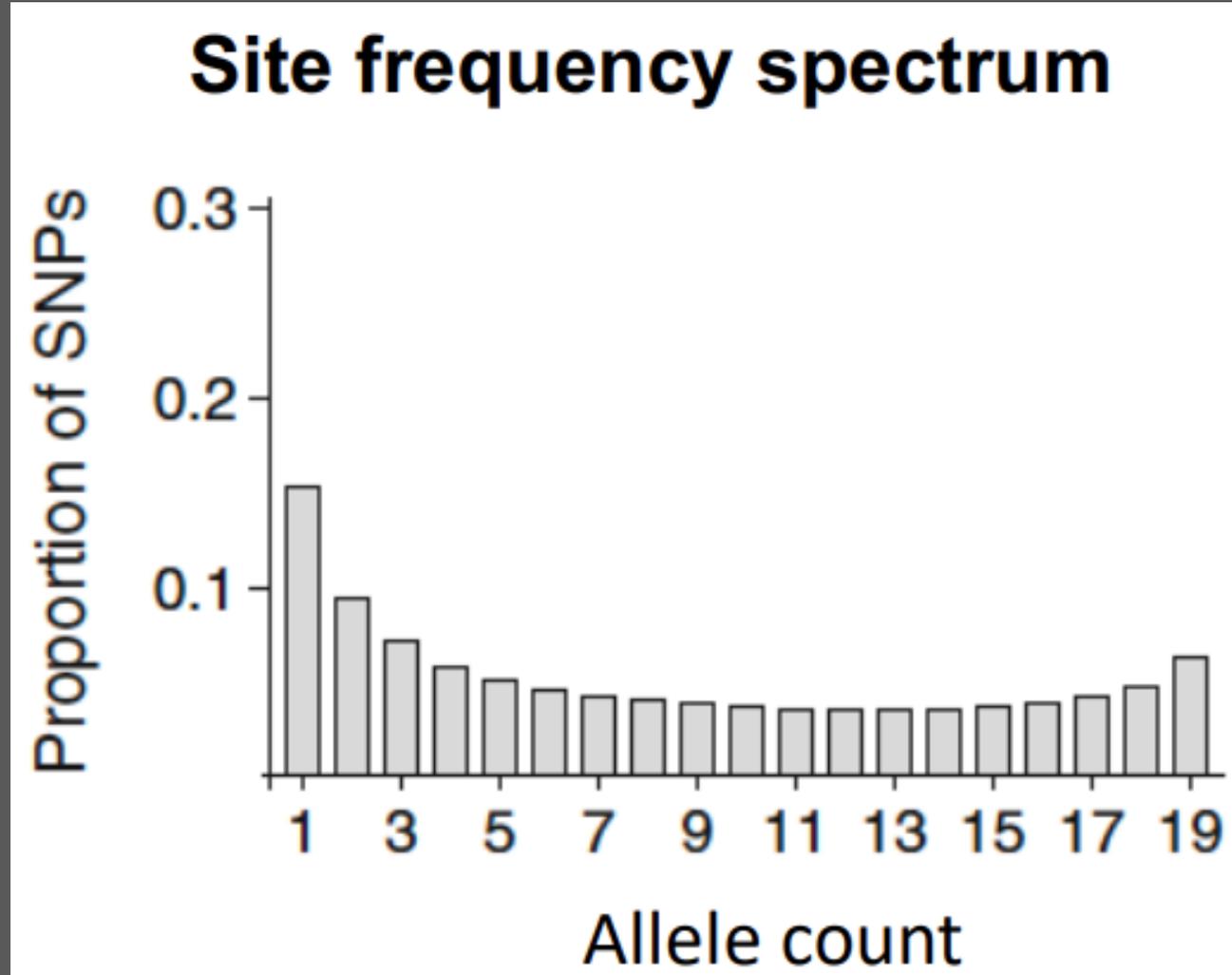
Crecimiento Populacional



Estrutura Populacional



Deriva Genética

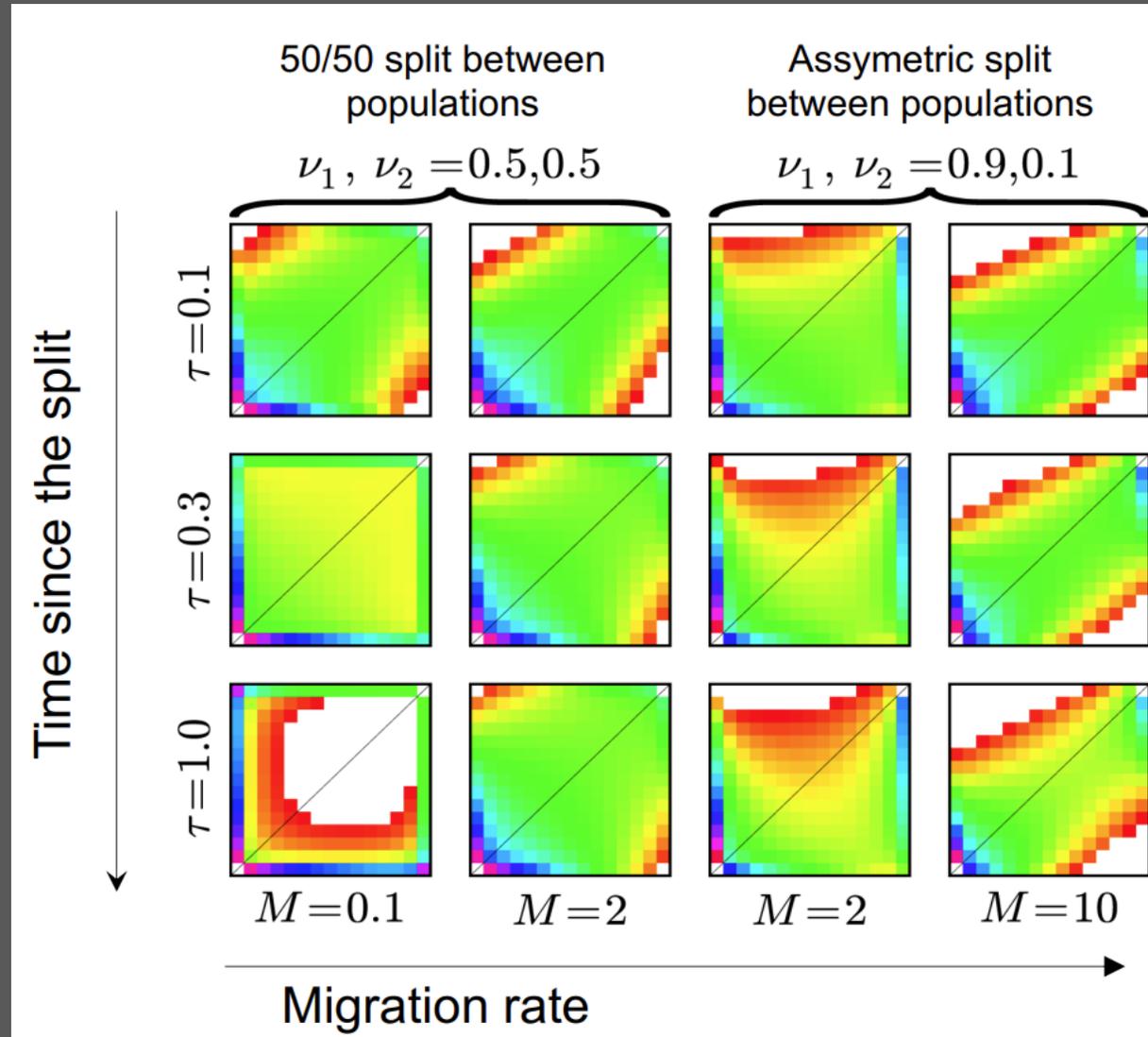


Importância do SFS

É possível inferir pela frequência alélica a que tipo de forças determinada população esteve sujeita.

É também possível compará-lo entre duas ou mais populações distintas.

Joint Site Frequency Spectrum (JSFS)



Fim!