

**Measures of sequence diversity**

sequence	sites																							
ancestor	·	C	·	·	C	·	·	A	·	·	T	·	·	A	·	·	G	·	·	C	·	·	G	·
1	C		T		A			G			C			G			C			T				
2	C		T		T			G			C			T			G			T				
3	G		T		A			T			C			G			G			T				
4	G		C		A			T			A			G			C			G				
5	C		T		A			G			C			G			G			T				
6	C		C		A			T			C			G			G			T				

- (a) What is  $S$  (number of segregating sites) for this dataset? What is  $n$  (sample size)?
  
- (b) Compute the site frequency spectrum (SFS):  $\xi_i$  is the number of sites with  $i$  copies of the mutant/derived allele and  $n - i$  copies of the ancestral allele.
  
- (c) Use the SFS to compute the *folded* SFS:  $\eta_i$  is the number of sites with  $i$  copies of the minor allele.
  
- (d) Use the folded SFS to compute the average pairwise heterozygosity  $\pi$  for this dataset.

**Partners (random):**

Arcadia & Em  
Hee & Chloe  
Karen & Mara