Measures of sequence diversity

sequence	sites							
ancestor	· C ·	· C ·	· A ·	· T ·	· A ·	· G ·	· C ·	· G ·
a	C	T	A	G	C	G	C	T
b	C	T	T	G	C	T	G	T
c	G	T	A	T	C	G	G	T
d	G	C	A	T	A	G	C	G
e	C	T	A	G	C	G	G	T
f	C	C	A	T	C	G	G	T

1. What is S (number of segregating sites) for this dataset? What is n (sample size)?

2. Compute the site frequency spectrum (SFS): ξ_i is the number of sites with i copies of the mutant/derived allele and $n - i$ copies of the ancestral allele.

3. Use the SFS to compute the *folded* SFS: η_i is the number of sites with i copies of the minor allele.

4. Find a formula for computing the average pairwise heterozygosity π from the folded SFS. Use your formula to compute π for this dataset.