

# Exercise (Kentaro Shimizu) $(\theta_\pi, \theta_T)$

## Nucleotide diversity $\pi$

Average proportion of pairwise differences between the sequences  
Basic index of molecular population genetics:

		. * . . . . * . *	(1)	(2)	(3)
(1) AGGCTGCATC		AGGCTGCATC	(1) ---		
(2) .A . . . . .	=	AAGCTGCATC	(2) ---	---	
(3) . . . . .T.C.		AGGCTGTACC	(3) ---	---	---

$$\pi = \sum_{i < j} \pi_{ij} / n_c = ?$$

$$n_c = \frac{n(n-1)}{2} \quad \text{number of pairwise comparisons}$$

$\pi_{ij}$  proportion of differences between i-th and j-th

## $\theta$ , another index

(nucleotide polymorphism based on polymorphic site)

		. * . . . . * . *	(1)	(2)	(3)
(1) AGGCTGCATC		AGGCTGCATC	(1) ---		
(2) .A . . . . .	=	AAGCTGCATC	(2) ---	---	
(3) . . . . .T.C.		AGGCTGTACC	(3) ---	---	---

$$\pi = \sum_{i < j} \pi_{ij} / n_c = ?$$

$$\theta = s / \sum_{k=1}^{n-1} \frac{1}{k} = ?$$

S: the proportion of polymorphic sites (or nucleotide polymorphism) observed in the sample, n: the number of sequences

# Test of neutrality: Tajima's $D$

example: balancing selection

Tajima Genetics 123, p. 229, 1989

(1) AGGCTGCATC

(2) .....

(3) .A.....C.

(4) .A.....C.

$$\pi = \sum_{i < j} \pi_{ij} / n_c = ?$$

$$\theta = s / \sum_{k=1}^{n-1} \frac{1}{k} = ?$$

$$D = \frac{\pi - \theta}{\text{standard\_deviation\_of\_}(\pi - \theta)} \quad \text{positive or negative?}$$

Intuitively: singleton vs. shared mutation

# Test of neutrality: Tajima's $D$

example: positive selection

Tajima Genetics 123, p. 229, 1989

(1) AGGCTGCATC

(2) .....

(3) .....

(4) .A.....C.

$$\pi = \sum_{i < j} \pi_{ij} / n_c = ?$$

$$\theta = s / \sum_{k=1}^{n-1} \frac{1}{k} = ?$$

$$D = \frac{\pi - \theta}{\text{standard\_deviation\_of\_}(\pi - \theta)} \quad \text{positive or negative?}$$

Intuitively: singleton vs. shared mutation