Short INDELS: genetic markers for adaptive divergence

Questions:

1. Can we incorporate short INDELs as genetic markers in studies of adaptive divergence?
2. What is their distribution across the genome?
3. What is the proportion of selectively neutral, deleterious and beneficial INDELs?

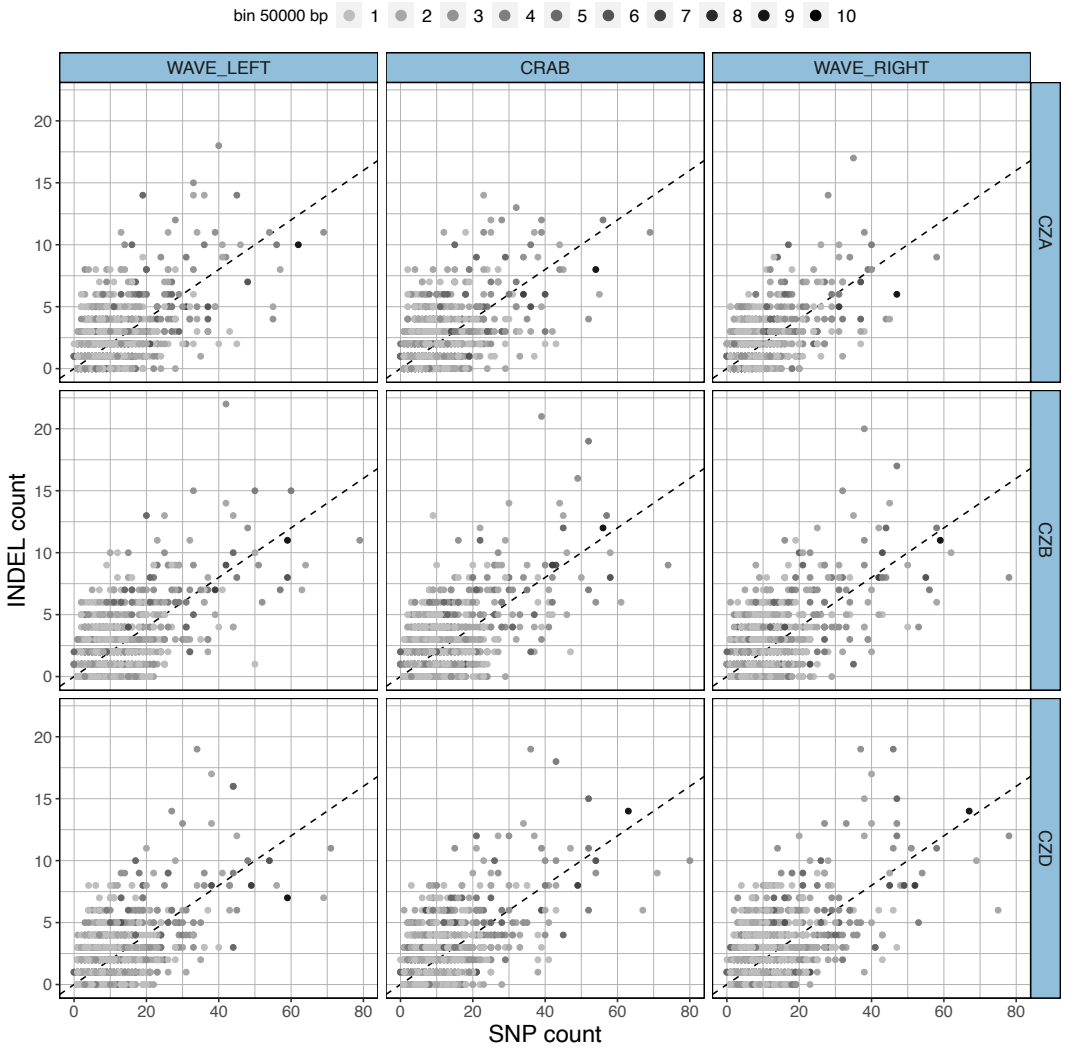
Original aspects:

- Divergent natural selection vs neutral processes
- Species with high diversity
- Systems with imperfect genomes can still contain useful functional information

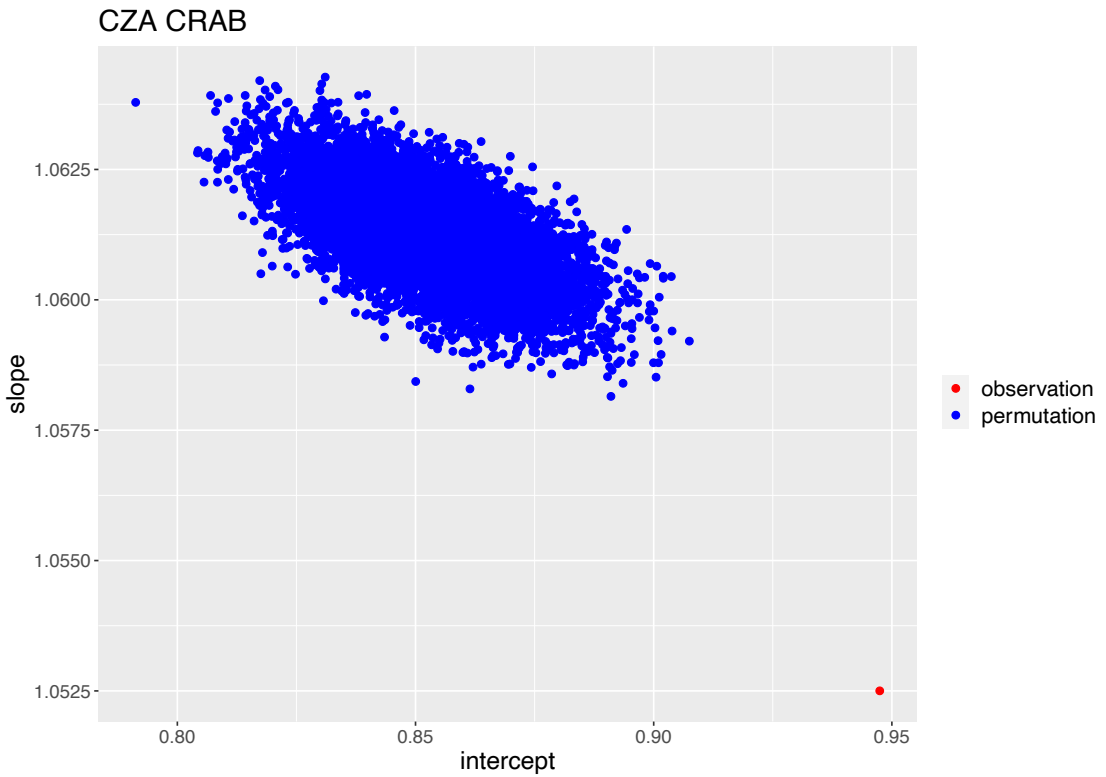
INDEL-SNP comparisons:

1. Clustering of INDEL and SNP markers
2. Unfolded allele frequency spectra (uAFS)
3. Outlier sharing
4. Distribution of cline parameters

1. Clustering of INDEL and SNP markers



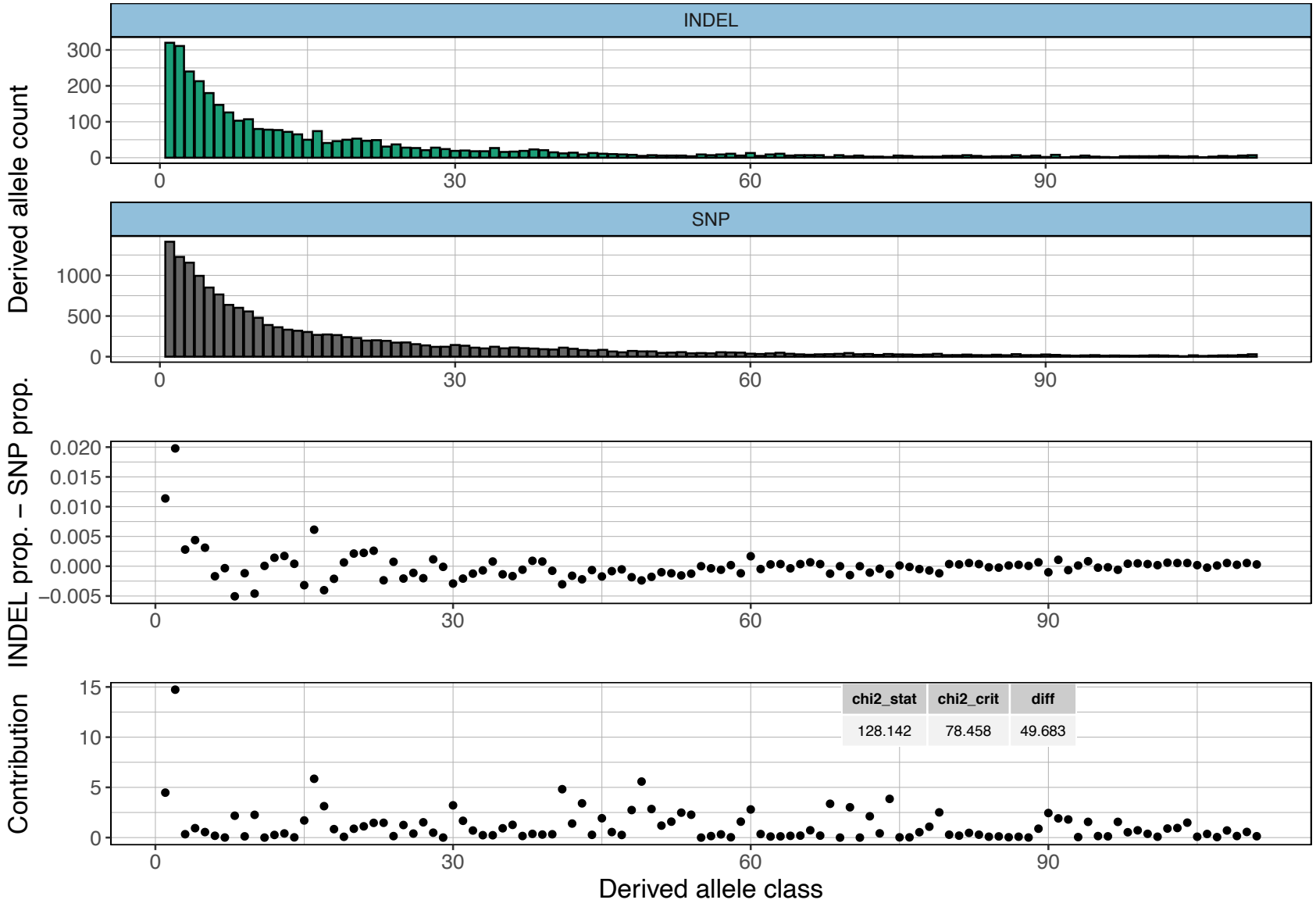
Dashed line = 0.2 which is obtained from the ratio of the total number of INDELs and the total number of SNPs for a given ecotype in each island



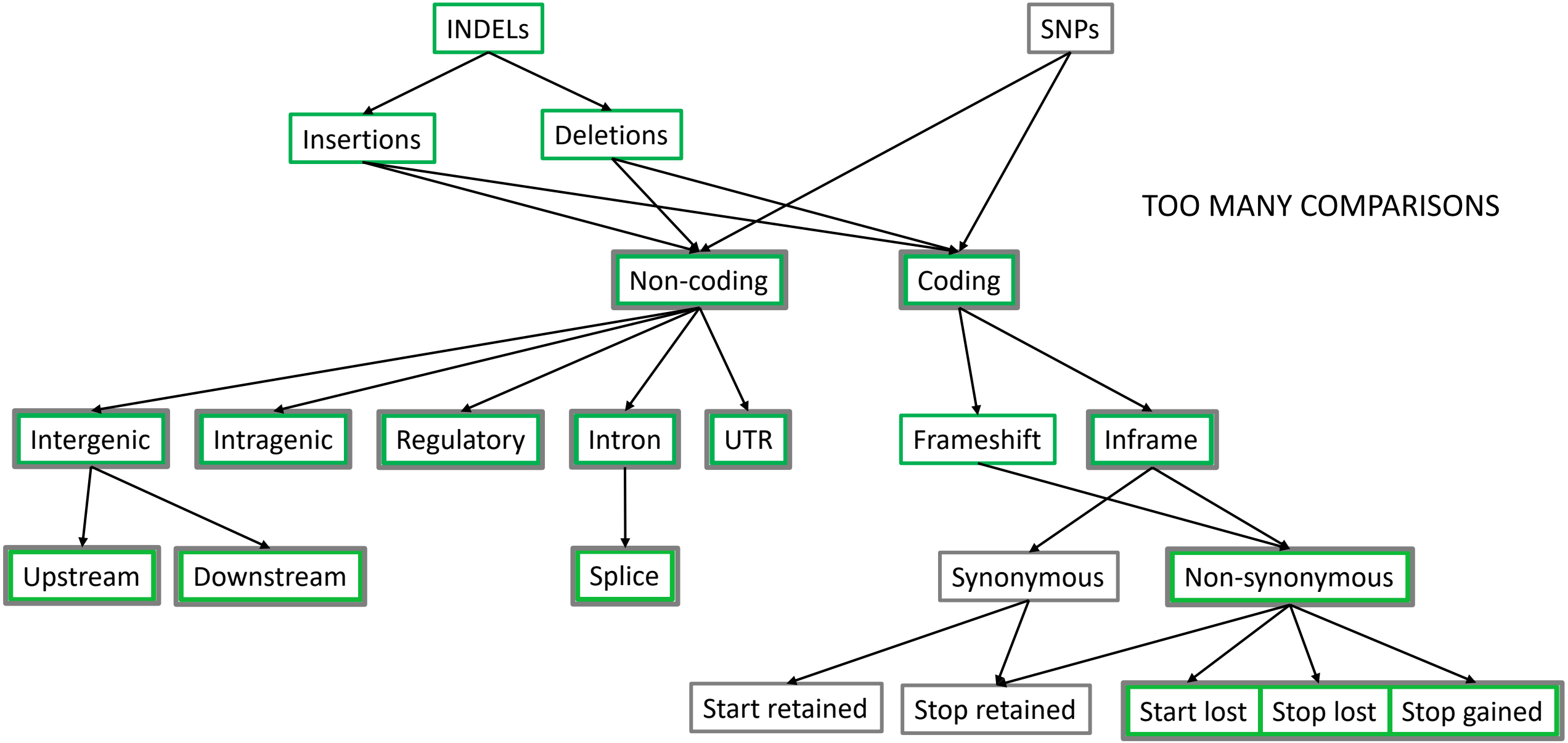
Blue cloud: expected relationship given the ratio between INDEL and SNP count

Red dot: the observed relationship consists of a higher concentration of INDELs in fewer regions

2. Unfolded allele frequency spectra

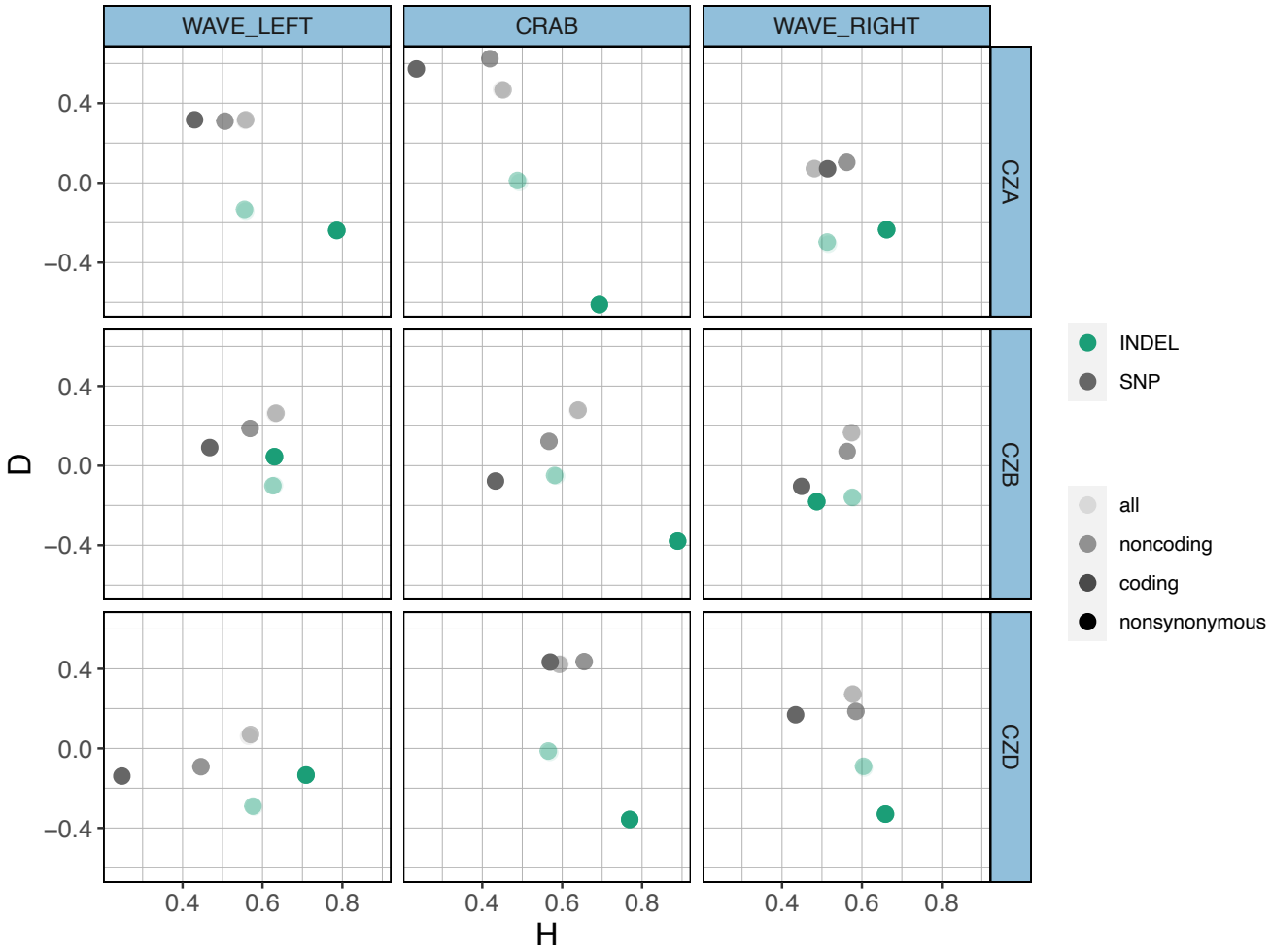


2. Unfolded allele frequency spectra



2. Unfolded allele frequency spectra

A way to summarise...



Tajima's D < 0: excess of low frequencies

Fay and Wu's H > 0: deficit of moderate- and high-frequencies

2. Unfolded allele frequency spectra

... And test for similarity of the patterns

One linear model for each of the four category where
y = D or H
x = factor with two levels, INDELs and SNPs

* = slope is significantly different from 0

INDELs and SNPs differed significantly for D and/or H

		SNPs			
	D,H	All	Non-coding	Coding	Non-synonymous
INDELs	All	*,			
	Non-coding		*,		
	Coding			*, *	
	Non-synonymous				*, *

2. Unfolded allele frequency spectra – GC-biased gene conversion

S = Strong = G or C W = Weak = A or T

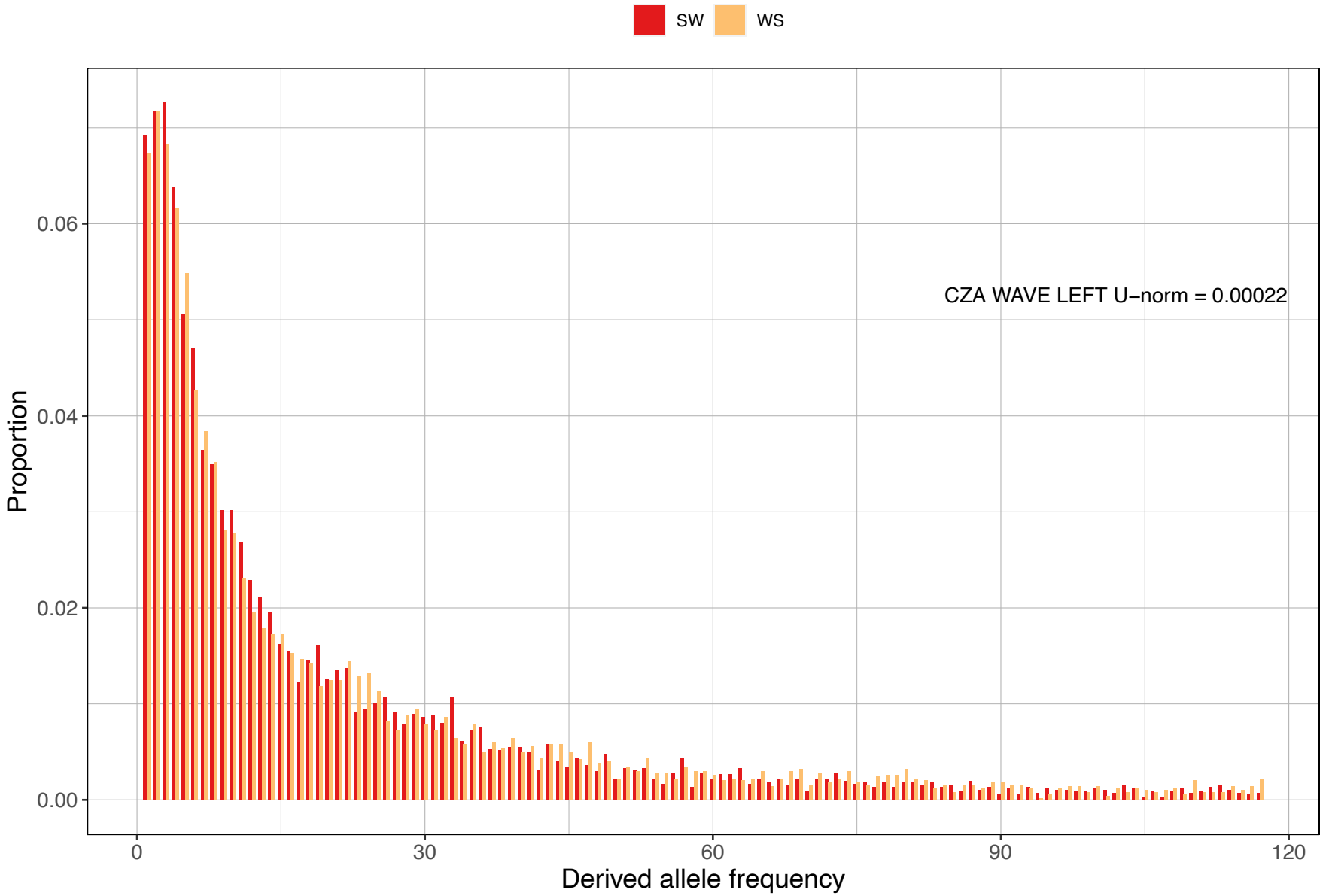
WS = Weak ancestral, Strong derived SW = Strong ancestral, Weak derived

D: WS > SW

H: WS < SW

WS = less low and more high frequencies SW = more low and less high frequencies

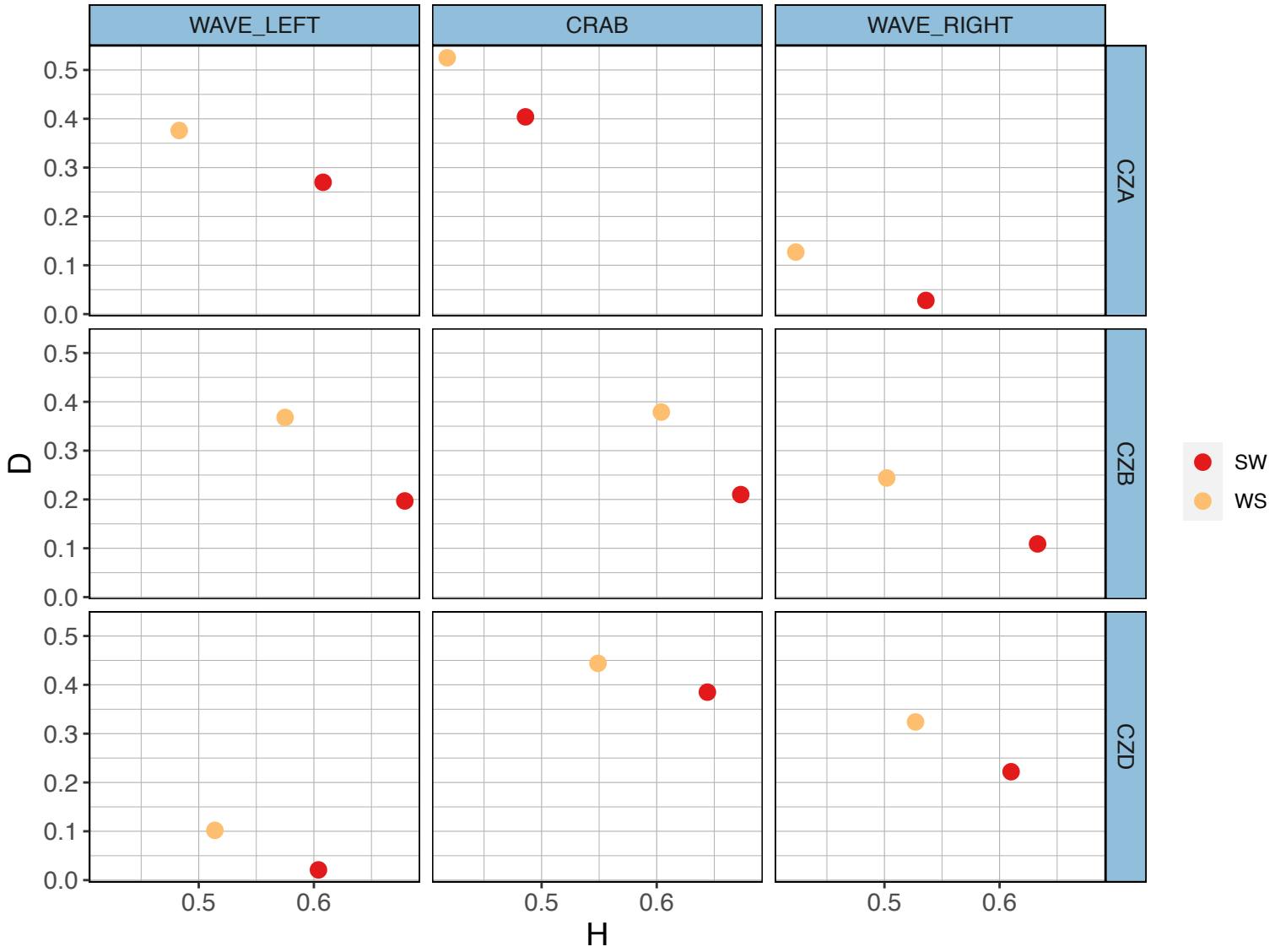
2. Unfolded allele frequency spectra – GC-biased gene conversion



CZA WAVE LEFT just as an example

U-norm = normalised U statistic as in Katzman et al 2011 and Lachance and Tishkoff 2014

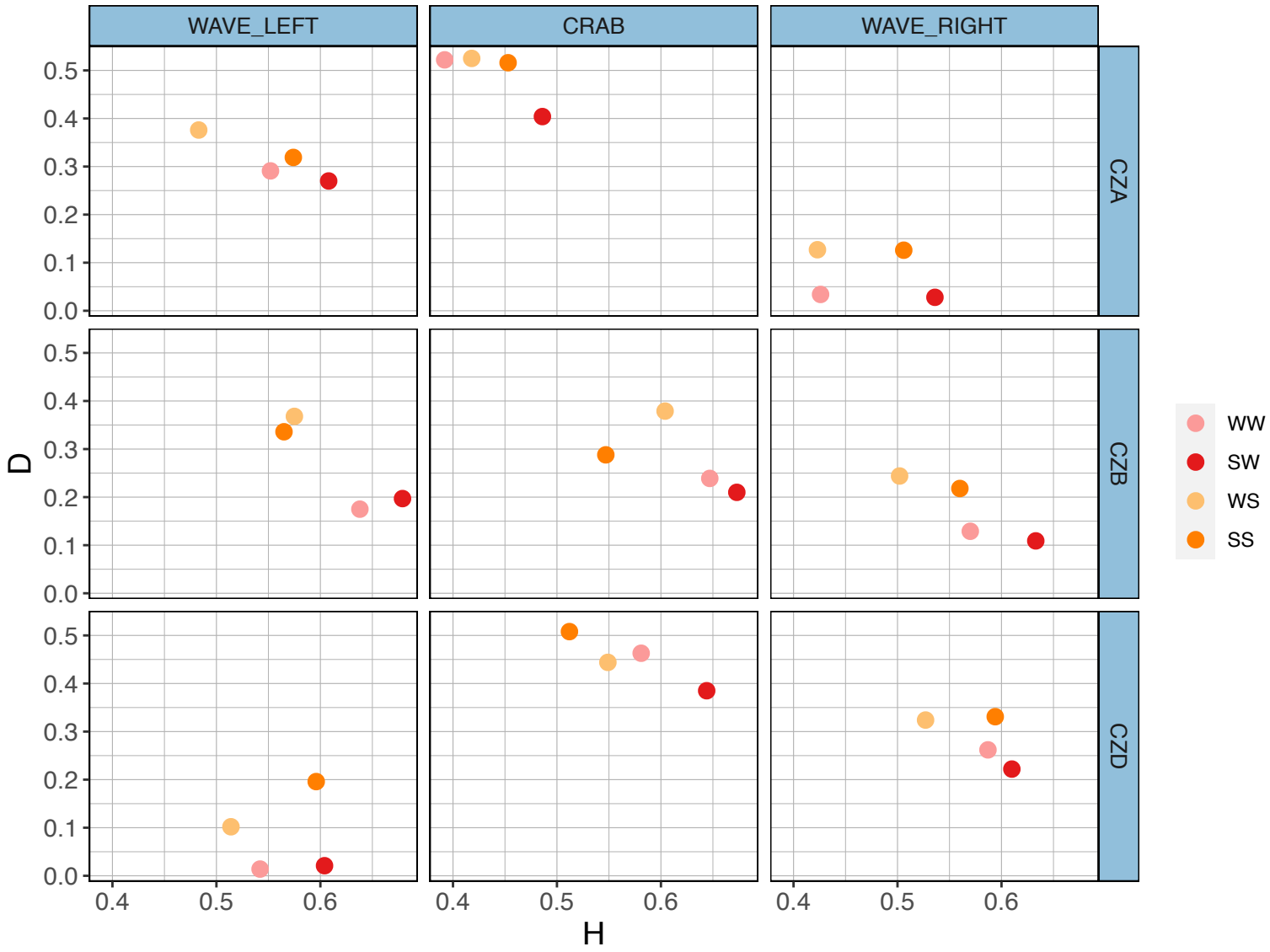
2. Unfolded allele frequency spectra – GC-biased gene conversion



D: WS > SW

H: WS < SW

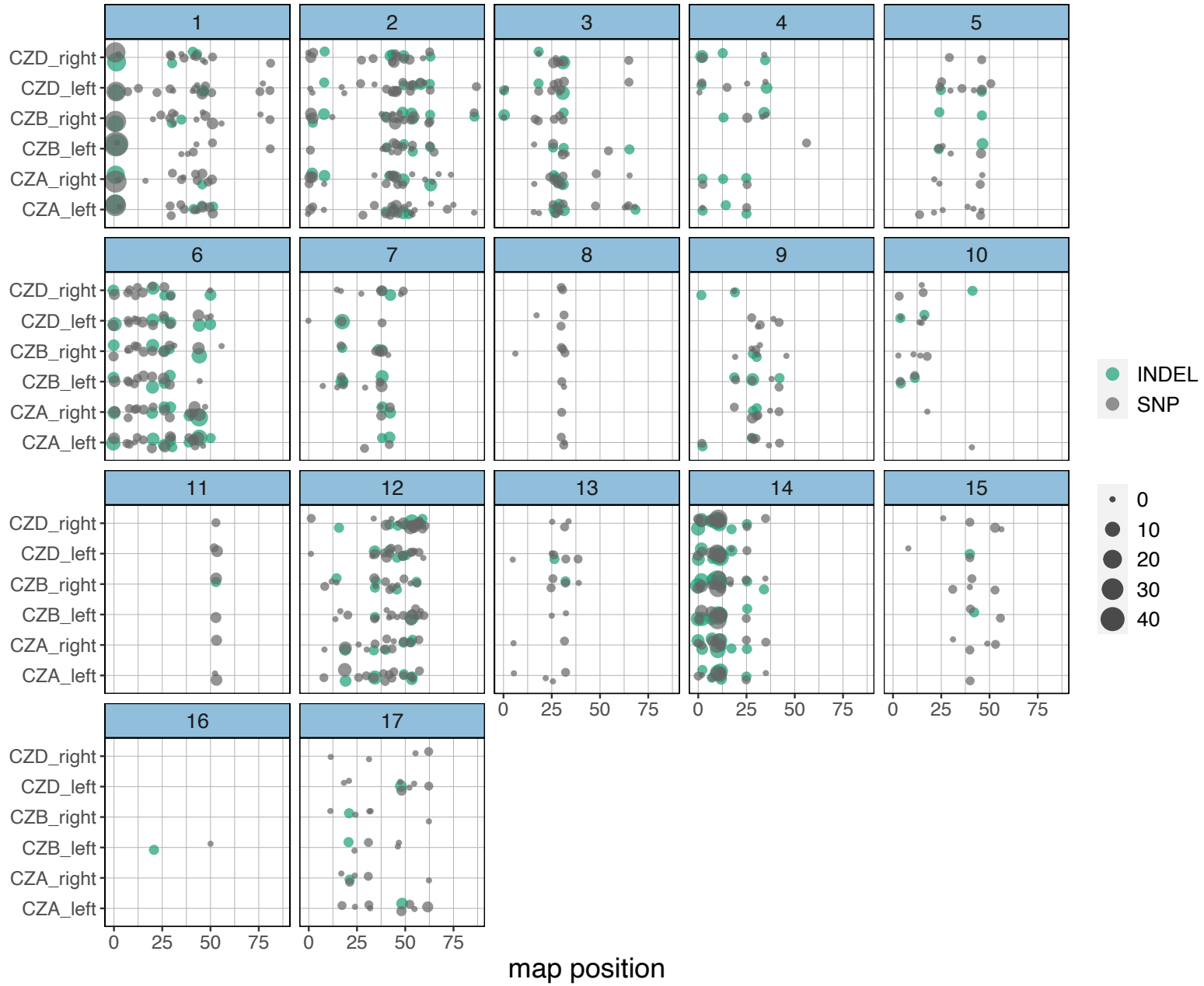
2. Unfolded allele frequency spectra – GC-biased gene conversion



D: WS > SW WW SS

H: WS < SW WW SS

3. Outlier sharing



4. Distribution of cline parameters