

Short INDELS: genetic markers  
for adaptive divergence

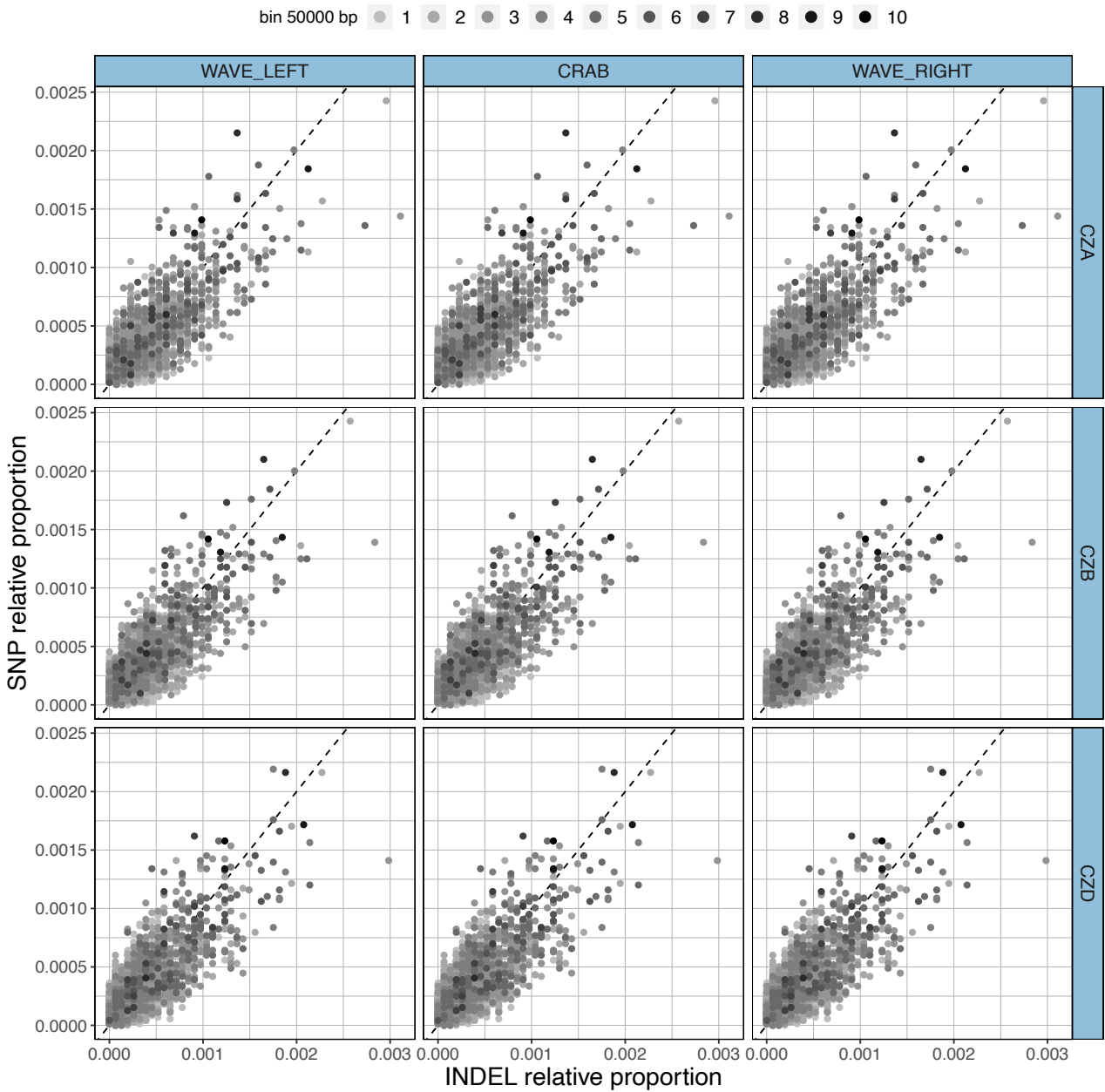
Original aspects of the short INDELs paper:

- 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 (different types) markers
2. Unfolded allele frequency spectra (uAFS)
3. Outlier sharing
4. Distributions of cline parameters

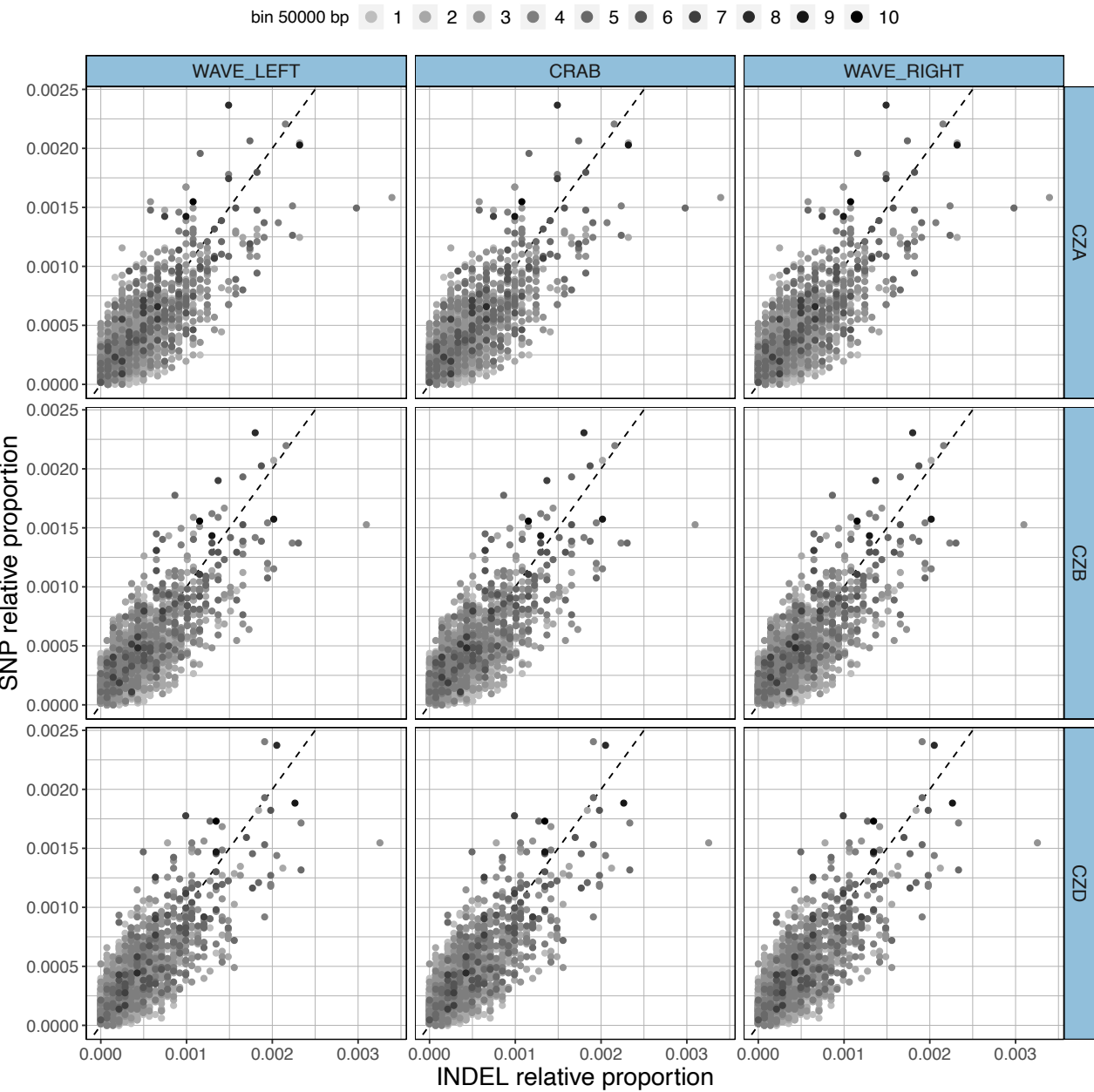
# 1. Clustering of (different types) markers



- INDELs and SNPs after filtering but before cline analysis
- Kept variants inside and outside inversions  
Unique INDELs = 20005  
Unique SNPs = 89770
- With map position  
Unique INDELs = 7209  
Unique SNPs = 34290
- Without map position  
Unique INDELs = 12796  
Unique SNPs = 55480

WAVE LEFT	CRAB	WAVE RIGHT	INDEL-SNP CORRELATION
0.77	0.77	0.77	CZA
0.79	0.79	0.79	CZB
0.79	0.79	0.79	CZD

# 1. Clustering of (different types) markers



- INDELs and SNPs after filtering but before cline analysis
- Kept only variants outside inversions  
Unique INDELs = 18373  
Unique SNPs = 82087
- With map position  
Unique INDELs = 5577  
Unique SNPs = 26607
- Without map position  
Unique INDELs = 12796  
Unique SNPs = 55480

WAVE LEFT	CRAB	WAVE RIGHT	INDEL-SNP CORRELATION
0.78	0.78	0.78	CZA
0.79	0.79	0.79	CZB
0.79	0.79	0.79	CZD

## 2. Unfolded allele frequency spectra (uAFS)

- Ancestral state was inferred from called genotypes:
  - Reference allele = ancestral allele = ref\_anc  
*compressa* is homo for the reference allele (0)
  - Alternative allele = ancestral allele = alt\_anc  
*compressa* is homo for the alternative allele (2)
  - Unknown ancestry = het  
*compressa* is het (1)
- Fixed differences removed:
 

SNPs = 148; INDELs = 42; ratio = 3.5
- Total number of polymorphic unique variants before cline analysis:
 

SNPs = 89770; INDELs = 20005; ratio = 4.5

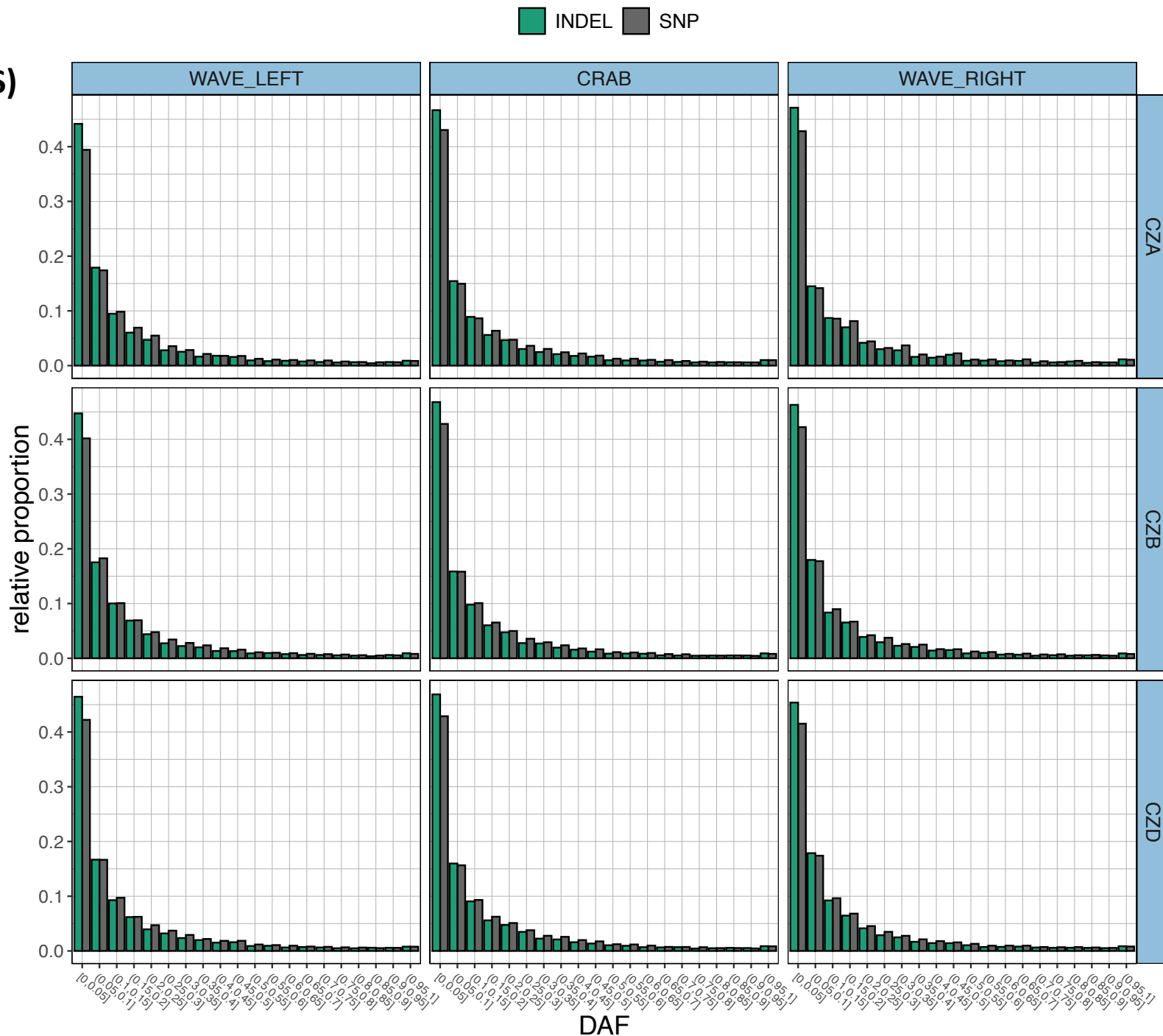
**For the rest of the results, kept only variants highlighted in green that were polymorphic.**

*Table 1. Count of INDELs and SNPs for each combination of possible allelic states given one outgroup (*L. compressa*) with two samples (NE and W). There are two combinations in which the allelic state is concordant in both samples (in green), eight in which the allelic state can only be retrieved from one sample (in yellow) and finally, five in which the allelic state cannot be inferred (in red).*

NE_Lcomp	W_Lcomp	INDEL	SNP
alt_anc	alt_anc	5305	27188
alt_anc	het	528	3543
alt_anc	NA	245	1097
alt_anc	ref_anc	511	2195
het	alt_anc	2231	12439
het	het	1577	9691
het	NA	151	627
het	ref_anc	3120	17198
NA	alt_anc	158	765
NA	het	33	267
NA	ref_anc	449	1831
ref_anc	alt_anc	693	3292
ref_anc	het	1422	7462
ref_anc	NA	1003	3675
ref_anc	ref_anc	38884	178715

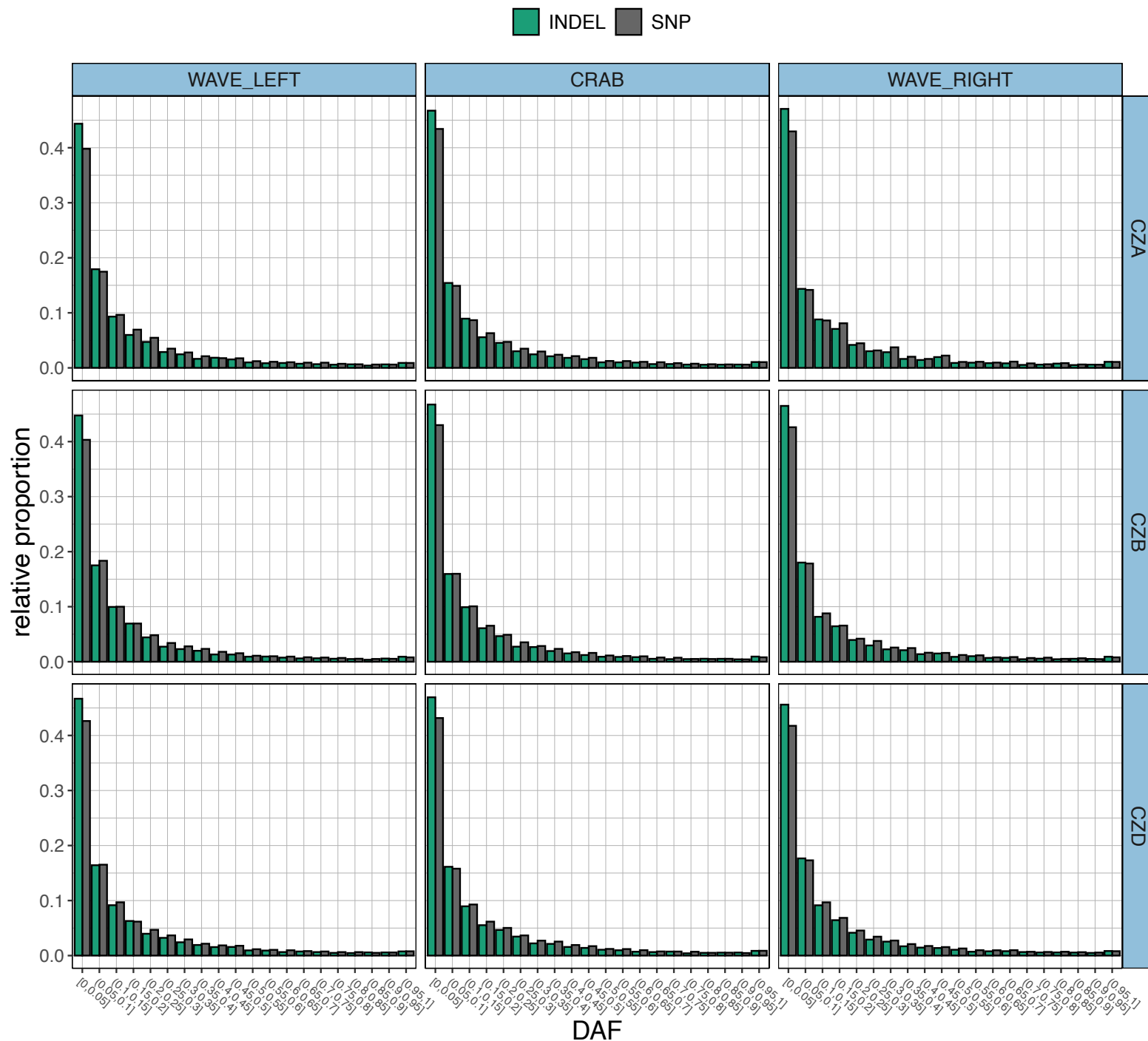
## 2. Unfolded allele frequency spectra (uAFS)

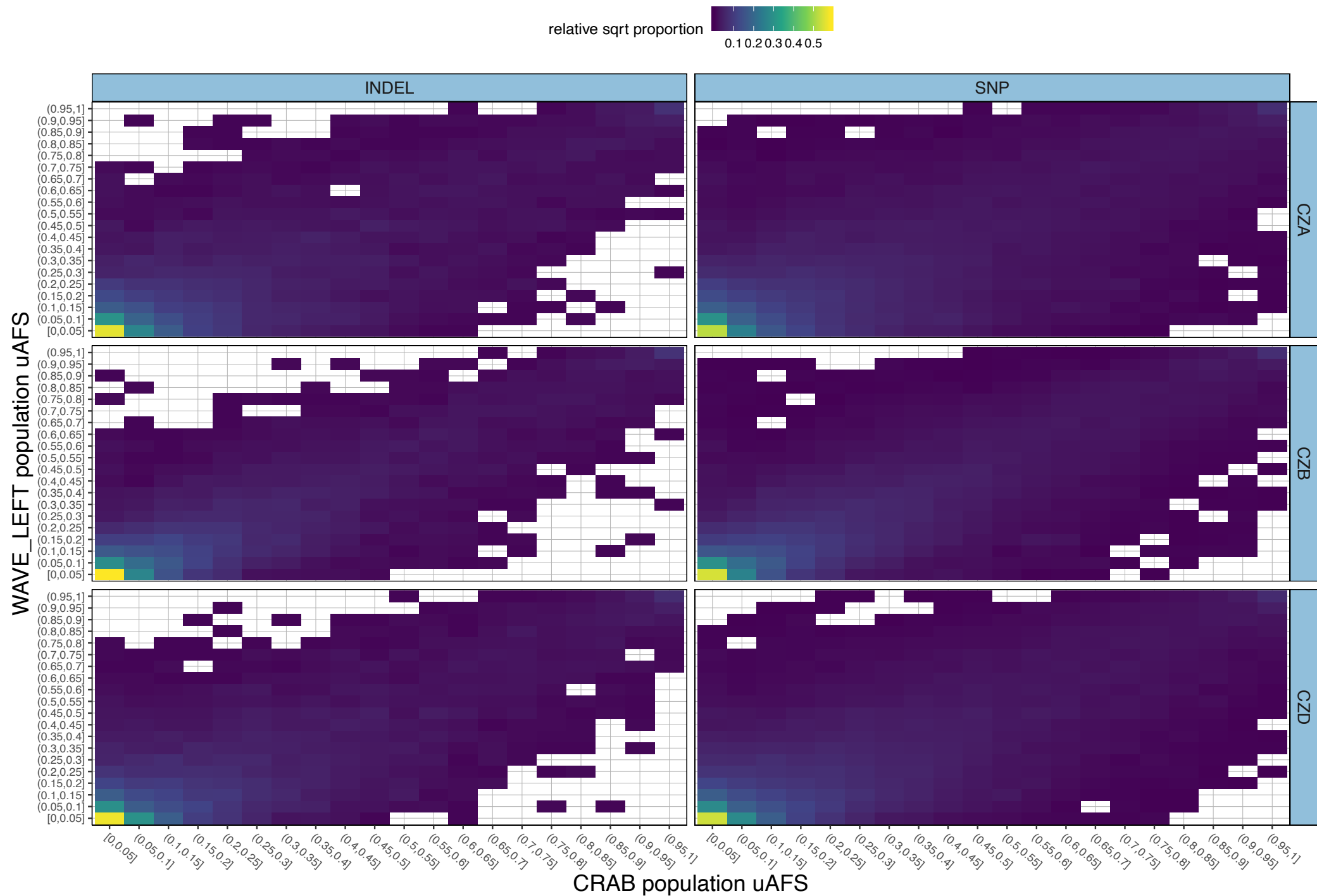
- Relative proportion of INDELs and SNPs after filtering but before cline analysis
- Kept variants inside and outside inversions
- Two-sample Kolmogorov-Smirnov test  
Null hypothesis:  
INDELs and SNPs are drawn from the same continuous distribution
  - Using relative proportions, the null cannot be rejected
  - Using square root counts, the null can be rejected



## 2. Unfolded allele frequency spectra (uAFS)

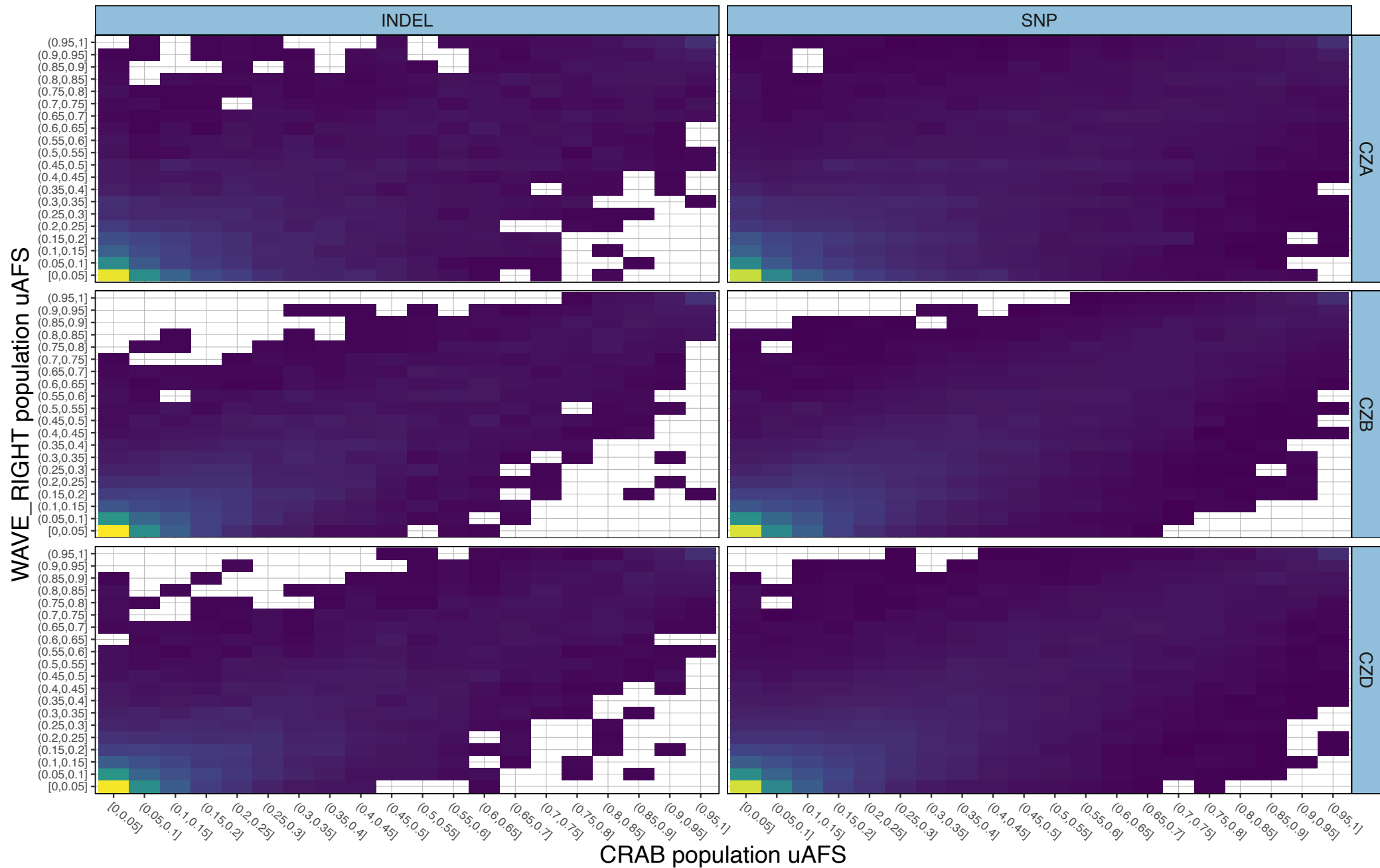
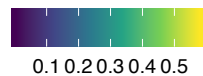
- Relative proportion of INDELs and SNPs after filtering but before cline analysis
- Kept only variants outside inversions
- Two-sample Kolmogorov-Smirnov test  
Null hypothesis:  
INDELs and SNPs are drawn from the same continuous distribution
  - Using relative proportions, the null cannot be rejected
  - Using square root counts, the null can be rejected

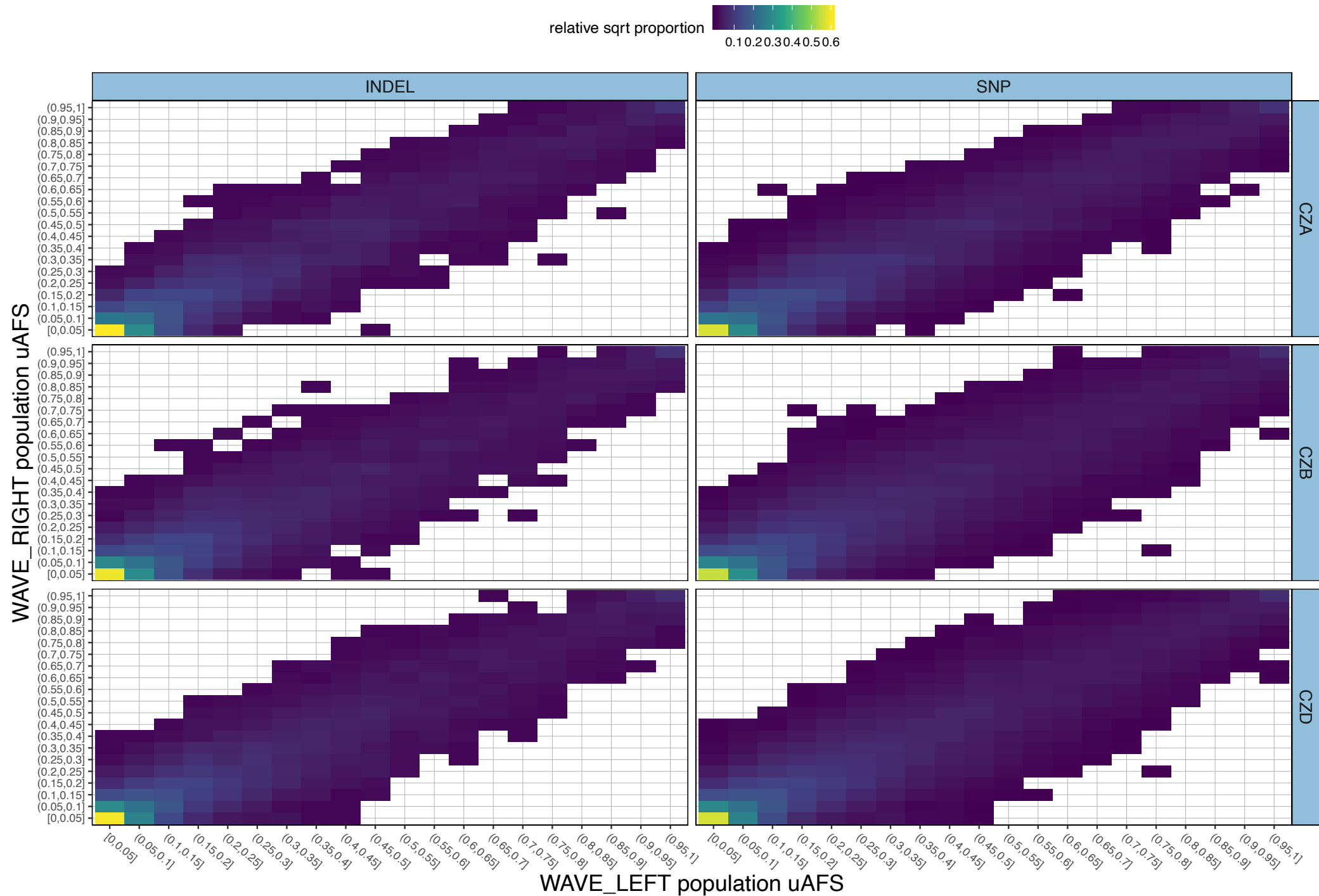






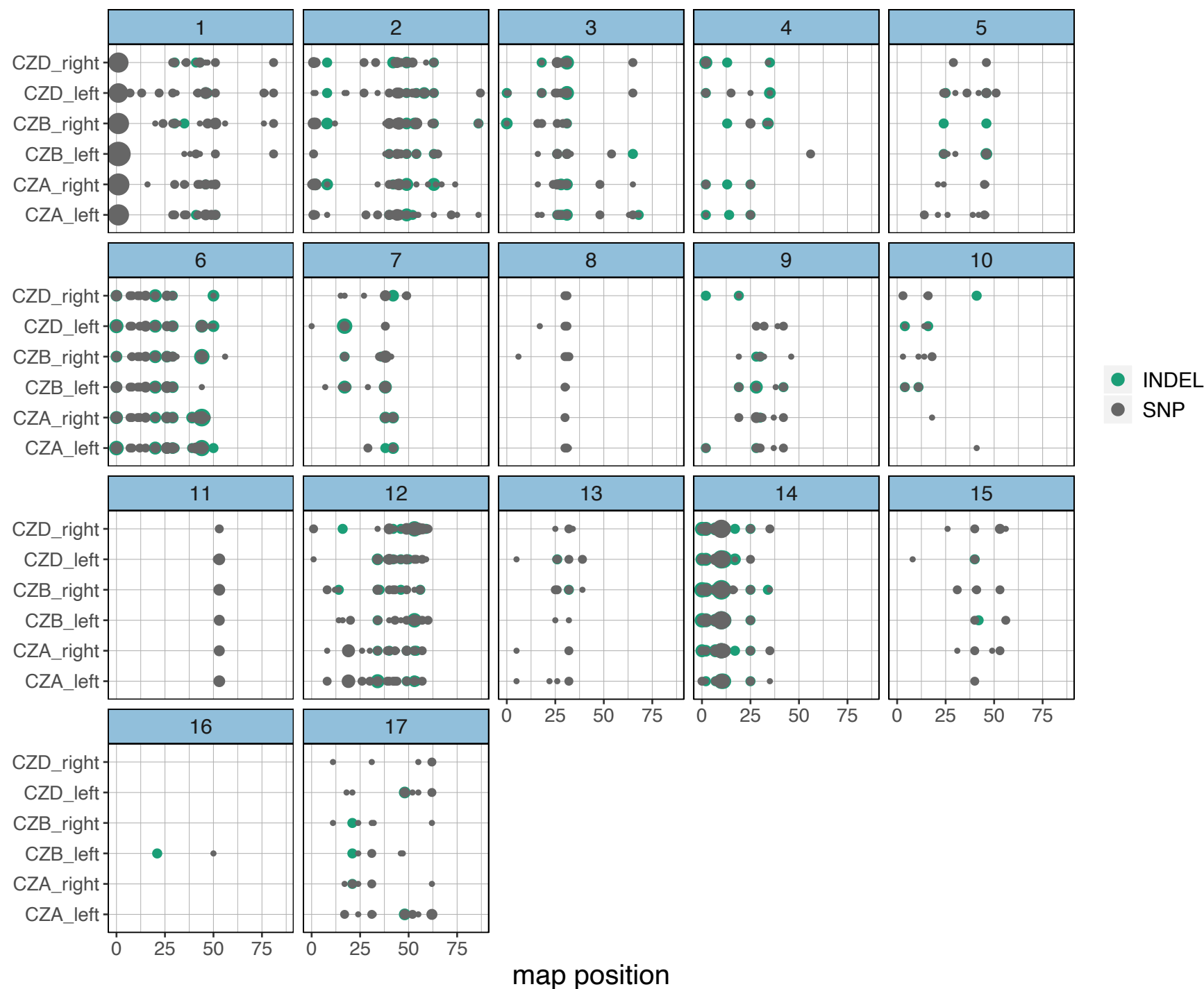
relative sqrt proportion





### 3. Outlier sharing

- Variants after cline analysis (maf filter 0.1)



INDEL SNP

