

Short INDELS: genetic markers
for adaptive divergence

Original aspects of the short INDELs paper:

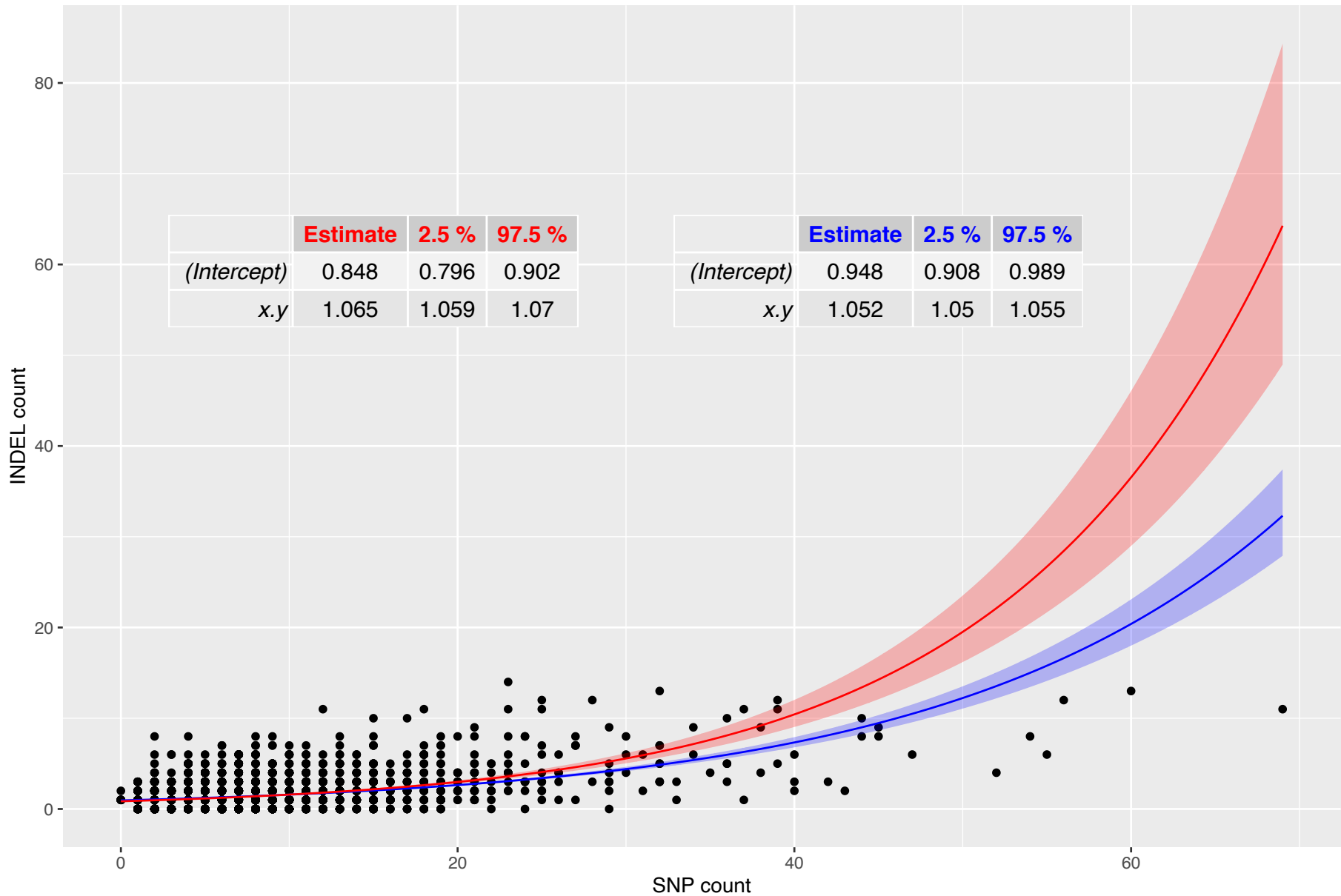
- Divergent natural selection vs neutral processes
- Species with high diversity
- Systems with imperfect genomes can still contain useful functional information

INDEL-SNP comparisons:

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

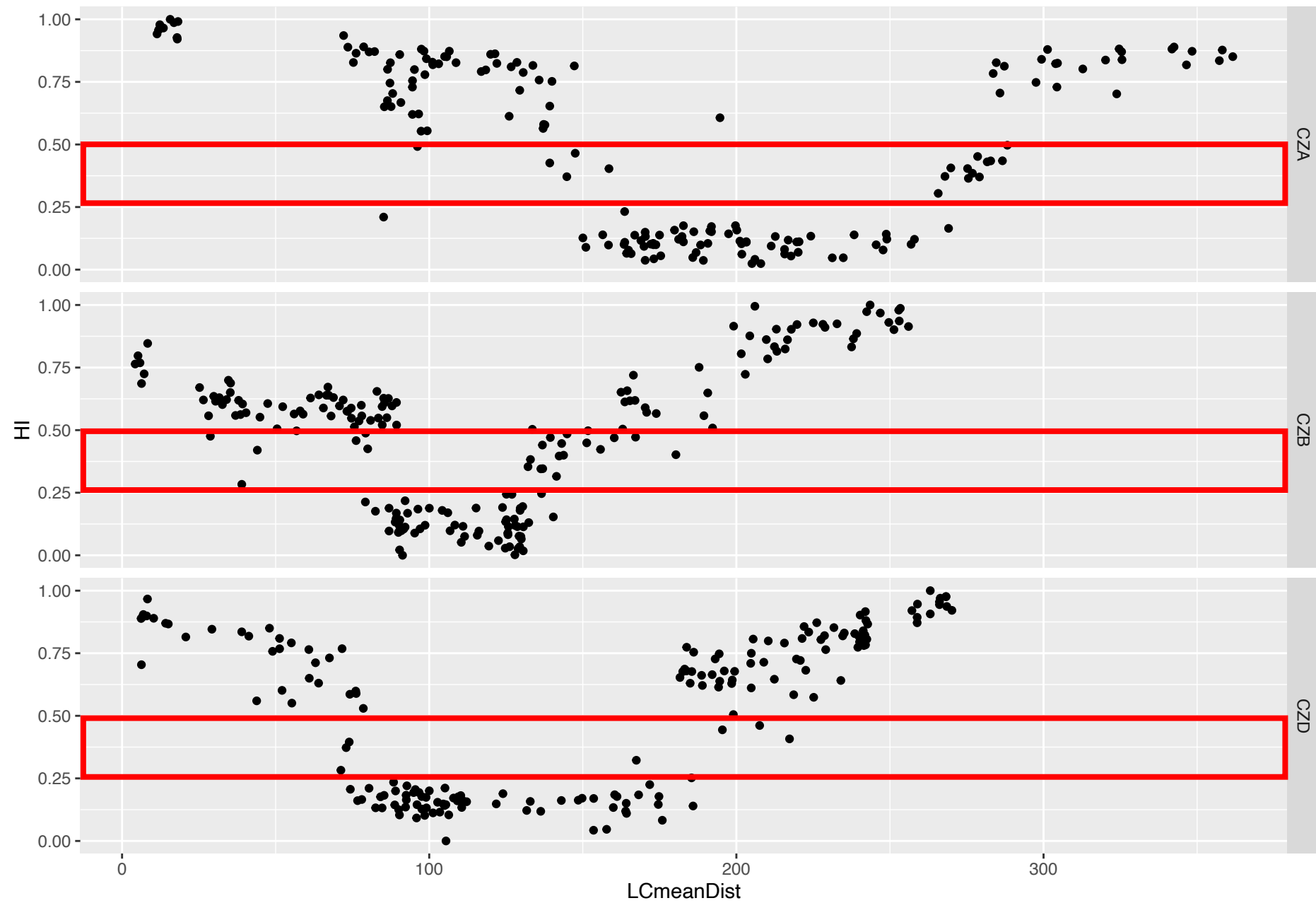
1. Clustering of (different types) markers

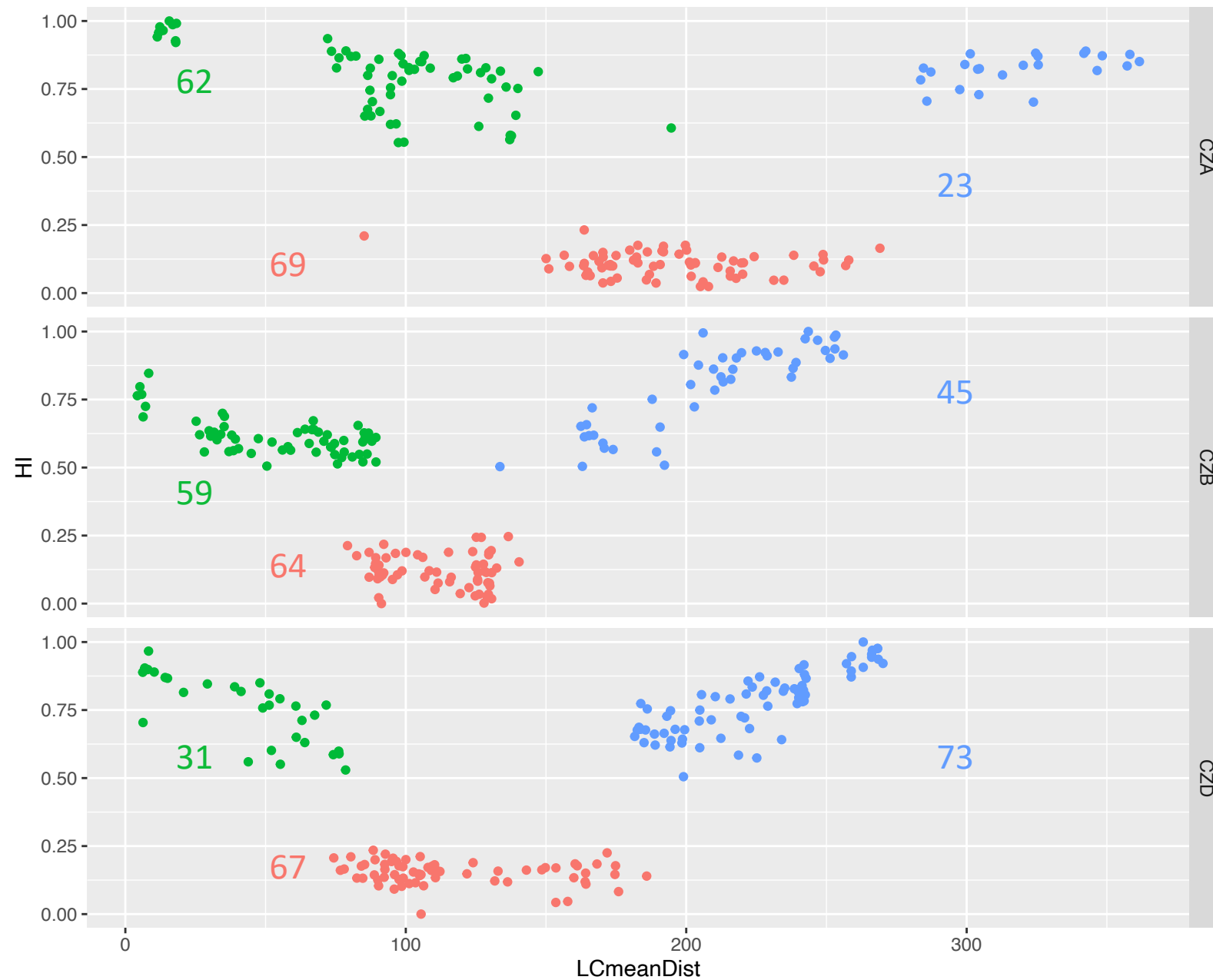
CZA CRAB



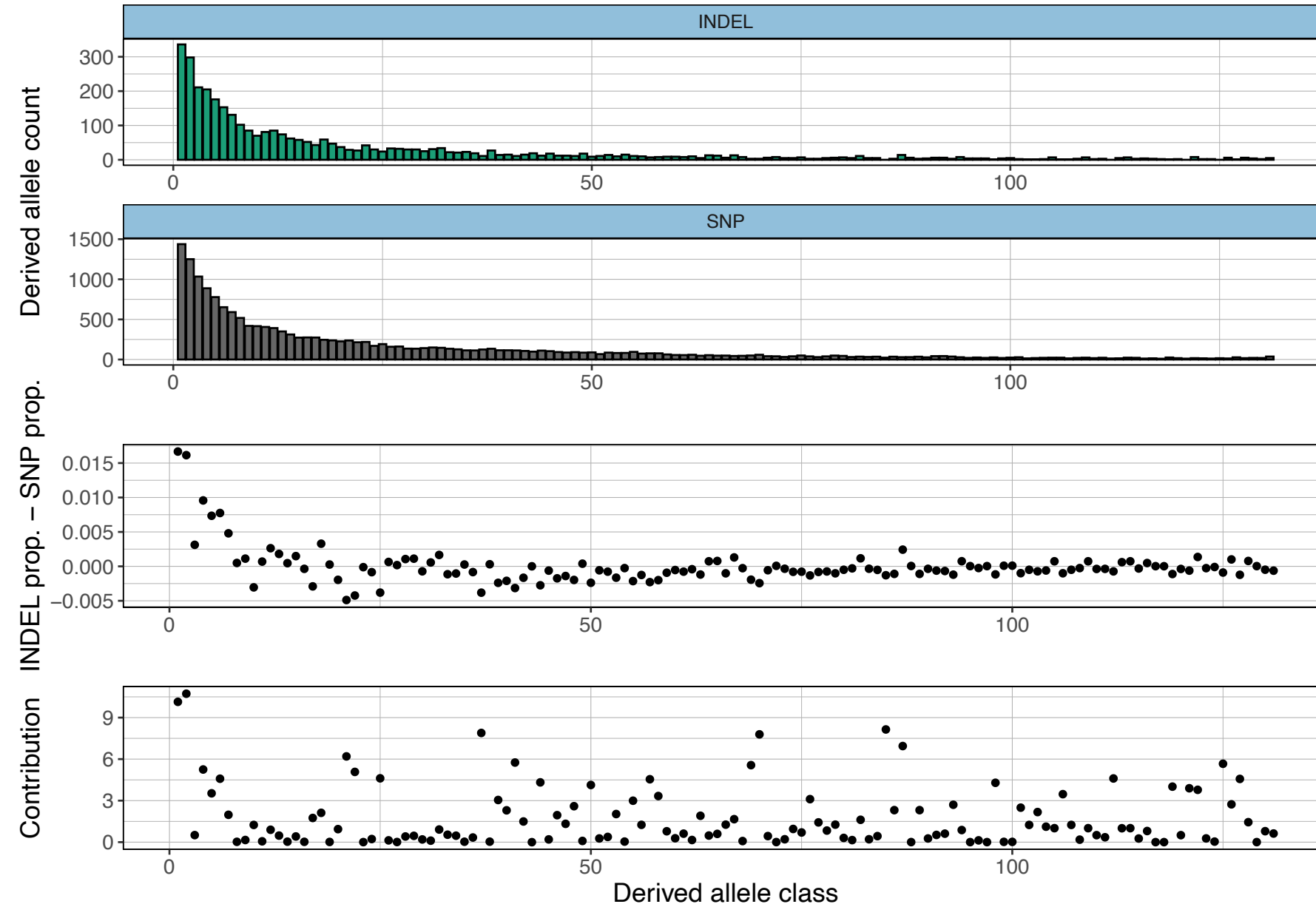
- INDELs and SNPs after filtering but before cline analysis
- Kept variants inside and outside inversions
Unique INDELs = 20005
Unique SNPs = 89770
- With map position and inversions
Unique INDELs = 7206
Unique SNPs = 34287
- Without map position
Unique INDELs = 12796
Unique SNPs = 55480

WAVE LEFT	CRAB	WAVE RIGHT	INDEL-SNP CORRELATION
0.56	0.53	0.55	CZA
0.59	0.59	0.59	CZB
0.57	0.57	0.58	CZD





2. Unfolded allele frequency spectra (uAFS)

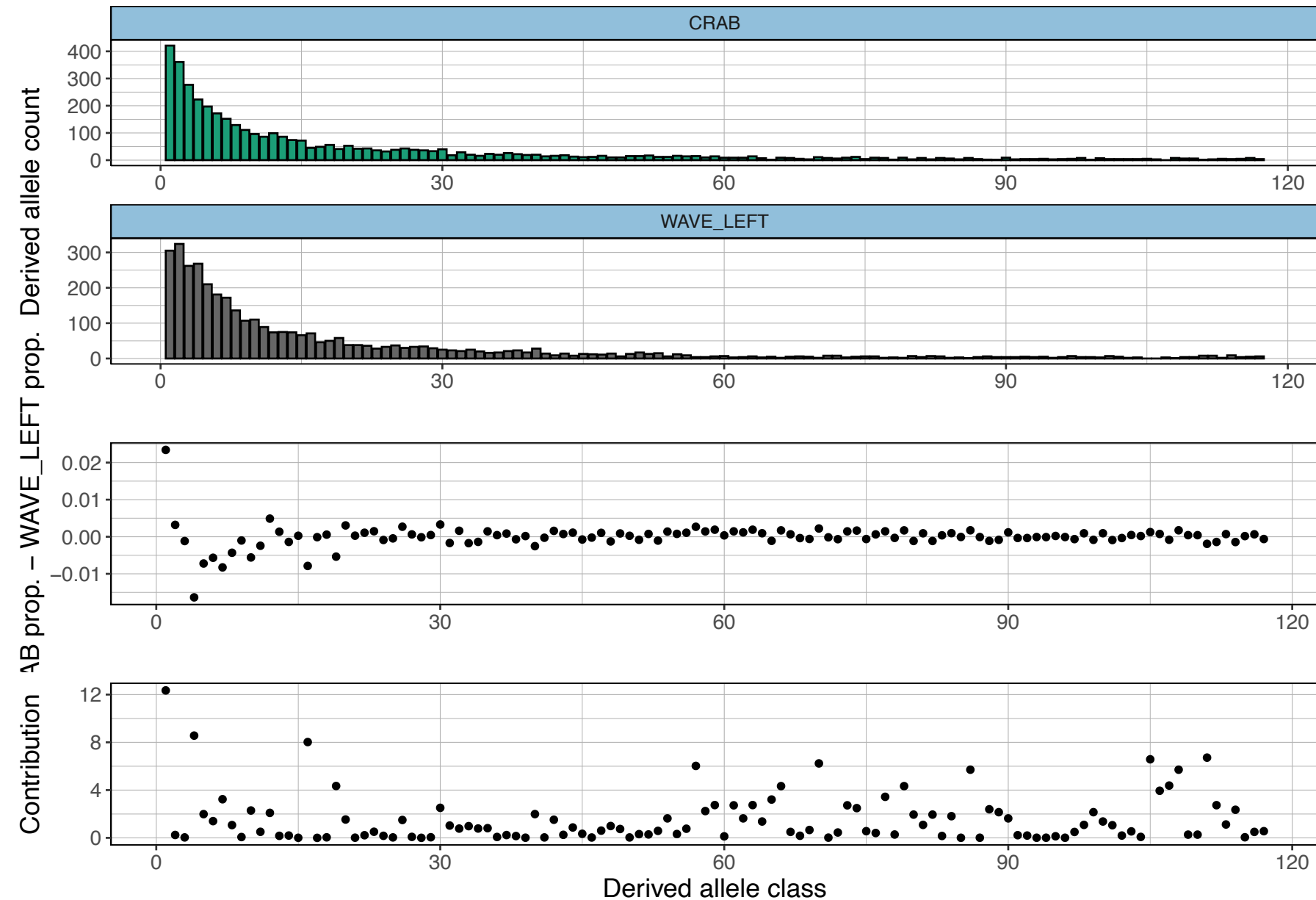


- Kept variants inside and outside inversions, present in at least $N - 3$ individuals. When the sample size of a variant was higher than $N - 3$, $N - 3$ individuals/genotypes were randomly sampled.

• CZA CRAB INDELs vs SNPs.

Chi-square test statistic = 224.3352
Chi-square critical value at 0.05 with 130 df = 104.6622
Test statistic - critical value = 119.6729

2. Unfolded allele frequency spectra (uAFS)

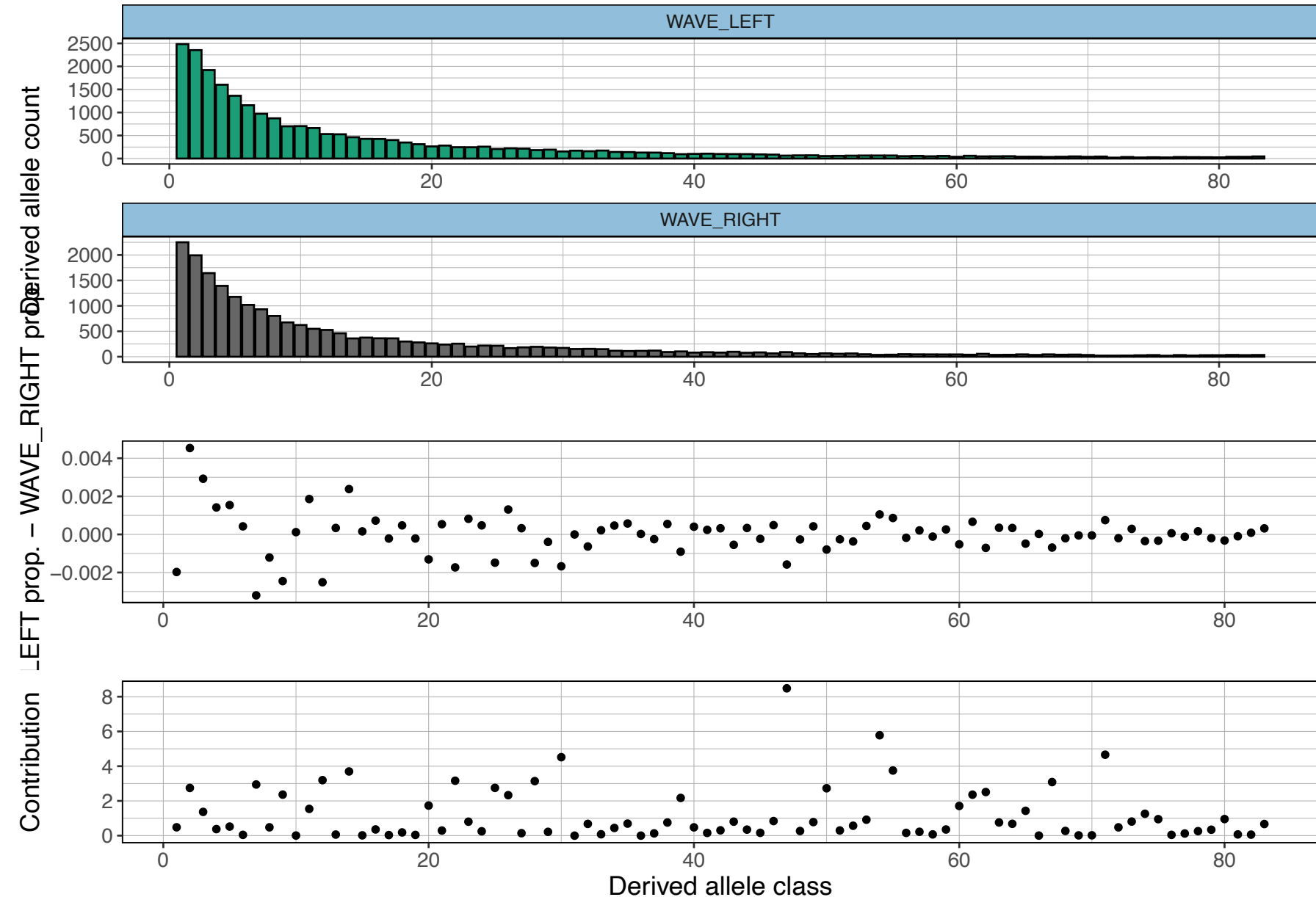


- Kept variants inside and outside inversions, present in at least $N - 3$ individuals of the population with the lowest N . When the sample size of a variant was higher than $N - 3$, $N - 3$ individuals/genotypes were randomly sampled.

- CZA INDELs CRAB vs WAVE LEFT.

Chi-square test statistic = 177.9672
 Chi-square critical value at 0.05 with 116 df = 92.13376
 Test statistic - critical value = 85.83344

2. Unfolded allele frequency spectra (uAFS)



- Kept variants inside and outside inversions, present in at least $N - 3$ individuals of the population with the lowest N . When the sample size of a variant was higher than $N - 3$, $N - 3$ individuals/genotypes were randomly sampled.
- CZB SNPs WAVE LEFT vs WAVE RIGHT.

Chi-square test statistic = 94.57866
 Chi-square critical value at 0.05 with 82 df = 62.13229
 Test statistic - critical value = 32.44637

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 and no inversions
Unique INDELs = 5576
Unique SNPs = 26604
- Without map position
Unique INDELs = 12796
Unique SNPs = 55480

WAVE LEFT	CRAB	WAVE RIGHT	INDEL-SNP CORRELATION
0.55	0.53	0.54	CZA
0.59	0.59	0.59	CZB
0.57	0.58	0.58	CZD

2. Unfolded allele frequency spectra (uAFS)

- Ancestral state was inferred from called genotypes:
 1. Reference allele = ancestral allele = ref_anc
compressa is homo for the reference allele (0)
 2. Alternative allele = ancestral allele = alt_anc
compressa is homo for the alternative allele (2)
 3. Unknown ancestry = het
compressa is het (1)

	Fixed	Polymorphic	Ratio
INDELs	10511	20000	1.9
SNPs	43355	89745	2.1

*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).*

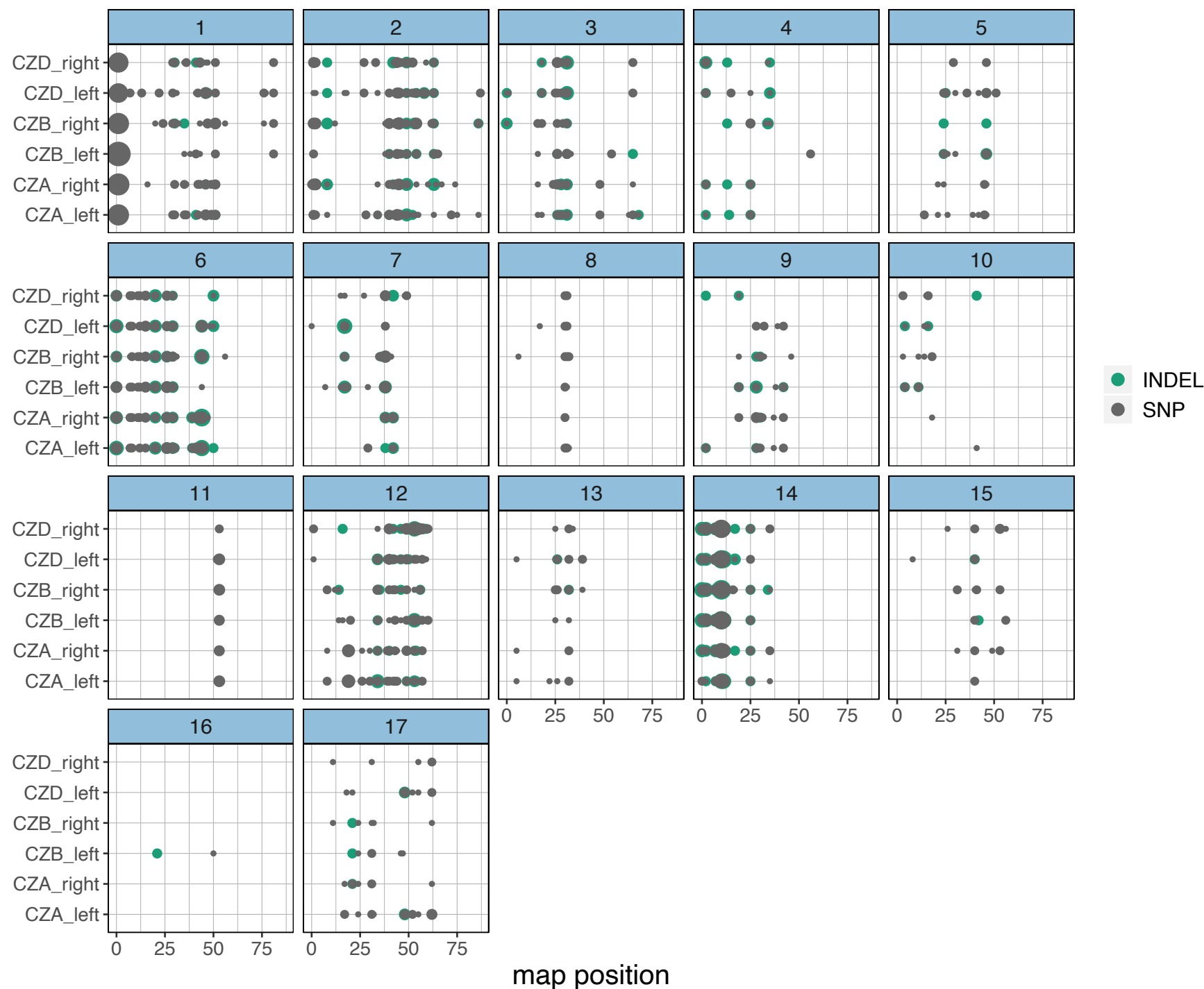
NE_Lcomp	W_Lcomp	INDEL	SNP
alt_anc	alt_anc	5305	27188
alt_anc	het	528	3543
alt_anc	NA	245	1097
alt_anc	ref_anc	511	2195
het	alt_anc	2231	12439
het	het	1577	9691
het	NA	151	627
het	ref_anc	3120	17198
NA	alt_anc	158	765
NA	het	33	267
NA	ref_anc	449	1831
ref_anc	alt_anc	693	3292
ref_anc	het	1422	7462
ref_anc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

- Fixed + polymorphic

Island	Ecotype	N	INDELs	SNPs	TOT	Island	Ecotype	N-1	INDELs	SNPs	TOT
CZA	CRAB	69	7977	38304	46281	CZA	CRAB	68	9180	43755	52935
CZA	WAVE L	62	7836	37843	45679	CZA	WAVE L	61	9114	43474	52588
CZA	WAVE R	23	9352	44548	53900	CZA	WAVE R	22	10859	51438	62297
CZB	CRAB	64	9751	46233	55984	CZB	CRAB	63	11021	52120	63141
CZB	WAVE L	59	9394	44480	53874	CZB	WAVE L	58	10969	51690	62659
CZB	WAVE R	45	10311	48558	58869	CZB	WAVE R	44	11611	54820	66431
CZD	CRAB	67	9764	46659	56423	CZD	CRAB	66	11066	52425	63491
CZD	WAVE L	31	10864	51446	62310	CZD	WAVE L	30	12327	58256	70583
CZD	WAVE R	73	9732	46449	56181	CZD	WAVE R	72	10993	52283	63276

3. Outlier sharing

- Variants after cline analysis (maf filter 0.1)



INDEL SNP

