CRISPR-cas_Systems_analysis

Thomas Nicholson

7/14/2017

Table 1: Genome analysis summary

	Counts
Number of genomes	32331
Number of genomes with cas genes	20498
Number of genomes with at least one identifiable subtype	6567
Number of genomes with a single system	5395
Number of contigs with arrays	5965
Number of arrays	8015
Number of spacers	140579
Number of genomes (protospacer results)	5289
Number of arrays (protospacer results)	7574
Number of spacers (protospacer results)	71190
Number of target genomes (protospacer results)	25191
Number of protospacers (protospacer results)	655054
Number of host/target pairs (protospacer results)	640936

Table 2: Subtypes Summary

Subtype	# genomes	# arrays	# spacers	arrays/genome	spacers/genome	spacers/array
I-A	73.00	151.00	3435.00	2.07	47.05	22.75
I-B	528.00	1787.00	36667.00	3.38	69.45	20.52
I-C	842.00	1692.00	29069.00	2.01	34.52	17.18
I-D	74.00	266.00	5957.00	3.59	80.50	22.39
I-E	1301.00	1783.00	26101.00	1.37	20.06	14.64
I-F	732.00	264.00	6825.00	0.36	9.32	25.85
I-U	151.00	285.00	4722.00	1.89	31.27	16.57
II-A	409.00	428.00	6279.00	1.05	15.35	14.67
II-B	122.00	127.00	2161.00	1.04	17.71	17.02
II-C	839.00	666.00	9824.00	0.79	11.71	14.75
III-A	148.00	327.00	5370.00	2.21	36.28	16.42
III-B	1.00	3.00	174.00	3.00	174.00	58.00
III-C	2.00	8.00	328.00	4.00	164.00	41.00
III-D	81.00	158.00	2679.00	1.95	33.07	16.96
IV	64.00	34.00	377.00	0.53	5.89	11.09
V	28.00	36.00	611.00	1.29	21.82	16.97

Also found 12 phage genomes that contained CRISPR arrays using CRISPRDetect 2.2 on the phage.fa file updated on 22/06/2017

```
plot.adjusted.distr(sd.from.mean.to.keep= 0.5, Subtype.label = "I-B", sd.from.mean.to.keep = 2)


##import data
targets.dat <-  read.table("~/Desktop/Project/CRISPRClustering/refseq_79.swipe.one_target_genome.no.arra
                           comment.char = "", fill = T, sep = "\t", header = T)
targets.dat <- transform.df(targets.dat)
```

Table 3: Subtypes Summary (Protospacer results)

Subtype	# genomes	# arrays	# spacers	# target genomes	# protospacers	# host/target pairs
I-A	71	144	1876	7444	14019	12876
I-B	1103	1733	21780	21178	206150	203889
I-C	1058	1598	13504	19497	94916	98340
I-D	169	258	2859	10954	20414	19682
I-E	1193	1685	13215	17764	158769	148710
I-F	197	249	3274	10722	25587	26745
I-U	185	264	2586	8613	21490	24421
II-A	346	405	2717	7457	21602	21501
II-B	93	113	795	4378	5645	5333
II-C	529	588	3584	11012	25851	27480
III-A	194	310	3040	12201	32486	35204
III-B	1	3	112	750	857	750
III-C	3	8	165	1031	1086	1058
III-D	95	149	1247	6728	10319	9933
IV	30	33	232	1691	2066	1986
V	22	34	202	1958	2414	3028

```
targets.dat <- targets.dat%>%mutate(strand.plus.direction = paste(target.strand, five.three.prime.dir, s
targets.dat <- targets.dat%>%mutate(legend.labels = ifelse(strand.plus.direction == "n_3",
                                                           "Non-target 3' direction",
                                                           ifelse(strand.plus.direction == "n_5",
                                                                  "Non-target 5' direction",
                                                                  ifelse(strand.plus.direction == "t_3"
                                                                         "Target 3' direction",
                                                                         "Target 5' direction"))))



###bars
subtypes <- c("I-A", "I-B", "I-C", "I-D", "I-E", "I-F", "II-A", "II-C", "III-A", "III-D")
rh <- random.hits.generate(set.seed.val = F)
for(i in subtypes){
  plot.adjusted.distr(sd.from.mean.to.keep= 0.5, Subtype.label = "I-A", sd.from.mean.to.keep = 2)
}
```

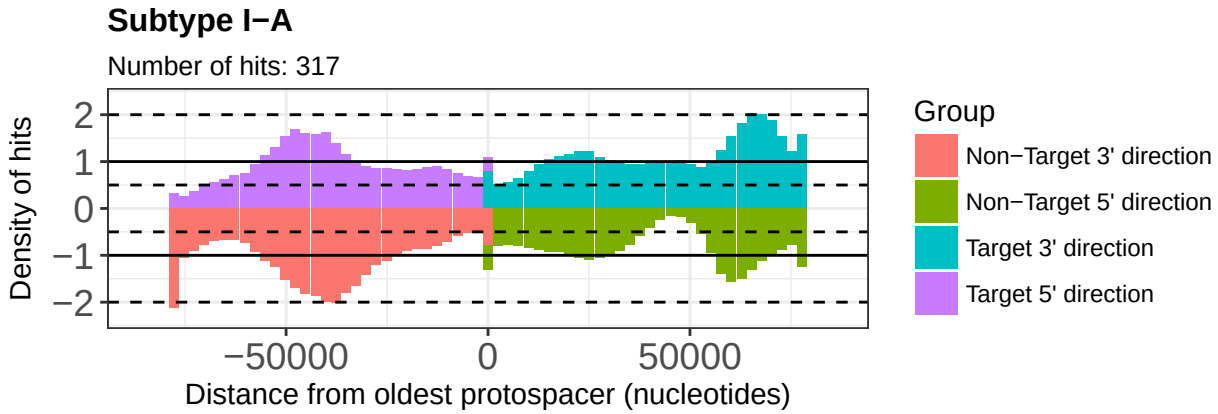
The random plots that are produced underneath the actual data take 100 iterations of the random data and then plot each of these iterations on the same plot. The result should be the expected distribution for a naive system, with a similar number of hits in terms of the height of the bars.

```
##import data
targets.dat <-  read.table("~/Desktop/Project/CRISPRClustering/refseq_79.swipe.one_target_genome.no.arra
targets.dat <- transform.df(targets.dat)

targets.dat <- targets.dat%>%mutate(strand.plus.direction = paste(target.strand, five.three.prime.dir, s
targets.dat <- targets.dat%>%mutate(legend.labels = ifelse(strand.plus.direction == "n_3", "Non-target 3

subtypes <- c("I-A", "I-B", "I-C", "I-D",
              "I-E", "I-F", "II-A", "II-B",
              "III-A", "III-D")
i <- "I-F"
for(i in subtypes){
```

Table 4: Subtypes comparison (Protospacer results)

Subtype	arrays/genome	spacers/genome	spacers/array	protospacers/target	protospacers/spacer	targets/host
I-A	2.03	26.42	13.03	1.88	7.47	104.85
I-B	1.57	19.75	12.57	9.73	9.47	19.20
I-C	1.51	12.76	8.45	4.87	7.03	18.43
I-D	1.53	16.92	11.08	1.86	7.14	64.82
I-E	1.41	11.08	7.84	8.94	12.01	14.89
I-F	1.26	16.62	13.15	2.39	7.82	54.43
I-U	1.43	13.98	9.80	2.50	8.31	46.56
II-A	1.17	7.85	6.71	2.90	7.95	21.55
II-B	1.22	8.55	7.04	1.29	7.10	47.08
II-C	1.11	6.78	6.10	2.35	7.21	20.82
III-A	1.60	15.67	9.81	2.66	10.69	62.89
III-B	3.00	112.00	37.33	1.14	7.65	750.00
III-C	2.67	55.00	20.62	1.05	6.58	343.67
III-D	1.57	13.13	8.37	1.53	8.28	70.82
IV	1.10	7.73	7.03	1.22	8.91	56.37
V	1.55	9.18	5.94	1.23	11.95	89.00

```
##I-A
plot.adjusted.distr(sd.from.mean.to.keep = 0.5, Subtype.label = i, plot.quadrants = F)

plot.adjusted.distr(sd.from.mean.to.keep = 1, Subtype.label = i, plot.quadrants = F)
plot.adjusted.distr(sd.from.mean.to.keep = 2, Subtype.label = i, plot.quadrants = F)
}
```



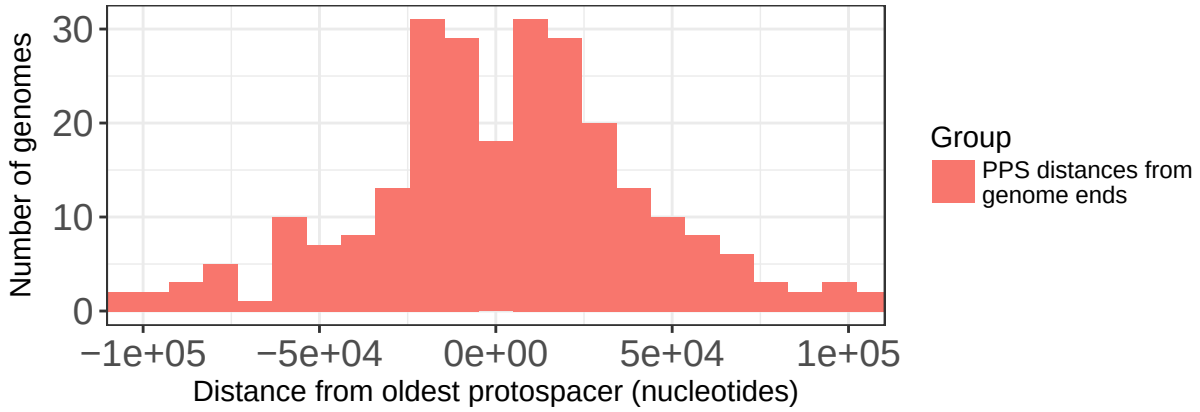
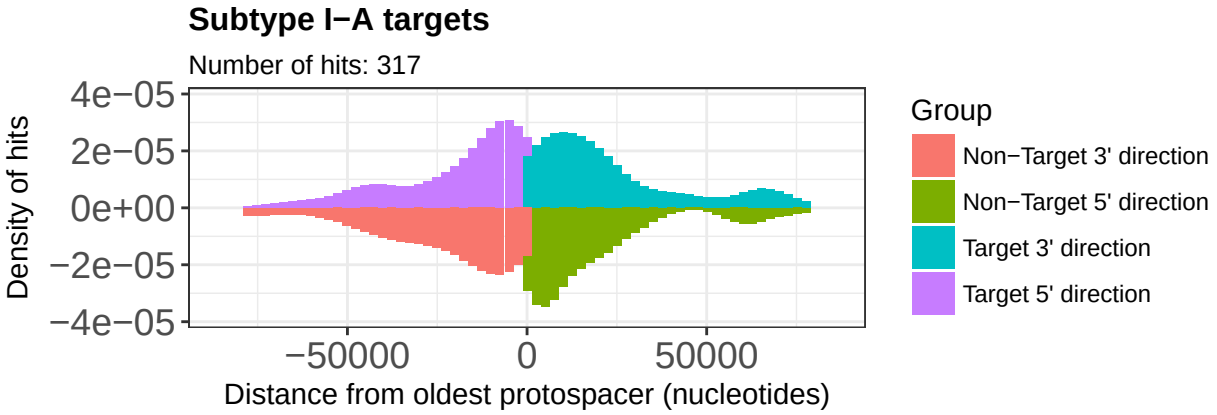


Table 5: Subtypes Summary (Results before filtering)

Subtype	# genomes	# arrays	# spacers	# target genomes	# protospacers	# host/target pairs
I-A	36	75	455	521	1267	596
I-B	561	896	5344	4576	22891	11163
I-C	353	488	2013	3026	9038	4414
I-D	67	100	380	641	1446	702
I-E	416	613	2294	4967	29480	15005
I-F	75	98	804	1636	5384	2942
I-U	67	84	487	614	1646	903
II-A	113	136	700	1319	4443	1972
II-B	21	27	109	184	401	189
II-C	130	146	561	1344	2986	1642
III-A	76	114	501	2070	4696	2723
III-B	1	3	51	65	158	65
III-C	2	6	30	26	53	26
III-D	26	42	156	316	626	340
IV	8	9	25	47	118	55
V	8	14	68	98	209	105



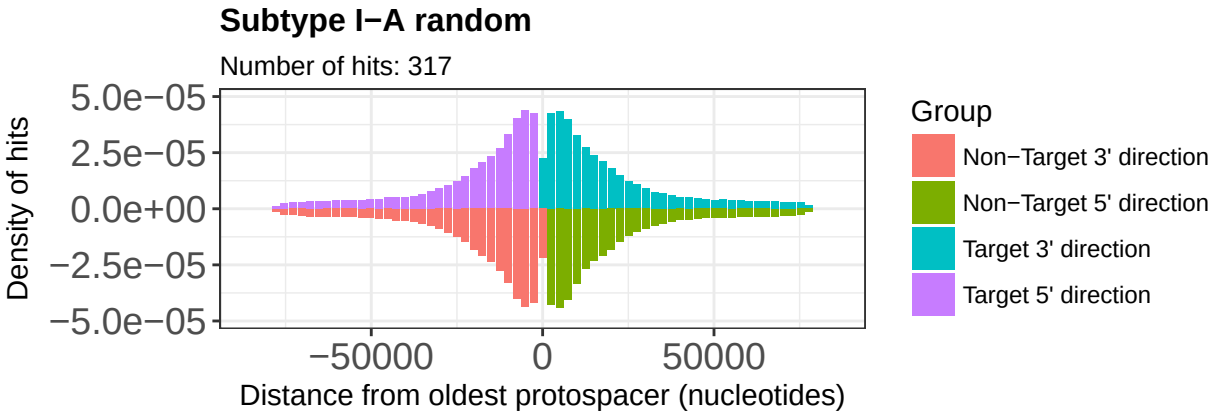


Table 6: Subtypes comparison (Results before filtering)

Subtype	arrays/genome	spacers/genome	spacers/array	protospacers/target	protospacers/spacer	targets/host
I-A	2.08	12.64	6.07	2.43	2.78	14.47
I-B	1.60	9.53	5.96	5.00	4.28	8.16
I-C	1.38	5.70	4.12	2.99	4.49	8.57
I-D	1.49	5.67	3.80	2.26	3.81	9.57
I-E	1.47	5.51	3.74	5.94	12.85	11.94
I-F	1.31	10.72	8.20	3.29	6.70	21.81
I-U	1.25	7.27	5.80	2.68	3.38	9.16
II-A	1.20	6.19	5.15	3.37	6.35	11.67
II-B	1.29	5.19	4.04	2.18	3.68	8.76
II-C	1.12	4.32	3.84	2.22	5.32	10.34
III-A	1.50	6.59	4.39	2.27	9.37	27.24
III-B	3.00	51.00	17.00	2.43	3.10	65.00
III-C	3.00	15.00	5.00	2.04	1.77	13.00
III-D	1.62	6.00	3.71	1.98	4.01	12.15
IV	1.12	3.12	2.78	2.51	4.72	5.88
V	1.75	8.50	4.86	2.13	3.07	12.25

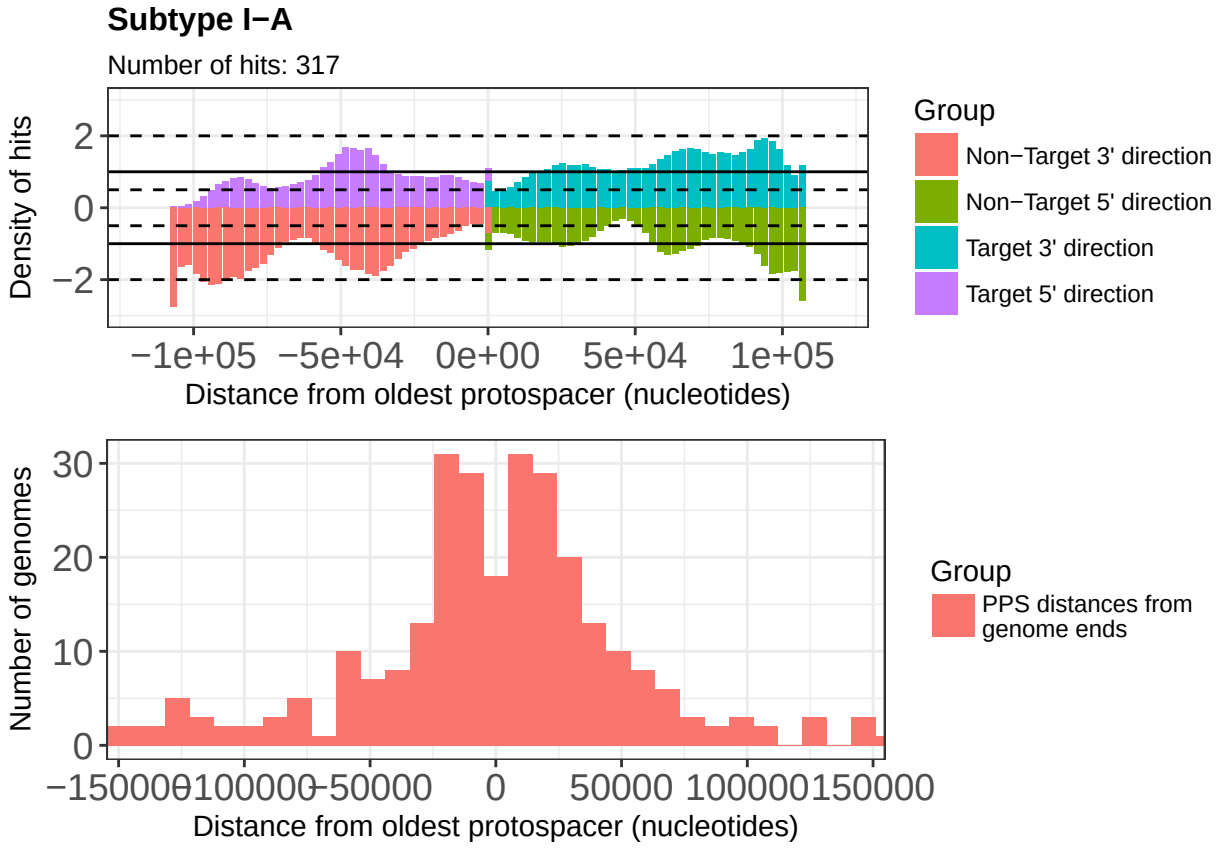
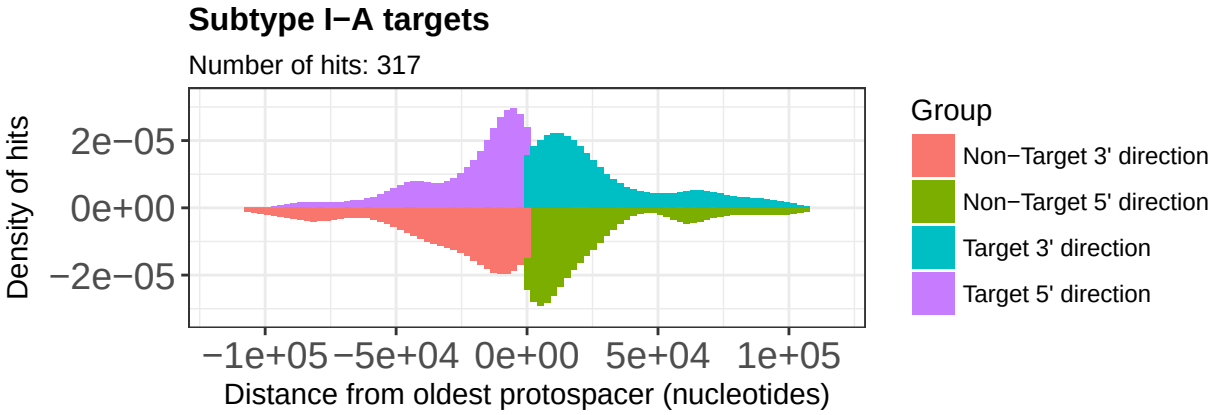


Table 7: Subtypes Summary (Filtered results)

Subtype	# genomes	# arrays	# spacers	# target genomes	# protospacers	# host/target pairs
I-A	31	64	293	138	317	155
I-B	459	722	3141	1191	3490	1682
I-C	262	347	1121	471	1198	544
I-D	48	72	244	130	272	135
I-E	302	465	1439	539	1578	747
I-F	60	77	430	168	455	189
I-U	56	68	305	151	338	167
II-A	76	87	270	96	284	117
II-B	16	21	50	23	52	24
II-C	82	94	251	110	264	126
III-A	56	85	299	152	334	158
III-B	1	3	32	16	34	16
III-C	2	6	25	13	27	13
III-D	23	37	115	58	127	60
IV	4	5	11	5	11	5
V	6	12	36	17	39	18



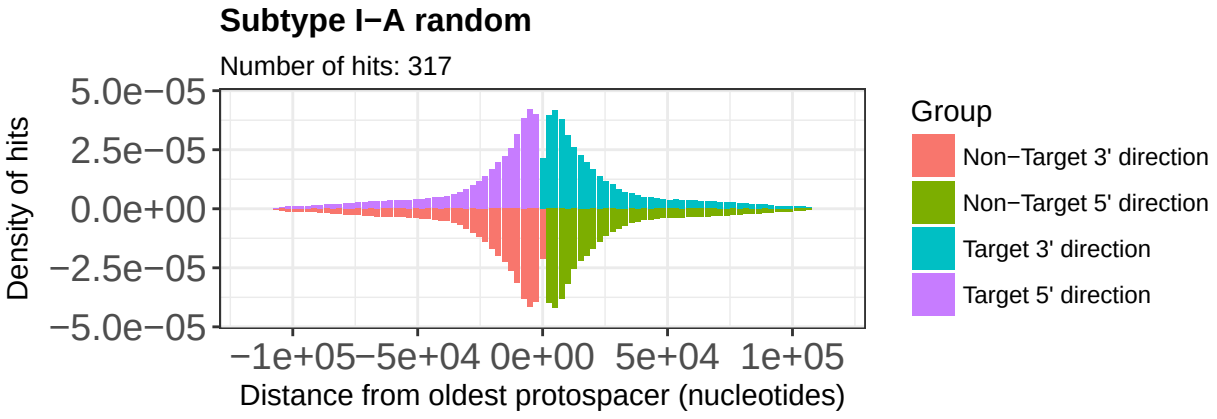
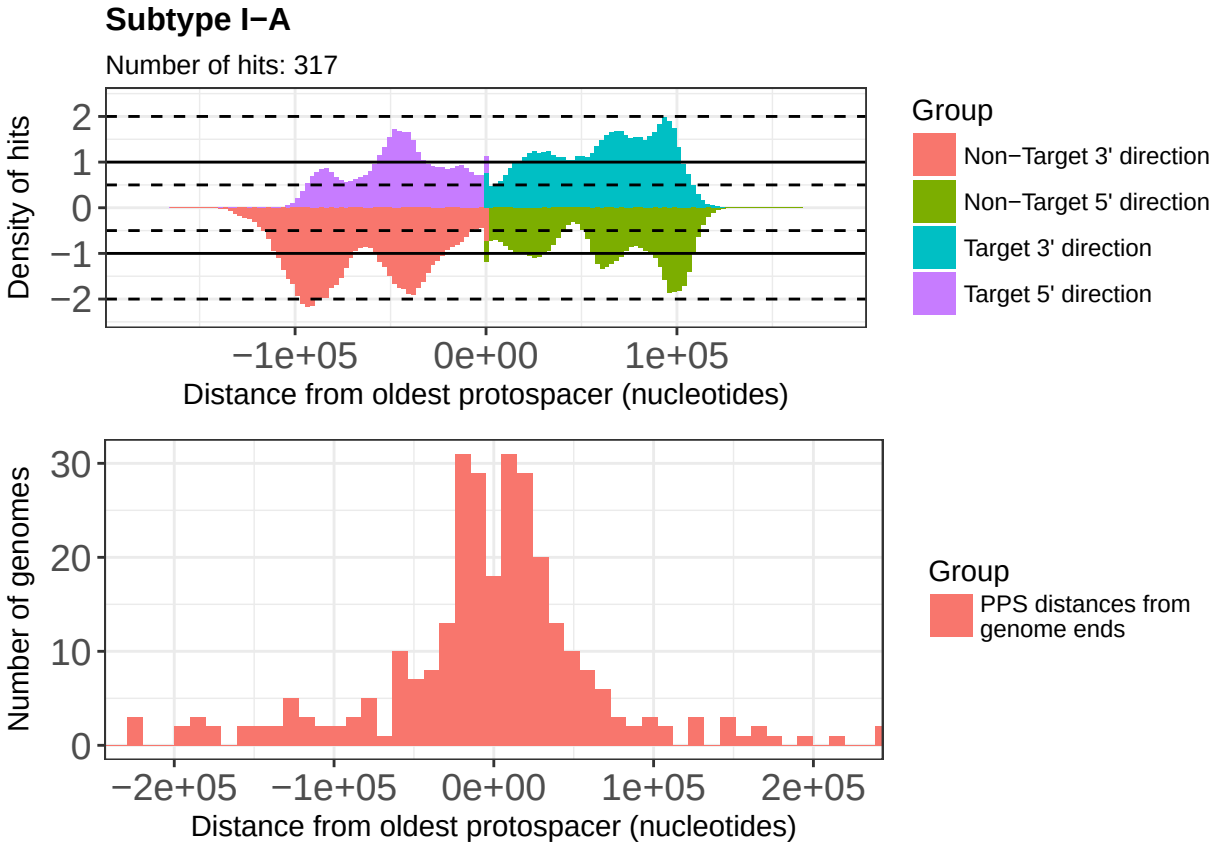
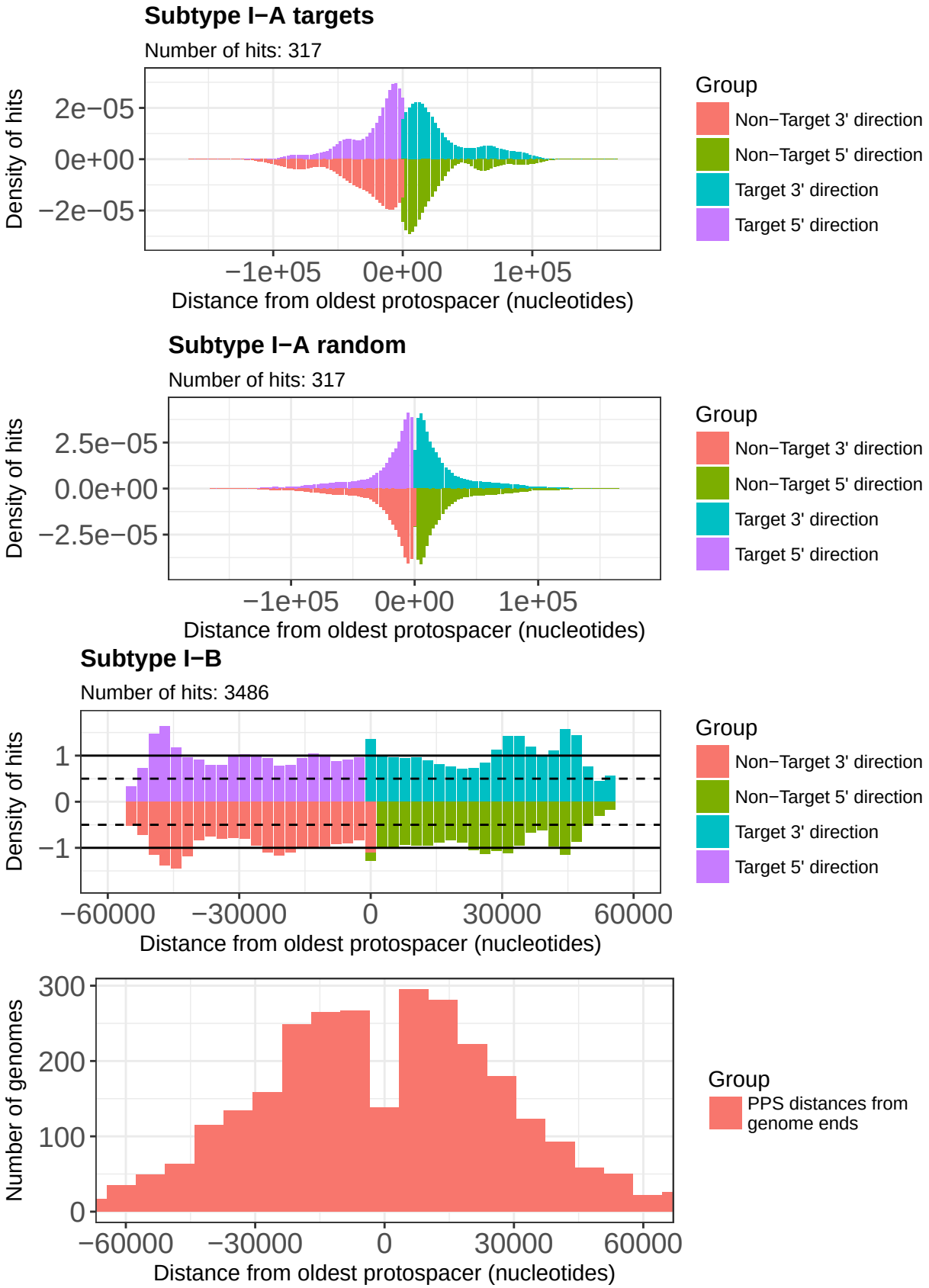
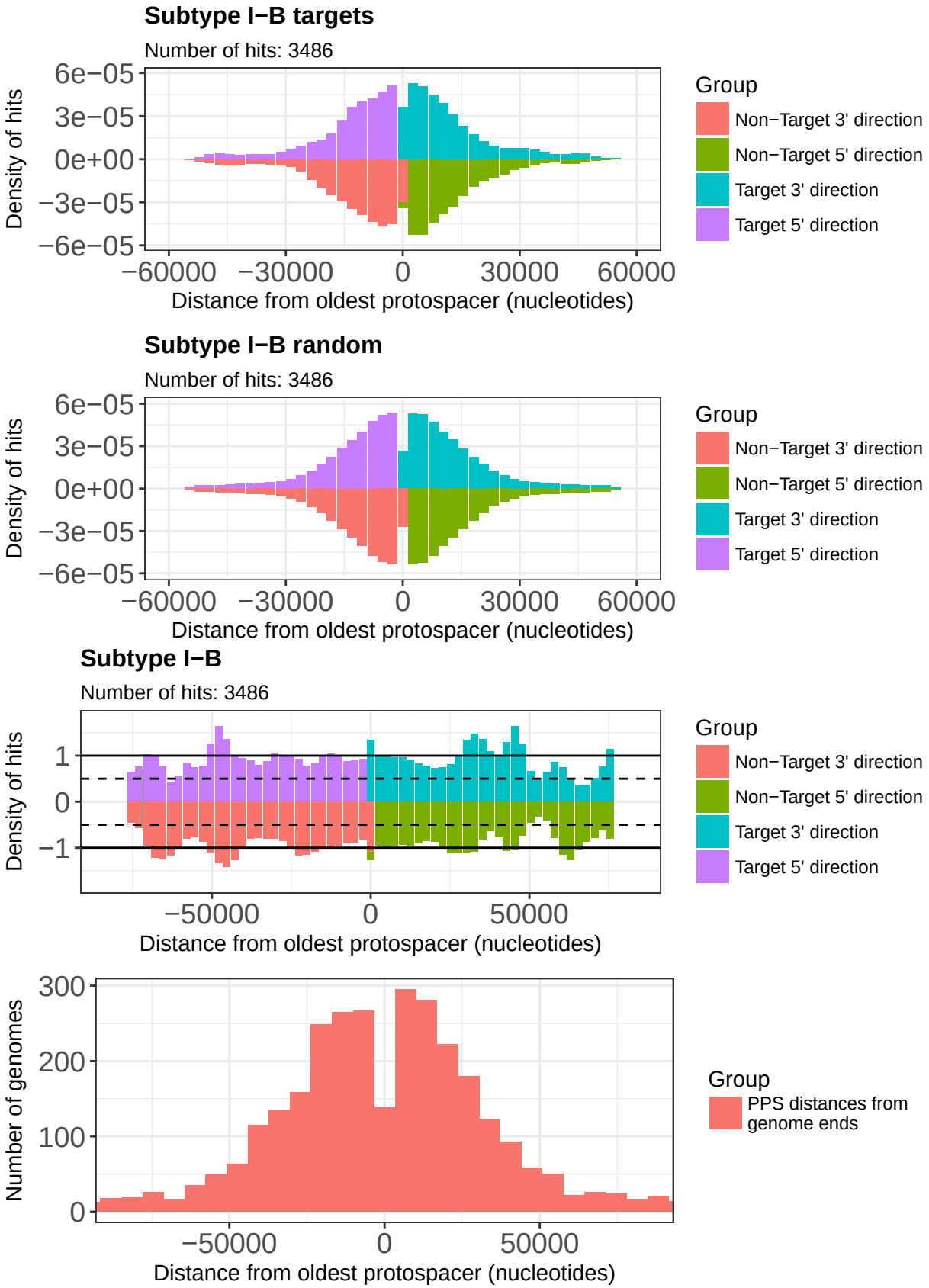


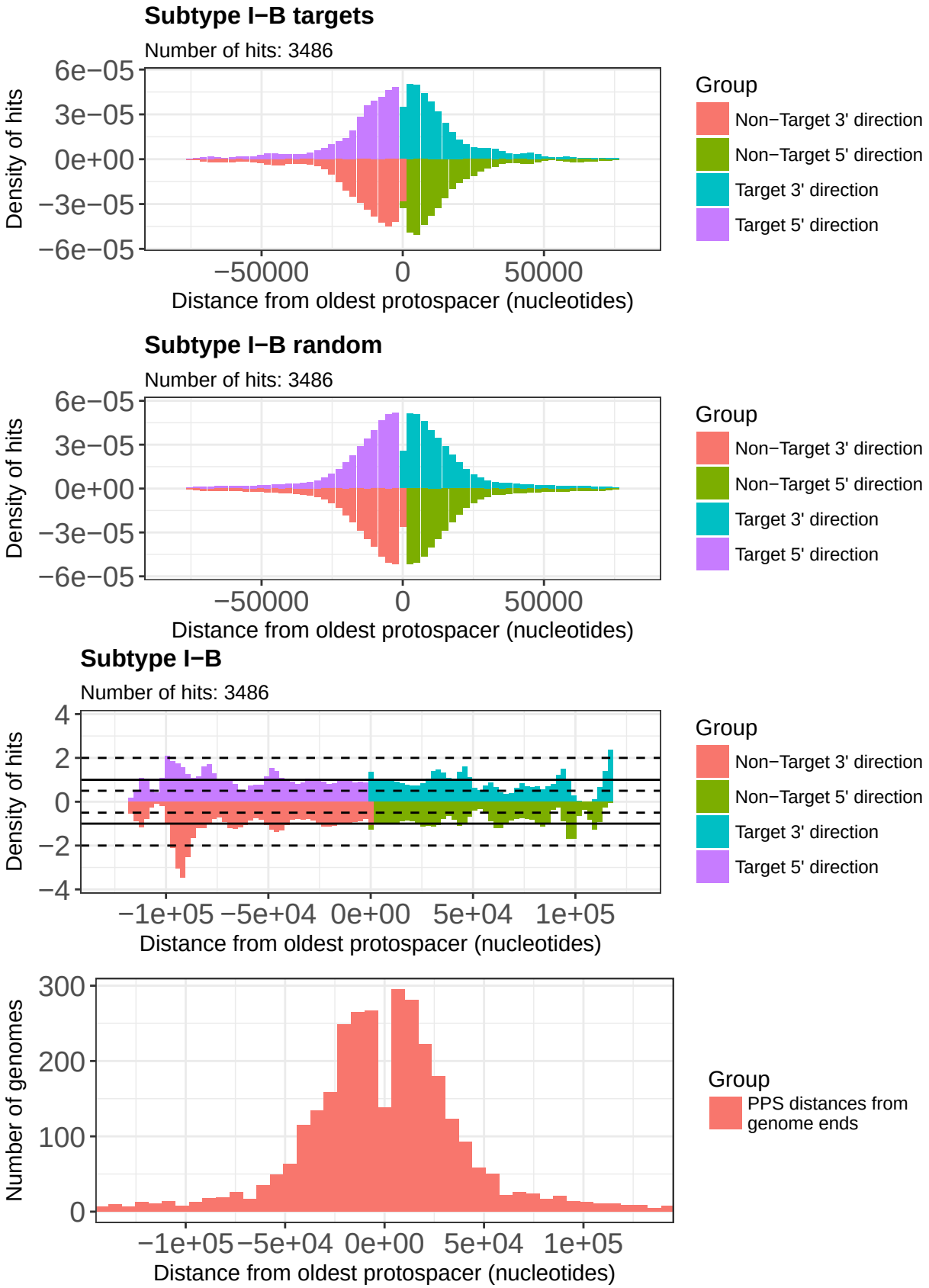
Table 8: Subtypes comparison (Filtered results)

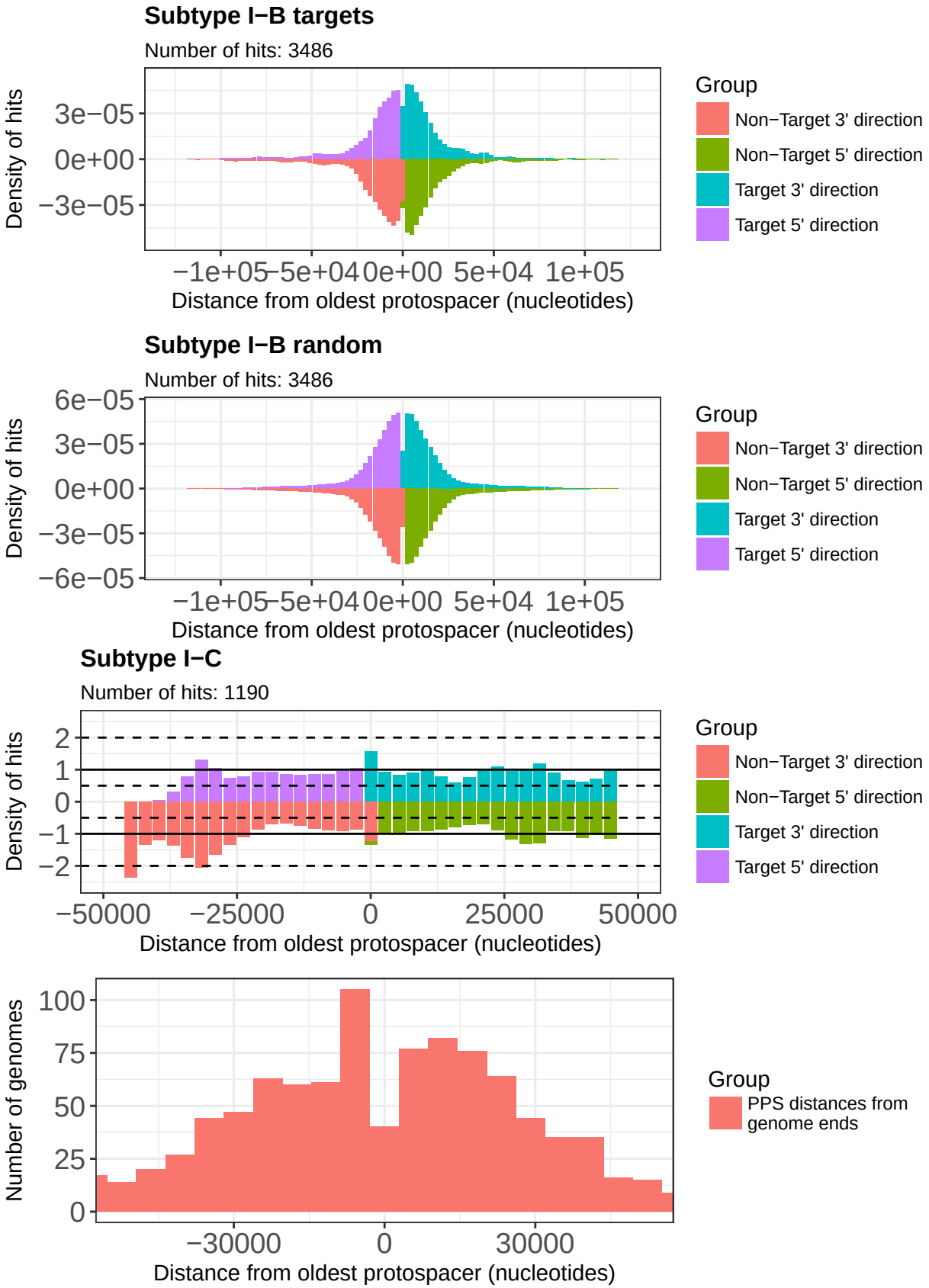
Subtype	arrays/genome	spacers/genome	spacers/array	protospacers/target	protospacers/spacer	targets/host
I-A	2.06	9.45	4.58	2.30	1.08	4.45
I-B	1.57	6.84	4.35	2.93	1.11	2.59
I-C	1.32	4.28	3.23	2.54	1.07	1.80
I-D	1.50	5.08	3.39	2.09	1.11	2.71
I-E	1.54	4.76	3.09	2.93	1.10	1.78
I-F	1.28	7.17	5.58	2.71	1.06	2.80
I-U	1.21	5.45	4.49	2.24	1.11	2.70
II-A	1.14	3.55	3.10	2.96	1.05	1.26
II-B	1.31	3.12	2.38	2.26	1.04	1.44
II-C	1.15	3.06	2.67	2.40	1.05	1.34
III-A	1.52	5.34	3.52	2.20	1.12	2.71
III-B	3.00	32.00	10.67	2.12	1.06	16.00
III-C	3.00	12.50	4.17	2.08	1.08	6.50
III-D	1.61	5.00	3.11	2.19	1.10	2.52
IV	1.25	2.75	2.20	2.20	1.00	1.25
V	2.00	6.00	3.00	2.29	1.08	2.83

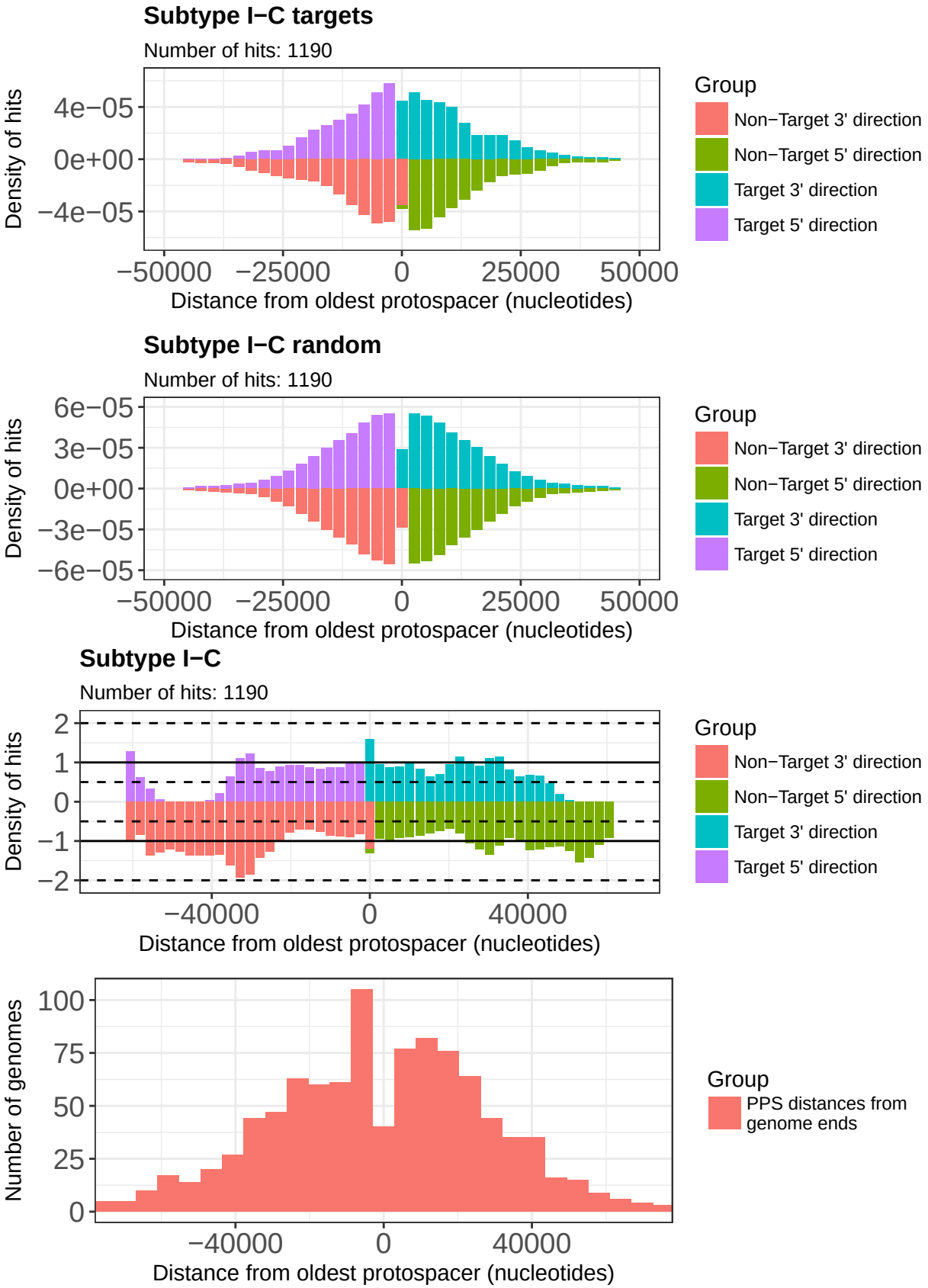


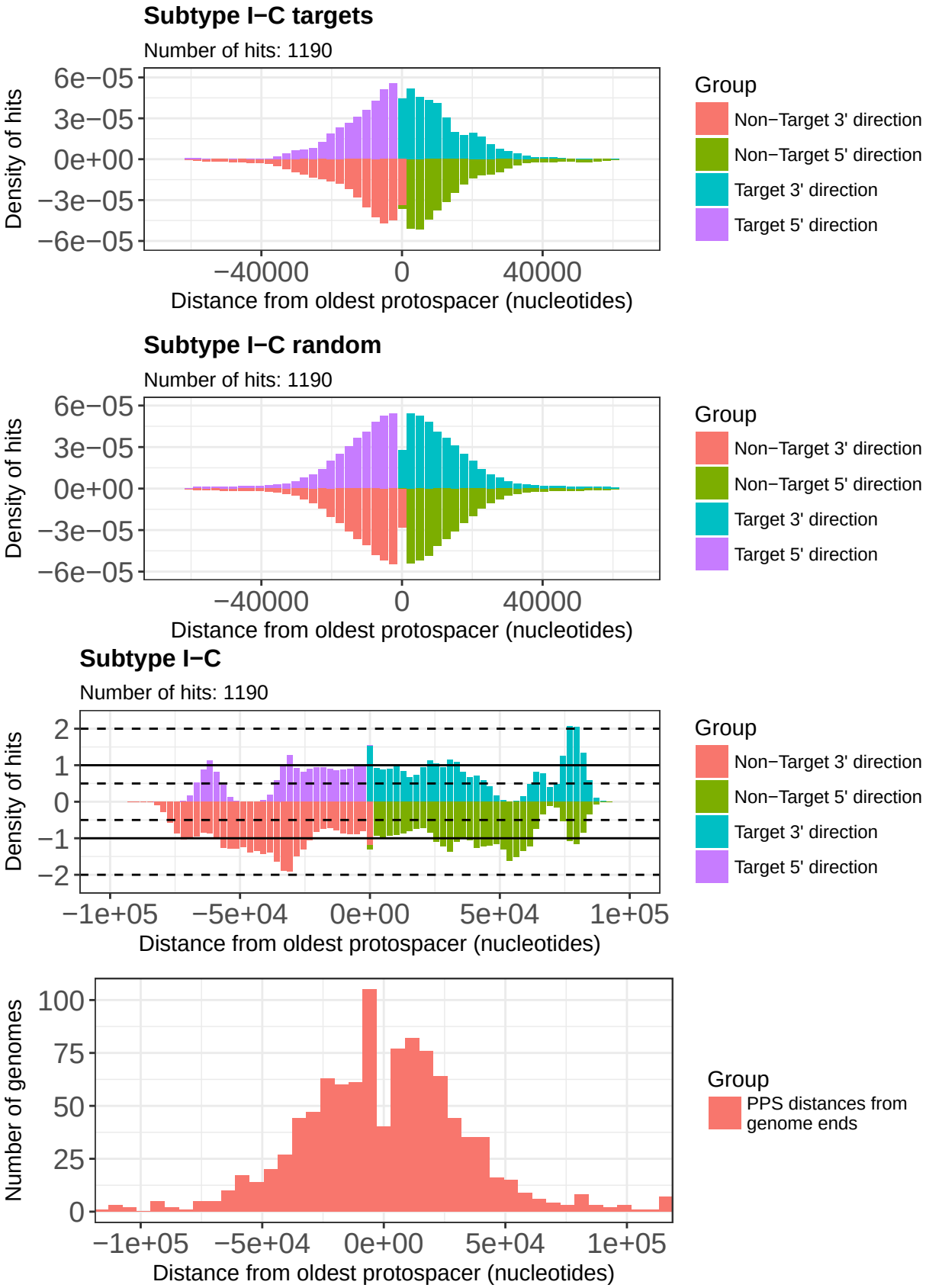

Subtype I−A targets
Number of hits: 317
Density of hits
2e−05
0e+00
−2e−05
−1e+05
0e+00
1e+05
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Subtype I−A random
Number of hits: 317
Density of hits
2.5e−05
0.0e+00
−2.5e−05
−1e+05
0e+00
1e+05
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Subtype I−B
Number of hits: 3486
Density of hits
1
0
−1
−60000
−30000
0
30000
60000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Number of genomes
300
200
100
0
−60000
−30000
0
30000
60000
Distance from oldest protospacer (nucleotides)
Group
PPS distances from genome ends

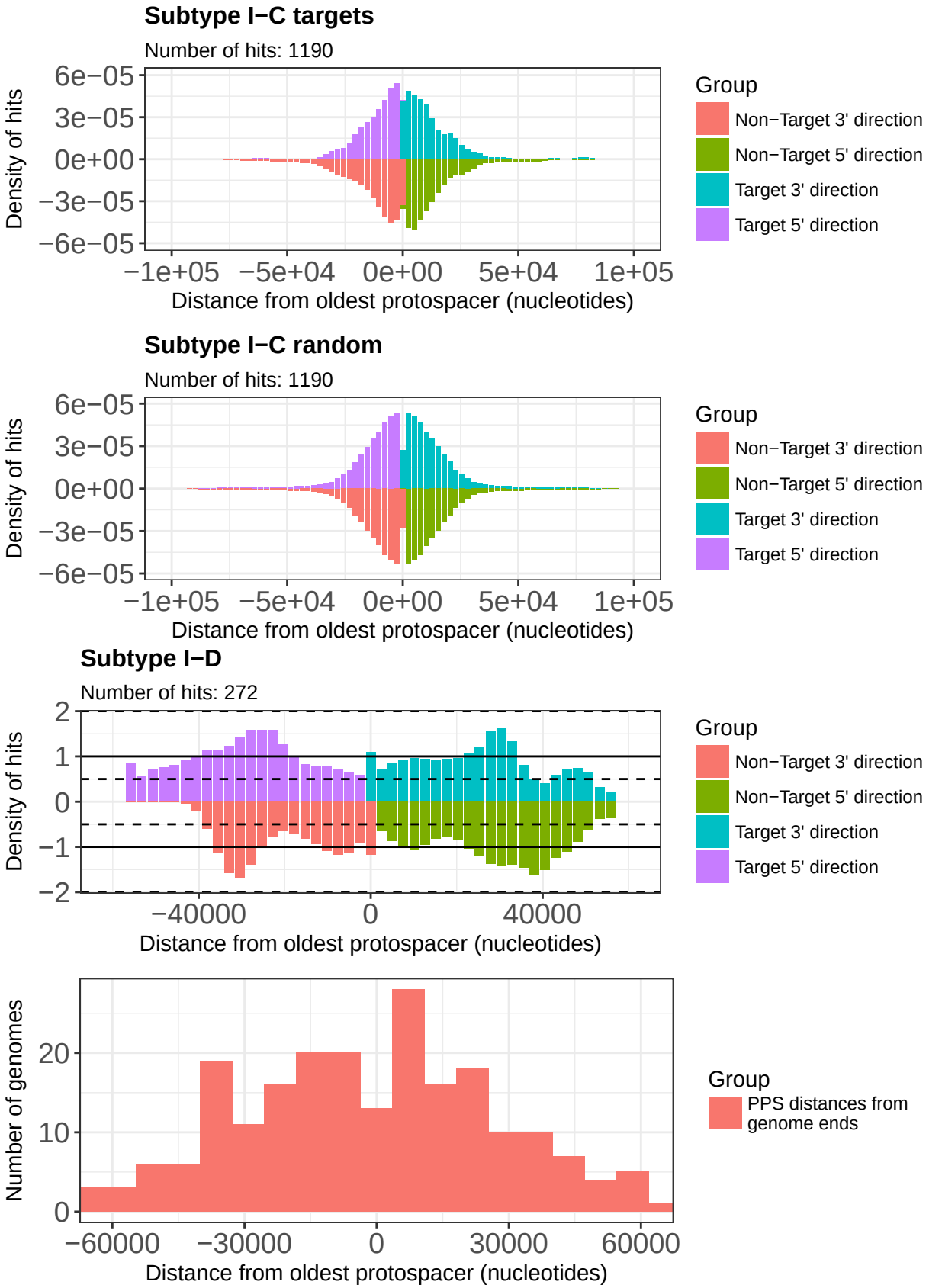

Subtype I−B targets
Number of hits: 3486
Density of hits
6e−05
3e−05
0e+00
−3e−05
−6e−05
−60000
−30000
0
30000
60000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Subtype I−B random
Number of hits: 3486
Density of hits
6e−05
3e−05
0e+00
−3e−05
−6e−05
−60000
−30000
0
30000
60000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Subtype I−B
Number of hits: 3486
Density of hits
1
0
−1
−50000
0
50000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Number of genomes
300
200
100
0
−50000
0
50000
Distance from oldest protospacer (nucleotides)
Group
PPS distances from genome ends

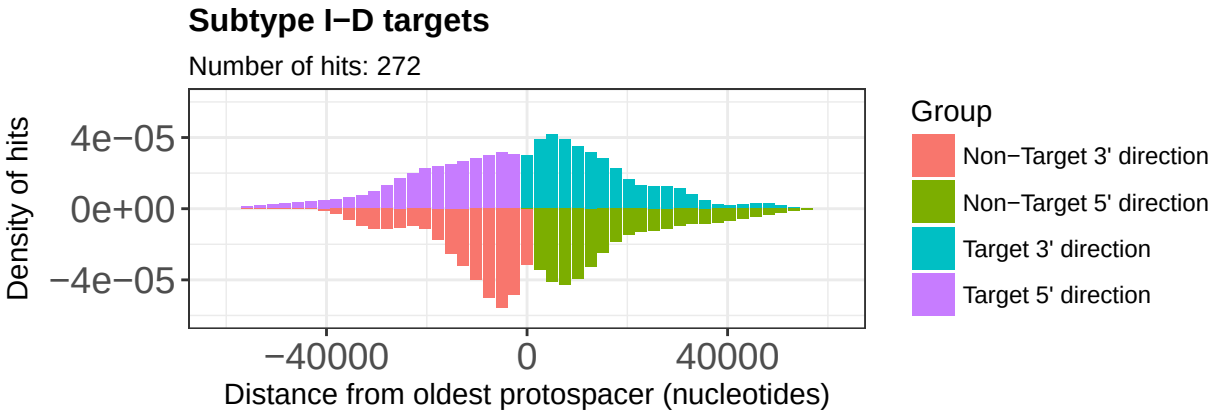

Subtype I−B targets
Number of hits: 3486
Density of hits
6e−05
3e−05
0e+00
−3e−05
−6e−05
−50000
0
50000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Subtype I−B random
Number of hits: 3486
Density of hits
6e−05
3e−05
0e+00
−3e−05
−6e−05
−50000
0
50000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Subtype I−B
Number of hits: 3486
Density of hits
4
2
0
−2
−4
−1e+05
−5e+04
0e+00
5e+04
1e+05
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Number of genomes
300
200
100
0
−1e+05
−5e+04
0e+00
5e+04
1e+05
Distance from oldest protospacer (nucleotides)
Group
PPS distances from genome ends

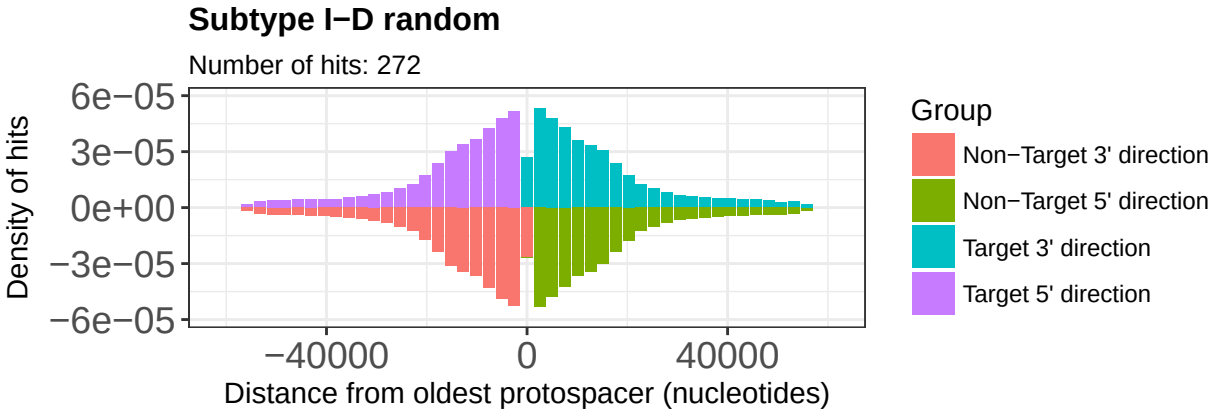

Subtype I−B targets
Number of hits: 3486
Density of hits
3e−05
0e+00
−3e−05
−1e+05
−5e+04
0e+00
5e+04
1e+05
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Subtype I−B random
Number of hits: 3486
Density of hits
6e−05
3e−05
0e+00
−3e−05
−6e−05
−1e+05
−5e+04
0e+00
5e+04
1e+05
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Subtype I−C
Number of hits: 1190
Density of hits
2
1
0
−1
−2
−50000
−25000
0
25000
50000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Number of genomes
100
75
50
25
0
−30000
0
30000
Distance from oldest protospacer (nucleotides)
Group
PPS distances from genome ends

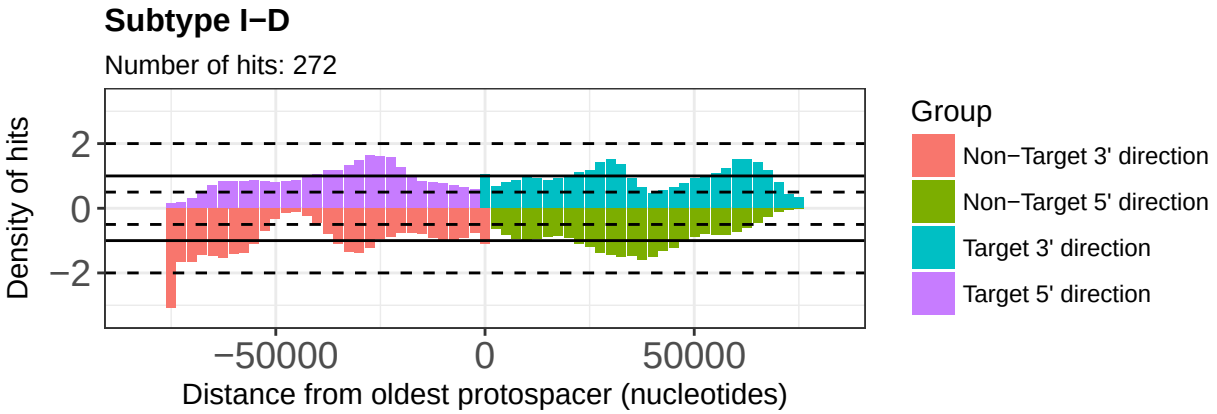

Subtype I−C targets
Number of hits: 1190
Density of hits
4e−05
0e+00
−4e−05
−50000
−25000
0
25000
50000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Subtype I−C random
Number of hits: 1190
Density of hits
6e−05
3e−05
0e+00
−3e−05
−6e−05
−50000
−25000
0
25000
50000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Subtype I−C
Number of hits: 1190
Density of hits
2
1
0
−1
−2
−40000
0
40000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Number of genomes
100
75
50
25
0
−40000
0
40000
Distance from oldest protospacer (nucleotides)
Group
PPS distances from genome ends

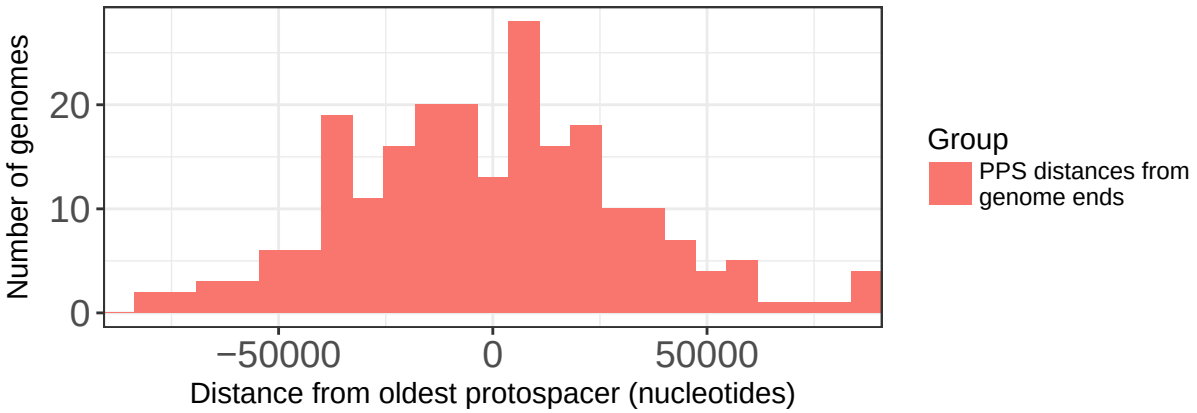

Subtype I−C targets
Number of hits: 1190
Density of hits
6e−05
3e−05
0e+00
−3e−05
−6e−05
−40000
0
40000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Subtype I−C random
Number of hits: 1190
Density of hits
6e−05
3e−05
0e+00
−3e−05
−6e−05
−40000
0
40000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Subtype I−C
Number of hits: 1190
Density of hits
2
1
0
−1
−2
−1e+05
−5e+04
0e+00
5e+04
1e+05
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Number of genomes
100
75
50
25
0
−1e+05
−5e+04
0e+00
5e+04
1e+05
Distance from oldest protospacer (nucleotides)
Group
PPS distances from genome ends

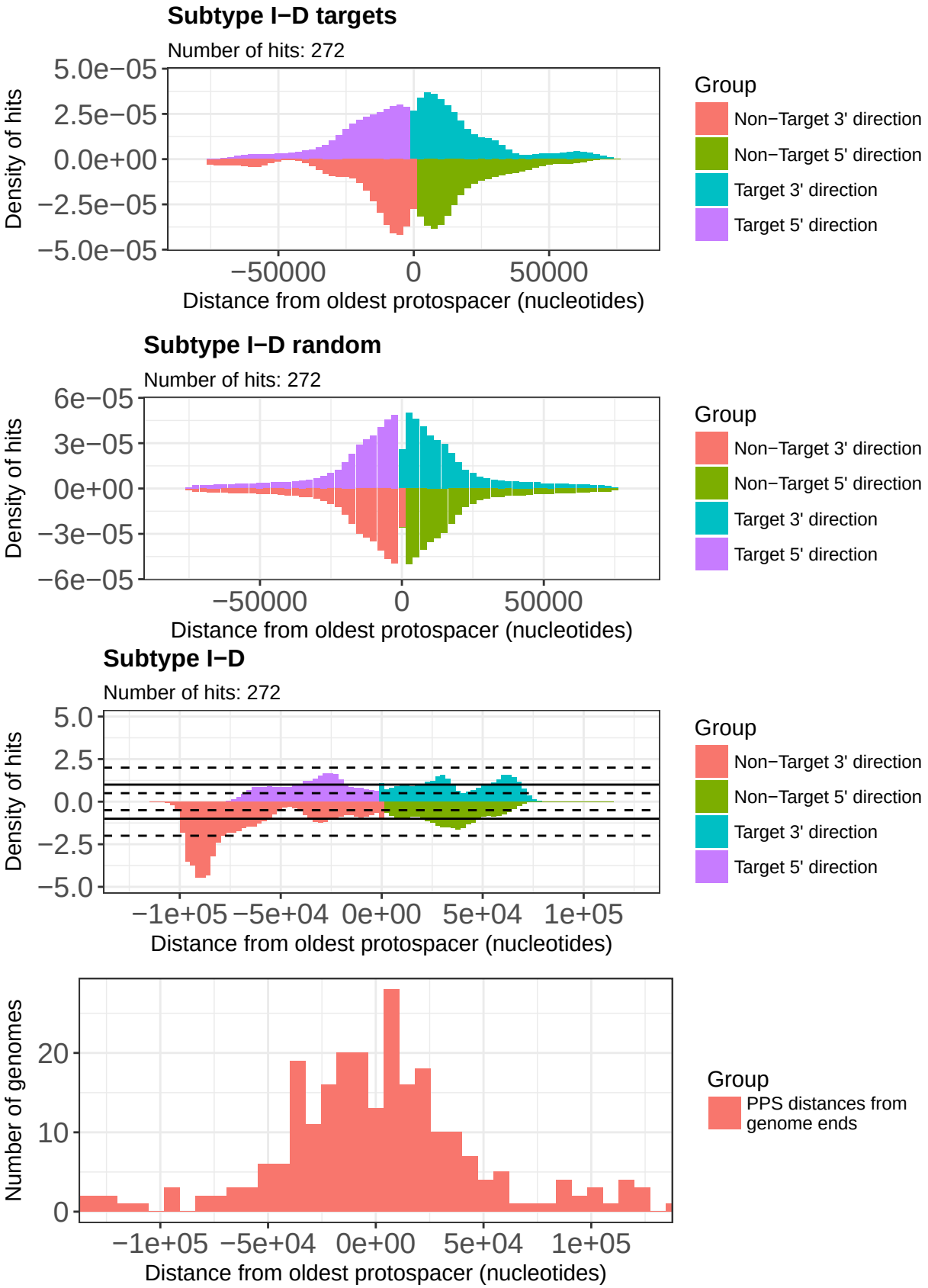

Subtype I−C targets
Number of hits: 1190
Density of hits
6e−05
3e−05
0e+00
−3e−05
−6e−05
−1e+05
−5e+04
0e+00
5e+04
1e+05
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Subtype I−C random
Number of hits: 1190
Density of hits
6e−05
3e−05
0e+00
−3e−05
−6e−05
−1e+05
−5e+04
0e+00
5e+04
1e+05
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Subtype I−D
Number of hits: 272
Density of hits
2
1
0
−1
−2
−40000
0
40000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Number of genomes
20
10
0
−60000
−30000
0
30000
60000
Distance from oldest protospacer (nucleotides)
Group
PPS distances from genome ends

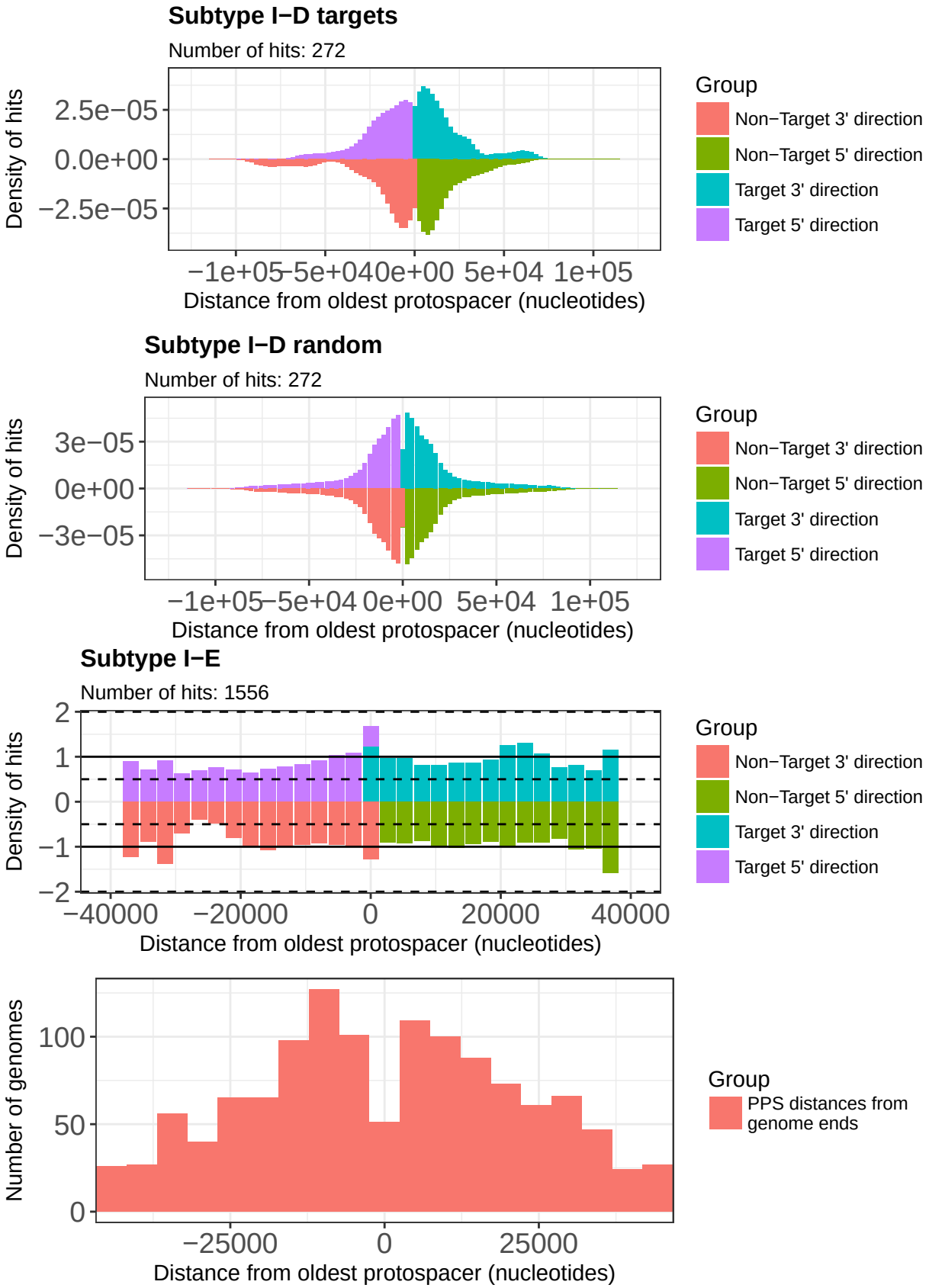

Subtype I−D targets
Number of hits: 272
Density of hits
4e−05
0e+00
−4e−05
−40000
0
40000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction

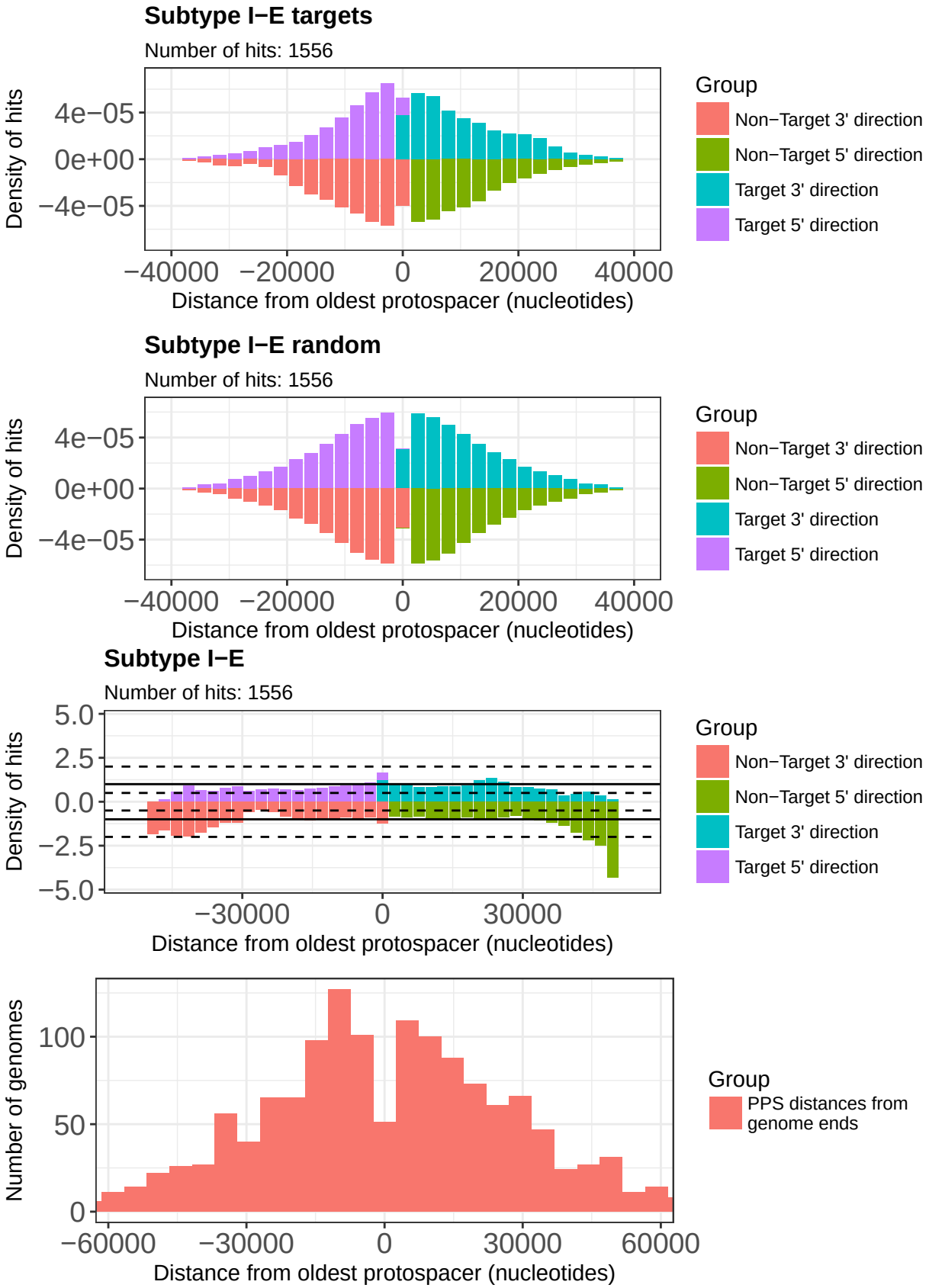

Subtype I−D random
Number of hits: 272
Density of hits
6e−05
3e−05
0e+00
−3e−05
−6e−05
−40000
0
40000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction

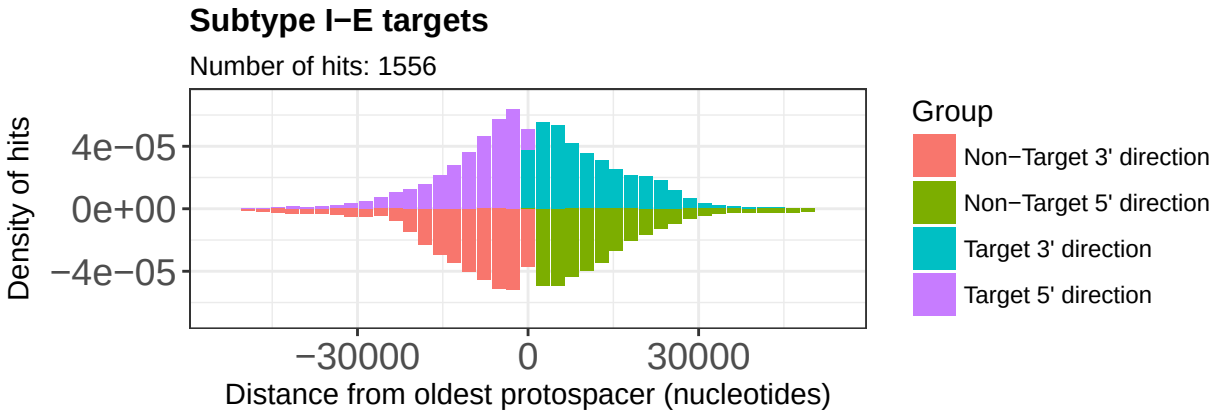

Subtype I−D
Number of hits: 272
Density of hits
2
0
−2
−50000
0
50000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction

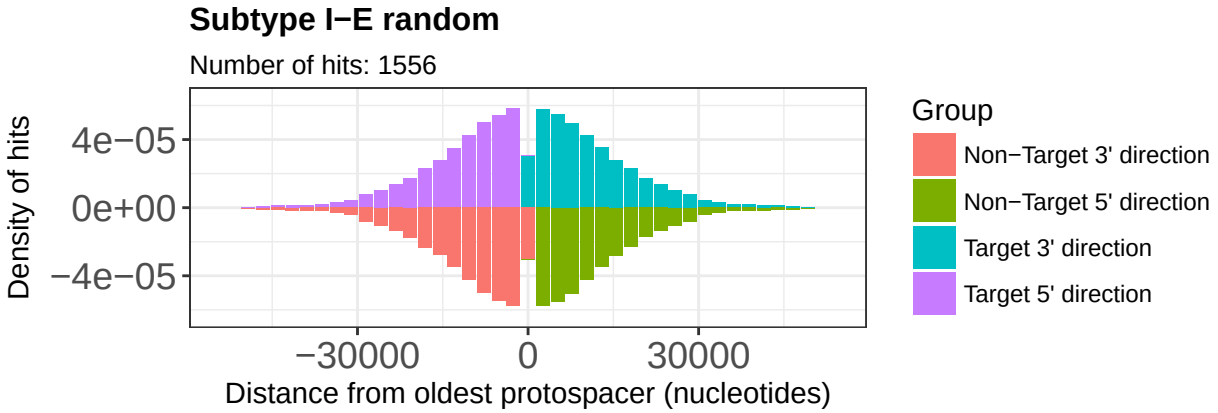

Number of genomes
20
10
0
−50000
0
50000
Distance from oldest protospacer (nucleotides)
Group
PPS distances from genome ends

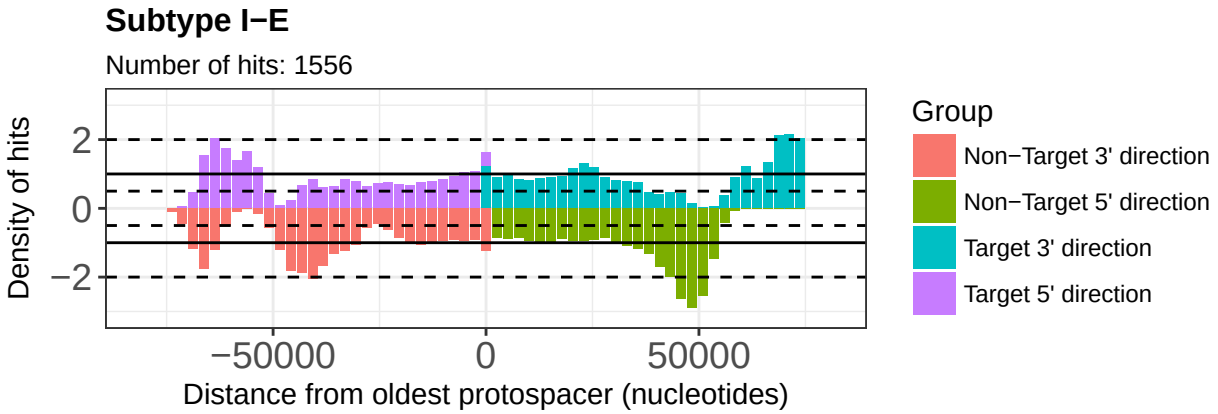

Subtype I−D targets
Number of hits: 272
Density of hits
5.0e−05
2.5e−05
0.0e+00
−2.5e−05
−5.0e−05
−50000
0
50000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Subtype I−D random
Number of hits: 272
Density of hits
6e−05
3e−05
0e+00
−3e−05
−6e−05
−50000
0
50000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Subtype I−D
Number of hits: 272
Density of hits
5.0
2.5
0.0
−2.5
−5.0
−1e+05
−5e+04
0e+00
5e+04
1e+05
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Number of genomes
20
10
0
−1e+05
−5e+04
0e+00
5e+04
1e+05
Distance from oldest protospacer (nucleotides)
Group
PPS distances from genome ends

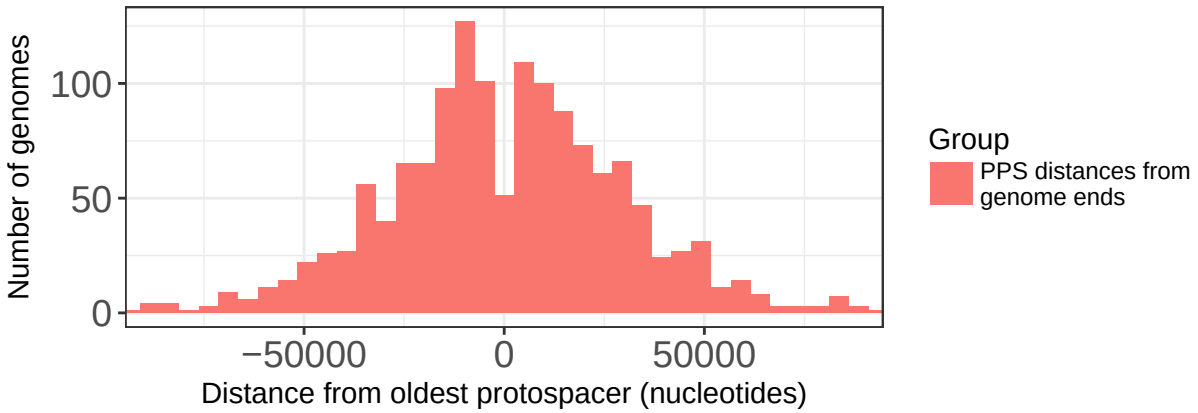

Subtype I−D targets
Number of hits: 272
Density of hits
2.5e−05
0.0e+00
−2.5e−05
−1e+05
−5e+04
0e+00
5e+04
1e+05
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Subtype I−D random
Number of hits: 272
Density of hits
3e−05
0e+00
−3e−05
−1e+05
−5e+04
0e+00
5e+04
1e+05
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Subtype I−E
Number of hits: 1556
Density of hits
2
1
0
−1
−2
−40000
−20000
0
20000
40000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Number of genomes
100
50
0
−25000
0
25000
Distance from oldest protospacer (nucleotides)
Group
PPS distances from genome ends

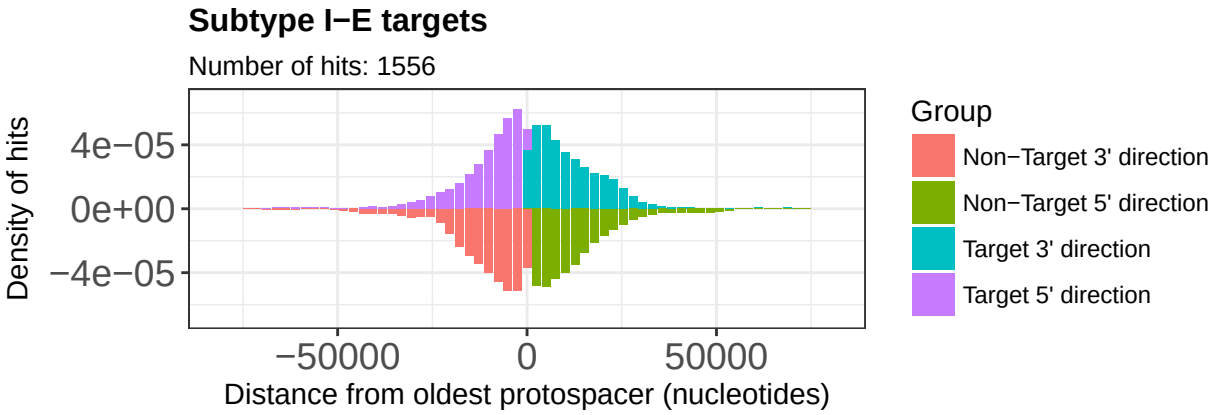

Subtype I−E targets
Number of hits: 1556
Density of hits
4e−05
0e+00
−4e−05
−40000
−20000
0
20000
40000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Subtype I−E random
Number of hits: 1556
Density of hits
4e−05
0e+00
−4e−05
−40000
−20000
0
20000
40000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Subtype I−E
Number of hits: 1556
Density of hits
5.0
2.5
0.0
−2.5
−5.0
−30000
0
30000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Number of genomes
100
50
0
−60000
−30000
0
30000
60000
Distance from oldest protospacer (nucleotides)
Group
PPS distances from genome ends

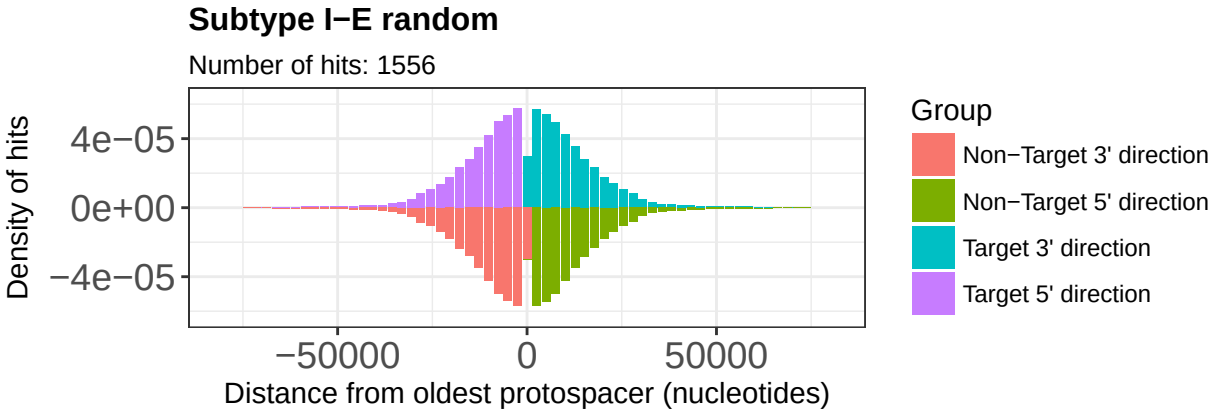

Subtype I−E targets
Number of hits: 1556
Density of hits
4e−05
0e+00
−4e−05
−30000
0
30000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction

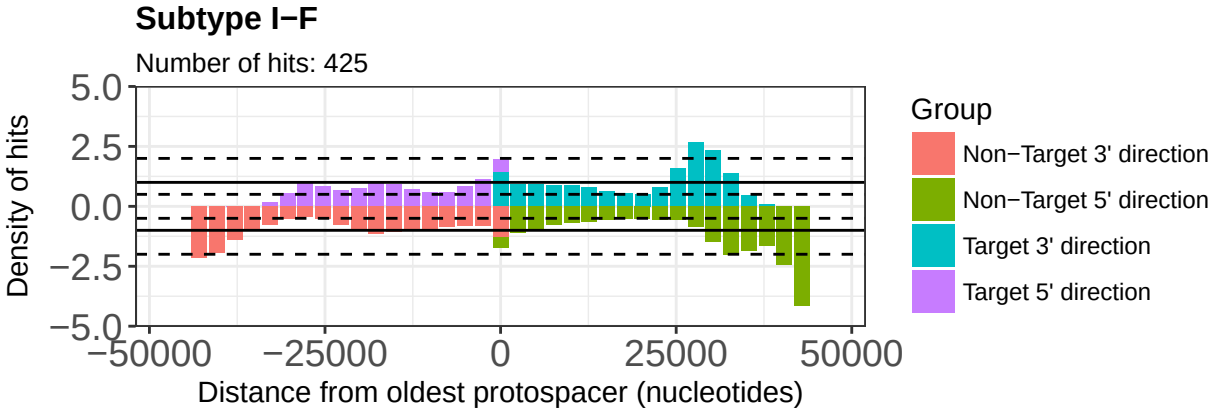

Subtype I−E random
Number of hits: 1556
Density of hits
4e−05
0e+00
−4e−05
−30000
0
30000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction

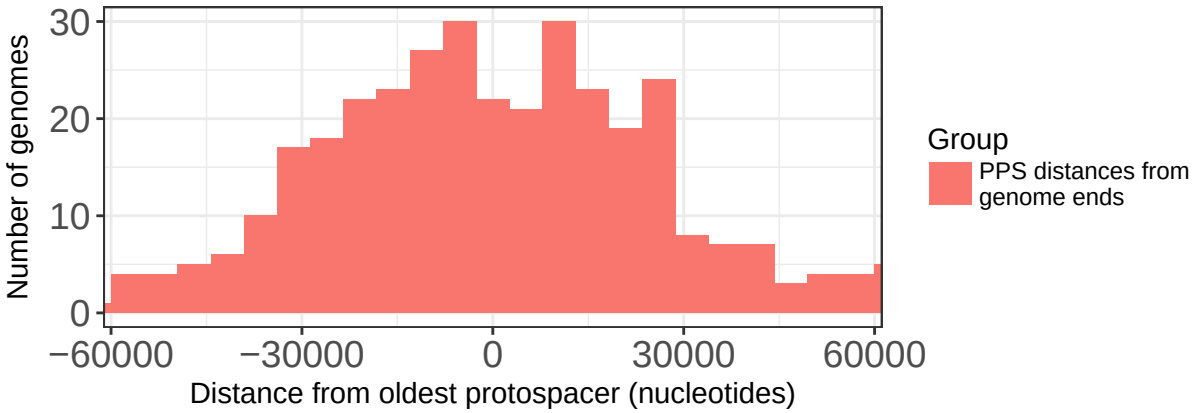

Subtype I−E
Number of hits: 1556
Density of hits
2
0
−2
−50000
0
50000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction

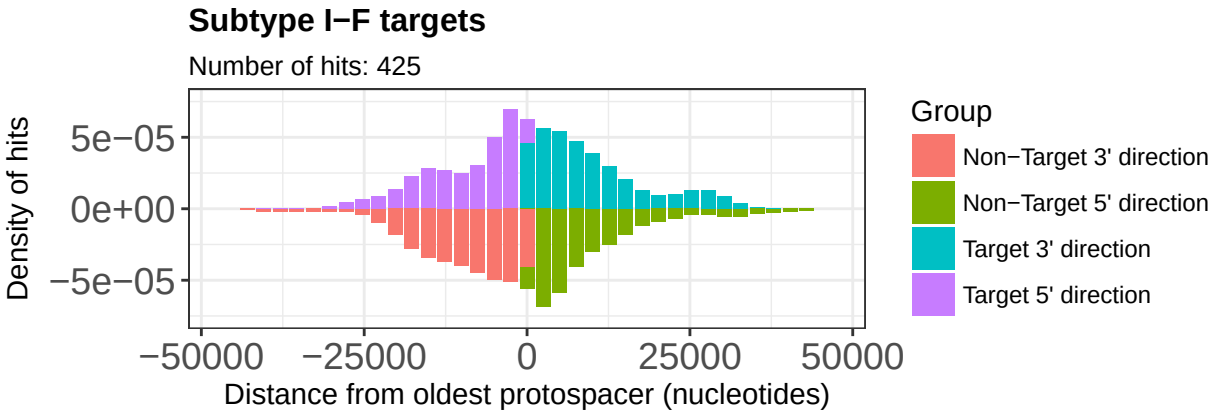

Number of genomes
100
50
0
−50000
0
50000
Distance from oldest protospacer (nucleotides)
Group
PPS distances from genome ends

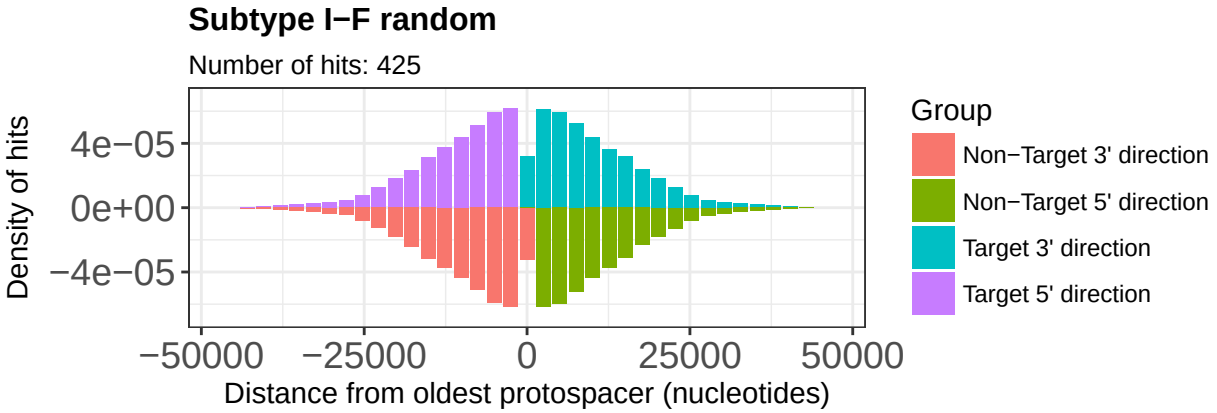

Subtype I−E targets
Number of hits: 1556
Density of hits
4e−05
0e+00
−4e−05
−50000
0
50000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction

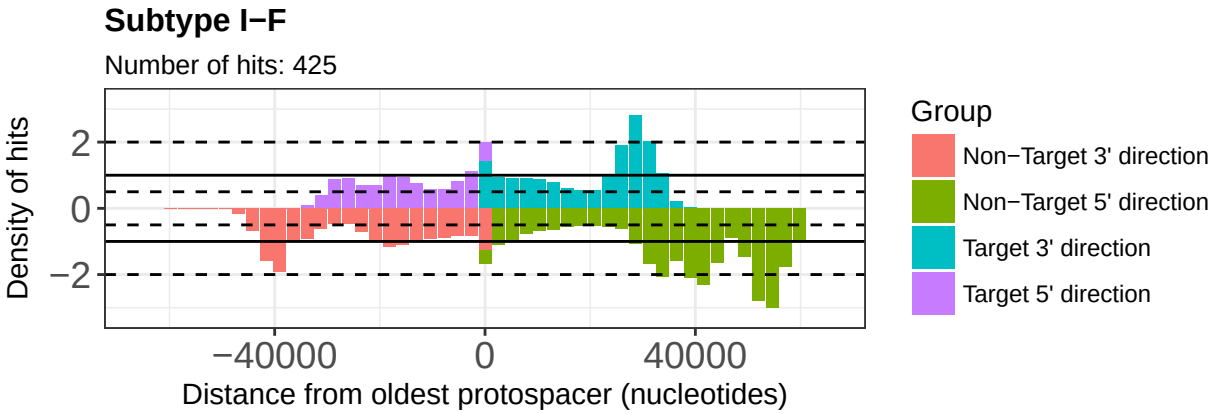

Subtype I−E random
Number of hits: 1556
Density of hits
4e−05
0e+00
−4e−05
−50000
0
50000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction

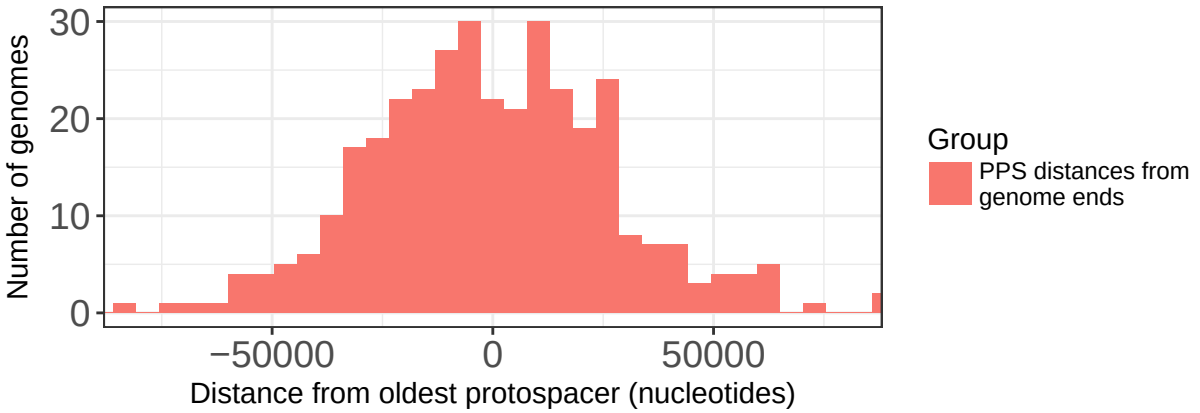

Subtype I−F
Number of hits: 425
Density of hits
5.0
2.5
0.0
−2.5
−5.0
−50000
−25000
0
25000
50000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction

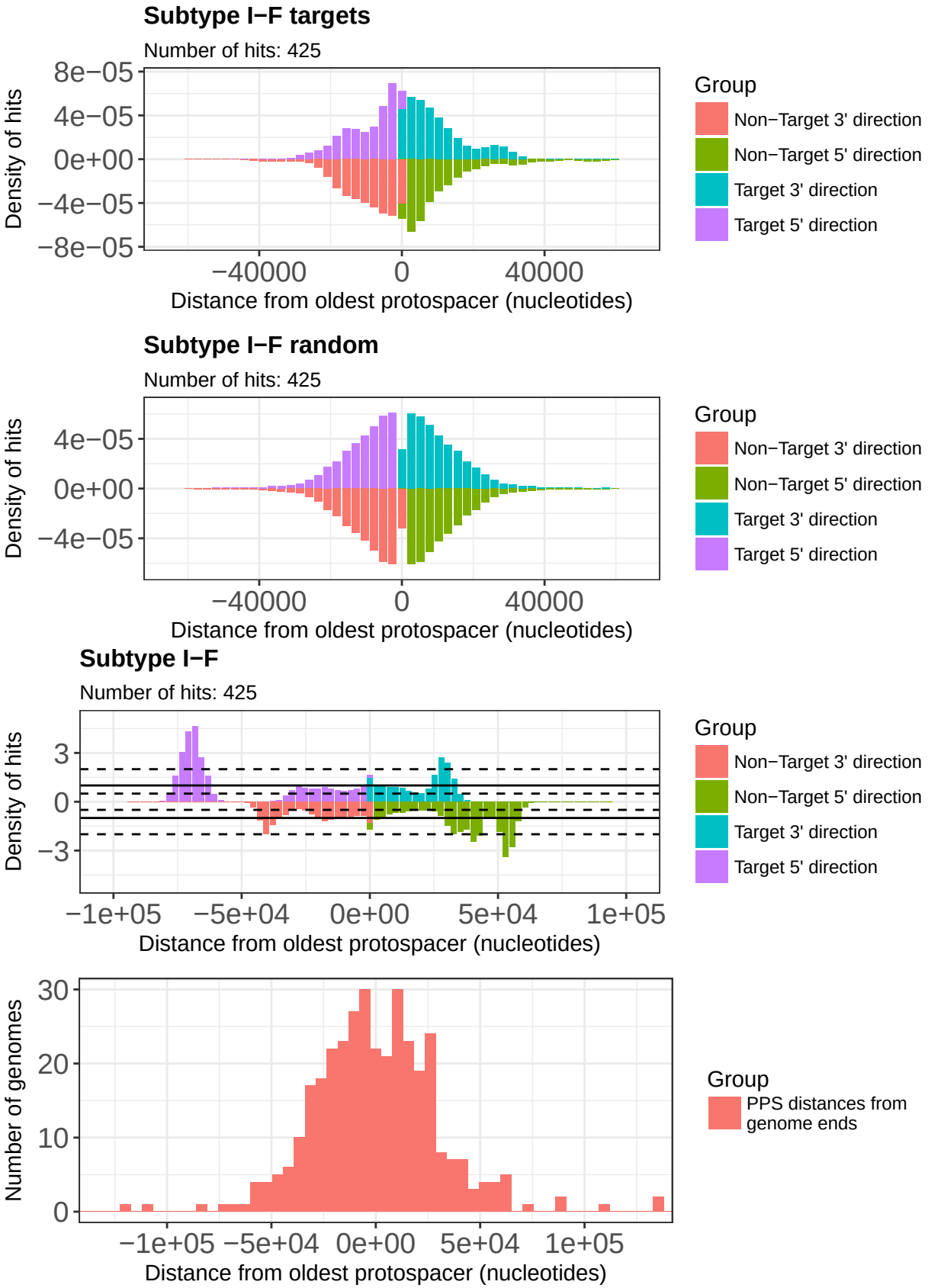

Number of genomes
30
20
10
0
−60000
−30000
0
30000
60000
Distance from oldest protospacer (nucleotides)
Group
PPS distances from genome ends

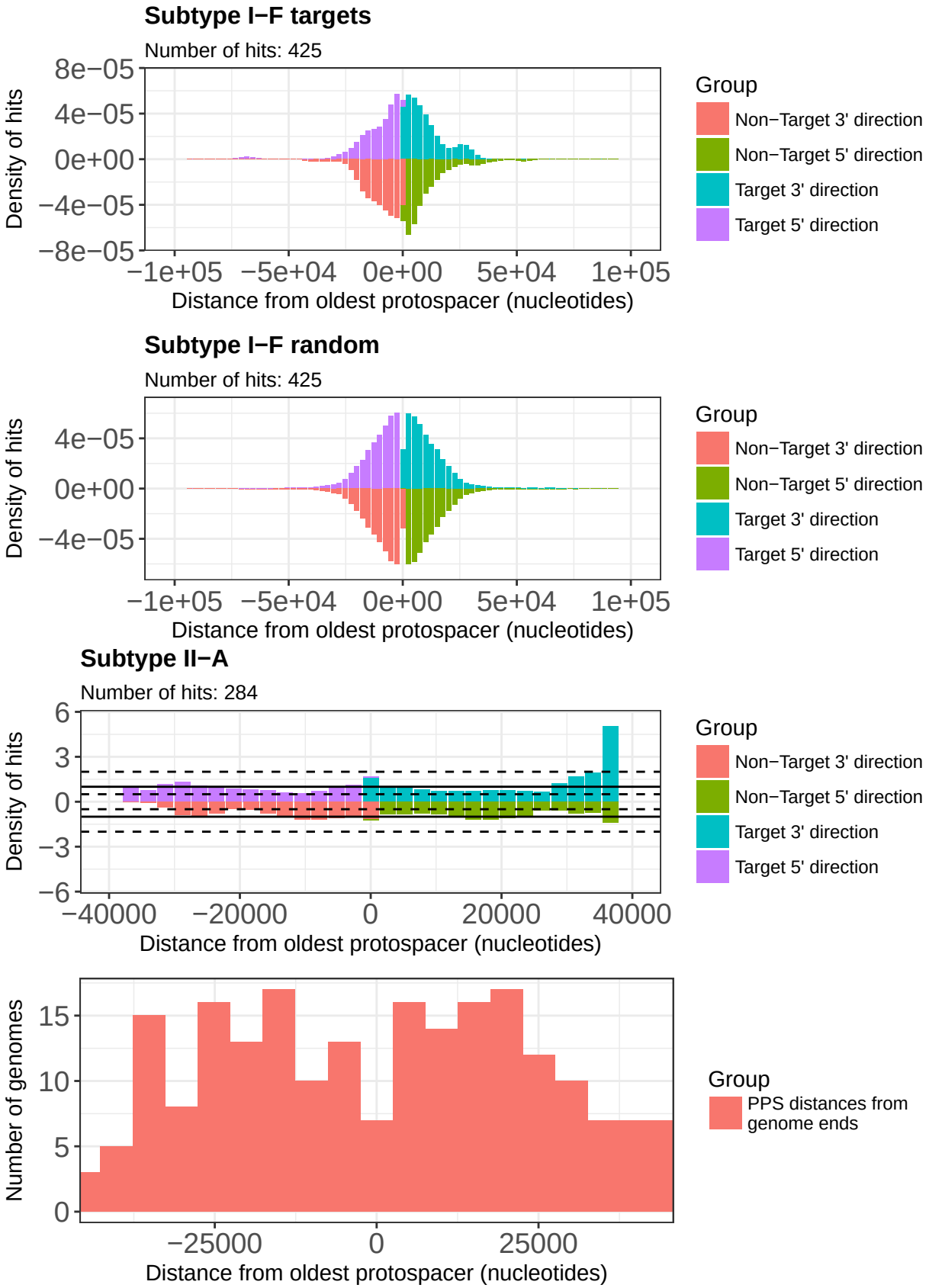

Subtype I−F targets
Number of hits: 425
Density of hits
5e−05
0e+00
−5e−05
−50000
−25000
0
25000
50000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction

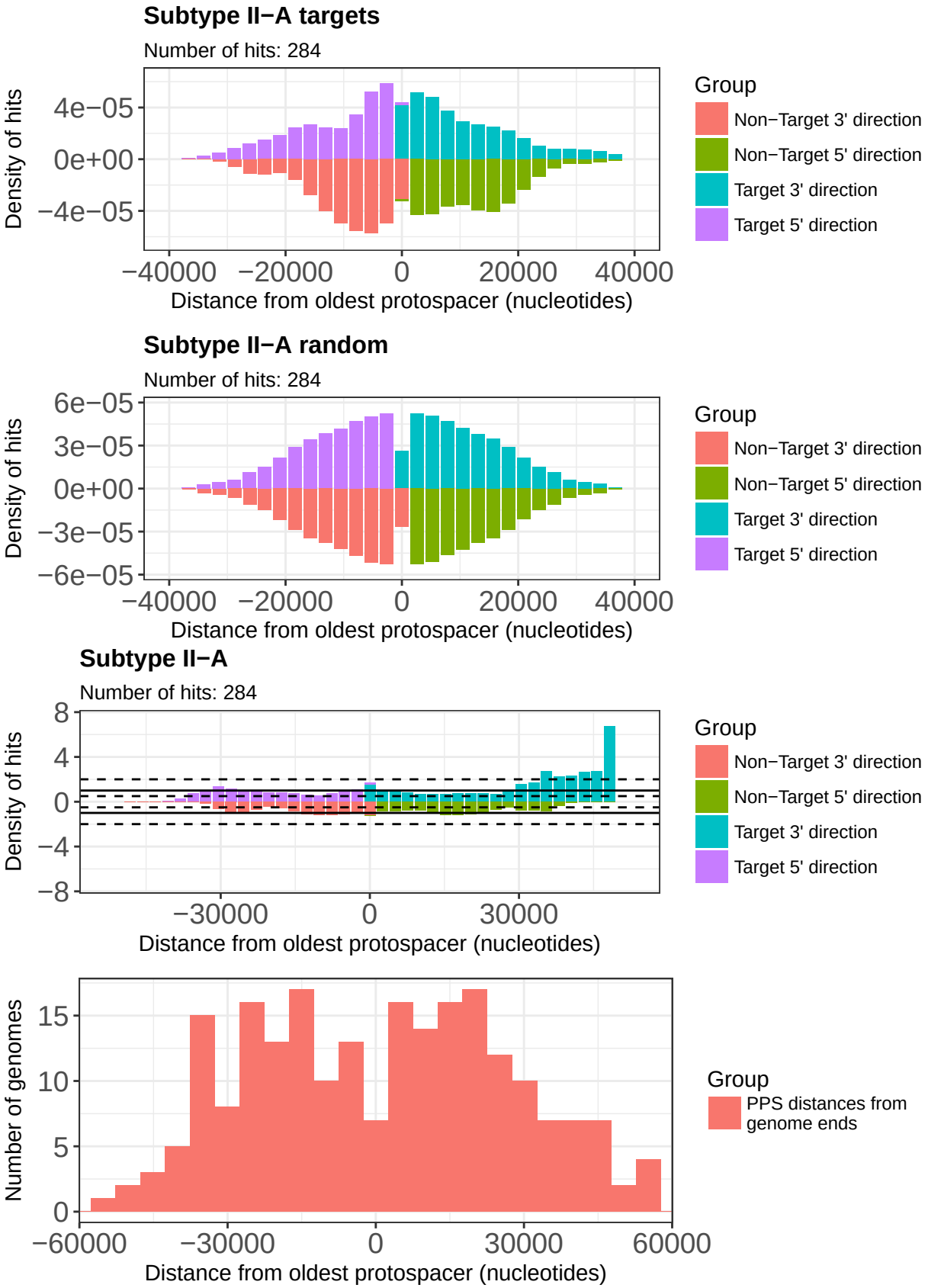

Subtype I−F random
Number of hits: 425
Density of hits
4e−05
0e+00
−4e−05
−50000
−25000
0
25000
50000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction

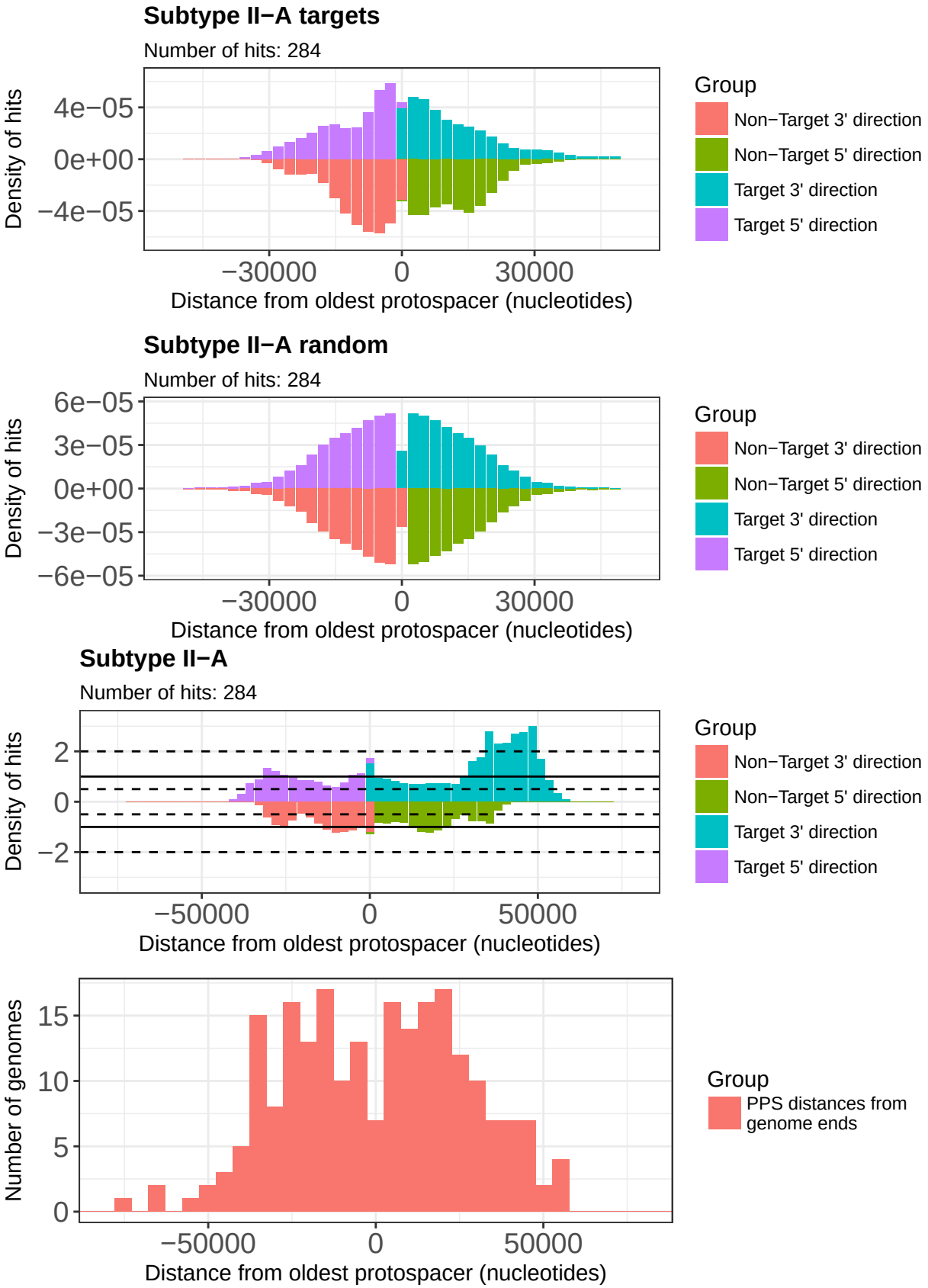

Subtype I−F
Number of hits: 425
Density of hits
2
0
−2
−40000
0
40000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction

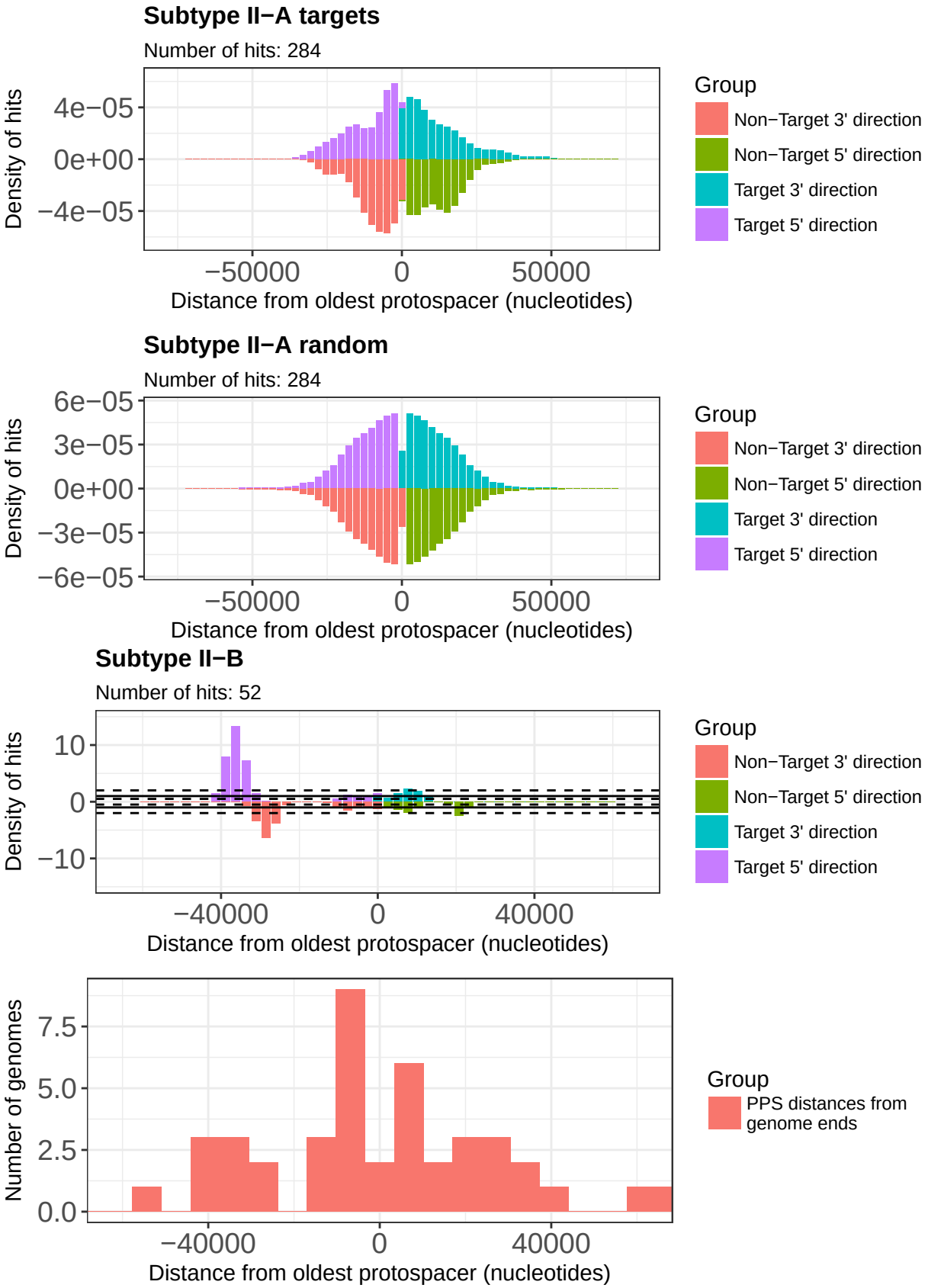

Number of genomes
30
20
10
0
−50000
0
50000
Distance from oldest protospacer (nucleotides)
Group
PPS distances from genome ends

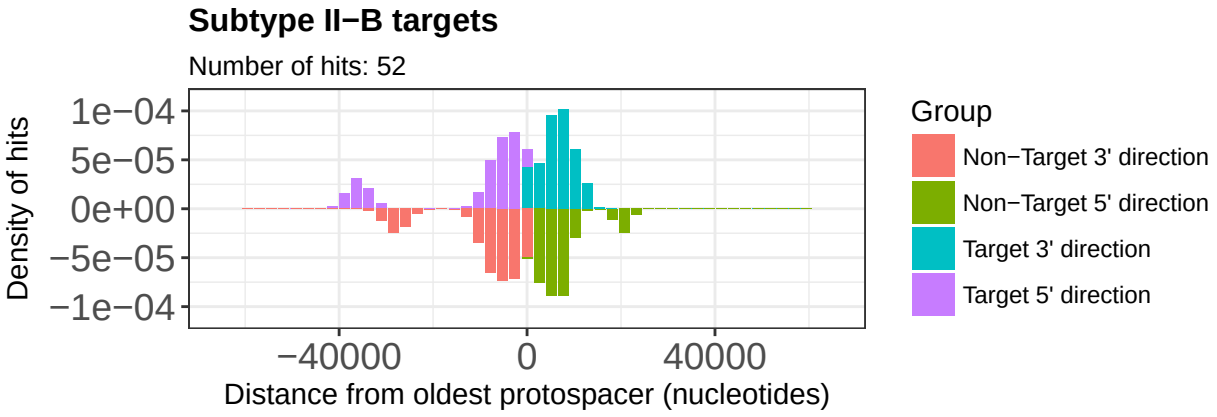

Subtype I−F targets
Number of hits: 425
Density of hits
8e−05
4e−05
0e+00
−4e−05
−8e−05
−40000
0
40000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Subtype I−F random
Number of hits: 425
Density of hits
4e−05
0e+00
−4e−05
−40000
0
40000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Subtype I−F
Number of hits: 425
Density of hits
3
0
−3
−1e+05
−5e+04
0e+00
5e+04
1e+05
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Number of genomes
30
20
10
0
−1e+05
−5e+04
0e+00
5e+04
1e+05
Distance from oldest protospacer (nucleotides)
Group
PPS distances from genome ends

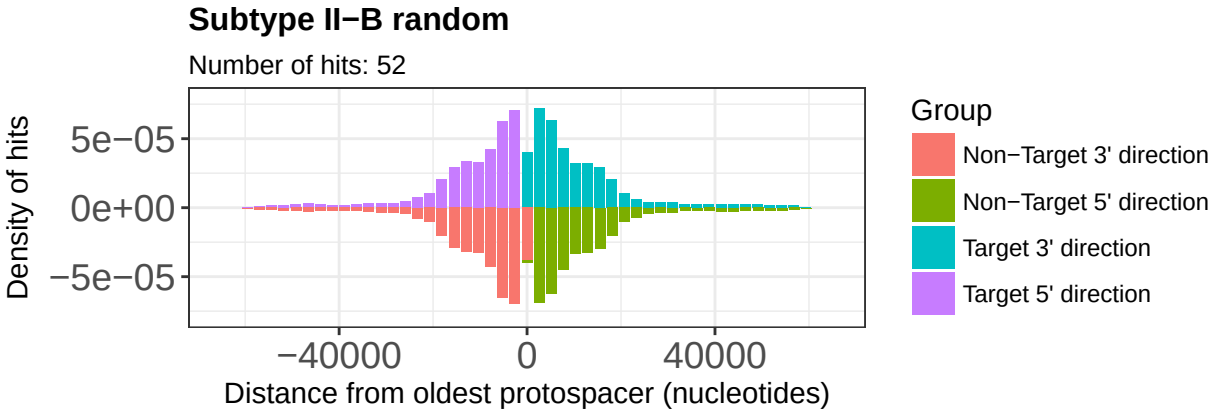

Subtype I−F targets
Number of hits: 425
Density of hits
8e−05
4e−05
0e+00
−4e−05
−8e−05
−1e+05
−5e+04
0e+00
5e+04
1e+05
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Subtype I−F random
Number of hits: 425
Density of hits
4e−05
0e+00
−4e−05
−1e+05
−5e+04
0e+00
5e+04
1e+05
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Subtype II−A
Number of hits: 284
Density of hits
6
3
0
−3
−6
−40000
−20000
0
20000
40000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Number of genomes
15
10
5
0
−25000
0
25000
Distance from oldest protospacer (nucleotides)
Group
PPS distances from genome ends

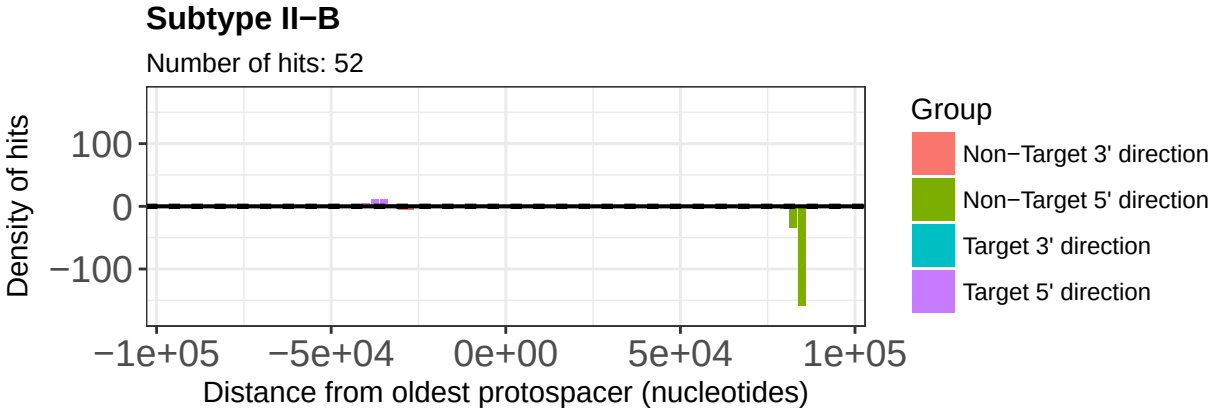

Subtype II−A targets
Number of hits: 284
Density of hits
4e−05
0e+00
−4e−05
−40000
−20000
0
20000
40000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Subtype II−A random
Number of hits: 284
Density of hits
6e−05
3e−05
0e+00
−3e−05
−6e−05
−40000
−20000
0
20000
40000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Subtype II−A
Number of hits: 284
Density of hits
8
4
0
−4
−8
−30000
0
30000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Number of genomes
15
10
5
0
−60000
−30000
0
30000
60000
Distance from oldest protospacer (nucleotides)
Group
PPS distances from genome ends

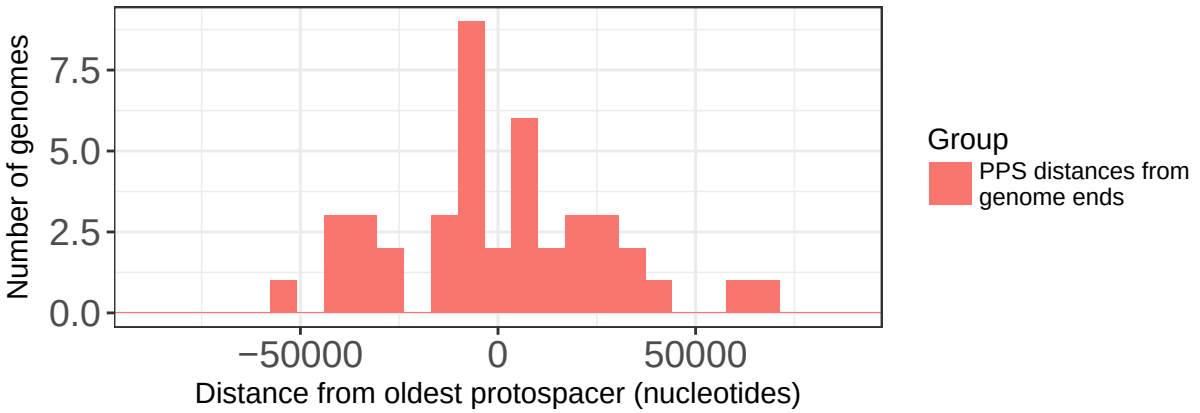

Subtype II−A targets
Number of hits: 284
Density of hits
4e−05
0e+00
−4e−05
−30000
0
30000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Subtype II−A random
Number of hits: 284
Density of hits
6e−05
3e−05
0e+00
−3e−05
−6e−05
−30000
0
30000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Subtype II−A
Number of hits: 284
Density of hits
2
0
−2
−50000
0
50000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Number of genomes
15
10
5
0
−50000
0
50000
Distance from oldest protospacer (nucleotides)
Group
PPS distances from genome ends

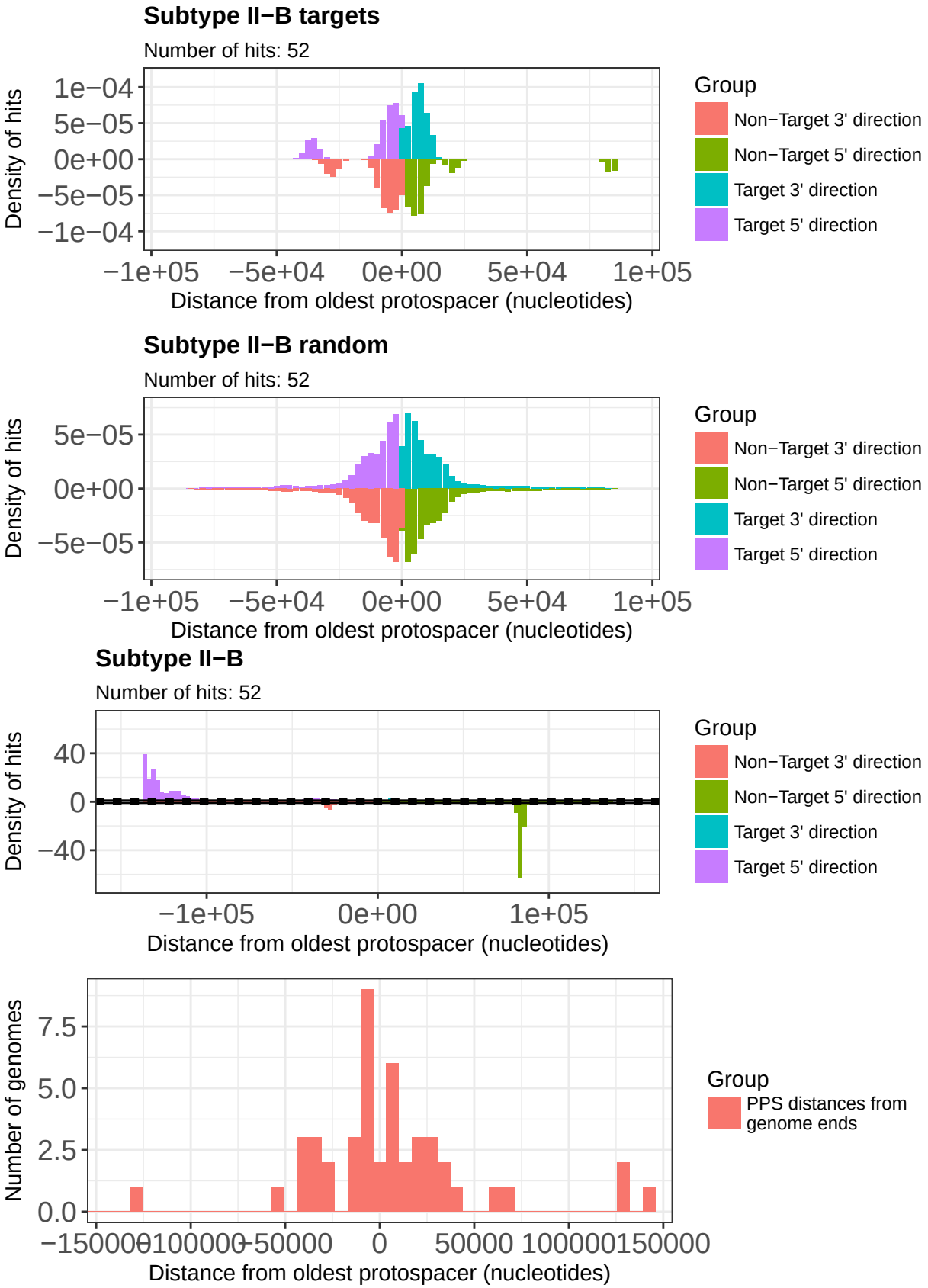

Subtype II−A targets
Number of hits: 284
Density of hits
4e−05
0e+00
−4e−05
−50000
0
50000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Subtype II−A random
Number of hits: 284
Density of hits
6e−05
3e−05
0e+00
−3e−05
−6e−05
−50000
0
50000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Subtype II−B
Number of hits: 52
Density of hits
10
0
−10
−40000
0
40000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Number of genomes
7.5
5.0
2.5
0.0
−40000
0
40000
Distance from oldest protospacer (nucleotides)
Group
PPS distances from genome ends

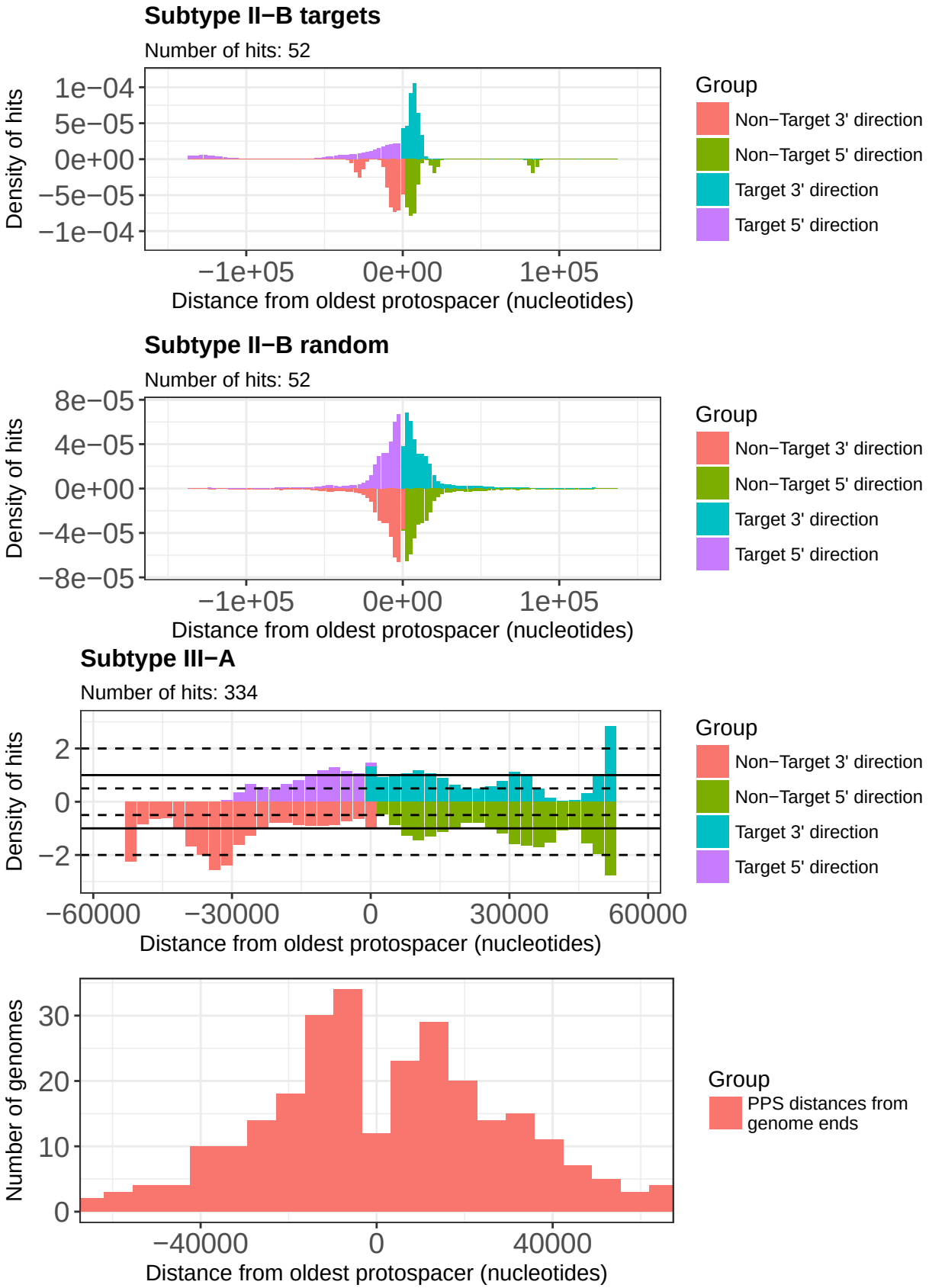

Subtype II−B targets
Number of hits: 52
Density of hits
1e−04
5e−05
0e+00
−5e−05
−1e−04
−40000
0
40000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction

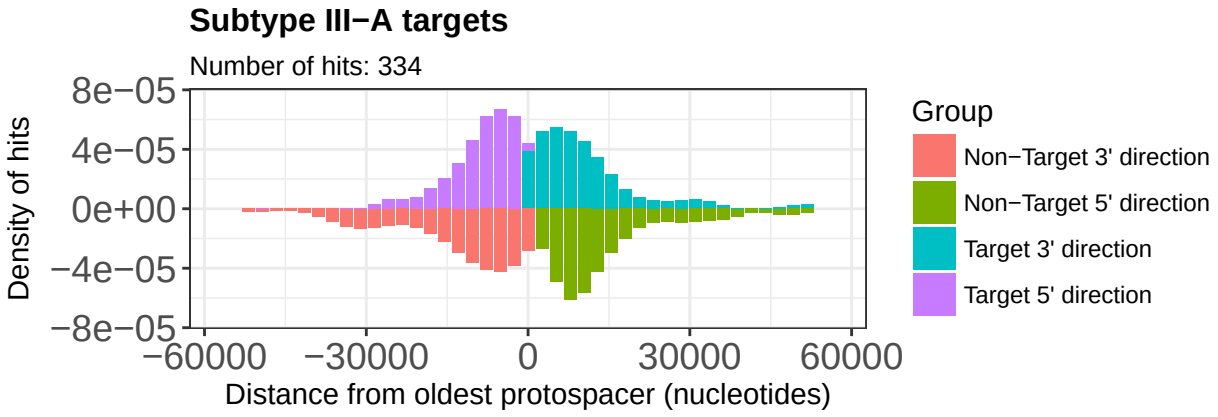

Subtype II−B random
Number of hits: 52
Density of hits
5e−05
0e+00
−5e−05
−40000
0
40000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction

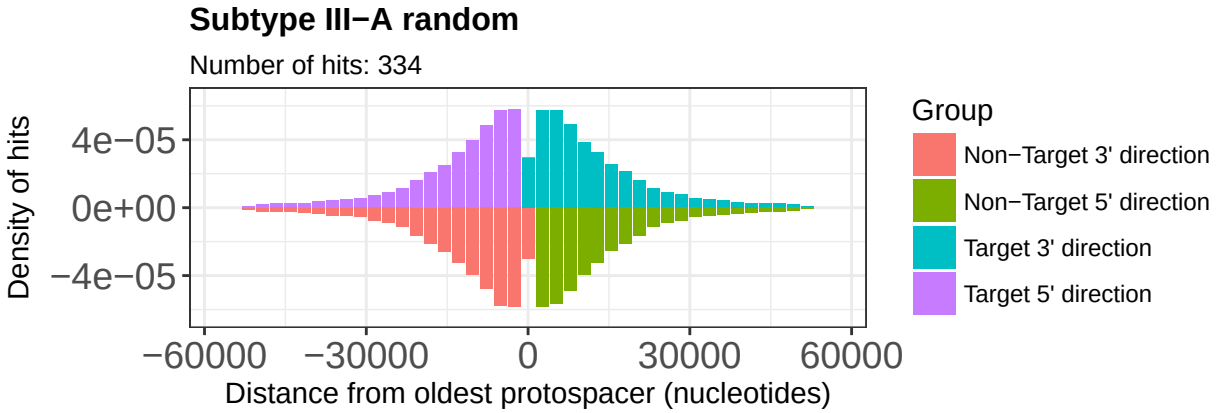

Subtype II−B
Number of hits: 52
Density of hits
100
0
−100
−1e+05
−5e+04
0e+00
5e+04
1e+05
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction

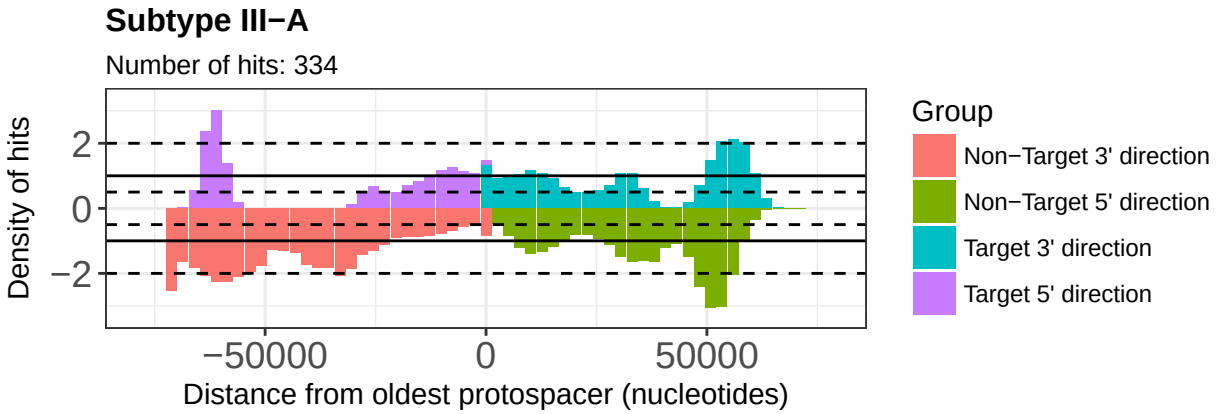

Number of genomes
7.5
5.0
2.5
0.0
−50000
0
50000
Distance from oldest protospacer (nucleotides)
Group
PPS distances from genome ends

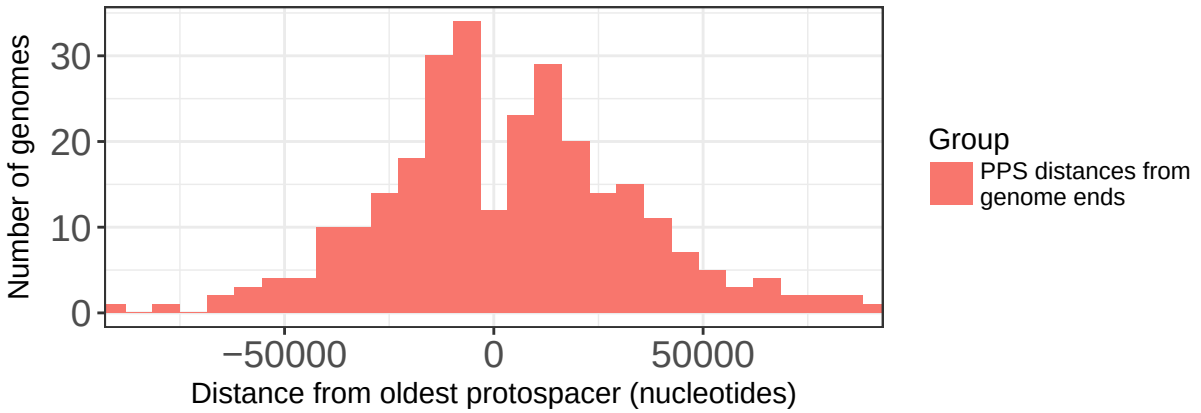

Subtype II−B targets
Number of hits: 52
Density of hits
1e−04
5e−05
0e+00
−5e−05
−1e−04
−1e+05
−5e+04
0e+00
5e+04
1e+05
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Subtype II−B random
Number of hits: 52
Density of hits
5e−05
0e+00
−5e−05
−1e+05
−5e+04
0e+00
5e+04
1e+05
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Subtype II−B
Number of hits: 52
Density of hits
40
0
−40
−1e+05
0e+00
1e+05
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Number of genomes
7.5
5.0
2.5
0.0
−150000
−100000
−50000
0
50000
100000
150000
Distance from oldest protospacer (nucleotides)
Group
PPS distances from genome ends

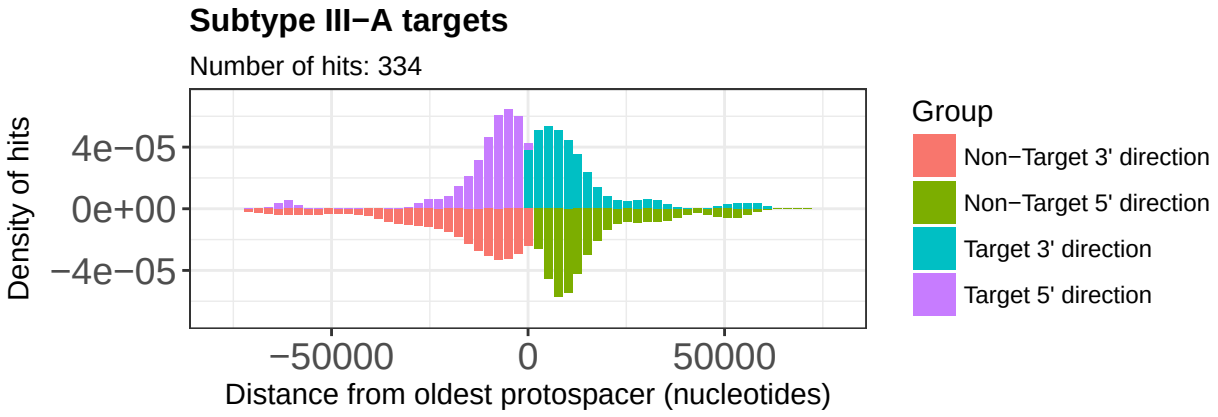

Subtype II−B targets
Number of hits: 52
Density of hits
1e−04
5e−05
0e+00
−5e−05
−1e−04
−1e+05
0e+00
1e+05
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Subtype II−B random
Number of hits: 52
Density of hits
8e−05
4e−05
0e+00
−4e−05
−8e−05
−1e+05
0e+00
1e+05
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Subtype III−A
Number of hits: 334
Density of hits
2
0
−2
−60000
−30000
0
30000
60000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Number of genomes
30
20
10
0
−40000
0
40000
Distance from oldest protospacer (nucleotides)
Group
PPS distances from genome ends

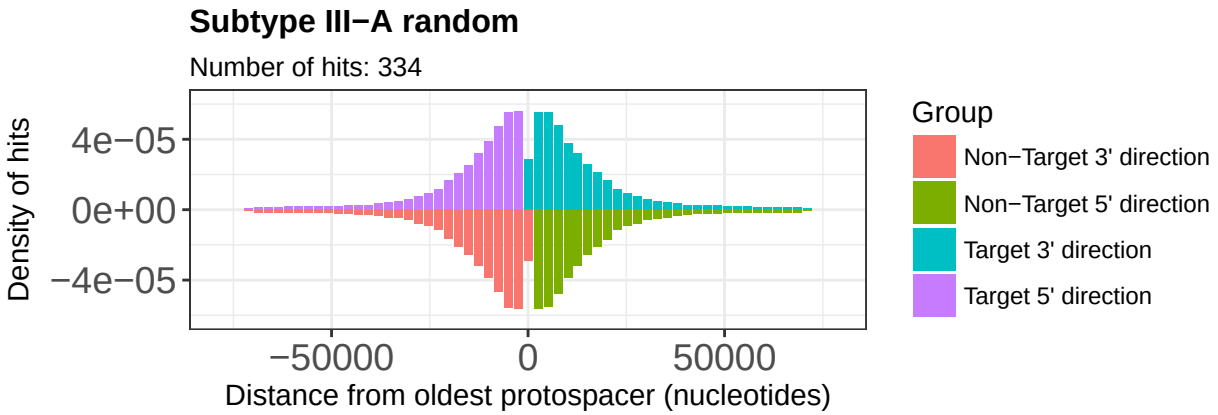

Subtype III−A targets
Number of hits: 334
Density of hits
8e−05
4e−05
0e+00
−4e−05
−8e−05
−60000
−30000
0
30000
60000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction

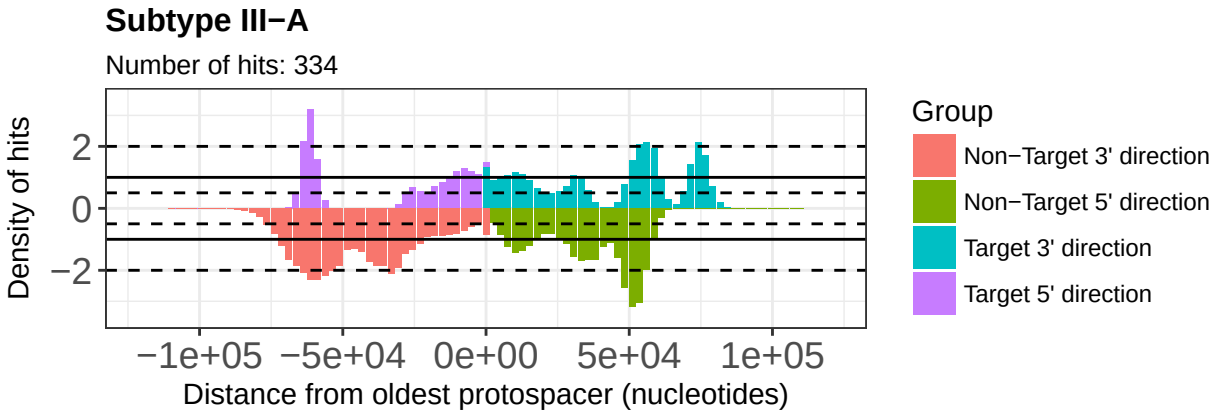

Subtype III−A random
Number of hits: 334
Density of hits
4e−05
0e+00
−4e−05
−60000
−30000
0
30000
60000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction

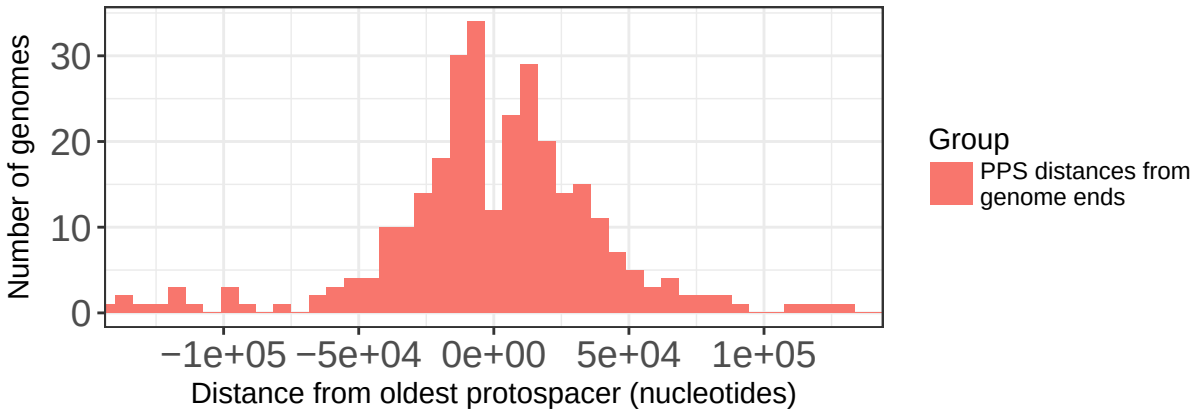

Subtype III−A
Number of hits: 334
Density of hits
2
0
−2
−50000
0
50000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction

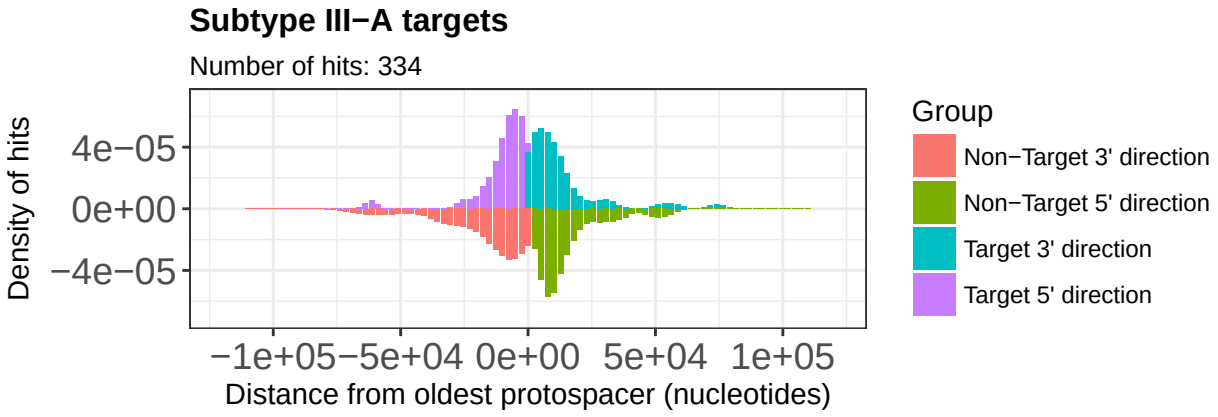

Number of genomes
30
20
10
0
−50000
0
50000
Distance from oldest protospacer (nucleotides)
Group
PPS distances from genome ends

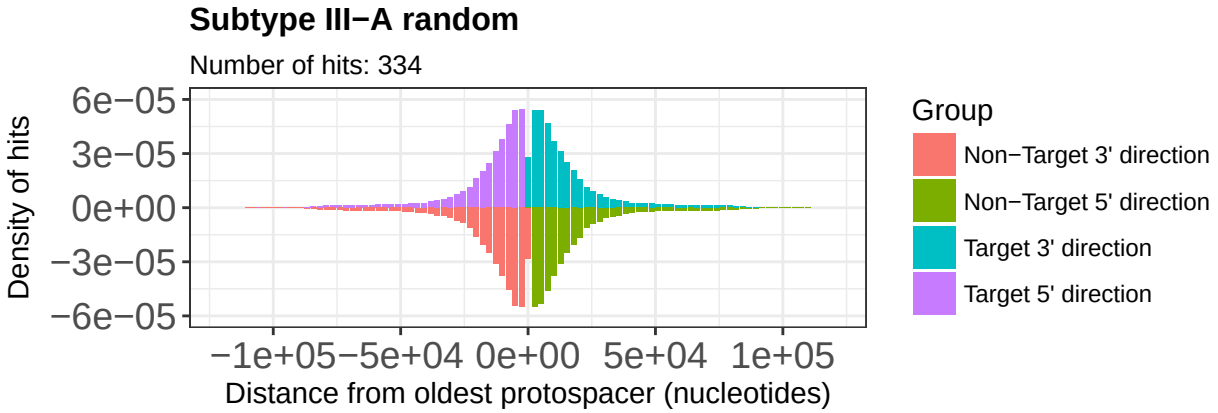

Subtype III−A targets
Number of hits: 334
Density of hits
4e−05
0e+00
−4e−05
−50000
0
50000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction

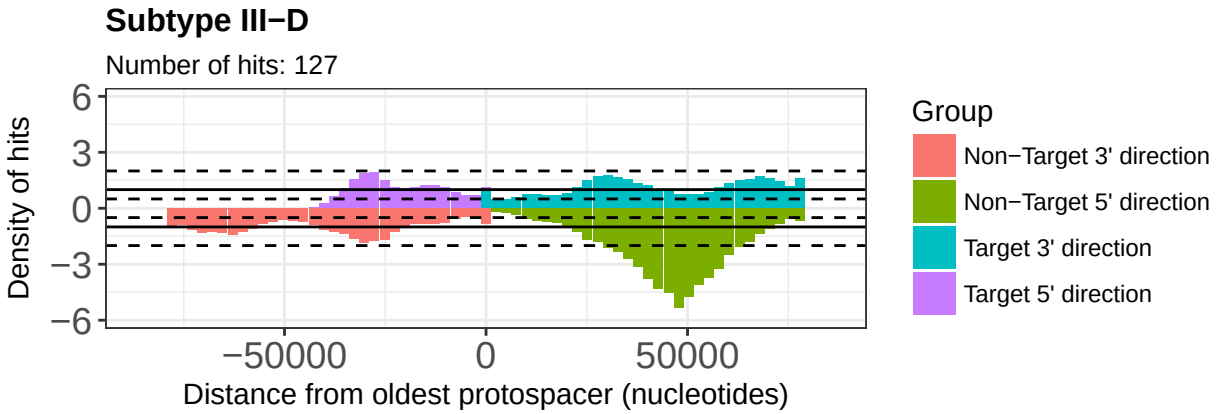

Subtype III−A random
Number of hits: 334
Density of hits
4e−05
0e+00
−4e−05
−50000
0
50000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction

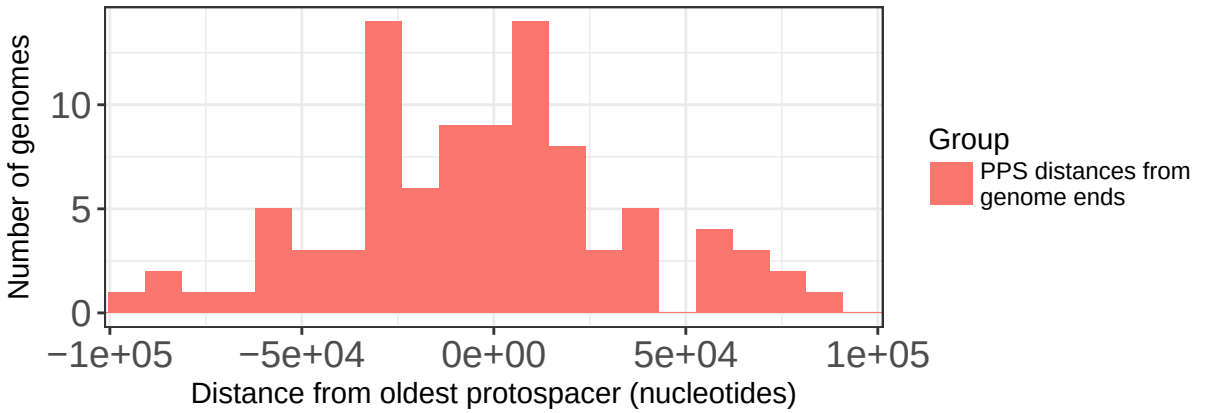

Subtype III−A
Number of hits: 334
Density of hits
2
0
−2
−1e+05
−5e+04
0e+00
5e+04
1e+05
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction

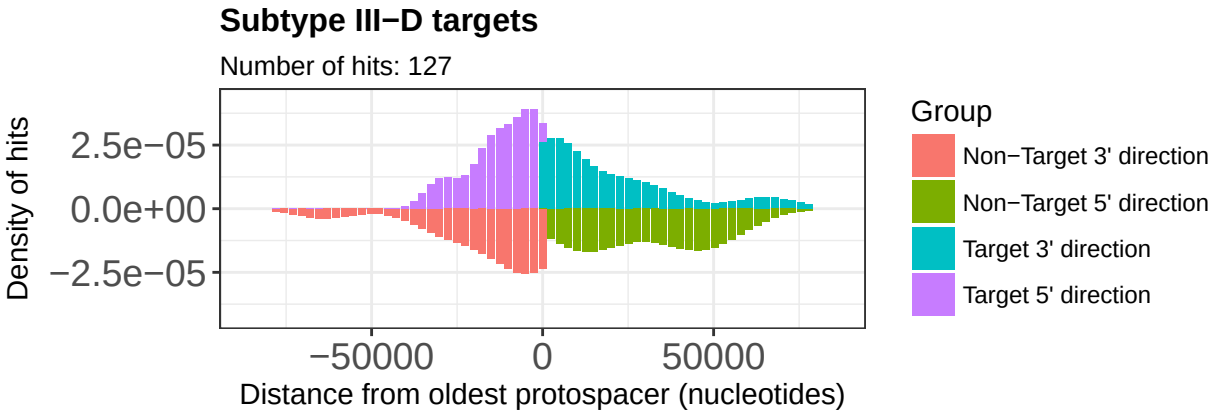

Number of genomes
30
20
10
0
−1e+05
−5e+04
0e+00
5e+04
1e+05
Distance from oldest protospacer (nucleotides)
Group
PPS distances from genome ends

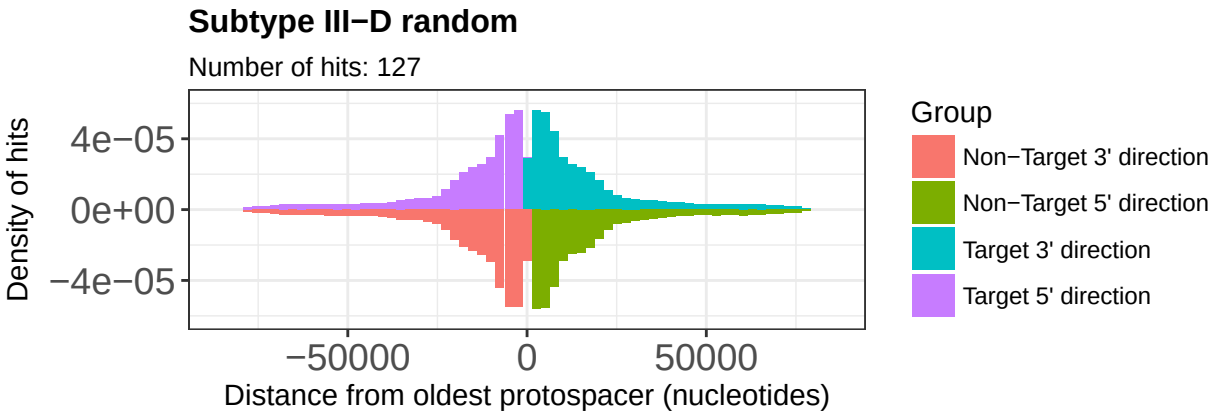

Subtype III−A targets
Number of hits: 334
Density of hits
4e−05
0e+00
−4e−05
−1e+05
−5e+04
0e+00
5e+04
1e+05
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction

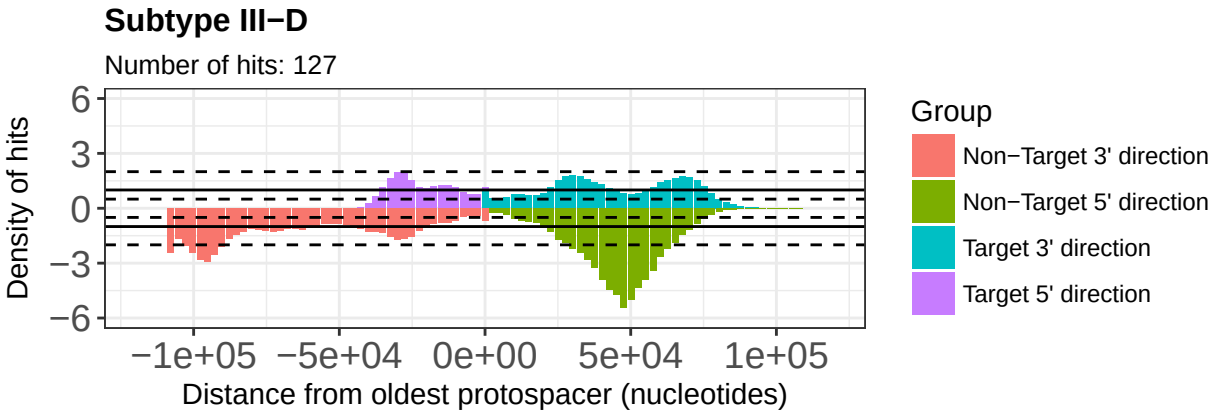

Subtype III−A random
Number of hits: 334
Density of hits
6e−05
3e−05
0e+00
−3e−05
−6e−05
−1e+05
−5e+04
0e+00
5e+04
1e+05
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction

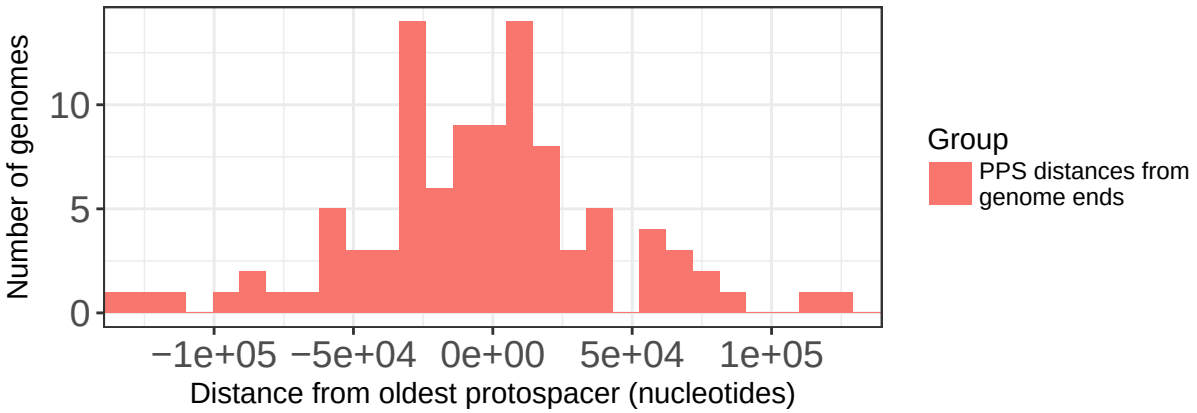

Subtype III−D
Number of hits: 127
Density of hits
6
3
0
−3
−6
−50000
0
50000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction

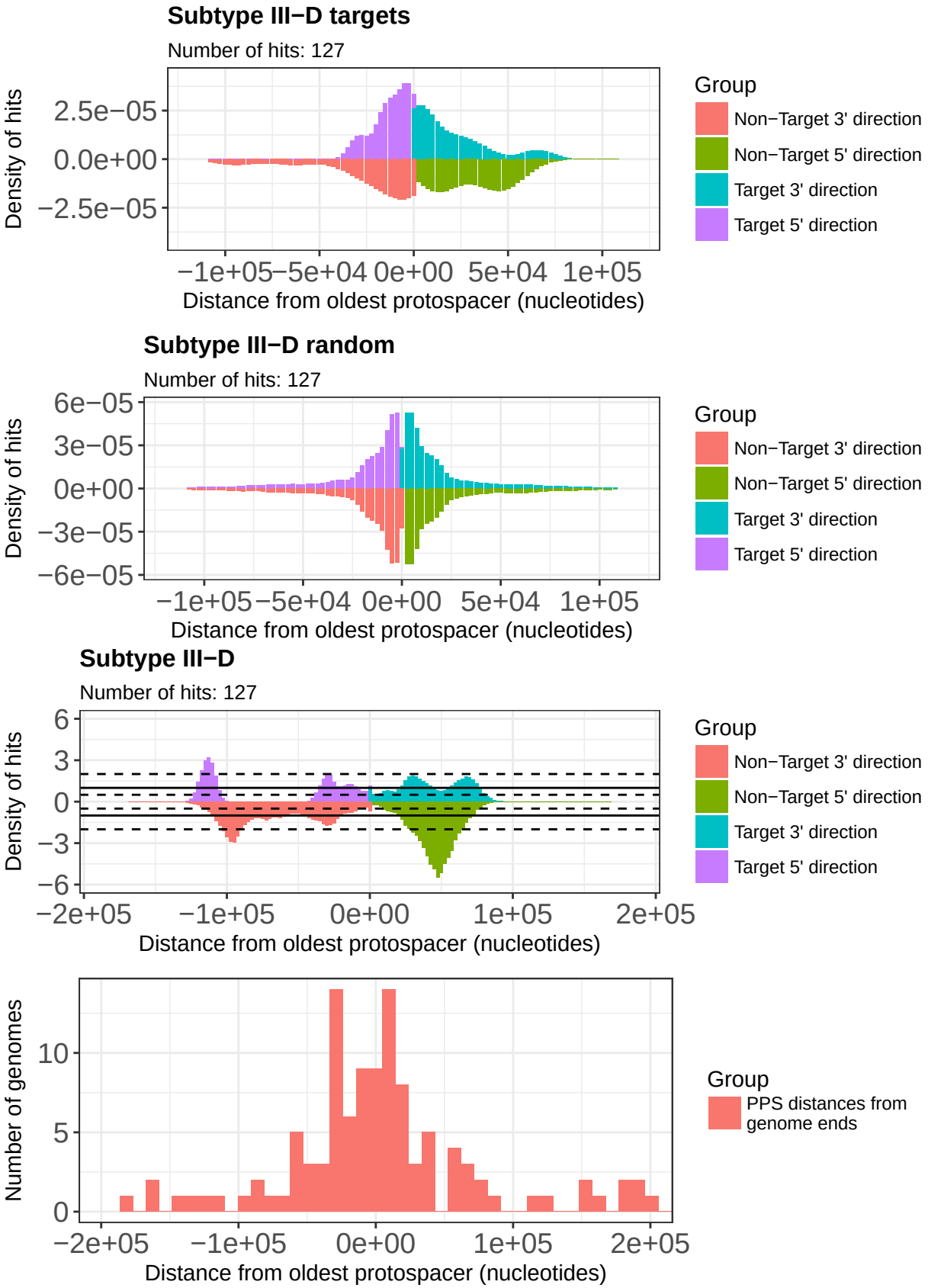

Number of genomes
10
5
0
−1e+05
−5e+04
0e+00
5e+04
1e+05
Distance from oldest protospacer (nucleotides)
Group
PPS distances from genome ends

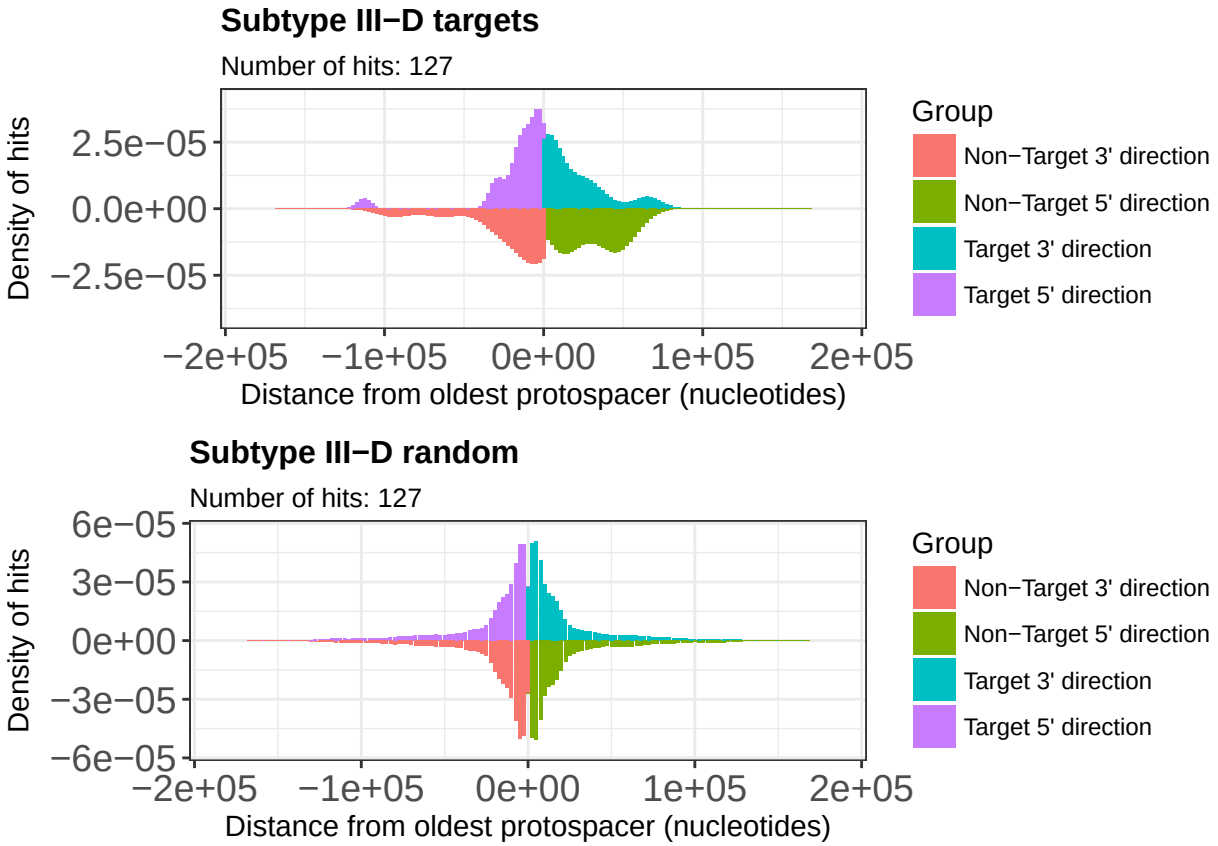

Subtype III−D targets
Number of hits: 127
Density of hits
2.5e−05
0.0e+00
−2.5e−05
−50000
0
50000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction


Subtype III−D random
Number of hits: 127
Density of hits
4e−05
0e+00
−4e−05
−50000
0
50000
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction


Subtype III−D
Number of hits: 127
Density of hits
6
3
0
−3
−6
−1e+05
−5e+04
0e+00
5e+04
1e+05
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction


Number of genomes
10
5
0
−1e+05
−5e+04
0e+00
5e+04
1e+05
Distance from oldest protospacer (nucleotides)
Group
PPS distances from genome ends


Subtype III−D targets
Number of hits: 127
Density of hits
2.5e−05
0.0e+00
−2.5e−05
−1e+05
−5e+04
0e+00
5e+04
1e+05
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Subtype III−D random
Number of hits: 127
Density of hits
6e−05
3e−05
0e+00
−3e−05
−6e−05
−1e+05
−5e+04
0e+00
5e+04
1e+05
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Subtype III−D
Number of hits: 127
Density of hits
6
3
0
−3
−6
−2e+05
−1e+05
0e+00
1e+05
2e+05
Distance from oldest protospacer (nucleotides)
Group
Non−Target 3' direction
Non−Target 5' direction
Target 3' direction
Target 5' direction
Number of genomes
10
5
0
−2e+05
−1e+05
0e+00
1e+05
2e+05
Distance from oldest protospacer (nucleotides)
Group
PPS distances from genome ends



I-A Figures

I-B Figures

I-C Figures

I-D Figures

I-E Figures

I-F Figures

II-A Figures

II-C Figures

III-A Figures

III-D Figures