Accurate multiplexing and filtering for high-throughput amplicon-sequencing

Esling Philippe[1,2,*], Lejzerowicz Franck[1] and Pawlowski Jan[1]

[1] Department of Genetics and Evolution, University of Geneva, Switzerland
[2] IRCAM, UMR 9912, Université Pierre et Marie Curie, Paris, France

* To whom correspondence should be addressed. Tel: +41223793077; Fax: +41223793340; Email: philippe.esling@unige.ch

Present Address: [Philippe Esling], Department of Genetics and Evolution, University of Geneva, Sciences 3, 30, Quai Ernest Ansermet, CH-1211 Geneva 4, Switzerland

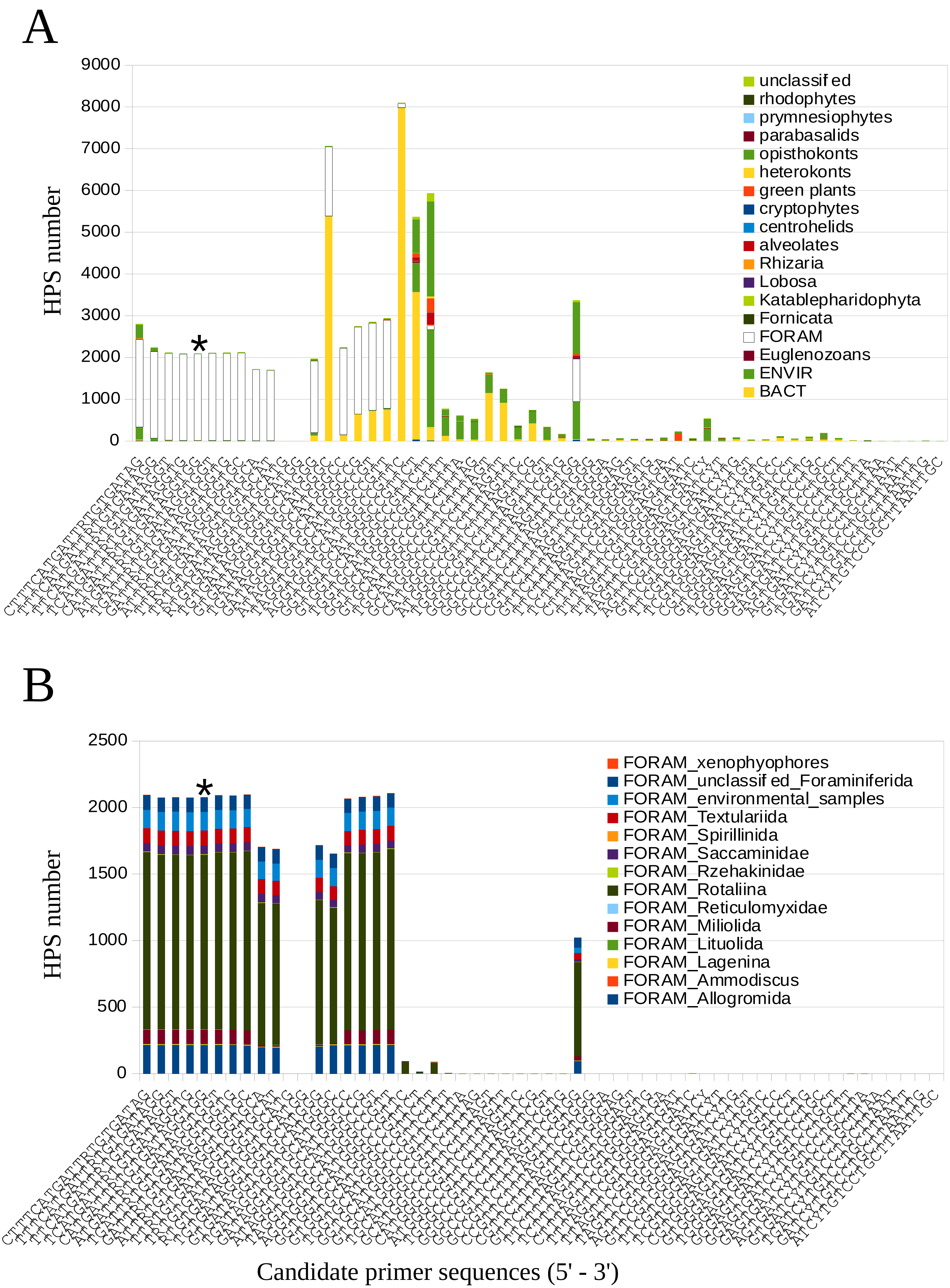
Supplementary Methods

Selection of the clone sample sets

We calculated all Needleman-Wunsch pairwise distances among the sequences of the single-sequence clone samples obtained as previously explained. Based on the resulting distance matrix, we performed clusters using average-linkage hierarchical clustering at decreasing sequence dissimilarity threshold ranging from 20 to 4 % dissimilarity. For each of the two sequencing runs, we manually assigned cluster reference sequences to the run libraries. We started by distributing the cluster reference sequences obtained at the 20 % divergence thresholds. If too few clusters exist at 20 % to bin enough sequences for our experiments, we continued the sequence distribution at 19 % dissimilarity, and continued until 4 %. This way, we ensure that we only put together samples (i.e. sequences) divergent enough to allow unambiguous assignment during analysis. We display inter- and intra-library samples divergences (Supplementary Figure 3), showing how we optimized the selection of the samples to be multiplexed per library experiment.

Supplementary Figures

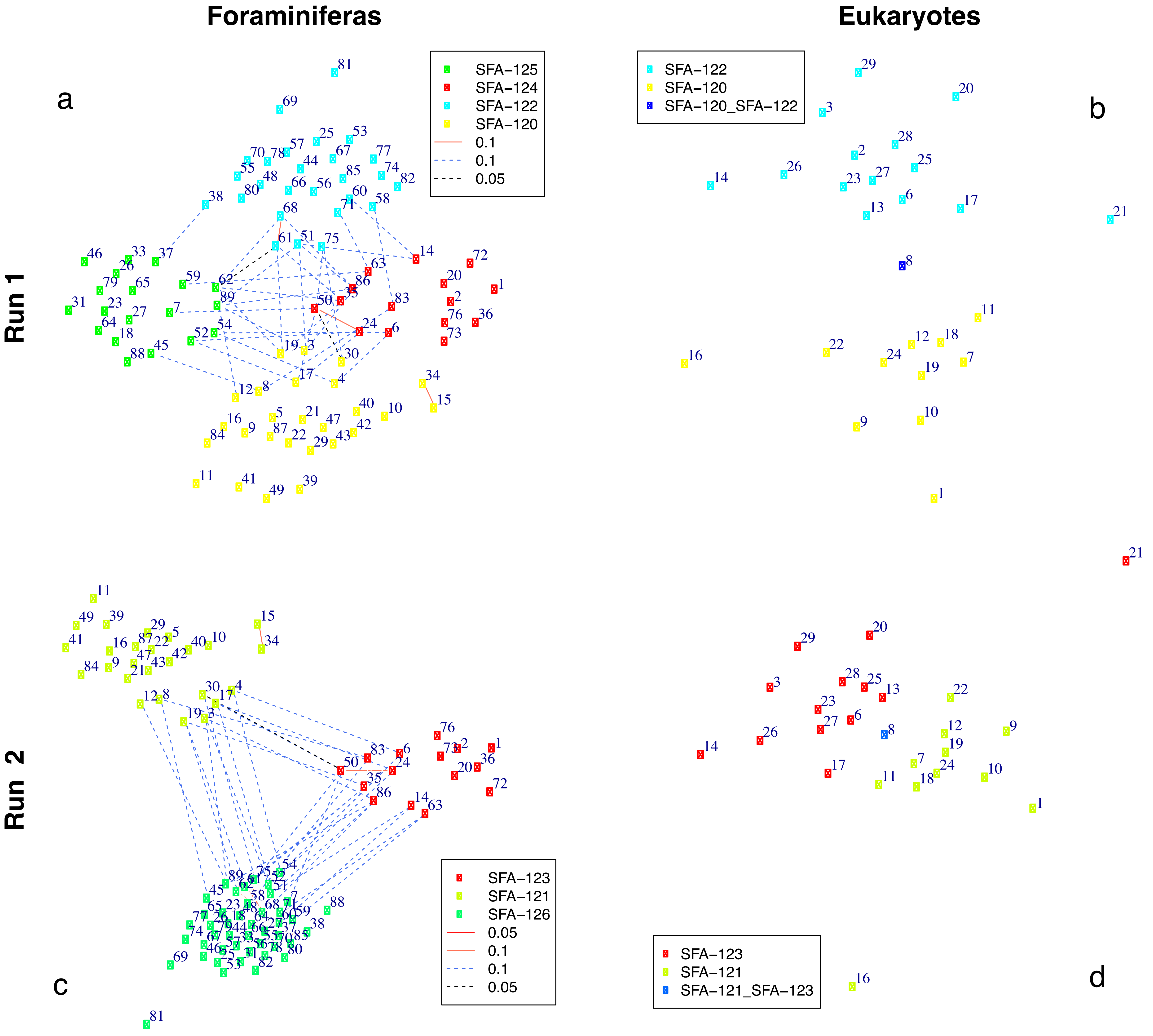
Supplementary Figure 1. Taxonomic specificity of the reverse foraminiferal primer s15. For each 20-nucleotide long candidate sequence, we show the results of extensive BLASTn searches against the NBCI's nt database (see online methods). The taxonomy retrieved for each HSP is displayed both at the phylum level (A) and at the foraminifera level (B). The s15 primer covering most of the foraminiferal diversity while avoiding most of the other phyla is indicated by a star.


A
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
HPS number
*
unclassified
rhodophytes
prymnesiophytes
parabasalids
opisthokonts
heterokonts
green plants
cryptophytes
centrohelids
alveolates
Rhizaria
Lobosa
Katablepharidophyta
Fornicata
FORAM
Euglenozoans
ENVIR
BACT
CTTTCATGATTRTGTGATAG
TTTCATGATTRTGTGATAGG
TTCATGATTRTGTGATAGGT
TCATGATTRTGTGATAGGTG
CATGATTRTGTGATAGGTGG
ATGATTRTGTGATAGGTGGT
TGATTRTGTGATAGGTGGTG
GATTRTGTGATAGGTGGTGC
ATTRTGTGATAGGTGGTGCA
TTRTGTGATAGGTGGTGCAT
TRTGTGATAGGTGGTGCATG
RTGTGATAGGTGGTGCATGG
TGTGATAGGTGGTGCATGGC
GTGATAGGTGGTGCATGGCC
TGATAGGTGGTGCATGGCCC
GATAGGTGGTGCATGGCCCG
ATAGGTGGTGCATGGCCCGT
TAGGTGGTGCATGGCCCGTT
AGGTGGTGCATGGCCCGTTC
GGTGGTGCATGGCCCGTTCT
GTGGTGCATGGCCCGTTCTT
TGGTGCATGGCCCGTTCTTT
GGTGCATGGCCCGTTCTTTA
GTGCATGGCCCGTTCTTTAG
TGCATGGCCCGTTCTTTAGT
GCATGGCCCGTTCTTTAGTT
CATGGCCCGTTCTTTAGTTC
ATGGCCCGTTCTTTAGTTCG
TGGCCCGTTCTTTAGTTCGT
GGCCCGTTCTTTAGTTCGTG
GCCCGTTCTTTAGTTCGTGG
CCCGTTCTTTAGTTCGTGGG
CCGTTCTTTAGTTCGTGGGA
CGTTCTTTAGTTCGTGGGAG
GTTCTTTAGTTCGTGGGAGT
TTCTTTAGTTCGTGGGAGTG
TCTTTAGTTCGTGGGAGTGA
CTTTAGTTCGTGGGAGTGAT
TTTAGTTCGTGGGAGTGATC
TTAGTTCGTGGGAGTGATCY
TAGTTCGTGGGAGTGATCYT
AGTTCGTGGGAGTGATCYTG
GTTCGTGGGAGTGATCYTGT
TTCGTGGGAGTGATCYTGTC
TCGTGGGAGTGATCYTGTCC
CGTGGGAGTGATCYTGTCCT
GTGGGAGTGATCYTGTCCTG
TGGGAGTGATCYTGTCCTGC
GGGAGTGATCYTGTCCTGCT
GGAGTGATCYTGTCCTGCTT
GAGTGATCYTGTCCTGCTTA
AGTGATCYTGTCCTGCTTAA
GTGATCYTGTCCTGCTTAAT
TGATCYTGTCCTGCTTAATT
GATCYTGTCCTGCTTAATTG
ATCYTGTCCTGCTTAATTGC
B
2500
2000
1500
1000
500
0
HPS number
*
FORAM_xenophyophores
FORAM_unclassified_Foraminiferida
FORAM_environmental_samples
FORAM_Textulariida
FORAM_Spirillinida
FORAM_Saccaminidae
FORAM_Rzehakinidae
FORAM_Rotaliina
FORAM_Reticulomyxidae
FORAM_Miliolida
FORAM_Lituolida
FORAM_Lagenina
FORAM_Ammodiscus
FORAM_Allogromida
CTTTCATGATTRTGTGATAG
TTTCATGATTRTGTGATAGG
TTCATGATTRTGTGATAGGT
TCATGATTRTGTGATAGGTG
CATGATTRTGTGATAGGTGG
ATGATTRTGTGATAGGTGGT
TGATTRTGTGATAGGTGGTG
GATTRTGTGATAGGTGGTGC
ATTRTGTGATAGGTGGTGCA
TTRTGTGATAGGTGGTGCAT
TRTGTGATAGGTGGTGCATG
RTGTGATAGGTGGTGCATGG
TGTGATAGGTGGTGCATGGC
GTGATAGGTGGTGCATGGCC
TGATAGGTGGTGCATGGCCC
GATAGGTGGTGCATGGCCCG
ATAGGTGGTGCATGGCCCGT
TAGGTGGTGCATGGCCCGTT
AGGTGGTGCATGGCCCGTTC
GGTGGTGCATGGCCCGTTCT
GTGGTGCATGGCCCGTTCTT
TGGTGCATGGCCCGTTCTTT
GGTGCATGGCCCGTTCTTTA
GTGCATGGCCCGTTCTTTAG
TGCATGGCCCGTTCTTTAGT
GCATGGCCCGTTCTTTAGTT
CATGGCCCGTTCTTTAGTTC
ATGGCCCGTTCTTTAGTTCG
TGGCCCGTTCTTTAGTTCGT
GGCCCGTTCTTTAGTTCGTG
GCCCGTTCTTTAGTTCGTGG
CCCGTTCTTTAGTTCGTGGG
CCGTTCTTTAGTTCGTGGGA
CGTTCTTTAGTTCGTGGGAG
GTTCTTTAGTTCGTGGGAGT
TTCTTTAGTTCGTGGGAGTG
TCTTTAGTTCGTGGGAGTGA
CTTTAGTTCGTGGGAGTGAT
TTTAGTTCGTGGGAGTGATC
TTAGTTCGTGGGAGTGATCY
TAGTTCGTGGGAGTGATCYT
AGTTCGTGGGAGTGATCYTG
GTTCGTGGGAGTGATCYTGT
TTCGTGGGAGTGATCYTGTC
TCGTGGGAGTGATCYTGTCC
CGTGGGAGTGATCYTGTCCT
GTGGGAGTGATCYTGTCCTG
TGGGAGTGATCYTGTCCTGC
GGGAGTGATCYTGTCCTGCT
GGAGTGATCYTGTCCTGCTT
GAGTGATCYTGTCCTGCTTA
AGTGATCYTGTCCTGCTTAA
GTGATCYTGTCCTGCTTAAT
TGATCYTGTCCTGCTTAATT
GATCYTGTCCTGCTTAATTG
ATCYTGTCCTGCTTAATTGC
Candidate primer sequences (5' - 3')

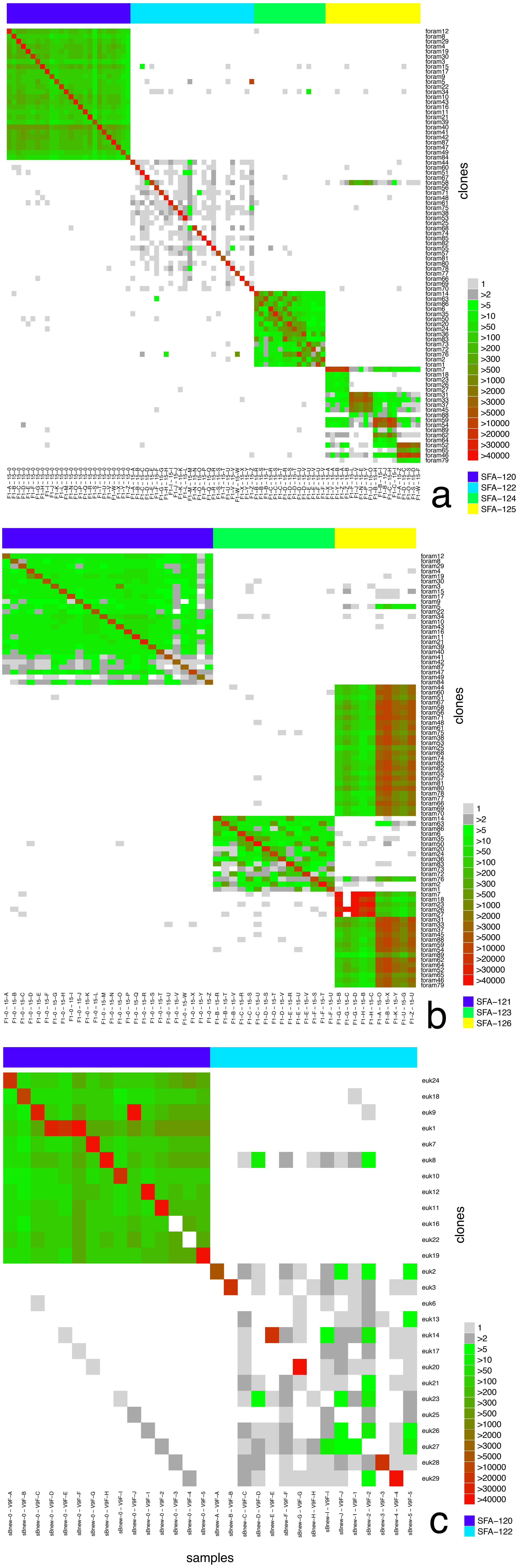
Supplementary Figure 2. Experimental designs and molecular workflow. We use a library of Sanger-sequenced clones to provide either *Single-sequence* samples or *Mock communities* samples corresponding to single-sequence clone amplicons pooled in controlled ratios. These samples are labelled by PCR amplification either using one out of the two primers tagged (*Single*) or the two primers tagged (*Double*). We deployed these tagged primers according to each *Experimental design*, represented by the vectors and matrices (rows: forward primers, columns: reverse primers). The samples labelled by the deployed combinations of tagged primers are indicated both for single-sequence samples (colored blocks) and mock communities (black symbols). After the tagging PCR, the labelled samples are pooled in equimolar ratios (*Sample pooling*) and a TruSeq Nano sequencing library (from SFA-120 to SFA-126) is prepared for each pool (*Library PCR*). The resulting libraries are then distributed in two mixes as indicated (*Library pooling*) and sequenced (*MiSeq sequencing*).

Clones library
Single-sequences
Mock communities
Samples
Single
Double
Tag
Primer
Mock community
Replicates
Tagging PCR
A ... ... Z
A 5
0
Foram
Euk
Experimental design
Sample Pooling
SFA-120
SFA-121
SFA-122
SFA-123
SFA-124
SFA-125
SFA-126
Library barcode
Adapter
Library PCR
Library Pooling
Run 1
Run 2
MiSeq sequencing

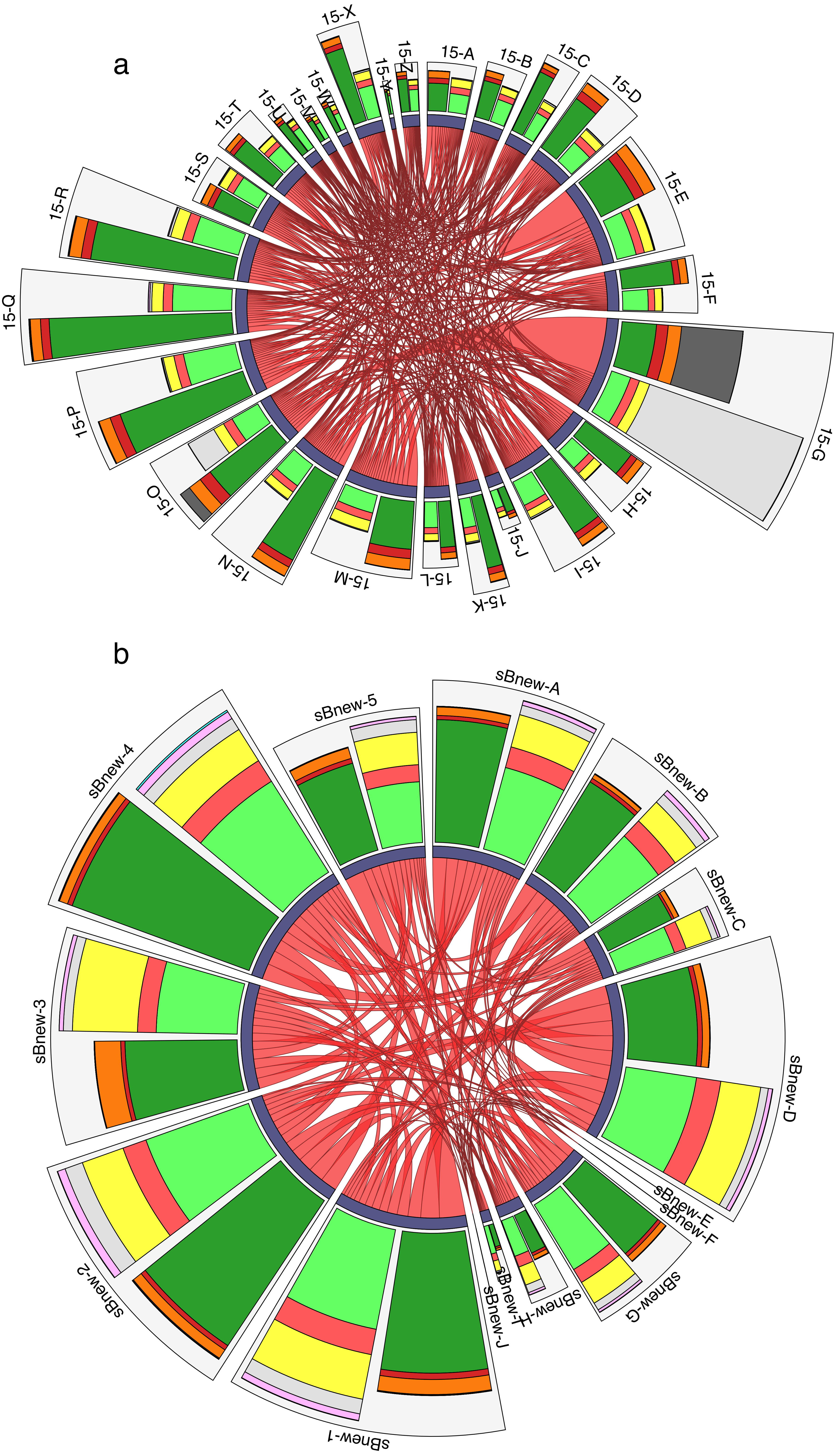
Supplementary Figure 3. Pairwise distance networks among Sanger-sequenced clone sequences per taxon and per run libraries. The distances among Foraminiferal clones are displayed according to their deployment in both the run 1 (a) and the run 2 (c). The distances among Eukaryotic clones are also displayed according to their deployment in both the run 1 (b) and the run 2 (d). The clones are represented by labeled vertices colored according to the library where they are used. The pairwise distances are represented by edges. Intra- and inter-library distances are materialized by plain and dotted edges, respectively. All distances were measured from exact, pairwise Needleman-Wunsch global alignments and by counting end gaps as well as each internal gaps as differences. Note that the minimum distance between clone samples pooled in a same library was always above 4 %.


Foraminiferas
Eukaryotes
Run 1
Run 2
a
b
c
d
SFA−125
SFA−124
SFA−122
SFA−120
0.1
0.1
0.05
SFA−122
SFA−120
SFA−120_SFA−122
SFA−123
SFA−121
SFA−126
0.05
0.1
0.1
0.05
SFA−123
SFA−121
SFA−121_SFA−123

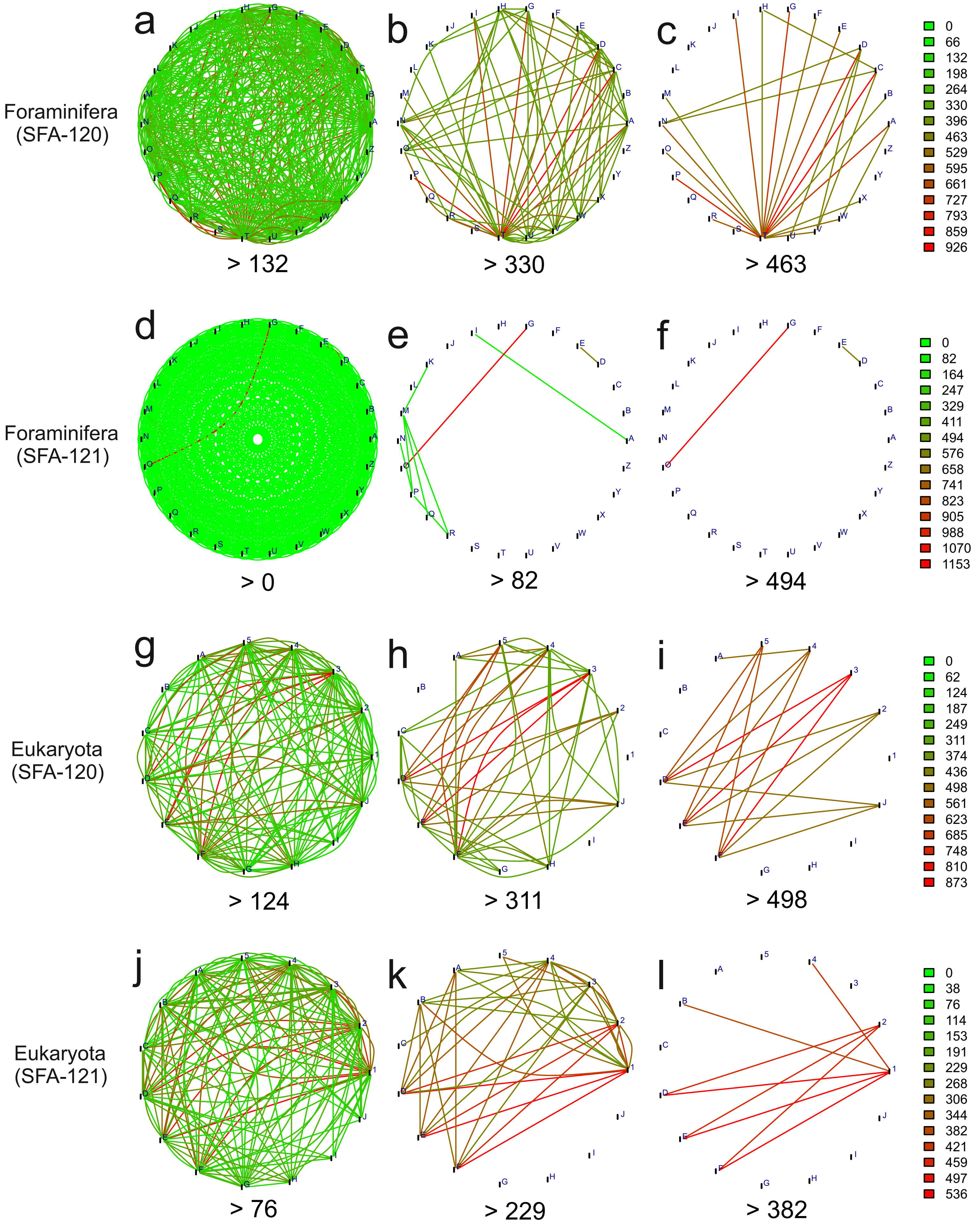
Supplementary Figure 4. Clone-to-sample heat maps for per taxon and run. The numbers of reads associated with all the sequences assigned to each foraminiferal clone used in the first (a) run and second run (b) as well as to each eukaryotic clone in the first run (c) are displayed. Only the true samples are presented labeled with the primers combinations. The samples are grouped by library (color code in upper bars and legends) and the libraries are sorted according to their incremental order of preparation. The clones are sorted according to the samples.

a
clones
foram12
foram8
foram29
foram4
foram19
foram30
foram3
foram15
foram17
foram9
foram5
foram22
foram34
foram10
foram43
foram16
foram11
foram21
foram39
foram40
foram41
foram42
foram87
foram47
foram49
foram84
foram44
foram60
foram51
foram67
foram58
foram56
foram71
foram48
foram61
foram75
foram38
foram53
foram25
foram68
foram74
foram85
foram82
foram55
foram57
foram81
foram80
foram78
foram77
foram66
foram69
foram70
foram14
foram63
foram86
foram6
foram35
foram50
foram20
foram24
foram36
foram83
foram73
foram72
foram76
foram2
foram1
foram7
foram18
foram23
foram26
foram27
foram31
foram33
foram37
foram45
foram88
foram59
foram54
foram89
foram62
foram64
foram52
foram65
foram46
foram79
F1–A – 15–0
F1–B – 15–0
F1–C – 15–0
F1–D – 15–0
F1–E – 15–0
F1–F – 15–0
F1–G – 15–0
F1–H – 15–0
F1–I – 15–0
F1–J – 15–0
F1–K – 15–0
F1–L – 15–0
F1–M – 15–0
F1–N – 15–0
F1–O – 15–0
F1–P – 15–0
F1–Q – 15–0
F1–R – 15–0
F1–S – 15–0
F1–T – 15–0
F1–U – 15–0
F1–V – 15–0
F1–W – 15–0
F1–X – 15–0
F1–Y – 15–0
F1–Z – 15–0
F1–A – 15–A
F1–B – 15–B
F1–C – 15–C
F1–D – 15–D
F1–E – 15–E
F1–F – 15–F
F1–G – 15–G
F1–H – 15–H
F1–I – 15–I
F1–J – 15–J
F1–K – 15–K
F1–L – 15–L
F1–M – 15–M
F1–N – 15–N
F1–O – 15–O
F1–P – 15–P
F1–Q – 15–Q
F1–R – 15–R
F1–S – 15–S
F1–T – 15–T
F1–U – 15–U
F1–V – 15–V
F1–W – 15–W
F1–X – 15–X
F1–Y – 15–Y
F1–Z – 15–Z
F1–B – 15–R
F1–B – 15–S
F1–B – 15–T
F1–C – 15–R
F1–C – 15–S
F1–C – 15–T
F1–D – 15–R
F1–D – 15–S
F1–D – 15–T
F1–D – 15–U
F1–D – 15–V
F1–E – 15–U
F1–E – 15–V
F1–F – 15–U
F1–F – 15–V
F1–X – 15–A
F1–Y – 15–A
F1–Y – 15–B
F1–Z – 15–A
F1–Z – 15–B
F1–F – 15–L
F1–J – 15–V
F1–N – 15–E
F1–P – 15–Y
F1–U – 15–O
F1–B – 15–H
F1–B – 15–I
F1–B – 15–J
F1–C – 15–H
F1–C – 15–I
F1–A – 15–Z
F1–D – 15–C
F1–O – 15–S
F1–S – 15–D
F1–W – 15–F
1
>2
>5
>10
>50
>100
>200
>300
>500
>1000
>2000
>3000
>5000
>10000
>20000
>30000
>40000
SFA–120
SFA–122
SFA–124
SFA–125
b
clones
foram12
foram8
foram29
foram4
foram19
foram30
foram3
foram15
foram17
foram9
foram5
foram22
foram34
foram10
foram43
foram16
foram11
foram21
foram39
foram40
foram41
foram42
foram87
foram47
foram49
foram84
foram44
foram60
foram51
foram67
foram58
foram56
foram71
foram48
foram61
foram75
foram38
foram53
foram25
foram68
foram74
foram85
foram82
foram55
foram57
foram81
foram80
foram78
foram77
foram66
foram69
foram70
foram14
foram63
foram86
foram6
foram35
foram50
foram20
foram24
foram36
foram83
foram73
foram72
foram76
foram2
foram1
foram7
foram18
foram23
foram26
foram27
foram31
foram33
foram37
foram45
foram88
foram59
foram54
foram89
foram62
foram64
foram52
foram65
foram46
foram79
F1–0 – 15–A
F1–0 – 15–B
F1–0 – 15–C
F1–0 – 15–D
F1–0 – 15–E
F1–0 – 15–F
F1–0 – 15–G
F1–0 – 15–H
F1–0 – 15–I
F1–0 – 15–J
F1–0 – 15–K
F1–0 – 15–L
F1–0 – 15–M
F1–0 – 15–N
F1–0 – 15–O
F1–0 – 15–P
F1–0 – 15–Q
F1–0 – 15–R
F1–0 – 15–S
F1–0 – 15–T
F1–0 – 15–U
F1–0 – 15–V
F1–0 – 15–W
F1–0 – 15–X
F1–0 – 15–Y
F1–0 – 15–Z
F1–B – 15–R
F1–B – 15–T
F1–B – 15–V
F1–C – 15–R
F1–C – 15–S
F1–C – 15–U
F1–D – 15–S
F1–D – 15–T
F1–D – 15–V
F1–E – 15–R
F1–E – 15–U
F1–E – 15–V
F1–F – 15–S
F1–F – 15–T
F1–F – 15–U
F1–G – 15–B
F1–G – 15–C
F1–G – 15–D
F1–H – 15–B
F1–H – 15–C
F1–A – 15–O
F1–B – 15–X
F1–K – 15–Y
F1–U – 15–G
F1–Z – 15–U
1
>2
>5
>10
>50
>100
>200
>300
>500
>1000
>2000
>3000
>5000
>10000
>20000
>30000
>40000
SFA–121
SFA–123
SFA–126
c
clones
euk24
euk18
euk9
euk1
euk7
euk8
euk10
euk12
euk11
euk16
euk22
euk19
euk2
euk3
euk6
euk13
euk14
euk17
euk20
euk21
euk23
euk25
euk26
euk27
euk28
euk29
sBrew–0 – V9F–A
sBrew–0 – V9F–B
sBrew–0 – V9F–C
sBrew–0 – V9F–D
sBrew–0 – V9F–E
sBrew–0 – V9F–F
sBrew–0 – V9F–G
sBrew–0 – V9F–H
sBrew–0 – V9F–I
sBrew–0 – V9F–J
sBrew–0 – V9F–1
sBrew–0 – V9F–2
sBrew–0 – V9F–3
sBrew–0 – V9F–4
sBrew–0 – V9F–5
sBrew–A – V9F–A
sBrew–B – V9F–B
sBrew–C – V9F–C
sBrew–D – V9F–D
sBrew–E – V9F–E
sBrew–F – V9F–F
sBrew–G – V9F–G
sBrew–H – V9F–H
sBrew–I – V9F–I
sBrew–J – V9F–J
sBrew–1 – V9F–1
sBrew–2 – V9F–2
sBrew–3 – V9F–3
sBrew–4 – V9F–4
sBrew–5 – V9F–5
samples
1
>2
>5
>10
>50
>100
>200
>300
>500
>1000
>2000
>3000
>5000
>10000
>20000
>30000
>40000
SFA–120
SFA–122

Supplementary Figure 5. Single tagging mayhem (SFA-121). Mistagging events are displayed in the chord diagrams separately for foraminiferal (a) and eukaryotic (b) data. The central parts represent critical mistags as red links indicating the amount of reads when a sample targeted by a specific tag (one extremity of the string) is found labelled with another tag (other extremity). These central parts would be completely empty in the absence of mistags. For each expected tagged primer, joint barplots indicate the amounts of ISUs (light colors) and reads (dark colors) binned into several categories, including good (expected sample), critical (unexpected sample), non-critical (spurious combination), chimera, dimers and unknown, sequences. The legend to the colors is the same used for Figure 2.

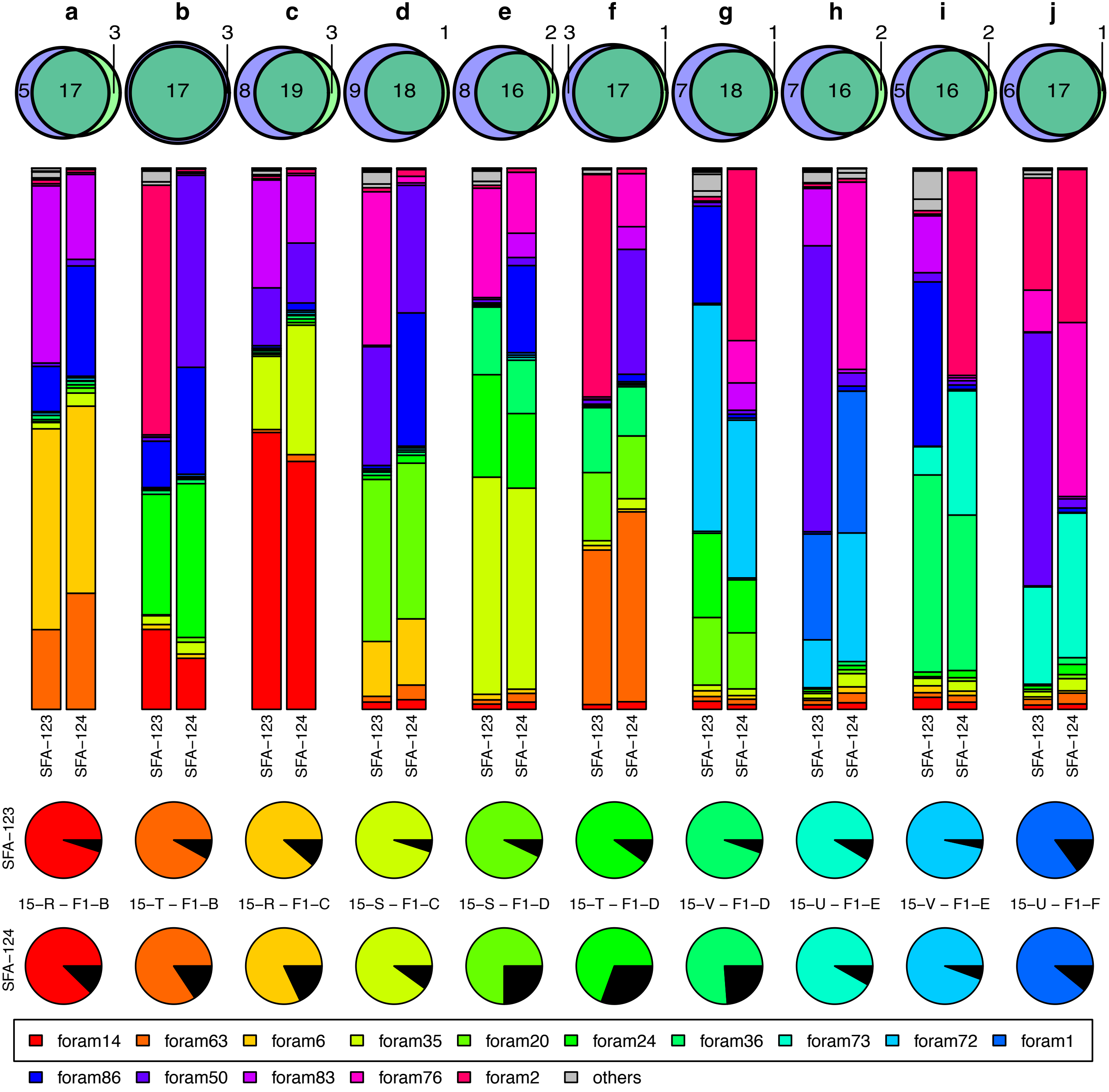

a
15-A
15-B
15-C
15-D
15-E
15-F
15-G
15-H
15-I
15-J
15-K
15-L
15-M
15-N
15-O
15-P
15-Q
15-R
15-S
15-T
15-U
15-V
15-W
15-X
15-Y
15-Z
b
sBnew-A
sBnew-B
sBnew-C
sBnew-D
sBnew-E
sBnew-F
sBnew-G
sBnew-H
sBnew-I
sBnew-J
sBnew-1
sBnew-2
sBnew-3
sBnew-4
sBnew-5

Supplementary Figure 6. Primer-to-primer mistagging events for each taxon in each single-tagging library. For three sequence abundance thresholds, three networks displaying the numbers of mistagged reads above each threshold are displayed for Foraminifera in SFA-120 (a, b, c) and in SFA-121 (d, e, f) as well as for Eukaryota in SFA-120 (g, h, i) and in SFA-121 (j, k, l). The threshold values associated with each network are indicated at the bottom.

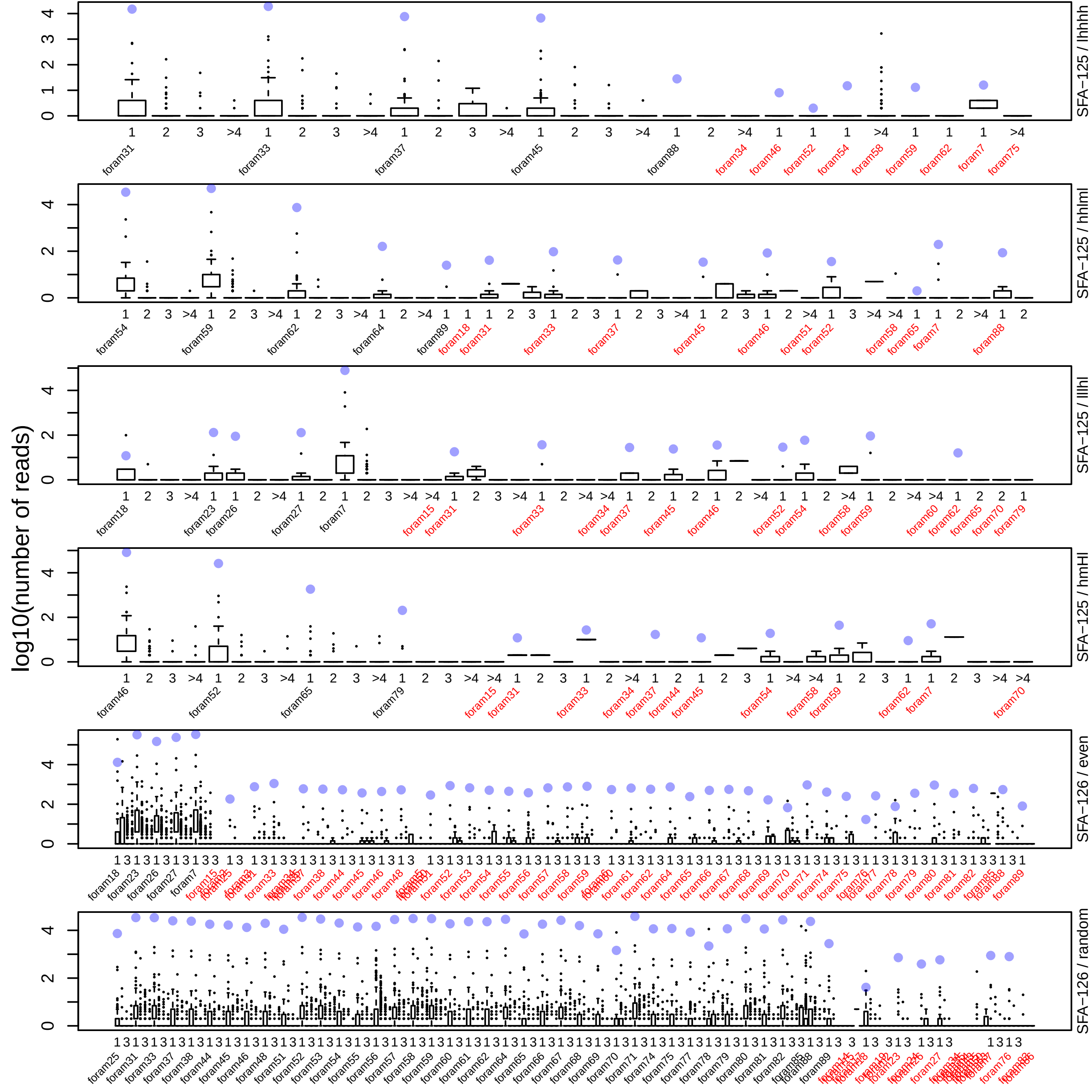


Foraminifera
(SFA-120)
a
> 132
b
> 330
c
> 463
0
66
132
198
264
330
396
463
529
595
661
727
793
859
926
Foraminifera
(SFA-121)
d
> 0
e
> 82
f
> 494
0
82
164
247
329
411
494
576
658
741
823
905
988
1070
1153
Eukaryota
(SFA-120)
g
> 124
h
> 311
i
> 498
0
62
124
187
249
311
374
436
498
561
623
685
748
810
873
Eukaryota
(SFA-121)
j
> 76
k
> 229
l
> 382
0
38
76
114
153
191
229
268
306
344
382
421
459
497
536

Supplementary Figure 7. Comparison of the 10 PCR product samples sequenced on two separate runs. Each of the 10 PCR products re-sequenced in either a LSD (SFA-123) or a Saturated Design (SFA-124) correspond to one sample and a clone (**a**: F1–B +15–R, **b**: F1–B + 15–T, **c**: F1–C + 15–R, **d**: F1–C + 15–S, **e**: F1–D + 15–S, **f**: F1–D + 15–T, **g**: F1–D + 15–V, **h**: F1–E + 15–U, **i**: F1–E + 15–V, **j**: F1–F + 15–U). The top row displays venn-euler diagrams of the assignments recovered in each sample (purple circle: SFA-123, green circle: SFA-124). The compositions of the re-sequenced sample in terms of relative read abundance are detailed in the vertical bars. For each sample, the correct clone used as template is not included in the bars. The read abundance of these clones are displayed in the pie charts (upper: SFA-123, below: SFA-124) relatively to all the other reads (black). The correct clones are boxed in the legend.

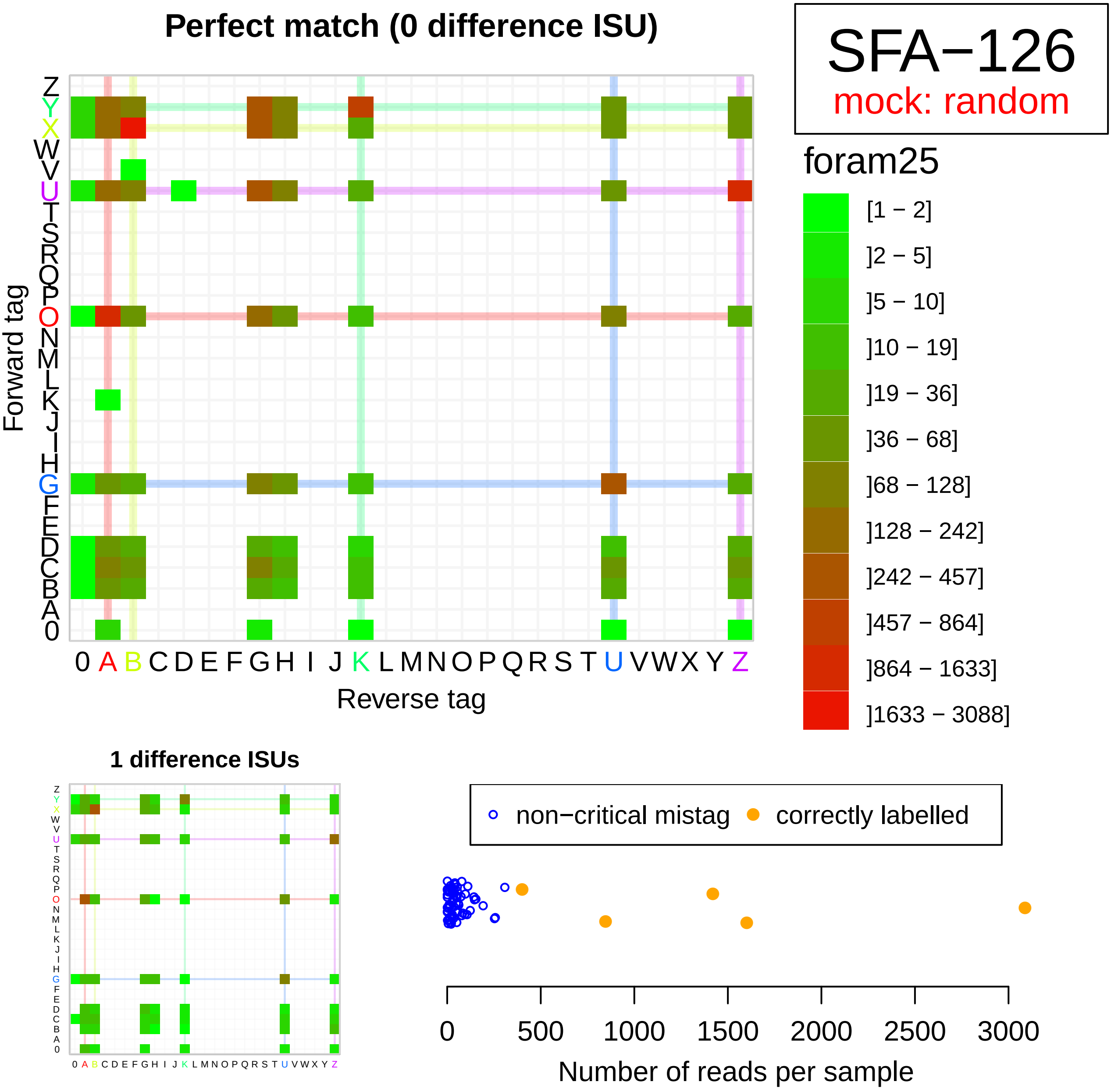

a
b
c
d
e
f
g
h
i
j
5 17 3
17 3
8 19 3
9 18 1
8 16 2
3 17 1
7 18 1
7 16 2
5 16 2
6 17 1
SFA−123
SFA−124
15−R − F1−B
15−T − F1−B
15−R − F1−C
15−S − F1−C
15−S − F1−D
15−T − F1−D
15−V − F1−D
15−U − F1−E
15−V − F1−E
15−U − F1−F
foram14
foram63
foram6
foram35
foram20
foram24
foram36
foram73
foram72
foram1
foram86
foram50
foram83
foram76
foram2
others

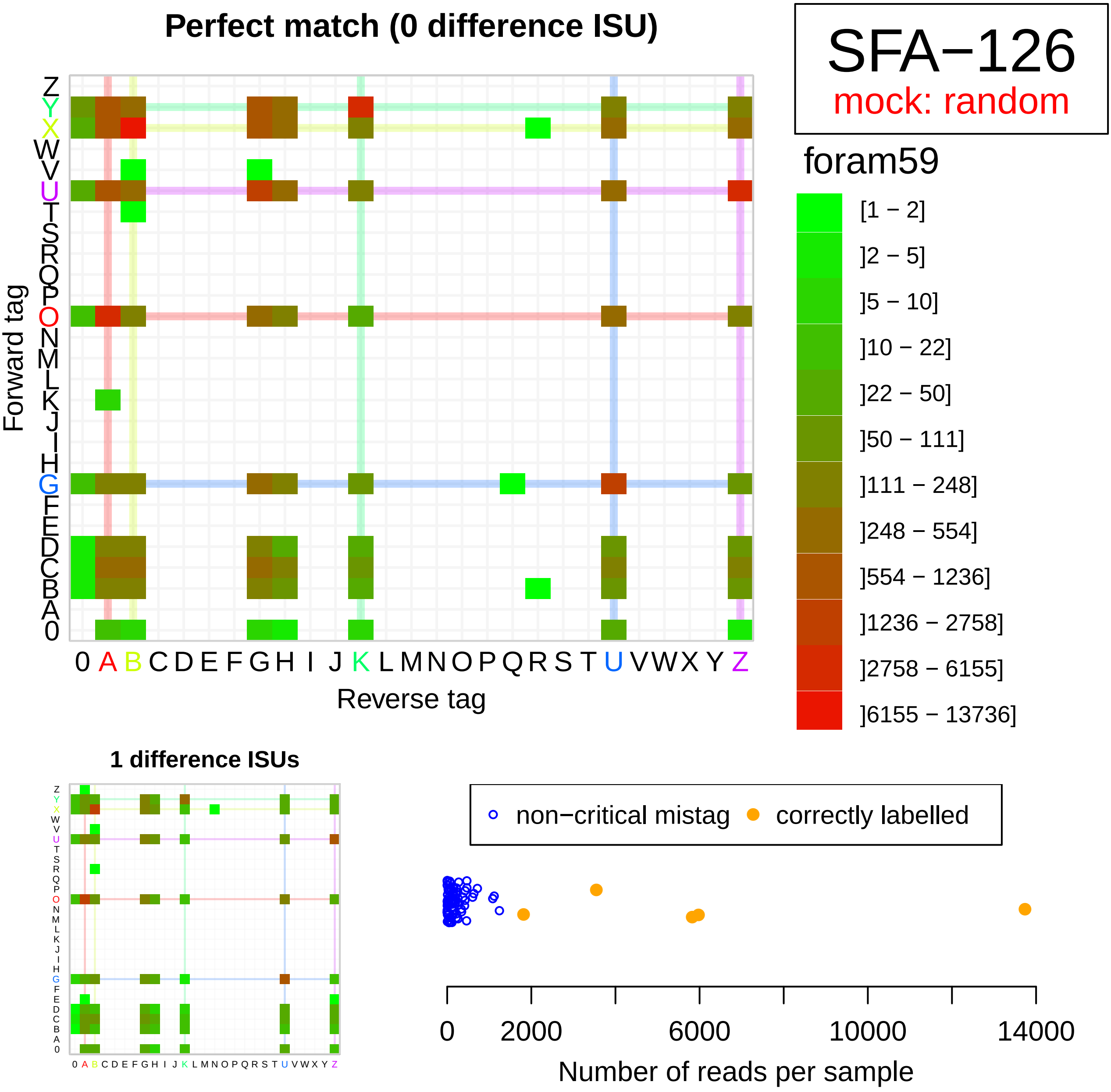
Supplementary Figure 8. Box plots of the number of reads per ISU assigned to a clone with 1, 2, 3 or more than 4 differences. The results are shown separately for each library and mock community. At the position of each clone name corresponds the group of ISU with the lowest number of difference(s) to this clone and the number of reads in the ISU perfectly matching this clone (blue dot). All the clones found in each mock community are displayed, including the expected clones (black) and the clones resulting from a critical mistagging event (red). The numbers of reads are displayed on a $\log_{10}$ scale.

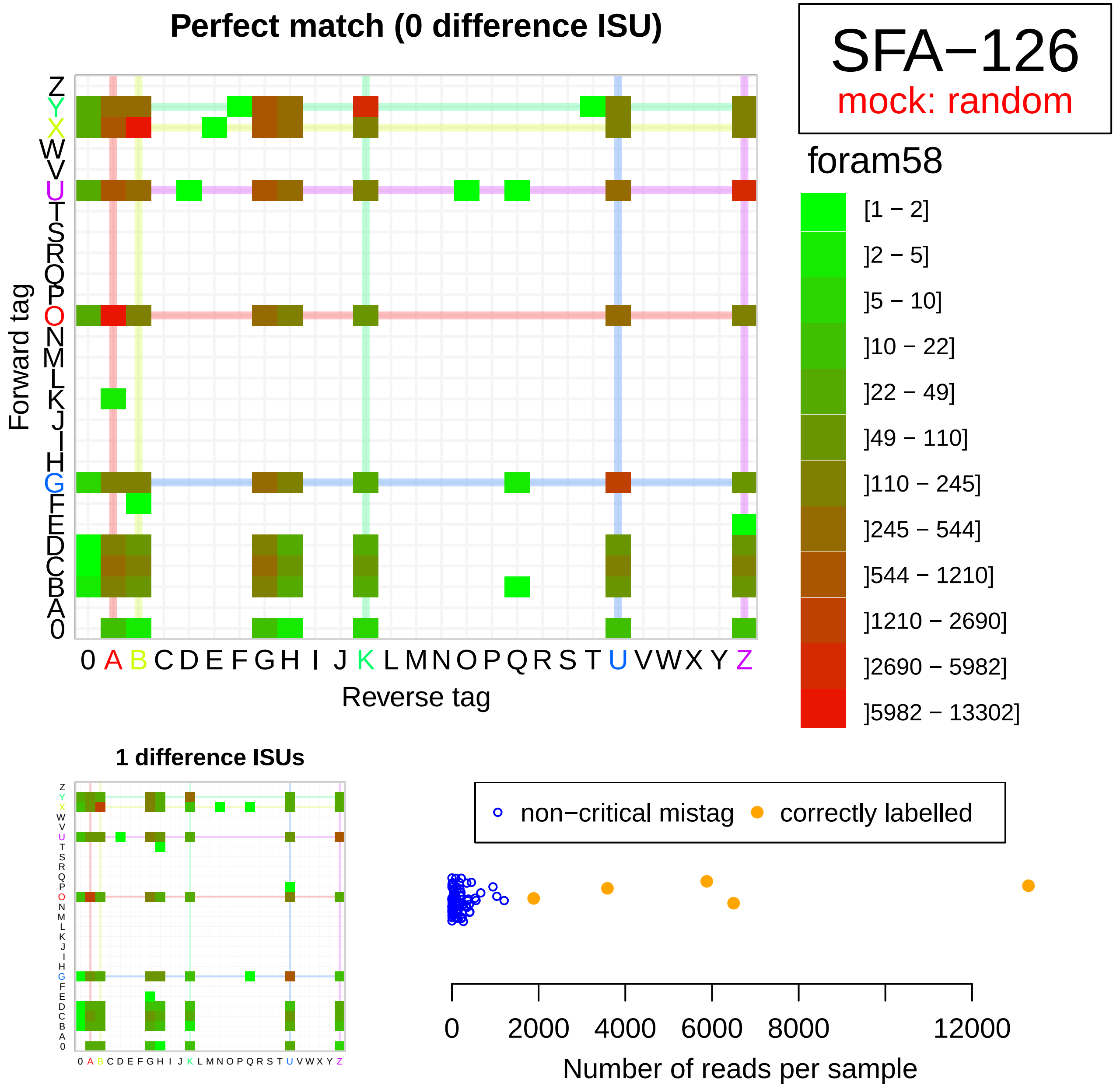


log10(number of reads)
SFA−125 / lhhhh
SFA−125 / hhlml
SFA−125 / lllhl
SFA−125 / hmHl
SFA−126 / even
SFA−126 / random
0
1
2
3
4
1 2 3 >4 1 2 3 >4 1 2 3 >4 1 2 3 >4 1 2 >4 1 1 1 >4 1 1 1 >4
foram31
foram33
foram37
foram45
foram88
foram34
foram46
foram52
foram54
foram58
foram59
foram62
foram7
foram75
1 2 3 >4 1 2 3 >4 1 2 3 >4 1 2 >4 1 1 1 2 3 1 2 3 1 2 3 >4 1 2 3 1 2 >4 1 3 >4 >4 1 1 2 >4 1 2
foram54
foram59
foram62
foram64
foram89
foram18
foram31
foram33
foram37
foram45
foram46
foram51
foram52
foram58
foram65
foram7
foram88
1 2 3 >4 1 1 2 >4 1 2 1 2 3 >4 >4 1 2 3 >4 1 2 >4 >4 1 2 1 2 1 2 >4 1 1 2 >4 1 2 >4 >4 1 2 2 1
foram18
foram23
foram26
foram27
foram7
foram15
foram31
foram33
foram34
foram37
foram45
foram46
foram52
foram54
foram58
foram59
foram60
foram62
foram65
foram70
foram79
1 2 3 >4 1 2 3 >4 1 2 3 >4 1 2 3 >4 >4 1 2 3 1 2 >4 1 2 1 2 3 1 >4 >4 1 2 3 1 1 2 3 >4 >4
foram46
foram52
foram65
foram79
foram15
foram31
foram33
foram34
foram37
foram44
foram45
foram54
foram58
foram59
foram62
foram7
foram70
foram18
foram23
foram26
foram27
foram7
foram15
foram25
foram31
foram33
foram34
foram37
foram38
foram44
foram45
foram46
foram48
foram50
foram51
foram52
foram53
foram54
foram55
foram56
foram57
foram58
foram59
foram60
foram61
foram62
foram64
foram65
foram66
foram67
foram68
foram69
foram70
foram71
foram74
foram75
foram76
foram77
foram78
foram79
foram80
foram81
foram82
foram85
foram88
foram89
foram25
foram31
foram33
foram37
foram38
foram44
foram45
foram46
foram48
foram51
foram52
foram53
foram54
foram55
foram56
foram57
foram58
foram59
foram60
foram61
foram62
foram64
foram65
foram66
foram67
foram68
foram69
foram70
foram71
foram74
foram75
foram77
foram78
foram79
foram80
foram81
foram82
foram85
foram88
foram89
foram15
foram17
foram18
foram19
foram23
foram24
foram26
foram27
foram7
foram76
foram83
foram86

Supplementary Figure 9. Mistagging cohorts for each individual sequence unit (ISU) assigned with less than 2 differences to each expected clone of each mock community sample. For each ISU, the distributions are shown in two separate heat maps in the framework of their correct samples. One heat map shows the ISU perfectly corresponding to the original clone sequence (large, top heat map) and all ISUs matching this sequence with 1 difference (lower left). The numbers of reads are indicated according to a green-to-red scale. Each clone name is indicated above this scale. The tagged primer pairs used for the replicate PCRs of the library indicated in the upper-right box are colored per combination. The mock community into which the clone is expected is indicated in the box in red. The relative abundance of each clone belonging to the mock communities of SFA-125 are indicated by a red letter in parentheses (“l”: low; “m”: medium; “h”, high and “H”: very high relative abundances). The numbers of reads per correct and non-critical mistag ISU are indicated on the lower right panel. Further details on mock community compositions are provided in Supplementary Table 2.

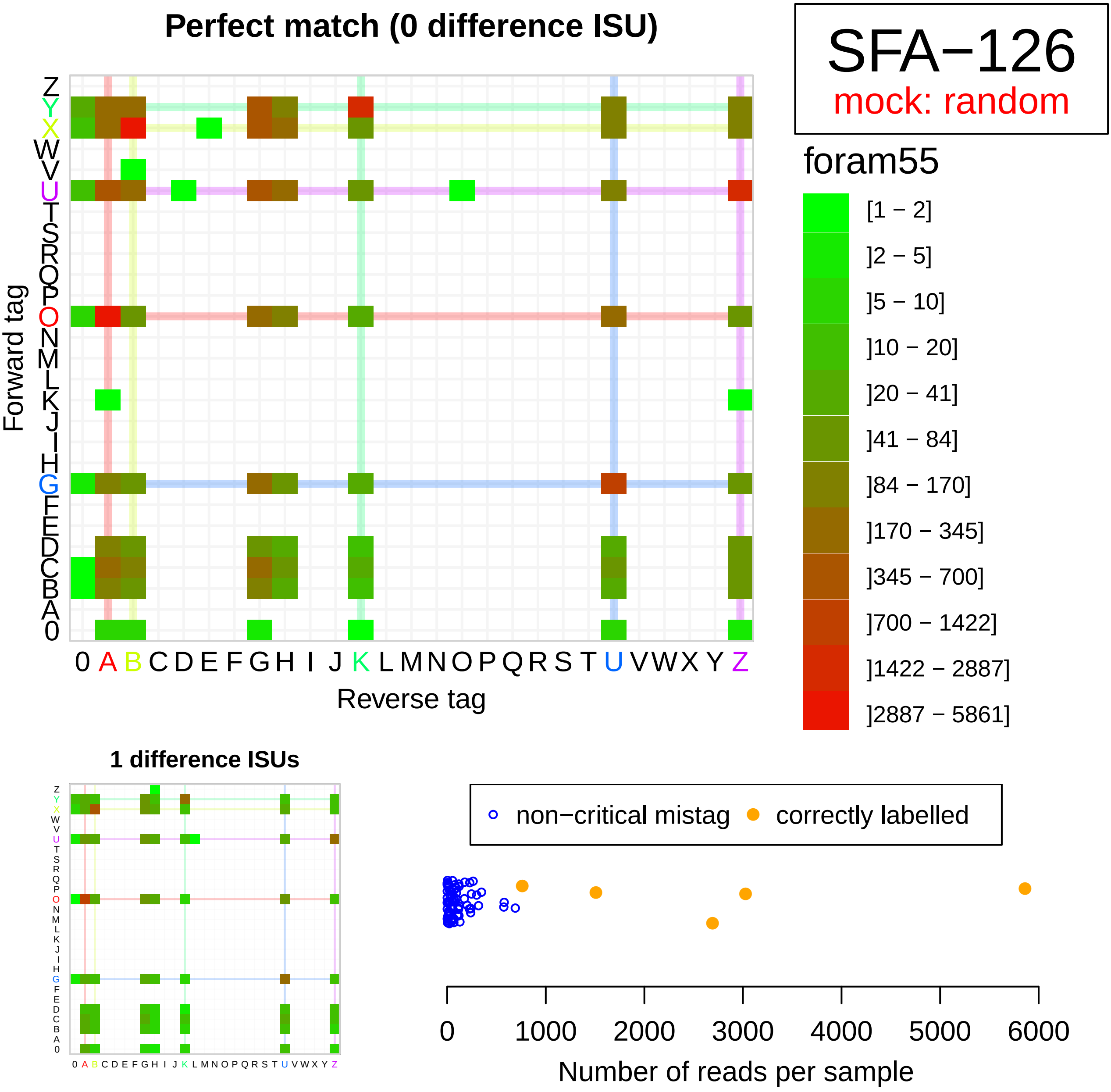

Perfect match (0 difference ISU)
SFA−126
mock: random
foram25
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 19]
]19 − 36]
]36 − 68]
]68 − 128]
]128 − 242]
]242 − 457]
]457 − 864]
]864 − 1633]
]1633 − 3088]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
○ non−critical mistag
● correctly labelled
0
500
1000
1500
2000
2500
3000
Number of reads per sample


Perfect match (0 difference ISU)
SFA−126
mock: random
foram59
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 22]
]22 − 50]
]50 − 111]
]111 − 248]
]248 − 554]
]554 − 1236]
]1236 − 2758]
]2758 − 6155]
]6155 − 13736]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
non−critical mistag
correctly labelled
0
2000
6000
10000
14000
Number of reads per sample


SFA−126
mock: random
Perfect match (0 difference ISU)
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
Reverse tag
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
foram58
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 22]
]22 − 49]
]49 − 110]
]110 − 245]
]245 − 544]
]544 − 1210]
]1210 − 2690]
]2690 − 5982]
]5982 − 13302]
1 difference ISUs
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
non−critical mistag
correctly labelled
0
2000
4000
6000
8000
12000
Number of reads per sample

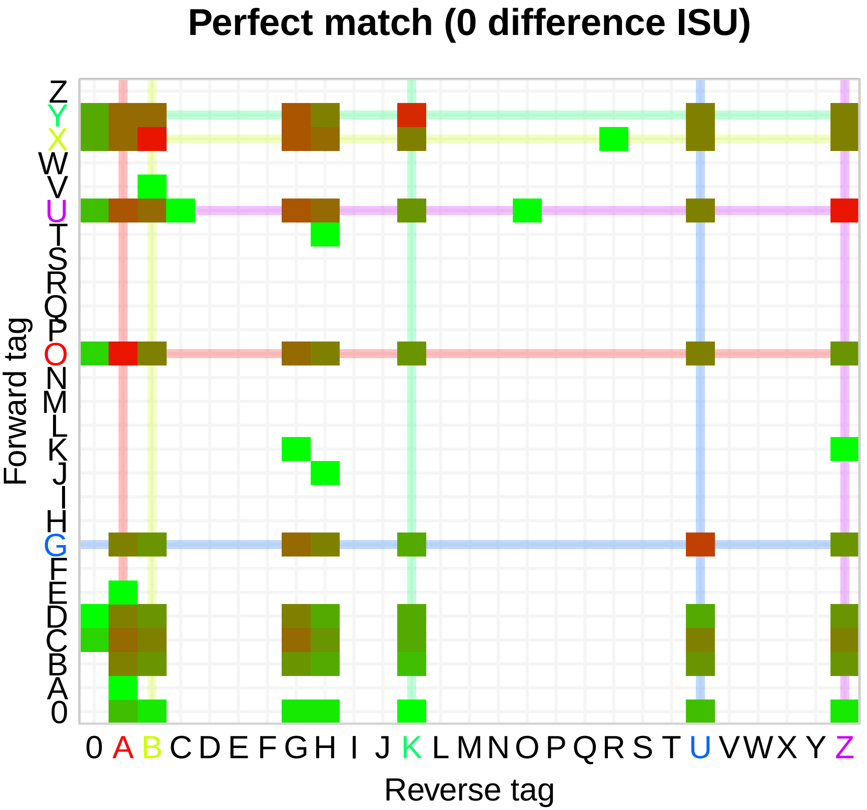
SFA−126

mock: random



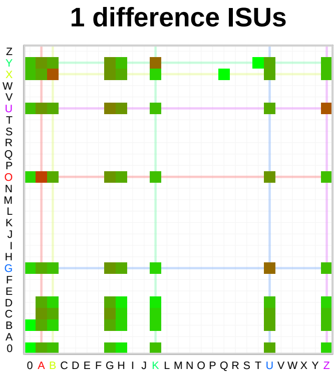
SFA−126

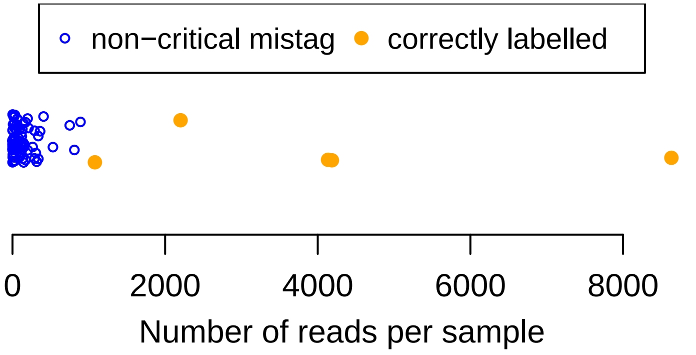
mock: random



foram54

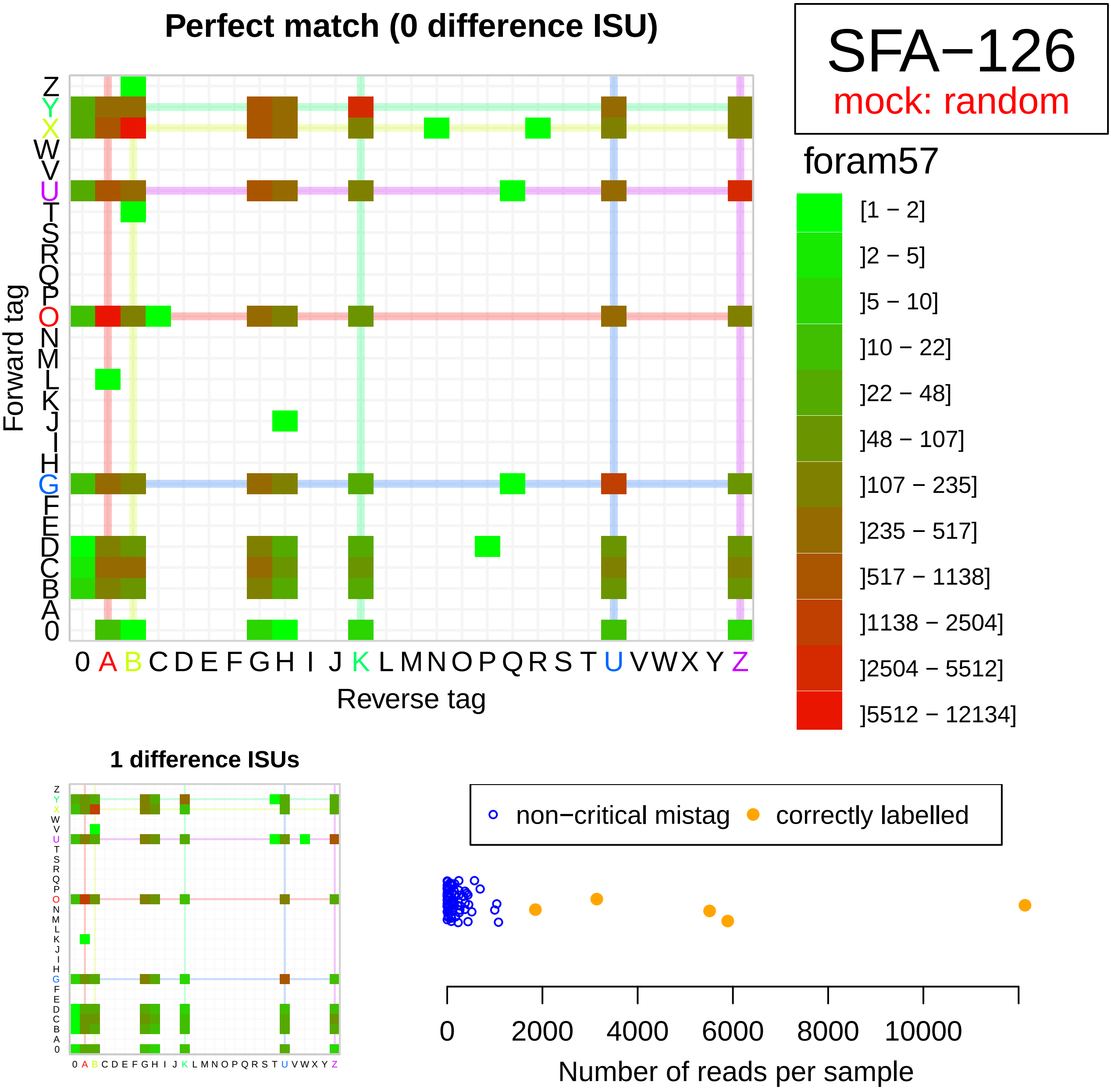
- 🟩 [1 − 2]
- 🟩]2 − 5]
- 🟩]5 − 10]
- 🟩]10 − 21]
- 🟩]21 − 45]
- 🟩]45 − 95]
- 🟫]95 − 202]
- 🟫]202 − 428]
- 🟫]428 − 907]
- 🟥]907 − 1922]
- 🟥]1922 − 4073]
- 🟥]4073 − 8634]





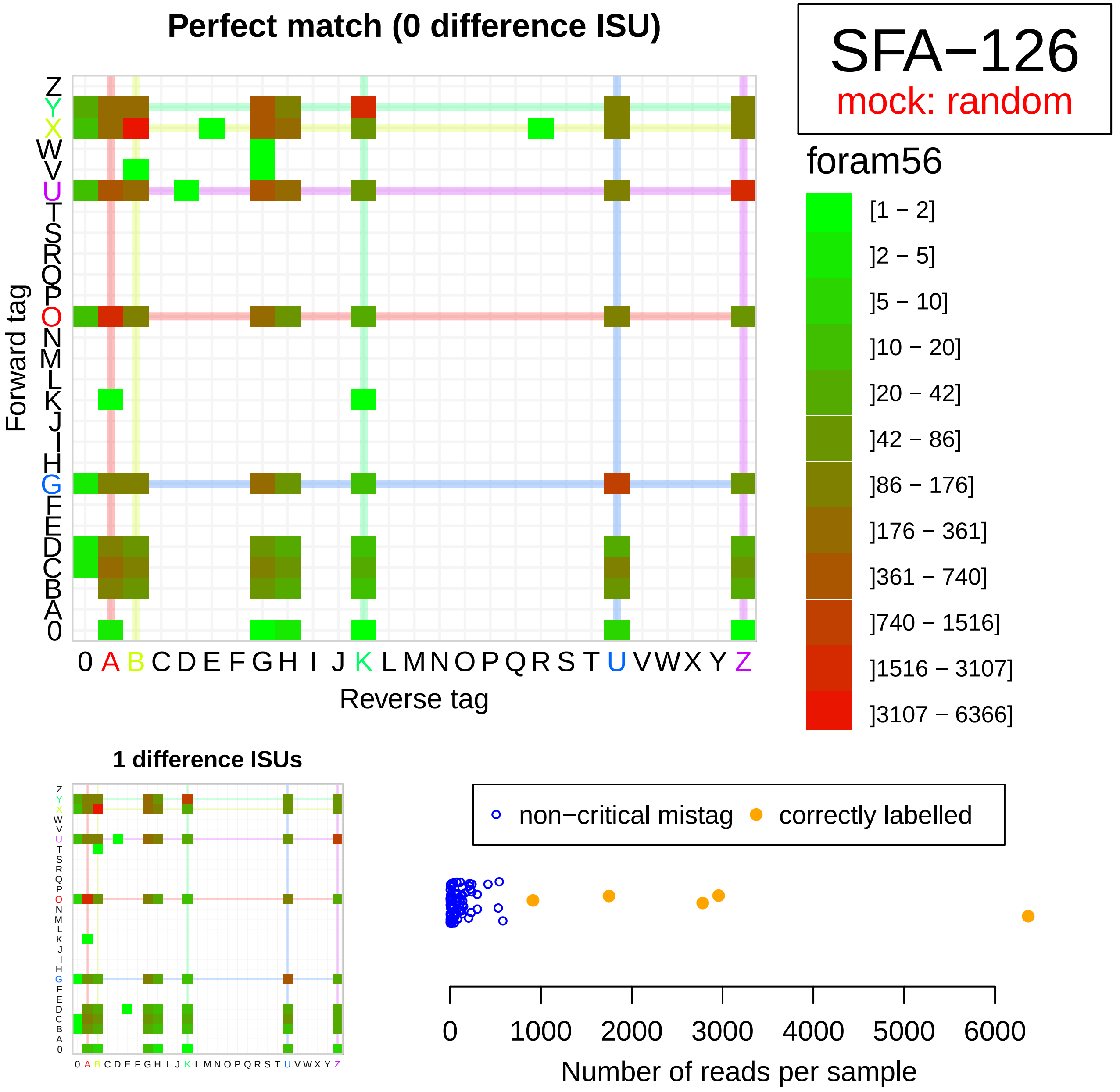
SFA−126

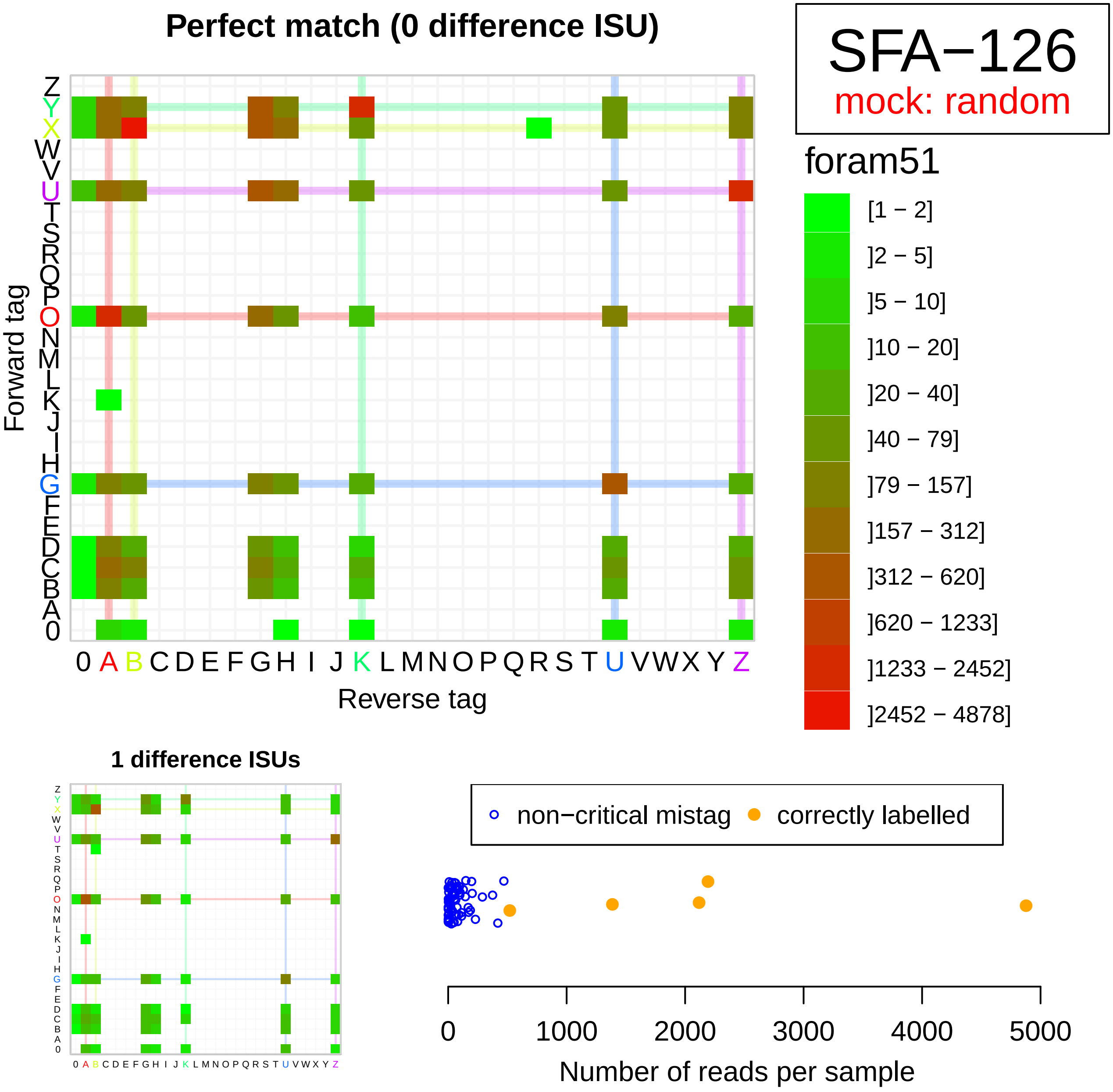
mock: random

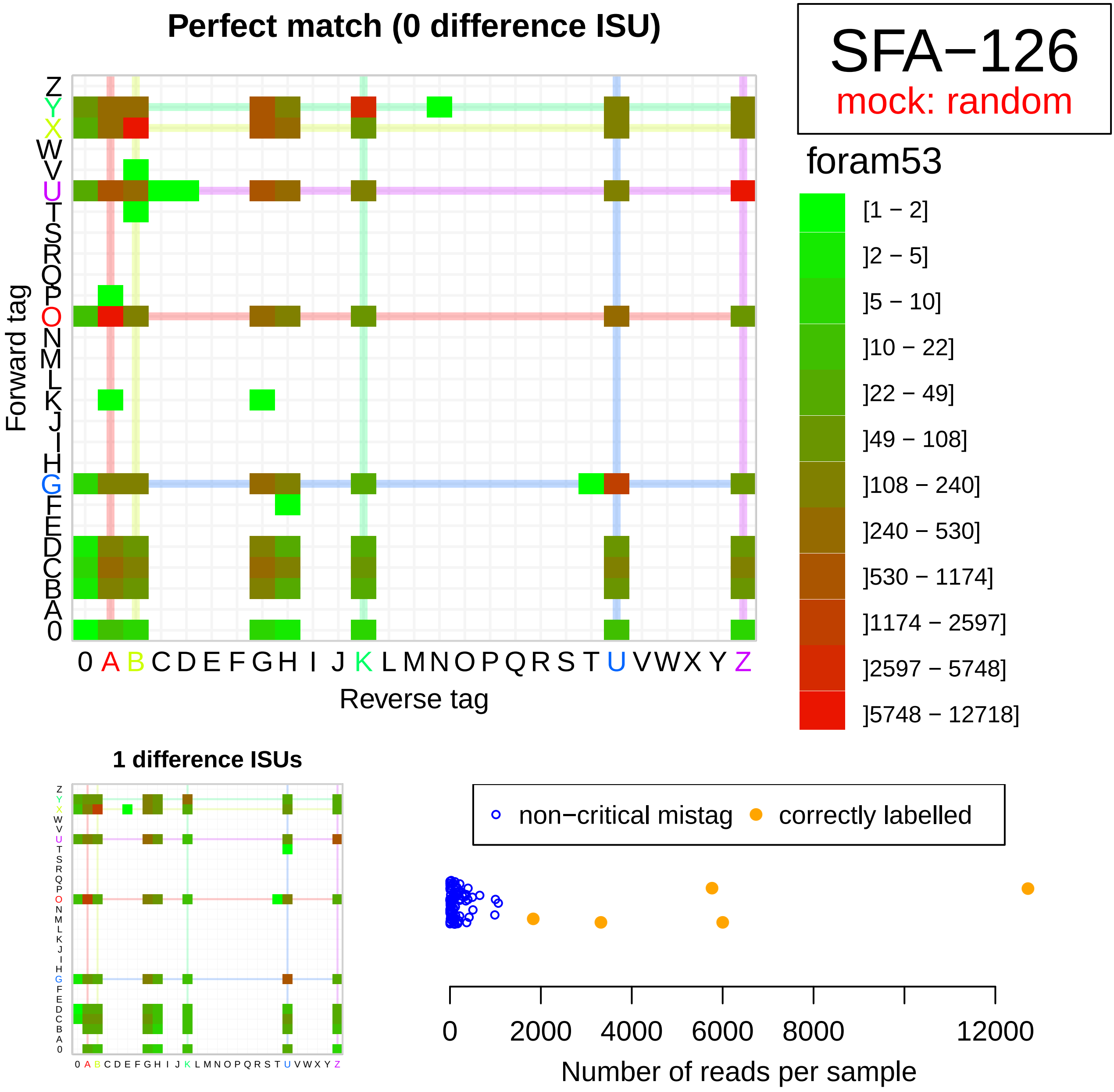


Perfect match (0 difference ISU)

1 difference ISUs


Perfect match (0 difference ISU)
SFA−126
mock: random
foram56
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 20]
]20 − 42]
]42 − 86]
]86 − 176]
]176 − 361]
]361 − 740]
]740 − 1516]
]1516 − 3107]
]3107 − 6366]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
non−critical mistag
correctly labelled
0
1000
2000
3000
4000
5000
6000
Number of reads per sample

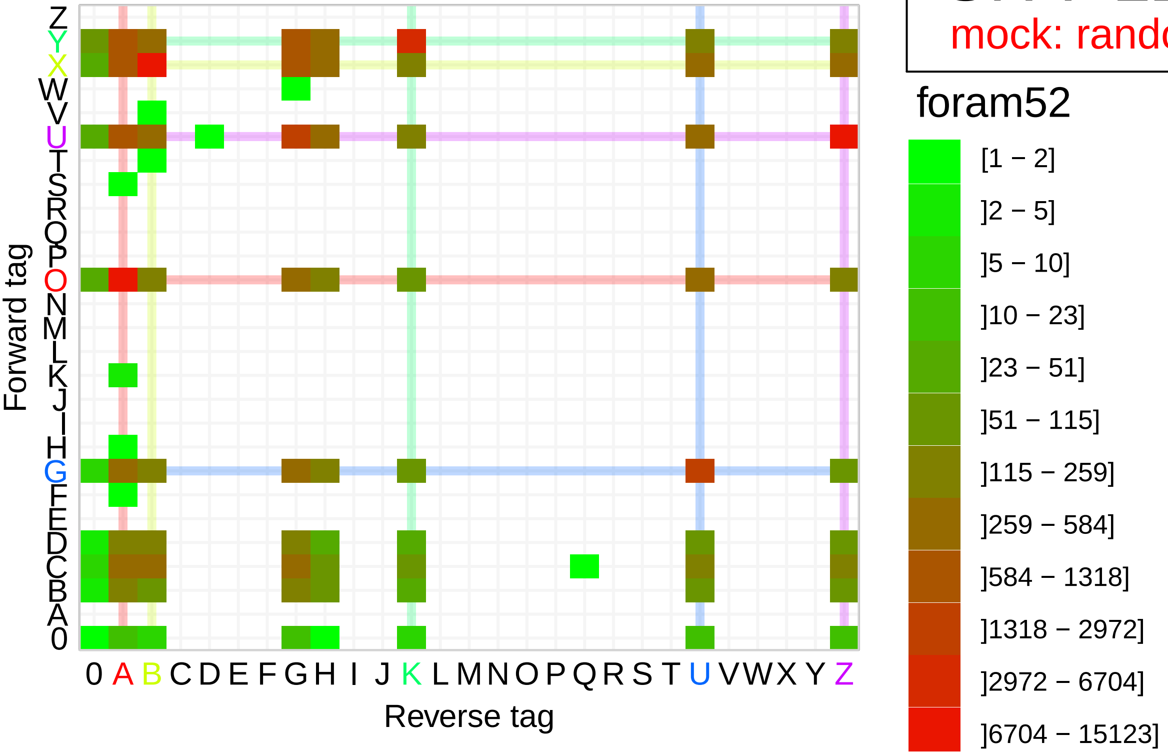

Perfect match (0 difference ISU)
SFA−126
mock: random
foram51
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 20]
]20 − 40]
]40 − 79]
]79 − 157]
]157 − 312]
]312 − 620]
]620 − 1233]
]1233 − 2452]
]2452 − 4878]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
non−critical mistag
correctly labelled
0
1000
2000
3000
4000
5000
Number of reads per sample


Perfect match (0 difference ISU)
SFA−126
mock: random
foram53
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 22]
]22 − 49]
]49 − 108]
]108 − 240]
]240 − 530]
]530 − 1174]
]1174 − 2597]
]2597 − 5748]
]5748 − 12718]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
non−critical mistag
correctly labelled
0
2000
4000
6000
8000
12000
Number of reads per sample

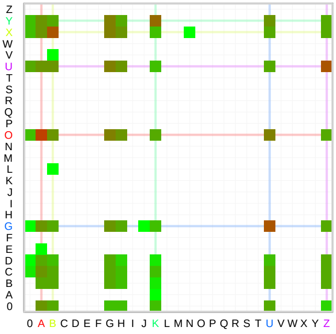
SFA−126

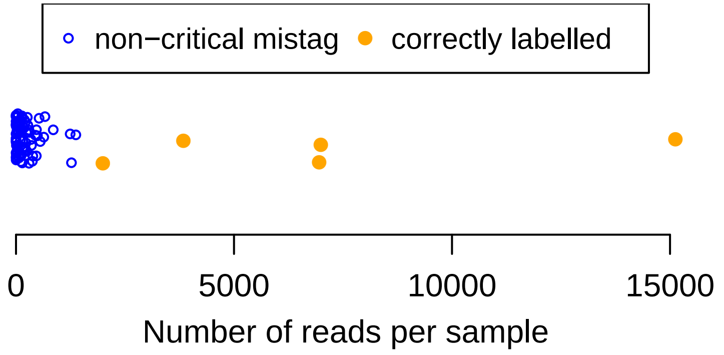
mock: random

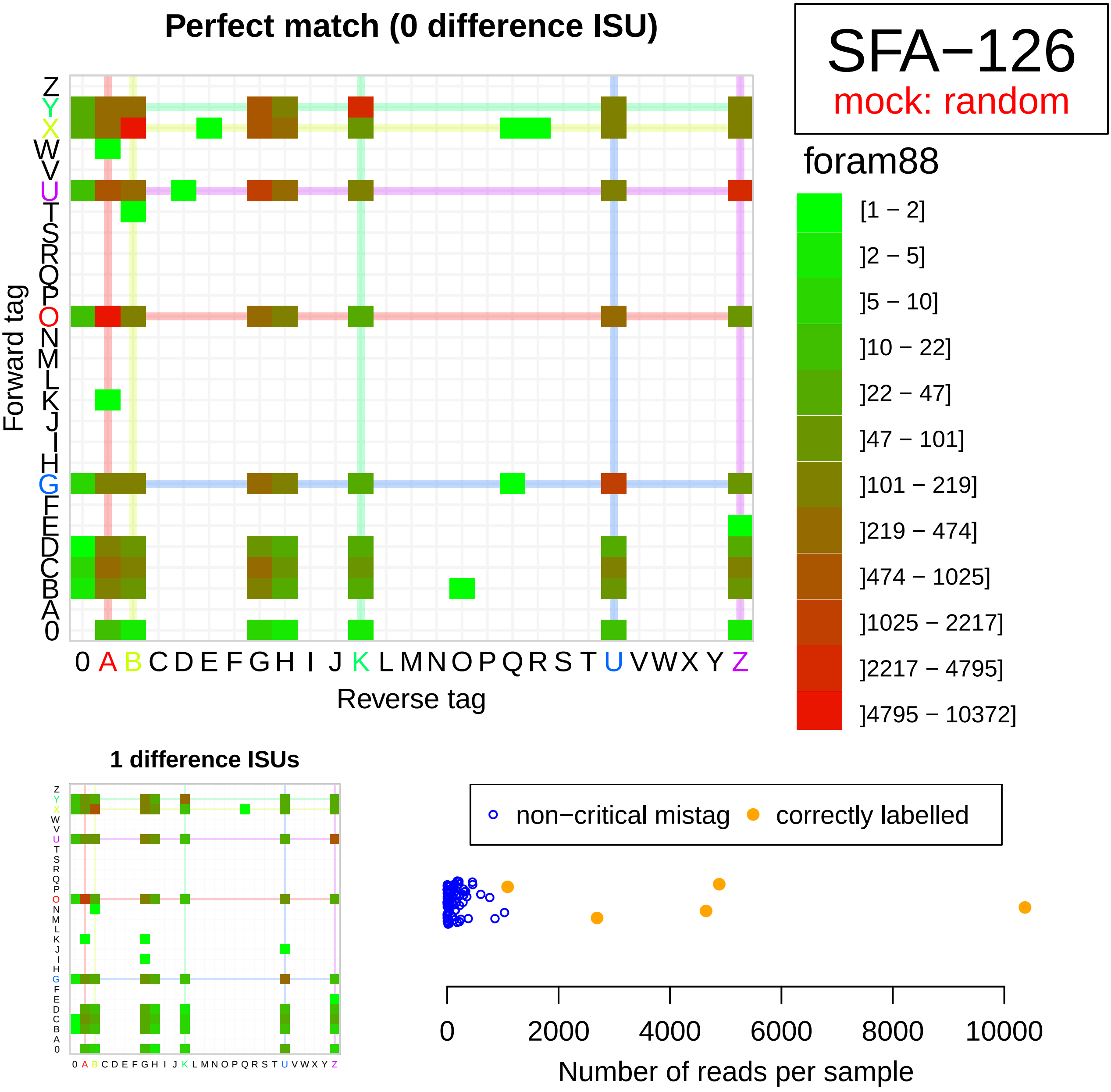
Perfect match (0 difference ISU)

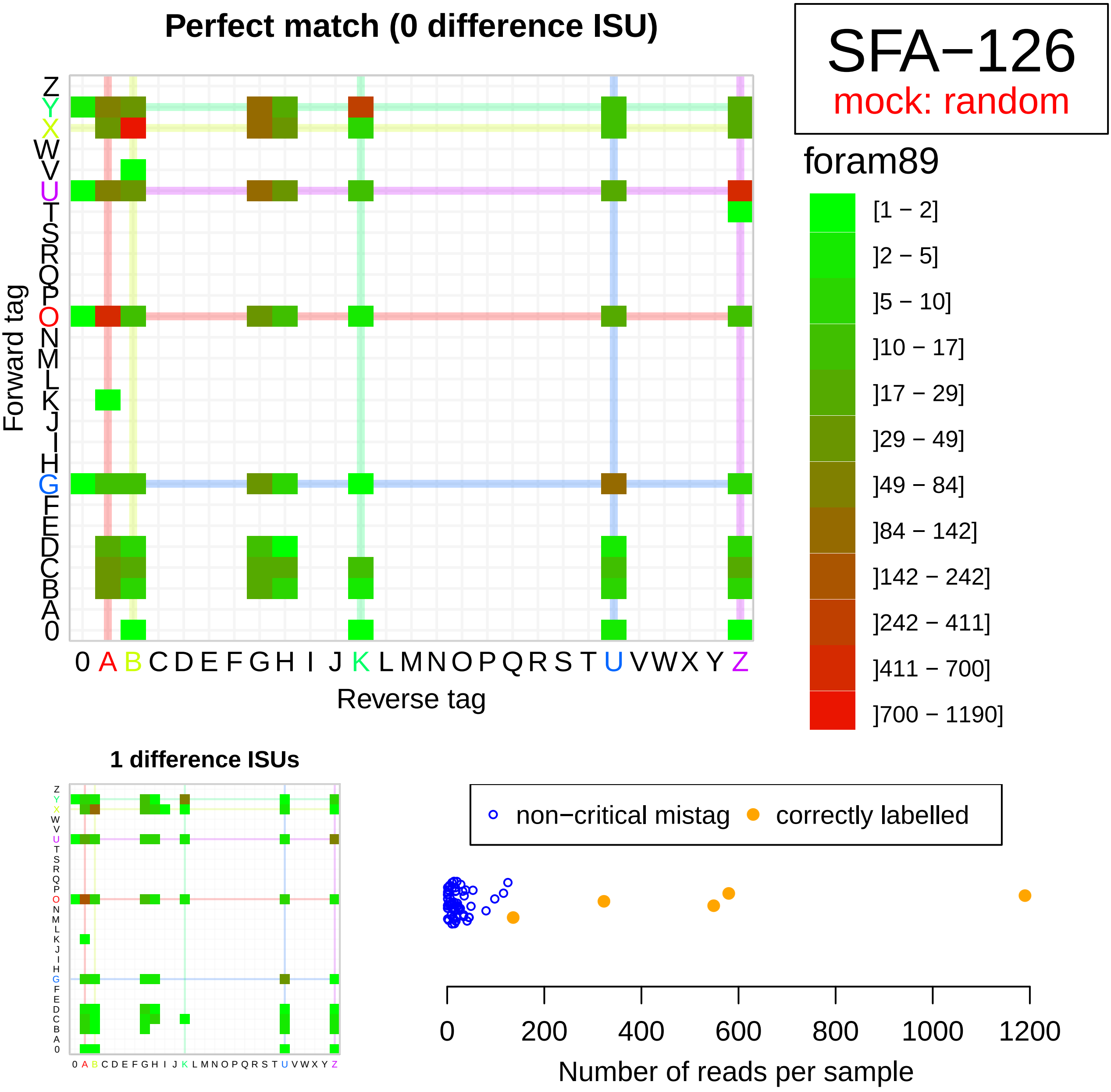


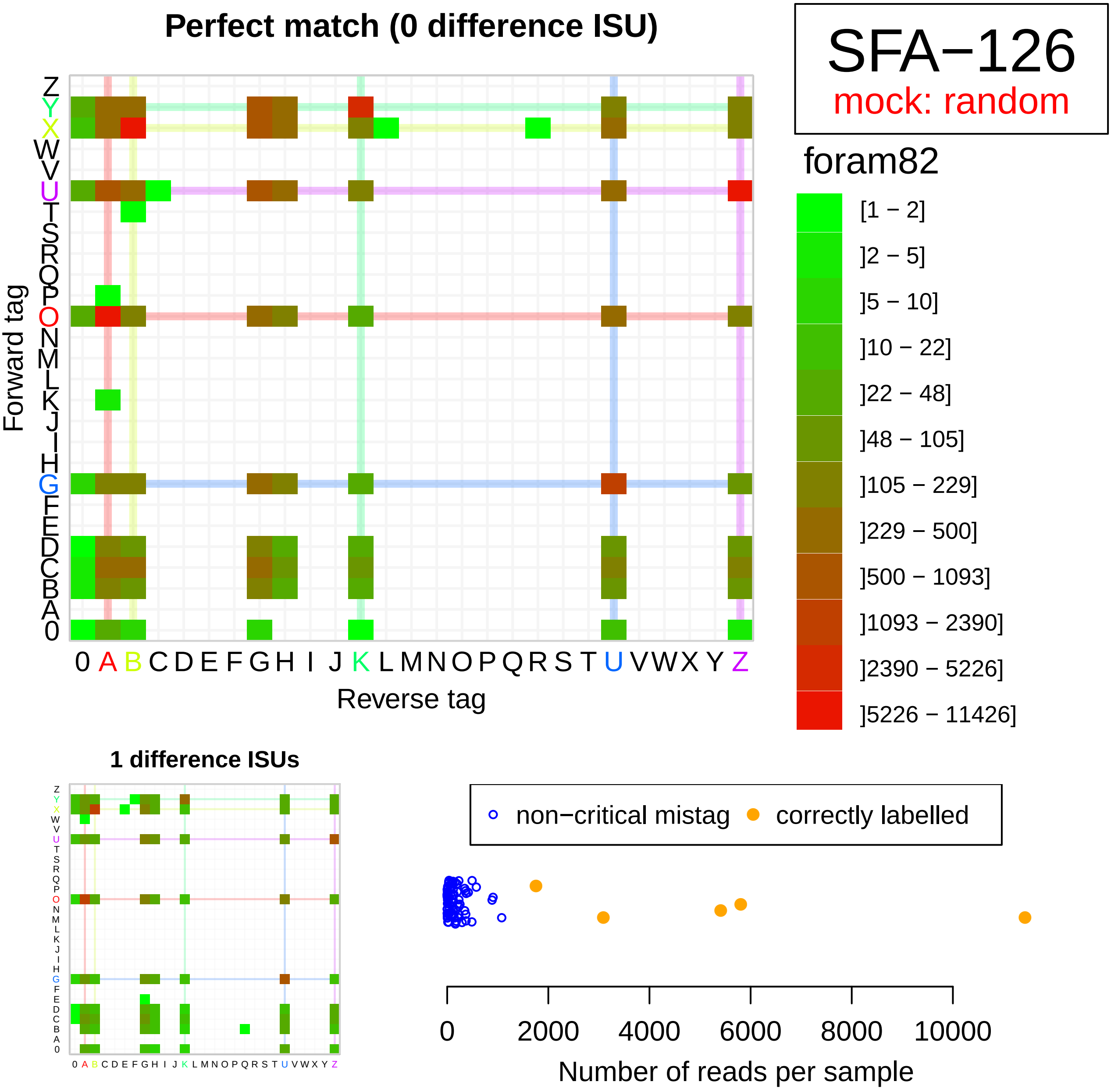
1 difference ISUs

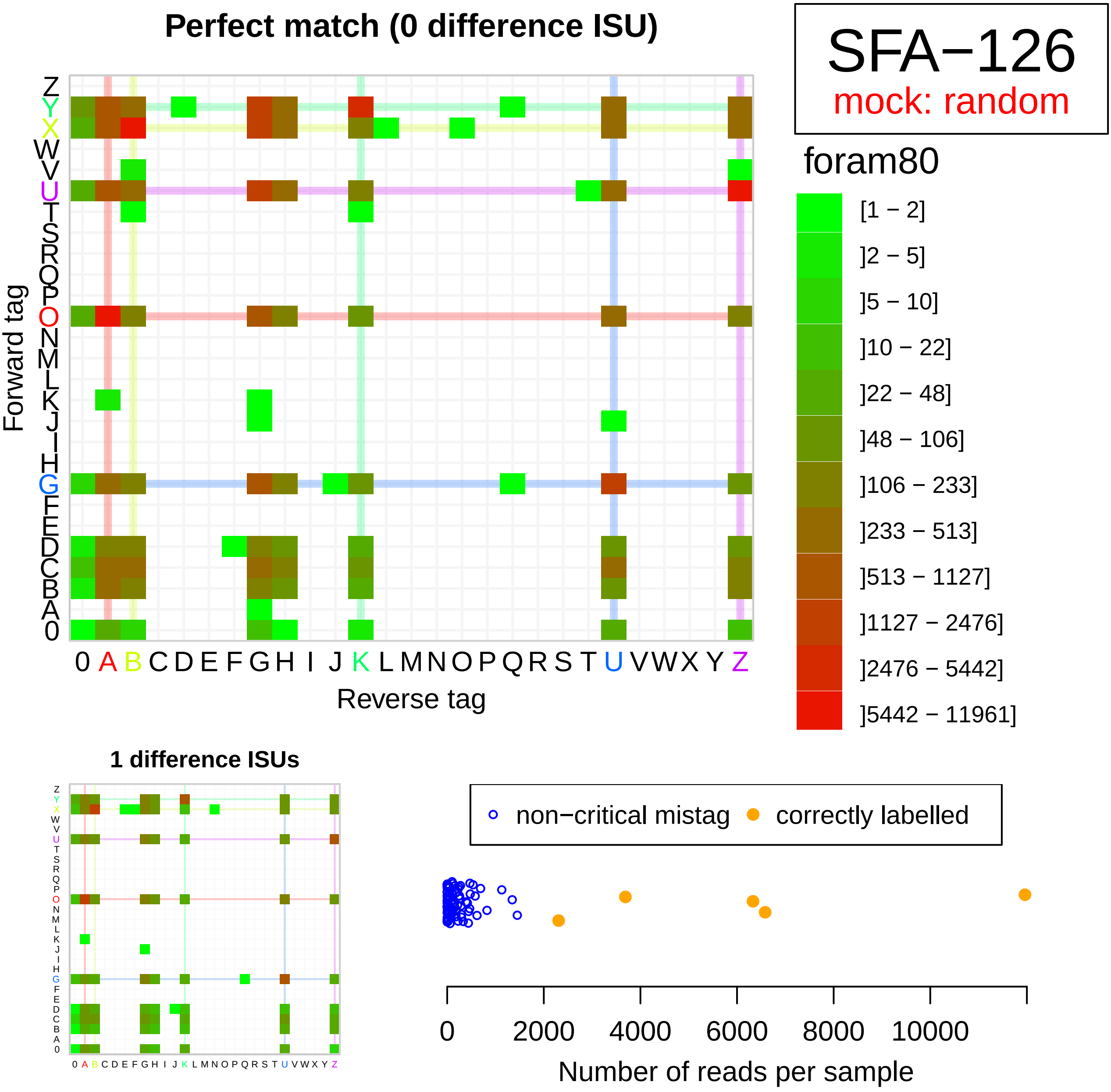


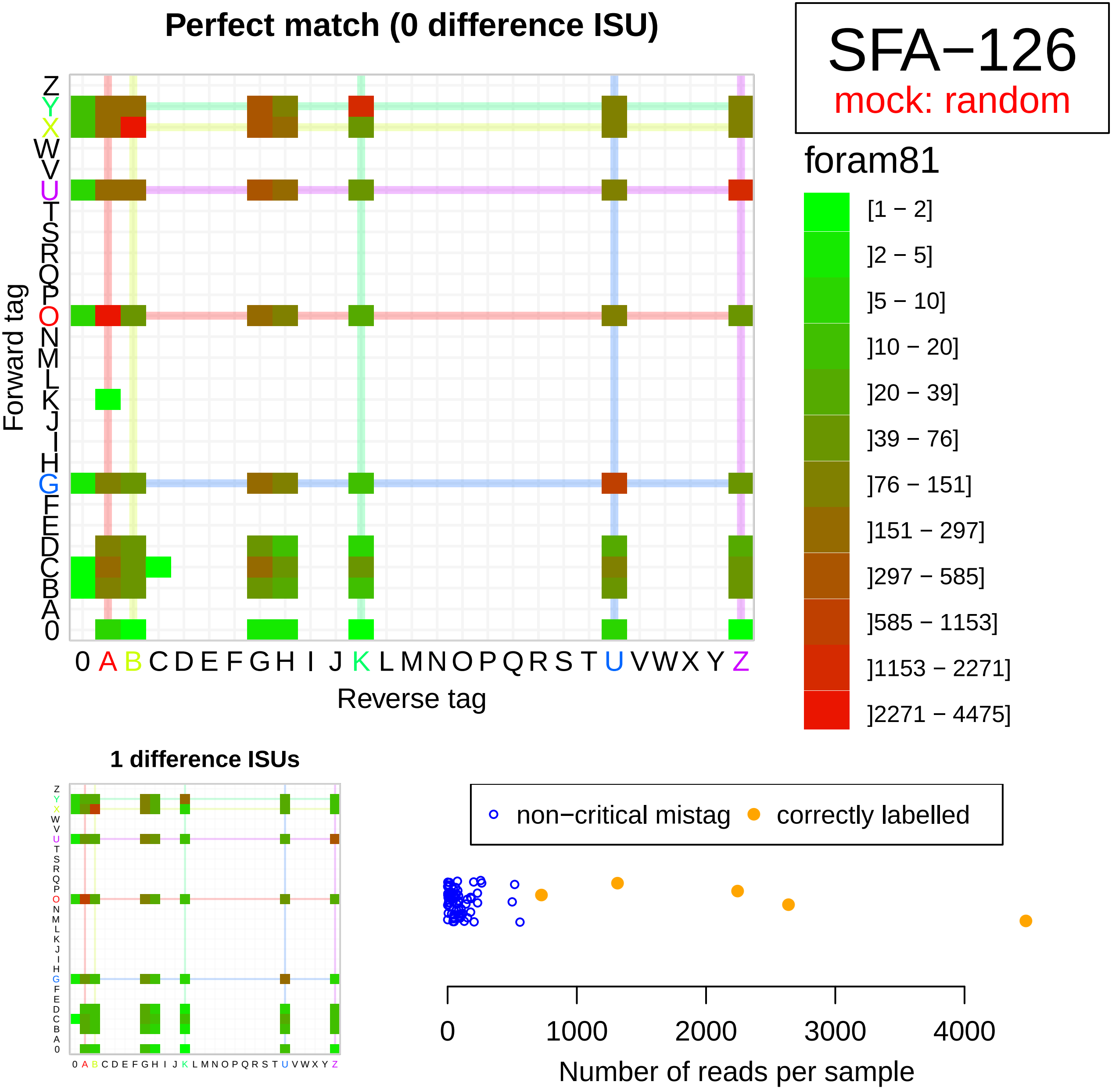



Perfect match (0 difference ISU)
SFA−126
mock: random
foram88
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 22]
]22 − 47]
]47 − 101]
]101 − 219]
]219 − 474]
]474 − 1025]
]1025 − 2217]
]2217 − 4795]
]4795 − 10372]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
non−critical mistag
correctly labelled
0
2000
4000
6000
8000
10000
Number of reads per sample


Perfect match (0 difference ISU)
SFA−126
mock: random
foram89
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 17]
]17 − 29]
]29 − 49]
]49 − 84]
]84 − 142]
]142 − 242]
]242 − 411]
]411 − 700]
]700 − 1190]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
non−critical mistag
correctly labelled
0
200
400
600
800
1000
1200
Number of reads per sample


Perfect match (0 difference ISU)
SFA−126
mock: random
foram82
Forward tag
Reverse tag
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 22]
]22 − 48]
]48 − 105]
]105 − 229]
]229 − 500]
]500 − 1093]
]1093 − 2390]
]2390 − 5226]
]5226 − 11426]
1 difference ISUs
non−critical mistag
correctly labelled
0
2000
4000
6000
8000
10000
Number of reads per sample


Perfect match (0 difference ISU)
SFA−126
mock: random
foram80
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 22]
]22 − 48]
]48 − 106]
]106 − 233]
]233 − 513]
]513 − 1127]
]1127 − 2476]
]2476 − 5442]
]5442 − 11961]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
non−critical mistag
correctly labelled
0
2000
4000
6000
8000
10000
Number of reads per sample

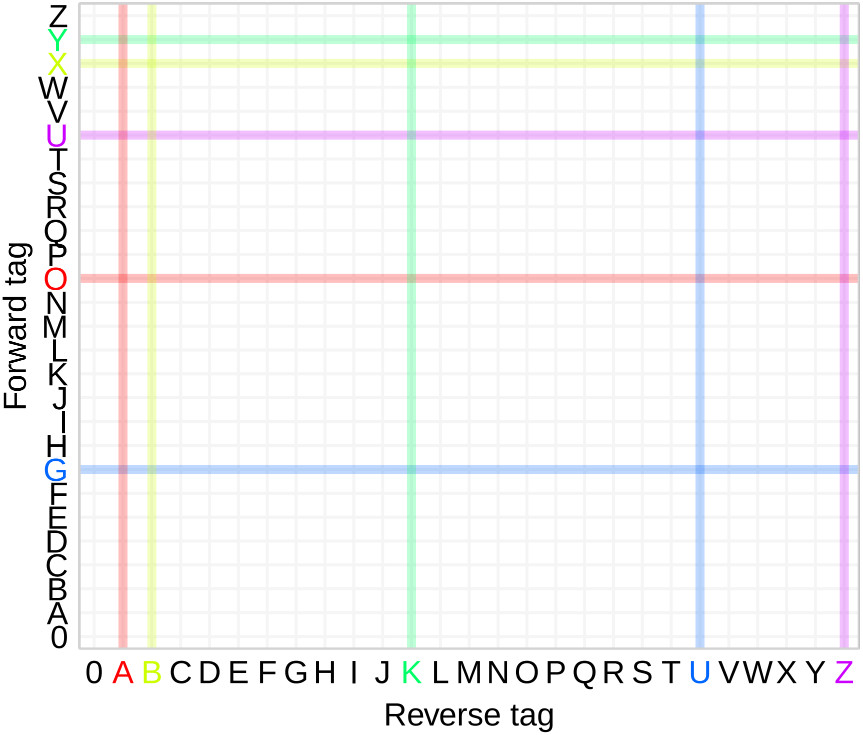

Perfect match (0 difference ISU)
SFA−126
mock: random
foram81
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 20]
]20 − 39]
]39 − 76]
]76 − 151]
]151 − 297]
]297 − 585]
]585 − 1153]
]1153 − 2271]
]2271 − 4475]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
non−critical mistag
correctly labelled
0
1000
2000
3000
4000
Number of reads per sample

Perfect match (0 difference ISU)

SFA−126

mock: random

foram85

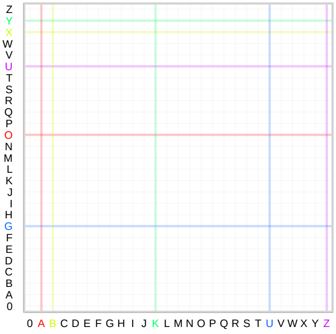


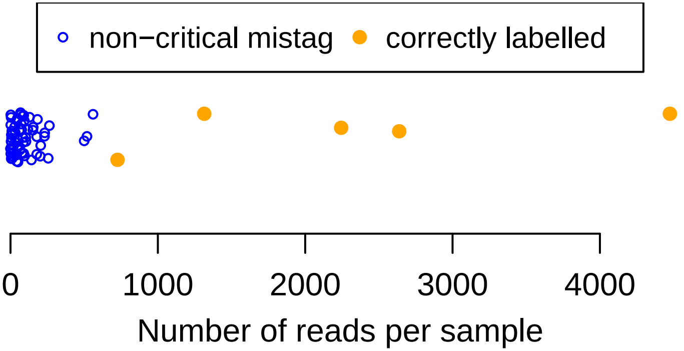
🟩 [1 − 2]

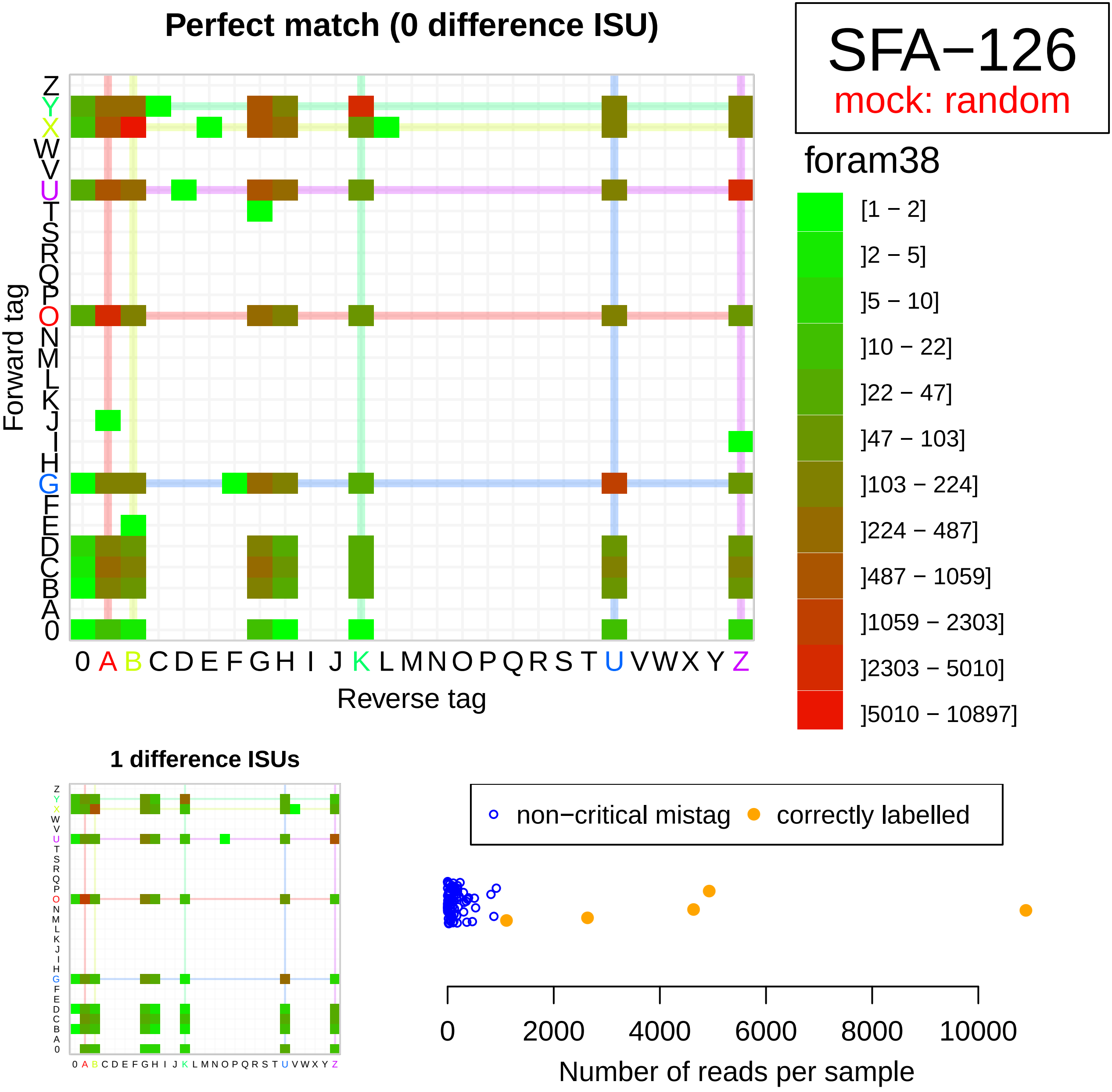
🟩]2 − 5]

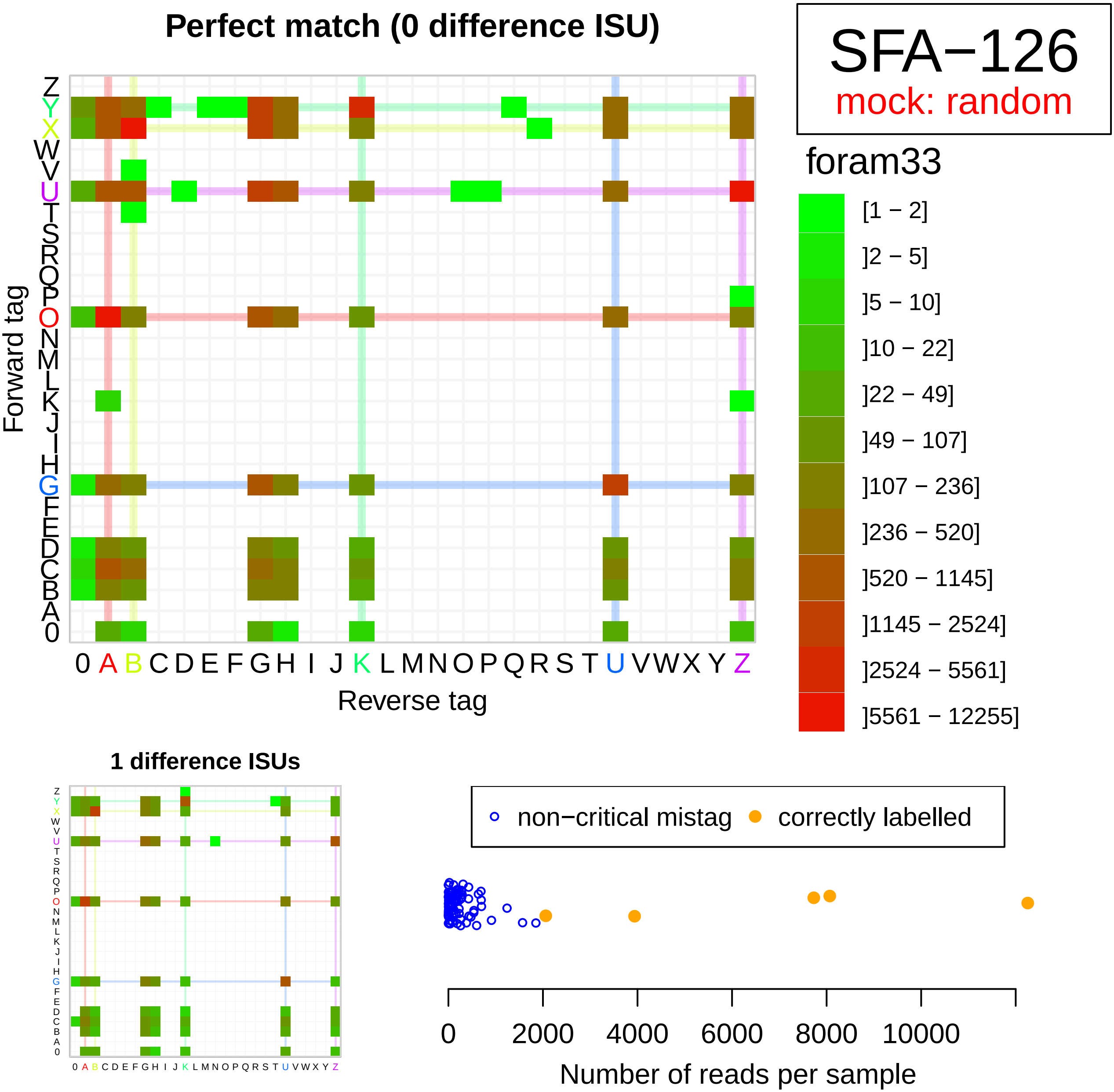
🟫]5 − 10]

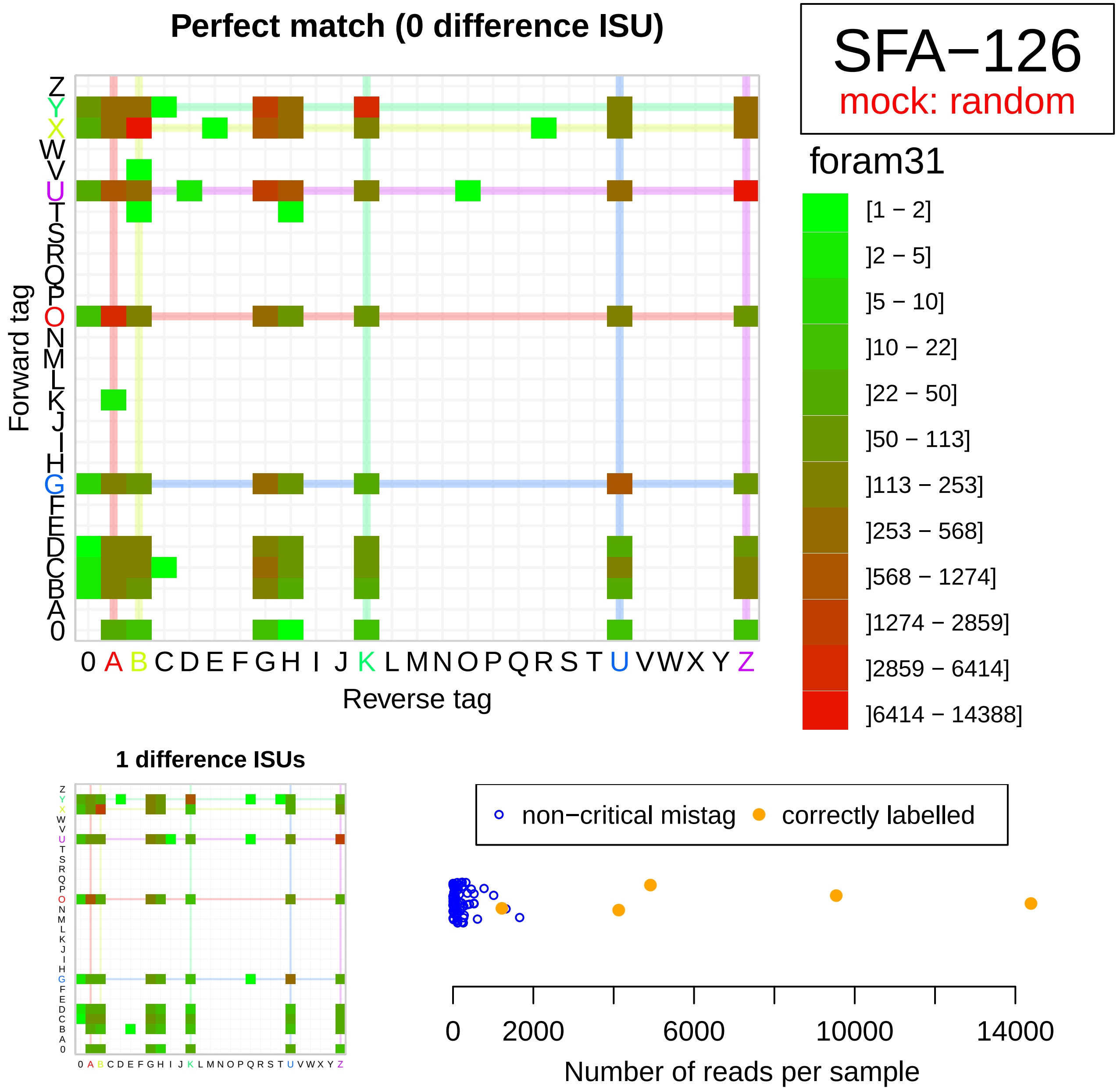
1 difference ISUs

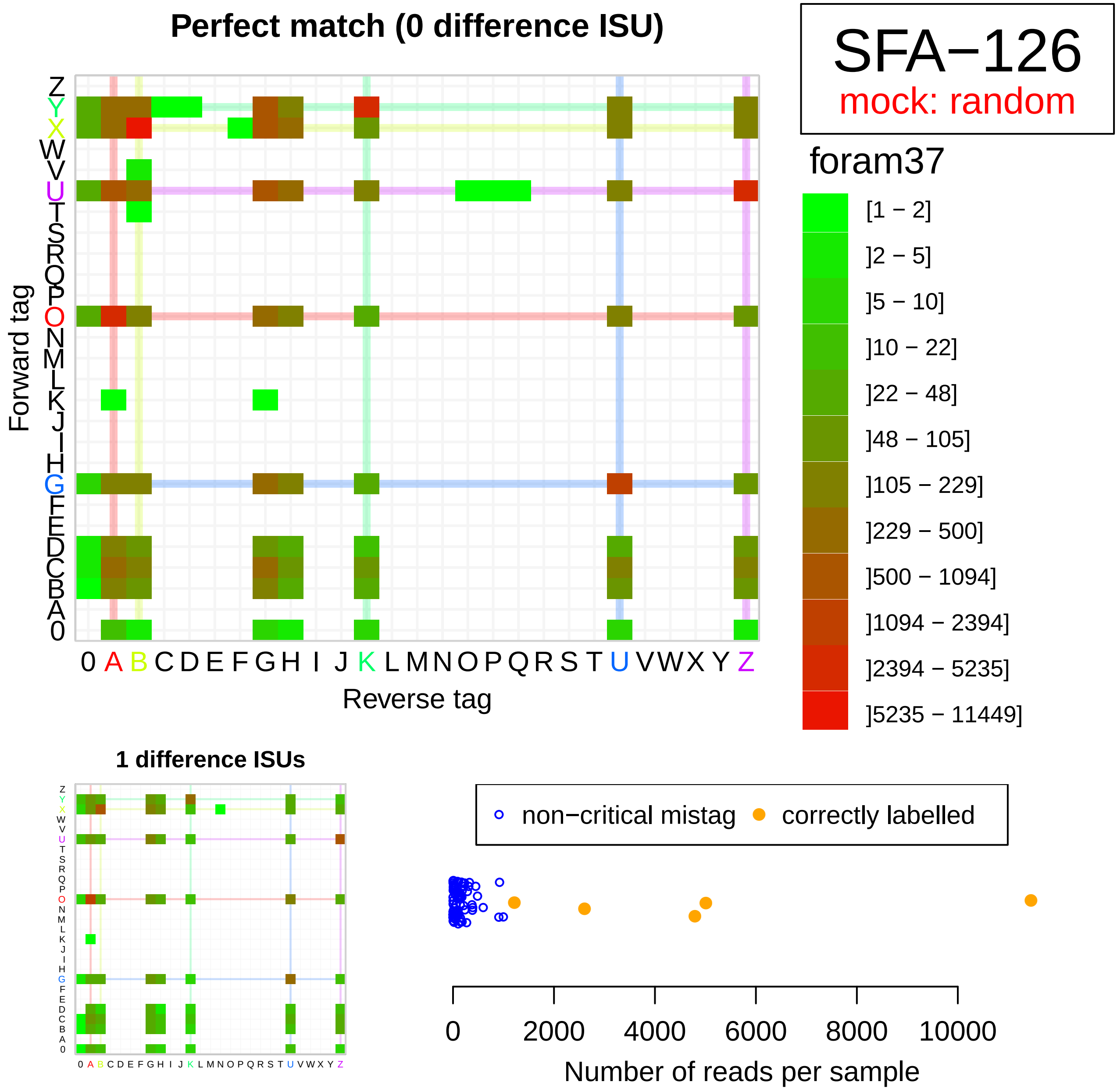





Perfect match (0 difference ISU)
SFA−126
mock: random
foram38
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 22]
]22 − 47]
]47 − 103]
]103 − 224]
]224 − 487]
]487 − 1059]
]1059 − 2303]
]2303 − 5010]
]5010 − 10897]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
non−critical mistag
correctly labelled
0
2000
4000
6000
8000
10000
Number of reads per sample


Perfect match (0 difference ISU)
SFA−126
mock: random
foram33
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 22]
]22 − 49]
]49 − 107]
]107 − 236]
]236 − 520]
]520 − 1145]
]1145 − 2524]
]2524 − 5561]
]5561 − 12255]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
non−critical mistag
correctly labelled
0
2000
4000
6000
8000
10000
Number of reads per sample


Perfect match (0 difference ISU)
SFA−126
mock: random
foram31
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 22]
]22 − 50]
]50 − 113]
]113 − 253]
]253 − 568]
]568 − 1274]
]1274 − 2859]
]2859 − 6414]
]6414 − 14388]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
non−critical mistag
correctly labelled
0
2000
6000
10000
14000
Number of reads per sample

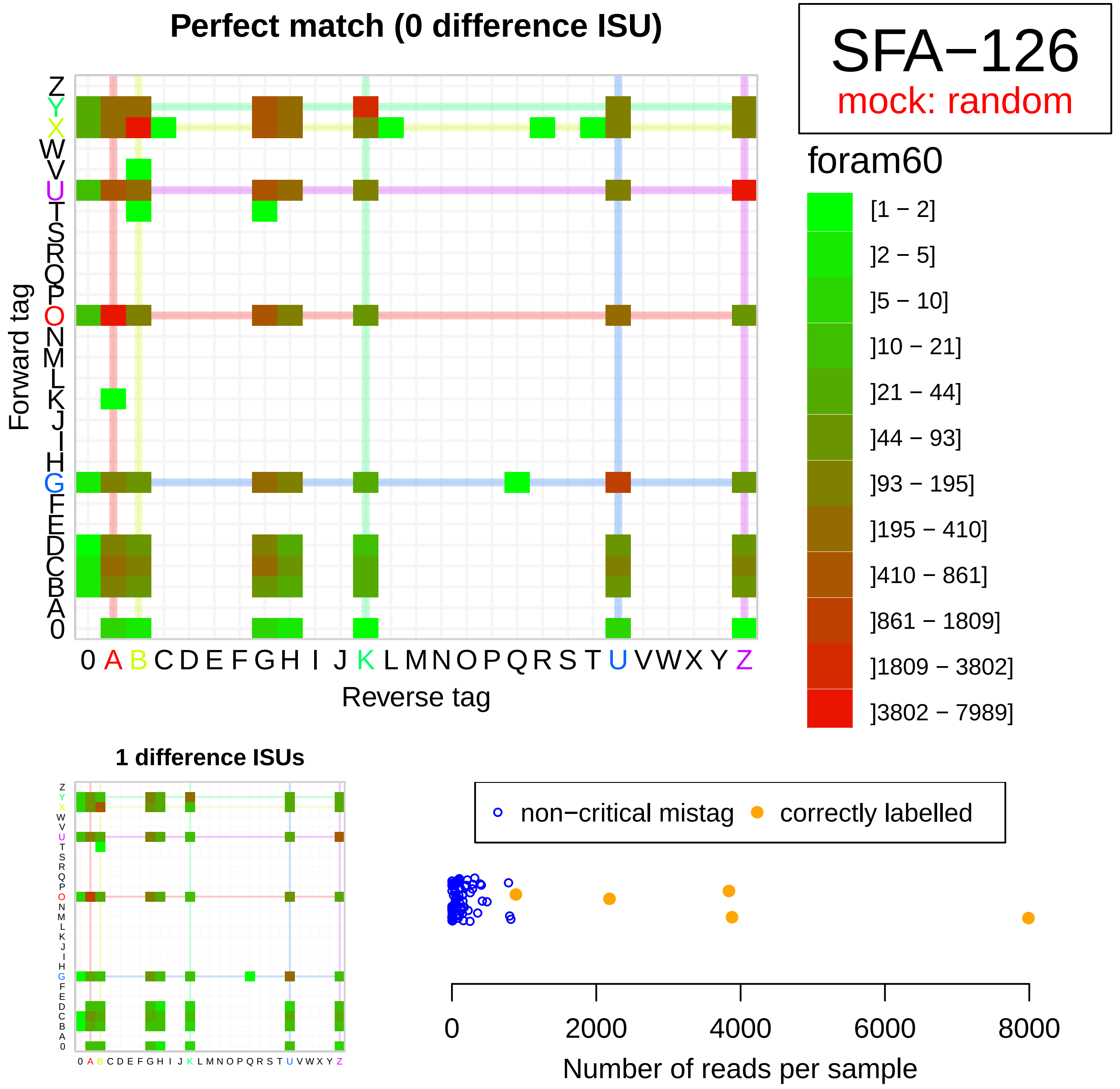


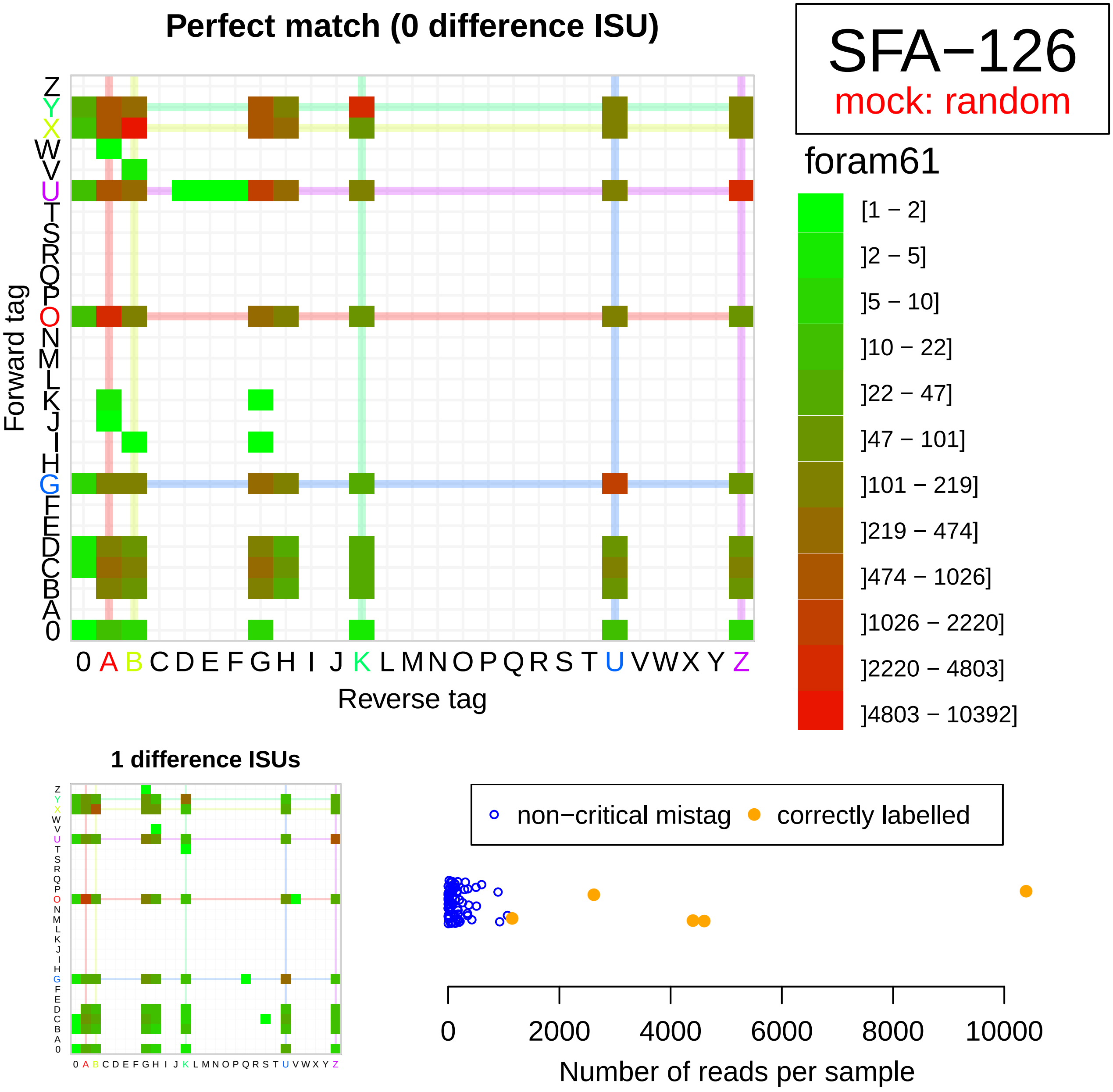
SFA−126

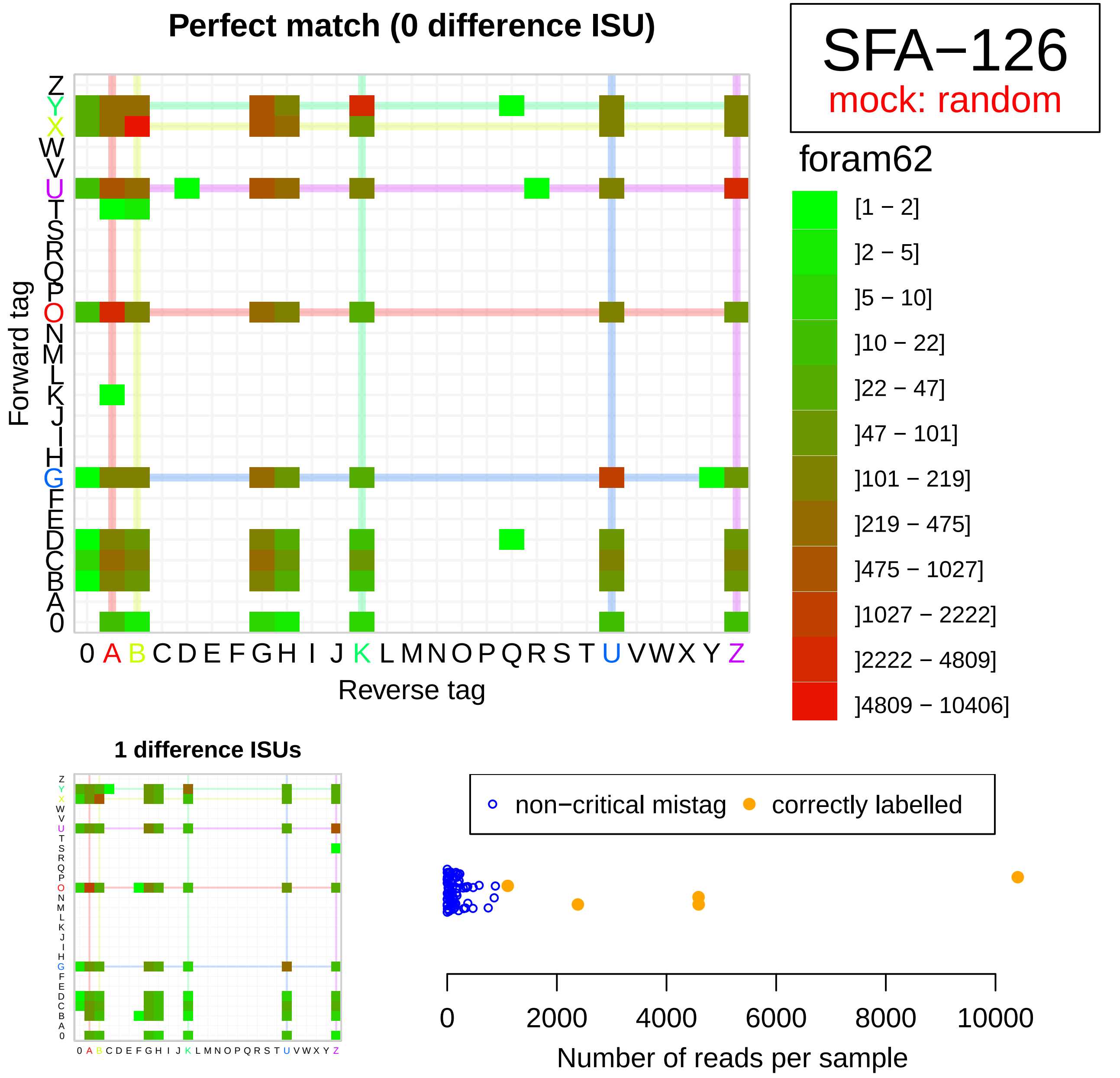
mock: random

Perfect match (0 difference ISU)

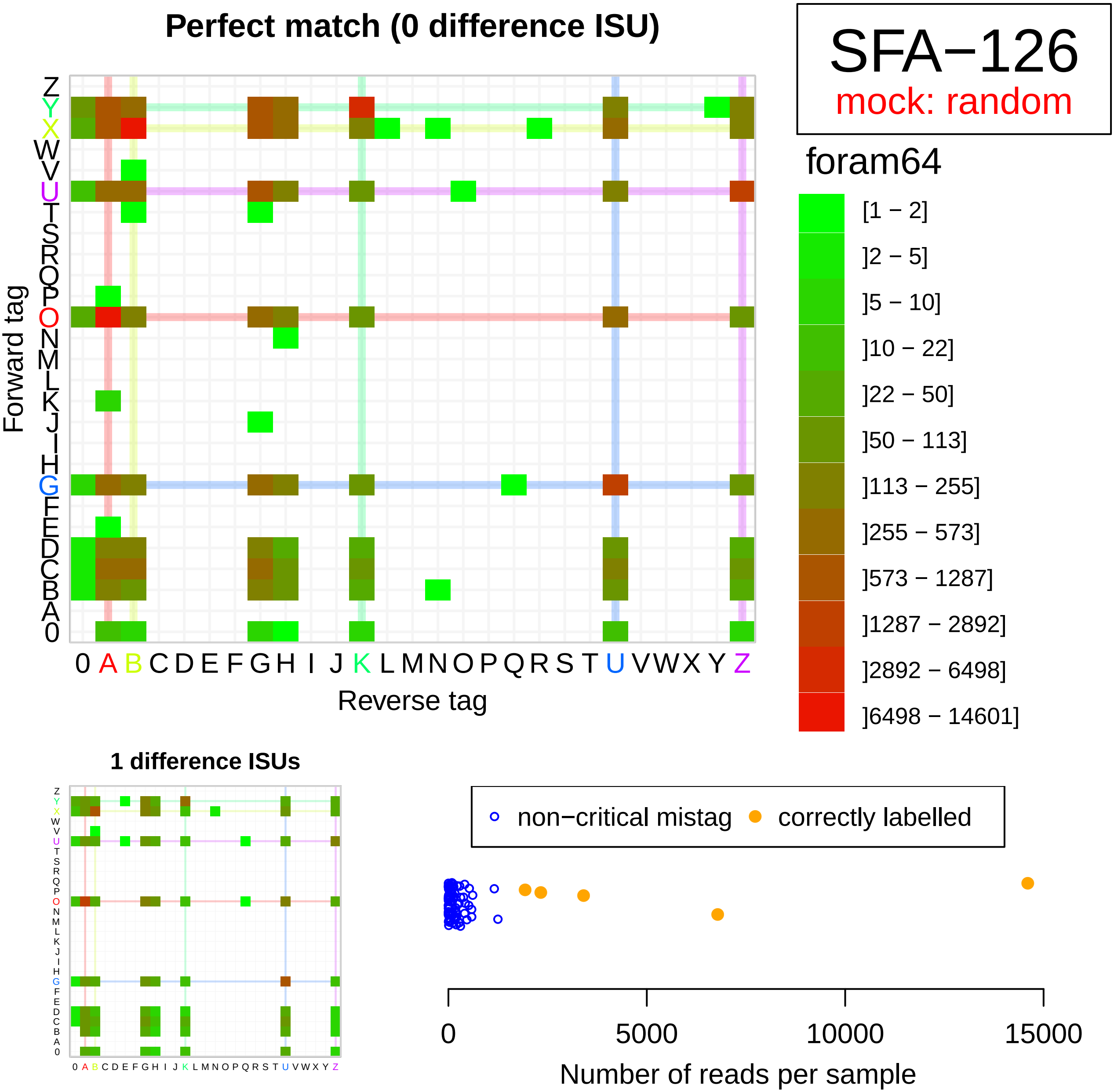
1 difference ISUs


Perfect match (0 difference ISU)
SFA−126
mock: random
foram60
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 21]
]21 − 44]
]44 − 93]
]93 − 195]
]195 − 410]
]410 − 861]
]861 − 1809]
]1809 − 3802]
]3802 − 7989]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
○ non−critical mistag
● correctly labelled
0
2000
4000
6000
8000
Number of reads per sample

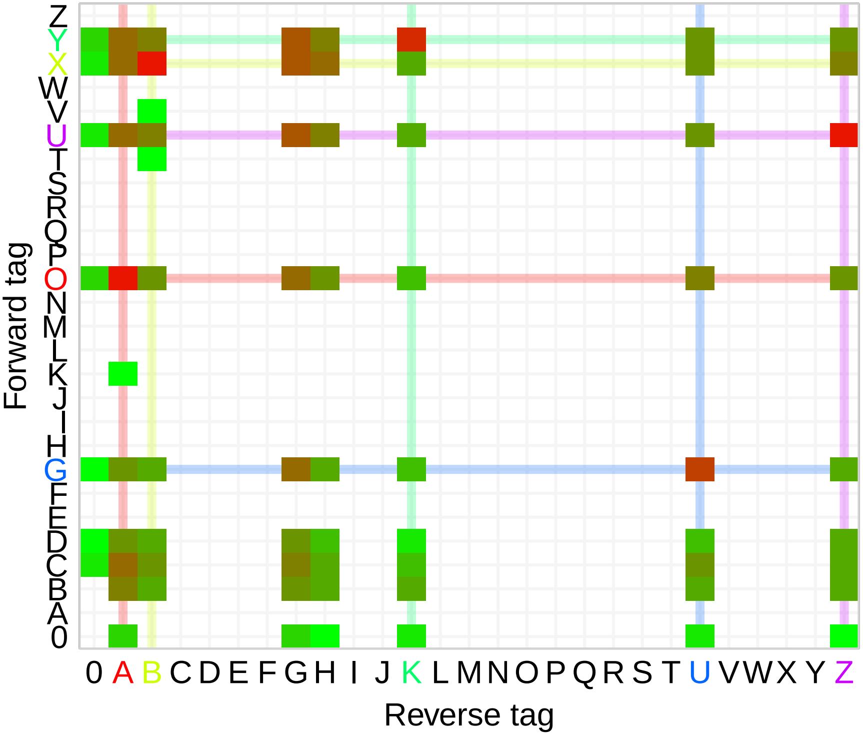

Perfect match (0 difference ISU)
SFA−126
mock: random
foram61
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 22]
]22 − 47]
]47 − 101]
]101 − 219]
]219 − 474]
]474 − 1026]
]1026 − 2220]
]2220 − 4803]
]4803 − 10392]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
non−critical mistag
correctly labelled
0
2000
4000
6000
8000
10000
Number of reads per sample



Perfect match (0 difference ISU)
SFA−126
mock: random
foram62
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 22]
]22 − 47]
]47 − 101]
]101 − 219]
]219 − 475]
]475 − 1027]
]1027 − 2222]
]2222 − 4809]
]4809 − 10406]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
non−critical mistag
correctly labelled
0
2000
4000
6000
8000
10000
Number of reads per sample

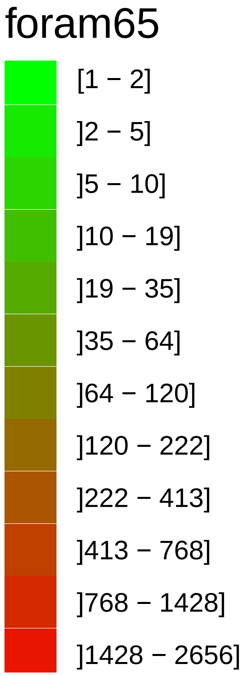

Perfect match (0 difference ISU)
SFA−126
mock: random
foram64
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 22]
]22 − 50]
]50 − 113]
]113 − 255]
]255 − 573]
]573 − 1287]
]1287 − 2892]
]2892 − 6498]
]6498 − 14601]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
non−critical mistag
correctly labelled
0
5000
10000
15000
Number of reads per sample

Perfect match (0 difference ISU)

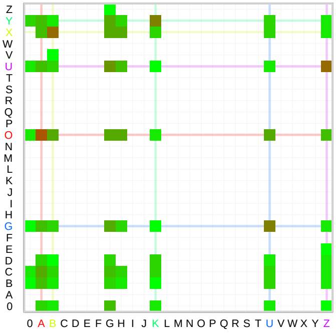


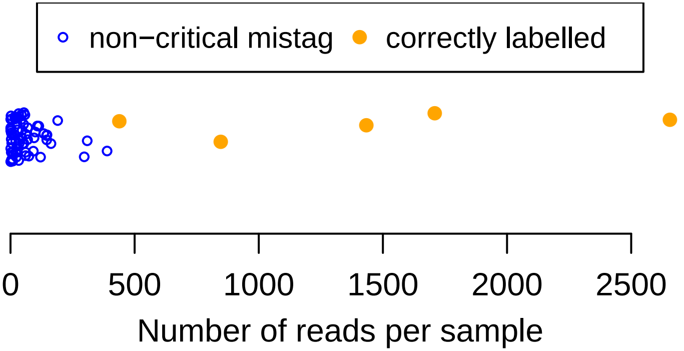
SFA−126

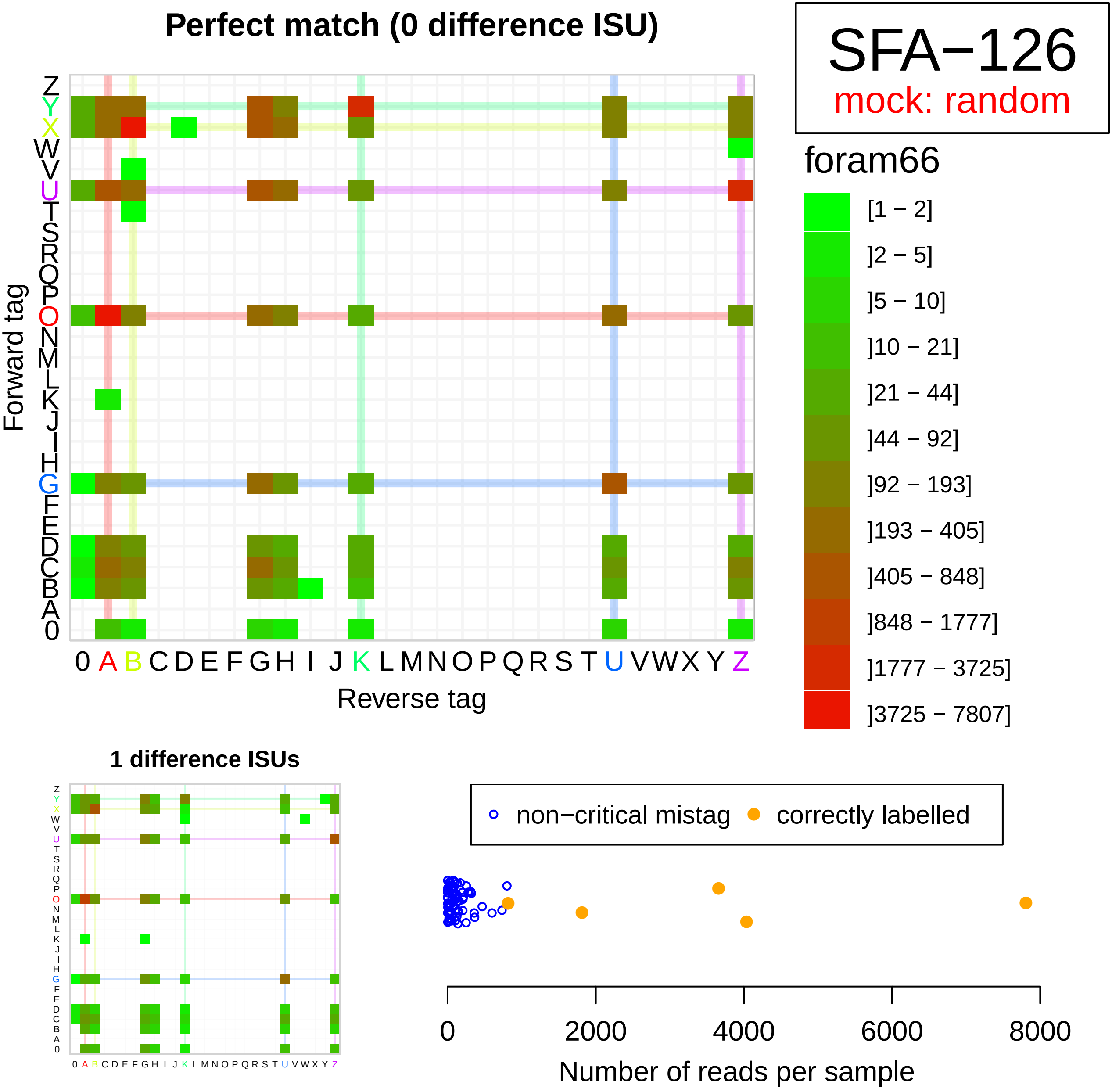
mock: random

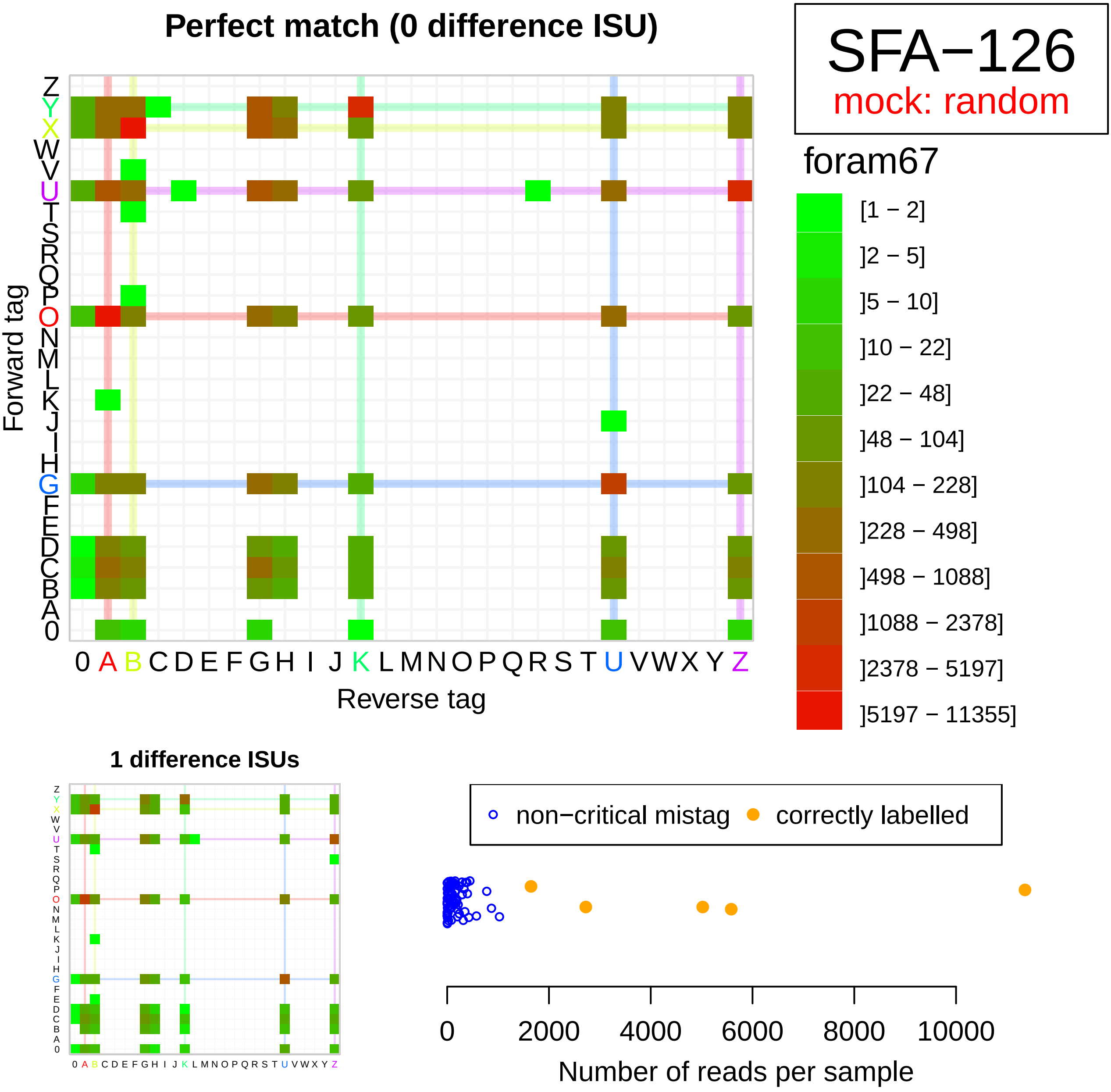


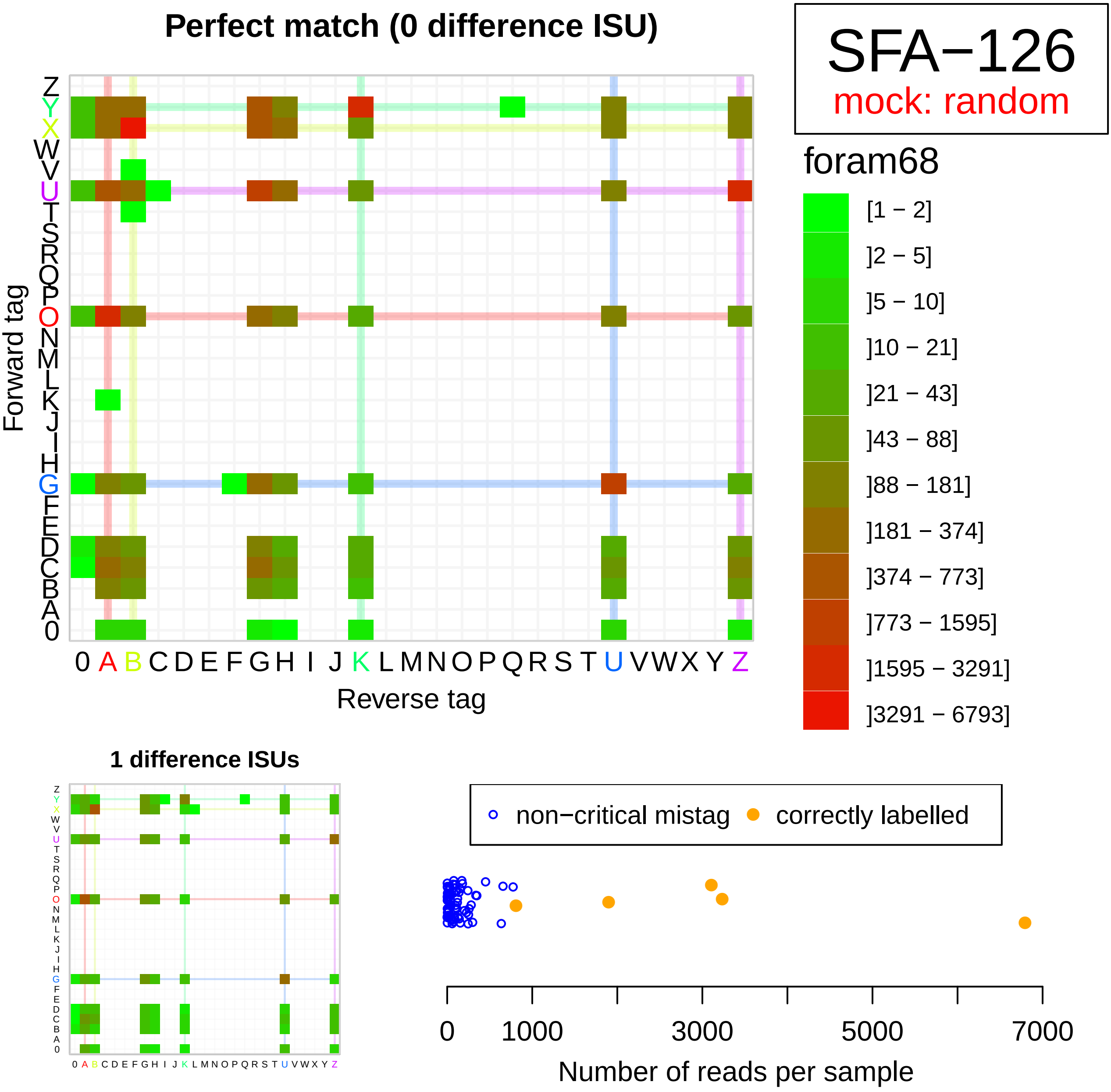
1 difference ISUs

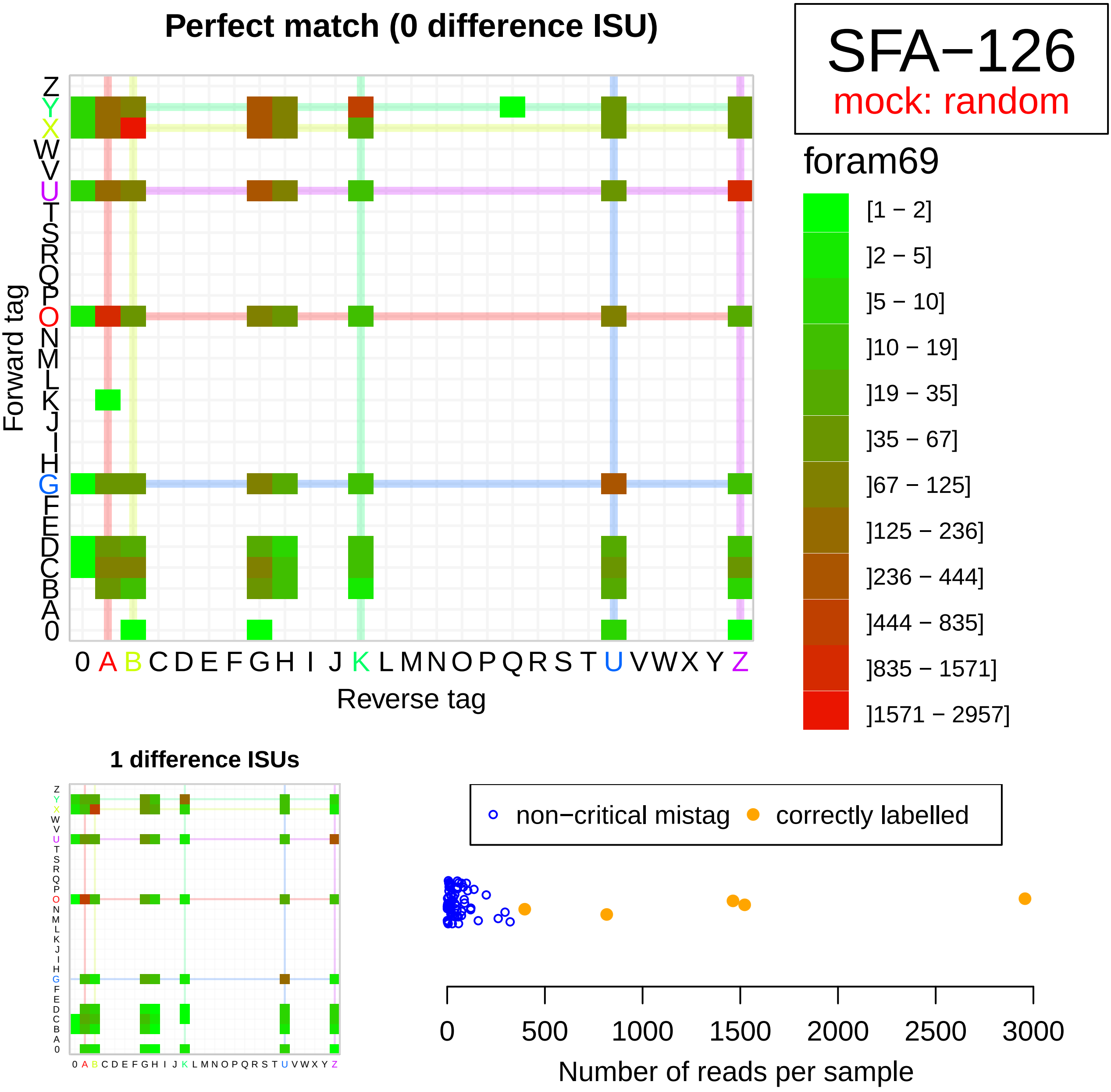


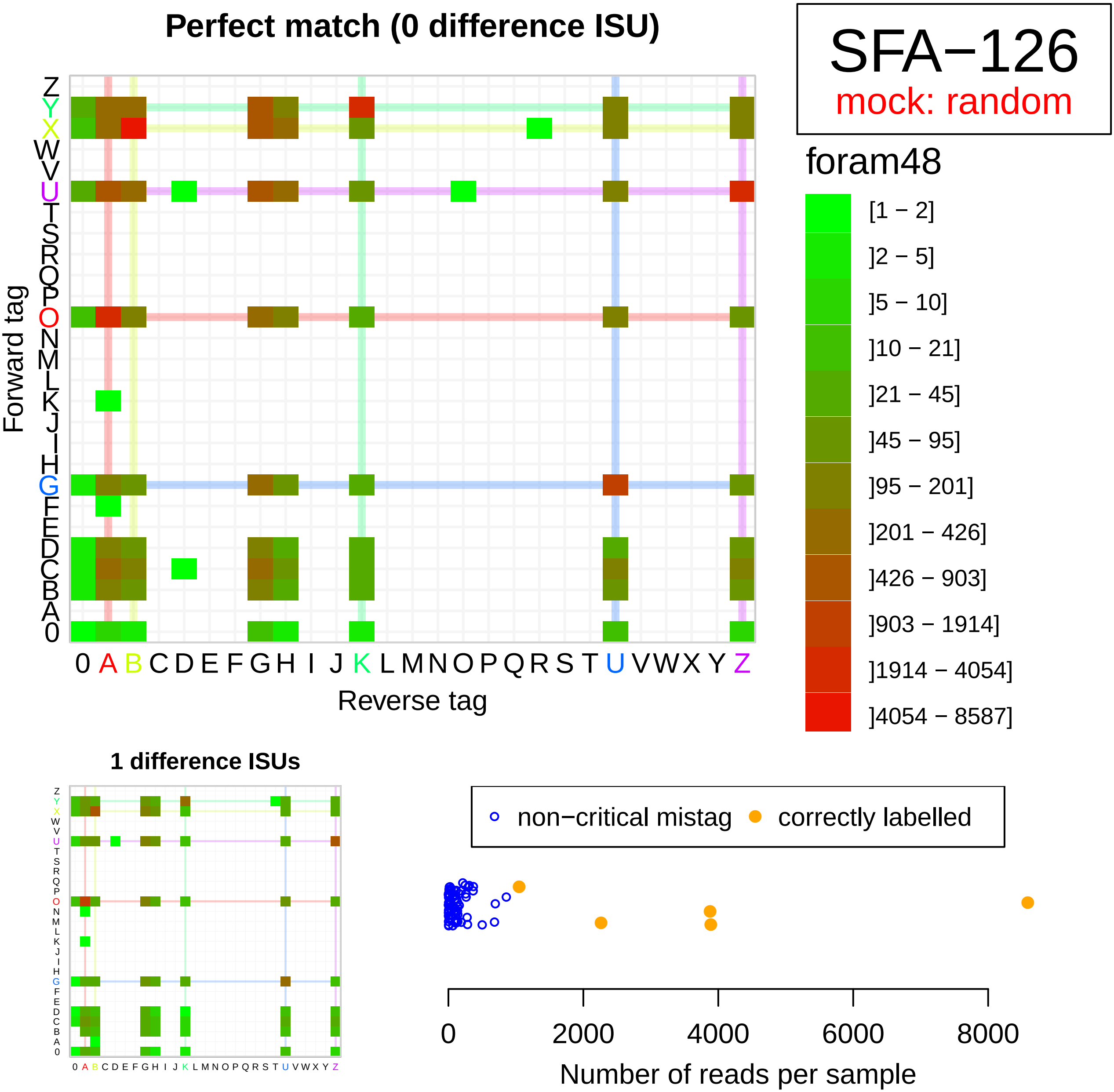


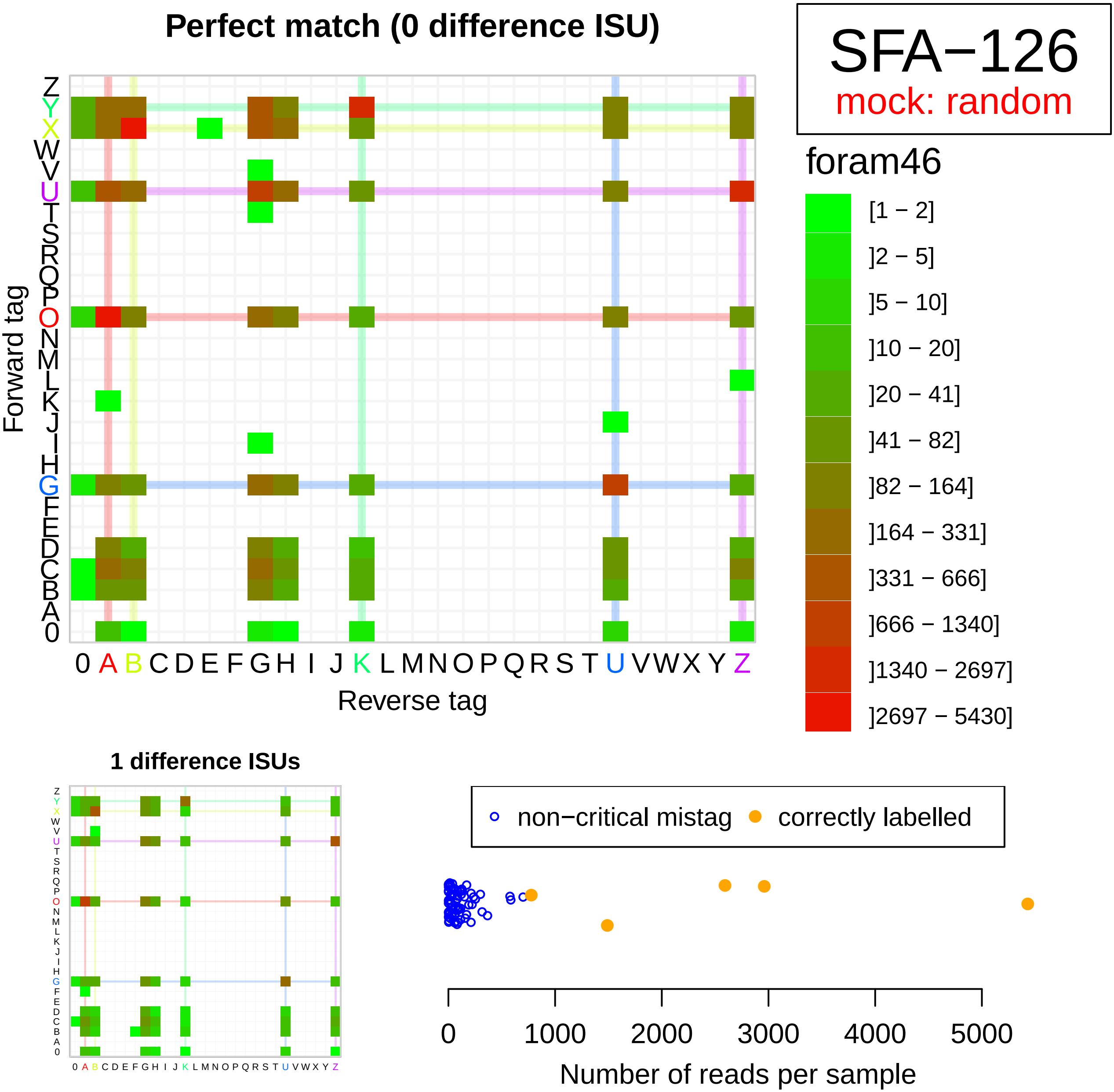

Perfect match (0 difference ISU)
SFA−126
mock: random
foram66
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 21]
]21 − 44]
]44 − 92]
]92 − 193]
]193 − 405]
]405 − 848]
]848 − 1777]
]1777 − 3725]
]3725 − 7807]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
non−critical mistag
correctly labelled
0
2000
4000
6000
8000
Number of reads per sample

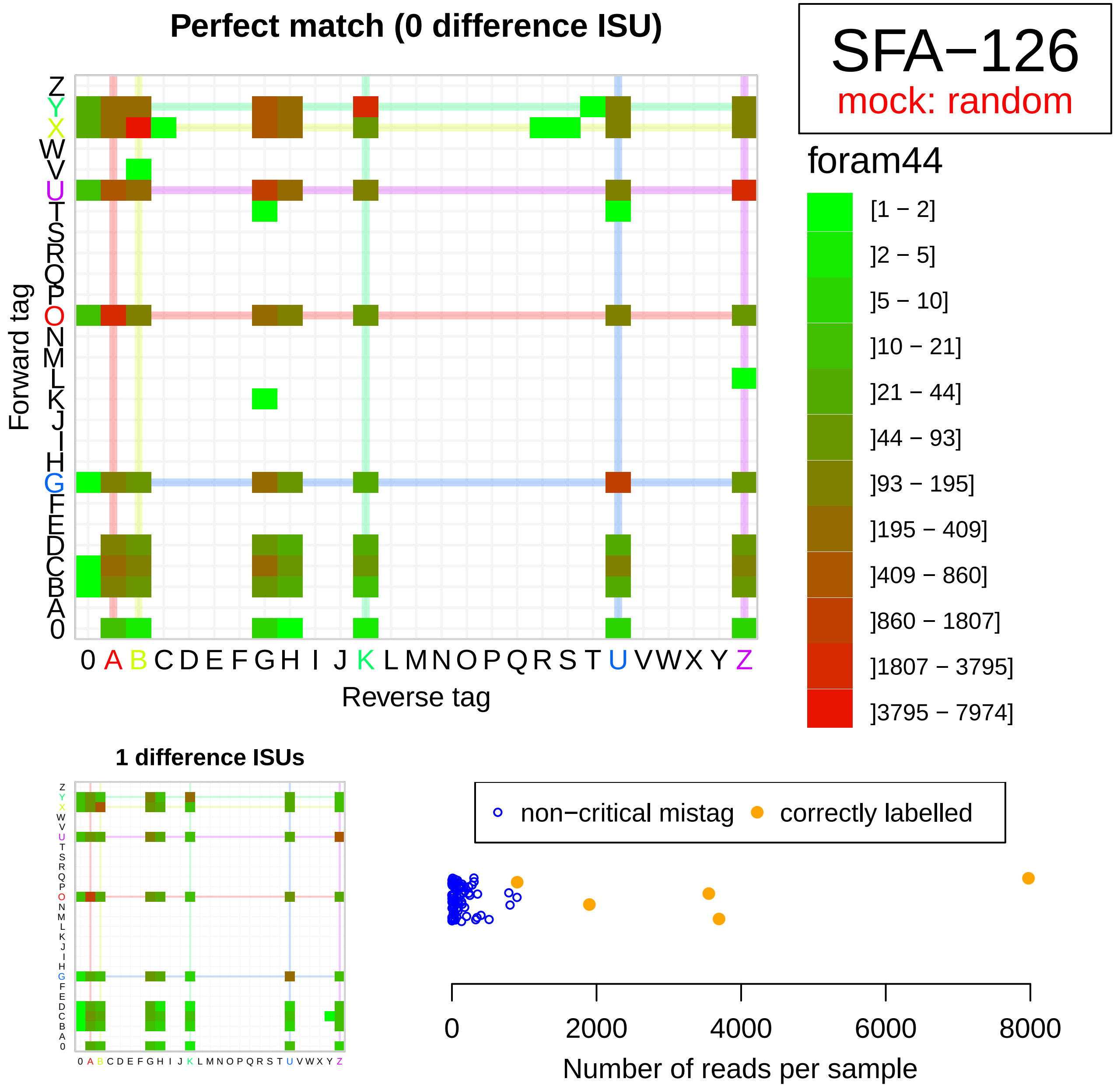

Perfect match (0 difference ISU)
SFA−126
mock: random
foram67
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 22]
]22 − 48]
]48 − 104]
]104 − 228]
]228 − 498]
]498 − 1088]
]1088 − 2378]
]2378 − 5197]
]5197 − 11355]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
non−critical mistag
correctly labelled
0
2000
4000
6000
8000
10000
Number of reads per sample

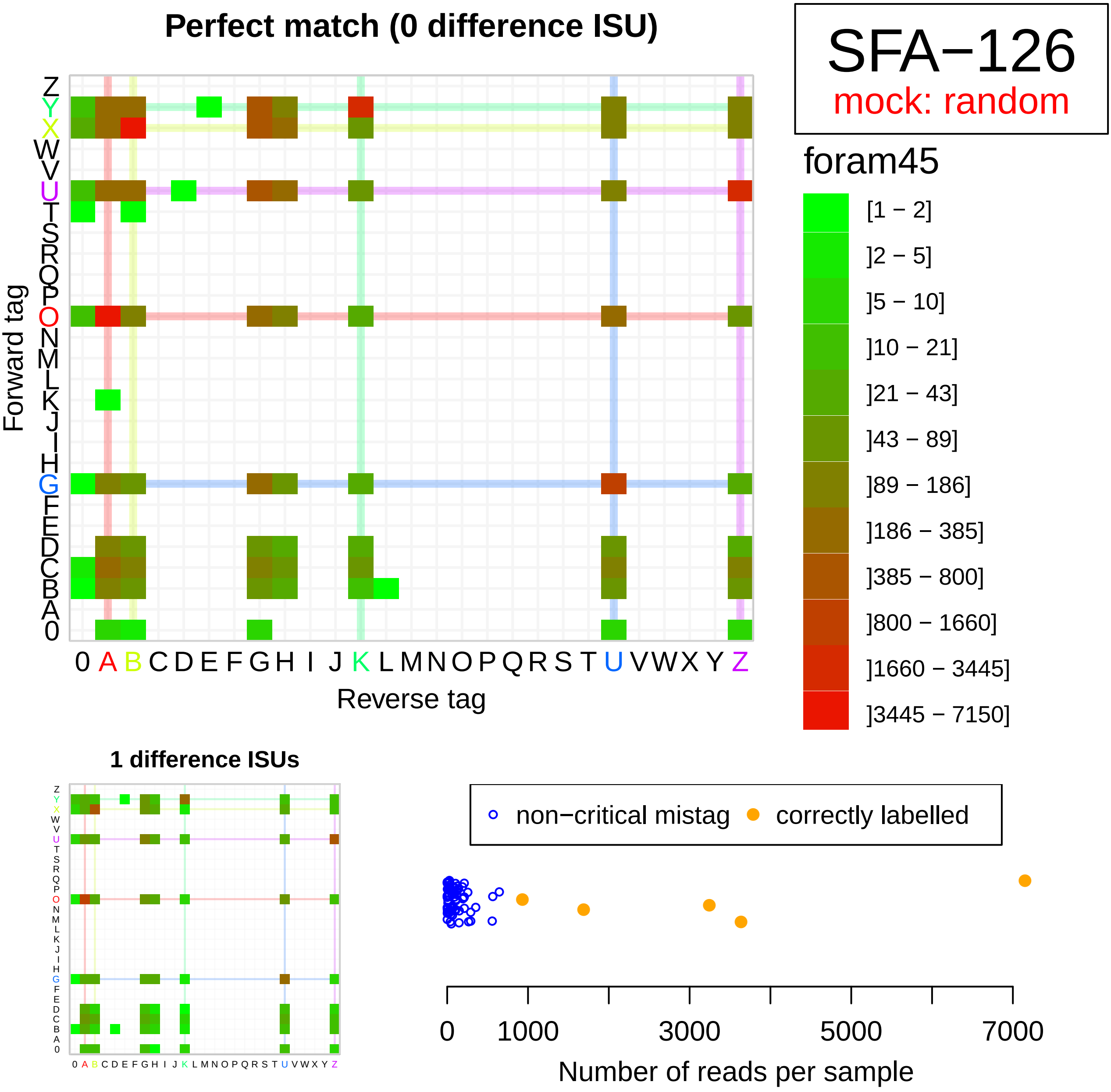

Perfect match (0 difference ISU)
SFA−126
mock: random
foram68
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 21]
]21 − 43]
]43 − 88]
]88 − 181]
]181 − 374]
]374 − 773]
]773 − 1595]
]1595 − 3291]
]3291 − 6793]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
○ non−critical mistag ● correctly labelled
0 1000 3000 5000 7000
Number of reads per sample

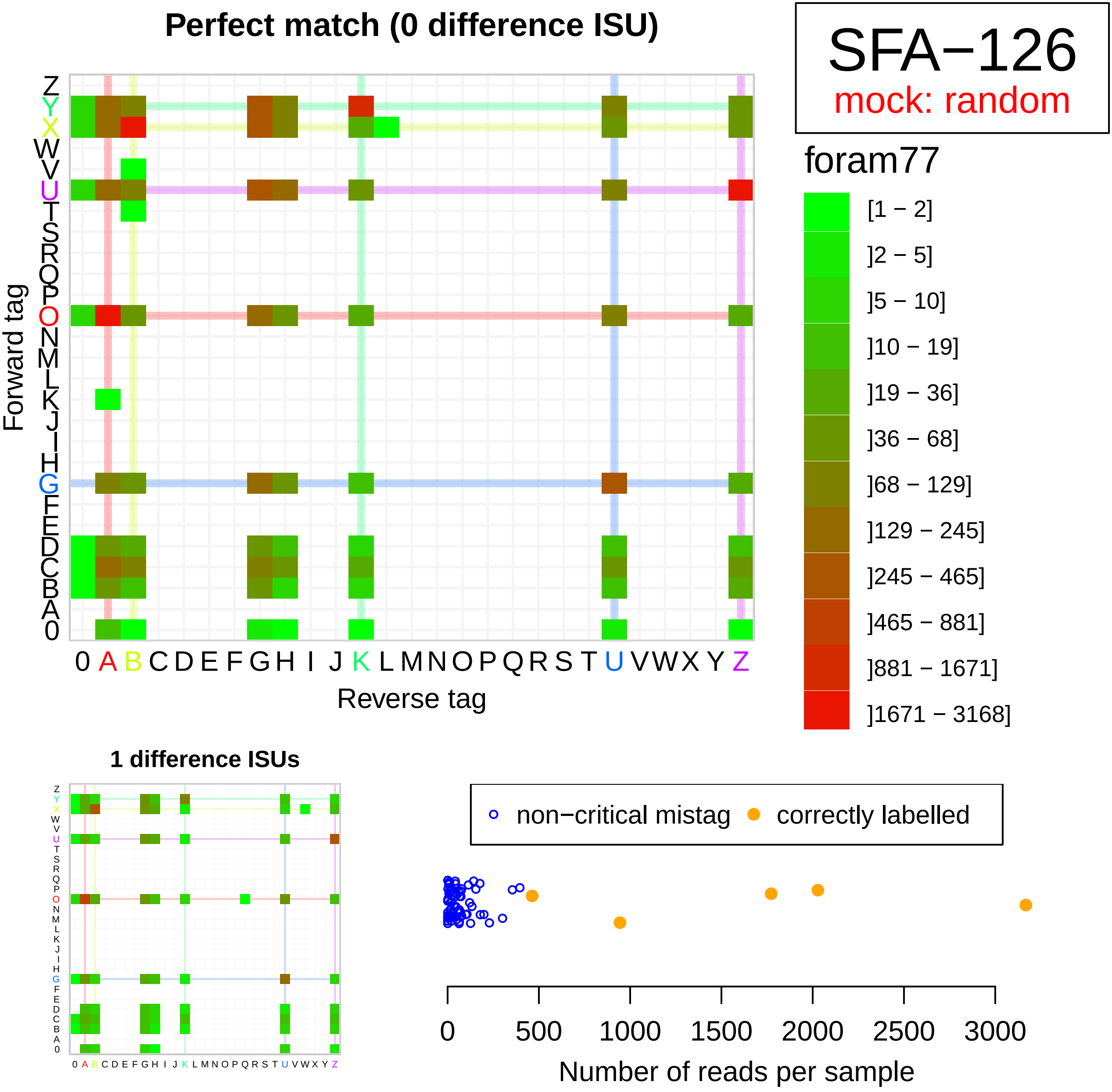

SFA−126
mock: random
Perfect match (0 difference ISU)
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
Reverse tag
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
foram69
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 19]
]19 − 35]
]35 − 67]
]67 − 125]
]125 − 236]
]236 − 444]
]444 − 835]
]835 − 1571]
]1571 − 2957]
1 difference ISUs
non−critical mistag
correctly labelled
0
500
1000
1500
2000
2500
3000
Number of reads per sample

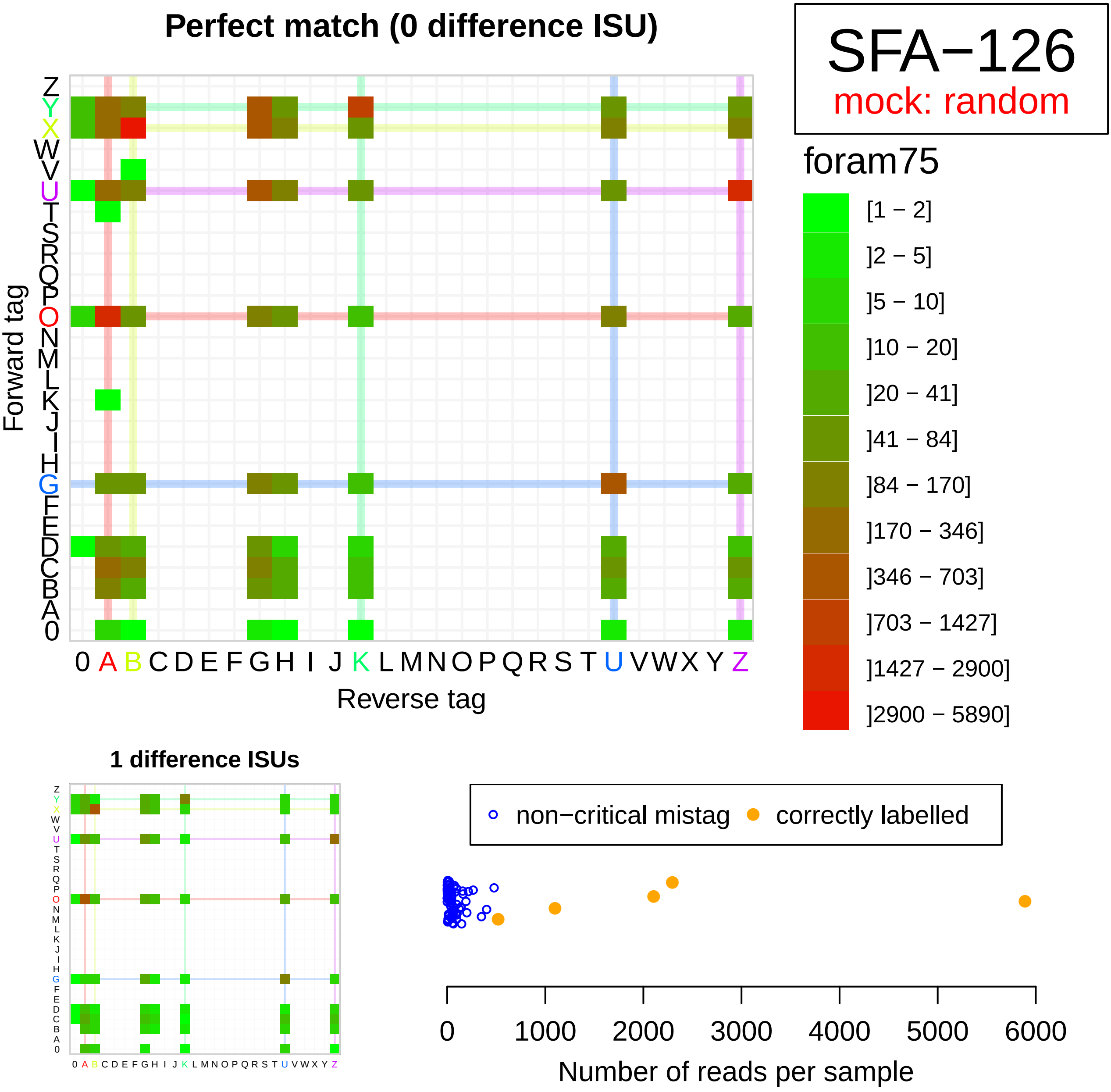

Perfect match (0 difference ISU)
SFA−126
mock: random
foram48
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 21]
]21 − 45]
]45 − 95]
]95 − 201]
]201 − 426]
]426 − 903]
]903 − 1914]
]1914 − 4054]
]4054 − 8587]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
○ non−critical mistag
● correctly labelled
0
2000
4000
6000
8000
Number of reads per sample


SFA−126
mock: random
Perfect match (0 difference ISU)
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
foram46
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 20]
]20 − 41]
]41 − 82]
]82 − 164]
]164 − 331]
]331 − 666]
]666 − 1340]
]1340 − 2697]
]2697 − 5430]
1 difference ISUs
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
○ non−critical mistag ● correctly labelled
0
1000
2000
3000
4000
5000
Number of reads per sample


Perfect match (0 difference ISU)
SFA−126
mock: random
foram44
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 21]
]21 − 44]
]44 − 93]
]93 − 195]
]195 − 409]
]409 − 860]
]860 − 1807]
]1807 − 3795]
]3795 − 7974]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
non−critical mistag
correctly labelled
0
2000
4000
6000
8000
Number of reads per sample

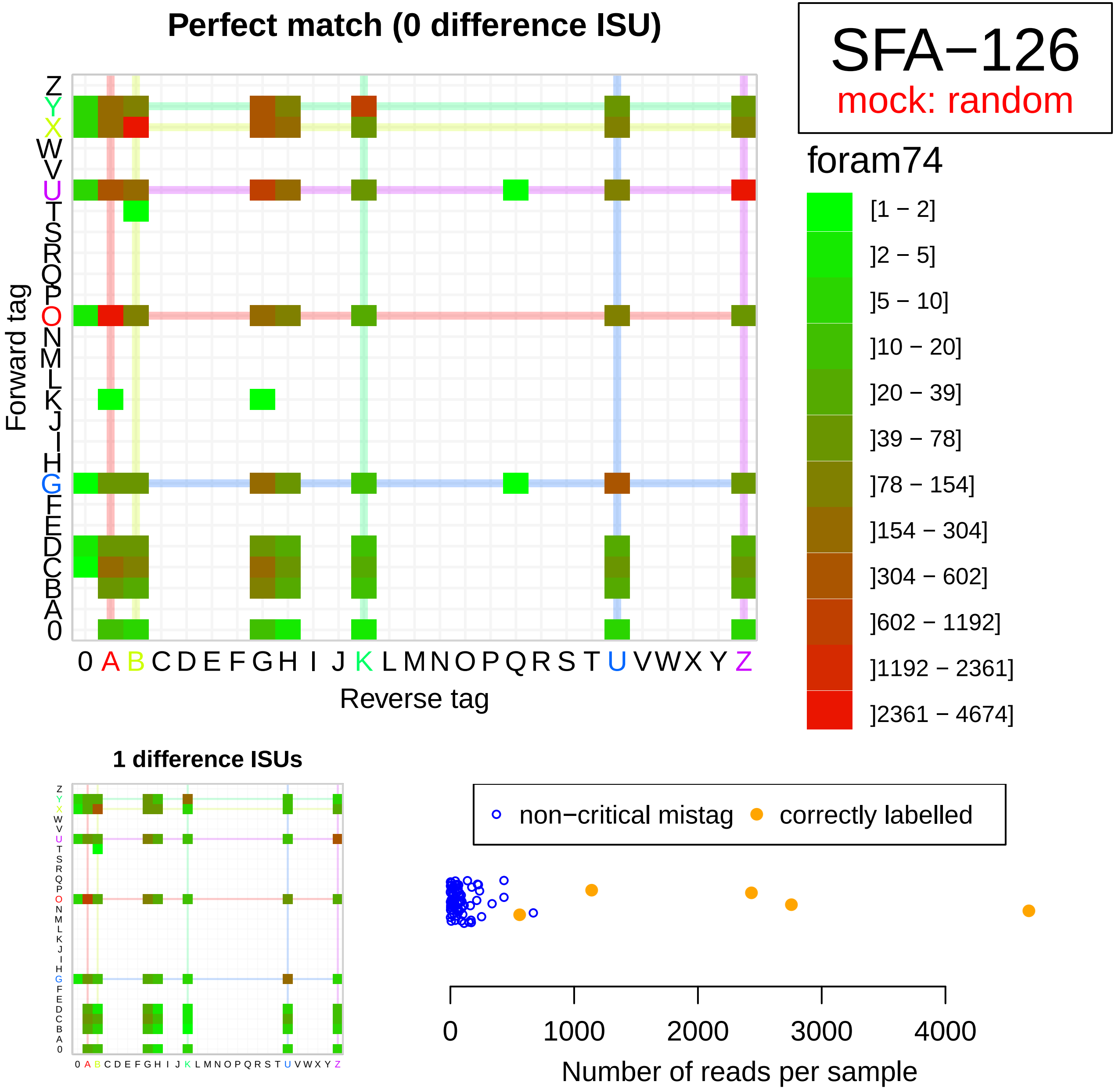

SFA−126
mock: random
Perfect match (0 difference ISU)
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
foram45
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 21]
]21 − 43]
]43 − 89]
]89 − 186]
]186 − 385]
]385 − 800]
]800 − 1660]
]1660 − 3445]
]3445 − 7150]
1 difference ISUs
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
○ non−critical mistag
● correctly labelled
0
1000
3000
5000
7000
Number of reads per sample

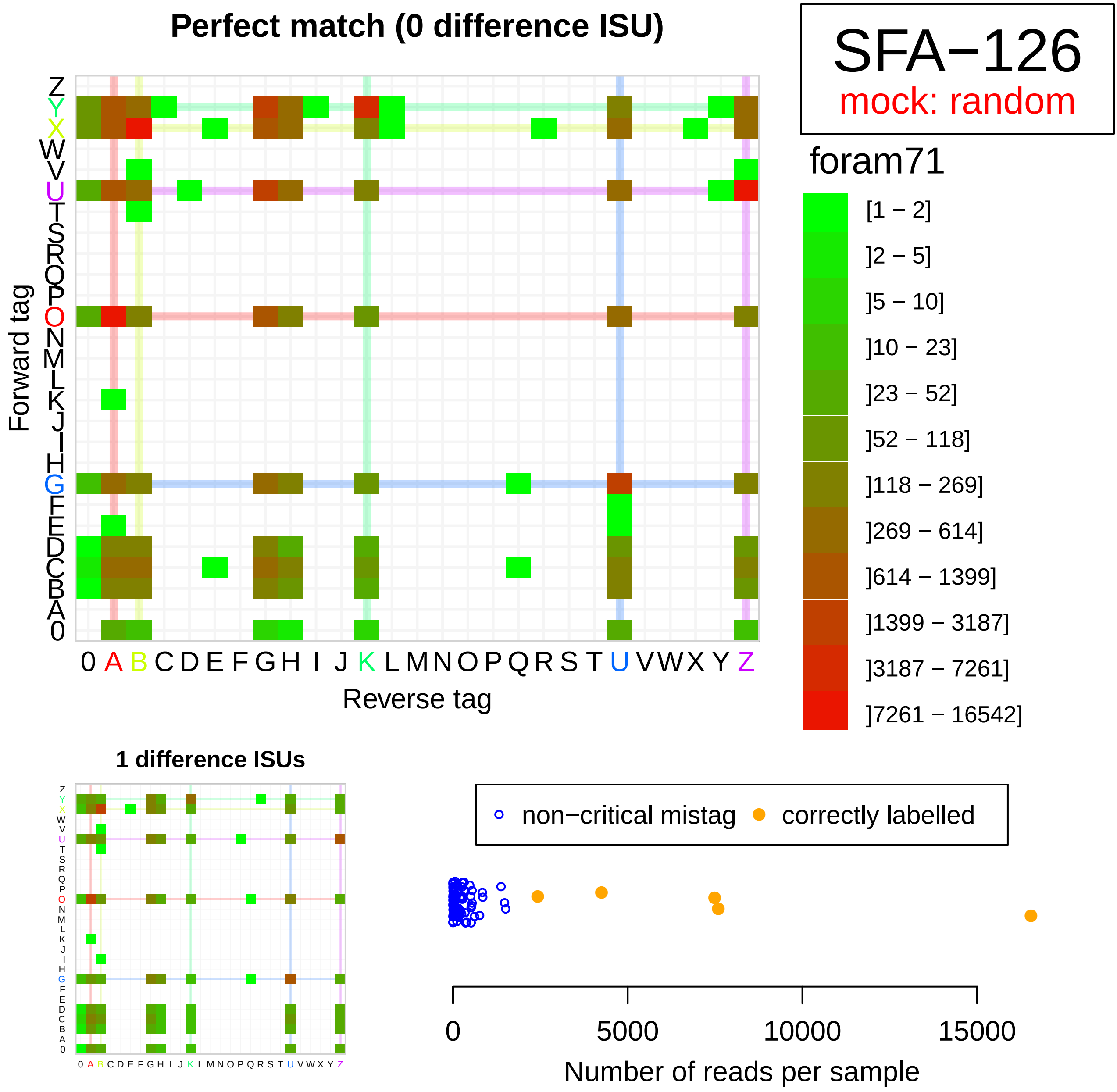

Perfect match (0 difference ISU)
SFA−126
mock: random
foram77
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 19]
]19 − 36]
]36 − 68]
]68 − 129]
]129 − 245]
]245 − 465]
]465 − 881]
]881 − 1671]
]1671 − 3168]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
non−critical mistag
correctly labelled
0
500
1000
1500
2000
2500
3000
Number of reads per sample


Perfect match (0 difference ISU)
SFA−126
mock: random
foram75
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 20]
]20 − 41]
]41 − 84]
]84 − 170]
]170 − 346]
]346 − 703]
]703 − 1427]
]1427 − 2900]
]2900 − 5890]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
non−critical mistag
correctly labelled
0
1000
2000
3000
4000
5000
6000
Number of reads per sample

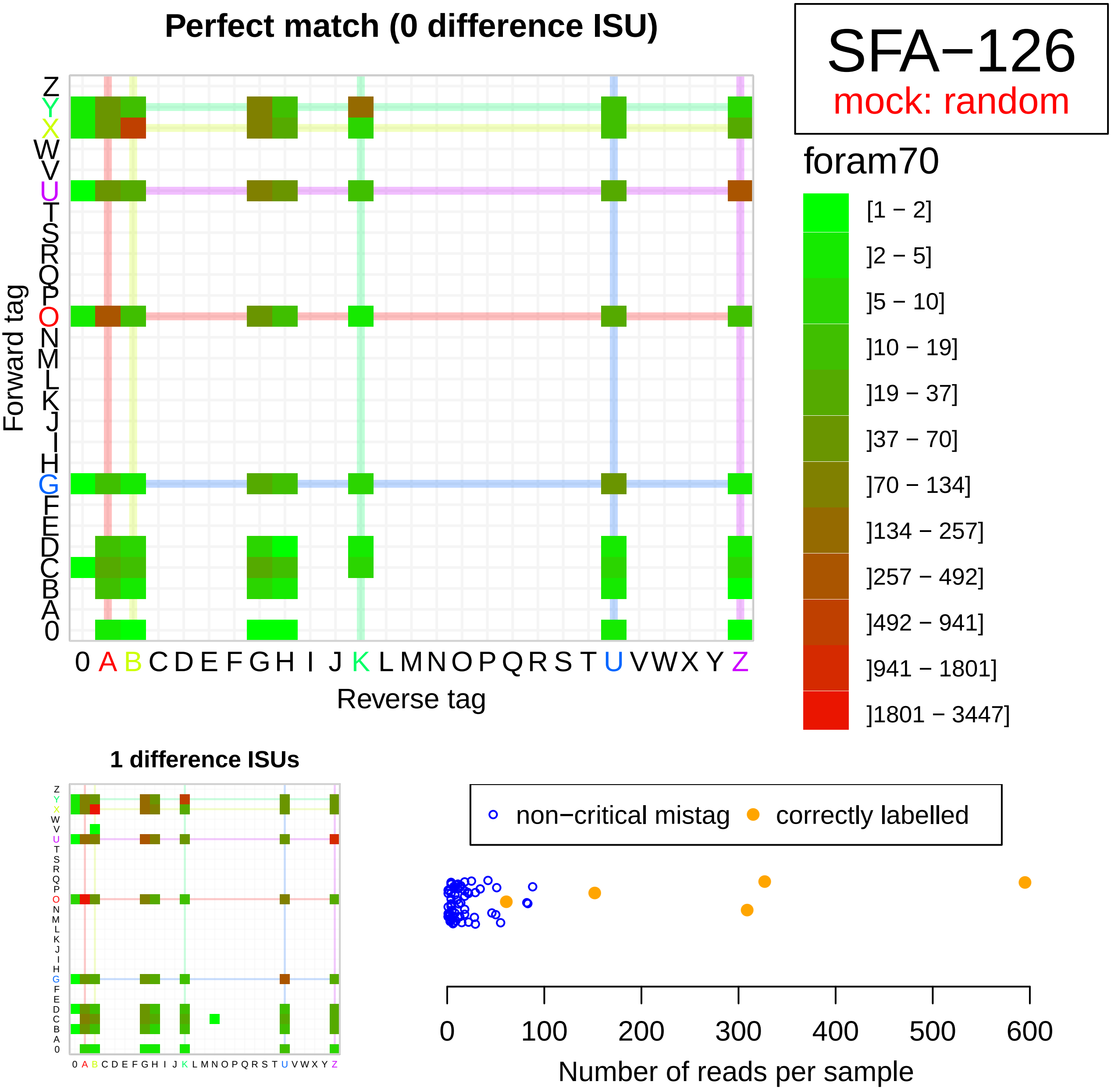
SFA−126

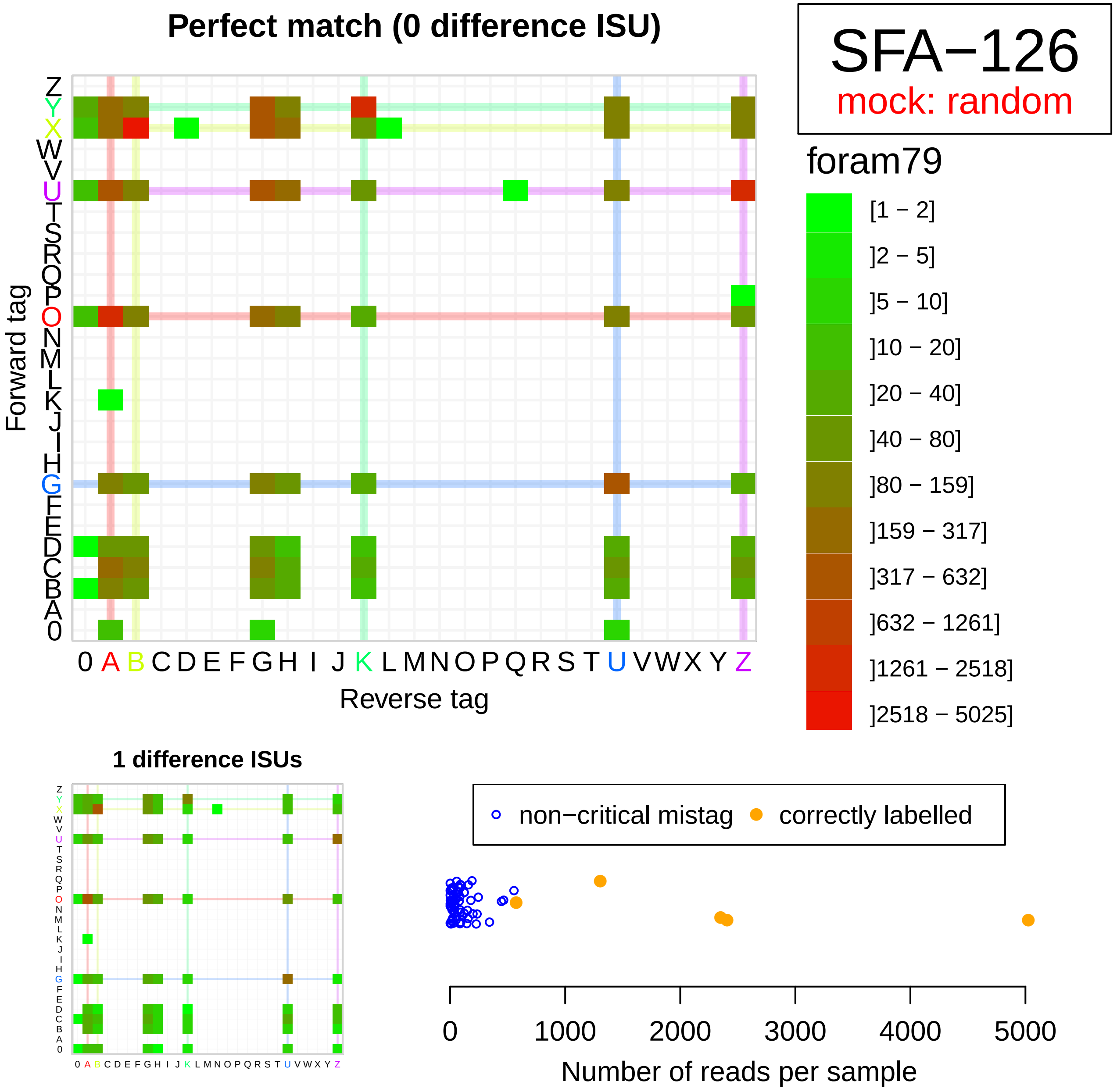
mock: random

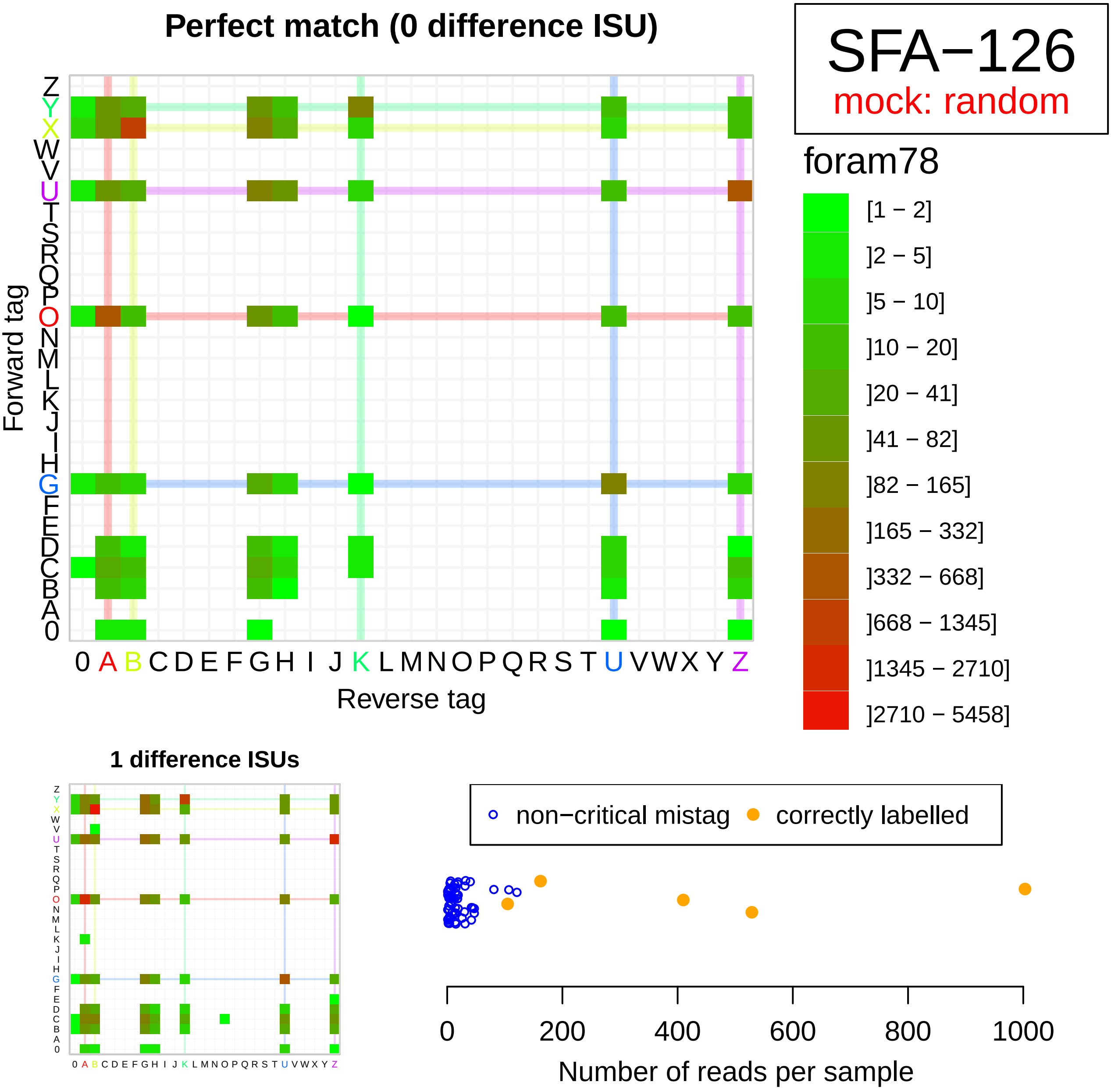


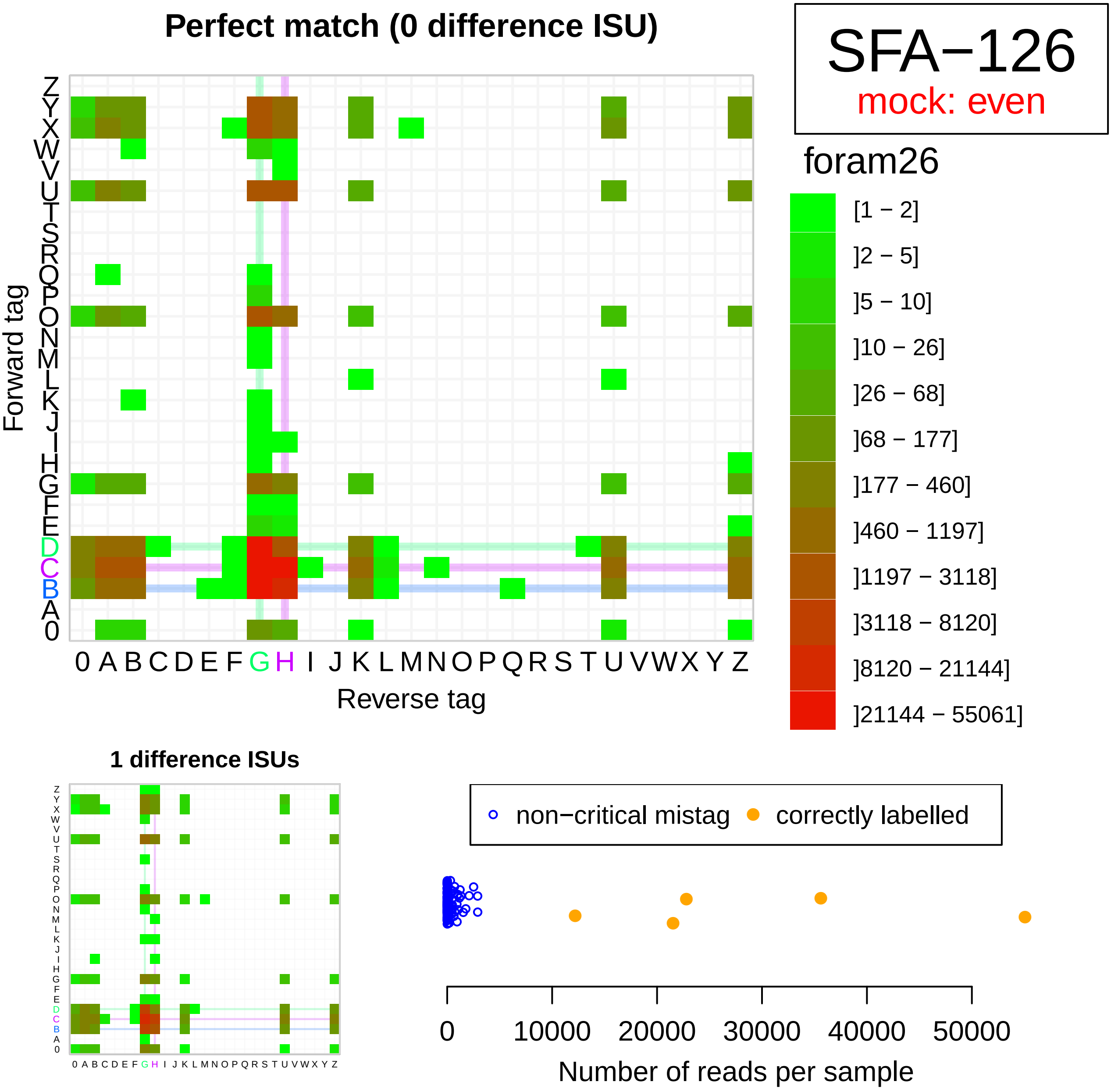


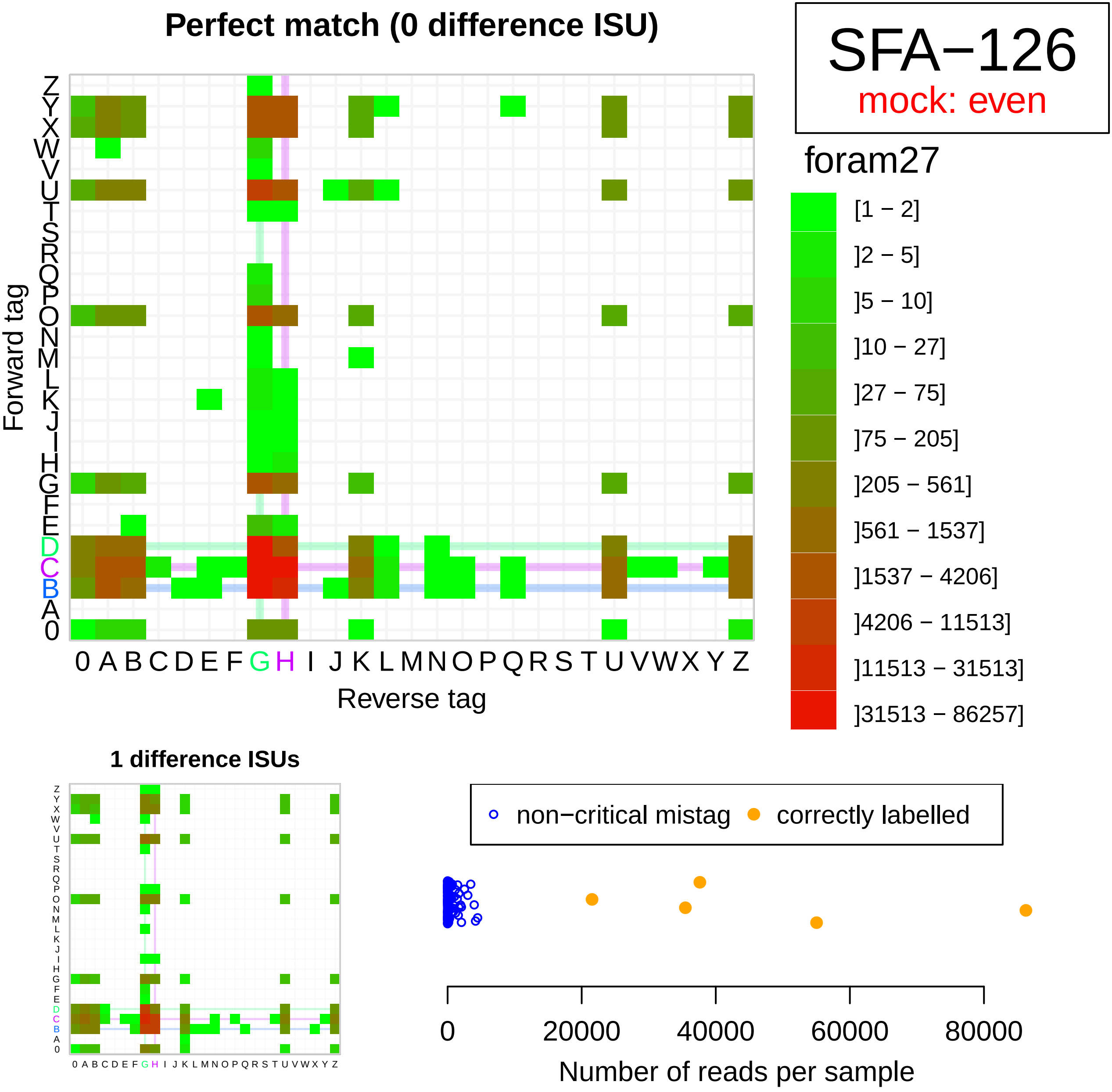
Perfect match (0 difference ISU)
SFA−126
mock: random
foram71
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 23]
]23 − 52]
]52 − 118]
]118 − 269]
]269 − 614]
]614 − 1399]
]1399 − 3187]
]3187 − 7261]
]7261 − 16542]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
non−critical mistag
correctly labelled
0
5000
10000
15000
Number of reads per sample

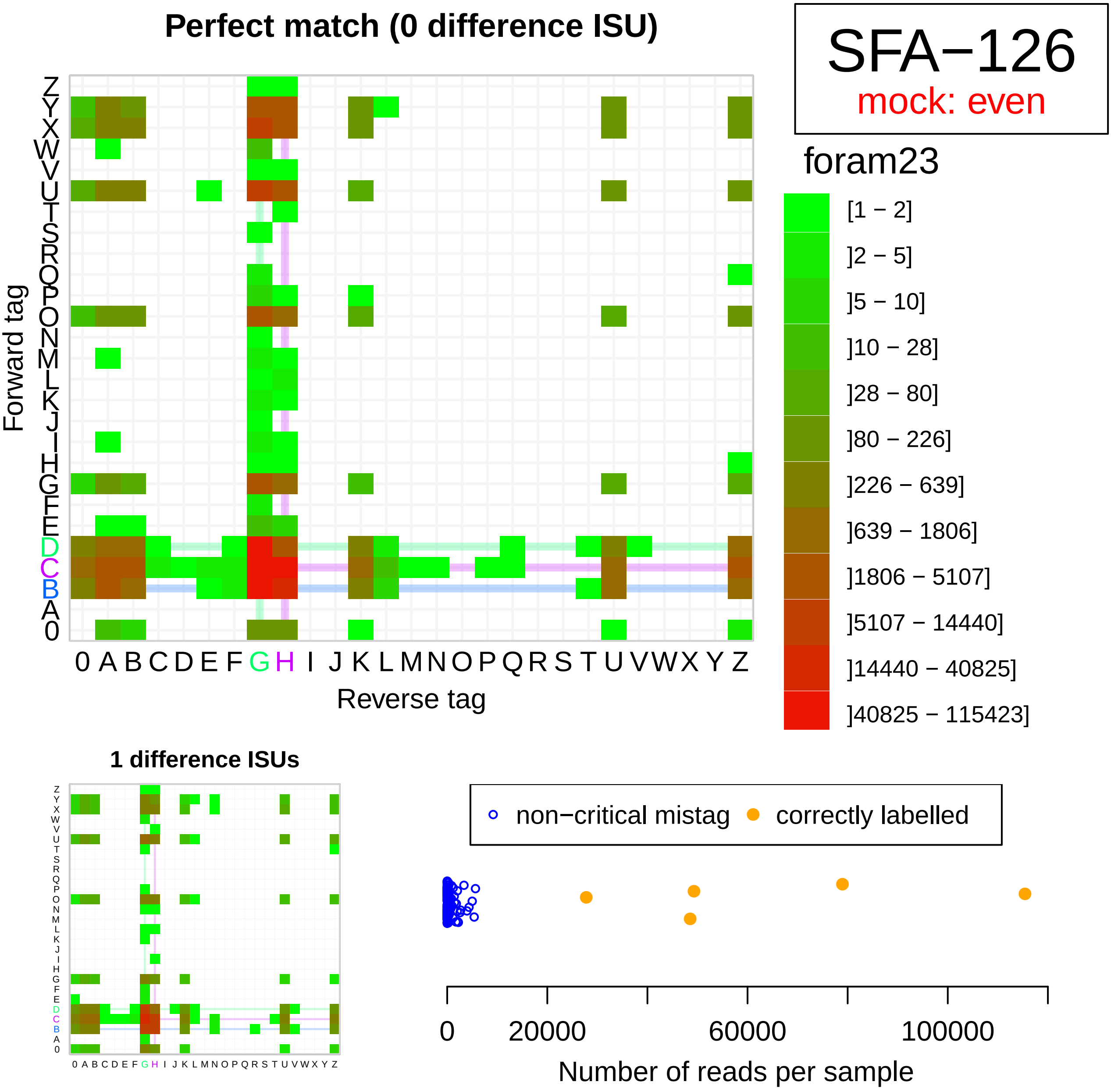

Perfect match (0 difference ISU)
SFA−126
mock: random
foram70
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 19]
]19 − 37]
]37 − 70]
]70 − 134]
]134 − 257]
]257 − 492]
]492 − 941]
]941 − 1801]
]1801 − 3447]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
non−critical mistag
correctly labelled
0
100
200
300
400
500
600
Number of reads per sample

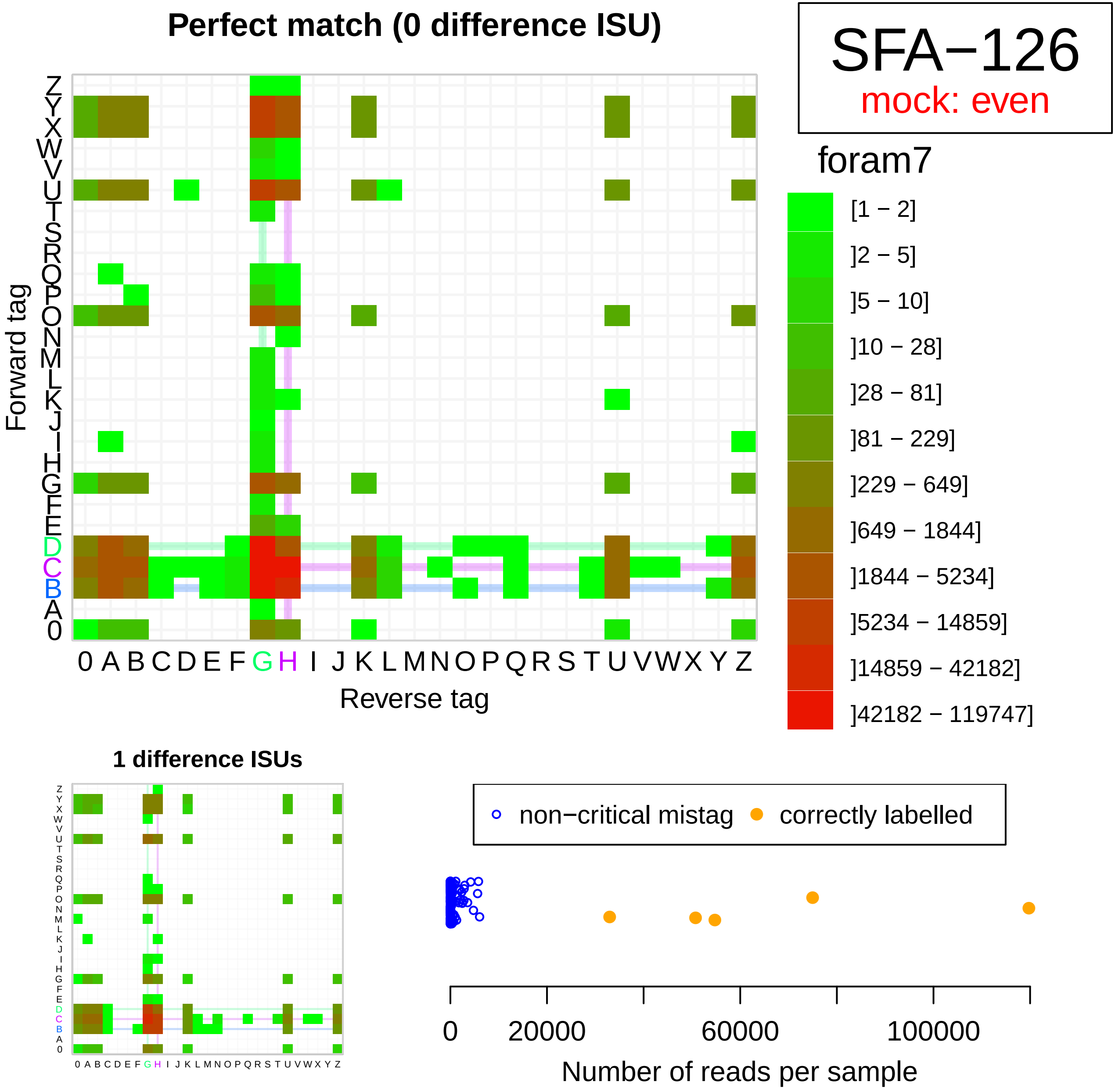

Perfect match (0 difference ISU)
SFA−126
mock: random
foram79
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 20]
]20 − 40]
]40 − 80]
]80 − 159]
]159 − 317]
]317 − 632]
]632 − 1261]
]1261 − 2518]
]2518 − 5025]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
non−critical mistag
correctly labelled
0
1000
2000
3000
4000
5000
Number of reads per sample


Perfect match (0 difference ISU)
SFA−126
mock: random
foram78
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 20]
]20 − 41]
]41 − 82]
]82 − 165]
]165 − 332]
]332 − 668]
]668 − 1345]
]1345 − 2710]
]2710 − 5458]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
non−critical mistag
correctly labelled
0
200
400
600
800
1000
Number of reads per sample


Perfect match (0 difference ISU)
SFA−126
mock: even
foram26
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 26]
]26 − 68]
]68 − 177]
]177 − 460]
]460 − 1197]
]1197 − 3118]
]3118 − 8120]
]8120 − 21144]
]21144 − 55061]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
non−critical mistag
correctly labelled
0
10000
20000
30000
40000
50000
Number of reads per sample


Perfect match (0 difference ISU)
SFA−126
mock: even
foram27
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 27]
]27 − 75]
]75 − 205]
]205 − 561]
]561 − 1537]
]1537 − 4206]
]4206 − 11513]
]11513 − 31513]
]31513 − 86257]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
non−critical mistag
correctly labelled
0
20000
40000
60000
80000
Number of reads per sample

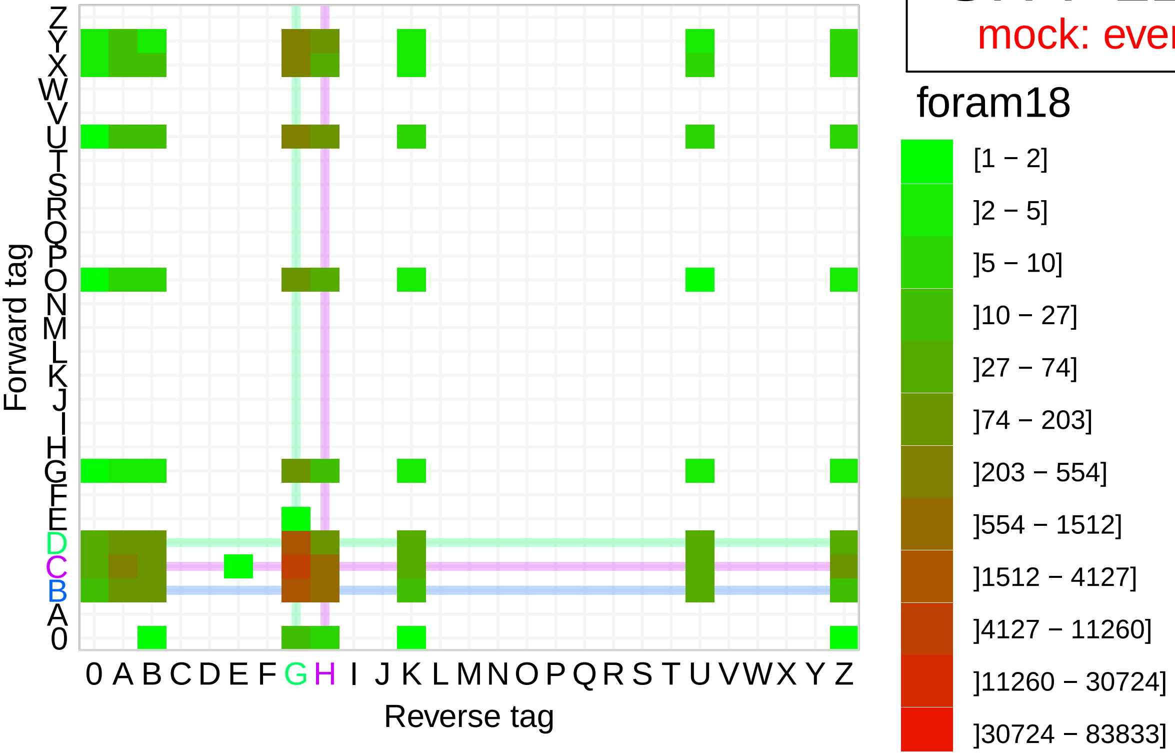

Perfect match (0 difference ISU)
SFA−126
mock: even
foram23
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 28]
]28 − 80]
]80 − 226]
]226 − 639]
]639 − 1806]
]1806 − 5107]
]5107 − 14440]
]14440 − 40825]
]40825 − 115423]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
non−critical mistag
correctly labelled
0
20000
60000
100000
Number of reads per sample


Perfect match (0 difference ISU)
SFA−126
mock: even
foram7
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 28]
]28 − 81]
]81 − 229]
]229 − 649]
]649 − 1844]
]1844 − 5234]
]5234 − 14859]
]14859 − 42182]
]42182 − 119747]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
non−critical mistag
correctly labelled
0
20000
60000
100000
Number of reads per sample

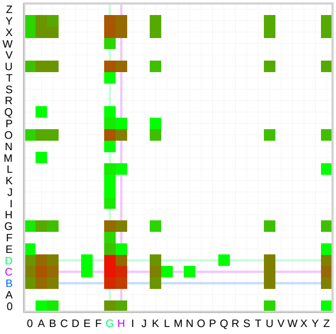
SFA−126

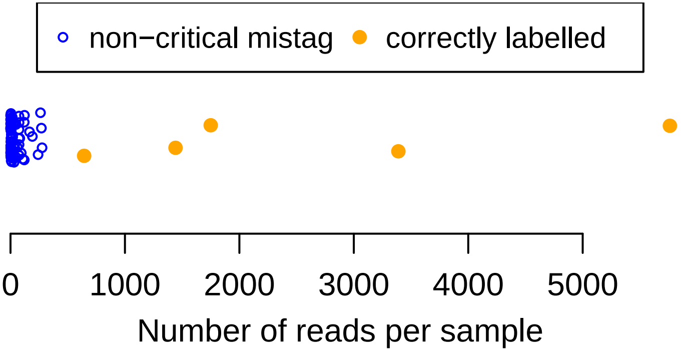
mock: even

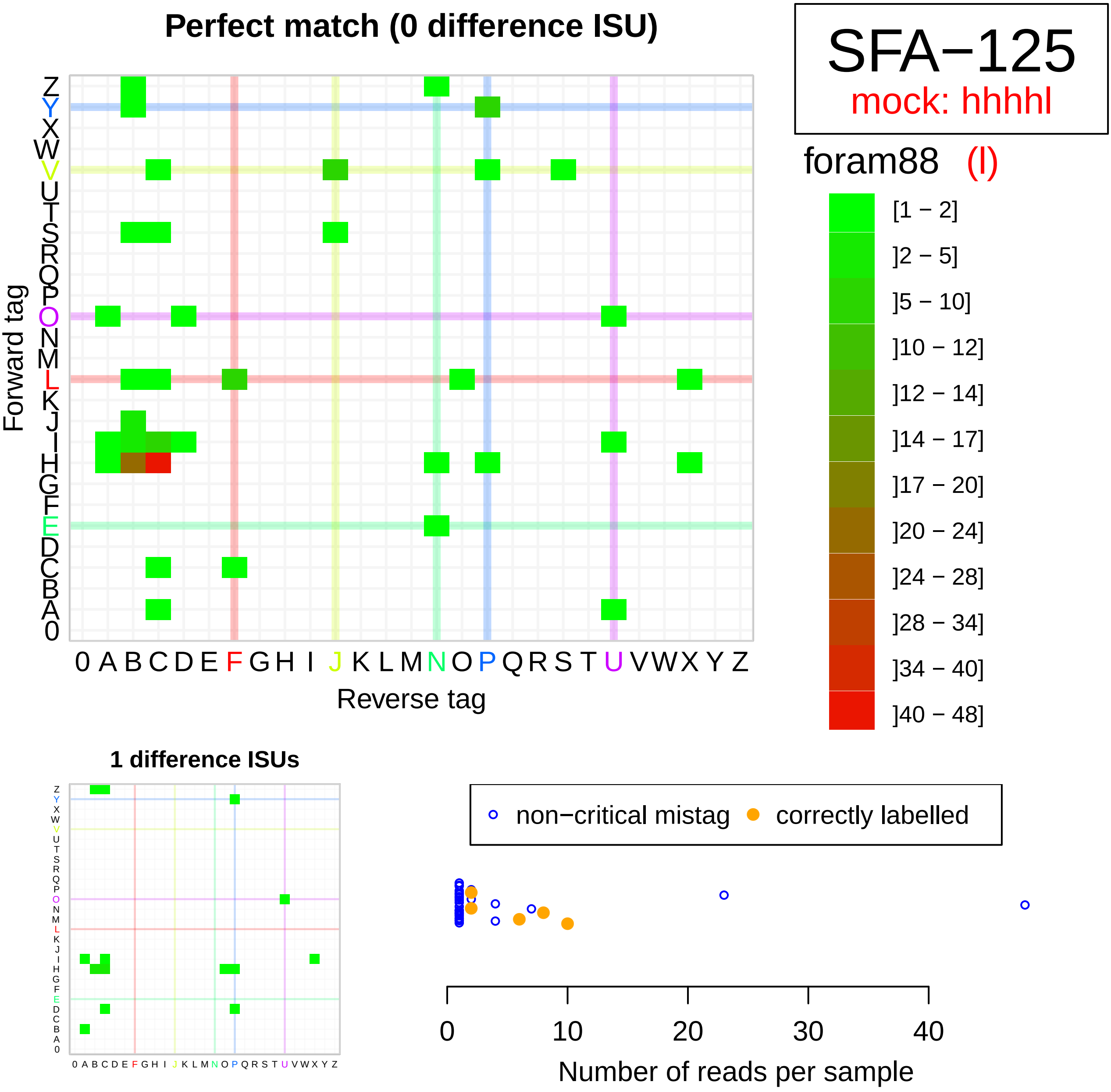
Perfect match (0 difference ISU)

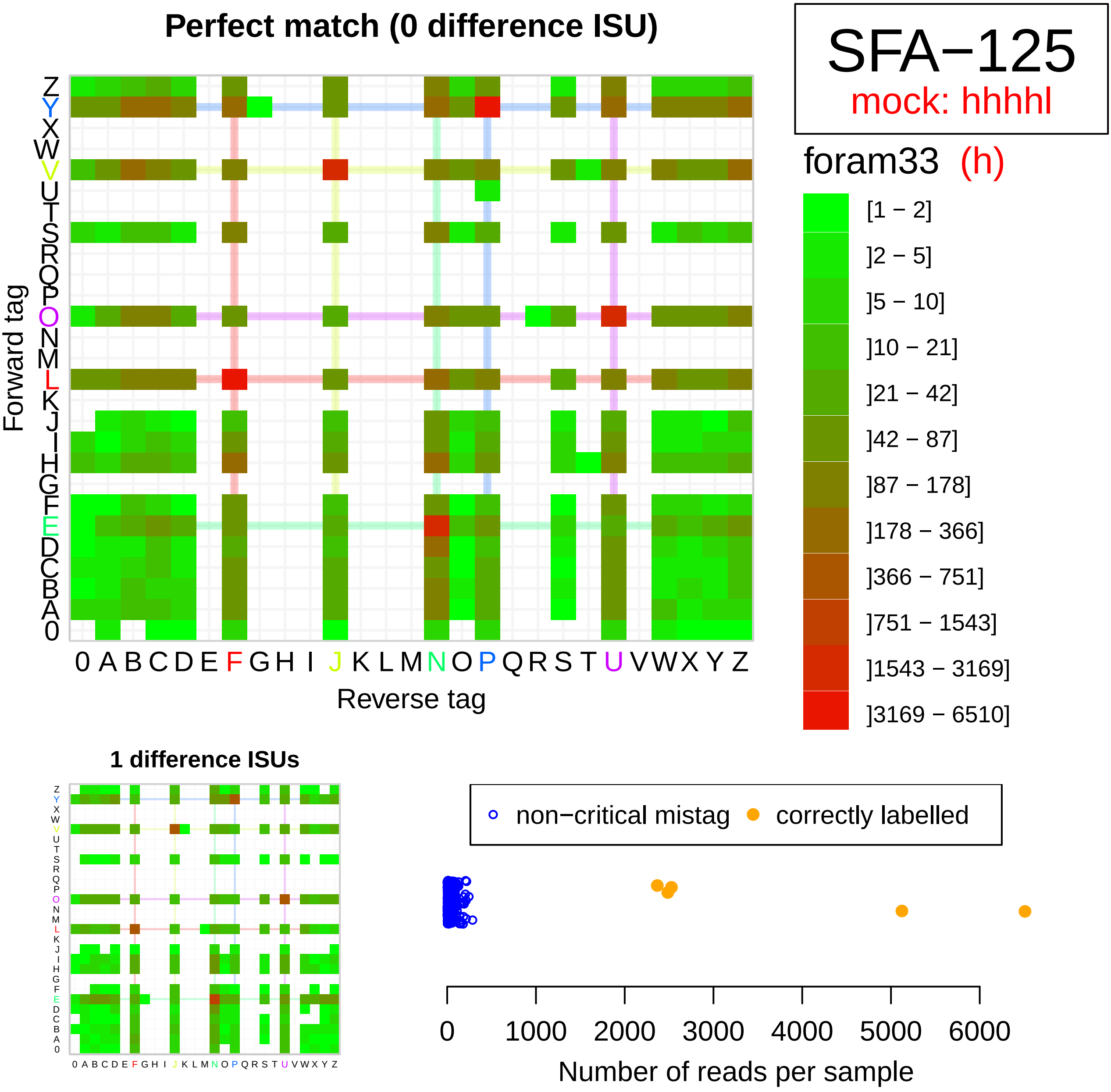


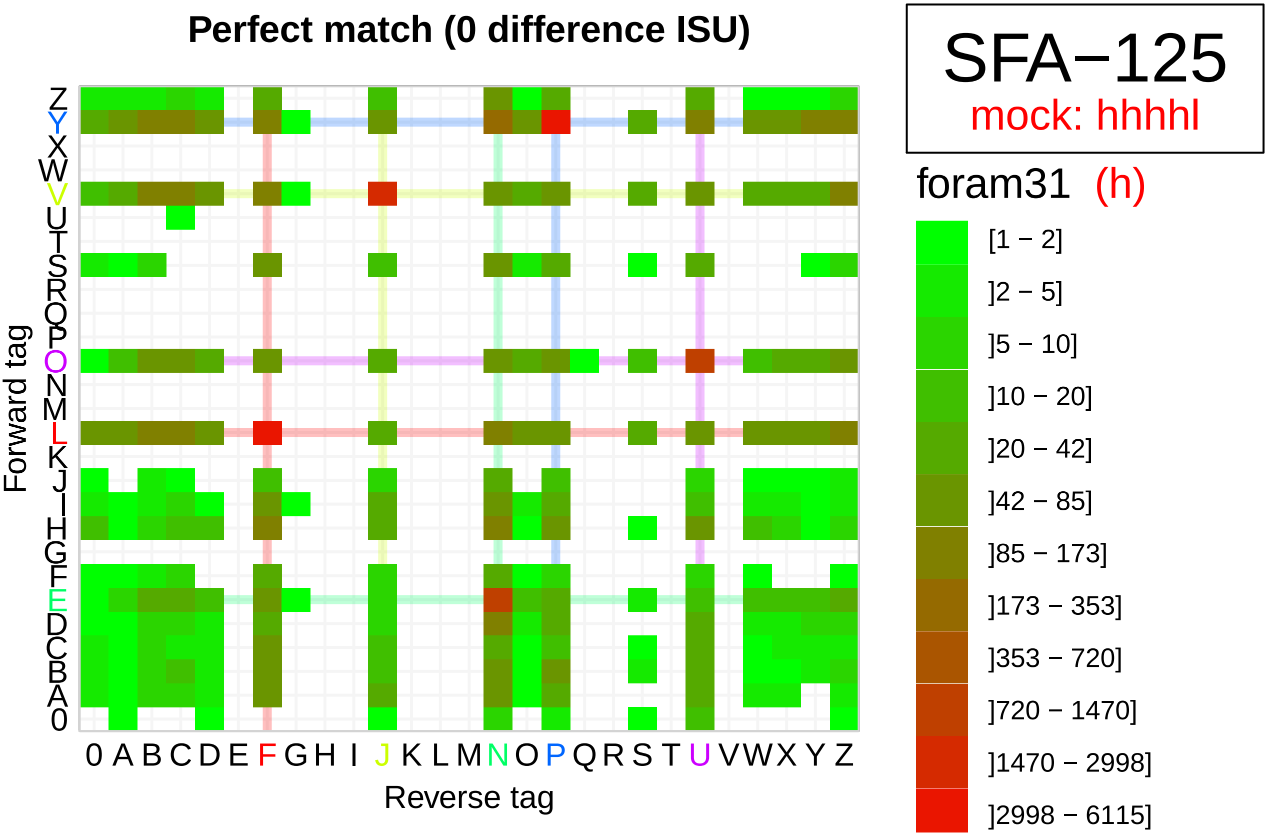
1 difference ISUs

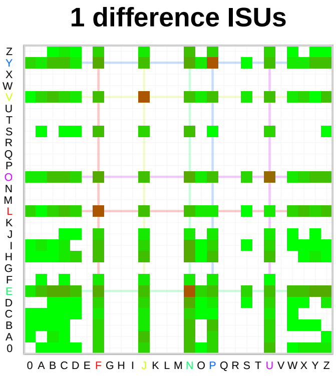


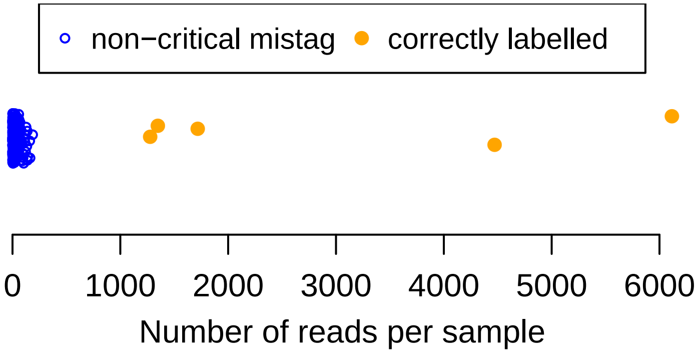


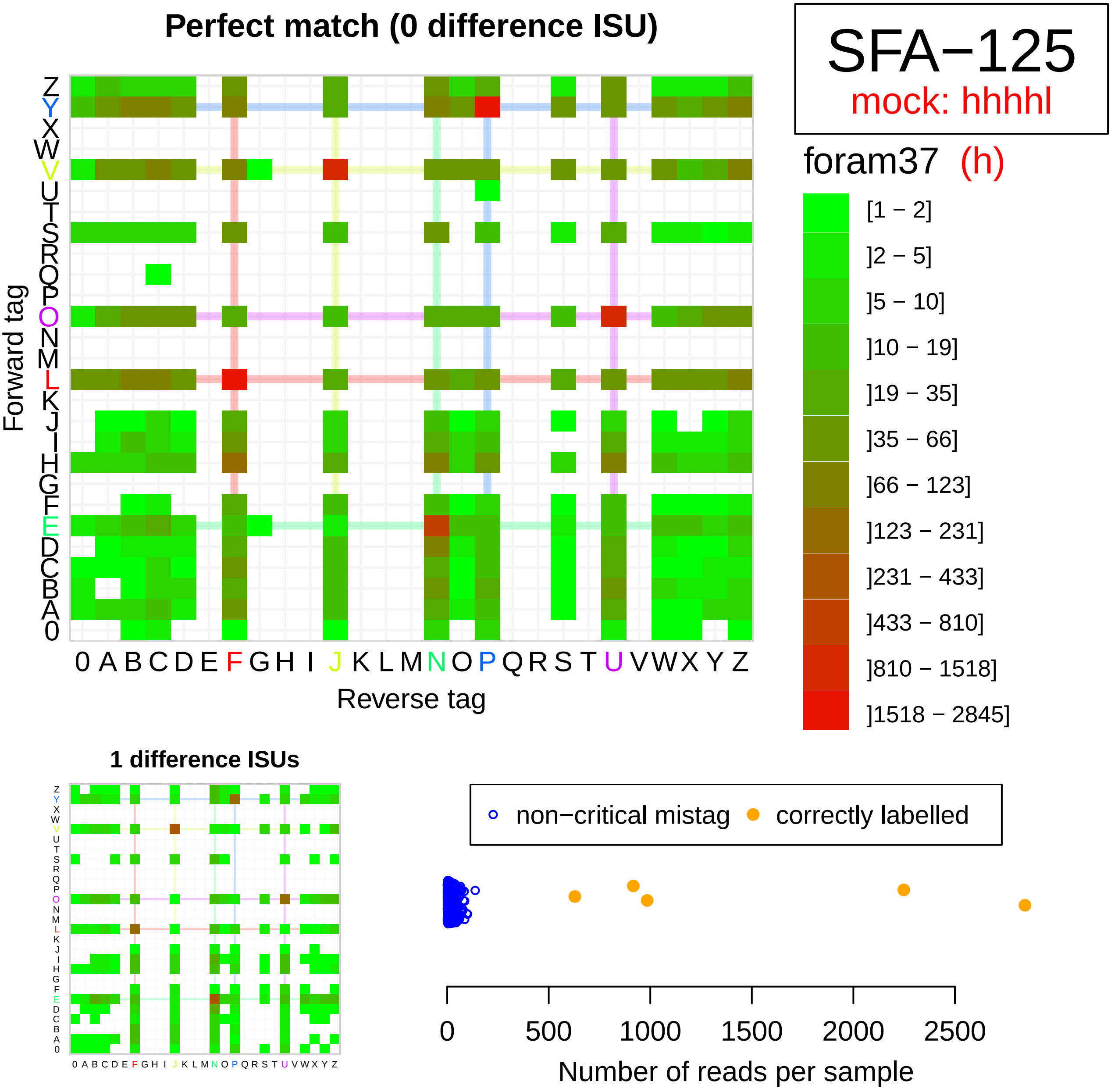

Perfect match (0 difference ISU)
SFA−125
mock: hhhhl
foram88 (l)
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 12]
]12 − 14]
]14 − 17]
]17 − 20]
]20 − 24]
]24 − 28]
]28 − 34]
]34 − 40]
]40 − 48]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
non−critical mistag
correctly labelled
0
10
20
30
40
Number of reads per sample


Perfect match (0 difference ISU)
SFA−125
mock: hhhhl
foram33 (h)
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 21]
]21 − 42]
]42 − 87]
]87 − 178]
]178 − 366]
]366 − 751]
]751 − 1543]
]1543 − 3169]
]3169 − 6510]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
non−critical mistag
correctly labelled
0
1000
2000
3000
4000
5000
6000
Number of reads per sample

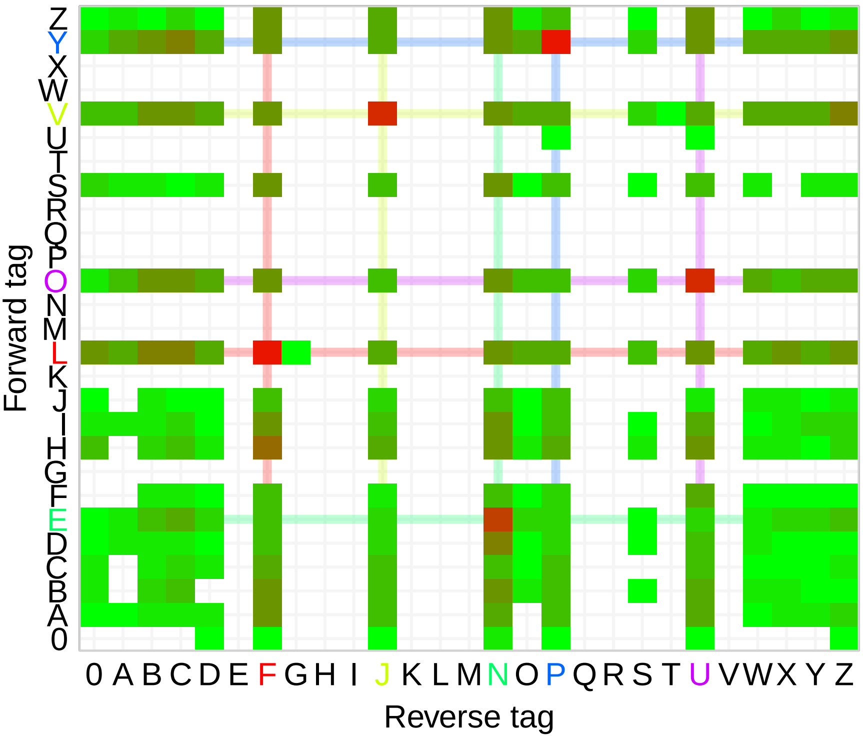

Perfect match (0 difference ISU)
SFA−125
mock: hhhhl
foram31 (h)
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 20]
]20 − 42]
]42 − 85]
]85 − 173]
]173 − 353]
]353 − 720]
]720 − 1470]
]1470 − 2998]
]2998 − 6115]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag


1 difference ISUs
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z


non−critical mistag
correctly labelled
0
1000
2000
3000
4000
5000
6000
Number of reads per sample


Perfect match (0 difference ISU)
SFA−125
mock: hhhhl
foram37 (h)
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 19]
]19 − 35]
]35 − 66]
]66 − 123]
]123 − 231]
]231 − 433]
]433 − 810]
]810 − 1518]
]1518 − 2845]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
non−critical mistag
correctly labelled
0
500
1000
1500
2000
2500
Number of reads per sample

Perfect match (0 difference ISU)

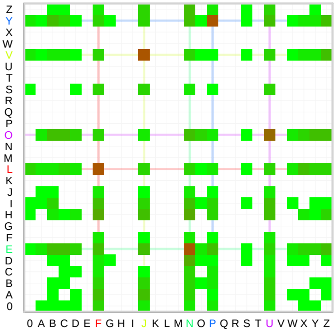


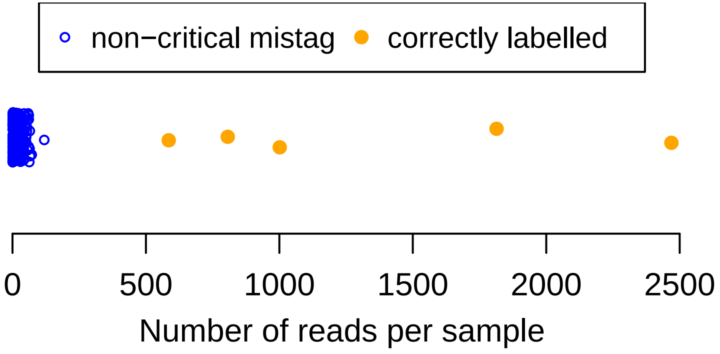
SFA−125
mock: hhhhl

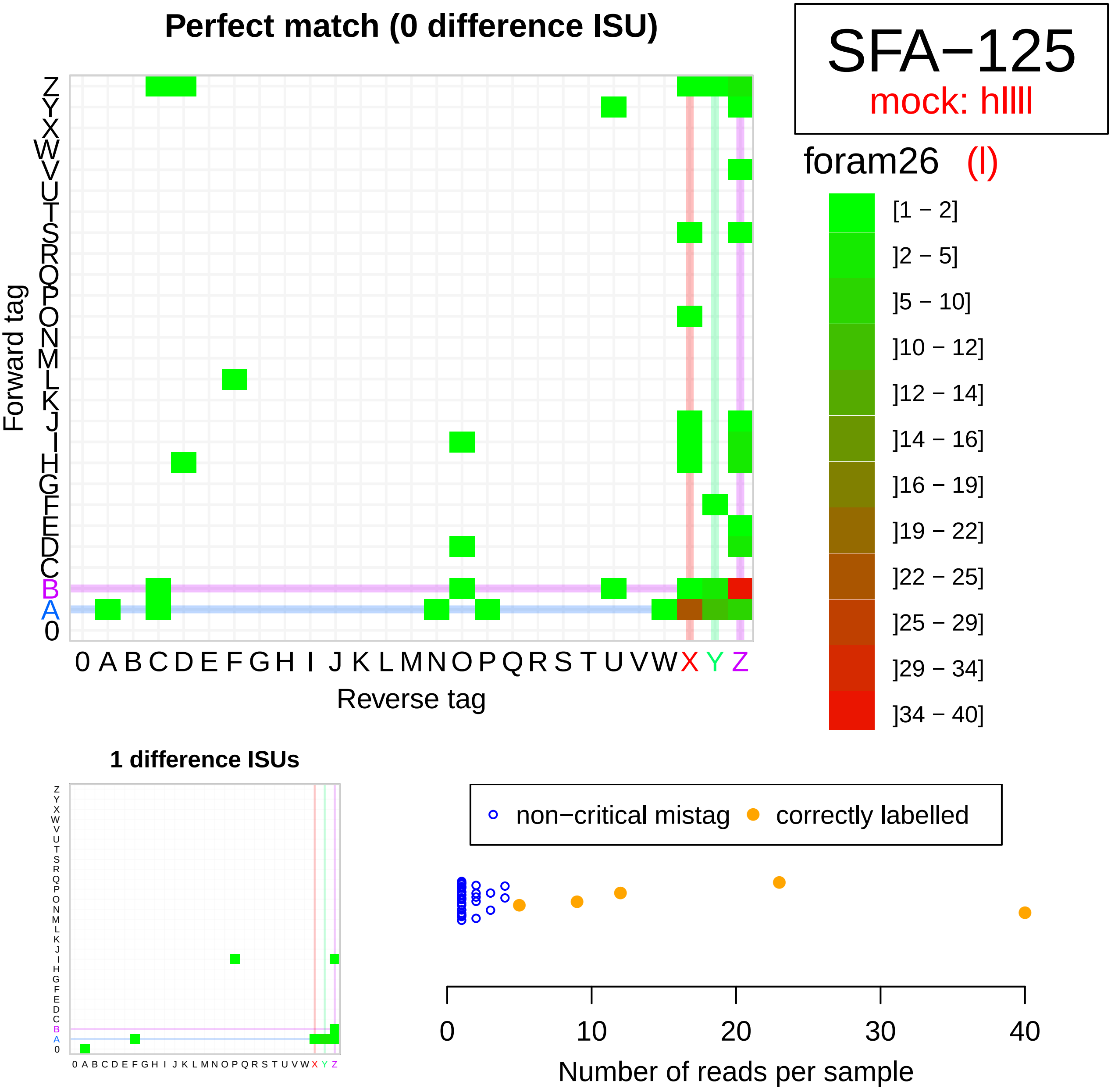
foram45 (h)

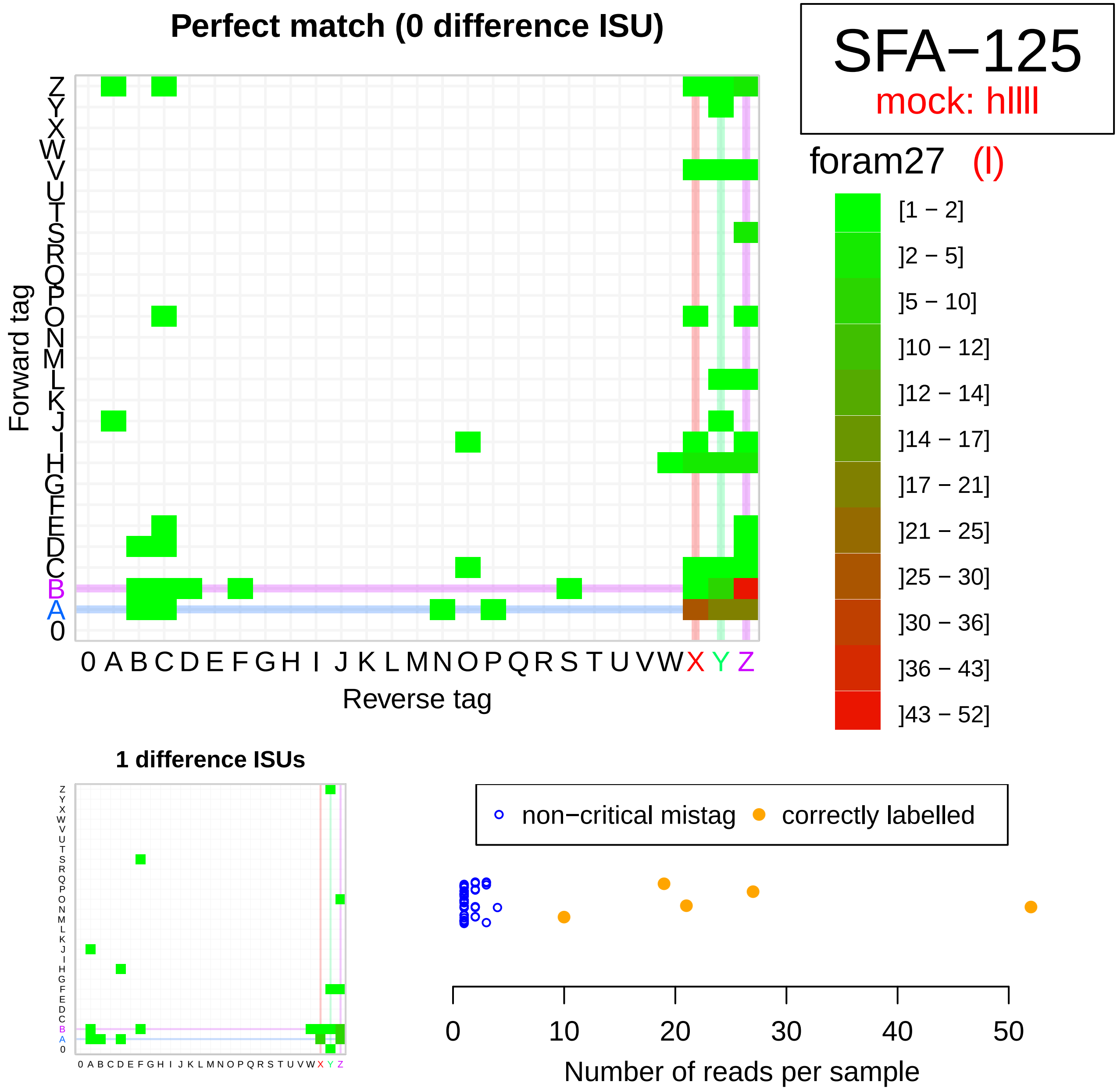
- 🟩 [1 − 2]
- 🟩]2 − 5]
- 🟩]5 − 10]
- 🟩]10 − 18]
- 🟩]18 − 34]
- 🟩]34 − 63]
- 🟫]63 − 116]
- 🟫]116 − 213]
- 🟫]213 − 394]
- 🟥]394 − 726]
- 🟥]726 − 1339]
- 🟥]1339 − 2469]

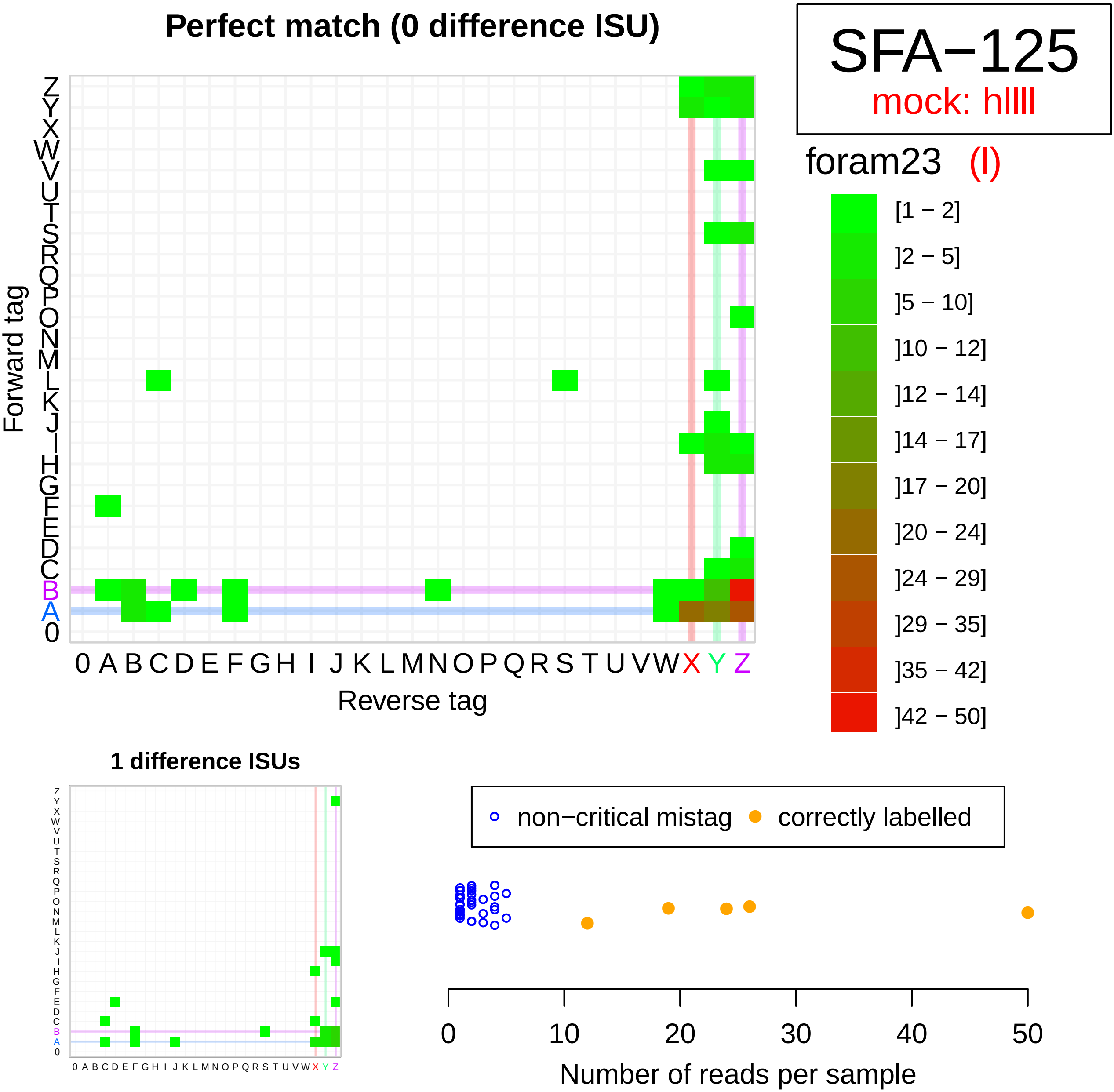
1 difference ISUs

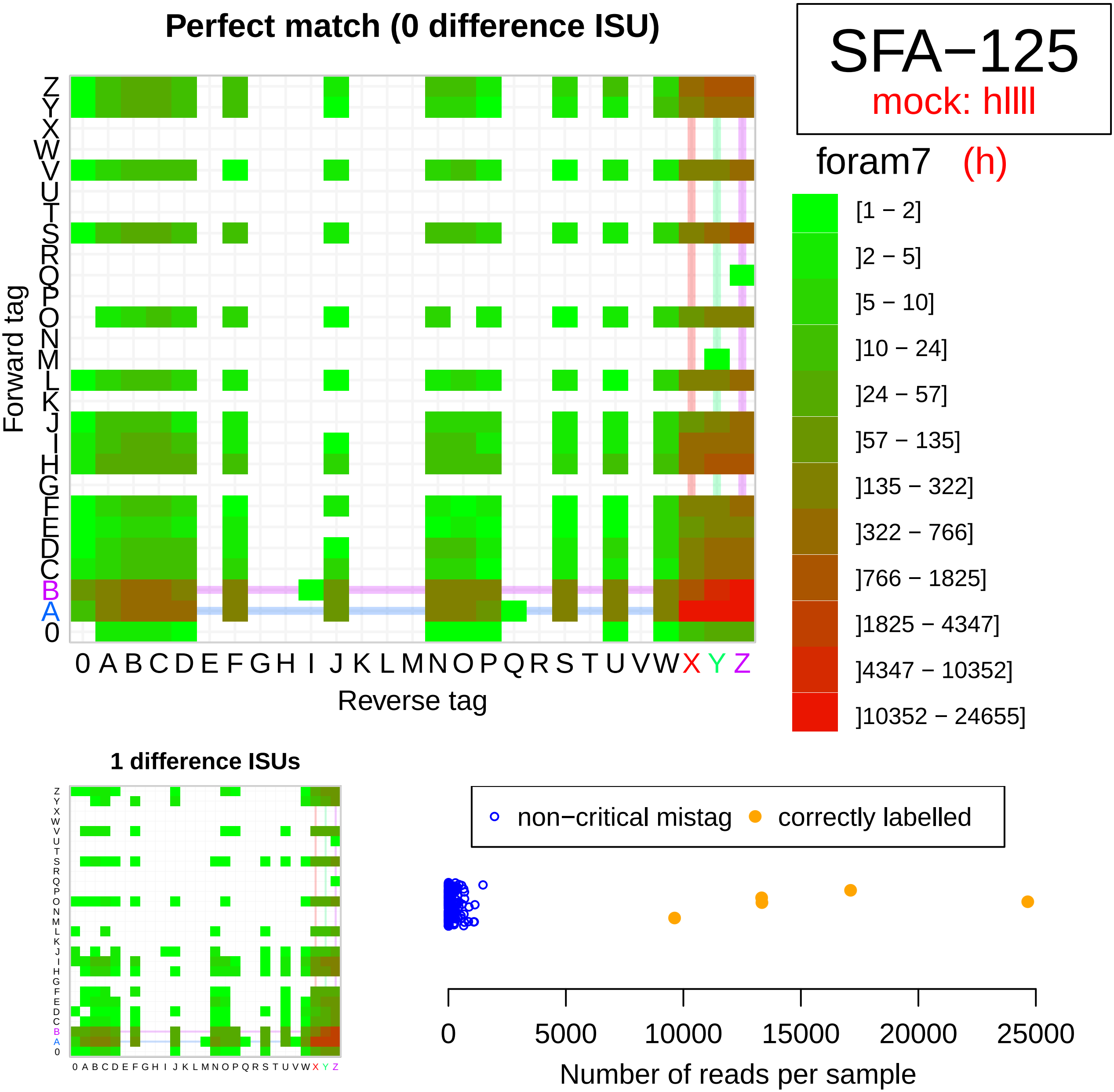





Perfect match (0 difference ISU)
SFA−125
mock: hllll
foram26 (l)
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 12]
]12 − 14]
]14 − 16]
]16 − 19]
]19 − 22]
]22 − 25]
]25 − 29]
]29 − 34]
]34 − 40]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
non−critical mistag
correctly labelled
0
10
20
30
40
Number of reads per sample


Perfect match (0 difference ISU)
SFA−125
mock: hllll
foram27 (l)
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 12]
]12 − 14]
]14 − 17]
]17 − 21]
]21 − 25]
]25 − 30]
]30 − 36]
]36 − 43]
]43 − 52]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
non−critical mistag
correctly labelled
0
10
20
30
40
50
Number of reads per sample


Perfect match (0 difference ISU)
SFA−125
mock: hllll
foram23 (l)
Forward tag
Reverse tag
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 12]
]12 − 14]
]14 − 17]
]17 − 20]
]20 − 24]
]24 − 29]
]29 − 35]
]35 − 42]
]42 − 50]
1 difference ISUs
non−critical mistag
correctly labelled
0
10
20
30
40
50
Number of reads per sample



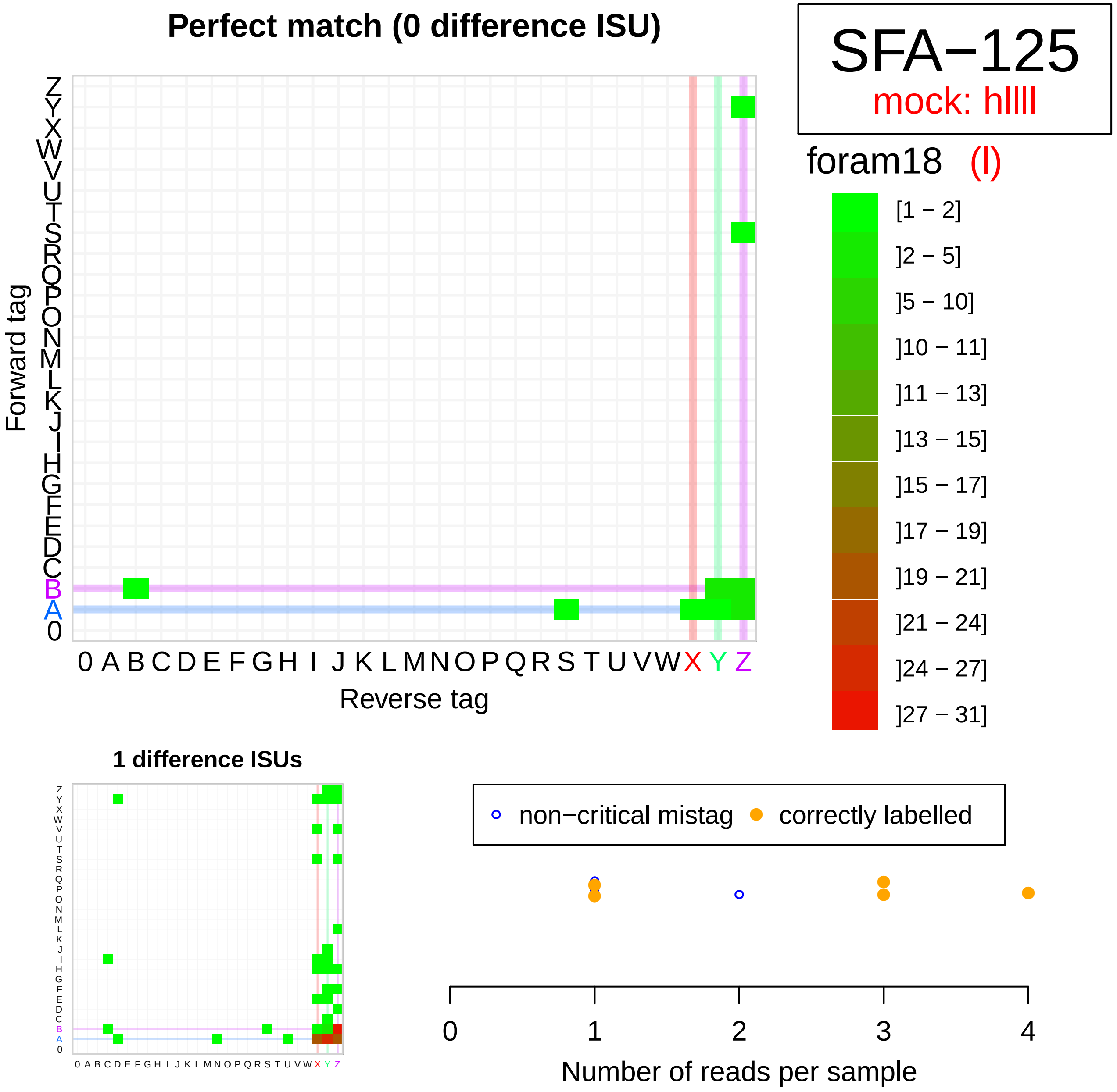
Perfect match (0 difference ISU)

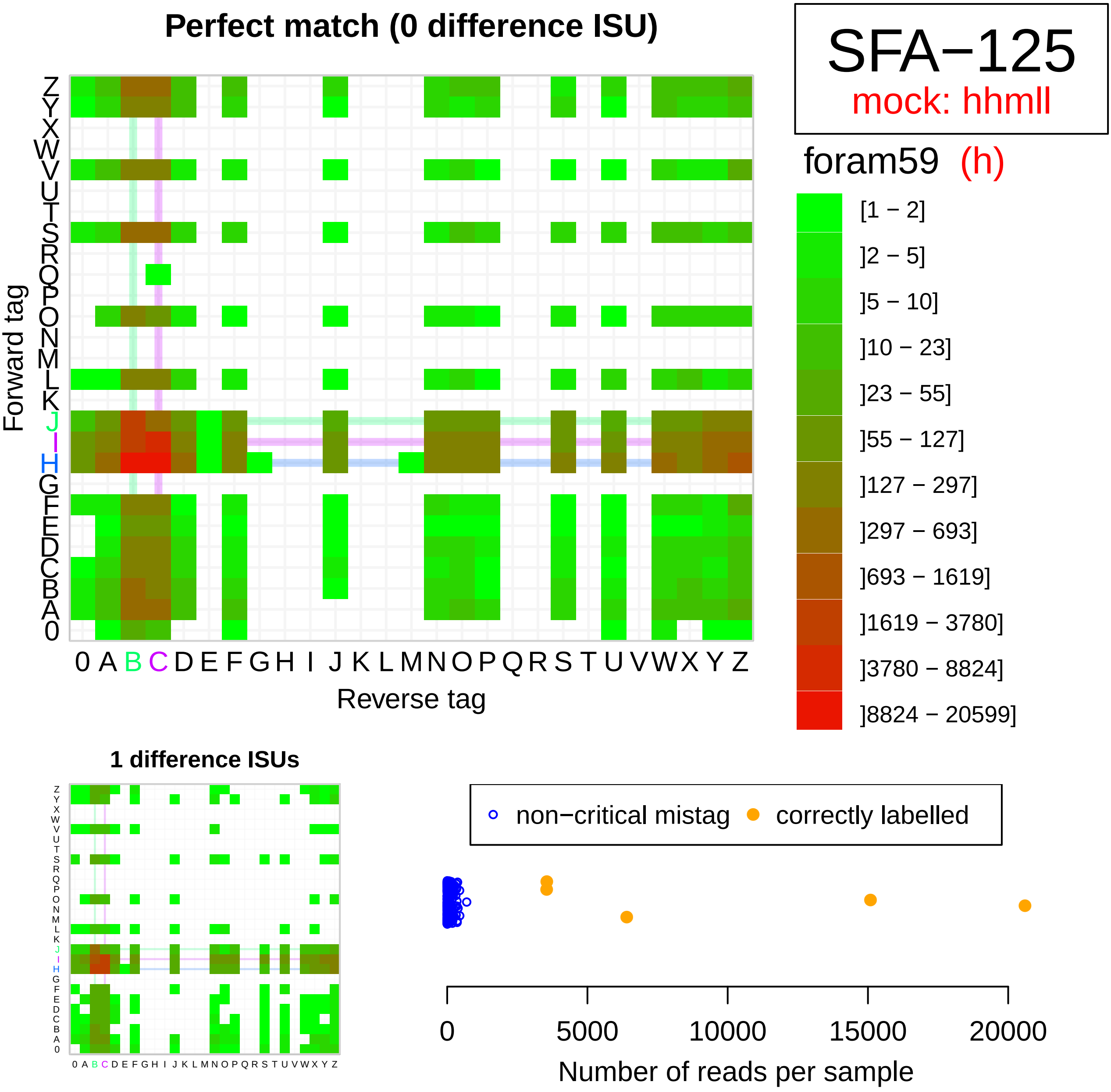
SFA−125

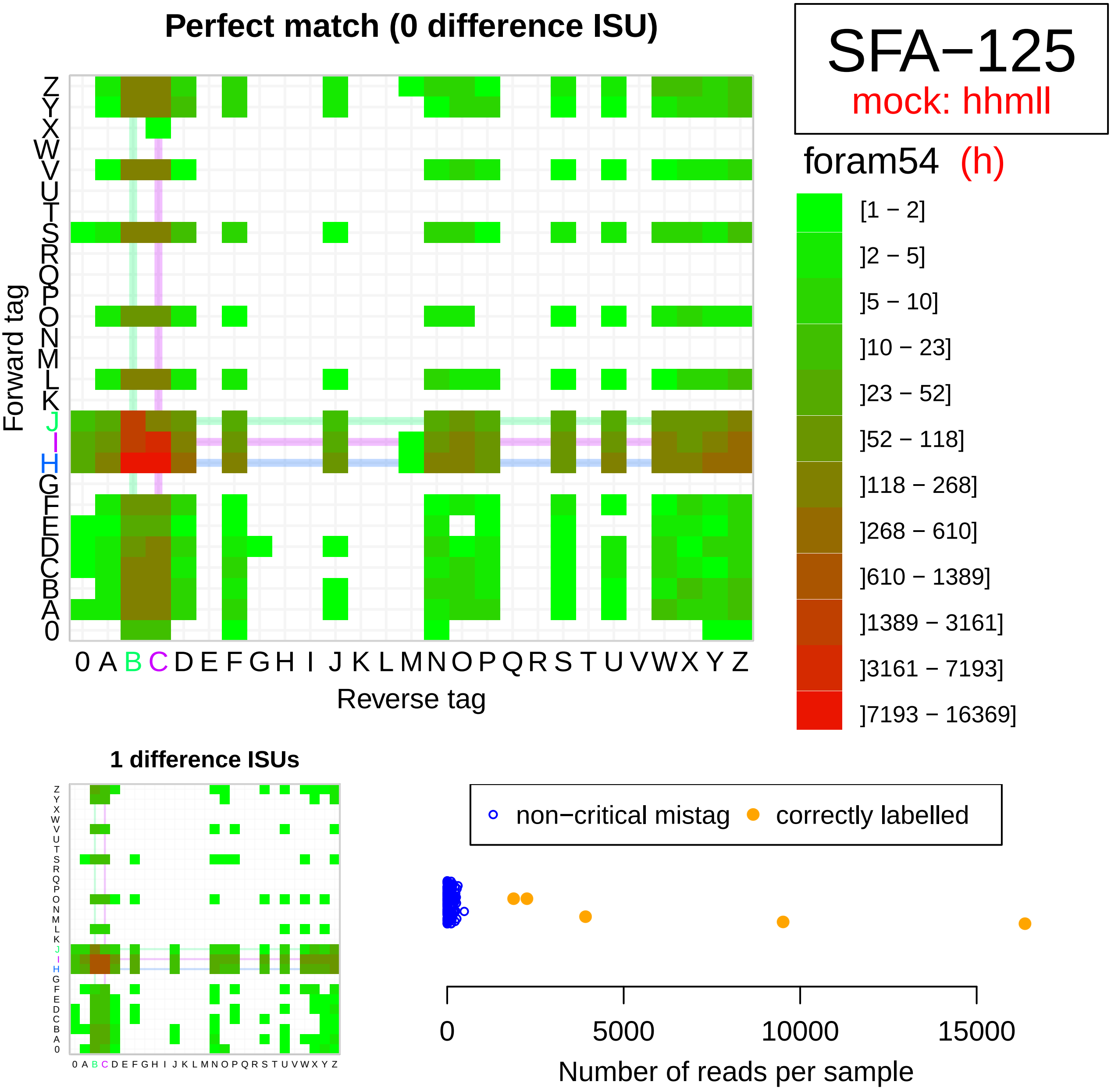
mock: hllll

foram7 (h)

1 difference ISUs


Perfect match (0 difference ISU)
SFA−125
mock: hllll
foram18 (l)
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 11]
]11 − 13]
]13 − 15]
]15 − 17]
]17 − 19]
]19 − 21]
]21 − 24]
]24 − 27]
]27 − 31]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
non−critical mistag
correctly labelled
0
1
2
3
4
Number of reads per sample

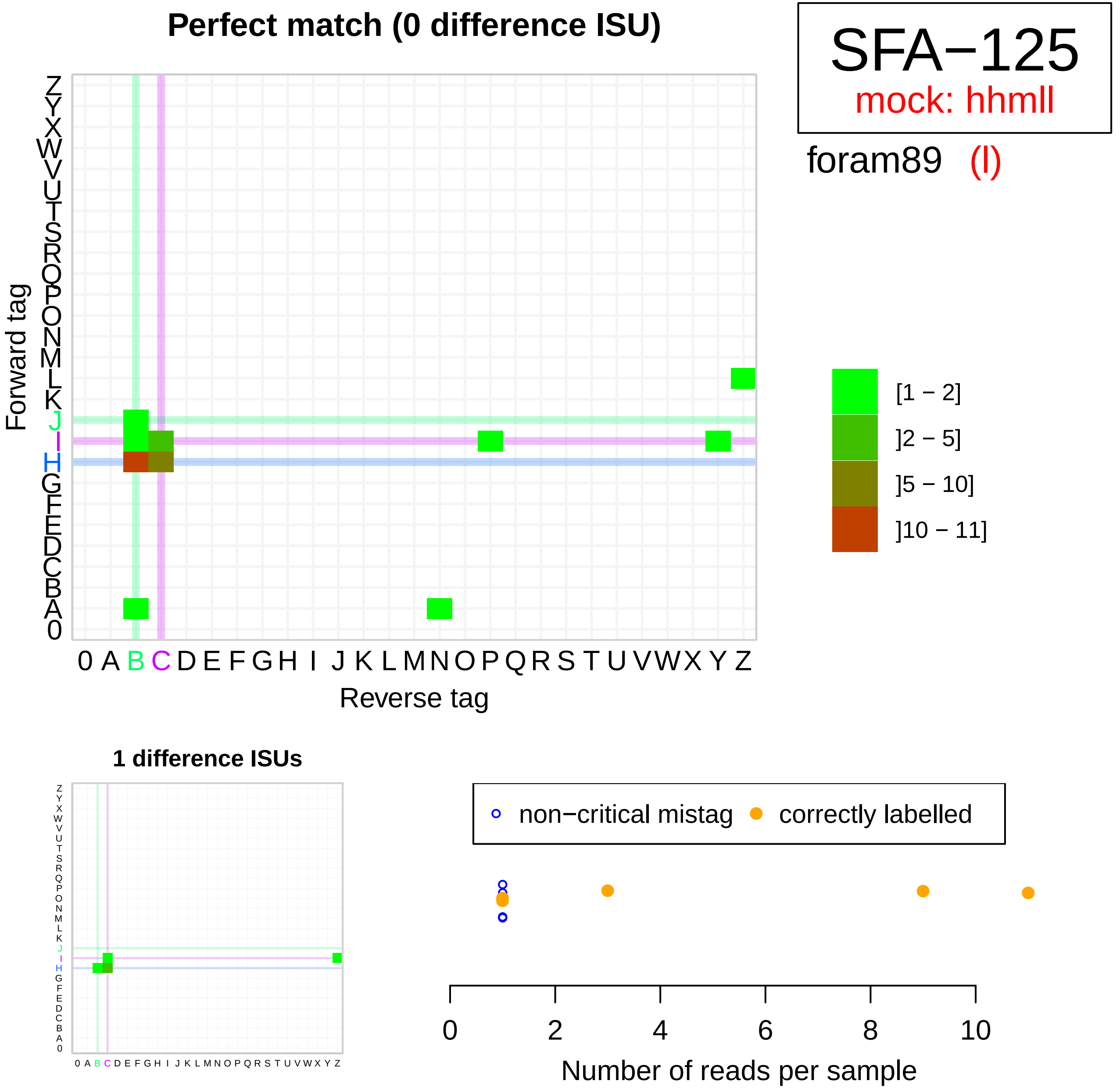

Perfect match (0 difference ISU)
SFA−125
mock: hhmll
foram59 (h)
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 23]
]23 − 55]
]55 − 127]
]127 − 297]
]297 − 693]
]693 − 1619]
]1619 − 3780]
]3780 − 8824]
]8824 − 20599]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
non−critical mistag
correctly labelled
0
5000
10000
15000
20000
Number of reads per sample


Perfect match (0 difference ISU)
SFA−125
mock: hhmll
foram54 (h)
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 23]
]23 − 52]
]52 − 118]
]118 − 268]
]268 − 610]
]610 − 1389]
]1389 − 3161]
]3161 − 7193]
]7193 − 16369]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
non−critical mistag
correctly labelled
0
5000
10000
15000
Number of reads per sample

SFA−125

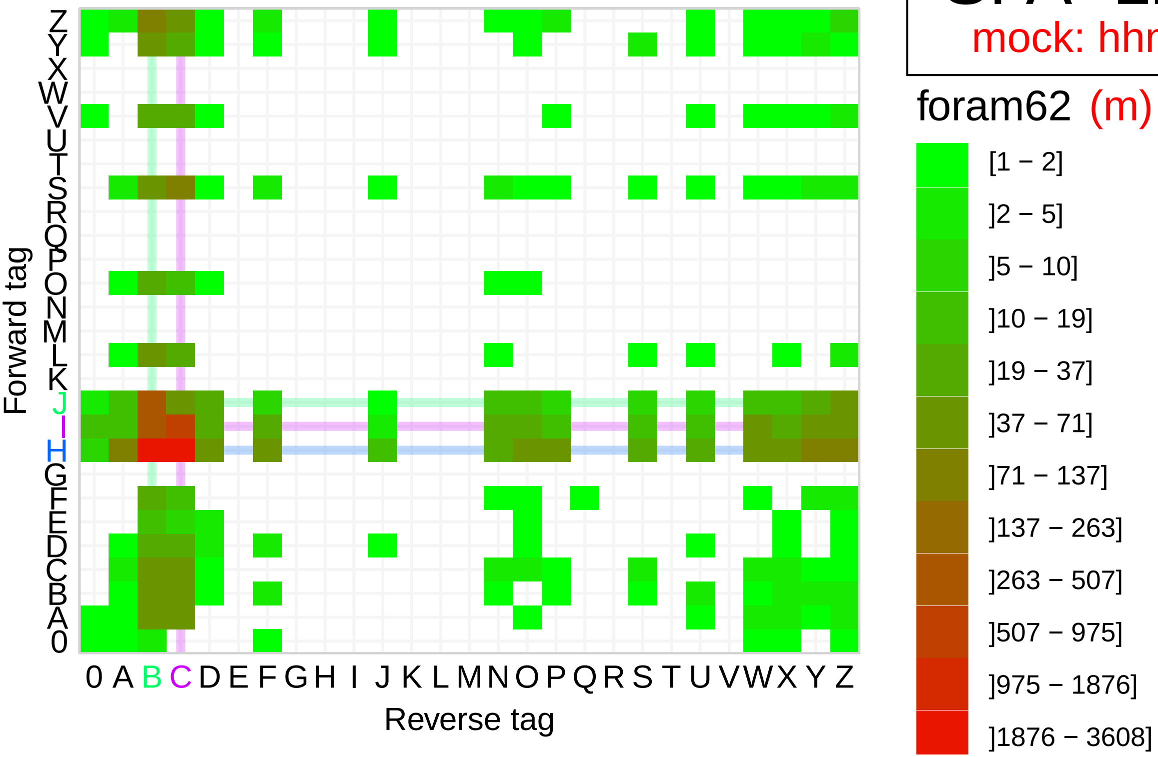
mock: hhmll

foram89 (I)

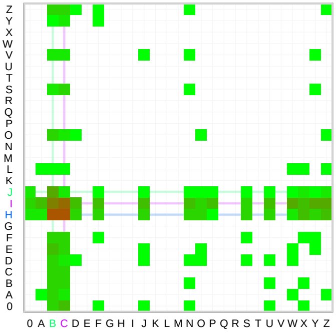


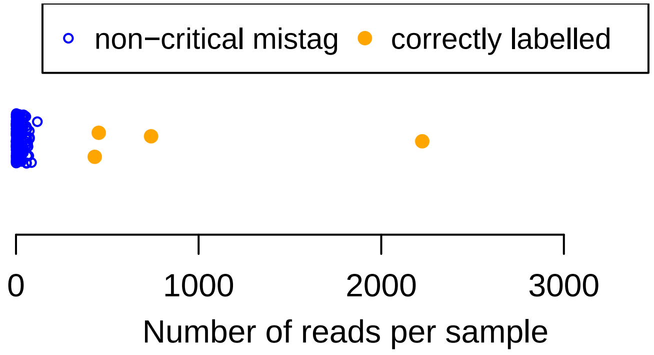
SFA−125
mock: hhmll

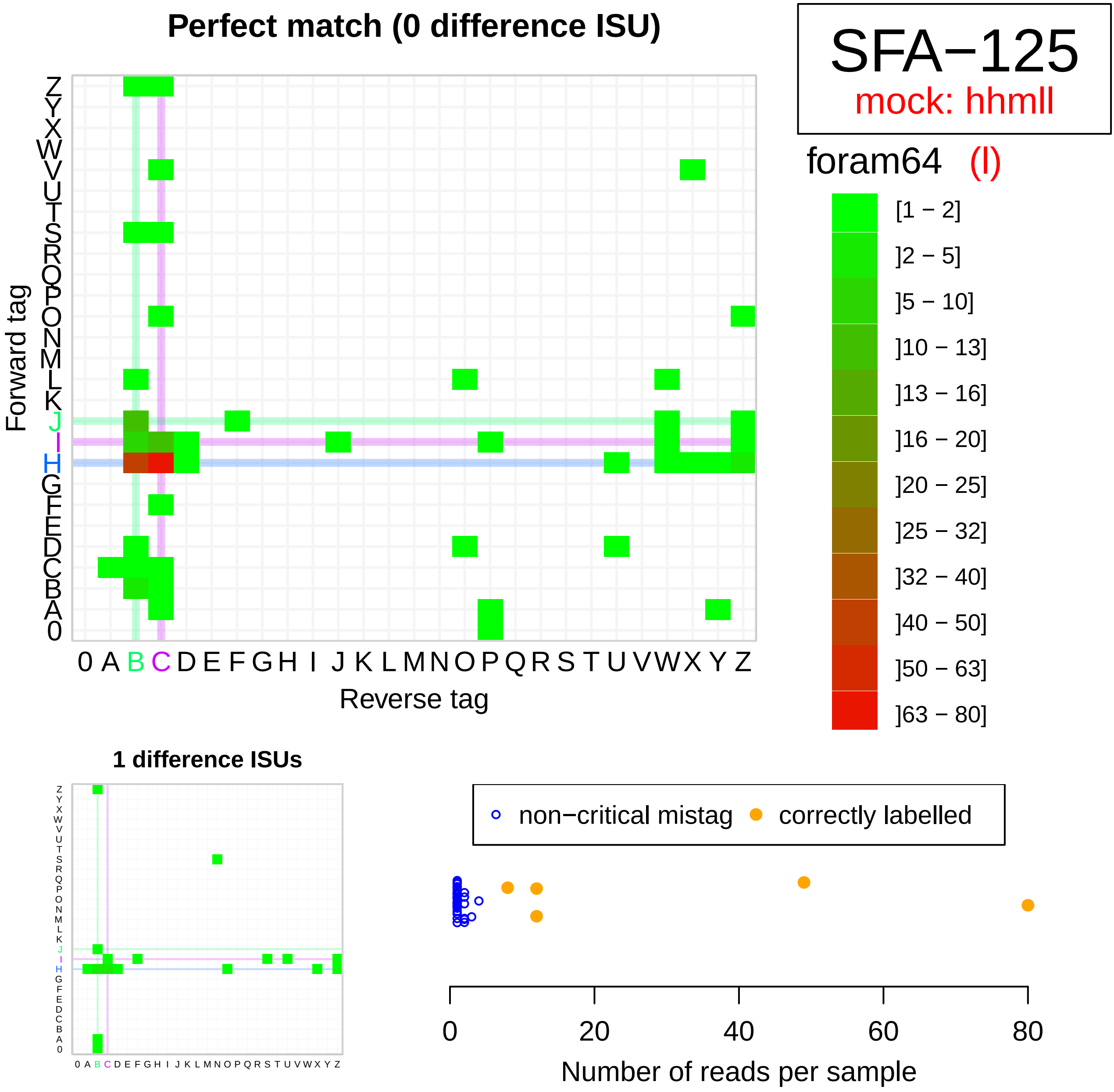
Perfect match (0 difference ISU)

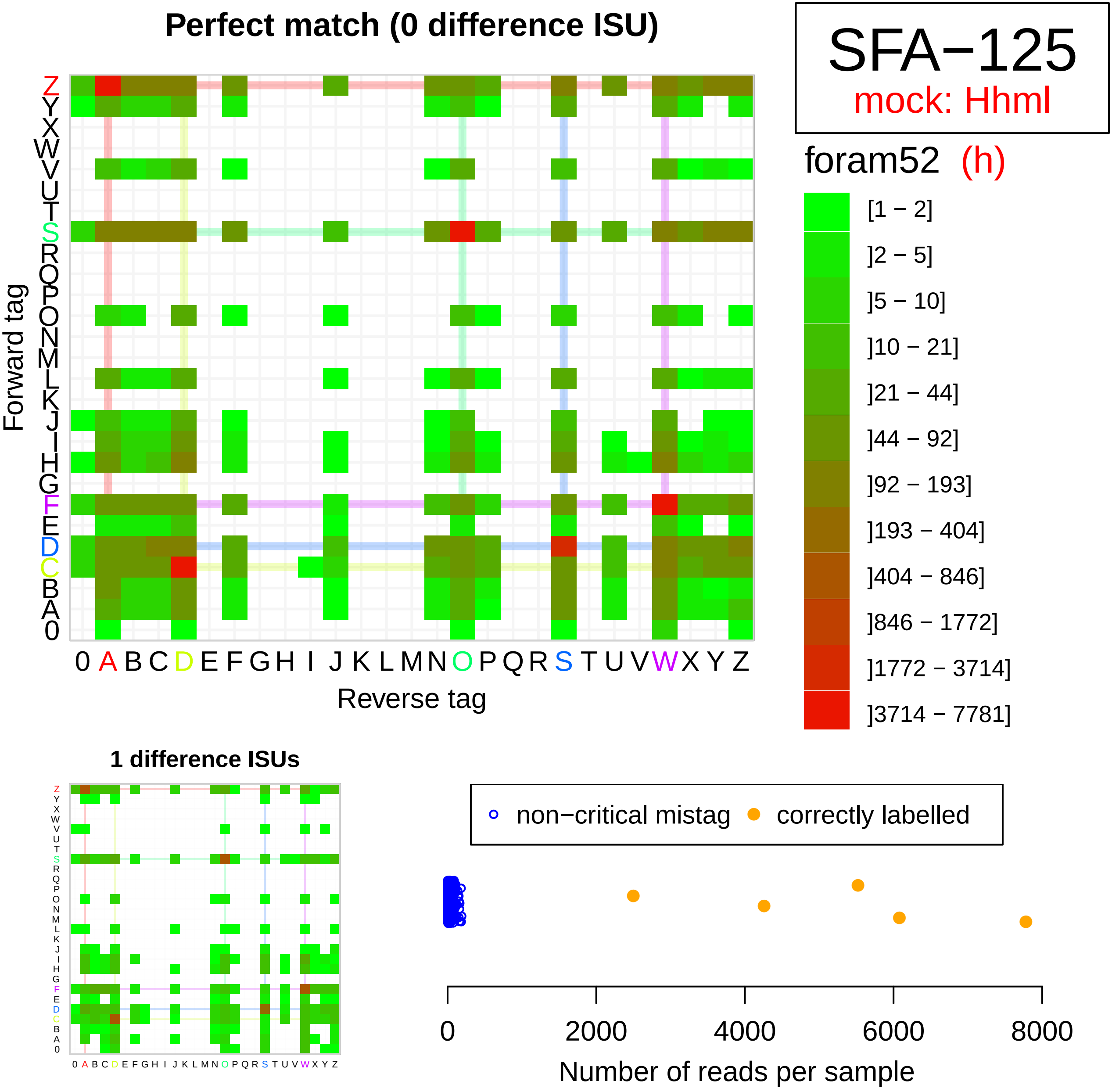


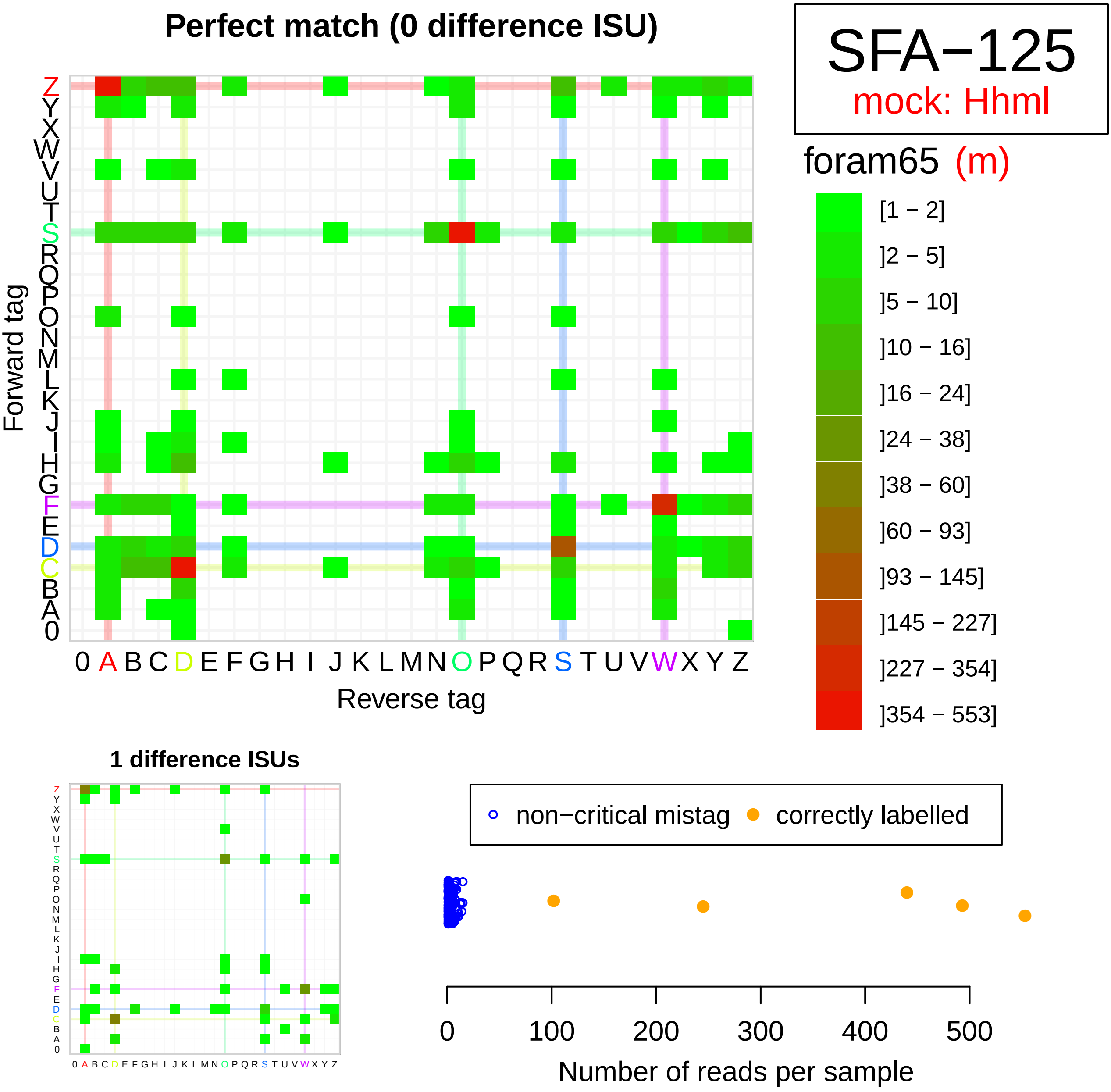
1 difference ISUs

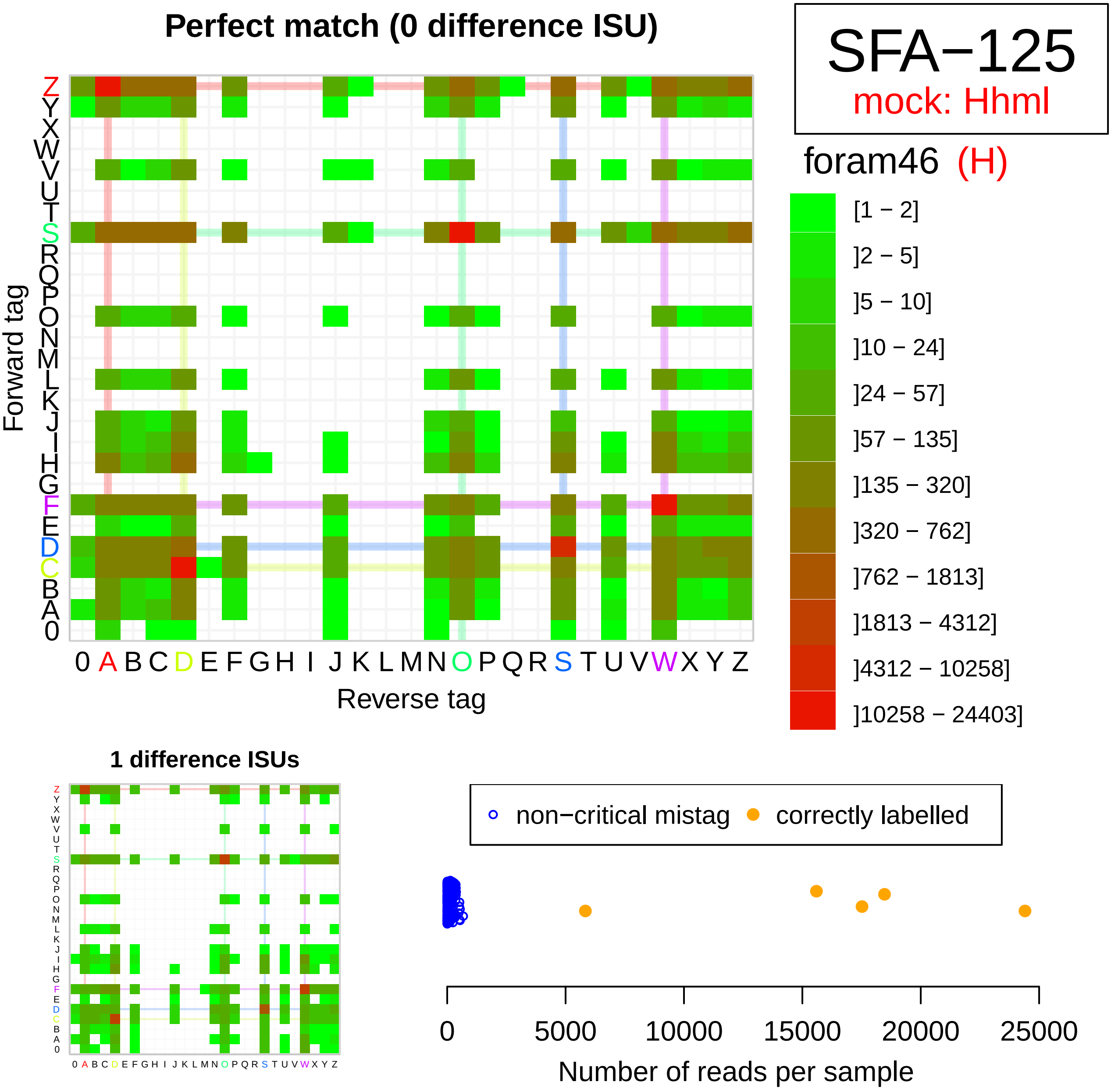





SFA−125
mock: hhmll
foram64 (l)
Perfect match (0 difference ISU)
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
Reverse tag
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 13]
]13 − 16]
]16 − 20]
]20 − 25]
]25 − 32]
]32 − 40]
]40 − 50]
]50 − 63]
]63 − 80]
1 difference ISUs
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
○ non−critical mistag
● correctly labelled
0
20
40
60
80
Number of reads per sample

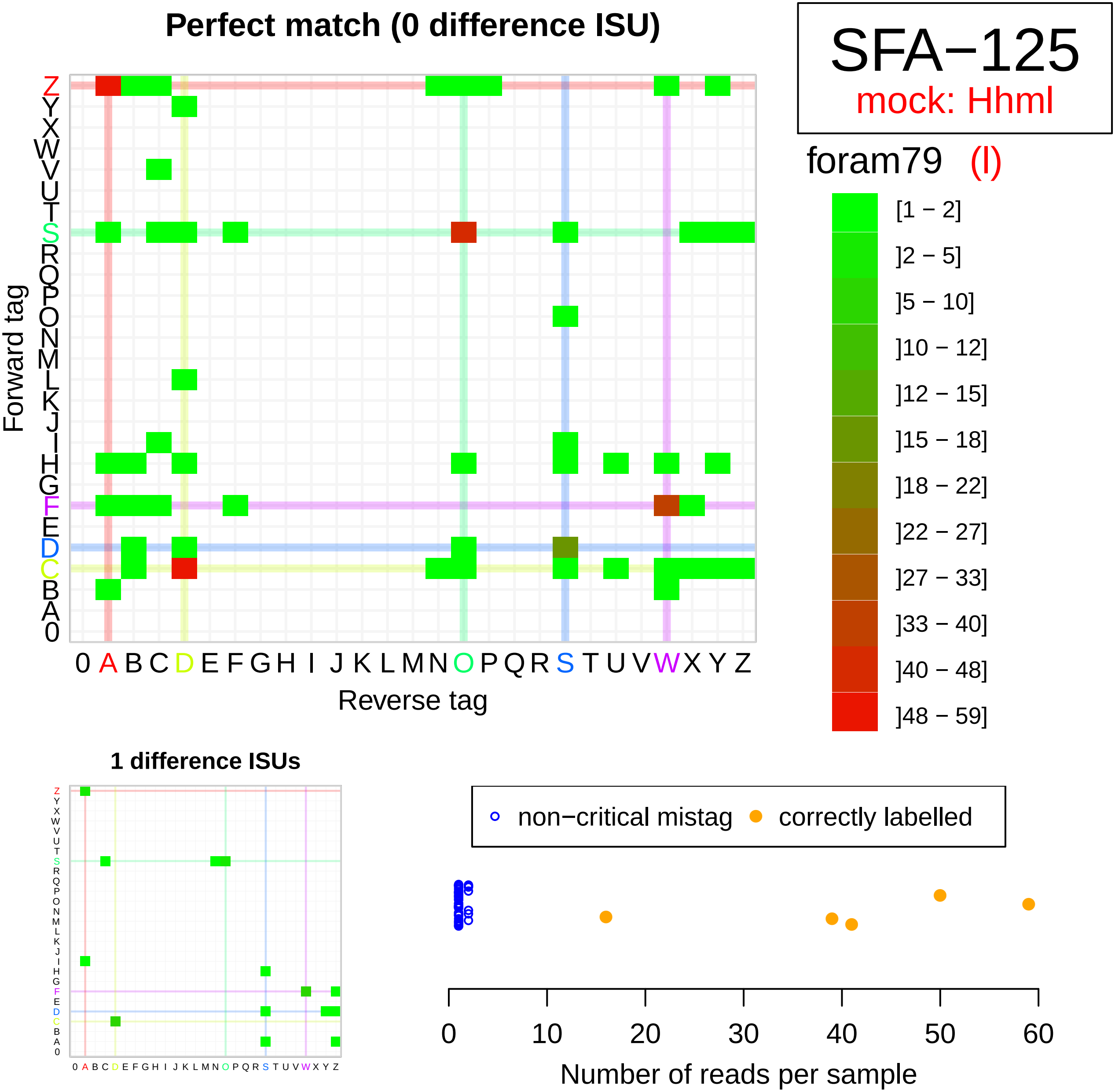

Perfect match (0 difference ISU)
Z
Y
X
W
V
U
T
S
R
Q
P
O
N
M
L
K
J
I
H
G
F
E
D
C
B
A
0
Forward tag
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
SFA−125
mock: Hhml
foram52 (h)
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 21]
]21 − 44]
]44 − 92]
]92 − 193]
]193 − 404]
]404 − 846]
]846 − 1772]
]1772 − 3714]
]3714 − 7781]
1 difference ISUs
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
non−critical mistag
correctly labelled
0
2000
4000
6000
8000
Number of reads per sample


Perfect match (0 difference ISU)
SFA−125
mock: Hhml
foram65 (m)
🟩 [1 − 2]
🟩]2 − 5]
🟩]5 − 10]
🟩]10 − 16]
🟩]16 − 24]
🟩]24 − 38]
🟫]38 − 60]
🟫]60 − 93]
🟫]93 − 145]
🟥]145 − 227]
🟥]227 − 354]
🟥]354 − 553]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
○ non−critical mistag
● correctly labelled
0
100
200
300
400
500
Number of reads per sample

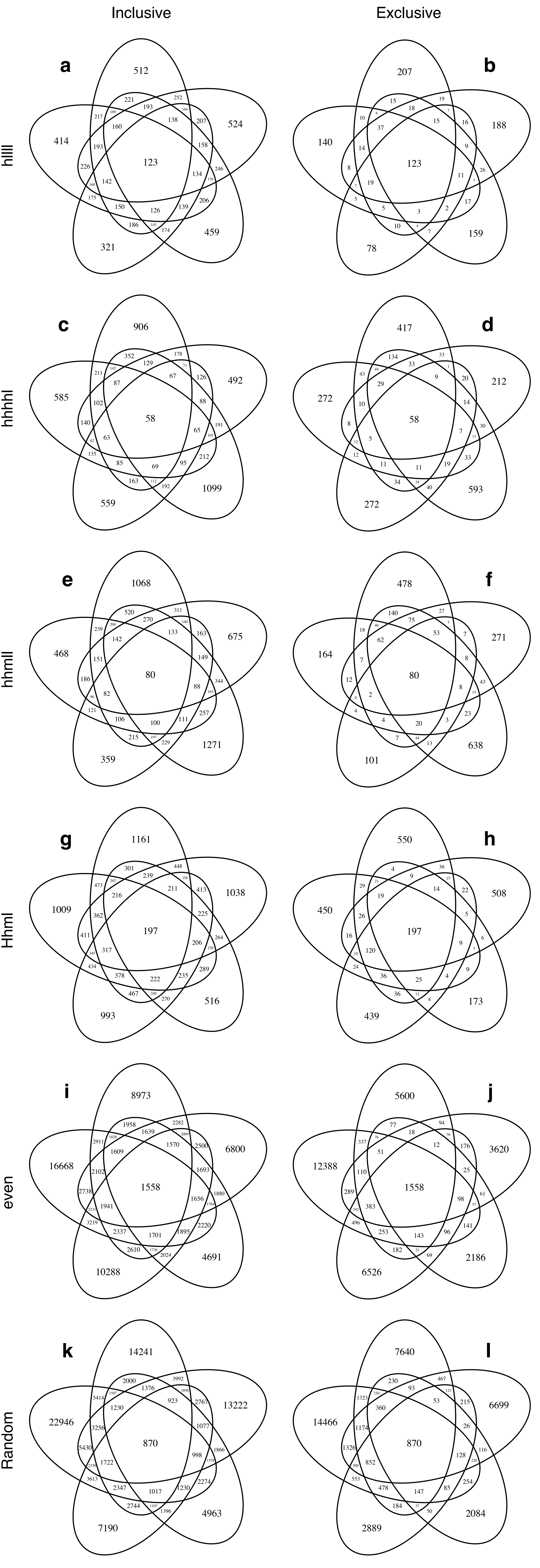


Perfect match (0 difference ISU)

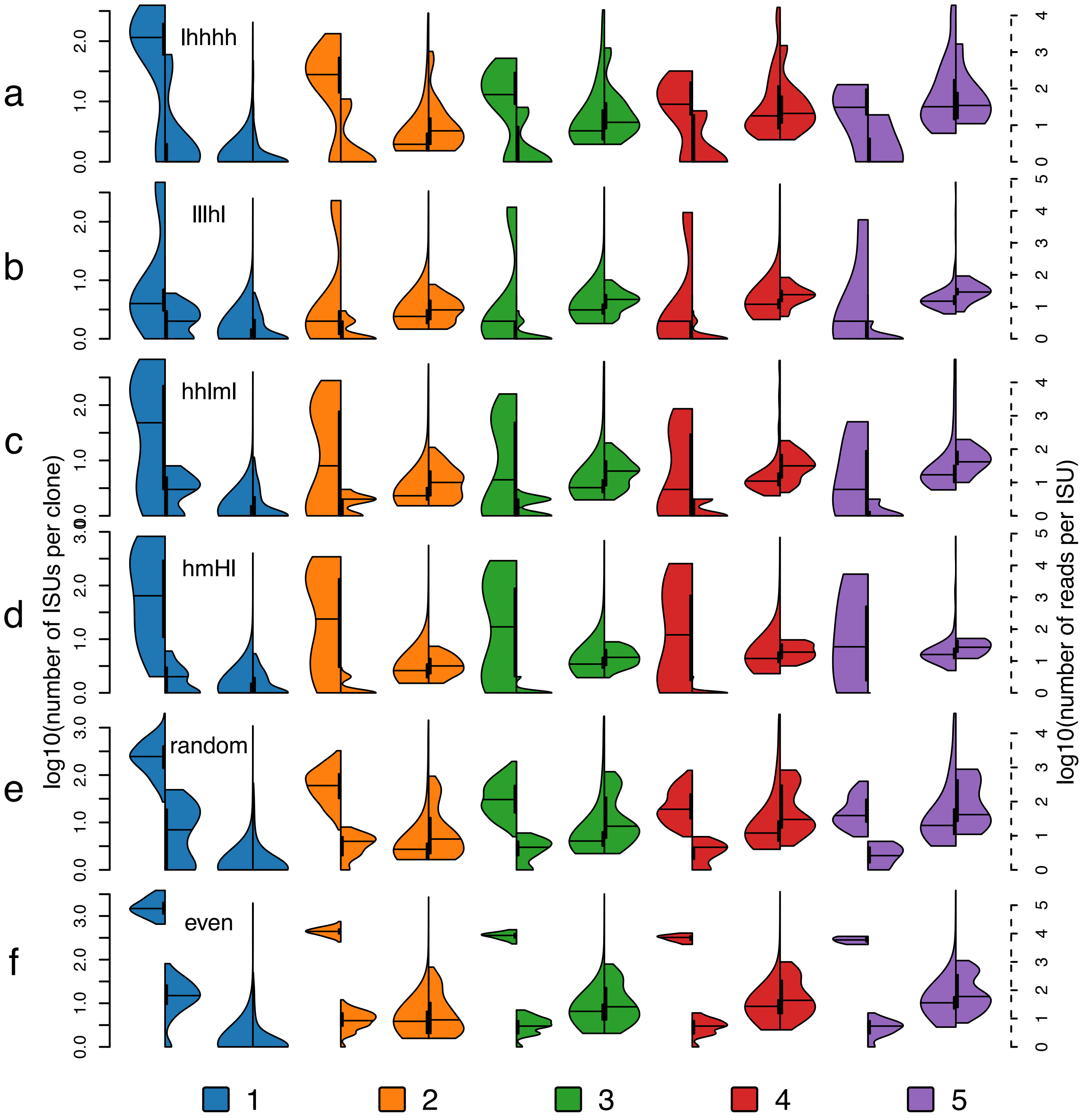
1 difference ISUs


Perfect match (0 difference ISU)
SFA−125
mock: Hhml
foram79 (I)
[1 − 2]
]2 − 5]
]5 − 10]
]10 − 12]
]12 − 15]
]15 − 18]
]18 − 22]
]22 − 27]
]27 − 33]
]33 − 40]
]40 − 48]
]48 − 59]
Forward tag
Z Y X W V U T S R Q P O N M L K J I H G F E D C B A 0
0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Reverse tag
1 difference ISUs
non−critical mistag
correctly labelled
0
10
20
30
40
50
60
Number of reads per sample

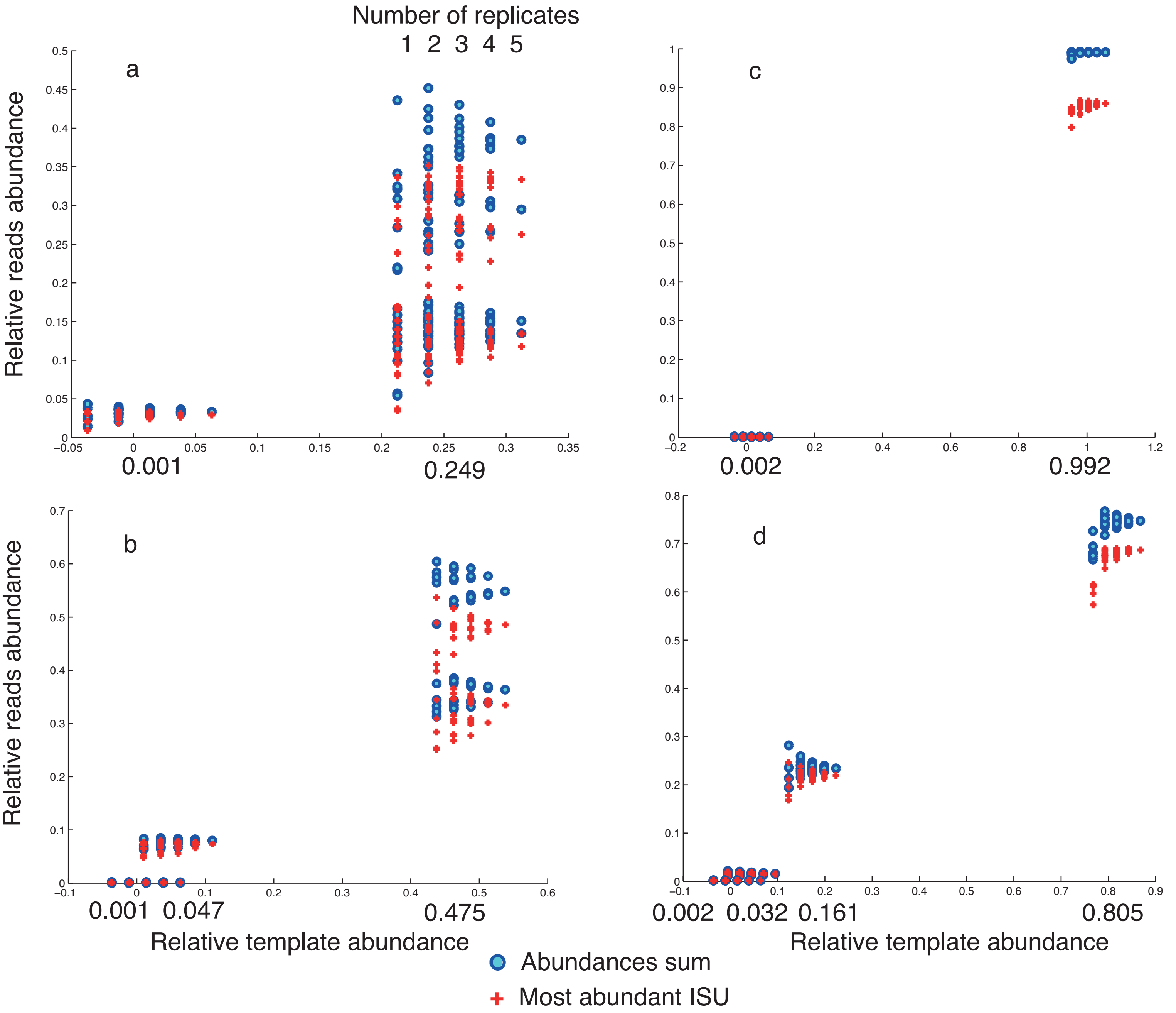
Supplementary Figure 10. Distribution of ISUs in each possible intersection of PCR replicate samples. The ISU distributions are indicated by two 5-way Venn diagrams for each mock community: *hllll* (a, b), *hhhhl* (c, d), *hhmll* (e, f), *Hhml* (g, h), *even* (i, j) and *random* (k, l). In each intersection area of the “Inclusive” diagrams (left side), the numbers correspond to the total numbers of ISUs that would be found, including those that would also be found using more replicates. In the “Exclusive” diagrams (right side), an ISU already counted in a given intersection area is not re-counted in the intersection areas involving the same replicates.


Inclusive
Exclusive
hllll
hhhhl
hhmll
Hhml
even
Random
a
512
221 252
169 193 164
217 160 138 207
414 524
193 158
226 123
246
160 142 134 178
175 206
150 126 139
186 145 174
459
321
b
207
15 19
6 18 7
10 37 15 16
140 188
14 9
8 123
26
7 19 11 7
5 17
5 3 2
10 4 7
159
78
c
906
352 178
142 129 73
211 87 67 126
585 492
102 88
140 58
191
82 63 65 105
135 212
85 69 95
163 112 192
1099
559
d
417
134 33
44 33 1
43 29 9 20
272 212
10 14
8 58
30
12 5 7 11
12 33
11 11 19
34 34 40
593
272
e
1068
520 311
208 270 140
239 142 133 163
468 675
151 149
186 80
344
90 82 88 165
121 257
106 100 111
215 197 229
1271
359
f
478
140 27
46 75 5
18 62 53 7
164 271
7 8
12 80
43
0 2 8 15
4 23
4 20 3
7 44 13
638
101
g
1161
301 448
262 239 358
473 216 211 413
1009 1038
362 225
411 197
264
345 317 206 230
434 289
378 222 235
467 248 270
516
993
h
550
4 36
21 9 27
29 19 14 22
450 508
26 5
16 197
6
19 120 9 5
24 9
36 25 4
36 12 4
173
439
i
8973
1958 2282
1828 1639 2009
2911 1609 1570 2500
16668 6800
2102 1693
2738 1558
1880
2231 1941 1656 1764
3219 2220
2337 1701 1895
2610 1736 2024
4691
10288
j
5600
77 94
76 18 56
337 51 12 176
12388 3620
110 25
289 1558
61
192 383 98 57
496 141
253 143 96
182 23 69
2186
6526
k
14241
2000 3992
1587 1376 1898
5414 1230 923 2767
22946 13222
3256 1077
5430 870
1866
2350 1722 998 1578
3613 2274
2347 1017 1230
2744 1107 1396
4963
7190
l
7640
230 467
210 93 123
1323 360 53 215
14466 6699
1174 26
1326 870
116
500 852 128 220
553 254
478 147 85
184 37 50
2084
2889

Supplementary Figure 11. Distributions of ISUs and reads per category of number of replicates and per mock community. Each series of violin plots represents the data collected for each mock community: *hllll* (a), *hhhhl* (b), *hhmll* (c), *Hhml* (d), *random* (e) and *even* (f). The series are split per color according to the number of replicates category (bottom colored box legends). For each category are represented both the density distribution of the number of ISUs per clone (left violin) and of the number of reads per ISU (right violin). The $\log_{10}$-transformed y-axis for the number of ISUs and for the number of reads are situated on the left (plain line) and right (dotted line) sides of each plot, respectively, on the Each violin separates the expected (left side) from the mistagging (right side) data. The median of each distribution is indicated (horizontal bars).


a
lhhhh
b
lllhl
c
hhlml
d
hmHl
e
random
f
even
log10(number of ISUs per clone)
log10(number of reads per ISU)
0.0
1.0
2.0
3.0
0
1
2
3
4
5
1
2
3
4
5

Supplementary Figure 12. Abundances comparisons between mock community sequence templates and the resulting ISUs. Abundances are displayed for each clone, but separately for each of the 4 mock communities of SFA-125, including (a) *hhhhl* containing 4 clones at high abundance and 1 clone at low abundance, (b) *hhmll* containing 2 clones at high abundance, 1 clone at medium abundance and 2 clones at low abundance, (c) *hllll* containing 1 clone at high abundance and 4 clones at low abundance and (d) *Hhml* containing 1 clone at very high abundance, 1 clone at high abundance, 1 clone at medium abundance and 1 clone at low abundance (see Supplementary Table 2). The clones (blue dots) and ISUs (red crosses) are organized in five columns according to the number of replicates intersection where it is found simultaneously. In each case, the exact value of the template abundance is located at the “3 replicates” position.


Number of replicates
1 2 3 4 5
a
b
c
d
Relative reads abundance
Relative reads abundance
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
–0.05
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.001
0.249
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
–0.2
0
0.2
0.4
0.6
0.8
1
1.2
0.002
0.992
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
–0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.001
0.047
0.475
Relative template abundance
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
–0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
0.002
0.032
0.161
0.805
Relative template abundance
● Abundances sum
+ Most abundant ISU