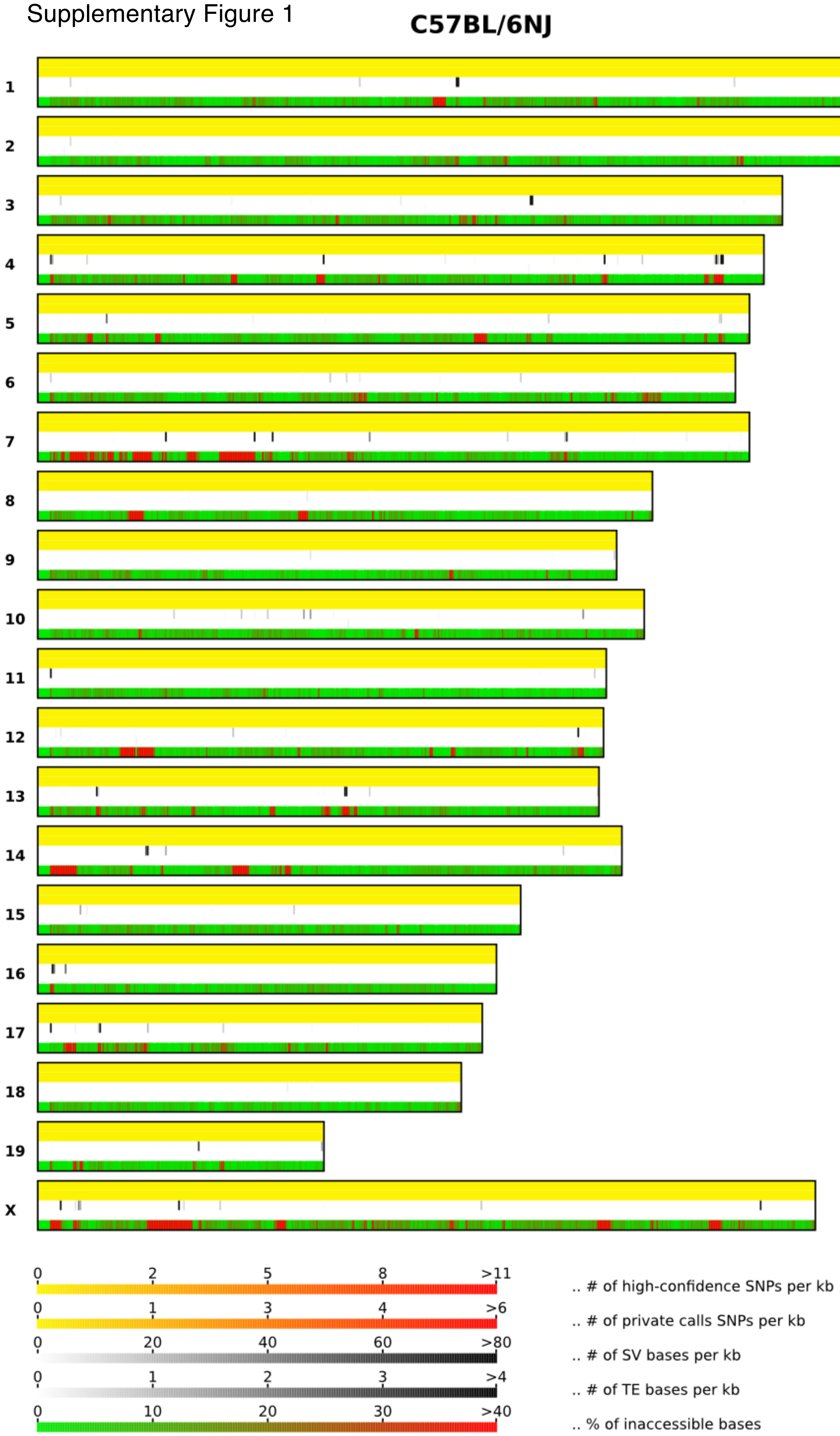
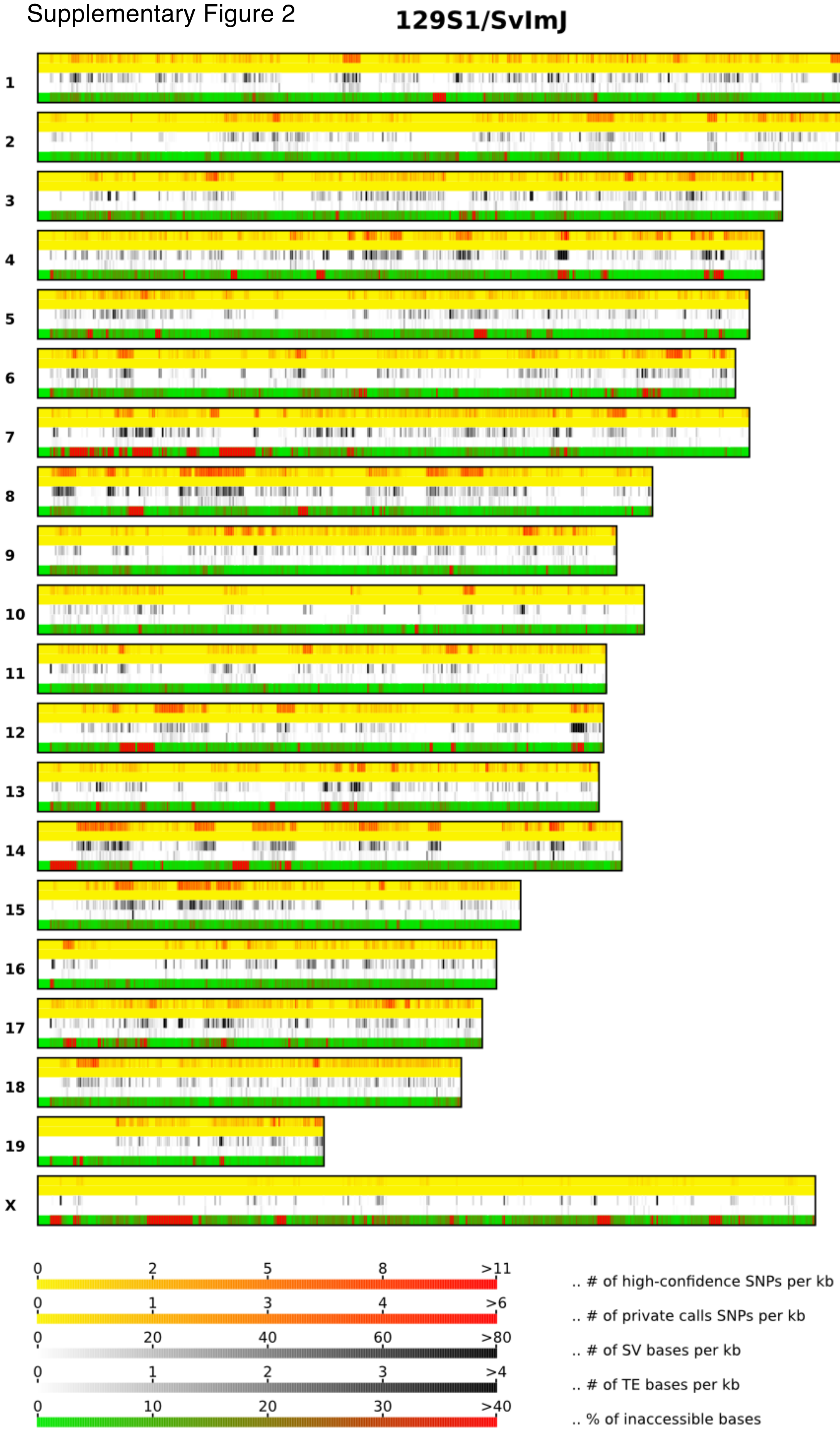
Supplementary Figure 1



Supplementary Figure 1: An overview of the variation identified in C57BL6NJ.

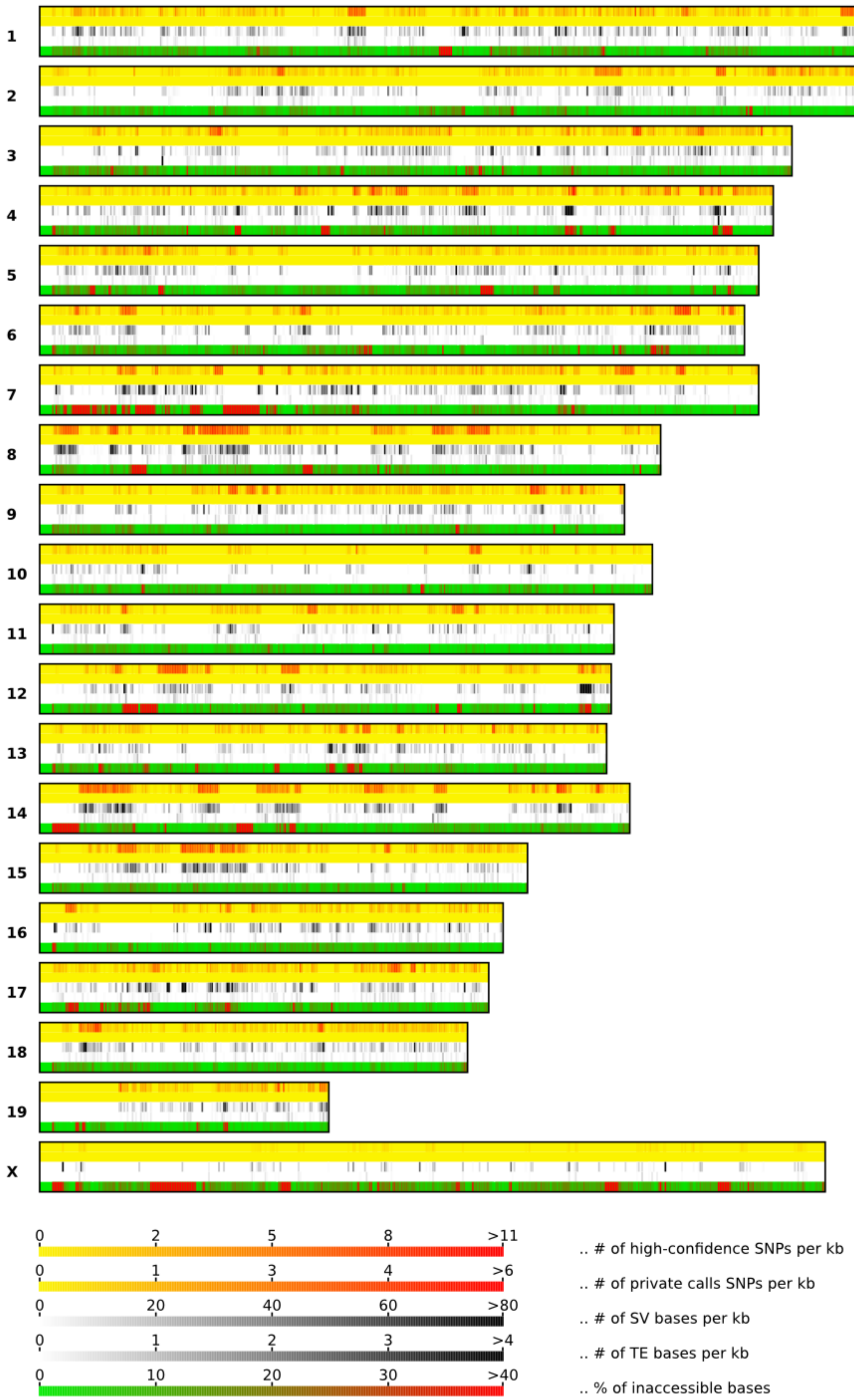
Supplementary Figure 2



Supplementary Figure 2: An overview of the variation identified in 129S1/SvImj.

Supplementary Figure 3

129S5/SvEvBrd

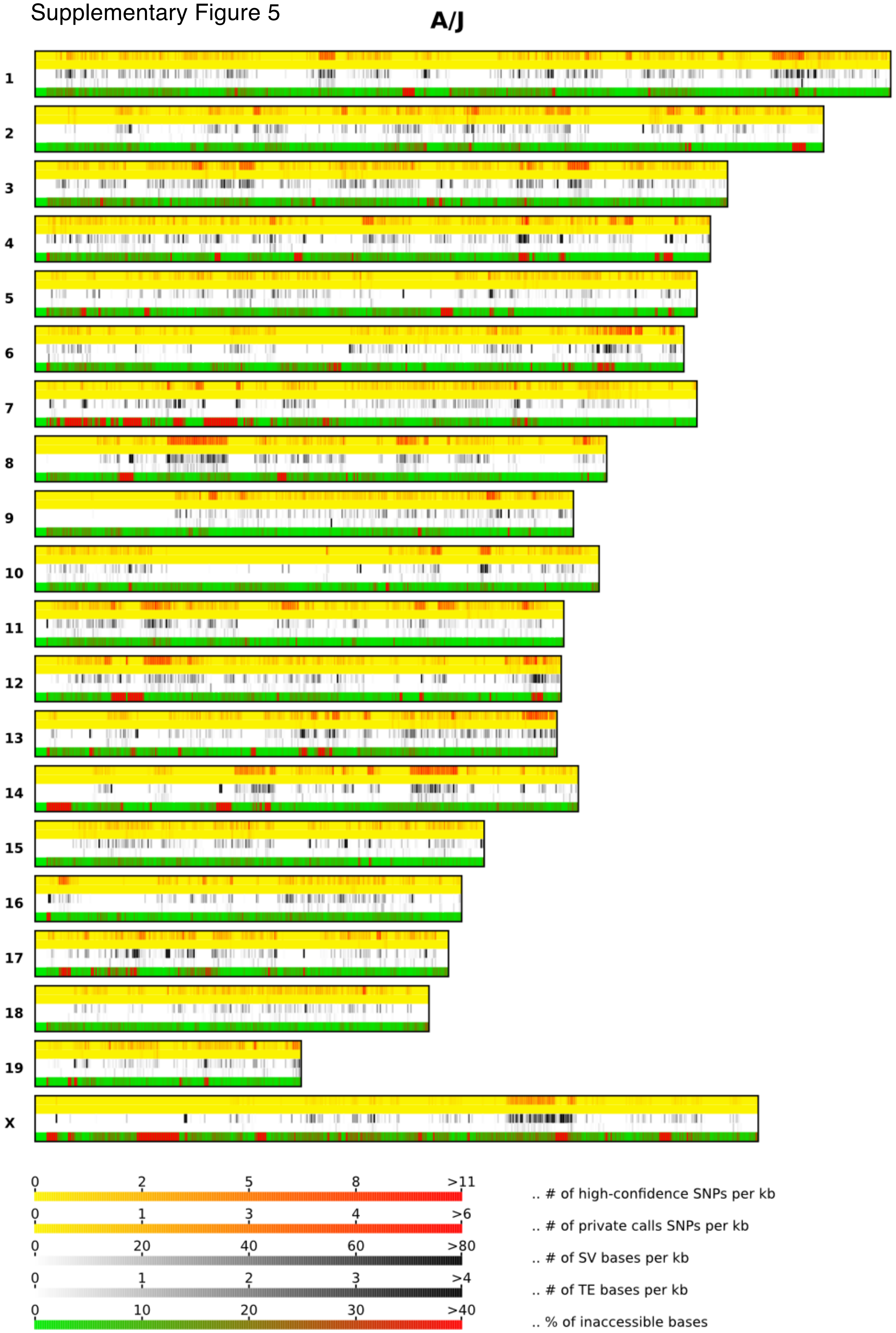


Supplementary Figure 3: An overview of the variation identified in 129S5/SvEvBrd.

Supplementary Figure 4

Supplementary Figure 4: An overview of the variation identified in 129P2/OlaHsd.

Supplementary Figure 5



Supplementary Figure 5: An overview of the variation identified in A/J.

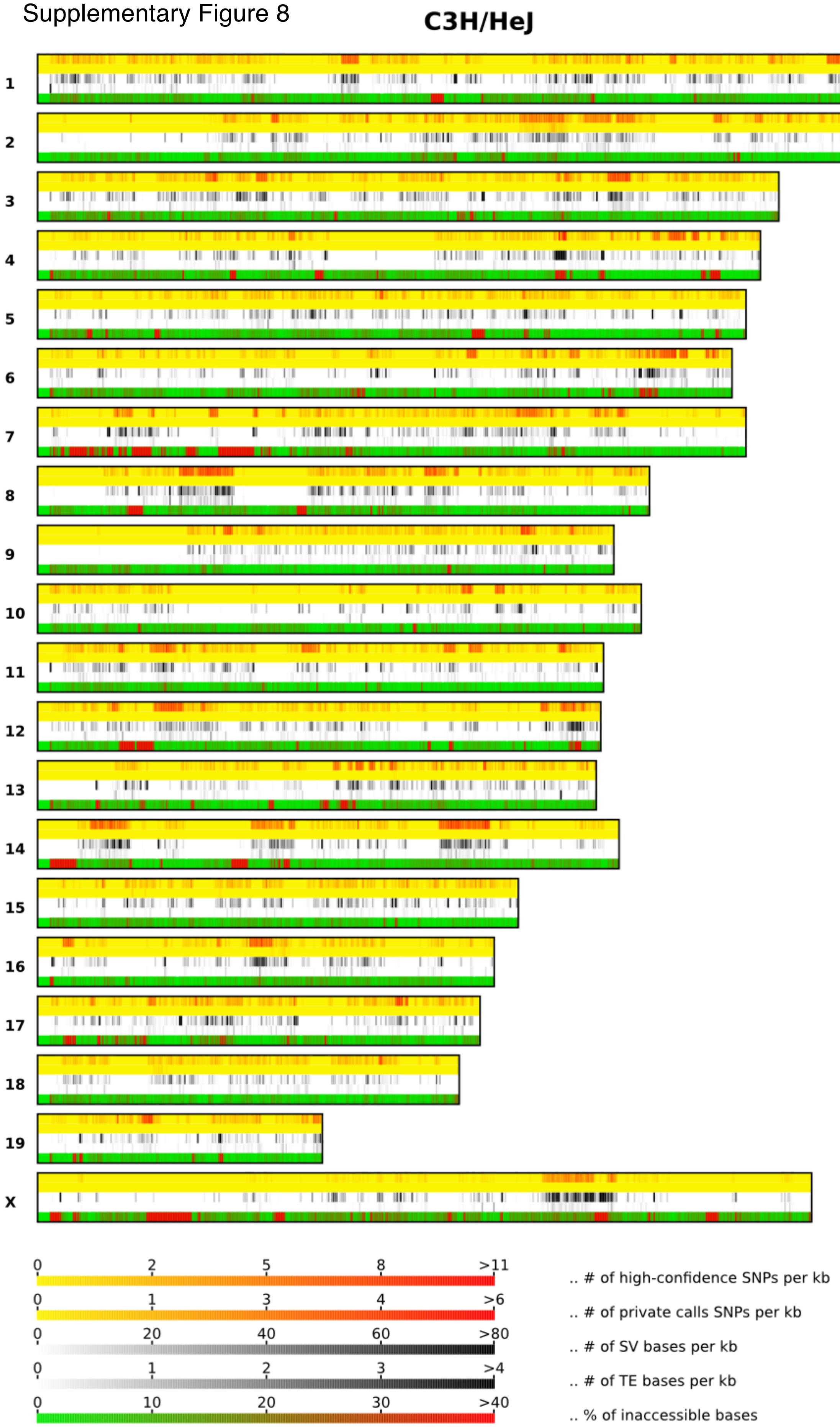
Supplementary Figure 6

Supplementary Figure 6: An overview of the variation identified in AKR/J.

Supplementary Figure 7

Supplementary Figure 7: An overview of the variation identified in BALB/cJ.

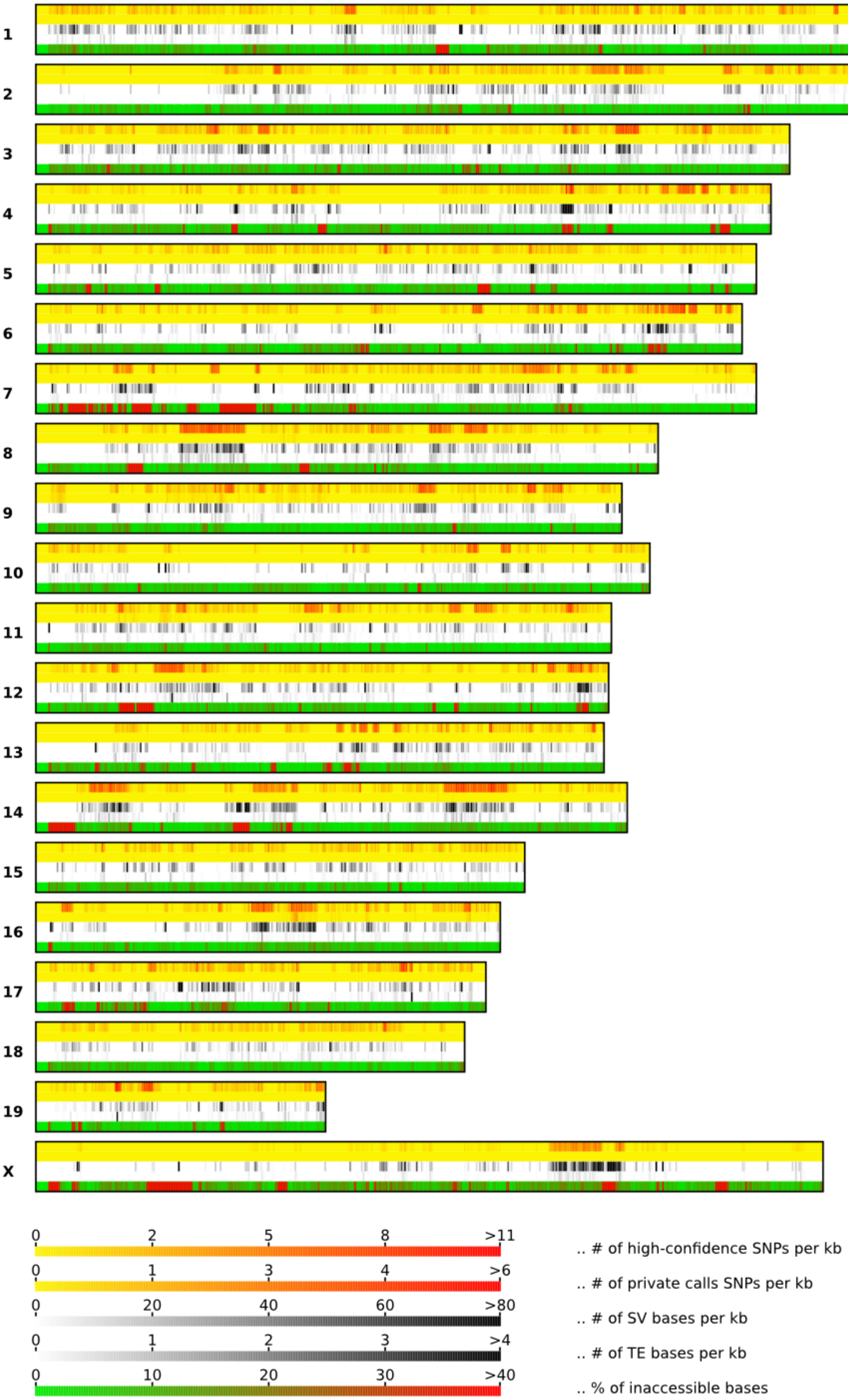
Supplementary Figure 8



Supplementary Figure 8: An overview of the variation identified in C3H/HeJ.

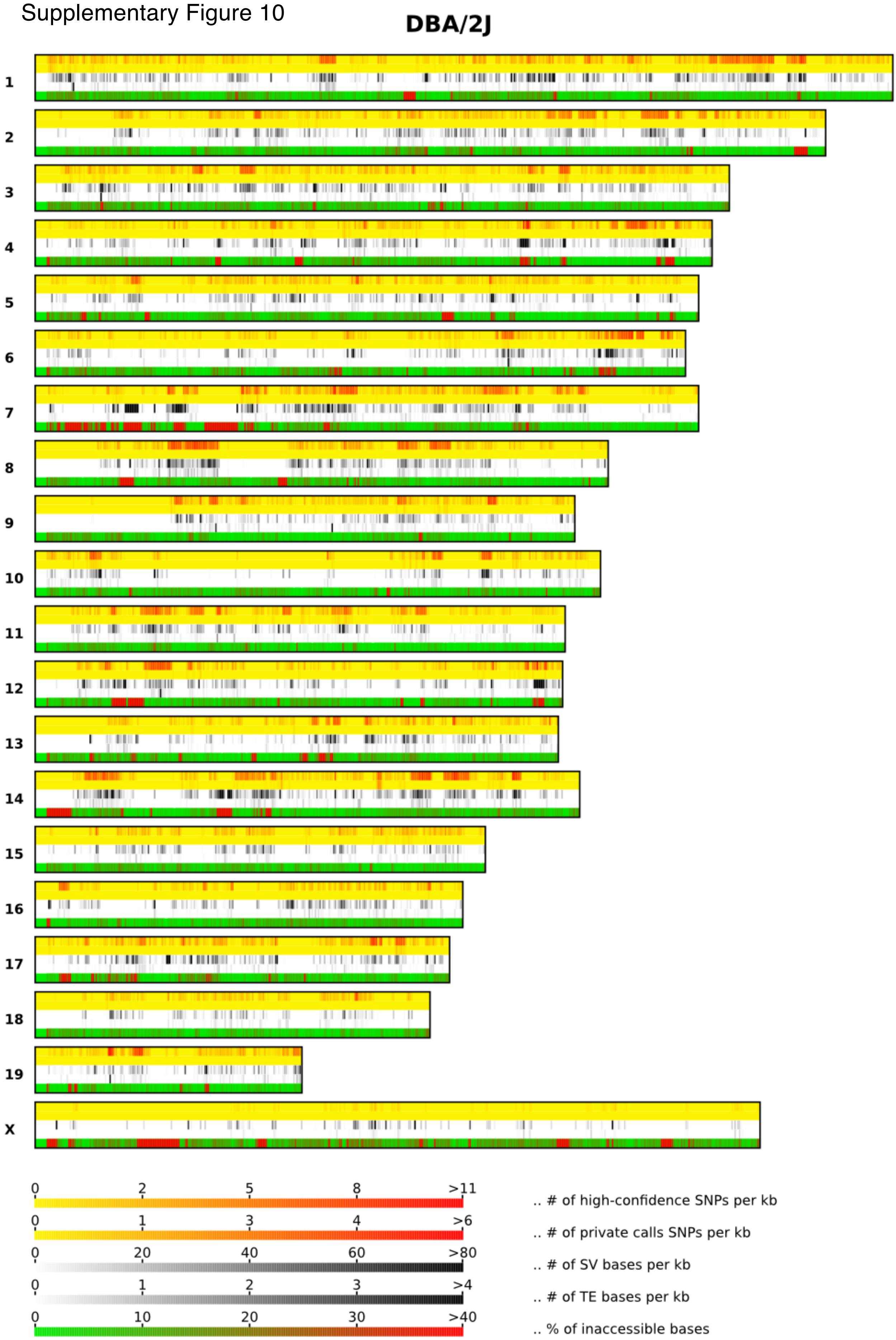
Supplementary Figure 9

CBA/J



Supplementary Figure 9: An overview of the variation identified in CBA/J.

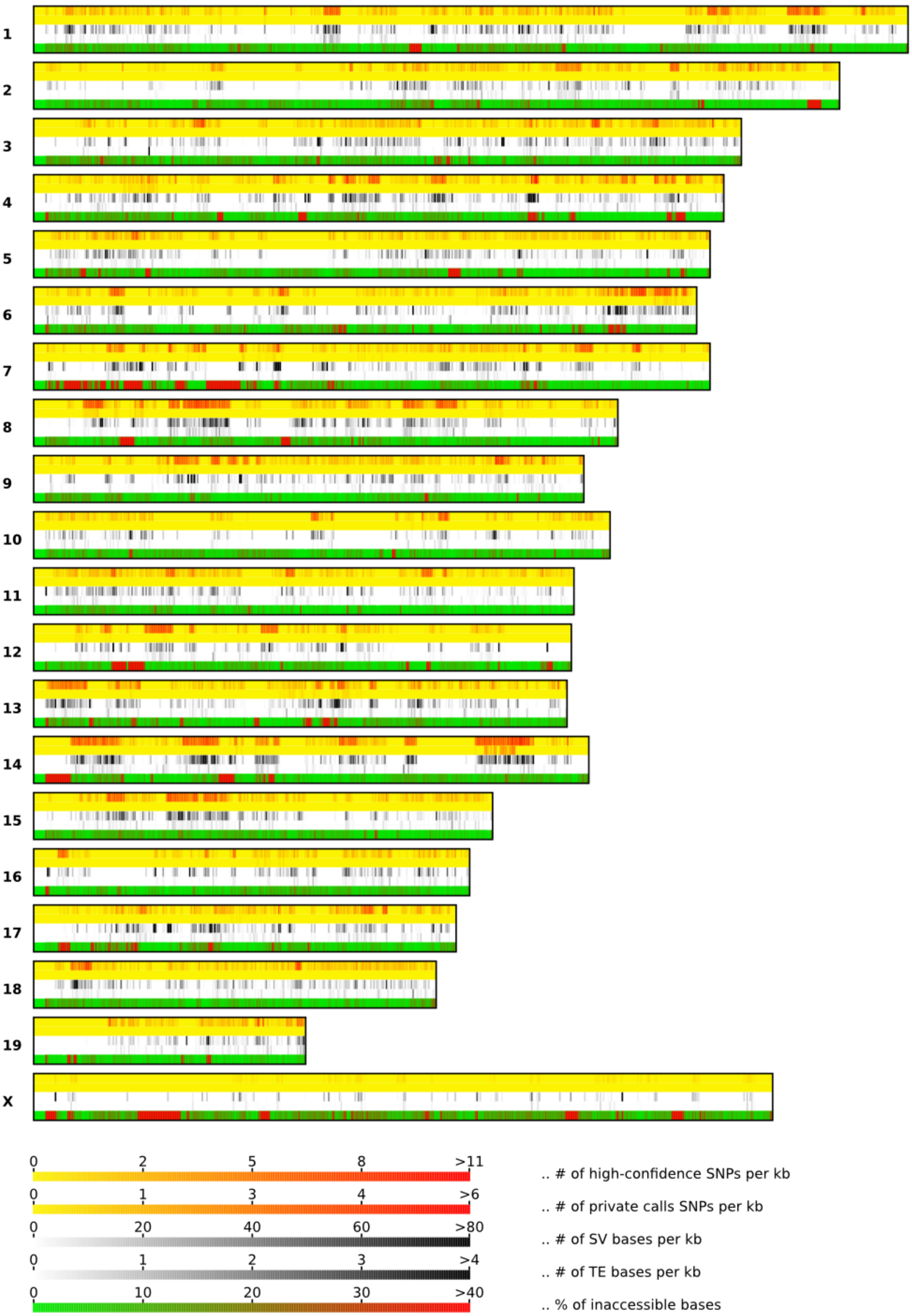
Supplementary Figure 10



Supplementary Figure 10: An overview of the variation identified in DBA/2J.

Supplementary Figure 11

LP/J

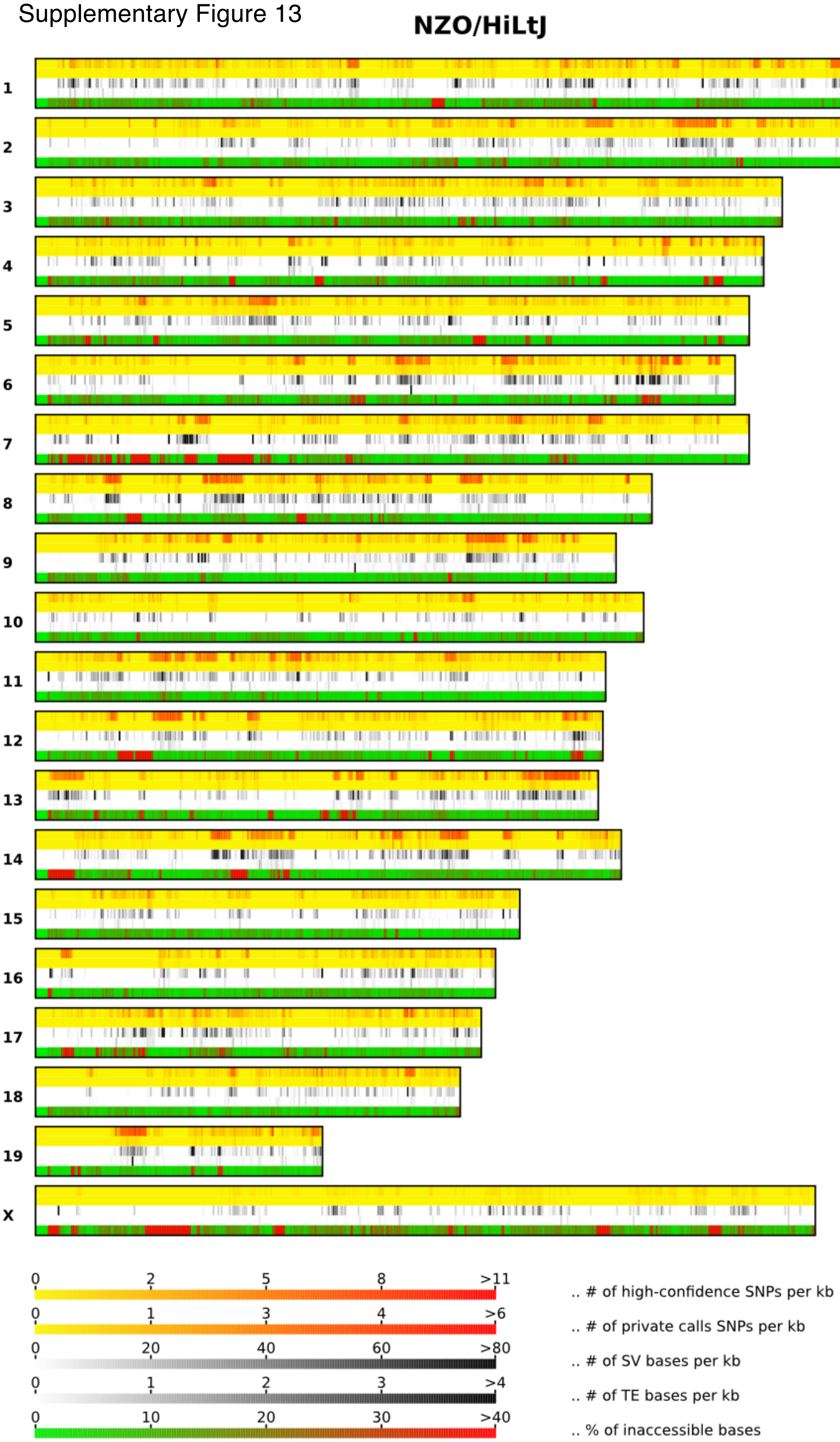


Supplementary Figure 11: An overview of the variation identified in LP/J.

Supplementary Figure 12

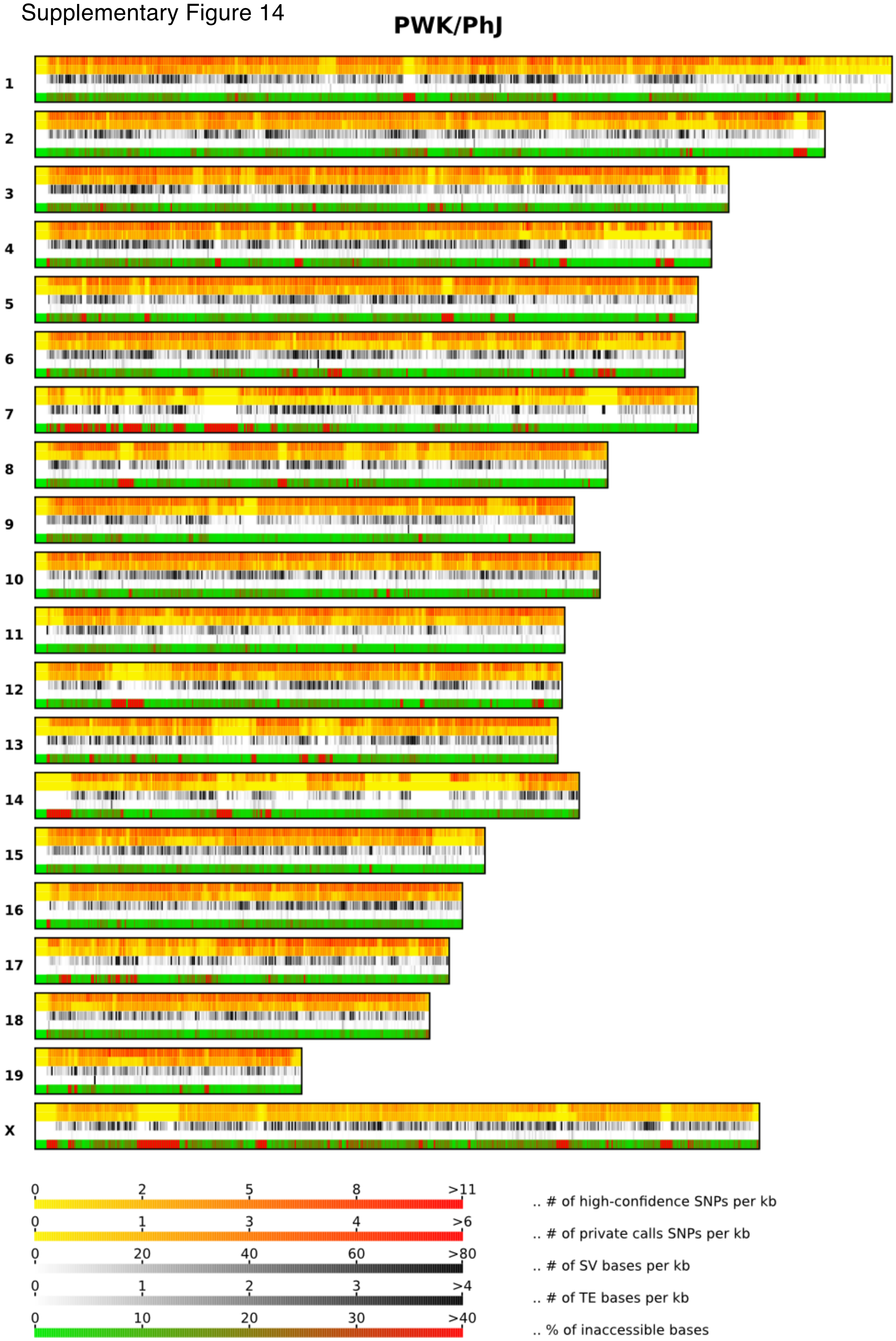
Supplementary Figure 12: An overview of the variation identified in NOD/ShiLtJ.

Supplementary Figure 13



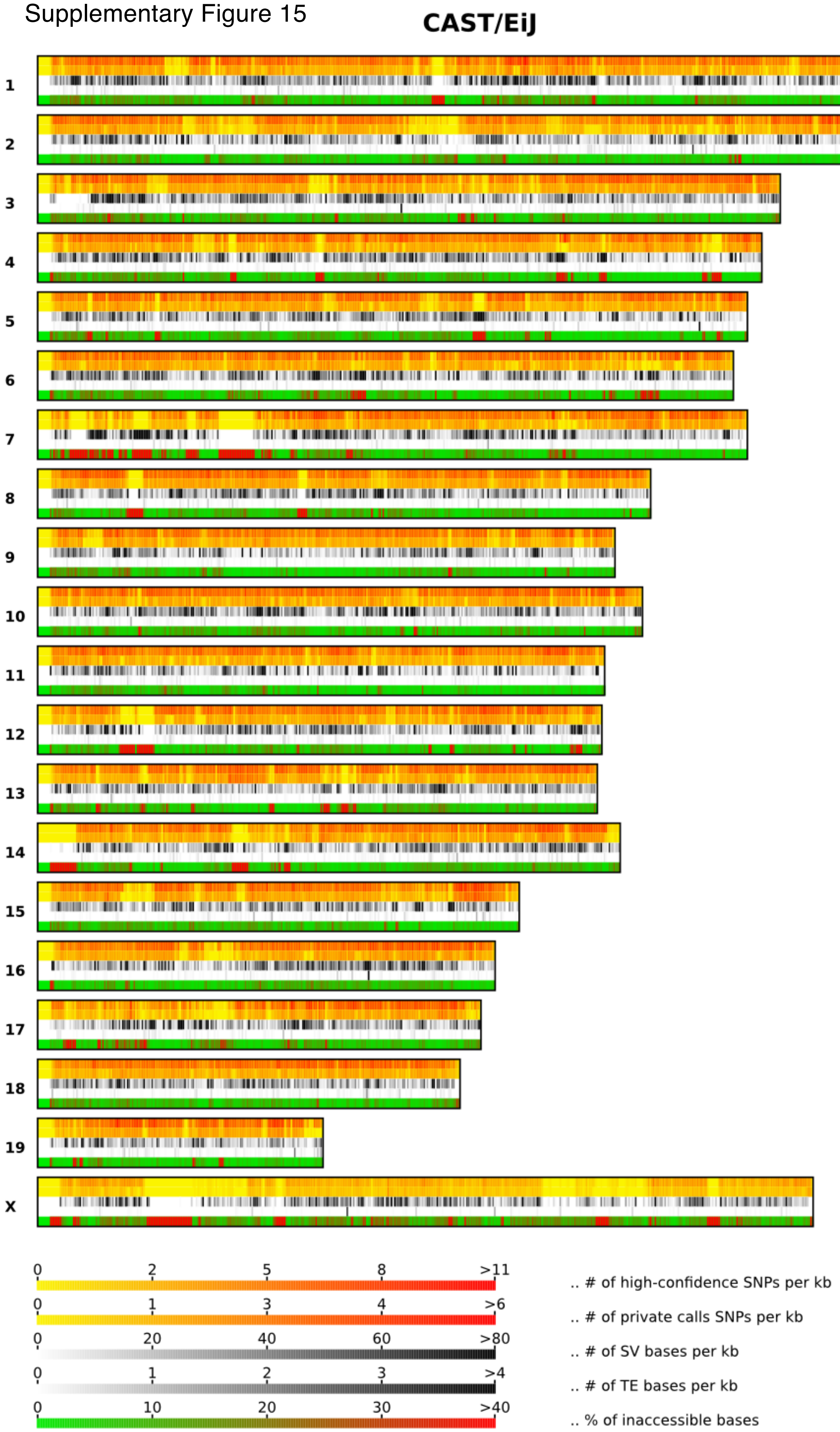
Supplementary Figure 13: An overview of the variation identified in NZO/HiLtJ.

Supplementary Figure 14



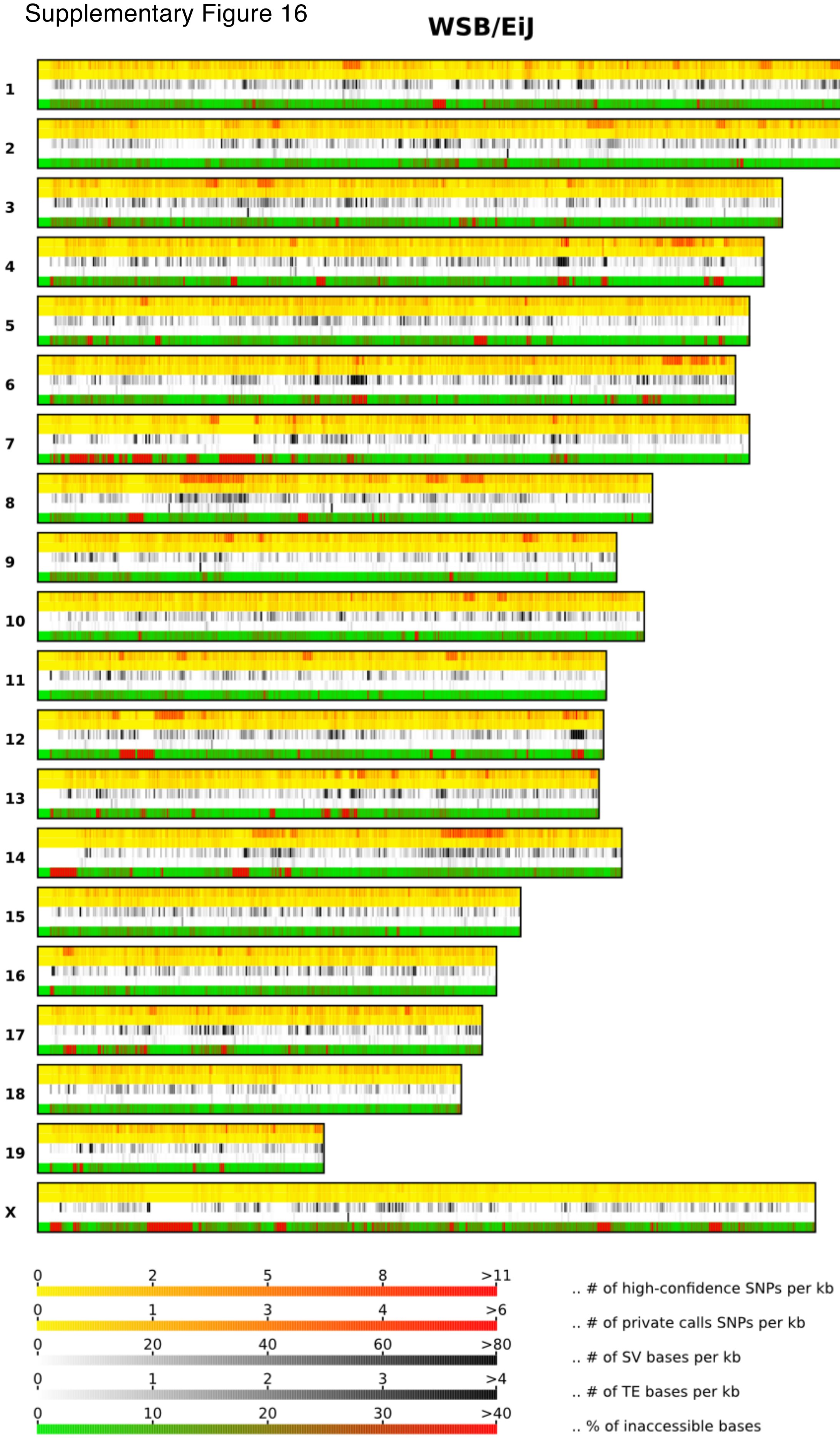
Supplementary Figure 14: An overview of the variation identified in PWK/PhJ

Supplementary Figure 15



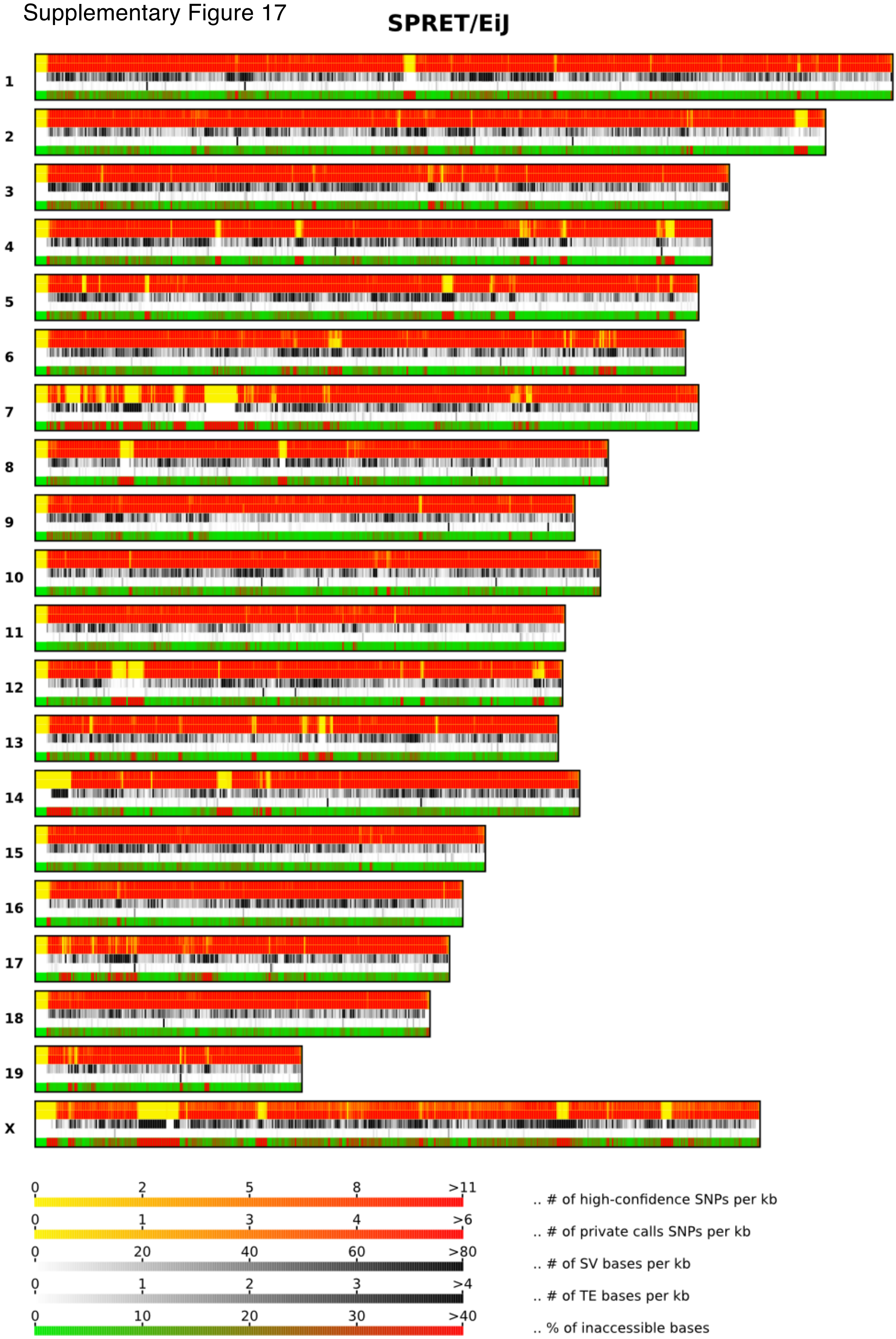
Supplementary Figure 15: An overview of the variation identified in CAST/EiJ.

Supplementary Figure 16



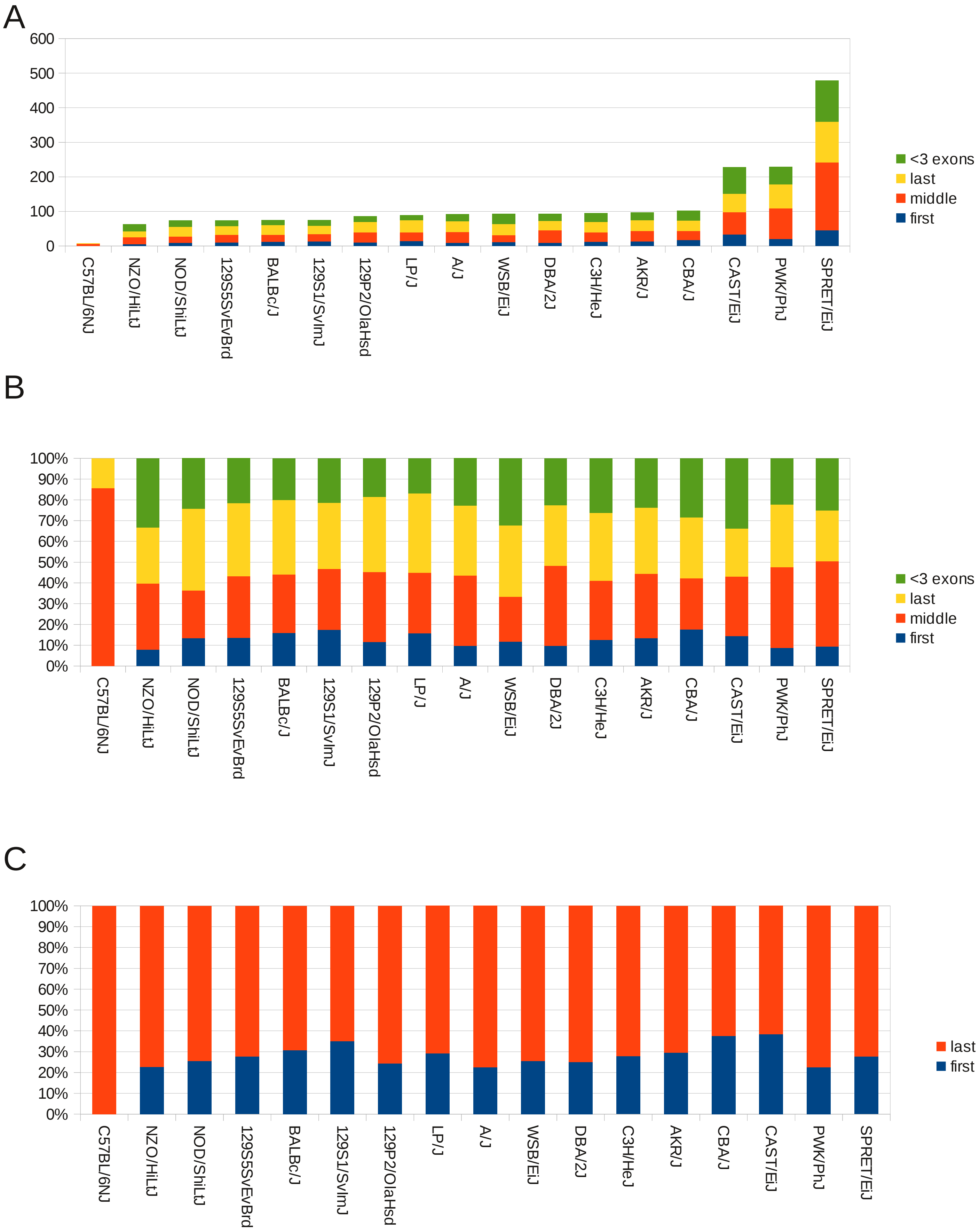
Supplementary Figure 16: An overview of the variation identified in WSB/EiJ.

Supplementary Figure 17



Supplementary Figure 17: An overview of the variation identified in SPRET/EiJ.

Supplementary Figure 18

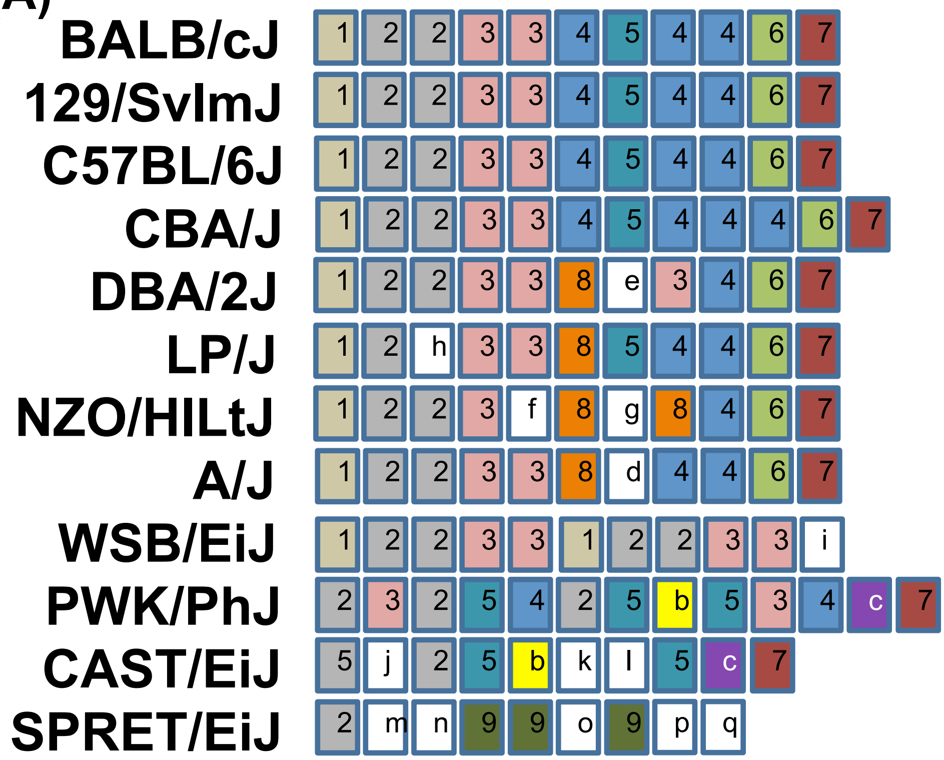


Supplementary Figure 18: Stop codons per strain. The figure shows the location of stop codons in transcripts. A: Shown are the number of stop codons in first, last, and middle exons of transcripts that contain a stop-codon with respect to the reference sequence. Transcripts with less than three exons are summarized as "<3 exons". B and C show normalized data .

Supplementary Figure 19

Supplemental Figure (Prdm9)

(A)



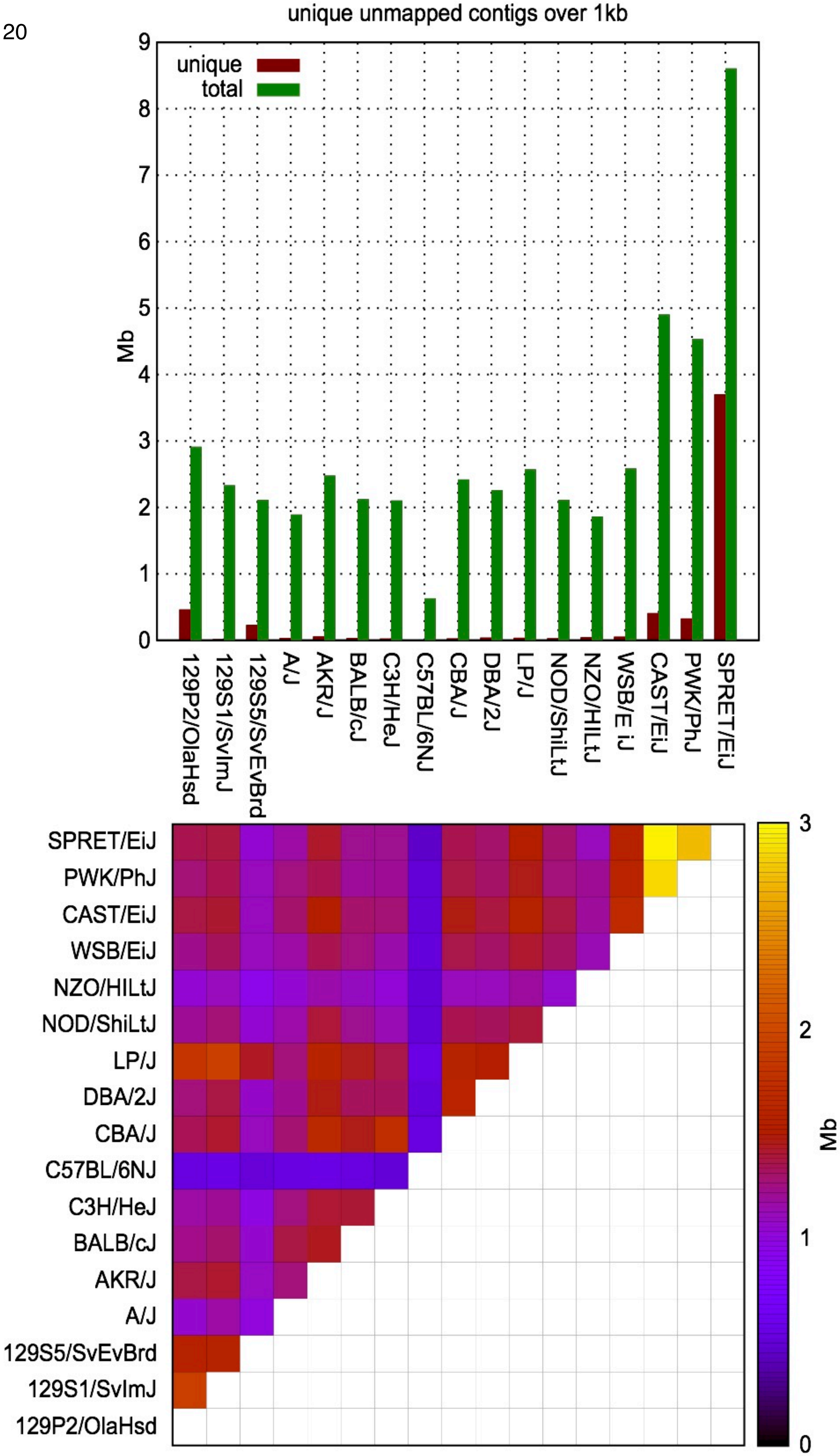
BALB/cJ 1 2 2 3 3 4 5 4 4 6 7
129/SvImJ 1 2 2 3 3 4 5 4 4 6 7
C57BL/6J 1 2 2 3 3 4 5 4 4 6 7
CBA/J 1 2 2 3 3 4 5 4 4 4 6 7
DBA/2J 1 2 2 3 3 8 e 3 4 6 7
LP/J 1 2 h 3 3 8 5 4 4 6 7
NZO/HILtJ 1 2 2 3 f 8 g 8 4 6 7
A/J 1 2 2 3 3 8 d 4 4 6 7
WSB/EiJ 1 2 2 3 3 1 2 2 3 3 i
PWK/PhJ 2 3 2 5 4 2 5 b 5 3 4 c 7
CAST/EiJ 5 j 2 5 b k l 5 c 7
SPRET/EiJ 2 m n 9 9 o 9 p q

(B)

```
TGCAGGGAGTGTGGGCGGGGCTTTACACAGAACTCACACCTCATCCAGCACCAGAGGACACACACAGGGGAGAAGCCCTATGTT 1
TGCAGGGAGTGTGGGCGGGGCTTTACACAGAAGTCAGACCTCATCAAGCACCAGAGGACACACACAGGGGAGAAGCCCTATGTT 2
TGCAGGGAGTGTGGGCGGGGCTTTACACAGAAGTCAGTCCTCATCAAGCACCAGAGGACACACACAGGGGAGAAGCCCTATGTT 3
TGCAGGGAGTGTGGGCGGGGCTTTACAGCGAAGTCAGTCCTCATCCAGCACCAGAGGACACACACAGGGGAGAAGCCCTATGTT 4
TGCAGGGAGTGTGGGCGGGGCTTTACAGCGAAGTCAAACCTCATCCAGCACCAGAGGACACACACAGGGGAGAAGCCCTATGTT 5
TGCAGGGAGTGTGGGCGGGGCTTTACACAGAAGTCAAACCTCATCAAGCACCAGAGGACACACACAGGGGAGAAGCCCTATGTT 6
TGCAGGGAGTGTGGGTGGGGCTTTACACAGAAGTCAGACCTCATCCAGCACCAGAGGACACATACAAGAGAGAAGTAA------ 7
TGCAGGGAGTGTGGGCGGGGCTTTACAGAGAAGTCAGTCCTCATCCAGCACCAGAGGACACACACAGGGGAGAAGCCCTATGTT 8
TGCAGGGAGTGTGGGCGGGGCTTTACAGCGAAGTCAGACCTCATCAAGCACCAGAGGACACACACAGGGGAGAAGCCCTATGTT 9
TGCAGGGAGTGTGGGCGGGGCTTTACAGAGAAGTCAAGCCTCATCAAGCACCAGAGGACACACACAGGGGAGAAGCCCTATGTT b
TGCAGGGAGTGTGGGTGGGGCTTTACACAGAAGTCAAACCTCATCAAGCACCAGAGGACACACACAGGGGAGAAGCCCTATGTT c
TGCAGGGAGTGTGGGCGGGGCTTTACAGAGAAGTCAATCCTCATCCAGCACCAGAGGACACACACAGGGGAGAAGCCCTATGTT d
TGCAGGGAGTGTGGGCGGGGCTTTACACAGAAGTCAATCCTCATCAAGCACCAGAGGACACACACAGGGGAGAAGCCCTATGTT e
TGCAGGGAGTGTGGGCGGGGCTTTACACAGAAGTCATTCCTCATCAAGCACCAGAGGACACACACAGGGGAGAAGCCCTATGTT f
TGCAGGGAGTGTGGGCGGGGCTTTACAGAGAAGTCAAACCTCATCCAGCACCAGAGGACACACACAGGGGAGAAGCCCTATGTT g
TGCAGGGAGTGTGGGCGGGGCTTTACACGGAAGTCAGACCTCATCAAGCACCAGAGGACACACACAGGGGAGAAGCCCTATGTT h
TGCAGGGAGTGTGGGCGGGGCTTTACACAGAAGTCAGACCTCATCCAGCACCAGAGGACACATACAAGAGAGAAGTAA------ i
TGCAGGGAGTGTGGGCGGGGCTTTACACAGAAGTCAGTCCTCATCCAGCACCAGAGGACACACACAGGGGAGAAGCCCTATGTT j
TGCAGGGAGTGTGGGTGGGGCTTTACAGCGAAGTCAAACCTCATCCAGCACCAGAGGACACACACAGGGGAGAAGCCCTATGTT k
TGCAGGGAGTGTGGGCGGGGCTTTACACAGAAGTCAAGCCTCATCAAGCACCAGAGGACACACACAGGGGAGAAGCCCTATGTT l
TGCAGGGAGTGTGGGCGGGGCTTTACACAGAAGTCAGACCTCATCAAGCACCAGAGGACACACACAGGGGAGAAGCCTTATGTT m
TGCAGGGAGTGTGGGCGGGGCTTTACACAGAAGTCAAACCTCATCCAGCACCAGAGGACACACACAGGGGAGAAGCCCTATGTT n
TGCAGGGAGTGTGGGCGGGGCTTTACAACGAAGTCAAGCCTCATCCAGCACCAGAGGACACACACAGGGGAGAAGCCCTATGTT o
TGCAGGGAGTGTGGGCGGGGCTTTACAGTGCAGTCAAACCTCATCCAGCACCAGAGGACACACACAGGGGAGAAGCCCTATGTT p
TGCAGGGAGTGTGGGTGGGGTTTTAAACAGAAGTCAGACCTCATCCAGCACCAAAGGACACACACAGGGGAGAAGTAA------ q
*************** *********  *  * * ***  ******* ******* ******** *** * ******
```

Supplementary Figure 19: The zinc finger sequences of the *Prdm9* 'speciation gene' for 12 mouse strains. (A) Each zinc finger is represented as a box and nucleotide sequence-identical zinc fingers are labeled with the same number or letter. These sequences are identical for three laboratory strains (BALB/cJ, 129S1/SvImJ, and C57BL/6NJ), but otherwise there are substantial differences in *Prdm9* zinc finger number and sequence, especially for the wild-derived strains. Unlike the other strains, the zinc finger repertoire of Prdm9 in WSB/EiJ consists of a 5 zinc-finger tandem repeat. (B) The nucleotide sequences of the different *Prdm9* zinc finger sequences with 100% identical positions are marked with an asterisk.

Supplementary Figure 20



Supplementary Figure 20: The identification of non-reference sequence. Unmapped reads (which did not map the mouse reference sequence) for each strain were assembled using ABYSS and contigs over 1 Kb long were selected. The histogram shows the total length of the unmapped contigs for each strain generated by this method (shown in green on the histogram). The resulting contigs were aligned (using Exonerate) against the contigs from other strains to estimate the amount of unmapped sequence that was unique to each strain (shown in red on the histogram). The amount of shared sequence between strains is shown as a heatmap. The wild-derived strains, CAST/EiJ, PWK/PhJ and SPRET/EiJ share more unmapped sequence than the other strains, while C57BL/6NJ contains very little unmapped sequence due to its close similarity to the reference strain C57BL/6J.