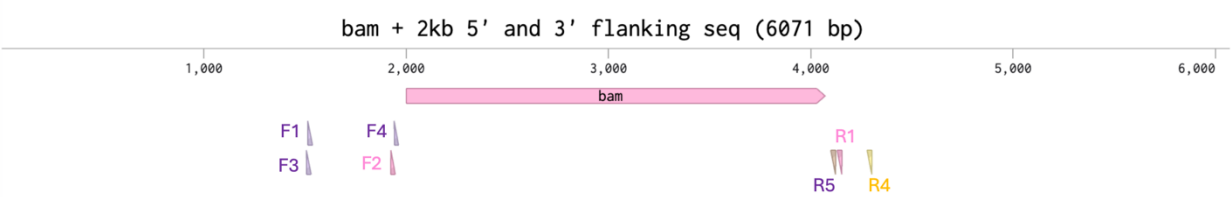
bam86 Mutant Sequencing Results Analysis

bam86

PCR products sequenced: F1+R1, F1+R5

- 4 samples total: bam86_F1R1_F1, bam86_F1R1_R1, bam86_F1R5_F1, bam86_F1R5_R5

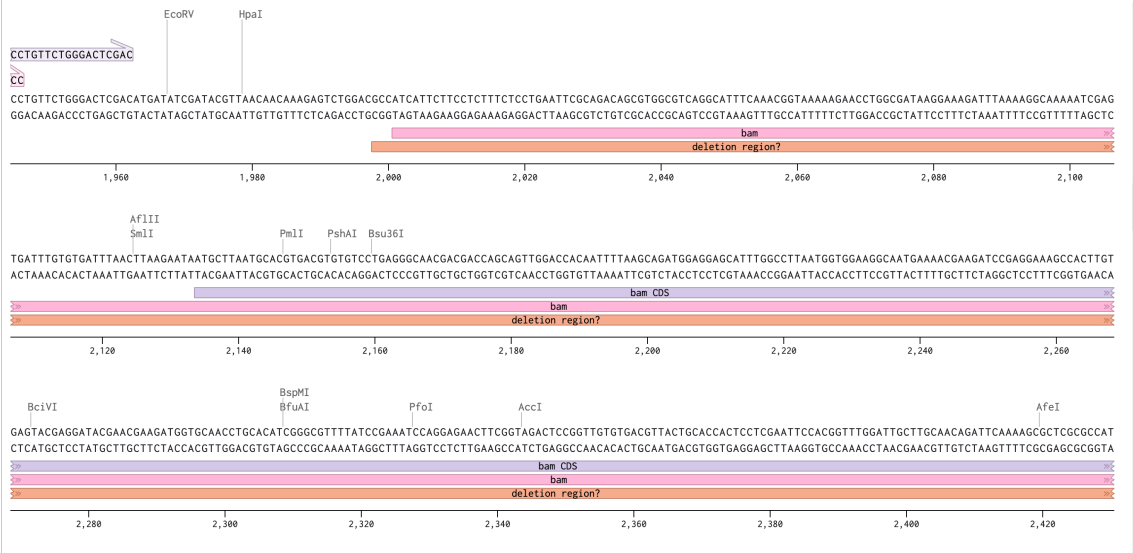


Benchling alignment:

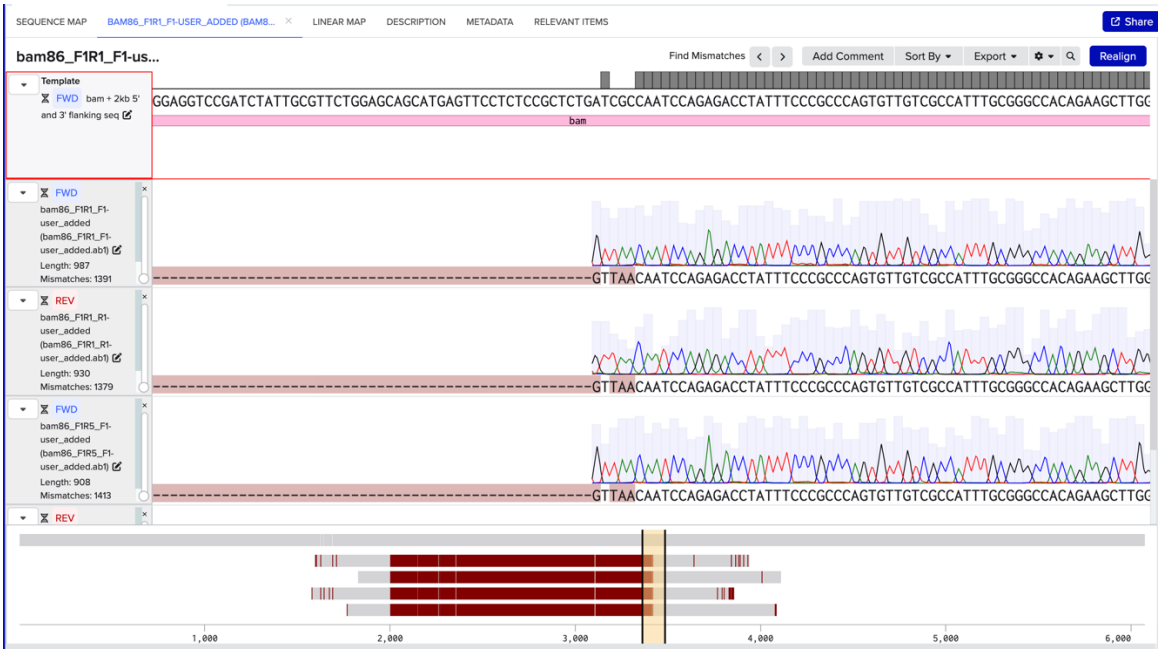
Observations:

- A ~1.4kb deletion is present in bam86 compared to reference
- Mismatches are present near the ends of each amplicon
  → acceptable?

- Start site of deletion is clear: 3bp upstream of bam gene region, or 136bp upstream of bam CDS.



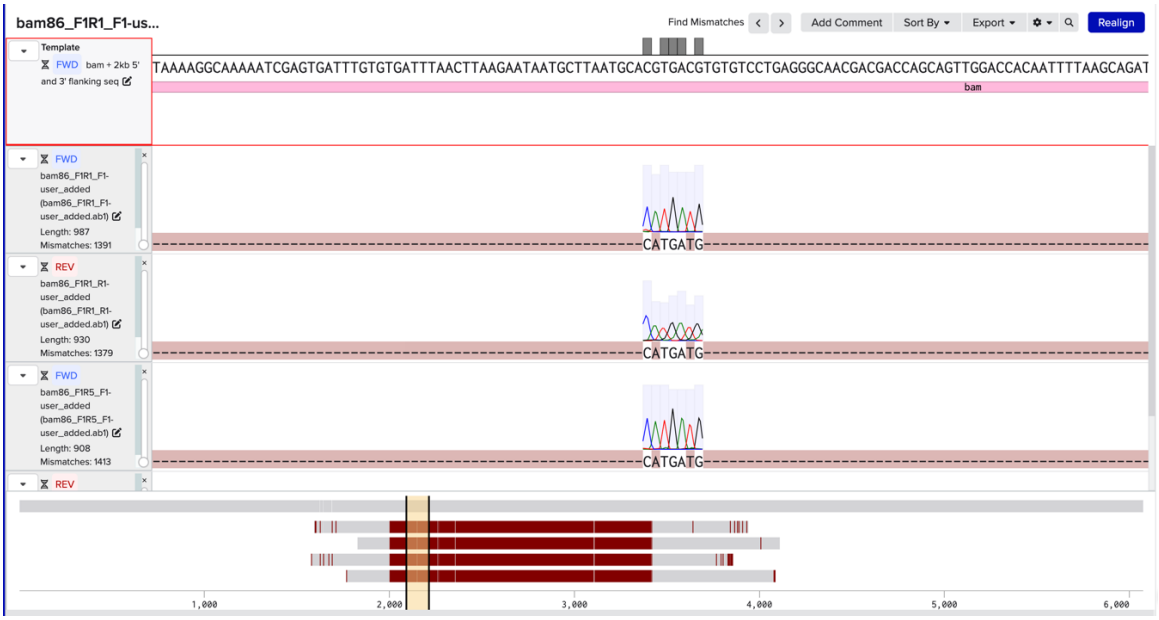
- Endpoint of deletion is unclear:



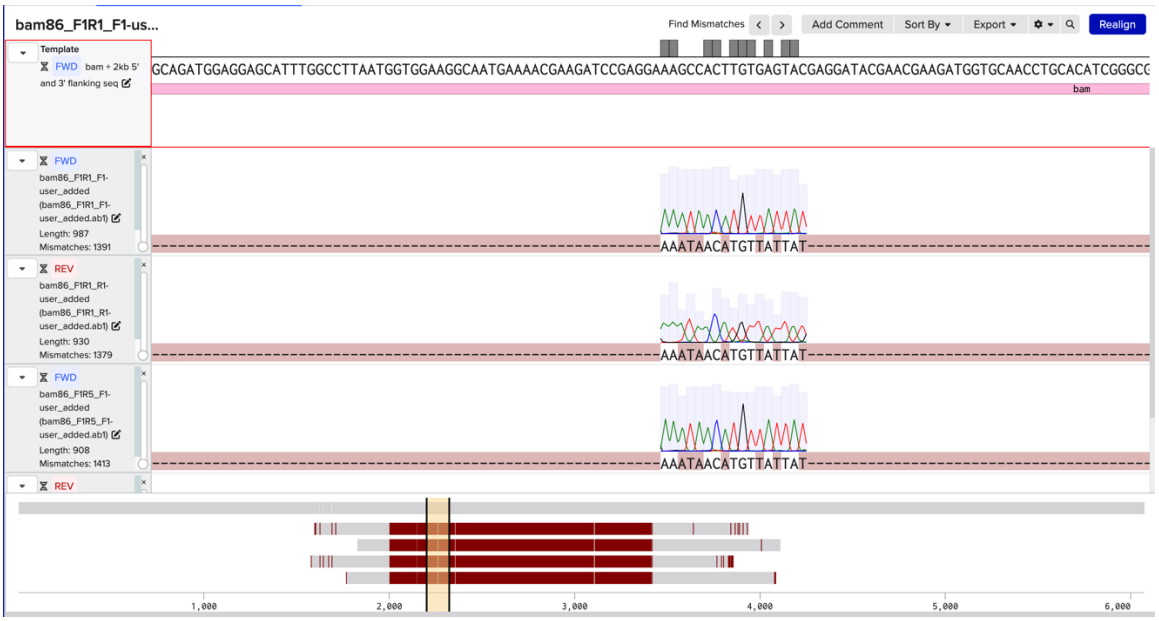
  - Several mismatches between all sequencing products and reference (GTTAA)
  - Good thing is that all sequencing products agree with each other, so we still have a confident sequence, we just don't know why there are mismatches
  - If we regard this GTTAA as part of the deletion region, then the end point of the deletion is 1282bp into the CDS (counting from the start codon), making the deletion 1418bp in total.
- Aside from the end, there are 4 mismatch regions within the deletion region, manifesting as groups of bases mapping within the deletion region, where we expect

nothing to map to. However, the alignment of these bases to the reference is very imperfect, so I don't believe they are part of the bam genomic sequence.

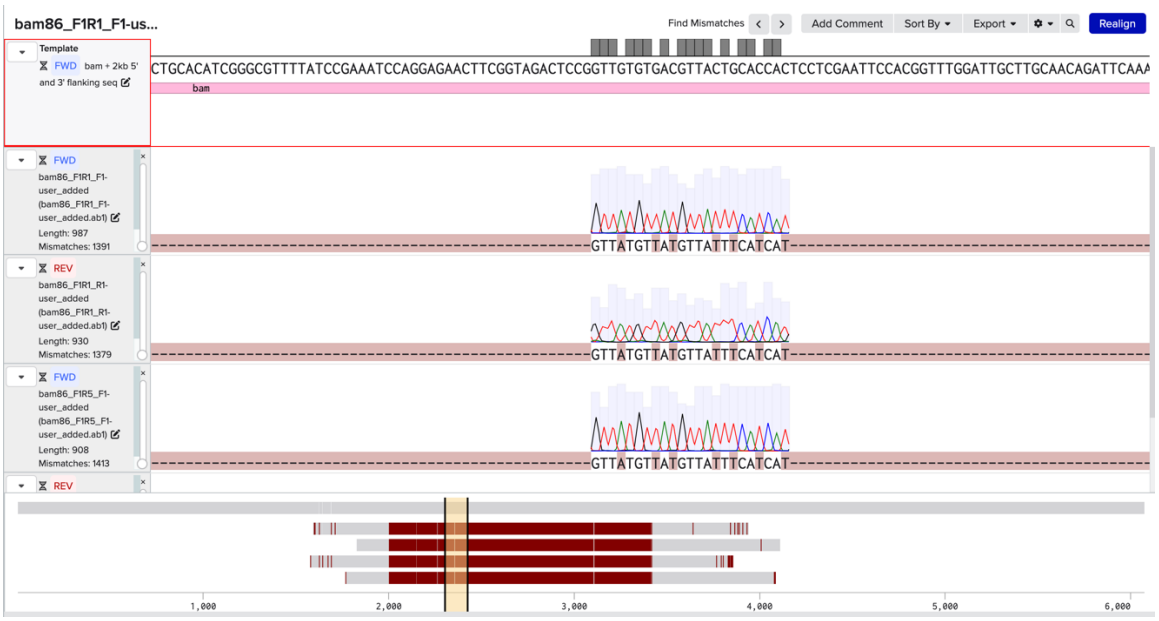
Mismatch region 1:



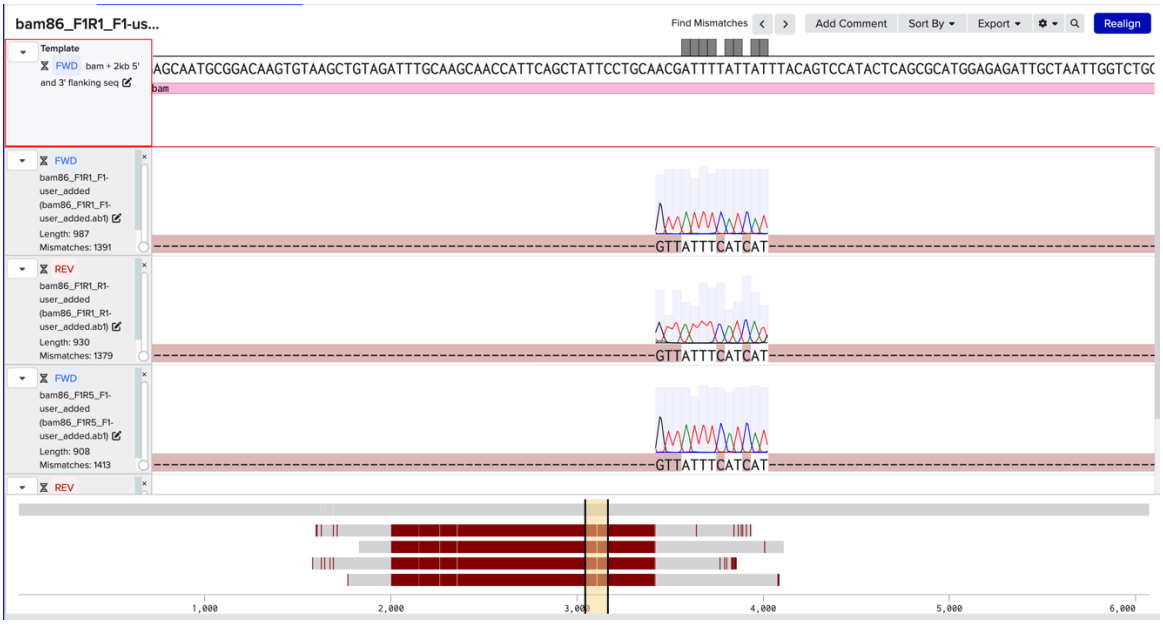
Mismatch region 2:



Mismatch region 3:



Mismatch region 4:



Possible origins of these clusters of bases:

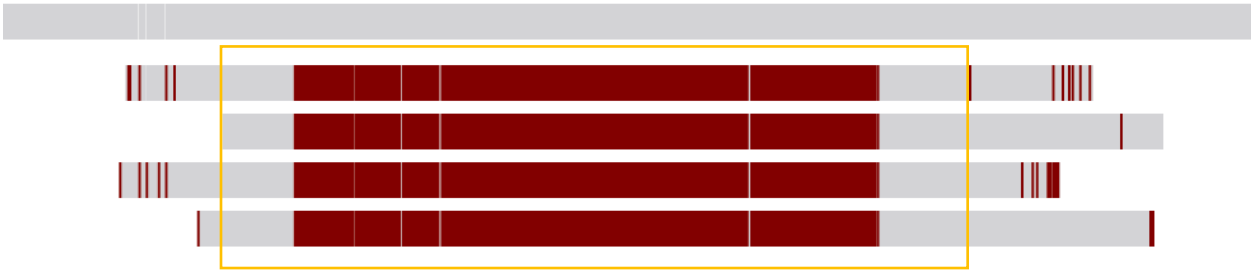
- P-element excision is not clean → remnant of P-element or adjacent genomic seq?
- Random bases introduced by the DNA excision repair process?

→ Searched for all 4 groups of mismatched bases in bam gene region + genomic flanking region + P{hsneo} sequence

- Mismatch region 1 (CATGATG): not found in bam + 2kb 5' and 3' flanking seq; found 3 times in P{hsneo} – once at both ends of the P-element, once in the middle → remnant of P-element?
- Mismatch region 2 (AAATAACATGTTATTAT): not found in either bam + 2kb 5' and 3' flanking seq or P{hsneo} → origin unknown
- Mismatch region 3 (GTTATGTTATGTTATTTCATCAT): not found in either bam + 2kb 5' and 3' flanking seq or P{hsneo} → origin unknown
- Mismatch region 4 (GTTATTTCATCAT): not found in bam + 2kb 5' and 3' flanking seq; found twice in P{hsneo}, once at either end of the P-element → remnant of P-element?

  The ends of the P-element actually have GTTATTTCATCATG – and the first mismatched base at the end of the deletion region is G. Since mismatch region 4 is connected to the end of the deletion region, I speculate that the G at the end of the deletion region is also part of this remnant P-element sequence.

- Sequence supported by all sequencing products:



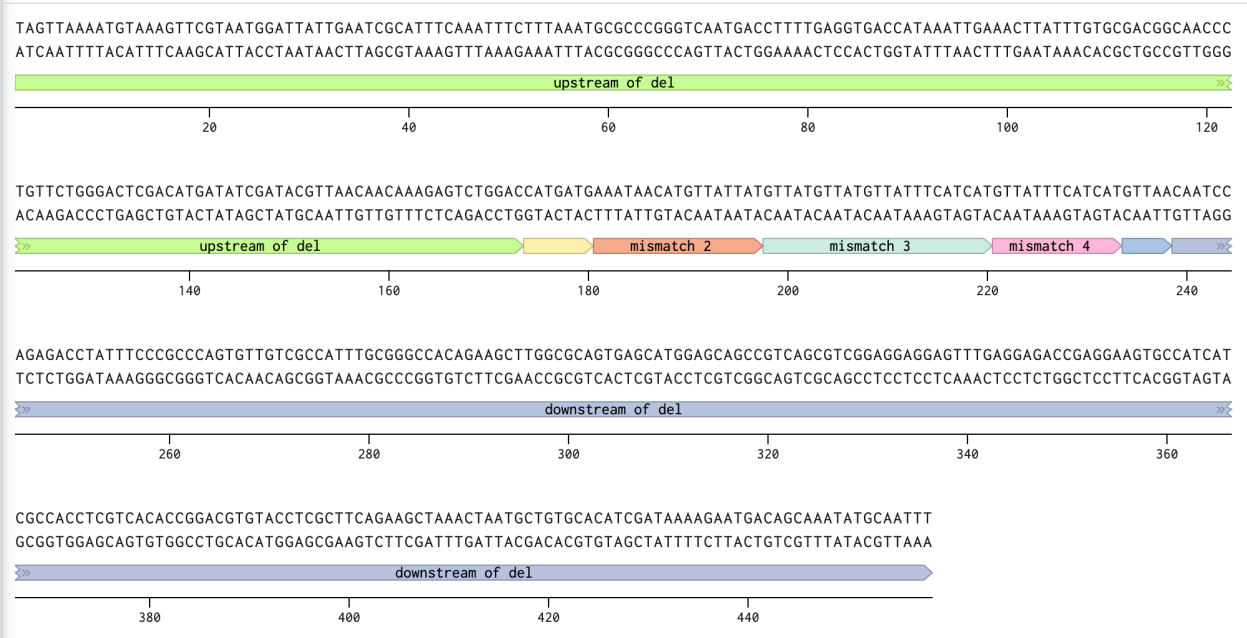
TAGTTAAAATGTAAAGTTCGTAATGGATTATTGAATCGCATTTCAAATTTCTTTAAATGCGCCCGGGT
CAATGACCTTTTGAGGTGACCATAAATTGAAACTTATTTGTGCGACGGCAACCCTGTTCTGGGACT
CGACATGATATCGATACGTTAACAACAAAGAGTCTGGAC----------

CATGATG----------
---AAATAACATGTTATTAT----------
GTTATGTTATGTTATTTCATCAT----------

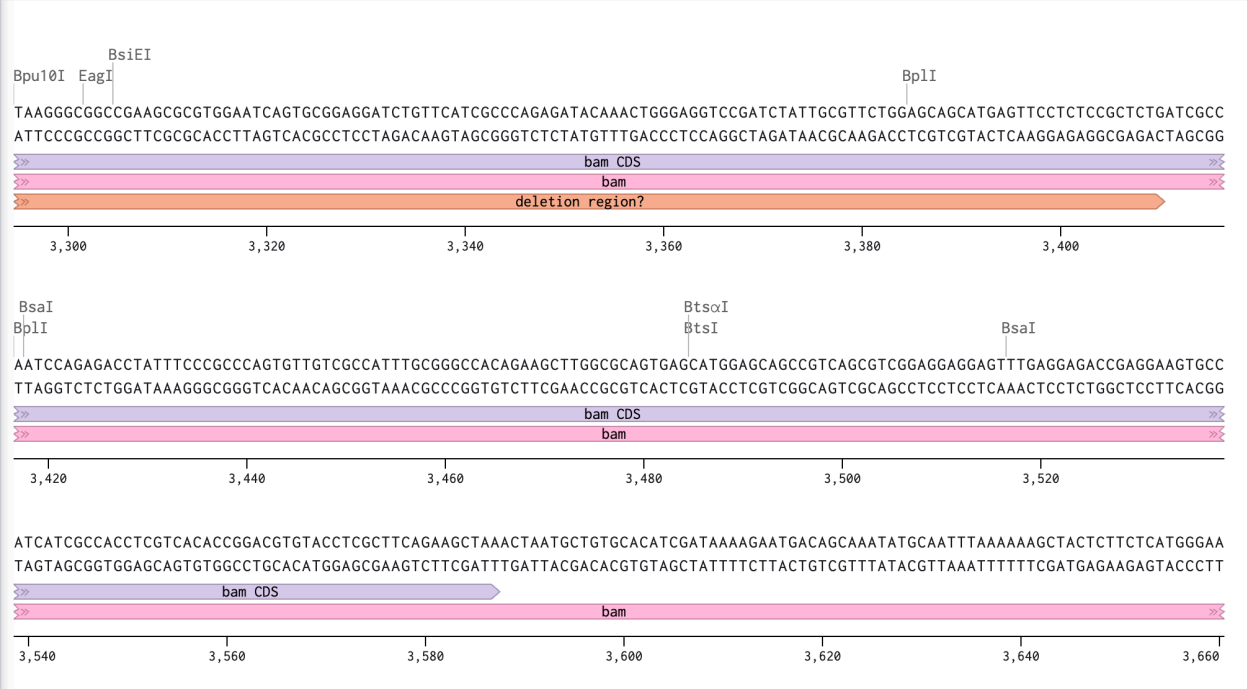
----------GTTATTTCATCAT----------

GTTAACAATCCAGAGACCTATTTCCCGCCCAGTGTTGTCGCCATTTGCGGGCCACAGAAGCTTG
GCGCAGTGAGCATGGAGCAGCCGTCAGCGTCGGAGGAGGAGTTTGAGGAGACCGAGGAAGT
GCCATCATCGCCACCTCGTCACACCGGACGTGTACCTCGCTTCAGAAGCTAAACTAATGCTGTG
CACATCGATAAAAGAATGACAGCAAATATGCAATTT

- Annotated in Benchling:



- Not sure how to analyze/predict the effect on the protein:



- Seems like there could be more than 6 residues left at the C-terminus?
- I don't know where translation will start in this case, since the original start codon was deleted.
  There is one possible start codon on the forward strand and several on the reverse strand in the C-terminal region of the CDS that is not deleted: