

Akkermansia BAA-835: reference sequence and annotations

Error in reference sequence (Amuc_1422, recG)

- >> G at 1704819 is not real (run of 3Gs, actually 2Gs)
- >> results in frame shift in recG, and missing annotation

Sequence info

- 2Gs found in all Illumina runs on spontaneous GNfast mutants
- 2Gs found in Illumina run of our BAA-835
- 2Gs confirmed by Sanger sequencing in multiple BAA-835 examined (see below)
- 2Gs also found in Akkermansia muciniphila YL44 (and many others) - not just our isolate!!!
 - [https://www.ncbi.nlm.nih.gov/nucleotide/CP015409?report=genbank&log\\$=nuclalign&blast_rank=3&RID=8SCE8WYU015](https://www.ncbi.nlm.nih.gov/nucleotide/CP015409?report=genbank&log$=nuclalign&blast_rank=3&RID=8SCE8WYU015)

```
gene      398652..400751
          /locus_tag="A4V05_01725"
CDS       398652..400751
          /locus_tag="A4V05_01725"
          /inference="COORDINATES: protein motif:HMM:PRK10917.2-2"
          /note="Derived by automated computational analysis using
          gene prediction method: Protein Homology."
          /codon_start=1
          /transl_table=11
          /product="ATP-dependent DNA helicase RecG"
          /protein_id="ANU60261.2"
```

Rubritalea marina

WP_018969780

V. spinosum

WP_009959636

A.glycaniphila (NZ_LT629973.1) recG homolog:

```
gene      complement(2436790..2439987)
          /locus_tag="PYTT_RS10210"
          /old_locus_tag="PYTT_2070"
CDS       complement(2436790..2439987)
          /locus_tag="PYTT_RS10210"
          /old_locus_tag="PYTT_2070"
          /inference="COORDINATES: similar to AA
          sequence:RefSeq:WP_006978940.1"
          /note="Derived by automated computational analysis using
          gene prediction method: Protein Homology."
          /codon_start=1
          /transl_table=11
          /product="transcription-repair coupling factor"
          /protein_id="WP_067775434.1"
```

Annotation issues resulting from sequence error in BAA-835

- 1707412..1704235 (3178 bp) with extra G
 - originally annotated as Amuc_1422 (hypothetical)
 - even though there are 3Gs and premature STOP?!? weird
- Am1422 changed at some point to be:
 - 1704801..1704235 (566 bp) RS12060 (hypothetical)
 - 1707412..1704722 (2690 bp) RS07610 (hypothetical)
 - genbank file downloaded Feb. 2018 has these annotations (and fasta has 3Gs)

Fix for INseq

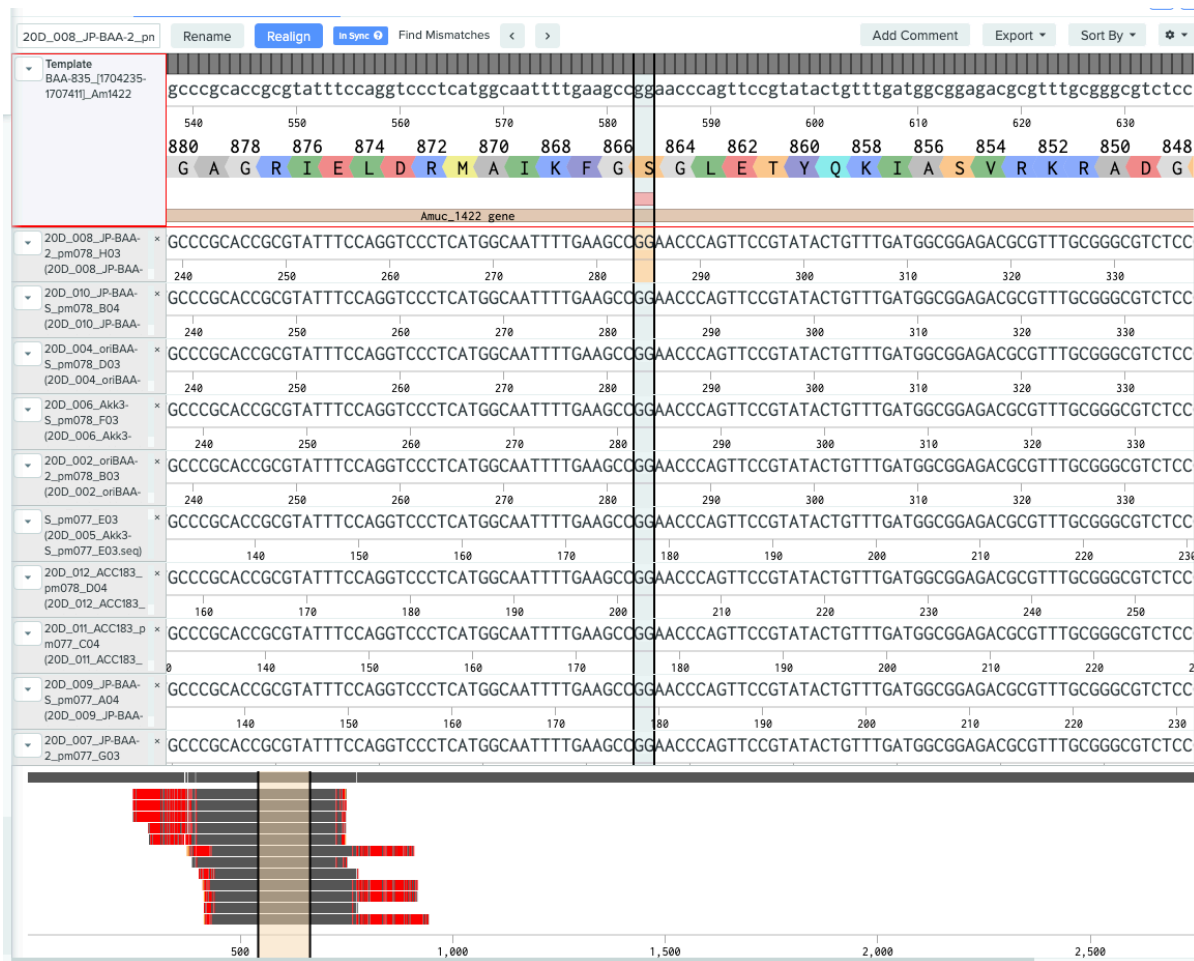
- used refseq annotation file (NC_010655.1)
- changed fasta file to have 2 Gs
- changed .ptt annotation file
 - removed RS12060 and RS07610
 - adjusted indices after 1704819 (all -1 bp)
 - inserted Amuc_1422: 1704235..1707411 (3717 bp)
- new files used for mapping INseq data
 - Akkermansia_BAA835_rv.fna
 - Akkermansia_BAA835_rv.ptt

>>> Communicate with NCBI to fix their information?

>>> KEGG has extra base and pseudogene (3178 nucleotides) for Amuc_1422
https://www.kegg.jp/dbget-bin/www_bget?amu:Amuc_1422

Sanger sequencing Amuc_1422 in multiple BAA-835

- oligos
 - pm-077
 - pm-088
- templates
 - two early frozen stocks made after receiving strain from ATCC
 - later stock, fast grower 18-3
 - JP's BAA-835 (Jon Clardy lab)
- PCR and send for sequencing using same primers used for PCR
- Benchling alignment for all templates, sequenced in both directions... YES, just 2Gs. NCBI genome is wrong
 - Oct 26, 2018 (folder name == 2018-10-2520D_786068)
- CAATTTTGAAGCCGGAACCCAGTTCCGT



History of CP001071 genbank entries at

<https://www.ncbi.nlm.nih.gov/nuccore/CP001071.1?report=girevhist>

Searched 2008, 2010, and 2011 revisions and did NOT find full annotation information for 1422. It's there, but just listed as a pseudogene. 2008 record is 2664102 nucleotides long, so appears to have extra G.

Mar-01-2021

Most recent version of RefSeq annotation NC_010655.1 (2021-07-31) has Amuc_1422

/complement(1704722..1707412)

/old_locus_tag="Amuc_1422"

/locus_tag="AMUC_RS07610"

/product="DEAD/DEAH box helicase"

/protein_id="WP_052294467.1"

but sequencing error is still present, causing frameshift at aa# 865 and termination at 896
correct sequence has 1058 aa

*Older version of RefSeq NC_010655.1 (21-OCT-2017) also has this annotation, but no 'old_locus_tag' field