Possible intron in Coco12 lysin A

potential splice junctions and effect on AA sequence

Aligned DNA to see insertion. A duplication of AACGAG is part of the DNA insertion relative to non-intron relatives. So DNA insertion boundary is ambiguous within the duplication.

```
                     26,300     26,306
                          |     |
Coco12_lysinA     451 GACTTCAACGAGATGCGGGCGTGCATTAAAGACGGGCGCGGGTGAGACGC    500
                      ||||||
Phanphagia_ly     451 GACTTC----------------------------------------    456

                                                    26,884    26,889
                                                         |    |
Coco12_lysinA    1001 AGGTGGGCCCTGAACCCCATGACAAACCGGCCTCCGTATAACGAGTTTCC   1050
                                                         ||||||||||
Phanphagia_ly     457 ------------------------------------AACGAGTTTCC    467

Coco12_lysinA    1051 GATCTGGTCGAACAACAACAGCGCCCGCAGCGGCAAGCCCACCATGTTCC   1100
                      ||||||||||||||||||||||||||||||||||||||||||||||||||
Phanphagia_ly     468 GATCTGGTCGAACAACAACAGCGCCCGCAGCGGCAAGCCCACCATGTTCC    517
```

If DNA insertion is removed, Coco12 protein is identical in this region to other lysin As. The duplicated region is the two codons for Asn-Glu (NE), which is at the boundary between N- and C- domains of lysin As in cluster F.

Welkin suggested to choose splice junctions that (a) preserved the AA sequence and (b) at least had a 3'T at the end of the upstream exon (3'U in the RNA) if we couldn't find a 3'T in the upstream exon and a 3'G for the intron. (Craig Peebles emphasized importance of the 3'U in the exon.)

If we demand that the AA sequence remain unchanged. 26,300C is the left-most nt possible, and 26306G is the rightmost possible, and there are no Ts present in this range, as seen here:

```
F   N  E                  N  E  F
TTC^aacgagatg.........tat^AACGAGTTT
TTCA^acgagatg.........tata^ACGAGTTT
TTCAA^cgagatg.........tataa^CGAGTTT
TTCAAC^gagatg.........tataac^GAGTTT
TTCAACG^agatg.........tataacg^AGTTT
TTCAACGA^gatg.........tataacga^GTTT
TTCAACGAG^atg.........tataacgag^TTT
```

But if we shifted one upstream, we could make a 3'T for the upstream exon. This changes the TTC codon to TTT, both code for F.

```
TT^Caacgagatg.........ta^tAACGAGTTT
```

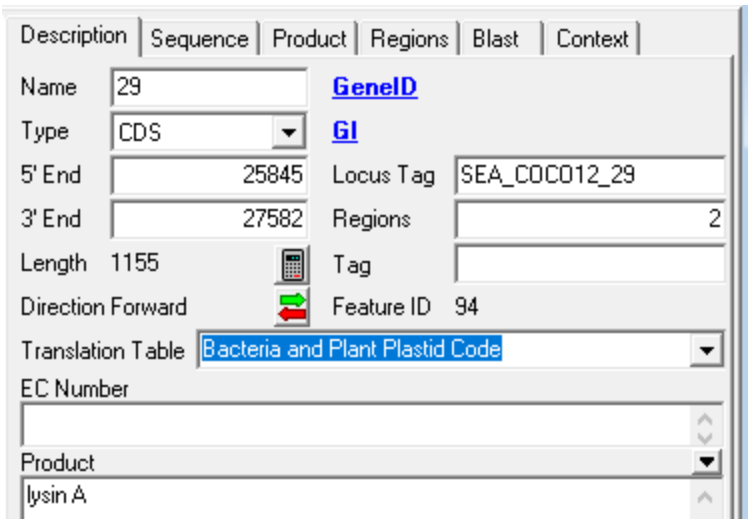
We chose this set for the splice junctions.

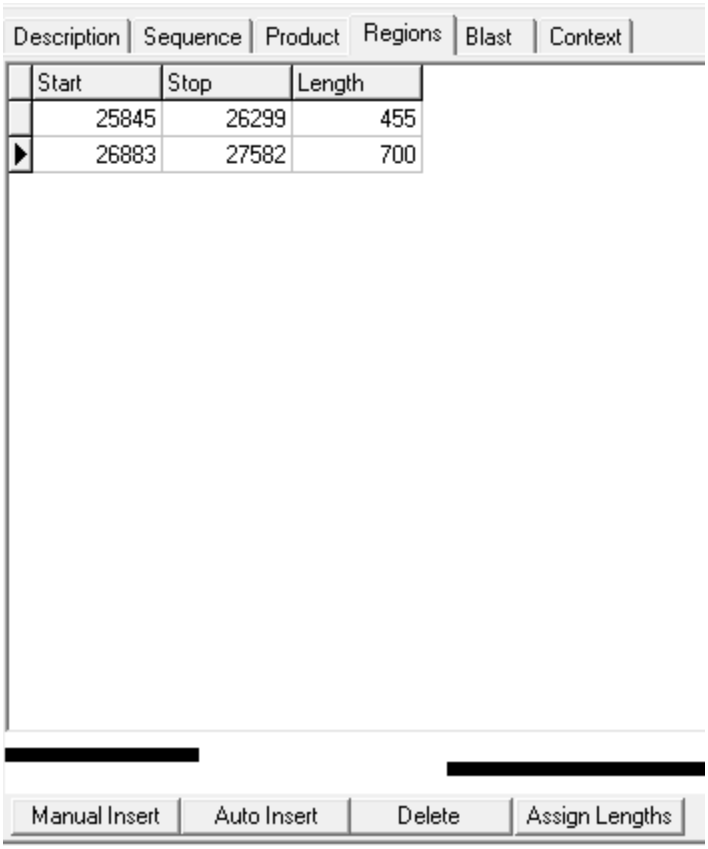
```
----showing intron in red and duplicated bit in blue ---------------------
        H  T  S  T  S  T  G  P  P  L  E  V  P  P  C  P  T  N  H  A  P  T  S
      P  I  P  R  Q  V  L  A  H  H  W  R  Y  R  R  A  R  R  T  T  P  R  L  Q
       P  Y  L  D  K  Y  W  P  T  T  G  G  T  A  V  P  D  E  P  R  P  D  F
26231 CCCATACCTCGACAAGTACTGGCCCACCACTGGAGGTACCGCCGTGCCCGACGAACCACGCCCCGACTTC
      ....:....|....:....|....:....|....:....|....:....|....:....|....:....|

       T  R  C  G  R  A  L  K  T  G  A  G  E  T  L  G  I  P  C  P  R  A  L
        R  D  A  G  V  H  *  R  R  A  R  V  R  R  L  A  Y  R  A  R  A  R  S
      N  E  M  R  A  C  I  K  D  G  R  G  *  D  A  W  H  T  V  P  A  R  A  R
26301 AACGAGATGCGGGCGTGCATTAAAGACGGGCGCGGGTGAGACGCTTGGCATACCGTGCCCGCGCGCGCTC
      ....:....|....:....|....:....|....:....|....:....|....:....|....:....|

        I  L  V  E  R  W  A  L  N  P  M  T  N  R  P  P  Y  N  E  F  P  I  W
      P  F  *  S  K  G  G  P  *  T  P  *  Q  T  G  L  R  I  T  S  F  R  S  G
       H  F  S  R  K  V  G  P  E  P  H  D  K  P  A  S  V  *  R  V  S  D  L
26831 CCATTTTAGTCGAAAGGTGGGCCCTGAACCCCATGACAAACCGGCCTCCGTATAACGAGTTTCCGATCTG
      ....:....|....:....|....:....|....:....|....:....|....:....|....:....|
```

Amino acid sequence of lysin A is preserved

```
PRPDFNEFPIW
||||||||||
PRPDFNEFPIW
```

Here is how we entered it into DNA Master.





The documentation ended up looking like this:

```
CDS join(25845. .26299,26883. .27582)
  /gene="29"
  /product="lysin A"
  /locus tag="SEA_COCO12_29"
    /note=lysin A
```