Amino acid sequence for ORF with stop at 42523 (Ciao) and 42306 (Pippin):

```
Ciao_66:       MKSS     RALWEAAPYGRPFGWSVHPNEEMRDHYRARASKLLDQFEIKERS
Pippin_68:     MDIVEKLA RALWEAAPYGRPFGWSVHPNEEMRDHYRARASKLLDQFEIKERS
```

note that I added the space above to highlight the identical sequence of this translation after the first 4-8 AA's

Here's where it gets interesting – I was trying to verify the start of the Ciao gene ending at 42523 and Ciao is missing the start that is called by 47 of 49 phamily members. I had also noticed that Ciao and Pippen have extremely high nucleotide identity (985 of 988nt) in most of this region, and that Ciao is missing the start found in Pippin.

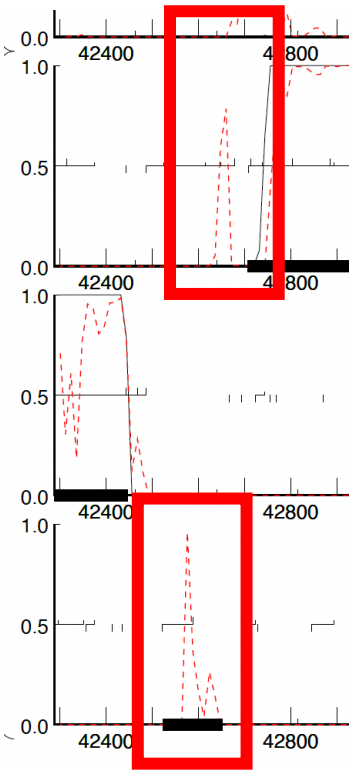
Also, when I initially went through the GenemarkS data eyeballing the coding potential, I had put a note in this region that there is some coding potential in another reading frame than what was called.

So down a rabbit-hole I went, and here's what I found:

1. Nucleotide alignment – red is region where Ciao lost one of the "GC" repeats. Turquoise is the start codon used in Ciao, Magenta is the start used in Pippin_68.

   This is a zoom on region where deletion is – the rest is 100% identical in a region much larger than just this one gene.

2. The GenemarkS file shows coding potential in this region is split (at right) between two frames – the top frame contains the original start codon (in frame) found in Pippin, while the bottom frame is the large piece of the gene that's identical above.



QUESTION:

Wouldn't the original start site override the greater coding potential in the "bottom" frame here since the ribosome is cued into elements around/upstream of the promoter?

Based on this, I'd call the gene in a different reading frame (the top one, with the magenta start from Pippin), which will destroy most of the AA similarity due to the frameshift generated after the 2bp deletion and likely generate an orpham.

The RBS sites for the two (turquoise and magenta) possible starts are on the next page.

The original/magenta site has slightly (significantly better final score?) better RBS characteristics and is ATG rather than TTG.

Internal Start Analysis

ORF Start : 42669
ORF Stop : 42523
ORF Length : 147

Idx	Raw Score	Z	Final Score	Spacer	Start Pos'n	Start Codon	Sequence
1	-2.273	2.7870	-4.273	17	42669	TTG	GAGGAGGAAGCAATGGACATCG

Internal Start Analysis

ORF Start : 42679
ORF Stop : 42485
ORF Length : 195

Idx	Raw Score	Z	Final Score	Spacer	Start Pos'n	Start Codon	Sequence
2	-2.071	2.8835	-2.765	10	42679	ATG	GATTCGTAACGAGGAGGAAGCA