

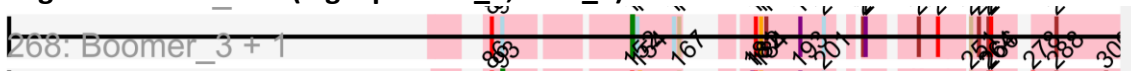
## Second opinion: Cluster F1 gene start site

Hello everyone,

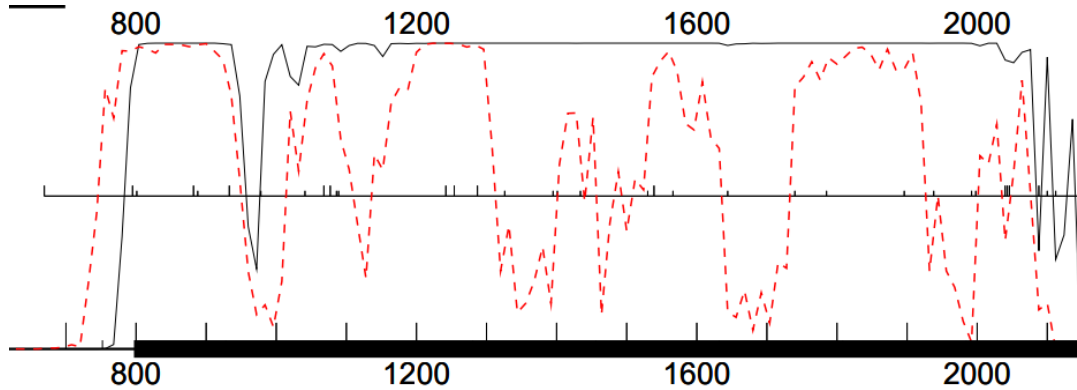
I'd welcome some thoughts on our JalFarm20\_3 start site. Thank you!

1. Glimmer and GeneMark agree on start: nucleotide 797. Not longest ORF. Could be nucleotide 671 or 659, but that would create overlap with JalFarm20\_3 by 28 or 40 bp.

2. The start is not conserved among all members of the pham (Starterator). JalFarm20\_3 start is 152, same as UncleRicky\_3, UPitt annotation-sequenced in 2017; most called is 103; similar phages earlier start 93 (e.g. SpikeLee\_3; Bobi\_3).



3. Currently predicted start site does not include all coding potential in GeneMark in +2 frame.



**4. For genes with functional predictions, start bp 797 includes the full-length protein using BLASTP, terminase large subunit.**

Sequences producing significant alignments:	Score (bits)	E Value
Tootsieroll_Draft_3, function unknown, 495	<a href="#">915</a>	0.0
Spikelee_3, terminase, 495	<a href="#">915</a>	0.0
Ramsey_3, terminase, 495	<a href="#">915</a>	0.0
JalFarm20_Draft_3, function unknown, 453	<a href="#">915</a>	0.0
Bob_3, terminase, 495	<a href="#">915</a>	0.0
Cornucopia_3, terminase large subunit, 495	<a href="#">914</a>	0.0
SimranZ1_3, terminase large subunit, 495	<a href="#">913</a>	0.0
Leozinho_Draft_3, function unknown, 453	<a href="#">913</a>	0.0
BuzzLyseyear_3, terminase large subunit, 495	<a href="#">913</a>	0.0
Boomer_3, terminase, 453	<a href="#">913</a>	0.0
Alexphander_3, terminase large subunit, 495	<a href="#">913</a>	0.0
UncleRicky_3, terminase large subunit, 453	<a href="#">912</a>	0.0
Rockyhorror_3, terminase, 495	<a href="#">912</a>	0.0
Pollywog_3, terminase large subunit, 495	<a href="#">912</a>	0.0
Lorde_3, terminase large subunit, 495	<a href="#">912</a>	0.0
Enby_3, terminase large subunit, 495	<a href="#">912</a>	0.0
Drago_3, terminase large subunit, 453	<a href="#">912</a>	0.0
DaWorst_3, terminase large subunit, 495	<a href="#">912</a>	0.0
Job42_3, terminase, 495	<a href="#">912</a>	0.0
Velveteen_3, function unknown, 495	<a href="#">910</a>	0.0
Sarma624_Draft_3, function unknown, 453	<a href="#">910</a>	0.0

>Boomer\_3, terminase, 453  
Length = 453

Score = 913 bits (2359), Expect = 0.0  
Identities = 452/453 (99%), Positives = 452/453 (99%)

Query: 1 MGVEFDRWQEDIWYAALGLREDGTLACDVMGVTLSIARQAGKTWGIMVGLIAICLSRPGT 60  
MGV FDRWQEDIWYAALGLREDGTLACDVMGVTLSIARQAGKTWGIMVGLIAICLSRPGT  
Sbjct: 1 MGVAFDRWQEDIWYAALGLREDGTLACDVMGVTLSIARQAGKTWGIMVGLIAICLSRPGT 60

>Spikelee\_3, terminase, 495  
Length = 495

Score = 915 bits (2365), Expect = 0.0  
Identities = 453/453 (100%), Positives = 453/453 (100%)

Query: 1 MGVEFDRWQEDIWYAALGLREDGTLACDVMGVTLSIARQAGKTWGIMVGLIAICLSRPGT 60  
MGVEFDRWQEDIWYAALGLREDGTLACDVMGVTLSIARQAGKTWGIMVGLIAICLSRPGT  
Sbjct: 43 MGVEFDRWQEDIWYAALGLREDGTLACDVMGVTLSIARQAGKTWGIMVGLIAICLSRPGT 102

Query: 61 LVVWSSHHDRTSSETLTKIAGIVEKPAIRPKMRPMHPVVQSDNDRGVHVFANGSRILFGAR 120  
LVVWSSHHDRTSSETLTKIAGIVEKPAIRPKMRPMHPVVQSDNDRGVHVFANGSRILFGAR  
Sbjct: 103 LVVWSSHHDRTSSETLTKIAGIVEKPAIRPKMRPMHPVVQSDNDRGVHVFANGSRILFGAR 162

BLASTP with start bp 671 (overlaps with JalFarm20\_2: by 28 bp) produces alignments as well.

BLAST search for Translation of 671 - 2158

Score	Description
2482	gp3 [Mycobacterium virus Ramsey] > ref YP_008408960.1  terminase [Mycobacterium phage Bob] > ref YP_009962456.1  terminase [Mycobacterium phage Spikelee] > gb AC12615.1  terminase [Mycobacterium virus Ramsey]
2479	terminase large subunit [Mycobacterium phage Cornucopia] > gb Q8P31481.1  terminase large subunit [Mycobacterium phage Cornucopia]
2478	terminase [Mycobacterium virus Rockyhorror] > ref YP_009955918.1  terminase large subunit [Mycobacterium phage Dal'fort] > gb AEK06711.1  terminase [Mycobacterium virus Rockyhorror] > gb ATN88495.1  terminase large subunit [Mycobacterium phage Buzzl'ryeyes] > ref YP_009954378.1  terminase large subunit [Mycobacterium phage Aleghande] > ref YP_009962353.1  terminase large subunit [Mycobacterium phage Simran]
2472	terminase large subunit [Mycobacterium phage Eruby] > ref YP_009961516.1  terminase large subunit [Mycobacterium phage Pollywog] > gb GA110933.1  terminase large subunit [Mycobacterium phage Pollywog] > gb QJL895C
2471	terminase large subunit [Mycobacterium phage Piper2020] > gb Q8P31683.1  terminase large subunit [Mycobacterium phage Piper2020]
2470	terminase [Mycobacterium phage Job42] > gb AGM61419.1  terminase [Mycobacterium phage Job42]
2468	putative terminase [Mycobacterium phage WIVsmall] > gb AGK88210.1  putative terminase [Mycobacterium phage WIVsmall]
2466	terminase large subunit [Mycobacterium phage Squarty] > gb AHM40950.1  terminase large subunit [Mycobacterium phage Squarty]

High-Scoring Pair (HSP)

B#	Score	Query	1-495	Positives	495
	Score	Target	1-495	Identity	495
E-Value	0.0E0	Length	495	Similarity	100.0
		% Aligned	100.0	Gaps	0

```

1  MPTGGSRVW PAVLVDIARQ CFVDDISHT RYELIAPEL PGMVVEFRM
1  |||
1  MPTGGSRVW PAVLVDIARQ CFVDDISHT RYELIAPEL PGMVVEFRM
51  QEDWVAALG LREDGTLCD VNVVLSIAR QAGHTWGMV GLIATCLSRP
51  |||
51  QEDWVAALG LREDGTLCD VNVVLSIAR QAGHTWGMV GLIATCLSRP
101  GTLVVSSHK DRTSSETLK IAGIVEKPAI RPIGRHHPV VQSDNRGVW
101  |||
101  GTLVVSSHK DRTSSETLK IAGIVEKPAI RPIGRHHPV VQSDNRGVW
151  FANGSRILXK XXXXXXXXXXXX SEVDIQVYDE CQNKESALT DMLAANVSE
151  |||
151  FANGSRILFG ARAQGGGRGF SEVDIQVYDE CQNKESALT DMLAANVSE
201  IGLAFFHGTG PRPQVALGV HDAFKRRDR ALQKRRGF KGVVYEFAPF
201  |||
201  IGLAFFHGTG PRPQVALGV HDAFKRRDR ALQKRRGF KGVVYEFAPF
251  SPDDVVADID APGFWRDLAE ANPFGHRVQ KSAIERLVEN HSPEDVRRV
251  |||
251  SPDDVVADID APGFWRDLAE ANPFGHRVQ KSAIERLVEN HSPEDVRRV

```

### 5. Start bp 671 has a better final RBS score.

Starts: 50    ORF Start: 797    Cdn1 Cdn2 Cdn3 Length

Selected: 1    ORF Stop: 2158    5' End 25.0 75.0 75.0 12    SD Scoring Matrix Kibler6    Explore

ORF Length: 1362    3' End 61.9 47.8 67.3 1488    Spacing Weight Matrix Karlin Medium    Document

Sta#	Raw Score	SD	Genomic Z Value	Spacer Distance	Final Score	Sequence of the Region	Start Codon	Start Position	ORF Length
1	-4.870	1.572	7	-6.393	GCTCCGACTGAGTCCGGCRAAG	TTG	659	1500	
2	-3.282	2.340	14	-4.629	TCCGGCAAGGTTGTGGCGTTCG	ATG	671	1488	
3	-5.454	1.290	13	-6.500	GATTGCCCCAGAGCTCCCGGT	ATG	797	1362	
4	-4.169	1.911	9	-4.944	CCCAGAGCTCCCGGTATGGGT	GTG	803	1356	
5	-5.111	1.456	11	-5.868	GGAAGACATTGGTACGCGGCC	TTG	845	1314	
6	-5.647	1.197	12	-6.482	CATTGGTACGCGCCTTGGGT	TTG	851	1308	
7	-4.910	1.553	13	-5.955	CGGCACGCTGGCGTGGACGTC	ATG	884	1275	
8	-4.651	1.679	5	-6.651	GCTGGCGTGGCAGCTCATGGT	GTG	890	1269	

## 6. Synteny evident.

