ESTAVE1 coversheet:

This genome is dominated overwhelmingly by tightly packed rightward transcribed genes. Only about ten percent of called features are left-transcribed. The predominating genes utilize all three rightward reading frames seemingly equally and many regions are very dense. A common motif is 4bp overlap left and right. There are sizeable regions in which the genes seemingly stair-step across the frames. No tRNAs or tmRNAs are called by any of the three software tools: internal Aragorn, external Aragorn or tRNAscanSE.

The only hugely troublesome thing here is probably what to do with the possible frameshift upstream from tapemeasure (start@9502). It’s been left alone because 1) there seems to be no overlapping coding potential 2) guidance from related genomes is unclear. It’s noteworthy that the BLAST results for gene 13 are significantly mis-aligned. Other than that:

1. largish gaps: We’ve tolerated the following in part due to ORF length/ SD issues and also because it just seems to go along with the overall plan of this genome:

13bp overlaps: betw. genes 30 and 40; and 320/330; 10bp overlap 680/690; 8bp overlaps: 540/550,760/79.5,810/820 and 107/108

2. Deletion of genes: nothing controversial here, deleted originally called features starting @ 31849, 37340 (REV), 48103, 58042 (REV), 59860 (REV) and 60250 (REV). All of these either completely overlapped with better-supported features, or had no BLAST results, or both. In addition, the software called an inappropriate wrap-around gene (genome has defined physical ends) that was also deleted.

3. Addition of genes: In order to fill sizeable gaps that, although most contained no coding potential did return meaningful homology results on NCBI-BLASTx, the following genes were added: 46.5 starting at 33787 (REV), 79.5 starting at 49101, [101.3 @ 53340](mailto:101.3%20@%2053340), [101.5 @ 55459](mailto:101.5@55459) and 112 @ 60259. A sizable gap in the middle of the genome, just downstream from gene 530, is being left alone because in addition to having no discernible coding potential, NCBI-BLASTx returns nothing meaningful. Gene 112 is the only one that NCBI-BLASTx actually located a functional domain in (HNH-endonuclease). This was corroborated by BLASTp upon inserting the gene; it correlates well with the maps of closely related ShiLan and Mutaforma , both of which end with a gene with these domains (per phamerator). It’s also the only added gene that covers noteworthy GeneMark coding potential.

4.The region centered around gene 470 is problematic. Crux of the matter is how to accommodate promoters for genes transcribed in opposite directions without unduly messing with genes that may have interesting function. The upstream gene (leftward transcribing) was shortened by favoring the GeneMark call to accommodate two promoters for transcription in opposite directions, but functional info. is murky. The leftward gene may or may not code for a repressor (several BLAST hits but not well aligned, HHPRED finds homology over limited alignment to transcriptional regulators), so one is reluctant to mess with it. Another approach would have been to delete the 34529 ORF, but there are indications that this may be a cro type protein gene. …..

5. Several extreme right hand features have good cp and SD scores but BLAST very poorly both at NCBI and locally @ phagesdb. This includes a couple with just one, two or three BLAST hits (genes 850, 920, 980) and one (gene 880) actually with none—retained because of SD/cp considerations, they don’t overlap anything heavily and they seem to fit the general pattern of other genomes. Genes 220 and 230 also don’t BLAST well and as leftward genes seem “out of place” where they are , but they cover significant coding potential.