

MGF 360 Figure Legend

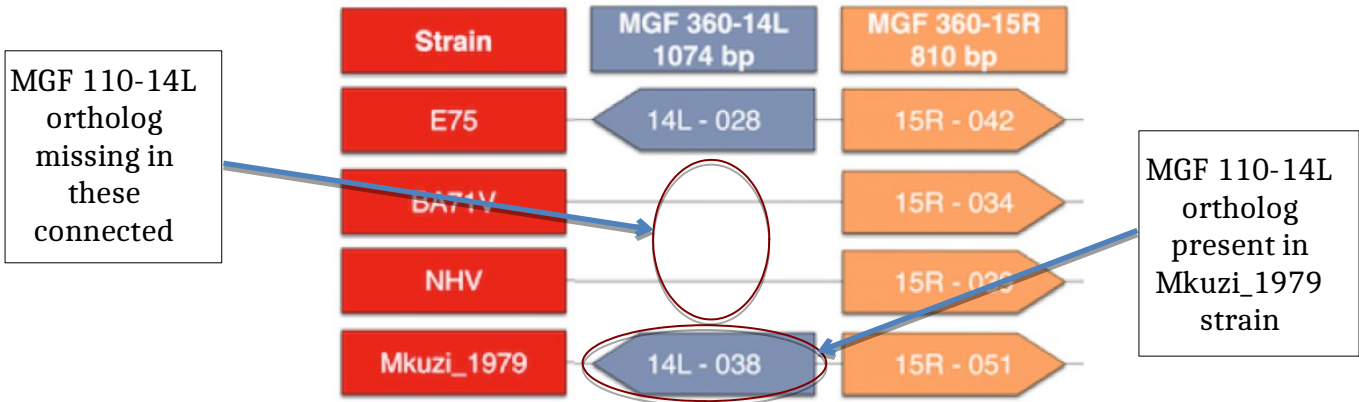
MGF 360 ortholog group

Approximate size of genes with this ortholog



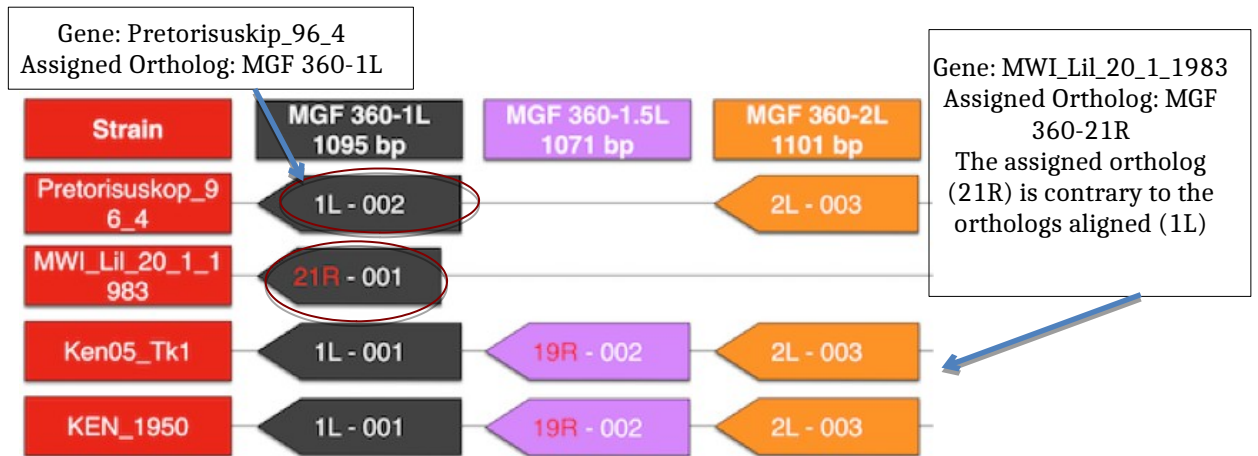
Ortholog columns and gene size:

All genes of same width as the heading are approximately the length of the reference length stated.



Genes:

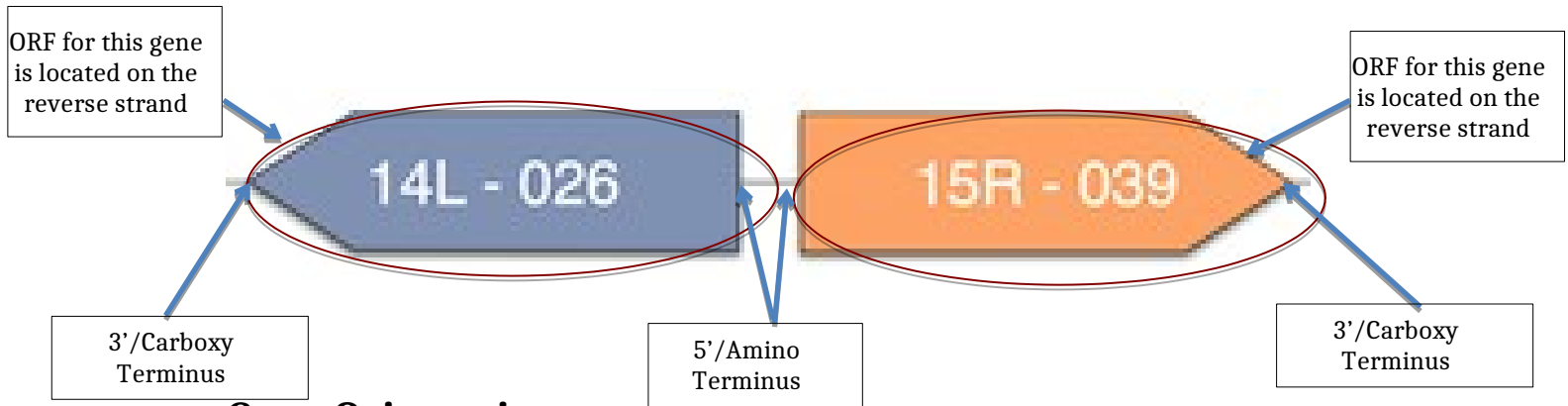
Each of the boxes below a MGF heading indicate the presence of that ortholog in the connected strain. The above figure indicates MGF 110-14L ortholog is missing in strains BA71V and NHV, but present in E75 and Mkuzi_1979



Gene labels:

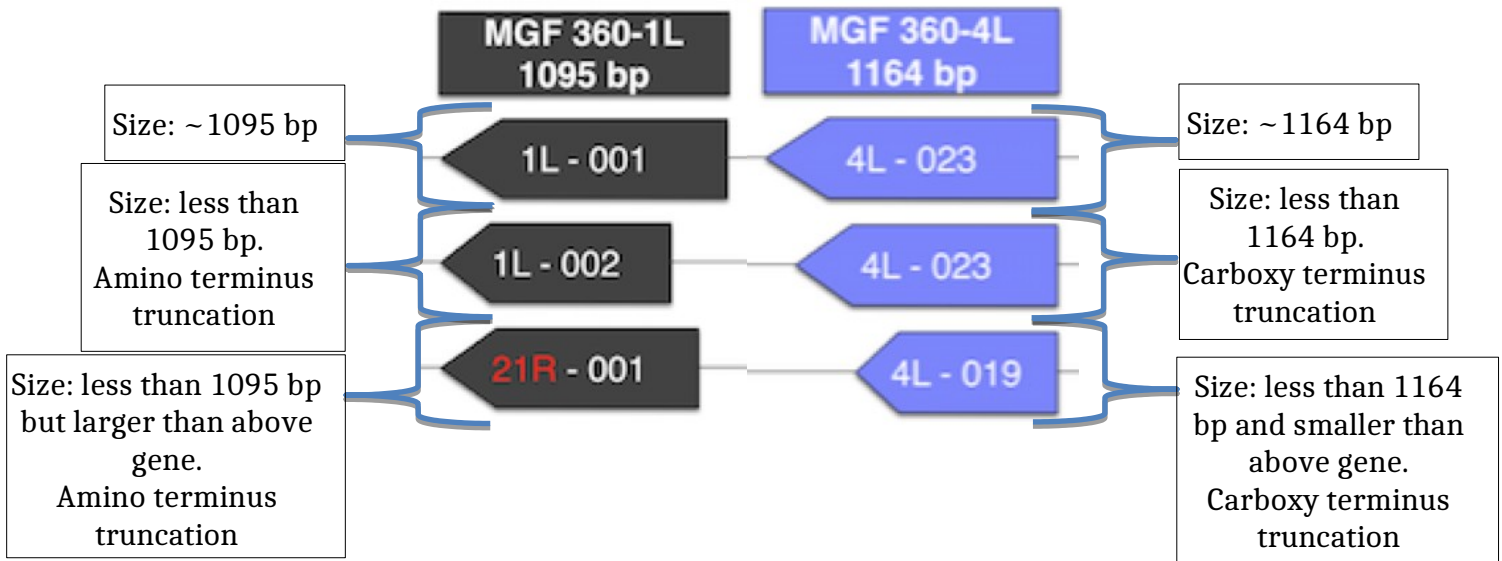
The annotation of each gene is in two parts: first, the ortholog group originally assigned during initial annotation, followed by the corresponding gene number of the connected strain. If the currently assigned ortholog group is labelled in **red**, this indicates that the currently assigned ortholog is contrary to orthologs it aligns to within this diagram.

MGF 360 Figure Legend



Gene Orientation:

MGF 100 has “R” orthologs that are transcribed on the forward strand (5' → 3') and “L” orthologs that are transcribed on the reverse strand (3' ← 5'). “R” gene boxes are pointed to the right and “L” gene boxes are pointed to the left.



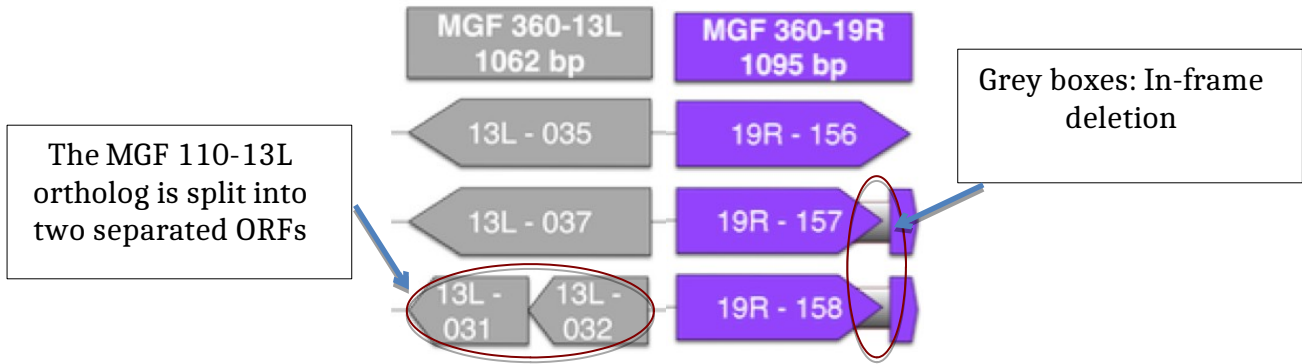
Relative gene size and truncations:

A smaller or larger gene box indicates the size difference of the gene relative to other genes of the same ortholog

A 5'/amino terminus that is not aligned with the amino terminus of the full-length genes indicates an amino terminus truncation of this gene.

A 3'/carboxy terminus that is not aligned with the carboxy terminus of the full-length genes indicates a carboxy terminus truncation of this gene.

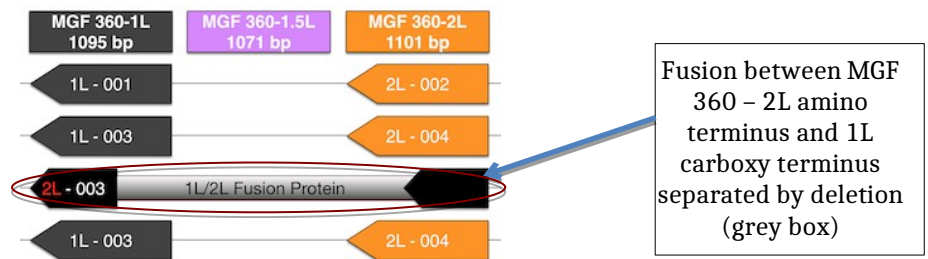
MGF 360 Figure Legend



Special note Gene Features:

The fragmentation of an ortholog into two smaller ORFs is represented by two smaller gene boxes underneath a single heading.

For the bottom two genes of the 19R ortholog in the above diagram is shown to have a large in frame deletion in the gene when compared to the aligned genomes.

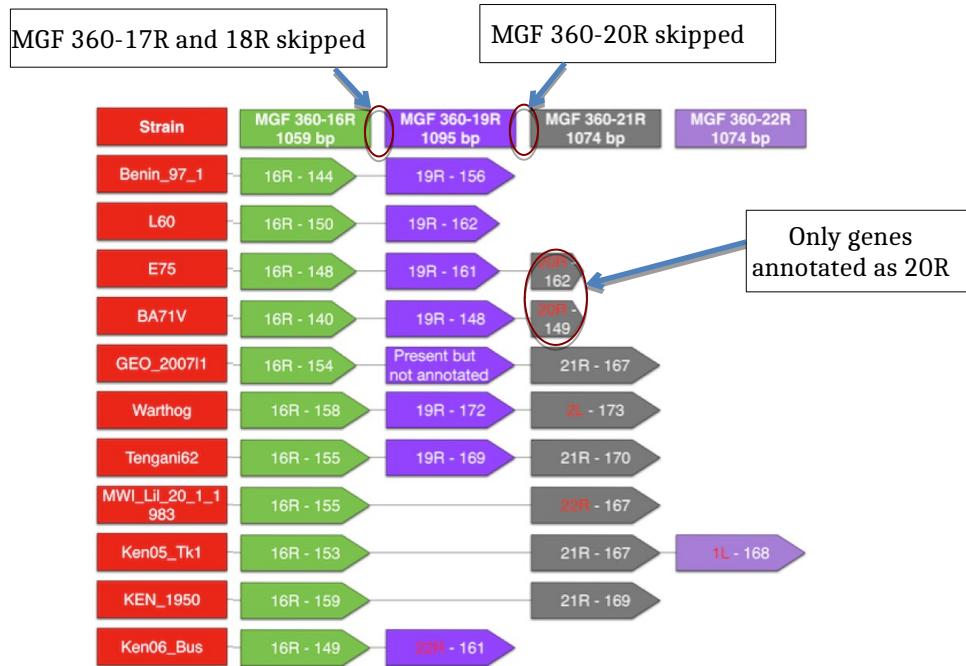


MGF 360 Fusion:

Genes encoded by an open reading frame that aligns across multiple ortholog loci due to genomic deletions are labelled in the above diagram as “fusion” genes and are represented by black gene boxes. The size and alignment of the black gene boxes to a given ortholog group is representative of how much of the ortholog is fused and which region.

The fused ortholog groups are labelled along the deletion box connecting the fragments.

MGF 360 Figure Legend



MGF 360-17R Missing:

MGF 360-17R is not annotated in any of these genomes and was likely removed due to no longer being classified as a MGF 360 ortholog.

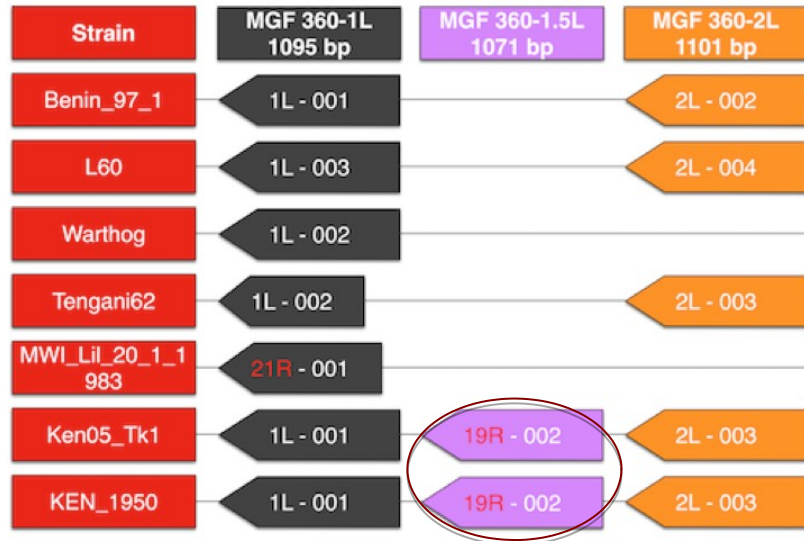
MGF 360-18R Missing:

The MGF 360-18R ortholog is present in most of these genomes however it has been since declared that this gene is not an MGF so it was not included in this diagram.

MGF 360-20R Missing:

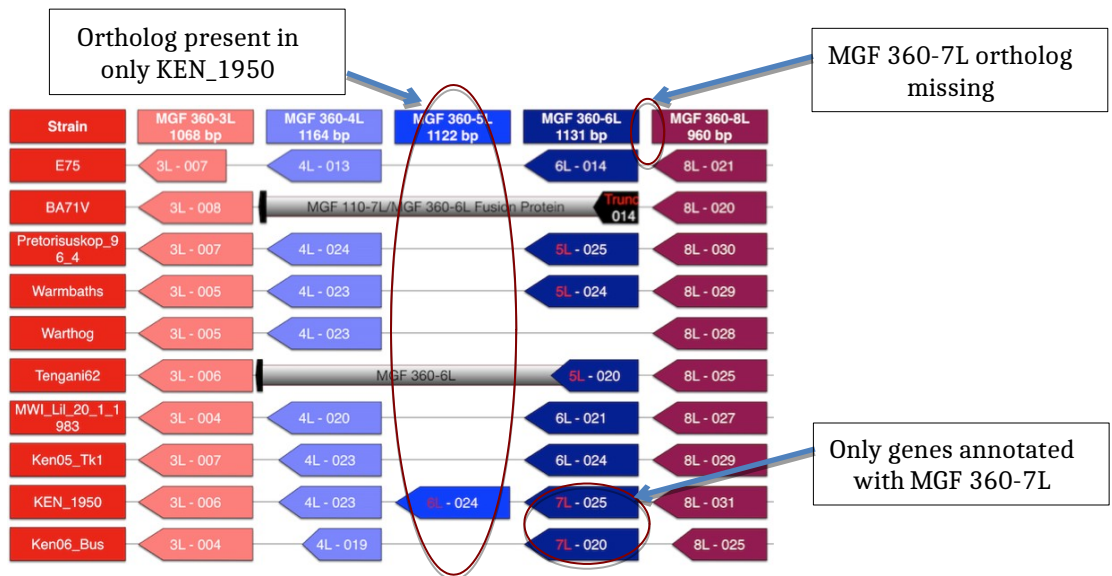
The MGF 360-20R ortholog is annotated in only E75 and BA71V strains which both appear to be truncated 21R orthologs. These two genes are quite short due to a carboxy terminus truncation. It is possible that when comparing to these genes they appeared to form their own ortholog, causing any subsequent assignments to be a new ortholog.

MGF 360 Figure Legend



MGF 360-1.5L:

The MGF 360-1.5L ortholog is present due to a large insertion present only in strains Ken05_Tk1 and KEN_1950 that aligns between the 1L and 2L orthologs.

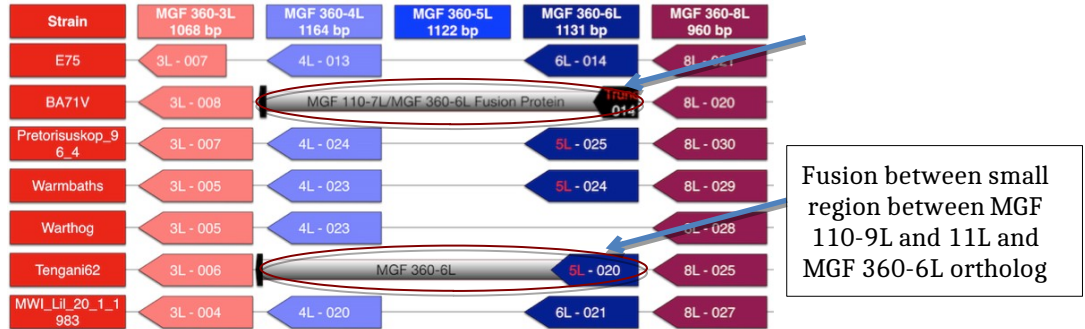


MGF 360-7L Missing:

The alignment between the MGF360-4L, 5L, and 6L orthologs is very complicated due to the unique insertion present in KEN_1950. Removal of the KEN_1950 strain from the alignment shows two distinct 4L and 6L orthologs. This is likely the reason for the annotation of genes under the 6L ortholog as 5L, 6L or 7L. Of all the aligned genes in this region, three genes are annotated as 5L, seven as 6L, and two as 7L. Majority classifies this ortholog as 6L so this is the annotation adopted in this diagram.

Fusion between MGF 110-7L and MGF 360-6L orthologs

MGF 360 Figure Legend



Cross-Diagram Fusions:

Trunc - 014 [MGF 110-7L/MGF 360-6L Fusion Protein]:

This gene is a fusion between the MGF 110-7L ortholog and MGF 360-6L. The carboxy terminus of this fusion is shown as a small box between the 3L and 4L orthologs as the MGF 110 orthologs are outside of the scope of this diagram.

The annotated ortholog for this gene is: "Truncated MGF 360 protein" which has been shortened to "Trunc", however this gene is most likely a fusion between the two MGF orthologs.

MGF 360-5L → MGF 360-6L:

This gene is a fusion between the MGF 360-6L amino terminus and a short non-MGF sequence between MGF 110-9L and 11L. Due to this gene having mostly MGF 360-6L character and only a short region aligned to the mid MGF 110-9L/11L sequence, the gene is given a MGF 360-6L ortholog assignment. The carboxy terminus of this fusion is shown as a small box between the 3L and 4L orthologs as the MGF 110 orthologs are not covered in this diagram.



GEO_2007|1 MGF 360-19R:

The GEO_2007|1 strain has an ORF that aligns with the 19R orthologs but was not annotated.