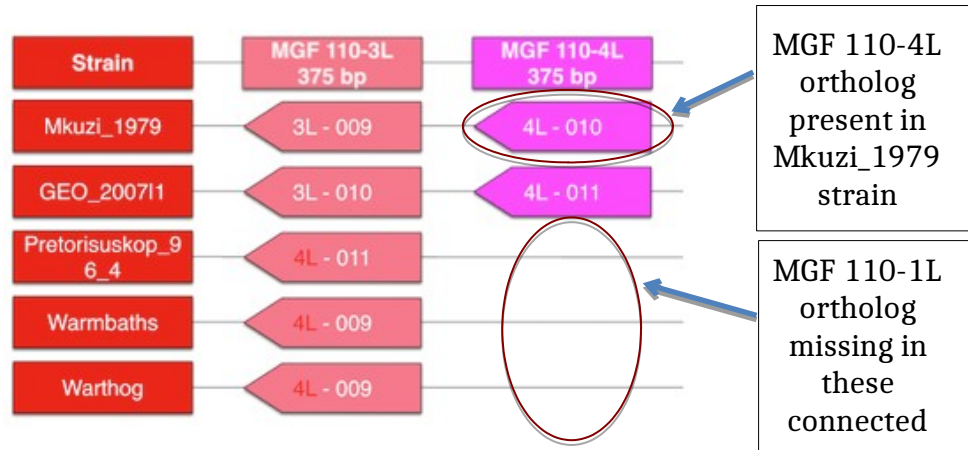


MGF 110 Figure Legend



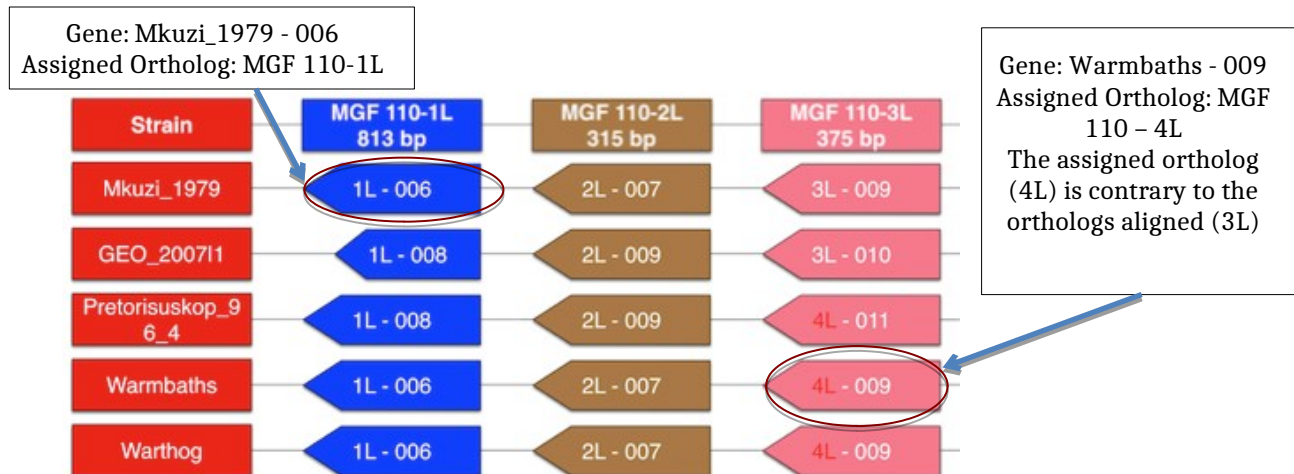
Ortholog columns and gene size

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



Genes:

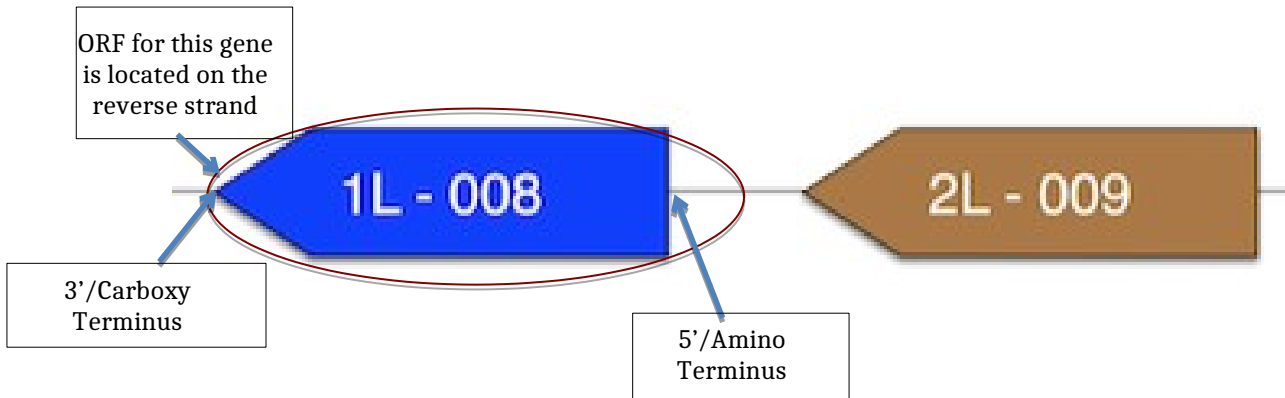
The above figure indicates MGF 110-4L ortholog is missing in strains Pretorisuskop_96_4, Warmbaths, and Warthog, but present in Mkuzi_1979 and GEO_2007|1.



Gene labels:

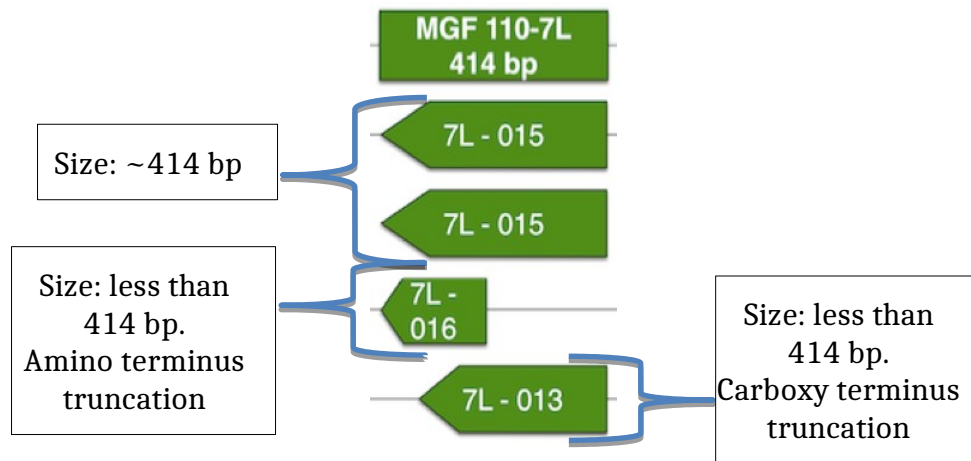
The annotation of each gene is in two parts: first, the ortholog group originally assigned during initial annotation, followed by the corresponding gene number of the connected strain. If the currently assigned ortholog group is labelled in red, this indicates that the currently assigned ortholog is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

MGF 110 Figure Legend



Gene Orientation:

MGF 110 has only “L” orthologs that are transcribed on the reverse strand (3' ← 5'). Their gene boxes are pointed to the left.



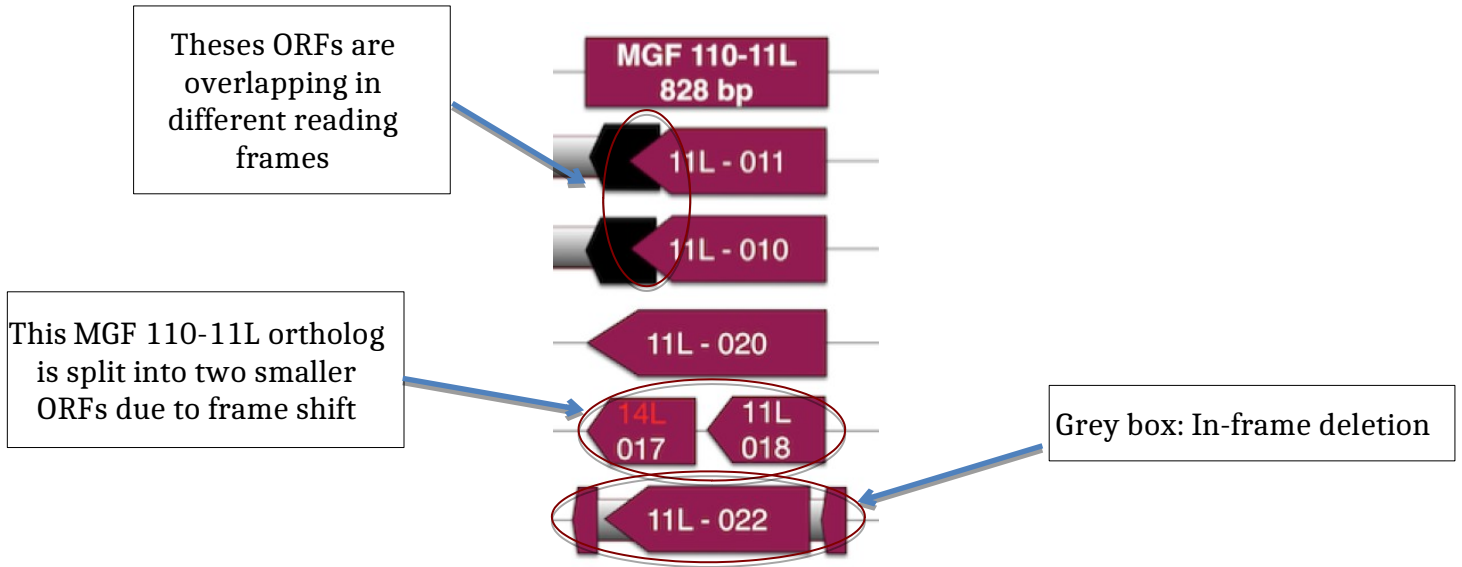
Relative gene size and truncations:

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 110 Figure Legend

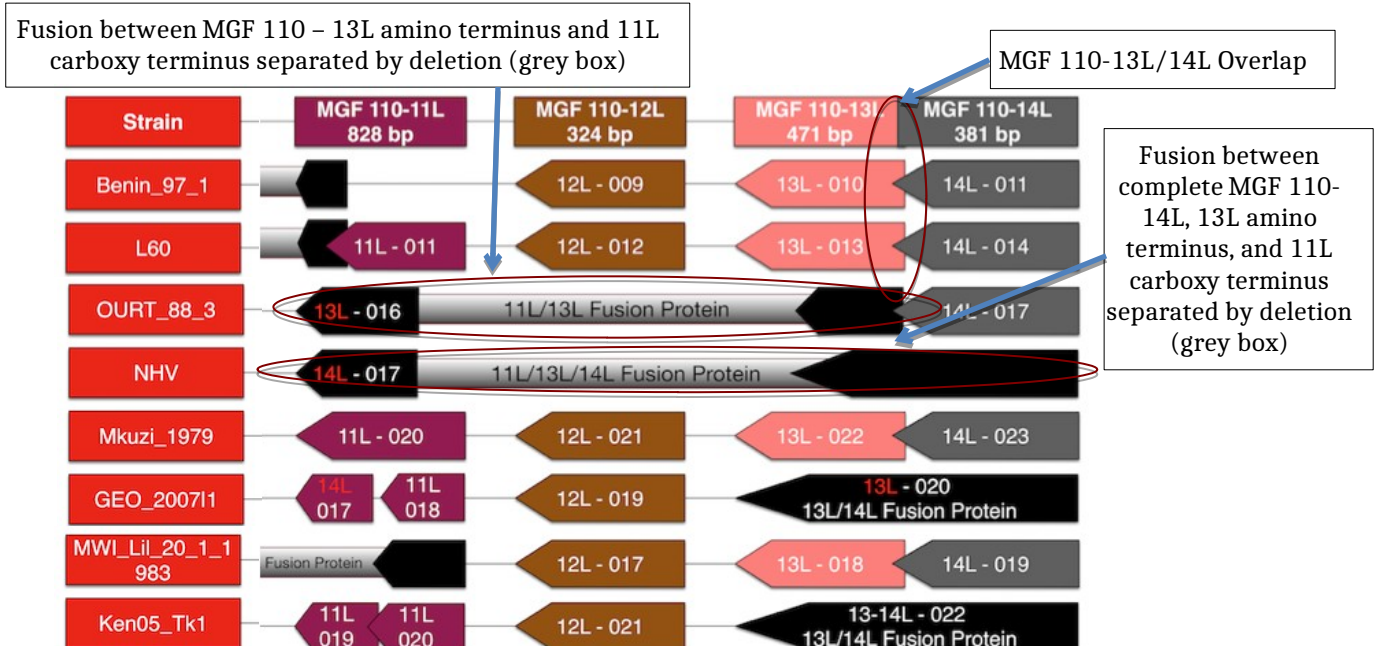


Special Note Gene Features:

The 5' end of the ORF for the top two 11L genes overlaps with the 3' end of the ORF for L60-011 in different reading frames.

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

For the bottom most 11L ortholog showed in the above diagram is showed to have several large in-frame deletions in the gene when compared to the aligned genomes.



MGF 110 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

MGF 110 Figure Legend

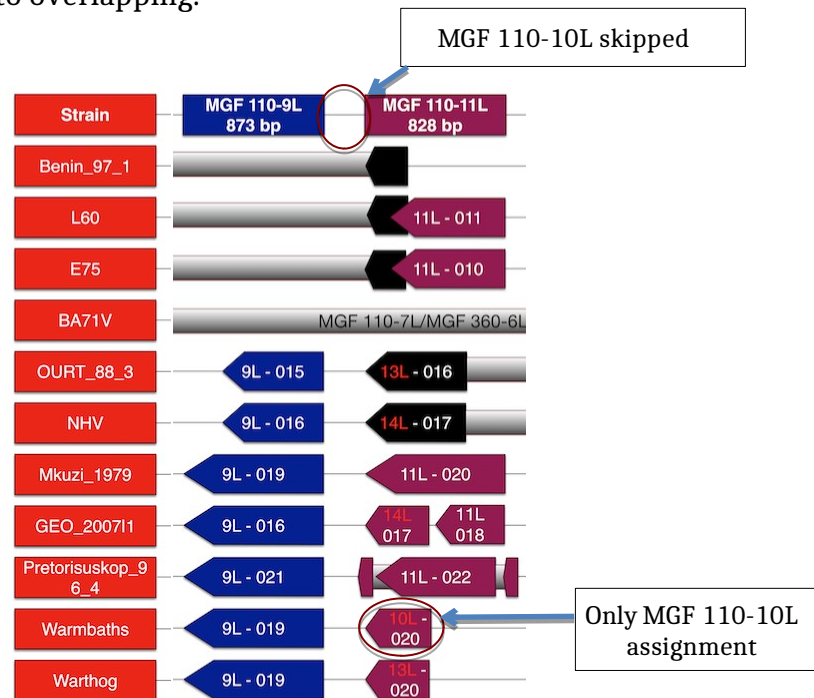
boxes to a given ortholog group is representative of how much of the ortholog is fused and which region.

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

The absence of a deletion box in a gene fusion indicates that the deletion that connects the two ORFs is only a few base pairs that would be too small to resolve on this diagram.

MGF 110-13L/14L overlap:

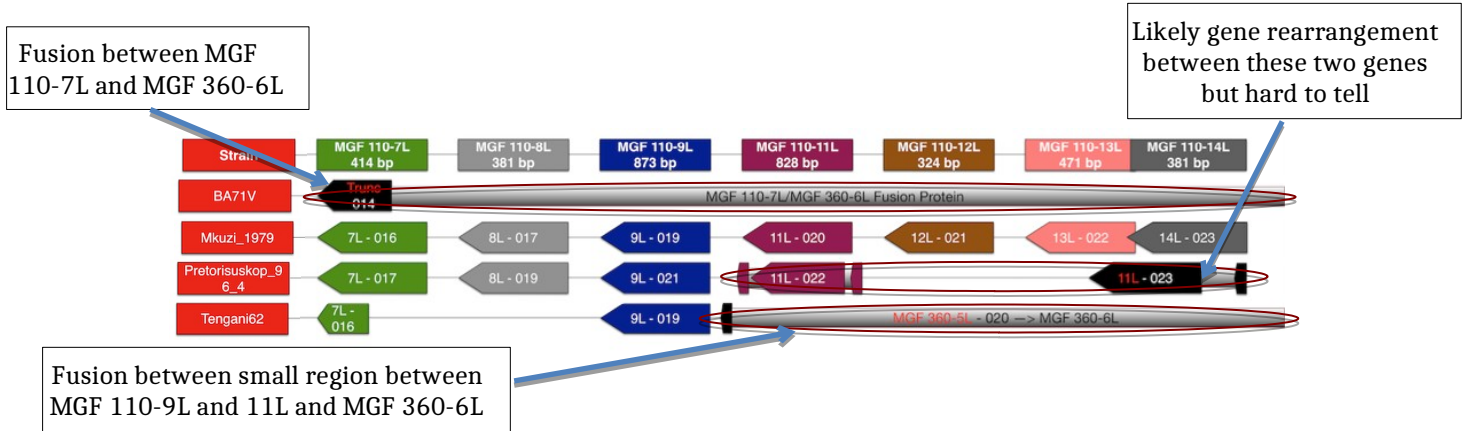
The 3' end of the 14L ORF overlaps with the 5' end of the 13L ORF. This is represented by no space between the ortholog headings and all genes under these headings to be overlapping.



MGF 110-10L Missing:

The MGF 110-10L ortholog is annotated in only the Warmbaths strain which appears to be a truncated 11L ortholog. This gene is quite short due to an amino terminus truncation. It is possible that when comparing this gene, it appeared to form its own ortholog, causing any subsequent assignments to be a new ortholog.

MGF 110 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 amino terminus of this fusion is not shown since it is outside the scope of this diagram.

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

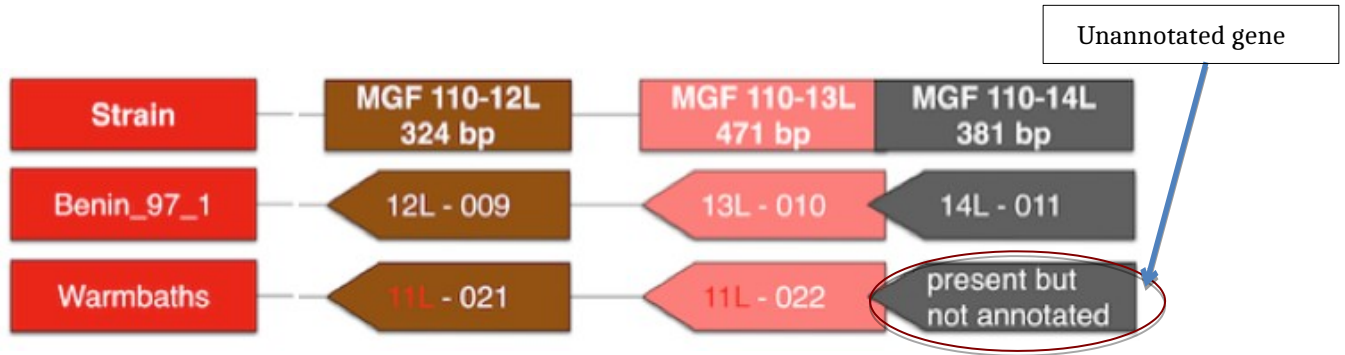
MGF 360-5L → MGF 360-6L:

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. Due to the small carboxy terminus fragment, the gene was labelled on the connecting deletion box. The amino terminus of this fusion is not shown since it is outside of the scope of this diagram.

Pretorisuskop 11L and 13/14L alignment:

These two Pretorisuskop genes do not align very well in the region they are positioned in the above diagram. In a dotplot analysis, it appears that 022 aligns best with the 13L/14L orthologs and 023 aligns best with the 11L ortholog. This supports a potential gene rearrangement in this genome but due to the high similarity between the 11L and 13L/14L orthologs it is hard to determine.

MGF 110 Figure Legend



Warmbaths MGF 110-14L ortholog:

The Warmbaths strain has an ORF that that aligns with the 14L orthologs but was not annotated.