Table S1 Classification Criteria

Absent Annotations

Best e-value and % identity from annotated gene alignment
Alignment Subject from a different replicon
Alignment >=80% coverage for query and subject

Genomic Artifacts

Best e-value and % identity from align. to ORF that overlaps a real gene
Alignment Subject from a different replicon
Alignment >=80% coverage for query and subject

Potentially Missing

Best e-value and % identity from intergenic ORF alignment
Alignment Subject from a different taxonomic family (defined by NCBI)
Alignment >=80% coverage for query and subject
20% margin for average coverage