Supplementary Data 1: Operon Structure Similarity Calculation

Consider the reference operon, O_{ref}: $\{ABCDE\}$, where A, B, C, D and E are genes.

Consider also the following detected operons in a target *GenomeAssembly*:

Operon, O_j, from *NucleotideRecord*, $f_{i=1}$, from *GenomeAssembly*, $S_{k=1}$:
$$O_{j=1}:\{ABC\} \ and \ O_{j=2}:\{DE\}$$
Operon, O_j, from *NucleotideRecord*, $f_{i=2}$, from *GenomeAssembly*, $S_{k=1}$:
$$O_{j=1}:\{DE\}$$

The gene-pairs, P_{ref}, in the reference operon O_{ref} are:
$$P_{ref}:[AB, BC, CD, DE]$$

The observed gene-pairs, P_i, for the nucleotide records *NucleotideRecord*, i, mapping to genome assembly *GenomeAssembly*, $S_{k=1}$ are:
$$P_{i=1}:[AB, BC, DE]$$
$$P_{i=2}:[DE]$$

The maximum number of times each reference pair could *possibly* occur based off the number of occurrences of each gene, M, in *GenomeAssembly*, $S_{k=1}$ is:
$$M_{[AB]} \rightarrow 1$$
$$M_{[BC]} \rightarrow 1$$
$$M_{[CD]} \rightarrow 1$$
$$M_{[DE]} \rightarrow 2$$

The *observed* number of occurrences, T, for each gene-pair P, in *GenomeAssembly*, $S_{k=1}$:
$$T_{[AB]} \rightarrow 1$$
$$T_{[BC]} \rightarrow 1$$
$$T_{[CD]} \rightarrow 0$$
$$T_{[DE]} \rightarrow 2$$

Therefore, the relative number of occurrences, R, for each pair is T/M:
$$R_{[AB]} = \frac{T_{[AB]}}{M_{[AB]}} = 1$$
$$R_{[BC]} = \frac{T_{[BC]}}{M_{[BC]}} = 1$$

$$R_{[CD]} = \frac{T_{[CD]}}{M_{[CD]}} = 0$$

$$R_{[DE]} = \frac{T_{[DE]}}{M_{[DE]}} = 1$$

The total number of relative occurrences is then:

$$\sum R = R_{[AB]} + R_{[BC]} + R_{[CD]} + R_{[DE]} = 3$$

The proportion of relative occurrences to the number of gene-pairs P_{ref} in the reference operon is the structural conservation score Y:

$$Y = \frac{\sum R}{P_{ref}} = \frac{3}{4} = 0.75$$