**Model Expansion Setup**

For a given rate matrix Q, with qij representing instantaneous transition rate from state i to state j, the gene conversion expansion yields a rate matrix R, with rii’, jj’instantaneous transition rate from state ii’ to jj’.



where diff() function returns the total number of different nucleotides at corresponding position in the state. For example, for nucleotide models, diff() returns 1 if the two nucleotides are different and 0 otherwise. For codon models, diff() counts for number of different nucleotides at the same position in two codon states, diff(ACT, AGC) = 2.

**Data Preparation**

Sequence data:

Sequence data of S. castellii and S. kluyveri were extracted from Fungal Orthogroups Repository’s database (<http://www.broadinstitute.org/regev/orthogroups/>).

Sequence data of S. paradoxus, S. mikatae, S. bayanus, S. kudriavzevii were extracted from the paper “The awesome power of Yeast evolutionary genetics”.

Orthology mapping:

Six species (S. cerevisiae, S. paradoxus, S. mikatae, S. bayanus, S. castellii, S. kluyveri) orthology mapping used Fungal Orthogroups Repository from Broad Institute(<http://www.broadinstitute.org/regev/orthogroups/>). The data in the debase was described in their paper “A natural history and evolutionary principles of gene duplication in fungi”.

Five species (S. cerevisiae, S. kudriavzevii, S. bayanus, S. paradoxus, S. mikatae) orthology mapping used the orthology determined by Scannell et al. in their paper “The awesome power of Yeast evolutionary genetics”.

Only gene pairs extent in all six ingroup species with only one gene extent in outgroup species S. kluyveri were kept.

Filters:

A multiple sequence alignment was performed at codon level using MAFFT for all pairs of genes. Pairs with gaps more than 10% in length were removed.

Pairs with less than 3.0 sum of branch length according to conventional MG94 model were removed.

Phylogenetic Tree:

