Non-Backtracking Pathways

Given k possible permutations for a genome of any fixed length, how many pathways are possible that traverse through these permutations from one given point to another and do not backtrack (i.e. that never go through the same point twice)?

To answer this question, we can assume a starting genome point g_a, and an endpoint g_b. The first move could be to any of the remaining genomes, except g_a and g_b. There are therefore $k-2$ possible locations to choose the first move.

The second move could be any of the remaining, non-chosen points, so there would be $k-3$ locations. Multiplying by the first move, we have $(k-2)(k-3)$ pathways available. If we follow this pattern, we can easily see that there will be $(k-2)!$ pathways with maximum length.

For shorter pathways, we use the same technique, but stop sooner. For a four-step path, we would have:

$$n_p(k,4) = \frac{(k-2)!}{(k-5)!}$$

pathways available. Where $n_p(k,4)$ is the number of pathways for a genome with k possible permutations and a four-step path.

The total number of pathways possible will thus be the sum of each of the lengths possible, thus:

$$\boxed{n_p(k) = \sum_0^{k-2} \frac{(k-2)!}{i!}}$$

Where $n_p(k)$ is the total number of pathways for a genome with k possible permutations.

To simplify this for an upper boundary, we can use the fact that:

$$n_p(k) = (k-2)! \cdot \sum_0^{k-2} \frac{1}{i!}$$

along with the additional fact that:

$$\sum_{i=0}^{\infty} \frac{1}{i!} = e$$

and therefore,

$$\sum_{i=0}^{k-2} \frac{1}{i!} < e$$

To get the boundary equation of:

$$n_p(k) = (k-2)! \cdot \sum_{0}^{k-2} \frac{1}{i!} < e(k-2)!$$

$$\boxed{n_p(k) < e(k-2)!}$$

And thus, using Gosper's Approximation of Stirling's Approximation, we have:

$$\boxed{n_p(k) < \frac{(k-2)^{(k-2)}}{e^{(k-1)}} \sqrt{\left(2k - \frac{11}{3}\right)\pi}}$$

This establishes the upper boundary for the number of pathways, and improves for k large.