E Appendix: Analytical Results for Higher Moments

Higher moments of the distribution, defined as $H_n(t) = \sum_m m^n F(m,t)$, for $n \geq 2$ in the extended model statisfy the following differential equation:

$$G(t)\frac{\partial H_n}{\partial t} = RG(t) + \sum_{m=1}^{\infty} F(m)\left[m(m+1)^n - (1+Q)m^{n+1} + Q(m-1)^n m\right] \quad (44)$$

In particular, the equations for the second and third moment are:

$$G(t)\frac{\partial H_2}{\partial t} = RG(t) + 2(1-Q)H_2(t) + (1+Q)G(t) \quad (45)$$

$$G(t)\frac{\partial H_3}{\partial t} = RG(t) + 3(1-Q)H_3(t) + 3(1+Q)3H_2(t) + (1-Q)G(t) \quad (46)$$

Higher moments depend on all lower moments except for the zeroth moment, the expected number of folds $F(t)$ This is fortuitous, since equation 11 for $F(t)$ could not be solved analytically due to its explicit dependence on the population of smallest folds: $F(1,t)$.

The solution for the second moment is given by:

$$H_2(t) = \begin{cases} N_0 \exp\left(\frac{2(1-Q)}{1+R-Q}u(t)\right) + N_0\frac{1+R+Q}{R-1+Q}\left[\exp u(t) - \exp\left(\frac{2(1-Q)}{1+R-Q}u(t)\right)\right] & R \neq 1-Q \\ N_0 \exp\left(u(t)\right)\left[1 + \frac{u(t)}{2R}\right] & R = 1-Q \end{cases} \quad (47)$$

where the variable $u(t)$ is related to the expected number of genes:

$$u(t) = \log\left[1 + \frac{(R+1-Q)t}{N_0}\right] \quad (48)$$

This result will be important in fitting actual genomic data to the models.