D Appendix: Solution to the Extended Model When $0 < Q < 1$ and $R = 0$

As one done in the solution for the minimal model, define $\phi(t)$:

$$\phi(t) = 1 + \frac{(1-Q)t}{N_0}, \tag{36}$$

and keep the association: $u = \log \phi(t)$. In terms of the time-like variable u, the fundamental evolution equations (10) now are:

$$\begin{aligned}(1-Q)\frac{\partial F(m,u)}{\partial u} &= (m-1)F(m-1,t) - (1+Q)mF(m,t) + Q(m+1)F(m+1,u) \quad (m>1)\\ (1-Q)\frac{\partial F(1,u)}{\partial u} &= -(1+Q)F(1,u) + QF(2,t)\end{aligned} \tag{37}$$

Substituting the ansatz: $F(m,u) = f(u)g^{m-1}(u)$ into the equation for $m > 1$ leads to the following relation:

$$(1-Q)\left[\frac{\partial \log f}{\partial u}g + (m-1)\frac{\partial g}{\partial u}\right] = (m-1) + (1+Q)mg + Q(m+1)g^2 \tag{38}$$

Since neither $g(u)$ nor $f(u)$ depend on m, this identity can only be satisfied if:

$$\begin{aligned}(1-Q)\frac{\partial g}{\partial u} &= 1 - (1+Q)g + Qg^2\\ (1-Q)\frac{\partial \log f}{\partial u} &= -(1+Q) + 2Qg\end{aligned} \tag{39}$$

These equations can be solved by integration, together with the restriction that $f(t=0) = 1$ and $g(t=0) = 0$. It is easy to verify that the ansatz also works when $m = 1$.

$$\begin{aligned}F(m,t) &= N_0 f(t) g^{m-1}(t)\\ f(t) &= \phi^{-1}\left[\frac{1-Q}{1-Q\phi^{-1}}\right]^2 = \phi\left[\frac{N_0}{N_0+t}\right]^2\\ g(t) &= \frac{1-\phi^{-1}}{1-Q\phi^{-1}} = \frac{t}{N_0+t}\\ \phi(t) &= 1 + \frac{(1-Q)t}{N_0}\end{aligned} \tag{40}$$

In fact, it is easy to solve for $F(t)$ in this special case:

$$\frac{\partial F(t)}{\partial t} = -Q\frac{F(1,t)}{G(t)} = -Q\frac{f(t)}{\phi(t)} \tag{41}$$

which can be integrated directly:

$$F(t) = N_0\frac{N_0 + (1-Q)t}{N_0+t} \tag{42}$$

The large-time asymptotic limit for $F(t)$ is $(1-Q)N_0$ folds, which reflects the fact that some of the initial N_0 folds will ultimately be lost due to gene deletion.

Equation (42) leads to a simple relation between the number of folds and the number of genes:

$$F(t) = \frac{G(t)}{1 + t/N_0} \tag{43}$$

Although $F(t)$ and $G(t)$ both depend on Q, their ratio does not.

The solution (40) we have derived for $0 < Q < 1$ is also the solution for $Q = 1$, which means that gene deletion and duplication occur at the same rate. Equations (42) and (43) for the total number of folds $F(t)$ are still valid for $Q = 1$, but now the expected number of genes is constant: $G(t) = N_0$. Although we will not do so here, analytic solutions can be derived when deletion dominates duplication so the genome shrinks in size.