Supporting Information

Callahan et al. 10.1073/pnas.0902364106

SI Text

Parameterizing Dynamics of Collinearity

Dynamics of collinearity distribution described by Eq. **1** involves a conditional probability $\Gamma(y|y',y'')$ that parents with collinearities y',y'' generate an offspring with y. This "transition probability" must allow an equilibrium distribution $\rho(y)$ corresponding to random genomic order. Hence:

$$\rho(y) = \int dy' \int dy'' \rho(y')\rho(y'')\Gamma(y|y',y''). \quad \textbf{[S1]}$$

For a sufficiently long genome, $\rho(y)$ is well approximated by a Gaussian:

$$\rho(y) = \frac{e^{-\frac{\kappa y^2}{2}}}{\sqrt{2\pi\kappa^{-1}}}. \quad \textbf{[S2]}$$

A reasonable phenomenological model for Γ must generate relaxation to a Gaussian equilibrium distribution while retaining some correlation between the offspring and the parents. Both effects can be captured by the following simple but nontrivial form:

$$\Gamma(y|y',y'') = \frac{e^{-\frac{\kappa[y-\alpha(y'+y'')]^2}{2(1-2\alpha^2)}}}{\sqrt{2\pi\kappa^{-1}(1-2\alpha^2)}}, \quad \textbf{[S3]}$$

which provides a whole family of models with the free parameter $\alpha < 1/2$ controlling the degree of parent–offspring correlation. For $\alpha = 0$, equilibrium is attained in a single step, $\Gamma(y|y',y'') = \rho(y)$, which is the case discussed in the main text. Qualitatively similar results—e.g. existence of the fixed point y^*—are obtained for any value of α.

Collinearity and the Success of Recombinant Offspring. The contribution of collinearity to the success of recombination in creating high-fitness offspring is quantified in our paper by use of the function $q(y)$. As presented, this function depends only on the recombinant's ("child's") collinearity. However, the likelihood of a recombination producing a high-fitness offspring is in fact a more complex mapping which depends in important ways on the parental genotypes. The function $q(y)$ implicitly is a population average over parental genotypes conditioned on (*i*) offspring collinearity and (*ii*) offspring encoding a novel L^* pathway. To be explicit, we define the probability of a recombinant offspring to have collinearity y and encode a novel L^* pathway

$$\chi_*(y) = N^{-2} \sum_{g_1} \sum_{g_2} n(g_1)n(g_2)K_*(y|g_1,g_2), \quad \textbf{[S4]}$$

where $n(g)$ is the number of genotypes g in the population. The probability of a recombinant offspring to have collinearity y whether or not it encodes an L^* pathway is

$$\chi(y) = N^{-2} \sum_{g_1} \sum_{g_2} n(g_1)n(g_2)K(y|g_1,g_2) \quad \textbf{[S5]}$$

$$= \int dy' \int dy'' \xi(y')\xi(y'')\Gamma(y|y',y''). \quad \textbf{[S6]}$$

We then have $q(y) = \chi_*(y)/\chi(y)$.

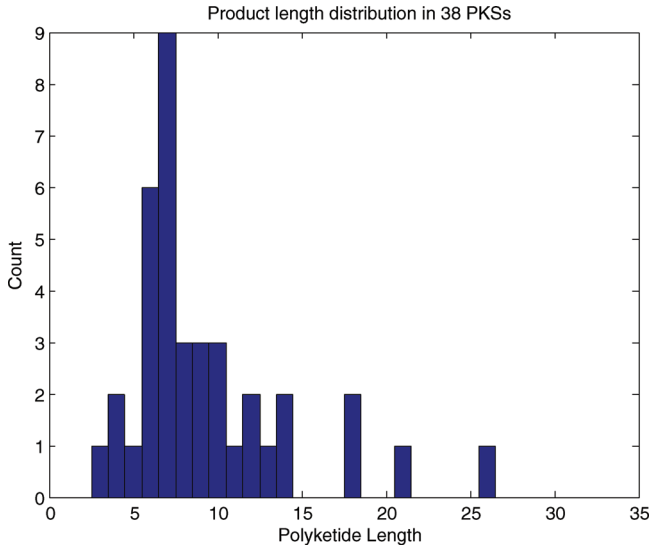

Fig. S1. Distribution of lengths of modular PKS pathways in bacteria. A histogram of pathway length for the set of 38 modular PKS pathways exhibits a peak length 7, suggesting a threshold level of complexity. Data from Thattai M, Burak Y, Shraiman B (2007) The origins of specificity in polyketide synthase protein interactions. *PLoS Comput Biol* 3:e186. (Dataset S1 found at doi:10.1371/journal.pcbi.0030186.sd001).