1 Mike Questions

1. What is the difference between $\vec{c_i}$ and c_i? Should we just use $\vec{c}$ and c_i?
2. Why does a $\vec{c_i}$ term appear in the denominator? is ω_j a function of $\vec{c_i}$ in this notation?
3. $\vec{c}$ denotes the entire genome, correct?
4. If $\vec{c}$ does denote the whole genome, should ϕ have some sort of subscript relating it to the current gene in the genome?
5. Is it alright if I use ν to denote the number of synonyms to c_i?
6. Where do the mutation terms go? I think I'll include them for completeness. Just turn $\exp\left[-\sum_j \beta_j \omega_j(\vec{c}_\iota)\phi\right]$ into $\exp\left[-\sum_j (\mu_j - \beta_j \omega_j(\vec{c}_\iota))\phi\right]$?
7. $\mathbb{C}$ only contains synonymous sequences, correct?

2 Notation

$\vec{c}$ is the vector of codons (of length n, the length of the ??genome/gene??)

??$\vec{c}_\iota / c_\iota$?? is the i^{th} codon of $\vec{c}$

$\vec{\theta}$ is our given conditions, including β, ϕ, ω, and the effective population size

$\sum_j$ is shorthand for $\sum_{j=1}^{n}$, the sum over the whole ??genome/gene??

β_j is the cost of codon translation. In the NSE model, this is constant. In the ROC model, this is the main variable.

ω_j is the odds of a nonsense error at codon j, $\frac{p_j}{1-p_j}$

ϕ is the expression level of the gene

$\mathbb{C}$ is the set of all possible ??synonymous?? sequences

ν is the number of synonymous codons to $\vec{c_i}$

3 Math

"The probability of seeing codon $\vec{c_i}$ at position i (given our conditions) is approximately the fitness of that sequence divided by the sum of the fitness of all possible synonymous sequences"

$$P(\vec{c}_\iota|\vec{\theta}) \approx \frac{\exp\left[-\sum_j \beta_j \omega_j(\vec{c}_\iota)\phi\right]}{\sum_{K\in\mathbb{C}} \exp\left[-\sum_j \beta_j \omega_j(\vec{c}_\iota)\phi\right]}$$

Rewrite the sum as the sum across the synonyms. Added notation:

ν is the number of synonyms for the given codon i

$$\approx \frac{\sum_{K \in \mathbb{C} | c_k = c_\iota} \exp\left[-\sum_j \beta_j \omega_j(\vec{c}_\iota)\phi\right]}{\sum_{\ell=1}^{\nu} \sum_{K \in \mathbb{C} | c_k = c_\ell} \exp\left[-\sum_j \beta_j \omega_j(\vec{c}_\iota)\phi\right]}$$

Pull out the term for the current codon chain, where $c_j = c_i$

$$\approx \frac{\left[\sum_{K \in \mathbb{C}} \exp\left[-\sum_{j \neq k} \beta_j \omega_j(\vec{c}_\iota)\phi\right]\right] \left(\exp\left[-\beta_k \omega_k(\vec{c}_\iota)\phi\right]\right)}{\sum_{\ell=1}^{\nu} \sum_{K \in \mathbb{C} | c_k = c_\ell} \exp\left[-\sum_j \beta_j \omega_j(\vec{c}_\iota)\phi\right]}$$

Which cancels to

$$\approx \frac{\exp\left[-\beta_k \omega_k(\vec{c}_\iota)\phi\right]}{\sum_{\ell=1}^{\nu} \exp\left[-\sum_j \beta_j \omega_j(\vec{c}_\iota)\phi\right]}$$