Midterm Exam

Zachary Foster

May 2, 2014

1

Assuming that the 8bp motif can be located at any position in the intergenic spacer and can be located on either strand, there is $1000 \times 2 \times (100 - 8 + 1) = 186000$ positions at which an 8bp motif can occur. The chances of a random 8bp motif being any one permutation of base pairs is $4^{-8} \approx 1.53 \times 10^{-5}$. For a particular permutation of motif locations, the chances of exactly n number of 8bp motifs being identical is...

$$\left(4^{-8}\right)^{n} \times \left(1 - 4^{-8}\right)^{186000-n}$$

and the number of possible permutations of locations that the motifs can appear in is...

$$\frac{186000!}{n!\,(186000-n)!}$$

Therefore, the chances that 5 or more identical 8bp motifs is...

$$1 - \sum_{n=0}^{4} \left[\left(\frac{186000!}{n!\,(186000-n)!} \right) \times \left(4^{-8}\right)^{n} \times \left(1 - 4^{-8}\right)^{186000-n} \right] = 0.1583097$$

Below is the R code I used to do the calculation:

```
1 - sum(choose(186000, 0:4) * ((4^(-8))^(0:4)) * (1 - 4^(-8))^(186000 - 0:4))
```

```
## [1] 0.1583
```