

Duh³: Doubley Dynamic DNA Programming

Cynthia Stallard
Anthony Macula
SUNY Geneseo

1st Things 1st What are we doing?

- Using the Maple programming language
 - Generating DNA strands
 - Comparing strands and scoring them
 - Creating a dynamic threshold
 - Accepting only strands that satisfy the threshold when compared to all other previously accepted strands and their complements
-

Why are we doing this?

- ❑ The ultimate goal of this project is to have a computer program capable of generating a DNA library
 - ❑ The strands we create should have minimal predicted interaction among other strands in the library
 - ❑ This is to ensure that in a lab a probe dropped into solution with all of the strands created will bind with only the specific strand it was intended to pull out.
-

Some Background Info

- ❑ DNA is composed of four bases, A,C,G, and T. Optimally they bind with other bases to form Watson Crick base pairs.
 - ❑ A binds with T and C binds with G.
 - ❑ A-T and C-G bonds have different strengths. A-T has two hydrogen bonds while C-G has three.
 - ❑ If two DNA strands have two complementary base pairs in a row, that pair of matches is called a stacked pair.
-

More Background Info

- Just as A-T and C-G bonds have different strengths, so do different stacked pairs.

- Stacked Pair Scores:

[A,A]:= 1	[A,C]:= 1.44	[A,G]:= 1.28	[A,T]:= 0.88
[C,A]:= 1.45	[C,C]:= 1.84	[C,G]:= 2.17	[C,T]:= 1.28
[G,A]:= 1.3	[G,C]:= 2.24	[G,G]:= 1.84	[G,T]:= 1.44
[T,A]:= 0.58	[T,C]:= 1.3	[T,G]:= 1.45	[T,T]:= 1

The Process

- Step 1: Generating a strand
 - Step 2: Weighted Stacked Pair Scores
 - Step 3: Creating a Threshold
 - Step 4: Rejecting strands
 - Step 5: Accepting Strands
-

Step 1: Generating a strand

- Using a function in Maple, a strand of length n is generated using input probabilities of selecting each of the bases, A,C,T,G, at each position in the strand
 - For our purposes the probability of each base is set at 0.25. This input can be easily changed by a user.
-

Example

```
with(stats);  
n:=6; N:=3; ap:=.25; cp:=.25; gp:=.25; tp:=.25;  
  
for i from 1 to N do  
(subs(1=A,2=C,3=G,4=T,[stats[random,empirical[ap,cp,gp,  
tp]](n))));  
end do;  
  
[G, A, T, A, C, C]  
[T, C, C, T, G, A]  
[T, G, T, T, T, G]
```

Example:

Strand 1 = 5'-AATCCAA-3'
Strand 2 = 5'-GTCACAT-3'

Scoring Matrix=

	GT	TC	CA	AC	CA	AT
AA	0	0	0	0	0	0
AT	0	0	0	0	0	0.88
TC	0	1.3	1.3	1.3	1.3	1.3
CC	0	1.3	1.3	1.3	1.3	1.3
CA	0	1.3	2.75	2.75	2.75	2.75
AA	0	1.3	2.75	2.75	2.75	2.75

Weighted stacked pair score of the best alignment.

Step 3: Creating a Threshold

- First we made the decision to make the threshold dynamic, rather than constant.
 - We did this because as strands get longer their score gets larger since there will be more stacked pairs common to the strands. For this reason we want the threshold to be smaller when the strands are short but larger as the strands get longer. The Partition Function shows us why.
-

Step 3: Creating a Threshold

□ The Partition Function

Probability of a strand binding with its complement

$$= \frac{e^{WC}}{e^{WC} + (n-1)e^{CH}}$$

$$= \frac{1}{1 + (n-1)e^{(CH-WC)}}$$

So we need to control (WC-CH).

Step 3: Creating a Threshold

□ The WC is the sum of all weighted stacked pairs of a strand X. This score increases linearly with the length of the strand.

□ Example: X=5'-ACTG3'

Y=5'-GTACATGC-3'

$$WC(X)=[AC]+[CT]+[TG]=4.17$$

$$WC(Y)=[GT]+[TA]+[AC]+[CA]+[AT]+[TG]+[GC]=11.81$$

Step 3: Creating a Threshold

- Since the WC is linear, we can allow the CH to be linear in order to control (WC-CH).
 - This is why our threshold of CH will be the equation of a plane.
-

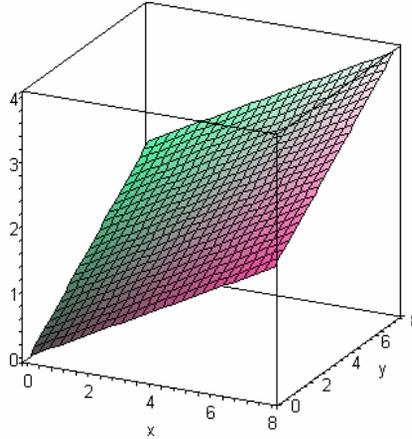
Step 3: Creating the Threshold

- The equation of the dynamic threshold will be a plane described by three points input by a user.
 - The input points are used to create a system of linear equations that can be solved by Maple to find the coefficients needed for an equation of a plane.
-

Example

Three points: (4,4,2)
(4,8,3)
(8,4,3)

Threshold = $\frac{1}{4}(x+y)$



Step 4: Rejecting strands

- Now that we have a threshold equation and the scores of the generated strand compared to all other accepted strands, we can check to see if each scoring matrix satisfies the threshold at each entry after a set starting point.
 - Since we want strands that are not likely to bind with each other, we will reject strands that are above the threshold. If we wanted strands to be highly interactive, we would reject strands below the threshold.
-

Step 5: Accepting Strands

- If a strand satisfies the requirements of the threshold with all other strands and with its own reverse complement, we union the strand and its reverse complement to the set of accepted strands.
-

Finally

- Once we either accept or reject a strand, we start from the beginning and generate a new strand.
 - The process continues until we have the desired number of strands.
-

Example

```
□ > GenerateStrands2(.25,.25,.25,.25,4,10,4,  
                    [4,4,2],[4,8,3],[8,4,3]);
```

So, ap=.25, cp=.25, gp=.25, tp=.25

n=4, N=10, start=4, A=[4,4,2], B=[4,8,3], C=[8,4,3]

{[C, C, C, C], [G, G, G, G], [G, G, A, G], [C, T, C, C], [T,
C, G, T], [G, G, C, C], [A, C, G, A], [T, A, A, T], [A, T, T,
A], [T, A, A, C], [G, T, T, A]}

Acknowledgements

Special thanks to Tony Macula, Jackie
Dresch, and Christy Callear