BIT 161B : Experiment 1
11 April 2018

Bioinformatics approach to designing degenerate primers

Aditya Nirgun 999723427

Abstract

This experiment is an investigation into the processes of constructing a degenerate primer for the purpose of identifying closely-related genes during a gene amplification treatment. Two kinds of primers were constructed: those that will identify members of a family and those that will identify specific genes. The methodology involved is reflective of typical bioinformatics where computer algorithms are deployed to sort through initial data that is farmed from the scientific community at large. We were able to produce the intended primer designs although their accuracy is in question due to large degeneracy number which could lead to many unaccounted possibilities. The biotechnology concepts of degeneracy, gene families, and conserved regions were heavily applied in this experiment and demanded full comprehension.

Introduction

Degeneracy in primers is used to describe the expanse of variability between any given genetic sequences that produce the same amino acid sequence after translation. This is possible because of the principle regarding redundancy of genetic sequence that allows several codon combinations that vary in the third position nucleotide to encode the same amino acids. This multiplicity of codons forms the basis of the relationship of between members in a gene family as it relates to this laboratory experiment.

On the 11th of April, we used a bioinformatics approach to identify members of a gene family and design PCR primers that can be used to isolate other members of the same family. The primers we designed will be relevant to our processes in Experiment 5. Arabidopsis Thaliana was the specimen of interest in this experiment and the gene family of interest was the family of six NHX genes that encode antiporters of which expression is unique to each gene. The databases from which the amino acid sequences for the NHX genes are sourced include Genbank and UniProtKB/SwissProt, each hosted on the official NCBI website.

Objectives

The procedure guided us from one step to another until we had taken the amino acid sequences of the six NHX genes gleaned from previous studies in Arabidopsis and, from that initial data, produced nucleotide sequences suitable for designing oligonucleotide primers that could identify potentially undiscovered members of the gene family or primers that could be specially designed to identify individual known genes within the family.

Methods

Database results on the NBCI presented two protein sequence sets for each of the six members of the gene family. One set came from the Genbank database and one came from UniProtKB/SwissProt. For each gene, the longer of the two options was chosen in the interest of having more initial data prior to the alignment build. The selected FASTA sequence entries can be seen in Figure 1.

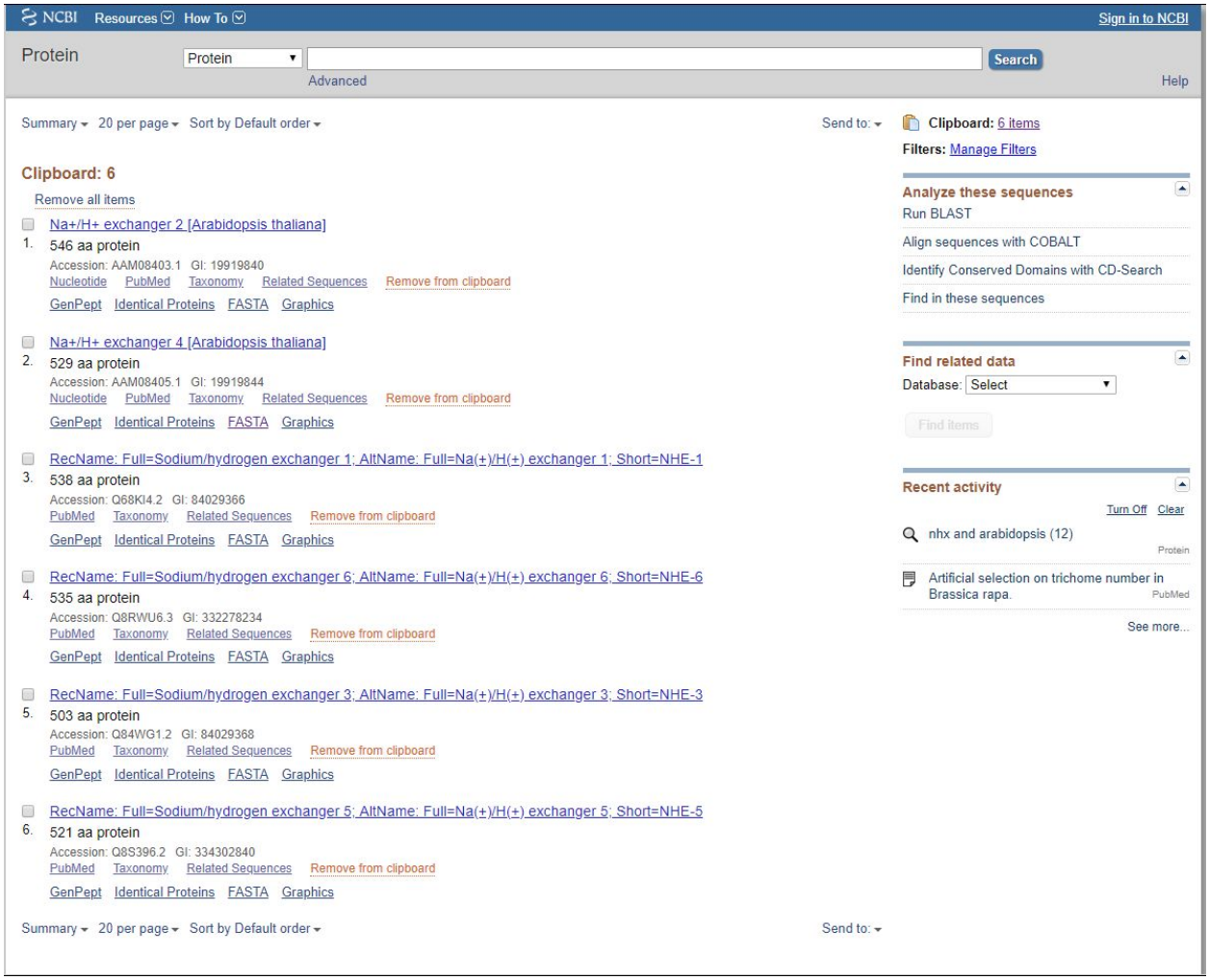

Figure 1. Six selected protein sequences of genes in the NHX gene family

Results

From the amino acid sequences of the six NHX genes gleaned from previous studies in Arabidopsis, we aligned pairs of sequences that revealed conserved domains among the gene family. Using the identified conserved regions, we produced a consensus sequence that will allow us to design oligonucleotide primers to be used in PCR to identify members within a gene family.

The output of the program ClustalW is displayed below.

```
AAM08403.1                 ---MTMFASLTSKMLSVSTSDHASVVSLNLFVALLCACIVIGHLLE--EN
sp|Q68KI4.2|NHX1_ARATH     -----MLDSLVSKLPSLSTSDHASVVALNLFVALLCACIVLGHLLE--EN
sp|Q84WG1.2|NHX3_ARATH      --MVIGLSTMLEKTEALFASDHASVVSMNLFVALLCACIVLGHLLE--ET
AAM08405.1                  -----MSIGLTECVTNKLAAEHPQVIPISVFIAILCLCLVIGHLLE--EN
sp|Q8RWU6.3|NHX6_ARATH      MSSELQISPAIHDPQGQEKQQQAAGVGILLQIMMLVLSFVLGHVLRRHKF
sp|Q8S396.2|NHX5_ARATH      -MEEVMISPVEHDPQGQVKQQQAAGVGILLQIMMLVLSFVLGHVLRRHRF
                                                  ::.  : : : : :*  .:*:**:*.  .

AAM08403.1                 RWMNESITALLIGLGTGVVILLISRGKNS-HLLVFSEDLFFIYLLPPIIF
```

```
sp|Q68KI4.2|NHX1_ARATH       RWMNESITALLIGLGTGVTILLISKGKSS-HLLVFSEDLFFIYLLPPIIF
sp|Q84WG1.2|NHX3_ARATH       RWMNESITALIIGSCTGIVILLISGGKSS-RILVFSEDLFFIYLLPPIIF
AAM08405.1                   RWVNESITAILVGAASGTVILLISKGKSS-HILVFDEELFFIYLLPPIIF
sp|Q8RWU6.3|NHX6_ARATH       YYLPEASASLLIGLIVGGLANISNTETSIRTWFNFHDEFFFLFLLPPIIF
sp|Q8S396.2|NHX5_ARATH       HYLPEASGSLLIGLIVGILANISDTETSIRTWFNFHEEFFFLFLLPPIIF
                          :: *:  ::::*   *    : .  ..      : * :::**::*******

AAM08403.1                NAGFQVKKKQFFRNFVTIMAFGAIGTVVSCTIISLGAIQFFKKLDIGTFD
sp|Q68KI4.2|NHX1_ARATH       NAGFQVKKKQFFRNFVTIMLFGAVGTIISCTIISLGVTQFFKKLDIGTFD
sp|Q84WG1.2|NHX3_ARATH       NAGFQVKKKQFFRNFMTIMLFGAIGTLISFVIISFGAKHLFEKMNIGDLT
AAM08405.1                   NAGFQVKKKKFFHNFLTIMSFGVIGVFISTVIISFGTWWLFPKLGFKGLS
sp|Q8RWU6.3|NHX6_ARATH       QSGFSLQPKPFFSNFGAIVTFSVLGTFVASMVTGLLVYLGGVMFLMYRLP
sp|Q8S396.2|NHX5_ARATH       QSGFSLQPKPFFSNFGAIVTFAIIGTFVASVVTGGLVYLGGSMYLMYKLP
                             ::**.:: * ** ** :*: *. :*..::  : .  .        :  :

AAM08403.1                      LGDFLAIGAIFAATDSVCTLQVLNQD-ETPLLYSLVFGEGVVNDATSVVL
sp|Q68KI4.2|NHX1_ARATH          LGDYLAIGAIFAATDSVCTLQVLNQD-ETPLLYSLVFGEGVVNDATSVVV
sp|Q84WG1.2|NHX3_ARATH          IADYLAIGAIFSATDSVCTLQVLNQD-ETPLLYSLVFGEGVVNDATSVVL
AAM08405.1                      ARDYLAIGTIFSSTDTVCTLQILHQD-ETPLLYSLVFGEGVVNDATSVVL
sp|Q8RWU6.3|NHX6_ARATH          FVECLMFGSLISATDPVTVLSIFQELGSDVNLYALVFGESVLNDAMAISL
sp|Q8S396.2|NHX5_ARATH       FVECLMFGALISATDPVTVLSIFQDVGTDVNLYALVFGESVLNDAMAISL
                             : * :*:::::**.* .*.::::    **:*****.*:*** :: :

AAM08403.1                      FNAIQSFDLTHLNHEAAFQFLGNFFYLFLLSTGLGVATGLISAYVIKKLY
sp|Q68KI4.2|NHX1_ARATH          FNAIQSFDLTHLNHEAAFHLLGNFLYLFLLSTLLGAATGLISAYVIKKLY
sp|Q84WG1.2|NHX3_ARATH          FNAIQRFDLTNINSAIALEFAGNFFYLFILSTALGVAAGLLSAFVIKKLY
AAM08405.1                      FNAVQKIQFESLTGWTALQVFGNFLYLFSTSTLLGIGVGLITSFVLKTLY
sp|Q8RWU6.3|NHX6_ARATH          YRTMSLVRSHSS-GQNFFMVIVRFLETFVGSMSAGVGVGFTSALLFKYAG
sp|Q8S396.2|NHX5_ARATH          YRTMSLVNRQSSSGEHFFMVVIRFFETFAGSMSAGVGVGFTSALLFKYAG
                                :.::. .           : .  .*:  *  *   * ..*: :: ::*

AAM08403.1                      FGRHSTD-REVALMMLMAYLSYMLAELFALSGILTVFFCGIVMSHYTWHN
sp|Q68KI4.2|NHX1_ARATH          FGRHSTD-REVALMMLMAYLSYMLAELFDLSGILTVFFCGIVMSHYTWHN
sp|Q84WG1.2|NHX3_ARATH          IGRHSTD-REVALMMLLAYLSYMLAELFHLSSILTVFFCGIVMSHYTWHN
AAM08405.1                      FGRHSTT-RELAIMVLMAYLSYMLAELFSLSGILTVFFCGVLMSHYASYN
sp|Q8RWU6.3|NHX6_ARATH       LDVDNLQNLECCLFVLFPYFSYMLAEGLSLSGIVSILFTGIVMKHYTYSN
sp|Q8S396.2|NHX5_ARATH          LDTENLQNLECCLFVLFPYFSYMLAEGVGLSGIVSILFTGIVMKRYTFSN
                             :. ..  * .:::*:.*:****** . **.*::::* *::*.:*:  *

AAM08403.1                   VTESSRITTKHAFATLSFLAETFIFLYVGMDALDIEKWRFVSDSPGTSVA
sp|Q68KI4.2|NHX1_ARATH       VTESSRITTKHTFATLSFLAETFIFLYVGMDALDIDKWRSVSDTPGTSIA
sp|Q84WG1.2|NHX3_ARATH       VTDKSKVTTKHTFAAMSFLAEIFIFLYVGMDALDIEKWDVVRNSPGQSIG
AAM08405.1                   VTESSRITSRHVFAMLSFIAETFIFLYVGTDALDFTKWKTSSLSFGGTLG
sp|Q8RWU6.3|NHX6_ARATH       LSANSQRFVSAFFHLISSLAETFVFIYMGFDIAMEKHS-------WSHLG
sp|Q8S396.2|NHX5_ARATH       LSEASQSFVSSFFHLISSLAETFTFIYMGFDIAMEQHS-------WSHVG
                             ::  *:      *  :* :** * *:*:* *        :           :.

AAM08403.1                      VSSILMGLVMLGRAAFVFPLSFLSNLAKKHQ--SEKISIKQQVVIWWAGL
sp|Q68KI4.2|NHX1_ARATH          VSSILMGLVMVGRAAFVFPLSFLSNLAKKNQ--SEKINFNMQVVIWWSGL
sp|Q84WG1.2|NHX3_ARATH          VSSILLGLILLGRAAFVFPLSFLSNLTKSSP--DEKIDLKKQVTIWWAGL
AAM08405.1                      VSGVITALVLLGRAAFVFPLSVLTNFMNRHTERNESITFKHQVIIWWAGL
sp|Q8RWU6.3|NHX6_ARATH          FIFFSILFIVIARAANVFGCGYLVNLARPAH---RKIPMTHQKALWYSGL
sp|Q8S396.2|NHX5_ARATH          FILFSILFIGVARAVNVFGCAYLVNLFRQEN---QKIPMKHQKALWYSGL
                                .  .   :: :.**. **  . * *: .      ..* :. *  :*::**

AAM08403.1                      MRGAVSMALAYNKFTRSGHTELRGNAIMITSTITVCLFSTMVFGMLTKPL
sp|Q68KI4.2|NHX1_ARATH          MRGAVSMALAYNKFTRAGHTDVRGNAIMITSTITVCLFSTVVFGMLTKPL
sp|Q84WG1.2|NHX3_ARATH          MRGAVSMALAYNQFTTSGHTKVLGNAIMITSTITVVLFSTVVFGLLTKPL
AAM08405.1                      MRGAVSIALAFKQFTYSGVTLDPVNAAMVTNTTIVVLFTTLVFGFLTKPL
sp|Q8RWU6.3|NHX6_ARATH          RG---AMAFALALQSVHDLPEGHGQTIFTATTAIVVLTVLLIGGSTGTML
sp|Q8S396.2|NHX5_ARATH          RG---AMAFALALQSLHDLPEGHGQIIFTATTTIVVVTVLLIGGSTGKML
```

```
                                    ::*:*  :  . .    :  : :.*  * :   :: *   . *

AAM08403.1                      IRYLMP--HQKATTSTTSMLSDDSTPKSIHIPLLDGEQLDSFELPGSHQD
sp|Q68KI4.2|NHX1_ARATH          ISYLLP--HQNATT---SMLSDDNTPKSIHIPLLDQ---DSFIEPSGNHN
sp|Q84WG1.2|NHX3_ARATH          VKHLQPSSKQSSTTALQITLRSSFHDPILHEPLLSTQ-------GQSEYD
AAM08405.1                      VNYLLPQDASHNTGNRGKRTEPGSPKEDATLPLLSFD------ESASTNF
sp|Q8RWU6.3|NHX6_ARATH          EALEVVGDSHDTSLGDGFEVVNSRYMT------------------SYDDE
sp|Q8S396.2|NHX5_ARATH          EALEVVGDDLDDSMSEGFEESDHQYVPPPFS-----------IGASSDED
                                                :

AAM08403.1                      VPRPNSLRGFLMRPTRTVHYYWRQFDDAFMRPVFGGRGFVPFVPGSPTER
sp|Q68KI4.2|NHX1_ARATH          VPRPDSIRGFLTRPTRTVHYYWRQFDDSFMRPVFGGRGFVPFVPGSPTER
sp|Q84WG1.2|NHX3_ARATH          PEQHVSFRMFWKSPSRFTH------------------------------
AAM08405.1                      NRAKDSISLLMEQPVYTIHRYWRKFDDTYMRPIFGGPRRENQPEC-----
sp|Q8RWU6.3|NHX6_ARATH          DTPPGSGFRTKLREFHKSAASFTELDRNYLTPFFTSNNGDYDDEGNMEQH
sp|Q8S396.2|NHX5_ARATH          TSSSGSRFKMKLKEFHKTTTSFTALDKNFLTPFFTTNSGDGDGDGE----
                                     *

AAM08403.1                      SSHDLSKP---------
sp|Q68KI4.2|NHX1_ARATH        NPPDLSKA---------
sp|Q84WG1.2|NHX3_ARATH          -----------------
AAM08405.1                      -----------------
sp|Q8RWU6.3|NHX6_ARATH          HEERIPFTRRGNLNNRG
sp|Q8S396.2|NHX5_ARATH        -----------------
```

We can see that multiple alignment sequence ClustalW returned an output in which three sections of protein sequences revealed conserved regions that were of appropriate length for our purposes (highlighted in the output above). The three regions as described by their protein FASTA sequence, their degeneracy computations, and reverse-transcripted nucleotide sequence are featured in Figure 2.

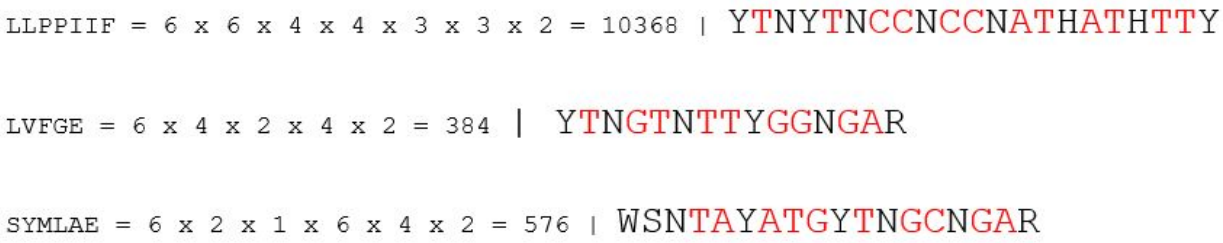

Figure 2. Conserved Regions

Even with the information that one pair of genes in the family shows approximately 79% similarity in sequence, the genes are not identifiable from the data output at this point in a strictly computational sense since the multiple sequence alignment reports only what is conserved among every gene in the family. Visually however, the pair can be pretty readily identified as genes 5 and 6. If the said pair was removed prior to alignment, we would likely find more alignment regions since the remaining genes are only 22% similar to the pair and 56% similar among each other. It is feasible that there are situations in which four genes share a conserved region but since the other two vary at that location the region is not identified by ClustalW's algorithm. At this stage however, the output does allow us to obtain degeneracy values that can predict the effectiveness of primer design based on these regions. In our case, the degeneracy values are very large and do not indicate ideal conditions for primer design for

the purpose of identifying these six genes in PCR. In fact, they indicate that even with the least degenerate sequence, there are 384 nucleotide possibilities that could present the shared region in the family which allows for a great deal of risk for including genes that express that region but have no other relation to the gene family in terms of function or organization. The conserved regions must be considered in conjunction with each other to improve our state of risk in this sense. The CODEHOP program and its algorithm attempt to produce oligonucleotide primers with that factor in mind. The CODEHOP output can be found below:

```
Oligos
Degenerate alphabet
Block ExperimenB
         F  H  E  E  F  F
oligo:5'-TTCCACganganytntt-3' degen=128 temp=9.3   Extend clamp
         F  H  E  E  F  F  F
oligo:5'-TTCCACGAGganytnttytt-3' degen=64 temp=33.4   Extend clamp
         F  H  E  E  F  F  F  I
oligo:5'-TTCCACGAGGAGytnttyttyht-3' degen=96 temp=37.1   Extend clamp
           F  F  I  Y  L  L  P  P  I  I
oligo:5'-TCTTCTTCATCTACCTGCTGccnccnathat-3' degen=48 temp=60.9
         F  I  Y  L  L  P  P  I  I  F
oligo:5'-TTCATCTACCTGCTGCCTccnathathtt-3' degen=36 temp=60.4
           Y  L  L  P  P  I  I  F  N
oligo:5'-TCTACCTGCTGCCTCCCathathttyma-3' degen=36 temp=60.8
          G  F  Q  V  K  K  K  Q  F  F
oligo:5'-CGGCTTCCAGGTGAAGAAGaarmmnttytt-3' degen=64 temp=63.9
Complement of Block ExperimenB
E  E  F  F  F  I  Y  L  L  P
ctnctnranaaGAAGTAGATGGACGACGGG oligo:5'-GGGCAGCAGGTAGATGAAGaanarntcntc-3' degen=128
temp=61.6
E  F  F  F  I  Y  L  L  P  P
ctnranaaraaGTAGATGGACGACGGGG oligo:5'-GGGGCAGCAGGTAGATGaaraanarntc-3' degen=64
temp=61.2
F  F  I  Y  L  L  P  P  I  I
aaraardanawGGACGACGGGGGATAGT oligo:5'-TGATAGGGGGCAGCAGGwanadraaraa-3' degen=96
temp=60.7
P  P  I  I  F  N  A  G  F
ggnggntadtaGAAGTTGCGGCCGAAG oligo:5'-GAAGCCGGCGTTGAAGatdatnggngg-3' degen=48 temp=61.3
P  I  I  F  N  A  G  F  Q
ggntadtadaaGTTGCGGCCGAAGG oligo:5'-GGAAGCCGGCGTTGaadatdatngg-3' degen=36 temp=62.0
I  I  F  N  A  G  F  Q  V
tadtadaarktGCGGCCGAAGGTCC oligo:5'-CCTGGAAGCCGGCGtkraadatdat-3' degen=36 temp=62.0
K  Q  F  F  R  N  F  M  T
ttykknaaraaGGCGTTGAAGCACTGG oligo:5'-GGTCACGAAGTTGCGGaaraankkytt-3' degen=64 temp=60.8
Block ExperimenD
          M  M  L  M  A  Y  F  S  Y  M  L
oligo:5'-GATGATGCTGATGGCCTACTTCwsntayatgyt-3' degen=64 temp=61.0
           M  L  M  A  Y  F  S  Y  M  L  A
oligo:5'-TGATGCTGATGGCCTACTTCTCntayatgytng-3' degen=64 temp=61.7
           L  M  A  Y  F  S  Y  M  L  A
oligo:5'-TGCTGATGGCCTACTTCTCCtayatgytngc-3' degen=16 temp=61.3
           L  M  A  Y  F  S  Y  M  L  A  E
oligo:5'-TGCTGATGGCCTACTTCTCCTAyatgytngcng-3' degen=64 temp=61.3
           M  A  Y  F  S  Y  M  L  A  E
oligo:5'-TGATGGCCTACTTCTCCTACatgytngcnga-3' degen=32 temp=60.5
Complement of Block ExperimenD
Y  M  L  A  E  L  F  H  L
atrtacrancgGCTCCACAAGGTGGAC oligo:5'-CAGGTGGAACACCTCGgcnarcatrta-3' degen=16 temp=62.7
  M  L  A  E  L  F  H  L  S  G
```

```
trtacrancgncTCCACAAGGTGGACAGGC oligo:5'-CGGACAGGTGGAACACCTcngcnarcatrt-3' degen=64
temp=61.6
M  L  A  E  L  F  H  L  S  G
tacrancgnctCCACAAGGTGGACAGGC oligo:5'-CGGACAGGTGGAACACCtcngcnarcat-3' degen=32
temp=61.6
Block ExperimenE
         H  M  M  S  F  L  A  E  T  F
oligo:5'-CACATGATGTCCTTCCTGgcngarayntt-3' degen=64 temp=61.5
           A  E  T  F  I  F  I  Y  M
oligo:5'-TGGCCGAGACCTTCATCttyhtntayrt-3' degen=96 temp=60.8
Complement of Block ExperimenE
A  E  T  F  I  F  I  Y  M  G  M
cgnctytrnaaGTAGAAGTAGATGTACCCGT oligo:5'-TGCCCATGTAGATGAAGATGaanrtytcngc-3' degen=64
temp=60.5
E  T  F  I  F  I  Y  M  G  M  D
ctytrnaaryrGAAGTAGATGTACCCGTACCT oligo:5'-TCCATGCCCATGTAGATGAAGryraanrtytc-3'
degen=128 temp=60.9
F  I  Y  M  G  M  D
aardanatryaCCCGTACCTG oligo:5'-GTCCATGCCCayrtanadraa-3' degen=96 temp=38.2   Extend
clamp
```

In this CODEHOP output, not every block from the BlockMaker format input has a corresponding set of predicted oligonucleotide primers. This is because only three conserved regions were of the appropriate length for primer design, hence three blocks are presented with predicted primers. The requirements of PCR that will allow us to select the most ideal pair of primers involves finding similarities in melting temperature and degeneracy. The best candidates are primers 1 and 4 from Block D which, looking at the BlockMaker format, corresponds to the conserved region identified earlier, designated by the protein sequence SYMLAE.
The Primer3 program was then used on two genes of our choosing, in my case those being genes 4 and 2. The outputs of the program can be found below:

Gene 4 Primer3 Output

```
KEYS (in order of precedence):
>>>>>> left primer
<<<<<< right primer

ADDITIONAL OLIGOS
                    start  len      tm     gc%   any   3'  seq

1 LEFT PRIMER       1370   20   60.32   55.00  4.00  2.00 gtaaacgcactgagccaggt
   RIGHT PRIMER     1567   20   60.07   55.00  4.00  2.00 gacgaggtccaccgaagata
   PRODUCT SIZE: 198, PAIR ANY COMPL: 3.00, PAIR 3' COMPL: 1.00

2 LEFT PRIMER       999   20   60.49   55.00  6.00  0.00 aagctttgggggtactctgg
   RIGHT PRIMER     1198   20   60.14   50.00  6.00  0.00 ccagagcaattgagacagca
   PRODUCT SIZE: 200, PAIR ANY COMPL: 6.00, PAIR 3' COMPL: 1.00

3 LEFT PRIMER       1192   20   59.76   50.00  6.00  2.00 gctctggctttcaagcagtt
   RIGHT PRIMER     1389   20   60.32   55.00  4.00  2.00 acctggctcagtgcgtttac
   PRODUCT SIZE: 198, PAIR ANY COMPL: 4.00, PAIR 3' COMPL: 1.00

4 LEFT PRIMER       152   20   60.22   50.00  3.00  0.00 ccgccattttagtaggagca
```

```
   RIGHT PRIMER     345   20   59.81   50.00  5.00  3.00 cacaccaaaggacatgatgg
   PRODUCT SIZE: 194, PAIR ANY COMPL: 4.00, PAIR 3' COMPL: 1.00

Statistics
       con   too    in     in           no     tm     tm  high  high      high
         sid  many   tar  excl   bad    GC   too   too   any       3'  poly   end
        ered Ns   get   reg   GC% clamp   low  high compl compl    X  stab      ok
Left   13648 0     0      0      33     0  5856  3387 26     95     22   157  4072
Right  13544 0     0      0      35     0  5705  3531       2      10     22   169
4070
Pair Stats:
considered 2703, unacceptable product size 1730, high any compl 1, high end compl 65,
ok 907
primer3 release 1.1.4
```

Gene 2 Primer3 Output

```
KEYS (in order of precedence):
>>>>>> left primer
<<<<<< right primer

ADDITIONAL OLIGOS
                    start  len      tm     gc%   any   3' seq

1 LEFT PRIMER       873   20   59.98   50.00  6.00  2.00 caatgtcaccgagagctcaa
   RIGHT PRIMER     1074   20   60.27   45.00  7.00  3.00 aaaagctgctcttccaagca
   PRODUCT SIZE: 202, PAIR ANY COMPL: 5.00, PAIR 3' COMPL: 0.00

2 LEFT PRIMER       873   20   59.98   50.00  6.00  2.00 caatgtcaccgagagctcaa
   RIGHT PRIMER     1075   20   60.27   50.00  7.00  2.00 caaaagctgctcttccaagc
   PRODUCT SIZE: 203, PAIR ANY COMPL: 5.00, PAIR 3' COMPL: 0.00

3 LEFT PRIMER       353   20   60.11   50.00  3.00  2.00 ttgggaccgtagtttcttgc
   RIGHT PRIMER     552   20   59.57   45.00  2.00  0.00 aacgccctctccaaatacaa
   PRODUCT SIZE: 200, PAIR ANY COMPL: 4.00, PAIR 3' COMPL: 2.00

4 LEFT PRIMER       1159   20   60.21   55.00  6.00  2.00 tgggctggtctaatgagagg
   RIGHT PRIMER     1364   20   60.04   55.00  4.00  0.00 gtggtactggtggtcgcttt
   PRODUCT SIZE: 206, PAIR ANY COMPL: 3.00, PAIR 3' COMPL: 1.00

Statistics
       con   too    in     in           no     tm     tm  high  high      high
         sid  many   tar  excl   bad    GC   too   too   any       3'  poly   end
        ered Ns   get   reg   GC% clamp   low  high compl compl    X  stab      ok
Left   13859 0     0      0      19     0  5102  4567       6   103      41   193
3828
Right  13708 0     0      0      19     0  4881  4715       1     14     45   192
3841
Pair Stats:
considered 5532, unacceptable product size 4229, high any compl 3, high end compl 97,
ok 1203
primer3 release 1.1.4
```

These output results were produced with default settings. Variables such as length, GC content, and melting temperature are critical in gene amplification so the best primers were selected based on these criteria. Ideal primer length was 18 and 22bp which was well with our primers of 20bp each. Each primer pair also satisfied the GC content levels we needed (between 40 and

60% was ideal). The melting temperature was ideally 54-58 degrees Celsius and absolutely less than 65 degrees Celsius which allowed us to distinguish Primer 4 in Gene 4 and Primer 1 in Gene 2 as the most ideal pairs. These primers were built specifically for each gene since their construction was based on the FASTA sequence unique to the gene rather than the CODEHOP primers that were constructed from conserved regions among the gene family. A second primer for the genes in the Primer3 output might be useful in the case of degeneracy where a protein with identical amino acid sequence or function might fail to be amplified during PCR due to a silent mutation or otherwise ultimately inconsequential change in the nucleotide sequence.

Discussion

Overall, the results were demonstrative of the conceptual framework that we prepared for in this lab. Since this experiment didn't require the scientific method with regard to the hypothesis or the data to support it, the success of our result is precedent upon the construction of our degenerate primers as output from our online resources. The objectives of the lab were successfully met to my estimation and will be further verified for accuracy when the primers we've designed are employed in Experiment 5.

Conclusion

This lab provided us with the opportunity to apply computational algorithms in a procedure that is directly related to biotechnology application. The PCR gene amplification that was implied downstream of our work here was a helpful contextual background that allowed us to identify the importance of various variables we encountered. I had previously not been aware that these resources were freely available online and it begs the question of whether the application of such heavy bioinformatics work that can be performed around the world for relatively little cost will lead to novel biotechnology integration at a greater rate. The experiment has personally solidified my knowledge of conserved regions, gene families, and degenerate primer construction through involved, hands-on application.

References

National Center for Biotechnology Information https://www.ncbi.nlm.nih.gov/

ClustalW http://www.genome.jp/tools-bin/clustalw

Protein to DNA Reverse Translation www.biophp.org/minitools/protein_to_dna/demo.php

BlockMaker http://blocks.fhcrc.org/blocks/make_blocks.html

CODEHOP http://blocks.fhcrc.org/codehop.html

Primer3 http://biotools.umassmed.edu/bioapps/primer3_www.cgi

PremierBiosoft “PCR Primer Design Guidelines” (1994-2018)
http://www.premierbiosoft.com/tech_notes/PCR_Primer_Design.html