Primer Design Workshop

École d'été en génétique des champignons 2012
Dr. Will Hintz
University of Victoria

Scenario

- You have discovered the presence of a novel endophytic organism living inside the cells of a tropical palm species
- They synthesize alkaloids that give some protection to the plant against insect attack
- The organism cannot be cultured but is thought to be related to the Clavicitaceae, Ascomycetes (a fungus)

Bio-mining

- Several genes from endophytic fungi related to toxin production are now being cloned and studied in depth
- i.e. (DMAT Synthase) dimetilail tryptophane synthase – the first step in ergot synthesis in *Claviceps purpurea*
- Mycotoxin genes are often clustered
- First you have to identify the endophyte

Challenges

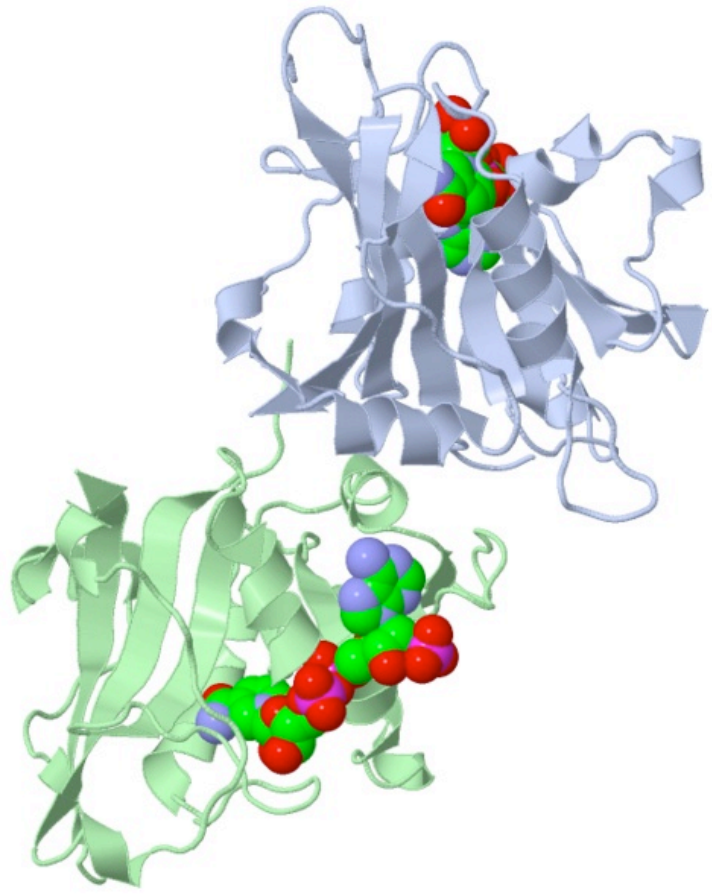
- The endophyte may be difficult to dissociate from the host and may unculturable
- You need to distinguish between the host genome and the endophyte genome
- You have absolutely no genetic information on the endophyte nor any idea of its identity
- First step: How is this organism related to others in its phylum?

Approach: Reverse Translation from compiled consensus peptide sequence

- How can you derive the DNA sequence for a gene when you don't have any peptide or DNA sequence information?
- Infer the polypeptide sequence by alignment of similar sequences from other closely related organisms
- Reverse translate with optimal conditions
- Today we will use as an example the gene for dihydrofolate reductase

Why DHFR?

- Ubiquitous enzyme found in all organisms
- Dihydrofolate reductase catalyzes reduction of dihydrofolic acid to tetrahydrofolic acid (folate) – an essential vitamin B9
- http://proteopedia.org/wiki/index.php/



In Class Assignment

Locate two regions of the DHFR gene suitable for primer design – support your choice

Derive oligonucleotide primers specific to these regions for PCR amplification of the intervening DNA

Give the anticipated length of the amplification product and explain how you would confirm its identity

Reverse Translation Steps

- Align several DHFR genes from database
- Identify conserved domains
- Reverse translate the protein into short DNA oligomers
- Develop (optimized) PCR primers to amplify gene from the unknown organism (and possibly host)
- Do Blast search and construct phylogenetic tree for DHRF sequences to possibly gain identity of unknown organism

Compile DHFR Sequence List

```
>Candida albicans
MLKPNVAIIVAALKPALGIGYKGKMPWRLRKEIRYFKDVTTRTTKPNTRNAVIMGRKTWESIPQKFRPLPDRLNIILSRSYENEIIDDNIIHASSIESSLNLVSDV
ERVFIIGGAEIYNELINNSLVSHLLITEIEHPSPESIEMDTFLKFPLESWTKQPKSELQKFVGDTVLEDDIKEGDFTYNYTLWTRK
>Mus musculus
MVRPLNCIVAVSQDMGIGKNGDLPWPPLRNEWKYFQRMTTTSSVEGKQNLVIMGRKTWFSIPEKNRPLKDRINIVLSRELKEPPRGAHFLAKSLDDALRLIEQPEL
ASKVDMVWIVGGSSVYQEAMNQPGHLRLFVTRIMQEFESDTFFPEIDLGKYKLLPEYPGVLSEVQEEKGIKYKFEVYEKKD
>Saccharomyces cerevisiae
MAGGKIPIVGIVACLQPEMGIGFRGGLPWRLPSEMKYFRQVTSLTKDPNKKNALIMGRKTWESIPPKFRPLPNRMNVIISRSFKGDFVHDKERSIVQSNSLANAIT
NLESNFKEHLERIYVIGGGEVYSQIFSITDHWLITKINPLDKNATPAMDTFLDAKKLEEVFSEQDPAQLKEFLPPKVELPETDCDQRYSLEEKGYCFEFTLYNRK
>Aspergillus oryzae
MPPTNPLTLIVATTPIPTREKTLLGIGLNGTLPWPRIKADMSFFARVTTRPPRPGTTNAMIMGRKTYDSVPKSLRPLGKRINVIVTRDVEGVSKRVAEELKEKRAK
MAAAAAAATSAGENKEEGPITDAIVSSGLEAALEDVEEKFKGGLGSVFVIGGAEIYATALGLGGDRPVRIVMTNVEKKGVDGEKAVFECDTFFPIDEELLMEKGWR
KVSAEEVTEWVGEPVSGEWKDEGEVRIQMVGYERVN
>Pneumocystis carinii
MNQQKSLTLIVALTTSYGIGRSNSLPWKLKKEISYFKRVTSFVPTFDSFESMNVVLMGRKTWESIPLQFRPLKGRINVVITRNESLDLGNGIHSAKSLDHALELLY
RTYGSESSVQINRIFVIGGAQLYKAAMDHPKLDRIMATIIYKDIHCDVFFPLKFRDKEWSSVWKKEKHSDLESWVGTKVPHGKINEDGFDYEFEMWTDL
>Cordyceps militaris
MQSLELTLVVAATRSMGIGAGGTMPWNGLRNEMKYFARVTTQLASPCPSGAVNAVIMGRKTWDSIPPKFRPLKGRLNIIISRAAATTPPPPPGVQGPVVRVASVEA
ALQYAGAHCGGGRIFVIGGGQVYKAVLRRPEVRRVLLTRIETEYDCDTFFPIKLGADDGTEPGWTRRSDEQWRAWTGEPDTEKGRREEAGVKYEFQMWERE
>[Trichophyton rubrum CBS]
MTTSIANPRMPAKLPPLTLVVATTPITTPTNPGILKLGIGKEGTLPWPRIKKDMSFFARVTTRPPATATASGSASPAINAVIMGRKTYDSIPAKFRPLSKRLNVII
TRDESGSVKERAIADWNASRKRELEKQADQDTDKAAATSTSTEEPEVIVSSSLEDALSTLQRNFVISSSSDVQGGKKRLGNVYIMGGSEIYASSLRLTADALGENN
PLRIVMTDIRRRADGNAQCDVEDLVDGFECDTCFPLDGKGLKEGWNKVPSEKLAEWVGEAVSSDWTWEGDIAMKISGYEKL
>Puccinia graminis
MIFWKQKNLIRYYEGRRLTKLEQPTRIPIPRQPRPIQAKPAEGDMIPLNLIVCATKSNGIGQAGRLPWRLKEDMNFFKSVTTLAPSGCKNVVIMGRKTWLSIPSKF
RPLANRINIVVSRQSKDPAALDIHQQQDSYLVNSIESACHLIRTLDSPDNNTQQETTTSVLKERKTGGEPLVNKVFVIGGSEIYKSVLDSQPDNNRLYKPSTILMT
RILSEHPAIETSLDAFFPEFRASKHWSKSQNPDLLNQFLVLPDHQNSSGSVPEPHYQLPFNFDQLIIENSFMYKFELWTYNNPL
>Cryptococcus neoformans
MQTTAKSSTPSITAVVAATAENGIGLNGGLPWRLPGEMKYFARVTTGETPSSDPSEQNVVIMGRKTWESIPSRFRPLKNRRNVVISGKGVDLGTAENSTVYTDIPS
ALSALRSTTESGHSPRIFLIGGATLYTSSLLPSSVPSLNSSTSTSPLPFSRPLIDRILLTRILSPFECDAYLEDFAAHTKPDGSKVWKKASIKEFREWIGWDIEEQ
VEEKGVKYIFEMWVLNQ
```

Align all known peptide sequences

http://www.ebi.ac.uk/Tools/msa/clustalw2/

```
Candida          ---------MLKPN----------------------------------------VAIIVAA 12
Saccharomyces    -------MAGGKIP----------------------------------------IVGIVAC 14
Mus              ---------------MVR-----------------------------------PLNCIVA 10
                                                                        :..

Candida          LKPA----------LGIGYKGKMPWR-LRKEIRYFKDVTT----RTTKPNT----RNAVI 53
Saccharomyces    LQPE----------MGIGFRGGLPWR-LPSEMKYFRQVTS----LTKDPNK----KNALI 55
Mus              VSQD----------MGIGKNGDLPWPPLRNEWKYFQRMTT----TSSVEGK----QNLVI 52
                               ***  . :**  :  :  :*  :*:               * ::

Candida          MGRKTWESIPQKFRPLPDRLNIILSRSYENEIIDDN-----IIHASSIESSLN------- 101
Saccharomyces    MGRKTWESIPPKFRPLPNRMNVIISRSFKGDFVHDKE--RSIVQSNSLANAITNLES--- 110
Mus              MGRKTWFSIPEKNRPLKDRINIVLSRELKEPP------RGAHFLAKSLDDALRLIE---- 102
                 *****: *:*   ***  * *::::

Candida          ------------------------------------LVSDVERVFIIGGAEIYNELIN 123
Saccharomyces    ------------------------------------NFKEHLERIYVIGGGEVYSQIFS 133
Mus              -----QPELAS-------------------------KVDMVWIVGGSSVYQEAMN 127
                                                     ::::**. :*

Candida          N--------------SLVSHLLITEIEHPSP-ESIEMDTFLKFP--LESWTKQP-KSELQ 165
Saccharomyces    ----------------ITDHWLITKINPLDKNATPAMDTFLDAKKLEEVFSEQD-PAQLK 176
Mus              --------QPGHLRLF------VTRIMQEFES-----DTFFPEID-------------LG 155
                              : .

Candida          KFVG--------DTVLEDD----------IKEGDFTYNYTLWTRK--- 192
Saccharomyces    EFLPPKVELPETDCDQRYS----------LEEKGYCFEFTLYNRK--- 211
Mus              KYKLLPEYPGVLSEVQE-------------EKGIKYKFEVYEKKD-- 187
                           .                 *        :
```

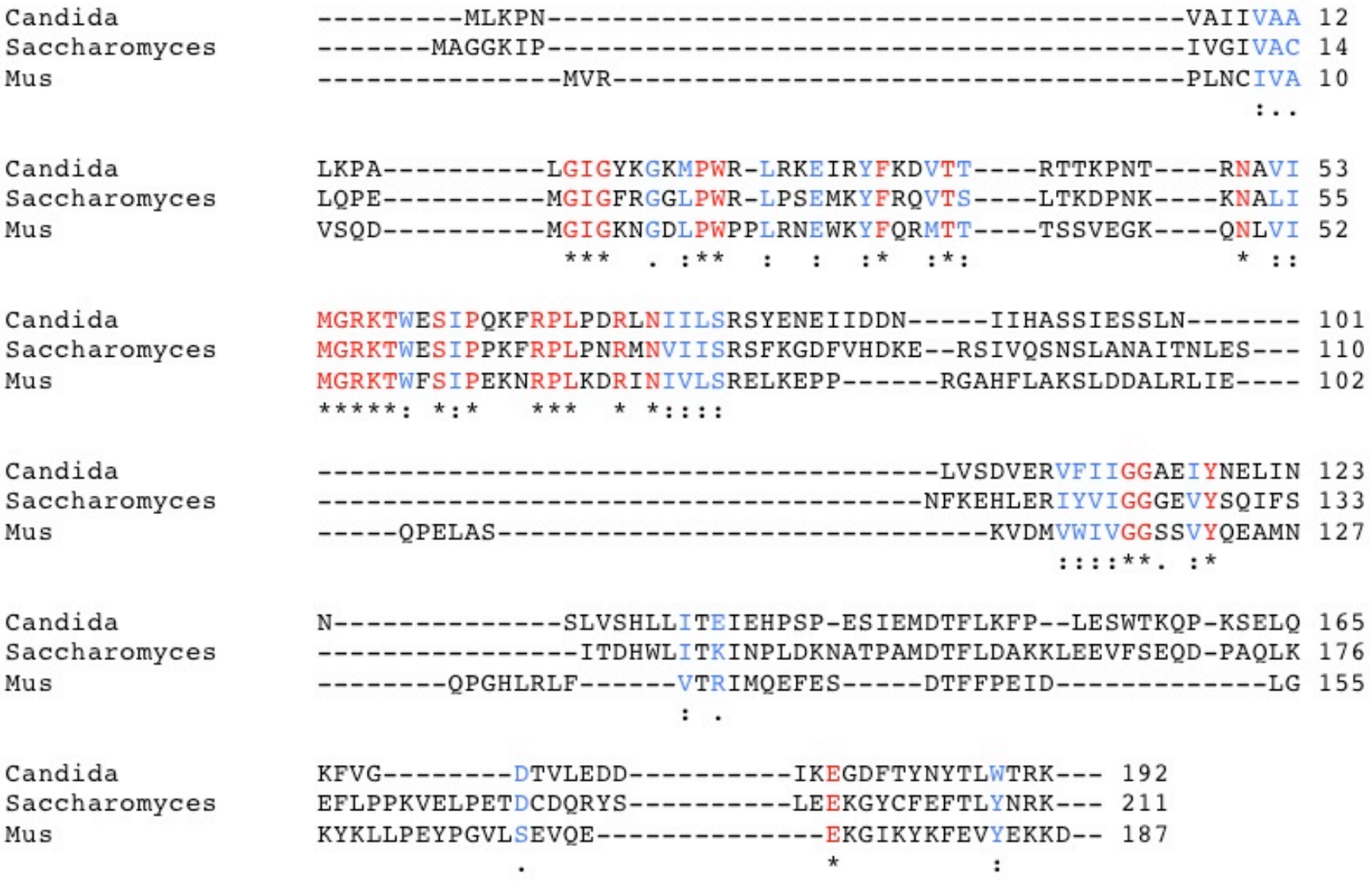
- Note – only three sequences shown

Identify highly conserved amino acids

```
Candida          ---------MLKPN----------------------------------------VAIIVAA 12
Saccharomyces    -------MAGGKIP----------------------------------------IVGIVAC 14
Mus              ---------------MVR-----------------------------------PLNCIVA 10
                                                                        :..

Candida          LKPA----------LGIGYKGKMPWR-LRKEIRYFKDVTT----RTTKPNT----RNAVI 53
Saccharomyces    LQPE----------MGIGFRGGLPWR-LPSEMKYFRQVTS----LTKDPNK----KNALI 55
Mus              VSQD----------MGIGKNGDLPWPPLRNEWKYFQRMTT----TSSVEGK----QNLVI 52
                               ***  . :**  :  :  :*  :*:               * ::

Candida          MGRKTWESIPQKFRPLPDRLNIILSRSYENEIIDDN-----IIHASSIESSLN------- 101
Saccharomyces    MGRKTWESIPPKFRPLPNRMNVIISRSFKGDFVHDKE--RSIVQSNSLANAITNLES--- 110
Mus              MGRKTWFSIPEKNRPLKDRINIVLSRELKEPP------RGAHFLAKSLDDALRLIE---- 102
                 *****: *:*   ***  * *::::

Candida          ------------------------------------LVSDVERVFIIGGAEIYNELIN 123
Saccharomyces    ------------------------------------NFKEHLERIYVIGGGEVYSQIFS 133
Mus              -----QPELAS-------------------------KVDMVWIVGGSSVYQEAMN 127
                                                     ::::**. :*

Candida          N--------------SLVSHLLITEIEHPSP-ESIEMDTFLKFP--LESWTKQP-KSELQ 165
Saccharomyces    ----------------ITDHWLITKINPLDKNATPAMDTFLDAKKLEEVFSEQD-PAQLK 176
Mus              --------QPGHLRLF------VTRIMQEFES-----DTFFPEID-------------LG 155
                               : .

Candida          KFVG--------DTVLEDD----------IKEGDFTYNYTLWTRK--- 192
Saccharomyces    EFLPPKVELPETDCDQRYS----------LEEKGYCFEFTLYNRK--- 211
Mus              KYKLLPEYPGVLSEVQE--------------EKGIKYKFEVYEKKD-- 187
                           .                  *         :
```

- Note – only three sequences shown

Also look at less conserved regions

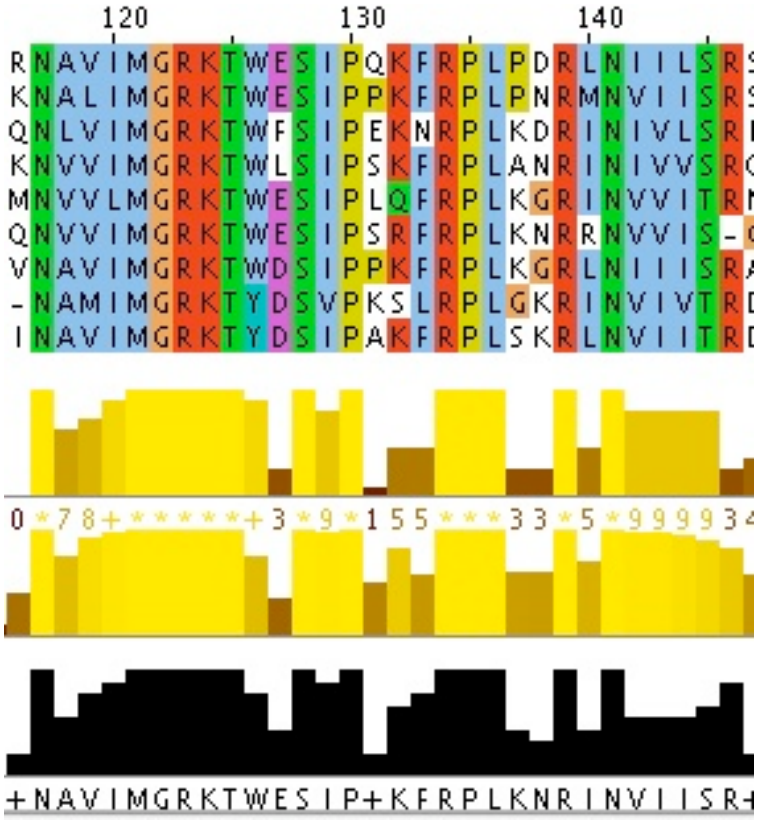


- Note – only three sequences shown

Re-check

- You can also recheck the alignment using other programs to see if you get consistent results
- Checking with COBALT http://www.ncbi.nlm.nih.gov/tools/cobalt/ gave much the same result
- Next derive a consensus sequence for the aligned regions

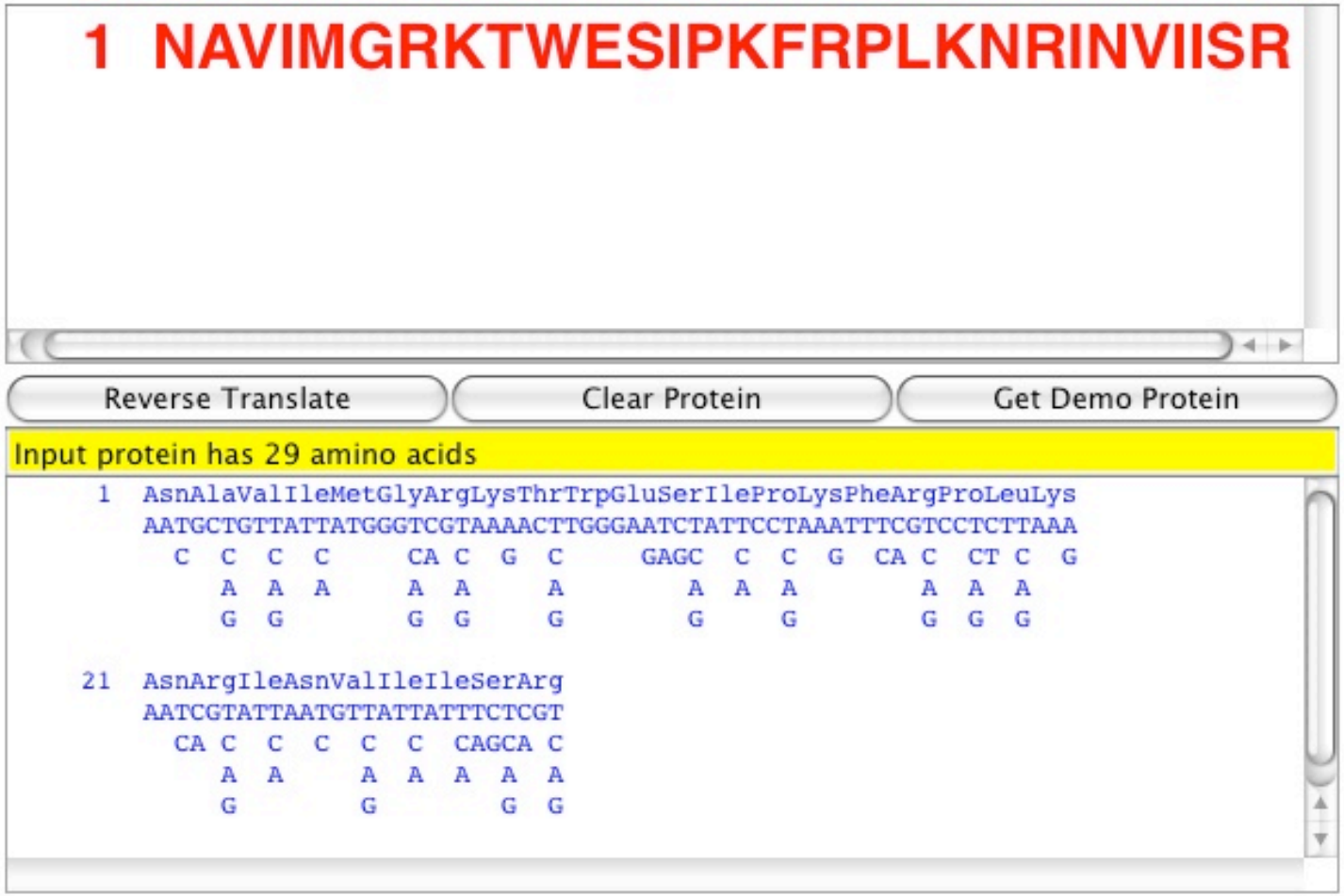
Consensus sequence



- You can draw a consensus sequence from the alignment or have Clustal do it for you

Reverse translate the peptide

- Many simple software tools to do this (i.e. http://www.vivo.colostate.edu/molkit/rtranslate/index.html



Codon Usage Database

- http://www.kazusa.or.jp/codon/

***Aspergillus niger* [gbpln]: 13905 CDS's (6130423 codons)**

fields: [triplet] [frequency: **per thousand**] ([number])

UUU 12.8(78285)	UCU 14.0(86042)	UAU 12.3(75100)	UGU 5.8(35261)
UUC 23.8(145663)	UCC 19.0(116511)	UAC 17.1(104725)	UGC 8.3(50889)
UUA 5.1(31278)	UCA 10.6(65095)	UAA 0.6(3916)	UGA 0.9(5689)
UUG 16.4(100325)	UCG 14.0(85723)	UAG 0.7(4444)	UGG 15.3(93788)
CUU 15.4(94490)	CCU 15.2(93422)	CAU 12.3(75191)	CGU 10.1(61861)
CUC 22.4(137051)	CCC 17.8(109010)	CAC 12.8(78227)	CGC 15.9(97559)
CUA 8.9(54445)	CCA 13.1(80260)	CAA 15.8(96611)	CGA 9.3(57162)
CUG 23.0(141009)	CCG 14.2(87312)	CAG 24.2(148296)	CGG 11.2(68379)
AUU 16.7(102080)	ACU 13.6(83396)	AAU 14.9(91080)	AGU 10.6(64937)
AUC 26.2(160337)	ACC 21.3(130788)	AAC 21.0(129023)	AGC 15.4(94652)
AUA 6.9(42354)	ACA 12.4(76029)	AAA 14.1(86447)	AGA 7.9(48671)
AUG 22.0(134646)	ACG 12.9(79289)	AAG 29.6(181722)	AGG 7.6(46656)
GUU 14.7(89946)	GCU 21.7(132918)	GAU 27.8(170550)	GGU 17.5(107274)
GUC 21.8(133437)	GCC 27.2(166685)	GAC 27.2(166885)	GGC 22.5(137915)
GUA 6.9(42490)	GCA 17.4(106738)	GAA 24.8(151769)	GGA 16.1(98975)
GUG 18.6(114235)	GCG 17.2(105173)	GAG 34.7(212844)	GGG 12.6(77463)

- Search the database for a codon preference for an organism which you suspect is related to your unknown

G	Gly	GGG	0.16	W	Trp	TGG	1.00
G	Gly	GGA	0.20	*	End	TGA	0.44
G	Gly	GGT	0.30	C	Cys	TGT	0.31
G	Gly	GGC	0.34	C	Cys	TGC	0.72
E	Glu	GAG	0.62	*	End	TAG	0.17
E	Glu	GAA	0.38	*	End	TAA	0.39
D	Asp	GAT	0.50	Y	Tyr	TAT	0.34
D	Asp	GAC	0.50	Y	Tyr	TAC	0.66
V	Val	GTG	0.21	L	Leu	TTG	0.13
V	Val	GTA	0.09	L	Leu	TTA	0.07
V	Val	GTT	0.29	F	Phe	TTT	0.30
V	Val	GTC	0.37	F	Phe	TTC	0.70
A	Ala	GCG	0.19	S	Ser	TCG	0.16
A	Ala	GCA	0.17	S	Ser	TCA	0.14
A	Ala	GCT	0.29	S	Ser	TCT	0.21
A	Ala	GCC	0.35	S	Ser	TCC	0.23
R	Arg	AGG	0.10	R	Arg	CGG	0.18
R	Arg	AGA	0.08	R	Arg	CGA	0.17
S	Ser	AGT	0.09	R	Arg	CGT	0.20
S	Ser	AGC	0.17	R	Arg	CGC	0.27
K	Lys	AAG	0.70	Q	Gln	CAG	0.61
K	Lys	AAA	0.30	Q	Gln	CAA	0.39
N	Asn	AAT	0.29	H	His	CAT	0.45
N	Asn	AAC	0.71	H	His	CAC	0.55
M	Met	ATG	1.00	L	Leu	CTG	0.24
I	Ile	ATA	0.11	L	Leu	CTA	0.10
I	Ile	ATT	0.31	L	Leu	CTT	0.20
I	Ile	ATC	0.58	L	Leu	CTC	0.26
T	Thr	ACG	0.22	P	Pro	CCG	0.22
T	Thr	ACA	0.19	P	Pro	CCA	0.22
T	Thr	ACT	0.25	P	Pro	CCT	0.27
T	Thr	ACC	0.34	P	Pro	CCC	0.28

- Construct codon preference table for highly expressed genes

G	Gly	GGG	0.16	W	Trp	TGG	1.00
G	Gly	GGA	0.20	*	End	TGA	0.44
G	Gly	GGT	0.30	C	Cys	TGT	0.31
G	Gly	GGC	0.34	C	Cys	TGC	0.72
E	Glu	GAG	0.62	*	End	TAG	0.17
E	Glu	GAA	0.38	*	End	TAA	0.39
D	Asp	GAT	0.50	Y	Tyr	TAT	0.34
D	Asp	GAC	0.50	Y	Tyr	TAC	0.66
V	Val	GTG	0.21	L	Leu	TTG	0.13
V	Val	GTA	0.09	L	Leu	TTA	0.07
V	Val	GTT	0.29	F	Phe	TTT	0.30
V	Val	GTC	0.37	F	Phe	TTC	0.70
A	Ala	GCG	0.19	S	Ser	TCG	0.16
A	Ala	GCA	0.17	S	Ser	TCA	0.14
A	Ala	GCT	0.29	S	Ser	TCT	0.21
A	Ala	GCC	0.35	S	Ser	TCC	0.23
R	Arg	AGG	0.10	R	Arg	CGG	0.18
R	Arg	AGA	0.08	R	Arg	CGA	0.17
S	Ser	AGT	0.09	R	Arg	CGT	0.20
S	Ser	AGC	0.17	R	Arg	CGC	0.27
K	Lys	AAG	0.70	Q	Gln	CAG	0.61
K	Lys	AAA	0.30	Q	Gln	CAA	0.39
N	Asn	AAT	0.29	H	His	CAT	0.45
N	Asn	AAC	0.71	H	His	CAC	0.55
M	Met	ATG	1.00	L	Leu	CTG	0.24
I	Ile	ATA	0.11	L	Leu	CTA	0.10
I	Ile	ATT	0.31	L	Leu	CTT	0.20
I	Ile	ATC	0.58	L	Leu	CTC	0.26
T	Thr	ACG	0.22	P	Pro	CCG	0.22
T	Thr	ACA	0.19	P	Pro	CCA	0.22
T	Thr	ACT	0.25	P	Pro	CCT	0.27
T	Thr	ACC	0.34	P	Pro	CCC	0.28

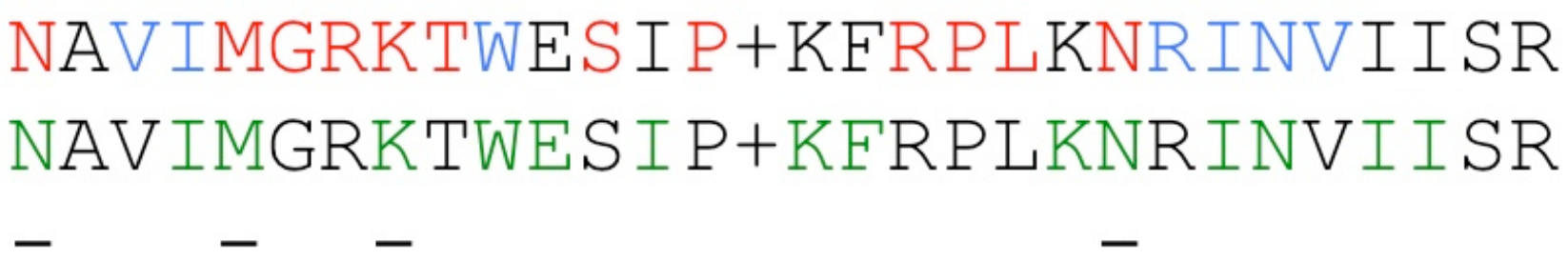
- Identify low redundancy codons and highly preferred codons
(> 58 % usage)

G	Gly	GGG	0.16	W	Trp	TGG	1.00
G	Gly	GGA	0.20	*	End	TGA	0.44
G	Gly	GGT	0.30	C	Cys	TGT	0.31
G	Gly	GGC	0.34	C	Cys	TGC	0.72
E	Glu	GAG	0.62	*	End	TAG	0.17
E	Glu	GAA	0.38	*	End	TAA	0.39
D	Asp	GAT	0.50	Y	Tyr	TAT	0.34
D	Asp	GAC	0.50	Y	Tyr	TAC	0.66
V	Val	GTG	0.21	L	Leu	TTG	0.13
V	Val	GTA	0.09	L	Leu	TTA	0.07
V	Val	GTT	0.29	F	Phe	TTT	0.30
V	Val	GTC	0.37	F	Phe	TTC	0.70
A	Ala	GCG	0.19	S	Ser	TCG	0.16
A	Ala	GCA	0.17	S	Ser	TCA	0.14
A	Ala	GCT	0.29	S	Ser	TCT	0.21
A	Ala	GCC	0.35	S	Ser	TCC	0.23
R	Arg	AGG	0.10	R	Arg	CGG	0.18
R	Arg	AGA	0.08	R	Arg	CGA	0.17
S	Ser	AGT	0.09	R	Arg	CGT	0.20
S	Ser	AGC	0.17	R	Arg	CGC	0.27
K	Lys	AAG	0.70	Q	Gln	CAG	0.61
K	Lys	AAA	0.30	Q	Gln	CAA	0.39
N	Asn	AAT	0.29	H	His	CAT	0.45
N	Asn	AAC	0.71	H	His	CAC	0.55
M	Met	ATG	1.00	L	Leu	CTG	0.24
I	Ile	ATA	0.11	L	Leu	CTA	0.10
I	Ile	ATT	0.31	L	Leu	CTT	0.20
I	Ile	ATC	0.58	L	Leu	CTC	0.26
T	Thr	ACG	0.22	P	Pro	CCG	0.22
T	Thr	ACA	0.19	P	Pro	CCA	0.22
T	Thr	ACT	0.25	P	Pro	CCT	0.27
T	Thr	ACC	0.34	P	Pro	CCC	0.28

- Also look at medium redundancy codons or lower prererences (45 - 55% usage)

M	Met	ATG	1.00
W	Trp	TGG	1.00
C	Cys	TGC	0.72
N	Asn	AAC	0.71
F	Phe	TTC	0.70
K	Lys	AAG	0.70
Y	Tyr	TAC	0.66
E	Glu	GAG	0.62
Q	Gln	CAG	0.61
I	Ile	ATC	0.58
H	His	CAC	0.55
D	Asp	GAT	0.50

- Create ranked list and begin to scan sequence for optimal codons



- Integrate both sources of information

Red = highly conserved

Blue = less conserved

Green = low redundancy

Tips for Primer Design

- Roughly calculate Tm by adding 4° C for every G or C and 2° C for every A or T
- Aim for a Tm of 60° – 62° C
- Increase length when you suspect a mismatch might occur
- End primer after two bases of the triplet
- Take reverse complement of the sequence for the return primer
- Check for snap-backs, self-annealing, or cross annealing between primers

Example Primer 01 Designs

```
 N   A   V   I   M   G   R   K   W   E   S   I   P   +
                 ATG GGA CGT AAG ACC TGG GA
AAC GCC GTC ATC ATG GGC
                       T CGT AAA ACT TAT GAT TCT ATT CC
            ATC ATG GGC CGC AAG ACC TG
                 ATG GGC CGC AAG ACC TGG GAG TCC ATC CC
  C GCC GTC ATC ATG GGT CG
            ATC GGC AAG AAC GGT ACC CTC CC
                 ATG GGC CGC AAG ACC TGG
                 ATG GGC CGC AAG TGG GAG TCC ATC CC
AAC GCC GTC ATC ATG GGC CG
AAC GCC GTC ATC ATC GGC
```

Confirmation of design

```
N   A   V   I   M   G   R   K   W   E   S   I   P   +
            ATG GGC CGC AAG ACC TGG GAG TCC ATC CC
```

BLASTn search with one primer gave the following results (of many)

> gb|CP000294.1| Cryptococcus gattii WM276 chromosome I, complete sequence Length=989306
Features in this part of subject sequence: dihydrofolate reductase, putative
Score = 42.1 bits (21), Expect = 0.070 Identities = 27/29 (93%), Gaps = 0/29 (0%) Strand=Plus/Minus

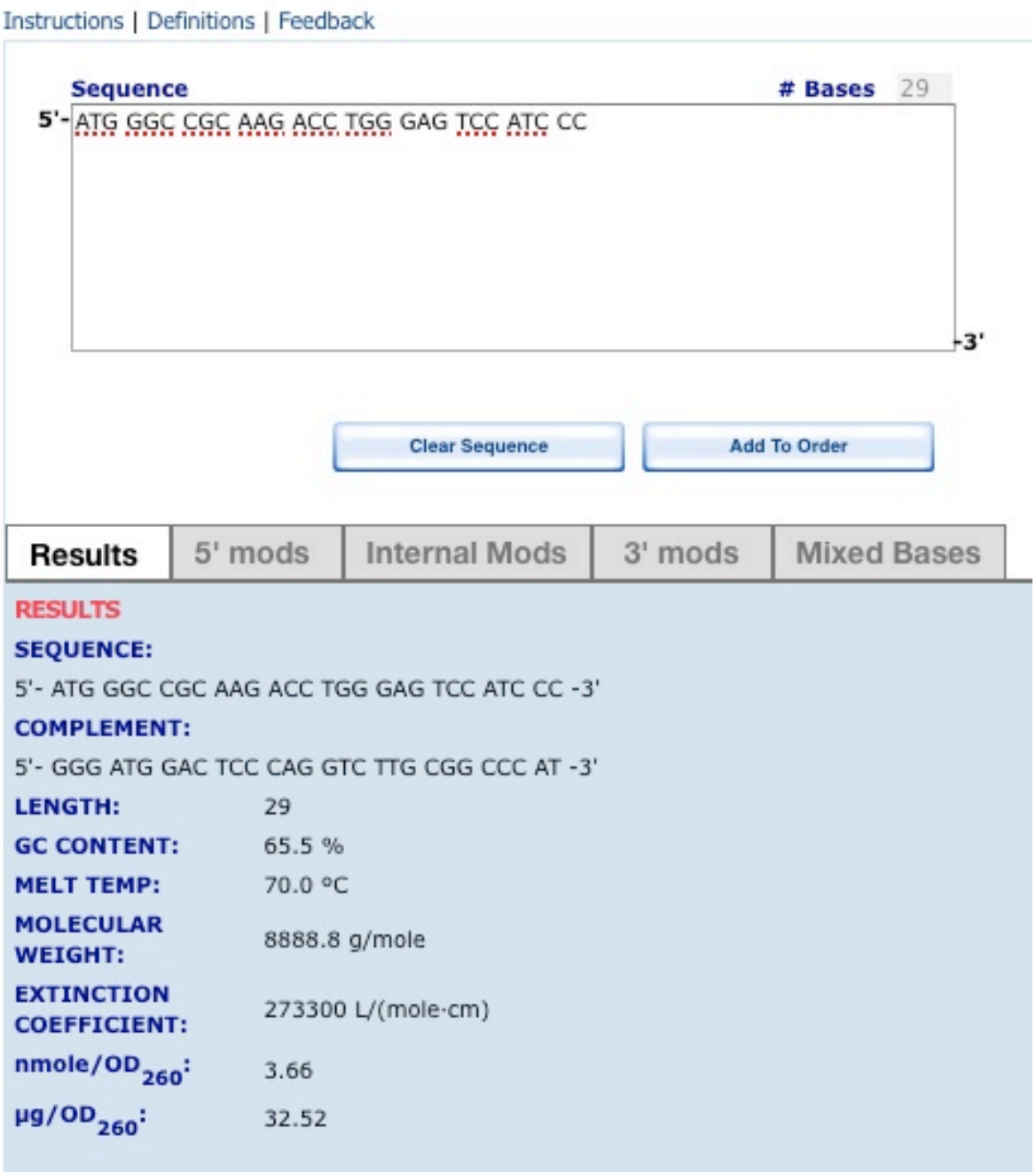
```
Query  1       ATGGGCCGCAAGACCTGGGAGTCCATCCC  29
               ||||||||||||||| ||||| ||||||||
Sbjct  549916  ATGGGCCGCAAGACATGGGAATCCATCCC  549888
```

http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastHome

http://www.humgen.nl/primer_design.html

OligoAnalyzer 3.1

http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/



OligoAnalyzer 3.1

Maximum Change in the Gibbs Free Energy
ΔG is -64 kcal.mol^{-1}

HOMO-DIMER ANALYSIS ?

Dimer Sequence

```
5'- ATGGGCCGCAAGACCTGGGAGTCCATCCC -3'
```

Maximum Delta G -64 kcal/mole

Delta G -9.28 kcal/mole
Base Pairs 4

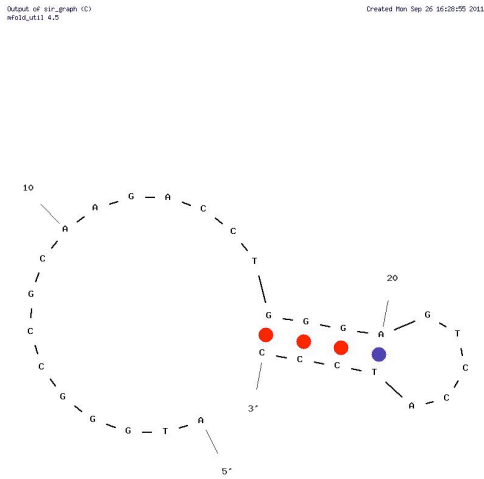
```
5'                    ATGGGCCGCAAGACCTGGGAGTCCATCCC
                        ||||
3' CCCTACCTGAGGGTCCAGAACGCCGGGTA
```

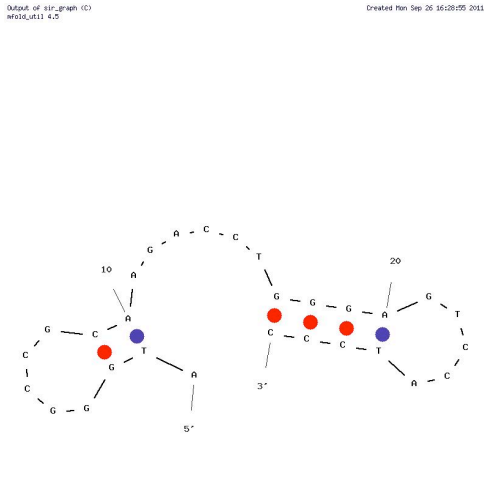
Delta G -7.71 kcal/mole
Base Pairs 4

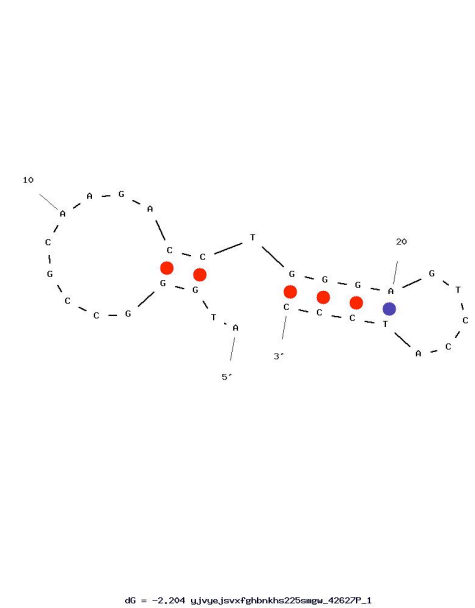
```
5' ATGGGCCGCAAGACCTGGGAGTCCATCCC
                  ||||     ::::
3'                CCCTACCTGAGGGTCCAGAACGCCGGGTA
```

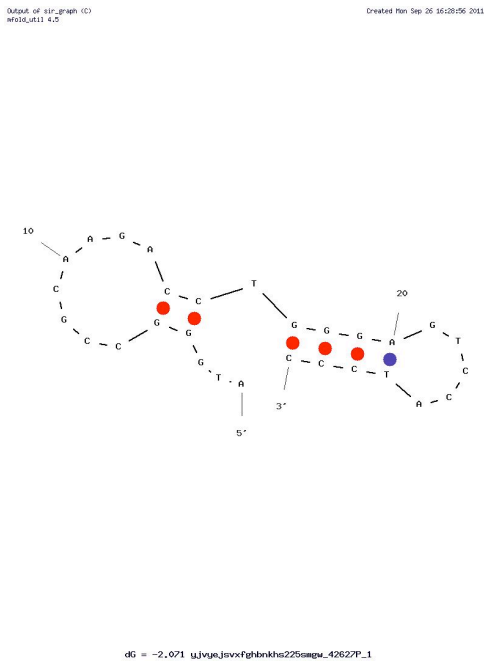
OligoAnalyzer 3.1

Change in the Gibbs Free Energy ΔG (kcal.mol^{-1})	Melt temp Tm (°C)
-2.86	54.5
-2.54	42.3
-2.20	39.1
-2.07	38.9









First Primer Analysis

- Run oligo analysis on your own primer sequence to determine the Tm, potential for snap-back and self annealing, and finally BLAST primer
- Note that this region can be used as either a forward or return primer and that the complement is listed on the results page

Second Primer Analysis

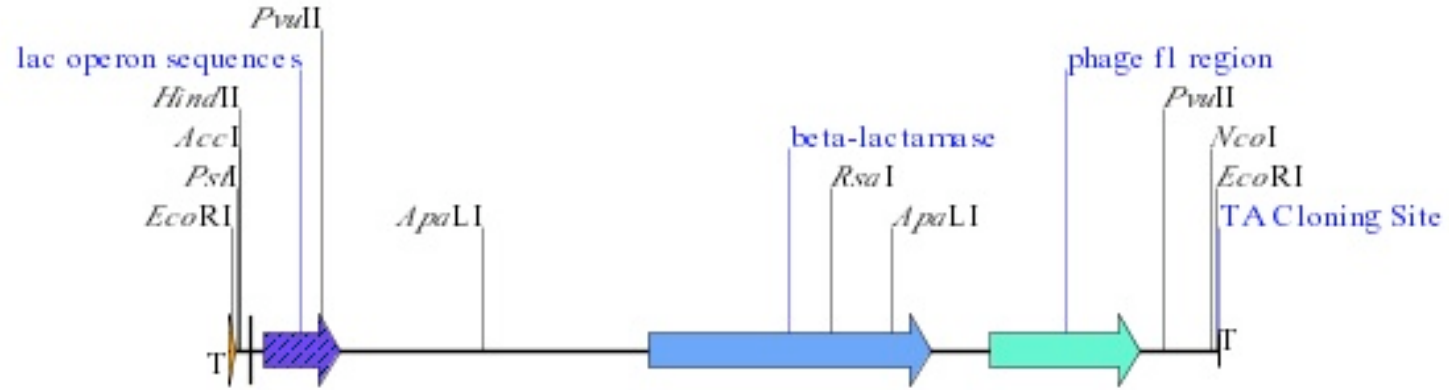
- Determine which region will be used for the forward primer design and which will be used for the return primer design
- Design return primer and analyze as before
- Run oligo analysis on both primer sequences to determine whether there is a potential for annealing between primers
- From the DHFR peptide alignment predict the size of amplification product

Next steps

- Amplify from environmental sample, clone the PCR products (using pGEM or equivalent) and determine the DNA sequence
- Match the sequence between the primers with derived polypeptide sequence to determine whether the authentic locus has been amplified
- Use authentic sequence to walk upstream and downstream to capture the entire gene

TA Cloning

PCR Product

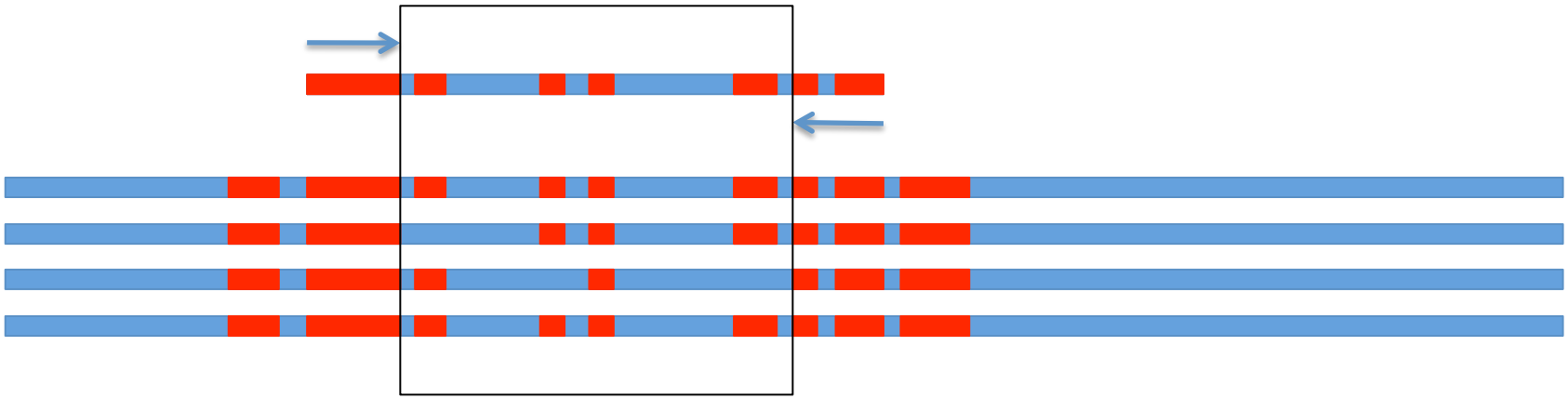


pGEM-T Easy T-Vector1
3015 bp

Many Taq polymerases preferentially add an adenine to the 3' end of the product
Vector tailed with a single 3’ overhang thymine residue on each end
T and A are paired and gap DNA linked by Ligase
Transform into E. coli and determine sequence using universal primers

Analysis

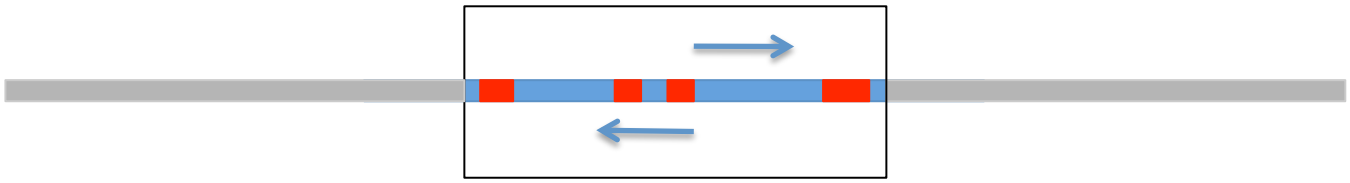
- Determine sequence of your PCR product and compare to original alignment



- If the sequence between the priming sites is more similar to fungal DNAs than plant DNAs you have your first piece of endophyte DHFR DNA – What now?

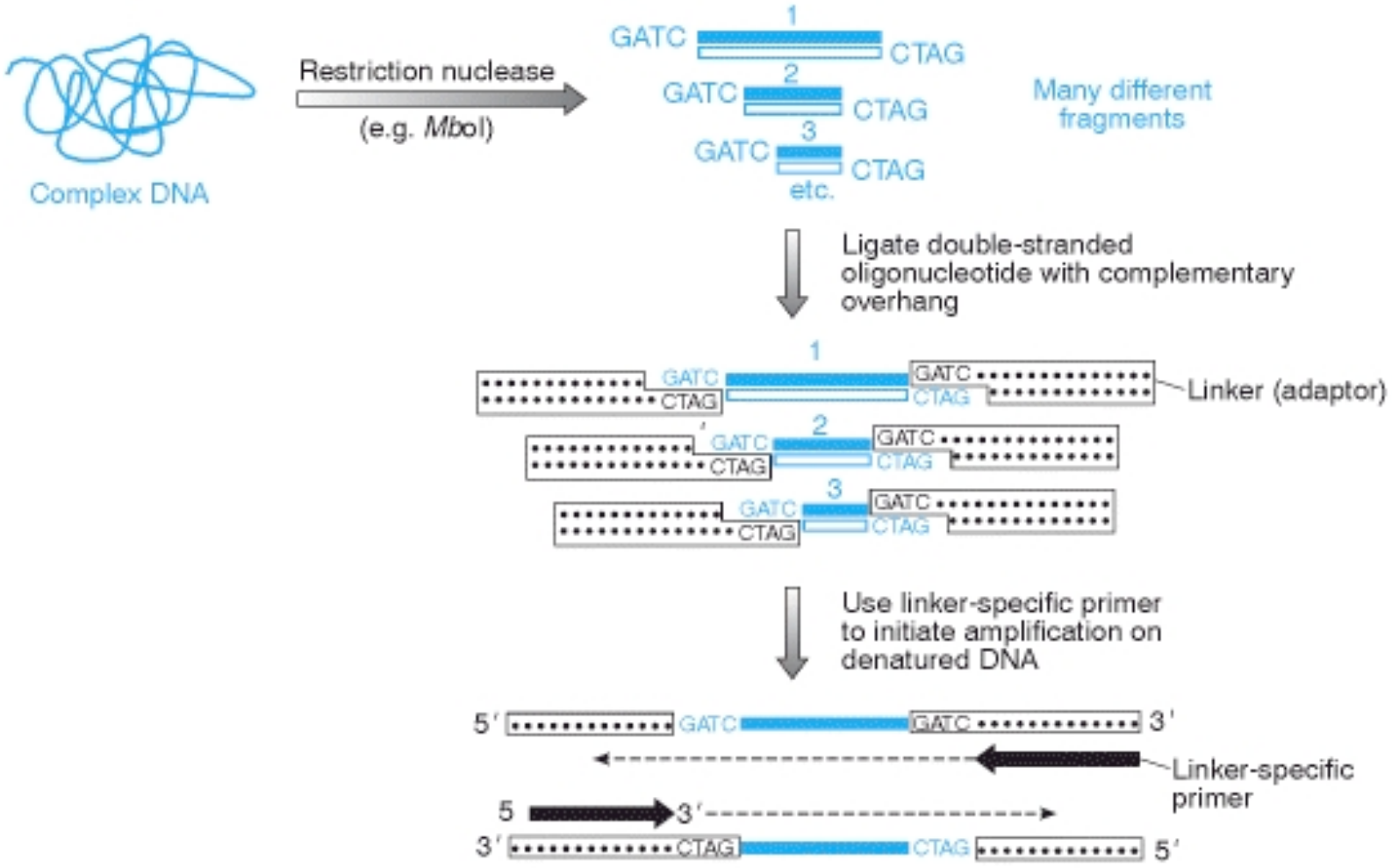
Take a walk into the unknown

- Chromosome walking is a technique for cloning everything in the genome around a known piece of DNA (the starting probe)

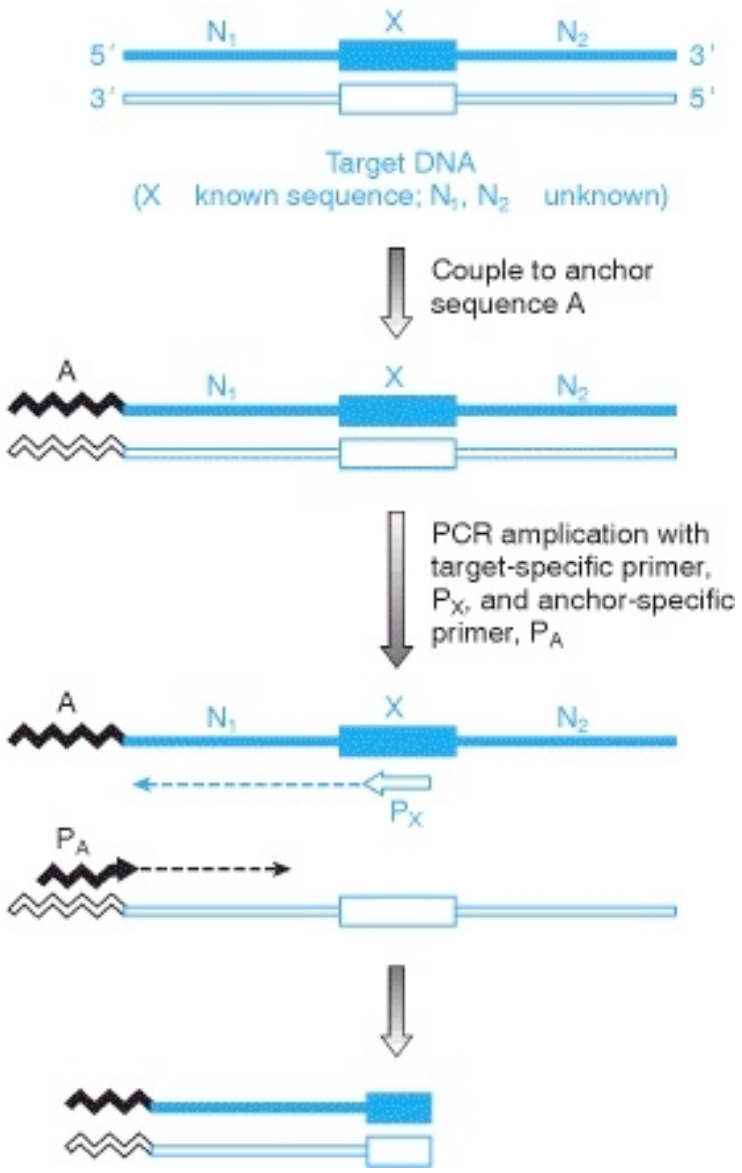


- Using defined unique primers to your sequence begin walking upstream and downstream of your endophyte DHFR DNA

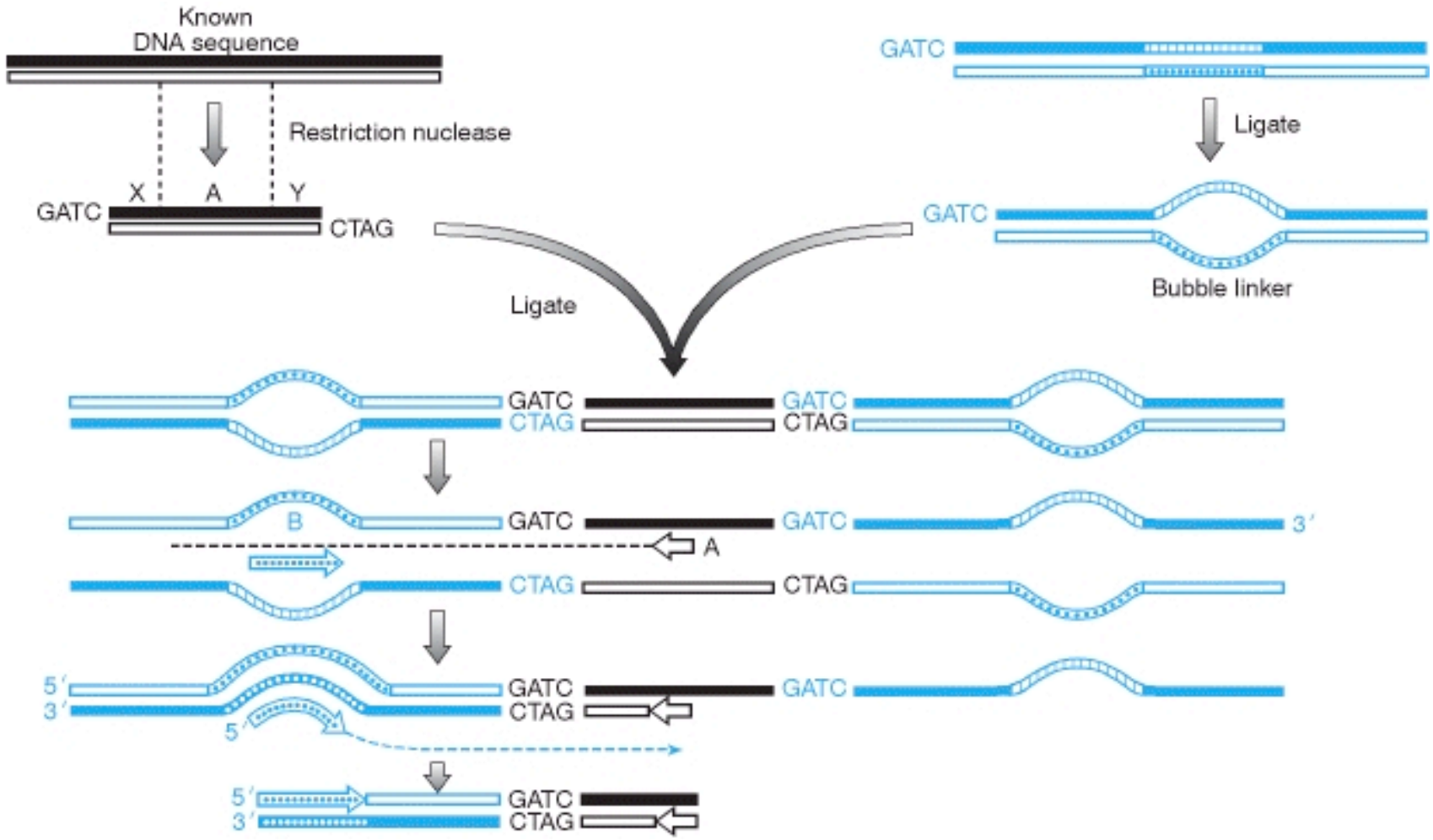
Construct Linker Library



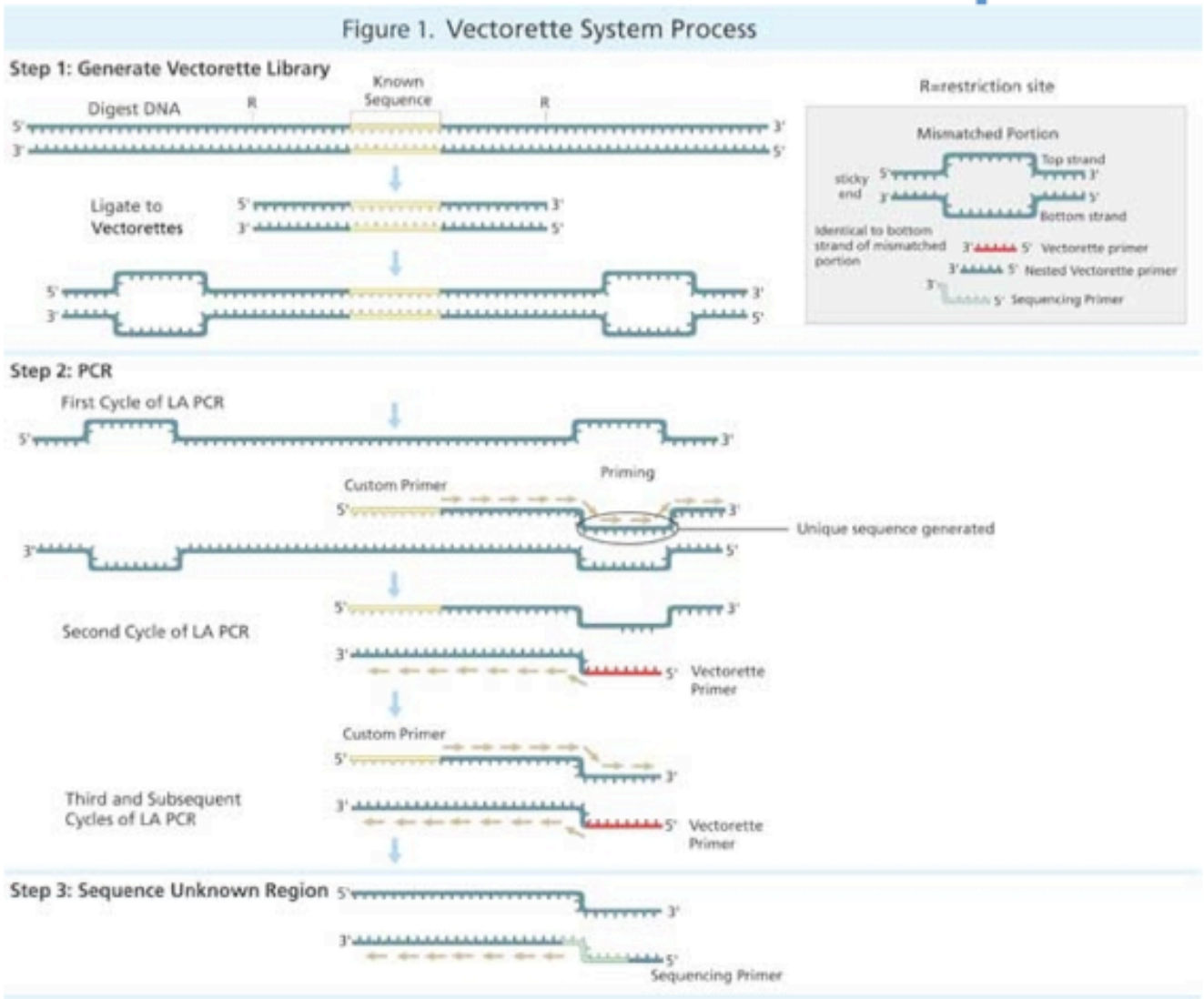
Anchored PCR



Bubble Linker

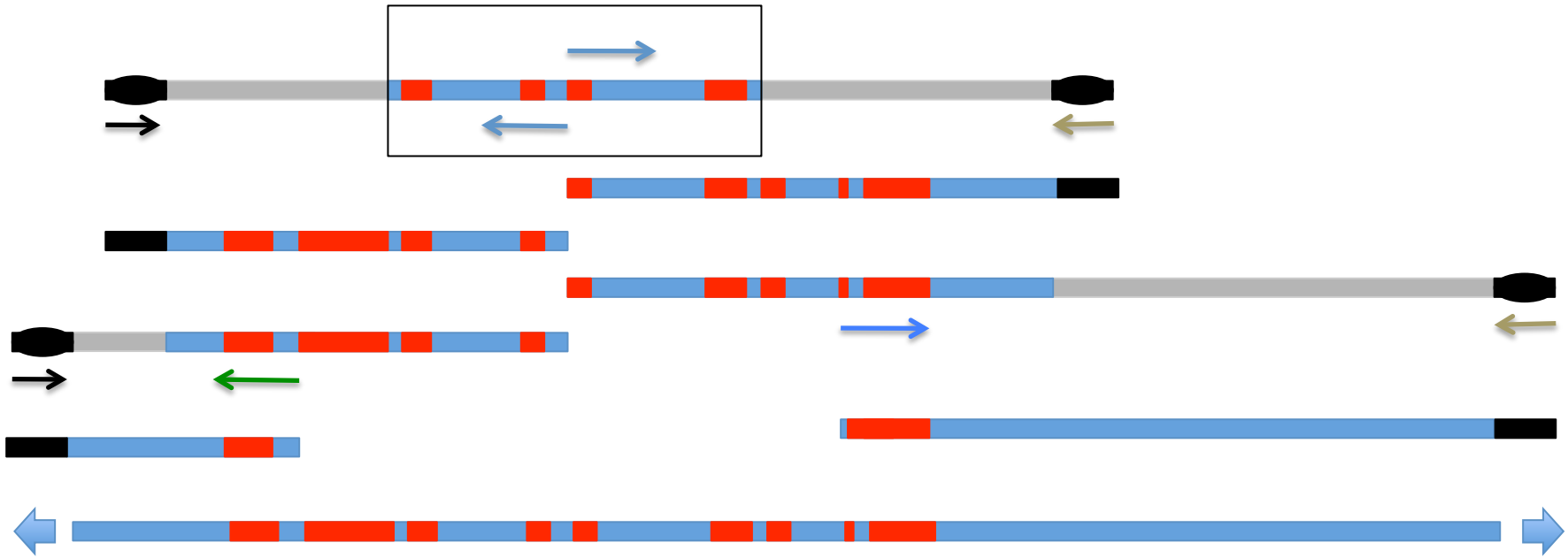


Vectorette – "Bubble" Adapter

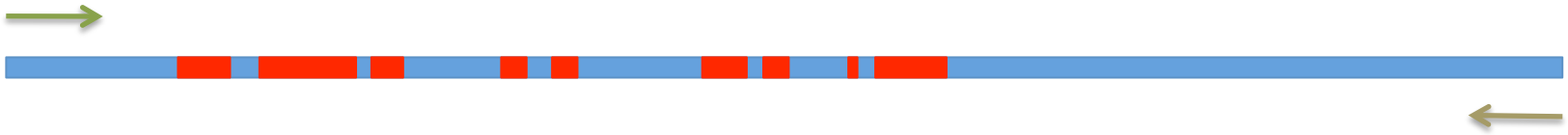


Analysis

- This process is repeated using new specific primers until you have recovered the entire coding region



Analysis



- Now you have the entire coding region of the endophyte DHFR gene
- Conduct a phylogenetic analysis to find its nearest neighbor
- Design primers (to non-conserved regions) to specifically screen for the presence of this endophyte in other host species