Science Applications International Corporation
An Employee-Owned Company

Rapid Design and Selection of PCR Primers for Advanced DNA Analysis

February 3, 2004

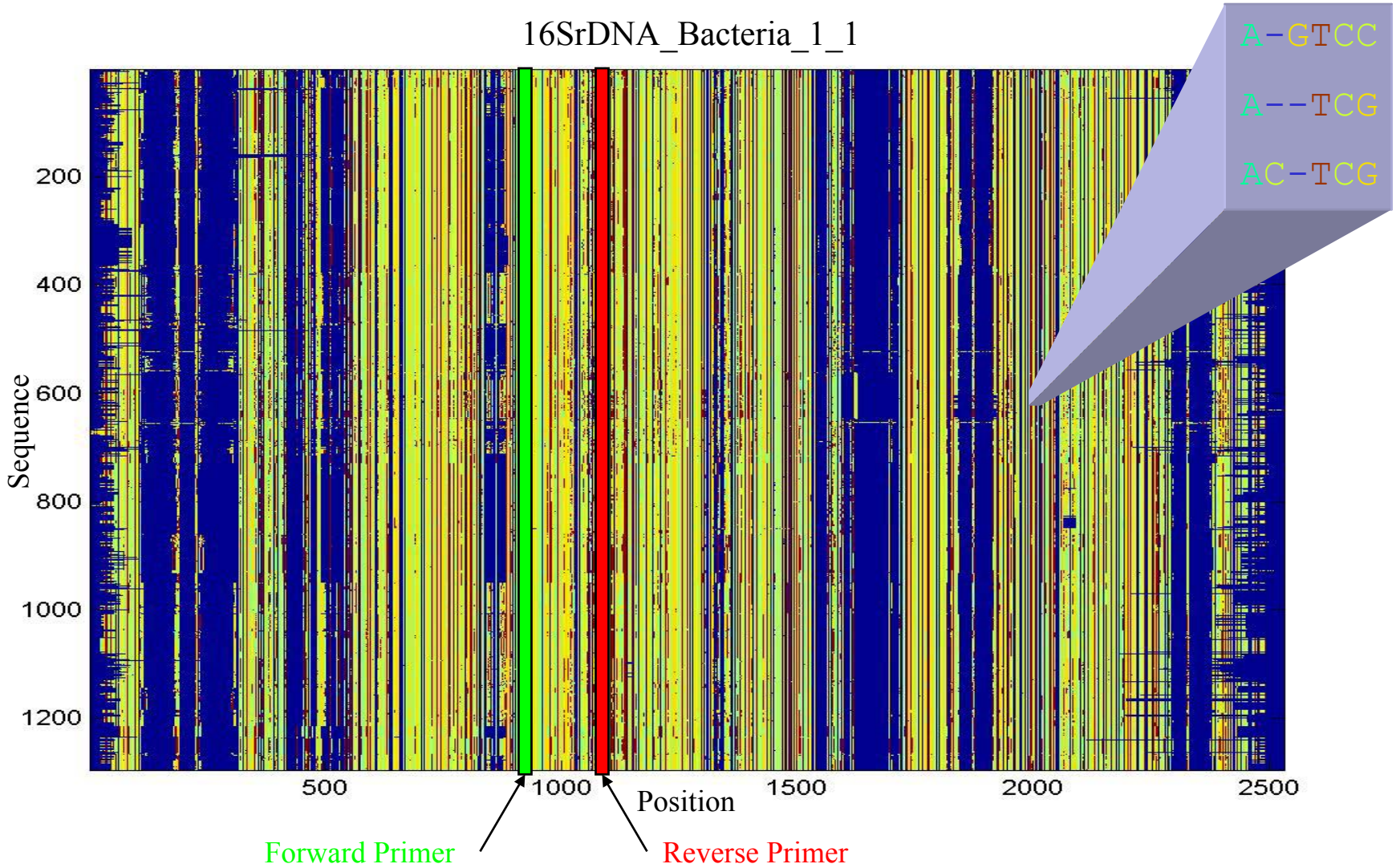
Roland Stoughton – GHC
Cliff Lewis – SAIC
Brons Larson - SAIC

Outline

- **Goal:**
  - **Develop an economical process for searching genomes for two highly similar regions, each of which can be significantly covered by a limited number of primer pairs, yet flanking a highly variable region.**
  - **Focus on high coverage regions in order to capture and classify unknown or unsequenced bacteria, viruses or toxins.**
  - **Find amplicons that can be differentiated, minding the resolution constraints of the developed sensor system**
    - FTICR mass spec
    - TOF mass spec
    - Microarray chip
    - Capillary Gel Electrophoresis
- **Two approaches**
  - **Alignment based**
  - **Non-alignment based**

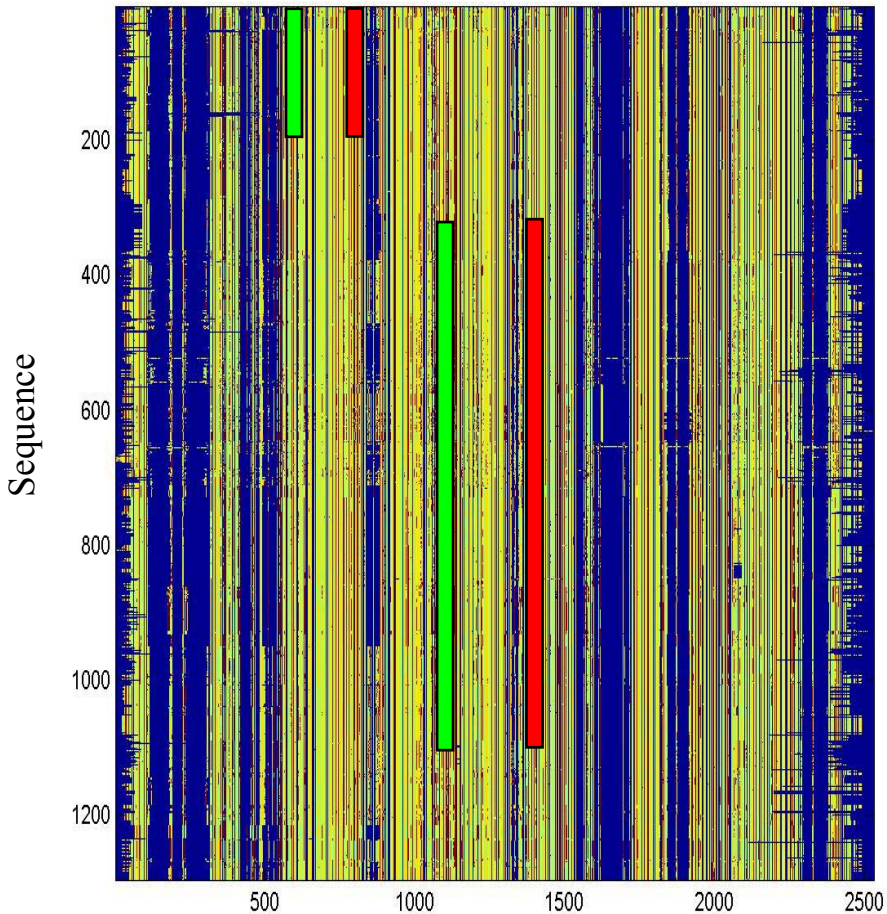
Funded under SAIC IR&D

Alignment Approach

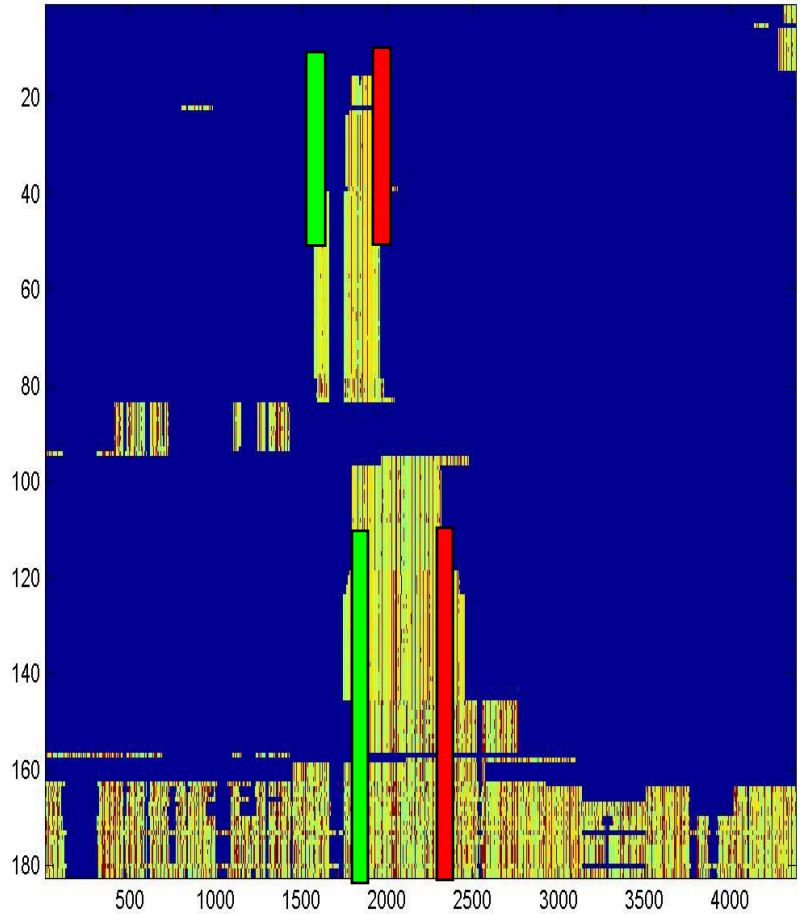


Multiple Alignments

16SrDNA_Bacteria

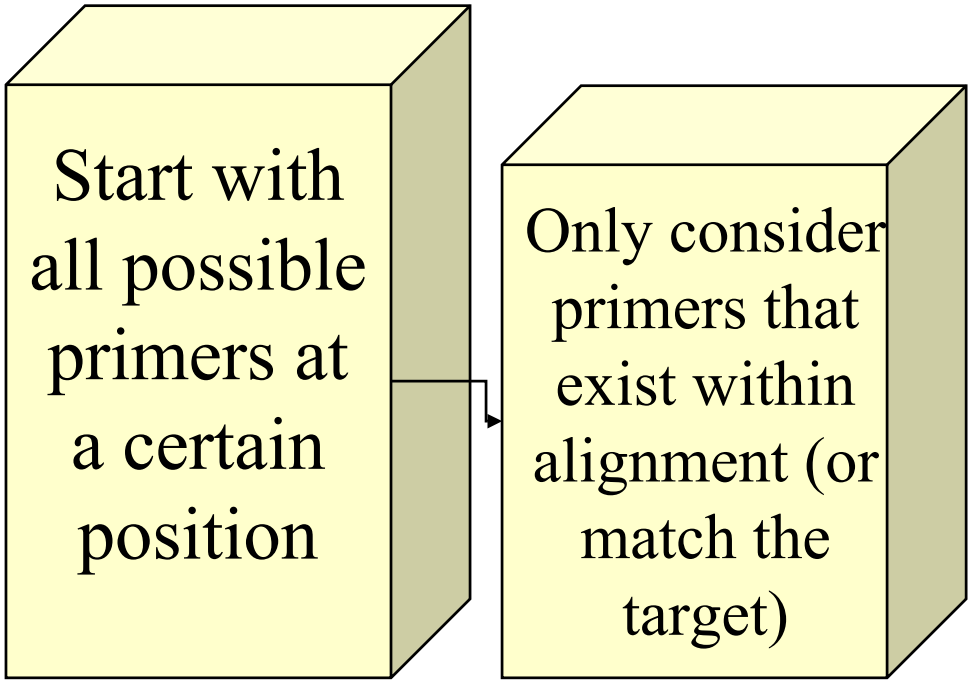


polB



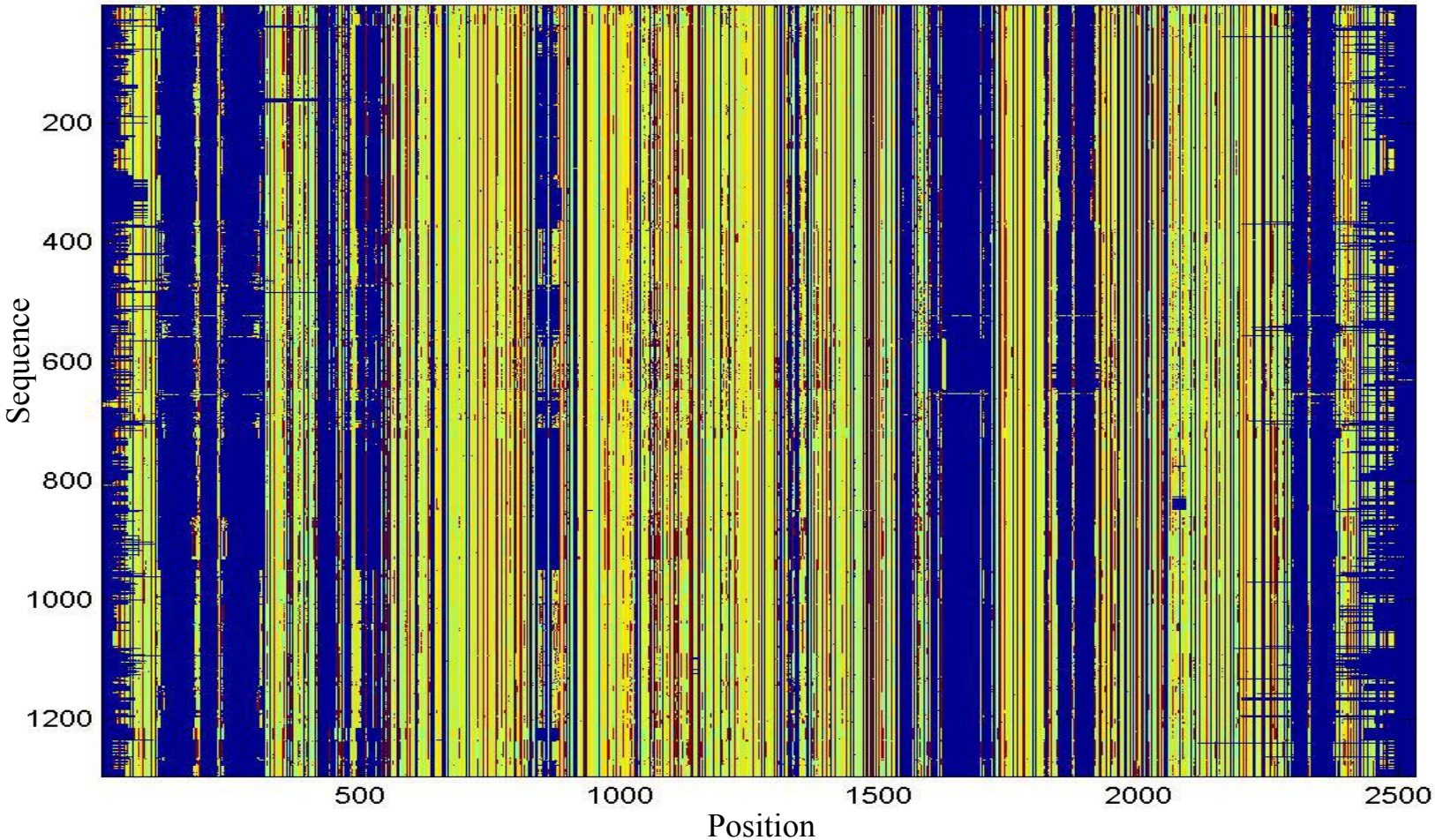
Multiple primer pairs over multiple regions

Primer Selection Outline

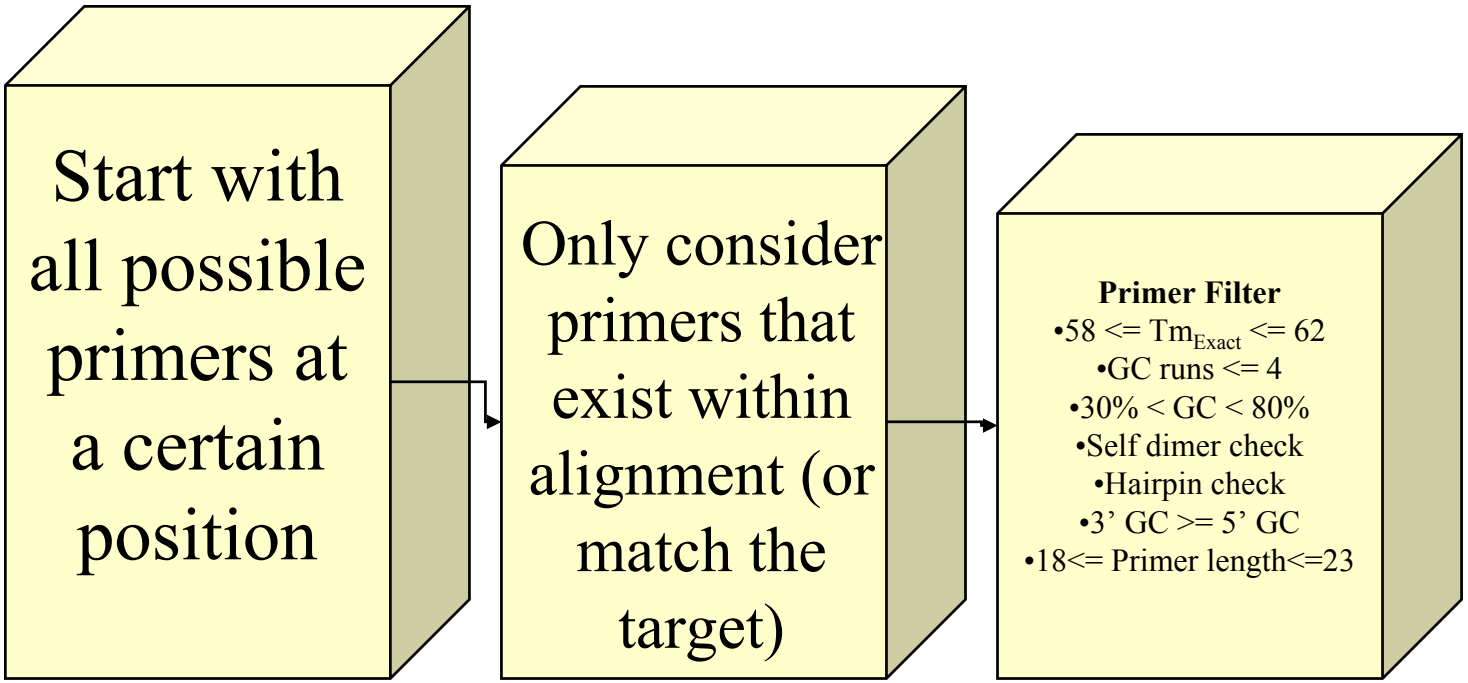


Candidate Primer Set

16SrDNA_Bacteria_1_1

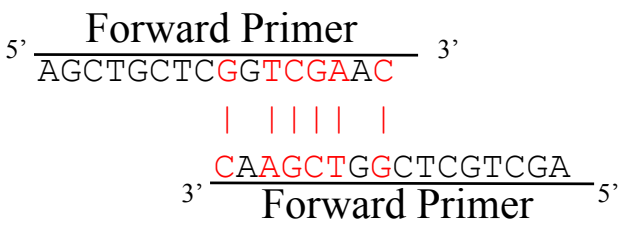


Primer Selection Outline

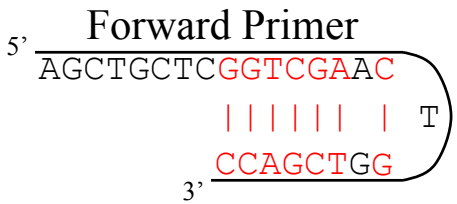


Primer Filter

- **Self dimerization**
  - **Two forward primers**



  - **Hair-pinning**



- **58 <= Tm_{Exact} <= 62**
- **30% < GC < 80%**
- **18 < primer length < 23**
- **3' GC >= 5' GC**

- **Limit CG runs <= 4**

Primer Selection Outline

Start with all possible primers at a certain position

Only consider primers that exist within alignment (or match the target)

Primer Filter
- 58 <= Tm_{Exact} <= 62
- GC runs <= 4
- 30% < GC < 80%
- Self dimer checks
- Hairpin check
- 3' GC >= 5' GC
- 18<= Primer length<=23

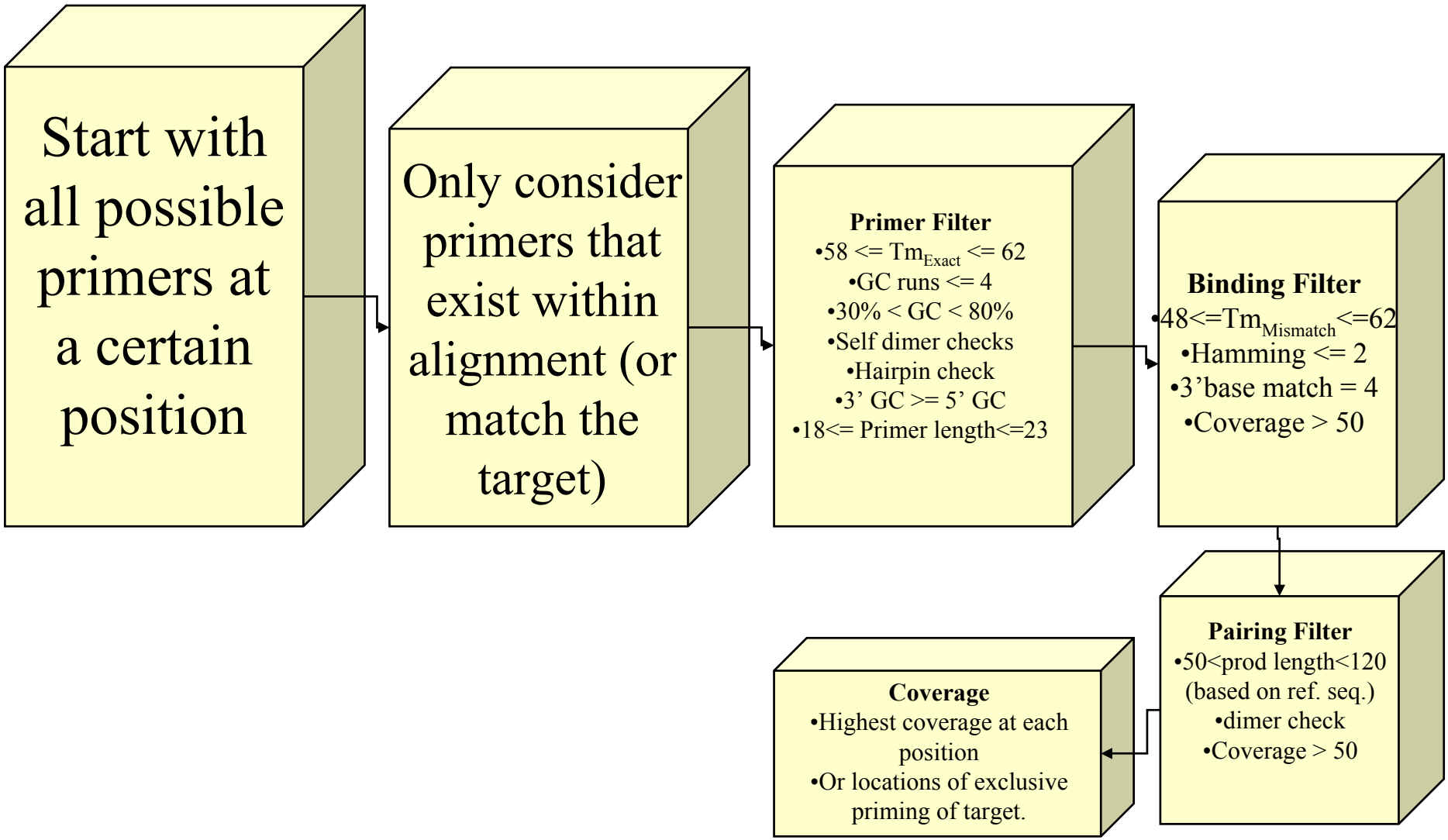
Binding Filter
- 48<=$Tm_{Mismatch}$<=62
- Hamming <= 2
- 3'base match = 4
- Coverage > 50

Binding Filter

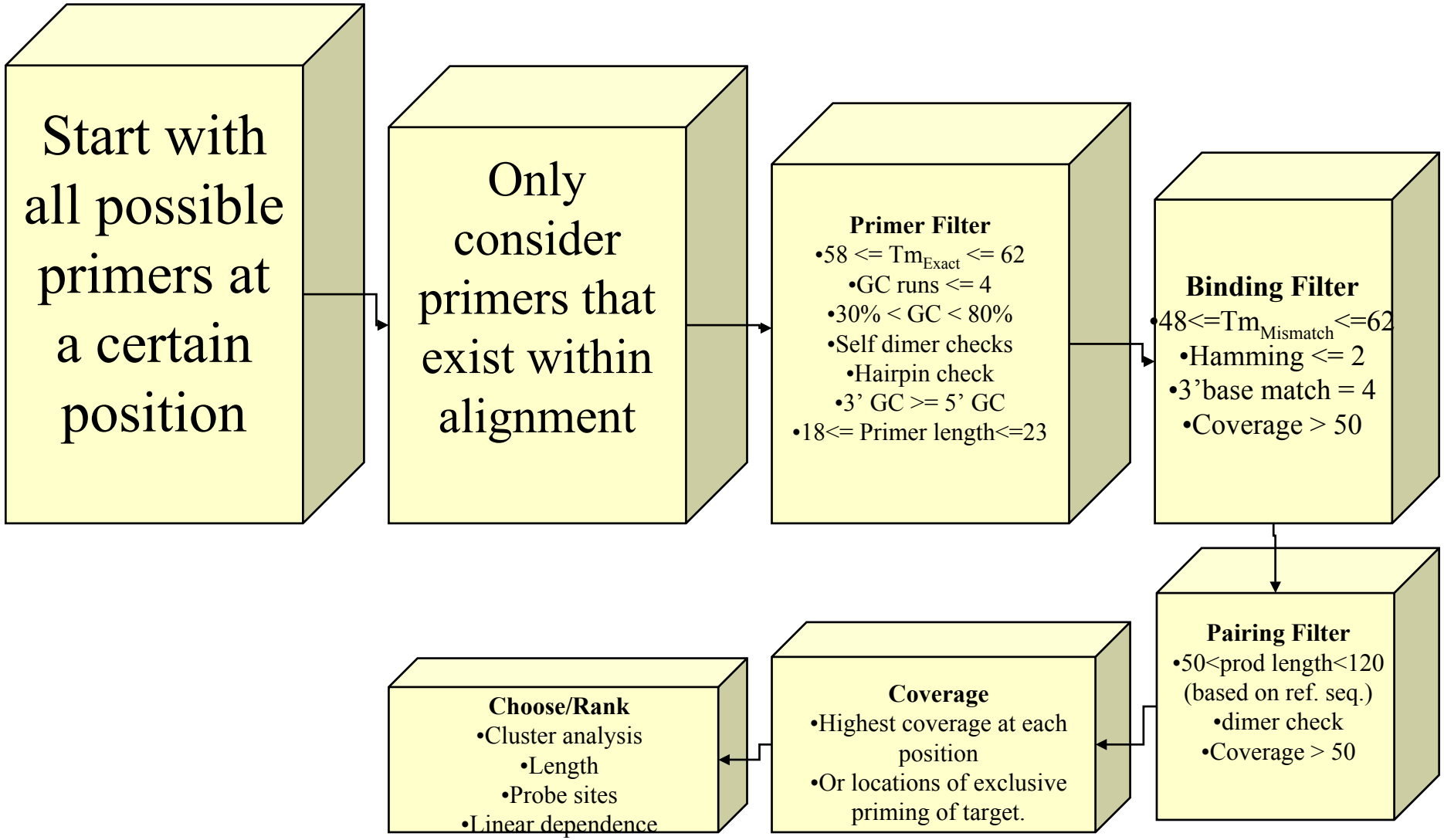
- 48 <= $Tm_{Mismatch}$ <= 62 (with weighting of 3' or 5' end)
- Coverage > 50
- Hamming distance <= 2
  - No mismatch in last 4 bases of the 3' end

```
5'        Forward Primer        3'
   TTACGATTCGAAGCTACGCG
   |||||||  |||||||||||  |
AGCTGCTCGAATGCTAGGCTTCGATGCACTACAACTAC
3'                Sequence                5'
```

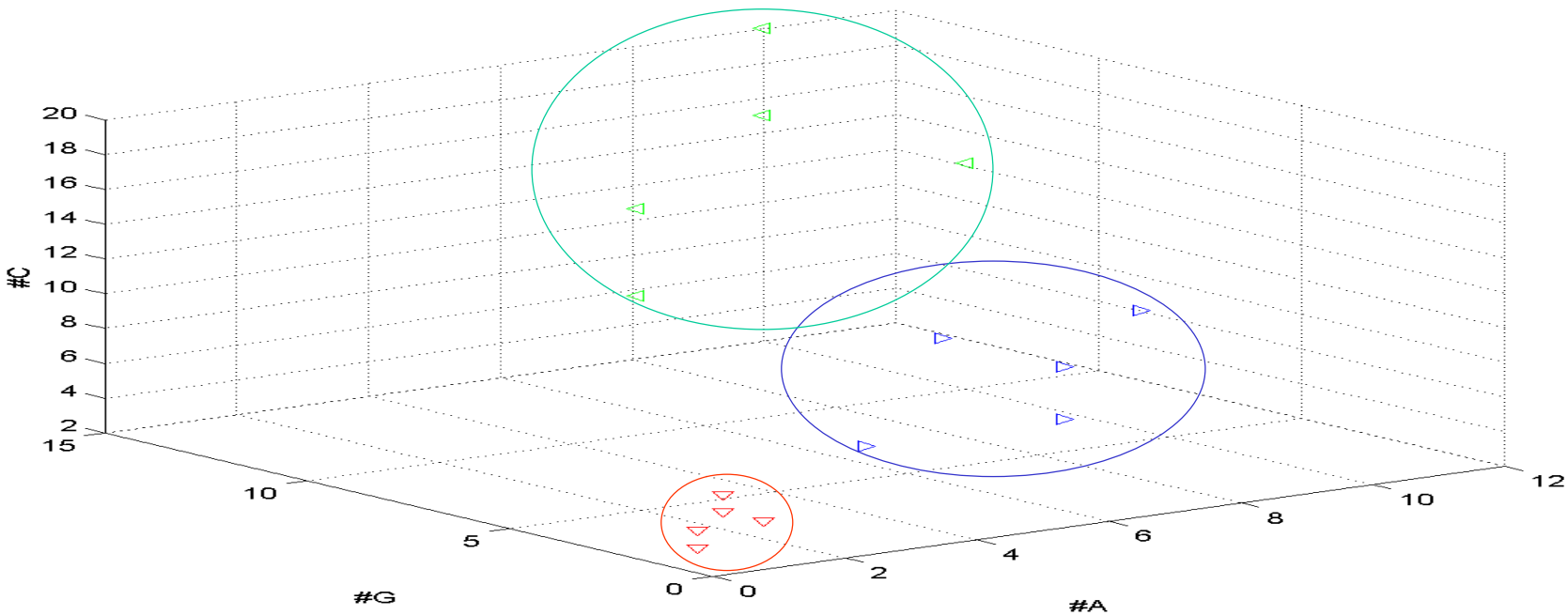
Primer Selection Outline



Primer Selection Outline

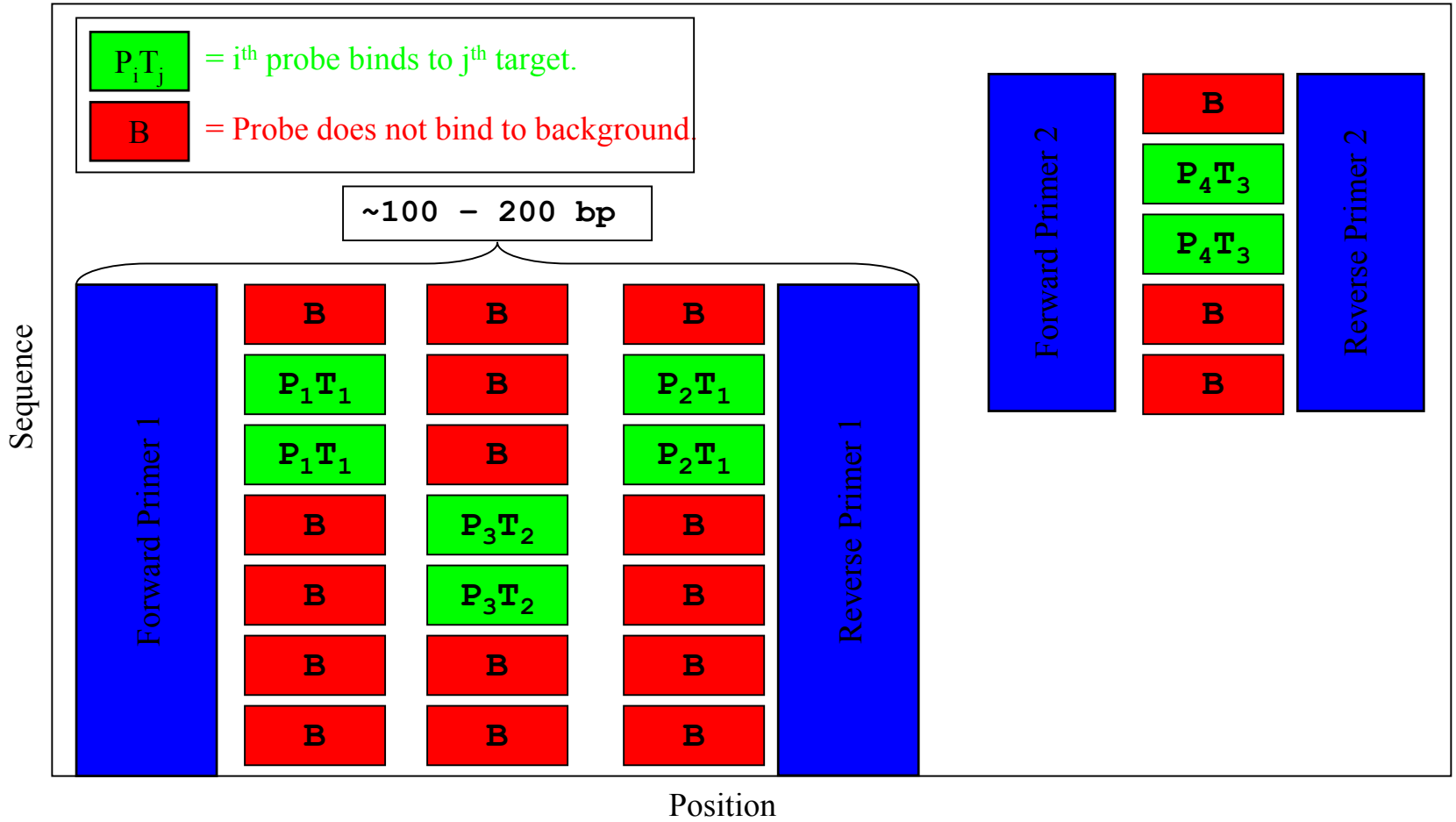


Resolution – Mass Spec Detector



- **Resolution:**
  - **Base count space,**
    - **Probability of mutation cloud,**
    - **Euclidian distance,**
  - **Mass space (similar masses can confuse the maximum likelihood processor).**

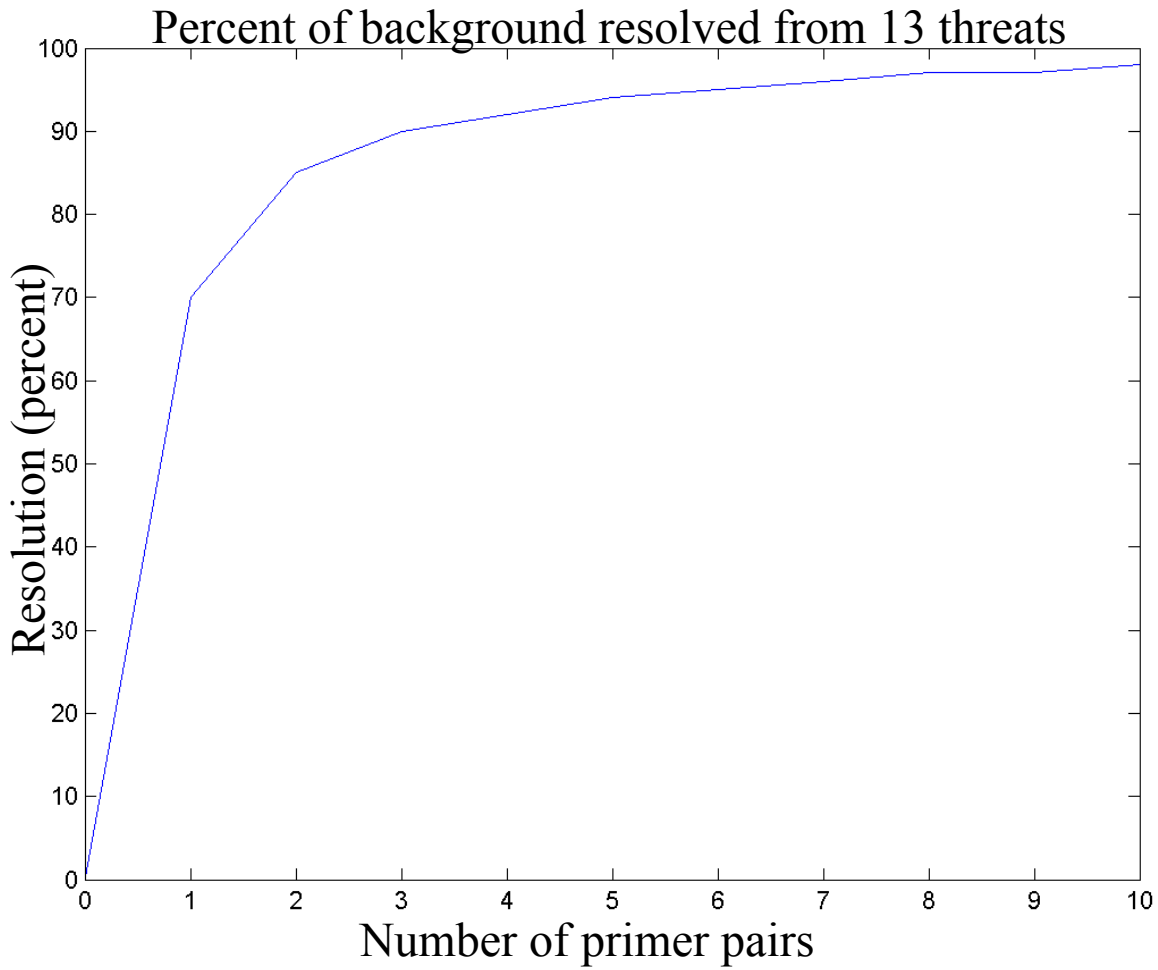
Resolution – Microarray Chips



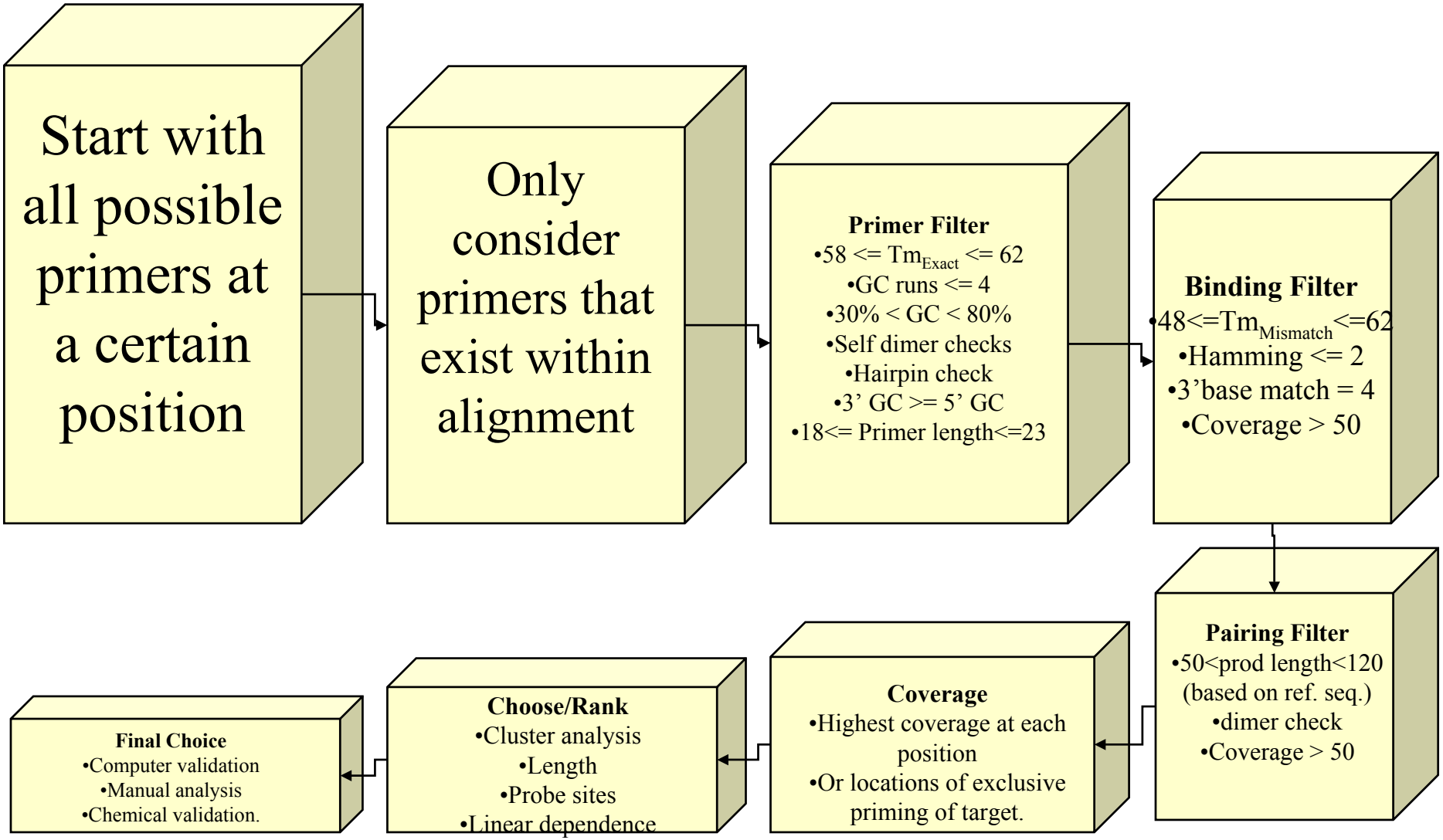
- **Current implementation finds primer pairs which:**
  - **Maximize the number of discriminating probes per target,**
  - **Maximize the number of discriminated targets.**

Resolution – Bioanalyzer Detector

- Use length as discriminating metric.



Primer Selection Outline



Science Applications International Corporation
An Employee-Owned Company

Primer Design Tool (PDT)

New design has more options: adenylation reduction, length and chip resolution, etc.

Primer Design Tool

Help

Use Input File: D:\work\SIMBAD\PDT_CM\INPUT\pdtl — Browse

Save Input File — Browse

ALIGNMENT SELECTION

Bacterial / Viral

Available Alignments: grpE_1_nuclaln_phy, gyrA_1_nuclaln_phy, gyrB_1_nuclaln_phy, hemK_1_nuclaln_phy, hemN_1_nuclaln_phy, hflB_1_102802, hflB_1_nuclaln_phy, himA_1_nuclaln_phy, hisS_1_nuclaln_phy, holA_1_nuclaln_phy, ileS_1_nuclaln_phy, infA_1_nuclaln_phy, infB_1_102802, infB_1_nuclaln_phy, infC_1_nuclaln_phy, ksgA_1_nuclaln_phy, lepA_1_nuclaln_phy, leuS_1_nuclaln_phy, lgt_1_nuclaln_phy, lig_1_nuclaln_phy, lspA_1_nuclaln_phy

Add ->

<- Remove

Chosen Alignments: 16S_genomes_old_groups_062, 23S_genomes_old_groups_062, hflB_1_102802, infB_1_102802, rpoB_1_102802, rpoC_1_102802, valS_1_102802

INPUTS

Parameter	Value
Min primer length	18
Max primer length	25
Min product length	60
Max product length	120
Na concentration	50
Mg concentration	0
Primer concentration	200
Robustness	1
Min coverage	0.4
Min Tm	50
Max Tm	60
Max # primers	20

Parameter	Value
Cloud threshold	
Euclidian threshold	2
Mass threshold	8
GC Trend	
Hamming distance	
3' match	2
Primer dimer	
GC run	
GC lower bound	
GC upper bound	
Load/Use Primer Set	
Load/Rank Primer Set	

SCENARIO

Target Species List: Bacillus_anthracis, Brucella_melitensis, Burkholderia_mallei, Burkholderia_pseudomallei, Clostridium_botulinum, Coxiella_burnetii, Francisella_tularensis, Rickettsia_prowazekii, Salmonella_enterica_subsp._enterica_serovar_typh, Vibrio_cholerae, Yersinia_pestis, Yersinia_pestis_KIM

<- Add

Remove ->

Available Species: Acidithiobacillus_ferrooxidans, Acidithiobacillus_ferroxidans, Acinetobacter_sp._ADP1, Actinobacillus_actinomycetemcomitans, Agrobacterium_tumefaciens_C58_Cereon, Agrobacterium_tumefaciens_C58_UW, Aquifex_aeolicus, Bacillus_anthracis, Bacillus_cereus_ATCC10987, Bacillus_cereus_ATCC14579, Bacillus_halodurans, Bacillus_subtilis, Bacteroides_forsythus, Bacteroides_fragilis, Bordetella_bronchiseptica, Bordetella_parapertussis, Bordetella_pertussis, Borrelia_burgdorferi, Brucella_melitensis, Buchnera_sp._APS, Burkholderia_fungorum, Burkholderia_mallei, Burkholderia_pseudomallei, Campylobacter_jejuni_NCTC11168

Add ->

<- Remove

Background Species List: Acidithiobacillus_ferrooxidans, Acidithiobacillus_ferroxidans, Acinetobacter_sp._ADP1, Actinobacillus_actinomycetemcomitans, Agrobacterium_tumefaciens_C58_Cereon, Agrobacterium_tumefaciens_C58_UW, Aquifex_aeolicus, Bacillus_anthracis, Bacillus_cereus_ATCC10987, Bacillus_cereus_ATCC14579, Bacillus_halodurans, Bacillus_subtilis, Bacteroides_forsythus, Bacteroides_fragilis, Bordetella_bronchiseptica, Bordetella_parapertussis, Bordetella_pertussis, Borrelia_burgdorferi, Brucella_melitensis, Buchnera_sp._APS, Burkholderia_fungorum, Burkholderia_mallei, Burkholderia_pseudomallei, Campylobacter_jejuni_NCTC11168

Clear

Cancel

Run

Science Applications International Corporation
An Employee-Owned Company

Protein Problem Cluster

Target

- Escherichia_coli_K12
- Escherichia_coli_O157_H7
- Escherichia_coli_O157_H7_EDL933
- Klebsiella_pneumoniae
- Salmonella_enterica_subsp._enterica_serovar_typhi
- Salmonella_typhimurium_LT2
- Yersinia_enterocolitica
- Yersinia_pestis
- Yersinia_pestis_KIM
- Yersinia_pseudotuberculosis

Background

- Escherichia_coli_K12
- Escherichia_coli_O157_H7
- Escherichia_coli_O157_H7_EDL933
- Klebsiella_pneumoniae
- Salmonella_enterica_subsp._enterica_serovar_typhi
- Salmonella_typhimurium_LT2
- Yersinia_enterocolitica
- Yersinia_pestis
- Yersinia_pestis_KIM
- Yersinia_pseudotuberculosis

Escherichia_coli_K12	- Escherichia_coli_O157_H7
Escherichia_coli_K12	- Escherichia_coli_O157_H7_EDL933
Escherichia_coli_K12	- Klebsiella_pneumoniae
Escherichia_coli_K12	- Salmonella_enterica_subsp._enterica_serovar_typhi
Escherichia_coli_K12	- Salmonella_typhimurium_LT2
Escherichia_coli_K12	- Yersinia_enterocolitica
Escherichia_coli_K12	- Yersinia_pestis
Escherichia_coli_K12	- Yersinia_pestis_KIM
Escherichia_coli_K12	- Yersinia_pseudotuberculosis
Escherichia_coli_O157_H7	- Escherichia_coli_O157_H7_EDL933
Escherichia_coli_O157_H7	- Klebsiella_pneumoniae
Escherichia_coli_O157_H7	- Salmonella_enterica_subsp._enterica_serovar_typhi
Escherichia_coli_O157_H7	- Salmonella_typhimurium_LT2
Escherichia_coli_O157_H7	- Yersinia_enterocolitica
Escherichia_coli_O157_H7	- Yersinia_pestis
Escherichia_coli_O157_H7	- Yersinia_pestis_KIM
Escherichia_coli_O157_H7	- Yersinia_pseudotuberculosis
Escherichia_coli_O157_H7_EDL933	- Klebsiella_pneumoniae
Escherichia_coli_O157_H7_EDL933	- Salmonella_enterica_subsp._enterica_serovar_typhi
Escherichia_coli_O157_H7_EDL933	- Salmonella_typhimurium_LT2
Escherichia_coli_O157_H7_EDL933	- Yersinia_enterocolitica
Escherichia_coli_O157_H7_EDL933	- Yersinia_pestis
Escherichia_coli_O157_H7_EDL933	- Yersinia_pestis_KIM
Escherichia_coli_O157_H7_EDL933	- Yersinia_pseudotuberculosis
Klebsiella_pneumoniae	- Salmonella_enterica_subsp._enterica_serovar_typhi
Klebsiella_pneumoniae	- Salmonella_typhimurium_LT2
Klebsiella_pneumoniae	- Yersinia_enterocolitica
Klebsiella_pneumoniae	- Yersinia_pestis
Klebsiella_pneumoniae	- Yersinia_pestis_KIM
Klebsiella_pneumoniae	- Yersinia_pseudotuberculosis
Salmonella_enterica_subsp._enterica_serovar_typhi	- Salmonella_typhimurium_LT2
Salmonella_enterica_subsp._enterica_serovar_typhi	- Yersinia_enterocolitica
Salmonella_enterica_subsp._enterica_serovar_typhi	- Yersinia_pestis
Salmonella_enterica_subsp._enterica_serovar_typhi	- Yersinia_pestis_KIM
Salmonella_enterica_subsp._enterica_serovar_typhi	- Yersinia_pseudotuberculosis
Salmonella_typhimurium_LT2	- Yersinia_enterocolitica
Salmonella_typhimurium_LT2	- Yersinia_pestis
Salmonella_typhimurium_LT2	- Yersinia_pestis_KIM
Salmonella_typhimurium_LT2	- Yersinia_pseudotuberculosis
Yersinia_enterocolitica	- Yersinia_pestis
Yersinia_enterocolitica	- Yersinia_pestis_KIM

Science Applications International Corporation
An Employee-Owned Company

Ribosome resolution

Target

- Escherichia_coli_K12
- Escherichia_coli_O157_H7
- Escherichia_coli_O157_H7_EDL933
- Klebsiella_pneumoniae
- Salmonella_enterica_subsp._enterica_serovar_typhi
- Salmonella_typhimurium_LT2
- Yersinia_enterocolitica
- Yersinia_pestis
- Yersinia_pestis_KIM
- Yersinia_pseudotuberculosis

Background

- Escherichia_coli_K12
- Escherichia_coli_O157_H7
- Escherichia_coli_O157_H7_EDL933
- Klebsiella_pneumoniae
- Salmonella_enterica_subsp._enterica_serovar_typhi
- Salmonella_typhimurium_LT2
- Yersinia_enterocolitica
- Yersinia_pestis
- Yersinia_pestis_KIM
- Yersinia_pseudotuberculosis

- **Resolved by ribosomal primers.**

Escherichia_coli_K12	- Escherichia_coli_O157_H7
Escherichia_coli_K12	- Escherichia_coli_O157_H7_EDL933
Escherichia_coli_K12	- Klebsiella_pneumoniae
Escherichia_coli_K12	- Salmonella_enterica_subsp._enterica_serovar_typhi
Escherichia_coli_K12	- Salmonella_typhimurium_LT2
Escherichia_coli_K12	- Yersinia_enterocolitica
Escherichia_coli_K12	- Yersinia_pestis
Escherichia_coli_K12	- Yersinia_pestis_KIM
Escherichia_coli_K12	- Yersinia_pseudotuberculosis
Escherichia_coli_O157_H7	- Escherichia_coli_O157_H7_EDL933
Escherichia_coli_O157_H7	- Klebsiella_pneumoniae
Escherichia_coli_O157_H7	- Salmonella_enterica_subsp._enterica_serovar_typhi
Escherichia_coli_O157_H7	- Salmonella_typhimurium_LT2
Escherichia_coli_O157_H7	- Yersinia_enterocolitica
Escherichia_coli_O157_H7	- Yersinia_pestis
Escherichia_coli_O157_H7	- Yersinia_pestis_KIM
Escherichia_coli_O157_H7	- Yersinia_pseudotuberculosis
Escherichia_coli_O157_H7_EDL933	- Klebsiella_pneumoniae
Escherichia_coli_O157_H7_EDL933	- Salmonella_enterica_subsp._enterica_serovar_typhi
Escherichia_coli_O157_H7_EDL933	- Salmonella_typhimurium_LT2
Escherichia_coli_O157_H7_EDL933	- Yersinia_enterocolitica
Escherichia_coli_O157_H7_EDL933	- Yersinia_pestis
Escherichia_coli_O157_H7_EDL933	- Yersinia_pestis_KIM
Escherichia_coli_O157_H7_EDL933	- Yersinia_pseudotuberculosis
Klebsiella_pneumoniae	- Salmonella_enterica_subsp._enterica_serovar_typhi
Klebsiella_pneumoniae	- Salmonella_typhimurium_LT2
Klebsiella_pneumoniae	- Yersinia_enterocolitica
Klebsiella_pneumoniae	- Yersinia_pestis
Klebsiella_pneumoniae	- Yersinia_pestis_KIM
Klebsiella_pneumoniae	- Yersinia_pseudotuberculosis
Salmonella_enterica_subsp._enterica_serovar_typhi	- Salmonella_typhimurium_LT2
Salmonella_enterica_subsp._enterica_serovar_typhi	- Yersinia_enterocolitica
Salmonella_enterica_subsp._enterica_serovar_typhi	- Yersinia_pestis
Salmonella_enterica_subsp._enterica_serovar_typhi	- Yersinia_pestis_KIM
Salmonella_enterica_subsp._enterica_serovar_typhi	- Yersinia_pseudotuberculosis
Salmonella_typhimurium_LT2	- Yersinia_enterocolitica
Salmonella_typhimurium_LT2	- Yersinia_pestis
Salmonella_typhimurium_LT2	- Yersinia_pestis_KIM
Salmonella_typhimurium_LT2	- Yersinia_pseudotuberculosis
Yersinia_enterocolitica	- Yersinia_pestis
Yersinia_enterocolitica	- Yersinia_pestis_KIM

Science Applications International Corporation
An Employee-Owned Company

Protein Resolution

Target

- Escherichia_coli_K12
- Escherichia_coli_O157_H7
- Escherichia_coli_O157_H7_EDL933
- Klebsiella_pneumoniae
- Salmonella_enterica_subsp._enterica_serovar_typhi
- Salmonella_typhimurium_LT2
- Yersinia_enterocolitica
- Yersinia_pestis
- Yersinia_pestis_KIM
- Yersinia_pseudotuberculosis

Background

- Escherichia_coli_K12
- Escherichia_coli_O157_H7
- Escherichia_coli_O157_H7_EDL933
- Klebsiella_pneumoniae
- Salmonella_enterica_subsp._enterica_serovar_typhi
- Salmonella_typhimurium_LT2
- Yersinia_enterocolitica
- Yersinia_pestis
- Yersinia_pestis_KIM
- Yersinia_pseudotuberculosis

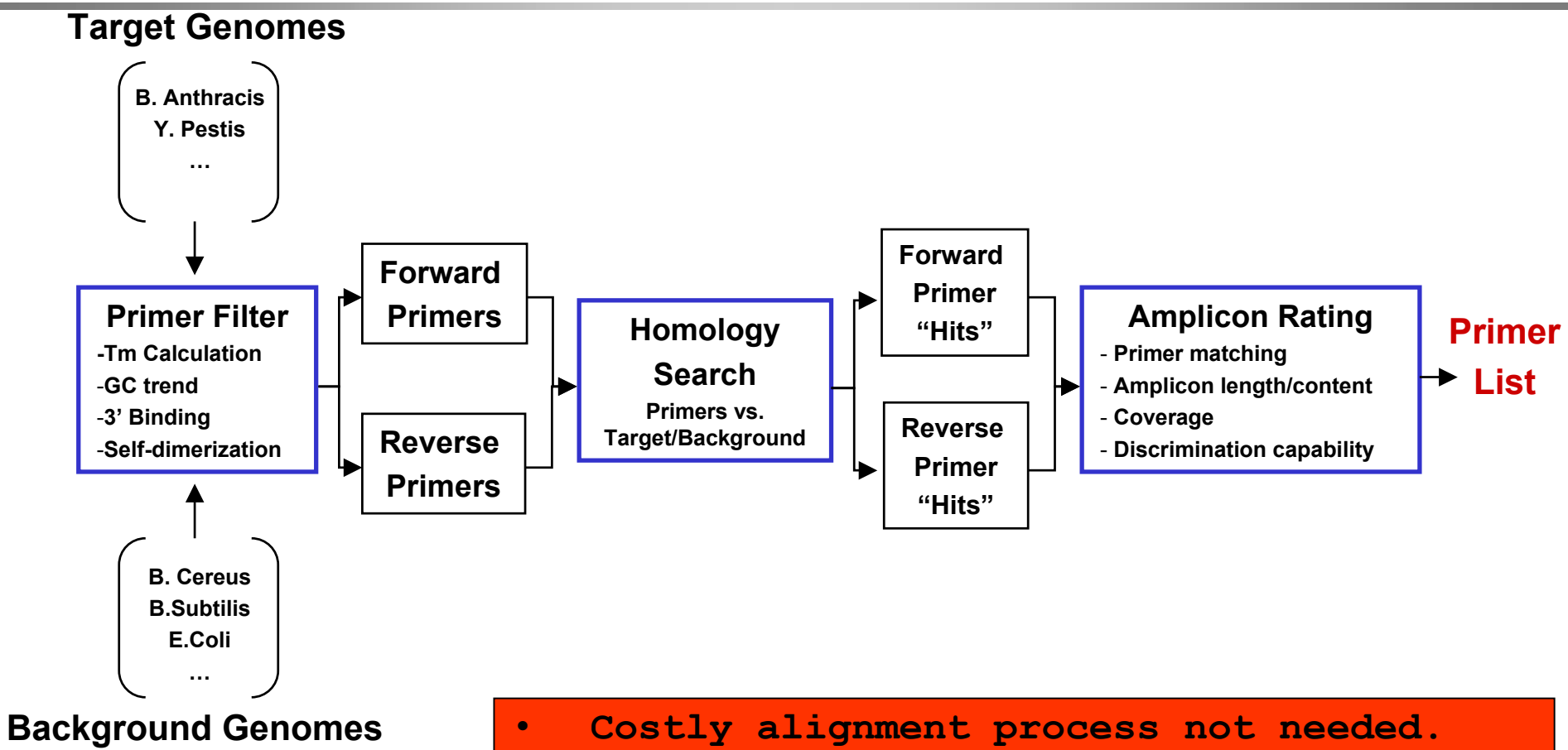
- **Resolved by ribosomal primers.**

- **Resolved by the following protein primers:**
  - **groL**
  - **infB**

Escherichia_coli_K12 - Escherichia_coli_O157_H7
Escherichia_coli_K12 - Escherichia_coli_O157_H7_EDL933
Escherichia_coli_K12 - Klebsiella_pneumoniae
Escherichia_coli_K12 - Salmonella_enterica_subsp._enterica_serovar_typhi
Escherichia_coli_K12 - Salmonella_typhimurium_LT2
Escherichia_coli_K12 - Yersinia_enterocolitica
Escherichia_coli_K12 - Yersinia_pestis
Escherichia_coli_K12 - Yersinia_pestis_KIM
Escherichia_coli_K12 - Yersinia_pseudotuberculosis
Escherichia_coli_O157_H7 - Escherichia_coli_O157_H7_EDL933
Escherichia_coli_O157_H7 - Klebsiella_pneumoniae
Escherichia_coli_O157_H7 - Salmonella_enterica_subsp._enterica_serovar_typhi
Escherichia_coli_O157_H7 - Salmonella_typhimurium_LT2
Escherichia_coli_O157_H7 - Yersinia_enterocolitica
Escherichia_coli_O157_H7 - Yersinia_pestis
Escherichia_coli_O157_H7 - Yersinia_pestis_KIM
Escherichia_coli_O157_H7 - Yersinia_pseudotuberculosis
Escherichia_coli_O157_H7_EDL933 - Klebsiella_pneumoniae
Escherichia_coli_O157_H7_EDL933 - Salmonella_enterica_subsp._enterica_serovar_typhi
Escherichia_coli_O157_H7_EDL933 - Salmonella_typhimurium_LT2
Escherichia_coli_O157_H7_EDL933 - Yersinia_enterocolitica
Escherichia_coli_O157_H7_EDL933 - Yersinia_pestis
Escherichia_coli_O157_H7_EDL933 - Yersinia_pestis_KIM
Escherichia_coli_O157_H7_EDL933 - Yersinia_pseudotuberculosis
Klebsiella_pneumoniae - Salmonella_enterica_subsp._enterica_serovar_typhi
Klebsiella_pneumoniae - Salmonella_typhimurium_LT2
Klebsiella_pneumoniae - Yersinia_enterocolitica
Klebsiella_pneumoniae - Yersinia_pestis
Klebsiella_pneumoniae - Yersinia_pestis_KIM
Klebsiella_pneumoniae - Yersinia_pseudotuberculosis
Salmonella_enterica_subsp._enterica_serovar_typhi - Salmonella_typhimurium_LT2
Salmonella_enterica_subsp._enterica_serovar_typhi - Yersinia_enterocolitica
Salmonella_enterica_subsp._enterica_serovar_typhi - Yersinia_pestis
Salmonella_enterica_subsp._enterica_serovar_typhi - Yersinia_pestis_KIM
Salmonella_enterica_subsp._enterica_serovar_typhi - Yersinia_pseudotuberculosis
Salmonella_typhimurium_LT2 - Yersinia_enterocolitica
Salmonella_typhimurium_LT2 - Yersinia_pestis
Salmonella_typhimurium_LT2 - Yersinia_pestis_KIM
Salmonella_typhimurium_LT2 - Yersinia_pseudotuberculosis
Yersinia_enterocolitica - Yersinia_pestis
Yersinia_enterocolitica - Yersinia_pestis_KIM

Science Applications International Corporation
An Employee-Owned Company

Whole-Genome, Non-Alignment Based Primer Design



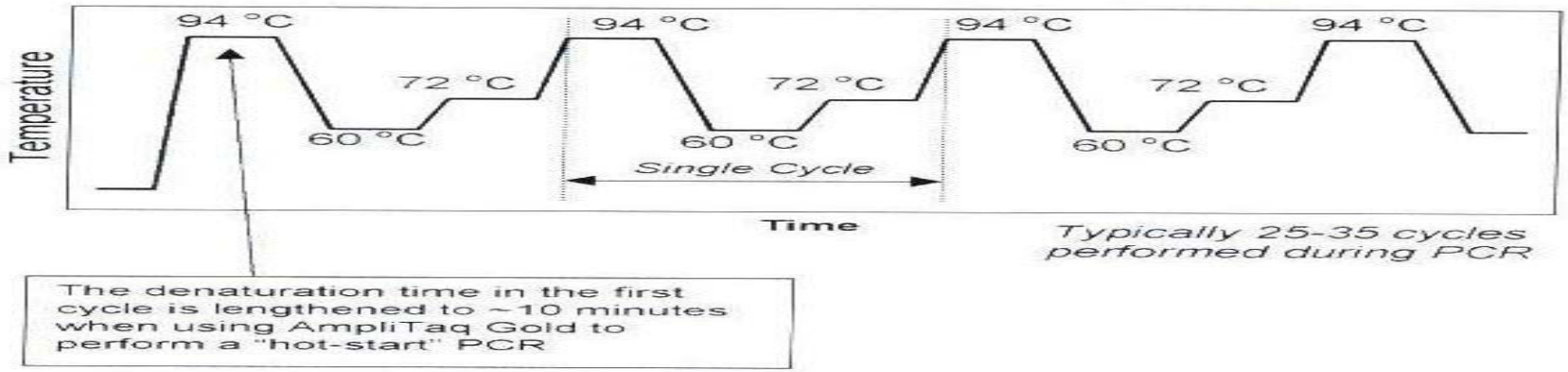
- Costly alignment process not needed.
- Amlicons not limited in length; forward primer may lie in different genes than reverse primer.

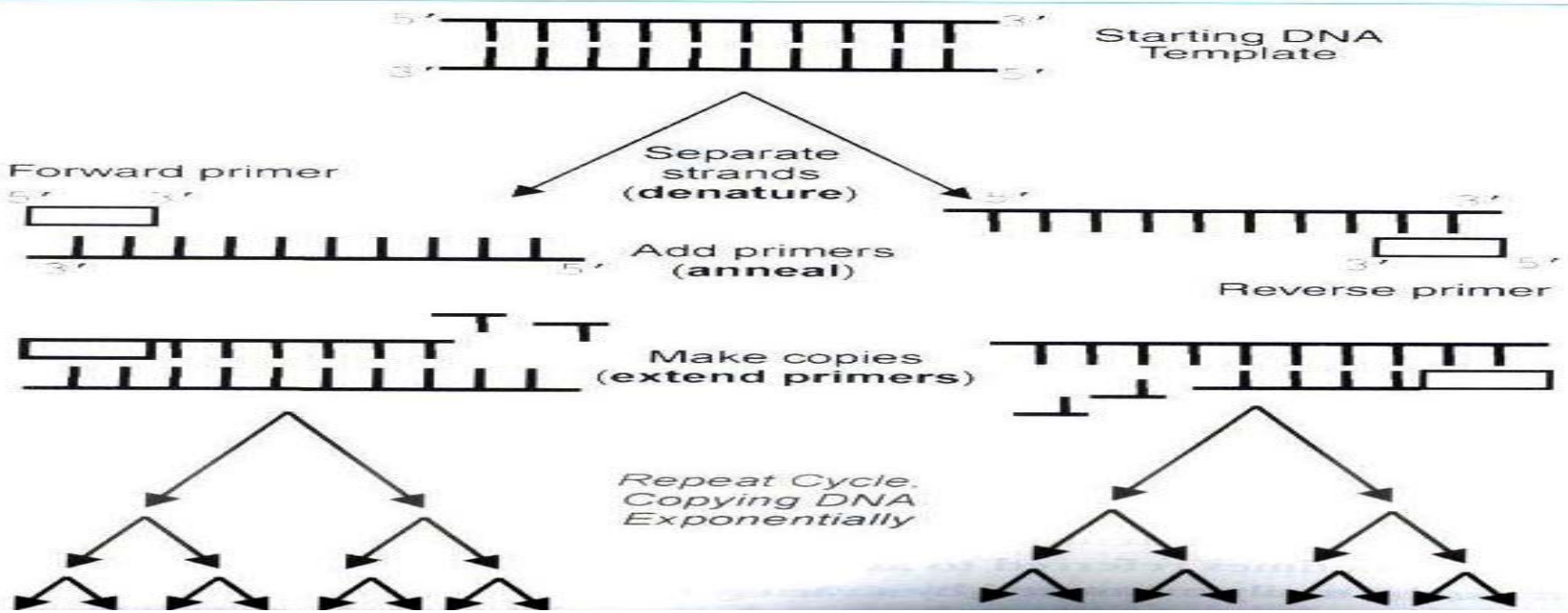
Conclusions

- Current software and algorithms can be easily adapted to find primers for a wide range of applications; e.g., universal primers, exclusive primers.
- We also have non-alignment based primer algorithms and software for finding universal and exclusive primers.
- We have also developed a huge library of utility functions to aid in analysis; e.g.,
  - Tools that allow the evaluation of hand designed primers to be analyzed against any scenario.
  - Tools that modify primers to have certain traits (e.g., tighter Tm between forward and reverse primer pair), while maintaining (optimizing) coverage and resolution for specific scenarios,
  - Visualization tools to analyze results,
- Tools have allowed us to rank the genes that are most informative for specific applications.

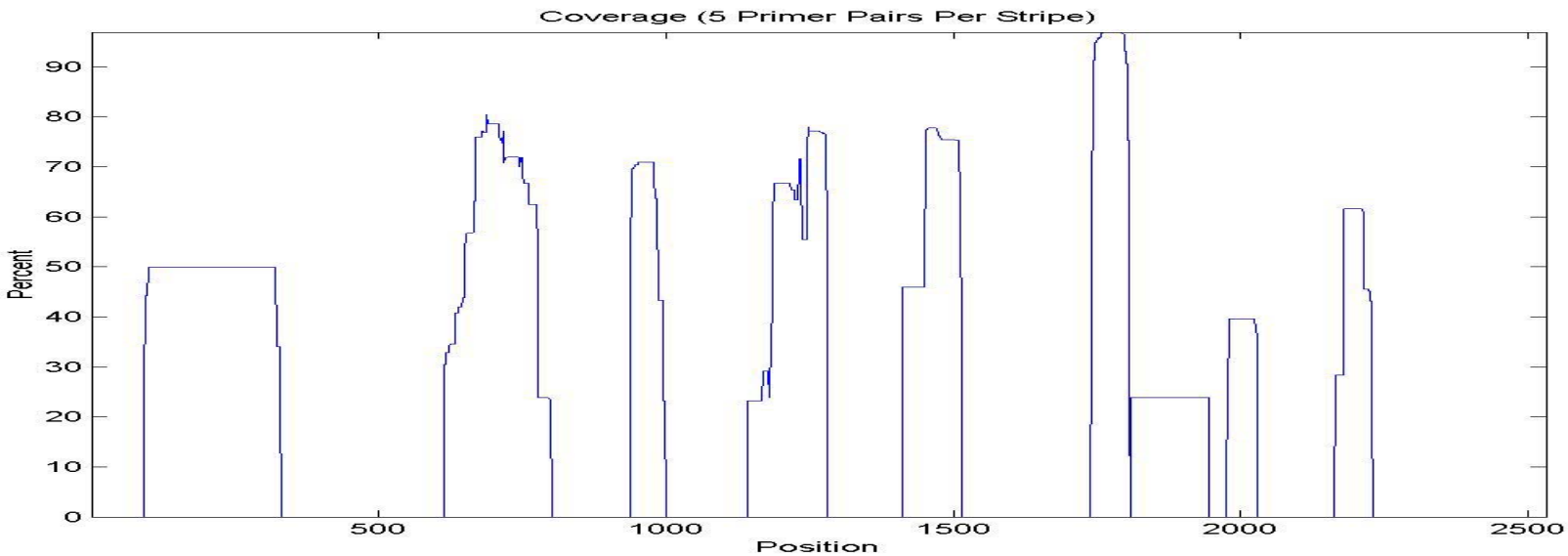
Supplemental Slides

Polymerase Chain Reaction(PCR)





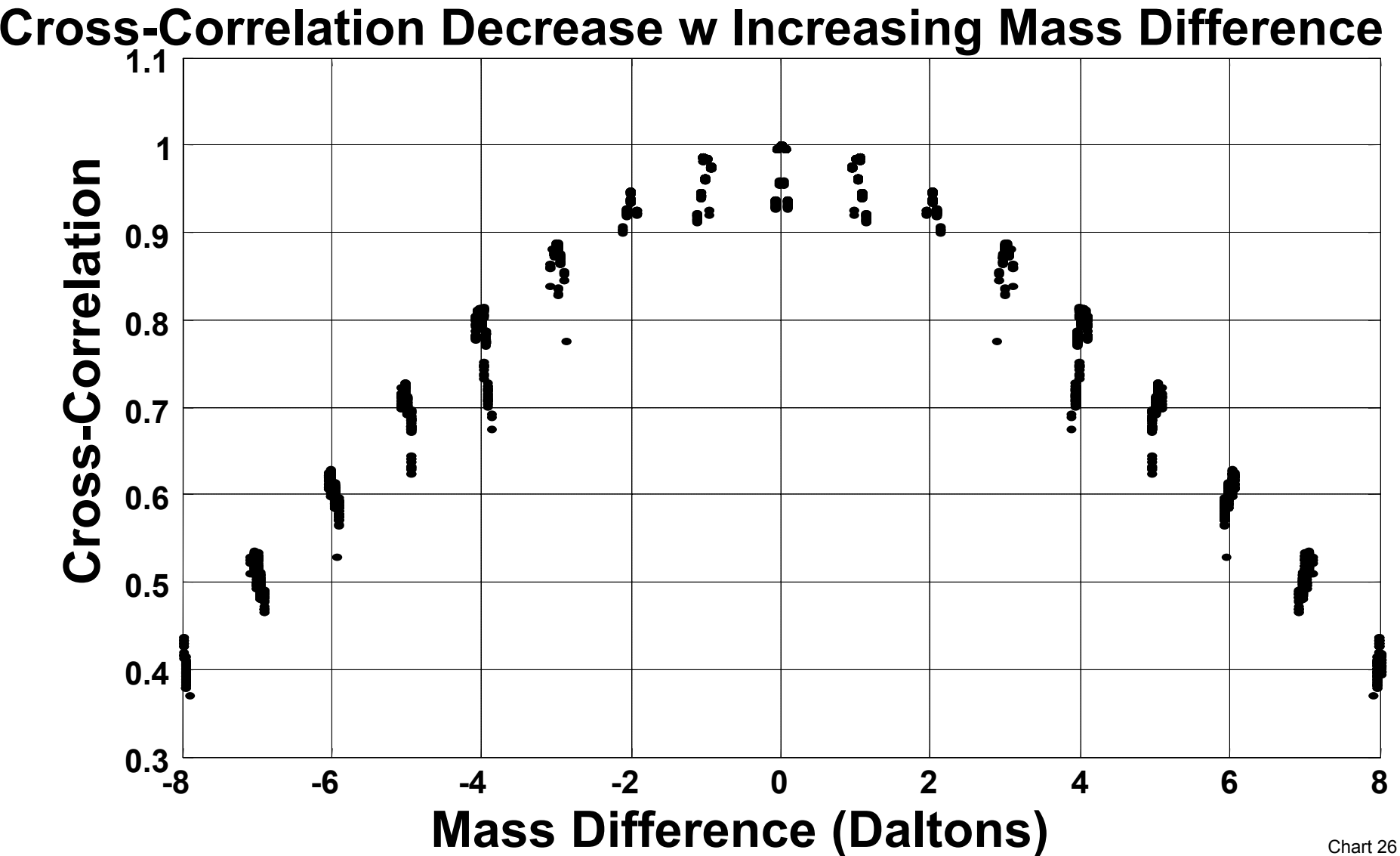
Coverage Of 16SrDNA_Bacteria



Predicted Coverage (previous approach)

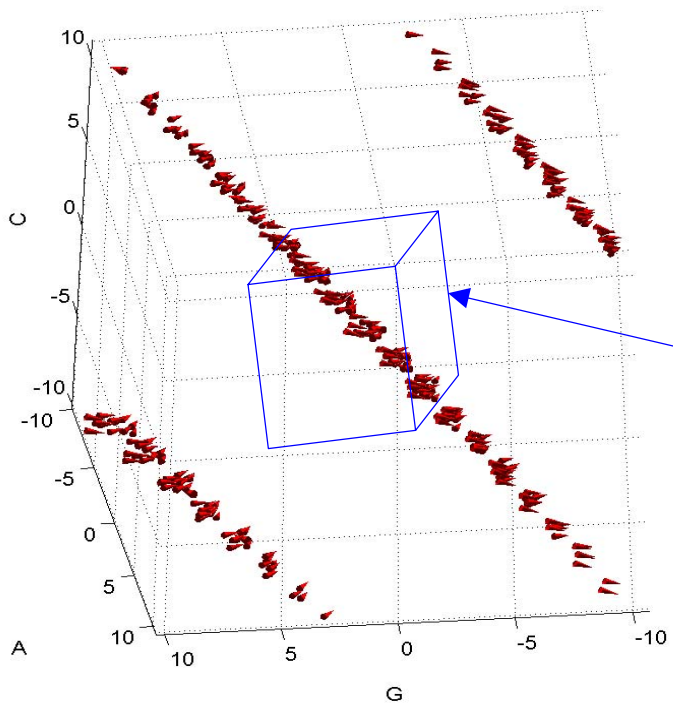
16SrDNA_Bacteria_CVS_1_1.xml coordinates			
Entire 'Band'	Forward Primer	Reverse Primer	Rating
56_376	56_104	318_376	*good region*
352_655	352_416	536_655	*unlikely*
579_726	579_620	680_726	*unlikely*
692_838	692_738	804_838	*unlikely*
753_1003	753_838	919_1003	*unlikely*
980_1154	980_1017	1094_1154	*unlikely*
1124_1298	1124_1189	1243_1298	*good region*
1243_1467	1243_1298	1384_1467	*good region*
1449_1815	1449_1543	1728_1815	*good region*
1745_2007	1745_1815	1955_2007	*good region*
2007_2174	2007_2045	2110_2174	*good region*
2200_2493	2200_2295	2380_2493	*unlikely*

Resolution – Mass Spec Detector Correlation problems



Resolution – Mass Spec Detector Correlation problems(2)

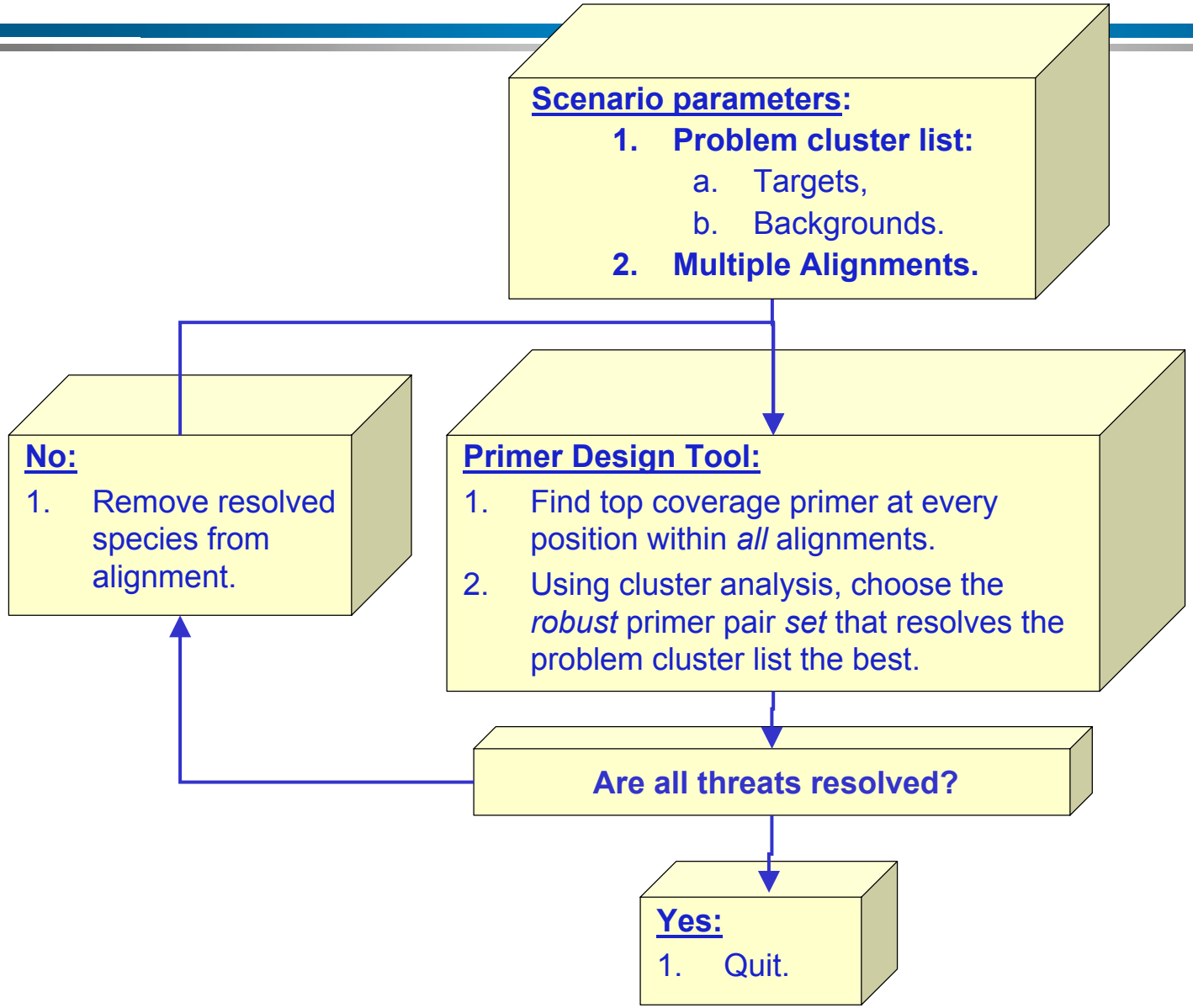
Mass distance of 8 Daltons.



Ambiguous delta base counts

(4d cube with side length 7)

2	0	1	-3
0	1	2	-3
-3	3	3	-3
3	-1	0	-2
1	0	1	-2
-2	2	2	-2
2	-1	0	-1
-1	1	1	-1
-3	2	2	-1
3	-2	-1	0
0	0	0	0
-3	2	1	0
3	-2	-2	1
1	-1	-1	1
-2	1	0	1
2	-2	-2	2
-1	0	-1	2
-3	1	0	2
3	-3	-3	3
0	-1	-2	3
-2	0	-1	3

- [3 –3 –3 3] is far in euclidian distance but close in mass.
- [-1 0 0 1] is close in euclidian distance but far in mass.
- <¼ % have mass difference within 8 Daltons of each other, therefore it is best to simply check the mass, rather than use it as a distance metric to separate species. (also, although the ambiguous delta base count list is very symmetric, both nominal and reverse complement base counts must be separated to avoid correlations).

Primer Design