PCR Primer Validation Report

Exploring the potential for false negatives in PCR design

Aaron C. Robinson

July 11, 2013

Send Blast

Blast bait sequences against NCBI

Utilizes in-house tool suite called blast2tree.

Filtering & Bounding:

- 1500 hits are collected
- Sequences longer than 1000 nucleotides are downloaded from GenBank
- Only results from our target organism are considered

Trim Database

Filter and build Fastas from Blast results

By this point,

1. Blast results have been converted to csv
2. The csv results have been imported into SQLite database

Given a database containing Blast results, this script builds FASTA files for pair-wise alignment:

- A filter of 1e-9 (or $10^{(} - 10)$) is used to screen results
- Sequences that pass the filter are converted to FASTA

Pairwise Alignment

Forward and reverse primers are pairwise-aligned using SSearch36

At this point, we need to perform a pairwise alignment to determine the primer region of our result sequences.

- A pairwise alignment is performed against the screened Blast results and each primer
- SSearch36 is part of the Fasta36 alignment tools, using a Smith-Waterman algorithm

This returns a format that is parsed into CSV and imported into SQLite3.

Outline

Generate Report

Using Python and Jinja2, the results are parsed into HTML report

This process...

- encompasses the majority of the report generation
- is standalone given the SQLite databases are built

By now the SQLite database should contain three tables:

- blast_results
- 'ssearch_forward'
- 'ssearch_reverse'

Generate Report

Using Python and Jinja2, the results are parsed into HTML report

We will count 100% matches but they will not be iterated for reporting purposes.

Therefore for the result of the process refer only to mismatches.

Filtering:

- A filter of 1e-4 (or $10^{(} - 5)$) is used to screen results
- We verify the alignment is over the full primer region

Report HTML

1 Send Blast

2 Trim Database

3 Pairwise Alignment

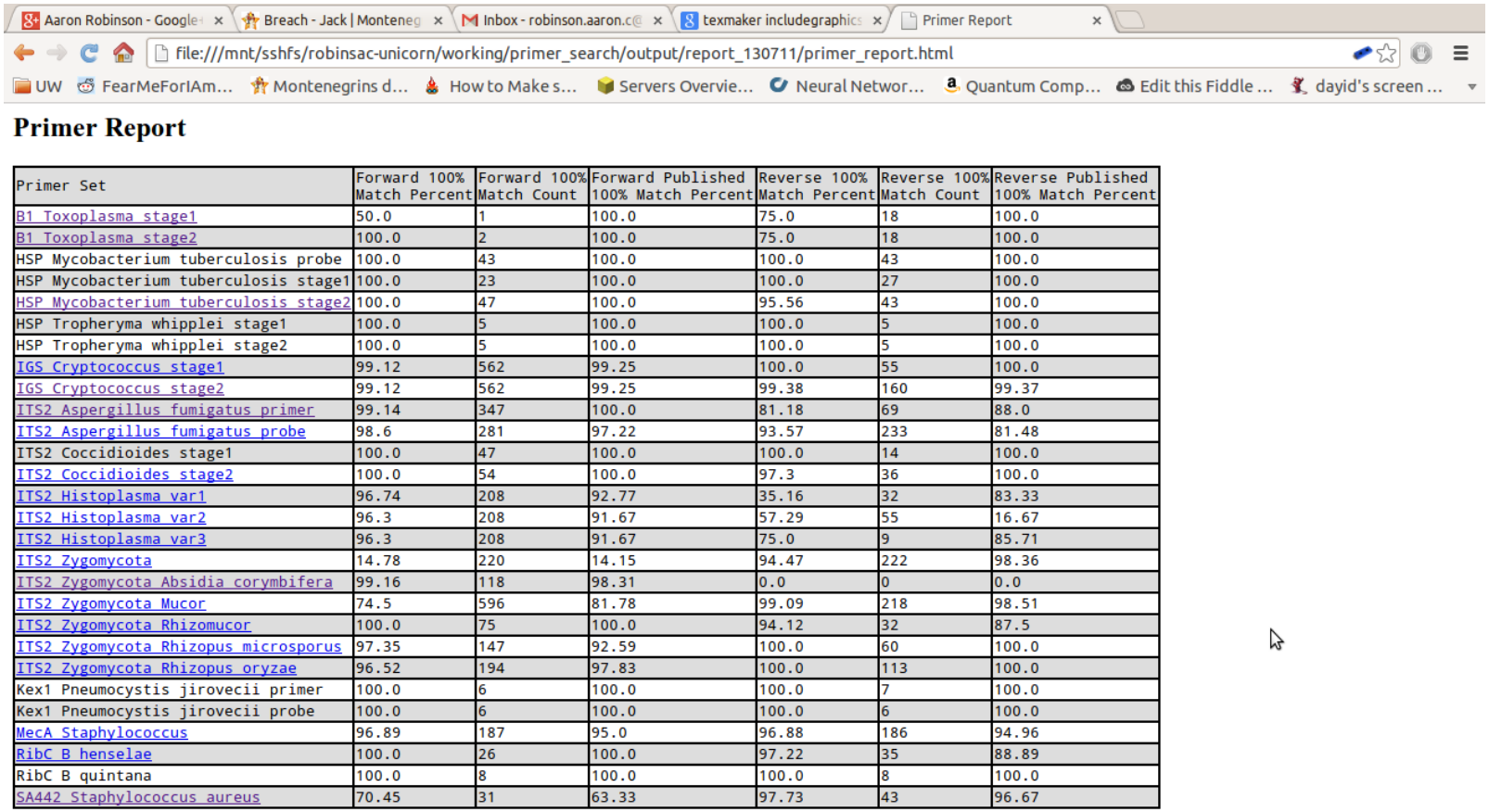
4 Generate Report
- Outline
- Report HTML
- Organism HTML

5 BASH it

6 Questions

Report HTML

This is the HTML output from the Primer Report



Primer Set	Forward 100% Match Percent	Forward 100% Match Count	Forward Published 100% Match Percent	Reverse 100% Match Percent	Reverse 100% Match Count	Reverse Published 100% Match Percent
B1_Toxoplasma_stage1	50.0	1	100.0	75.0	18	100.0
B1_Toxoplasma_stage2	100.0	2	100.0	75.0	18	100.0
HSP_Mycobacterium_tuberculosis_probe	100.0	43	100.0	100.0	43	100.0
HSP_Mycobacterium_tuberculosis_stage1	100.0	23	100.0	100.0	27	100.0
HSP_Mycobacterium_tuberculosis_stage2	100.0	47	100.0	95.56	43	100.0
HSP_Tropheryma_whipplei_stage1	100.0	5	100.0	100.0	5	100.0
HSP_Tropheryma_whipplei_stage2	100.0	5	100.0	100.0	5	100.0
IGS_Cryptococcus_stage1	99.12	562	99.25	100.0	55	100.0
IGS_Cryptococcus_stage2	99.12	562	99.25	99.38	160	99.37
ITS2_Aspergillus_fumigatus_primer	99.14	347	100.0	81.18	69	88.0
ITS2_Aspergillus_fumigatus_probe	98.6	281	97.22	93.57	233	81.48
ITS2_Coccidioides_stage1	100.0	47	100.0	100.0	14	100.0
ITS2_Coccidioides_stage2	100.0	54	100.0	97.3	36	100.0
ITS2_Histoplasma_var1	96.74	208	92.77	35.16	32	83.33
ITS2_Histoplasma_var2	96.3	208	91.67	57.29	55	16.67
ITS2_Histoplasma_var3	96.3	208	91.67	75.0	9	85.71
ITS2_Zygomycota	14.78	220	14.15	94.47	222	98.36
ITS2_Zygomycota_Absidia_corymbifera	99.16	118	98.31	0.0	0	0.0
ITS2_Zygomycota_Mucor	74.5	596	81.78	99.09	218	98.51
ITS2_Zygomycota_Rhizomucor	100.0	75	100.0	94.12	32	87.5
ITS2_Zygomycota_Rhizopus_microsporus	97.35	147	92.59	100.0	60	100.0
ITS2_Zygomycota_Rhizopus_oryzae	96.52	194	97.83	100.0	113	100.0
Kex1_Pneumocystis_jirovecii_primer	100.0	6	100.0	100.0	7	100.0
Kex1_Pneumocystis_jirovecii_probe	100.0	6	100.0	100.0	6	100.0
MecA_Staphylococcus	96.89	187	95.0	96.88	186	94.96
RibC_B_henselae	100.0	26	100.0	97.22	35	88.89
RibC_B_quintana	100.0	8	100.0	100.0	8	100.0
SA442_Staphylococcus_aureus	70.45	31	63.33	97.73	43	96.67

Figure: This is an example of 'primer_report.html'

Organism HTML

1 Send Blast

2 Trim Database

3 Pairwise Alignment

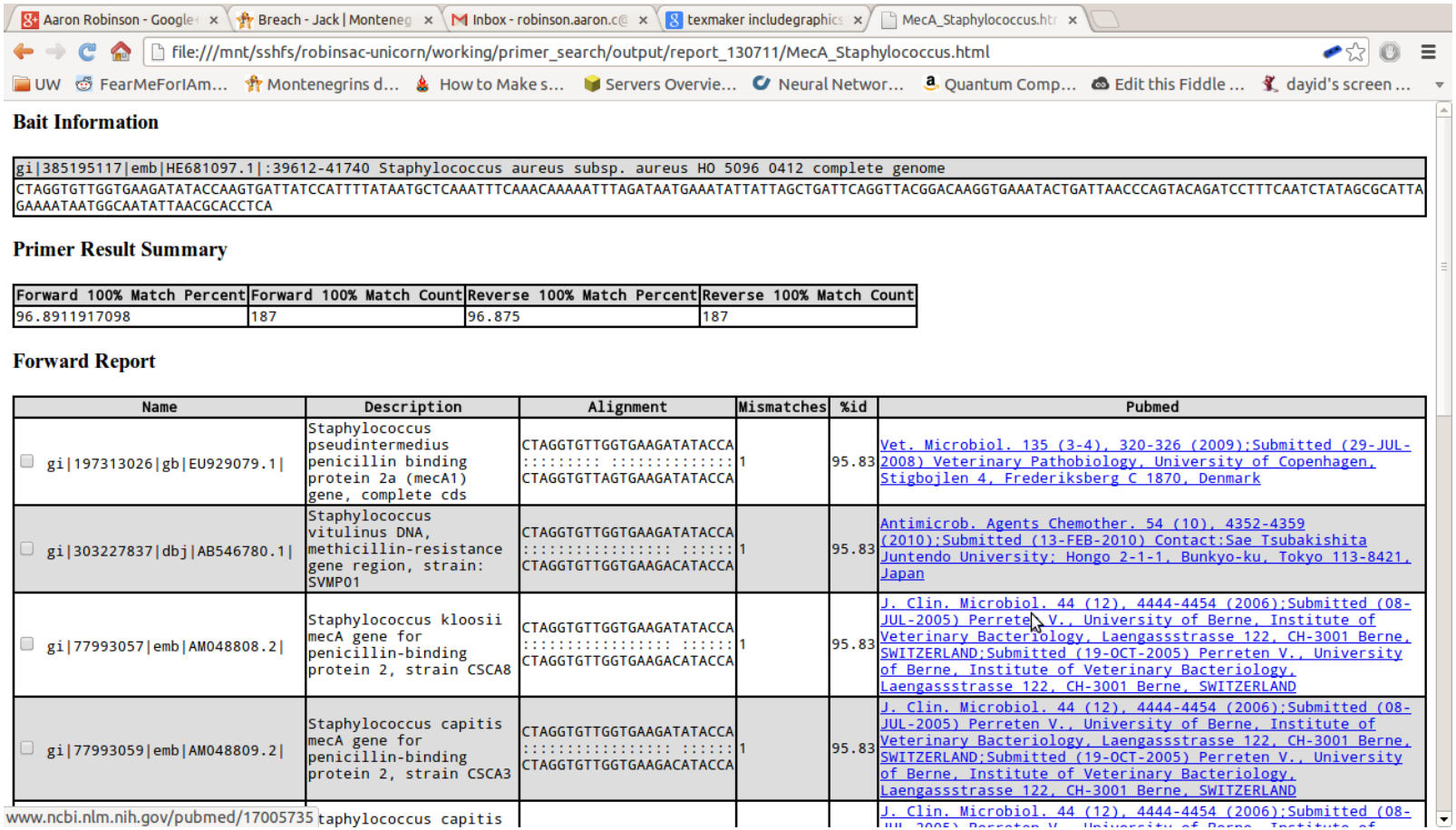
4 Generate Report
- Outline
- Report HTML
- Organism HTML

5 BASH it

6 Questions

Organism HTML

This is the organism specific report



Aaron Robinson - Google+ | Breach - Jack | Monteneg | Inbox - robinson.aaron.c@ | texmaker includegraphics | MecA_Staphylococcus.htr

file:///mnt/sshfs/robinsac-unicorn/working/primer_search/output/report_130711/MecA_Staphylococcus.html

UW FearMeForIAm... Montenegrins d... How to Make s... Servers Overvie... Neural Networ... Quantum Comp... Edit this Fiddle ... dayid's screen ...

Bait Information

gi|385195117|emb|HE681097.1|:39612-41740 Staphylococcus aureus subsp. aureus HO 5096 0412 complete genome

CTAGGTGTTGGTGAAGATATACCAAGTGATTATCCATTTTATAATGCTCAAATTTCAAACAAAAATTTAGATAATGAAATATTATTAGCTGATTCAGGTTACGGACAAGGTGAAATACTGATTAACCCAGTACAGATCCTTTCAATCTATAGCGCATTA
GAAAATAATGGCAATATTAACGCACCTCA

Primer Result Summary

Forward 100% Match Percent	Forward 100% Match Count	Reverse 100% Match Percent	Reverse 100% Match Count
96.8911917098	187	96.875	187

Forward Report

Name	Description	Alignment	Mismatches	%id	Pubmed
gi|197313026|gb|EU929079.1|	Staphylococcus pseudintermedius penicillin binding protein 2a (mecA1) gene, complete cds	CTAGGTGTTGGTGAAGATATACCA ::::::::: :::::::::::::: CTAGGTGTTAGTGAAGATATACCA	1	95.83	Vet. Microbiol. 135 (3-4), 320-326 (2009):Submitted (29-JUL-2008) Veterinary Pathobiology, University of Copenhagen, Stigbojlen 4, Frederiksberg C 1870, Denmark
gi|303227837|dbj|AB546780.1|	Staphylococcus vitulinus DNA, methicillin-resistance gene region, strain: SVMP01	CTAGGTGTTGGTGAAGATATACCA :::::::::::::::: :::::: CTAGGTGTTGGTGAAGACATACCA	1	95.83	Antimicrob. Agents Chemother. 54 (10), 4352-4359 (2010):Submitted (13-FEB-2010) Contact:Sae Tsubakishita Juntendo University; Hongo 2-1-1, Bunkyo-ku, Tokyo 113-8421, Japan
gi|77993057|emb|AM048808.2|	Staphylococcus kloosii mecA gene for penicillin-binding protein 2, strain CSCA8	CTAGGTGTTGGTGAAGATATACCA :::::::::::::::: :::::: CTAGGTGTTGGTGAAGACATACCA	1	95.83	J. Clin. Microbiol. 44 (12), 4444-4454 (2006):Submitted (08-JUL-2005) Perreten V., University of Berne, Institute of Veterinary Bacteriology, Laengassstrasse 122, CH-3001 Berne, SWITZERLAND;Submitted (19-OCT-2005) Perreten V., University of Berne, Institute of Veterinary Bacteriology, Laengassstrasse 122, CH-3001 Berne, SWITZERLAND
gi|77993059|emb|AM048809.2|	Staphylococcus capitis mecA gene for penicillin-binding protein 2, strain CSCA3	CTAGGTGTTGGTGAAGATATACCA :::::::::::::::: :::::: CTAGGTGTTGGTGAAGACATACCA	1	95.83	J. Clin. Microbiol. 44 (12), 4444-4454 (2006):Submitted (08-JUL-2005) Perreten V., University of Berne, Institute of Veterinary Bacteriology, Laengassstrasse 122, CH-3001 Berne, SWITZERLAND;Submitted (19-OCT-2005) Perreten V., University of Berne, Institute of Veterinary Bacteriology, Laengassstrasse 122, CH-3001 Berne, SWITZERLAND
	taphylococcus capitis				J. Clin. Microbiol. 44 (12), 4444-4454 (2006):Submitted (08-

www.ncbi.nlm.nih.gov/pubmed/17005735

Figure: This is an example output for the MecA gene given S. aurous

BASH it

Pipe it all together using a BASH script

- There are many scripting options for connecting all the peices.
- A BASH script was used for it's ability to facilitate a fast path to solution.
- Moving forward, a Python wrapper would be a more elegant solution.

Questions

Please feel free to critique how this process could be improved