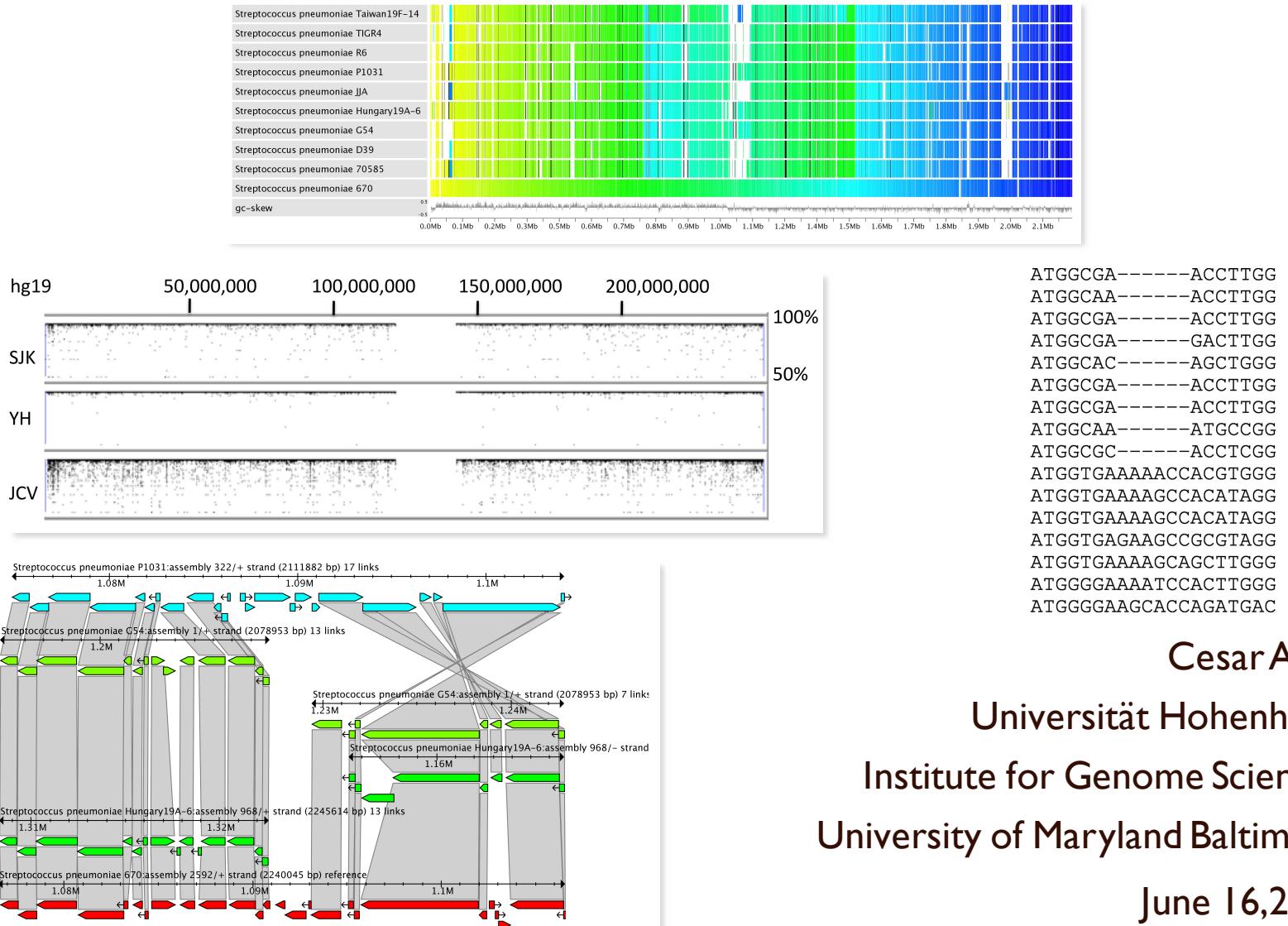


CloVR Comparative Pipeline



Cesar Arze
 Universität Hohenheim
 Institute for Genome Sciences
 University of Maryland Baltimore

June 16, 2016

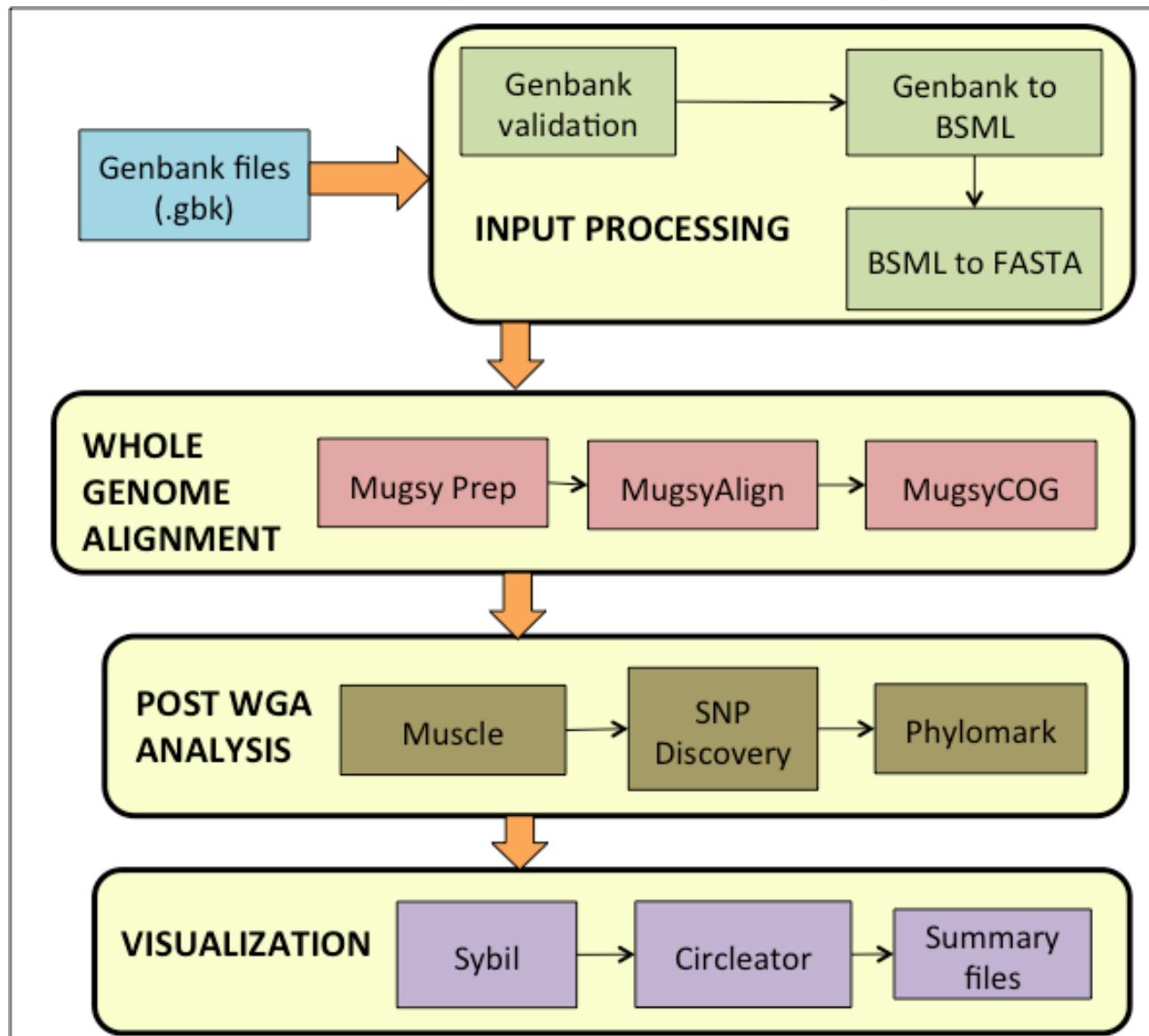
Outline

- Pipeline Overview
- Whole Genome Alignment (WGA)
- Mugsy-Annotator
- Pipeline Outputs
 - Summary report
 - Mugsy clusters
 - SNP table
 - Visualization
 - Alignment
 - Phylomark
 - Circleator
 - Sybil

Motivation

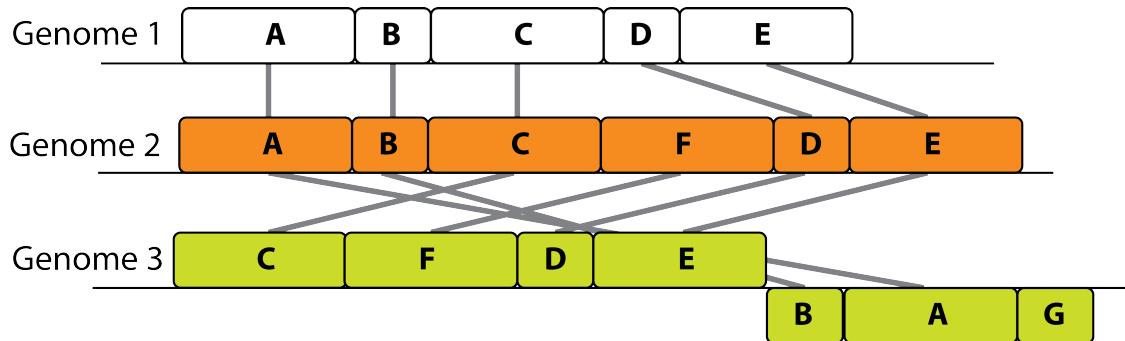
- Functional comparison of several bacterial genomes
- Reference independent analysis
- Phylogenetic trees
- SNP calling
- High throughput, completely automated

CloVR Comparative Pipeline



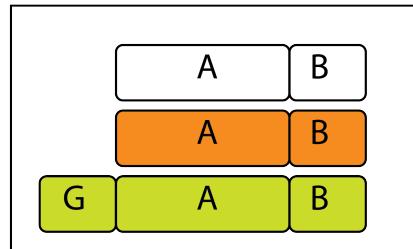
Whole Genome Alignment

- True global alignment not possible

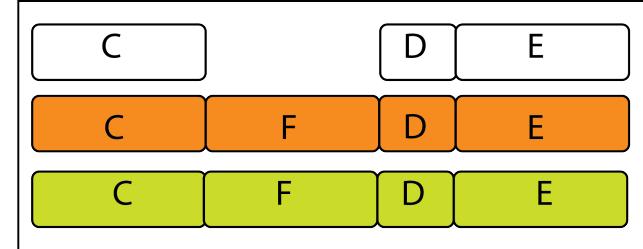


- Set of local alignments (LCBs)

Locally collinear blocks

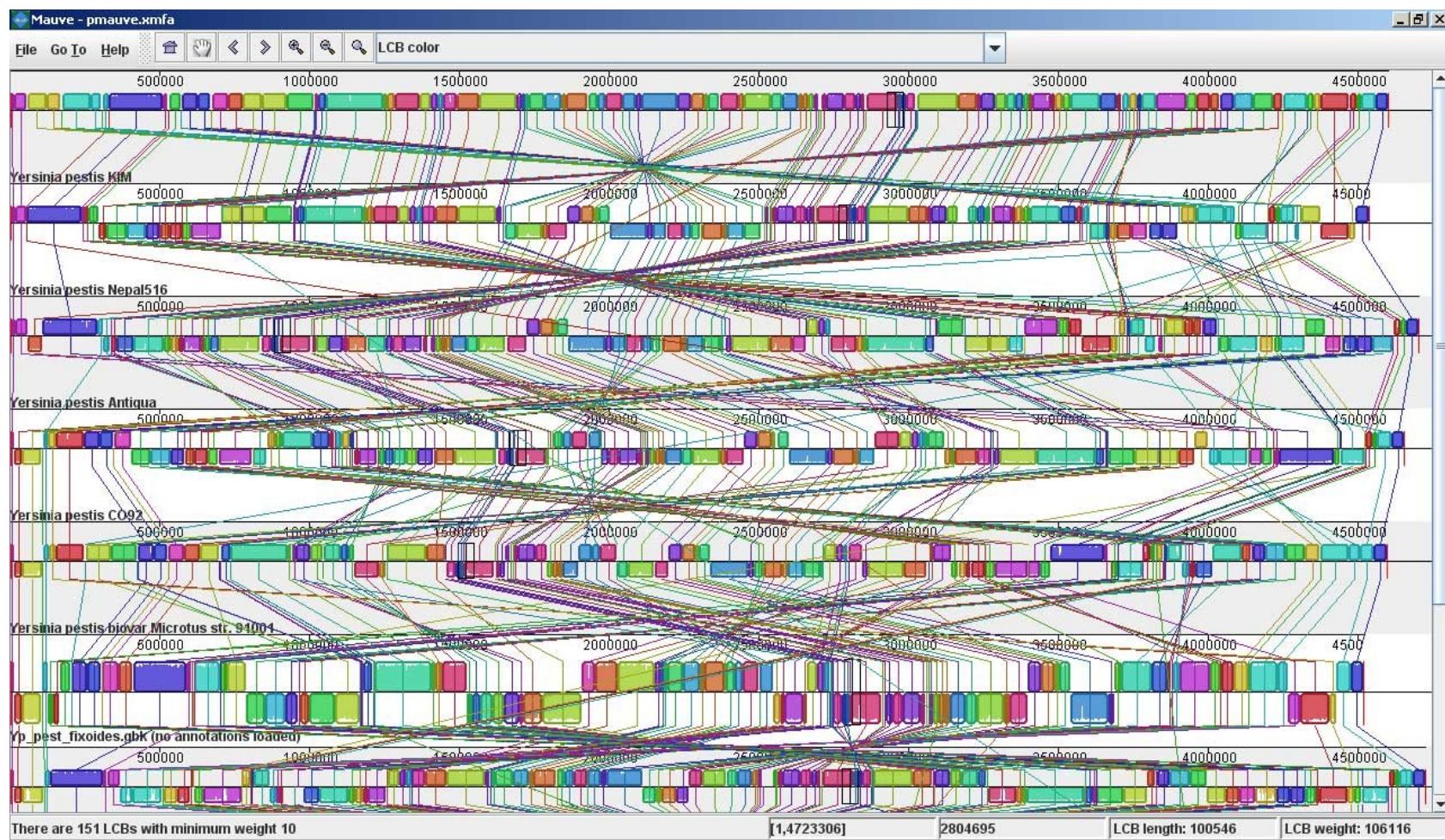


LCB 1



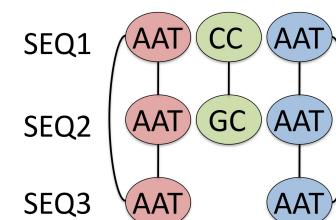
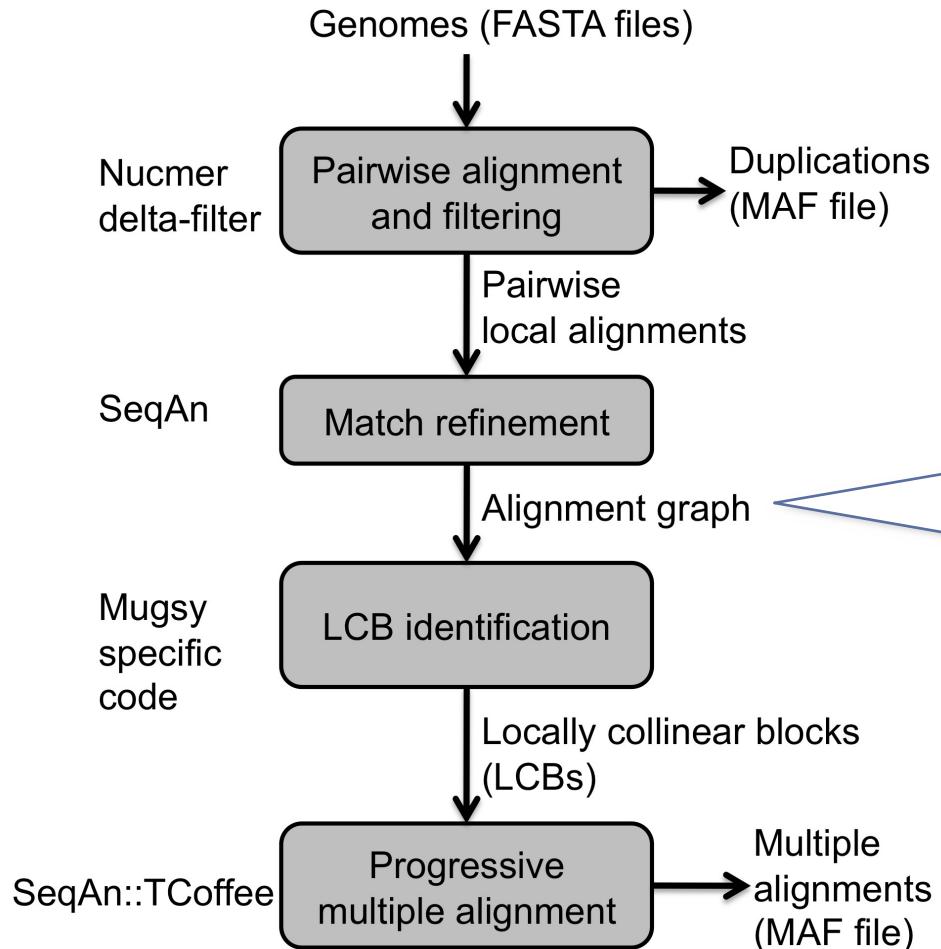
LCB 2

Whole Genome Alignment

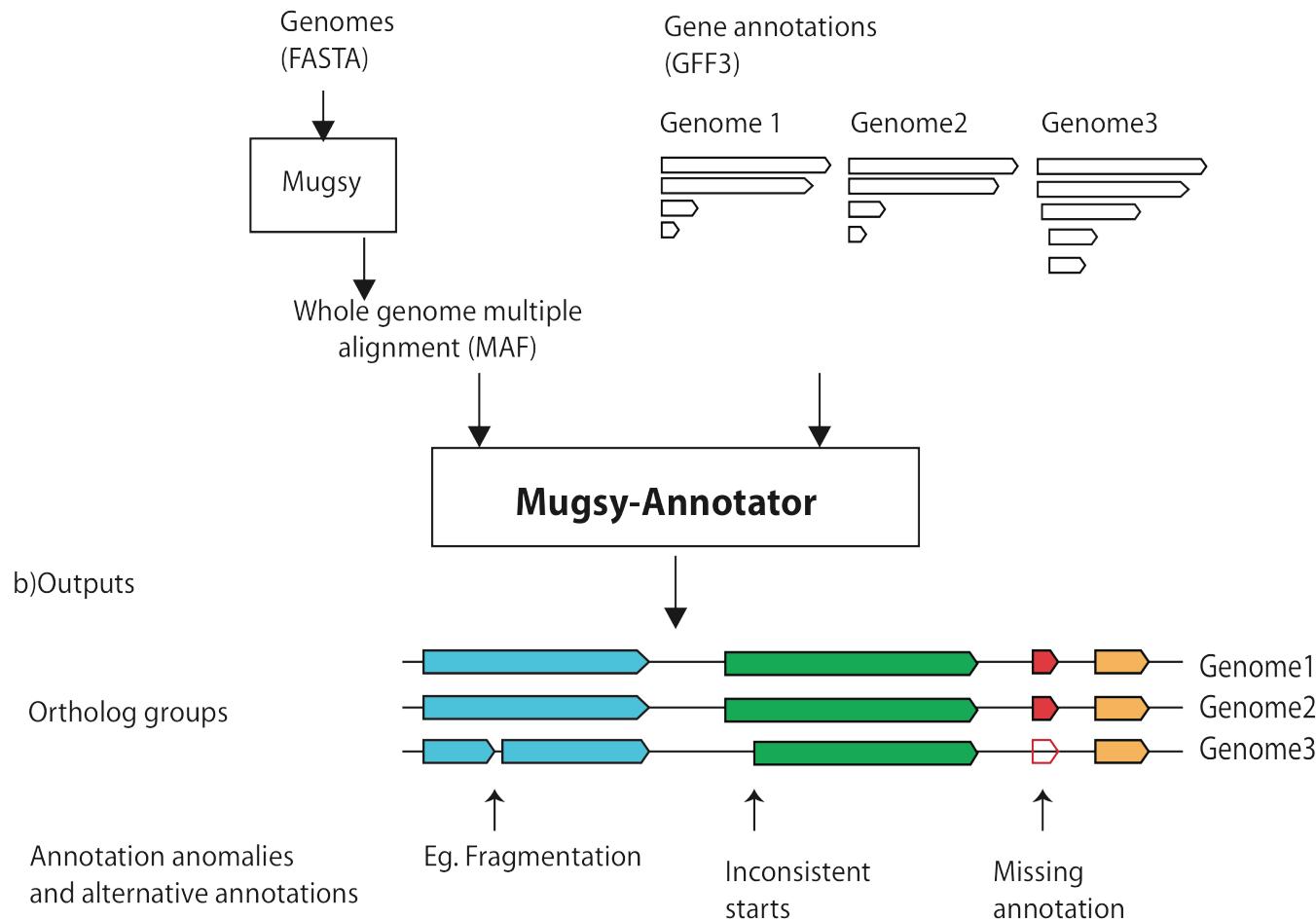


Aaron C.E. Darling, Bob Mau, Frederick R. Blatter, and Nicole T. Perna. *Genome Research.* (2004)

Mugsy



Mugsy Annotator



Summary Report

Locus Id	Species	Strain	Genome Length(in Mbp)	#scaffolds/contigs	#CDS	#uniqueCDS
-						
NC_003116	Neisseria meningitidis Z2491	Z2491	2.18	1	1909	127
NC_008767	Neisseria meningitidis FAM18	FAM18	2.19	1	1917	83
NC_003112	Neisseria meningitidis MC58	MC58	2.27	1	2063	265
NC_010120	Neisseria meningitidis 053442	053442	2.15	1	2020	149
NC_013016	Neisseria meningitidis alpha14	alpha14	2.15	1	1872	135

Core genome length (in Kbp) : 1775.67

Number of core clusters : 1460

References:

-
1. CloVR Angiuoli S.V., Matalka M., Gussman A., Galens K., Vangala M., Riley D., Arze C., White J.R., White O., and Fricke W.F. (2011). CloVR: A virtual machine for automated and portable sequence analysis from the desktop using cloud computing. *BMC Bioinformatics*, 12(1):356. PMID: 21878105
 2. Mugsy Angiuoli S.V., Salzberg S.L. (2011). Mugsy: fast multiple alignment of closely related whole genomes. *Bioinformatics*, 27(3):334-42. doi: 10.1093/bioinformatics/btq665. PMID: 21148543
 3. Pan-genome annotation Angiuoli S.V., Dunning Hotopp J.C., Salzberg S.L., Tettelin H. (2011). Improving pan-genome annotation using whole genome multiple alignment. *BMC Bioinformatics*, 12:272. doi: 10.1186/1471-2105-12-272. PMID: 21718539
 4. Phylomark, phylogenetic tree generation Sahl J.W., Matalka M.N., Rasko D.A. (2012). Phylomark, a tool to identify conserved phylogenetic markers from whole-genome alignments. *Appl Environ Microbiol*, 78(14):4884-92. doi: 10.1128/AEM.00929-12. PMID: 22582056

Mugsy Clusters

COG = 2, size 2, connections = 0, perfect = 0;

NC_003112.NMB1768

NC_010120.NMCC_1333

COG = 4, size 4, connections = 0, perfect = 0;

NC_008767.NMC0527

NC_003112.NMB0585

NC_010120.NMCC_0531

NC_013016.NMO_0470

COG = 6, size 2, connections = 0, perfect = 0;

NC_003116.NMA0688

NC_008767.NMC0450

COG = 8, size 8, connections = 0, perfect = 0;

NC_013016.NMO_0398

NC_003112.NMB0504

NC_003112.NMB0514

NC_003112.NMB0515

NC_003116.NMA0695

NC_008767.NMC0444

NC_008767.NMC0456

NC_010120.NMCC_0457

COG = 10, size 5, connections = 0, perfect = 0;

NC_013016.NMO_0592

NC_003112.NMB0700

NC_003116.NMA0905

NC_008767.NMC0651

NC_010120.NMCC_0659

COG = 12, size 2, connections = 0, perfect = 0;

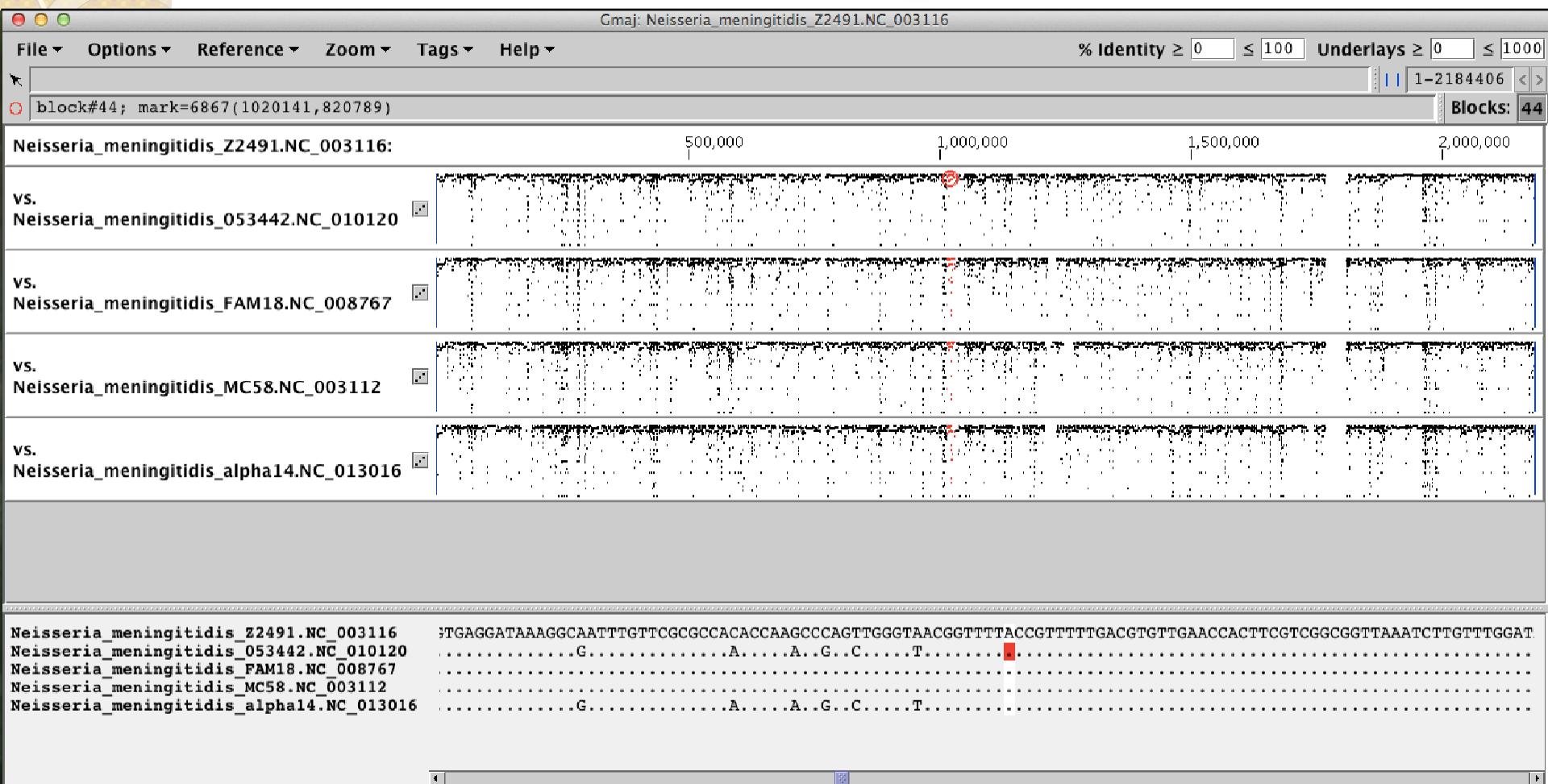
NC_013016.NMO_0169

NC_003112.NMB1998

SNP Identification

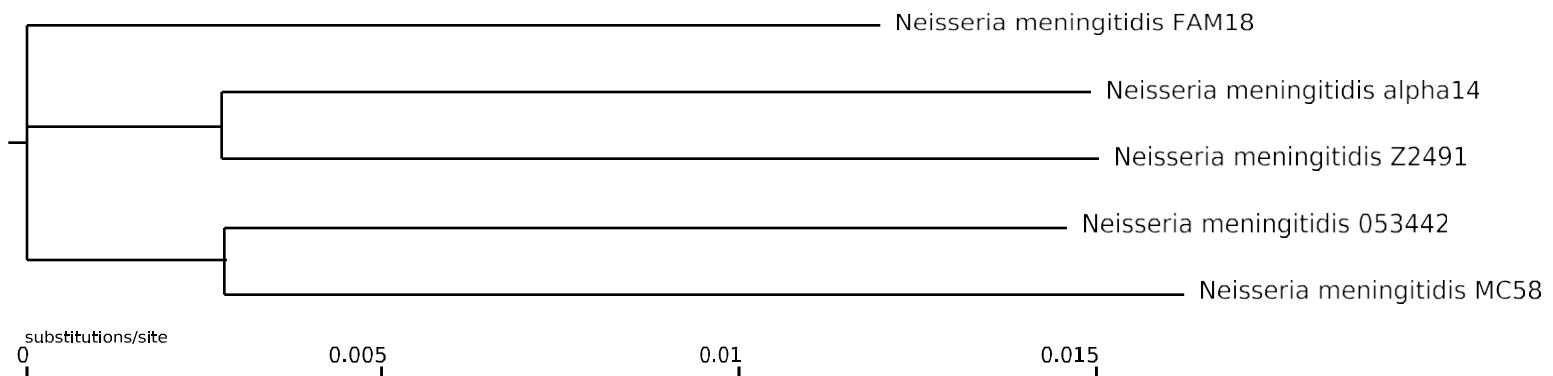
| Neisseria_meningitidis |
|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| NC_010120 | 305725 T | NC_008767 | 306034 C | NC_003112 | | | |
| NC_010120 | 305767 T | NC_008767 | 306076 G | NC_003112 | | | |
| NC_010120 | 305844 C | NC_008767 | 306154 G | NC_003112 | | | |
| NC_010120 | 305914 G | NC_008767 | 306223 G | NC_003112 | | | |
| NC_010120 | 305942 T | NC_008767 | 306251 T | NC_003112 | | | |
| NC_010120 | 305948 A | NC_008767 | 306257 G | NC_003112 | | | |
| NC_010120 | 306105 C | NC_008767 | 306414 G | NC_003112 | | | |
| NC_010120 | 306106 C | NC_008767 | 306415 G | NC_003112 | | | |
| NC_010120 | 306107 T | NC_008767 | 306416 G | NC_003112 | | | |
| NC_010120 | 306108 T | NC_008767 | 306417 C | NC_003112 | | | |
| NC_010120 | 306121 G | NC_008767 | 306430 T | NC_003112 | | | |
| NC_010120 | 306124 T | NC_008767 | 306433 C | NC_003112 | | | |
| NC_010120 | 306125 A | NC_008767 | 306434 G | NC_003112 | | | |
| NC_010120 | 306126 A | NC_008767 | 306435 G | NC_003112 | | | |
| NC_010120 | 306127 A | NC_008767 | 306436 T | NC_003112 | | | |
| NC_010120 | 306129 A | NC_008767 | 306438 C | NC_003112 | | | |
| NC_010120 | 306131 C | NC_008767 | 306440 G | NC_003112 | | | |
| NC_010120 | 306135 T | NC_008767 | 306444 C | NC_003112 | | | |
| NC_010120 | 306154 A | NC_008767 | 306463 G | NC_003112 | | | |
| NC_010120 | 306175 C | NC_008767 | 306484 T | NC_003112 | | | |
| NC_010120 | 306193 C | NC_008767 | 306502 T | NC_003112 | | | |
| NC_010120 | 306208 G | NC_008767 | 306517 C | NC_003112 | | | |
| NC_010120 | 306210 G | NC_008767 | 306519 A | NC_003112 | | | |

Visualization - Alignment



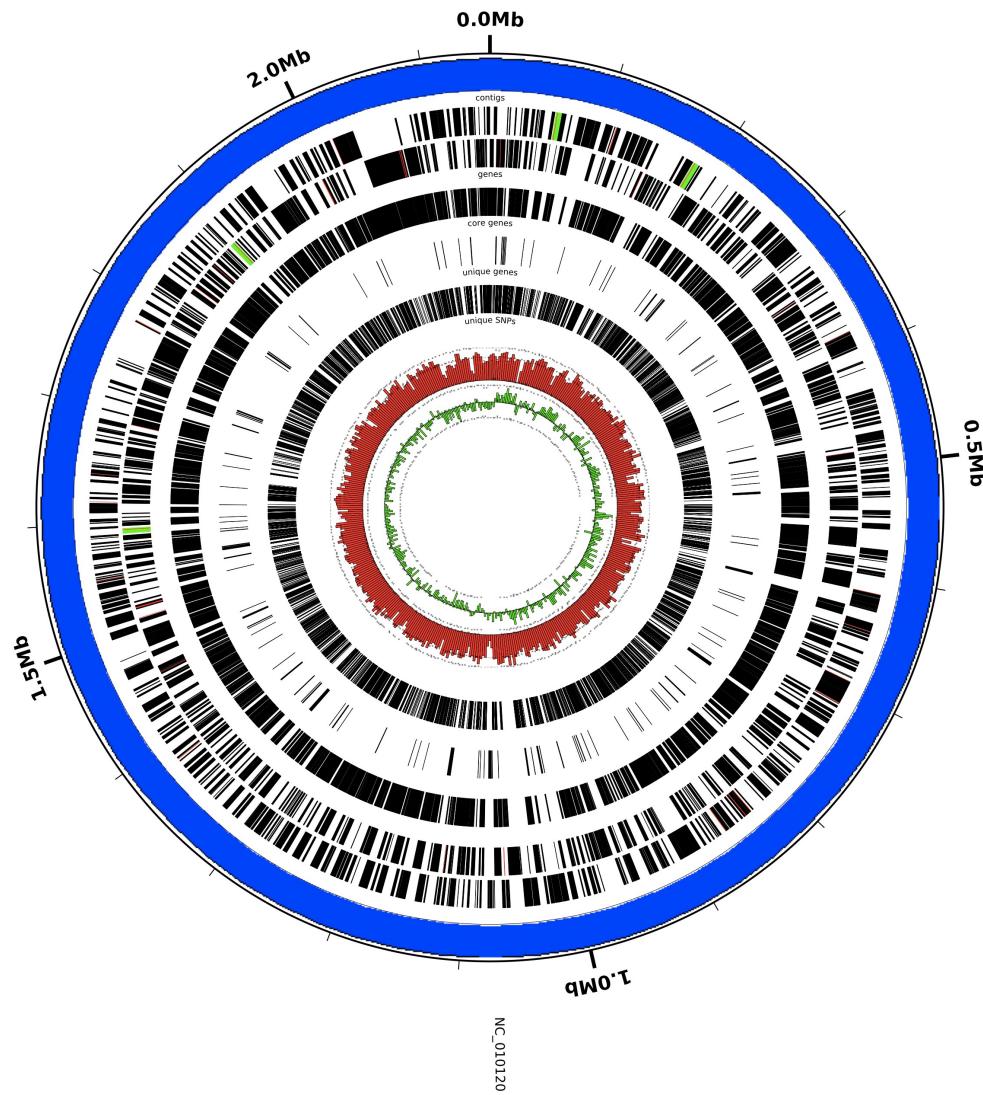
- gmaj (<http://globin.bx.psu.edu/dist/gmaj/>)
- Patched version available (<http://cb2.igs.umaryland.edu/gmaj.zip>)

Visualization - Phylomark/FastTree

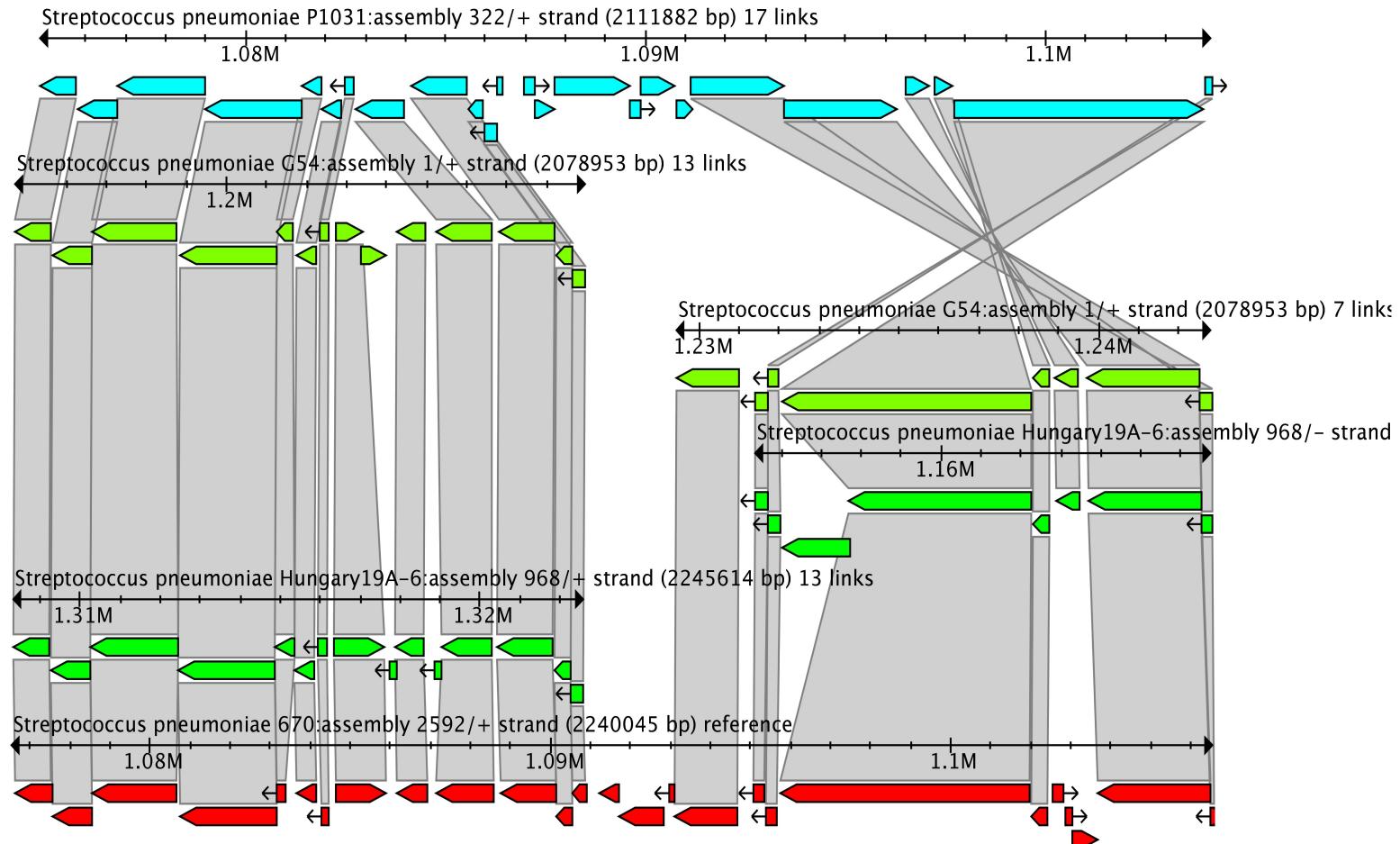


Sahl JW, Matalka MN, Rasko DA. *Appl Environ Microbiol.* (2012)
Price M.N., Dehal P.S., Arkin A. *PLoS ONE.* (2010)

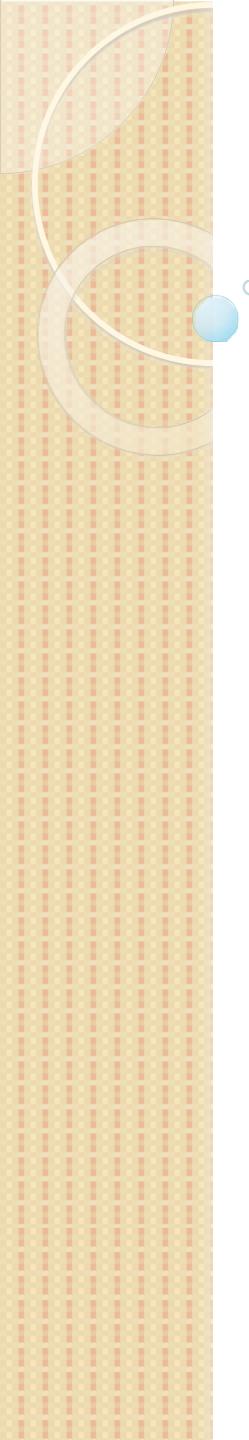
Visualization - Circleator



Visualization - Sybil



Riley DR, Angiuoli SV, Crabtree J, Dunning Hotopp JC, Tettelin H. *Bioinformatics* (2012)
Crabtree J, Angiuoli SV, Wortman JR, White OR. *Methods Mol. Biol.* (2007)



Thank You

Questions??

Herve Tettlin
tettelin@som.umaryland.edu