Supplementary Material for "Circleator: Flexible Circular Visualization of Genome-Associated Data with BioPerl and SVG" by J. Crabtree *et al.*

May 20, 2014

1 Example Circleator Figures

1.1 Fig. 1: *H. influenzae* Rd KW20 genes, GC-content and GC-skew

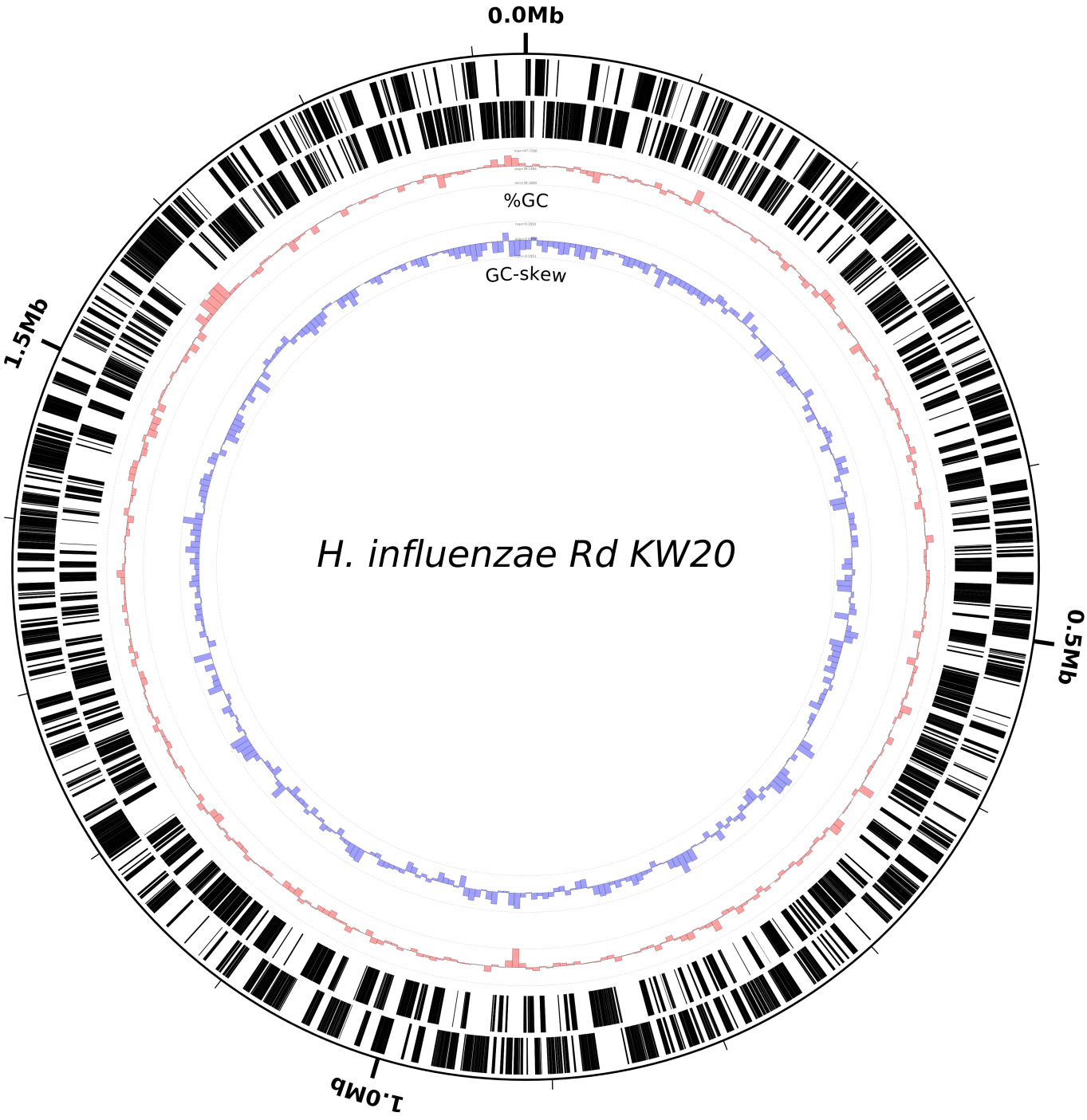

Figure 1: A simple Circleator figure for *Haemophilus influenzae* Rd KW20 showing, from outside to inside: coordinate labels, forward and reverse strand genes, percent GC content, GC-skew, and the organism name.

Fig. 1 depicts the 1,830,138 bp genome sequence of *Haemophilus influenzae* Rd KW20, the first free-living organism to have its genome completely sequenced. The outermost ring/track displays the scale, with a tick mark every 0.1 Mb and a label every 0.5 Mb. The next track shows all gene features from the GenBank flat file that are annotated on the forward strand, including any rRNA or tRNA genes. The next track shows all gene features annotated on the *reverse* strand. Next, in pink/light red, is a percent GC-content graph using nonoverlapping windows of length 5kb and showing the deviation from the computed average value ($\backsim$ 38%) up to a maximum of $\backsim$ 48% and down to a minimum of $\backsim$ 28%. The next track, in

light blue, is a GC-skew graph, also using nonoverlapping windows of length 5kb and showing the deviation of the average value (for the window) from the average value computed over the entire sequence (0.005). In this example all minimum and maximum graph values are determined dynamically based on the input data, but this behavior is configurable. Both graph tracks are also labeled in the figure for convenience. The final track displays the organism name in the center of the circle and uses a horizontal label rather than the circular labels used by the %GC and GC-content tracks.

1.1.1 Fig. 1 Inputs

- The GenBank flat file for accession L42023.1, downloaded from `http://www.ncbi.nlm.nih.gov/nuccore/L42023.1`
- The Circleator configuration file shown in subsection 1.1.2

1.1.2 Fig. 1 Configuration File (conf/Hi_RdKW20.cfg)

This is the entire configuration file for Fig. 1:

```
# coordinate labels
coords

# genes, forward and reverse strand
tiny-cgap
genes-fwd
tiny-cgap
genes-rev
small-cgap

# %GC
%GCmin-max-dfa color1=#ffa0a0
medium-label label-text=%GC
small-cgap
# GC-skew
GCskew-min-max-df0 color1=#a0a0ff
medium-label label-text=GC-skew

# caption
large-label innerf=0,label-text=H. influenzae Rd KW20,label-type=horizontal,font-style=italic
```

1.1.3 Fig. 1 Command Line

The following commands were used to produce Fig. 1 in SVG and PNG format:

```
circleator --data=data/L42023.1.gb --config=conf/Hi_RdKW20.cfg > Hi_RdKW20.svg
rasterize-svg Hi_RdKW20.svg png 5000 5000
```

1.2 Fig. 2: *C. caviae* GPIC: BSR Comparison with *C. muridarum* and *C. pneumoniae*

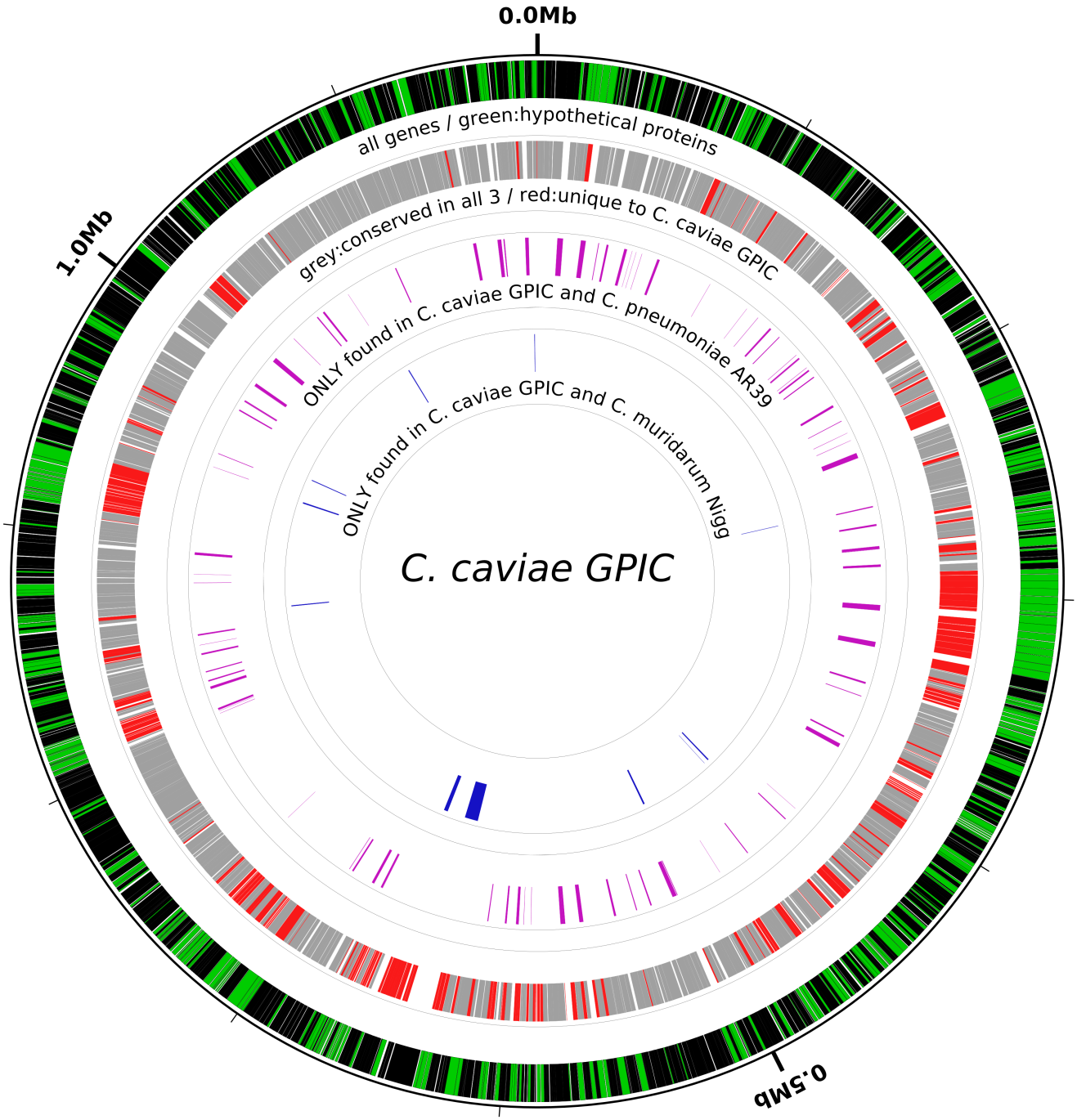

Figure 2: A 3-way genome comparison based on output from the BSR (BLAST Score Ratio) tool [3]. The outermost ring shows all genes in the reference, *Chlamydophila caviae* GPIC, with hypothetical proteins in green. The next ring shows genes conserved (with BSR score ≥ 0.4) in all 3 genomes (grey) plus those that are unique to the reference (red). The inner two rings show genes conserved only in the reference and exactly one of the two query genomes, either *Chlamydophila pneumoniae* AR39, or *Chlamydia muridarum* Nigg, respectively.

Fig. 2 depicts the 1,173,390 bp genome sequence of *Chlamydophila caviae* GPIC and illustrates which of its genes are conserved in the genomes of two related organisms, *Chlamydophila pneumoniae* AR39 and *Chlamydia muridarum* Nigg. A 3-way gene conservation analysis was performed using the BLAST Score Ratio (BSR) tool [3] and genes with a BSR score of 0.4 or higher were marked as conserved (relative to the reference genome, that of *C. caviae* GPIC.) The outermost ring/track displays the scale, with a tick mark every 0.1 Mb and a label every 0.5 Mb. The next track shows all gene features from the GenBank flat file (both forward and reverse strand annotations), and those genes whose annotated gene product contains the string "hypothetical protein" are highlighted in green. The next track also displays genes, but shows only those that are either:

1. conserved in all 3 genomes according to the BSR analysis (shown in grey), or 2. present in the reference genome but *not* conserved in either of the others (red.) The next track shows genes that are conserved in the reference and *C. pneumoniae* AR39 (but *not C. muridarum* Nigg) and the final track before the label in the center shows genes that are conserved in the reference and *C. muridarum* Nigg (but *not C. pneumoniae* AR39). The final track displays the organism name in the center of the circle and uses a horizontal label rather than the circular labels used by the other tracks.

1.2.1 Fig. 2 Inputs

- The RefSeq flat file for accession NC_03361.3, downloaded from `http://www.ncbi.nlm.nih.gov/nuccore/NC_003361`
- The Circleator configuration file shown in subsection 1.2.2

1.2.2 Fig. 2 Configuration File (conf/Cc_GPIC_BSR.cfg)

This is the entire configuration file for Fig. 2. The string \\\ is used to indicate places where a single line in the configuration file must be wrapped over two or more lines in this document, but it is not valid Circleator syntax:

```
# read output from BLAST Score Ratio (BSR) tool but don't display anything yet
new load_bsr1 load-bsr heightf=0,bsr-file=./data/BSR_dir/Cc_GPIC_Cm_Nigg_Cp_AR39.txt,genome1=Cm,genome2=Cp

# coordinate labels
coords

# genes, not separated by strand
tiny-cgap
genes

# highlight conserved hypotheticals, anything with "hypothetical protein" in the gene product name
# innerf=same, outerf=same means this track will be overlaid on the previous one
new HCH rectangle feat-type=CDS,feat-tag=product,feat-tag-regex=hypothetical\sprotein,innerf=same,outerf=same,\\\
opacity=0.8,color1=#00ff00
medium-label label-text=all genes / green:hypothetical proteins
small-cgap

# thin grey circle and very small space
new CH1 rectangle color1=none,color2=#000000,heightf=0.14,feat-type=contig
tiny-cgap outerf=same

# genes conserved in all 3 genomes (phylogenetic signature = 11)
new bsr_all bsr 0.07 threshold=0.4,genomes=Cm|Cp,signature=11,color1=#a0a0a0,color2=#a0a0a0
# genes unique to the reference
new bsr_ref bsr 0.07 threshold=0.4,genomes=Cm|Cp,signature=00,color1=#f91919,color2=#f91919,innerf=same
medium-label label-text=grey:conserved in all 3 / red:unique to C. caviae GPIC
medium-cgap

# thin grey circle and very small space
new CH2 rectangle color1=none,color2=#000000,heightf=0.14,feat-type=contig
tiny-cgap outerf=same

# genes ONLY in the reference and Chlamydophila pneumoniae AR39 (signature = 01)
new bsr_cp bsr 0.07 threshold=0.4,genomes=Cm|Cp,signature=01,color1=#c411be,color2=#c411be
medium-label label-text=ONLY found in C. caviae GPIC and C. pneumoniae AR39m
edium-cgap

# thin grey circle and very small space
new CH3 rectangle color1=none,color2=#000000,heightf=0.14,feat-type=contig
tiny-cgap outerf=same

# genes ONLY in the reference and Chlamydia muridarum Nigg (signature = 10)
new bsr_cm bsr 0.07 threshold=0.4,genomes=Cm|Cp,signature=10,color1=#1511c4,color2=#1511c4
medium-label label-text=ONLY found in C. caviae GPIC and C. muridarum Nigg

# horizontal caption in the center
large-label innerf=0,label-text=C. caviae GPIC,label-type=horizontal,font-style=italic
```

1.2.3 Fig. 2 Command Line

The following commands were used to produce Fig. 2 in SVG and PNG format:

```
circleator --data=data/NC_003361.3.gbk --config=./conf/Cc_GPIC_BSR.cfg Cc_GPIC_BSR.svg
rasterize-svg Cc_GPIC_BSR.svg png 5000 5000
```

1.3 Fig. 3: *G. vaginalis* HMP9231: Human Microbiome Project Example

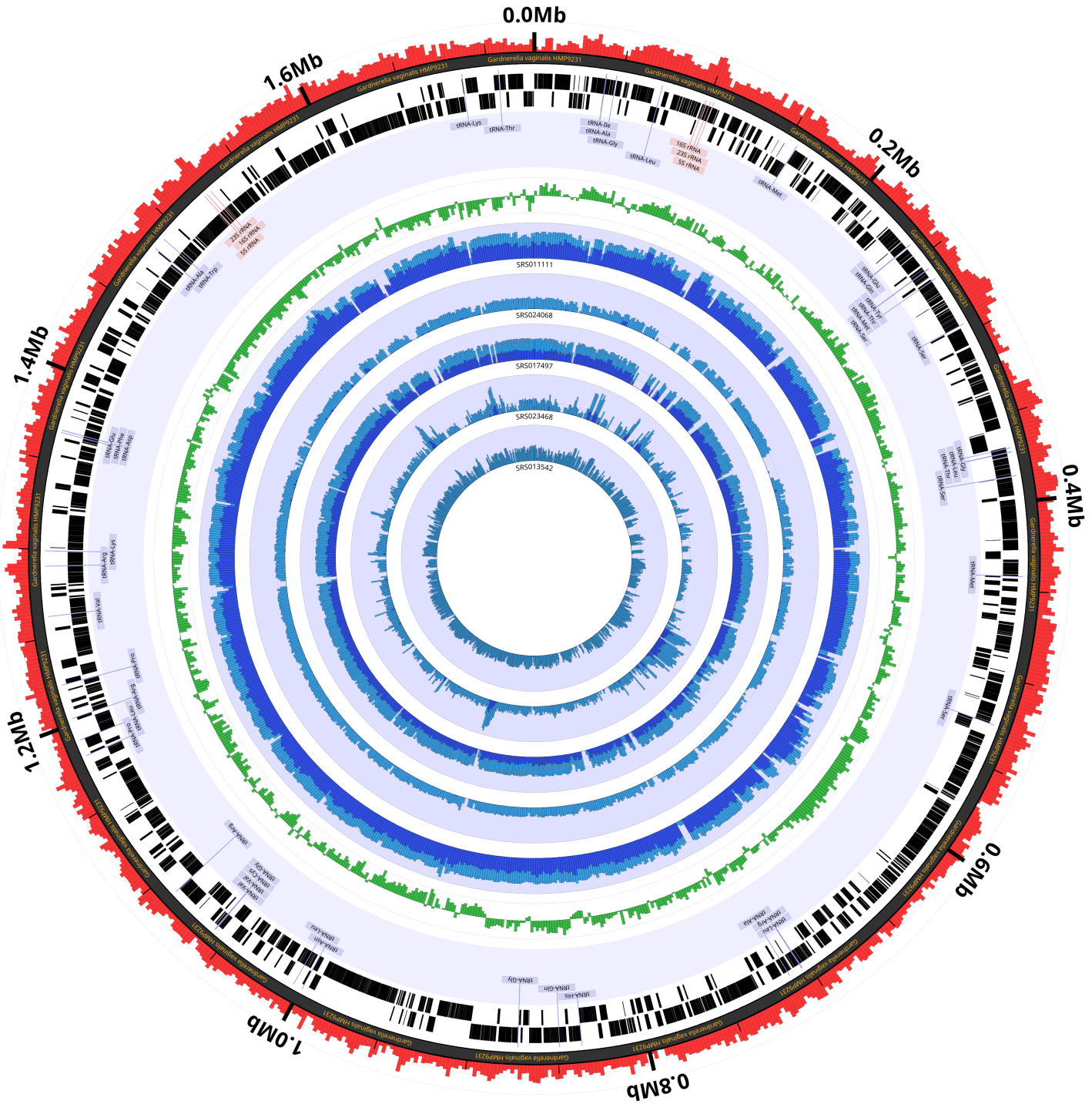

Figure 3: Fig. 1A from the paper. The genome of *Gardnerella vaginalis* HMP9231 annotated with percent GC content (red), genes, GC-skew (green), and read coverage (blue) from 5 human metagenomic samples.

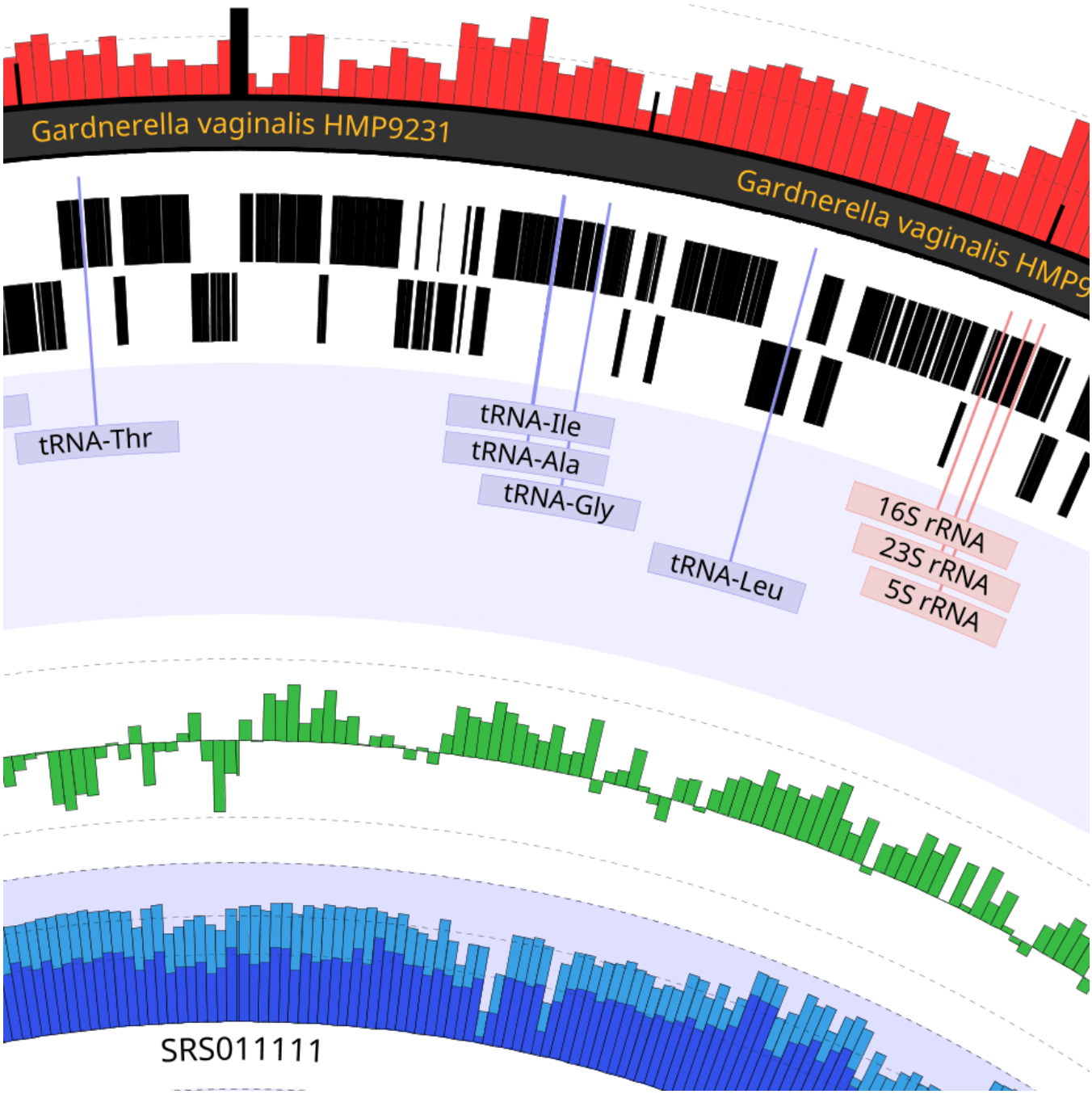

Figure 4: Detail from Fig. 3.

Fig. 3 depicts the 1,726,519 bp genome of *Gardnerella vaginalis* HMP9231, one of the reference genomes sequenced as part of the Human Microbiome Project [4]. The outermost ring/track displays the scale, with a tick mark every 0.05 Mb and a label every 0.2 Mb. Overlaid on the scale track is a percent GC-content track (red) using a nonoverlapping window size of 2kb and minimum and maximum y-axis values determined by the data. The next track is simply a black circle overlaid with the organism's strain name, repeated every 87kb. Inside that are tracks displaying the forward and reverse strand genes, respectively. Individual tRNA (light blue) and rRNA (peach) genes are shown in the next track along with lines indicating their positions relative to the protein-coding genes. The next track, in green, is a GC-skew graph, also using a window size of 2 kb. The remaining 5 tracks display read alignment histograms for 5 WMS (Whole Metagenomic Shotgun) sequencing runs using data from the Human Microbiome Project's Healthy Human Subjects (HHS) study [5]. Paired and singleton reads from the five indicated samples (NCBI Short Read Archive ids SRS011111, SRS024068, SRS017497, SRS023468, and SRS013542) were aligned to the *Gardnerella vaginalis* HMP9231 reference genome using Bowtie [1]. The resulting SAM files were converted to sorted BAM files and then summarized and log-transformed before being displayed by Circleator. Each set of alignments is displayed as a log-scaled read coverage histogram showing the average number of reads per nonoverlapping 2 kb window. Each singleton read alignment histogram is light blue and the corresponding paired read histogram is dark blue and overlaid on top of the graph for the singleton reads.

1.3.1 Fig. 3 Inputs

- The GenBank flat file for CP002725.1, downloaded from `http://www.ncbi.nlm.nih.gov/nuccore/CP002725`
- The Circleator configuration file shown in subsection 1.3.2
- Log-scaled tab-delimited singleton read coverage summary file for each of the 5 samples
- Log-scaled tab-delimited paired read coverage summary file for each of the 5 samples

1.3.2 Fig. 3 Configuration File (conf/CP002725-2.cfg)

This is the entire configuration file for Fig. 3. The string \\\ is used to indicate places where a single line in the configuration file must be wrapped over two or more lines in this document, but it is not valid Circleator syntax:

```
# %GC will be overlaid on coordinates because both have innerf=1.0
%GCmin-max opacity=0.8,heightf=0.06,innerf=1.0,window-size=2000,no-labels=1
coords innerf=1.0,label-interval=200000,tick-interval=50000

# filled circle overlaid with the genome name
new r8 rectangle 0.022 color1=#303030,color2=#000000,stroke-width=2
# label-repeat=87000 repeats the same label every 87 kb
new fl1 label 0.018 outerf=same,text-color=#ffba22,label-repeat=87000,packer=none,\\\
label-text=Gardnerella vaginalis HMP9231
new n1 none 0.004

# create invisible tRNA and rRNA features for subsequent linking
tRNAs trnas heightf=0.01,color1=none,color2=none
rRNAs rrnas innerf=same,outerf=same,color1=none,color2=none

# forward and reverse strand genes
tiny-cgap
genes-fwd heightf=0.03
tiny-cgap heightf=0.005
genes-rev heightf=0.03
tiny-cgap

# individual labeled rRNAs and tRNAs
new r1 rectangle heightf=0.005,color1=#efefff,color2=#efefff
new r2 rectangle heightf=0.1,color1=#efefff,color2=#efefff
# draw-link=1 links the individually-labeled features to the invisible ones in the previous track
# style=signpost specifies a label with a filled background and a line linking it to the thing that it labels
large-label heightf=0.05,outerf=same,feat-track=rrnas,feat-type=rRNA,style=signpost,label-function=rRNA_product,\\\
draw-link=1,color1=#f0d0d0,color2=#f09090,link-color=#f09090,stroke-width=1.5
large-label heightf=0.1,outerf=same,feat-track=trnas,feat-type=tRNA,style=signpost,label-function=product,\\\
draw-link=1,color1=#d0d0f0,color2=#9090f0,link-color=#9090f0,stroke-width=1.5,font-width-frac=1.5
new r3 rectangle heightf=0.005,color1=#efefff,color2=#efefff

# green GC-skew graph
small-cgap
GCskew-min-max-df0 heightf=0.07,color1=#38ba44,window-size=2000,no-labels=1

# 5 WMS samples

# SRS011111
small-cgap
# background shading
new r5 rectangle 0.07 color1=#dfdfff,color2=#dfdfff
# overlay paired and singleton alignment results
new bam2p graph 0.07 outerf=same,graph-function=FlatFile,file=data/SRS011111-p-cov-log10.txt,seq-id-regex=CP002725,\\\
graph-min=0,graph-max=3.5,window-size=2000,color1=#3aa0e5,no-labels=1
new bam2s graph outerf=same,innerf=same,graph-function=FlatFile,file=data/SRS011111-s-cov-log10.txt,\\\
seq-id-regex=CP002725,graph-min=0,graph-max=3.5,window-size=2000,color1=#2e2eec,no-labels=1,opacity=0.7
new fl label 0.02 label-text=SRS011111

# SRS024068
tiny-cgap
new r5 rectangle 0.07 color1=#dfdfff,color2=#dfdfff
# overlay paired and singleton alignment results
new bam2p graph 0.07 outerf=same,graph-function=FlatFile,file=data/SRS024068-p-cov-log10.txt,seq-id-regex=CP002725,\\\
graph-min=0,graph-max=3.5,window-size=2000,color1=#3aa0e5,no-labels=1
new bam2s graph outerf=same,innerf=same,graph-function=FlatFile,file=data/SRS024068-s-cov-log10.txt,\\\
seq-id-regex=CP002725,graph-min=0,graph-max=3.5,window-size=2000,color1=#2e2eec,no-labels=1,opacity=0.7
new fl label 0.02 label-text=SRS024068
```

```
# SRS017497
tiny-cgap
new r5 rectangle 0.07 color1=#dfdfff,color2=#dfdfff
# overlay paired and singleton alignment results
new bam2p graph 0.07 outerf=same,graph-function=FlatFile,file=data/SRS017497-p-cov-log10.txt,seq-id-regex=CP002725,\\\
graph-min=0,graph-max=3.5,window-size=2000,color1=#3aa0e5,no-labels=1
new bam2s graph outerf=same,innerf=same,graph-function=FlatFile,file=data/SRS017497-s-cov-log10.txt,\\\
seq-id-regex=CP002725,graph-min=0,graph-max=3.5,window-size=2000,color1=#2e2eec,no-labels=1,opacity=0.7
new fl label 0.02 label-text=SRS017497

# SRS023468
tiny-cgap
new r5 rectangle 0.07 color1=#dfdfff,color2=#dfdfff
# overlay paired and singleton alignment results
new bam2p graph 0.07 outerf=same,graph-function=FlatFile,file=data/SRS023468-p-cov-log10.txt,seq-id-regex=CP002725,\\\
graph-min=0,graph-max=3.5,window-size=2000,color1=#3aa0e5,no-labels=1
new bam2s graph outerf=same,innerf=same,graph-function=FlatFile,file=data/SRS023468-s-cov-log10.txt,\\\
seq-id-regex=CP002725,graph-min=0,graph-max=3.5,window-size=2000,color1=#2e2eec,no-labels=1,opacity=0.7
new fl label 0.02 label-text=SRS023468

# SRS013542
tiny-cgap
new r5 rectangle 0.07 color1=#dfdfff,color2=#dfdfff
# overlay paired and singleton alignment results
new bam2p graph 0.07 outerf=same,graph-function=FlatFile,file=data/SRS013542-p-cov-log10.txt,seq-id-regex=CP002725,\\\
graph-min=0,graph-max=3.5,window-size=2000,color1=#3aa0e5,no-labels=1
new bam2s graph outerf=same,innerf=same,graph-function=FlatFile,file=data/SRS013542-s-cov-log10.txt,\\\
seq-id-regex=CP002725,graph-min=0,graph-max=3.5,window-size=2000,color1=#2e2eec,no-labels=1,opacity=0.7
new fl label 0.02 label-text=SRS013542
```

1.3.3 Fig. 3 Command Line

The following commands were used to produce Fig. 3 in SVG and PNG format:

```
circleator.pl --data=data/CP002725.1.gb --config=conf/CP002725-2.cfg >CP002725-2.svg
rasterize-svg.pl CP002725-2.svg png 5000 5000
```

1.4 Fig. 5: *Yersinia pestis* CO92: An 80-genome SNP panel

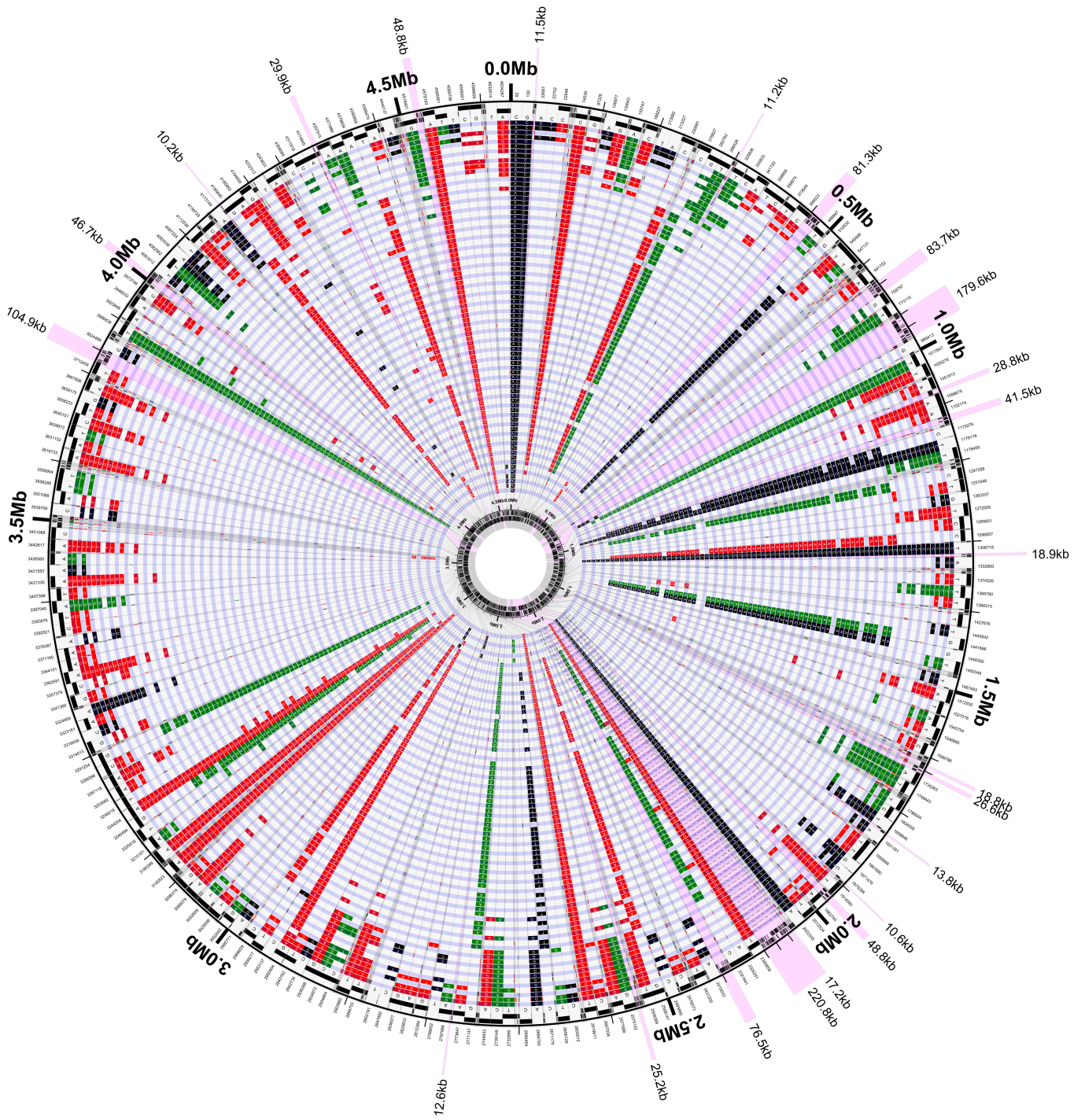

Figure 5: Fig. 1B from the paper. SNPs from an 80-genome *Yersinia pestis* SNP panel with the scale in the outer rings expanded to show the affected bases. The reference base and position is shown on the outside and SNPs are color-coded according to their predicted type.

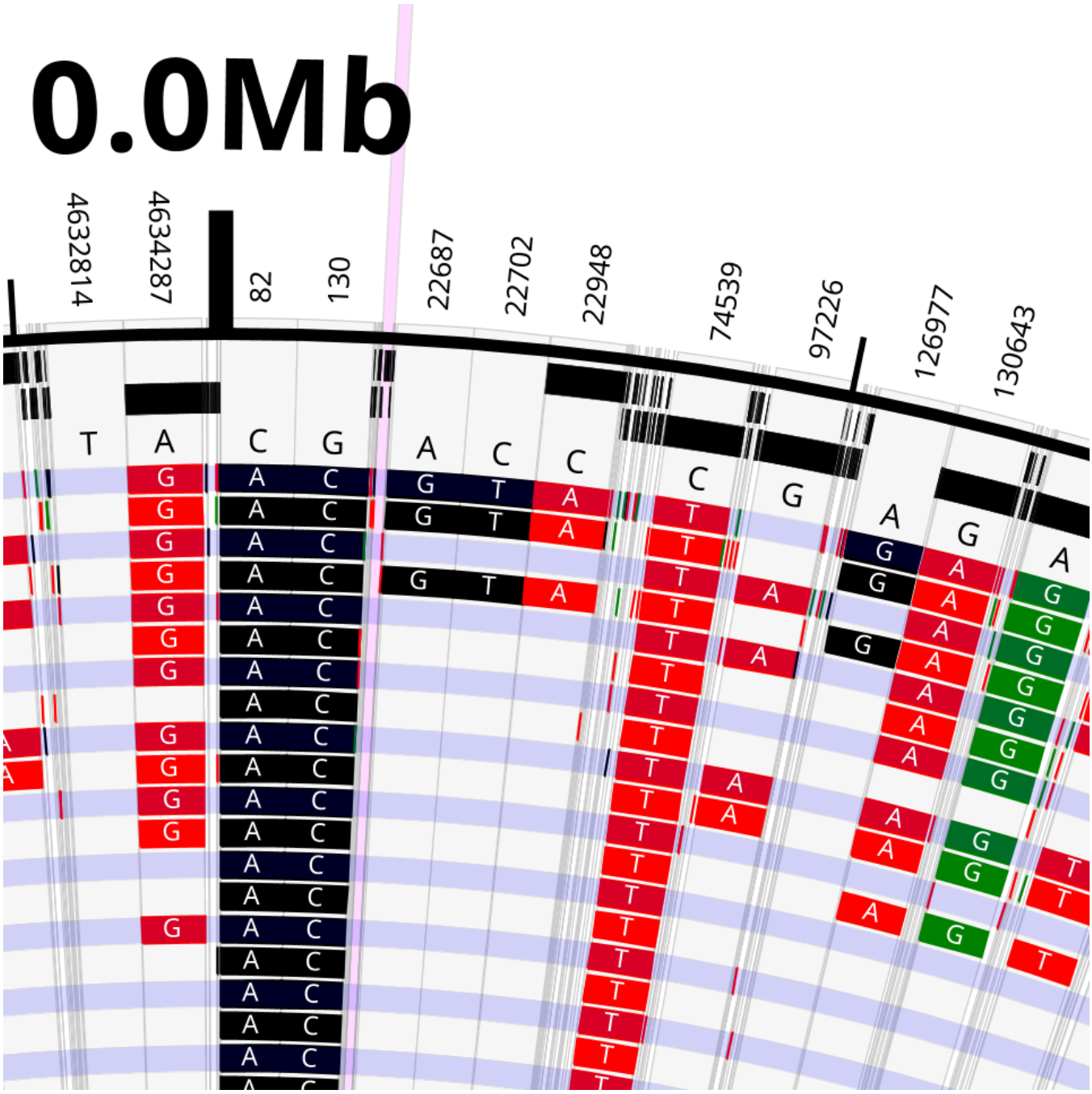

Figure 6: Detail from Fig. 5. Note the narrow red, green, and black SNP features without sequence information overlaid on them: these are SNP positions that were not expanded in Fig. 5 because fewer than 3 strains differed from the reference at those positions.

Fig. 5 depicts the 4,653,728 bp genome sequence of *Yersinia pestis* CO92. The outermost track displays the positions of SNP "deserts" in the reference genome (highlighted in pink with the length of the desert indicated next to each one.) These are regions of length 10 kb or greater in which no SNPs have been predicted (with respect to this particular set of 80 strains.) The next track displays the sequence scale, with a tick mark every 0.1 Mb and a label every 0.5 Mb. Note that the tick marks and labels are not evenly distributed around the circle because the sequence scale has been expanded (by a factor of 18,000) for a selected subset of the SNP positions. Regions that contain more of these SNPs will take up more of the circumference of the circle than those that contain fewer. The next track/set of labels, which is overlaid above the scale track, shows the base position of each expanded SNP. The next track shows the position of each annotated forward strand gene, and the next shows the reverse strand genes. The next track (the black bases with white background) shows the reference strain sequence at the indicated SNP position. Each of the subsequent tracks shows the corresponding non-reference strain sequence/base for that SNP. The background color of each sequence base indicates the category into which it has been placed: green for SNPs that overlap a protein-coding gene but are predicted to be synonymous (when considered in isolation), red for potentially nonsynonymous SNPs, and black for intergenic SNPs (all with respect to the reference strain's gene annotation.) Every

other SNP track has a light purple background to make it easier to discern adjacent tracks. Finally, inside all of the SNP tracks there is a second sequence scale and forward and reverse gene annotation. This second and much smaller figure shows the positions of the SNPs and SNP deserts *without* distorting the coordinate system; the lines between the inner and outer figures illustrate the distortion caused by the expansion of the SNP loci. In this particular figure only SNP loci where at least 3 of the strains differ from the reference have been expanded to show the base-level details.

1.4.1 Fig. 5 Inputs

- The RefSeq flat file for NC_003143.1, downloaded from `http://www.ncbi.nlm.nih.gov/nuccore/NC_003143.1`
- The Circleator configuration file shown in subsection 1.4.2
- A tab-delimited file of SNPs/indels in the **MergedTable** format supported by the Ergatis bioinformatics workflow system [2].

1.4.2 Fig. 5 Configuration File (conf/yp-CO92-fig1b.cfg)

This is the entire configuration file for Fig. 5. The string \\\ is used to indicate places where a single line in the configuration file must be wrapped over two or more lines in this document, but it is not valid Circleator syntax:

```
# change the sequence scale, expanding each SNP location where 3 or more strains differ from the
# reference (snp-min-diffs=3) by a factor of 18000X (scale=18000)
new SNP1 scaled-segment-list 0 scale=18000,feat-type=SNP,feat-file=./filtered_labeled_merged_table.txt,\\\
feat-file-type=merged-table-snp,snp-min-diffs=3
# load _all_ the SNPs (not just those with 3 or more differences) but don't display anything yet
new SNP_ALL none 0 feat-type=SNP,feat-file=./filtered_labeled_merged_table.txt,feat-file-type=merged-table-snp

# compute SNP deserts of 10kb or more, but don't display them yet
new SNP_D compute-deserts desert-min-length=10000,desert-feat-type=SNP_desert,feat-track=SNP_ALL,feat-type=SNP

# highlight the SNP deserts in light pink/purple
new SD_L2 rectangle outerf=0.1595,innerf=0.12,opacity=0.15,color1=#ff00ff,inner-scale=none,outer-scale=default,\\\
feat-type=SNP_desert
new SD_L3 rectangle outerf=0.12,innerf=0.08,opacity=0.15,color1=#ff00ff,inner-scale=none,outer-scale=none,\\\
feat-type=SNP_desert
new SD_L4 rectangle innerf=0.1595,outerf=1.15,opacity=0.15,color1=#ff00ff,feat-type=SNP_desert
new SDL1 label innerf=1.16,heightf=0.04,label-type=spoke,label-text-anchor=start,feat-type=SNP_desert,\\\
label-function=length_kb,packer=none

# highlight all SNP loci in light grey
new SNP_HL1 rectangle innerf=0.1595,outerf=1.035,feat-track=SNP_ALL,opacity=0.2,color1=#d0d0d0
# label the positions of only the expanded SNP loci (those with 3 or more differences)
# feat-track=SNP1 is a reference to the previous track that defined this set of features
new SNP_L1 label innerf=1.04,heightf=0.04,feat-track=SNP1,label-function=position,label-type=spoke,\\\
color1=#000000,label-text-anchor=start

# outer coordinate circle and forward and reverse strand genes
coords outerf=1.05
new miniscule-cgap gap 0.002
miniscule-cgap
genes-fwd heightf=0.01
miniscule-cgap
genes-rev heightf=0.01

# display the reference base sequence (A,C,T,G, etc.) at each expanded SNP locus
new SNP_RB1 label 0.015 feat-track=SNP1,label-function=snp_ref_base,text-color=#000000,packer=none

################################
# LOOP BEGIN
################################

# loop over 79 non-reference SNP strains
# This is long, but not as long as repeating the 5 tracks in the loop 79 times!
new LS1 loop-start loop-var=TARGETS,loop-values=Angola_<84_Hum_Ang_0PE3|PestoidesF_<84_Hum_USSR_0PE2|\\\
Harbin35_NA_Hum_Chi_NA|91001_70_Hum_Chi_0PE4|KIM_68_Hum_Kur_2MED1|UG05-0454_04_Hum_Uga_1ANT1|\\\
Antiqua_65_Hum_Con_1ANT1|D182038_82_AChe_Chi_NA|Nepal516_67?_Hum_Nep_2ANT1|K1973002_73_Hum_Chi_2MED2|\\\
D106004_06_Hum_Chi_NA|B42003004_03_Hum_Chi_0ANT2|PY66_10_RatN_Lib_Tru_Sal_Ena|PY54_10_Dog_Lib_Asc_Chi_Pea|\\\
Z176003_76_Mar_Tib_NA|PY103_10_Did_Lib_Asc_Chi_Lib|PY03_06_RatR_Caj_Jae_Dom_Cha|PY14_09_RatR_Lib_Asc_Cas_Tom|\\\
E1979001_79_EotM_Chi_1IN3ANT|PY64_10_RatN_Lib_Tru_Sal_Ena|PY02_06_RatR_Caj_Jae_Dom_Cha|\\\
PY58_10_RatN_Lib_Asc_Chi_Lib|PY05_06_RatR_Caj_Cho_Lic_Pay|PY63_10_RatN_Lib_Tru_Sal_Ena|Pexu2_66_Hum_Bra_ORI|\\\
PY36_10_Mus_Lib_Asc_Cas_Moc|PY89_10_RatN_Lib_Pas_Gua_Mar|PY94_10_RatR_Caj_Cho_NA_NA|\\\
PY01_06_Mus_Caj_Cho_Lic_Por|PY32_10_Hum_Lib_Asc_Chi_Chi|PY88_10_RatR_Lib_Pas_Gua_Mar|\\\
PY52_10_Dog_Lib_Asc_Chi_Pea|PY29_10_Mus_Lib_Asc_Chi_Lib|PY90_10_RatN_Lib_Pas_Gua_Mar|\\\
```

```
PY15_09_Hum_Lib_Asc_Cas_Tom|PY96_10_Dog_Lib_Asc_Chi_Sau|PY93_10_RatR_Caj_Cho_NA_NA|\\\
PY99_10_RatN_Lib_Asc_Chi_Lib|PY16_09_Hum_Lib_Asc_Cas_Tom|PY47_10_Cat_Lib_Asc_Cas_Moc|\\\
PY95_10_RatR_Caj_Cho_NA_NA|PY56_10_Dog_Lib_Asc_Chi_Lib|PY48_10_Mus_Lib_Asc_Chi_Lib|\\\
PY65_10_RatN_Lib_Tru_Sal_Ena|PY59_10_RatR_Lib_Asc_NA_NA|PY34_10_Cat_Lib_Asc_Chi_Lib|\\\
PY25_10_Mus_Lib_Asc_Chi_Lib|PY98_10_Hum_Lib_Asc_NA_NA|PY61_10_RatN_Lib_Tru_Mil_S9|\\\
PY45_10_Mus_Lib_Asc_Chi_Lib|PY53_10_Dog_Lib_Asc_Chi_Pea|PY11_07_Hum_Caj_Cho_NA_NA|\\\
IP275_95_Hum_Mad_1ORI3|F1991016_91_RatF_Chi_1ORI2|PY92_10_RatR_Caj_Cho_NA_NA|PY13_08_Hum_Caj_Cru_NA_NA|\\\
PY10_06_Hum_Caj_Jae_NA_NA|PY08_06_Hum_Caj_Jae_NA_NA|PY113_10_Hum_Lib_Asc_NA_NA|PY04_06_RatR_Caj_Cho_Lic_Pay|\\\
PY102_10_Did_Lib_Asc_Chi_Sau|MG05-1020_05_Hum_Mad_1ORI3|PY55_10_Dog_Lib_Asc_Chi_Pea|\\\
PY100_10_RatN_Lib_Asc_Chi_Lib|PY101_10_Hum_Lib_Asc_Cao_Car|PY46_10_Cat_Lib_Asc_Asc_Ant|\\\
PY91_10_RatR_Caj_Cho_NA_NA|PY12_08_Hum_Caj_Jae_NA_NA|PY09_06_Hum_Caj_Jae_NA_NA|PY42_10_Mus_Lib_Asc_Chi_Lib|\\\
PY60_10_RatN_Lib_Tru_Mil_S9|PY07_06_RatR_Caj_Cho_Lic_Pay|PY06_06_Gui_Caj_Cho_Lic_Pay|\\\
PY72_10_RatR_Lib_Asc_Chi_Lib|India195_NA_Hum_Ind_ORI|PY71_10_RatR_Lib_Asc_Chi_Lib|\\\
PY19_10_Aeg_Lib_Asc_Cas_Tom|PY76_10_Sig_Lib_Asc_Chi_Lib|CA88-4125_88_Hum_USA_1ORI1

miniscule-cgap
# draw all SNPs for the specified target genome
new SNP rectangle 0.0085 feat-track=SNP_ALL,featd-type=SNP,stroke-width=1,color1=snp_type,color2=snp_type,\\\
snp-query=<TARGETS>,color1=snp_type,color2=snp_type
# label track with target genome name
small-label label-text=<TARGETS>,innerf=same,outerf=same,label-position=2180000,label-text-anchor=middle,\\\
color1=#ffffff,opacity=0.6,packer=none
# highlight every other track, regardless of target genome:
new SNPC rectangle innerf=same,outerf=same,feat-type=contig,color1=#0000ff,opacity=0.15,skip-track=<LOOP_EVEN>
# label only expanded/highlighted SNPs with the target base at each position
new SNPL label 0.012 innerf=same,feat-track=SNP1,label-function=snp_base,text-color=snp_text,\\\
snp-query=<TARGETS>,packer=none

# the end of the loop: the tracks in the middle will be repeated 79 times, once for each target strain
# listed in the loop-start line
new LS1 loop-end

################################
# LOOP END
################################

# SNP loci - connect outer scaled circle to inner unscaled circle
new SNP_HL2 rectangle innerf=0.12,outerf=0.1595,feat-track=SNP_ALL,opacity=0.2,color1=#d0d0d0,\\\
color2=#000000,inner-scale=none,outer-scale=default

# restore unscaled coordinate system
new SSLR scaled-segment-list scale=1,feat-type=dummy

# display unscaled SNPs
new SNPS_inner rectangle outerf=0.12,innerf=0.08,feat-track=SNP_ALL,feat-type=SNP,stroke-width=1,\\\
color1=#d0d0d0,color2=#d0d0d0

# unscaled inner coordinate circle and genes (the figure within a figure)
coords innerf=0.12,font-size=12,heightf=0.006
miniscule-cgap
genes-fwd heightf=0.01
miniscule-cgap
genes-rev heightf=0.01
tiny-cgap
```

1.4.3 Fig. 5 Command Line

The following commands were used to produce Fig. 5 in SVG and PNG format:

```
circleator --data=./data/NC_003143.1.gbk --config=./conf/yp-CO92-fig1b.cfg > yp-CO92-fig1b.svg
rasterize-svg yp-CO92-fig1b.svg png 5000 5000
```

References

[1] Langmead, B. *et al.* (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, *Genome Biology*, **10:R25**.

[2] Orvis, J. *et al.* (2010) Ergatis: a web interface and scalable software system for bioinformatics workflows. *Bioinformatics*, **26:12**, 1488-1492.

[3] Rasko, D. *et al.* (2005) Visualization of comparative genomic analyses by BLAST score ratio, *BMC Bioinformatics*, **6**, 2.

[4] Human Microbiome Jumpstart Reference Strains Consortium. (2010) A catalog of reference genomes from the human microbiome. *Science*, **328:5981**, 994-999.

[5] The Human Microbiome Project Consortium. (2012) Structure, function and diversity of the healthy human microbiome. *Nature*, **486**, 207-214.