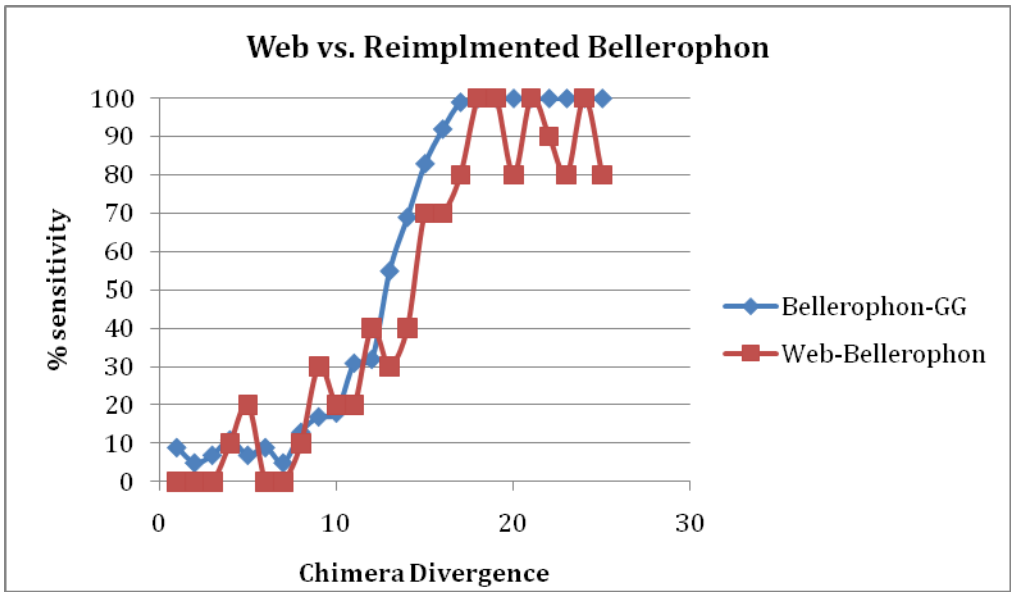
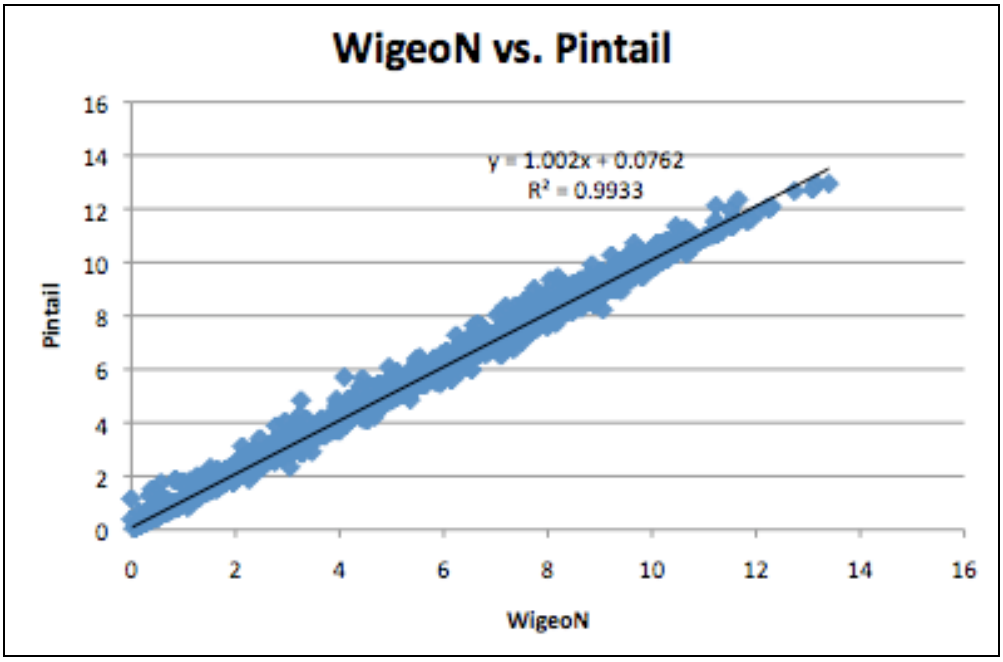
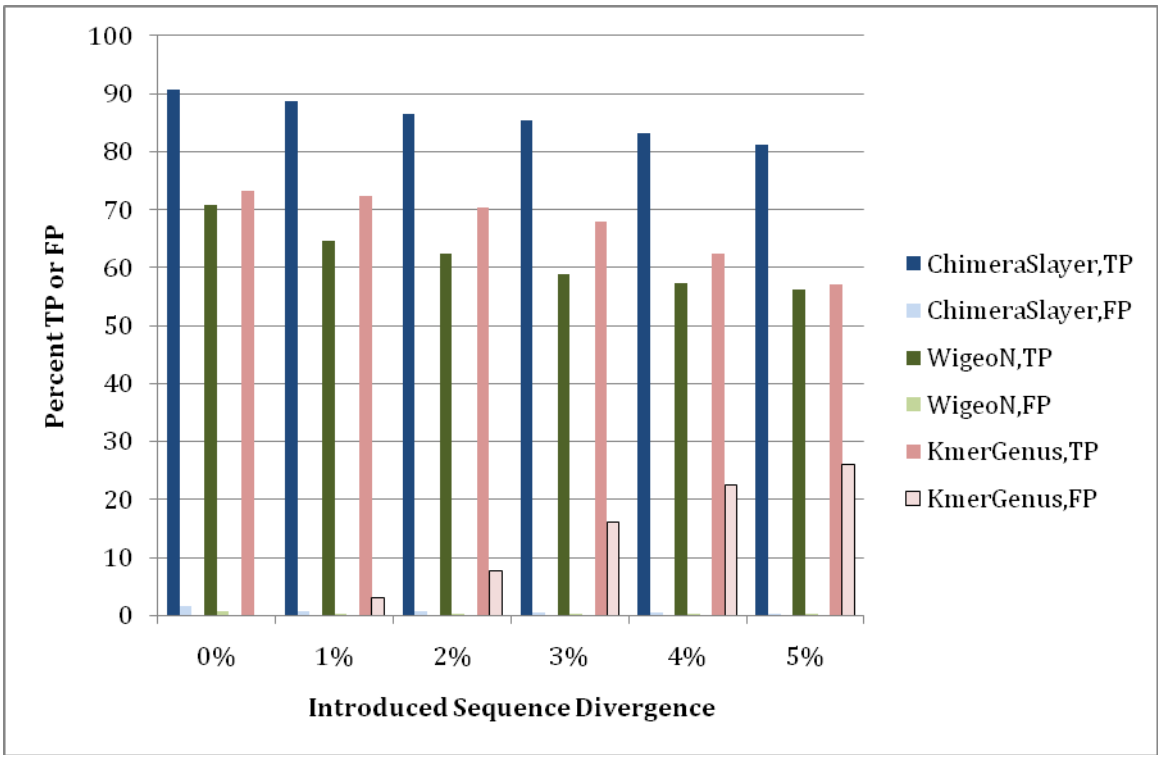
a.



b.

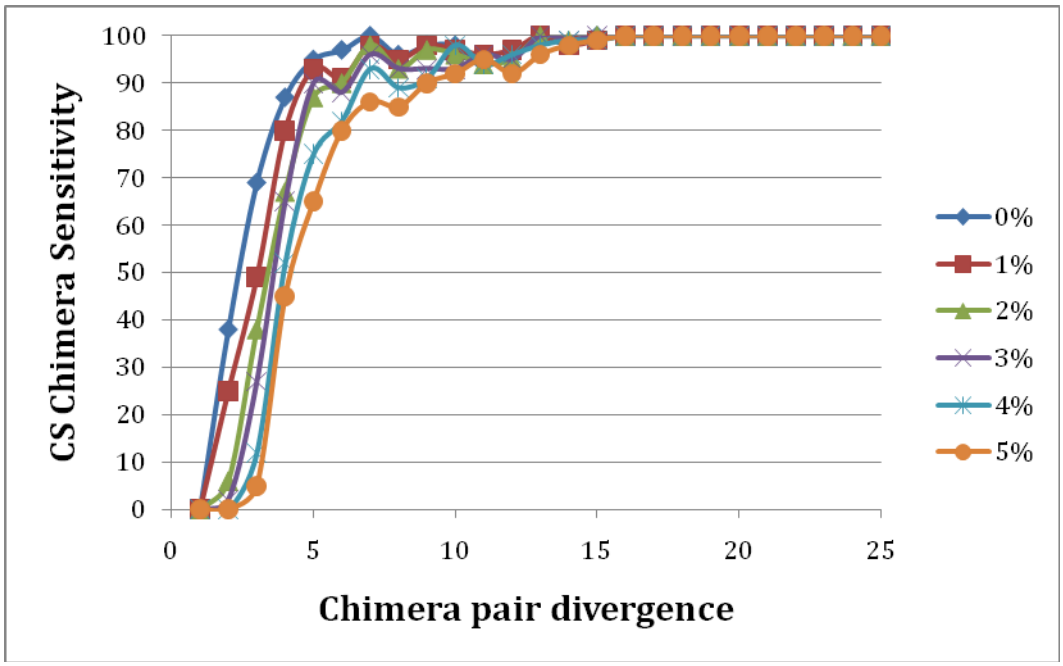


Supplementary Figure S1: Comparison of reimplemented Bellerophon (GreenGenes) and Pintail to outputs derived from the original tools. (a) Comparison of chimera detection sensitivity according to chimera pair divergence for the GreenGenes Bellerophon run from the GreenGenes website and the reimplemented algorithm in a utility named BellerophonGG. (b) Deviation from expected (DE) values are plotted for Pintail and WigeoN based on synthetic chimeric sequences included in the test regime.

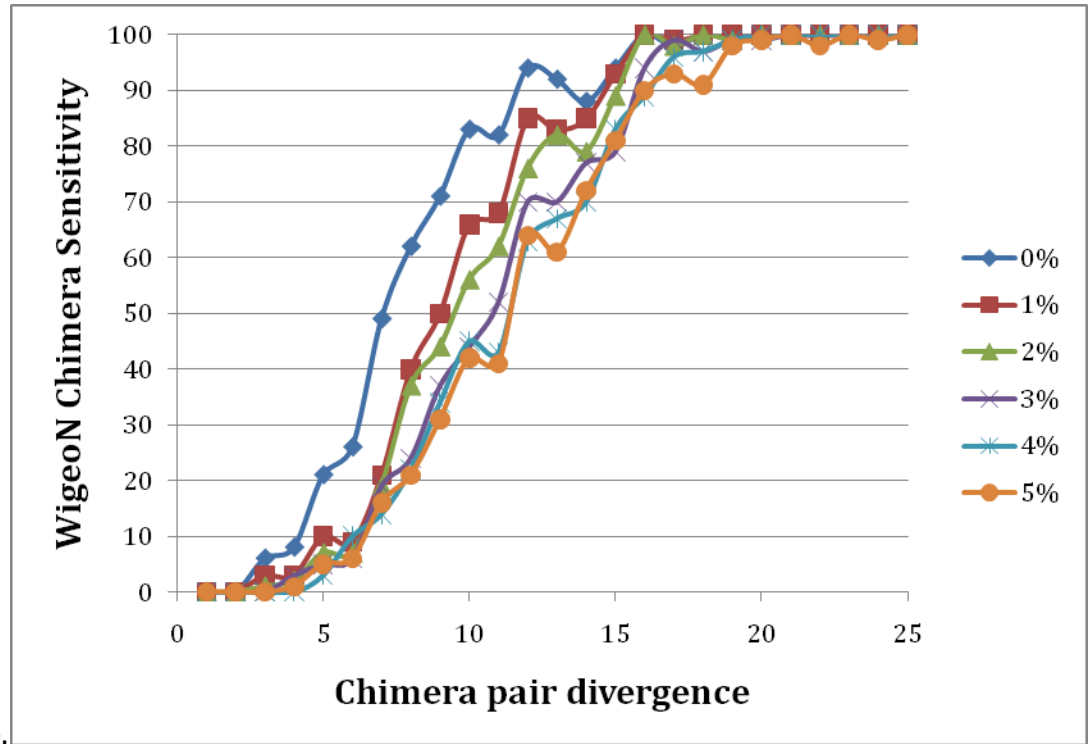


Supplementary Figure S2: Chimera detection accuracy with chimera sequences diverged from parental reference sequences. Cumulative true positive (TP) and false positive (FP) rates are shown for methods CS, WigeoN, and KmerGenus. KmerGenus exhibits higher FP with chimera sequences increasingly diverged from parental sequences.

a.

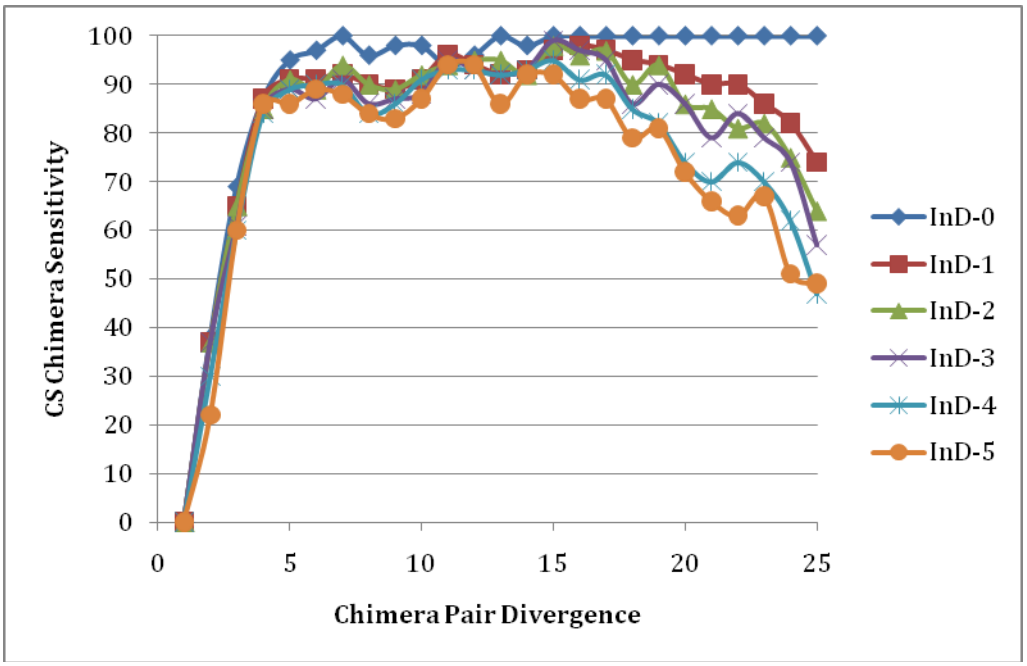


b.

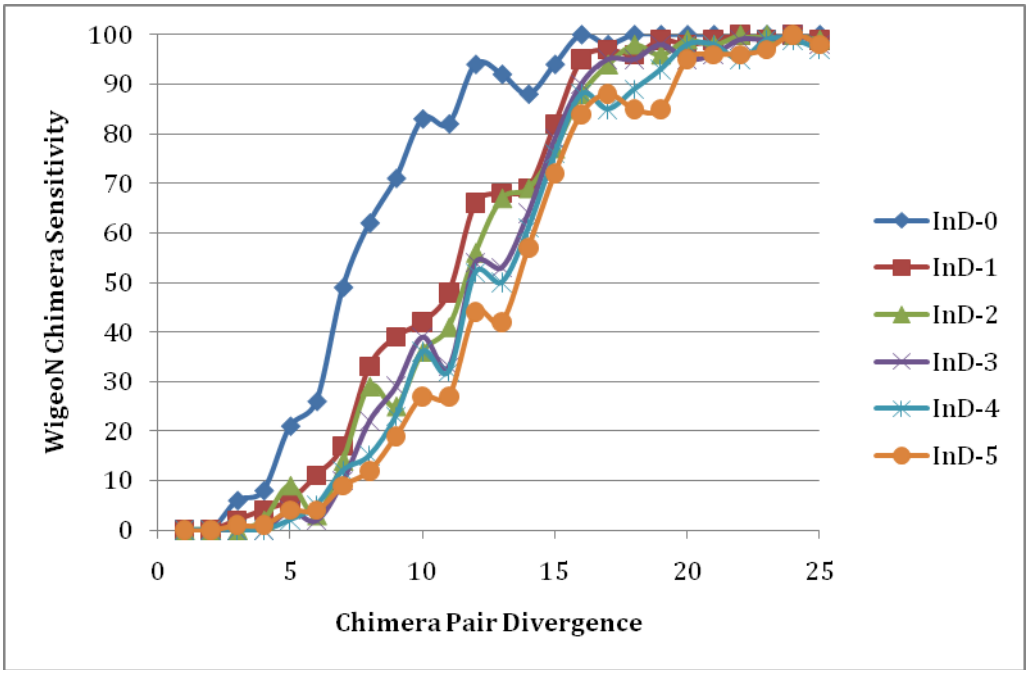


Supplementary Figure S3: Chimera detection sensitivity for (a) CS and (b) WigeoN according to chimera pair divergence and included sequence variation resulting from random **nucleotide substitutions**.

a.

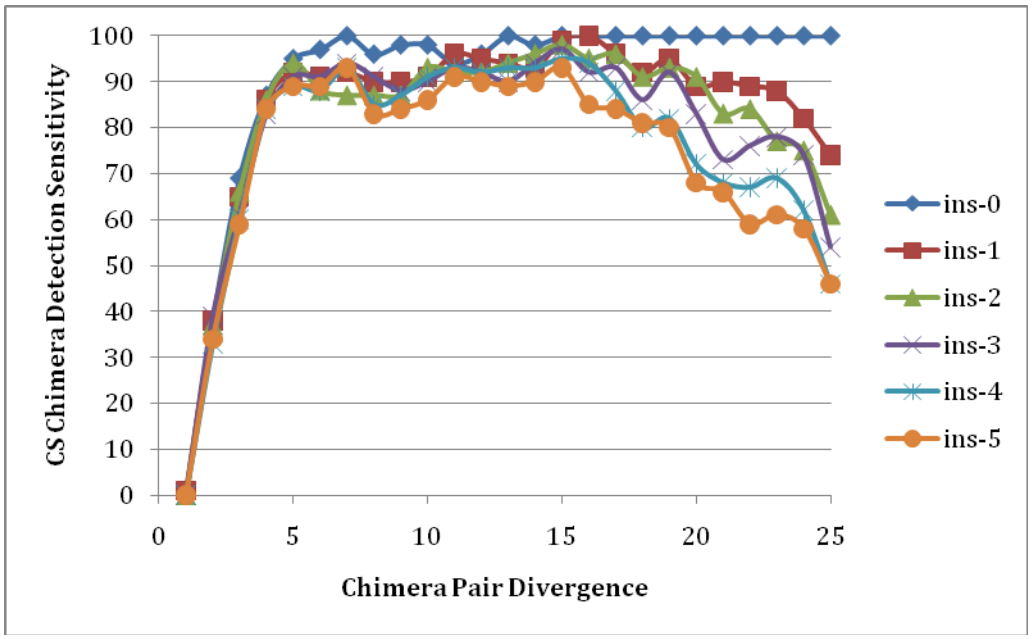


b.

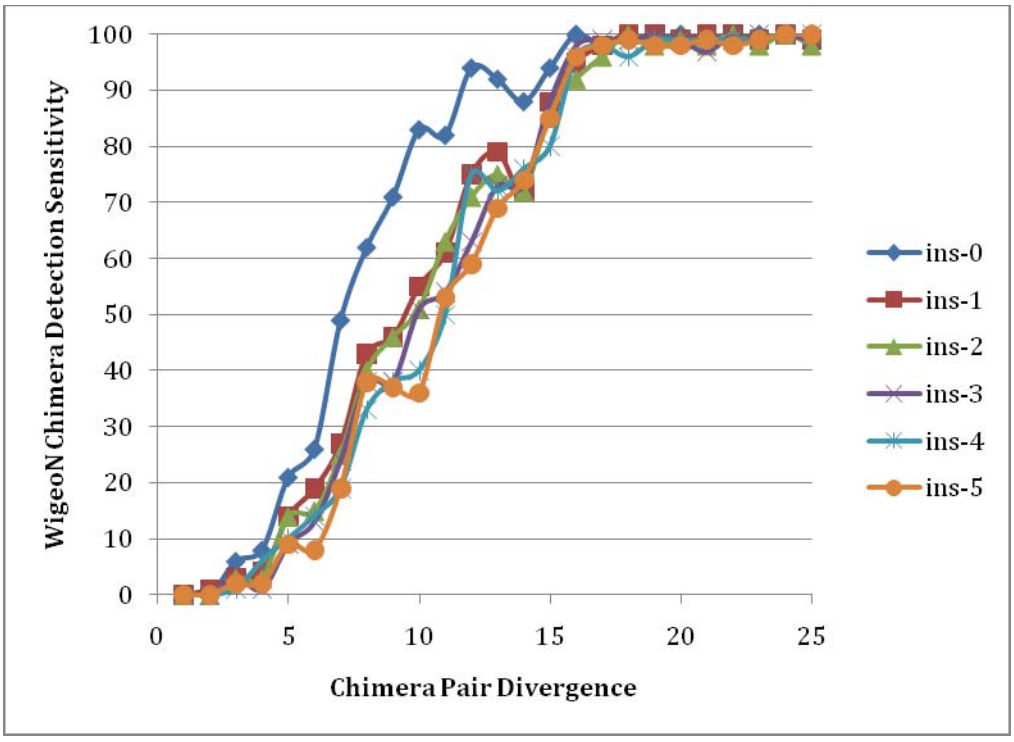


Supplementary Figure S4: Chimera detection sensitivity for (a) CS and (b) WigeoN according to chimera pair divergence and included sequence variation resulting from random **insertion and deletion (InD)** mutations. Mutation percentages are shown as InD-0% through 5%.

a.

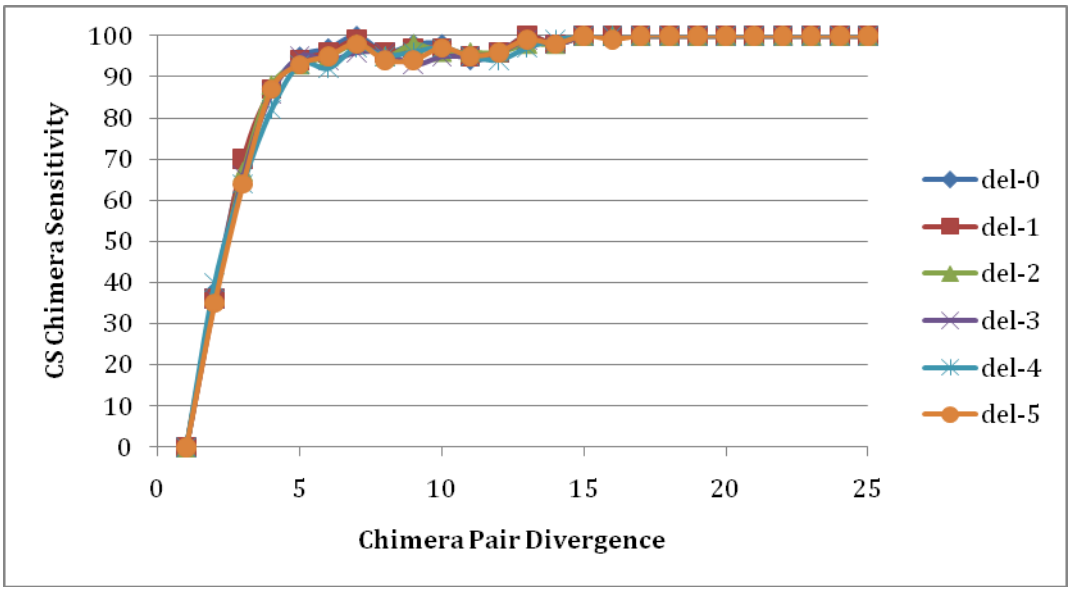


b.

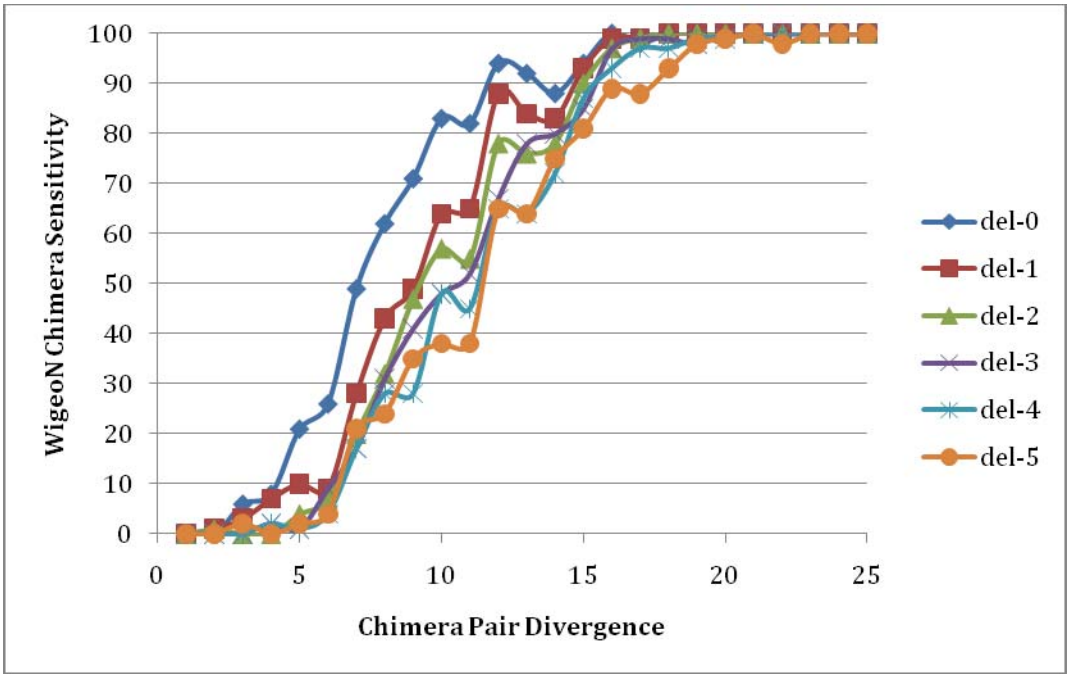


Supplementary Figure S5: Chimera detection sensitivity for (a) CS and (b) WigeoN according to chimera pair divergence and included sequence variation resulting from random **insertion** mutations. Insertion mutation percentages are shown as ins-0% through 5%.

a.



b.

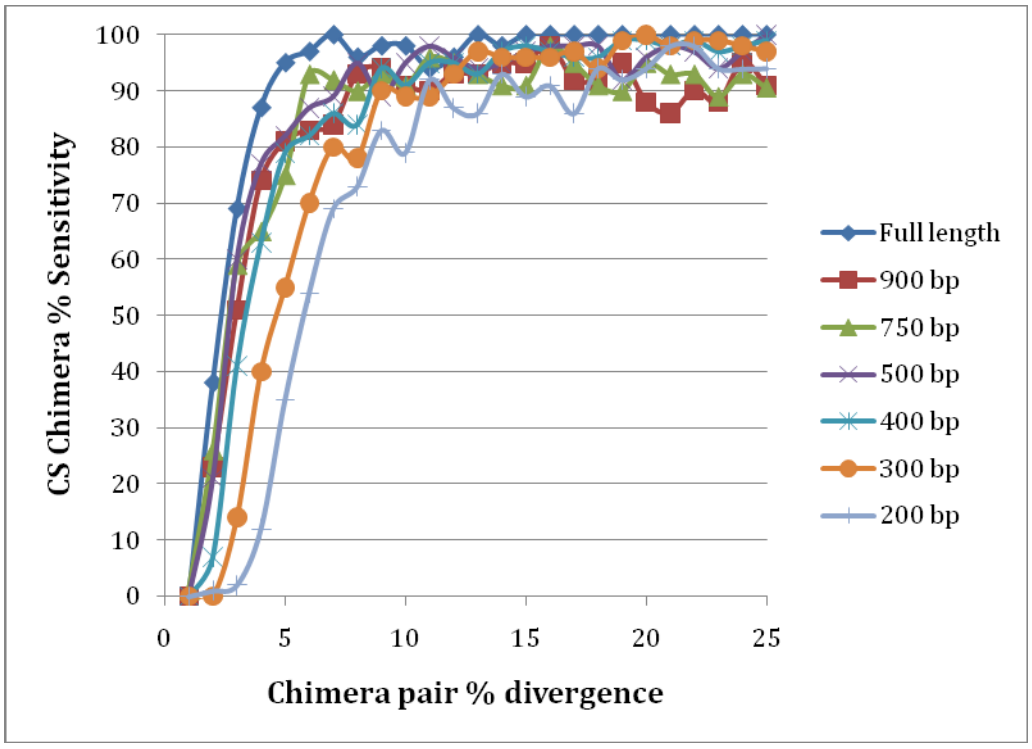


Supplementary Figure S6: Chimera detection sensitivity for (a) CS and (b) WigeoN according to chimera pair divergence and included sequence variation resulting from random **deletion** mutations. Deletion mutations are shown for del-0% through 5%.

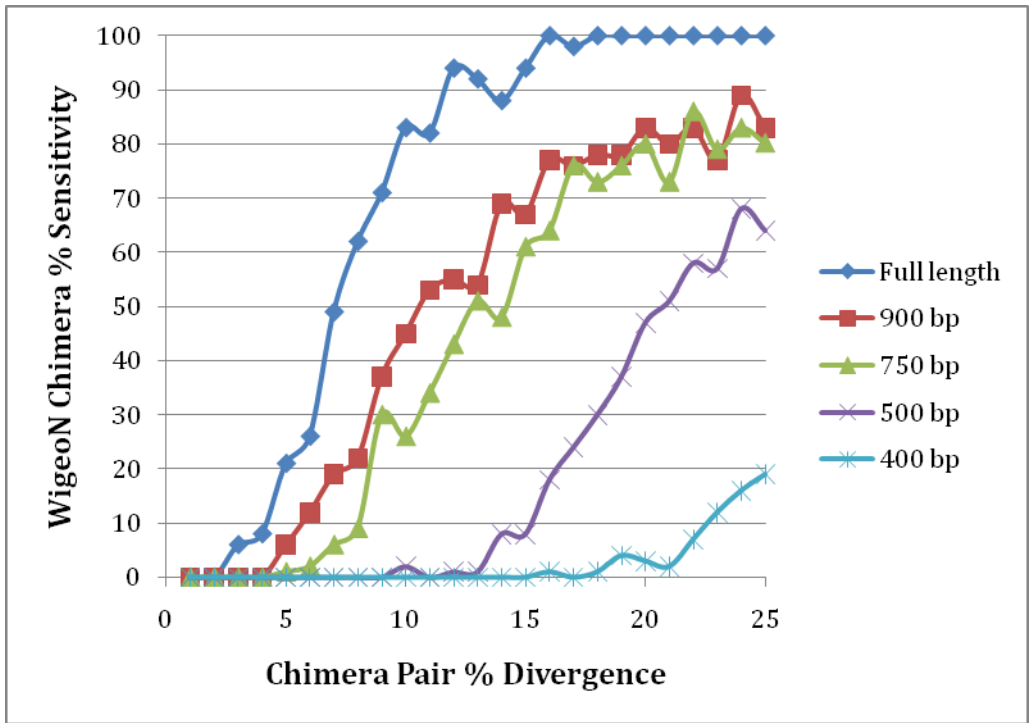
```
ChimeraSlayer
S000004105    # query accession (Mycobacterium pulveris)
S000002743    # parent A
S000015160    # parent B

# Parent A information:
Mycobacterium
Mycobacterium elephantis (T); AJ010747 Mycobacterium elephantis

# Parent B information:
Mycobacterium
Mycobacterium rhodesiae (T); DSM 44223T; AJ429047 Mycobacterium rhodesiae

# intra-taxon chimera type
GENUS

            Per_id parents: 97.09

           Per_id(Q,A): 99.05
-------------------------------------------------- A: S000002743
 99.79                                  97.68
~~~~~~~~~~~~~~~~~~~~~~~~\ /~~~~~~~~~~~~~~~~~~~~~~~~ Q: S000004105
DivR: 1.008 BS: 100.00   |
Per_id(QLA,QRB): 99.80   |
                         |
   (L-AB: 96.67)         |      (R-AB: 97.87)
   WinL:0-959            |      WinR:960-1476
                         |
Per_id(QLB,QRA): 97.02   |
DivR: 0.979 BS: 0.00     |
~~~~~~~~~~~~~~~~~~~~~~~~/ \~~~~~~~~~~~~~~~~~~~~~~~~ Q: S000004105
 96.67                                  99.81
-------------------------------------------------- B: S000015160
            Per_id(Q,B): 97.77

DeltaL: 3.12                   DeltaR: -2.13

!!
CGCACCTCTGTCCCGGGTGCTCCGTGGCCGTGG  A: S000002743
TACACCTCTGTCCCGGGTGCTCCGTGGCCGTGG  Q: S000004105
CAGGATATGCCTGTTTAAATGGGCACCTGCCTA  B: S000015160
! !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

             ** Breakpoint **

!!!!!!!!!!!!
TGTTCTCGCAAA  A: S000002743
GTCGGGGTTGTT  Q: S000004105
GTCGGGGTTGTA  B: S000015160
           !
```

Supplementary Figure S7: ChimeraSlayer (CS) prediction of the 16S reference sequence of *Mycobacterium pulvis* as a chimera between *Mycobacterium elephantis* and *Mycobacterium rhodesiae* sequences. The alignment provided by CS includes only the informative variations within the multiply aligned sequences. The bootstrap (BS) value of 100% supports a chimera formed by joining the upper left (A,Q) and lower right (B,Q) of the centrally illustrated breakpoint region.

a.

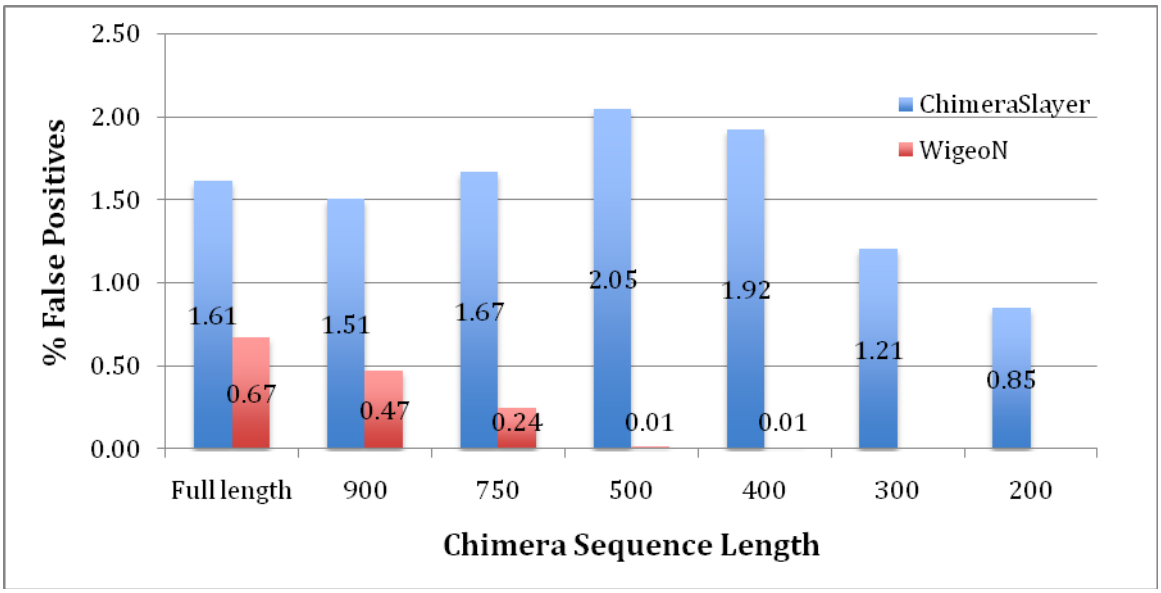


b.

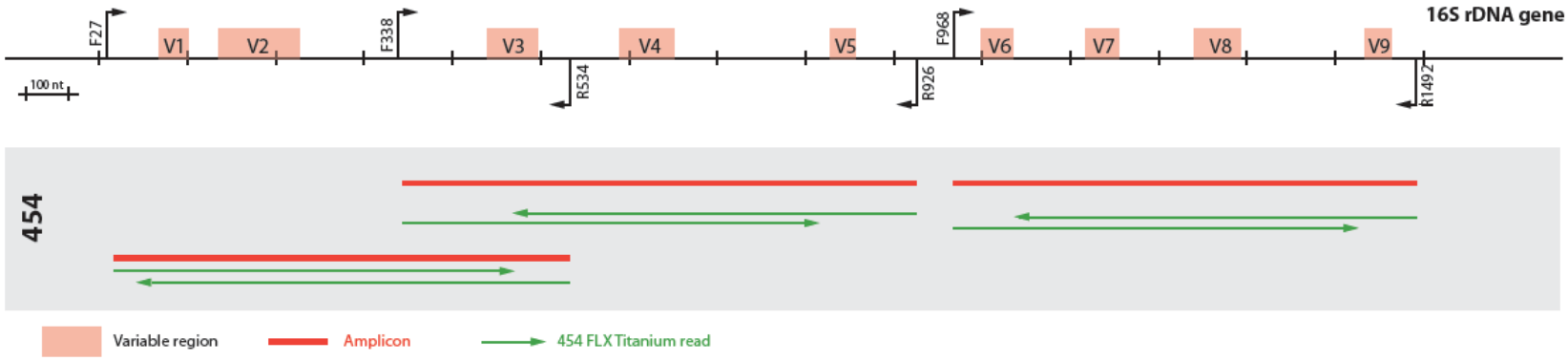


Supplementary Figure S8 continued...

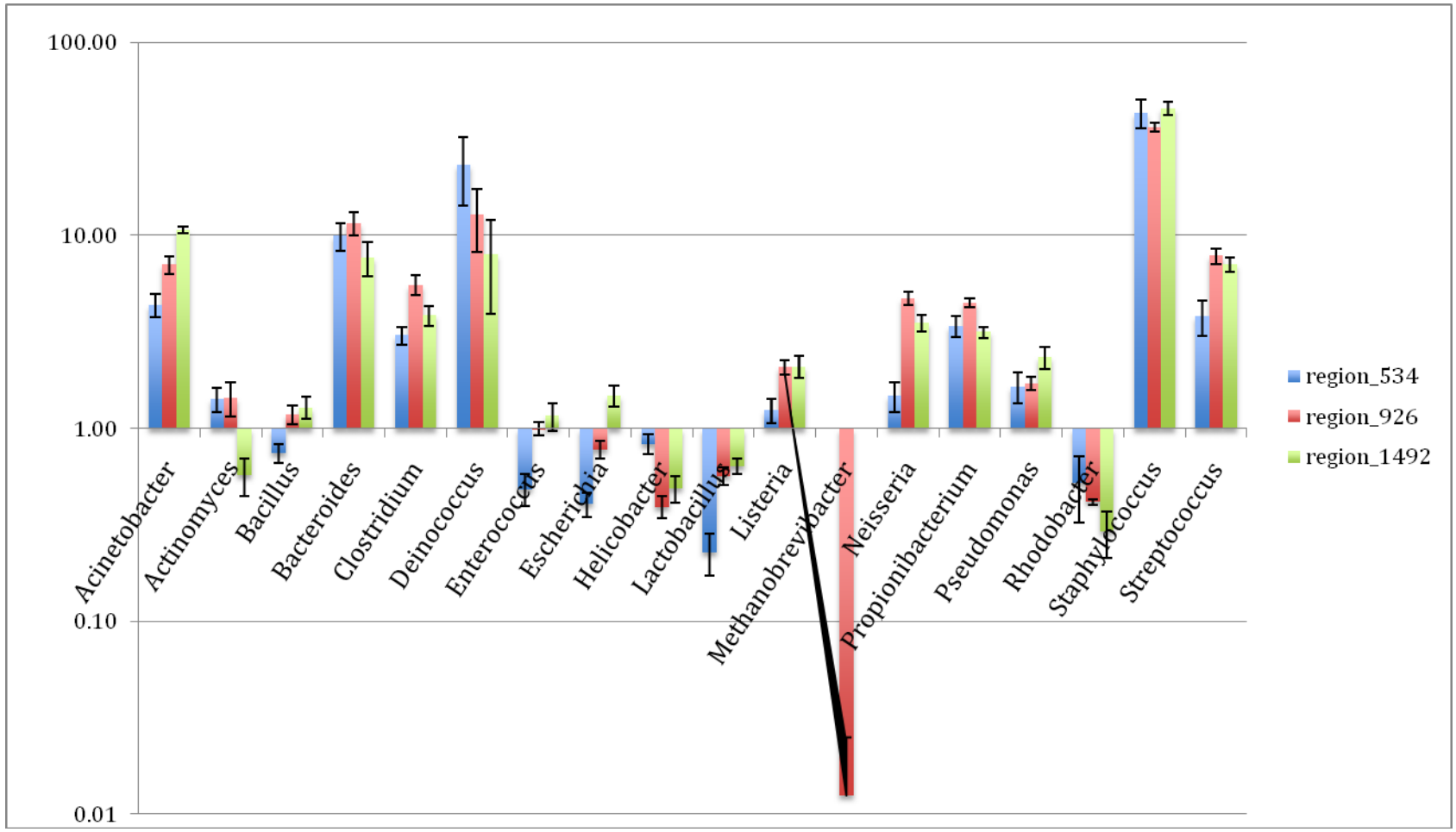
c.



Supplementary Figure S8: Chimera detection accuracy for variable length chimeras. Sensitivity according to chimera divergence and sequence length for (a) CS and (b) WigeoN. (c) Cumulative false positives for CS and WigeoN according to sequence length.

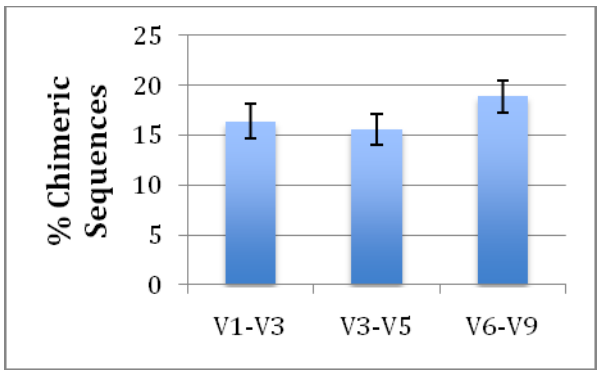


Supplementary Figure S9: Three regions targeted for PCR and subsequent 454 pyrosequencing are shown. Primer sequences and references are as follows: primers: **63f** GCCTAACACATGCAAGTC;**357f** CCTACGGGAGGCAGCAG; **968F** AACGCGAAGAACCTTAC as described in Yu, Z., and M. Morrison. 2004; Primer: **1525R** AAGGAGGTGWTCCARCC as described in Larkin, MJ, Osborn, AM, & Fairly, D. 2005; primers: **27F** AGAGTTTGATCCTGGCTCAG**; 338F** TGCTGCCTCCCGTAGGAGT as described in Hamady M, W. J., Harris JK, Gold NJ, Knight R. 2008; primer: **926R**: (modified primer 907r) CCGTCAATTCMTTTRAGT based on Lane, D. J. 1991. 16S/23S rRNA sequencing; and primers: **515F** GTGCCAGCMGCCGCGGTAA; **519F**: CAGCMGCCGCGGTAATWC as described in Baker, G. C., J. J. Smith, and D. A. Cowan**.** 2003.

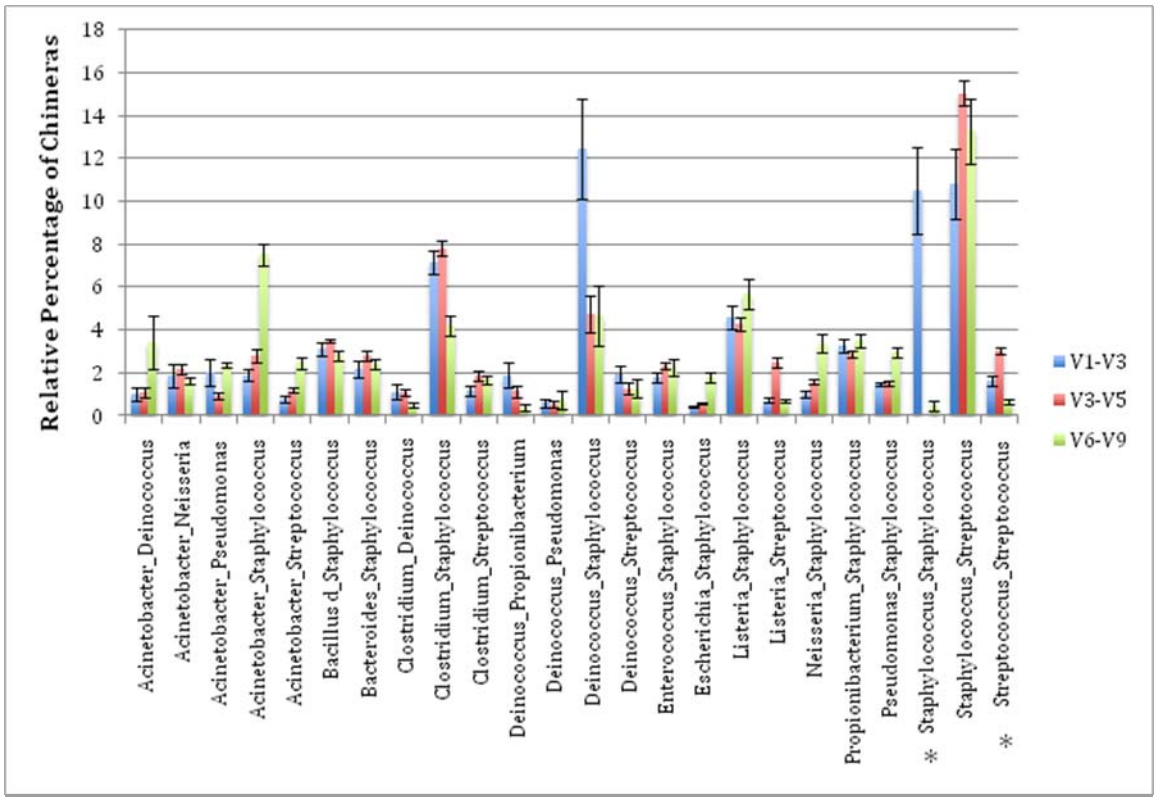


Supplementary Figure S10: Inferred abundance of organisms in the uniform mock community assayed by PCR-targeted regions and 454 pyrosequencing. Average values of four replicates are shown. Error bars correspond to standard error from the mean.

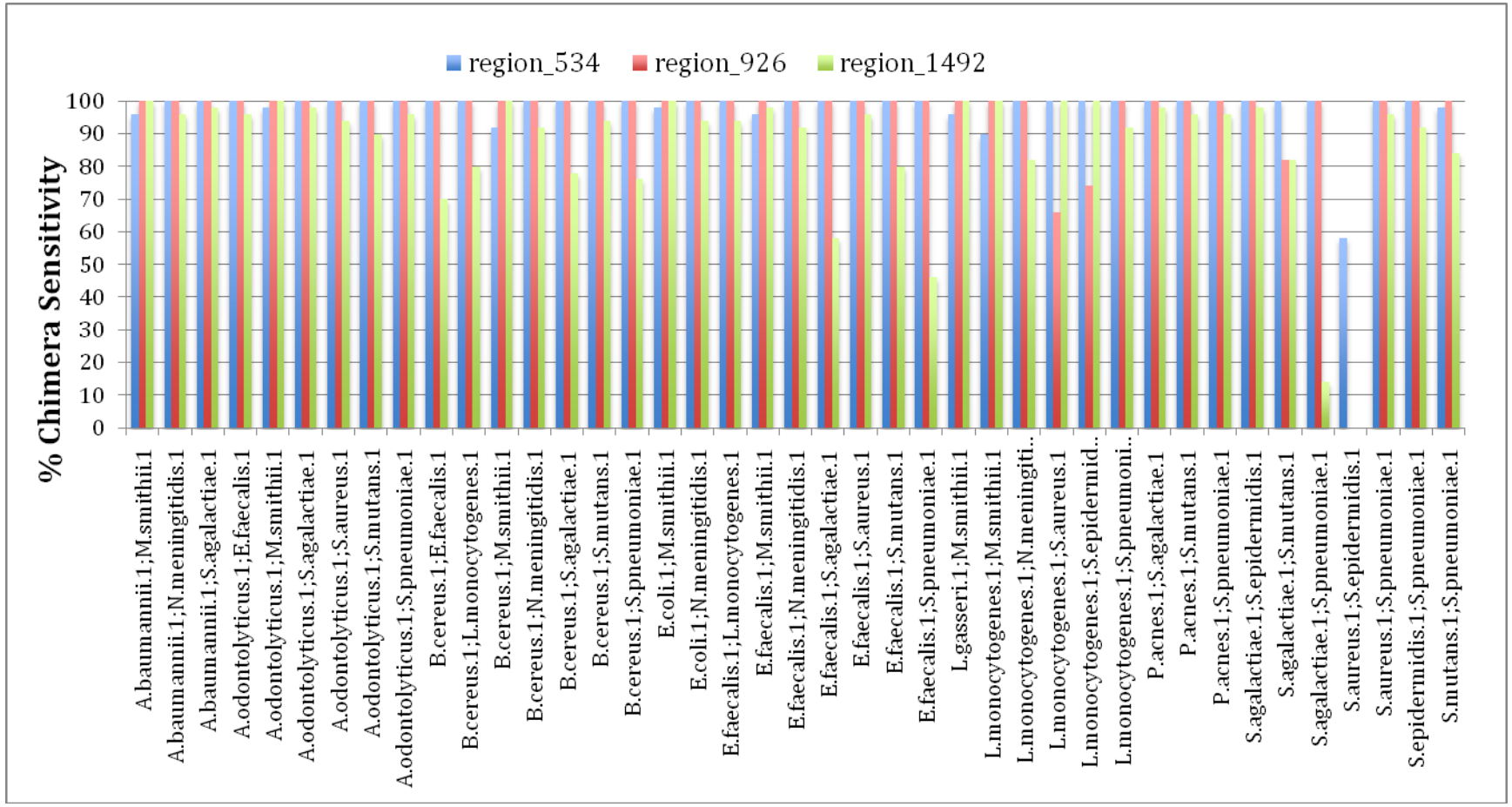
a.



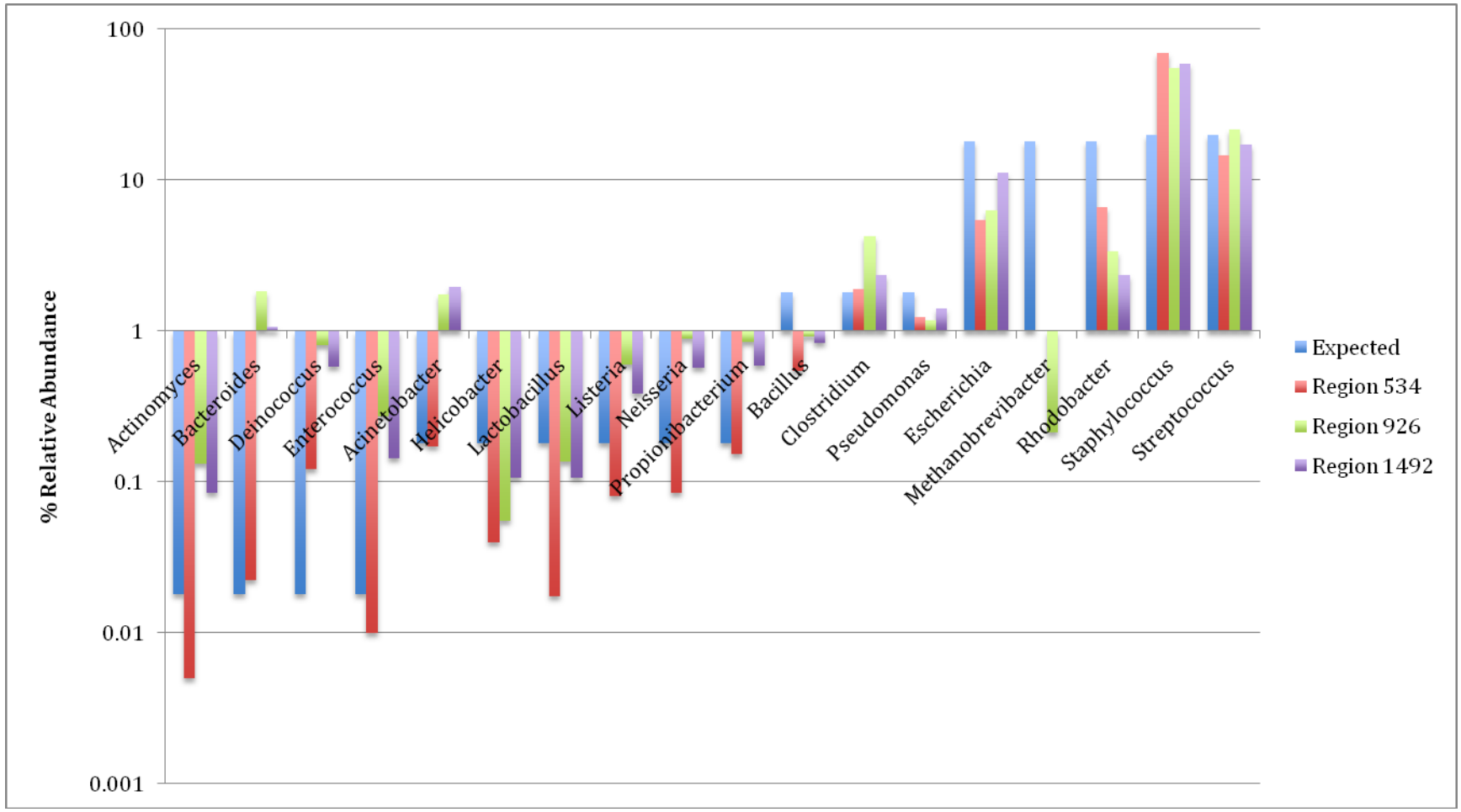
b.



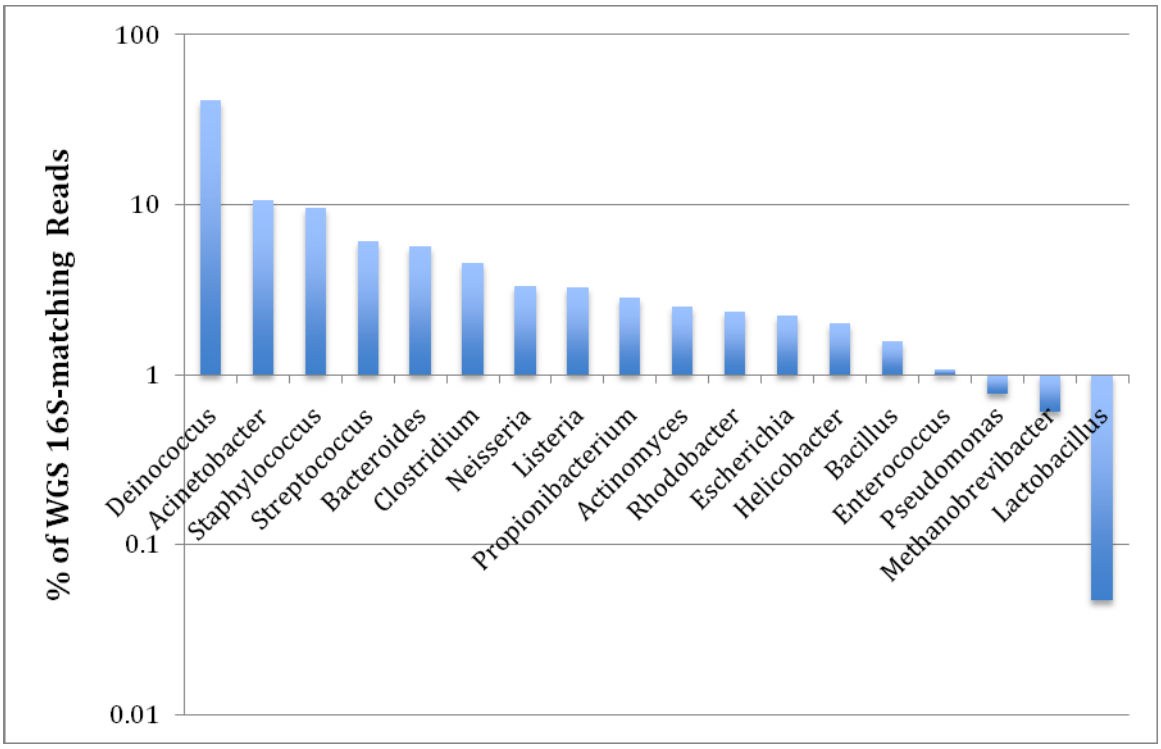
Supplementary Figure S11: 16S rRNA Chimera content of a 454-sequenced uniform mock community. (a) Percent chimera content according to sequenced region. (b) Relative abundance of chimeras according to genus-level pairs, showing only those pairs with at least 2% chimeras in any one replicate. (*) These organism pairs with region-specific chimeras are likely under-represented due to decreased sensitivity for detection.



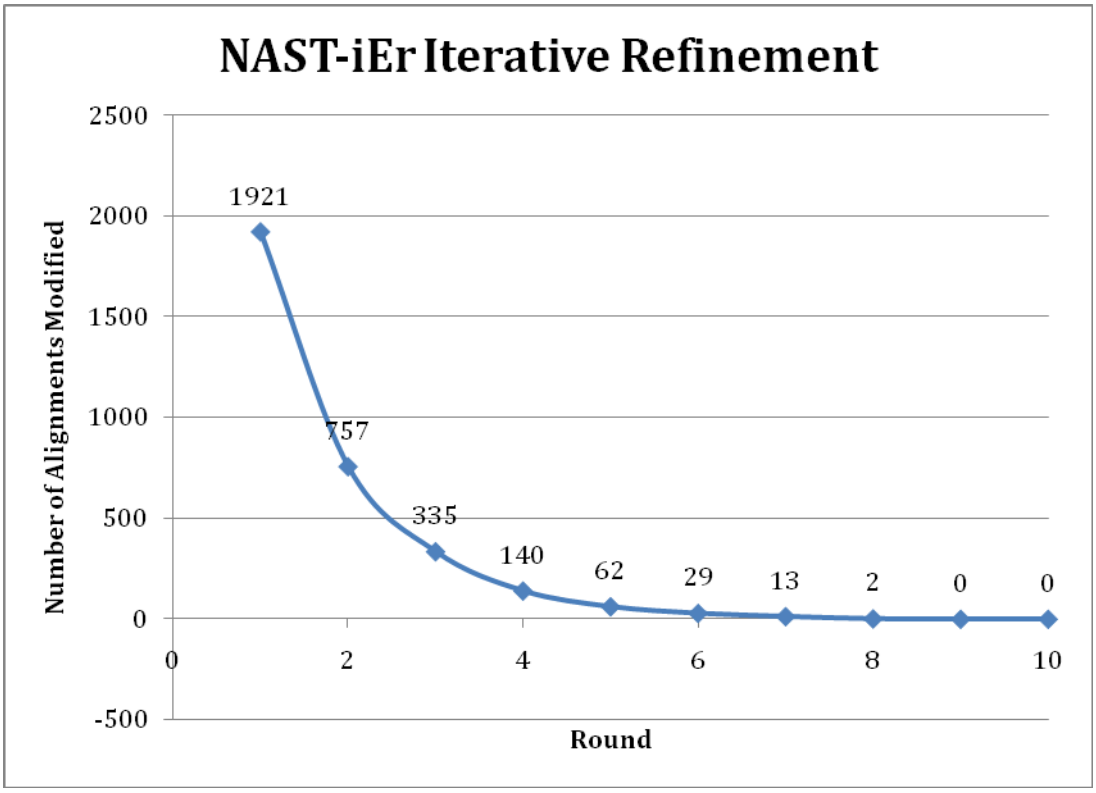
Supplementary Figure S12: Organism pairs exhibiting less than 100% chimera detection sensitivity with CMCS and simulated chimeras for at least one V-region. CS exhibits high sensitivity for chimera detection with all species pairs and regions except between *S. aureus* and *S. epidermis*. Region_534 = V1-V3, region_926 = V3-V5, and region_1492 = V6-V9.



Supplementary Figure S13: Organism abundance inferred from PCR of targeted regions of 16S within the staggered community and assayed by 454 pyrosequencing. Expected values for the staggered target concentrations are shown adjacent to observed values. *Methanobrevibacter smithii* is detected only with region 926 primers. Average values of four replicates are shown. Region_534 = V1-V3, region_926 = V3-V5, and region_1492 = V6-V9.



Supplementary Figure S14: Relative abundance of organisms' 16S sequences in the WGS 454 sequencing of the even mock community. Sequences matching known reference 16S sequences were identified by BLASTN (E<=1e-10).



Supplementary Figure S15: Iterative rounds of NAST alignments lead to more consistent alignments. After the eighth round, no further changes within NAST alignments were observed. Note, however, that although the alignments are more consistent, misalignments between homologous positions continue to persist and are not fully rectified by the current reiteration method.