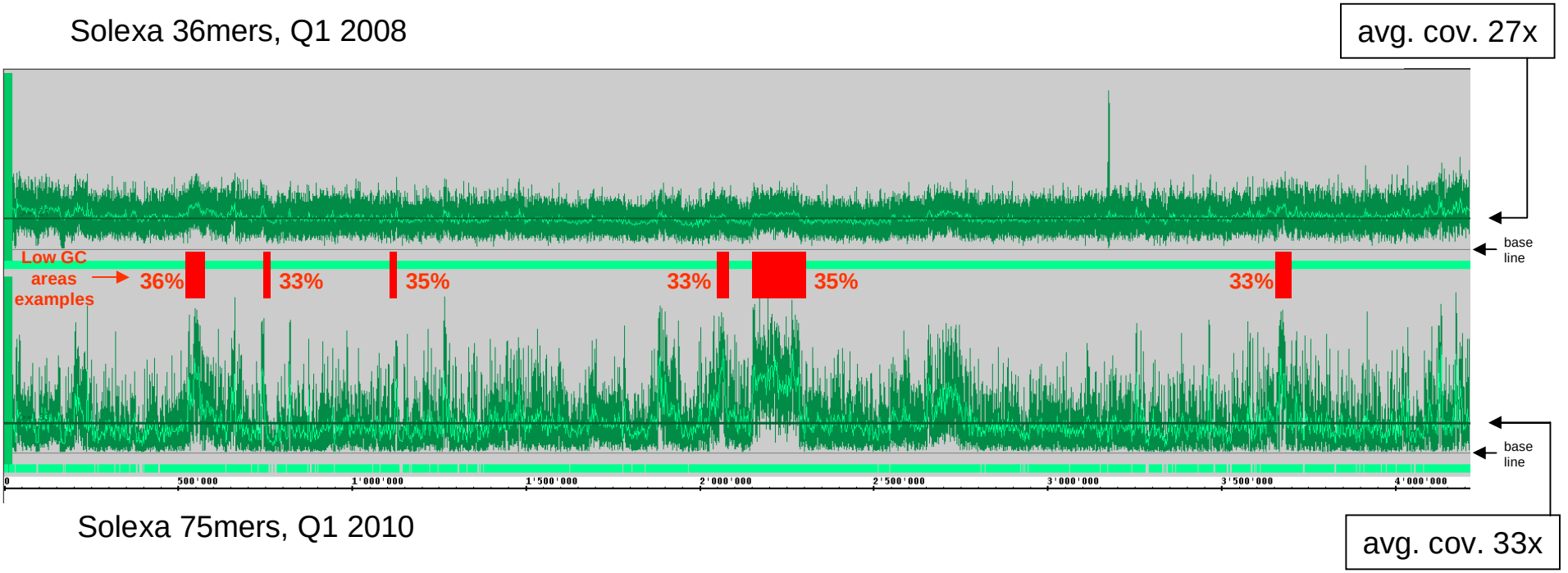
Comparison Solexa sequencing 2008 / 2010
Bacterium, 45%GC average, mapped on perfect reference sequence

Analysis of coverage uniformity



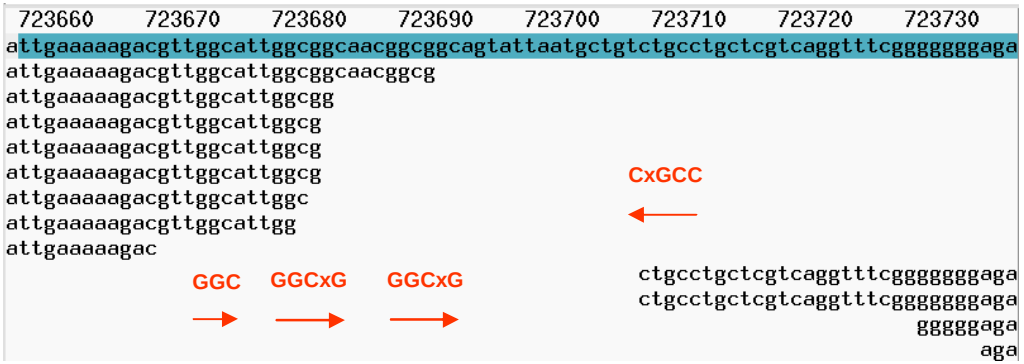
- 2008 coverage uniformity much better than 2010
- bias: higher coverage in low GC regions
  - bias barely noticeable in 2008 data
  - in 2010, bias strong enough in low GC regions (red bars in image) to look like genome duplication!

Comparison Solexa sequencing 2008 / 2010
Bacterium, 45%GC average, mapped on perfect reference sequence

Analysis of holes in coverage

- 2008 data (27x avg. cov., 36mers):
  - zero unexplained holes in the genome, everything covered
  - perfect

- 2010 data (33x avg. cov., 75mers)
  - 237 unexplained holes in the genome which are not repetitive areas like rRNA etc.
  - many more areas only with coverage of 1 to 5
  - often at sites with
    GGC
    or stronger even with
    GGCxG
    in 5' to 3' direction
  - worst with multiple local sites, or even when in both directions



→ For 75mers in 2010, avg. coverage of >=70x to 80x needs to be present get complete coverage (i.e., no holes)