EAGLEVIEW: A GENOME ASSEMBLY VIEWER FOR NEW-GENERATION SEQUENCING TECHNOLOGIES

Weichun Huang and Gabor Marth
Department of Biology, Boston College

SUPPLEMENTARY MATERIALS

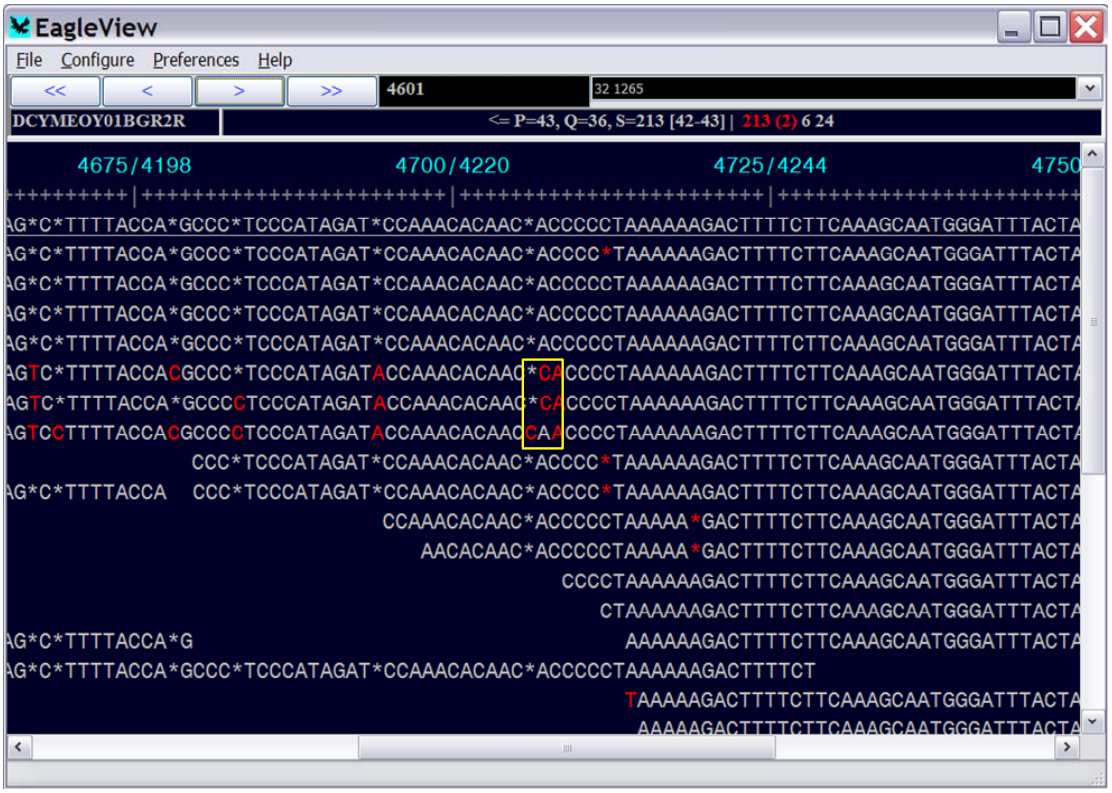

Figure 1: Detailed assembly view. The figure shows a pinpoint view using EagleView zooming capability for exanimation of individual bases in assembly. The yellow box indicates an area of local misalignment of 454 reads because of base overcall errors in three reads where the three "C"s are results of over-call errors and two "AA" in the 3rd line should be one "A".

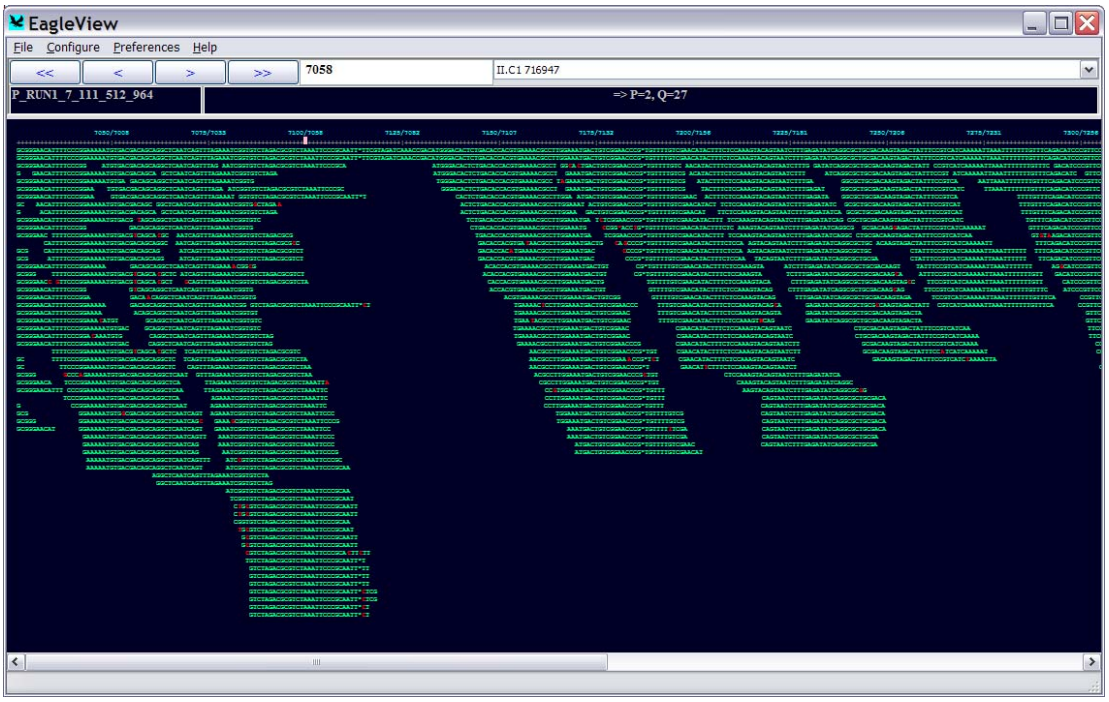

Figure 2: Bird-eye assembly view. A zoomed-out view of a section of an assembly of 32-base *C. elegans* Illumina reads is shown. A small region without single read coverage is shown in the middle of the window.

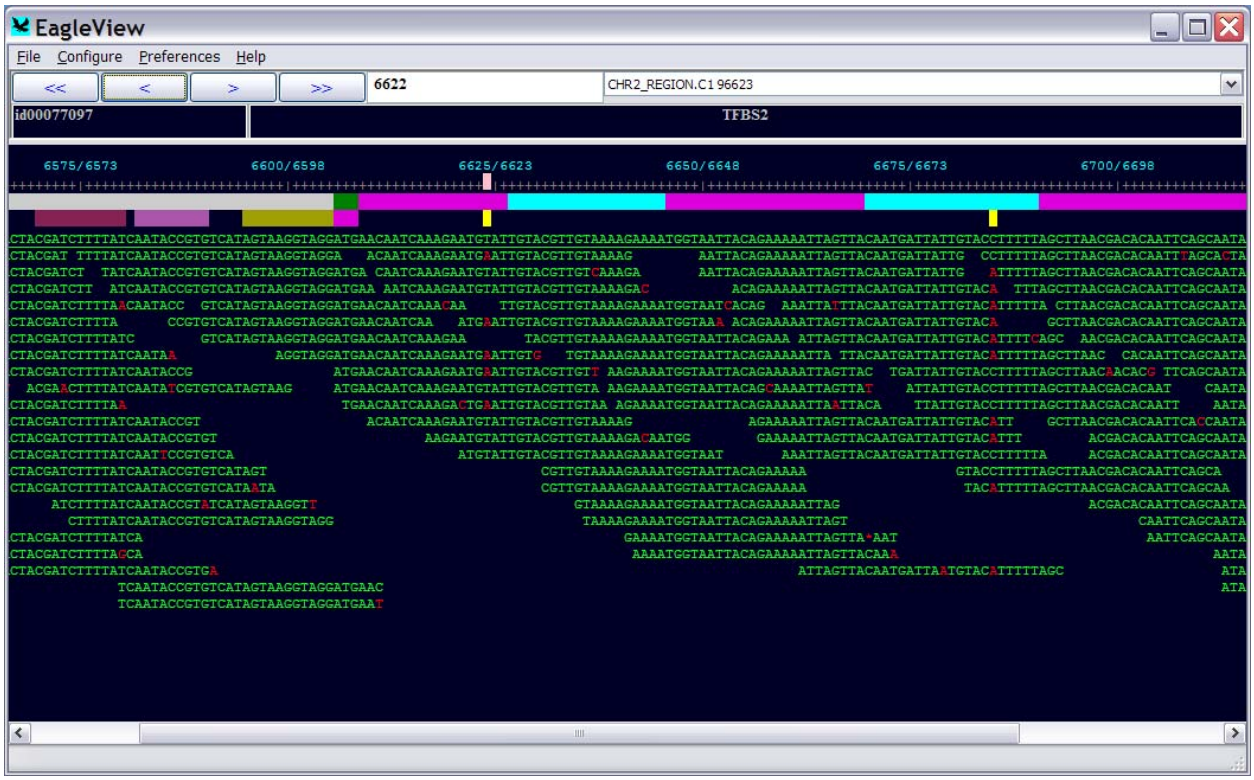

Figure 3: Genome feature visualization. The figure shows non-overlapping view of genomic features including promoter (grey), exon (pink), intron (cyan), translation start site (green) and SNP candidate (yellow) and three regulatory elements in different colors. For each feature, its name if available will show in read-information window on mouse-over.

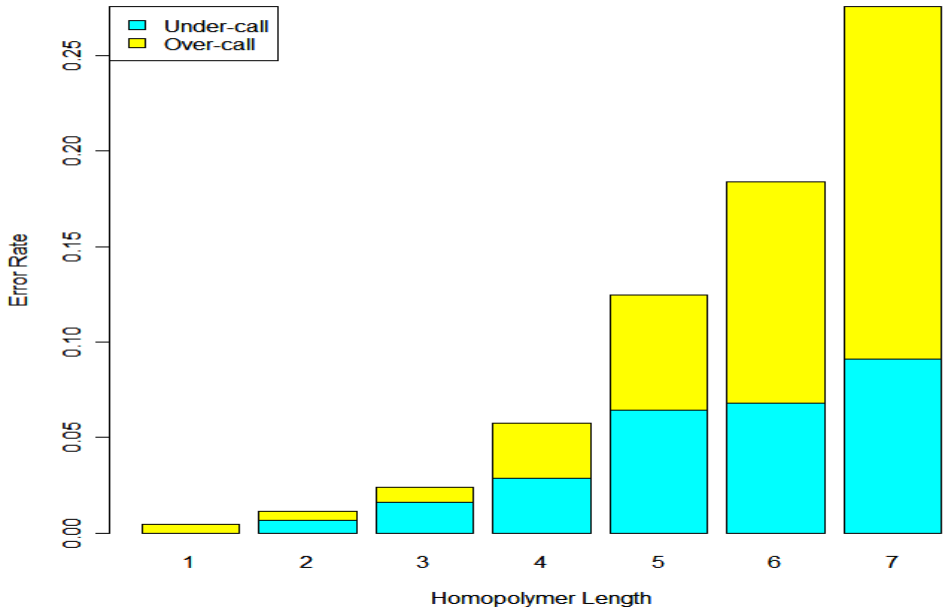

Figure 4: The sequencing error profile of 454 GS20 Pyrosequencer. The figure shows that the average error rate is markedly increased as the length of homo-polymer run increases, and the over-call error rate increases more rapidly than the under-call error rate.

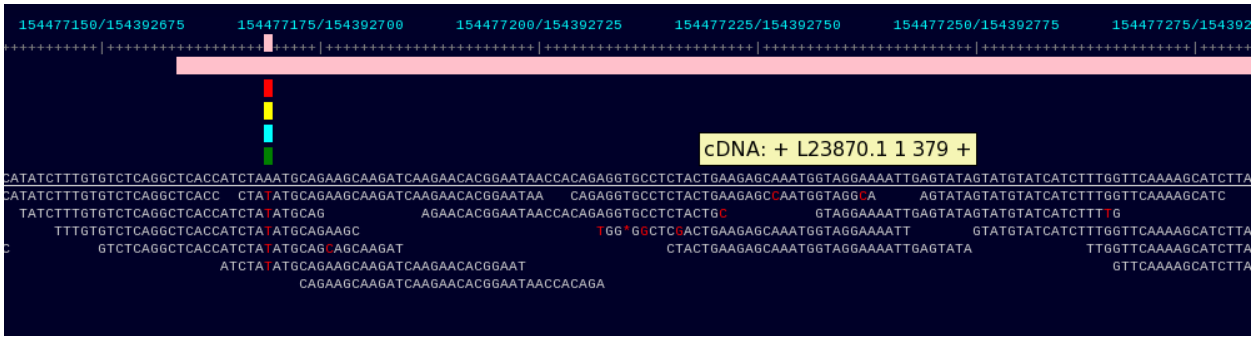

Figure 5: Integrated visualization of a human genome assembly with the HapMap SNP data and NCBI genome feature annotation. The figure show a common SNP site with known genotype A/T (SNP ID: rs989711) among four human subpopulations CEU (red), CHB(yellow), JPT(cyan), and YRI (green). This SNP is in at beginning region of mRNA L23870, and is confirmed by the new assembly as shown in the figure.