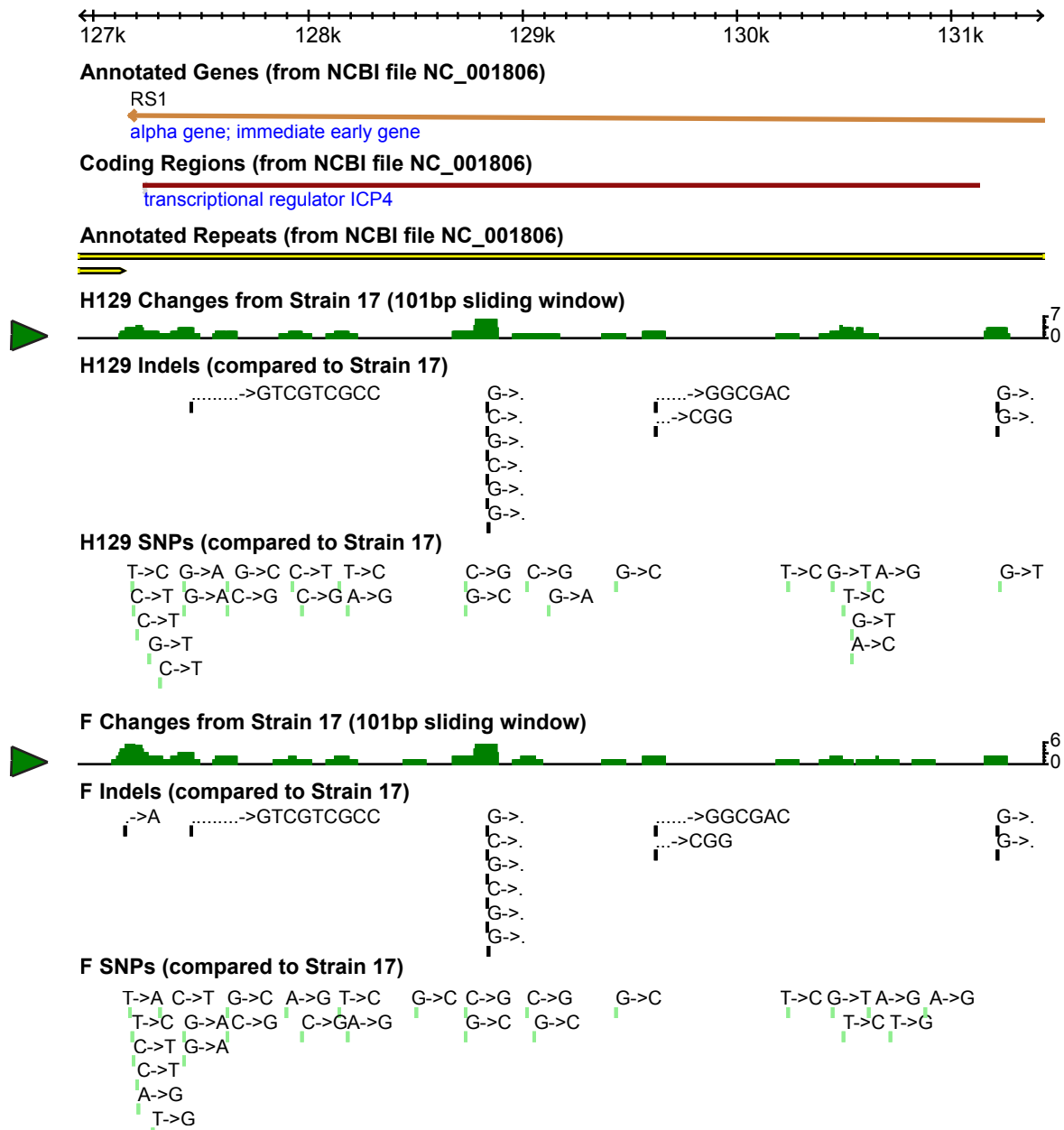


**Supp. Figure 1. Genome length and difference summary for HSV-1 strains 17, F, and H129.**

Note that the genome length for strains F and H129 may under-estimate changes in reiteration length. There are three HSV-1 genes located in the terminal repeats and thus present in two copies in each genome; in determining the number of amino acid (AA)-differences between strains, we only counted one copy for each gene.

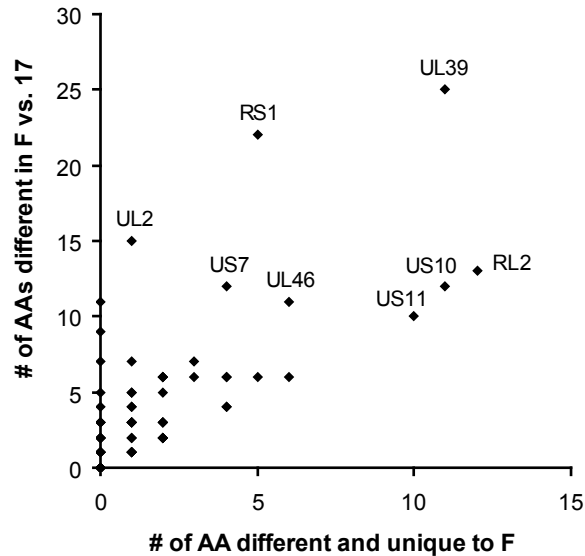
### Genome browser view of Immediate early gene ICP4 (RS1) (<http://viro-genome.princeton.edu>)



**Supp. Figure 2. Genome browser view of differences in ICP4 (RS1), an immediate-early transcriptional regulator.** This genome browser view displays the location of ICP4, one of the five immediate-early genes in HSV-1 and the one with the greatest amount of sequence variation among strains 17, F, and H129. Histogram tracks (green arrowheads) depict the location of differences between each new strain and the reference genome. The histograms summarize both insertion-deletion (indel) differences and single-nucleotide polymorphism (SNP) differences, each of which are plotted beneath the histogram at their respective genomic positions. Similar plots can be called up for any gene in the HSV-1 genome, at <http://viro-genome.princeton.edu>.

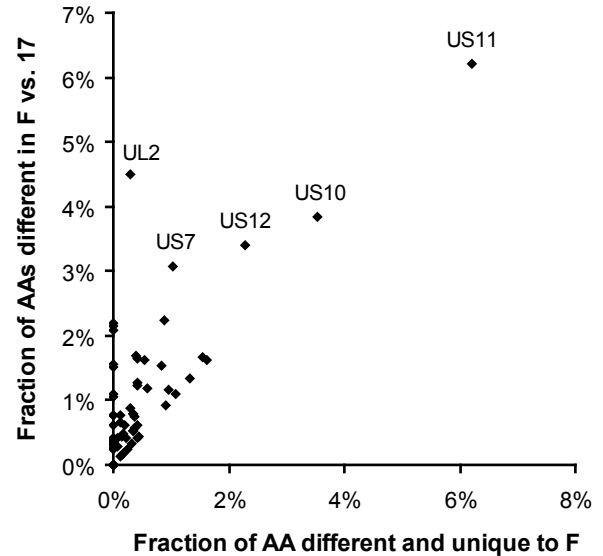
A.

Large number of amino acid changes in strain F



B.

Number of amino acid changes in F, normalized for gene length



**Supp. Figure 3. Genes with a large number or proportion of amino acid coding changes in HSV-1 strain F.** (A) Scatter plot of the strain F proteins with the largest overall number of amino acid differences from the wild-type reference strain 17 (y-axis), vs. the subset of these changes that are also different from H129 (x-axis) and are thus unique to strain F. The most extreme points are labeled to identify the proteins. (B) Since gene length affects the total possible number of observed mutations, the same data were normalized for gene length. The scatter plot presents strain F proteins with the largest overall percentage of amino acids differing from the reference strain 17 (y-axis), vs. the subset of these changes that are also different from strain H129 (x-axis).

**Supplementary Table 1. Sequence read statistics for Strains F and H129.**

Strain	Total combined # reads	Total # reads (per run)	% MAQ aligns to HSV-1*	% MAQ aligns to human**
HSV1 H129	17,741,445	9,790,374 (36 bp) <sup>1</sup>	70.5 <sup>1</sup>	15.4 <sup>1</sup>
		7,951,071 (75 bp) <sup>2</sup>	64.4 <sup>2</sup>	2.7 <sup>2</sup>
HSV1 F	14,127,481	6,715,897 (75 bp) <sup>1</sup>	86.0 <sup>1</sup>	0.4 <sup>2</sup>
		7,411,584 (75 bp) <sup>2</sup>	87.4 <sup>1</sup>	0.4 <sup>2</sup>

<sup>1</sup> First sequencing run for each strain. <sup>2</sup> Second sequencing run for each strain.

\* using strain 17+ (NC\_001806) as the reference.

\*\* human sequence is proxy for monkey Vero cell line used to grow virus stock.