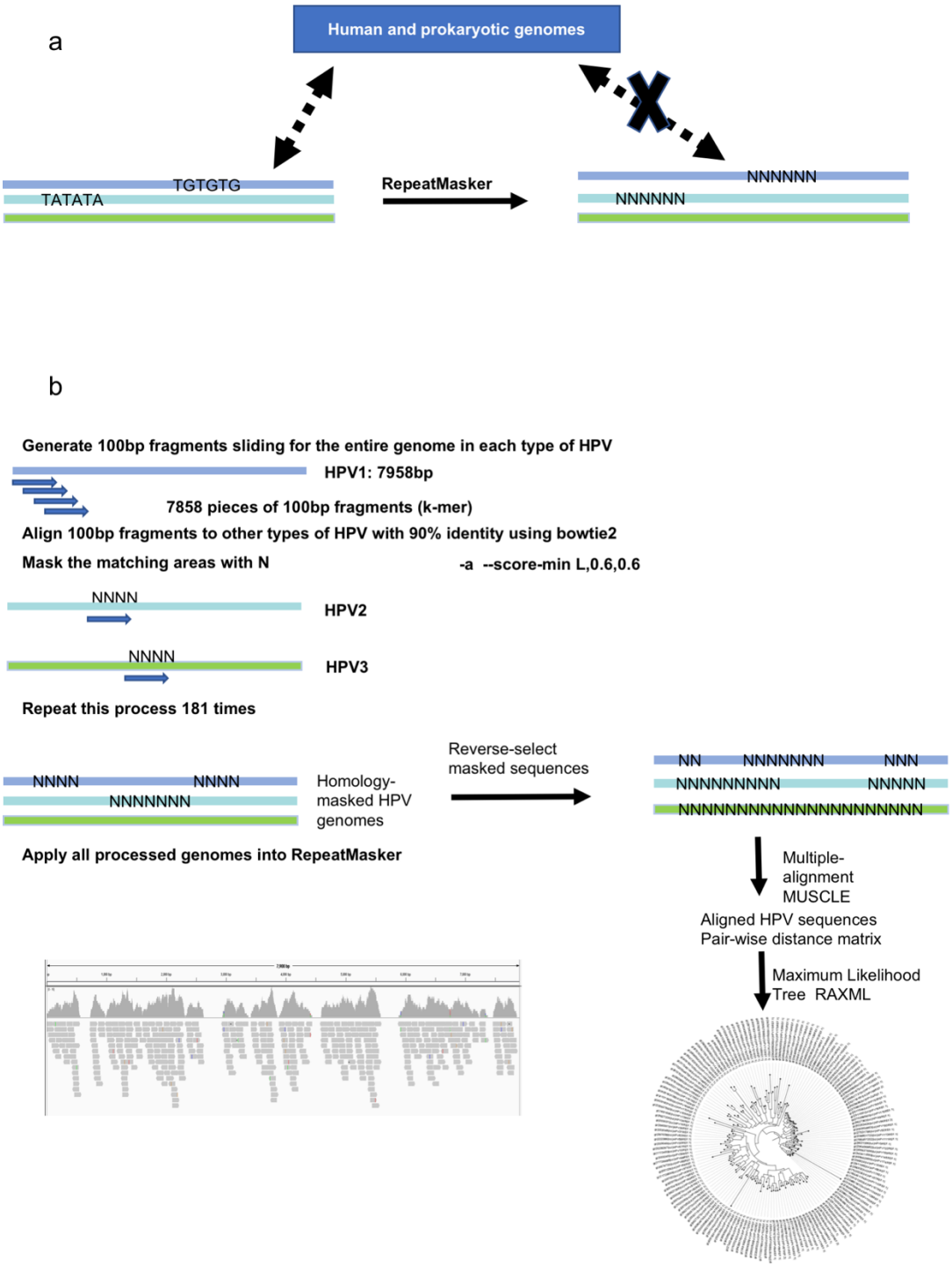
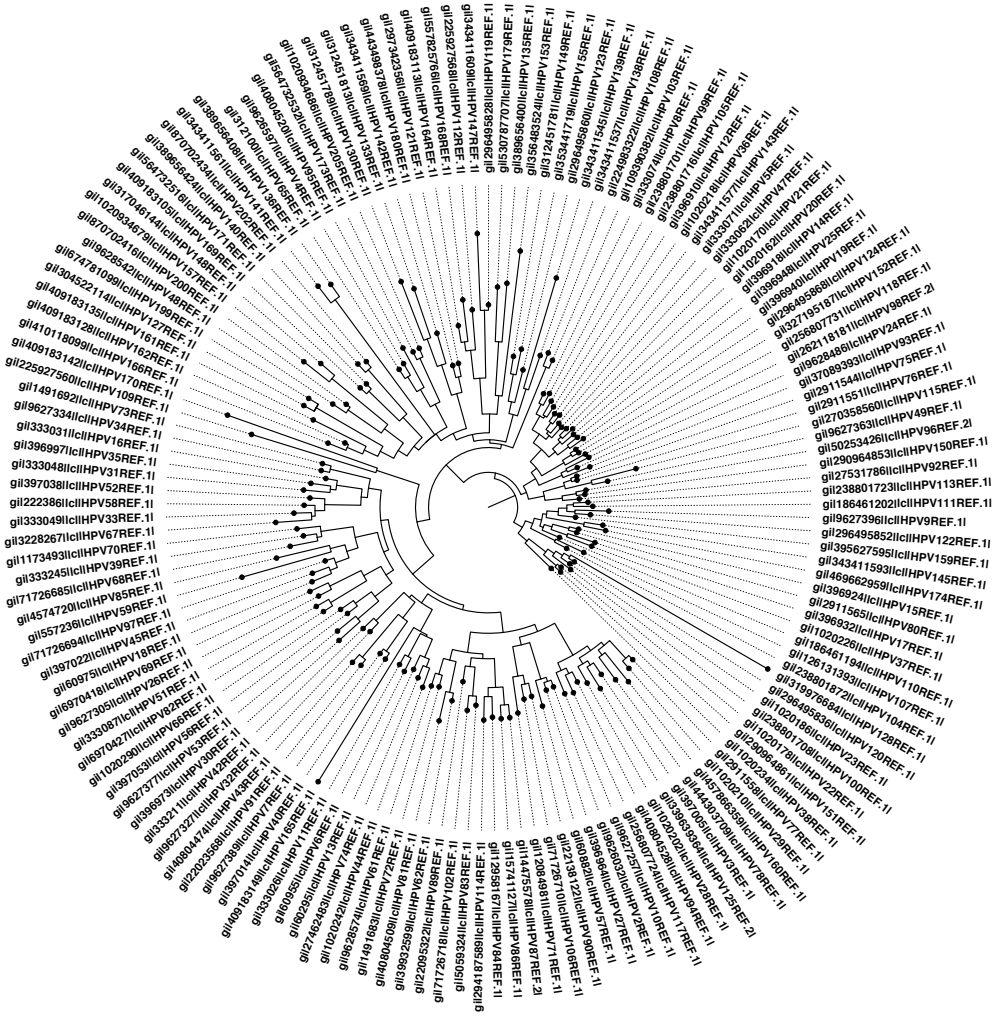


Supplementary Figure 1. Construction of repeat-mask (a), homology-mask database and the homology distance matrix and tree (b).



Supplementary Figure 2. The Maximum Likelihood tree of HPV only using their homology sequences.

c

Homo sapiens CREB binding protein (CREBBP), RefSeqGene on chromosome 16

Sequence ID: NG_009873.1 **Length:** 162067 **Number of Matches:** 38

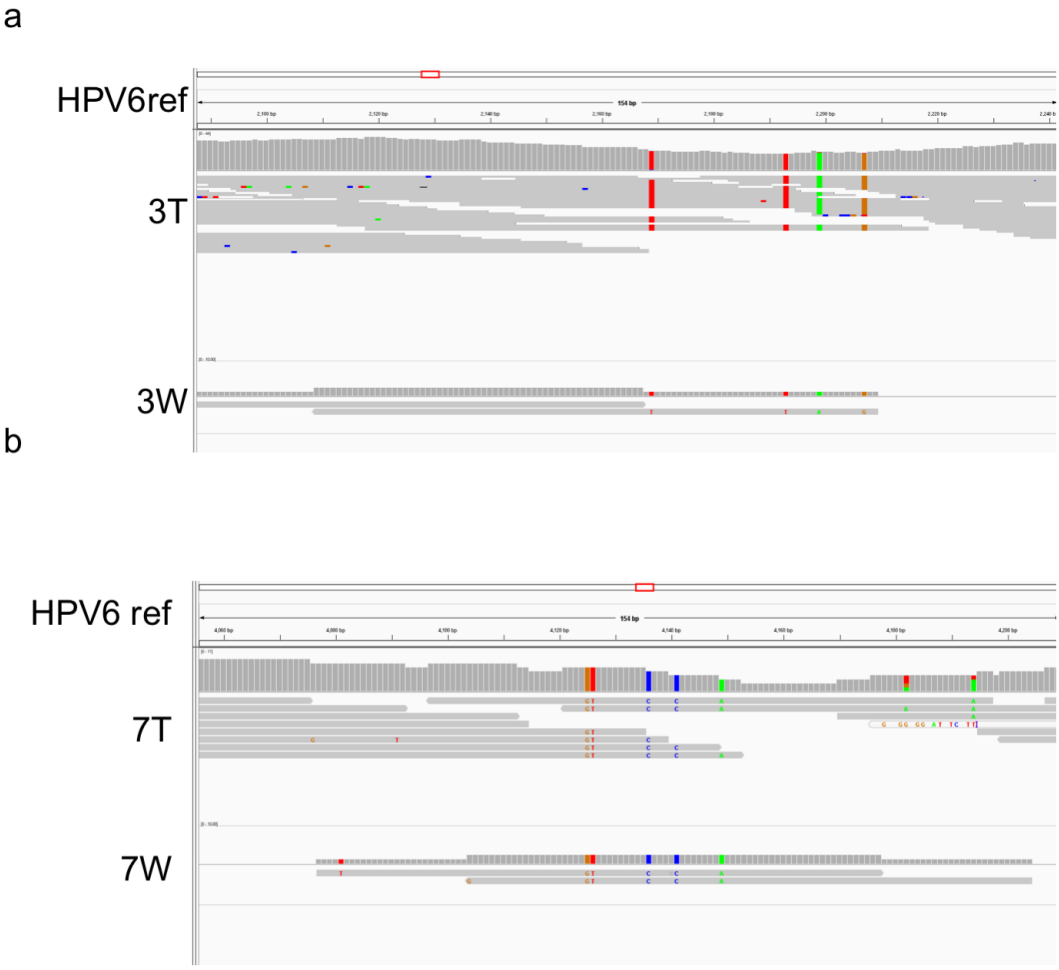
Range 1: 35627 to 35725 GenBank Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
150 bits(81)	2e-34	93/99(94%)	0/99(0%)	Plus/Minus

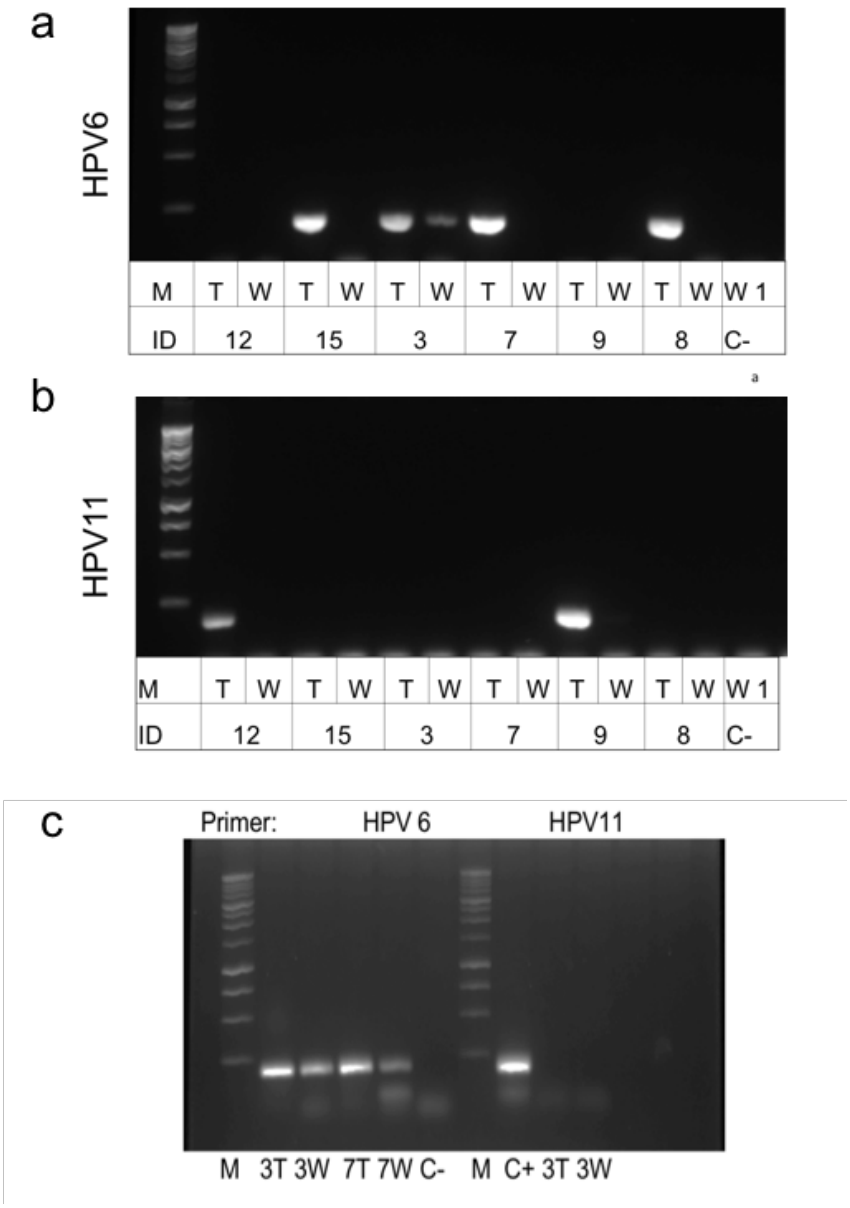
```
Query  1      TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTATC  60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  35725  TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG  35666

Query  61     TGTGTGTGTGCGTGTGTATGTGTGTGTGTGAGTGAGGTG  99
              |||||||||| ||||||| |||||||||||| |||| |||
Sbjct  35665  TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGATGTG  35627
```

Supplementary Figure 3.
Visualization of HPV71 alignments from 9 samples in Integrative Genomics Viewer (IGV) (a) and MEGA (b). It was found that 3 laryngeal tissues and 6 oral wash samples had short reads aligning to HPV71 and they were all aligned to the same simple repeats (TG repeats) region of HPV71 genome. The results of BLASTn searching the read of 3T against human genome (c).



Supplementary Figure 4.
Visualization of HPV6 alignment from sample 3T,3W (a) and sample 7T,7W (b) in IGV. Even oral wash samples only had two reads (one pair) aligned to HPV6, they shared the SNP pattern and it indicated that they present in the oral cavity were from tumor tissues.



Supplementary Figure 5. PCR validation of HPV6 and HPV11 in 12 shotgun sequencing samples. **(a,b)** Validation of presence of HPV6/11 in Recurrent Respiratory Papillomatosis. Samples from laryngeal tissues all had HPV6 or HPV11. We used 5ng DNA/reaction as template and PCR program: 95ºC 15', 62ºC 15' and 72ºC 30' for 45 cycles for HPV 6 and 11. (**c**) HPV PCR validation for low abundant HPV in sample 3W, 7W detected by HPViewer. Here we decreased PCR annealing temperature from 62 to 58ºC and other conditions were the same. Sample 3T,3W and 7T,7W were all HPV6 positive and both 3T and 3W were HPV11 negative. Though only two reads were detected in sample 3W and 7W for HPV6, they were true positive. M: marker; T: tumor tissue; W: oral wash; C-: negative control; C+: positive control.