Assessing Recombination in HPV, Part II

Andrew Farmer and Gerald Myers

MS K710, Los Alamos National Laboratory, Los Alamos, NM 87545

In this section, a second computational method for detecting viral recombinant molecules, specifically HPV hybrid proteins, is described. The program known as RIP is described in the previous section of Part III. No evidence of recombination among known HPV types is found by either method, however the analysis can be of interest in its own right for defining type-specific protein patterns.

Signature patterns for "super-group A" PVs (as defined in Part I) are generated herein from aligned E1, E2, L1 and L2 amino acid sequences using the VESPA program (Korber and Myers, *AIDS Res Hum Retroviruses* **8**:1549–1560, 1992). VESPA (viral epidemiology signature pattern analysis) is a program designed to discover characters – nucleotides or amino acids – that differentiate one sequence from all other sequences except those that are closely related. It was originally written to assist in the analysis of HIV sequences that were strongly shared by a Florida dentist and six of his patients (Ou et al., *Science* **256**:1165–1171,1992; Korber and Myers, *AIDS Res Hum Retroviruses* **8**: 1549–1560, 1992). In that investigation, we were struck by the uncanny sharing of atypical amino acid residues among the dentist and the patients he appeared to have infected. Thus a signature pattern was defined to be the set of amino acids (or nucleotides), which will usually be non-contiguous, that are sufficiently rare in a background set of sequences to differentiate a query sequence from that background. In short, it is a kind of "fingerprint."The VESPA program objectively determines signatures and provides the frequencies for the rarity of characters, therefore statistical analysis can be conducted. It runs on PC-DOS, Macintosh and SUN-Unix platforms, and is available at no cost (contact Kersti MacInnes, kam@t10.lanl.gov or Chuck Calef, cxc@t10.lanl.gov).

Starting with completely aligned E1, E2, L1 and L2 PV amino acid sequences from super-group A, approximately 2200 residues treated as a continuous stretch, some initial pruning was conducted. Because many variant sequences of HPV-16 are available (Part I), positions at which intratype variation is observed are stripped out. The resulting HPV-16 signature, then, should be virtually type-specific. As variant sequences become known for other PVs, this same procedure will apply to the determination of type-specific signatures.

The VESPA program took the pruned alignment (in so-called "table" format) and determined signatures according to specified criteria. In the analyses that follow, a signature position was required to have a 75% or better consensus in the background set; a total of 38 PVs constituted the background, thus the signature for each was determined against the 37 others. Furthermore, to qualify as a signature site, no more than two sequences sharing the same atypical (nonconsensus) residue were allowed. The first criterion tends to eliminate highly variable sites, which produce softer signatures, while the second eliminates clutter of the analysis due to mere phylogenetic effects, as will become clear below. These parameters can be adjusted so as to optimize the selection of differentiating characters: with a large data set, they can be fairly stringent; with smaller data sets, the stringency may have to be relaxed. Nucleotides can be used instead of amino acids, with an increase in the number of characters but also an increase in the number of homoplasies (chance occurrence of shared characters among taxa of different lineages). Gaps can be signature characters by either approach.

Note what has been eliminated by the program in order to improve the signal to noise ratio: On the one hand, all sites that are invariant are not part of the analysis. On the other hand, all highly variable sites for which a sufficient consensus does not exist are also eliminated. Finally, sites for which strong phylogenetic relationships dictate the sharing of a rare residue between more than two sequences are eliminated. Hence, low information content has been relinquished in order to enhance the signal-to-noise ratio. In the matrix table below, a numerical summary of the signatures determined for the 38 PVs (diagonal running from upper left to lower right) and the occurrence of shared signature residues are shown. The signature patterns for HPV-16, 18, 6, and 11 comprise 55, 33, 14, 15, characters respectively, for example; the actual signatures for these four PVs are shown at the end of this section and the remaining 34 signatures are available from the HPV Web server (`http://hpv-web.lanl.gov`). The shorter signatures result from sequences having several other closely related sequences (6b, 11, etc.) in the survey.

	54	32	42	3	28	10	29	61	2a	27	57	26	51	30	53	56	66	18	45	39	70	59	7	40	16	35h	31	52	33	58	Rh	6b	11	44	55	13	PC	34
54	69	1	2	2	1	1	0	3	1	0	2	2	1	2	0	0	0	0	0	0	1	0	0	1	1	1	1	0	0	1	1	1	2	1	2	1	0	3
32	1	74	37	0	2	0	1	1	0	1	0	0	0	0	0	0	1	0	3	0	1	0	1	3	2	2	2	3	0	0	0	1	0	0	0	0	1	4
42	2	37	62	1	0	0	2	1	0	1	0	0	1	0	0	0	0	0	4	0	2	0	0	0	0	2	2	1	0	0	1	0	0	0	0	0	2	3
3	2	0	1	36	15	9	9	1	0	1	0	0	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1
28	1	2	0	15	33	11	4	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
10	1	0	0	9	11	38	6	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	1	0	0	0	0	1	0	0	0	0	0	0
29	0	1	2	9	4	6	34	2	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1
61	3	1	1	1	0	1	2	63	1	0	1	0	2	1	0	1	1	0	2	1	2	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1
2a	1	0	0	0	0	0	0	1	40	30	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1
27	0	1	1	1	0	0	0	0	30	39	28	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
57	2	0	0	0	0	0	0	1	28	28	42	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	2	0	1	0	0	0	0	0
26	2	0	0	0	1	0	0	0	0	0	0	41	16	0	0	0	0	1	2	1	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	2
51	1	0	1	0	1	0	0	2	0	0	0	16	53	1	1	0	0	1	0	1	2	1	1	1	0	0	0	1	0	0	0	1	0	0	0	1	0	3
30	2	0	0	0	0	0	0	1	0	0	0	0	1	24	6	3	2	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0	1
53	0	0	0	1	0	0	0	0	0	0	0	0	1	6	23	5	5	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0
56	0	0	0	0	0	0	0	1	0	0	0	0	0	3	5	27	17	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
66	0	1	0	0	0	0	0	1	0	0	0	0	0	2	5	17	28	2	1	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	3
18	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	1	2	33	14	1	1	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
45	0	3	4	1	0	0	1	2	0	0	0	2	0	0	0	0	1	14	29	2	2	3	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
39	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	1	2	38	13	3	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1	1
70	1	1	2	0	0	0	0	2	0	0	0	1	2	0	0	0	0	1	2	13	37	2	1	1	0	0	0	0	0	1	0	0	0	0	1	0	0	1
59	0	0	0	1	0	0	0	0	0	1	0	1	1	0	0	0	0	2	3	3	2	30	0	0	0	1	1	0	1	2	1	1	1	0	0	0	1	1
7	0	1	0	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	1	0	66	53	2	1	0	0	1	0	1	0	0	0	0	0	0	1
40	1	3	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	1	0	53	70	4	1	1	0	1	0	1	0	0	0	0	0	0	0
16	1	2	0	0	1	2	1	0	0	0	0	0	0	1	1	0	2	0	0	0	0	0	2	4	55	7	8	1	1	1	1	0	1	0	0	0	0	1
35h	1	2	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	1	1	0	1	1	1	7	43	8	1	0	1	1	0	0	0	0	0	0	3
31	1	2	2	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	8	8	35	1	1	1	0	0	0	0	0	0	0	0
52	0	3	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	54	13	11	0	0	0	0	0	0	1	1
33	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	1	1	1	1	0	1	13	43	23	0	0	1	0	0	0	0	2
58	1	0	0	0	0	0	0	0	1	0	1	1	0	1	0	0	0	0	0	0	1	2	0	0	1	1	1	11	23	45	2	0	0	0	0	0	0	0
Rh	1	0	1	0	0	0	1	0	0	0	2	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	2	48	0	0	0	0	0	1	2
6b	1	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	14	5	1	1	0	0	0
11	2	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0	5	15	0	0	3	3	0
44	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	12	10	2	0	0
55	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	10	11	2	0	0
13	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	2	2	16	6	0
PC	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	1	0	3	0	0	6	18	0
34	3	4	3	1	0	0	1	1	1	0	0	2	3	1	0	2	3	0	1	1	1	1	1	0	1	3	0	1	2	0	2	0	0	0	0	0	0	74

For one of the PV types to be a recombinant of two or more of the other 37 types, a significantly large number of characters must be shared with two or more other signatures. What is a significantly large number? As a rule of thumb, any two proteins can be about 20% similar by chance (Doolittle, *Of Urfs and Orfs*, University Science Books, Mill Valley, CA., 1987), therefore 20% of the “unknown ” type signature might be a meaningful estimate. In fact, most signatures in the Table are related to all other signatures by less than 10% (weak convergence), with exception of i) phylogenetically related sequences and ii) potential recombinants. The stringency of the VESPA criteria discussed above helps resolve any uncertainty, as can be seen in the example of HPV-57 relative to HPV-27 and HPV-2a. HPV-57 has a signature of 42 residues, of which 28 are shared with HPV-27 and 28 are shared with HPV-2a. If the same, or nearly the same, signatures are in common among the three PVs, 57, 27, and 2a, these sequences must reflect a shared evolutionary history. On the other hand, if the 28 characters shared by HPV-27 with HPV-57 are clustered and are different from the 28 characters shared between HPV-2a and HPV-57, then recombination seems likely. In the table below, it is evident that these are evenly shared signature characters among the three PVs, indicating phylogenetic (i.e., cladistic) connectedness and not recombination. The same conclusion could have been reached through tree analysis over the separate genes; however, the simplicity of the VESPA analysis and its reduction of homoplasy make this approach superior to tree analysis for this purpose.

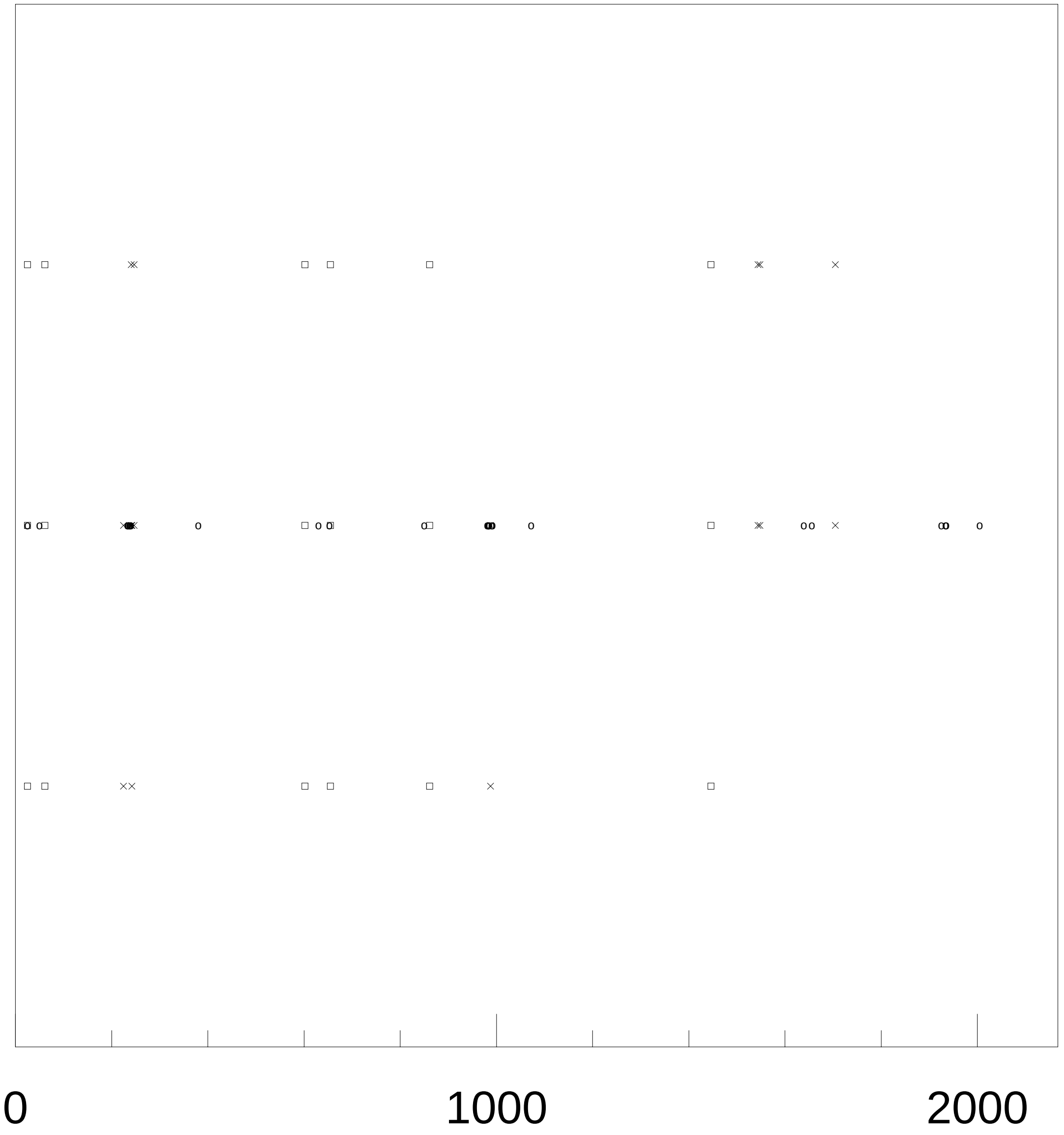
HPV2a **Q**R.**MVE**Y.**QSF**S.**S**H**APCTASRD**.**N**SA**QA**SA**V**S**PR**..**D**..S.**VHL**.**F**.L**MR**.
HPV57 **Q**SR**MVE**.R**QSF**..**S**.**APCTTSRD**F**N**..**QA**AT**V**H**PR**..**D**T**V**AS**VHL**Y**F**..**MRI**
HPV27 **Q**R.**MVE**....**F**.F**S**.**APCTTSRD**.**N**SA**QA**TA.N**PR**LK**D**.**V**S.**VHL**.**F**R.**MRI**

In the alignment above, signature pattern residues that are shared by HPV-57 and at least one of the other types (2a and 27) are shown in boldface. Note that although the signature pattern residues appear to be contiguous, the alignment has been stripped of all positions for which none of the sequences in the alignment had a signature pattern residue.

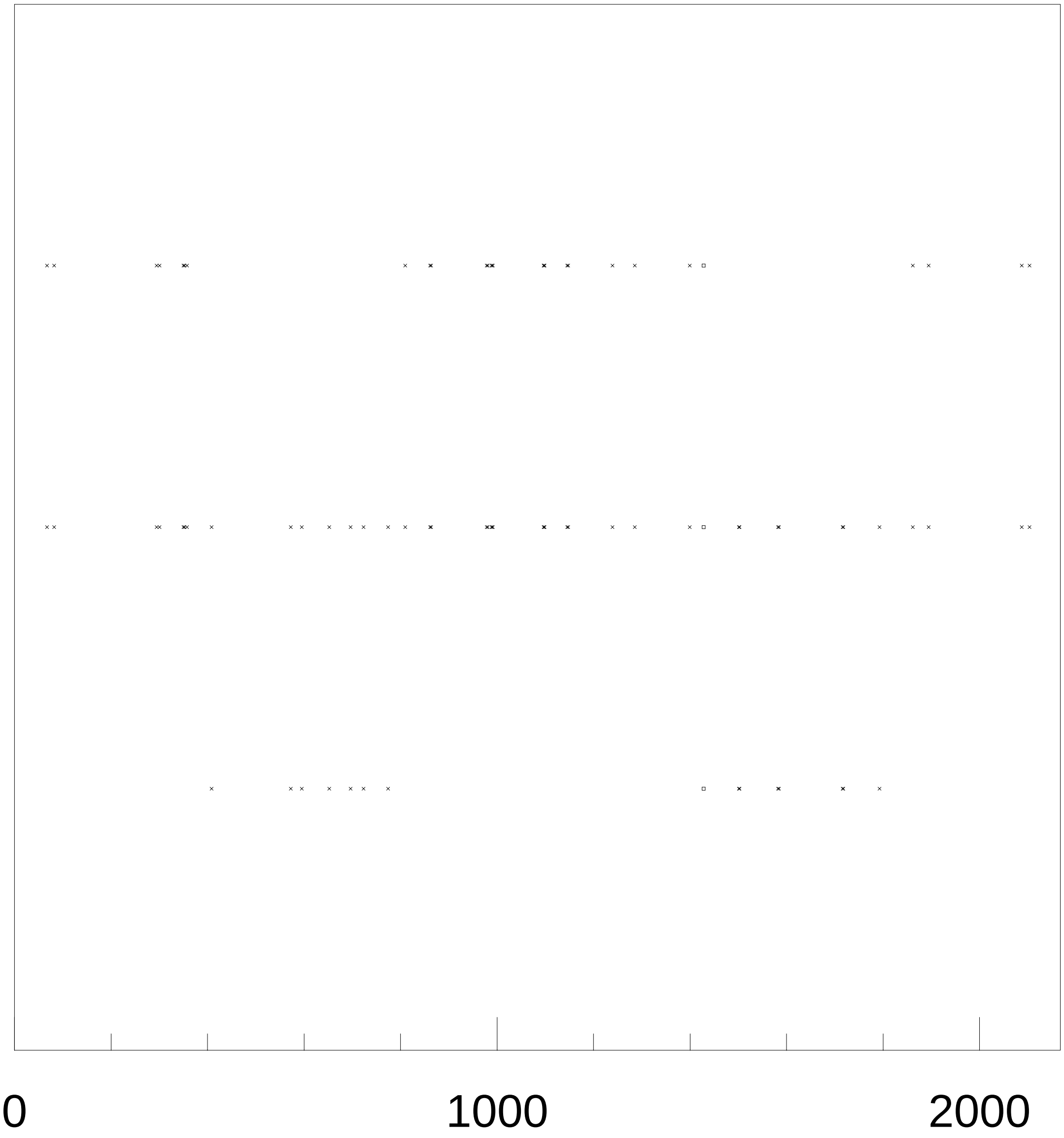
The issue becomes more complicated when, hypothetically, one HPV type is a recombinant of another type in the table and something unknown. In this instance, it will be the distribution of shared signature characters that will distinguish a hybrid sequence from a phylogenetically-related sequence: in the recombination case, the distribution would be uneven across the four proteins, E1, E2, L1 and L2, whereas in the phylogenetic, and more likely, case the distribution should be approximately uniform. We scanned the above Table for evidence of uneven distributions of shared signature characters and found none. A rigourous approach to this problem might include statistical analysis of the distributions and separate tree analyses over the four different coding regions.

Recombination in HPV

In the previous analysis in this section by Halpern, the RIP program was used to analyze the relatedness of HPV-10 to HPV-28 and HPV-3. Could HPV-10 be a cryptic hybrid? We analyzed the same three sequences using the VESPA approach, which greatly increases the signal-to-noise ratio, as can be seen by comparing the following figure with the RIP analysis. Here, VESPA clearly offers no ground for thinking HPV-10 (middle pattern) is a hybrid of 28 (top) and 3 (bottom). In the figure, squares indicate signature characters shared by all three sequences, Xs indicate characters shared between HPV-10 and either HPV-3 or HPV-28, and ovals represent characters unique to HPV-10. The resolution of the figure is such that symbols representing characters in close proximity may overlap one another.



If an HPV hybrid existed, how would the analysis present itself? To address this question, a chimeric sequence of HPV-16 and HPV-52, with four crossover points, was constructed. The following figure illustrates what the recombinant pattern (shown in the middle) would be for this construct.



The discussion up to now has focused upon VESPA as a rapid screening method for semi-quantitatively evaluating sequences. Under the assumption of independence of sites, which may be reasonable for signature characters from different coding regions (E1, E2, etc.), a more rigorous statistical analysis of shared characters and their distributions can be generated from VESPA. We will not take up those aspects of the analysis herein. They are best evaluated by working directly with the program. The applicability of VESPA to other questions, for example differential variability in coding regions, should also become apparent.

Recombination in HPV

The signatures for HPV-16, 18, 6, and 11 follow.

```
   E1 ->
HPV16 .............................................................V..............H..      2
HPV18 ...............................................................................      0
HPV6b ...............................................................................      0
HPV11 ...............................................................................      0

HPV16 ...............................................................................      2
HPV18 ..........................N....................................................      1
HPV6b ...............................................................................      0
HPV11 ...............................................................................      0

HPV16 ...............................................................................      2
HPV18 .......T.............................................C.........................      3
HPV6b ...............................................................................      0
HPV11 ...............................................................................      0

HPV16 ...................................................L.....D.....................      4
HPV18 .............G....V............................................................      5
HPV6b ...............................................................................      0
HPV11 .............S...........................................D.....................      2

HPV16 ............................C.S....M...........................................      7
HPV18 ...............................................................................      5
HPV6b .................................................................A.............      1
HPV11 .................................................................A.............      3

HPV16 ...............................................................................      7
HPV18 .........................M.....................................................      6
HPV6b ...............................................................................      1
HPV11 ...............................................................................      3

HPV16 ............Q....................................A.....................C.......     10
HPV18 .............................................................F...I.............      8
HPV6b ...............................................................................      1
HPV11 ...............................................................................      3

HPV16 ...........................N........L....................................N.....     13
HPV18 ......................................................I......H.................     10
HPV6b .........................................................................T.....      2
HPV11 ...............................................................................      3

                                                                        E1  E2
                                                                         <-->
HPV16 ................................................................N..............     14
HPV18 .............................................A.............................K...     12
HPV6b ...............................................................................      2
HPV11 ...............................................................................      3

HPV16 .............................................................................I.     15
HPV18 ...........I.............................................Y.....................     14
HPV6b .........................V.........................M.........E.................      5
HPV11 ...................................................L.........E.................      5

HPV16 .............V...................................................V.V...........     18
HPV18 ......................................................................T........     15
HPV6b ...............................................................................      5
HPV11 ...............................................................................      5

HPV16 ...............................................................................     18
HPV18 .................C.............................................................     16
HPV6b ...............................................................................      5
HPV11 ...............................................................................      5
```

```
HPV16 ........................T.T.......IQ.P.........................................     23
HPV18 ..................................T..G.........................................     18
HPV6b ...............................................................................      5
HPV11 ...............................................................................      5

HPV16 ...............................................................HKS.............     26
HPV18 ...........................................H...................................     19
HPV6b .....................F..............N.R........................................      8
HPV11 ....................................N..........................................      6

                             E2  L2
                              <-->
HPV16 .................................K.T...........................................     28
HPV18 .......A...........................................V...........................     21
HPV6b ...............................................................................      8
HPV11 ...............................................................................      6

HPV16 ................................................V..............................     29
HPV18 ...............................................................................     21
HPV6b ....................................A..........................................      9
HPV11 ...............................................................................      6

HPV16 ...............D...............................................................     30
HPV18 ........................A....................................................G.     23
HPV6b ...............................................................................      9
HPV11 ...............................................................................      6

HPV16 ..................................................T............................     31
HPV18 ...............................................................................     23
HPV6b ...............................................................................      9
HPV11 .......Q...D.......................R...........................................      9

HPV16 D..........IN.......................T.......................K.......T..........     37
HPV18 ...............................................................................     23
HPV6b ...............................A...............................................     10
HPV11 .....................................................................V.........     10

HPV16 ......TI...TPSTYTT.............................................................     46
HPV18 ...............................................................................     23
HPV6b ...............................................................................     10
HPV11 ...............................................................................     10

                                                                            L2  L1
                                                                             <-->
HPV16 ...................N....S............................................L.......L.     50
HPV18 .............................A.................................................     24
HPV6b ...............................................................................     10
HPV11 ...............................................................................     10

HPV16 .................................T.............................................     51
HPV18 ........................P......................................................     25
HPV6b ...............................................................................     10
HPV11 ...............................................................................     10

HPV16 ...............................................................................     51
HPV18 ...............................................................................     25
HPV6b .................................................................G.............     11
HPV11 .................................................................G.............     11

HPV16 ......................................H................................I.......     53
HPV18 ...............................L......................C........................     27
HPV6b ...............................................................................     11
HPV11 ...............................................................................     11

HPV16 ...............................................................................     53
HPV18 .....................................G.........................................     28
HPV6b ................................I..............................................     12
HPV11 ................................L..............................................     12
```

Recombination in HPV

```
HPV16 ...............................................................................     53
HPV18 ...V..............P............................................................     30
HPV6b ...............................................................................     12
HPV11 ........H......................................................................     13

HPV16 ...........................................D...................................     54
HPV18 .........................................N.................................V...     32
HPV6b ...............................................................................     12
HPV11 ...............................................................................     13

                                       <- L1
HPV16 ...................................                                                 54
HPV18 .........................A.........                                                 33
HPV6b .........V.........................                                                 13
HPV11 .........I.........................                                                 14
```