

Alignment of HERV-K(HML-2) proviral sequences acquired using GAPS

4q321 1
 ATTAATGGTGAGTGTGGAATAATTCACCTTATCTTGTGTATTTGATGATTATAACAAGTCAGACTCTACATACTTTACATTTTTAAGAAAATAGTGCTAAAGTTACTTTTAGAATGAGTGG 120

4q321 121
 GACTTTCCAACCTCACATATTCAGTCATAGCAGGTAACACTTACACAGAGATACAGATTATAAGACGAGTATATAAATGGCTGTTTCAATTCTAACAAAAGTACCAGGAGTTATATAGTAG 240

4q321 241
 TGTCTCTAGTGTTTACAATACTCAGGCAATTGAAAGAAAAAAGACTTTGTTTGTAGTCTAACTAAAATGTGTTTGGAGAATACATACTACTTTGTAAGTAGGCTTGGATATCTTTTCAGA 360
 4q323long ATTAATCTAGAAGACTGGATACCCTACAGATTGTTTCAGAACACATCTATTTTCAATTTATAGTATTTCCAGAGAGAATCTTTCTCA
 K107 ATTAATCTTATAAGTG

4q321 361
 AAAAGTTAAGACAGATGTAAATAATATGAAAAGAGAAATGCATATAGATTTTTTAGATGAAAGAGGGGAGCACACAGCATCCCAAATTGTGATATCGTTTTTGCCTAAGCACCAGGGGTT 480
 4q323long AATTGGGTAATCCATTAAAAAATAGCCCGAAGGCTGGGCGCGGTGGCTCACACTTGTAAATACCAGCACCTTGACAGGCCGAGGTAGGCGGATCACCTGAGGTGAGGAGTTCAAGACCAGC
 K107 CAAAACAAAGCTAAAACTTTAAGTACATGGATTGCTGGGGTGTGTGGGAAGTGGAGAGAAGTAGAGAGGAAGACAATTACAGAACATGTTTGCAACTGTTCAAAAGGGAGAAAAATT

4q321 481
 TTAGGGAAAAAGACACGTTTCAGTGAATAACACCGTGAAATAAATTCATTGAATGTTTGTGTTTTAAATTTTTAAATAAATCCTGAAAGTTTCTAGTGAAGTATACAAAGAATATTGTCTA 600
 4q323long CTGGCCAACATGGTGAACTTCGTCTCTACTAAAAATACAAAAATTGGCTTGGCTTGGTGGGCACCTGTAATCTCAGCTACTGGAGTGGCTGAGGTAGGAGAATTGCTTGAACCCAGGAG
 K107 ACGGATTTCTTCCACAATGGTATCAGTCATACTGCCCCACATACCAAGTACCTACAGCTAAGAAGCTCAATATACACTCCTCCTTGTGTCTGTCTGTCTCTCTCTTTGTATGTGTC
 1p311a TAATTAAGGAAGAAGGTAAAGGAGAG.ATGAAGAGAGTAAAAAGAAGAAAAACAACAAA
 K104 ATTAATAGCCTACCACTAACCAAGTCC

4q321 601
 TTCTAAGGCAGCAGAGAGCATGCCTGGAACAGAAATTGTCTGCTGACTACGGGGTTCAAGGTTTCTTACTCAACTGTATGGAAGAAATAATCTGACTGGAGAGAAGACTGAGATGAGG 720
 4q323long GAAGAGGTTGCAGTGAGCCGAGATTGCGCCACTGCACTCCAGCCTGGGTGACAGAGCAAGATTCGCTCTAAAAAAGAGCCTGAAGTATATATAGTGATAT
 K107 TTTCTCACTTTCAATATATATGTTTATACAGAATGAGGATCCTTTATCTGAAATAAGGGACCAGAAGTGTTTTGAAATATGAATTTTTCCAGATTTTGAAATACCTGCATTACAGGTTGA
 1p311a TTAAAAGTACCTACTTTAAAGAGTACTTTAGCAATTCAATCAAGTTTTTGATACATTCTTACAATAACCGTAGGAAGCAGAATTATTATGATGTTTTACTACAGTTGAAAAATTAGATTT
 K104 AGGACTAGATGGATTACAGCCAAATCTACCAGAGGTACAAAGAGGAGCTGGTACCATTCTTCTGAACTATTCCAATCAATAGAAAAAGAGTTATCCTCCCTAAGTATGTTGA
 K106 ATTAATGTCTACAGCCAGTTGATTTTATGTAAAGGGGATTATT

	721		840
4q321	AAAGAAATTTCTGAATCAGTACCATGCCTTTTGGTTTGGTTTATTTATTTGTTTTCTTCCTTCTCAGAATTACTAAAAGAATAGAATACCAAAAAGGGAACAATGATGGCCGAGTTT		
4q323long	AGGACAAAAATAAATGATAAGTAGGCAGTTGCAGCATATGCAAGTAGCTCCCATTTGGAGATAAAACAAAGGGGACCATTCTATACATAACCTATATTCATATAAATGTTTTATTTTGTGTA		
K107	GCGCCCCAAATCTGAAAATCTGAGATTGAAATGCTCCAATGAGTATTTCCCTTTGAGCTTTATGTGTCAGTCCTCAAAAAGTTTCAATTTTAGAGCATTTTGGATTTGGATTTTGGATT		
1p311a	AGAGAGGCTAAATGTTTTACTTTTCTCCTAGTGACATAAATTATGAACTGTAGAAAAGAGCCCAACTCTTAGATTTTCTCCCCAAACACACGGGCTGGTCCACTTTACCCAATGGTAGA		
K104	AGCCAGCATCATCCGGACACCAAAGCCCGGAGAGACACACAAAAAAGAGAATTTTAGACAAATATCCCTGATGAACATCAATGTGAAAATCCTCAATAAAATACTGGCAAACAGAA		
K106	CTGGGTAATGTGGGTGAGCCTGATCCAATCAGTTGAAAGGCCCTAAGAATAGAACCTAGGTTTTCTGAAGATGTGAAGAAAGAAGAAATTCACCTGCGGACTTCCTCTTCAGCTCCTG		
7q222	ATTAATCCCTTCCTTCACCAGCTAATGCACACCCACTAAGTGACACCCACAGCAAGCCTTATCATTATTATATGCAATAATAAATTTTCAGGTAGTAGCTATGGAGCTTCT		
12q13		ATTAATGGCTGG	

	841		960
4q321	CCTCTTAGAGGACAGTACTGATCCATCTGTGTCTTTGCGTATGGTTGTGAAGGTGATACAGGCCAGGTTTACCATCCATCTGACCCTTCATCATAACAAACAAACCTATGTTCTTTGTGG		
4q323long	TATGCATAGACCATTGGGAATACTTTACAAAAAGCCATTAAACAGCAATTATCCTTTAGAAGGAAAATGAGTGACAGCAAGTACTTTGTAACCTGTTGTGCCAGGTAGGGCTTTCTT		
K107	TGGGACACCCAACCTGCATGGTTTTTTTAATTTTAATTTTTTATTATTAGTTTTTGAGACAAAGTCACCTGTCAACCCAGGCTGGGGTGCAGTGGCATGATCTCAACTCACTG		
1p311a	TCACAAAATCCCATGATGGGTAACAGTGAAGAACCAGTGACTGCCTCAGGTTTCCATTTGGGGTCTCATGAAGCAGAATGTCTCTTCATCTTTTGGGAGATCAACAACCTCATGT		
K104	TCCAGCAGCATATCAGAAAGGTTATCCACCACGATCAAGTCAGCTTCATCCCTGGGATGCAAGTCTGTGTTAACATACACAAATCAATAAACATAATCCATCATATAAACAGAAC		
K106	CCTGAGAGTTTCCAGCCTTATGAATTTTCAGACTTTTCTATCTGACCTCTACAATGGTATAAGTCAATTTCTTTCTGTAAATCCCTTTACATATTCTACTGTTTCTCTTTGGCTGG		
7q222	AAGAGCATAATATTTACTTTGGGATACTTATCTTAAATGCAAATTTCTGACTTCAGTAGTTGTAAGGAACAGCTCTGTAATCTTTTTTTCTTCAATTAATAAATTGTATTTTAGAG		
12q13	ACTATCGGAGAATTCCACCTTCAAATGTTAATTTTACTCTCATGTTCTCCCGAATGATGTTGATAAAAAAAGCTTTAGGTCTAGCACACTTTATCTGTAAATTTGTATACTGCTAT		
4q323short	ATTAATAGCAATTATCCTTTAGAAGGAAAATGAGTGACAGCAAGTAAGTGTGTAAGTGTGTGCCAGGTAGGGCTTTCTT		
K109	ATTAATATTCTATGACACTCACTGCCAGTCTCAGGTGTTTGGATCTTCCACATATATATGC		

	961		1028
4q321	GGAAAAGCAAGAGAGATCAGATTGTTACCGTGTCTGTGTAGAAAAGAAGTAGACATAGGAGACTCCATT		
4q323long	A.....--.....T.....		
K107	A.....T.....		
1p311a	A.....TA.....		
K104	T.....		
K106	T.....		
7q222	T.....		
12q13	T.....		
4q323short	A.....--.....T.....		
K109	T.....		

B

	1		120
Pan8q	GATCCTCCATATGCTGAACGCTGGTTCCCCGGGTCCCTTATTTCTTTCTCTATACTTTGTCTCTGTGTCTTTTTCTTTCTTAAGTCTCTCGTTCCACCTTACGAGAAACACCCACAGGT		
Pan2Ap	G.....		
K107	T.C..A.....C.....		
K110	T.....C.....T.C.....C.....T.....		
2q211	C..T.....A.....		
K104	A.....A.....		
K115	T.....T.C..A.....C.....		
19p12c	T.....T.....T.....T.C..A.....C.....A.....		
	121		240
Pan8q	GTGGAGGGGCAACCCACCCCTACAAGGAGATACCATCTCACACCAGTTAGAATGGCAATCATTAAGTCAGGAACAACAGGTGCTGGAGAGGATGTGGAGAAATAGGAACACTTTTACA		
Pan2Ap	...T.....GCTGATGGGCGAGAACTGTGAGAATCTTTCATTACGGCATGCACAGCCCACAAGTGTGGGGCTTAACACTCCACTAGACCAGTGGGGTGCAGGA		
K107	...T.....ACTGCAACCTCCGCCTCCTAGTTCAGTGATTCTCCTGCCTCAGCCTCCCACTAGCTGGGATTACCGGCATGCACCACCACACCCAGCTAATTTT		
K110	T...TGAGACAATATTTAAAGGTTTGGGGAAATCCTGTAAGGCAGTAATCACAGCAATTAAGTCCGCCTTTTGAGCAGAAGTATAAGAGGTAGAAATAAG		
2q211	T.....T...AGAACTTAATAGTAGGCCCAATTGATAAACTAACAGCGTTTCTATATTCAACATCTTTATGATTCCATGCAGTCTAAGCATGGTTTAAAGCCACG		
K104	CAGAACCAATGACAAAAACCATGATTATCTCAATAGATGCAGAAAAGGCCTTCGACAAAATTCAATAGCCATTAAT		
K115	.G.T.....CCTTTTATATTGATCATTAAATTCACTTCTAACTCTGAGATTCTATGTGATTAAAT		
19p12c	...T.....T...TCTTTGACATCTCTGCTCTAATTTAATTAAT		
	241		360
Pan8q	CTGTTGGTGGGACTGTAAACTAGTTCAACCATTGTGGAAGTCAGTGTGGCAATTCCTCAGGGATCTAGAAGTAAATACCATTTGACCCAGCCATCCCATTACTGGGTATATACCCAAA		
Pan2Ap	GCCAATGAATACATGCTTCTCCCTTTGAACCTGTGCAAAGTCCTGAGATGCATTTCTCCTCAAAGATCTCATGATCCTCACATGTCATGATCCCATAAGACTGGGAGCTGGTTGCCTCTAG		
K107	TGTATTTTGTAGTAGAGATGGGTTTACCATGGTGGCCAGGCTGATCTGGAACCTCTGAGCTCAGGTGATCTCCTGCCTCAGCCTCCCAAAGTGCTGGGATCATAGGCATGAGCTACCC		
K110	TTTGTCTGTAGGACTTACATAGCCAGCATTGCCATTACTGGAGCCATCAGTGAACACTGTAACGGCCTCAGGAATGGGTTGATTTTTGGTTAATCGAGGAACCACCTAAGAAGTCATTTT		
2q211	GTGTTAACATAAGCCTGAGTTAGGTTTAAGAGGTTATATTTTCTGATGTGTAAAGAGGCTTTTGTAGCAAAACAACCTGCAATCTCAGACAGCCCTTATATCTTCACACTGAGGGTCTGAGA		
	361		480
Pan8q	GGACTATAAATCATGCTGCTATAAAGACACATGCACACGCATGTTTATTGCAGCACTATTACAAATAGCTAAGACTTGGAAACCAACCCAAATGTCCAACAATGATAGACTGGATTAAGAA		
Pan2Ap	TGGTAGCTGTTGGGGAAATATAATTAATAAATCAAAATCTCCAAACAGAAATCTTCTCTACAGTGGTAATATAGGAAGAAAACACCTTTACTAATGAATAAGAAACAACTAGAAATATCA		
K107	CACACCTGGCCCTCAACCTGTATGTTTTTAATGAGAACACATGGACACAGGAAGGGGAACATCAC.CTC.GGGGACTACTGTGGGGTGGGGGGAGGGGGAGGGGATAGCATTAGGAGATAT		
K110	TATAAAATCAATAAATTTGTTTTTTGGGATAATGATTGTCAATAACGCCAATAAAATCAGCCAAGTGAATTTGCCACAGTACAGAATGTTGAAAGGCAGCTTGAACCTTGTAGCCAATTTAA		
2q211	ACATGAAAGAGAAGTCAAAGAACCTGCAACTGAGGTGAGAGACAGTGCATTCACATTTTTATCTTGCTAGGGCCCTTCTGAGTCTCAATTAAT		
	481		600
Pan8q	AATGTGGCACATATACACCATACTATGCAGCCATAAAATACTATGCAGCCATAAAAAATGATGAGTTCATGTCCTTTGTAGGGACATGGATGAAACTGGAAATCATCATTTCTCAGTAAACA		
Pan2Ap	TGCACATCACACGGCCATCATACAGGATTGCAAAAGGCAGAAATAATGCTCACCTCTTTATGTACCCAAGCATATACGACCCATCATATACATGTTCTCAAGATAAACAATAAATCATCCTC		
K107	ACCTAATGCTAAATGACCAGTTAATGGGTGCAGCACACCAACATGGCACATGTATACAATATGTAACAAACCTGCACATTGTGCACATGTACCCTAAACCTAATGTATAATAATAATAAA		
K110	AGGAACACAAATTATATTTGGATCAATCCAGAAATTTGAAGTATTCTGCACCAAGCCATTCTAATTAAT		

	601		720
Pan8q	TATCGCAAGGTCAAAAAACAAACACCGCATGTTCTCACTCATAGATGGGAATTGAACAATGAGAATACATGGACACAGGAAGGGGAACATCACACTCTGGGGACTGTTGTGGGGTGGGG		
Pan2Ap	AAGTAAAAGGACTTGACAGCACCATTGCCACATGTAGTTCATCCTCTTAAATTCATCTGGTAATTGTGGTAACCATCTAGGTTAGCTTATTAGCTTTATCCAGTGGAAAAATTAAT		
K107	ATAAAATAAAGAGAGAGGAAAAAAAAAAGTCAGACTAGTTATCCAATATTTTGGTACTGTAACGTGCAGTGGTGTGCGGAGAATTACCAAGCCTCTAATTAAT		
	721		840
Pan8q	GGAAGGGGAGGGATAGCATTAGGAGATATACCTAATGCTAAATGACGAGTTAATGGGTGCAGCACACCAGCATGGCACATGTATACATATGTAACCTAACCTGCACATTGTGCACATGTA		
	841		960
Pan8q	CCCTAAACTTAAAGTATAATAATAATAAAAAATTTGAAAGAAGTGACTTACACCAGATGCACAGAAATCAAGGTAAGGCACAAAAACATTATAAAAAATATATATAAAAAACAAGG		
	961		1080
Pan8q	AAACATGAGACTTCCAAAGGAACAAAATAATTCTTCATCAGTAAATTCCAATAAAAAATGTATAAACTGCCTAAAAGATAATTGAAAACAATGATATTGAATAAGCTCAGTGAGACACAA		
	1081	1125	
Pan8q	GAGAACATAGATAAACAGTATGAAGAAATCAGAAAAACAATTAAT		