# Chapter 4

# Results and Discussion

## 4.1 Amplification of NCCR, VP1 and T-Ag:

JCPyV DNA was extracted from urine, blood and CSF samples from different groups included in the study. The JCPyV non-coding control region (NCCR) was amplified using SDJ1 (5′-CCCTATTCAGCACTTTGTCC-3′) and SDJ2 (5′-CAAACCACTGTGTCTCTGTC-3′) primers that amplified a 586 bp size fragment of NCCR. The amplified products were analyzed by agarose gel electrophoresis. **Fig.s 1, 2 and 3** are the representative gel images of the amplified NCCR. A 100 bp ladder (1$^{st}$ lane in each agarose gel image) was used for estimating the size of the amplified product. Both positive and negative controls were taken during each PCR reaction and the last two lanes in each agarose gel image signify them. pMITC-BSMKS plasmid containing the full-length genome JCPyV Mad1 strain was used as positive control. **Fig.5** represents the NCCR amplified from Rabha and Mech tribes, **Fig.6** represents NCCR from Oraon and Munda tribes and **Fig.7** represents NCCR amplified from pregnant women group. JCPyV positive samples showed sharp bands of the amplified products. The lanes with no bands are JCPyV negative samples. The viral protein 1 region (VP1) was amplified using JCVP1 (5′-TTTTGGGACACTAACAGGAG-3′) and JCVPR (5′-AAAACCAAAGACCCCTC-3′) primers in JCPyV positive individuals. The size of the PCR product amplified with these primers is about 500 bp. **Fig.8** is the representative gel image of amplified VP1 region from tribal group. JCVT1 (5′-GAATAGGGAGGAATCCATGG-3′) and JCVT2 (5′-GGAATGCATGCAGATCTACAGG-3′) primers were used to amplify the large T-antigen (T-Ag) coding region that amplified a product of about 770 bp. **Fig.9** represents the amplified T-Ag from tribal groups.

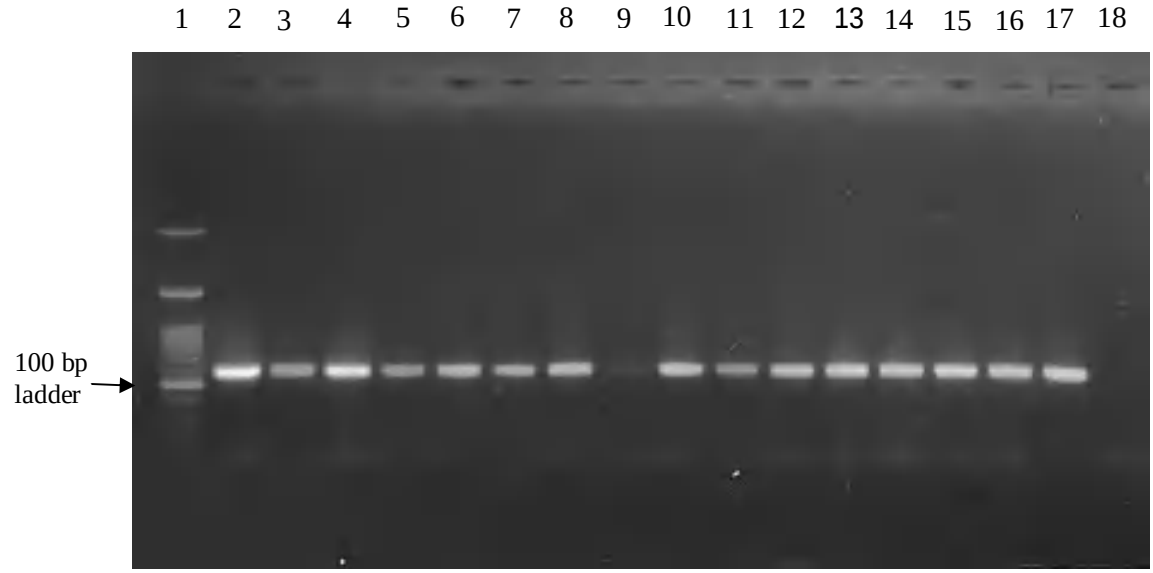



**Figure 5:** The NCCRs of endemic JCPyV in 1% agarose gel amplified from Rabha and Mech tribal populations using SDJ1 and SDJ2 primers. Lane 1: 100 bp DNA ladder; lanes 2–16: amplified NCCR products, light bands were observed in the lane number 9; lane 17: positive control and lane 18: negative control.

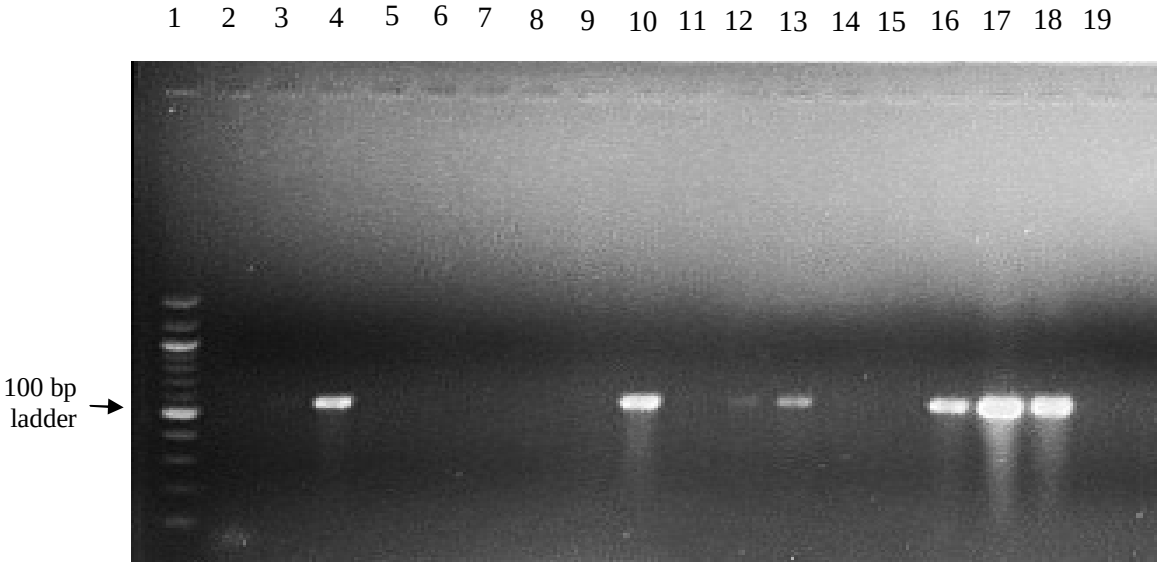



**Figure 6:** The NCCR of JCPyV in 1% agarose gel amplified from Oraon and Munda tribal population using SDJ1 and SDJ2 primers. Lane 1: 100 bp DNA ladder; lanes 4, 10, 12, 13, 16, 17: amplified NCCR products; lanes 2, 3, 5, 6, 7, 8, 9, 11, 14, 15: JCPyV negative samples; lane 18: positive control and lane 19: negative control.

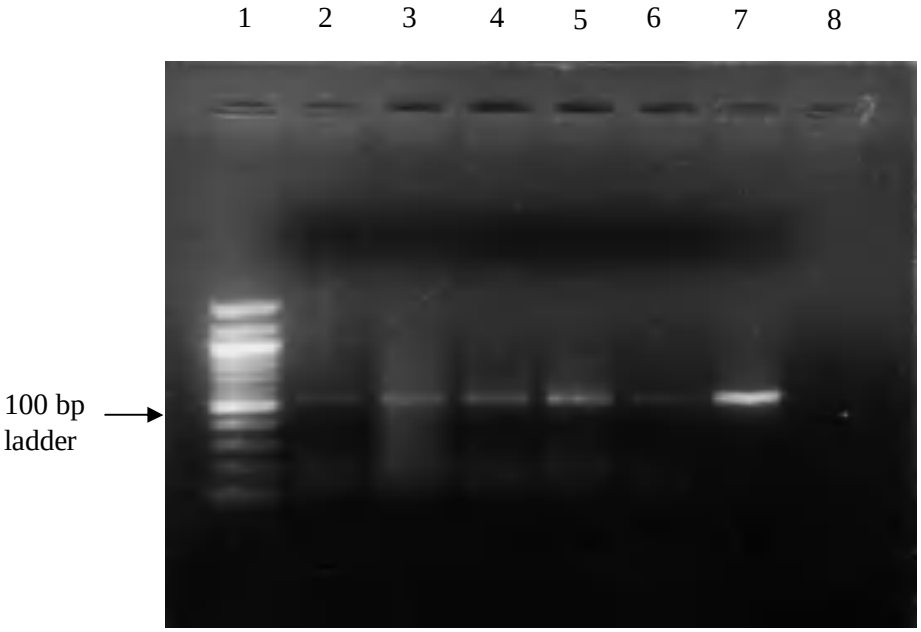



**Figure 7:** The NCCR of JCPyV in 1% agarose gel from the pregnant women using SDJ1 and SDJ2 primers. Lane 1: 100 bp DNA ladder; lanes 2, 3, 4, 5, 6: amplified NCCR products; lane 7: positive control and lane 8: negative control.

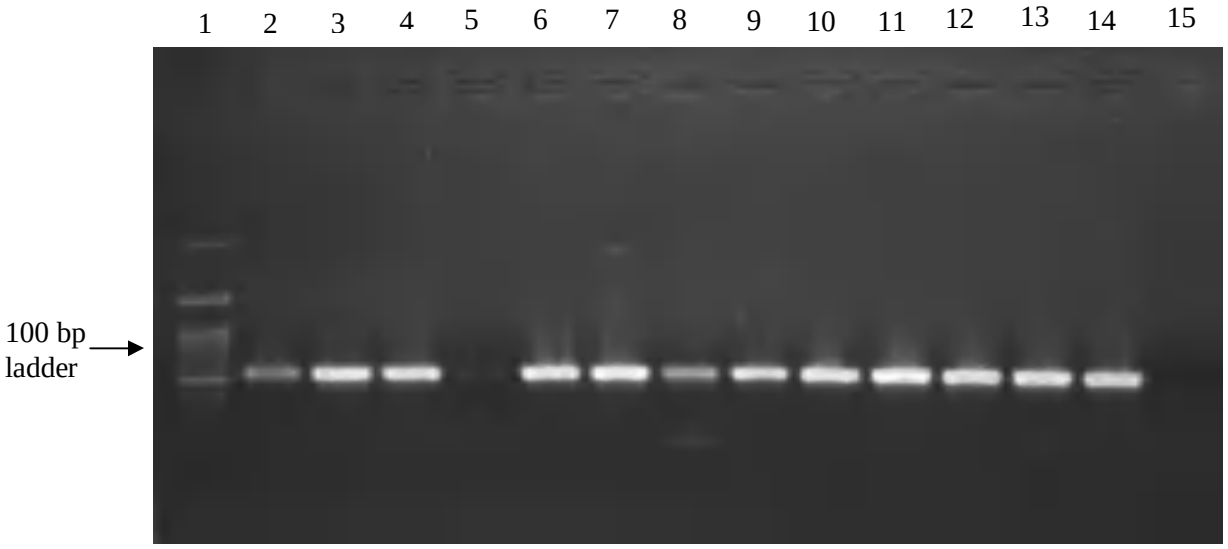



**Figure 8:** Amplified VP1 region of JCPyV in 1% agarose gel from tribal groups using JCVP1 and JCVPR primers. Lane 1: 100 bp DNA ladder; lanes 2-13: amplified VP1 products; amplified bands are light in lane 5; lane 14: positive control and lane 15: negative control.

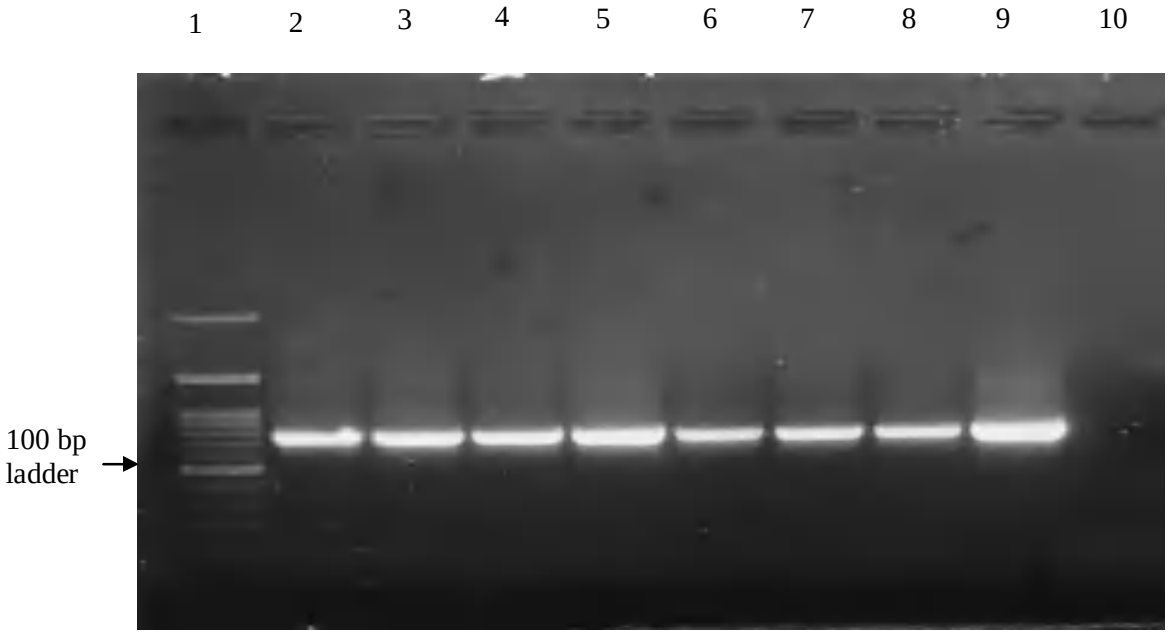



**Figure 9:** Amplified T-Ag of JCPyV in 1% agarose gel from tribal groups using JCVT1 and JCVT2 primers. Lane 1: 100 bp DNA ladder, lanes 2-8: amplified T-Ag products, lane 9: positive control and lane 10: negative control.

## 4.2 Prevalence of JC virus:

The present study was conducted in the sub-Himalayan part of West Bengal, a region from where no such genetic epidemiological documentation has so far been reported. A total of 613 samples were collected from different health centres/regions of Jalpaiguri, Darjeeling, Coochbehar and Alipurduar districts of West Bengal. The study consisted of two groups: one group containing only immunocompromised subjects and the other one containing non-immunocompromised individuals. The group containing immunocompromised subjects consisted of women who were pregnant in their 3$^{rd}$ trimester, patients receiving steroids and cancer patients undergoing chemotherapy (**Table 16**). Changes in immune system occur during 2$^{nd}$ and 3$^{rd}$ trimester of pregnancy. Hence, samples from pregnant women in their 3$^{rd}$ trimester were considered for the study, as it has been observed in earlier studies that reactivation of JCPyV can occur during 2$^{nd}$ and 3$^{rd}$ trimester of pregnancy (Markowitz *et al*., 1991; Markowitz *et al*., 1993). Samples from patients receiving steroid drugs such as Prednisolone, Deflazacort etc., and cancer patients undergoing chemotherapy were examined for the presence of viral DNA.

Prednisolone is a steroid drug used in treatment of inflammatory and autoimmune conditions. Deflazacort is a glucocorticoid used as an anti-inflammatory and imunosuppressant. These steroid drugs when consumed for a longer period suppress immunity. Chemotherapy is responsible for weakened immune system in cancer patients.

| **Table 16: Demographic characteristics of the Immunocompromised individuals.** | | |
|---|---|---|
| Pregnant Women (3rd trimester) | Number of cases | 197 |
| | Age (Range) | 18-40 |
| | Gender | Female:197 |
| | Specimen Type | Urine:83<br>Blood:114 |
| Steroid drug receiving patient | Number of cases | 56 |
| | Age (Range) | 3-75 |
| | Gender | Male:40<br>Female:16 |
| | Specimen Type | Urine:56 |
| Cancer Patients | Number of cases | 11 |
| | Age (Range) | 4-58 |
| | Gender | Male:7<br>Female:4 |
| | Specimen Type | CSF:11 |

The group containing non-immunocompromised group subjects consisted of individuals with healthy immune systems. Some of them belonged to the tribal groups of this region and a few non-tribal individuals have also been included in this group (**Table 17**).

| **Table 17: Demographic characteristics of the non-immunocompromised individuals** | | | |
|---|---|---|---|
| Tribal Group | Oraon and Munda tribes | Number of cases | 186 |
| | | Age (Median age) | 15-70 (32) |
| | | Gender | Male:73,<br>Female:113 |
| | | Specimen Type | Urine:74,<br>Blood:112 |
| | Rabha tribes | Number of cases | 103 |
| | | Age (Median age) | 18-75 (38) |
| | | Gender | Male:64,<br>Female:39 |
| | | Specimen Type | Urine:50,<br>Blood:53 |
| | Mech tribes | Number of cases | 36 |
| | | Age (Median age) | 6-60 (33) |
| | | Gender | Male:16,<br>Female:20 |
| | | Specimen Type | Urine:36 |
| Normal healthy individuals | | Number of cases | 24 |
| | | Age (Median age) | 23-58 (27) |
| | | Gender | Male:14,<br>Female:10 |
| | | Specimen Type | Urine:24 |

Of the 613 samples, 50 samples tested positive for JCPyV virus with an overall incidence rate of 8.15% in this region (**Table 18**). One hundred and ninety-seven (197) samples were from pregnant women, of which 27 samples tested positive for JCPyV with an incidence rate of 13.70% in this group. Out of 11 cancer patients undergoing chemotherapy, only one individual showed presence of JCPyV DNA with an incidence rate of 9.09% in the group. A total of 56 samples were collected from patients receiving steroid medicines, none of them showed presence of JCPyV genome.

Of the 24 non-tribal individuals, none showed presence of JC viral DNA in them. Four tribal groups included in the present study are: Oraon, Munda, Rabha and Mech. A total of 325 samples were collected in this group including 22 samples that tested positive for JCPyV with an incidence rate of 6.76%. Details of the samples from each tribal group are given in **Table 19**. Of this 22 people, 7 (3.76%) individuals were from Oraon and Munda tribes, 13 (12.62%) individuals were from Rabha tribe and 2 (5.55%) individuals were of Mech tribal group. The sub-Himalayan parts of West Bengal have long been inhabited by several tribal groups. The Oraons are the second largest and the Mundas are the third largest tribal population constituting about 14% and 7.8% respectively of total tribal population of West Bengal (Census of India, 2001).

**Table 18: Total number of urines, blood and CSF samples screened and their status in each group**

| Sl .No. | Samples | Specimen Type | Positive | Negative | Total | Prevalence rate (%) |
|---|---|---|---|---|---|---|
| 1 | Pregnant Women | Urine | 18 (21.68%) | 65 | 83 | 27 (13.70 %) |
| | | Blood | 9 (7.89%) | 105 | 114 | |
| 2 | Steroid drug receiving patient | Urine | 0 (0 %) | 56 | 56 | 0 (0%) |
| 3 | Normal healthy people | Urine | 0 (0 %) | 24 | 24 | 0 (0%) |
| 4 | Tribal Group | Urine | 12 (7.5 %) | 148 | 160 | 22 (6.76%) |
| | | Blood | 10 (6.06 %) | 155 | 165 | |
| 5 | Cancer Patients | CSF | 1 (9.09%) | 10 | 11 | 1 (9.09 %) |
| **Total** | | | **50 (8.15 %)** | **563** | **613** | **50 (8.15 %)** |

**Table 19: Summary of results of JCPyV detection in different tribal groups**

| S. No. | Tribal Groups | Collection site | Total No of Individuals | Sex | Number of Positive Samples (%) | JCV Positive Specimen Type |
|---|---|---|---|---|---|---|
| 1 | Oraon and Munda | Naxalbari, Darjeeling District | 186 | F=113 M=73 | 7 (3.76 %) F=03 M=04 | Urine=7 |
| 2 | Rabha | Poro Busty, Coochbehar District | 103 | F=39 M= 64 | 13 (12.62 %) F= 01 M=12 | Urine=3 Blood=10 |
| 3 | Mech | Mendabari, Jalpaiguri District | 36 | F=20 M=16 | 2 (5.55%) F=01 M=01 | Urine=2 |
| | | Total | 325 | F=172 M=153 | 22 (6.76 %) F=05 M=17 | Urine=12 Blood=10 |

Prevalence of JCPyV has been recorded in the tribal population worldwide showing variations in the incidence range of the virus in different groups such as 56% to 66% in Native Americans (Agostini *et al.*, 1997), 20% to 22% in the tribal people of Africa (Chima *et al*., 1998), 47% to 55 % in Bunun tribes of Taiwan (Chang *et al*., 1999) and 48% to 67% in Myanmar tribals (Saruwatari *et al*., 2002).

The JCPyV status in both the groups was verified statistically. The incidence rate of JCPyV in the immunocompromised group was 10.60% i.e., 28 individuals among the 264 individuals tested positive for JCPyV. JCPyV status within the non-

immunocompromised group i.e., in individuals with a healthy immune system was 6.30%. Twenty-two (22) out of 349 samples tested positive for JCPyV in this group. The incidence rate of JCPyV in the immunocompromised group of people was higher compared to the group consisting of non-immunocompromised individuals (**Fig. 10**) which was in agreement with the results reported earlier (Azzi *et al*., 1996; Zanotta *et al*., 2013; Boukoum *et al*., 2016).

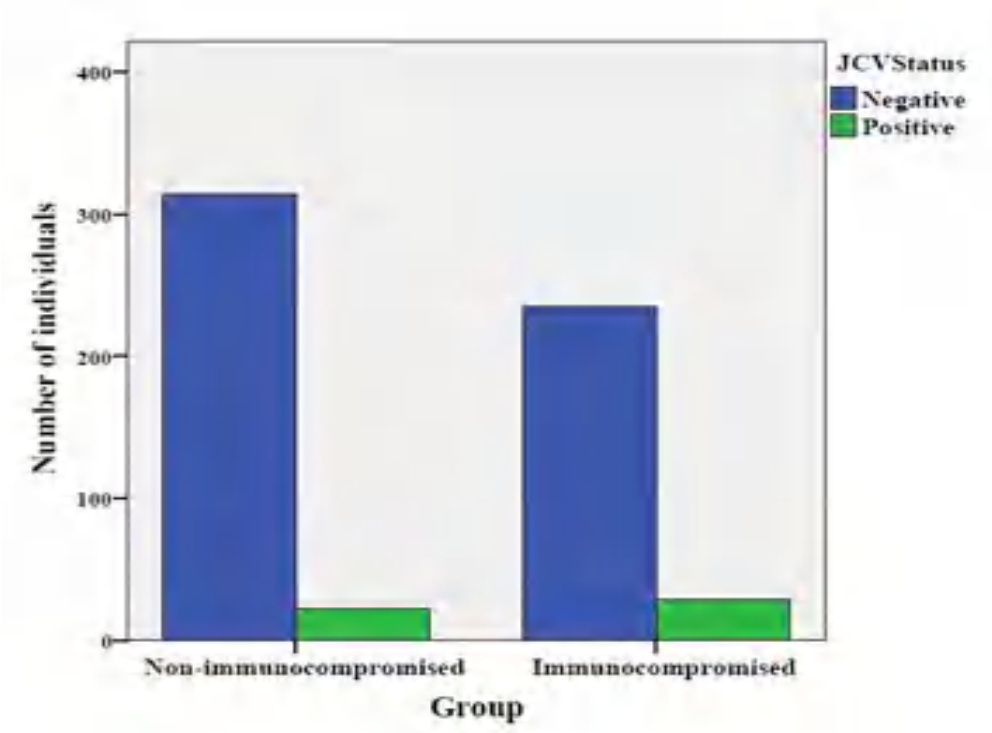



**Figure 10:** JCPyV status in immunocompromised and non-immunocompromised groups.

The possible association of prevalence of the virus with the gender of the subjects was analyzed statistically. The frequency of JCPyV detection appears to differ between the genders and was found to be higher in males compared to the females in the tribes of this region with a *P* value of 0.003 ($\chi^2$ =8.636, df=1, N=325) which is in accordance with the previous studies done on Native Americans (Agostini *et al.,* 1997) and the tribal

people of Africa (Chima *et al*., 1998). In the case of the Mech tribe, due to the low positive detection of JCPyV (5.55%) among the random sample, the association between virus status and sex was not significant enough. Of 172 women, 5 (2.9%) were found to be JCPyV positive whereas, in males, 17 out of 153 were positive with an incidence rate of about 11.11% (**Fig. 11**).

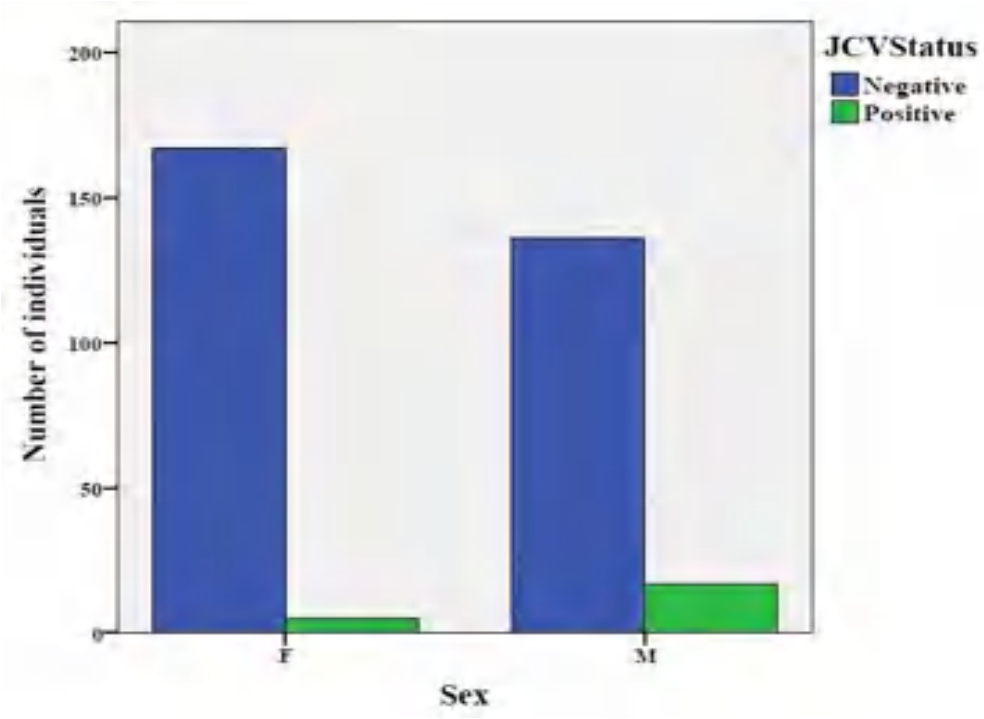



**Figure 11:** JCPyV status based on gender in tribal groups

## 4.3 Sequences submission in GenBank with their accession numbers:

Nine (9) NCCR region sequences and their GenBank accession number (**Table 20**), ten VP1 sequences with their GenBank accession numbers (**Table 21**) and six T-antigen sequences and their GenBank accession numbers (**Table 22**) are listed below.

| Table 20: NCCR sequences of JCPyV with their accession number | | |
|---|---|---|
| **S.No.** | **Sequence name and accession number** | **Nucleotide sequence of Non coding control region** |
| 1 | **JX294575.1**<br><br>**JC polyomavirus isolate NB1 T antigen and agnoprotein genes, partial cds** | CCCTATTCAGCACTTTGTCCATTTTAGCTTTTTGCAG<br>CAAAAAATTACTGCAAAAAAGGGAAAAACAAGGG<br>AATTTCCCTGGCCTCCTAAAAAGCCTCCACGCCCTT<br>ACTACTTCTGAGTAAGCTTGGAGGCGGAGGCGGCC<br>TCGGCCTCCTGTATATATAAAAAAAAGGGAAGGTA<br>GGGAGGAGCTGGCTGGCTGCCAGCCAAGCATGAGC<br>TCATACCTAGGGAGCCAACCAGCTGCCAGCCAGAG<br>GGAGCCCTGGCTGCATGCCACTGGCAGTTATAGTG<br>AAACCCCTCCCATAGTCCTTAATCACAAGTAAACA<br>AAGCACAAGGGGAAGTGGAAAGCAGCCAGGAGCA<br>TGTTTTGCGAGCCAGAGCTGTTTTGGCTTGTCACCA<br>GCTGGCCATGGTTCTTCGCCAGCTGTCACGTAAGGC<br>TTCTGTGAAAGTTAGTAAAACCTGGAGTGGAACTA<br>AAAAAAGAGCTCAGAGGATTTTAATTTTTTTGTTAG<br>AATTTTTGCTGGATTTTTGCACAGGTGAAGACAGTG<br>TAGACGGGAAAAAAAGACAGAGACACACAGTGGT<br>TTG |
| 2 | **JX534216.1**<br><br>**JC polyomavirus strain NB2 agnoprotein gene, partial cds** | TTTTGCTTTTTGTAGCAAAAAATTACAGCAAAAAAG<br>GGAAAAACAAGGGAATTTCCCTGGCCTCCTAAAAA<br>GCCTCCACGCCCTTACTACTTCTGAGTAAGCTTGGA<br>GGCGGAGGCGGCCTCGGCCTCCTGTATATATAAAA<br>AAAAGGGAAGGTAGGGAGGAGCTGGCTGGCTGCC<br>AGCCAAGCATGAGCTCATATCTAGGGAGCCAACCA<br>GCTGCCAGCCAGAGGGAGCCCTGGCTGCATGCCAC<br>TGGCAGTTATAGTGAAACCCCTCCCATAGTCCTTAA<br>TCACAAGTAAACAAAGCACAAGGGGAAGTGGAAA<br>GCAGCCAGGAGCATGTTTTGCGAGCCAGAGCTGTT<br>TTGGCTTGTCACCAGCTGGCCATGGTTCTTCGCCAG<br>CTGTCACGTAAGGCTTCTGTGAAAGTTAGTAAAAC<br>CTGGAGTGGAACTAAAAAAAGAGCTCAGAGGATTT<br>TAATTTTTTTGTTAGAATTTTTGCTGGATTTTTGCAC<br>AGGTGAAGACAGTGTAGACGGG |
| 3 | **JX534217.1**<br><br>**JC polyomavirus strain NB3 agnoprotein gene, partial cds** | TTTTGCTTTTTGTAGCAAAAAATTAGAGCAAAAAA<br>GGGAAAAACAAGGGAATTTCCCTGGCCTCCTAAAA<br>AGCCTCCACGCCCTTACTACTTCTGAGTAAGCTTGG<br>AGGCGGAGGCGGCCTCGGCCTCCTGTATATATAAA<br>AAAAAGGGAAGGTAGGGAGGAGCTGGCTGGCTGC<br>CAGCCAAGCATGAGCTCATACCTAGGGAGCCAACC<br>AGCTGCCAGCCAGAGGGAGCCCTGGCTGCATGCCA<br>CTGGCAGTTATAGTGAAACCCCTCCCATAGTCCTTA<br>ATCACAAGTAAACAAAGCACAAGGGGAAGTGGAA |

| | | |
|---|---|---|
| | | AGCAGCCAGGAGCATGTTTTGCGAGCCAGAGCTGT<br>TTTGGCTTGTCACCAGCTGGCCATGGTTCTTCGCCA<br>GCTGTCACGTAAGGCTTCTGTGAAAGTTAGTAAAA<br>CCTGGAGTGGAACTAAAAAAAGAGCTCAGAGGATT<br>TTAATTTTTTTGTTAGAATTTTTGCTGAATTTTTGCA<br>CAGGTGAAGACAGTGTAGACGGGAAAAAAAGACA<br>GAGA |
| 4 | **JX534218.1 JC polyomavirus strain NB4 agnoprotein gene, partial cds** | TCCATTTTTGCTTTTTGTAGCAAAAAATTAGAGCAA<br>AAAAGGGAAAAACAAGGGAATTTCCCTGGCCTCCT<br>AAAAAGCCTCCACGCCCTTACTACTTCTGAGTAAGC<br>TTGGAGGCGGAGGCGGCCTCGGCCTCCTGTATATA<br>TAAAAAAAAGGGAAGGTAGGGAGGAGCTGGCTGG<br>CTGCCAGCCAAGCATGAGCTCATACCTAGGGAGCC<br>AACCAGCTGCCAGCCAGAGGGAGCCCTGGCTGCAT<br>GCCACTGGCAGTTATAGTGAAACCCCTCCCATAGTC<br>CTTAATCACAAGTAAACAAAGCACAAGGGGAAGTG<br>GAAAGCAGCCAGGAGCATGTTTTGCGAGCCAGAGC<br>TGTTTTGGCTTGTCACCAGCTGGCCATGGTTCTTCG<br>CCAGCTGTCACGTAAGGCTTCTGTGAAAGTTAGTA<br>AAACCTGGAGTGGAACTAAAAAAAGAGCTCAGAG<br>GATTTTAATTTTTTTGTTAGAATTTTTGCTGGATTTT<br>TGCACAGGTGAAGACAGTGTAGACGGGAAAAAAA<br>G |
| 5 | **JX534219.1 JC polyomavirus strain NB5 agnoprotein gene, partial cds** | TTGTCCATTTTTGCTTTTTGTAGCAAAAAATTAGAG<br>CAAAAAAGGGAAAAACAAGGGAATTTCCCTGGCCT<br>CCTAAAAAGCCTCCACGCCCTTACTACTTCTGAGTA<br>AGCTTGGAGGCGGAGGCGGCCTCGGCCTCCTGTAT<br>ATATAAAAAAAAGGGAAGGTAGGGAGGAGCTGGC<br>TGGCTGCCAGCCAAGCATGAGCTCATACCTAGGGA<br>GCCAACCAGCTGCCAGCCAGAGGGAGCCCTGGCTG<br>CATGCCACTGGCAGTTATAGTGAAACCCCTCCCATA<br>GTCCTTAATCACAAGTAAACAAAGCACAAGGGGAA<br>GTGGAAAGCAGCCAGGAGCATGTTTTGCGAGCCAG<br>AGCTGTTTTGGCTTGTCACCAGCTGGCCATGGTTCT<br>TCGCCAGCTGTCACGTAAGGCTTCTGTGAAAGTTAG<br>TAAAACCTGGAGTGGAACTAAAAAAAGAGCTCAGA<br>GGATTTTAATTTCTTTGTTAAAATTTTTGCTGGATTT<br>TTGCACAGGTGAAGACAGTGTATACGGGAAAAAAA<br>GAC |
| 6 | **JX534220.1 JC polyomavirus strain NB6 agnoprotein gene, partial cds** | TTTTGCTTTTTGTAGCAAAAAATTACAGCAAAAAAG<br>GGAAAAACAAGGAATTTCCCTGGCCTCCTAAGAAG<br>CCTCCACGCCCTTACTACTTCTGAGTAAGCTTGGAG<br>GCGGAGGCGGCCTCGGCCTCCTGTATATATAAAAA<br>AAAGGGAAGGTAGGGAGGAGCTGGCTAAAACTGG |

| | | |
|---|---|---|
| | | ATGGCTGCCAGCCAACCATGAGCTCATACCTAGGG<br>AGCCAACCAGCTGACAGCCAGAGGGAGCCCTGGCT<br>GCATGCCACTGGCAGTTATAGTGAAACCCCTCCCAT<br>AGTCCTTAATCACAAGTAAACAAAGCACAAGGGGA<br>AGTGGAAAGCAGCCAGGGGAACATGTTTTGCGAGC<br>CAGAGCTGTTTTGGCTTGTCACCAGCTGGCCATGGT<br>TCTTCGCCAGCTGTCACGTAAGGCTTCTGTGAAAGT<br>TAGTAAAACCTGGAGTGGAACTAAAAAAAGAGCTC<br>AAAGGATTTTAATTTTTTTGTTAGAATTTTTGCTGG<br>A |
| 7 | **KF739298.1**<br><br>**JC polyomavirus isolate NB-PU1 agnoprotein gene, partial cds** | AGCACTTTGTCCATTTTAGCTTTTTGTAGCAAAAAA<br>TTAGCGCAAAAAAGGGAAAAACAAGGGAATTTCCC<br>TGGCCTCCTAAAAAGCCTCCACGCCCTTACTACTTC<br>TGAGTAAGCTTGGAGGCGGAGGCGGCCTCGGCCTC<br>CTGTATATATAAAAAAAAGGGAAGGTAGGGAGGA<br>GCTGGCTGGCTGCCAGCCAAGCATGAGCTCATACC<br>TAGGGAGCCAACCAGCTGCCAGCCAGAGGGAGCCC<br>TGGCTGCATGCCACTGGCAGTTATAGTGAAACCCCT<br>CCCATAGTCCTTAATCACAAGTAAACAAAGCACAA<br>GGGGAAGTGGAAAGCAGCCAGGAGCATGTTTTGCG<br>AGCCAGAGCTGTTTTGGCTTGTCACCAGCTGGCCAT<br>GGTTCTTCGCCAGCTGTCACGTAAGGCTTCTGTGAA<br>AGTTAGTAAAACCTGGAGTGGAACTAAAAAAAGAG<br>CTCAAAGGATTTTAATTTTTTTGGTAAAATTTTTGCT<br>GATTTTT |
| 8 | **KF739299.1**<br><br>**JC polyomavirus isolate NB-PU2 agnoprotein gene, partial cds** | AGCACTTTGTCCATTTTAGCTTTTTGTAGCAAAAAA<br>TTAGCTGAAAAAAGGGAAAAACAAGGGTATTTCCC<br>TGGCCTCCTAAAAAGCCTCCACGCCCTTACTACTTC<br>TGAGTAAGCTTGGAGGCGGAGGCGGCCTCGGCCTC<br>CTGTATATATAAAAAAAAGGGAAGGTAGGGAGGA<br>GCTGGCTGGCTGCCAGCCAAGCATGAGCTCATACC<br>TAGGGAGCCAACCAGCTGCCAGCCAGAGGGAGCCC<br>TGGCTGCATGCCACTGGCAGTTATAGTGAAACCCCT<br>CCCATAGTCCTTAATCACAAGTAAACAAAGCACAA<br>GGGGAAGTGGAAAGCAGCCAGGAGCATGTTTTGCG<br>AGCCAGAGCTGTTTTGGCTTGTCACCAGCTGGCCAT<br>GGTTCTTCGCCAGCTGTCACGTAAGGCTTCTGTGAA<br>AGTTAGTAAAACCTGGAGTGGAACTAAAAAAAGA |
| 9 | **MK904563**<br><br>**JC polyomavirus NB-PU3 agnoprotein gene,** | TTGTCCATTTTTGCTTTTTGTAGCAAAAAATTAGAG<br>CAAAAAAGGGAAAAACAAGGGAATTTCCCTGGCCT<br>CCTAAAAAGCCTCCACGCCCTTACTACTTCTGAGTA<br>AGCTTGGAGGCGGAGGCGGCCTCGGCCTCCTGTAT<br>ATATAAAAAAAAGGGAAGGTAGGGAGGAGCTGGC<br>TGGCTGCCAGCCAAGCATGAGCTCATACCTAGGGA |

| | | |
|---|---|---|
| | **partial cds** | GCCAACCAGCTGCCAGCCAGAGGGAGCCCTGGCTG<br>CATGCCACTGGCAGTTATAGTGAAACCCCTCCCATA<br>GTCCTTAATCACAAGTAAACAAAGCACAAGGGGAA<br>GTGGAAAGCAGCCAGGAGCATGTTTTGCGAGCCAG<br>AGCTGTTTTGGCTTGTCACCAGCTGGCCATGGTTCT<br>TCGCCAGCTGTCACGTAAGGCTTCTGTGAAAGTTAG<br>TAAAAC |

| **Table 21: VP1 sequences of JCPyV with their accession number** | | |
|---|---|---|
| **S.No.** | **Sequence name and accession number** | **Nucleotide sequence of VP1 region** |
| 1 | **MK728940.1**<br><br>**JC polyomavirus isolate NB2 VP1 gene, partial cds** | TAACAGGAGGAGAAAATGTTCCTCCAGTTCTTCATA<br>TAACAAACACTGCCACAACAGTGCTGCTTGATGAAT<br>TTGGTGTTGGGCCACTTTGCAAAGGTGACAACTTGT<br>ATTTGTCAGCTGTTGATGTTTGTGGCATGTTTACTAA<br>CAGATCTGGTTCCCAGCAGTGGAGAGGACTGTCCA<br>GATATTTTAAGGTTCAGCTAAGAAAAAGGAGGGTT<br>AAAAACCCCTACCCAATTTCTTTCCTTCTTACTGATT<br>TAATTAACAGAAGGACCCCTAGAGTTGATGGGCAG<br>CCTATGTATGGCATGGATGCTCAGGTAGAGGAGGTT<br>AGAGTGTTTGAGGGGACAGAGGAACTTCCAGGGGA<br>CCCAGACATGATGAGATATGTTGACAGATATGGAC<br>AGTTGCAGACAAAGATGCTGTAATCAAAGGCCTTTA<br>TTGTAATATGCAGTACAATTTAATAAAGTATAACCA<br>GCTTTACTTTACAGTTGCAGTTATTTT |
| 2 | **MH671352.1 JC polyomavirus isolate NB4 VP1 gene, partial cds** | CTAACAGGAGGAGAAAATGTTCCTCCAGTTCTTCAT<br>ATAACAAACACTGCCACAACAGTGCTGCTTGATGA<br>ATTTGGTGTTGGGCCACTTTGCAAAGGTGACAACTT<br>GTATTTGTCAGCTGTTGATGTTTGTGGCATGTTTACT<br>AACAGATCTGGTTCCCAGCAGTGGAGAGGACTGTC<br>CAGATATTTTAAGGTTCAGCTAAGAAAAAGGAGGG<br>TTAAAAACCCCTACCCAATTTCTTTCCTTCTTACTGA<br>TTTAATTAACAGAAGGACCCCTAGAGTTGATGGGCA<br>GCCTATGTATGGCATGGATGCTCAGGTAGAGGAGG<br>TTAGAGTGTTTGAGGGGACAGAGGAACTTCCAGGG<br>GACCCAGACATGATGAGATATGTTGACAGATATGG<br>ACAGTTGCAGACAAAGATGCTGTAATCAAAGGCCT<br>TTATTGTAATATGCAGTACAATTTAATAAAGTATAA<br>CCAGCTTTACTTTACAGTTGCAGTTATTTTGGGGGA<br>GGG |

| | | |
|---|---|---|
| 3 | **MH708168.1**<br><br>**JC polyomavirus isolate NB7 VP1 gene, partial cds** | TAACAGGAGGAGAAAATGTTCCTCCAGTTCTTCATA<br>TAACAAACACTGCCACAACAGTGTTGCTTGATGAAT<br>TTGGTGTTGGGCCACTTTGCAAAGGTGACAACTTAT<br>ACTTGTCAGCTGTTGATGTCTGTGGCATGTTTACAA<br>ACAGGTCTGGTTCCCAGCAGTGGAGAGGACTCTCCA<br>GATATTTTAAGGTGCAGCTAAGGAAAAGGAGGGTT<br>AAAAACCCCTACCCAATTTCTTTCCTTCTTACTGATT<br>TAATTAACAGAAGGACTCCTAGAGTTGATGGGCAG<br>CCTATGTATGGCATGGATGCTCAAGTAGAGGAGGTT<br>AGAGTTTTTGAGGGAACAGAGGAGCTTCCAGGGGA<br>CCCAGACATGATGAGATACGTTGACAAATATGGAC<br>AGTTGCAGACAAAAATGCTGTAATCAAAAGCCTTTA<br>TTGTAATATGCAGTACATTTTAATAAAGTATAACCA<br>GCTTTACTTAACAGTTGCAGTTATTTTGGGGGAGGG<br>G |
| 4 | **MH744484.1**<br><br>**JC polyomavirus isolate NB8 VP1 gene, partial cds** | CTAACAGGAGGAGAAAATGTTCCTCCAGTTCTTCAT<br>ATAACAAACACTGCCACAACAGTGTTGCTTGATGAA<br>TTTGGTGTTGGGCCACTTTGCAAAGGTGACAACTTA<br>TACTTGTCAGCTGTTGATGTCTGTGGCATGTTTACA<br>AACAGGTCTGGTTCCCAGCAGTGGAGAGGACTCTCC<br>AGATATTTTAAGGTGCAGCTAAGGAAAAGGAGGGT<br>TAAAAACCCCTACCCAATTTCTTTCCTTCTTACTGAT<br>TTAATTAACAGAAGGACTCCTAGAGTTGATGGGCA<br>GCCTATGTATGGCATGGATGCTCAAGTAGAGGAGG<br>TTAGAGTTTTTGAGGGAACAGAGGAGCTTCCAGGG<br>GACCCAGACATGATGAGATACGTTGACAAATATGG<br>ACAGTTGCAGACAAAAATGCTGTAATCAAAAGCCT<br>TTATTGTAATATGCAGTACATTTTAATAAAGTATAA<br>CCAGCTTTACTTAACAGTTGCAGTTATTTTGGGGGA<br>GGGG |
| 5 | **MH744485.1 JC polyomavirus isolate NB9 VP1 gene, partial cds** | CTAACAGGAGGAGAAAATGTTCCTCCAGTTCTTCAT<br>ATAACAAACACTGCCACAACAGTGTTGCTTGATGAA<br>TTTGGTGTTGGGCCACTTTGCAAAGGTGACAACTTA<br>TACTTGTCAGCTGTTGATGTCTGTGGCATGTTTACA<br>AACAGGTCTGGTTCCCAGCAGTGGAGAGGACTCTCC<br>AGATATTTTAAGGTGCAGCTAAGGAAAAGGAGGGT<br>TAAAAACCCCTACCCAATTTCTTTCCTTCTTACTGAT<br>TTAATTAACAGAAGGACTCCTAGAGTTGATGGGCA<br>GCCTATGTATGGCATGGATGCTCAAGTAGAGGAGG<br>TTAGAGTTTTTGAGGGAACAGAGGAGCTTCCAGGG<br>GACCCAGACATGATGAGATACGTTGACAAATATGG<br>ACAGTTGCAGACAAAAATGCTGTAATCAAAAGCCT<br>TTATTGTAATATGCAGTACATTTTAATAAAGTATAA<br>CCAGCTTTACTTAACAGTTGCAGTTATTTTGG |

| | | |
|---|---|---|
| 6 | **MH758781.1 JC polyomavirus isolate NB10 VP1 gene, partial cds** | AGGAGGAGAAAATGTTCCTCCAGTTCTTCATATAAC<br>AAACACTGCCACAACAGTGTTGCTTGATGAATTTGG<br>TGTTGGGCCACTTTGCAAAGGTGACAACTTATACTT<br>GTCAGCTGTTGATGTCTGTGGCATGTTTACAAACAG<br>GTCTGGTTCCCAGCAGTGGAGAGGACTCTCCAGATA<br>TTTTAAGGTGCAGCTAAGGAAAAGGAGGGTTAAAA<br>ACCCCTACCCAATTTCTTTCCTTCTTACTGATTTAAT<br>TAACAGAAGGACTCCTAGAGTTGATGGGCAGCCTA<br>TGTATGGCATGGATGCTCAAGTAGAGGAGGTTAGA<br>GTTTTTGAGGGAACAGAGGAGCTTCCAGGGGACCC<br>AGACATGATGAGATACGTTGACAAATATGGACAGT<br>TGCAGACAAAAATGCTGTAATCAAAAGCCTTTATTG<br>TAATATGCAGTACATTTTAATAAAGTATAACCAGCT<br>TTACTTAACAGTTGCAGTTATTTTGGGGGAGGGGT |
| 7 | **MK728936.1**<br><br>**JC polyomavirus isolate NB11 VP1 gene, partial cds** | CAGGAGGAGAAAATGTTCCTCCAGTTCTTCATATAA<br>CAAACACTGCCACAACAGTGTTGCTTGATGAATTTG<br>GTGTTGGGCCACTTTGCAAAGGTGACAACTTATACT<br>TGTCAGCTGTTGATGTCTGTGGCATGTTTACAAACA<br>GGTCTGGTTCCCAGCAGTGGAGAGGACTCTCCAGAT<br>ATTTTAAGGTGCAGCTAAGGAAAAGGAGGGTTAAA<br>AACCCCTACCCAATTTCTTTCCTTCTTACTGATTTAA<br>TTAACAGAAGGACTCCTAGAGTTGATGGGCAGCCT<br>ATGTATGGCATGGATGCTCAAGTAGAGGAGGTTAG<br>AGTTTTTGAGGGAACAGAGGAGCTTCCAGGGGACC<br>CAGACATGATGAGATACGTTGACAAATATGGACAG<br>TTGCAGACAAAAATGCTGTAATCAAAAGCCTTTATT<br>GTAATATGCAGTACATTTTAATAAAGTATAACCAGC<br>TTTACTTAACAGTTGCAGTTATTTTGGGGGAGGGGT |
| 8 | **MK728937.1**<br><br>**JC polyomavirus isolate NB12 VP1 gene, partial cds** | CAGGAGGAGAAAATGTTCCTCCAGTTCTTCATATAA<br>CAAACACTGCCACAACAGTGTTGCTTGATGAATTTG<br>GTGTTGGGCCACTTTGCAAAGGTGACAACTTATACT<br>TGTCAGCTGTTGATGTCTGTGGCATGTTTACAAACA<br>GGTCTGGTTCCCAGCAGTGGAGAGGACTCTCCAGAT<br>ATTTTAAGGTGCAGCTAAGGAAAAGGAGGGTTAAA<br>AACCCCTACCCAATTTCTTTCCTTCTTACTGATTTAA<br>TTAACAGAAGGACTCCTAGAGTTGATGGGCAGCCT<br>ATGTATGGCATGGATGCTCAAGTAGAGGAGGTTAG<br>AGTTTTTGAGGGAACAGAGGAGCTTCCAGGGGACC<br>CAGACATGATGAGATACGTTGACAAATATGGACAG<br>TTGCAGACAAAAATGCTGTAATCAAAAGCCTTTATT<br>GTAATATGCAGTACATTTTAATAAAGTATAACCAGC<br>TTTACTTAACAGTTGCAGTTATTTTGGGGGAGGGGT |
| 9 | **MK728938.1 JC polyomavirus** | AGGAGAAAATGTTCCTCCAGTTCTTCATATAACAAA<br>CACTGCCACAACAGTGTTGCTTGATGAATTTGGTGT |

| | | |
|---|---|---|
| | **isolate NB13 VP1 gene, partial cds** | TGGGCCACTTTGCAAAGGTGACAACTTATACTTGTC<br>AGCTGTTGATGTCTGTGGCATGTTTACAAACAGGTC<br>TGGTTCCCAGCAGTGGAGAGGACTCTCCAGATATTT<br>TAAGGTGCAGCTAAGGAAAAGGAGGGTTAAAAACC<br>CCTACCCAATTTCTTTCCTTCTTACTGATTTAATTAA<br>CAGAAGGACTCCTAGAGTTGATGGGCAGCCTATGT<br>ATGGCATGGATGCTCAAGTAGAGGAGGTTAGAGTT<br>TTTGAGGGAACAGAGGAGCTTCCAGGGGACCCAGA<br>CATGATGAGATACGTTGACAAATATGGACAGTTGC<br>AGACAAAAATGCTGTAATCAAAAGCCTTTATTGTAA<br>TATGCAGTACATTTTAATAAAGTATAACCAGCTTTA<br>CTTAACAGTTGCAGTTATTTTGGGGGAGGGG |
| 10 | **MK728939.1 JC polyomavirus isolate NB14 VP1 gene, partial cds** | AGGAGAAAATGTTCCTCCAGTTCTTCATATAACAAA<br>CACTGCCACAACAGTGTTGCTTGATGAATTTGGTGT<br>TGGGCCACTTTGCAAAGGTGACAACTTATACTTGTC<br>AGCTGTTGATGTCTGTGGCATGTTTACAAACAGGTC<br>TGGTTCCCAGCAGTGGAGAGGACTCTCCAGATATTT<br>TAAGGTGCAGCTAAGGAAAAGGAGGGTTAAAAACC<br>CCTACCCAATTTCTTTCCTTCTTACTGATTTAATTAA<br>CAGAAGGACTCCTAGAGTTGATGGGCAGCCTATGT<br>ATGGCATGGATGCTCAAGTAGAGGAGGTTAGAGTT<br>TTTGAGGGAACAGAGGAGCTTCCAGGGGACCCAGA<br>CATGATGAGATACGTTGACAAATATGGACAGTTGC<br>AGACAAAAATGCTGTAATCAAAAGCCTTTATTGTAA<br>TATGCAGTACATTTTAATAAAGTATAACCAGCTTTA<br>CTTAACAGTTGCAGTTATTTTGGGGGAGGGG |

| **Table 22: T-Antigen sequences of JCPyV with their accession number** | | |
|---|---|---|
| **S.No.** | **Sequence name and accession number** | **Nucleotide sequence of T-Ag region** |
| 1 | **JC polyomavirus strain NB2 large T antigen gene, partial cds** | CTTTTTTTTCTTCTTAGGTGGGGTAGAGTGCTGGGAT<br>CCTGTGTTTTCATCATCACTGGCAAACATTTCTTCAT<br>GGCAAAACAGGTCTTCATCCCACTTCTCATTAAATG<br>TATTCCACCAGGATTCCCATTCATCTGTTCCATAGGT<br>TGGCACCTAAAAACAAAAAATTAAGTTTATTGTAAA<br>AAACAAAATGCCCTGCAAAAGAAAAATTGTGGTTT<br>ACCTTAAAGCTTTAGATCCCTGTAGGGGGTGTCTCC<br>AAGAACCTTCTCCCAGCAATGAAGAGCTTCTTGGGT<br>TAAGTCACACCCAAACCATTGTCTGAAGCAATCAAA<br>GCAATAGCAATCTATCCACACAAGTGGGCTGCTTCT<br>TAAAAATTTTCTGTTTTTATGCCTTAATTTTAGCATG |

| | | |
|---|---|---|
| | | CACATTAAACAGGGACAATGCACTGAAGGATTAGT<br>GGCACAATTAGGCCATTCCTTGCAATAAAGGGTATC<br>AGAATTAGGAGGAAAATCACAACCAACCTCTGAAC<br>TATTCCATGTACCAAAATCAGGCTGATGAGCAACTT<br>TTACACCTTGTTCCATTTTTTTATATAAAAAATTCAT<br>TCTCTTCATTTTGTCTTCGTCCCCACCTTTATCAGGG<br>TGAAGTTCTTTGCATTTTTTCA |
| 2 | **JX534221.1**<br><br>**JC polyomavirus strain NB6 large T antigen gene, partial cds** | GAATAGGGAGGAATCCATGGAGCTTATGGATTTATT<br>AGGCCTTGATAGGTCTGCATGGGGGAACATTCCTGT<br>CATGAGAAAAGCTTATCTGAAAAAATGCAAAGAAC<br>TTCACCCTGATAAAGGTGGGGACGAAGACAAAATG<br>AATAGAATGAATTTCTTATATAAAAAAATGGAACA<br>AGGTGTAAAAGTTGCTCATCAACCTGATTTTGGTAC<br>ATGGAATAGTTCAGAGGTTGGTTGTGATTTTCCTCC<br>TAATTCTGATACTCTTTATTGCAAGGAATGGCCTAA<br>TTGTGCCACTAATCCTTCAGTGCATTGCCCCTGTTTA<br>ATGTGCCTGCTAAAATTAAGGCATAGAAACAGAAA<br>ATTTTTAAGAAGCAGCCCACTTGTGTGGATAGATTG<br>CTATTGCTTTGATTGTTTCAGACAATGGGTTGGGGG<br>GGTGTTTACCCAAGAAGCTTTTCATTTCTGGGAGAA<br>GGGTTTTGGAGACACCCCCTACAGGGATCTAAAGCT<br>TTAAGGTAAACCCCAATTTTTTTTTTGCAGGGCATTT<br>TGTTTTTTACAATAAACTCAATTTTTTGTTTTTAGGT<br>GCCAACCTATGGAACAGATGAATGGGAGTCCTGGT<br>GGAATACATTTAATGAGAAGTGGGATGAAGACCTG<br>TTTTGCCATGAAGAAATGTTTGCCAGTGATGATGAA<br>AACACAGGATCCCATCACTCTACCCCACCTAAGAAG<br>AAAAAAAAGGTAGAAGACCCTAAAGACTTTCCTGT<br>AGATCTGCATGCATTCC |
| 3 | **MH744486.1**<br><br>**JC polyomavirus isolate NB9 large T antigen and small T antigen genes, partial cds** | CTTTTTTTTCTTTTTAGGTGGGGTAGAGTGTTGGGAT<br>CCTGTGTTTTCATCATCACTGGCAAACATTTCTTCAT<br>GGCAAAACAGGTCTTCATCCCACTTCTCATTAAATG<br>TATTCCACCAGGATTCCCATTCATCTGTTCCATAGGT<br>TGGCACCTAAAAAAAAACAATTAAGTTTATTGTAAA<br>AAACAAAATGCCCTGCAAAAGAAAAATAGTGGTTT<br>ACCTTAAAGCTTTAGATCCCTGTAGGGGGTGTCTCC<br>AAGAACTTTCTCCCAGCAATGAAGAGCTTCTTGGGT<br>TAAGTCACACCCAAACCATTGTCTGAAGCAATCAAA<br>GCAATAGCAATCTATCCACACAAGTGGGCTGCTTCT<br>TAAAAATTTTCTGTTTCTATGCCTTAATTTTAGCATG<br>CACATTAAACAGGGGCAATGCACTGAAGGATTAGT<br>GGCACAGTTAGGCCATTCCTTGCAATAAAGGGTATC<br>AGAATTAGGAGGAAAATCACAACCAACCTCTGAAC<br>TATTCCATGTACCAAAATCAGGCTGATGAGCAACTT |

| | | |
|---|---|---|
| | | TTACACCTTGTTCCATTTTTTTATATAAAAAATTCAT<br>TCTCTTCATCTTGTCTTCGTCCCCACCTTTATCAGGG<br>TGGAGTTCTTTGCATTTTTTCAGATAAGCTTTTCTCA<br>TGACAGGAATGTTCCCCCATGCAGACCTATCAAGGC<br>CTAATAAATCCATAA |
| 4 | **MH746212.1**<br><br>**JC polyomavirus isolate NB10 large T antigen and small T antigen genes, partial cds** | CTTTTTTTTCTTTTTAGGTGGGGTAGAGTGTTGGGAT<br>CCTGTGTTTTCATCATCACTGGCAAACATTTCTTCAT<br>GGCAAAACAGGTCTTCATCCCACTTCTCATTAAATG<br>TATTCCACCAGGATTCCCATTCATCTGTTCCATAGGT<br>TGGCACCTAAAAAAAAACAATTAAGTTTATTGTAAA<br>AAACAAAATGCCCTGCAAAAGAAAAATAGTGGTTT<br>ACCTTAAAGCTTTAGATCCCTGTAGGGGGTGTCTCC<br>AAGAACTTTCTCCCAGCAATGAAGAGCTTCTTGGGT<br>TAAGTCACACCCAAACCATTGTCTGAAGCAATCAAA<br>GCAATAGCAATCTATCCACACAAGTGGGCTGCTTCT<br>TAAAAATTTTCTGTTTCTATGCCTTAATTTTAGCATG<br>CACATTAAACAGGGGCAATGCACTGAAGGATTAGT<br>GGCACAGTTAGGCCATTCCTTGCAATAAAGGGTATC<br>AGAATTAGGAGGAAAATCACAACCAACCTCTGAAC<br>TATTCCATGTACCAAAATCAGGCTGATGAGCAACTT<br>TTACACCTTGTTCCATTTTTTTATATAAAAAATTCAT<br>TCTCTTCATCTTGTCTTCGTCCCCACCTTTATCAGGG<br>TGGAGTTCTTTGCATTTTTTCAGATAAGCTTTTCTCA<br>TGACAGGAATGTTCCCCCATGCAGACCTATCAAGGC<br>CTAATAAATCCATAAGCTCCATGGATTCCTCCC |
| 5 | **MK728941.1 JC polyomavirus isolate NB11 large T antigen and small T antigen genes, partial cds** | CTTTTTTTTCTTTTTAGGTGGGGTAGAGTGTTGGGAT<br>CCTGTGTTTTCATCATCACTGGCAAACATTTCTTCAT<br>GGCAAAACAGGTCTTCATCCCACTTCTCATTAAATG<br>TATTCCACCAGGATTCCCATTCATCTGTTCCATAGGT<br>TGGCACCTAAAAAAAAACAATTAAGTTTATTGTAAA<br>AAACAAAATGCCCTGCAAAAGAAAAATAGTGGTTT<br>ACCTTAAAGCTTTAGATCCCTGTAGGGGGTGTCTCC<br>AAGAACTTTCTCCCAGCAATGAAGAGCTTCTTGGGT<br>TAAGTCACACCCAAACCATTGTCTGAAGCAATCAAA<br>GCAATAGCAATCTATCCACACAAGTGGGCTGCTTCT<br>TAAAAATTTTCTGTTTCTATGCCTTAATTTTAGCATG<br>CACATTAAACAGGGGCAATGCACTGAAGGATTAGT<br>GGCACAGTTAGGCCATTCCTTGCAATAAAGGGTATC<br>AGAATTAGGAGGAAAATCACAACCAACCTCTGAAC<br>TATTCCATGTACCAAAATCAGGCTGATGAGCAACTT<br>TTACACCTTGTTCCATTTTTTTATATAAAAAATTCAT<br>TCTCTTCATCTTGTCTTCGTCCCCACCTTTATCAGGG<br>TGGAGTTCTTTGCATTTTTTCAGATAAGCTTTTCTCA<br>TGACAGGAATGTTCCCCCATGCAGACCTATCAAGGC |

| | | |
|---|---|---|
| | | CTAATAAATCCATAAGCTCCATGGATTCCTCCC |
| 6 | **MK728942.1 JC polyomavirus isolate NB13 large T antigen and small T antigen genes, partial cds** | CTTTTTTTTCTTTTTAGGTGGGGTAGAGTGTTGGGAT<br>CCTGTGTTTTCATCATCACTGGCAAACATTTCTTCAT<br>GGCAAAACAGGTCTTCATCCCACTTCTCATTAAATG<br>TATTCCACCAGGATTCCCATTCATCTGTTCCATAGGT<br>TGGCACCTAAAAAAAAACAATTAAGTTTATTGTAAA<br>AAACAAAATGCCCTGCAAAAGAAAAATAGTGGTTT<br>ACCTTAAAGCTTTAGATCCCTGTAGGGGGTGTCTCC<br>AAGAACTTTCTCCCAGCAATGAAGAGCTTCTTGGGT<br>TAAGTCACACCCAAACCATTGTCTGAAGCAATCAAA<br>GCAATAGCAATCTATCCACACAAGTGGGCTGCTTCT<br>TAAAAATTTTCTGTTTCTATGCCTTAATTTTAGCATG<br>CACATTAAACAGGGGCAATGCACTGAAGGATTAGT<br>GGCACAGTTAGGCCATTCCTTGCAATAAAGGGTATC<br>AGAATTAGGAGGAAAATCACAACCAACCTCTGAAC<br>TATTCCATGTACCAAAATCAGGCTGATGAGCAACTT<br>TTACACCTTGTTCCATTTTTTTATATAAAAAATTCAT<br>TCTCTTCATCTTGTCTTCGTCCCCACCTTTATCAGGG<br>TGGAGTTCTTTGCATTTTTTCAGATAAGCTTTTCTCA<br>TGACAGGAATGTTCCCCCATGCAGACCTATCAAGGC<br>CTAATAAATCCATAAGCTCCATGGATTCCTCCC |
| 7 | **MK728943.1 JC polyomavirus isolate NB14 large T antigen and small T antigen genes, partial cds** | CTTTTTTTTCTTTTTAGGTGGGGTAGAGTGTTGGGAT<br>CCTGTGTTTTCATCATCACTGGCAAACATTTCTTCAT<br>GGCAAAACAGGTCTTCATCCCACTTCTCATTAAATG<br>TATTCCACCAGGATTCCCATTCATCTGTTCCATAGGT<br>TGGCACCTAAAAAAAAACAATTAAGTTTATTGTAAA<br>AAACAAAATGCCCTGCAAAAGAAAAATAGTGGTTT<br>ACCTTAAAGCTTTAGATCCCTGTAGGGGGTGTCTCC<br>AAGAACTTTCTCCCAGCAATGAAGAGCTTCTTGGGT<br>TAAGTCACACCCAAACCATTGTCTGAAGCAATCAAA<br>GCAATAGCAATCTATCCACACAAGTGGGCTGCTTCT<br>TAAAAATTTTCTGTTTCTATGCCTTAATTTTAGCATG<br>CACATTAAACAGGGGCAATGCACTGAAGGATTAGT<br>GGCACAGTTAGGCCATTCCTTGCAATAAAGGGTATC<br>AGAATTAGGAGGAAAATCACAACCAACCTCTGAAC<br>TATTCCATGTACCAAAATCAGGCTGATGAGCAACTT<br>TTACACCTTGTTCCATTTTTTTATATAAAAAATTCAT<br>TCTCTTCATCTTGTCTTCGTCCCCACCTTTATCAGGG<br>TGGAGTTCTTTGCATTTTTTCAGATAAGCTTTTCTCA<br>TGACAGGAATGTTCCCCCATGCAGACCTATCAAGGC<br>CTAATAAATCCATAAGCTCCATGGATTCCTCC |

## 4.4 Pairwise and Multiple sequence alignment of NCCR sequence:

**Table 23** represents the pairwise similarities based on Non coding control regions (NCCRs) of endemic JCPyV isolates NB1, NB2, NB3, NB4, NB5, NB6, NB-PU1, NB-PU2 and NB-PU3 in comparison to archetypal CY, Mad1, LH3, Tai3 and IN8 NCCRs using DiAlign alignment software of Genomatix suite v2.5 GmbH (Cartharius *et al*., 2005). The NCCR region is the most variable region of the viral genome except the proximal part of the sequence which is highly conserved. The NB1, NB2, NB3, NB4 and NB5 isolates were obtained from the Oraon and Munda tribes. NB6 isolate was from the Rabha tribal group and the NB-PU1, NB-PU2 and NB-PU3 were isolated from the pregnant women group. All these sequences were compared to the archetypal sequence JCPyV-CY as well as other types of region-specific NCCR sequences such as LH3 (Tibet), Tai3 (Taiwan), IN8 (India) and Mad1. JCPyV-Mad1 is the prototype, first isolated from a patient with PML (Frisque *et al*., 1984). The naturally occurring variant of the NCCR is termed as archetype sequence which is mostly found in the kidney tissues and in urine and is rarely associated with PML. The archetype JCPyV strain consists of only a single copy of 98 bp tandem repeat in A, C and E blocks with 23 bp (B) and 66 bp (D) sequences present between them. The archetypal NCCR lacks transcription-factor binding sites essential for viral gene expression that include YB-1/Purα and NF-1. The prototype strain has been detected in brain, CSF and blood of PML patients and is characterized by the presence of a 98 bp tandem repeat resulting in duplication of TATA box and transcription factor binding sites (Frisque, 1983).

It was evident from the **Table 23** that the NB-PU3, NB3, NB4 and NB5 were identical. The NB1, NB2, NB3, NB4, NB5, NB-PU1, NB-PU2 and NB-PU3 NCCR sequences, except the NB6 sequence, paired closely with each other (97-100% similar). Pairwise comparison with other strains revealed that the NCCR sequences of NB1, NB2, NB3, NB4, NB5, NB PU1, NB PU2, NB PU3 were highly similar to the Tibetan LH3 (97-98% similar). The NCCR sequence of the NB6 isolate seemed almost identical to the archetype strain CY (98% similarity). The NCCR sequences were divergent from that of Tai3, IN8 and Mad1 control regions. All the endemic NCCR sequences have an archetype-type of NCCR architecture with few mutations and deletions in the sequence.

The 98 bp tandem repeat present in NCCR of Mad1 strain i.e., prototype strain was not found in any of the strains studied.

**Table 23: Pairwise similarities in percentage (relative to the maximum similarity) based on Non-Coding Control Regions (NCCRs) of endemic JCPyV strains NB1, NB2, NB3, NB4, NB5, NB6, NB-PU1, NB-PU2, NB-PU3 with archetype CY, Mad1, LH3, Tai3 and IN8 NCCRs using DiAlign alignment software of Genomatix suite v2.5 GmbH.**

| | **NB-PU2 KF739299** | **NB-PU3 MK904563** | **NB4 JX534218** | **NB2 JX534216** | **NB1 JX294575** | **NB3 JX534217** | **NB5 JX534219** | **NB6 JX534220** | **CY AB038249** | **LH3 AB262411** | **Tai3 U61771** | **IN8 AB126992** | **Mad1NC_001699** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **NB-PU1 KF739298** | 98 | 99 | 99 | 98 | 99 | 99 | 99 | 94 | 95 | 97 | 94 | 96 | 78 |
| **NB-PU2 KF739299** | | 98 | 98 | 97 | 97 | 98 | 98 | 95 | 96 | 97 | 93 | 95 | 76 |
| **NB-PU3 MK904563** | | | **100** | 99 | 98 | **100** | **100** | 96 | 97 | 98 | 94 | 97 | 78 |
| **NB4 JX534218)** | | | | 99 | 98 | **100** | **100** | 96 | 97 | 98 | 94 | 97 | 78 |
| **NB2 JX534216** | | | | | 98 | 99 | 99 | 96 | 97 | 98 | 93 | 97 | 80 |
| **NB1 JX294575** | | | | | | 98 | 98 | 94 | 95 | 97 | 94 | 95 | 79 |
| **NB3 JX534217** | | | | | | | **100** | 96 | 97 | 98 | 94 | 97 | 78 |
| **NB5 JX534219** | | | | | | | | 96 | 97 | 98 | 94 | 97 | 78 |
| **NB6 JX534220** | | | | | | | | | 98 | 95 | 97 | 97 | 74 |
| **CY AB038249** | | | | | | | | | | 97 | 98 | 99 | 76 |
| **LH3 AB262411** | | | | | | | | | | | 98 | 97 | 79 |
| **Tai3 U61771** | | | | | | | | | | | | 98 | 74 |
| **IN8 AB126992** | | | | | | | | | | | | | 79 |

A multiple sequence alignment of endemic JCPyV NCCRs and other previously reported strains is shown in **Fig. 12**. The NCCR rearrangements of these strains were of archetype type strain CY with A-B-C-D-E-F blocks. A 10 nucleotide (169-178) deletion in block B and one di-nucleotide (454-455) deletion in block F were observed in NB-PU1, NB-PU2, NB-PU3, NB1, NB2, NB3, NB4 and NB5 isolates. Point mutations in seven different sites within the NCCR of the endemic strains recorded are at 4, 13, 26, 27, 69, 226 and 452 nucleotide positions of the sequence alignment when compared with the archetype strain CY (**Fig. 12**). Isolate NB1 had four-point alterations at nucleotide 4 (T → A), nucleotide 13 (T → C), nucleotide 226 present at block C (A → C) and nucleotide 452 present at block F (G → A). NB2, NB3, NB4 and NB5 had three-point alterations at nucleotides 26 (T → A), 226 (A → C), 452 (G → A). NB6 had two-point alterations at nucleotides 26 (T → A) and 69 (A → G). NB-PU1 had four alterations at nucleotide positions 4 (T → A), 26 (T → C), 226 (A → C) and 452 (G → A). And NB-PU2 had five-point alterations at nucleotide 4 (T → A), 26 (T → C), 27 (G → T), 226 (A → C) and 452 (G → A).

Several transcription factors are implicated in the regulation of JCPyV gene expression which include NF-κB (Ranganathan and Khalili, 1993), NFAT4 (Wollebo *et al*., 2012), upstream Target or up-TAR (Chowdhury *et al*., 1993), Tst-1 (Wegner *et al*., 1993), Sp-1 (Henson *et al*., 1992), Spi-B (Marshall *et al*., 2010), GBP-i (Raj and Khalili, 1994), Y-box binding protein 1 (YB-1) and Purα (Chen and Khalili, 1995), NF-1 (Amemiya *et al*. 1989), CREB/ATF-1 (Lonze and Guinty, 2002), Activator Protein 1 (AP-1) family members (Sadowska *et al*., 2003), p53 (Ariza *et al*., 1994), Egr-1 (Romagnoli *et al*., 2008), BAG-1 (Devireddy *et al*., 2000) and C/EBPβ (Romagnoli *et al*., 2009). These transcription factors are proposed to bind to JCPyV NCCRs either individually or in cooperation with other cellular factors and/or with proteins encoded by the virus. The binding motifs of NF-κB, NFAT4, upstream Target or up-TAR, Tst-1, Sp-1, Sp1-β, GBp-1, Y-box binding protein 1, Purα, Nuclear factor 1, NF-1, CREB/ATF-1, Activator Protein 1(AP-1) family members, p53, early growth response-1 protein or Egr-1, Bcl-2- associated athano gene-1 or BAG-1 and CAAT/enhancer binding protein beta or C/EBPβ are shown in the aligned sequences (**Fig. 12**)

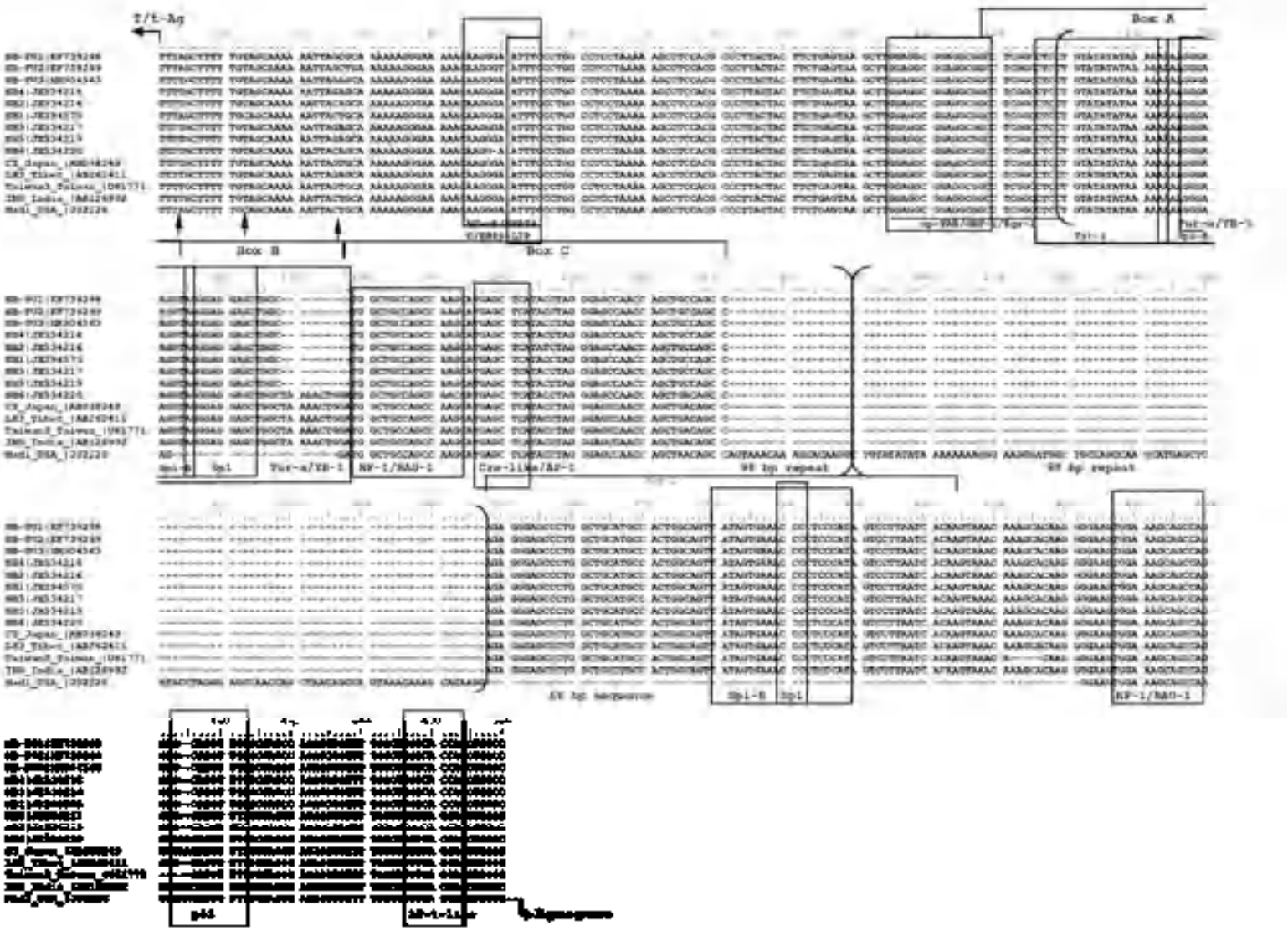


**Figure 12**: Non-Coding Control Regions (NCCRs) of JCPyV endemic strains NB1, NB2, NB3, NB4, NB5, NB6, NB-PU1, NB-PU2 and NB-PU3 were aligned against the CY, LH3, Tai3, IN-8 and Mad-1 strains, showing the common binding sites of known and reported transcription factors mentioned in square blocks. The 98 bp repeats of Mad-1 and 66 bp sequences of CY archetype strains are shown by large brackets and dotted arrows respectively. The NCCR is divided into six regions according to CY strain: Box A (36 bp), B (23 bp), C (55 bp), D (66 bp), E (18 bp) and F (69 bp). Nucleotide deletion (s) are marked by a dash (–) and nucleotide variation sites among endemic strains are marked with vertical arrows.

In most cases, binding of host transcription factors activates viral gene expression, however, repression of viral gene expression has been reported for AP-1 (Kim *et al*., 2003; Ravichandran *et al*., 2006), C/EBPβ (Romagnoli *et al*., 2009), and NF-1A (Ravichandran and Major, 2008). NF-κB is inducible by a wide variety of extracellular stimuli, like phorbol esters and cytokines. Since it is stimulated by extracellular cytokines, such as tumour necrosis factor alpha (TNF-α), it may have a role in determining latency of JCPyV and reactivation of it in pro-inflammatory situations (Ranganathan and Khalili, 1993; Mayreddy *et al*., 1996). The upTAR and GGA/C-rich sequences (GRS), present at the origin of replication has been shown to respond to HIV-1 Tat (Tada *et al.*, 1990; Chowdhury *et al*., 1993) and to GRS binding protein (GBPi) (Raj and Khalili, 1994) respectively. It has also been suggested that this sequence element might also bind Sp1 (Raj and Khalili, 1995). However, in a later study, it was demonstrated in a series of gel shift experiments that Sp1 does not bind to GRS (Romagnoli *et al.*, 2008). The Tat has been found to form a complex with the cellular transcription factor Purα to activate transcription through the GRS/upTAR element (Krachmarov *et al*., 1996). This interaction of Purα/Tat at the GRS/upTAR element can also affect T-antigen binding and enhance viral replication (Daniel *et al*., 2001). Since the majority of cases of PML occur in HIV-1 positive individuals, Tat-mediated upregulation of viral transcription and replication may be an important mechanism for JCPyV reactivation.

Studies also revealed the identity of GBPi to the early growth response protein 1 (Egr-1) (Romagnoli *et al.*, 2008). Egr-1 is synthesized rapidly in cells treated with cytokines or phorbol myristate acetate that binds to the GRS element and activates JCPyV late transcription. Egr-1 is upregulated during JCPyV infection of cultured astrocytes and in immunohistochemistry of the oligodendrocyte inclusion bodies and bizarre astrocytes of patients with PML, suggesting that Egr-1 induction may be important in JCPyV and PML pathogenesis (Romagnoli *et al*., 2008).

The late proximal NCCR of the Mad-1 strain contains a 98-bp tandem repeat that functions as an enhancer for transcription and is also responsible for the glial-cell tropism. Thus, reporter constructs containing the JCPyV 98-bp repeat are expressed

well in primary glial cells but not in HeLa cells or CV-1 cells (Kenney *et al*., 1984). Studies on Tst-1, AP-1, NFAT4, C/EBPβ, Purα, and YB-1 suggest a role of these proteins in glial cells. NF-1X and Spi-B are reported to have elevated protein expression in all cell types susceptible to JCPyV infection (Sumner *et al*., 1996; Shinohara *et al*., 1997; Marshall *et al*., 2009, Major, 2010).

Two prominent features within the endemic JCPyV NCCR Box B are the absence, unlike JCPyV Mad-1, of a full-length Purα/YB-1 binding pentanucleotide motif (5'-AGGGAAGGGA-3') (Chen and Khalili, 1995) and the presence of Sp1 binding site (GA Box) (5'-AGGGAGGAGC-3') (Henson *et al*., 1992) in the same region. These features can also be seen in archetypal CY, Tibetan LH3, Taiwanese Tai3 and the Central-North Indian IN-8 strains (**Fig. 12**). Sp1 has been shown to promote early gene transcription in both glial and non-glial cells (Henson, 1994) and TAg-mediated transactivation of viral late genes. Therefore, the presence of Sp1 binding site in the endemic strain may enable replication of the virus in both glial and non glial cells. In early stages of infection, Pur-α binds to the NCCR and stimulates early viral gene transcription. The LTAg binds to the ORI only after it is present in sufficient amount and recruits the cellular DNA polymerase. The T antigen/YB-1 interaction with the NCCR leads to the dissociation of Pur-α dissociation switching late gene transcription (Chen and Khalili, 1995; Chen *et al*., 1995). This interaction is considered as a switch system to regulate early and late viral gene transcription. The Pur-α/YB-1 binding is absent in endemic strains of this region like that of archetype strain CY.

Duplications and deletions in rearranged NCCR sequence of Mad-1 strain generate more transcription factor binding sites that confer advantages to JCPyV (Frisque *et al*., 1984). The archetype D region may be inhibitory to JCPyV growth in some cells and its deletion allows productive infection in cells (Gosert *et al*., 2010). The repeats of C region in Mad-1 NCCR consist of additional number of NF1 binding sites that are crucial to activate viral gene transcription in the brain and in lymphoid tissues (Amemiya *et al*., 1989; Monaco *et al*., 2001). These additional number of NF1 binding sites are absent in our strain like that of archetype strain CY. This deletion of

NF1 binding sites in the endemic strain may have effect on its ability to replicate in brain and lymphoid tissues.

We have found a di-nucleotide deletion (nucleotide position: 454-455) within this p53 binding site of endemic JCPyV strains from the Oraon/Munda group as well as NCCR sequences of pregnant women but the deletion was not present in NB6 NCCR which was isolated from the Rabha tribal group (**Fig. 12**). p53 gene is a tumour suppressor gene. Wild-type p53 have been found to directly arrest growth of proliferating cells at the G1–S boundary (Diller *et al.,* 1990; Martinez *et al.,* 1991). Inactivation of this gene plays a critical role in malignant transformation. Murine and human wild-type p53 binds to the SV40 LTAg that acts as helicase and also binds to the cellular DNA polymerase α (Gannon and Lane, 1987). Wild-type p53 inhibits initiation and DNA replication by preventing DNA unwinding by the LTAg helicase (Wang *et al.,* 1989). Cellular tumour suppressor protein p53 is reported to bind to JCPyV large T-antigen to repress viral replication (Staib *et al*., 1996). This deletion in the p53 binding site within the NCCR is presently not known, however this binding may have implications with respect to JCPyV DNA replication. However, the implications of these deletions need to be evaluated by *in vivo* studies.

## 4.5 Transcription factor binding analyses of NCCR sequences:

Putative transcription factor binding sites (TFBSs) on the non-coding control region was searched *in silico* using families of general core and vertebrate transcription factor matrix using the MatInspector program of the Genomatix software suite (Catharius *et al.*, 2005). The predicted TFBS-depicted putative binding sites for transcription factors corresponding to thirty-three matrix families were searched. The **Table 24** shows the putative binding sites of transcription factors in an abridged form. Matches for transcription factors that are reported to be active in cells/tissues such as antibody-producing cells, antigen-presenting cells, bone marrow cells, hematopoetic and immune system, leukocytes, lymphocytes, monocytes, myeloid cells, phagocytes, brain, CNS, endocrine system, kidney, nervouss system, urinogenital system are retained in the table.

The POU domain factors, EGR/nerve growth factor, human and murine ETS1 factors, fork head domain factors, Krueppel-like C2H2 zinc finger factors, homeodomain transcription factors, Nuclear factor 1, Nuclear factor κB, pleomorphic adenoma gene, Purα, SWI/SNF related nucleophosphoproteins, Spalt-like transcription factors, SP1 transcription factors are common to all the sequences (**Table 24**). Although the TFBS were common on a closer look, transcription factors such as Neuron-specific Olfactory factor or NOLF, TALE homeodomain class recognizing TG motifs or TALE, X-box binding factor or XBBF and Activator/repressor binding to transcription initiation site or YY1F were found to match exclusively to either positive (+) or negative (-) strands of NB6 NCCR owing to its sequence differences with that of other endemic NCCRs.

The POU is a family of proteins and have a role in function of the neuroendocrine system (Assa-Munt *et al*., 1993) and development of an organism (Andersen and Rosenfeld, 2001). Tst-1, a member of the POU-domain family, also referred to as SCIP or Oct-6 has been detected in myelinating glial cells. In adults, it is mainly expressed in the myelin-producing Schwann cells of the peripheral nervous system and the oligodendrocytes of the central nervous system (He *et al*., 1989; Monuki *et al*., 1991). Both oligodendrocytes and Schwann cells are permissive for infection by JCPyV in culture (Assouline and Major, 1991). Studies have shown that Tst-1 stimulates expression of small and large tumor antigen and also stimulates expression of the late viral capsid proteins (Wegner *et al*., 1993). The early growth response-1 protein (Egr-1) is a zinc fnger transcription factor. Origin of DNA replication contains a region with the GG(A/C) sequence trinucleotide repeats known as GG(A/C)-rich sequence (GRS). This region consists of potential binding sites for transcription factors such as Sp1 and Egr-1. Egr-1 has been found to stimulate transcription of JCPyV late promoter. Romagnoli and co-workers showed that Egr-1 is induced by JCpyV infection of primary astrocytes in culture. Mutated Egr-1 site showed reduced VP1 expression and DNA replication (Romagnoli *et al*., 2008). NF-1 is a family of transcription factors that contains four members: A, B, C, and X. NF-1A is expressed in several JCPyV non-permissive cell types and has been shown to

decrease viral late protein expression (Ravichandran and Major, 2008), while NF-1X increases viral gene expression and is highly expressed in JCPyV-permissive cells (Monaco *et al*., 2001). NF-1X is overexpressed in the brain where it binds to the NCCR region of JCPyV and affects both early and late viral transcription (Shinohara *et al*., 1997). Transcription factor binding sites that are responsible for replication of JCPyV in glial cells are present in all the endemic JCPyV strains (NB1, NB2, NB3, NB4, NB5, NB6, NB-PU1, NB-PU2 and NB-PU3) of this region (**Table 24**). Presence of this transcription factor binding sites may have a role in infection of endemic JCPyV in glial cells.

Several other transcription factor-binding sites have been found in the NCCRs of endemic JCPyV strains of this region (**Table 24**). But effect of these transcription factor binding sites on viral replication and tissue tropism is not known. Transcription factor such as ETS factors, Ikaros zinc finger family, homeodomain transcription factors, Myc associated zinc fingers, Krueppel-like C2H2 zinc finger factors are present in all of the endemic strains of this region. All these transcription factor binding sites are active in leucocytes, lymphocytes, hematopoetic system etc (**Table 24**). ETS is expressed at high levels mainly in immune tissues such as thymus, spleen, and lymph node. The expression of Ets1 blocks differentiation of B- and T-cells. The Krüppel-like family of transcription factors is a set of eukaryotic C2H2 zinc-finger DNA-binding proteins that regulate gene expression. The homeodomain protein products share a characteristic protein-fold structure that binds DNA to regulate expression of target genes (Gehring, 1992; Gehring, 1993). They regulate gene expression and cell differentiation during early embryonic development. The Ikaros family zinc finger protein displays crucial functions in the hematopoietic system and is a known regulator of immune cells development, mainly in early B cells, CD4+ T cells. It has been found to be a major tumour suppressor involved in human B-cell acute lymphoblastic leukemia (Kastner and Chan, 2011). Interferon regulatory factors (IRF) are proteins which regulate transcription of interferons and are used in the JAK-STAT signaling pathway (Paun and Pitha, 2007; Ikezu *et al*., 2008). Myc is a family of regulator genes and

proto-oncogenes. The bHLH transcription factor is a member of the myc family and encodes a nuclear phosphoprotein that plays a role in cell cycle progression, apoptosis and cellular transformation. The Myb genes are part of a large gene-family of transcription factors and may have a role in cell cycle regulation. NeuroD, also called Beta2, is a basic helix loop helix transcription factor expressed in certain parts of brain, beta pancreatic cells and enteroendocrine cells is involved in the differentiation of nervous system and development of pancreas. Hence, the presence of all these transcription factors in the endemic JCPyV strains may have a role replication of the virus in blood tissues. A high viral load has been detected in blood samples of individuals of this region. The presence of these sites in endemic JCPyV strains and its role in viral replication and infection needs further investigation in specific cell culture experiments.

Several transcription factor binding sites have been predicted to be present in the NCCR region of endemic JCPyV strains. The potential roles of some of the transcription factors listed in **Table 24** such as NF-κB (Ranganathan and Khalili, 1993), Sp-1 (Henson *et al*., 1992), Purα (Chen and Khalili, 1995), Nuclear factor 1 or NF-1 (Amemiya *et al*. 1989), p53 (Ariza *et al*., 1994), Early growth response-1 protein or Egr-1 (Romagnoli *et al*., 2008) have already been studied by several workers in the past. All these transcription factor binding sites have a role in viral replication and tissue tropism. Further investigation is needed to validate the effect of all these transcription factors on endemic JCPyV replication listed in the **Table 24.**

**Table 24:** P**redicted transcription factor binding sites (TFBS) in the NCCRs of endemic JCPyVstrains NB1, NB2, NB3, NB4, NB5,NB6, NB-PU1,NB-PU2, NB-PU3 derived using MatInspector Release Professional 8.0.5, March 2011 of Genomatix Software suite v2.5 GmbH. Selected TFBS matches are shown as alphabetically arranged vertebrate matrix families. TFBS search involved both general core promoters (0.75/Optimized) and vertebrate (0.75/Optimized) promoter element groups of MatInspector matrix family library version 8.4.**

| JCPyV strains (NCCRs) | Matrix Family | Detailed Family Information | Tissue | Optimized Similarity Scores | TF Binding Sequence (Core Sequence in capital) |
|---|---|---|---|---|---|
| NB1,NB2,NB3,NB4,NB5, NB6, NB-PU1, NB-PU2, NB-PU3 | O$VTBP | Vertebrate TATA binding protein factor | Ubiquitous | 0.90 | 131-gtataTATAaaaaaaag-147 |
| NB1,NB2,NB3,NB4,NB5, NB6, NB-PU1, NB-PU2, NB-PU3 | V$AP4R | AP4 and related proteins | Ubiquitous | 0.92 | 214-gctggcAGCTggttggc-230 (-) and 486-tggccAGCTggtgaca-500 (-) |
| NB1,NB2,NB3,NB4,NB5, NB6, NB-PU1, NB-PU2, NB-PU3 | V$BRNF | Brn POU domain factors | Brain, CNS, Endocrine System, Neuroglia, Neurons | 0.89 | 14-tttgctgTAATttttgct-32 (-) |
| NB1,NB2,NB3,NB4,NB5, NB6, NB-PU1, NB-PU2, NB-PU3 | V$CAAT | CCAAT binding factors | Ubiquitous | 0.81 | 185-ccagCCAAgcatgag-199 and 211-ggagCCAAccagctg-225 |
| NB1,NB2,NB3,N | V$E2FF | E2F-myc activator/cell | Ubiquitous | 0.84-0.85 | 458-ctctgGCTCgcaa |

| | | | | | |
|---|---|---|---|---|---|
| B4,NB5, NB6, NB-PU1, NB-PU2, NB-PU3 | | cycle regulator | | | aaca-474 (-) and<br><br>456-ctggctcgcAAAAcatg-472 (-) |
| NB1,NB2,NB3,NB4,NB5, NB6, NB-PU1, NB-PU2, NB-PU3 | V$EGRF | EGR/nerve growth factor induced protein C and related factors | Brain, CNS, Endocrine System Kidney, Nervous System, Urinogenital System | 0.88 | 105-ggaggcggAGGCggcct-121 and 387-gactatGGGAggggttt-403 (-) |
| NB1,NB2,NB3,NB4,NB5, NB6, NB-PU1, NB-PU2, NB-PU3 | V$ETSF | Human and murine ETS1 factors | Hematopoietic and Immune System, Leukocytes, Lymphocytes, Monocytes | 0.88-0.96 | 424-gcacaaggGGAAgtggaaagc-444<br>29-caaaaaagGGAAaaacaaggg-49<br>140-aaaaaaagGGAAggtagggag-160 |
| NB1,NB2,NB3,NB4,NB5, NB6, NB-PU1, NB-PU2, NB-PU3 | V$FKHD | Fork head domain factors | APCs, Blood Cells, Immune System, Leukocytes, Lymphocytes | 0.89 | 132-tatataTAAAaaaaagg-148 and<br><br>410-cacaagTAAAcaaagca-426 |
| NB2 | V$GATA | GATA binding factors | Blood and Bone Marrow Cells Hematopoietic System, Immune System, Leukocytes, Lymphocyte | 0.90 | 199-cctaGATAtgagc-211 |

| | | | | | |
|---|---|---|---|---|---|
| | | | s | | |
| NB1,NB2,NB3,NB4,NB5, NB6, NB-PU1, NB-PU2, NB-PU3 | V$HICF | Krueppel-like C2H2 zinc finger factors hypermethylated in cancer | Erythropoiesis, control of cell proliferation, monocyte activation | 0.88 | 180-ggcTGCCagccaa-192 and 221-agcTGCCagccag-349 |
| NB1,NB2,NB3,NB4,NB5, NB6, NB-PU1, NB-PU2, NB-PU3 | V$HOMF | Homeodomain transcription factors | Blood and Bone Marrow Cells, Endocrine System, Hematopoietic System, Immune System, Leukocytes, Lymphocytes | 0.88-0.95 | 13-cagcaaaaAATTactgcaa-31 and 14-tttgcagtAATTtttgct-32 (-) |
| NB1,NB2,NB3,NB4,NB5, NB-PU1, NB-PU2, NB-PU3 | V$IKRS | Ikaros zinc finger family | Antibody-Producing Cells, Blood Cells, Hematopoietic System, Immune System, Leukocytes, Lymphocytes | 0.84 | 43-acaagGGAAtttc-55 |
| NB1,2,3,6 | V$IRFF | Interferon regulatory factors | Antibody-Producing Cells, APCs, Blood, Bone Marrow Cells, Hematopoie | 0.85-0.87 | 376-cagttatagtGAAAcccctcc-396; 429-agggGAAGtggaaagcagcca-449; and 28- |

| | | | | | |
|---|---|---|---|---|---|
| | | | tic and Immune System, Leukocytes, Lymphocytes, Monocytes Myeloid Cells, Phagocytes | | gcaaaaaaggGAAAaacaagg-48 |
| NB1,2,3,6 | V$MAZF | Myc associated zinc fingers | Blood Cells, Immune System, Leukocytes | 0.90 | 387-atggGAGGggttt-399 (-) |
| NB1,2,3,6 | V$MYBL | Cellular and viral myb-like transcriptional regulators | Blood and Bone Marrow Cells, Hematopoietic and Immune System, Leukocytes Lymphocytes | 0.96 | 370-ctaTAACtgccagtg-384 (-) |
| NB1,2,3,6 | V$MZF1 | Myeloid zinc finger 1 factors | Blood and Bone Marrow Cells, Hematopoietic and Immune System, Leukocytes Myeloid Cells | 0.99 | 428-aaGGGGaagtg-438 |
| NB1,NB2,NB3,NB4,NB5, | V$NEUR | NeuroD, Beta2, HLH domain | Antibody-Producing Cells, Blood | 0.95 | 216-caaccaGCTGcca-228; |

| | | | | | |
|---|---|---|---|---|---|
| NB6, NB-PU1, NB-PU2, NB-PU3 | | | and Bone Marrow Cells, Brain, CNS, Hematopoietic System, Immune System, Leukocytes, Lymphocytes, Nervous System, Neuroglia, Neurons | | 488-tcacCAGCtggcc-500 and 489-ggcCAGCtggtg-500 (-) |
| NB1,NB2,NB3,NB4,NB5, NB6, NB-PU1, NB-PU2, NB-PU3 | V$NF1F | Nuclear factor 1 | Brain, Central Nervous System, Digestive System, Liver, Nervous System | 0.81-0.92 | 161-gagctggctggctGCCAgcca-191 (+); 212-gagccaaccagctGCCAgcca-348; 433-ctcCTGGctgctttccacttc-453 (-); 477-gttTTGGcttgtcaccagctg-497; 161-tggcTGGCagccagccagctc-191 (-); 165-tggcTGGCtgccagccaagca-195 (+); 433-gaagTGGAaagcagccaggag-453; 171-aaactggatggctG |

| | | | | | |
|---|---|---|---|---|---|
| | | | | | CCAgcca-191 (NB6); 171-tggCTGGcagccatccagttt-191 (-) (NB6); 175-tggTTGGctggcagccatcca-195 (-) (NB6); 175-tggaTGGCtgccagccaacca-195 (NB6); 355-gcccTGGCtgcatgccactgg-375; 355-ccagTGGCatgcagccagggc-375 (-); 477-cagcTGGTgacaagccaaaac-497 (-) |
| NB1,NB2,NB3,NB4,NB5, NB6, NB-PU1, NB-PU2, NB-PU3 | V$NFKB | Nuclear factor kappa B/c-rel | Blood and Bone Marrow Cells, Hematopoietic and Immune System, Leukocytes, Myeloid Cells, Phagocytes | 0.87 | 45-aagggaatTTCCctg-59 and 45-cagggaaaTTCCctt-59 (-) |
| NB6 | V$NOLF | Neuron-specific olfactory | Antibody-Producing Cells, | 0.88 | 460-acatgtTCCCctggctgctttcc-438 |

| | | | | | |
|---|---|---|---|---|---|
| | | factor | Blood, Bone Marrow Cells, Hematopoietic, Immune, Nervous System, Leukocytes, Lymphocytes, Neurons | | (-) |
| NB2,3,6 | V$P53F | p53 tumour suppressor | Ubiquitous | 0.92 | 441-aagcagccagggaaCATGtttt-463 (NB6) and 452-ctctggctcgcaaaaCATGttcc-474 (-) (NB6) |
| NB6 | V$PAX3 | PAX-3 binding sites | Embryonic Structures Muscle, Skeletal Muscles | 0.93 | 187-gagctCATGgttggctggc-202 (-) |
| NB6 | V$PAX5 | PAX-2/5/8 binding sites | Antibody-Producing Cells, Blood Cells, Endocrine System, Hematopoietic System, Immune System, Kidney, Leukocytes, Lymphocytes | 0.79 | 177-tatgagCTCAtggttggctggcagccatc-205 (-) |
| NB2,6 | V$PAXH | PAX homeodomai | Brain, CNS, Endocrine | 0.99 | 16-caaaaaATTAca |

| | | | | | |
|---|---|---|---|---|---|
| | | n binding sites | System, Nervous System Neurons | | gca-30 |
| NB1,NB 2,NB3,N B4,NB5, NB6, NB-PU1, NB-PU2, NB-PU3 | V$PLAG | Pleomorphic adenoma gene | Brain, CNS, Nervous System | 0.87 | 223-ccaggGCTCcct ctggctggcag-361 and 348-agaggGAGCcct ggctgcatgcc-370 |
| NB1,NB 2,NB3,N B4,NB5, NB6, NB-PU1, NB-PU2, NB-PU3 | V$PURA | Pur-alpha binds both single-stranded and double-stranded DNA in a sequence-specific manner | Brain, CNS, Nervous System, Neuroglia, Neurons | 0.97 | 105-ggAGGCggagg cg-117 |
| NB1,NB 2,NB3,N B4,NB5, NB6, NB-PU1, NB-PU2, NB-PU3 | V$RUSH | SWI/SNF related nucleophosp hoproteins with a RING finger DNA binding motif | Ubiquitous | 0.98 | 410-gtttACTTgtg-420 (-) |
| NB1,NB 2,NB3,N B4,NB5, NB6, NB-PU1, NB-PU2, NB-PU3 | V$SALL | Spalt-like transcription factors | Embryonic Structures, Kidney, Urogenital System | 0.96 | 135-atATAAaaaaaa g-147 |
| NB1,NB 2,NB3,N B4,NB5, NB6, | V$SP1F | GC-Box factors SP1/GC | Ubiquitous | 0.88 | 151-aggtagGGAGga gctgg-167 |

| | | | | | |
|---|---|---|---|---|---|
| NB-PU1, NB-PU2, NB-PU3 | | | | | 386-actatgGGAGgggtttc-402 (-) |
| NB6 | V$TALE | TALE homeodomain class recognizing TG motifs | Bone Marrow Cells, CNS, Embryonic Structures, Hematopoietic System, Myeloid Cells, Nervous System, Neurons | 0.95 | 219-ctctggcTGTCagctgg-351 (-)<br>217-aaccagctGACAgccag-349 |
| NB6 | V$XBBF | X-box binding factors | Antibody-Producing Cells, Blood Cells, Immune System, Leukocytes, Lymphocytes | 0.90 | 441-aagcagccaggGGAAcatg-459 |
| NB6 | V$YY1F | Activator/repressor binding to transcription initiation site | Embryonic Structures | 0.82 | 166-ggcagCCATccagttttagcc-186 (-) |
| NB1,2,3,6 | V$ZF02 | C2H2 zinc finger transcription factors 2 | Blood Cells, Immune System Leukocytes, Lymphocytes | 0.87 | 107-ggaggccgaggCCGCctccgcct-129 (-) |
| Start and end positions of TF binding sites are numbered according to the aligned NCCR sequences as shown in Figure 12. Matching sequences found in negative strands are indicated as (-); however, sequences are numbered along the plus strands. | | | | | |

## 4.6 Pairwise and Multiple sequence alignment of VP1:

Pairwise sequence comparison based on VP1 sequences of JCPyV isolates of this region with different JCPyV types which included Type 1A, Type 1B, Type 2A, Type 2B, Type 2C, Type 2D, Type 3, Type 4 and Type 6 were done using DiAlign alignment software of Genomatix suite v2.5 GmbH (**Table 25**).

**Reference sequences**: Genbank accession numbers are given in brackets. JCPyV Type 1A strains: Mad1 [J0227] and #124 [AF015526]. JCPyV Type 1B strain: #123 [AF015527]. JCPyV Type 2A/2C strains: #224 [AF015529], #225 [AF015530], #226 [AF015531], #228 [AF015534], #229 [AF015535] and Tokyo-1 [AF030085]. JCPyV Type 2B strains: #223 [AF015532] and #227 [AF015533]. JCPyV Type 2D strain: #230 [AF015536]. JCPyV Type 3 strains: #308 [U73500], #311 [U73501] and #312 [U73502]. JCPyV Type 1 strains: #123 [AF015527] and #124 [AF015526]. JCPyV Type 4 strain: #402 [AF015528]. JCPyV Type 6 strain: #601 [AF015537]. JCPyV Type 7 strain: Taiwan-3 [U61771]. JCPyV archetype strain CY [M35834], JCPyV MY strain [AB038250] and JCPyV Tky-2a strain [AB038255].

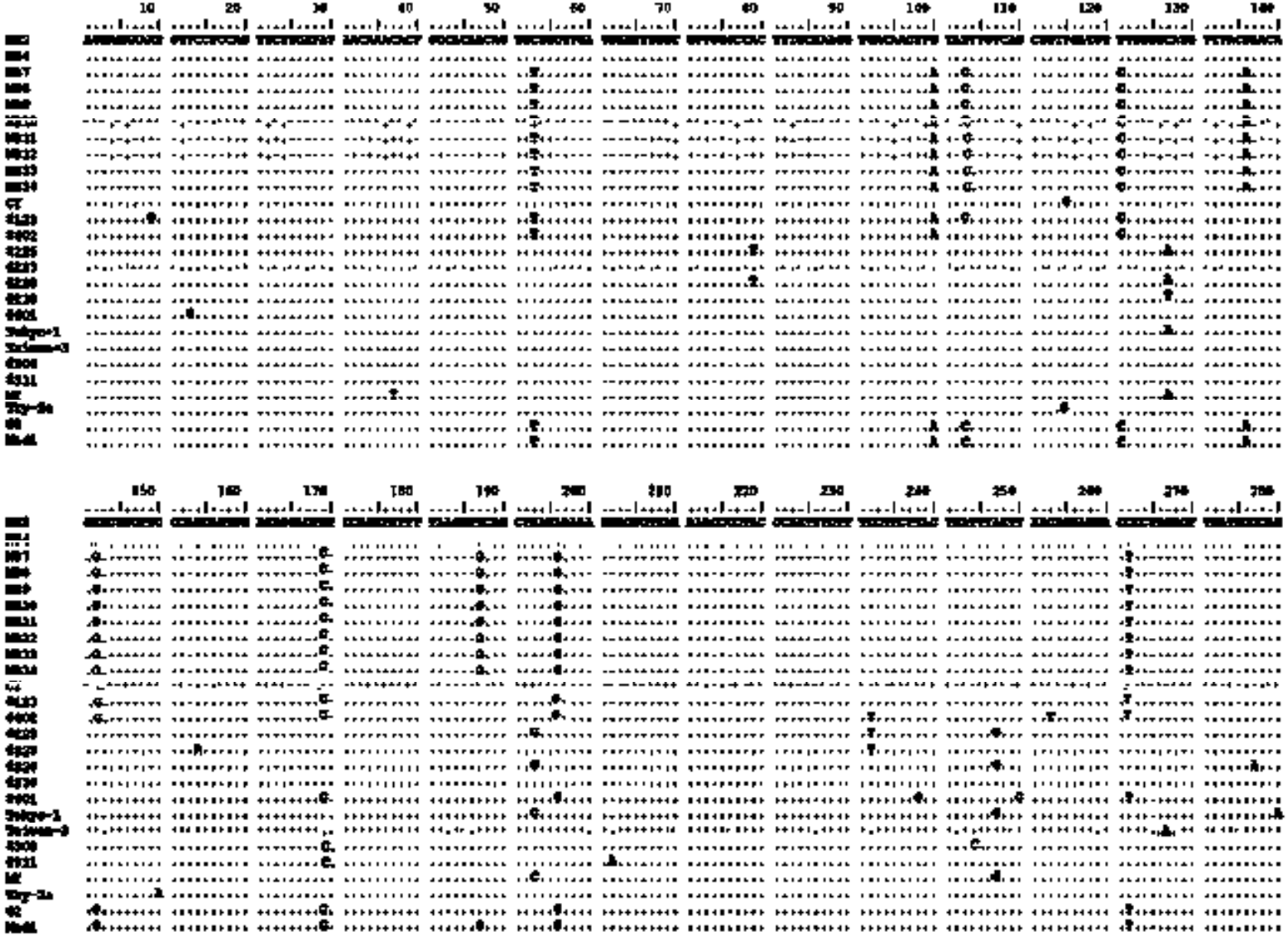
From the comparison between the VP1 sequences, the endemic JCPyV isolates from the Rabha tribes, NB7, NB8, NB9, NB10, NB11, NB12, NB13 and NB14 were found to be identical. Upon comparison with other JCPyV types, these isolates were found to be most similar to the Type 1B (Europe) (98.97% similarity). JCPyV Type 1 is commonly found in Europeans and European-Americans. But it has also been detetected in some Asian groups such as in Japan (Kitamura *et al*., 1998) and South Korea (Jeong *et al*., 2002). Type 1 exists as a minor group in both of Korean and Japanese population (Kitamura *et al*., 1998; Jeong *et al*., 2004). Presence of Type 1 in Asian population suggests the possibility of transmitting of this genotype from Europeans or European-Americans (Jeong *et al*., 2004). NB2 and NB4 VP1 sequences from Oraon/Munda tribes appeared almost identical to the Type 2D sequence (India) (99.59% similar). Type 2D is the Indian subtype and is mostly found in Asians and South Asians. The JCPyV sequences amplified from the Rabha tribe isolates differed from the JCPyV strains of Oraon/Munda tribes by 4.12%.

The majority of JCPyV strains analyzed do not differ in length from the prototype strain JCPyV (Mad1) within the overall coding region. Multiple sequence alignment of endemic JCPyV VP1 and other previously reported strains revealed point mutations in twenty different sites within the sequence. Several point mutations were recorded in the VP1 region of endemic strains of this region (**Fig.13** and **Table 26**). The sequences obtained from Rabha tribes differed from Type 1B strain (#123) at three nucleotide positions. Point mutations were recorded at 136, 187 and 378 nucleotide positions. Isolate NB7, NB8, NB9, NB10, NB11, NB12, NB13 and NB14 had alterations at nucleotide 136 (T→A), 187 (T→G) and 378 (G→A). And the sequences identified from the Oraon/Munda tribes differed from Type 2D strain (#230) at only one nucleotide position i.e., at 440. Isolate NB2 and NB4 had alteration at nucleotide 440 (T→A).

JCPyV capsid consists of 360 copies of VP1 arranged into 72 pentamers called "capsomers" in a T = 7 d icosahedral configuration. Each capsomer is further associated with a single copy of a minor capsid protein, VP2 or VP3. JCPyV binds to the sialylated oligosaccharide lactoseries tetrasaccharide c (LSTc), which interacts with residues in the BC and HI loops of VP1 (Neu *et al*., 2010; Stroh *et al*., 2015). VP1 mutations that disrupt the interactions with LSTc have effects on viral infectivity (Neu *et al*., 2010). PML related mutations are predominantly located in the external loops of VP1. Three of the most commonly mutated residues are L54, S266, and S268 which are located in sialic acid binding pocket of VP1 (Maginnis *et al*., 2013). Substitutions of these residues disrupt interactions of VP1 binding to LSTc (Neu *et al*., 2010; Maginnis *et al*., 2013). In the present study, due to absence of full length of JCPyV VP1 region the whole protein structure cannot be predicted *in silico*. Hence, the effect of the point mutations present in the VP1 region sequence of endemic JCPyV strains on viral tropism has not been determined.

**Table 25: Percentage pairwise similarities (relative to the maximum similarity) based on VP1 region of endemic JCPyV strains NB2, NB4, NB8, NB9, NB10, NB11, NB12, NB13, NB14 in comparison to Type 4, Type 1A, Type 1B, Type 2A, Type 2B, Type 2C, Type 2D, Type 6 and Type 3 using DiAlign alignment software of Genomatix suite v2.5 GmbH**

| | MH671352\|NB4 | MH708168\|NB7 | MH744484\|NB8 | MH744485\|NB9 | MH758781\|NB10 | MK728936\|NB11 | MK728937\|NB12 | MK728938\|NB13 | MK728939\|NB14 | MF662208\|Type4 | AF015526\|Type 1A | AF015527\|Type 1B | AF015528\|Type 4 | AF015529\|Type 2A | AF015532\|Type 2B | AF015534\|Type 2C | AF015536\|Type 2D | AF015537\|Type 6 | U73501\|Type 3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **MK728940\|NB2** | **100** | 95.88 | 95.88 | 95.88 | 95.88 | 95.88 | 95.88 | 95.88 | 95.88 | 96.08 | 95.46 | 96.08 | 95.88 | 97.53 | 98.56 | 97.73 | 99.59 | 97.11 | 97.94 |
| **MH671352\|NB4** | | 95.88 | 95.88 | 95.88 | 95.88 | 95.88 | 95.88 | 95.88 | 95.88 | 96.08 | 95.46 | 96.08 | 95.88 | 97.53 | 98.56 | 97.73 | 99.59 | 97.11 | 97.94 |
| **MH708168\|NB7** | | | **100** | **100** | **100** | **100** | **100** | **100** | **100** | 98.56 | 98.35 | 98.97 | 98.35 | 94.64 | 95.88 | 94.85 | 95.88 | 96.29 | 95.46 |
| **MH744484\|NB8** | | | | **100** | **100** | **100** | **100** | **100** | **100** | 98.56 | 98.35 | 98.97 | 98.35 | 94.64 | 95.88 | 94.85 | 95.88 | 96.29 | 95.46 |
| **MH744485\|NB9** | | | | | **100** | **100** | **100** | **100** | **100** | 98.56 | 98.35 | 98.97 | 98.35 | 94.64 | 95.88 | 94.85 | 95.88 | 96.29 | 95.46 |
| **MH758781\|NB10** | | | | | | **100** | **100** | **100** | **100** | 98.56 | 98.35 | 98.97 | 98.35 | 94.64 | 95.88 | 94.85 | 95.88 | 96.29 | 95.46 |
| **MK728936\|NB11** | | | | | | | **100** | **100** | **100** | 98.56 | 98.35 | 98.97 | 98.35 | 94.64 | 95.88 | 94.85 | 95.88 | 96.29 | 95.46 |
| **MK728937\|NB12** | | | | | | | | **100** | **100** | 98.56 | 98.35 | 98.97 | 98.35 | 94.64 | 95.88 | 94.85 | 95.88 | 96.29 | 95.46 |
| **MK728938\|NB13** | | | | | | | | | **100** | 98.56 | 98.35 | 98.97 | 98.35 | 94.64 | 95.88 | 94.85 | 95.88 | 96.29 | 95.46 |
| **MK728939\|NB14** | | | | | | | | | | 98.56 | 98.35 | 98.97 | 98.35 | 94.64 | 95.88 | 94.85 | 95.88 | 96.29 | 95.46 |
| **MF662208\|Type4** | | | | | | | | | | | 98.56 | 99.18 | 99.38 | 94.85 | 96.49 | 95.05 | 96.08 | 96.49 | 96.08 |
| **AF015526\|Type 1A** | | | | | | | | | | | | 98.56 | 97.94 | 94.23 | 95.46 | 94.43 | 95.46 | 95.88 | 95.46 |
| **AF015527\|Type 1B** | | | | | | | | | | | | | 98.97 | 94.85 | 96.08 | 95.05 | 96.08 | 96.49 | 96.08 |
| **AF015528\|Type 4** | | | | | | | | | | | | | | 94.64 | 96.29 | 94.85 | 95.88 | 96.29 | 95.88 |
| **AF015529\|Type 2A** | | | | | | | | | | | | | | | 97.32 | 98.97 | 97.73 | 96.08 | 96.70 |
| **AF015532\|Type 2B** | | | | | | | | | | | | | | | | 97.53 | 98.56 | 96.91 | 97.73 |
| **AF015534\|Type 2C** | | | | | | | | | | | | | | | | | 97.94 | 96.08 | 96.91 |
| **AF015536\|Type 2D** | | | | | | | | | | | | | | | | | | 97.11 | 97.94 |
| **AF015537\|Type 6** | | | | | | | | | | | | | | | | | | | 96.7 |



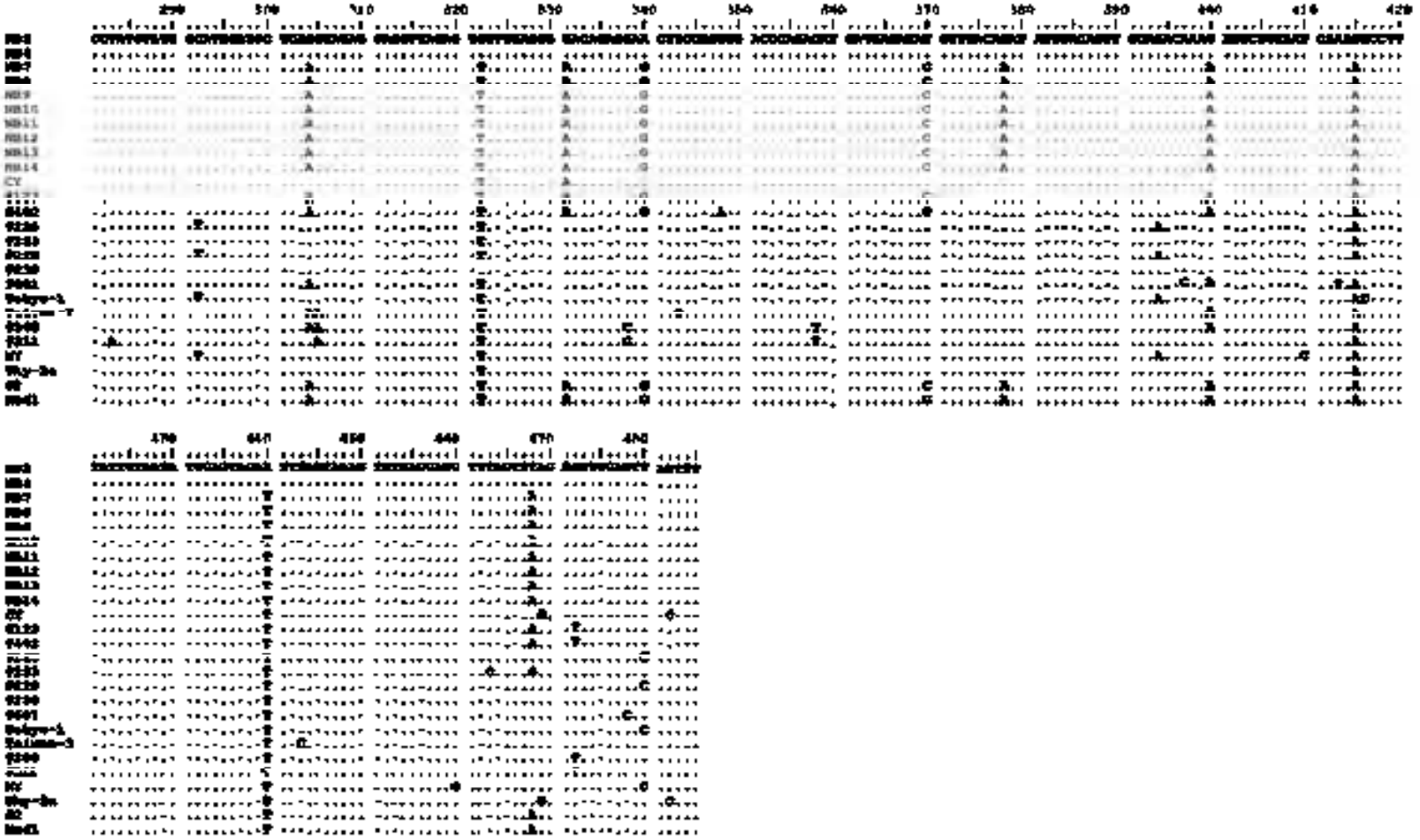


**Figure 13**: Multiple sequence alignment of VP1 region sequences of endemic JCPyV strains NB2, NB4, NB8, NB9, NB10, NB11, NB12, NB13, NB14 in comparison to Strains CY, #123, #402, #225, #223, #228, #230, #601, Tokyo-1, Taiwan-3, #308, #311, MY, Tky-2a, G2, Mad1 using Clustal X sequence alignment program.

| **Table 26: Strain specific nucleotide variation in VP1 region of JCPyV genome. Position of nucleotides is based on CY strain (archetype) numbering.** | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **JCPyV** | **53** | **100** | **103** | **121** | **136** | **142** | **169** | **187** | **196** | **262** | **304** | **322** | **331** | **340** | **370** | **378** | **400** | **415** | **440** | **468** |
| **NB2** | C | G | T | T | T | A | G | T | A | C | G | G | G | A | T | G | G | G | A | T |
| **NB4** | C | G | T | T | T | A | G | T | A | C | G | G | G | A | T | G | G | G | A | T |
| **NB7** | T | A | C | C | A | G | C | G | G | T | A | T | A | G | C | A | A | A | T | A |
| **NB8** | T | A | C | C | A | G | C | G | G | T | A | T | A | G | C | A | A | A | T | A |
| **NB9** | T | A | C | C | A | G | C | G | G | T | A | T | A | G | C | A | A | A | T | A |
| **NB10** | T | A | C | C | A | G | C | G | G | T | A | T | A | G | C | A | A | A | T | A |
| **NB11** | T | A | C | C | A | G | C | G | G | T | A | T | A | G | C | A | A | A | T | A |
| **NB12** | T | A | C | C | A | G | C | G | G | T | A | T | A | G | C | A | A | A | T | A |
| **NB13** | T | A | C | C | A | G | C | G | G | T | A | T | A | G | C | A | A | A | T | A |
| **NB14** | T | A | C | C | A | G | C | G | G | T | A | T | A | G | C | A | A | A | T | A |
| **CY** | C | G | T | T | T | A | G | T | A | C | G | T | A | A | T | G | G | A | T | T |
| **#124** | T | A | C | C | T | G | C | T | G | T | A | T | A | G | T | A | A | A | T | A |
| **#123** | T | A | C | C | T | G | C | T | G | T | A | T | A | G | C | G | A | A | T | A |
| **#402** | T | A | T | C | T | G | C | T | G | T | A | T | A | G | C | G | A | A | T | A |
| **#224** | C | G | T | T | T | A | G | T | A | C | G | T | G | A | T | G | G | A | T | T |
| **#225** | C | G | T | T | T | A | G | T | A | C | G | T | G | A | T | G | G | A | T | T |
| **#226** | C | G | T | T | T | A | G | T | A | C | G | T | G | A | T | G | G | A | T | T |
| **#223** | C | G | T | T | T | A | G | T | A | C | G | T | G | A | T | G | G | A | T | G |
| **#227** | C | G | T | T | T | A | G | T | A | C | G | T | G | A | T | G | G | A | T | G |
| **#228** | C | G | T | T | T | A | G | T | A | C | G | T | G | A | T | G | G | A | T | T |
| **#229** | C | G | T | T | T | A | G | T | A | C | G | T | G | A | T | G | G | A | T | T |
| **#230** | C | G | T | T | T | A | G | T | A | C | G | G | G | A | T | G | G | G | T | T |
| **#601** | C | G | T | T | T | A | C | T | G | T | A | T | G | A | T | G | A | A | T | T |
| **Tokyo-1** | C | G | T | T | T | A | G | T | A | C | G | T | G | A | T | G | G | A | T | T |
| **Taiwan-3** | C | G | T | T | T | A | G | T | A | C | A | T | G | A | T | G | A | A | T | T |
| **#308** | C | G | T | T | T | A | C | T | A | C | A | T | G | A | T | G | A | A | T | T |
| **#311** | C | G | T | T | T | A | C | T | A | C | G | T | G | A | T | G | G | A | T | T |
| **#312** | C | G | T | T | T | A | C | T | A | C | A | T | G | A | T | G | A | A | T | T |
| **MY** | C | G | T | T | T | A | G | T | A | C | G | T | G | A | T | G | G | A | T | T |
| **Tky-2a** | C | G | T | T | T | A | G | T | A | C | G | T | G | A | T | G | G | A | T | T |
| **G2** | T | A | C | C | A | G | C | T | G | T | A | T | A | G | C | A | A | A | T | A |
| **Mad1** | T | A | C | C | A | G | C | G | G | T | A | T | A | G | C | A | A | A | T | A |

## 4.7 Pairwise and Multiple sequence alignment of T-Ag sequences:

The JCPyV early coding region encodes two regulatory proteins, small t and large T-antigens. Little is known about the function of small t-antigen but the large T-Ag is a multifunctional phosphoprotein that mediates both initiation of viral DNA replication and T-Ag mediated activation of late genes (Khalili *et al*, 1987). Depending on the cell type, it may either support a lytic infection in permissive cells or contribute to the transformation in nonpermissive cells. Large T-Ag modulates many cellular functions by interacting with cellular regulators such as pRb and p53 to promote cell cycle progression.

Pairwise sequence comparison based on T-Ag sequence of JCPyV isolates of this region with different JCPyV types which included Type 1A, Type 1B, Type 2A, Type 2B, Type 2C, Type 2D, Type 3, Type 4 and Type 6 was done using DiAlign alignment software of Genomatix suite v2.5 GmbH (**Table 27**).

**Reference sequences**: GenBank accession numbers are given in brackets. JCPyV Type 1A strains: Mad1 [J0227] and #124 [AF015526]. JCPyV Type 1B strain: #123 [AF015527]. JCPyV Type 2A/2C strains: #224 [AF015529], #225 [AF015530], #226 [AF015531], #228 [AF015534], #229 [AF015535] and Tokyo-1 [AF030085]. JCPyV Type 2B strains: #223 [AF015532] and #227 [AF015533]. JCPyV Type 2D strain: #230 [AF015536]. JCPyV Type 3 strains: #308 [U73500], #311 [U73501] and #312[U73502]. JCPyV Type 1 strains: #123 [AF015527] and #124 [AF015526]. JCPyV Type 4 strain: #402 [AF015528]. JCPyV Type 6 strain: #601 [AF015537]. JCPyV Type 7 strain: Taiwan-3 [U61771]. JCPyV archetype strain CY [M35834], JCPyV MY strain [AB038250] and JCPyV Tky-2a strain [AB038255].

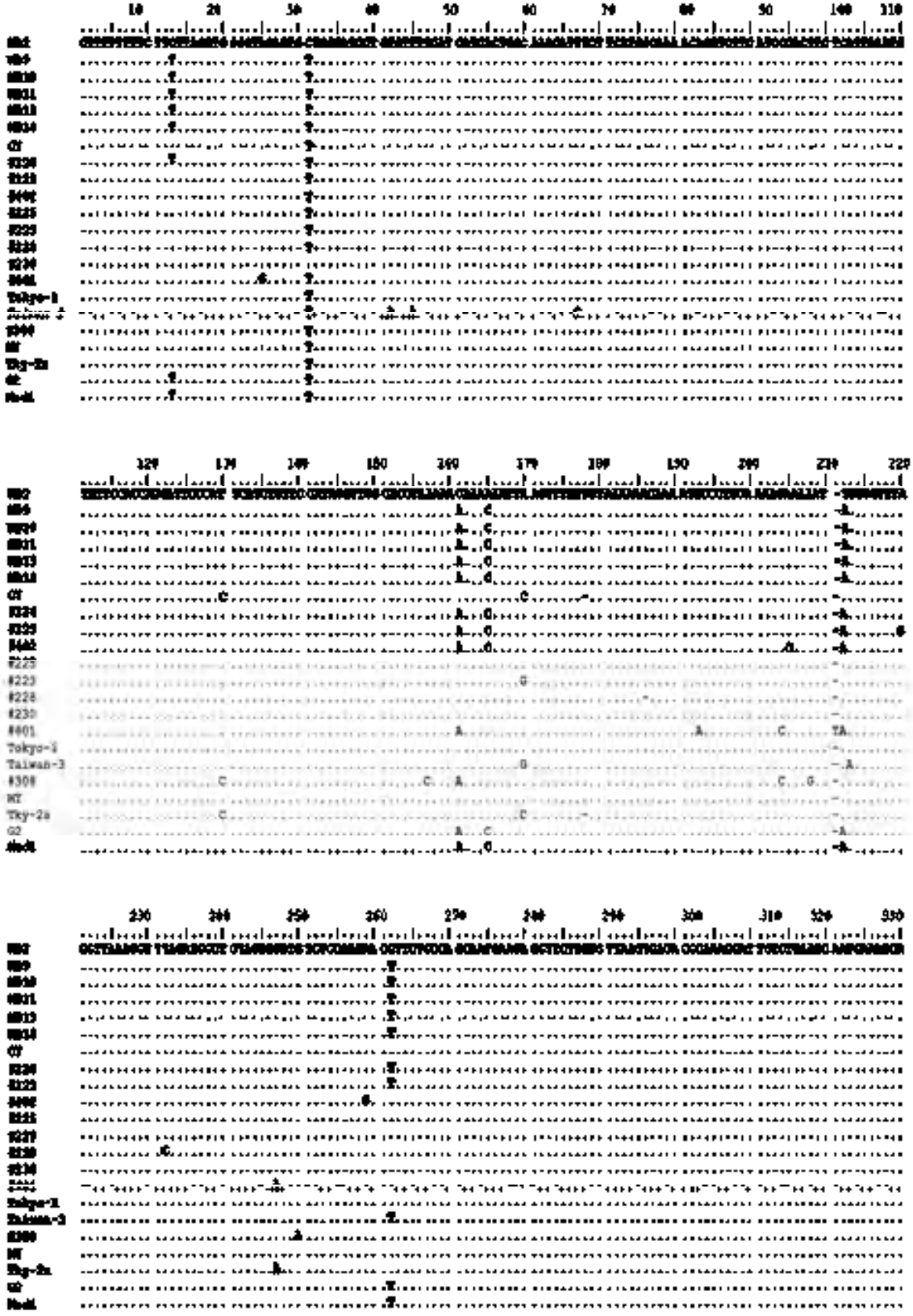
The results of pairwise sequence comparison of T-Ag sequences were similar to the results obtained in VP1 sequences. The isolates NB9, NB10, NB11, NB13 and NB14 from the Rabha tribes were 100% similar to each other and have 100% similarity to Type 1A and 99.69% to Type 1B (Europe). JCPyV Type 1 is mostly prevalent in Europeans and European-Americans. Type 1 has been in the populations

of Japan (Kitamura *et al*., 1998) and Korea (Jeong *et al*., 2002; Jeong *et al*., 2004) as a minor group. Type 1 has not been found in Asian groups other than Japanese and South Korean. Presence of Type 1 in this population suggests the possibility of transfer of this genotype from Europeans or European-Americans (Jeong *et al*., 2004). And the NB2 isolate was almost identical to the Type 2D (India) (99.84% similar). Type 2D is prevalent in both Asians and South Asians and is designated as the Indian subtype.The JCPyV strains isolated from Rabha tribes differ from the JCPyV strains of Oraon/Munda tribes by 1.88%.

The JCPyV sequences analyzed did not differ in lengthfrom the prototype strain JCPyV (Mad1). Multiple sequence alignment of endemic JCPyV T-Ag sequence and other previously reported strains revealed point mutations in eleven different sites within the region. Several point mutations have been recorded in the T-Ag sequence of endemic JCPyV strains (**Fig.14** and **Table 28**). The strains from Rabha tribes (NB9, NB10, NB11, NB13 and NB14) are identical to the Type 1A strain (#124, Mad1). The sequences isolated from Oraon/Munda tribes differed from Type 2D strain (#230) by only one nucleotide. Nucleotide alteration in the isolate NB2 was observed at nucleotide position 380 (C→T). The mutation observed in the endemic strain from oraon/munda group was present in the region that codes for small t antigen protein. Both large and small tumour antigen are produced by alternative splicing from early codng region of JCPyV. Not much is known about the function of small tAg. Small tAg has a binding site for PP2A (Khalili *et al*, 2008). PP2A is a serine/threonine phosphatase and a multifunction tumour suppressor gene (Zhang and Claret, 2012). Inhibition of PP2A is necessary for cell transformation (Janssens *et al*., 2005). Mutation in the stAg region may have effect on its interaction with PP2A and on viral replication. The effect of this point mutationpresentin the endemic strain of JCPyV in viral infection and tumorigenesis needs to be investigated. Full length T-Ag region was not sequenced and hence, *in silico* studies using whole TAg protein structure cannot be performed.

**Table 27: Pairwise similarities in percentage (relative to the maximum similarity) based on T-Antigen region of endemic JCPyV strains NB2, NB9, NB10, NB13, NB14 in comparison to Type 4, Type 1A, Type 1B, Type 2A, Type 2B, Type 2C, Type 2D, Type 6 and Type 3 using DiAlign alignment software of Genomatix suite v2.5 GmbH.**

| | **MH744486\|NB9** | **MH746212\|NB10** | **MK728941\|NB11** | **MK728942\|NB13** | **MK728943\|NB14** | **MF662208\|Type-4** | **AF015526\|Type 1A** | **AF015527\|Type 1B** | **AF015528\|Type 4** | **AF015529\|Type 2A** | **AF015532\|Type 2B** | **AF015534\|Type 2C** | **AF015536\|Type 2D** | **AF015537\|Type 6** | **U73501\|Type 3** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **NB2** | 98.12 | 98.12 | 98.12 | 98.12 | 98.12 | 98.28 | 98.12 | 98.12 | 98.28 | 99.53 | 99.37 | 99.22 | 99.84 | 98.28 | 98.59 |
| **MH744486\|NB9** | | **100** | **100** | **100** | **100** | 99.06 | **100** | 99.69 | 99.22 | 98.59 | 98.44 | 98.28 | 98.28 | 98.12 | 97.81 |
| **MH746212\|NB10** | | | **100** | **100** | **100** | 99.06 | **100** | 99.69 | 99.22 | 98.59 | 98.44 | 98.28 | 98.28 | 98.12 | 97.81 |
| **MK728941\|NB11** | | | | **100** | **100** | 99.06 | **100** | 99.69 | 99.22 | 98.59 | 98.44 | 98.28 | 98.28 | 98.12 | 97.81 |
| **MK728942\|NB13** | | | | | **100** | 99.06 | **100** | 99.69 | 99.22 | 98.59 | 98.44 | 98.28 | 98.28 | 98.12 | 97.81 |
| **MK728943\|NB14** | | | | | | 99.06 | **100** | 99.69 | 99.22 | 98.59 | 98.44 | 98.28 | 98.28 | 98.12 | 97.81 |
| **MF662208\|Type-4** | | | | | | | 99.06 | 99.06 | 99.53 | 98.75 | 98.59 | 98.44 | 98.44 | 98.44 | 98.12 |
| **AF015526\|Type 1A** | | | | | | | | 99.69 | 99.22 | 98.59 | 98.44 | 98.28 | 98.28 | 98.12 | 97.81 |
| **AF015527\|Type 1B** | | | | | | | | | 99.22 | 98.59 | 98.44 | 98.28 | 98.28 | 98.12 | 97.81 |
| **AF015528\|Type 4** | | | | | | | | | | 98.75 | 98.59 | 98.44 | 98.44 | 98.28 | 97.97 |
| **AF015529\|Type 2A** | | | | | | | | | | | 99.84 | 99.69 | 99.69 | 98.43 | 99.06 |
| **AF015532\|Type 2B** | | | | | | | | | | | | 99.53 | 99.53 | 98.28 | 98.9 |
| **AF015534\|Type 2C** | | | | | | | | | | | | | 99.37 | 98.12 | 98.74 |
| **AF015536\|Type 2D** | | | | | | | | | | | | | | 98.12 | 98.74 |
| **AF015537\|Type 6** | | | | | | | | | | | | | | | 98.12 |



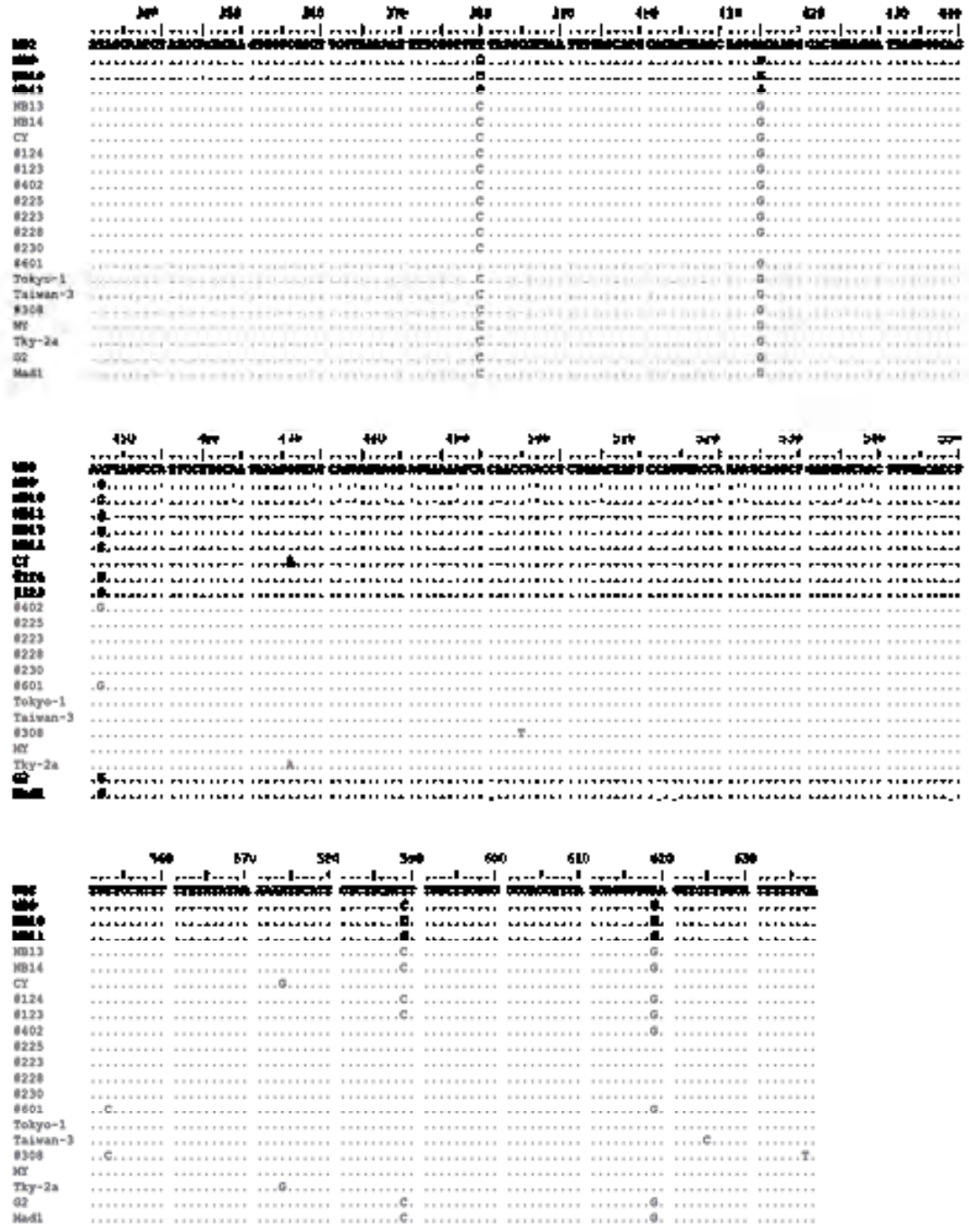


**Figure 14**: Multiple sequence alignment of T-antigen region of endemic JCPyV strains NB2, NB9, NB10, NB13, NB14 in comparison to Type 4, Type 1A, Type 1B, Type 2A, Type 2B, Type 2C, Type 2D, Type 6 and Type 3 using ClustalX sequence alignment program.

| Table 28: Strain specific nucleotide variation in T-Ag region of JCPyV genome. Position of nucleotides based on CY strain (archetype) numbering. | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **JCPyV** | **13** | **31** | **161** | **165** | **212** | **262** | **380** | **415** | **442** | **589** | **619** |
| **NB2** | C | C | C | A | T | C | T | A | A | T | A |
| **NB9** | T | T | A | C | A | T | C | G | G | C | G |
| **NB10** | T | T | A | C | A | T | C | G | G | C | G |
| **NB11** | T | T | A | C | A | T | C | G | G | C | G |
| **NB13** | T | T | A | C | A | T | C | G | G | C | G |
| **NB14** | T | T | A | C | A | T | C | G | G | C | G |
| **CY** | C | T | C | A | T | C | C | G | A | T | A |
| **#124** | T | T | A | C | A | T | C | G | G | C | G |
| **#123** | C | T | A | C | A | T | C | G | G | C | G |
| **#402** | C | T | A | C | A | C | C | G | G | T | G |
| **#224** | C | T | C | A | T | C | C | G | A | T | A |
| **#225** | C | T | C | A | T | C | C | G | A | T | A |
| **#226** | C | T | C | A | T | C | C | G | A | T | A |
| **#223** | C | T | C | A | T | C | C | G | A | T | A |
| **#227** | C | T | C | A | T | C | C | G | A | T | A |
| **#228** | C | T | C | A | T | C | C | G | A | T | A |
| **#229** | C | T | C | A | T | C | C | G | A | T | A |
| **#230** | C | C | C | A | T | C | C | A | A | T | A |
| **#601** | C | T | A | A | A | C | T | G | G | T | G |
| **Tokyo-1** | C | T | C | A | T | C | C | G | A | T | A |
| **Taiwan-3** | C | T | C | A | T | T | C | G | A | T | A |
| **#308** | C | T | A | A | T | C | C | G | A | T | A |
| **#311** | C | T | A | A | T | C | C | G | A | T | A |
| **#312** | C | T | A | A | T | C | C | G | A | T | A |
| **MY** | C | T | C | A | T | C | C | G | A | T | A |
| **Tky-2a** | C | T | C | A | T | C | C | G | A | T | A |
| **G2** | T | T | A | C | A | T | C | G | G | C | G |
| **Mad1** | T | T | A | C | A | T | C | G | G | C | G |

## 4.8 Phylogenetic analyses based on VP1 and T-Antigen:

The phylogenetic relationship of JCPyV was determined by maximum likelihood analyses based on VP1 and T-Ag sequences. **Fig.15** represents the phylogenetic tree of JCPyV based on VP1 region sequences and **Fig.16** represents phylogenetic tree based on T-Ag coding region sequences.

Because of the ubiquitous distribution of JCPyV and its association with particular ethnic groups, it can be used as a population marker. Like mtDNA, JCPyV mutates relatively rapidly compared to nuclear genes. But when compared to other viruses the changes are very slow. Low rate of mutation rate and long history of presence of JCPyV in the human population makes it useful in the human migration study. Millions of years ago, some from the original population of *Homo sapiens sapiens*, native to Africa migrated to the near East and further to West, East, and North, differentiating into Caucasoids and Mongoloids (Vigilant *et al.*, 1989). Distribution of JCPyV subtypes is compatible with the above mentioned concept.

JCPyV genome classification by creating a phylogenetic tree based on variation in the 610 bp intergenic region sequence has allowed identification of 12 genotypes- EU, Af1, Af2, Af3, B1-a, B1-b, B1-c, B1-d, B2, CY, MY and SC (Sugimoto *et al*., 1997; Guo *et al*., 1998). Some of them were further subdivided, such as EU was divided into three genotypes EU-a, EU-b and EU-c (Sugimoto *et al.*, 2002). Different types of JCPyV have been associated with different populations and have been used to study migration of human population. The Eu-a, Eu-b and B1-C are mainly spread over Europe and Mediterranean areas. Af2 is spread throughout Africa and in West and South Asia. Af1 and Af3 are localized in West and Central Africa, respectively. The B1-a, B1-b, B1-d, B2, CY, MY and SC are distributed throughout East Asia.

More recently, sequencing of full-length JCPyV sequences have identified 8 JCPyV types that were numbered 1 through 8 each having multiple subtypes (Cubitt *et al.*, 2001). Type 5 was found to be a minor member of type 3 (Agostini *et al.*, 1997). Pavesi hypothesized that initial interaction between human and JCPyV arose in Africa and Type 6 is the original JCPyV type. The synthetic map presented by him showed two early expansions of humans from Africa, each carrying a different virus lineage. Progenitor Type 6 gave rise to two independent lineages, one including Type 1/Type 4 and the other including Type 2, 3, 7 and 8 (Pavesi, 2003).

Type 1 and type 4 are generally found in Europeans and European-Americans, while type 2A in Asians and Native American populations. Types 3 and 6 are associated with Africans and African-Americans. Types 2D and 7C are found in both Asians and South Asians (Yanagihara *et al*., 2002; Cui *et al*., 2004). Types 2E, 8A, and 8B are found in Western Pacific populations (Yanagihara *et al*., 2002). Type 8A is detected only in Papua New Guinea populations (Jobes *et al*., 2001).

The present analysis based on both the genes showed that isolates from Rabha tribe form a distinct group and another such group was formed by the isolates from the Oraon and Munda tribal group. In the phylogenetic tree based on VP1 region of JCPyV (**Fig.15**), the isolates from Rabha tribal group NB7, NB8, NB9, NB10, NB11, NB12, NB13 and NB14 clustered with the strains #123, #124, #402, G2 and Mad1 suggesting a close affinity towards the EU/Type 1/ Type 4 group. The EU/Type 1/Type 4 is generally found in European and in the European-American population and is mainly spread over Europe and Mediterranean areas. Type 1/Type 4 strains have been detected in Japan (Kitamura *et al*., 1998) and in Korean population (Jeong *et al*., 2002), although it has not been detected in most of other Asian groups. Jeong and co-workers detected the JCPyV Type 1 genotype in two Korean PML patients with AIDS (Jeong *et al*., 2002). Type 1 exists as a minor group in Korean population (Jeong *et al*., 2004). The presence of Type 1 in Asian population suggests the possibility of transmitting of this genotype from Europeans or European-Americans (Jeong *et al*., 2004). Type EU has been found in sample from South Korea (Sugimoto *et al*., 1997). Sixteen EU isolates from Japan (Kitamura *et al*., 1998) and a single EU isolate from South Korea (Sugimoto *et al*., 1997) have been detected in earlier studies. No rational explanation can be given to the presence of EU in these samples. The occurrence of the minor JCPyV genotype EU in Japan indicates that the minor group may have migrated to Japan and contributed to the formation of modern Japanese population. Detection of EU in Japan can be the result of ancient colonization by Caucasians. These findings suggest that some distinct groups of Caucasian origin may have migrated to the East Asia during ancient times (Kitamura *et al*., 1998). Rabha are an indigenous Assamese community mostly residing in Assam, Meghalaya and West

Bengal. Rabhas are mostly concentrated in two districts of West Bengal, Jalpaiguri and Coochbehar. They exhibit some Mongoloid features. Study based on Y-chromosome haplogroup diversity showed that the Rabhas clustered with the North East Asians (Debnath *et al*., 2011). In another study based on KIR gene frequencies proximity of Rabha population towards mongoloid ethnicity was found (Guha *et al*., 2015). Chakraborty and co-workers have also shown the presence of mongoloid element in the Rabha gene pool (Chakraborty *et al*., 1986). Hence, the fact that JCPyV has been postulated to have been co-evolving wththe humans. The proximity of the Rabha population to the North East Asian population can be the reason forthe presence of Type 1/Type 4/EU genotype in this population.

The NB2 and NB4 isolates from the Oraon and Munda tribal group clustered together with the strain #230 suggesting an affinity towards Type B1-b/Type 2D group. Type 2D has been designated as the Indian subtype. Types B1-b/Type 2D are found in both Asians and South Asians and are distributed throughout Asia. The B1 could have originated in the near East and its extension eastward might have accompanied the migration of protoMongoloids.

Similar results were observed in the phylogenetic tree based on T-Ag coding regions (**Fig.16**). The isolates from Rabha tribes (NB9, NB10, NB11, NB13 and NB14) clustered with the EU/Type 1/Type 4 group and isolate from the Oraon and Munda tribes (NB2) clustered with the B1-b/Type 2D group. The results based on both the genes were in accordance with the results of pairwise comparison of the sequences.

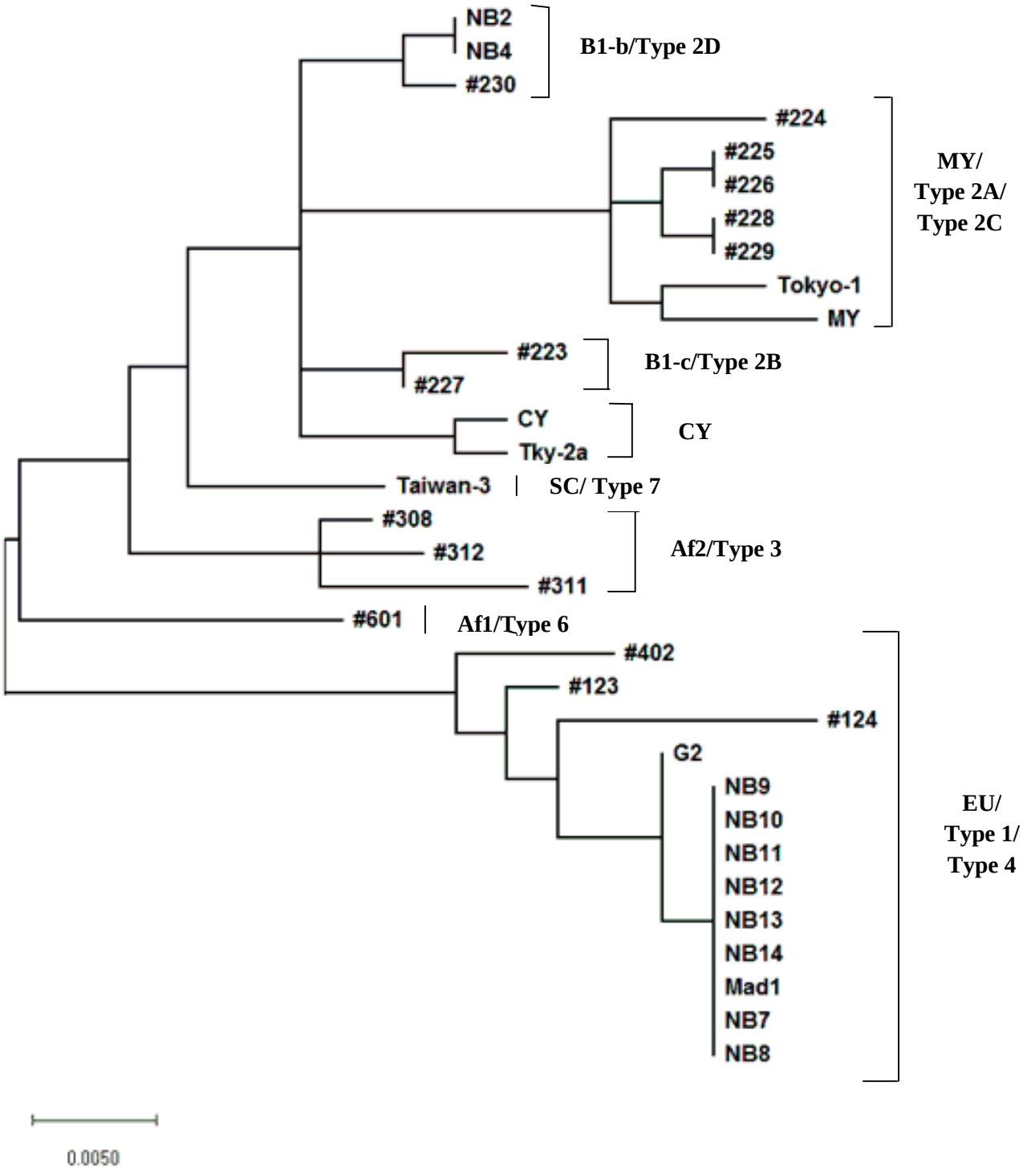

NB2
NB4
#230
B1-b/Type 2D
#224
#225
#226
#228
#229
Tokyo-1
MY
MY/
Type 2A/
Type 2C
#223
#227
B1-c/Type 2B
CY
Tky-2a
CY
Taiwan-3
SC/ Type 7
#308
#312
#311
Af2/Type 3
#601
Af1/Type 6
#402
#123
#124
G2
NB9
NB10
NB11
NB12
NB13
NB14
Mad1
NB7
NB8
EU/
Type 1/
Type 4
0.0050

**Figure 15:** Molecular Phylogenetic analysis of JCPyV based on VP1 sequences by Maximum Likelihood method. The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura 3-parameter model (Tamura, 1992). The tree with the highest log likelihood (-1217.77) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 32 nucleotide sequences. There were a total of 485 positions in the final dataset. Evolutionary analyses were conducted in MEGA X (Kumar *et al*., 2018).

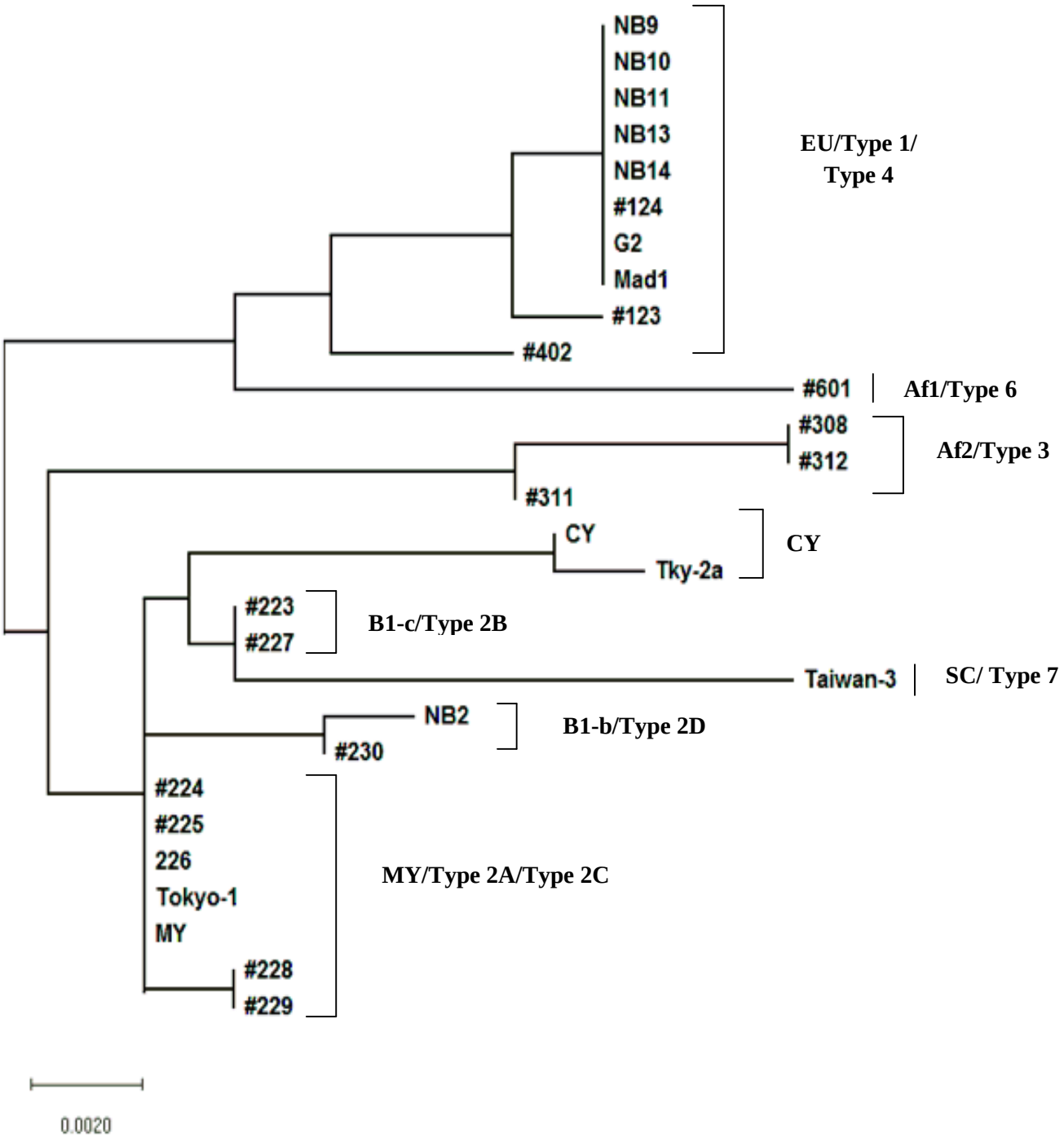

NB9
NB10
NB11
NB13
NB14
#124
G2
Mad1
#123
#402
EU/Type 1/
Type 4
#601
Af1/Type 6
#308
#312
Af2/Type 3
#311
CY
Tky-2a
CY
#223
#227
B1-c/Type 2B
Taiwan-3
SC/ Type 7
NB2
#230
B1-b/Type 2D
#224
#225
226
Tokyo-1
MY
#228
#229
MY/Type 2A/Type 2C
0.0020

**Figure 16:** Molecular Phylogenetic analysis of JCPyV based on T-Ag sequences by Maximum Likelihood method. The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura 3-parameter model (Tamura, 1992). The tree with the highest log likelihood (-1152.51) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 28 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 635 positions in the final dataset. Evolutionary analyses were conducted in MEGA X (Kumar *et al.*, 2018).

## 4.9 Viral load estimation by real time PCR:

This study, being the first ever report on prevalence and distribution pattern of JCPyV in sub-Himalayan part of West Bengal, also quantified viral load in JCPyV positive samples. The DNA was extracted from urine and whole blood and used for quantification of JCPyV by real time PCR. The viral load was estimated in the JCPyV positive samples by using JCVRVF (5'-TCAATGGATGTTGCCTTTACTTT-3') and JCVRVR (5'-ACGGGGTCCTTCCTTTCTC-3') primers specific for VP1 region that amplified a 109 bp fragment. Quantitative real-time PCR was performed using SYBR green dye-based kit from Roche (Switzerland). For detection of JCPyV viral load in urine and blood samples by qPCR, standard curve (**Fig. 17**) was plotted by using 10-fold serial dilutions of standard plasmid pMITC-BSMKS carrying the full length of JC viral genome (JCPyV Mad-1 strain; 5130 base pairs). Six dilutions of this standard DNA were taken during the reaction. **Fig.s 18** and **19** represent the amplification plot of tribal samples and pregnant women samples including the standard JC viral DNA in several dilutions. Amplification plot shows the change in fluorescence (ΔRn) in respect to PCR cycle number. The amplification curve shifted towards right with reduced quantity of DNA. **Fig.s 20** and **21** represent the melting curves of all the JCPyV DNA samples along with the standard plasmid.

Details regarding the copy number and $C_T$ value range of the JCPyV positive individuals are shown in **Table 29**, which also includes the mean and median values of viral DNA load in urine and blood samples from the study population. $C_T$ (cycle threshold) is defined as the first cycle in which amplification signal is detected over the mean base line. Mean base line is calculated from ΔRn values of 40 cycles. A wide range of viral load was found in the samples ranging from $3.52 \times 10^2$ to $6.71 \times 10^6$ copies/ml of sample. The mean and median copy number values were found to be $8.67 \times 10^5$ copies/ml of sample and $9.47 \times 10^5$ copies/ml of sample respectively. The mean viral load in urine ($2.19 \times 10^6$) was higher than that in the blood ($7.77 \times 10^5$) samples. Different studies have described variable range of JCPyV viral loads in different parts of the world (Delbue *et al*., 2010; Husseiny *et al*., 2010). The mean ($1.28 \times 10^6$) and

the median ($9.47 \times 10^5$) viral copy number recorded in the present study are comparatively higher than those observed in the some of the studies performed previously. The mean and median viral load in the samples were comparatively higher than normal healthy subjects of Switzerland (Egli *et al.*, 2009), USA (Husseiny *et al*., 2010), Portugal (Matos *et al*., 2010), Kuwait (Chehadeh *et al.*, 2013) and Pakistan (Hussain *et al*., 2017). However, it was lower than observed in three Italian immunocompetent subjects (Rossi *et al.*, 2007) and healthy adult women of USA (Kling *et al*., 2012).

Based on these very few reports available on JCPyV viral load, it may be hypothesized that it is comaparitively higher in this region than other parts of the world. However, more studies may be required to substantiate these findings. It has long been suggested that nutritional deficiency might result in immunodeficient conditions which in turn enhances susceptibility towards infections (Harbige, 1996). Majority of population in India cannot meet their dietary requirements due to poor socio-economic conditions. The nutritional deficiencies couldalso explain the high viral load in this region in healthy individuals. However, more studies will be needed to understand the contribution of nutritional deficiencies in immunosuppression.

**Table 29: Details of copy number value and CT value range of JCPyV viral DNA in both urine and blood samples of the population studied.**

| S.No. | Specimen Type | CT Value Range | Mean CT value | Copies/ml of sample | Mean Copy No. | Median Copy No. |
|---|---|---|---|---|---|---|
| 1 | Urine | 22.48 - 39.19 | 27.76 | $3.52 \times 10^2$ – $6.71 \times 10^6$ | $2.19 \times 10^6$ | $1.13 \times 10^6$ |
| 2 | Blood | 27.05 - 31.19 | 28.48 | $1.71 \times 10^5$ - $1.18 \times 10^6$ | $7.77 \times 10^5$ | $7.69 \times 10^5$ |
| Total | | 22.48 - 39.19 | 28.20 | $3.52 \times 10^2$ – $6.71 \times 10^6$ | $1.28 \times 10^6$ | $9.47 \times 10^5$ |

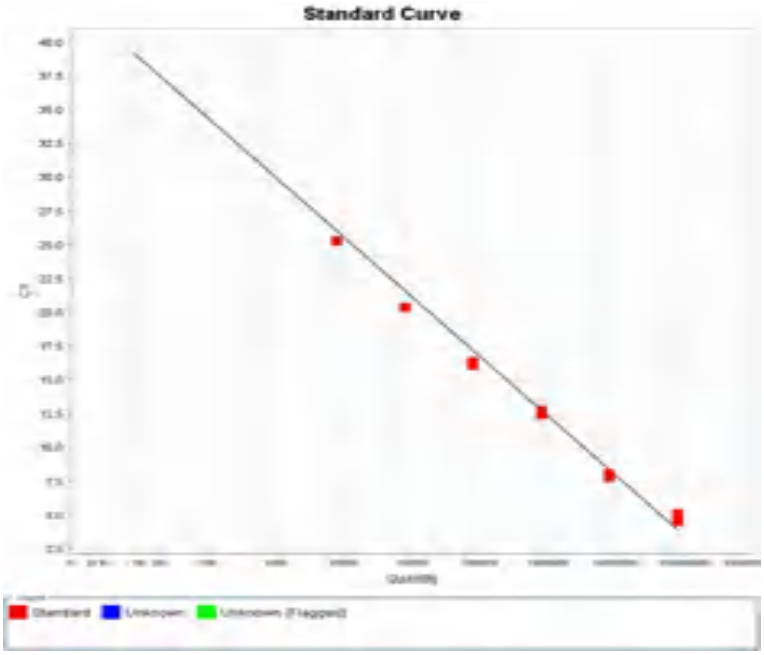



**Figure 17:** Standard curve plot of the JCPyV DNA quantity against cycle threshold (CT), where CT is defined as the first cycle in which amplification signal is detected over the mean base line. Mean base line is calculated from ΔRn values of 40 cycles.

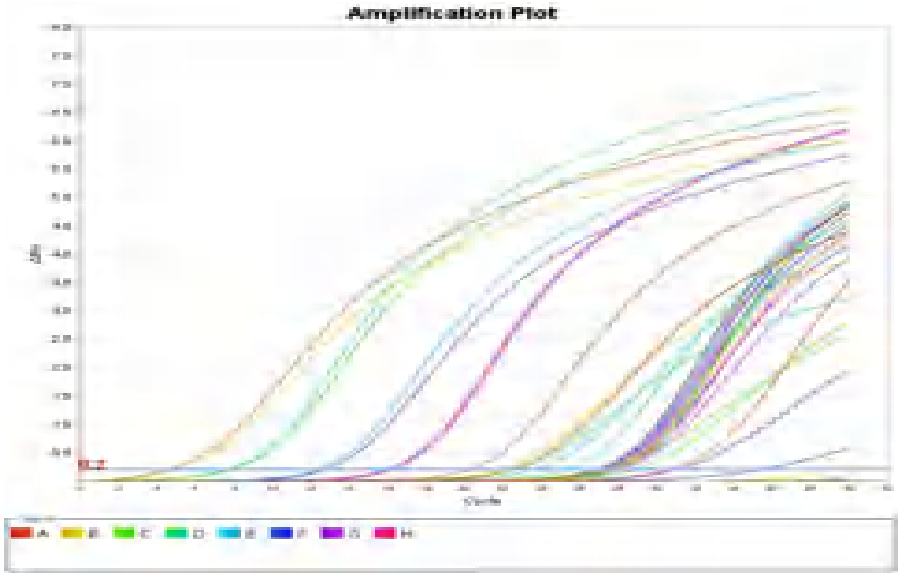



**Figure 18:** Amplification plot of tribal samples along with Standard JCPyV DNA (in 1:10 serial dilutions) using the JCPyV VP1 region specific primers. Amplification plot shows the change in fluorescence (ΔRn) in respect to PCR cycle numbe

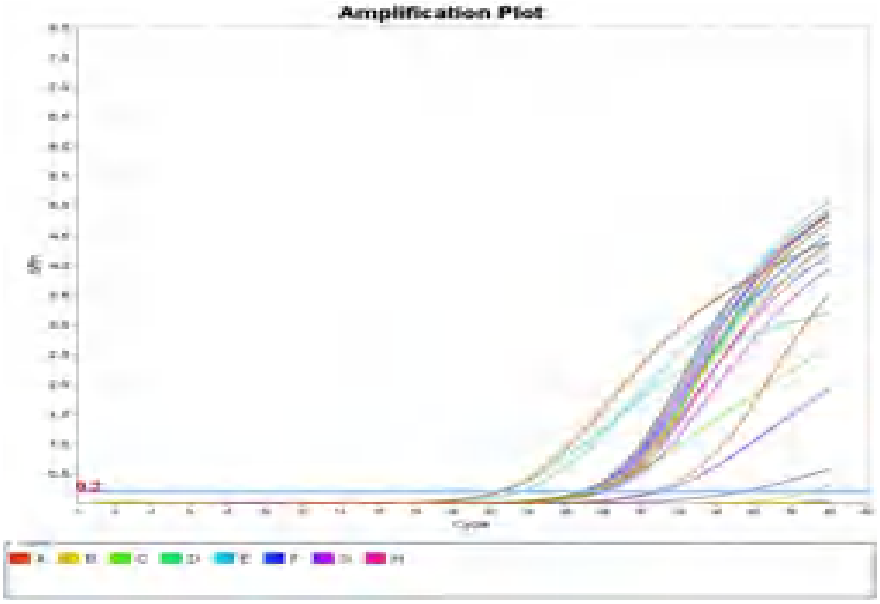



**Figure 19:** Amplification plot of pregnant women samples and standard JCPyV DNA (in 1:10 serial dilutions) using the JCPyV VP1 region specific primers. Amplification plot shows the change in fluorescence (ΔRn) in respect to PCR cycle number.

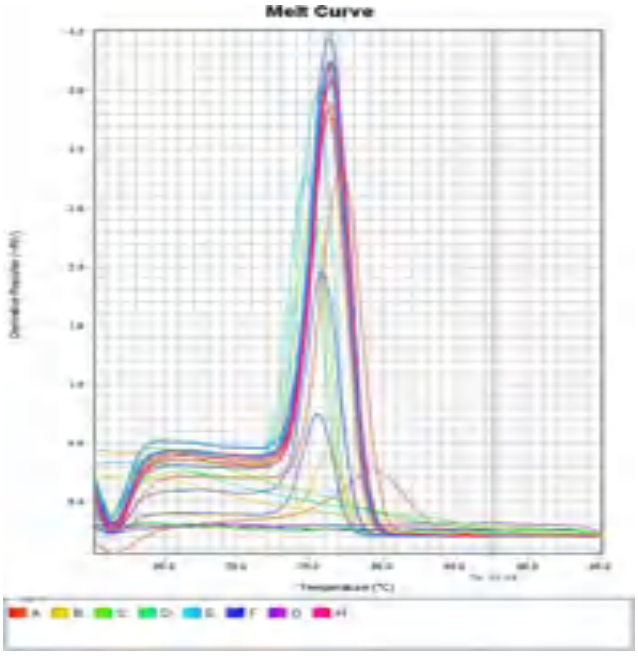



**Figure 20:** Melting curve of tribal JCPyV DNA including the standard JCPyV DNA

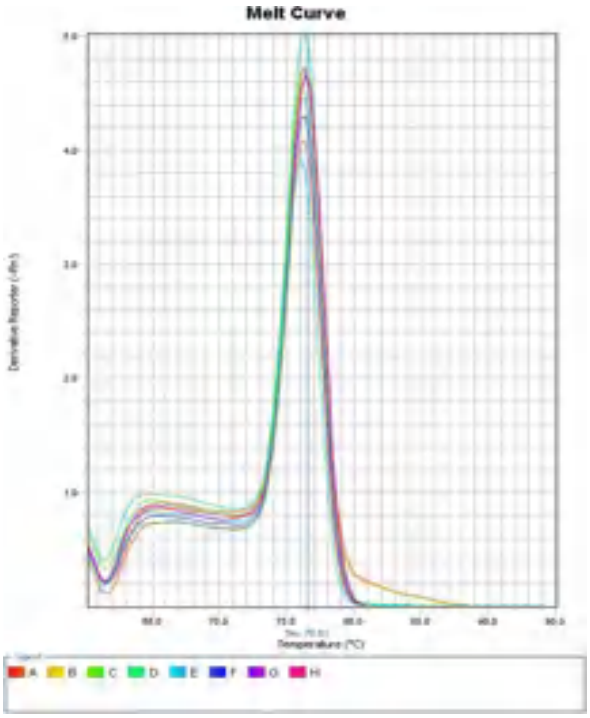



**Figure 21:** Melting curve of JCPyV DNA isolated from pregnant women group and standard JCPyV