International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor (2018): 7.426

Study of Five Complete Genomes of *Vibrio Parahaemolyticus*: Focusing Non-Pandemic to Pandemic Development

Sushanta Kumar Bhowmik

Abstract: ***Five complete genome sequences of Vibrio parahaemolyticus strains, FDA_R31, CDC_K4557, UCM-V493, BB22OP& RIMD 2210633were available in the database, and RIMD 2210633 was the first published and well documented pandemic strain. We performed the comparative study of these genome sequences and their genomic level development from environmental as well as non-pandemic to the pandemic character of this bacterium. The non-pandemic genomes of V. parahaemolyticus use a freedom of Genomic DNA-Segmental arrangement and pandemic genome is rather distinct and fixed. Major pandemic development may happen even in environment with change of essentialities. Thermostablehemolysin or thermostablehemolysin delta-VPH gene for chromosome-1 and thermolabilehemolysin or lecithin-dependent hemolysin (LDH) gene for chromosome-2 is global in V. parahaemolyticus genomes. The nucleotide sequences and genetic sequences of chromosome-1s are rather conserved against that of the chromosome-2s whose are more sensitive to pandemic development. Among the five full genomes, UCM-V493 strain beholds a midterm position in the series of development. The classified genes of hemolysins and other pathogen related genes clearly divided these five strains in two, one of FDA_R31 & CDC_K4557 strains and second of UCM-V493, BB220P & RIMD 2210633strains. The genome of FDA_R31 strain is the ancestral, holding total 39 transposase genes that showed more susceptibility to receive foreign genetic element. The genomes of environmental BB220P (1980) and pandemic RIMD 2210633 strains are almost similar that proves that the pandemic development happened in an environment and long before 1995 also. Chrmosome-2s of two stool isolated genomes of non-pandemic FDA_R31 and pandemic RIMD 2210633strains carried toxin-location of very different genomic character and arrangement. The toxin location of chromosome-2 of FDA_R31 strain is in very special feature with 7-ureage gene cassette and three hemolysin genes, without trh gene that may be an intermediate development of this pathogen. This study revealed that even in the environment, V. parahaemolyticus have a sequential development towards pandemic form in genome sequence level of two chromosomes, whereas chromosome-2 is more responsible towards a confined pandemic arrangement, especially, in toxins and other pathogenic genes of toxin location.***

Keywords: *V. Parahaemolyticus* (Vp), toxin location (TL), hemolysins and other pathogen related genes, genomic segment (GS)

1. Introduction

Vibrio parahaemolyticus (Vp) the worldwide agent of gastroenteritis is a major cause of seafood contamination and human health hazard and a global fear to the seafood eaters (Fujino T 1951). The researchers are hunting about the information to have its' details of mystery in pathogenic developments in its genomic level. A comparative genomic analysis of incomplete genome sequences of six strains of *V. parahaemolyticus* isolated from Asia and Peru was performed in order to advance knowledge concerning the evolution of *V. parahaemolyticus* (Chen Y et al., 2011). This need of full genomic information results the availability of five full genome sequences in NCBI data base from researchers. It is very frequently asked about the changes in the genome sequences of bacteria from environmental strains of the pandemic developments. This study is to invade into and find out significant analysis of Vp-genomes-FDA_R31, CDC_K4557, UCM-V493, BB22OP & RIMD 2210633. The FDA_R31 and CDC_K4557 strains of *V. Parahaemolyticus* are collected from human stool sample (2006) and oyster (4007) respectively, and their complete genome sequences directly submitted on May 2013 to NCBI database. UCM-V493 (Kalburge et al., 2014) & BB220P (Jensen et al., 2013) strains are environmental, from sea-sediment of Spain (2002) and environment of Bangladesh (place not mentioned) in early 1980 respectively, while RIMD 2210633 Vp-strain is the only designated pandemic strain of the clinical isolate of 1996 and first published full genomic nucleotide sequence (Makino et al., 2003, Nair et al., 2007) with descriptive studies (Table No 1).

RIMD 2210633 Vp-genome sequence published in 2003 was a pandemic strain of traveliers' recourse was well documented having two major toxin genes-thermostable direct hemolysin A (*tdh*A) and thermostable direct hemolysin S (*tdh*S) with genes of 'Type Three Secretion System' (TTSS) proteins, *tox*R regulatory gene and genes of transposase in 80 kb flanking region called pathogenicity island in chromosome-2 (Makina K et al., 2003; Park et al 2004; Bhowmik et al. 2014). Such integration of pathogenicity island in genomes whether universal or not was a big interrogation in *V. parahaemolyticus* genome development. Toxin rich such pathogenicity islands are a modification in the cause of disease development in *V. parahaemolyticus* infection and the study of such toxin location in chromosome level is a vital point of interest in this study.

Now the availability of more four full genome sequences of *V. Parahaemolyticus* from different sources creates a scope to solve this question.

Generally *V. Parahaemolyticus* does not carry a plasmid like extra-chromosomal sequence, but UCM-V493 strain carried a plasmid of 88,530 bp (Kalburge et al., 2014) which is not in any lead to the major pathogenic development of this strain (in this study). *V. Parahaemolyticus* specific major pandemic toxin gene is thermosable direct hemolysin (*tdh*) and *tdh*-related hemolysin (*trh*) gene also infamous with gastroenteritis infection of this pathogen (Takeda 1976, Tada 1992).

The *trh* and *tdh* positive strains had different types of pathogenicity islands and mobile elements as well as major

structural differences between the *tdh* pathogenicity islands of the pre-pandemic and pandemic strains (Chen Y at al., 2011). The *trh* gene is associated with an ureage gene cassette that has been studied well from major seafood isolated Vp-strains (Park et al., 2000). Therefore, *trh* gene positive strains of *V. Parahaemolyticus* showed ureage positive in biochemical test result. From the complete genome sequences of Vp-FDA_R31 strain, chromosome-2 carries a 7-ureage gene cassette, but no *trh* gene (in this study). The *tox* R gene is global toxin regulatory gene of *vibrio*s and pandemic Vp-strains carries *tdh* gene accompanied with the *tox*R gene (Lin Z et al., 1993). At least twenty proteins of 'type three secretion system' (TTSS) are well recommended for the role of extraintestinal infection of *V. parahaemolyticus* while pandemic strains are enriched with additional TTSS2 genes essentially (Park et al., 2006). Transposase genes play vital role in pathogenic gene transferring through lateral gene transfer of pathogenic development of bacteria. *V. parahaemolyticus* also carry transposase genes in the genome that have a good role in its' pandemic development (Terai A et al., 1990, 1991). All these hemolysins and other pathogen related genes are being well orchestrated in defining chromosomal regions are essentially assessed to evaluate the evolutionary direction of this pathogen.

Here we are trying to identify the respective as well as comparative developments or changes in the genome level of *V. parahaemolyticus* strains of different resources.

2. Material and Method

The five full genome sequences and the respective genes' sequences of *V. Parahaemolyticus* are being collected from the NCBI database in FASTA and GenBank formats. Nucleotide sequences of FASTA formats are used to perform phylogeny studies by MEGA6 from an evolutionary point of view. GeneBank formats are used mainly in Mauve 2.3.1 program (source details) for pointing the genes in chromosome level. The most reference and documented genome sequence of RIMD 2210633 strain is used as the pandemic strain (Makino et al., 2003). Strain BB22OP of *V. parahaemolyticus* is a pre-pandemic strain of 1980 and the other three strains are non-pandemic category as per presence or absence of two thermostabe direct hemolysin genes (*tdh*A & *tdh*S) in the genome sequence (Makino et al., 2003). In these genomes, other than *tdh*A and *tdh*S genes, non-pandemic hemolysin genes, TTSS genes, transposase genes and *tox*R genes are very much involved in toxin-activities and pandemic gene transferring. But it is not well defined yet. The DNA sequences of hemolysins and other pathogen related genes are being selected from five full genomes. They are total 83 hemolysin genes, 167 genes of TTSS genes, 77 transposase genes and 9 *tox*R genes from chromosomes (Table No 2 A& B). The evolutionary pattern has been studied with MEGA6.

Total used toxin and pathogen related genes are collected from respective GenBank data and are listed in Table No 2 and depicted respective positions of genes in 100% concentric circles (Figure No 3 A & B) with the help of Microsoft Excel. The presence of chromosomal depicted genome circles of the respective strains placed from out mostly in the core position - RIMD, BB220P, UCM-V493, CDC_K4557 & FDA_R31 strains of the comparative concentric circles as per the result of evolutionary steps (Figure No 1 C). The circles are been made with the formula:

$$\frac{(\text{end nucleotide number of gene - initiation nucleotide number of gene})+1}{\text{length of chromosome}} \times 100$$

The gaps between two serial genes are not spared from this calculation. After circle formation the respective gene-serial numbers & positions are edited manually.

3. Result

This study performed in three stages and chromosomal sequence wise. First, it was to check the evolutionary point of views with whole genome sequences and individual gene hemolysins and other pathogen related gene level. Second, it was to find out the placement situation of selected genes in comparative chromosomal genome circles and to point out the accumulated gene-location as Toxin Location (TL). Third, it was to evaluate the arrangements of Genome Segments (GSs) following the 'First' evolutionary findings and to do the exclusive study of the TL region of 'Second' study with the selected gene sequences in respective chromosome.

The comparative genomic studies are performed in three consecutive stages. 1. Evolutionary phylogeny study, 2. Study of Hemolysins and other pathogen related genes in comparative genome circles, and 3. Rearrangement study of genomic DNA-segments and toxin related genes

3.1. Evolutionary phylogeny study performed withMEGA6 software

Study of chromosomal genome sequences-
The chromosomal nucleotide sequences in FASTA format were used in the phylogeny study to get an evolutionary result in case of a complete chromosomal sequence based with individual hemolysins and on other pathogen related gene (Figure No 1). The neighbour-joining result of chromosome-1s showed BB220P & UCM-V493 strains were homologous and reported pandemic RIMD 2210633 strain was midterm stage while FDA_R31 strain was the ancestor among these fives (Figure No 1 A). The neighbor-joining figure of chromosome-2s (Figure No 1 B) differed from that of UPGMA figure (Figure No 1 C). UPGMA figure was the better evolutionary result towards pandemic development of *V. parahaemolyticus.* The sum was that the chromosome-2s of RIMD 2210633 & BB220P strains and UCM-V493 & CDC_K4557 strain were nearest, while chromosome-2 of FDA_R31 strain was the ancestor as well.

FDA_R31 strain was the representation of non-pandemic Vp-genome of environmental clone; UCM-V493 strain was a midterm form toward pandemic genome of RIMD 2210633 like strains.

3.2. Study of hemolysins and other pathogen related genes

3.2.1. Hemolysins

Total 83 nucleotide sequences of hemolysin genes were used in this study to check their evolutionary result within five Vp-strains (Figure No 2 A). Two pandemic thermostable direct hemolysin genes *tdh*A & *tdh*S were present in chromosome-2s of BB220P & RIMD 2210633 genomes; *tdh*-related hemolysin (*trh*) gene was not found among the fives including genome of FDA_R31 strain also, which carried a 7-ureage gene cassette. Three groups of hemolysins were found in the phylogeny tree. Inherited (In) group of hemolysin genes were distinct from the other two groups of hemolysin genes. Among these, one In-group of thermolabilehemolysin and lecithin-dependent hemolysin genes were the *V. parahaemolytcus* marker gene (Taniguchi et al., 1985) and were found in all chromosome-2s as a single copy in each chromosome. Acquired (Ac) group of hemolysin genes, responsible for pandemicity was distributed in genomes of RIMD 2210633, BB220P and UCM-V493 strain (Figure 2. A). Non-grouped hemolysin genes were distributed without any lineage among all fives. These hemolysin genes were mostly strain specific. Few homologous and paired hemolysin genes of this group were present in genomes of FDA_R31 & CDC_K4557 strains. Two major toxin genes, *tdh*A & *tdh*S whose were pandemic responsible (Makino et al., 2003) were limited in RIMD 2210633 and BB220P strains. Acquired thermostablehemolysin delta-VPH genes were present in each chromosome-1 of RIMD 2210633, BB220P and UCM-V493 strains.

Hence, hemolysin gene insertion and deletion happened in both chromosomal nucleotide sequences. UCM-V493 strain showed a place in this development as a midterm one and BB220P strain was the pre-pandemic with pandemic RIMD 2210633 strain in acquiring hemolysin genes concern.

3.2.2. Type three secretion system (TTSS)

167 TTSS genes found in the genome sequences of these five Vp-strains and the resulted evolutionary tree showed that no TTSS gene carried in all steps of pathogenic developing in this species. These TTSS genes were divided into clearly two groups- the TTSS genes of FDA_R31 & CDC_K4557 in one and UCM-V493, BB220P & RIMD 2210633 strains' TTSS genes in another group. Such division depended mainly on TTSS genes of chromosome-1. Seven TTSS genes of Chromosome-2 of RIMD 2210633 were designated as TTSS2 (Makino K et al., 2003; Park et al., 2004) were homologous with the TTSS genes of Chromosome-2 of BB220P (Figure No 2 B). It was noted that 4 TTSS2 genes of RIMD 2210633 were distantly homolog to that of FDA_R31, the ancestral Vp-strain. Therefore, it was distinct that the TTSS (TTSS1 & TTSS2) genes were much dissimilar between TTSS1 genes of FDA_R31 & CDC_K4557 isolates and TTSS2 genes of UCM-V493, BB220P & RIMD 2210633 Vp-strains.

3.2.3. Transposase genes

Total 77 transposase genes showed the evolutionary character in Figure No 2 C. Only FDA_R31 strain contained major (39) transposase genes and these showed in 6 clusters. These >95% identical (CLUSTALW result) gene-clusters pointed out that those genes were repeated or copied intra and inter chromosomally within the genome. Three pairs of transposase genes appeared homologous and distributed within chromosome-2 of RIMD 2210633 and BB220P strains. One asterisk (*) marked cluster of these transposase homologous genes were distributed in RIMD, BB220P and FDA_R31 strains that was remarkable.

3.2.4. *tox*R regulatory genes

The Figure No 2 D yielded the evolutionary motive of total 9 *tox*R genes. It seemed that *tox*R genes were originated from chromosome-1 of UCM-V493 strain and spread to both chromosome-1s of BB220P and RIMD 2210633 strains. The sequence homology (not shown) showed no acceptable liaison between the *tox*R genes of tree branches, 4-8% (CLUSTALW result) homology among branches means the gene pairs were very much insignificant and acquired as distant origin, might be for different purposes. Appearance of *tox*R genes in genome might be the pathogenic indication with the presence of the respective hemolysin gene of this bacterium.

3.3. Study of Hemolysins and other pathogen related genes in comparative genome circles

The toxins and other pathogen-related genes were oriented, very individual mannerin comparative genome circles (Figure No 3). In concentric circles of chromosome-1s, hemolysin genes were scattered, while TTSS genes and transposase genes were mainly clustered. FDA_R31 strain carries more than one such transposase gene clusters in chromosome-1 circle. The hemolysins and other pathogen related genes concerted genomic parts of the respective genome-circle were mentioned as a toxin location (TL) indicated with red arrows (Figure No 3 A& B). Chromosome-1s of five Vp-strains showed that each chromosome contains one TL region containing one thermostablehemolysin gene, 22 TTSS genes and at least 1 transposase gene, except UCM-V493 strain. *Tox*R genes present in distant from TL regions of chromosome-1 of respective circle. The TL regions of RIMD, BB220P and UCM-V493 were at an almost similar position of chromosome-1 (Figure No 3 A).

The chromosome-2 concentric genome circles (Figure No 3 B) also showed the concern gene's position. One TL regions of chromosome-2 of RIMD 2210633 strain was the super-integron or Vpal-7 mentioned in earlier research (Makino et al., 2003, Hurley et al., 2006), and that found in RIMD 2210633 and BB220P strains only. The long 140 kb part was the extended part of 75045 bpTL region of RIMD 2210633 chromosome-2, and it is loaded with two major hemolysin genes (*tdh*A & *tdh*S), two *tox*R genes, 7 TTSS genes and two bunches of terminal transposase genes-4-8 & 9-14. Similar to fragment, enriched with pathogenic genes was also seen in chromosome-2 of BB220P strains. There was one bunch of three hemolysin genes present in the first quarter of chromosome-2s circles of UCM-V493, BB220P and RIMD 2210633 strains. Twelve/thirteen hemolysin genes scattered in chromosome-2s circles of CDC_K4557 and FDA_R31 strains respectively. Chromosome-2 of FDA_R31 strain pointed a TL region on its' 3'-end point that included three hemolysin genes and one cluster of 7 ureage genes (Figure No 5 B). Chromosome-2 of CDC_K4557 strain contained only two TTSS genes and

with no transposase gene, when chromosome-2 of FDA_R31 strain contained 10 TTSS genes and 17 transposase genes mainly at end-terminal position of TL region. The toxin regulatory *tox*R genes were pointed with hemolysin concentric circles. Among total 9 *tox*R genes, 8 genes were limited in RIMD 2210633 and BB220P strains in both chromosomes; each chromosome carried two *tox*R genes, while UCM-V493 strain carried one *tox*R gene on chromosome-1 (Table No 2). The *tox*R genes of chromosome-1s were of no-relation with the respective TL region.

Such presence of hemolysin and pathogen related genes highlighted that the chromosome-1s were some-extend organized manner and these were probably inherited character, whereas chromosome -2 of pandemic RIMD 2210633 and prepandemic BB220P strains acquired especial TL region enriched with more effective toxin genes. The stool isolated FDA_R31 strain was non pandemic nature, but modified for better pathogenic ability with a different TL region containing unknown functional feature of three hemolysin and seven ureage genes at chromosome-2 end point (Figure No 3). This genome-circle study has been illustrated more in the next Mauve study of genetic arrangement.

3.4. Rearrangement study of chromosomal segments and toxin related genes with Mauve 2.3.1

This study performed with the genomic nucleotide sequence in FASTA format and GenBank format by Mauve software. The Mauve results of chromosome-1s and chromosome-2s were figured out in Figure No 4 (A & B) & Figure No 5 (A & B) respectively. The 'Locally Collinear Block's of Mauve were representing as Genome Segments (GSs) in different color to match analog GSs of collinear parts in other chromosomes. The link lines connected between analog GSs of all five strains according to evolutionary serial (RIMD, BB220P, UCM-V493, CDC_K4557 & FDA_R31 strains) of this study. RIMD 2210633 strain was used as referring nucleotide sequence.

3.4.1. Mauve study of chromosome-1s

The nucleotide sequences in FASTA format were used for Mauve alignment of chromosome-1s of these five Vp-strains. Resulted colored blocks were GSs and analogs were linked with lines between. Very small GSs were not visible for their very short length, but magnified results identify them (figure not published). Total 36 GSs were identified in chromosome-1 of RIMD 2210633 strain (Figure No 4 A). The genome sequence of BB220P strain was almost analogous with a reference strain RIMD 2210633 in GS blocks arrangement and following FDA_R31 strain was the third similar. Next, UCM-V493 strain was the fourth similar and CDC_K4557 strain's result was the special, reversely oriented (below the horizontal line) that was most interesting in a genomic arrangement among these five chromosomse-1 sequences. GS blocks were randomly arranged in three non-pandemic strains- FDA_R31, UCM-V493 and CDC_K4557. In FDA_R31 and UCM-V493 strain, A few GS blocks were reversely arranged while most of other GSs were in leading strand. Thus, there were big chaismata formations between nucleotide sequences of chromosome-1s of three non-pandemic strains of *V. parahaemolyticus*.

The TLs of chromosome-1s were identified by red brackets and red arrows were marked for thermostablehemolysin gene (Figure No 4 A). Each chromosome-1 marked with one TL. When Mauve ran with GenBank data, they were more highlighted in zoomed figure to see the nucleotide sequence in the respective TL region, (Figure No 4 B). The colored GSs were mentioning the regional similarity of respective chromosomal segments and serial genomic arrangements were visible. We were interested in hemolysin and pathogen related genes' arrangement, mentioned. Each *V. parahaemolyticus*chromosome-1 contained one TL region with one thermostablehemolysin gene, flaking 122597 bp, 163098 bp, 72487 bp, 126306 bp and 93705 bp region of RIMD 2210633, BB220P, UCM-V493, CDC_K4557 and FDA_R31 strains respectively, as per order of comparative chromosomal circles in Figure No 3 A. TL region of chromosome-1 of RIMD 2210633 strain contained 23 TTSS genes and two transposase genes (Figure 4 B). TTSS genes were located in close cluster up-stream position of thermostablehemolysin delta-VPH gene and two transposase genes were in the down-stream position of TL region. TL region of chromosome-1 of BB220P strain was matched with reference pandemic strain in hemolysin and pathogen related genes' arrangement; 36 TTSS genes were in one close cluster and almost the same position, while two transposase genes were incomplete genes as per BLAST result. UCM-V493 strain carried 34 TTSS genes, but no transposase gene in TL region. TL region of chromosome-1 of CDC_K4557 strain was in 2nd quarter position in comparative circle figure (Figure No 3 A) and it noticed in Mauve study due to the chromosomal inversely arrangement. TL region of CDC_K4557 strain carried 2-23 TTSS genes and only one transposase gene placed in opposite orientation of central thermostablehemolysin gene. TL region of FDA_R31 strain of *V. parahaemolyticus* carried 22 TTSS genes in cluster like other four strains, when 6-10 and 11, 12 transposase genes were placed opposite end part of central thermostablehemolysin gene position (Figure No 4 B).

These hemolysins and pathogen related genes' arrangements were mostly similar pattern, except that of CDC_K4557 strain, due to the reversely arrangement of chromosomal-1 nucleotide sequence. Five TL regions carry only essential toxin genes might be to survive in an environment.

3.4.2. Mauve study of chromosome-2s

The Mauve study of chromosome-2s nucleotide sequences of five *V. Parahaemolyticus*in genetic formation and arrangement were very interesting because of the major pandemic hemolysins or hemolysins and other pathogen related genes were placed in respective chromosomal part. Even non-pandemic Vp-strains were also of great importance to observe these arrangements of GSs. The chromosome-2s Mauve result showed a much less number of chaismata, formed by the GS-link lines (Figure No 5 A). Chromosome-2s of RIMD 2210633 and BB220P were almost homologus in GS arrangement pattern. Chromosome-2 of RIMD 2210633 strain was segmented into 19 parts, only one segment was invisible at first Mauve figure (Figure

No.-5 A). The chromosomal-2 sequence of CDC_K4557 strain, unlikely with chromosome-1 sequence was with leading strand and the GSs was more similarly arranged as per GS sequences of pandemic RIMD 2210633 strain (result not shown). GSs of chromosome-2s sequences of UCM-V493 and FDA_R31 strains were mostly in leading strand, except a few.

The interesting TL regions were identified on chromosome-2s of RIMD 2210633 and BB220P strains in the same position with 75045 bp and 76367 bp flanking region respectively (Figure No 5 B). RIMD 2210633 and BB220P strains of *V. Parahaemolyticus* carry only specific TL regions containing each two thermostable direct hemolysin genes (*tdh*A&*tdh*S) with two *tox*R or *tox*R activated genes, whereas FDA_R31 strain carries a special TL-like region (61853 bp flanking region) at 3'-end position of chromosome-2. This flanking region contains three hemolysin genes, seven ureage genes in a single cluster (without *trh*gene), 6-10 TTSS genes, 13-17 transposase genes and two transcriptional regulatory genes (genes not indicated in Figure No 5 B).

This region does not contain any major *tdh/trh* gene, instead there were three hemolysin genes whose importance were yet to be prove in pathogenesis of *V. parahaemolyticus* .

The Mauve observation indicated that the colored GSs were arranged and turned to specific pandemic form. Chromosome-1s' GSs were more fragmented and more random-arranged than that of the chromosome-2s' GSs of non-pandemic strains. GSs of chromosome-1 of CDC_K4557 strain was inversely oriented was also especial case. On the other hand, chromosome-2 of it was closely related to pandemic orientation in GS pattern.

The zoomed Mauve figures of two chromosomes were more interesting in the study of hemolysins and other pathogen related genes' arrangement following same colored GSs. The toxin and pathogen related genes maintained more or less similar sequence in case of chromosome-1s and carrying one thermostablehemolysin/thermostablehemolysin delta-VPH gene in middle position. The respective minor genomic sequence changes were notable, might be due to the evolutionary effect to pandemic conversion. TTSS genes were point of importance to consider that genomic arrangement sequence was not very different, but in phylogeny result yielded their distant homology, especially in pandemic and non-pandemic division (Figure No 2 B).

The hemolysins and other pathogen related genes of chromosome-2s were limited within the TL region of RIMD 2210633 and BB220P strains in confined pattern. But the 3'-end part of chromosome-2 of FDA_R31 strain carried a special TL region with non-interrupted seven ureage genes, and three hemolysin genes of unknown function in flanking region. It seems that the genomic sequence of the LT region of chromosome-2 of RIMD 2210633 strain was just copy and paste in BB220P strain with all said pandemic genes (Figure No 5 B). Therefore, it might be the best development of pandemicity of this pathogen in the environment.

4. Discussion

In this study, we selected five full genome sequences of *V. parahaemolyticus* with the genomic toxins and pathogen related genes' sequential arrangement and modification in a view of non-pandemic form of pandemic development in this bacterium. The phylogeny study of both chromosomal sequences pointed a clear evolutionary development that initiated from FDA_R31 strain isolated from stool sample from a USA patient. UCM-V493 strain of sea sediment of Spain carried intermediate modification. This genome placed in the middle of chromosome-2s' UPGMA tree and top-second in neighbor-joining tree of chromosome-1s of *V. parahaemolyticus.* The two chromosomal pathogenic developments of *V. parahaemolyticus* were not parallel happened. Chromosome-1s were comparatively conservative than chromosome-2s for pandemic adaptation. The comparative chromosomal figures of chromosome-1 of these five strains (Figure No 3) provided the same result that TL regions were placed relatively in the same position except CDC_K4557 strain's chromosome-1 sequence, due to the reverse complementary genomic arrangement (Figure No 4 A). The colored GSs result following GS-link lines showed in Mauve of both chromosomes were identical without major loss or gain, but very randomly arranged in non-pandemic UCM-V493, CDC_K4557 and FDA_R31 strains. The genomic arrangements of chromosome-1 of CDC_K4557 strain, an oyster isolate showed a very interesting pattern of reverse complement, which is the cause of major chiasmata formation of GS-link lines at both sides with other chromosomes (Figure 4 A). Pandemic RIMD 2210633 and prepandemic BB220P strains were homologous in nucleotide sequence and identical in GS matching in the serial arrangement (Figure No 4 A& 5 A). This study has been revealed that the two Vp-strains of stool sample isolated (RIMD 2210633 & FDA_R31) are in a same genomic situation, but opposite position in phylogeny tree. Genome of FDA_R31 strain carried the non-pandemic genomic position like environmental strain, while the genome of RIMD 2210633 strain, the compact pathogenic development against host, named as pandemic. The both strains might cause human health hazard, but in different potential. In both genomic developments of RIMD 2210633 and FDA_R31 strains, there were very distinct in pandemicity, but well conserves position of housekeeping genes in genome segments. The genome sequence of BB220P strain, may named as pre-pandemic is an environmental isolated from Bangladesh in early 1980 (Jensen et al., 2013). It carried very identical pattern in DNA sequence of both chromosomes with pandemic and reference strain of *V. Parahaemolyticus* RIMD 2210633. Itis also an example of close pandemic development in nature happened before 1980, far beyond of pandemic identification of *V. Parahaemolyticus* on and after 1995 (Nair G B et al., 2007).

The TL regions were detected in all chromosome-1s sequences in mostly similar pattern of toxin genes, including hemolysin, TTSS and transposase genes. Therefore, these special regions may perform some global role in the bacterial defence mechanism that is essential to survive for *V. parahaemolyticus*in nature. The evolutionary study results that thermostablehemolysin or thermostablehemolysin delta-VPH and transposase genes are carried out from

environmental strains to the pandemic strain, likely inherited way. TTSS genes are clearly divided in two groups, one of non-pandemic category (CDC_K4557 and FDA_R31 strains), and other, close to pandemic category (UCM-V493, BB220P and RIMD 2210633 strains) (Figure No 2 B). The pandemic category TTSS genes may be best DNA-level modified for better toxicity against the host. The higher number of transposase genes and the concern homologs spread in both chromosomes of FDA_R31 strain (Figure No 2 C) may be to increase the acceptability of foreign genes into the genome. The lower number of transposase gene present in another four genomes may be due to gradual saturation of foreign gene insertion for essential development, following pandemicity. The *tox*R, toxin regulatory genes of chromosome-1s of RIMD 2210633 and BB220P strains present in distant position from the respective TL region (Figure 3 A) that may be later inserted into genomes.

Comparative studies of the GS arrangement pattern of chromosomes are very significant to understand concern development at the gene level even. For the development of pathogenicity or pandemicity of *V. parahaemolyticus*, chromosome-2s play the major role and that observed in this study. Chromosoe-2s of pandemic RIMD 2210633 strain is very much close with BB220P strain, including the TL region and GSs of other three strains are randomly arranged (Figure 5 A). The TL regions were pointed in RIMD 2210633 and BB220P strains with extended 140391 bp (extended TL region of 75045 bp) and 76367 bp region respectively. Other stool isolated FDA_R31 strain carries a 61853 bp TL region with different genetic inclusion from that of RIMD 2210633 strain. Earlier research had been found out seven pathogenic islands (VpaI 1-7) in RIMD 2210633 pandemic strain of *V. Parahaemolyticus* (Hurley et al., 2006) and this VpaI-7 was the responsible super integron caused for the bacterial pandemic character (Makino K et al., 2003). In this study, the expanded Vpal-7, TL region of RIMD 2210633 strain carries two hemolysin genes, two *tox*R/*tox*R activated genes, 7 TTSS genes and 11 transposase genes of which 8 genes are noted as part transposase gene, while the LT region of BB220P strain carries 7 complete transposase genes with other genes as that of RIMD 2210633 strain (Figure No 5 B). To find out the functional cause of this minor variation in TL regions of chromosome-2s of RIMD 2210633 and BB220P strains, it needs genomic comparison of more new pandemic strains. Our previous study provided few different characters of major pandemic toxin (*tdh*) genes present in *V. parahaemolyticus* of different source and years (Bhowmik et al., 2014). We also explained the special genomic level developments in the promoter region of thermostable direct hemolysin S (*tdh*S) gene of RIMD 2210633 strain with that genre of BB220P strain. Although, the pre-pandemic BB220P strain is almost pandemic in genome development, still more basic developments might be essential in *tdh*A/*tdh*S gene level modification to achieve pandemic standard (Bhowmik et al., 2014).

The other TL region that pointed in chromosome-2 of FDA_R31 strain (Figure 5 B) is a point of our interest. This region carries three hemolysin genes of unknown function and 7 ureage genes in one cluster with 4 TTSS and 4 transposase genes. There is no presence of *tdh*-related hemolysin (*trh*) gene in this FDA_R31 genome, when *trh* gene always accompanied with this cluster of ureagegenes reported earlier study (Park et al., 2000). This stool isolated Vp-strain may be a representative of *trh* positive strain's character (ureage+) without this gene. It is may be a stage of *V. Parahaemolyticus* of elimination or to be inserted of *trh* gene. This very noble region found in the terminal region of 3'-end of chromosome-2 of FDA_R31 strain of *V. Parahaemolyticus* also needs for more study.

5. Conclusion

5.1. Phylogeny study

The complete and whole genome sequences of *V. parahaemolyticus* are used chromosome-wise to follow the chronological genomic flow.

1) Chromosome-2s are the major pandemic importance to genomic development and Figure 1 C is more acceptable for phylogenies chronology. FDA_R31 strain may be the ancestor and RIMD 2210633 strain is up-to date product of these fives. Chromosome-1 of FDA_R31 strain is also the ancestor and Chromosome-1s of all fives are closer with minor diversions.
2) There are three groups of hemolysin-genes in Vp, 1) 'Inherited', the genes are present in pandemic, pre-pandemic and non-pandemic strains, 2) 'Acquired' hemolysin genes are present in pandemic/pre-pandemic genomes only, and 3) other hemolysin genes represent individually without any lineage.
3) TTSSs are fixed and specified, unlikely inherited hemolysin genes. TTSS1 genes are carryed on chromosome-1s, but in two sub-sets. Chromosome-1s of RIMD, BB220P and UCM-V493 strains carry one set of TTSS1 genes, and CDC_K4557 and FDA_R31 strains carry another set of TTSS1 genes. Chromosome-2s of RIMD, BB220P & UCM-V493 strains contain TTSS2 that is absent in other two strains. It is a point of interest that RIMD 2210633, BB220P & UCM-V493 strains carry two sets of TTSS genes and these genes are absent in the ansector two strains.
4) The ancestral FDA_R31 strain carries more or less six clusters of transposase genes whose are >95% (CLUSTALW result) homologous within the cluster and one copy of this gene may be the origin of intra and extra-chromosomal rearrangement. The asterisk (*) group of transposase genes is noted because of their inherited presence in TL of chromosome-2s of pandemic or pre-pandemic strains with the ancestral FDA_R31 strain only.

5.2. The study with comparative chromosome circles, the supportive data of Mauve that yields-

1) The TL regions are identified with closely located hemolysins and other pathogen related genes.
2) TL region of chromosome-1 of CDC_K4557 strain has shown in different places in genome circle due to inversely orient genome sequence (Mauve study).

3) TLs are marked in only chromosome-2s of RIMD 2210633 and BB220P strains and it is newly appeared in comparison with other genomes (Mauve study).
4) The presence of special TL region at terminal end of chromosome-2 of FDA_R31 strain, presence of 7 ureage genes without *trh* gene.

5.3. The Mauve study yields

1) There are two major findings- pandemic RIMD 2210633 and pre-pandemic as well as environmental BB220P strains are much closer in GSs arrangement, whereas non-pandemic other three strains are very much random in GS rearrangement.
2) The GSs are moving within non-pandemic chromosomal positions and insertion or deletion of GS is not prominent in both chromosomes. But, the TL regions in chromosome-2s of RIMD 2210633 and BB220P strains are the newly adapted genomic parts, maybe for pandemic development of this species. There is a noble TL region at terminal end of chromosome-2 of FDA_R31 strain.
3) The chromosome-2s of pandemic or pre-pandemic strains only carry special TL regions probably acquired and also guarded by two sets of terminal transpsase genes, and these genes are homologous.
4) The GSs of CDC_K4557 strain are identified inversely oriented, that is very much remarkable in a Genomic Rearrangement event within this species, while GSs of chromosome-2 is on the leading strand as well.

6. Acknowledgement

The author would like to thank to scientists of National Institute of Cholera and Enteric Diseases (ICMR), Dr. T. Ramamurthy, and Dr. Ashish Mukherjee for helping me a lot from this study. I am especially indebted to Dr. Nabendu Chaterjee of Biochemistry division for his immense help and contribution in language correction of this article.

References

[1] Bhowmik SK, Pazhani GP, Ramamurthy T. Phylogenic and in silico functional analyses of thermostable-direct hemolysin and tdh-related encoding genes in *Vibrio parahaemolyticus* and other Gram-negative bacteria. Biomed Res Int. 2014; 2014:576528. doi: 10.1155/2014/576528. Epub 2014 Jul 8.
[2] Chen Y, Stine OC, Badger JH, Gil AI, Nair GB, Nishibuchi M, Fouts DE. Comparative genomic analysis of *Vibrio parahaemolyticus*: serotype conversion and virulence. BMC Genomics, 2011 Jun 6; 12:294. doi: 10.1186/1471-2164-12-294.
[3] Fujino T, 1951. Bacterial food poisoning. SaishinIgaku, 6: 263-271
[4] Hurley CC, Quirke A, Reen FJ, Boyd EF. Four genomic islands that mark post-1995 pandemic *Vibrio parahaemolyticus* isolates. BMC Genomics, 2006 May 3; 7:104.
[5] Jensen RV, Depasquale SM, Harbolick EA, Hong T, Kernell AL, Kruchko DH, Modise T, Smith CE, McCarter LL, Stevens AM. Complete Genome Sequence of Prepandemic*Vibrio parahaemolyticus* BB22OP, Genome Announc., 2013 Jan; 1(1). pii: e00002-12. doi: 10.1128/genomeA.00002-12. Epub 2013 Feb 21.
[6] Kalburge SS, Polson SW, Boyd Crotty K, Katz L, Turnsek M, Tarr CL, Martinez-Urtaza J, Boyd EF. Complete Genome Sequence of *Vibrio parahaemolyticus* Environmental Strain UCM-V493. Genome Announc. 2014 Mar 13; 2(2). pii: e00159-14. doi: 10.1128/genomeA.00159-14.
[7] Lin Z, Kumagai K, Baba K, Mekalanos JJ, Nishibuchi M. *Vibrio parahaemolyticus* has a homolog of the *Vibrio choleraetox*RS operon that mediates environmentally induced regulation of the thermostable direct hemolysin gene. J Bacteriol., 1993 Jun; 175(12):3844-55.
[8] Makino K, Oshima K, Kurokawa K, Yokoyama K, Uda T, Tagomori K, Iijima Y, Najima M, Nakano M, Yamashita A, Kubota Y, Kimura S, Yasunaga T, Honda T, Shinagawa H, Hattori M, Iida T. Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism distinct from that of V cholerae. Lancet. 2003 Mar 1; 361(9359):743-9.
[9] Nair GB, Ramamurthy T, Bhattacharya SK, Dutta B, Takeda Y, Sack DA. Global dissemination of *Vibrio parahaemolyticus* serotype O3:K6 and its serovariants.ClinMicrobiol Rev. 2007 Jan; 20(1):39-48. Review.
[10] Ono T, Park KS, Ueta M, Iida T, Honda T. Identification of proteins secreted via *Vibrio parahaemolyticus* type III secretion system 1., Infect Immun., 2006 Feb; 74(2):1032-42.
[11] Park KS, Iida T, Yamaichi Y, Oyagi T, Yamamoto K, Honda T. Genetic characterization of DNA region containing the *trh* and *ure* genes of *Vibrio parahaemolyticus*. Infect Immun., 2000 Oct; 68(10):5742-8.
[12] Park KS, Ono T, Rokuda M, Jang MH, Okada K, Iida T, Honda T. Functional characterization of two type III secretion systems of *Vibrio parahaemolyticus*. Infect Immun., 2004 Nov; 72(11):6659-65.
[13] Tada J, Ohashi T, Nishimura N, Shirasaki Y, Ozaki H, Fukushima S, Takano J, Nishibuchi M, Takeda Y. Detection of the thermostable direct hemolysin gene (*tdh*) and the thermostable direct hemolysin-related hemolysin gene (*trh*) of *Vibrio parahaemolyticus* by polymerase chain reaction. Mol Cell Probes., 1992 Dec; 6(6):477-87.
[14] Takeda Y, Honda T, Miwatani T. [Biological activity and receptor of thermostable direct hemolysin of *Vibrio parahaemolyticus* (author's transl)]TanpakushitsuKakusanKoso. 1976 Nov; Suppl: 109-20. Review.
[15] Taniguchi H, Ohta H, Ogawa M, Mizuguchi Y. Cloning and expression in Escherichia coli of *Vibrio parahaemolyticus*thermostable direct hemolysin and thermolabilehemolysin genes. J Bacteriol., 1985 May; 162(2):510-5.
[16] Terai A, Baba K, Shirai H, Yoshida O, Takeda Y, Nishibuchi M. Evidence for insertion sequence-mediated spread of the thermostable direct hemolysin gene among Vibrio species. J Bacteriol., 1991 Aug; 173(16):5036-46.

[17] Terai A, Shirai H, Yoshida O, Takeda Y, Nishibuchi M. Nucleotide sequence of the thermostable direct hemolysin gene (*tdh* gene) of *Vibrio mimicus* and its evolutionary relationship with the *tdh* genes of *Vibrio parahaemolyticus*. FEMS MicrobiolLett., 1990 Sep 15; 59(3):31923.

Table No 1

Strain	Source/year	Length of sequence (bp)		Hemolysin genes (83)		TTSS (1&2) genes (167)		Transposase genes (77)		Toxr genes (9)	
		Chro.-1	Chro.-2	Chro.-1	Chro.-2	Chro.-1	Chro.-2	Chro.-1	Chro.-2	Chro.-1	Chro.-2
RIMD 2210633	Clinical (stool)/1996	3288558	1877212	7	7	23	7	2	14	2	2
BB22OP	Environmental/1980	3297305	1806219	8	6	36	8	3	9	2	2
*UCM-V493	Sea sediment/2002	3445790	1698252	6	3	34	1	5	1	1	0
CDC K4557	Oyster /2007	3331580	1806995	11	12	23	2	4	0	0	0
FDA_R31	Clinical (stool)/2006	3362228	1861334	10	13	23	10	22	17	0	0
Total				42	41	139	28	36	41	5	4

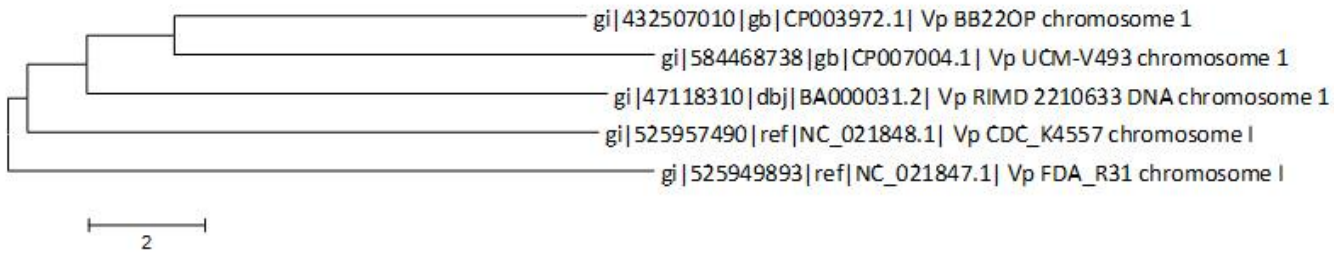
Genomes *Vibrio parahaemolyticus* strains with numbers of toxins and other pathogen related genes
*UCM Vp strain only carried a plasmid of 88530 bp nucleotide sequence that contained no hemolysin or other pathogen related gene.

Table No 2 List of hemolysin and other pathogen related genes of five *Vibrio parahaemolyticus* genomes
A Chromosome-1s

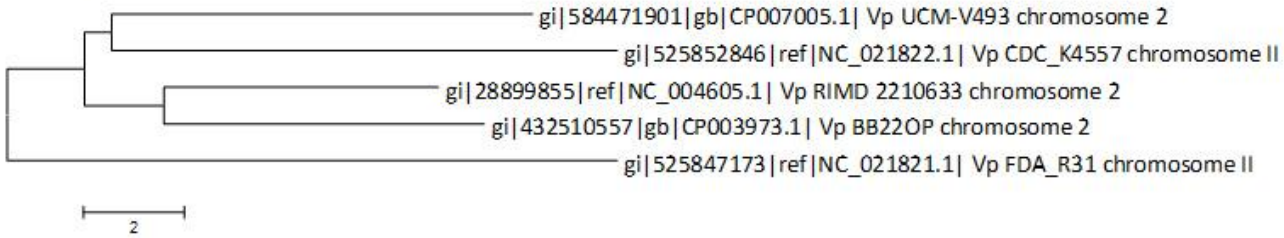
Name of strains → / Category of genes ↓	RIMD(gi47118310) Flanking region & Gene's product (Toxin Location: 1779438-1902035=122597 bp)	BB220P(gb|CP003972.1) Flanking region & Gene's product (Toxin Location: 1722306-1885404=163098 bp)	UCM(gb|CP007004.1) Flanking region & Gene's product (Toxin Location: 1869299-1941786 =72487 bp)	CDC(gb|CP006008.1) Flanking region & Gene's product (Toxin Location: 1013491-1139797 = 126306 bp)	FDA(gb|CP006004.1) Flanking region & Gene's product (Toxin Location: 1818889-1912594=93705 bp)
Hemolysin genes	1: 372883-374631 "putative hemolysin" 2: 447419-448120 "putative hemolysin" 3: 766326-767225 "putative hemolysin" 4: 933540-934892 "putative hemolysin" 5: 1844875-1845486 "thermostable hemolysin delta-VPH" 6: 2674301-2675581 "putative hemolysin" 7: 3255312-3255968 "putative hemolysin III"	1: 392177-393925 "Putative hemolysin" 2: 443740-444342 "21 kDa hemolysin precursor" 3: 764007-764906 " putative hemolysin" 4: 937519-938871 " putative hemolysin" 5: 1792603-1793214 "thermostable hemolysin delta-VPH" 6: 2639613-2640893 "Hemolysins-related proteins containing CBS domains" 7: 2885810-2886412 "Hemolysins-related proteins containing CBS domains" 8: 3238172-3238825 "putative membrane protein hemolysin III-like protein"	1: 411435-413183 "Putative hemolysin" 2: 461077-461673 "21 kDa hemolysin precursor" 3: 1090060-1091379 "Hemolysins-related protein containing CBS domains" 4: 1941175-1941786 "Thermostable hemolysin delta-VPH" 5: 2837016-2838296 "Hemolysins-related protein containing CBS domains" 6: 3412547-3413200 "putative membrane protein hemolysin III – like protein"	1: 288439-289554 "hemolysin D" 2: 1073737-1074348 "thermostable hemolysin" 3: :1453647-1454510 "hemolysin D" 4: 1662446-1663531 "hemolysin secretion protein D" 5: 1894288-1895382 "hemolysin D" 6: 1939782-1941134 "hemolysin" 7: 2456391-2457056 "hemolysin" 8: 2516882-2517988 "hemolysin D" 9: 2529320-2531068 "hemolysin" 10: :2955492-2956145 "hemolysin" 11: 3130758-3131885 "hemolysin"	1: 205205-206332 "hemolysin" 2: 407668-408321 "hemolysin" 3: 761385-763133 "hemolysin" 4: 822486-823193 "hemolysin" 5: 1331086-1332438 "hemolysin" 6: 1376853-1377947 "hemolysin D" 7: 1613253-1614338 "hemolysin secretion protein D" 8: 1849595-1850206 "thermostable hemolysin" 9: 2223297-2224157 "hemolysin D" 10: 3064158-3065273 "hemolysin D"

Figure No 1

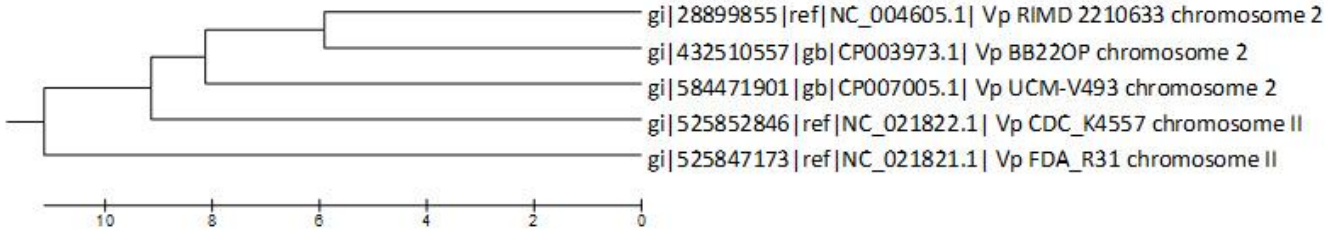
A Neighbour-joining tree of Vp-1 (5 chromosome-1s)



B Neighbour-joining tree of Vp-2 (5 chromosome-2s)

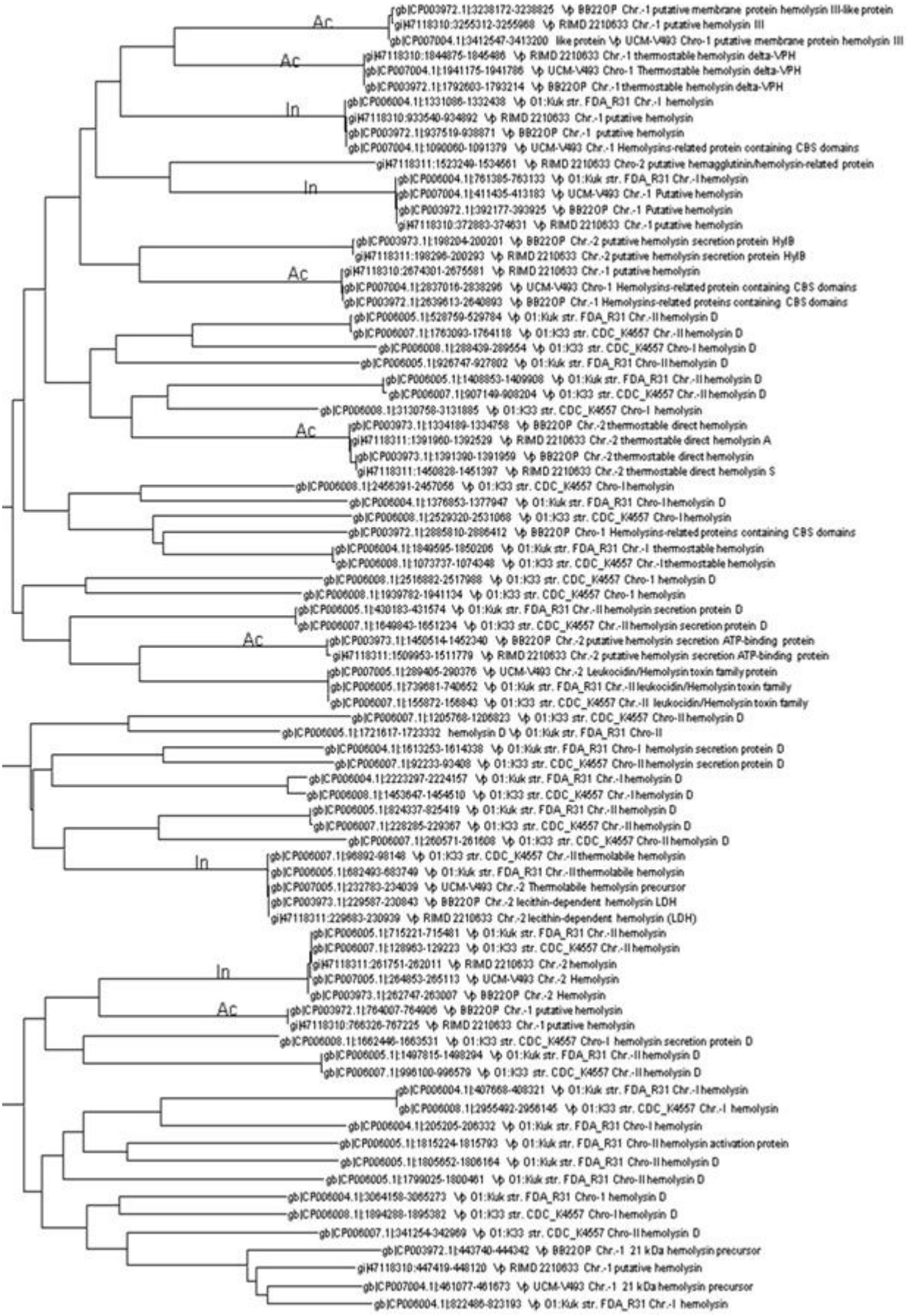


C UPGMA tree of Vp-2 (5 chromosome-2s)



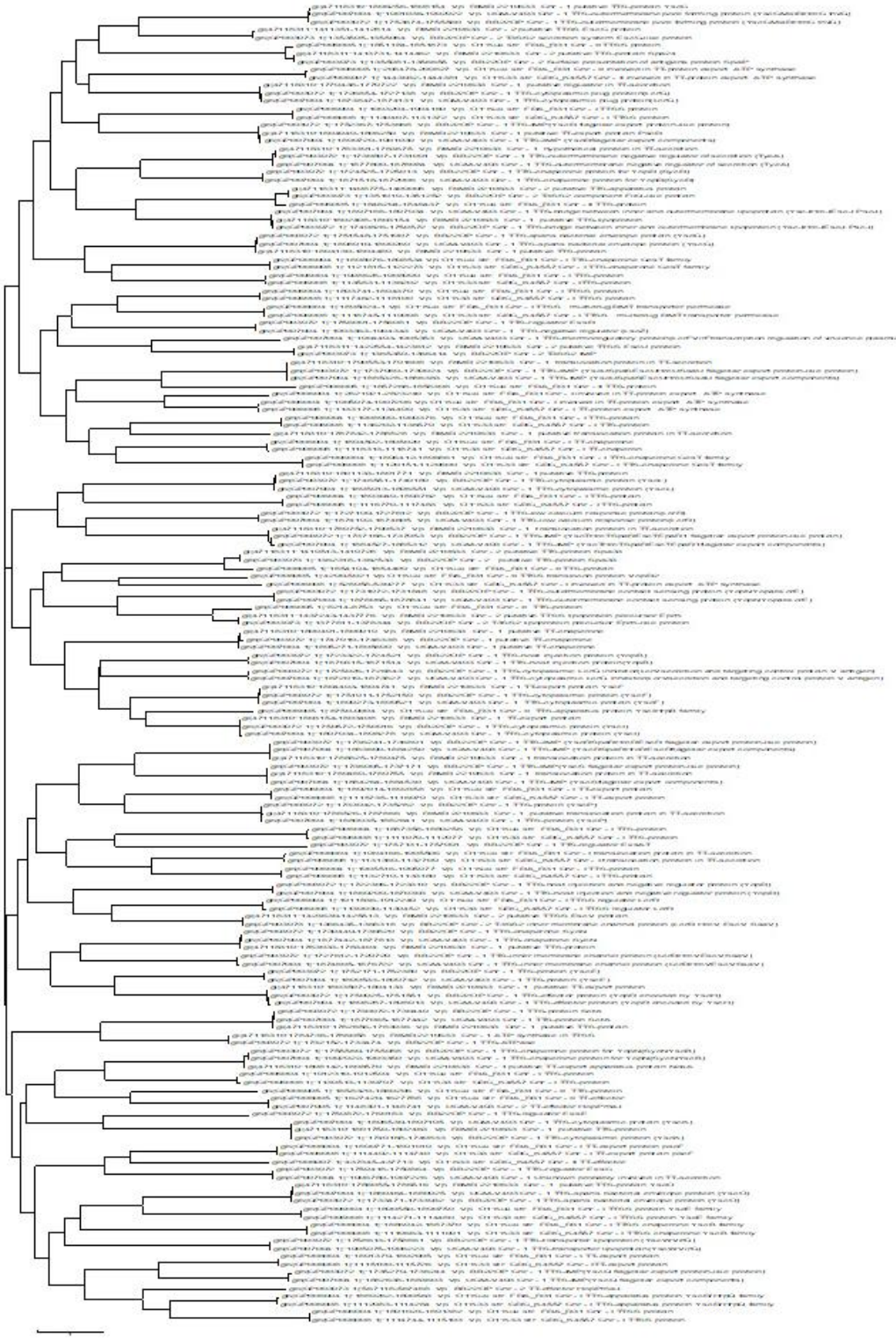
Neighbour-joining tree of A & B result is similar as chromosome-1s and chromosome-2s of FDA_R31 and RIMD 2210633 are the evolutionary resource and mid-term respectively in development. UPGMA tree result of C of chromosome-2s makes a serial profile of fives as RIMD 2210633, BB220P, UCM-V493, CDC_K4557 & FDA_R31 in a descending order respectively.

Figure 2 A: Neighbor-Joining tree of 83 hemolysin genes



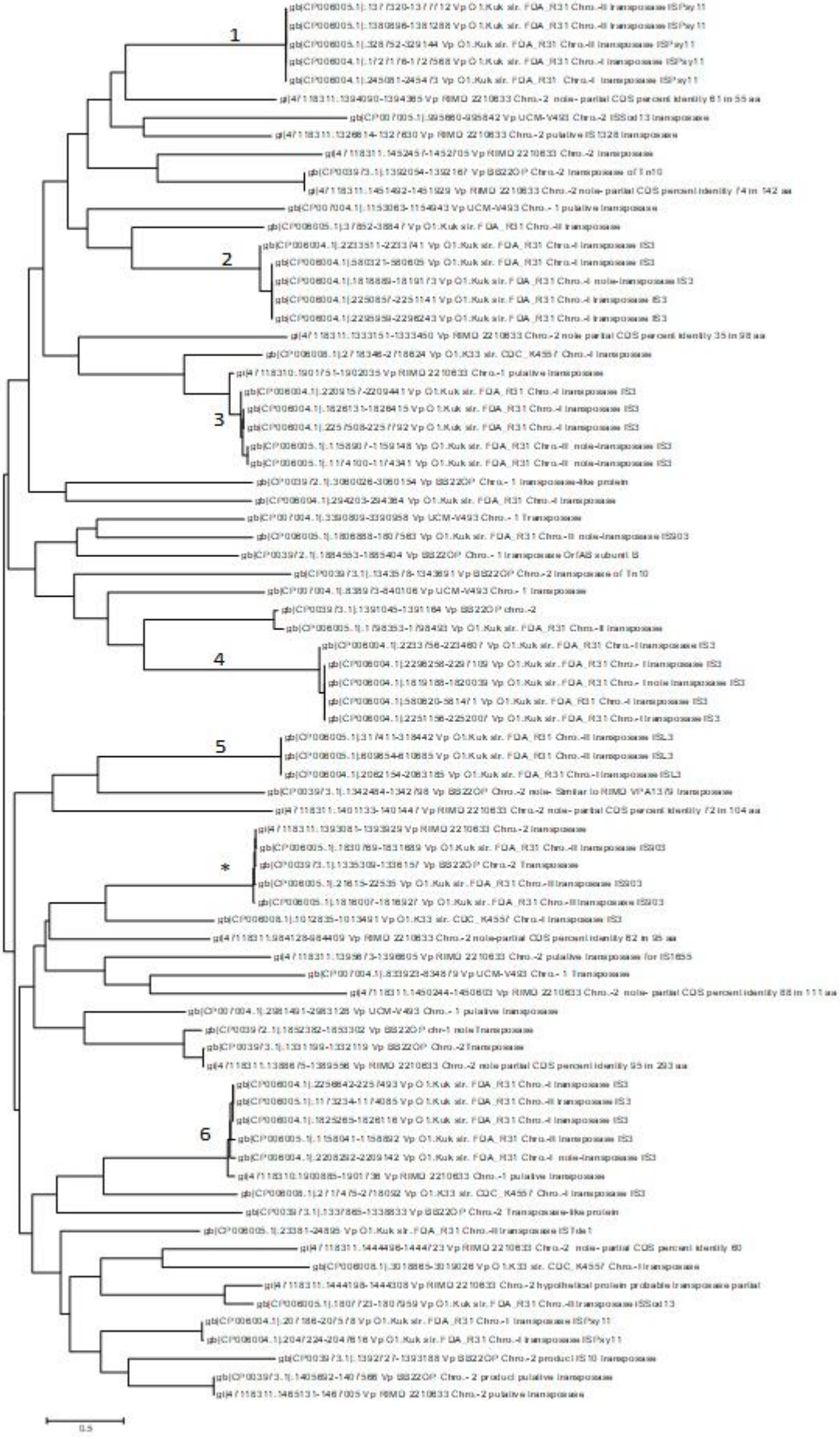
This phylogenetic tree reveals three group of hemolysin-genes in Vp- Inherited (In), the genes are present in both pandemic and nonpandemic strains, Acquired (Ac) genes are placed in pandemic genomes only and numbers of hemolysin genes are present in mostly in single strains' genome without any lineage.

Figure 2 B: Neighbour-Joining tree of 167 TTSS genes



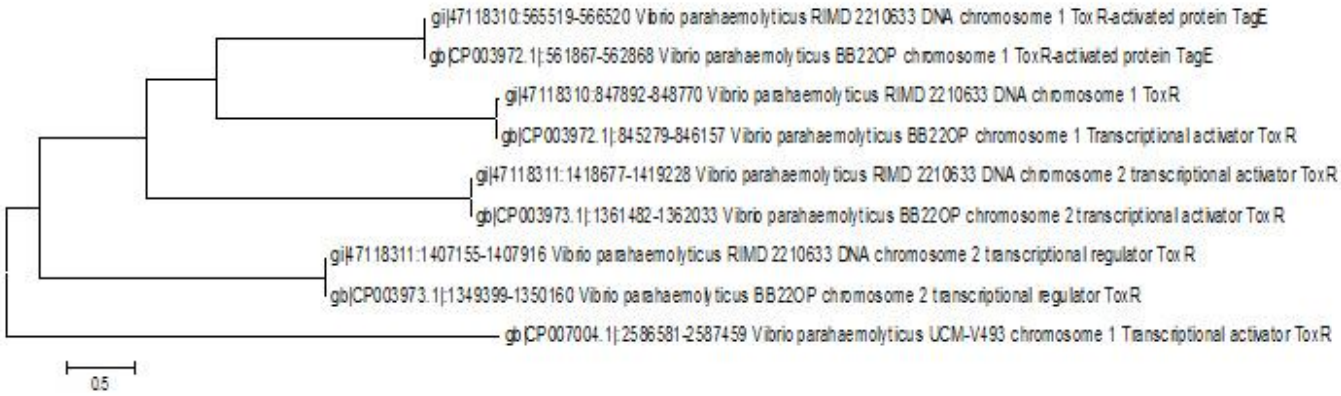
TTSS genes of chromosome-1s and chromosome-2s of CDC & FDA strains' genomes are almost distict from TTSS genes of other thee genomes. The TTSS1 & TTSS2 genes of pandemic strain RIMD are almost homologus with TTSS genes of BB220P & UCM genomes.

Figure 2 C: Neighbour-joining tree of 77 TRPS genes



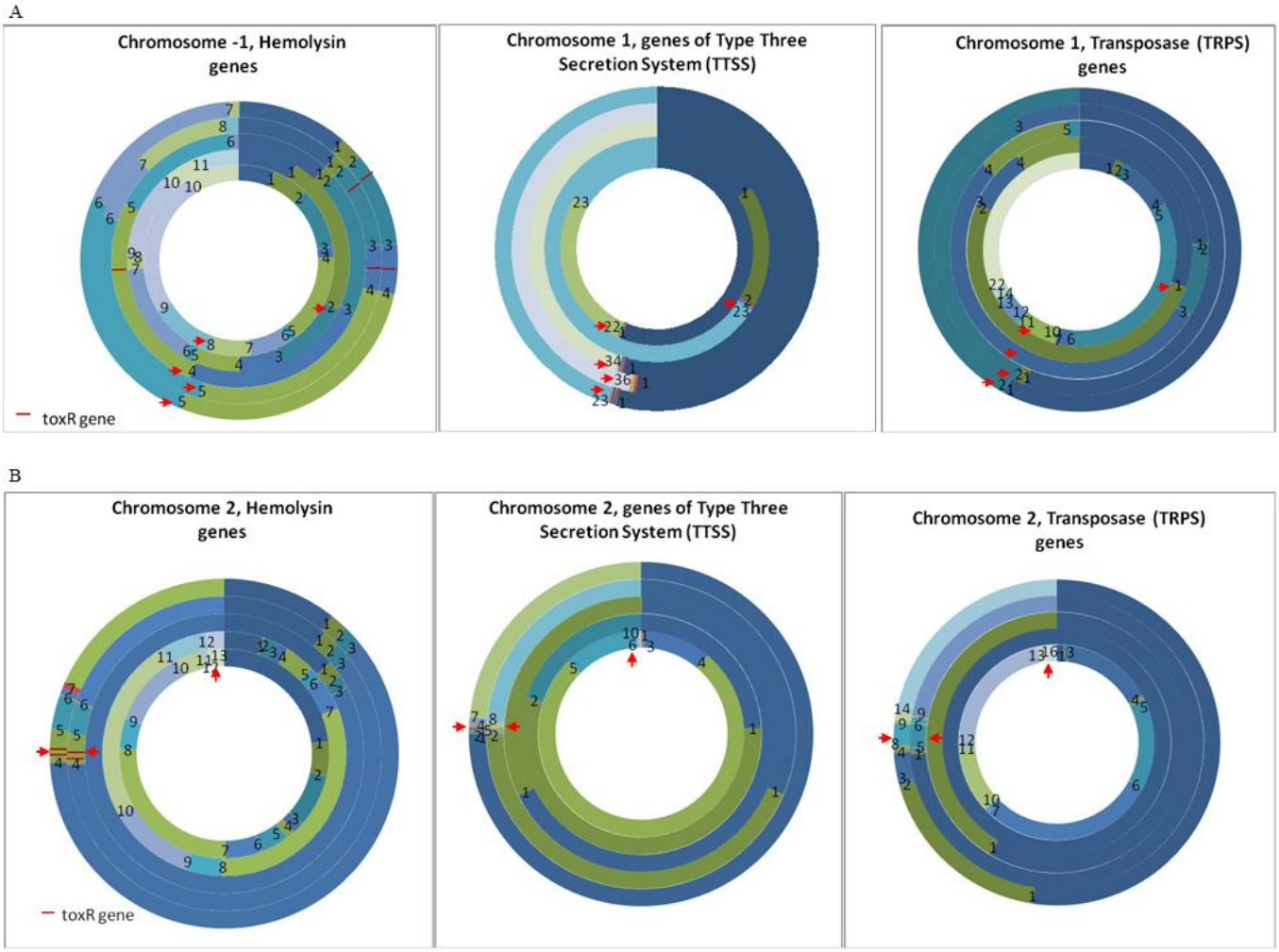
Vp FDA strain carries six cluster of TRPS genes whose are >95% nucleotide sequence homologous within cluster and they are spread intra and extra-chromosomally. Other TRPS genes of asterisk (*) group are noted because of their presence in TL region of RIMD & BB220P and in environmental FDA strain also.

Figure 2 D: Neighbour-joining tree of 9 *tox*R genes



*Tox*R gene of UCM strain has no significant homology with that of RIMD & BB220P strains whose are may be laterally transferred to both chromosomes.

Figure 3: Comparative circular presentation of hemolysin and other pathogen related genes of RIMD 2210633, BB220P, UCM-V493, CDC_K4557 and FDA_R31 strains are from the outside inwards in co-centric circles

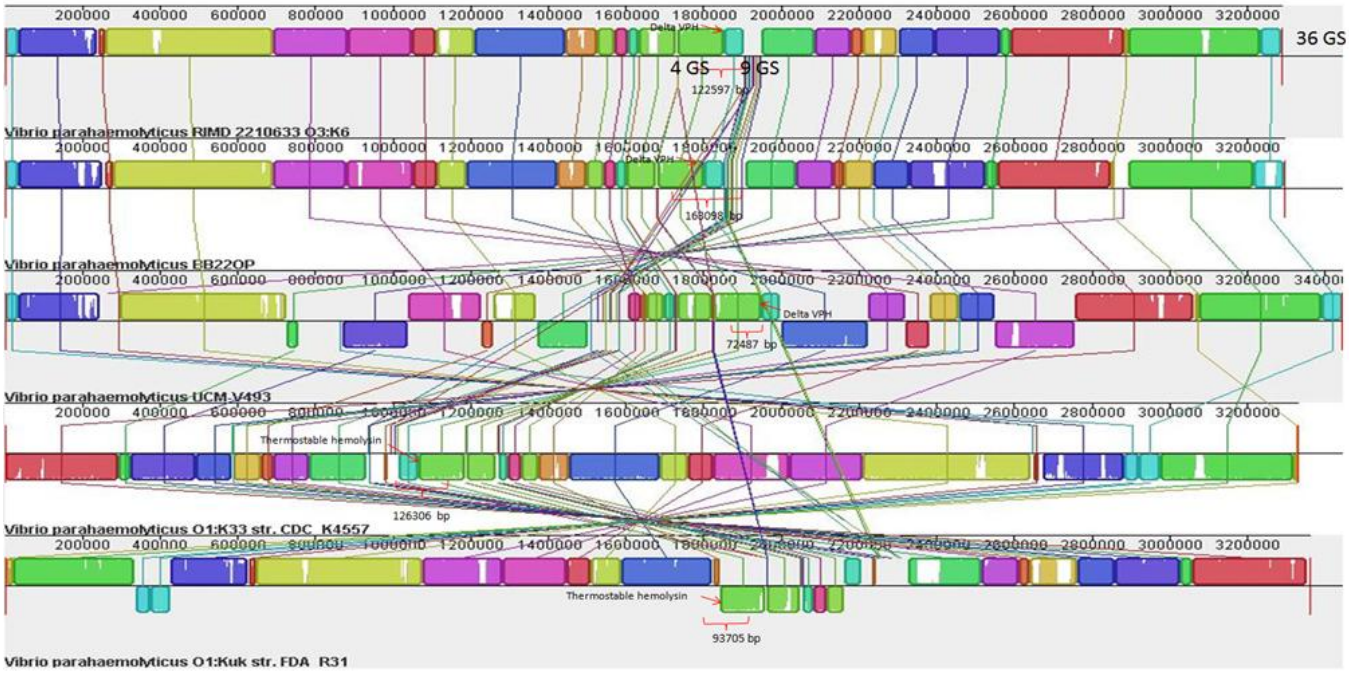


The respective genes' serial numbers pointed the position. In congestions of genes' position the first and last numbers are pointed. Changes of colours are as per excel-program. Red-arrow indicates the point of TL region. Red-line indicates the position of *tox*R gene. A & B –The co-centric circulars of chromosome-1s & chromosome-2s respectively.

(The original format is 'Landscape' size)

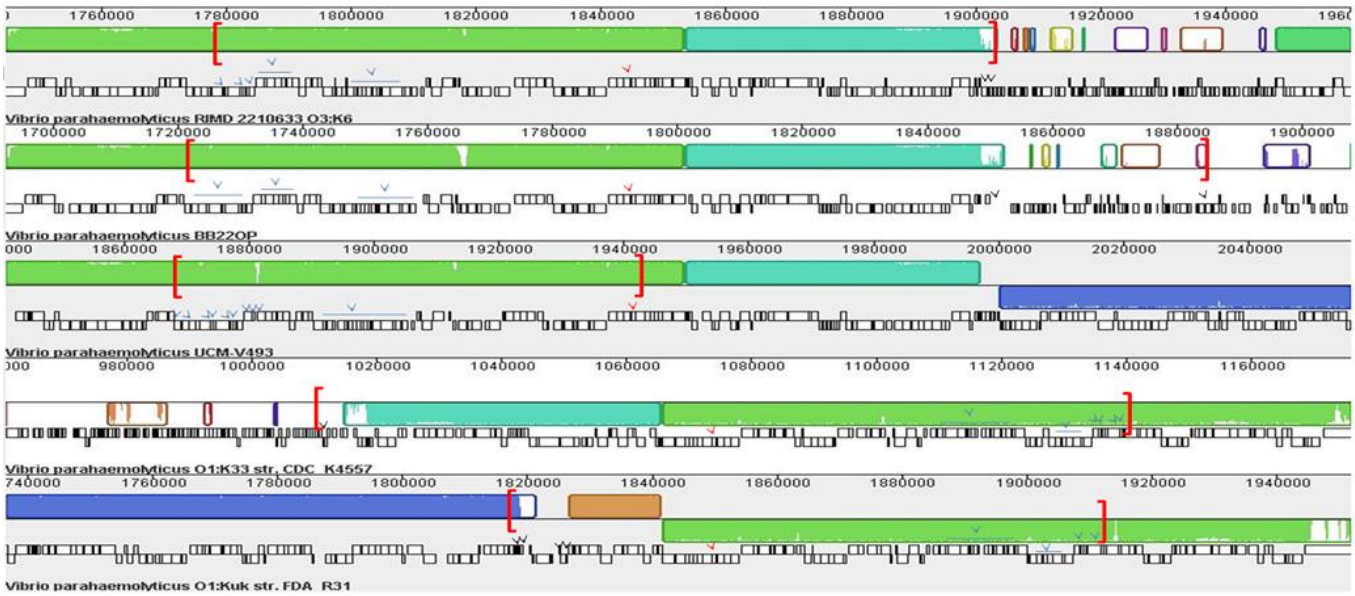
Figure 4: Arrangements of Genomic segment (GS) and the toxin and other pathogen related genes in TL regions with Mauve

A Vp chromosome-1s



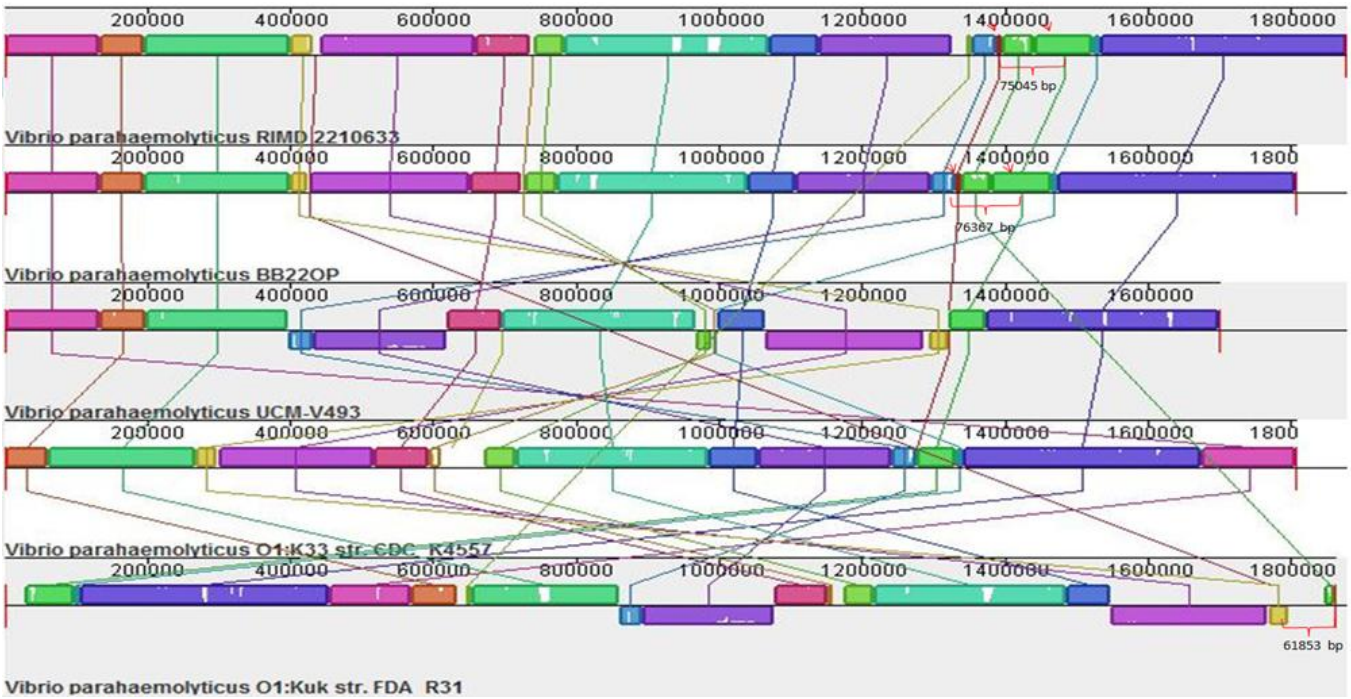
Mauve result with nucleotide sequences in FASTA-format. GSs are depicted in colors. The positions of GSs are placed according to genetic leading and lagging strand. Genome of CDC strain shows in lagging strand in full (below the horizotal line) and massive chaismata formation of GS link lines are due to the genome of CDC strain. TLs are indicated by red-brakets with flanking region and thermostable hemolysin/thermostable Delta VPH (Delta VPH) in red-arrows.

B Vp chromosome 1s expanded to show gene of TL regions marked with red brakets



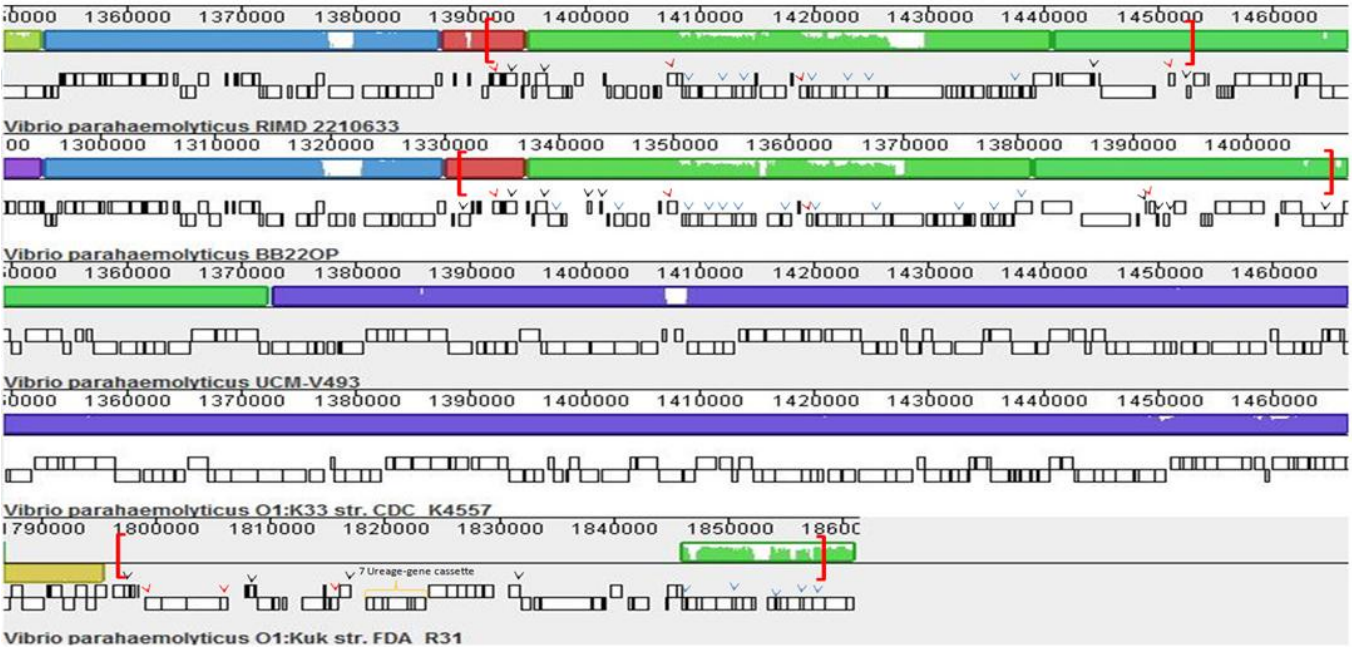
Mauve result with GenBank files. The genomic span TL regions are 122597 bp-RIMD, 163098 bp-BB220P, 72487 bp-UCM, 126306 bp-CDC & 93704 bp-FDA with one hemolysin gene (red arrow), only two TRPS genes (black arrow) are in away of downstream stream position of RIMD and BB220P strains. UCM strain carries no transposes gene. TTSS genes (green arrow) are in closely situated in major three sub sets.
(The original figure is 'Landscape' size)

Figure 5: Arrangements of Genomic segment (GS) and the toxin and other pathogen related genes in TL regions with Mauve
A Vp chromosome-2s



Mauve result with nucleotide sequences in FASTA-format. GSs are depicted in colors. The positions of GSs are placed according to genetic leading and lagging strand. Genome of CDC strain shows in leading strand in full (above the horizotal line) and lesser chaismata formation of GS link lines. TLs are indicated by red-brakets with flanking region and thermostable direct hemolysin genes by red-arrows.

B Vp chromosome-2s expanded to show gene of TL regions marked with red brakets



Mauve result with GenBank files. The genomic span of TL regions are 75045 bp-RIMD & 76367 bp-BB220P in two pandemic strains only and mostly identical with two hemolysin (*tdh*A/S) genes (red arrows) are in opposite position, TRPS genes in scarttered position, TTSS (black arrows) and two toxR-regulatory genes (pink arrows). The special type of 61853 bp -TL is obsrved in end chromosomal position of FDA strain. It carries three hemolysin genes (red arrows) and a 7-Ureage gene cassete, four transposases and five TTSS genes (black & green arrow respectively) are included.

(The original figure is 'Landscape' size.)