Indian J Microbiol (Oct–Dec 2015) 55(4):375–383
DOI 10.1007/s12088-015-0553-5

ORIGINAL ARTICLE

Genome Wide Analysis for Rapid Identification of *Vibrio* Species

Vipin Chandra Kalia[1,2] · Prasun Kumar[1] · Ravi Kumar[1] · Anjali Mishra[1] · Shikha Koul[1,2]

Received: 26 August 2015 / Accepted: 1 September 2015 / Published online: 7 September 2015

Abstract The highly conserved 16S rRNA (*rrs*) gene is generally used for bacterial identification. In organisms possessing multiple copies of *rrs*, high intra-genomic heterogeneity does not allow easy distinction among different species. In order to identify *Vibrio* species, a wide range of genes have been employed. There is an urgent requirement of a consensus gene, which can be used as biomarker for rapid identification. Eight sequenced genomes of *Vibrio* species were screened for selecting genes which were common among all the genomes. Out of 108 common genes, 24 genes of sizes varying from 0.11 to 3.94 kb were subjected to *in silico* digestion with 10 type II restriction endonucleases (RE). A few unique genes—*dapF*, *fadA*, *hisD*, *ilvH*, *lpxC*, *recF*, *recR*, *rph* and *ruvB* in combination with certain REs provided unique digestion patterns, which can be used as biomarkers. This protocol can be exploited for rapid diagnosis of *Vibrio* species.

Keywords Biomarkers · Diagnosis · Genome · *In silico* · Restriction endonuclease · *Vibrio*

Electronic supplementary material The online version of this article (doi:10.1007/s12088-015-0553-5) contains supplementary material, which is available to authorized users.

✉ Vipin Chandra Kalia
vckalia@igib.res.in; vc_kalia@yahoo.co.in

[1] Microbial Biotechnology and Genomics, CSIR - Institute of Genomics and Integrative Biology (IGIB), Delhi University Campus, Mall Road, New Delhi 110007, India

[2] Academy of Scientific and Innovative Research (AcSIR), 2, Rafi Marg, Anusandhan Bhawan, New Delhi 110001, India

Introduction

The genus *Vibrio* consists of 103 species [1]. Of these, only ten species have been implicated to cause gastrointestinal and extra-intestinal diseases in human beings. *Vibrio* species are generally inhabited in marine niches. In humans, *Vibrio* species has been isolated from stool, vomitus, blood, or wound infections and also from environmental niches such as seawater, sediments, plankton, shellfish (oysters, clams and crabs) [2, 3]. *Vibrio* species which have great medical implications include: *V. alginolyticus*, *V. carchariae*, *V. cholerae*, *V. cincinnatiensis*, *V. fluvialis*, *V. furnissii*, *V. metschnikovii*, *V. mimicus*, *V. parahaemolyticus*, and *V. vulnificus* [4]. *V. parahaemolyticus* spreads into humans through contaminated sea food leading to acute gastroenteritis with diarrhea [2]. *V. cholerae* and *V. vulnificus* are responsible for other serious life-threatening infections in humans [2, 5].

Identification of Vibrios

Vibrio cultures are identified by colonial appearance, Gram stain, serology, and biochemical tests: Oxidase test, Voges–Proskauer test, sensitivity to pteridine O129, serology (agglutination with specific antisera), etc. [6, 7]. For species level identification, Matrix Assisted Laser Desorption/Ionisation—Time of Flight Mass Spectrometry is being employed [8]. This approach is effective in distinguishing very closely related species: *Photobacterium damselae* and *Grimontia hollisae* isolates from *Vibrio* species [8]. The highly conserved gene such as *rrs* (16S rRNA) is most widely used for detecting bacteria. Although, quite effective and precise, it does have some limitations. Species specific genes allow distinction between pathogenic and non-pathogenic strains.

Amplification and sequencing of *dnaJ* gene has been instrumental in identifying *Vibrio* species—*V. alginolyticus*, *V. cholerae*, *V. mimicus*, *V. parahaemolyticus*, and *V. vulnificus* whereas *toxR* amplified using real-time quantitative PCR was found to be useful for detecting *V. vulnificus* in patients with skin and soft tissue infections [9, 10]. For distinguishing *Vibrio* from *Aeromonas* species in patients showing cholera-like symptoms, a duplex- PCR directed at genes—*rrs* and *gcat* (encoding cholesterol acyltransferase) has been used [11]. *notI* and *sfiI* genes have also proved helpful in distinguishing different species of *Vibrio* [12, 13]. Multiplex PCR sequencing of *rpoB* along with *hsp60*, and *sodB* and *flaE* genes was employed to distinguish four species of *Vibrio*: *V. cholerae*, *V. mimicus*, *V. parahaemolyticus*, and *V. vulnificus* [14]. pPCR assay to simultaneously detect virulent and non-virulent strains of *V. vulnificus* and *V. parahaemolyticus* was based on *viuB*, *tdh*, *trh*, *vvhA* and *tlh* genes as biomarkers [15].

Loop-mediated isothermal amplification (LAMP) protocol designed for amplification of *ompW* gene encoding outer membrane protein was targeted to detect *V. cholerae*, where ascytolysin/hemolysin gene (*vvhA*), could rapidly identify *V. vulnificus* with tenfold higher sensitivity than conventional PCR method [16, 17]. Innovative combination of LAMP method and Lateral Flow Dipstick to target *vhhP2* and *rpoX* genes allowed rapid and sensitive detection of *V. alginolyticus* and *V. harveyi* [18, 19]. LAMP based detection was targeted on *rpoS* and *vcgC* genes of *V. vulnificus* on α subunit gene of RNA polymerase of *Vibrio corallilyticus* and thermolabile hemolysin gene (*tlh*) of *V. parahaemolyticus* [20–23].

A few other methods which are employed for typing clinical isolates are: (i) Multi-locus sequence typing, (ii) Multiple-locus variable number tandem repeat analysis, and (iii) Whole genome sequencing [24–26]. In spite of their high accuracy, discriminatory power, and reproducibility, these are limited to reference laboratories only and are not easy to implement for routine assays. The methods are costly, time-consuming and require special equipments [27]. Whole Genome Sequencing is relatively more promising as a rapid, accurate, and comprehensive technique with much wider implications and utility [28]. Rapid and accurate identification of pathogenic bacteria has always been a challenge. Molecular tools have proven helpful in meeting this challenge. A range of novel genomic tools, developed recently have enabled elucidation of the latent features of the highly conserved gene—*rrs* [29–34]. However, this gene could not prove effective in identifying organisms, which possess multiple copies of *rrs* e.g., in *Clostridium* and *Yersinia* [35, 36]. Identification of *Vibrio* has been quite a tough task. Different researchers have used a variety of genes including *rrs*, as biomarkers for distinguishing *Vibrio* species. However, *rrs* alone has not proved very effective in identifying *Vibrio* species. The need is to identify a consensus gene, with unique features to be used as biomarker for rapid diagnosis. Here, we segregated the genes which were common to all species within a genus and digested them *in silico* with various type II restriction endonucleases (RE). Species within each genus could be segregated by different sets of gene-RE combinations.

Materials and Methods

Sequence Data and Comparative Genome Analysis

Completely sequenced genomes of the eight species of *Vibrio* were retrieved (http://www.ncbi.nlm.nih.gov/): *V. anguillarum*, *V. cholerae*, *V. fischeri*, *V. nigripulchritudo*, *V. parahaemolyticus*, *V. tasmaniensis*, *V. tubiashii*, and *V. vulnificus* (Table S1). Characteristics of *Vibrio* genomes have been presented in Table S1. Genes which were common to all the *Vibrio* genomes were elucidated by Pairwise comparisons (Table S2). Among the 8 genomes, 108 protein encoding common genes could be segregated. Of these 108, we selected 24 to represent the whole range of gene sizes, in the range of 113 nucleotides (nts) to 3494 nts (Tables S2 and S3). The highly conserved non-protein coding gene, *rrs* was taken as reference, because it is used widely for identifying bacteria.

Restriction Endonuclease Analysis of Common Genes

All the selected genes were subjected to digestion with ten Type II REs: (i) four base cutters *AluI* (AG'CT), *BfaI* (C'TA_G), *BfuCI* (_GATC'), *CviAII* (C_AT'G), *HpyCH4V* (TG'CA), *RsaI* (GT'AC), *TaqI* (T_CG'A), *Tru9I* (T_TA'A), and (ii) 6 base cutters *HaeI* (WGG'CCW), *Hin1I* (GR_CG'YC) [36]. RE digestion patterns of all the 24 genes sequences along with *rrs* (Table S3) were analysed through Cleaver (http://cleaver.sourceforge.net/). Data matrices of REs generating 5–15 fragments were considered for consensus RE patterns [35, 36]. *Vibrio* species were then identified on the basis of unique gene-RE combinations.

Results

The completely sequenced genomes of *Vibrio* spp.: *V. anguillarum*, *V. cholerae*, *V. fischeri*, *V. nigripulchritudo*, *V. parahaemolyticus*, *V. tasmaniensis*, *V. tubiashii*, and *V. vulnificus* (Table S1) were found to vary from 4.03 to

6.32 Mb. Each genome is composed of 3656 to 5807 genes with an overall GC content in the range of 43.87–47.49 mol% (Table S1).

In Silico rrs Gene Analysis of *Vibrio* Species

The frequency of occurrence of the *rrs* gene per genome of *Vibrio* strains varied from 7 to 11. Within each genome, the *rrs* copies showed high similarity. Multiple sequence alignments of 69 copies of *rrs* from eight *Vibrio* genomes allowed us to conclude that these can be represented by ten groups containing 1–11 copies i.e., 67 copies are highly similar among themselves. RE digestion of *rrs* sequences showed that only a few in each species can be designated as unique: *V. anguillarum* (3/7 copies), *V. cholerae* (1/8 copies), *V. fischeri* (3/11 copies), *V. nigripulchritudo* (4/8 copies), *V. parahaemolyticus* (1/10 copies), *V. tasmaniensis* (2/7 copies), *V. tubiashii* (2/10 copies), and *V. vulnificus* (1/8 copies) (Table S4). It may be stated that *rrs* is not a good candidate gene for distinguishing *Vibrio* species unless all its copies are sequenced. It implies that we may need to resort to other gene sequences for deriving meaningful conclusions.

In Silico RE Digestion Patterns of Common Genes

In view of the fact that unique RE digestion patterns in *rrs* could not be deduced from any of the *Vibrio* genomes, genes which were common among them were analyzed. Genome wide comparison leads to the identification of 108 common genes in these 8 *Vibrio* genomes. Out of these 108 genes, we selected 24 genes, which varied in size from 113 to 3494 nts, in such a manner that genes of all sizes were represented (Tables S2 and S3).

In silico RE digestion patterns of 24 common genes with 10 different REs revealed some very interesting features in them. Of these 24 genes, 9 could be used for distinguishing most of the genomes—*dapF*, *fadA*, *hisD*, *ilvH*, *lpxC*, *recF*, *recR*, *rph* and *ruvB* (Tables S5–S13). The information on RE digestion patterns of the rest 15 genes has been presented as supplementary material (Tables S14–S28). However, due to the generation of a large number (ranging from 10 to 40) of small sized fragments, it became difficult to deduce meaningful conclusions. Hence, these were not considered significant enough for further evaluation.

A comparative analysis of all the nine genes and their RE digestion patterns revealed that *fadA*, *hisD*, and *recF* are the potential candidate genes, which can be used as biomarkers. These three genes had unique RE digestion patterns with REs: *AluI*, *BfuCI*, *CviAII*, *HpyCH4V*, *RsaI*, *TaqI* and *Tru9I*. *HaeI*, *Hin1I* and *BfaI* did not prove very effective, as they scarcely cleave these nine genes.

(i) *hisD*, *recF* and *fadA* genes

In silico digestion of *hisD* gene with REs—*AluI*, *HpyCH4V*, *RsaI*, and *TaqI* resulted in generation of unique digestion patterns with all the eight *Vibrio* genomes, where as *BfuCI* and *Tru9I* were successful in providing information which allowed identification of seven species of *Vibrio*. On the other hand, digestion of *recF* gene with REs—*AluI*, *CviAII* and *Tru9I* resulted in unique digestion patterns with all the eight *Vibrio* genomes, where as *BfuCI*, *HpyCH4V* and *TaqI* were helpful in distinguishing seven species of *Vibrio*. It is interesting to note that these two genes showed contrasting behavior with different REs. The RE digestion patterns of *fadA* gene with REs—*BfuCI*, *CviAII*, *HpyCH4V*, and *Tru9I* were unique and thus could be used as distinct biomarkers, where as with REs—*AluI*, *RsaI* and *TaqI* were effective in distinguishing 5–6 species of *Vibrio*. The three *Vibrio* genomes, which showed resistance to digestion with certain REs: *V. cholerae* (AE003852) to *AluI*, *Tru9I* and *TaqI*, *V. tasmaniensis* (FM954972) to *TaqI* and *Tru9I*, and *V. vulnificus* (BA000037) to *BfuCI* and *RsaI* (Tables 1, 2, 3).

(ii) *dapF*, *ilvH*, *lpxC*, *recR*, *rph* and *ruvB* genes

The digestion of these genes allowed segregation of 5–8 genomes of *Vibrio*. The genomes which could not be digested with most of the REs were: *V. anguillarum* (CP002284) and *V. fischeri* (CP000020) (Tables 4, 5, 6, 7, 8, 9).

(iii) The rest of the genes (15)

The rest 15 genes were effective in distinguishing certain genes with low frequency. The information on their RE digestion pattern can be used to supplement that generated with other genes. Thus, though occasional, these genes have some potential as biomarkers (Tables S14–28).

Discussion

Bacterial identification based on *rrs* gene has turned out to be quite effective. However, organisms having multiple copies of this gene show high Intra-genomic heterogeneity, which may lead to over estimation of the existing variability [37, 38]. A high level of similarity among the different copies of *rrs* present in different *Vibrio* strains further complicates the issue of closely related organisms. In the case of *Vibrio* species, a host of genes have been employed for their identification from time to time: *dnaJ*, *flaE*, *hsp60*, *notI*, *ompW*, *recA*, *rpoA*, *rpoB*, *rpoX*, *rpoS*, *sfiI*, *sodB*, *tdh*, *tlh*, *toxR*, *toxR*, *trh*, *vcgC*, *vhhP2*, *viuB*, and *vvhA*. It indicates that no consensus gene is available so far. It further highlights that *rrs* has not been very fruitful for

Table 1 Unique *in silico* restriction endonuclease digestion pattern (5′–3′) of *hisD* gene of *Vibrio* genomes

Vibrio genome	Restriction endonucleases		
	AluI	*BfuCI*	*Tru9I*
AE003852	34·479·342·275·43·123	165·194·394·72·83·388	
BA000031	769·233·276·18	107·625·93·471	209·102·449·213·323
BA000037	513·392·387·4	–	173·56·82·617·45·306·17
CP000020	262·225·396·9·386·42	222·383·127·21·72·495	110·41·103·602·318·15·106·25
FM954972	70·198·30·348·129·95·316·116·18	212·399·148·173·388	31·118·168·51·812·105·35
CP002284	139·123·38·213·256·86·8·139·61·78·87·71	308·445·72·83·391	142·9·211·56·555·326
FO203526	250·42·221·351·10·299·105·21	222·510·21·546	362·56·771·110
CP009354	150·363·342·147·171·112·8	107·58·317·12·331·468	25·148·684·50·386
	HpyCH4V	*RsaI*	*TaqI*
AE003852	691·155·135·186·24·105	416·119·44·504·213	46·478·419·73·250·30
BA000031	46·48·342·51·78·153·156·107·82·104·88·41	123·379·77·540·177	409·492·337·58
BA000037	94·141·162·333·93·473	845·266·8·47·130	146·38·5·20·108·584·95·300
CP000020	445·246·57·171·156·92·153	502·609·209	189·7·459·611·54
FM954972	85·33·606·123·222·251	125·312·735·148	215·315·401·24·335·30
CP002284	303·133·9·117·111·168·5·37·98·186·76·38·18	379·156·548·36·180	949·317·5·28
FO203526	397·39·9·117·186·75·60·141·51·153·53·18	50·69·964·83·133	524·75·417·283
CP009354	397·48·120·276·33·150·269	379·200·401·139·47·127	184·225·534·6·67·14·241·22

Symbol (·) indicates RE site in the gene sequences (–): no digestion

Additional unique patterns: (i) *BfaI*—1021·117·66·116 (FM954972); (ii) *Hin1I*—304·501·93·398 (BA000037) and (iii) *CviAII*—324·766·68·138 (AE003852), 224·68·281·747 (FM954972) and 218·22·978·75 (CP009354)

Table 2 Unique *in silico* restriction endonuclease digestion pattern (5′–3′) of *recF* gene of *Vibrio* genomes

Vibrio genome	Restriction endonucleases		
	AluI	*BfuCI*	*HpyCH4V*
AE003852	161·78·16·269·568	20·156·67·303·147·42·357	210·147·106·39·403·187
BA000031	161·82·269·10·241·65·164·88	231·303·83·30·76·357	345·58·48·442·59·128
BA000037	243·81·188·10·558	195·36·17·475·287·70	58·270·17·106·149·480
CP000020	161·40·127·184·142·78·348	–	439·12·39·294·57·52·122·8·57
FM954972	201·42·85·199·70·135·260·88	231·492·42·240·75	194·151·106·442·122·8·57
CP002284	201·42·579·54·204	231·138·165·189·357	210·135·58·168·80·190·52·122·65
FO203526	161·82·81·4·194·5·295·258	231·303·189·357	–
CP009354	161·40·42·72·9·184·4·10·561	125·106·303·83·106·360	623·218·75·167
	CviAII	*TaqI*	*Tru9I*
AE003852	142·69·253·568·60	422·36·198·6·16·31·19·189·175	16·25·33·81·114·30·177·191·121·304
BA000031	142·69·241·424·144·28·9·23	277·133·306·309·55	41·33·183·30·177·312·47·257
BA000037	142·69·241·568·60	368·31·595·73·13	74·81·102·30·89·88·365·217·34
CP000020	142·69·241·572·33·23	174·9·216·595·52·34	16·7·51·81·11·121·185·24·273·7·48·44·33
FM954972	142·69·241·572·24·9·23	–	16·7·51·81·102·30·312·57·305·119
CP002284	142·69·361·448·28·9·23	566·138·12·251·113	16·25·33·92·91·30·177·32·273·55·5·39·33
FO203526	142·69·241·379·217·32	56·340·14·256·38·12·309·55	74·183·30·177·305·192·119
CP009354	142·69·837·9·26	410·106·128·153·286	56·18·81·102·30·177·9·303·47·78·60·122

Symbol (·) indicates RE site in the gene sequences (–): no digestion

Additional unique patterns: *BfaI*—110·8·585·374·3 (BA000031)

Table 3 Unique *in silico* restriction endonuclease digestion pattern (5′–3′) of *fadA* gene of *Vibrio* genomes

Vibrio genome	Restriction endonucleases			
	BfuCI	*CviAII*	*HpyCH4V*	
AE003852	84·190·58·294·52·147·242·97	282·13·20·45·112·45·12·54·254·57·219·51	174·111·9·34·143·61·171·173·21·267	
BA000031	353·405·66·13·339	307·20·45·45·30·37·45·233·57·30·113·163·51	91·95·40·71·9·16·423·126·24·9·42·114·116	
BA000037	84·248·9·271·66·68·66·13·188·151	315·90·30·94·54·33·134·57·30·66·210·51	285·9·34·143·61·402·184·23·23	
CP000020	201·428·115·147·220·119	66·282·13·65·7·15·234·548	271·80·9·7·37·133·214·423·33·23	
FM954972	84·257·484·339	295·20·67·23·112·233·87·66·47·214	94·191·9·7·170·28·90·150·144·258·23	
CP002284	158·480·120·79·135·53·151	307·87·23·178·167·144·270	106·191·9·177·412·14·37·184·46	
FO203526	14·70·190·551·242·100	282·33·45·112·111·158·96·113·163·54	79·15·191·9·7·204·429·21·63·149	
CP009354	96·479·49·134·79·135·204	307·20·45·7·15·23·67·45·224·9·87·327	226·12·59·9·177·214·48·150·281	
	AluI	*RsaI*	*TaqI*	*Tru9I*
AE003852	–	33·821·193·117	–	197·357·227·19·89·275
BA000031	–	45·270·68·226·96·161·310	32·333·56·268·237·250	209·248·109·610
BA000037	468·306·144·246	–	20·55·205·129·140·343·272	134·63·248·355·75·126·69·91·3
CP000020	355·162·222·236·255	239·198·322·471	96·104·349·147·47·112·125·250	61·265·185·156·273·82·44·66·95·3
FM954972	61·156·72·246·45·83·55·200·246	33·140·198·337·146·310	–	209·63·185·336·138·76·75·94
CP002284	238·201·153·357·227	185·198·226·96·471	161·462·363·190	782·107·112·163
FO203526	451·212·103·152·19·230	–	30·323·294·245·165·103·7	134·311·430·126·74·92
CP009354	73·602·255·190·56	45·140·130·294·111·146·310	421·487·161·107	146·63·248·109·447·69·91·3

Symbol (·) indicates RE site in the gene sequences (–): no digestion

Additional unique patterns: (i) *BfaI*—50·476·285·419 (CP000020), 184·780·69·131 (FM954972), (ii) *Hin1I*—350·660·81·76 (FO203526)

Table 4 Unique *in silico* restriction endonuclease digestion pattern (5′–3′) of *ruvB* gene of *Vibrio* genomes

Vibrio genome	Restriction endonucleases		
	BfuCI	*CviAII*	*RsaI*
AE003852	12·51·45·12·51·285·513·36	112·305·378·210	258·122·315·250·60
BA000031	12·96·12·227·220·402·36	112·305·378·27·183	258·295·107·345
BA000037	12·96·12·402·45·305·97·36	393·222·147·33·27·39·144	116·544·35·310
CP000020	12·159·328·23·45·465·45	112·305·198·99·81·282	–
FM954972	12·51·45·63·176·105·565	417·345·33·66·156	258·127·168·107·357
CP002284	12·51·45·414·45·402·36	–	116·423·38·428
FO203526	2·118·51·627·153	112·305·345·33·156	–
CP009354	12·51·45·12·11·40·120·231·447·36	112·305·444·144	258·17·105·173·452
	HpyCH4V	*TaqI*	*Tru9I*
AE003852	439·57·44·45·10·57·332·21	66·344·45·158·241·151	–
BA000031	484·80·31·111·270·29	278·93·242·32·128·15·66·151	–
BA000037	232·207·45·89·12·151·269	170·285·312·11·173·54	–
CP000020	25·123·153·135·48·111·36·21·213·74·138	–	19·121·218·117·30·253·319
FM954972	–	170·285·193·125·15·169·38·22	–
CP002284	585·46·114·126·134	170·240·45·158·83·92·66·151	19·25·96·218·91·56·500
FO203526	391·93·12·89·149·134·71·12	5·84·81·171·30·402·178	140·335·88·41·347
CP009354	573·133·165·68·66	15·47·27·81·147·138·318·15·217	–

Symbol (·) indicates RE site in the gene sequences (–): no digestion

Additional unique patterns: (i) *AluI*—87·349·371·64·134 (AE003852), 10·77·58·87·66·779 (CP000020), 10·398·399·58·140 (CP002284), and 41·46·66·79·176·28·371·58·86 (FO203526), (ii) *BfaI*—644·170·48·143 (CP002284)

Table 5 Unique *in silico* restriction endonuclease digestion pattern (5′–3′) of *lpxC* gene of *Vibrio* genomes

Vibrio genome	Restriction endonucleases		
	RsaI	*HpyCH4V*	*AluI*
AE003852	15·108·66·26·192·163·348	292·53·28·230·315	–
BA000031	15·108·66·381·28·320	97·96·54·6·113·199·162·165·26	154·9·189·566
BA000037	15·174·254·127·348	91·6·276·120·72·353	94·538·209·77
CP000020	15·17·91·66·26·355·28·52·258·10	322·30·291·135·42·9·57·8·24	94·471·67·178·108
FM954972	15·108·66·26·355·80·174·94	97·195·81·462·83	94·60·93·385·209·77
CP002284	123·66·254·127·28·144·176	–	565·67·178·108
FO203526	189·381·28·144·176	603·108·109·98	91·261·213·67·209·77
CP009354	15·108·47·19·26·228·127·28·226·94	97·195·81·192·78·186·57·6·26	–
	CviAII	*TaqI*	*Tru9I*
AE003852	180·564·10·88·61·15	266·15·189·123·99·168·58	200·39·210·469
BA000031	744·98·61·15	266·93·111·123·42·225·58	395·54·33·317·119
BA000037	462·282·10·39·125	266·210·21·27·69·99·226	76·319·54·33·189·128·119
CP000020	–	–	131·69·168·27·87·436
FM954972	–	266·21·103·80·6·216·168·58	200·39·210·222·247
CP002284	180·78·258·66·162·49·125	266·15·109·18·89·421	76·162·157·404·116·3
FO203526	180·78·190·14·120·162·49·125	266·93·111·12·42·336·58	–
CP009354	309·153·120·162·49·49·61·15	266·93·31·80·165·225·58	200·38·157·153·123·247

Symbol (·) indicates RE site in the gene sequences (–): no digestion

Additional unique patterns: (i) *BfaI*—250·402·69·189·8 (FM954972) and 250·73·299·296 (CP009354), (ii) *BfuCI*—2·154·48·59·381·91·96·87 (AE003852), 2·124·30·572·103·87 (BA000037) and 2·154·143·619 (FM954972)

Table 6 Unique *in silico* restriction endonuclease digestion pattern (5′–3′) of *dapF* gene of *Vibrio* genomes

Vibrio genome	Restriction endonucleases			
	HpyCH4V	*CviAII*	*BfuCI*	*Tru9I*
AE003852	3·21·564·108·135	261·31·65·124·350	92·16·24·30·177·5·42·222·223	–
BA000031	3·21·35·529·108·51·84	45·312·69·42·13·187·163	–	–
BA000037	3·21·35·529·159·84	261·96·69·42·13·69·233·19·29	92·16·54·177·269·223	–
CP000020	3·21·211·353·108·51·49·35	–	92·16·24·30·177·492	194·77·51·73·436
FM954972	3·21·35·140·36·353·88·155	25·20·216·96·111·13·187·134·29	–	194·77·52·20·52·177·24·235
CP002284	3·21·35·529·243	45·312·69·55·69·35·217·29	92·16·24·30·177·410·82	194·128·250·259
FO203526	3·21·223·341·108·135	468·82·103·149·29	92·16·24·30·177·251·241	194·201·216·209·11
CP009354	3·21·35·374·155·79·164	45·216·207·13·69·118·134·29	87·5·16·54·177·251·241	194·129·30·42·177·153·21·85

Symbol (·) indicates RE site in the gene sequences (–): no digestion

Additional unique patterns: (i) *AluI*—138·16·601·64·12 (FM954972), (ii) *RsaI*—257·26·17·20·511 (AE003852) and, (iii) *TaqI*—249·125·132·325 (CP002284) and 131·251·139·310 (FO203526)

Table 7 Unique *in silico* restriction endonuclease digestion pattern (5′–3′) of *rph* gene of *Vibrio* genomes

Vibrio genome	Restriction endonucleases				
	CviAII	*RsaI*	*BfuCI*	*HpyCH4V*	*AluI*
AE003852	192·69·69·142·98·61·86	67·191·114·6·339	278·28·26·99·88·198	205·192·296·24	–
BA000031	192·138·228·73·86	219·39·120·151·188	–	298·14·85·9·12·299	316·45·86·192·78
BA000037	192·69·211·86·73·86	219·39·254·205	12·12·282·284·127	–	–
CP000020	429·43·86·12·78·69	117·70·71·114·6·339	–	64·57·177·99·203·117	–
FM954972	–	140·118·114·6·151·188	24·254·54·385	397·9·99·140·72	639·9·52·17
CP002284	192·69·211·159·86	–	–	–	–
FO203526	261·57·154·98·147	110·30·238·339	12·507·71·127	397·9·12·63·236	–
CP009354	192·138·258·129	140·47·71·114·157·188	12·12·282·284·127	–	447·100·92·78

Symbol (·) indicates RE site in the gene sequences (–): no digestion

Additional unique patterns: (i) *BfaI*—8·302·9·398 (CP000020) and 310·120·69·218 (CP009354), (ii) *TaqI*—128·177·129·219·64 (BA000031) and 385·115·93·124 (FO203526)

Table 8 Unique *in silico* restriction endonuclease digestion pattern (5′–3′) of *ilvH* gene of *Vibrio* genomes

Vibrio genome	Restriction endonucleases				
	AluI	*HpyCH4V*	*RsaI*	*TaqI*	*Tru9I*
AE003852	168·21·66·94·141·5	165·27·27·70·77·129	82·227·69·117	–	–
BA000031	168·21·180·21·105	165·54·147·129	209·100·49·137	291·101·42·61	–
BA000037	–	165·130·71·7·122	218·91·49·20·117	236·9·93·157	–
CP000020	–	–	–	203·88·153·51	97·39·63·103·182·11
FM954972	168·21·45·261	–	209·9·54·106·53·64	89·65·49·42·199·51	22·177·13·283
CP002284	–	121·44·27·27·147·129	–	–	22·114·76·273·7·3
FO203526	198·57·58·182	43·122·27·27·76·71·129	–	89·249·54·103	–
CP009354	168·30·36·261	–	309·49·7·13·117	–	190·22·90·190·3

Symbol (·) indicates RE site in the gene sequences (–): no digestion

accurate identification. Significantly low sequence similarity of the *dnaJ* gene (77.9 %) compared to 97.2 % of the *rrs* gene, implied its high discriminatory power for *Vibrio* species [9]. Our study has also shown that *rrs* alone cannot be used for identifying *Vibrio* up to the species level. In fact, *tlh* gene studied through LAMP, could identify 143 *V.*

Table 9 Unique *in silico* restriction endonuclease digestion pattern (5′–3′) of *recR* gene of *Vibrio* genomes

Vibrio genome	Restriction endonucleases			
	AluI	*HpyCH4V*	*RsaI*	*TaqI*
AE003852	–	–	–	–
BA000031	454·48·38·6·54	123·304·55·118	–	180·129·90·98·88·15
BA000037	126·414·6·57	123·13·79·388	–	257·117·9·100·24·96
CP000020	73·53·4·347·25·92·9	123·145·159·176	320·177·61·45	–
FM954972	136·341·63·6·54	123·145·9·323	6·177·33·384	–
CP002284	126·163·251·6·48·6	–	–	–
FO203526	246·231·63·54·6	–	–	257·126·7·195·15
CP009354	126·351·69·48·9	123·13·132·9·326	183·33·342·45	–

Symbol (·) indicates RE site in the gene sequences (–): no digestion

Additional unique patterns: (i) *BfaI*—19·108·376·100 (CP000020), (ii) *CviAII*—145·5·367·86 (BA000037) and (iii) *Tru9I*—217·177·203·3 (CP009354)

parahaemolyticus strains but was not able to identify 33 other *Vibrio* spp. and a large number of non-*Vibrio* strains [39]. LAMP assay targeting *toxR* gene was able to correctly detect 36 *V. parahaemolyticus* strains [40]. Multiplex PCR sequencing of *rpoB* along with *hsp60*, *sodB* and *flaE* genes was employed to distinguish four species of *Vibrio*: *V. cholerae*, *V. mimicus*, *V. parahaemolyticus*, and *V. vulnificus*. Here, *rrs* gene was used as positive internal control [14], which implies that this gene alone was not sufficient for identifying *Vibrio* species.

Some of the genes reported in literature are among the common genes detected in our study; these include *flaE*, *recA*, *rpoA*, *rpoB*, *sodB*. *In silico* RE digestion of these genes (Tables S29–S33) revealed that they can also be used to distinguish all eight *Vibrio* species except *rpoB*. RE digestion of *rpoB* gene leads to an unmanageable number of fragments (25–30), which are thus difficult to analyse. The later cannot be recommended as candidate gene also on account of the fact that its amplification is not easy due to its large size (4029 nts; Table S30). Our study allows us to conclude that genes—*fadA*, *hisD*, and *recF* varying in size between 1080 and 1296 nts (in combination with certain REs) are the most suitable candidates for identification of all 8 *Vibrio* species. Here, we need to amplify the specific gene by polymerase chain reaction and subject the amplicon to defined RE. The second category of genes—*dapF*, *ilvH*, *lpxC*, *recR*, *rph* and *ruvB* (495–1004 nts) provide reasonably good information, which can also be exploited for identification of *Vibrio*. As we can expect these genes to be present in other related genera, such as *hisD* gene in *Escherichia coli*, an analysis of RE digestions obtained with ten REs used here (data not shown), revealed clear cut differences between the two genera. This protocol can thus be exploited for rapid diagnosis of *Vibrio* species.

Acknowledgments We are thankful to the Director of CSIR-Institute of Genomics and Integrative Biology (IGIB), and CSIR project GENESIS (BSC0121) for providing the necessary funds, facilities and moral support. Authors are also thankful to Academy of Scientific & Innovative Research (AcSIR), New Delhi.

References

1. Euzeby JP (2013) List of prokaryotic names with standing in nomenclature—Genus *Vibrio*. 2013. http://www.bacterio.net/vibrio.html
2. Austin B (2010) Vibrios as causal agents of zoonoses. Vet Microbiol 140:310–317. doi:10.1016/j.vetmic.2009.03.015
3. Kongrueng J, Tansila N, Mitraparp-arthorn P, Nishibuchi M, Vora GJ, Vuddhakul V (2015) LAMP assay to detect *Vibrio parahaemolyticus* causing acute hepatopancreatic necrosis disease in shrimp. Aquac Int. doi:10.1007/s10499-014-9874-3
4. Reilly GD, Reilly CA, Smith EG, Baker-Austin C (2011) *Vibrio alginolyticus*-associated wound infection acquired in British waters, Guernsey. Euro Surveill 16:pii1994. http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19994
5. Ryu HH, Lee JY, Yun NR, Kim DM (2013) Necrotizing soft tissue infection with gas formation caused by *Vibrio vulnificus* and misdiagnosed as *Pseudomonas aeruginosa*. Am J Emerg Med 31:464–468. doi:10.1016/j.ajem.2012.09.001
6. Choopan N, Louis V, Huq A, Colwell RR (2002) Simple procedure for rapid identification of *Vibrio cholerae* from the aquatic environment. Appl Environ Microbiol 68:995–998. doi:10.1128/AEM.68.2.995-998.2002
7. Ottaviani D, Masini L, Bacchiocchi S (2003) A biochemical protocol for the isolation and identification of current species of *Vibrio* in seafood. J Appl Microbiol 95:1277–1284. doi:10.1046/j.1365-2672.2003.02105.x
8. Clark AE, Kaleta EJ, Arora A, Wolk DM (2013) Matrix-assisted laser desorption ionization-time of flight mass spectrometry: a fundamental shift in the routine practice of clinical microbiology. Clin Microbiol Rev 26:547–603. doi:10.1128/CMR.00072-12
9. Nhung PH, Ohkusu K, Miyasaka J, Sun XS, Ezaki T (2007) Rapid and specific identification of 5 human pathogenic *Vibrio* species by multiplex polymerase chain reaction targeted to *dnaJ* gene. Diagn Microbiol Infect Dis 59:271–275. doi:10.1016/j.syapm.2006.11.004
10. Kim HS, Kim DM, Neupane GP, Lee YM, Yang NW, Jang SJ, Jung SI, Park KH, Park HR, Lee CS, Lee SH (2008) Comparison of conventional, nested, and real-time PCR assays for rapid and accurate detection of *Vibrio vulnificus*. J Clin Microbiol 46:2992–2998. doi:10.1128/JCM.00027-08

11. Mendes-Marques CL, Hofer E, Leal NC (2013) Development of duplex-PCR for identification of *Aeromonas* species. Rev Soc Bras Med Trop 46:355–357. doi:10.1590/0037-8682-1344-2013
12. Kam KM, Luey CK, Tsang YM, Law CP, Chu MY, Cheung TL, Chiu AW (2003) Molecular subtyping of *Vibrio cholerae* O1 and O139 by pulsed-field gel electrophoresis in Hong Kong: correlation with epidemiological events from 1994 to 2002. J Clin Microbiol 41:4502–4511. doi:10.1128/JCM.41.10.4502-4511.2003
13. Taneja N, Sangar G, Chowdhury G, Ramamurthy T, Mishra A, Singh M, Sharma M (2012) Molecular epidemiology of *Vibrio cholerae* causing outbreaks & sporadic cholera in northern India. Indian J Med Res 136:656–663
14. Tarr CL, Patel JS, Puhr ND, Sowers EG, Bopp CA, Strockbine NA (2007) Identification of *Vibrio* isolates by a multiplex PCR assay and *rpoB* sequence determination. J Clin Microbiol 45:134–140. doi:10.1128/JCM.01544-06
15. Bhattacharyya N, Hou A (2013) A pentaplex PCR assay for detection and characterization of *Vibrio vulnificus* and *Vibrio parahaemolyticus* isolates. Lett Appl Microbiol 57:233–240. doi:10.1111/lam.12101
16. Han F, Ge B (2008) Evaluation of a loop-mediated isothermal amplification assay for detecting *Vibrio vulnificus* in raw oysters. Foodborne Pathog Dis 5:311–320. doi:10.1089/fpd.2008.0084
17. Srisuk C, Chaivisuthangkura P, Rukpratanporn S, Longyant S, Sridulyakul P, Sithigorngul P (2010) Rapid and sensitive detection of *Vibrio cholerae* by loop-mediated isothermal amplification targeted to the gene of outer membrane protein *ompW*. Lett Appl Microbiol 50:36–42. doi:10.1111/j.1472-765X.2009.02749.x
18. Plaon S, Longyant S, Sithigorngul P, Chaivisuthangkura P (2015) Rapid and sensitive detection of *Vibrio alginolyticus* by loop-mediated isothermal amplification combined with a lateral flow dipstick targeted to the *rpoX* gene. J Aquat Anim Health 27:156–163. doi:10.1080/08997659.2015.1037468
19. Thongkao K, Longyant S, Silprasit K, Sithigorngul P, Chaivisuthangkura P (2015) Rapid and sensitive detection of *Vibrio harveyi* by loop-mediated isothermal amplification combined with lateral flow dipstick targeted to *vhhP2* gene. Aquac Res 46:1122–1131. doi:10.1111/are.12266
20. Li Y, Zheng Z, Zhao Y, Wei X, Zhu L, Huang X (2010) A culture-free method for detection of *Vibrio vulnificus* from coastal seawater based on loop-mediated isothermal amplification targeting *vcgC* gene. Acta Oceanol Sin 29:93–97. doi:10.1007/s13131-010-0027-x
21. Liu GF, Wang JY, Xu LW, Ding X, Zhou SN (2010) Sensitive and rapid detection of *Vibrio corallilyticus* by loop-mediated isothermal amplification targeted to the alpha subunit gene of RNA polymerase. Lett Appl Microbiol 51:301–307. doi:10.1111/j.1472-765X.2010.02894.x
22. Sun X, Xu Q, Pan Y, Lan W, Zhao Y, Wu VCH (2012) A loop-mediated isothermal amplification method for rapid detection of *Vibrio parahaemolyticus* in seafood. Ann Microbiol 62:263–271. doi:10.1007/s13213-011-0255-0
23. Surasilp T, Longyant S, Rukpratanporn S, Sridulyakul P, Sithigorngul P, Chaivisuthangkura P (2011) Rapid and sensitive detection of *Vibrio vulnificus* by loop-mediated isothermal amplification combined with lateral flow dipstick targeted to *rpoS* gene. Mol Cell Probes 25:158–163. doi:10.1016/j.mcp.2011.04.001
24. Thompson FL, Gevers D, Thompson CC, Dawyndt P, Naser S, Hoste B, Munn CB, Swings J (2005) Phylogeny and molecular identification of vibrios on the basis of multilocus sequence analysis. Appl Environ Microbiol 71:5107–5115. doi:10.1128/AEM.71.9.5107-5115.2005
25. Gonzalez-Escalona N, Martinez-Urtaza J, Romero J, Espejo RT, Jaykus LA, DePaola A (2008) Determination of molecular phylogenetics of *Vibrio parahaemolyticus* strains by multilocus sequence typing. J Bacteriol 190:2831–2840. doi:10.1128/JB.01808-07
26. Teh CS, Chua KH, Thong KL (2010) Multiple-locus variable-number tandem repeat analysis of *Vibrio cholerae* in comparison with pulsed field gel electrophoresis and virulotyping. J Biomed Biotechnol 2010:817190. doi:10.1155/2010/817190
27. Liu D (2006) Identification, subtyping and virulence determination of *Listeria monocytogenes*, an important food borne pathogen. J Med Microbiol 55:645–659. doi:10.1099/jmm.0.46495-0
28. Heidelberg JF, Eisen JA, Nelson WC, Clayton RA, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Umayam L, Gill SR, Nelson KE, Read TD, Tettelin H, Richardson D, Ermolaeva MD, Vamathevan J, Bass S, Qin H, Dragoi I, Sellers P, McDonald L, Utterback T, Fleishmann RD, Nierman WC, White O, Salzberg SL, Smith HO, Colwell RR, Mekalanos JJ, Venter JC, Fraser CM (2000) DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. Nature 406:477–483. doi:10.1038/35020000
29. Porwal S, Lal S, Cheema S, Kalia VC (2009) Phylogeny in aid of the present and novel microbial lineages: diversity in *Bacillus*. PLoS One 4:e4438. doi:10.1371/journal.pone.0004438
30. Kalia VC, Mukherjee T, Bhushan A, Joshi J, Shankar P, Huma N (2011) Analysis of the unexplored features of *rrs* (16S rDNA) of the genus *Clostridium*. BMC Genom 12:18. doi:10.1186/1471-2164-12-18
31. Lal D, Verma M, Lal R (2011) Exploring internal features of 16S rRNA gene for identification of clinically relevant species of the genus *Streptococcus*. Ann Clin Microbiol Antimicrob 10:28. doi:10.1186/1476-0711-10-28
32. Bhushan A, Joshi J, Shankar P, Kushwah J, Raju SC, Purohit HJ, Kalia VC (2013) Development of genomic tools for the identification of certain *Pseudomonas* up to species level. Indian J Microbiol 53:253–263. doi:10.1007/s12088-013-0412-1
33. Bhushan A, Mukherjee T, Joshi J, Shankar P, Kalia VC (2015) Insights into the origin of *Clostridium botulinum* strains: evolution of distinct restriction endonuclease sites in *rrs* (16S rRNA gene). Indian J Microbiol 55:140–150. doi:10.1007/s12088-015-0514-z
34. Kalia VC (2015) Let's explore the latent features of genes to identify bacteria. J Mol Genet Med 9:e105. doi:10.4172/1747-0862.1000E105
35. Kalia VC, Kumar P (2015) Genome wide analysis for searching markers to diagnose *Yersinia* infections. Indian J Microbiol. doi:10.1007/s12088-015-0552-6
36. Kekre A, Bhushan A, Kumar P, Kalia VC (2015) Genome wide analysis for searching novel markers to rapidly identify *Clostridium* strains. Indian J Microbiol 55:250–257. doi:10.1007/s12088-015-0535-7
37. Pei AY, Oberdorf WE, Nossa CW, Agarwal A, Chokshi P, Gerz EA, Jin Z, Lee P, Yang L, Poles M, Brown SM, Sotero S, DeSantis T, Brodie E, Nelson K, Pei Z (2010) Diversity of 16S rRNA genes within individual prokaryotic genomes. Appl Environ Microbiol 76:3886–3897. doi:10.1128/AEM.02953-09AEM.02953-09
38. Chen J, Miao X, Xu M, He J, Xie J, Wu X, Chen G, Yu L, Zhang W (2015) Intra-genomic heterogeneity in 16S rRNA genes in strictly anaerobic clinical isolates from periodontal abscesses. PLoS ONE 10:e0130265. doi:10.1371/journal.pone.0130265
39. Yamazaki W, Ishibashi M, Kawahara R, Inoue K (2008) Development of a loop-mediated isothermal amplification assay for sensitive and rapid detection of *Vibrio parahaemolyticus*. BMC Microbiol 8:163. doi:10.1186/1471-2180-8-163
40. Chen S, Ge B (2010) Development of a *toxR*-based loop-mediated isothermal amplification assay for detecting *Vibrio parahaemolyticus*. BMC Microbiol 10:41. doi:10.1186/1471-2180-10-41