Evidence that the cytolethal distending toxin locus was once part of a genomic island in the periodontal pathogen *Aggregatibacter* (*Actinobacillus*) *actinomycetemcomitans* strain Y4

Sombhun Doungudomdacha,[†] Alla Volgina and Joseph M. DiRienzo

Correspondence
Joseph M. DiRienzo
dirienzo@pobox.upenn.edu

Department of Microbiology, School of Dental Medicine, University of Pennsylvania, Philadelphia, PA, USA

The authors have previously shown that the periodontal pathogen *Aggregatibacter* (formerly *Actinobacillus*) *actinomycetemcomitans* Y4 contains an operon for a genotoxin known as the cytolethal distending toxin (Cdt). The *cdt* locus in strain Y4 is flanked by remnants of heterologous plasmid and integrase sequences. In this study, the DNA sequence immediately downstream from the *cdt* locus on the Y4 chromosome was examined. The extended sequence contained a region that had all the characteristics of a typical bacterial pathogenicity or genomic island. The genomic island (GIY4-1) was approximately 22 kb long, was flanked by a bacteriophage attachment (*att*) sequence and contained a full-length integrase/resolvase gene (*xerD*). A total of 22 complete and partial ORFs represented putative DNA replication/DNA binding/conjugation proteins as well as hypothetical proteins. GIY4-1 was most closely related to putative genomic islands in *Haemophilus ducreyi* 35000HP and *Haemophilus influenzae* 86-028NP and to a chromosomal region in *Haemophilus somnus* 129PT. GIY4-1 was not present in HK1651, which was used as the prototype strain for genomic sequencing of *A. actinomycetemcomitans*. Several sequences in GIY4-1 were homologous to ORFs found on the *A. actinomycetemcomitans* plasmid pVT745. None of the identified ORFs in GIY4-1 appeared to encode potential virulence genes. However, several unique observations supported the possibility that the *cdt* locus of *A. actinomycetemcomitans* Y4 was originally contained within the genomic island.

Received 6 March 2007
Accepted 27 June 2007

INTRODUCTION

Actinobacillus actinomycetemcomitans is a facultative Gram-negative bacterium that has long been associated with localized aggressive periodontitis (LAP) and may also contribute to chronic forms of the disease. This bacterium has recently been reclassified as *Aggregatibacter actinomycetemcomitans* (Norskov-Lauritsen & Kilian, 2006). The bacterium expresses a number of genes whose products can be defined as virulence factors based on their *in vitro* biological activities or similarities to proteins produced by other pathogens (Henderson *et al.*, 2003). Most notable is the expression of two multi-gene cytotoxin loci, leukotoxin (*lkt*) and cytolethal distending toxin (*cdt*). While *Lkt* has been extensively studied (see Lally *et al.*, 1999 for a review), *Cdt* is a relatively newly described *A. actinomycetemcomitans* virulence factor. *Cdt* is a heterotrimer composed of three gene products (*CdtA*, *CdtB* and *CdtC*) that have

homologues in a handful of facultative or microaerophilic Gram-negative pathogenic bacterial genera (Mao & DiRienzo, 2002; Mayer *et al.*, 1999; Sugai *et al.*, 1998).

Although *Cdt* is not unique to *A. actinomycetemcomitans*, this bacterium is the only member of the oral microbial flora identified to date that carries and expresses the toxin locus (Yamano *et al.*, 2003). *Cdt* is prevalent in *A. actinomycetemcomitans* strains. Forty of 45 isolates from periodontitis patients examined by Yamano *et al.* (2003) exhibited various levels of *Cdt* activity. Ahmed *et al.* (2001) found that 43 of 50 strains from periodontitis patients contained all three *cdt* genes and displayed *Cdt* activity. In another study, PCR of subgingival plaque samples revealed that 13 of 106 diseased sites in 146 aggressive and chronic periodontitis patients contained *A. actinomycetemcomitans* expressing all three *cdt* genes (Tan *et al.*, 2002). Fabris *et al.* (2002) reported that 39 of 40 *A. actinomycetemcomitans* isolates from a mix of healthy and periodontal-diseased subjects displayed activity characteristic of *Cdt*.

Much of our work has been focused on characterization of the periodontopathic strain *A. actinomycetemcomitans* Y4. This strain was isolated from a LAP patient in 1979 at the

[†]Present address: Department of Physiology and Biochemistry, Faculty of Dentistry, Mahidol University, Bangkok, Thailand.

The GenBank/EMBL/DBJ accession number for the GIY4-1 sequence of *Aggregatibacter actinomycetemcomitans* is EF196803.

Forsyth Dental Institute (Boston, MA, USA) and has served for many years as a prototype strain for laboratory studies (Tanner *et al.*, 1979). Our initial description of Cdt in this strain revealed that the genetic locus for this toxin is flanked by incomplete sequences homologous to virulence-associated and integrating plasmids (Mayer *et al.*, 1999). In the same study, we also identified a bacteriophage integration (*att*) sequence immediately downstream from the *cdt* genes. These observations are consistent with the possibility that strain Y4 acquired the *cdt* gene as part of a recombination event (most probably horizontal gene transfer). To obtain more conclusive evidence in support of this hypothesis, we sequenced a region of the chromosome farther downstream from the *cdt* locus. Here we report the presence and characterization of a classical genomic island, designated GIY4-1, and discuss the implications of its location proximal to a major genotoxin locus.

METHODS

The nucleotide sequence of GIY4-1 was obtained using a combination of chromosome-walking strategies by employing oligonucleotide primers based on data from previous sequencing runs, restriction endonuclease sites and homologous sequences in the *Haemophilus ducreyi* 35000HP genomic sequence (GenBank accession no. AE017143) and genomic DNA from *A. actinomycetemcomitans* Y4 (GenBank accession no. AF006830; Mayer *et al.*, 1999). The sequence was obtained by walking along both DNA strands. Automated-cycle sequencing reactions were conducted by the Genetics Core Facility at the University of Pennsylvania using an Applied Biosystems 377 sequencer with dye primer chemistry.

The sequences were compiled and assembled with LaserGene (DNASTAR). ORFs were identified and examined with the European Molecular Biology Open Software Suite (EMBOSS, release 3.0; <http://emboss.sourceforge.net>). The Lipman–Pearson algorithm was used to make amino acid sequence alignments (Altschul *et al.*, 1997). The DNA and protein sequence databases were searched for homologous sequences using the TBLASTX and BLASTP algorithms (Gish & States,

Fig. 1. Genetic organization of GIY4-1 and the upstream *cdt* locus region. GIY4-1 is bounded by the *att* site (magenta). Genomic island ORFs are in black except for *xerD* (magenta) and ORF3 and *ssb* (green). Green ORFs have homologues on the *A. actinomycetemcomitans* plasmid pVT745. The *cdt* ORFs are in red and ORFs that reside outside of GIY4-1 are in grey (Mayer *et al.*, 1999). The 74 bp direct repeat is in yellow. Three ORF clusters are shown. The arrows indicate the direction of transcription.

Table 1. Features of GIY4-1 and upstream *cdt* gene locus region

ORF/genetic feature	Size (bp)	Position (bp)*	Deduced peptide (Da)	Putative gene product/function	Conserved Domain Database (CDD) match	G+C (mol%)†
<i>cafA</i>	Partial	1–729	Carboxy-terminus	RNases G and E	Translation, ribosomal structure and biogenesis	48.9
<i>grxB</i>	699	795–1493	24 396	Glutaredoxin 2	Post-translational modification, protein turnover, chaperones	48.3
<i>cobB</i>	639	1497–2135	22 861	NAD-dependent deacetylase	Enzymes which catalyse NAD ⁺ -dependent protein/histone deacetylation	51.5
<i>vagC</i>	279	2223–2501	10 366	Unknown	Virulence-associated proteins of unknown function	39.7
Repeat	74	2552–2625	—	—	—	17.2
<i>cdtA</i>	714	2675–3388	24 514	Cytolethal distending toxin subunit A	Ricin chain B and abrin	38.3
<i>cdtB</i>	858	3397–4254	31 494	Cytolethal distending toxin subunit B	Endonuclease/exonuclease/phosphatase family, which includes magnesium-dependent endonucleases and a large number of phosphatases involved in intracellular signalling	41.0
<i>cdtC</i>	558	4268–4825	20 708	Cytolethal distending toxin subunit C	No conserved domains	37.6
<i>atr</i>	159	5179–5021	Partial	DNA replication, recombination and repair	Antirestriction protein	ND
<i>soj (parA)</i>	681	5208–5888	Partial	Cell division and chromosome partitioning	ATPases involved in chromosome partitioning	ND
ORF1	294	6421–6128	Partial	Hypothetical protein	No conserved domains	43.8
<i>parA</i>	192	6788–6597	Partial	Plasmid replication and partition	Conserved family of bacterial proteins implicated in chromosome segregation.	ND
<i>att</i>	160	6859–7018	—	Bacteriophage attachment site	—	46.0
<i>xerD</i>	852	8069–7218	34 785	DNA breaking–rejoining	Involved in the site-specific integration and excision of lysogenic bacteriophage genomes, transposition of conjugative transposons, termination of chromosomal replication, and stable plasmid inheritance; DNA breaking–rejoining enzyme superfamily	32.1
ORF8	1911	10 033–8123	70 860	Hypothetical protein (putative relaxase)	No conserved domains	37.0
ORF2	195	10 267–10 461	7047	Hypothetical protein	No conserved domains	34.4
ORF3	312	10 425–10 736	10 860	Hypothetical protein (pVT745)	No conserved domains	35.9
ORF9	342	11 343–11 002	13 613	Hypothetical protein	No conserved domains	38.0
Repeat	74	11 333–11 406	—	—	—	23.0
<i>prm</i>	1266	14 651–13 386	36 633	DNA replication, recombination and repair (putative DNA primase)	Antirestriction protein	39.6
ORF10	333	14 812–14 480	12 880	Hypothetical protein	No conserved domains	38.4
ORF11	315	15 134–14 820	12 091	Hypothetical protein	No conserved domains	39.6
ORF4	453	15 513–15 965	16 752	Hypothetical protein	No conserved domains	38.4
ORF5	978	15 914–16 891	34 336	Hypothetical protein	No conserved domains	45.9
ORF6	1791	16 895–18 685	62 918	Hypothetical protein	No conserved domains	42.4
ORF12	594	19 216–18 623	18 428	Hypothetical protein	No conserved domains	39.1
<i>ssb</i>	432	19 721–19 290	15 047	Single-strand binding protein (pVT745)	Single-strand binding protein family	45.6
ORF7	231	19 850–20 080	8825	Hypothetical protein	No conserved domains	32.1
<i>hae</i>	219	20 311–20 093	8295	Hypothetical protein	No conserved domains	34.9
ORF13	363	21 102–20 740	Partial	Hypothetical protein	No conserved domains	ND
ORF14	789	22 197–21 409	29 026	Hypothetical protein	No conserved domains	48.0

Table 1. cont.

ORF/genetic feature	Size (bp)	Position (bp)*	Deduced peptide (Da)	Putative gene product/function	Conserved Domain Database (CDD) match	G + C (mol%)†
ORF15	1230	23 626–22 397	46 641	Hypothetical protein	No conserved domains	41.8
ORF16	582	24 357–23 776	21 127	Hypothetical protein	No conserved domains	42.8
<i>parB</i>	1785	26 104–24 320	66 167	Nuclease	Preferentially cleaves single-stranded DNA, nicks supercoiled plasmid DNA preferably at sites with potential single-stranded character (AT-rich regions and sequences that can form cruciform structures) and exhibits 5'→3' exonuclease activity	42.4
<i>dnaB-parA</i>	2214	28 268–26 055 [27 445–26 052 (<i>dnaB</i>) 28 265– 27 450 (<i>parA</i>)]	82 888 [30 206 (DnaB) 52 489 (<i>ParA</i>)]	DNA replication, recombination and repair (DnaB); plasmid replication and partition (<i>ParA</i>)	Replicative DNA helicase (DnaB); conserved family of bacterial proteins implicated in chromosome segregation (<i>ParA</i>)	42.7 [41.7 (<i>dnaB</i>) 43.2 (<i>parA</i>)]
<i>att</i>	185	28 353–28 537	—	Bacteriophage attachment site	—	46.0
<i>act</i>	168	28 636–28 469	Partial	Acetyltransferase	No conserved domains	ND
<i>pgsA</i>	Partial	28 996–28 799	Carboxy-terminus	Phosphotidylglycerophosphate synthetase	No conserved domains	ND

*Sequence ranges in the reverse direction indicate that the ORF was located on the reverse strand.

†G + C (mol%): *A. actinomycetemcomitans* HK1651=44.4; *H. ducreyi* 35000HP=38.8; *H. somnus* 129PT=37.0; *H. influenzae* 86-028NP=38.1. ND, Not determined.

1993) accessed through the National Center for Biotechnology Information (NCBI) website (<http://www.ncbi.nlm.nih.gov/>). Conserved domain matches were obtained from the Conserved Domain Database (CDD; Marchler-Bauer *et al.*, 2005) using the Conserved Domain Architecture Retrieval Tool (CDART; Geer *et al.*, 2002). Multiple sequence alignments were performed with CLUSTAL_X version 1.83 (Thompson *et al.*, 1997) using the default settings. Phylogenetic trees were constructed using DRAWGRAM in Phylogeny Inference Package (PHYLIP) version 3.65 (Felsenstein, 1989; <http://evolution.genetics.washington.edu/phylip.html>).

RESULTS AND DISCUSSION

The genetic organization of GIY4-1 and adjacent coding regions surrounding the *cdt* locus is shown in Fig. 1. The previously sequenced *cdt* locus (Mayer *et al.*, 1999) is included to orientate the genomic island and to provide updated annotation for several of the ORFs. GIY4-1 (bp 6859–28 537) is flanked by bacteriophage integration or attachment sequences (*att*) which delineate the boundaries of the genomic island. The sequence of the left *att* site (bp 6859–7018) has been reported previously (Mayer *et al.*, 1999) and is included in GenBank under the accession number AF006830. The sequence of the right *att* site (bp 28 353–28 537) is identical to that of the left *att* site. A total of 22 ORFs, each containing 50 or more codons and matching sequences in the GenBank database, reside between the *att* sites (Table 1). These ORFs include both partial and complete gene sequences, determined by interpretation of the results of TBLASTX, BLASTP and CDART search matches with sequences in the databases. Five of the ORFs could be assigned putative functions based on deduced amino acid matches in the CDD. These include genes for an integrase/resolvase (*xerD*; bp 8069–7218), DNA primase (*prm*; bp 14 651–13 386), single-strand DNA-binding protein (*ssb*; bp 19 721–19 290), nuclease (*parB*; bp 26 104–24 320) and a combined DNA helicase–plasmid partitioning protein (*dnaB-parA*; bp 28 268–26 055). All of these gene products have putative roles in the insertion, replication and partitioning of extrachromosomal DNA. The predicted functions of each CDD match are detailed in Table 1. It is interesting that although the *dnaB-parA* sequence contains a single ORF, it represents two full-length in-frame genes. It is not clear at this time whether both genes are expressed. The remaining 17 ORFs have significant matches to hypothetical protein gene sequences deposited in GenBank and have no conserved domains. One ORF, designated *hae*, matched to a *Haemophilus*-specific protein. The other ORFs are designated ORF1–16. Three ORF cluster groups could be arbitrarily assigned based on visual inspection of ORF positions. These are labelled integrase cluster, primase cluster and *par* cluster in Fig. 1. ORF3 and *ssb* are also present in the *A. actinomycetemcomitans* plasmid pVT745 (GenBank accession no. NC002579). Sequence matches to the *A. actinomycetemcomitans* temperate bacteriophage phi 23 were not found (Resch *et al.*, 2004). Many of the ORF sequences had a mol% G + C content well below that of the

Fig. 2. GIY4-1 ORF deduced amino acid sequence homologues in *A. actinomycetemcomitans* HK1651, *H. ducreyi* 35000HP, *H. influenzae* 86-028NP and *H. somnus* 129PT. Chromosome maps are in kilobase pairs. Genomic island maps (GIY4-1, GI35000HP and GI86-028NP) are in base pairs. All maps are drawn to relative scale. The colour scheme of the ORFs is the same as in the legend to Fig. 1. The three ORF clusters (integrase, primase and par) are labelled. ORF coordinate data are from GenBank [accession numbers: AF006830 and EF196803 (*cdt* locus and GIY4-1), AE017143 (*H. ducreyi* 35000HP), CP000057 (*H. influenzae* 86-028NP) and CP000436 (*H. somnus* 129PT)]. *A. actinomycetemcomitans* HK1651 sequence information was obtained from the *Actinobacillus* Genome Sequencing Project (<http://www.genome.ou.edu/act.html>).

Fig. 3. Comparison of XerD homologues. (a) CLUSTAL_X alignment of the *xerD* deduced amino acid sequences from GIY4-1, *H. ducreyi* 35000HP, *H. influenzae* 86-028NP and *H. somnus* 129PT. Identical and similar amino acids are marked with asterisks and colons, respectively. Sequence data were obtained from GenBank. Accession numbers are the same as those in the legend to Fig. 2 except for *xerD* from ICEhin1056 (GenBank accession no. AJ627386). (b) Rooted phylogenetic tree showing the relationships of the XerD homologues aligned in (a). The tree was constructed using PHYLIP 3.6.

A. actinomycetemcomitans genome. This characteristic is indicative that the DNA is from a foreign source (Table 1).

The properties of GIY4-1 just described are typical features of genomic islands. These features include (i) association with tRNA-encoding genes (*att* site), (ii) mol% G+C content differing from that of the host genome, (iii) flanking repeat structures (*att* site), (iv) a mosaic-like structure comprising a multitude of functional, truncated and non-functional putative ORFs with known or unknown functions and (v) the presence of many fragments of mobile genetic elements (Hacker & Kaper, 2000). GIY4-1 is clearly different from AAI-1 of *A. actinomycetemcomitans* HK1651 (Chen *et al.*, 2005) and the atypical pathogenicity island in *Porphyromonas gingivalis* (*rag* locus; Curtis *et al.*, 1999) reported previously. We searched the complete genomes of *A. actinomycetemcomitans* HK1651 (the only strain of this species for which there are complete genome sequencing data; <http://www.genome.ou.edu/act.html>), *H. ducreyi* 35000HP, *Haemophilus influenzae* 86-028NP and *Haemophilus somnus* 129PT for deduced amino acid sequence homologues to the GIY4-1 ORFs. Sequences matching those of the GIY4-1 ORFs were then mapped on each of the genomes with some very interesting results (Fig. 2). All maps are drawn to relative scale. *A. actinomycetemcomitans* HK1651 does not have a genomic island related to GIY4-1, lacks the *xerD* ORF and has only a single copy of the *att* sequence. The single copy of the *att* sequence is located immediately upstream of the *cdt* locus. There are two or more *att* copies in *H. ducreyi* 35000HP, *H. influenzae* 86-028NP and *H. somnus* 129PT. A genomic island very similar to GIY4-1 was found in *H. ducreyi* 35000HP (GI35000HP) and *H. influenzae* 86-028NP (GI86-028NP). To the best of

our knowledge, information about these two genomic islands has not been published. GI35000HP and GI86-028NP are flanked by the *att* sequence as in GIY4-1. The *cdt* locus is located within the boundaries of GI35000HP. The *H. influenzae* 86-028NP and *H. somnus* 129PT genomes do not contain *cdt* genes. The *cdt* gene order is reversed in *A. actinomycetemcomitans* HK1651 relative to that in GIY4-1 and GI35000HP. The organization of the integrase, primase and *par* ORF clusters in GIY4-1 is remarkably similar to those in *H. ducreyi* 35000HP (GI35000HP), *H. influenzae* 86-028NP (GI86-028NP) and *H. somnus* 129PT. There is one copy of the *ssb* ORF in *A. actinomycetemcomitans* HK1651 and it is relatively close to the *cdt* locus. There are two copies of *ssb* in *H. ducreyi* 35000HP, *H. influenzae* 86-028NP and *H. somnus* 129PT. One of the two copies of *ssb* in *H. ducreyi* 35000HP is in GI35000HP and one of the copies in *H. influenzae* 86-028NP is in GI86-028NP.

Based on the observations from the genetic map comparisons, it is very tempting to speculate about the origins of the *cdt* locus in *A. actinomycetemcomitans* Y4. It seems reasonable to suggest that a prototype genomic island originated from a region of the *H. somnus* 129PT chromosome that contained the integrase, primase and *par* ORF clusters. The *xerD* ORF, in proximity to these ORF clusters, in *H. somnus* 129PT could have contributed to the horizontal gene transfer of this region. This region could have been acquired by *H. influenzae* 86-028NP, where it became a true genomic island (flanked by an *att* sequence). The ORF clusters and *xerD* are bounded by the *att* sequence in GI86-028NP. A recombination event could have inserted the *cdt* locus in GI86-028NP either prior to or after transfer to *H. ducreyi* 35000HP, resulting in

Fig. 4. Alignment of att nucleotide sequences from GIY4-1, *H. ducreyi* 35000HP, *H. influenzae* 86-028NP and *H. somnus* 129PT. (a) Sequences that contain only the left conserved region. (b) Sequences that contain the complete att site. (c) Sequences that contain only the right conserved region. Multiple att sequences in a genome are designated by position in bp. Identical nucleotides are marked with an asterisk. The att sequences in GIY4 and *A. actinomycetemcomitans* HK1651 (457263) are identical.

GI35000HP. This modified genomic island could then have been passed to *A. actinomycetemcomitans*. This theory is supported by results from an examination of the phylogenetic relationships among the *xerD* gene products produced by *A. actinomycetemcomitans* Y4, *H. ducreyi* 35000HP, *H. influenzae* 86-028NP and *H. somnus* 129PT. The deduced amino acid sequences of these *xerD* genes, including one found on the *H. influenzae* plasmid ICEhin1056 (Mohd-Zain *et al.*, 2004), are approximately 40% identical (Fig. 3a). The XerD proteins comprise three distinct cluster groups that fit the predicted pattern of horizontal gene transfer of the genomic islands (Fig. 3b). It is important to note that our conclusions are limited by the number of currently available genome sequences. These interpretations may be subject to change as more genomes are sequenced.

Another supporting argument for the proposed order of genomic island passage among the species examined here is that prokaryotic genomes, especially those of pathogens, tend not to increase in size. Prokaryotic genomes are continuously evolving (Nilsson *et al.*, 2005) with the changes a consequence of horizontal gene transfer, duplications that lead to paralogous genes, and deletions (Mira *et al.*, 2001; Ochman, 2001; Rocha, 2004). Even though the size of genomes may increase due to some of these processes, bacterial genomes remain relatively small and generally lack non-functional sequences. The most likely explanation for this occurrence is a process known as deletional bias in which genetic change is skewed towards deletions rather than insertions. The current view is that evolution of pathogens differs from that of non-pathogenic bacteria. Interaction with a host leads to smaller effective bacterial population sizes, low genetic diversity and infrequent recombination (Lawrence, 2005; Lawrence & Hendrickson, 2005). Calculated gene loss rates indicate that genome size can be extensively reduced during a relatively short evolutionary time period (Nilsson *et al.*, 2005). It appears that various genetic rearrangements have occurred in *A. actinomycetemcomitans*, since most of the genomic island sequence appears to have been lost in strain HK1651 and reduced in size in strain Y4. In addition, restriction fragment length polymorphism analysis of the region of the chromosome containing the *cdt* locus revealed deletions, of various lengths, in members of a collection of clinical strains (DiRienzo & McKay, 1994; Mayer *et al.*, 1999). Evidence that the *cdt* locus in strain Y4 was once part of GIY4-1 is supported by the presence of a 74 bp repeat sequence immediately upstream of the *cdt* locus (bp 2552–2625) and within GIY4-1 (bp 11333–11406) (Table 1 and Fig. 1). Interestingly, there are no significant ORFs immediately downstream of the copy of the repeat sequence in GIY4-1. It will be interesting to see how widely GIY4-1 is distributed in *A. actinomycetemcomitans* and the extent of sequence variation.

Although possible, it does not seem likely that the *A. actinomycetemcomitans* Y4 *cdt* locus was obtained by a plasmid or bacteriophage recombination event. ORF3 and

ssb sequences are present on the *A. actinomycetemcomitans* plasmid pVT745 (Galli *et al.*, 2001) but are also widespread since they are found in each of the bacterial genomes shown in Fig. 2. Except for the *xerD* deduced amino acid sequence, no other sequences were shared by ICEhin1056 and GIY4-1. In addition, a complete att sequence, containing the right and left conserved and middle non-conserved regions, was found only in GIY4-1, GI35000HP, GI86-028NP and the *H. influenzae* 86-028NP genome (Fig. 4). A related att sequence was also found in the *H. influenzae* bacteriophage HP2 (Williams *et al.*, 2002). However, only the left and right conserved regions were present.

Thus we have identified a 22 kb genomic island (GIY4-1) in the periodontal pathogen *A. actinomycetemcomitans* Y4. GIY-4 is located immediately downstream from a *cdt* locus. Although no virulence genes were found in GIY4-1, the functions of 17 ORFs that encode hypothetical proteins have not yet been identified. Indirect evidence supports the possibility that the *cdt* locus was once part of the genomic island.

ACKNOWLEDGEMENTS

We thank Sam Cao and Jon Korostoff for critical reading of the manuscript and Paul Sniegowski (Department of Biology) for valuable discussions. This study was supported in part by USPHS grant DE012593 from the National Institute of Dental and Craniofacial Research.

REFERENCES

- Ahmed, H. J., Svensson, L. A., Cope, L. D., Latimer, J. L., Hansen, E. J., Ahlman, K., Bayat-Turk, J., Klamer, D. & Lagergard, T. (2001). Prevalence of *cdtABC* genes encoding cytolethal distending toxin among *Haemophilus ducreyi* and *Actinobacillus actinomycetemcomitans* strains. *J Med Microbiol* **50**, 860–864.
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.
- Chen, W., Wang, Y. & Chen, C. (2005). Identification of a genomic island of *Actinobacillus actinomycetemcomitans*. *J Periodontol* **76**, 2052–2060.
- Curtis, M. A., Hanley, S. A. & Aduse-Opoku, J. (1999). The *rag* locus of *Porphyromonas gingivalis*: a novel pathogenicity island. *J Periodontol Res* **34**, 400–405.
- DiRienzo, J. M. & McKay, T. L. (1994). Identification and characterization of genetic cluster groups of *Actinobacillus actinomycetemcomitans* isolated from the human oral cavity. *J Clin Microbiol* **32**, 75–81.
- Fabris, A. S., DiRienzo, J. M., Wikstrom, M. & Mayer, M. P. A. (2002). Detection of cytolethal distending toxin activity and *cdt* genes in *Actinobacillus actinomycetemcomitans* isolates from geographically diverse populations. *Oral Microbiol Immunol* **17**, 231–238.
- Felsenstein, J. (1989). PHYLIP – phylogeny inference package (version 3.2). *Cladistics* **5**, 164–166.
- Galli, D. M., Chen, J., Novak, K. F. & Leblanc, D. J. (2001). Nucleotide sequence and analysis of conjugative plasmid pVT745. *J Bacteriol* **183**, 1585–1594.
- Geer, L. Y., Domrachev, M., Lipman, D. J. & Bryant, S. H. (2002). CDART: protein homology by domain architecture. *Genome Res* **12**, 1619–1623.
- Gish, W. & States, D. J. (1993). Identification of protein coding regions by database similarity search. *Nat Genet* **3**, 266–272.
- Hacker, J. & Kaper, J. B. (2000). Pathogenicity islands and the evolution of microbes. *Annu Rev Microbiol* **54**, 641–679.
- Henderson, B., Nair, S. P., Ward, J. M. & Wilson, M. (2003). Molecular pathogenicity of the oral opportunistic pathogen *Actinobacillus actinomycetemcomitans*. *Annu Rev Microbiol* **57**, 29–55.
- Lally, E. T., Hill, R. B., Kieba, I. R. & Korostoff, J. (1999). The interaction between RTX toxins and target cells. *Trends Microbiol* **7**, 356–361.
- Lawrence, J. G. (2005). Horizontal and vertical gene transfer: the life history of pathogens. *Contrib Microbiol* **12**, 255–271.
- Lawrence, J. G. & Hendrickson, H. (2005). Genome evolution in bacteria: order beneath chaos. *Curr Opin Microbiol* **8**, 572–578.
- Mao, X. & DiRienzo, J. M. (2002). Functional studies of the recombinant subunits of a cytolethal distending holotoxin. *Cell Microbiol* **4**, 245–255.
- Marchler-Bauer, A., Anderson, J. B., Cherukuri, P. F., DeWeese-Scott, C., Geer, L. Y., Gwadz, M., He, S., Hurwitz, D. I., Jackson, J. D. & other authors (2005). CDD: a Conserved Domain Database for protein classification. *Nucleic Acids Res* **33**, D192–D196.
- Mayer, M. P. A., Bueno, L. C., Hansen, E. J. & DiRienzo, J. M. (1999). Identification of a cytolethal distending toxin gene locus and features of a virulence-associated region in *Actinobacillus actinomycetemcomitans*. *Infect Immun* **67**, 1227–1237.
- Mira, A., Ochman, H. & Moran, N. A. (2001). Deletional bias and the evolution of bacterial genomes. *Trends Genet* **17**, 589–596.
- Mohd-Zain, Z., Turner, S. L., Cerdeno-Tarraga, A. M., Lilley, A. K., Inzana, T. J., Duncan, A. J., Harding, R. M., Hood, D. W., Peto, T. E. & other authors (2004). Transferable antibiotic resistance elements in *Haemophilus influenzae* share a common evolutionary origin with a diverse family of syntenic genomic islands. *J Bacteriol* **186**, 8114–8122.
- Nilsson, A. I., Koskiniemi, S., Eriksson, S., Kugelberg, E., Hinton, J. C. & Andersson, D. I. (2005). Bacterial genome size reduction by experimental evolution. *Proc Natl Acad Sci U S A* **102**, 12112–12116.
- Norskov-Lauritsen, N. & Kilian, M. (2006). Reclassification of *Actinobacillus actinomycetemcomitans*, *Haemophilus aphrophilus*, *Haemophilus paraphrophilus* and *Haemophilus segnis* as *Aggregatibacter actinomycetemcomitans* gen. nov., comb. nov., *Aggregatibacter aphrophilus* comb. nov. and *Aggregatibacter segnis* comb. nov., and emended description of *Aggregatibacter aphrophilus* to include V factor-dependent and V factor-independent isolates. *Int J Syst Evol Microbiol* **56**, 2135–2146.
- Ochman, H. (2001). Lateral and oblique gene transfer. *Curr Opin Genet Dev* **11**, 616–619.
- Resch, G., Kulik, E. M., Dietrich, F. S. & Meyer, J. (2004). Complete genomic nucleotide sequence of the temperate bacteriophage AaΦ 23 of *Actinobacillus actinomycetemcomitans*. *J Bacteriol* **186**, 5523–5528.
- Rocha, E. P. (2004). Order and disorder in bacterial genomes. *Curr Opin Microbiol* **7**, 519–527.
- Sugai, M., Kawamoto, T., Pérès, S. Y., Ueno, Y., Komatsuzawa, H., Fujiwara, T., Kurihara, H., Suganaka, H. & Oswald, E. (1998). Cell cycle-specific growth-inhibitory factor produced by *Actinobacillus actinomycetemcomitans* is a cytolethal distending toxin. *Infect Immun* **66**, 5008–5019.
- Tan, K.-S., Song, K.-P. & Ong, G. (2002). Cytolethal distending toxin of *Actinobacillus actinomycetemcomitans*. Occurrence and association with periodontal disease. *J Periodont Res* **37**, 268–272.

Tanner, A. C., Haffer, C., Bratthall, G. T., Visconti, R. A. & Socransky, S. S. (1979). A study of the bacteria associated with advancing periodontitis in man. *J Clin Periodontol* **6**, 278–307.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The CLUSTAL_X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**, 4876–4882.

Williams, B. J., Golomb, M., Phillips, T., Brownlee, J., Olson, M. V. & Smith, A. L. (2002). Bacteriophage HP2 of *Haemophilus influenzae*. *J Bacteriol* **184**, 6893–6905.

Yamano, R., Ohara, M., Nishikubo, S., Fujiwara, T., Kawamoto, T., Ueno, Y., Komatsuzawa, H., Okuda, K., Kurihara, H. & other authors (2003). Prevalence of cytolethal distending toxin production in periodontopathogenic bacteria. *J Clin Microbiol* **41**, 1391–1398.