

***ILG1*: A New Integrase-like Gene that is a Marker of Bacterial Contamination by the Laboratory *Escherichia coli* Strain TOP10F'**

Wenzhi Tian,¹ Kevin Chua,² Warren Strober,² and Charles C. Chu,¹

¹Department of Medicine, New York University School of Medicine, and Department of Medicine, Division of Rheumatology and Allergy-Clinical Immunology, North Shore University Hospital, and the North Shore-LIJ Research Institute, Manhasset, NY, USA

²Mucosal Immunity Section, Laboratory of Clinical Investigation, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA

Accepted June 26, 2002

Abstract

Background: Identification of differentially expressed genes between normal and diseased states is an area of intense current medical research that can lead to the discovery of new therapeutic targets. However, isolation of differentially expressed genes by subtraction often suffers from unreported contamination of the resulting subtraction library with clones containing DNA sequences not from the original RNA samples.

Materials and Methods: Subtraction using cDNA representational difference analysis (RDA) was performed on human B cells from normal or common variable immunodeficiency patients. The material remaining after the subtraction was cloned and individual clones were sequenced. The sequence of one clone with similarity to integrases (*ILG1*, integrase-like gene-1) was used to obtain the full length cDNA sequence and as a probe for the presence of this sequence in RNA or genomic DNA samples.

Results: After five rounds of cDNA RDA, 23.3% of the clones from the resulting subtraction library contained *Escherichia coli* DNA. In addition, three clones contained the sequence of a new integrase, *ILG1*. The full length cDNA sequence of *ILG1* exhibits prokaryotic, but not eukaryotic, features. At the DNA level, *ILG1* is not similar to any known gene. At the protein level, *ILG1* has 58% similarity to integrases from the cryptic P4 bacteriophage family (S clade). The catalytic domain of *ILG1* contains the

conserved features found in site-specific recombinases. The critical residues that form the catalytic active site pocket are conserved, including the highly conserved R-H-R-Y hallmark of these recombinases. Interestingly, *ILG1* was not present in the original B cell populations. By probing genomic DNA, *ILG1* could only be detected in the *E. coli* TOP10F' strain used in our laboratory for molecular cloning, but not in any of its precursor strains, including TOP10. Furthermore, bacteria cultured from the mouth of the laboratory worker who performed cDNA RDA were also positive for *ILG1*.

Conclusions: In the course of our studies using cDNA RDA, we have isolated and identified *ILG1*, a likely active site-specific recombinase and new member of the bacteriophage P4 family of integrases. This family of integrases is implicated in the horizontal DNA transfer of pathogenic genes between bacterial species, such as those found in pathogenic strains of *E. coli*, *Shigella*, *Yersinia*, and *Vibrio cholera*. Using *ILG1* as a marker of our laboratory *E. coli* strain TOP10F', our evidence suggests that contaminating bacterial DNA in our subtraction experiment is due to this laboratory bacterial strain, which colonized exposed surfaces of the laboratory worker. Thus, identification of differentially expressed genes between normal and diseased states could be dramatically improved by using extra precaution to prevent bacterial contamination of samples.

Introduction

New medically important molecular targets have been identified by subtraction techniques that are able to isolate differentially expressed genes between different cell types, developmental stages or different treatments of the same cell type (1). Although not frequently reported, the resulting subtraction libraries are often contaminated with "junk" sequences, including bacteria (most often *Escherichia coli*), yeast, mitochondrial, viral, and vector DNA (2). Increasing reliance of subtraction techniques on the

polymerase chain reaction (PCR) to amplify small starting samples and the small amounts of material remaining after subtractive hybridization (3–6) may increase the chances of obtaining this "junk" DNA by amplification of laboratory contaminants. We have used the cDNA representational difference analysis (RDA) technique (6) to perform a subtraction between human B cells from healthy and common variable immunodeficiency (CVI) individuals. A large percentage of clones from the resulting subtraction libraries consisted of contaminating "junk" *E. coli* DNA. In addition, we isolated a new gene with greatest similarity to bacteriophage integrases, which we have named *ILG1* (integrase-like gene-1).

Bacteriophage integrases are site-specific recombinases that are critical for the bacteriophage life

Address correspondence and reprint requests to: Charles C. Chu, North Shore-LIJ Research Institute, 350 Community Dr., Manhasset, NY 11030. Phone: (516) 562-1085, fax: (516) 562-1683, e-mail: cchu@nshs.edu.

cycle in switching between lytic and lysogenic states (7). Integrases, such as the prototypic bacteriophage λ Integrase, are essential for the integration and excision of bacteriophage DNA from host genomic DNA. These enzymes catalyze the cutting and rejoining between two DNA duplexes at a short site-specific sequence (8). Although catalyzing a similar DNA recombination reaction, the different bacteriophage integrases can be separated by their different DNA sequence specificities.

Integrases belong to a large family of over 100 prokaryotic and eukaryotic site-specific recombinases that includes integrases, resolvases, transposases, invertases, and excisionases (9). These proteins play an important role in biology, including the movement of viral DNA into and out of host genomes, maintenance of plasmid copy number, alteration of cell surface components, and maintenance of monomeric chromosomes (10). Recently, site-specific recombinases, particularly those of the bacteriophage P4 integrase family, have been implicated in horizontal transfer of DNA among microbes, which can promote pathogenicity to humans (11). A conserved feature of this diverse group of site-specific recombinases is comprised of the amino acid residues that form the catalytic active pocket (9). Especially notable is the conserved R-H-R-Y tetrad hallmark spread between the Box I and Box II domains of this family.

ILG1 is most similar to integrases of the bacteriophage P4 family, suggesting *ILG1* is not of eukaryotic origin. The sequence of *ILG1* lacks characteristic eukaryotic features and contains prokaryotic features. Indeed, analysis of genomic DNA demonstrates that *ILG1* comes from the *Escherichia coli* strain, TOP10F'. Furthermore, *ILG1* is a marker of TOP10F', since it is not found in any precursor strains of *E. coli*. Thus, *ILG1* was employed as a marker of bacterial contamination from TOP10F'.

Materials and Methods

Oligonucleotides

Oligonucleotides were synthesized by the core facility at North Shore University Hospital (Table 1). Oligonucleotides used for cDNA RDA can make three adaptor sets: adaptor set 1 consists of annealed CCC24 and CCC25, adaptor set 2 consists of annealed CCC26 and CCC27, and adaptor set 3 consists of annealed CCC28 and CCC29. The remaining primers were used either for PCR analysis and sequencing, or for 5'/3' RACE (rapid amplification of cDNA ends) experiments.

Cell Culture

B cells were purified from peripheral blood mononuclear cells (PBMC) from CVI patients and normal control blood bank donors by Ficoll-Hypaque (Amersham Pharmacia Biotech, Piscataway, NJ) density gradient centrifugation (12) followed by

anti-CD19 monoclonal antibody coated magnetic beads (Dynal, Lake Success, NY) positive selection (13). B cells (>95% CD20⁺) were cultured in complete RPMI 1640 supplemented with 0.005% *Staphylococcus aureus* Cowan I (Calbiochem, San Diego, CA), 5 ng/ml interleukin-2 (R&D Systems, Minneapolis, MN), and 50 ng/ml interleukin-10 (PharMingen, San Diego, CA) for 48h at 37°C and 5% CO₂ (14).

Table 1. Oligonucleotides

Name	Sequence (5' → 3')
CCC24	AGC ACT CTC CAG CCT CTC ACC GAG
CCC25	GAT CCT CGG TGA
CCC26	ACC GAC GTC GAC TAT CCA TGA ACG
CCC27	GAT CCG TTC ATG
CCC28	AGG CAA CTG TGC TAT CCG AGG GAG
CCC29	GAT CCT CCC TCG
CCC34	ATT AAC CCT CAC TAA AG
CCC35	AAT ACG ACT CAC TAT AG
CCC48	GAC TCG AGT CGA CAT CGA TTT TTT TTT TTT TTT TT
M13R	GGA AAC AGC TAT GAC CAT G
CCC131	ATC CCT GGC GCT ACA GAG AAG C
CCC137	ATC TGG CAC CAC ACC TTC TAC AAT GAG CTG GG
CCC138	CGT CAT ACT CCT GCT TGC TGA TCC ACA TCT GG
CCC155	GCT TGA GGC AGA CGT GTA TCC G
CCC159	CGA GCG AAA CAC GAT GGC AGA C
CCC160	GAT GCA GCG CGT TGA CCG TAT G
CCC161	GCA AAC GAA TCA GGA TGG AGC G
CCC162	CTG CGC CCA CCA ATT CAT CAT C
CCC165	CCC ACC GTC CAG CCT GAT G
CCC166	TAT TCG CTG CGT CCT GTT C
CCC169	AGG GGG TAG ACG CCG AAA G
CCC170	CCG TGG CGA TAT TTC ATG C
CCC173	GAG GCT GCG TAC TTT GAG G
CCC174	CCC CTA GAG TTT ATG CAC C
CCC175	CGA TTT CCA TGC CAC TGA C
CCC176	TAC TTT TCT TAC GTC GCA G
CCC403	GGG GGA TGT AGA AAC TCA A
CCC404	GGA GAA CGT CAG GAG AGG C
CCC417	AAC AAG CGA CAG AGC GTG C
CCC418	GGT GTA GAG CAG GTC GGT G
Oligo d(T)- anchor	GAC CAC GCG TAT CGA TGT CGA CTT TTT TTT TTT TTT TTV
PCR anchor	GAC CAC GCG TAT CGA TGT CGA C

RNA Preparation

Total RNA was prepared from 5×10^6 pelleted cells that were resuspended and lysed in RNA STAT-60 (Tel-Test Inc., Friendswood, TX) using the protocol provided by the manufacturer. The aqueous RNA fraction was precipitated by isopropanol, washed, resuspended in diethyl-pyrocyanate-treated H_2O , and quantitated by UV adsorption at 260 nm. Poly(A)⁺ mRNA was prepared from 1×10^8 cells using a FastTrack mRNA Isolation Kit (InVitrogen, San Diego, CA) according to manufacturer's directions. Approximately 1 μ g poly(A)⁺ mRNA was purified from the Oligo (dT) Cellulose tablets.

cDNA Representational Difference Analysis (RDA)

cDNA RDA was performed as previously described (6). cDNA was prepared from 1 μ g poly(A)⁺ mRNA using a RiboClone cDNA Synthesis System (Promega Corp., Madison, WI). Briefly, after conversion of mRNA into single stranded cDNA using Avian Myeloblastosis Virus reverse transcriptase and oligo(dT)₁₅ primer, second strand synthesis was performed with RNaseH and DNA polymerase I resulting in approximately 2 μ g double stranded cDNA between 200 to 2000 bp in length. The double-stranded cDNA products were digested with *Sau3AI* (New England Biolabs (NEB), Beverly, MA), ligated to adaptor set 1, and PCR amplified using CCC24 to produce amplicon representations of healthy (tester) and CVI patient (driver) B cell RNA. The driver amplicon was biotinylated by PCR amplification with 5' biotinylated CCC24. The tester amplicon was modified by replacing adaptor set 1 ends with adaptor set 3 by digestion with *Sau3AI*, purification by ChromaSpin-100 columns (Clontech Laboratories, Palo Alto, CA), and ligation with adaptor set 3. For the first round of subtraction, the modified tester amplicon was hybridized overnight at 65°C with excess biotinylated driver amplicon (20:1 driver:tester). The hybridization mixture was depleted with streptavidin magnetic beads (Dynal) to remove driver DNA strands. CCC28 of adaptor set 3 was used to PCR amplify tester DNA hybrids from the depleted mixture. The resulting amplicon was modified for the subsequent subtraction by replacing adaptor set 3 with adaptor set 2. The second round of hybridization and subtraction was performed as above except using a higher driver:tester ratio (100:1) and using CCC26 for PCR amplification. For subsequent rounds of subtraction, adaptor set 2 and adaptor set 3 were interchanged on the tester amplicon as above, the hybridizations were performed at increasing driver:tester ratios (200:1, 1000:1, 2000:1), and the subtraction repeated.

Cloning, DNA Sequencing, and Sequence Analysis

Products of the last round of cDNA RDA were either directly cloned into pCR2.1 using the Original TA Cloning Kit (InVitrogen) or first digested with *Sau3AI*

and cloned into the *Bam*HI site of pBluescriptSK+ (Stratagene, La Jolla, CA). After transformation of TOP10F' *E. coli* (InVitrogen) to ampicillin resistance, individual bacterial colonies were boiled and tested by PCR for inserts in the cloning plasmid (6). Inserts in pCR2.1 were PCR amplified using CCC35 and M13R primers, purified by Chromaspin-100 columns, and sequenced using the same primers. Inserts in pBluescriptSK+ were analyzed similarly using CCC34 and CCC35 primers. DNA sequencing was performed on an ABI 373 DNA Sequencer using the ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kit with AmpliTaq DNA Polymerase, FS (PE Applied Biosystems, Foster City, CA). Sequences were assembled using AssemblyLIGN (Oxford Molecular, Campbell, CA) and analyzed by BLAST (15) using default settings provided by the National Center for Biotechnology Information (Bethesda, MD) to compare to public sequence databases or by MacVector (Oxford Molecular) on Macintosh computers (Apple, Cupertino, CA).

Determination of Full Length *ILG1* cDNA Sequence

From the sequence of the cDNA RDA clone (wz1#1), we employed 5'/3' RACE to obtain the complete *ILG1* cDNA sequence. Seven RACE clones were isolated that spanned the entire *ILG1* cDNA (Fig. 1). 3' RACE was performed using ϵ LONGase Enzyme Mix (Life Technologies, Rockville, MD) according to manufacturer's instructions. After first strand cDNA synthesis with CCC48, the 3' end was PCR amplified using CCC131 and CCC48. The resulting 1180 bp PCR product was cloned into pCR2.1 resulting in plasmids pTIAN1, pTIAN2, and pTIAN3. These plasmids are identical with the exception that the insert in pTIAN1 is in the opposite orientation. The inserts in pTIAN1-3 were sequenced with CCC35, CCC155, CCC159, CCC160, CCC161, CCC162, and M13R primers. 5' RACE was performed using a 5'/3' RACE Kit (Boehringer Mannheim, Indianapolis, IN). 5' RACE clone pTIAN7 was obtained after first strand cDNA synthesis with CCC160 primer, poly(A) tailing

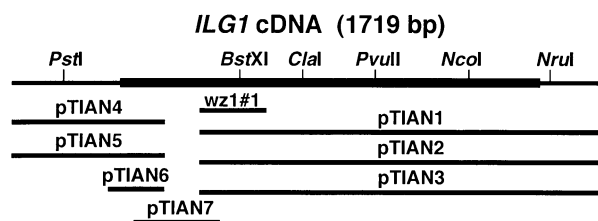


Fig. 1. *ILG1* cDNA clone map. The top line diagrams the full length *ILG1* cDNA with the thin lines indicating the 5' and 3' untranslated regions on the left and right, respectively. The large dark box indicates the coding sequence. The locations of selected restriction enzyme sites are indicated above the line. Underneath, the locations of the *ILG1* cDNA fragments contained in the original cDNA RDA clone (wz1#1) and the subsequent RACE clones (pTIAN1-7) are indicated by dark lines. Diagram is shown to scale.

with terminal transferase, PCR amplification with CCC165 and oligo d(T)-anchor primers, followed by a second PCR amplification with CCC166 and PCR anchor primers. A 282 bp PCR product was inserted into pCR2.1 resulting in pTIAN7. pTIAN7 was sequenced with oligo d(T)-anchor and M13R primers. 5' RACE clones pTIAN4, pTIAN5, and pTIAN6 were obtained after first strand cDNA synthesis with CCC166 primer, poly(A) tailing, PCR amplification with CCC169 and oligo d(T)-anchor primers, followed by a second PCR amplification with CCC170 and PCR anchor primers. A 470 bp, 473 bp, or 199 bp PCR product was inserted into pCR2.1 resulting in pTIAN4, pTIAN5, and pTIAN6, respectively. pTIAN4 and 5 were sequenced with oligo d(T)-anchor, M13R, CCC170, CCC173, CCC174, and CCC176 primers. pTIAN6 was sequenced with CCC35 and CCC175 primers. Assembly of all the resulting sequences and removal of terminal poly(A) stretches resulted in the full length *ILG1* cDNA.

Northern Blot Analysis

After denaturation, total RNA samples (10 μ g) were electrophoresed in 1% agarose/formaldehyde gels, transferred to Hybond-N nylon membrane (Amersham Pharmacia Biotech), and UV-crosslinked. Membrane was prehybridized at 68°C for 1 h in ExpressHyb Hybridization Solution (Clontech), hybridized with probe at 68°C for 16 h, washed following the manufacturer's instructions, and then analyzed by PhosphorImager (Molecular Dynamics, Sunnyvale, CA). Probe DNA fragment (50 ng) was radiolabeled by random primer labeling kit (Stratagene) in the presence of [α -³²P]dCTP (3000 Ci/mmol, Amersham Pharmacia Biotech). The *ILG1* probe was plasmid pTIAN1. The human β -actin probe (838 bp) was PCR amplified from cDNA using primers CCC137/CCC138. The 18S rRNA probe was PCR amplified using CCC34/CCC35 primers from clone wz2#3 containing a 556 bp 18S rRNA cDNA RDA *Sau3AI* fragment inserted in pBluescriptSK+. Membrane was reprobed after being stripped by washing in boiling 0.5% SDS.

Genomic DNA Preparation

Human genomic DNA was prepared from peripheral blood mononuclear cells using a Puregene DNA isolation kit for buccal cells (Gentra Systems, Minneapolis, MN). Briefly, 20×10^6 cells were lysed in Cell Lysis Solution, extracted with the Protein Precipitation Solution, and precipitated with isopropanol. The resulting DNA was resuspended in TE buffer (10 mM tris(hydroxymethyl)aminomethane (Life Technologies), 1 mM ethylenediaminetetraacetic acid (Sigma, St. Louis, MO), pH 8.0). Bacterial genomic DNA from *E. coli* strains TOP10, TOP10F' (Invitrogen), M182, MC1000, MC1060, MC1061, MG1655 (*E. coli* Genetic Stock Center, New Haven, CT), or bacteria cultured from the laboratory worker were prepared using CTAB (hexadecyltrimethyl ammonium bromide) (Sigma) (16). Swabs from the mouth, face and, hand

were inoculated onto LB plates, incubated overnight at 37°C, and resulting single colonies were picked for genomic DNA preparation. Briefly, bacteria were cultured in LB medium for 16–18 hrs at 37°C with aeration, lysed, digested with proteinase K and precipitated with CTAB. High molecular weight DNA was recovered from the supernatant by isopropanol precipitation, resuspended in TE buffer, and quantitated by UV absorption at 260 nm.

Southern Blot and Dot Blot Analysis

Genomic DNA was digested with *Sau3AI* at 37°C overnight. 10 μ g DNA was electrophoresed in 0.7% agarose gels, transferred by Southern blot to Hybond-N, membranes and probed as above (see Northern blot analysis).

For dot blots, 1 μ g DNA was directly spotted onto membranes and treated as above. Like the 18S rRNA probe, the *lacI* probe was PCR amplified from clone wz1#15 containing a 108 bp *lacI* *Sau3AI* fragment. The *frdA* probe was PCR amplified either from cDNA RDA clone wz1#103 (contains a 278 bp *frdA* fragment plus 72 bp of adaptor inserted in the TA cloning site of pCR2.1) using M13R/CCC35 primers or from *E. coli* genomic DNA using CCC403/CCC404 primers (183 bp product). The *selC* probe was PCR amplified from *E. coli* genomic DNA using CCC417/CCC418 primers to generate a 293 bp product.

Results

Contamination of cDNA RDA Subtraction Clones

To isolate genes expressed by B cells during isotype class switch, subtraction by cDNA RDA was performed between healthy (tester) and CVI patient (driver) B cells cultured under conditions where only the healthy B cells undergo class switching (14). After five rounds of subtraction, Southern blot analysis indicated that tester DNA fragments had been enriched (data not shown). This material was cloned and 103 randomly picked clones were sequenced. After sequence comparison, these clones could be grouped into 52 genes (many genes were represented by multiple clones) (Table 2). Of these genes, 15 could be functionally categorized based on sequence similarity to genes involved in enzyme reactions (6), signal transduction (4), transcription (2), secreted products (1), cell cycle (1), and DNA recombination (1). A surprising number of unknown genes (39 clones representing 21 genes) were isolated that did not match any gene or protein of known biological function. Finally, a large number of identified clones were not of human origin, but represented contaminating *E. coli* (24 clones representing 11 genes) and plasmid (10 clones representing 5 genes) DNA fragments. When this cDNA RDA experiment was repeated a second time using three rounds of subtraction with higher driver to tester amplicon ratios (100:1, 1000:1, and 10,000:1), a significant number of clones containing *E. coli* DNA were still obtained (12.9% of 93 clones).

Table 2. Sequence identification of cDNA RDA clones

Function	Similarity* or Name	Genbank Similarity*	No. of Clones	Freq.(%)
cell cycle control	HRAD9	HSU53174	1	1.0%
enzyme	ATP citrate lyase	HSU18197	8	7.8%
enzyme	lysyl hydroxylase 3' splice variant	AF046889	2	1.9%
enzyme	aldehyde reductase	HUMALR	1	1.0%
enzyme	(alanine aminotransferase)	(461498)	1	1.0%
enzyme	(HMG CoA lyase)	(14714839)	1	1.0%
enzyme	(maoC dehydrogenase)	(6458985)	1	1.0%
product	SP-D	HUMSPD02	1	1.0%
recombination	(integrase) " <i>ILG1</i> "	(1788783)	3	2.9%
signal transduction	Maxp1	AF002251	3	2.9%
signal transduction	MAP kinase kinase 3	HUMMKK3A	2	1.9%
signal transduction	PI ₂ 5-phosphatase	HSU45973	2	1.9%
signal transduction	casein kinase I epsilon	HUMCSNK1E	1	1.0%
transcription	TLE3 Enhancer of split	HUMTLEIII	2	1.9%
transcription	RNPS1	AF015608	1	1.0%
unknown	21 gene fragments		39	37.9%
<i>E. coli</i>	11 gene fragments	various	24	23.3%
plasmid	5 gene fragments	various	10	9.7%
Total			103	100.0%

*Similarity at only the amino acid level is indicated by parentheses.

Isolation of a New Integrase Gene, *ILG1*

The clones containing the DNA recombination gene obtained in our first experiment had some similarity to integrases at the amino acid level (45% identical over 44 residues), but had no similarity at the DNA level (data not shown). Therefore, this gene was named *ILG1*. Using the RACE technique, we obtained the full-length cDNA sequence of *ILG1* (Fig. 2). The cDNA sequence contains at least six features that are not eukaryotic, but more prokaryotic in nature. First, *ILG1* has a 311 bp 5' untranslated region (UTR). Eukaryotic cDNA characteristically have a short 5' UTR of 20–100 bp (17). Second, the *ILG1* translation start is not the first methionine (Met) codon, but the second Met codon from the 5' end. Typically, the first Met codon from the 5' end is the start codon in eukaryotic cDNA (18). Third, *ILG1* cDNA exhibits a typical *E. coli* ribosomal binding site (5/8 bp match) that exhibits the appropriate spacing from the start codon (19). This feature allows *E. coli* ribosomes to start translation at internal start codons in a mRNA without having to scan the 5' end and begin at the first start codon. Thus, *E. coli* cDNA may have long 5' UTRs with many upstream Met codons. Fourth, *ILG1* does not have a typical eukaryotic Kozak sequence surrounding the start codon for optimal

translation (18). The most conserved Kozak sequence features, a purine nucleotide at position -3 and a guanine nucleotide at position +4 from the first nucleotide of the start codon, are not found. Fifth, *ILG1* does not have a typical eukaryotic polyadenylation site (AATAAA) to signal transcription termination (20). Instead, a possible *E. coli* rho-independent transcription termination site is found at the 3' end, a small GC-rich stem-loop structure (21). Sixth, *ILG1* codon usage is most optimal using bacterial codon usage tables (22) (data not shown).

ILG1 Belongs to the Bacteriophage P4 Integrase Family

The *ILG1* coding sequence encodes a predicted 401 amino acid protein of 45 kDa. BLAST comparison of the predicted protein sequence to the public protein databases detected similarities to several integrases, with highest similarity (~58%) to the bacteriophage P4 family of integrases. Ten of these are shown in Table 3. These include those from *E. coli* bacteriophage P4 (23) and relatives (bacteriophage ϕ -R73 (24), cryptic P4 prophages (25,26)), as well as integrases from other bacterial strains that are associated with horizontal gene transfer of large DNA segments. These DNA segments are associated with pathogenicity in humans (*Shigella flexneri* (27,28), *Yersinia pseudotuberculosis* (29),



Fig. 2. *ILG1* cDNA sequence. The complete *ILG1* cDNA sequence (Genbank accession no. AF251435) is shown in the 5' to 3' direction with nucleotide numbering of the final nucleotide and the first nucleotides in the left hand position of each line. Half arrows (labeled by numbers without the CCC prefix) underline the locations of oligonucleotides (with the arrowhead indicating the 3' end). Selected restriction enzyme sites are indicated with underlined sequence labeled above with the enzyme name. Boxed bold sequences indicate the translation start (ATG) and stop (TAA) codons. The RBS labeled boxed sequence indicates the putative bacterial ribosomal binding site, a 5/8 bp match to the ideal sequence TAAGGAGG located 5–10 nucleotides upstream of the start codon (19). A potential 3' stem loop structure involved in bacterial transcription termination is indicated by overlined inverted half arrows. Sequence obtained from the original cDNA RDA clone ranged from position 540 to 731. Sequence from pTIAN1, 2, and 3 began at position 540 and ended at the 3' end of the cDNA. Sequence from pTIAN4, 5, and 6 began at position 436 and ended at the 5' end of the cDNA (pTIAN4 and 5) or at position 276 (pTIAN6). Sequence from pTIAN7 ranged from position 351 to 597.

Y. pestis (30), *Vibrio cholera* (31)) or colonization in plants (*Mesorhizobium loti* (32)). Despite similarity that extends the entire length of the protein (Fig. 3), BLAST comparison to the public nucleotide databases did not detect any DNA sequence similarities (data not shown).

ILG1 Belongs to the Integrase Family of Site-specific Recombinases

ILG1 similarity to bacteriophage P4 family integrases includes both the N-terminal domain that aids in DNA-binding, and the C-terminal catalytic domain (Fig. 3). The C-terminal catalytic domain has many

conserved features among a diverse group of site-specific recombinases (9). *ILG1* matches this consensus sequence (59% similar) as well as other bacteriophage P4 family integrases (Fig. 3). First, *ILG1* contains the hallmark of these site-specific recombinases, a nearly invariant R-H-R-Y tetrad (9), in the appropriate location (R240, H329, R332, and Y365). Second, the R-H-R-Y tetrad is found within two longer conserved regions called Box I (26 amino acids containing the first R) and Box II (41 amino acids containing H-R-Y) (9). *ILG1* is fairly similar to the consensus in both Box I (65%) and Box II (48%). Third, additional smaller

Table 3. *ILG1* Belongs to the bacteriophage P4 family of integrases

Organism	Element	Gene	Accession Number	Identity (%)	Similarity (%)
<i>E. coli</i> TOP10F'	*Cryptic P4 prophage ϕ -TOP10F	<i>ILG1</i>	AAK49816	100	100
<i>S. flexneri</i> M90T	Shigella pathogenicity island 2	<i>Int</i>	AAD44730	40	58
<i>E. coli</i> O157:H7 (EDL933)	Cryptic P4 prophage 933L	CP4 <i>Int</i>	AAC31482	41	58
<i>S. flexneri</i> SA100	Shigella pathogenicity island 2	Sai <i>Int</i>	AAD54663	40	58
<i>M. Loti</i> ICMP3153	Symbiosis island	<i>IntS</i>	AAC24508	42	58
-	Bacteriophage P4	<i>Int</i>	VINT_BPP4	41	57
<i>E. coli</i> K-12 MG1655	Cryptic P4 prophage	<i>IntB</i>	INTB_ECOLI	39	55
<i>E. coli</i> Cl-23	Bacteriophage ϕ -R73	<i>Int</i>	A42465	35	55
<i>Y. pseudotuberculosis</i> IP32637	High-pathogenicity island	<i>Int</i>	CAA08754	36	55
<i>Y. pestis</i> KIM6+	High-pathogenicity island	<i>Int</i>	AAC69581	35	54
<i>V. cholera</i> O395	Colonization gene cluster	<i>Int</i>	AAC44230	35	55

*Tentative proposal based on *ILG1* similarity to other bacteriophage P4 integrase family members shown here.

regions of conservation called Patches I, II, and III are found in these site-specific integrases (9). *ILG1* matches the three Patches fairly well (71%, 57%, and 100% similar, respectively).

Although prokaryotic and eukaryotic site-specific recombinases share many features, a few differences remain (9). *ILG1* contains these prokaryotic features. First, in Box II, the separation between the H and Y residues of the R-H-R-Y hallmark is typically 33–35 amino acids in prokaryotic recombinases, whereas eukaryotic recombinases have a 37–40 amino acid separation. *ILG1* has a 35 amino acid separation between these residues (Fig. 3). Second, in Box I, the consensus prokaryotic T-G-X-R motif is replaced by N-C-C-R in eukaryotic recombinases. *ILG1* has a more prokaryotic-like T-A-A-R (position 237–240) at this location. Third, in Box II, the four residue prokaryotic motif L-L-G-[H,W] is replaced by a five residue motif [L,V,S,P]-[Y,F,L,V]-G-N-W. *ILG1* has a more prokaryotic four residue motif (Q-L-A-H, position 352–355).

ILG1 is Not of Human Origin and is Present in Bacteria

To determine if *ILG1* is expressed in human cells, we examined its expression by Northern blot and found that *ILG1* is not expressed in human B cells (Fig. 4). To test if *ILG1* is of human origin, we probed human genomic DNA by Southern blot and found that *ILG1* could not be detected (Fig. 5A). Because of the high percentage of contaminating *E. coli* and plasmid sequences in our cDNA RDA library (Table 2), we reasoned that *ILG1* may represent a previously uncharacterized bacteriophage integrase found in the contaminating *E. coli*. Therefore, we probed genomic DNA from bacteria for the presence of *ILG1* (Fig. 5A). *ILG1* could easily be detected in genomic DNA from TOP10F'.

ILG1 was Recently Introduced into TOP10F'

To determine if *ILG1* is present in other *E. coli* strains, especially those that preceded TOP10F', we examined some of these strains. TOP10F' was generated by introducing an F' episome containing *lacI^q* and Tn10(*tet^R*) into strain TOP10 (InVitrogen). TOP10 comes from DH10B, which was derived from MC1061 (33). The derivation of MC1061 is complex, involving over 10 mutagenesis, transduction, and conjugation steps from the original *E. coli* K-12 strain (Mary Berlyn, pers. com.). In brief, MC1061 was derived from MC1060 conjugated with D7091 Hfr. MC1060 was derived from MC1000 conjugated with an F' strain TSM100. MC1000 was derived from a cross between M182 and D7091F. The derivation of D7091F involves at least seven more genetic manipulations from *E. coli* K-12. The original K-12 strain probably does not contain *ILG1*, because it is not found in the complete *E. coli* genomic sequence (26). This sequence was determined from *E. coli* strain MG1655, which is derived from the original *E. coli* K-12 strain by two direct steps (UV treatment and growth on blood agar to λ^- and *rph-1*, acridine orange treatment to F⁻) (34). We probed genomic DNA from MG1655 and immediate precursors of TOP10F' (Fig. 6). *ILG1* can only be detected in TOP10F' and not in any precursor strains, including TOP10. Confirming our sequence comparison with the *E. coli* genomic sequence, MG1655 does not contain *ILG1*. In contrast, *E. coli* genes, *selC* and *frdA*, are detected in all strains.

Laboratory Bacteria is Detectable on Laboratory Worker

In order to determine the origin of bacterial contamination, we prepared genomic DNA from bacteria

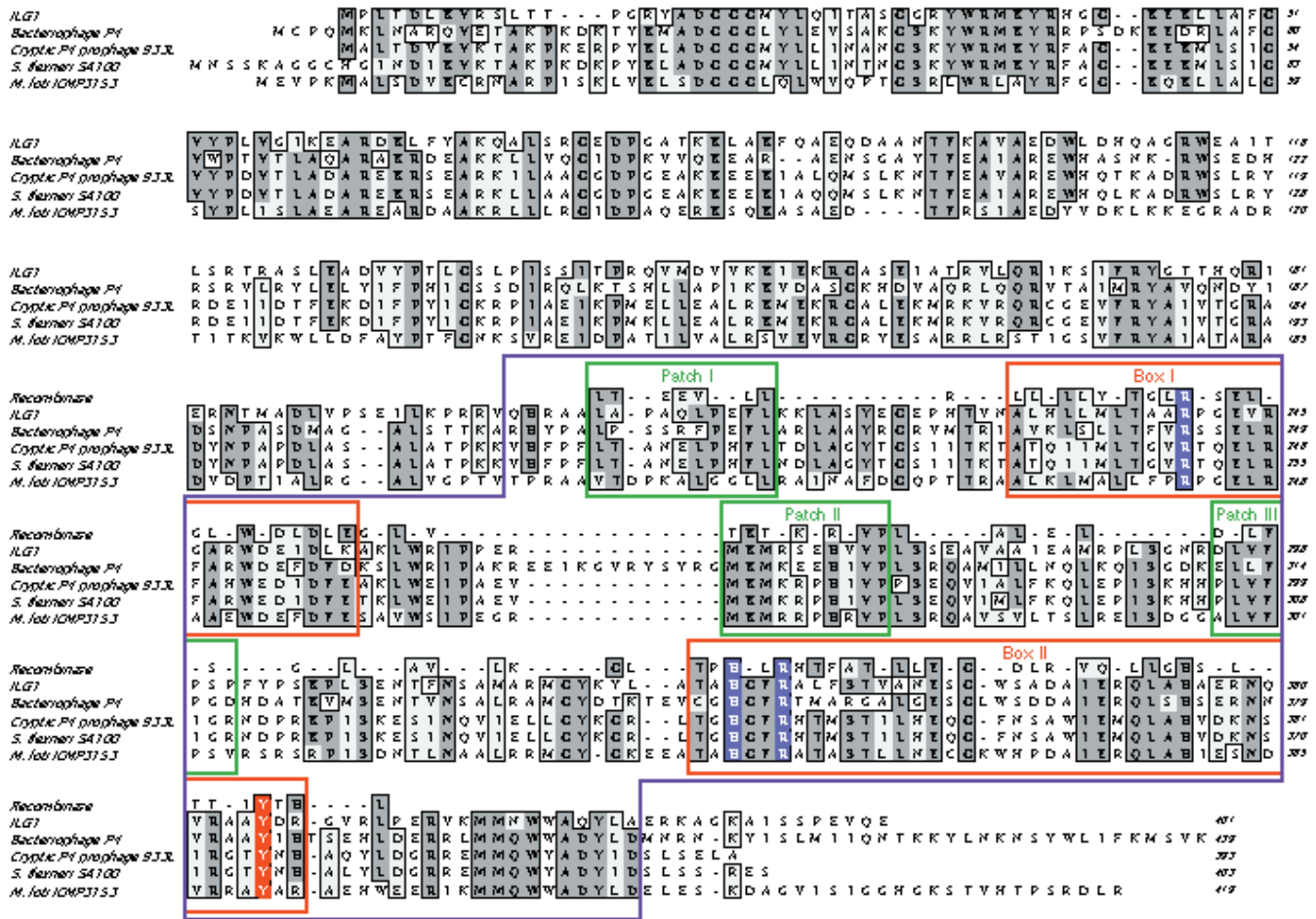


Fig. 3. *ILG1* is a site-specific recombinase belonging to the bacteriophage P4 integrase family. An alignment (ClustalW) of the complete amino acid sequences of *ILG1*, the four most similar integrases, and the site-specific recombinase consensus catalytic domain is shown using the BLOSUM30 similarity matrix. Amino acid position of the rightmost residue of each integrase is shown for each line. Among the aligned sequences, outlined groups of amino acids indicate identical matches (shaded) and conservative substitutions (>60%). The catalytic domain is outlined in purple. Highly conserved site-specific recombinase Box and Patch domains are boxed in red and green, respectively, with the nearly invariant R-H-R-Y hallmark shown in blue or red (active site) reverse text.

cultured from the mouth, face, and hands of the laboratory worker performing cDNA RDA. Probing these bacterial DNAs by Southern blot revealed that *ILG1* was present in TOP10F' and bacteria grown from the mouth and possibly face of the laboratory worker (Fig. 5B). We confirmed these results by dot blot, probing genomic DNA with *ILG1*, two *E. coli* sequences we obtained from the cDNA RDA subtraction (*frdA*, *lacl*), and 18S rRNA (Fig. 5B). *ILG1*, *frdA*, and *lacl* are found in TOP10F' and in bacteria cultured from the laboratory worker's mouth. 18S rRNA is found in human PBMC, but not in bacterial DNAs.

Discussion

ILG1 Belongs to the S Clade (Bacteriophage P4 Family) of Integrases

Using cDNA RDA, we have isolated *ILG1*, a new prokaryotic integrase-like gene with significant

amino acid sequence similarity to integrases of the bacteriophage P4 family (Table 3, Fig. 3), but not to other integrase families as detected by BLAST analysis. To confirm this conclusion and detect similarities to other integrase families, *ILG1* was compared directly to a collection of 24 different bacteriophage integrases by phylogenetic analysis (Alvin J. Clark, pers. com.). These integrases may be grouped by similarity into four different families: the D, Q, E, and S clades (35). This phylogenetic analysis confirms that *ILG1* belongs to the S clade, containing bacteriophage P4 integrases (Alvin J. Clark, pers. com.).

Understanding this family of integrases may be of medical importance, because they often are involved in the horizontal transfer between bacteria of large segments of DNA that promote pathogenicity (11). The next most related family of integrases is the E clade, represented by bacteriophage λ and P21 integrases. The Q clade, represented by four integrases

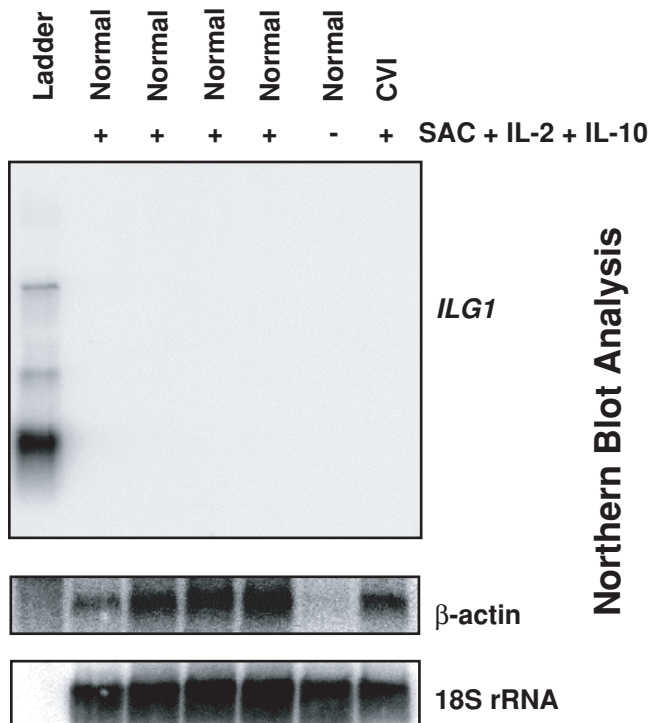


Fig. 4. *ILG1* RNA is not present in stimulated human B cells. RNA samples from CVI and normal B cells cultured with SAC, IL-2, and IL-10 were Northern blotted and probed for *ILG1* expression. The blot was stripped and reprobed for human β -actin and then 18S rRNA expression. Ladder is (0.24 to 9.5 kb) RNA ladder molecular size standards (Life Technologies).

including that from the cryptic *E. coli* RAC prophage, was the third most similar family. The least similar family was the D clade, represented by the bacteriophage P22 integrase.

ILG1 is a Site-specific Recombinase

Not only is *ILG1* similar to bacteriophage integrases, *ILG1* shares features with the integrase family of site-specific recombinases (Fig. 3). A comparison of these site-specific recombinases reveals a consensus sequence that preserves amino acid residues that generally also have a conserved position in those proteins whose crystal structures are available (9). These positions mostly comprise those in the hydrophobic core required for proper protein folding and those that comprise the catalytic active pocket. Mutational analysis supports the critical importance of the consensus amino acids for proper recombinase function. *ILG1* not only is similar to this recombinase consensus, which includes the box I and II, as well as patch I, II, and III regions, it exactly conserves the R-H-R-Y tetrad hallmark of this family (Fig. 3). The available crystal structures show that these critical residues form a catalytic pocket (36–39). The triad R-H-R residues form a cluster on the protein surface located at the center of interaction with the DNA substrate. The active site Y is found in the catalytic

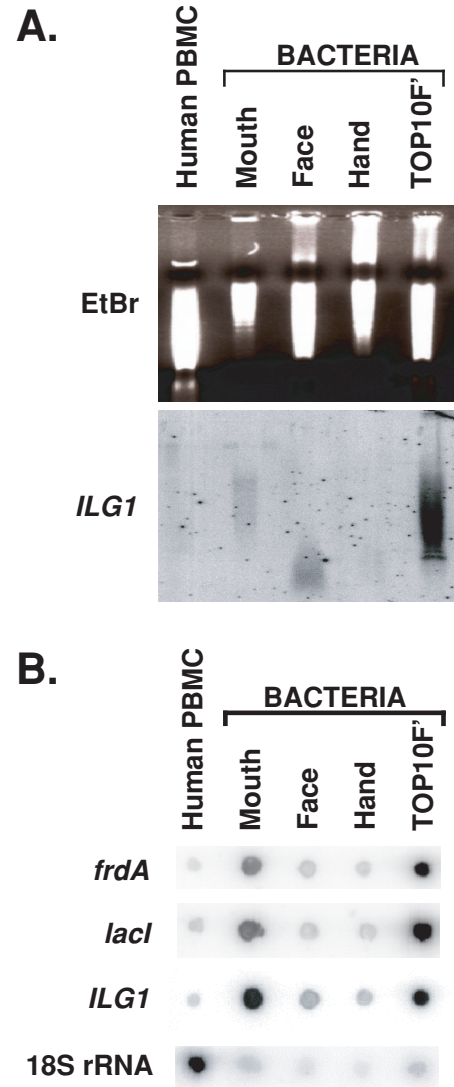


Fig. 5. *ILG1* is present in bacteria cultured from lab worker, but not in human genomic DNA. (A) Digested genomic DNA from human and bacterial samples was gel electrophoresed, stained with ethidium bromide (EtBr) and photographed. The gel was Southern blotted and probed with *ILG1*. (B) Digested genomic human and bacterial DNA was dot blotted and probed with bacteriophage *ILG1*. The blot was repeatedly stripped and probed with *E. coli frdA*, *lacI*, and human 18S rRNA.

pocket in a *cis* or *trans* conformation (8). In addition, *ILG1* has three additional conserved residues found in the region of the active site (9,38). 1) In box II, the H at position 355 in *ILG1* is strongly conserved among this family of recombinases. Based on crystal structures, this conserved position may contribute a hydrogen bond to the stabilization of the DNA substrate in the active site. 2) The strongly conserved first K residue in Patch II (K267 in *ILG1*) contacts the DNA substrate two nucleotides adjacent to the cleavage site. 3) In Box I, the first R (position 240 in *ILG1*) of the R-H-R-Y hallmark forms a water bridge with an adjacent strongly conserved D

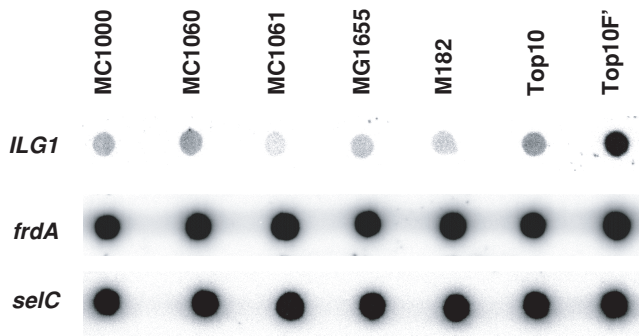


Fig. 6. *ILG1* was introduced recently into TOP10F'. Genomic DNA from indicated *E. coli* strains were dot blotted and probed with *ILG1*, *frdA*, or *selC* DNA fragments.

or E residue (E243 in *ILG1*). Since *ILG1* has conserved the critical residues composing the active site region, *ILG1* is likely an active site-specific recombinase.

Other *ILG1* Sequence Features

ILG1 not only shares the conserved site-specific recombinase features that includes Box I, II, Patch I, II, III, and the active site, but also shows some similarity in the intervening sequence between these regions (Fig. 3). In this site-specific recombinase family, the intervening sequences generally represent loops in the protein structure and show the greatest variability, including in length (9). For example, the separation between Box I and Patch II can vary from 6 to 48 amino acids. *ILG1* contains 10 amino acids separating these two regions, whereas bacteriophage P4 integrase contains 22. In this aspect, *ILG1* may be more similar to cryptic P4 prophage integrases. Despite this variation in length, the separation between all conserved Box and Patch regions in *ILG1* falls within previously determined ranges for the recombinase consensus (9).

Sequence variability of these intervening sequences may serve to modify the recombinase reaction. For example, the sequence between Patch III and Box II is critical for sequence-specific recognition of the particular DNA site for cleavage. Mutations in this region of the bacteriophage λ Integrase alter the preference for a particular DNA sequence (9). Unlike most of the intervening sequence, this particular area forms an α -helix. Hydrophilic residues (λ Integrase S282, G283, R287) of this amphipathic helix form the critical DNA recognition component. *ILG1* conserves hydrophobic residues that may be involved in formation of this helical structure (Fig. 3.). However, *ILG1* contains different hydrophilic residues (*ILG1* E310, N311, S315). Thus, *ILG1* is predicted to have a different DNA recognition sequence than that of bacteriophage λ Integrase.

Recent Introduction of *ILG1* into TOP10F'

ILG1 appears in TOP10F' after the introduction of the F' episome into TOP10 (Fig. 6). Since *ILG1* is most similar to bacteriophage P4 integrase family members that are involved in horizontal DNA transfer of large DNA modules (Table 3), we tentatively propose that *ILG1* resides on a cryptic prophage ϕ -TOP10F. Cryptic prophages become part of the DNA element involved in transfer of these large DNA modules between bacteria (11). In these elements, the cryptic prophage integrase gene is located next to the site of insertion, which is typically a tRNA gene. Thus, ϕ -TOP10F may reside on the F' episome or may have entered TOP10 by an independent ϕ -TOP10F bacteriophage infection or by a ϕ -TOP10F mediated horizontal transfer of DNA between bacteria. The origin of the F' episome is unknown (Invitrogen). To preliminarily test for independent ϕ -TOP10F insertion in the *E. coli* genome, we examined tRNA sites typically used by the cryptic bacteriophage P4 family. The *selC* tRNA gene is the site of insertion for cryptic P4 prophage 933L, bacteriophage ϕ -R73, and *Shigella* pathogenicity island 2 (24,25,27,28). The Phe and Asn tRNA genes are sites of insertion for *M. loti* symbiosis island and *Yersinia* High-pathogenicity islands, respectively (29,32). The *leuX* tRNA gene and the *ssrA* gene, which are sites of insertion for bacteriophage P4 and *V. cholera* colonization gene cluster, already contain prophage insertions as judged from the *E. coli* MG1655 genomic sequence. The F plasmid does not contain tRNA genes (Genbank Accession No. NC_002483). Therefore, we examined TOP10F' genomic DNA for insertions in the *selC*, *pheU*, *pheV*, *asnT*, *asnU*, *asnV*, and *asnW* by PCR across these genes. No insertions were found in these locations (data not shown). Thus, the proposed ϕ -TOP10F bacteriophage likely resides on the introduced F' episome or possibly another infrequently used tRNA insertion site.

Bacterial Contamination of cDNA RDA

Subtraction techniques often are challenged by contaminating "junk" DNA. In two separate cDNA RDA experiments, the frequency of this contamination was 35.9% and 12.9%. In one of these experiments, *ILG1*, a unique bacteriophage integrase-like gene found only in TOP10F', was isolated. This suggests that the contamination originated from this laboratory strain of *E. coli* that we use for routine cloning. TOP10F' appears to have colonized the laboratory worker, because *E. coli* genes and *ILG1* were detected in bacteria sampled from the laboratory worker. Thus, it appears that TOP10F' colonizing the laboratory worker contaminated the material that was eventually PCR amplified into amplicons. These amplicons were further amplified during five rounds of cDNA RDA, possibly increasing the frequency of contaminating DNA fragments. These multiple rounds of PCR have the potential to amplify significant

amounts of DNA fragments from even a small amount of bacterial contamination in the initial sample. Therefore, to prevent bacterial DNA contamination, extreme caution should be exercised when performing PCR based subtraction techniques, such as cDNA RDA. Prevention of bacterial DNA contamination in such subtractions will enrich the yield of relevant genes, which could result in the identification of new therapeutic targets.

Acknowledgements

Craig Gawel, Dorothy Guzowski, Colleen Millan and Dr. Robert Pergolizzi of the core facility at NSUH supplied oligonucleotides and performed DNA sequencing. We thank Drs. Nicholas Chiorazzi, James M. Mason, Mamta Naidu, Sangeeta Chavan, Yih-Pai Chu, Mary Berlyn, and Alvin J. Clark for discussion, encouragement, and critical review. This work was supported by the Muriel Fوسفeld Foundation, Leonard Wagner Autoimmunity Fund, S.L.E. Foundation, Inc., and the National Institutes of Health (AI44837-01).

References

- Klickstein LB. (1998) Production of a subtracted cDNA library. In: Ausubel FM, Brent R, Kingston RE, et al., eds., *Current Protocols in Molecular Biology*. New York: John Wiley & Sons, Inc., pp. 5.8.9–5.8.15.
- Orr SL, Gese E, Hood L. (1992) A new approach to understanding T cell development: the isolation and characterization of immature CD4-, CD8-, CD3- T cell cDNAs by subtraction cloning. *Mol. Biol. Cell* **3**: 761–773.
- Zeng J, Gorski RA, Hamer D. (1994) Differential cDNA cloning by enzymatic degrading subtraction (EDS). *Nucleic Acids Res.* **22**: 4381–4385.
- Yang M, Sytkowski AJ. (1996) Cloning differentially expressed genes by linker capture subtraction. *Anal. Biochem.* **237**: 109–114.
- Patel M, Sive H. (1998) PCR-based subtractive cDNA cloning. In: Ausubel FM, Brent R, Kingston RE, et al., eds., *Current Protocols in Molecular Biology*. New York: John Wiley & Sons, Inc., pp. 5.9.1–5.9.20.
- Chu CC, Paul WE. (1998) Expressed genes in interleukin-4 treated B cells identified by cDNA representational difference analysis. *Mol. Immunol.* **35**: 487–502.
- Argos P, Landy A, Abremski K, et al. (1986) The integrase family of site-specific recombinases: regional similarities and global diversity. *EMBO J.* **5**: 433–440.
- Gopaul DN, Duyne GD. (1999) Structure and mechanism in site-specific recombination. *Curr. Opin. Struct. Biol.* **9**: 14–20.
- Nunes-Duby SE, Kwon HJ, Tirumalai RS, et al. (1998) Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucleic Acids Res.* **26**: 391–406.
- Esposito D, Scocca JJ. (1997) The integrase family of tyrosine recombinases: evolution of a conserved active site domain. *Nucleic Acids Res.* **25**: 3605–3614.
- Ochman H, Lawrence JG, Groisman EA. (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**: 299–304.
- Murakawa Y, Strober W, James SP. (1991) Monoclonal antibody against the human peripheral lymph node homing receptor homologue (Leu 8) inhibits B cell differentiation but not B cell proliferation. *J. Immunol.* **146**: 40–46.
- Funderud S, Erikstein B, Asheim HC, et al. (1990) Functional properties of CD19+ B lymphocytes positively selected from buffy coats by immunomagnetic separation. *Eur. J. Immunol.* **20**: 201–206.
- Eisenstein EM, Chua K, Strober W. (1994) B cell differentiation defects in common variable immunodeficiency are ameliorated after stimulation with anti-CD40 antibody and IL-10. *J. Immunol.* **152**: 5957–5968.
- Altschul SF, Madden TL, Schaffer AA, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**: 3389–3402.
- Wilson K. (1998) Preparation of genomic DNA from bacteria. In: Ausubel FM, Brent R, Kingston RE, et al., eds., *Current Protocols in Molecular Biology*. New York: John Wiley & Sons, Inc., pp. 2.4.1–2.4.5.
- Kozak M. (1987) An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **15**: 8125–8148.
- Kozak M. (1991) An analysis of vertebrate mRNA sequences: intimations of translational control. *J. Cell Biol.* **115**: 887–903.
- Jacob WF, Santer M, Dahlberg AE. (1987) A single base change in the Shine-Dalgarno region of 16S rRNA of *Escherichia coli* affects translation of many proteins. *Proc. Natl. Acad. Sci. USA* **84**: 4757–4761.
- Barabino SM, Keller W. (1999) Last but not least: regulated poly(A) tail formation. *Cell* **99**: 9–11.
- Henkin TM. (2000) Transcription termination control in bacteria. *Curr. Opin. Microbiol.* **3**: 149–153.
- Nakamura Y, Gojobori T, Ikemura T. (2000) Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res.* **28**: 292.
- Halling C, Calendar R, Christie GE, et al. (1990) DNA sequence of satellite bacteriophage P4. *Nucleic Acids Res.* **18**: 1649.
- Sun J, Inouye M, Inouye S. (1991) Association of a retroelement with a P4-like cryptic prophage (retrophage phi R73) integrated into the selenocystyl tRNA gene of *Escherichia coli*. *J. Bacteriol.* **173**: 4171–4181.
- Perna NT, Mayhew GF, Posfai G, et al. (1998) Molecular evolution of a pathogenicity island from enterohemorrhagic *Escherichia coli* O157:H7. *Infect. Immun.* **66**: 3810–3817.
- Blattner FR, Plunkett G, 3rd, Bloch CA, et al. (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* **277**: 1453–1474.
- Moss JE, Cardozo TJ, Zychlinsky A, Groisman EA. (1999) The selC-associated SHI-2 pathogenicity island of *Shigella flexneri*. *Mol. Microbiol.* **33**: 74–83.
- Vokes SA, Reeves SA, Torres AG, Payne SM. (1999) The aerobactin iron transport system genes in *Shigella flexneri* are present within a pathogenicity island. *Mol. Microbiol.* **33**: 63–73.
- Buchrieser C, Brosch R, Bach S, et al. (1998) The high-pathogenicity island of *Yersinia pseudotuberculosis* can be inserted into any of the three chromosomal *asn* tRNA genes. *Mol. Microbiol.* **30**: 965–978.
- Gehring AM, DeMoll E, Fetherston JD, et al. (1998) Iron acquisition in plague: modular logic in enzymatic biogenesis of yersiniabactin by *Yersinia pestis*. *Chem. Biol.* **5**: 573–586.
- Kovach ME, Shaffer MD, Peterson KM. (1996) A putative integrase gene defines the distal end of a large cluster of ToxR-regulated colonization genes in *Vibrio cholerae*. *Microbiology* **142**: 2165–2174.
- Sullivan JT, Ronson CW. (1998) Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a *phe*-tRNA gene. *Proc. Natl. Acad. Sci. USA* **95**: 5145–5149.
- Grant SG, Jessee J, Bloom FR, Hanahan D. (1990) Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. *Proc. Natl. Acad. Sci. USA* **87**: 4645–4649.
- Bachmann BJ. (1996) Derivations and genotypes of some mutant derivatives of *Escherichia coli* K-12. In: Neidhardt FC,

- ed., *Escherichia coli and Salmonella: Cellular and Molecular Biology*. Washington, D.C.: ASM Press, pp. 2460–2488.
35. Clark AJ, Inwood W, Cloutier T, Dhillon TS. (2001) Nucleotide sequence of coliphage HK620 and the evolution of lambdoid phages. *J. Mol. Biol.* **311**: 657–679.
 36. Hickman AB, Waninger S, Scocca JJ, Dyda F. (1997) Molecular organization in site-specific recombination: the catalytic domain of bacteriophage HP1 integrase at 2.7 Å resolution. *Cell* **89**: 227–237.
 37. Kwon HJ, Tirumalai R, Landy A, Ellenberger T. (1997) Flexibility in DNA recombination: structure of the lambda integrase catalytic core. *Science* **276**: 126–131.
 38. Guo F, Gopaul DN, van Duyne GD. (1997) Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. *Nature* **389**: 40–46.
 39. Subramanya HS, Arciszewska LK, Baker RA, et al. (1997) Crystal structure of the site-specific recombinase, XerD. *EMBO J.* **16**: 5178–5187.