Revised Manuscript
No. JMM/2011/038398

Significant association of *dupA* of *Helicobacter pylori* with duodenal ulcer development in South East Indian Population

Jawed Alam[1], Sankar Maiti[2], Prachetash Ghosh[1], Ronita De[1], Abhijit Chowdhury[3], Suryasnata Das[4], Ragini Macaden[4], Harshad Devarbhavi[4], T. Ramamurthy[1], and Asish K. Mukhopadhyay[1]*

1) National Institute of Cholera and Enteric Diseases, Kolkata-700010, India; 2) IISER, Kolkata; 3) Centre for Liver Research, School of Digestive and Liver Diseases, Institute of Post Graduate Medical Education & Research, Kolkata; 4) St. John's Medical College Hospital, Bangalore

***Corresponding Author:** **Dr. Asish K. Mukhopadhyay**
Division of Bacteriology
National Institute of Cholera and Enteric Diseases,
P 33, CIT Road, Scheme XM, Beliaghata
Kolkata 700010
India
E-mail: asish_mukhopadhyay@yahoo.com
FAX: 91-33-2370-5066

Running Title: *dupA* of *H. pylori* associated with duodenal ulcer

Key words: *Helicobacter pylori*, duodenal ulcer, *dupA*, non-ulcer dyspepsia, disease association

Abstract

A novel virulence factor, duodenal ulcer promoting gene A (*dupA*) in *Helicobacter pylori* has been found to be associated with disease in certain population but not in others. The debate of relevance of *dupA* for the prediction of clinical outcome has prompted us to take this study in South East Indian population. A total of 140 *H. pylori* strains isolated from duodenal ulcer (DU) [n=83] and non-ulcer dyspepsia (NUD) subjects (n=57) were screened by PCR and Dot-Blot to determine the presence of *jhp0917* and *jhp0918*. Part of *jhp0917-0918* was sequenced to search for the C/T insertion that characterizes *dupA* and was also tested for *dupA* transcript. PCR and Dot-Blot results indicated presence of *jhp0917-0918* in 37.3% (31/83) and 12.2% (7/57) of *H. pylori* strains isolated from DU and NUD, respectively. Sequencing analysis showed an insertion of 'C' at position 1386 in 3'region of *jhp0917* forming *dupA* gene in 35 strains. RT-PCR analysis detected *dupA* transcript in 28 out of 35 strains. Expression level of *dupA* transcript varies from strain to strain as shown by Real Time PCR. Our study demonstrated that only PCR based analysis for *dupA* may furnish erroneous interpretation. Prevalence of *dupA* was significantly greater among strains isolated from patients with DU than NUD (P=0.001, OR=4.26, CI=1.60-11.74) in this population. Based on our finding, *dupA* can be considered as one of the biomarkers for DU patients in India. The reported discrepancy for this putative virulence-marker in different populations may be due to the genome plasticity of *H. pylori*.

Word Count: 250

Introduction

Helicobacter pylori is a gram-negative, spiral pathogen that infects more than 50% of the world's population (Brown *et al*., 2000). Infection with *H. pylori* plays an important role in development of peptic ulcer disease, distal gastric carcinoma and gastric mucosa-associated lymphoid tissue lymphoma (Megraud & Lamouliatte, 1992*;* Parsonnet *et al*., 1991; Wotherspoon *et al*., 1991). In India, around 65-70% populations are infected with the *H. pylori* (Graham *et al*., 1991; Singh *et al*., 2002). Overall, 15-20% of infected patients develop gastric or duodenal ulcer (DU) and less than 1% develop gastric adenocarcinoma. *H. pylori* infection is more prevalent in developing countries, and its incidence is decreasing in western countries (Czinn *et al*., 2005). The decisive factor(s) of *H. pylori*-mediated infection is still unclear. However, involvement of several virulence factors of the bacteria, host genetics and environmental influences are believed to determine the outcome of the infection. Among the host factors, pro-inflammatory cytokine gene polymorphisms have been associated with DU and gastric carcinoma (GC) (EI-Omar *et al*., 2000; Machado *et al*., 2001; Rocha *et al*., 2005).

Several bacterial virulence genes such as *vacA*, *cagA*, *babA* and *oipA* of *H. pylori* have been investigated to understand their association with gastroduodenal diseases (Covacci *et al*., 1993; Atherton *et al*., 1995; Yamoka *et al*., 1999, 2000, 2002; Argent *et al*., 2004). One possible problem that has complicated identification of definite disease-specific *H. pylori* virulence factors is the considerable geographic diversity in the prevalence of *H. pylori* virulence factors. Cytotoxin-associated gene (*cagA*) was the first reported gene that varies in *H. pylori* strains and considered as a marker for the presence of the *cag* Pathogenecity Island (*cag*-PAI), which include a number of other genes associated with

increased virulence (Broutet *et al*., 2001; Cenini *et al*., 1996; Rahman *et al*., 2003). However, none of the above mentioned virulence factors have exhibited any discriminating roles in the development of peptic ulcer versus GC. In addition to the *cag*-PAI, comparison of whole genome of two unrelated *H. pylori* (J99 and 26695) (Alm *et al*., 1999; Tomb *et al*., 1997), indicated presence of a hypervariable region called 'plasticity zone' with low G+C content along with strain specific open reading frames (ORFs). This plasticity region is 45 kb long, continuous in strain J99 and 68 kb discontinuous in strain 26695. As compared to 38 ORFs of the plasticity zone (*jhp0914-jhp0951)* in strain J99, 33 were absent in strain 26695) (Yamaoka., 2008; Pacheco *et al*., 2008; Yakoob *et al*., 2010; Kersulyte *et al*., 2003; Occhialini *et al*., 2000). Recently, a novel duodenal ulcer promoting gene (*dupA*) was described, which consists of two ORFs *jhp0917* and *jhp0918* and form one continuous gene by the insertion of a base T or C after the position 1385 of the *jhp0917* in the 3' region (Lu *et al*., 2005). This gene (homologues to *virB*4) is located in the plasticity region and is associated with increased risk of DU and protective against gastric atrophy, intestinal metaplasia and gastric carcinoma in Japan and Korea (Lu *et al*., 2005).

However, the role of *dupA* as a virulence marker is still controversial. Some researchers have supported the interpretations of Lu *et al.* (2005) but others did not find any association. Hussein *et al.* (2008) have reported that *dupA* gene is associated with peptic ulcer but they did not find any negative association with GC in Iraqi population. In Chinese and north Indian populations significant association of *dupA* with DU was established (Zhang *et al*., 2008; Arachchi *et al*., 2007). In contrast, Argent *et al.* (2007) showed no association of *dupA* gene with DU in population from Belgium, South Africa, China and the United States. Douraghi *et al.* (2008) showed no association of *dupA* gene with any clinical

outcome in Iranian population. Schmidt *et al.* (2009) identified no consistent association between *dupA* and DU or GC across the Swedish, Australian and ethnic Chinese, Indian and Malaysian population residing in Singapore and Malaysia. Meta-analysis based study by Shiota *et al.* (2010) has shown that the presence of *dupA* gene was significantly associated with DU. Another systematic review confirmed that *dupA* was associated with gastroduodenal diseases (Hussein, 2010).

There are also indications of significant geographic differences among strains. Indian *H. pylori* strains are genetically distinct than East Asian and Western strains (Mukhopadhyay *et al*., 2000). Moreover, our recent study showed that presence of strains with intact *cag* Pathogenicity Island was found more frequently in Kolkata than in Southern India indicating regional variation in the *H. pylori* gene pools (Patra *et al*., 2011). These considerations and our interest in the dynamics of genetic traits associated with *H. pylori* infection and disease association motivated us to conduct the present study for investigating the prevalence of *dupA* gene of *H. pylori* in duodenal ulcer and NUD patients isolated from South East Indian population and also to find out the association of *dupA* with the clinical outcome in a different setting.

Materials and Methods

Collection of Biopsy samples:

A total of 221 adult subjects of both genders (aged between 20 and 65 years) with upper gastrointestinal disorder underwent endoscopy at the hospital of the Institute of Post Graduate Medical Education and Research, Kolkata, and St. John's Medical College Hospital, Bangalore, India during the year 2006 to 2008. A detailed patient's history was taken, and a physical examination of each subject was carried out prior to endoscopy. The objective of the

study was explained to every individual and the informed consents were obtained from each individual under protocols approved by the ethical committees of the respective institutes based on the Helsinki Declaration. During endoscopy, two biopsies, one from antrum and the other from corpus of the stomach, were obtained from each subject. Biopsies taken in 0.6 ml of Brucella broth (Difco Laboratories, Detroit, MI) with 15% glycerol were transported to the laboratory in ice-cold condition and were stored at -70°C until culture.

***H. pylori* Culture:**

In the laboratory, Brucella broth containing the specimen was vortexed for 2 min and 200 µl of the mixture was streaked on Petri plates containing brain heart infusion (BHI) agar (Difco Laboratories) supplemented with 7% sheep blood, 0.4% IsoVitaleX, amphotricin B (8µg/ml) (Sigma Chemicals Co., St. Louis, MO), trimethoprim (5µg/ml), vanacomycin (6µg/ml) (Sigma Chemicals) and Nalidixic acid (8µg/ml) (all from Sigma). Plates were incubated for 3 days at 37^0C in a double gas incubator (Heraeus Instrument, Germany) which maintains an atmosphere of 5% O_2, 10% CO_2 and 85% N_2. . The *H. pylori* colonies were identified by their typical colony morphology, appearance on Gram staining and positive reactions in urease, catalase and oxidase tests along with the urease PCR. Bacteria were sub cultured at 37^0C on the above medium and under the same microaerophilic condition.

Extraction of genomic DNA:

Cells were harvested from the culture plates and washed with phosphate-buffer saline (pH 8.0) followed by centrifugation at 3000rpm for 1min. The pelleted cells were resuspended in 540µl of TE buffer (10mM Tris-HCL, 1mM EDTA), 60µl of 10% Sodium dodecylsulfate (SDS) (Sigma) and 9µl of Proteinase K (20mg/ml) (Invitrogen, Carlsbad, CA), mixture was incubated at 50ºC for 1 hour followed by addition of 100µl of 5M NaCl,

80µl of 10% CTAB solution and then again incubated at 65ºC for 10 minutes. The DNA was extracted according to the standard phenol-chloroform-method (Ausubel *et al*., 1993)

PCR amplification:

PCR amplification was performed in a final volume of 20 µl containing template DNA (2-20 ng), 2 µl of 10x Buffer (Roche, Germany), 2.5mM dNTPs (Roche) and 10 pmol of corresponding primers in the presence of 1U of Taq DNA Polymerase (Catalog no. 11435094001, Roche). The cycling program has the following condition: initial denaturation at 95ºC for 3 min followed by 30 cycles of denaturation at 94ºC for 1 min, annealing at 55ºC for 1 min, extension at 72ºC for 1 min, and final extension at 72ºC for 7 min. Genomic DNA from the strain J99 and 26695 were included as positive and negative control respectively. The PCR products were analyzed by 1.5% agarose gels (containing 0.5 µg of ethidium bromide per ml) in 1X TAE buffer. Gels were scanned under UV light and analyzed with Quantity One software (Bio-Rad, Hercules, CA). The size of product was confirmed by using DNA molecular size standard.

Dot-Blot Hybridization:

Dot-Blot was performed using the DNA extracted from all the strains to avoid false-negative results in PCR assay due to variation in the primer annealing sites. About 50 ng of purified DNA in 2 µl volume was spotted individually onto the Hybond N^+ membrane (Amersham Biosciences, U.K). The membrane was gently placed on 2ml of denaturation solution for 5 min; enough care was taken to avoid drowning of the membrane into the solution. The membrane was subsequently treated with 2 ml of neutralization solution for 5

min. Following neutralization, the membrane was air dried, and UV-cross linked (Bio-Rad). Parts of *jhp0917* and *jhp0918* were amplified by PCR using the primer sets Jhp0917F/jhp0917R and jhp0918F/jhp0918R respectively. (Table 1). The amplified fragments were purified with, QIAquick PCR Purification kit (QIAGEN, Germany) and used as a probe. The probes were labeled with alkaline phosphatase using the Gene image Alkphos Direct Labelling and Detection System (Amersham Biosciences). The membrane was then used for hybridization with DNA probes at 55ºC overnight and then washed with primary wash buffer having pH 7 and Magnesium salt 1 mM two times at 55°C for 10 mins each. Hybridization Blot was again washed with secondary wash buffer having pH 10 with 2 mM Magnesium salt twice at room temperature for 5 mins each and then finally developed. Based on the signal intensity as detected for positive and negative controls, presence or absence of genes in the test strains was assigned.

Nucleotide sequencing:

The intergenic region between *jhp0917* and *jhp0918* of 38 strains was amplified using four sets of primers: DupAsetIF/DupAsetIR; DupAsetIIF/DupAsetIR; DupAsetIIF/918R and DupAsetIF/DupAR (Table 1). The amplified products were purified using the QIAquick PCR Purification Kit (QIAGEN). The purified PCR product was quantified on 1% agarose Gel. The intensity of the band compared with λ *hind* III digest. The PCR purified products were sequenced with a Big Dye Terminator v3.1 cycle sequencing kit on an ABI PRISM 3100 genetic Analyzer (Applied Biosystem, USA). The sequences obtained in this study were deposited in GenBank under accession numbers: JN379045-JN379050.

Gene expression assay by RT-PCR

Total RNA of *H. pylori* was isolated with TRIzol reagent (Invitrogen, USA) according to the manufacturer's protocol and treated with DNase I (Ambion, USA) to remove DNA contamination. The absence of DNA contamination was checked by PCR with primer set ureBF/ureBR (Table 1) and quantified by measuring the absorbance at 260nm. 2µg of total RNA was reverse transcribed into cDNA with the RevertAid first strand cDNA synthesis kit (Fermentas, EU). The cDNA was then amplified with two different primer sets set5F/set5R, jhp0917F/jhp0917R (Table 1). All the cDNA samples were amplified with ureBF/ureBR (table1) to check the integrity of cDNA formed. Real time PCR was carried out in 12µl final volume with the step one plus (Applied Biosystem, USA) under the following the condition containing 75ng of cDNA, 10pmol each primer dupAsetIF/dupAsetIR, rpsTF/rpsTR (table 1) and Power SYBR Green master mix (Applied Biosystem, U.K). Polymerase activation at 95°C for 5 min followed by 40 cycle at 95°C for 15s, 55°Cfor 30s, and 62°C for 30s. Threshold cycle number (C_T) of triplicate reactions was determined using the stepone software v2.1, and the mean C_T of triplicate reactions was determined. The internal control gene *rpsT* was amplified simultaneously in separate reaction tubes under same condition. The levels of expression of the *dupA* genes were normalized against *rpsT* as $\Delta C_T = C_{T\ \mathrm{dupA}} - C_{T\ \mathrm{rpsT}}$. The relative expression of *dupA* gene in *H. pylori* strains were calculated as $2^{-\Delta\Delta C_T}$, where $-\Delta\Delta C_T = \Delta C_T$ (sample) - ΔC_T (reference). Strain I-77 was used as reference in the real time PCR assay.

Statistical analysis: A univariate analysis was performed to determine the risk of *dupA* in relation to clinical outcome. For univariate analysis, χ^2 test was used. A Probability levels (P) value of ≤ 0.05 was considered statistically significant.

Result:

A total of 140 *H. pylori* strains were isolated from the enrolled 221 subjects, who underwent endoscopy. Subject with abdominal discomfort, acidity, loss of appetite but no frank ulceration was considered as non-ulcer dyspepsia (NUD) but those have visible duodenal ulceration endoscopically were considered as duodenal ulcer (DU) patients. These strains were isolated from the following two groups: (i) 83 DU patients and (ii) 57 NUD. Out of 83 (51 male and 32 female) DU cases, the mean age difference was 46 ± 10.72 and 43.7 ± 9.36 and among 57 (36 male and 21 female) NUD, the mean age difference was 32.4 ± 7.22 and 33.13 ± 6.84 respectively. The genomic DNA from these 140 strains was used for further PCR based analysis.

Distribution of *jhp0917, jhp0918 , cagA and vacA*:

We first studied the presence of *jhp0917* and *jhp0918* in 140 strains from South East India, (83 with DU and 57 with NUD) using PCR and dot blot hybridization. *Jhp0917* was targeted with the specific primers jhp0917F/jhp0917R that yielded 307-bp amplicon. Similarly, *jhp0918* was also amplified with the gene specific primers jhp0918F/jhp0918R having amplicon size of 276-bp [Table 1 and Fig 1]. All strains that were positive for both the *jhp0917* and the *jhp0918* PCR were also positive in the dot blot hybridization. In addition, 5 PCR negative strains showed positive by dot blot due to strong binding of the probe (Fig 2). Hybridization results inferred that the interpretation of *jhp0917* and *jhp0918* positivity should not be considered only based on PCR result (Fig 2). All PCR and Dot-Blot Hybridization data indicated absence of both *jhp0917* and *jhp0918* in 68.5% (96/140) strains isolated from South-East India. Among the positive strains, three strains were positive only

for *jhp0917* and other three strains were positive only for *jhp0918*. Thus, PCR and Dot-Blot Hybridization results showed that 38 strains had both the ORFs.

The *cagA* and *vacA* status were determined using primers and protocols described earlier (Mukhopadhyay *et al*, 2000; Chattopadhyay *et al*. 2004). *cagA* was present in 92.1% (129/140) of the tested strains from this region. 70% (98/140) of the strains had *vacA* s1m1 allele. Other two alleles s1m2 and s2m2 of *vacA* were present in 17.1% (24/140) and 12.8% (18/140), respectively. Status of *cagA* and *vacA* gene in *jhp0917-0918* positive 38 strains isolated from South East Indian population was 34/38 (89.4%) and 22/38 (57.8%) respectively. Four strains isolated from DU patients were positive for *jhp0917-0918* but negative for *cagA* and had s2m2 allele of *vacA*.

Sequencing analysis

Four different primer sets were used for sequencing a small fragment of *jhp917* gene to search one base insertion that characterize *dupA* as single PCR set was unable to yield PCR fragment from all 38 strains. Using these four sets of primers, junction region of all the *jhp0917-jhp0918* positive 38 strains were sequenced and compared with the published sequence of strain J99 (GenBank access AE001439). We observed an insertion of C after position 1385 in the 3'region of *jhp0917* gene in 35 out of 38 clinical isolates. The remaining three isolates were same as J99 where no insertion of either T or C was found but still these strains were PCR and dot-blot positive with the primers described earlier indicating these were not the true *dupA* (Fig 3).

Expression of *dupA*

We performed the RT-PCR of 35 *dupA* positive strains and found that *dupA* transcript was present in 28 strains and absent in remaining 7 strains indicating that *dupA*

gene was not expressed in all strains (Fig 4A). Real time PCR were done with randomly selected *dupA* positive strains and we found that level of *dupA* transcript varies in different strains (Fig 4B).

dupA status from South-East Indian Population

In our study, 38 (31 DU and 7 NUD) clinical strains were positive for *jhp0917-0918* by PCR and dot-blot hybridization. We found by sequencing that 35 strains had insertion of C and three strains had no insertion of C/T like J99. During analysis by RT-PCR, we found that 28 strains out of 35 were positive in the RT-PCR. On the basis of sequencing and RT-PCR findings, we confirm that 28 (23 duodenal ulcer and 5 NUD) strains were *dupA* positive and 7 (5 duodenal ulcer and 2 NUD) strains were negative for *dupA*. Our study showed that the prevalence of *dupA* in duodenal ulcer patients 23/83 (27.7%) was significantly higher than the NUD 5/57 (8.7%). [P=0.001, OR= 6.49, 95% CI=1.71-28.94].

Discussion

Recent studies have proposed the possibility of using genetic markers in the plasticity zone as indicators of pathogenicity for *H. pylori* infection, in spite of a lack of credible knowledge regarding the functions of the putatively encoded proteins in this cluster. It seems that these determinants may play a key role in determining the virulence capacity of *H. pylori* strains either directly or by encoding factors that may lead to varying clinical outcomes. The association between some of the ORFs in the plasticity zone and various disease categories has been previously reported. For instance, Occhialini *et al.* (2000) found that two single ORFs (*jhp0940* and *jhp0947*) were more prevalent in strains isolated from patients with gastric adenocarcinoma in Costa Rica. However, Santos *et al.* (2003) showed the association between *jhp0947* and DU as well as GC in Brazilian patients. This was once

306 more confirmed for *jhp0947* and *jhp0949* genes in DU patients from the Netherlands (de
307 Jonge *et al*., 2004).

308 Our study in Southeast Indian population demonstrated that *dupA* gene was 6.5 times
309 more prevalent in duodenal ulcer patients than non-ulcer patients. Hence, *dupA* gene was
310 significantly associated (*P*=0.001) with DU. Associations between the presence of *dupA* and
311 *H. pylori*-diseases varies around the world (Nguyen *et al*., 2009, Argent *et al*., 2007, Gomes
312 *et al*., 2008, Hussein *et al*., 2010). Several issues starting from geographical variations to
313 study procedures have to be considered. In some studies, only one set of primer pairs for
314 *jhp0917* and *jhp0918* was used (Lu *et al*., 2005; Zhang *et al*., 2008; Douraghi *et al*., 2008;
315 Pacheco *et al*., 2008). Our study showed that 13% (5/38) dot blot positive strains for
316 *jhp0917-jhp0918* failed to provide any amplicon by initial set of primers. But later on,
317 different set of primers yielded amplicons of the same strains. Hence, use of multiple primer
318 pairs is recommended for detection of the *dupA* gene in future studies. Besides that,
319 sequenced based analysis showed that 7.8 % (3/38) *jhp0917-jhp0918* positive strains did not
320 have any insertion of C or T after position 1385 in the 3'region of *jhp0917* indicating that
321 they are not forming the *dupA.* This report is inconsistent with previous reports in other
322 populations, which indicated that all clinical isolates possessed a continuous *dupA* gene
323 (Douraghi *et al*., 2008; Schmidt *et al*., 2008; Gomes *et al*., 2008). Moreover, in our study not
324 a single strain was detected with the insertion of T after position 1385 of *jhp917*. RT-PCR
325 analysis showed that 20% (7/35) of the *dupA* positive strains did not show any *dupA*
326 transcript. This contradicted the findings of Nguyen *et al.* (2009) that *dupA* was always
327 expressed. Real time PCR analysis showed that expression level of *dupA* transcript varies
328 from strain to strain. A recent systematic review study demonstrated the importance of the

329 presence of the *dupA* gene for duodenal ulcer, especially in Asian countries (Shiota *et al.*
330 2010). Arachchi *et al*. (2007) showed that *dupA* gene was present in 37.5% and 22.8% of
331 DU and Functional Dyspepsia patients, respectively from North India, but in our study *dupA*
332 gene was present in 27.7% and 8.7% of DU and NUD patients, respectively. The reason of
333 this variation of *dupA* prevalence in India might be due to the fact that their study did not
334 include the sequencing of intergenic region of *jhp0917-918* to check the insertion of one
335 nucleotide after position 1385 and RNA expression profile of *dupA* gene or might be related
336 to the geographical genome variation of *H. pylori* as India is a big country with lots of
337 diversities. Some studies reported that *dupA* gene have single nucleotide polymorphism
338 (SNP) that created a premature stop codon and may have considerable effects on protein
339 expression or function (Gomes *et al*., 2008; Hussein *et al*., 2010; Queiroz *et al*., 2011;
340 Moura *et al*., 2012). Moreover, Douraghi *et al.* (2008) reported that *dupA* was inversely
341 associated with the histological feature dysplasia, a main pre-malignant and precancerous
342 lesion associated with increased incidence of cancer in Iranian population. As a result, *dupA*
343 gene may be applicable as a protective marker against GC development. But we were unable
344 to study this hypothesis as we did not have samples from gastric cancer patients. However, a
345 very recent study showed that the presence of a complete *dupA* cluster (type IV secretory
346 system with *vir* genes around *dupA*) seemed to be important in DU development (Jung *et al.*,
347 2012).

348 In conclusion, infection with the *dupA*-positive *H. pylori* increased the risk for DU
349 overall and this evidence was significant in Indian study. The gene *dupA* can be considered
350 as an important biomarker for DU in Indian population. However, further studies are
351 required to determine the functionality of *dupA* and its relationship with disease. The

discrepancy of *dupA* association with diseases outcome could be related to the limitation of PCR techniques for detecting the intact *dupA* gene or may be a consequence of the plasticity of *H. pylori*, which contributes to its genetic diversity and requires additional studies for a firm conclusion.

Acknowledgement

JA thanks Indian Council of Medical Research for a Senior Research Fellowship [No. 3/1/JRF/36/MPD/2007 (22588)]. The work was supported in part by the ICMR, Government of India, Program of Founding Research Center for Emerging and Reemerging Infectious Diseases, Ministry of Education, Culture, Sports, Science and Technology of Japan and Department of Biotechnology (No. BT/PR10407/BRB/10/604/2008).

References

Alm, R. A., Ling, L. S., Moir, D. T., King, B. L., Brown, E. D., Doig, P. C., Smith, D. R., Noonan, B., Guild, B. C. & other authors. (1999). Genome-sequence comparison of two unrelated isolates of human gastric pathogen *Helicobacter pylori*. *Nature* 397, 176-180.

Arachchi, H. S., Kalra, V., Lal, B., Bhatia, V., Baba, C. S., Chakarvarthy, S., Rohatgi, S., Sarma, P. M., Mishra, V. & other authors. (2007). Prevalence of Duodenal Ulcer-Promoting Gene (*dupA*) of *Helicobacter pylori* in patients with Duodenal Ulcer in North Indian Population. *Helicobacter* **12**, 591-597.

Argent, R. H., Burette, A., Miendje Deyi, V. Y. & Atherton, J. C. (2007). The presence of *dupA* in *Helicobacter pylori* is not significantly associated with duodenal

ulceration in Belgium, South Africa, China, or North America. *Clin Infect Dis* **45**, 1204-1206.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., & Struhl, K. (1993). *Current Protocols in Molecular Biology*, Greene Publishing and Wiley-Interscience, New York

Broutet, N., Marais, A., Lamouliatte, H., de Mascarel, A., Samoveau, R., Salamon, R. & Mégraud, F. (2001). CagA Status and eradication treatment outcome of anti-*Helicobacter pylori* triple therapies in patients with nonulcer dyspepsia. *J Clin Microbiol* **39,** 1319–1322.

Brown, L. M. (2000). *Helicobacter pylori*: epidemiology and routes of transmission. *Epidemiol Rev* **22**, 283-297

Censini, S., Lange, C., Xiang, Z., Crabtree, J. E., Ghiara, P., Borodovsky, M., Rappuoli, R. & Covacci, A. (1996). Cag, pathogenecity island of *Helicobacter pylori*, encodes typeI-specific and diseases associated virulence factors. *Proc NatI Acad Sci USA* **93**, 14648-14653.

Chattopadhyay, S., Patra, R., Ramamurthy, T., Chowdhury, A., Santra, A., Dhali, G. K., Bhattacharya, S. K., Berg, D. E., Nair, G. B. &. Mukhopadhyay, A. K. (2004). Multiplex PCR Assay for Rapid Detection and Genotypingof *Helicobacter pylori* Directly fromBiopsy Specimens. *J Clin Microbiol* **42,** 2821–282.

Czinn, S. J. (2005). *Helicobacter pylori* infection: detection, investigation, and management. *J Pediatr* **146**, S21-S26.

de Jonge, R., Kuipers, E. J., Langeveld, S. C., Loffeld, R. J., Stoof, J., van Vliet, A. H. & Kusters, J. G. (2004). The *Helicobacter pylori* plasticity region locus *jhp0947-*

jhp0949 is associated with duodenal ulcer disease and interleukin-12 production in monocyte cells. *FEMS Immunol Med Microbiol* **41**,161-167.

Douraghi, M., Mohammadi, M., Oghalaie, A., Abdirad, A., Mohagheghi, M. A., Hosseini, M. E., Zeraati, H., Ghasemi, A., Esmaieli, M. & other authors. (2008). *dupA* as a risk determined in *Helicobacter pylori* infection. *J Med Microbiol* **57**, 554-562.

EI-Omar, E. M., Carrington, M., Chow, W. H., McColl, K. E., Bream, J. H., Young, H. A., Herrera, J., Lissowska, J., Yuan, C. C. & other authors. (2000). Interleukin -1 polymorphisms associated with increased risk of gastric cancer. *Nature* **404**, 398-402.

Gomes, L. I., Rocha, G. A., Rocha, A. M., Soares, T. F., Oliveira, C. A., Bittencourt, P. F. & Queiroz, D. M. (2008). Lack of association between *Helicobacter pylori* infection with *dupA* positive strains and gastroduodenal diseases in Brazilian patients. *Int J Med Micribiol* **298,** 223-230.

Graham, D. Y., Adam, E., Reddy, G. T., Agarwal, J. P., Agarwal, R., Evans, D. J., Jr Malaty, H. M., & Evans, D. G. (1991). Seroepidemiology of *Helicobacter pylori* infection in India. Comparison of developing and developed countries. *Dig Dis Sc*i **36(8)**, 1084-1088.

Hussein, N. R. (2010). The association of *dupA* and *Helicobacter pylori*-related gastroduodenal diseases. *Eur J Clin Microbiol Infect Dis*. **29**, 817-821.

Hussein, N. R., Mohammadi, M., Talebkhan, Y., Doraghi, M., Letley, D. P., Muhammad, M. K., Argent, R. H & Atherton, J. C. (2008). Differences in Virulence Markers between *Helicobacter pylori* strains from Iraq and those from Iran: Potential importance of Regional differences in *H. Pylori*-Associated Diseases. *J Clin Microbiol* **46**, 1774-1779.

Jung, S. W., Sugimoto, M., Shiota, S., Graham, D. Y & Yamoka, Y. (2012). The intact *dupA* cluster is a more reliable *Helicobacter pylori* virulence marker than *dupA* alone. *Infect Immun* **80**, 381-387.

Kersulyte, D., Velapatino, B., Mukhopadhyay, A. K., Cahuayme, L., Bussalleu, A., Combe, J., Gilman, R. H. & Berg, D. E. (2003). Cluster of type IV secretion genes in *Helicobacter pylori's* plasticity zone. *J Bacteriol* **185**, 3764-3772.

Lu, H., Hsu, P. I., Graham, D. Y. & Yamoka, Y. (2005). Duodenal ulcer promoting gene of *Helicobacter pylori*. *Gastroenterology* **128**, 833-848.

Machado, J. C., Pharoah, P., Sousa, S., Carvalho, R., Oliveira, C., Figueiredo, C., Amorim, A., Seruca, R., Caldas, C. & other authers. (2001). Interlukin 1B and interleukin 1RN polymorphism are associated with increased risk of gastric carcinoma. *Gastroenterology* **121**, 823-829.

Megraud, F. & Lamouliatte, H. (1992). *Helicobacter pylori* and duodenal ulcer. Evidences suggesting causation. *Dig Dis Sci* **37**, 769-772.

Moura, S. B., Costa, R. F. A., Anacleto, C., Rocha, G. A., Rocha, A. M. C., &. Queiroz, D. M. M. (2012). Single Nucleotide Polymorphisms of *Helicobacter pylori dupA* that Lead to Premature Stop Codons. *Helicobacter* **17,** 176-180

Mukhopadhyay, A. K., Kersulyte, D., Jeong, J. Y., Datta, S., Ito, Y., Chowdhury, A., Chowdhury, S., Santra, A. & Bhattacharya, S. K. (2000). Distinctiveness of genotypes of *Helicobacter pylori* in Calcutta, India. *J Bacteriol* **182**,3219-3227.

Nguyen, L. T., Uchida, T., Tsukamoto, Y., Kuroda, A., Okimoto, T., Kodama, M., Murakami, K., Fujioka, T. & Moriyama, M. (2009). *Helicobacter pylori dupA* gene is

not associated with clinical outcome in the Japanese population. *Clin Microbiol Infect* **16**, 1264-1269.

Occhialini, A., Marais, A., Alm, R., Garcia, F., Sierra, R. & Megraud, F. (2000). Distribution of open reading frames of plasticity region of strain J99 in *Helicobacter pylori* strains isolated from gastric carcinoma and gastritis patients in Costa Rica. *Infect Immun* **68**, 6240-6249.

Pacheco, A. R., Proenca-Modena, J. L., Sales, A. I., Fukuhara, Y., da Silveira, W. D., Pimenta-Modena, J. L., de Oliveira, R. B. & Brocchi, M. (2008). Involvement of the *Helicobacter pylori* plasticity region and cag pathogenicity island genes in the development of gastroduodenal diseases. *Eur J Clin Microbiol Infect Dis* **27,** 1053–1059.

Parsonnet, J., Friedman, G. D., Vandersteen, D. P., Chang, Y., Vogelman, J. H., Orentreich, N. & Sibley, R. K. (1991). *Helicobacter pylori* infection and the risk of gastric carcinoma. *N Eng. J Med* **325**, 1127-1131.

Patra, R., Chattopadhyay, S., De, R., Datta, S., Chowdhury, A., Ramamurthy, T., Nair, G. B., Berg, D. E. & Mukhopadhyay, A. K. (2011). Intact *cag* Pathogenicity Island of *Helicobacter pylori* without Disease Association in Kolkata, India. *Int J Med Microbiol* **301,** 293-302**.**

Queiroz, D. M., Rocha, G. A., Rocha, A. M., Moura, S. B., Saraiva, I. E., Gomes, L. I., Soares, T. F., Melo, F. F., Cabral, M. M. & other authers. (2011). *dupA* polymorphisms and risk of *Helicobacter pylori*-associated diseases. *Int J Med Micriobiol* **301,** 225–228.

Rahman, M., Mukhopadhayay, A. K., Nahar, S., Dutta, S., Ahmed, M. M., Sarkar, S., Masud, I. M., Engstrand, L., Albert, M. J. & other authers. (2003). DNA level

characterization of *Helicobacter pylori* strains from patients with overt disease and with benign infection in Bangladesh. *J Clin Microbiol* **41,** 2008-2014.

Rocha, G. A., Guerra, J. B., Rocha, A. M., Saraiva, I. E., da Silva, D. A., de Oliveira, C. A, & Queiroz, D. M. (**2005).** IL1RN polymorphic gene and *cag*A-positive status independently increased the risk of non cardia gastric carcinoma. *Int. J. Cancer* **115**, 678-683.

Santos, A., Queiroz, D. M., Ménard, A., Marais, A., Rocha, G. A., Oliveira, C. A., Nogueira, A. M., Uzeda, M. & Mégraud, F. (2003). New pathogenicity marker found in the plasticity region of the *Helicobacter pylori* genome. *J Clin Microbiol* **41(4)**, 1651-1655.

Schmidt, H. M., Andres, S., Kaakoush, N. O., Engstrand, L., Eriksson, L., Goh, K. L., Fock, K. M., Hilmi, I., D., Dhamodaran, S. & other authers (2009). The prevalence of the duodenal ulcer promoting gene (*dupA*) in *Helicobacter pylori* isolates varies by ethnic group and is not universally associated with disease development: a case control study. *Gut Pathog* **1**, 5.

Shiota, S., Matsunari, O., Watada, M., Hanada, K. & Yamaoka Y. **(2010).** Systematic review and meta-analysis:the relationship between the *Helicobacter pylori dupA* gene and clinical outcomes. *Gut Pathog* **2**, 13.

Singh, V., Trikha, B., Nain, C. K., Singh, K., Vaiphei, K.(2002). **Epidemiology of *Helicobacter pylori* and peptic ulcer in India.** ***J Gastroenterol Hepatol*** 17(6)**, 659-65.**

Tomb, J. F., White. O., Kerlavage, A. R., Clayton, R. A., Sutton, G. G., Fleischmann, R. D., Ketchum, K. A., Klenk, H.-P., Gill, S. & other authors (1997). The complete genome of sequences of the gastric pathogen *Helicobacter pylori. Nature* **388,** 539-547.

Wotherspoon, A. C., Ortiz-Hidalgo, C., Falzon, M. R. & Isaacson, P. G. (1991). *Helicobacter pylori*-associated gastritis and primary B-cell gastric lymphoma. *Lancet* **338**, 1175-1176.

Yakoob, J., Abbas, Z., Naz, S., Islam, M., Abid, S. & Jafri, W. **(2010).** Associations between the Plasticity Region Genes of *Helicobacter pylori* and Gastroduodenal Diseases in a High-Prevalence Area. *Gut Liver* **4,** 345-350.

Yamaoka, Y. **(2008)**. Roles of the plasticity region of *Helicobacter pylori* in gastroduodenal pathogenesis. *J Med Microbiol* **57**, 545-553.

Zhang, Z., Zheng, Q., Chen, X., Xiao, S., Liu, W. & Lu, H. **(2008)**. The *Helicobacter pylori* duodenal ulcer promoting gene, *dupA* in China. *BMC Gastroenterol* **8**, 49.

Legends to Figures:

Figure 1: Schematic representation of *dupA* gene of clinical isolates with reference to strain j99 having region *jhp0917* and *jhp0918*. Different sets of primers pair with different location of *dupA* gene for the amplification of PCR products.

Figure 2: PCR and Dot-Blot Hybridization result (A) Lane1, strain j99 act as a positive control having amplicon size 307bp, lane 2-8 clinical isolates. (B) Lane 1strain J99 act as positive control 2, 5, 6 shows negative result by PCR but these same strains shows positive result by Dot-Blot Hybridization.

Figure 3: Nucleotide sequences alignment of partial sequences of *dupA* from position 1355bp to 1406bp, start codon of *dupA* gene is taken as position1. The asterisks show the positions where nucleotide sequences match and the hyphens represent deletions. Alignment was done by CLUSTAL 2.0.12 multiple sequence alignment tool. Shaded part shows variation.

Fig 4.

(A) Semiquantitative RT-PCR analysis of *dupA* gene of representative strains from India. 1st lane: 100bp marker. Lane 2-5 and lane 8-9 showed that *dupA* transcript was present but lanes 6-7 reflected absence of *dupA* transcript.

(B) *dupA* mRNA expression was analyzed in triplicate sample by qRT-PCR and data presented as mean (±S.D). Relative expression (*rpsT* normalized) of *dupA* gene in strains I-114, 217(4b), san77, 127(1a) compared with strain I-77.

532 **Table 1**: Primers used for the amplification and sequencing of *dupA* gene.

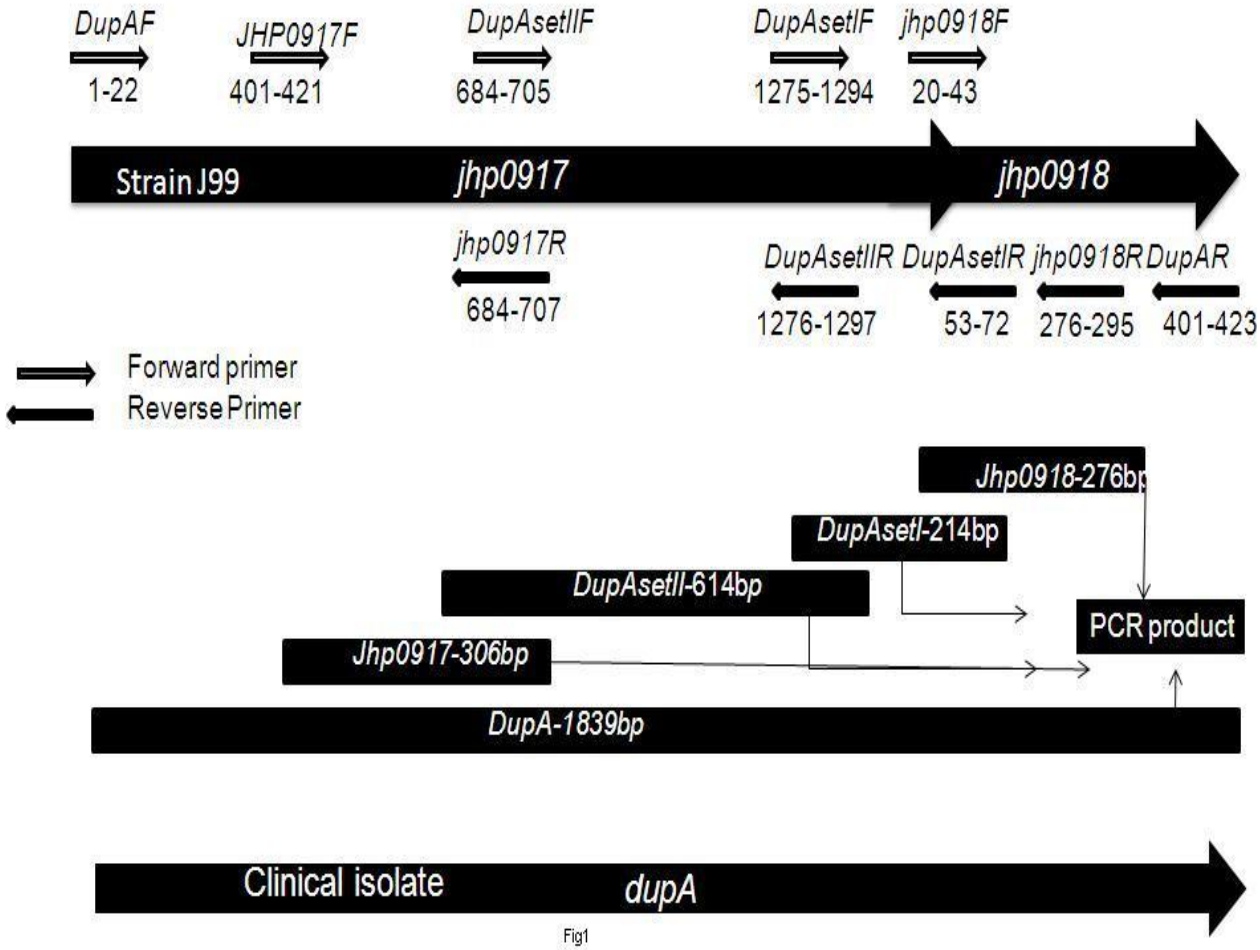
Primer	Sequence (5'-3')	Size (bp)	Annealing temp.(°C)	Reference
Jhp0917F Jhp0917R	TGGTTTCTACTGACAGAGCGC AACACGCTGACAGGACAATCTCCC	307	55	Lu et al.(2005)
Jhp0918F Jhp0918R	CCTATATCGCTAACGCGCGCTC AAGCTGAAGCGTTTGTAACG	276	55	Lu et al.(2005)
ureBF ureBR	CGTCCGGCAATAGCTGCCATAGT GTAGGTCCTGCTACTGAAGCCTTA	480	50	This study
set5F set5R	CTAGCGAACAAGATTTTAATGAGAT CCTAATTCTTTGACTTGAGATATT	350	56	This study
DupA setIF DupAsetIR	CGTGATCAATATGGATGCTT GCAAAGTGTTCCGTTGATCT	214	54	Gomes et al (2008)
rpsTF rpsTR	GGCAAATCATAAGTCCGCAGAA CTTTCCTAGAAGCGGTGTTTTTCT	217	55	This study
DupAsetIIF DupAsetIR	GGGAGATTGTCCTGTCAGCGTG GCAAAGTGTTCCGTTGATCT	805	58	This study
DupAsetIIF Jhp0918R	GGGAGATTGTCCTGTCAGCGTG AAGCTGAAGCGTTTGTAACG	1022	55	This study
DupA setIF DupAR	CGTGATCAATATGGATGCTT TTAAATACTCTTCCTTATAAGT	563	50	This study
cagA5cf cagA3cR	GTTGATAACGCTGTCGCTTCA GGGTTGTATGATATTTTCCATAA	350	55	Chattopadhyay et al (2004)

533
534 **Table 2:** Prevalence of *dupA*, *cagA* and *vacA* genes among the studied strains in India
535

	Total	DU	NUD
Number	140	83	57
dupA	28 (20%)	23 (27.7%)	5 (8.7%)
cagA	129 (92.1%)	76 (91.5%)	53 (92.9%)
vacA s1m1	98 (70%)	59 (71%)	39 (68.4%)

536
537

538



539 Fig1

540

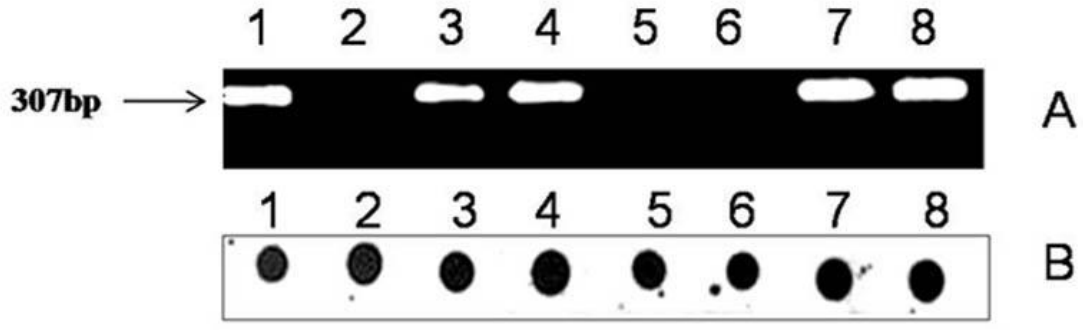


Fig2

541
542
543
544
545

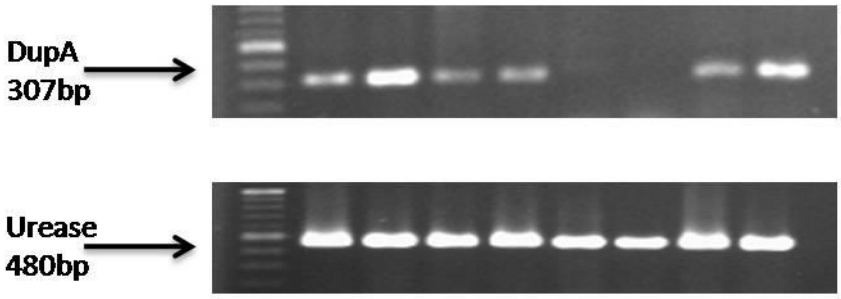
```
                1355bp                                                  1406bp

158(6b)         GAAGCAAAGGAGCACAACAAACCTTTTTTCCTATTCATTGATGAAACTAAA
I-121           GAAGCAAAGGAGCACAACAAACCTTTTTTCCTATTCATTGATGAAACTAAA
San74           GAAGCAAAGGAACACAACAAACCTTTTTTCCTATTCATTGATGAAACTAAA
216(1A)         GAAGCAAAGGAACACAACAAACCTTTTTTCCTATTCATTGATGAAACTAAA
osc17           GAAGCAAAGGAACACAACAAACCTTTTTTCCTATTCATTGATGAAACTAAA
157(1a)         GAAGCAAAGGAACACAACAAACCTTTTTTCCTATTCATTGATGAAACTAAA
135(1a)         GAAGCAAAGGAACACAACAAACCTTTTTTCCTATTCATTGATGAAACTAAA
156(2A)         GAAGCAAAGGAGCACAACAAACCTTTTTTC-TATTCATTGATGAAACTAAA
157(9b)         GAAGCAAAAGAACACAACAAACCTTTTTTC-TATTCATTGATGAAACTAAA
I-3             GAAGCAAAAGAACACAACAAACCTTTTTTC-TATTCATTGATGAAACTAAA
J99             GAAGCAAAAGAACACAACAAACCTTTTTTC-TATTCATTGATGAAACTAAA
                ******** ** ****************** ********************
```

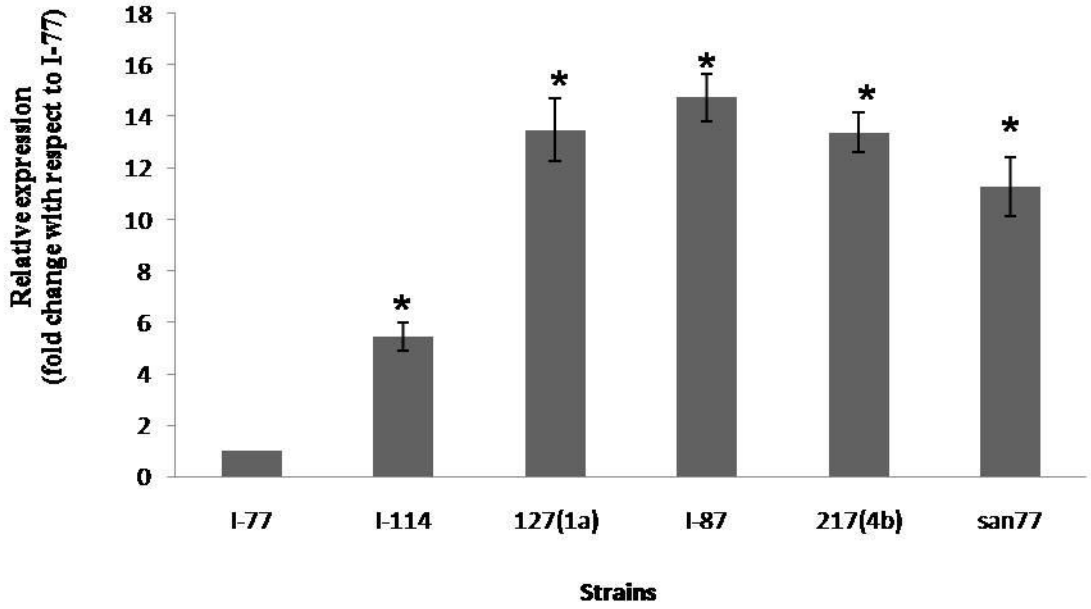
Fig.3

546
547
548
549

550
551



A



B

552
553 Figure 4