Supplemental Information

INVA PROTEIN: A NUDIX HYDROLASE REQUIRED FOR INFECTION BY PATHOGENIC *LEPTOSPIRA* IN CELL LINES AND ANIMALS*

Yihui Luo[1,2,7], Yan Liu[3,7], Dexter Sun[4], David M. Ojcius[5], Jinfang Zhao[1,2], Xuai Lin[1,2], Dong Wu[3], Rongguang Zhang[3], Ming Chen[6], Lanjuan Li[1], Jie Yan[1,2]

Supplemental Experimental Procedures

Leptospiral strains - Information about the leptospiral strains used in this study are presented in Table S1.

Primers- All the primers used in this study were synthesized by Invitrogen Co. Shanghai, China (Table S2).

Amplification and sequence analysis of invA genes in different leptospiral strains- Leptospiral DNA was extracted using a Bacterial Genomic DNA Extraction Kit (BioColor, Shanghai, China) and then dissolved in TE buffer. The density and purity of the extracted DNAs were measured by UV spectrophotometry. With a High Fidelity PCR Kit (TaKaRa, Dalian, China), in which the Taq-Pfu mixture was used as DNA polymerase, one pair of the primers N1F and N1R (Table S2) was used to amplify the entire *invA* gene. The product was detected in 1.5% ethidium bromide pre-stained agarose gels after electrophoresis. To obtain more accurate sequencing data, the PCR products were purified using a PCR Products Purification Kit (BioColor) and then ligated into plasmid pMD18-T using a T-A Cloning Kit (TaKaRa). The cloned *invA* genes from the leptospiral strains were sequenced by Invitrogen Co. (Shanghai, China). A phylogenetic tree from comparison among the actual InvA sequence of *L. interrogans* strain Lai and the sequences of other characterized microbial Nudix enzymes were analyzed using MEGA 4.0 software [1].

Supplemental References

1. Kumar, S., Nei, M., Dudley, J., and Tamura, K. (2008). MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform*. **9**(4), 299-306.

FIGURE LEGENDS

Fig. S1. Strategy for generation of the *invA*$^{-}$ mutant and *invA*^{com} revertant based on homologous recombination.
LA3978/P in the figure means that the DNA segment is composed of LA3978 gene and the promoter of *invA*-LA3976-LA3975 operon. See Methods for details.

Fig. S2. Distribution and sequencing of *invA* genes from *L. interrogans* species.
A: Detection of *invA* genes in different *Leptospira* species by PCR.
Lane M: DNA ladder (TaKaRa); Lane 1: blank control; Lanes 2 to 16: *invA* gene fragments amplified from the ten *L. interrogans* strains belonging to serovars Lai, Canicoda, Pyrogenes, Autumnalis, Australis, Pomona, Grippotyphosa, Hebdomadis, Paidjan and Wolffi, four *L. borgpetersenii* strains belonging to serovars Javanica, Ballum, Tarassovi and Mini, and one *L. weilii* strain belonging to serovar Manhao 2. Lanes 17 and 18: negative PCR results for the *L. biflexa* strains belonging to serovars Pactoc and Andamana.
B: Sequences of *invA* genes from different *Leptospira* species.
(1): Nucleotide and amino acid sequences of *invA* genes from the ten *L. interrogans* strains belonging to serovars Lai, Canicoda, Pyrogenes, Autumnalis, Australis, Pomona, Grippotyphosa, Hebdomadis, Paidjan and Wolffi, and four *L. borgpetersenii* strains belonging to serovars Javanica, Ballum, Tarassovi and Mini.
(2): Nucleotide and amino acid sequences of the *invA* gene from the *L. weilii* strain belonging to serovar Manhao 2. The underlined areas indicate the primer positions.

Fig. S3. Comparison of motif and phylogenesis among leptospiral InvA and other Nudix hydrolases.
A: Nudix motif and multiple alignments of leptospiral InvA and other bacterial Nudix hydrolases. The closest matches are to *R. prowazekii* InvA (No.Q9ZDT9 in NCBI), *B. bacilliformis* IalA (No. P35640 in NCBI), *H. pylori* NudA (No. O25826 in NCBI), *C. jejuni* dinucleoside oligophosphate hydrolase (No.Q9PHT5 in NCBI), *E. coli* YgdP (No. POA776 in NCBI), *P. aeruginosa* invasion protein homolog (No.Q9X4P2 in NCBI), *N. meningitidis* Ap_4A pyrophosphatase (No.Q9JT78 in NCBI), *H. influenzae* InvA (No.Q57045 in NCBI) and *S. typhimurium* hydrolase (No. P65555 in NCBI). The Nudix motif is indicated by "*", and identical amino acids are marked with a black background.
B: Molecular phylogenetic relationship between leptospiral InvA and other bacterial Nudix hydrolases.

Table S1. Leptospiral strains used in this study

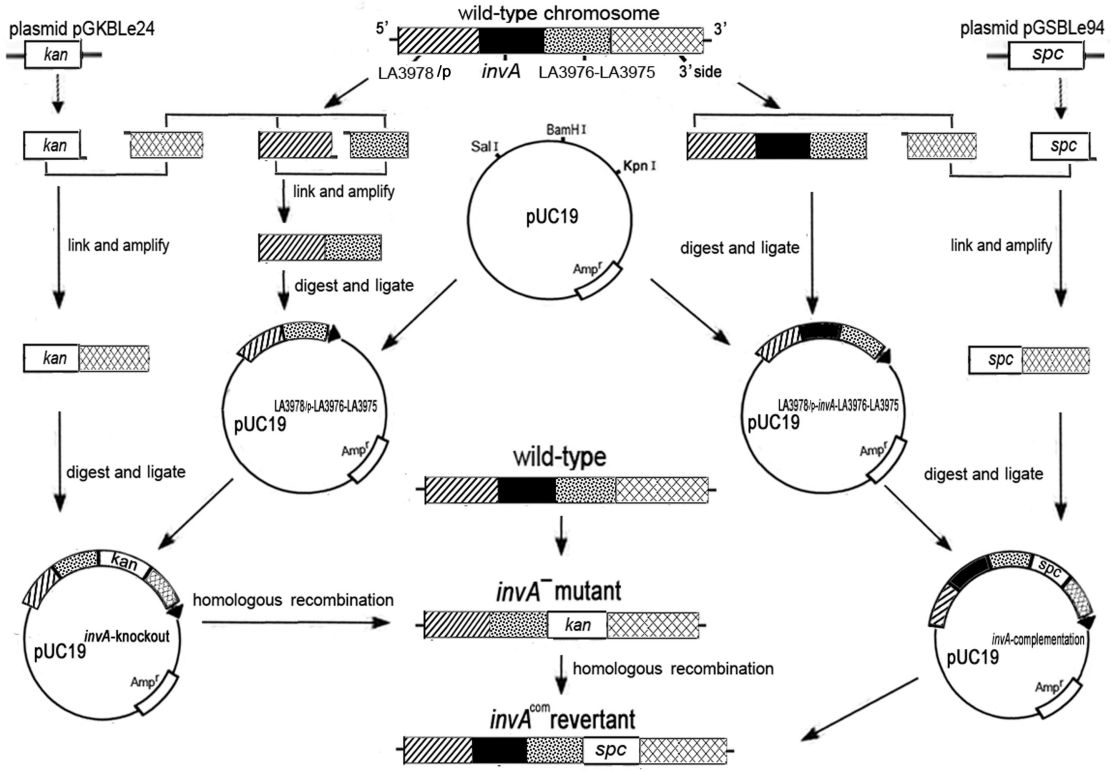
Strains	Serovars	Serogroups	Genospecies
Lai	Lai	Icterohaemorrhagiae	*L. interrogans*
Lin	Canicola	Canicola	*L. interrogans*
Tian	Pyrogenes	Pyrogenes	*L. interrogans*
Lin 4	Autumnalis	Autumnalis	*L. interrogans*
65-9	Australis	Australis	*L. interrogans*
Luo	Pomona	Pomona	*L. interrogans*
Lin 6	Grippotyphosa	Grippotyphosa	*L. interrogans*
P 7	Hebdomadis	Hebdomadis	*L. interrogans*
L 37	Paidjan	Bataviae	*L. interrogans*
L 183	Wolffi	Sejroe	*L. interrogans*
M 10	Javanica	Javanica	*L. borgpetersenii*
Pishu	Ballum	Ballum	*L. borgpetersenii*
55-52	Tarassovi	Tarassovi	*L. borgpetersenii*
Nan 10	Mini	Mini	*L. borgpetersenii*
L 105	Manhao 2	Manhao	*L. weilii*
Patoc I	Patoc	Semaraga	*L. biflexa*
CH11	Andamana	Andamana	*L. biflexa*

Table S2. Primers used in this study

Primer	Sequence (5' to 3')	Target	Size (bp)
N1	F: CGCCATATG (Nde I) GACAAACCCTACCGA R: CGCCTCGAG (Xho I) CGATCTATTTCCGATGTC	*invA* gene detection and cloning	486
N2	F: AAGATTCACTTTAAAATTAAA R: ATCATGAGAAATACGAAAACC	Identification of *invA*$^{-}$ mutant and *invA*^{com} revertant	4080 4741
N3	F: ATG GAC AAA CCC TAC CGA R: TGG AAA TTG CCA AGA ACC	*invA* mRNA detection by qPCR	150
K	F: CGCGGATCC(BamH I)ATCGGCTCCGTCGATACTATG R: agaatcgatattatacgtaggCATCAGAGTATGGACAGTTGC	Amplification of *kan* cassette	1062
L1	F: CGCGTCGAC(Sal I) CTAATTTCTTTCGATTGGATC R: aggtgggataaaaaaggacatACGAATAAGCTCCCGTTAACT	5'-homologous arm for *invA* gene knock-out	977
L2	F: agttaacgggagcttattcgtATGTCCTTTTTTATCCCACCT R: CGCGGATCC(BamH I)TCATTTCTTAATTTCAAAAGT	Amplification of LA3976 -LA3975 genes	715
S	F:CGCGGATCC(BamH I)AACGCGTCCCGAGCTTCAAGG R: agaatcgatattatacgtaggAACGCGTAAAGTAAGCACCTG	Amplification of *spc* cassette	1235
L3	F: gcaactgtccatactctgatgCCTACGTATAATATCGATTCT R: CGCGGTACC(Kpn I)GAAATCATACGACCCTGGAAG	3'-homologous arm for *invA* gene knock-out	1086
L4	F: caggtgcttactttacgcgttCCTACGTATAATATCGATTCT R: CGCGGTACC(Kpn I)GAAATCATACGACCCTGGAAG	3'-homologous arm for *invA* gene reversion	1086
16S	F: CTTTCGTGCCTCAGCGTCAGT R: CGCAGCCTGCACTTGAAACTA	16S rRNA for internal reference in qPCR	145
LA3976	F: CGGTCCACCTTTGTATTCTCAA R: AAACTTCCGCATCCCAATCGTC	Detection for mRNA level of LA3976 gene by qPCR	140
LA3975	F: GCTAAGCGGTTATACTCAA R: TCTTCCCAAGTAATCAGG	Detection for mRNA level of LA3975 gene by qPCR	147

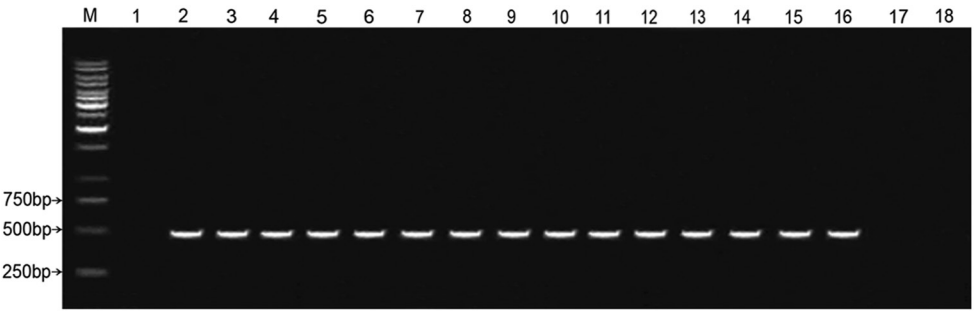
F: forward primer; R: reverse primer. Letters in lowercase indicate linking sequences.

Supplemental Figure S1



Supplemental Figure S2

A



B

Nucleotide sequence

```
(1)   1  ATGGACAAACCCTACCGAAAAAATGTCGGGATGGTCGTATTTAACTCTCGTGGAGAGGTT
(2)   1  ....................................................................

(1)  61  TTGGTTGGAGAAAGATTGAATTTTCTAGGTTCTTGGCAATTTCCACAAGGTGGAATTGAC
(2)  61  ....................................................................

(1) 121  GACGATGAAGATCCGATCAAGGCAGCCATGAGAGAATTATATGAAGAAGTCGGAATCGAT
(2) 121  ....................................................................

(1) 181  TCTGGAAAAATCGTAGCTGAATATCCAGATTGGATTTCCTATGACTTTCCCGAAAACCTT
(2) 181  ...................................T................................

(1) 241  CCTCTAAACCGTCATCTTCAAAAATATAGGGGGCAACTTCAAAAGTGGTTTCTTATCTAT
(2) 241  ....................................................................

(1) 301  TGGGACGGGGAAGTGGATCAATGTGATTTGGATATTCATGAAAGAGAATTTGGAACGGTT
(2) 301  ........A...........................................................

(1) 361  CGTTTTATTCCTATAAAAAACACGTTGAATACAGTCGTTCCTTTTAAAAAAGATGTATAT
(2) 361  ....................................................................

(1) 421  TATAAAATTGTAAATGACTTTGGGCCTAAGATCCAAAACTTTTTGCAAGACATCGGAAAT
(2) 421  ....................A...........................................

(1) 481  AGATCGTAA
(2) 481  .......
```

Amino acid sequence

```
(1)   1  MDKPYRKNVGMVVFNSRGEVLVGERLNFLGSWQFPQGGIDDDEDPIKAAMRELYEEVGID
(2)   1  .......................................................................

(1)  61  SGKIVAEYPDWISYDFPENLPLNRHLQKYRGQLQKWFLIYWDGEVDQCDLDIHEREFGTV
(2)  61  ............F..........................................................

(1) 121  RFIPIKNTLNTVVPFKKDVYYKIVNDFGPKIQNFLQDIGNRS
(2) 121  ...............................E.....................
```

Supplemental Figure S3