Research Journal of Applied Sciences 4 (2): 62-72, 2009
ISSN: 1815-932X

In silico Annotation of the Genes Involved in Biosynthesis of Lipopolysaccharide for *Burkholderia pseudomallei*

[1]Lai L. Suang, [1]Zamberi Sekawi, [1, 2]Nagi A. Al-Haj,
[1, 2]Mariana N. Shamsudin, [2]Rasedee Abdullah and [3]Rahmah Mohamed
[1]Department of Microbiology and Parasitology, Faculty of Medicine and Health Sciences,
University of Putra Malaysia, 43400 Serdang, Malaysia
[2]Laboratory of Immunotherapeutic and Vaccine,
[3]Department Marine Science and Aquaculture, Institute of Bioscience,
University of Putra Malaysia, 43400 Serdang, Malaysia

Abstract: *Burkholderia pseudomallei* is the causative agent of melioidosis, a serious disease of man and animals. The high mortality of *B. pseudomallei* infections may cause by lipopolysaccharides, an endotoxin. The biosynthesis of LPS is complex comprising three components, lipid A, core oligosaccharide and O-specific antigen. In the current study, by using the available *B. pseudomallei* genome database provided by Wellcome. The study demonstrated that the bioinformatics comparative technique was able to annotate LPS genes in *Burkholderia pseudomallei*. By developing a simple and easy flow chart including the using of Artemis software, total of 44 putative ORFs involved in biosynthesis of lipopolysaccharide for *B. pseudomallei* and the genetic mapping for the ORFs have been successfully determined using bioinformatics and laboratory approach. It is about 95.7% of success for annotation based on the 46 genes that act as references. In near future, a suitable vaccine or antimicrobial may be developed by targeting the genes encoding the various components essential in LPS biosynthesis and survival of the pathogen.

Key words: Melioidosis, *Burkholderia pseudomallei*, open reading frames, lipopolysaccharides, bioinformatics, data mining, genes

INTRODUCTION

Melioidosis is an infectious disease of humans and animals caused by *Burkholderia pseudomallei* (Dance, 1991). This organism is widely distributed in rice field soil and in stagnant water throughout the tropics. Although, a major disease in Southeast Asia and Northern Australia, melioidosis occurs sporadically throughout the world (often in patients with a history of residence in these disease-endemic areas) (Dance, 1991). Humans are usually infected by traumatic inoculation of the organism from the soil or, rarely, by inhalation or ingestion (Dance, 1991; Yee *et al*., 1988). LPS biosynthesis pathway involves a series of genes. Determining the genes function is a complicated task and is usually determined by comparative or homology analysis, which involves bioinformatics utilization. Bioinformatics is the analysis of biological information using computational approaches. The bioinformatics field has become world wide over the past few years, where it is a set of tools that may help researcher reach the ultimate truth with regards to the biological properties of living organisms (Christos, 2000). The bioinformatics approach involves determining the function of genes in LPS biosynthesis pathway of *B. pseudomallei*. The whole *B. pseudomallei* genome although, recently housed at the Sanger Institute in 2002 and annotated gene functions are published on the PNAS (Proceedings of the National Academy of Sciences, PNAS.0403302101) web site in 2004, the annotation was not complete when the present study was initiated. Therefore, the present study will present research outputs from manual annotation of genes involve in LPS biosynthesis of *B. pseudomallei*. Accurate annotation of the virulent genes is an essential element in supporting current drug discovery. Although, automatic annotate techniques is faster than manual process, traditionally manual annotation can attain high degree of accuracy (Alistair *et al*., 2002).

MATERIALS AND METHODS

Data mining: In this study, data mining process is carried to extract all possible genes involved in LPS biosynthesis pathway of *B. pseudomallei* by comparison from existed

Corresponding Author: Nagi Al-Haj, Laboratory Immunotherapeutic and Vaccine, Institute of Bioscience, University of Putra Malaysia, 43400, Serdang, Malaysia

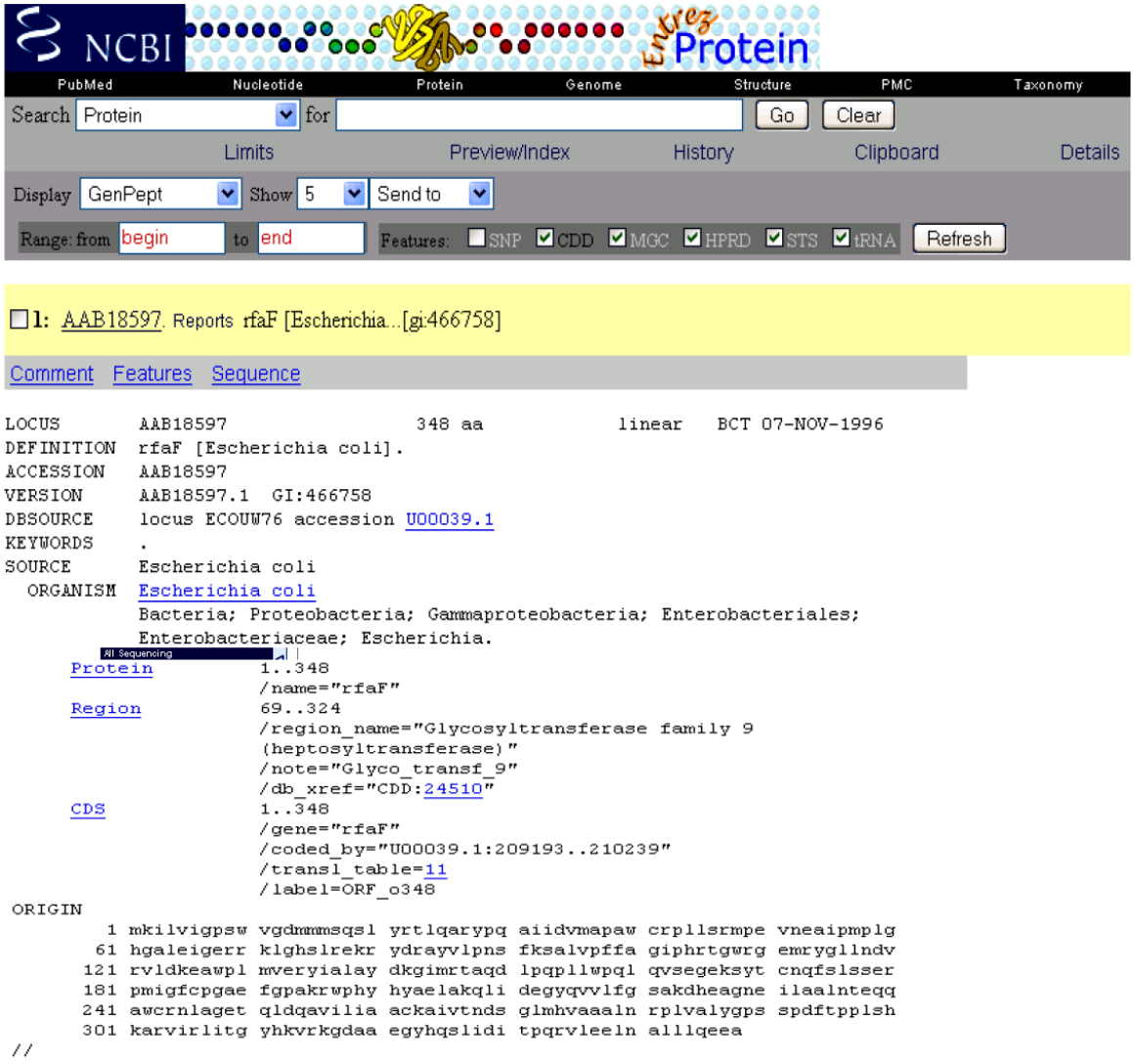

Fig. 1: GenBank® protein sequence database

databases. The initial stage of annotation is query search. Since, the whole genome of *B. pseudomallei* has been successfully sequenced, basically pair-wise sequence comparison method was used twice through out data mining to determine putative genes and annotated function in this study. The pair-wise comparison is a fundamental task in sequence analysis, providing the basis of database search algorithms, which seek to determine whether sequences are significantly similar and hence, whether, or not they are likely to be homologous. Prediction of genes and functions based on protein match and ab initio were included. By using Artemis software, DNA sequence of *B. pseudomallei* can be visualized. Artemis is an annotation tool that allows the results of any analysis or sets of analyses to be viewed in the context of the sequence and in six-frame translation. The query sequences were compared and identified homology matches to *B. pseudomallei* genome database (http://www.sanger.ac.uk/Projects/B_pseudomallei) using BLAST server (Fig. 1 and 2). From the BLAST result, those queries, which did not match or show low homologue to *B. pseudomallei* genome database were ignored (Fig. 3).

Homologue matching was done by BLAST the amino sequence of the query with *B. pseudomallei* genome database using BLAST server. The BLAST tool is prepared by The Wellcome Trust Sanger Institute through (http://www.sanger.ac.uk/Projects/B_pseudomallei).

The query here showed 43% identified to the subject. About 15-20 amino acid sequences of the subject was

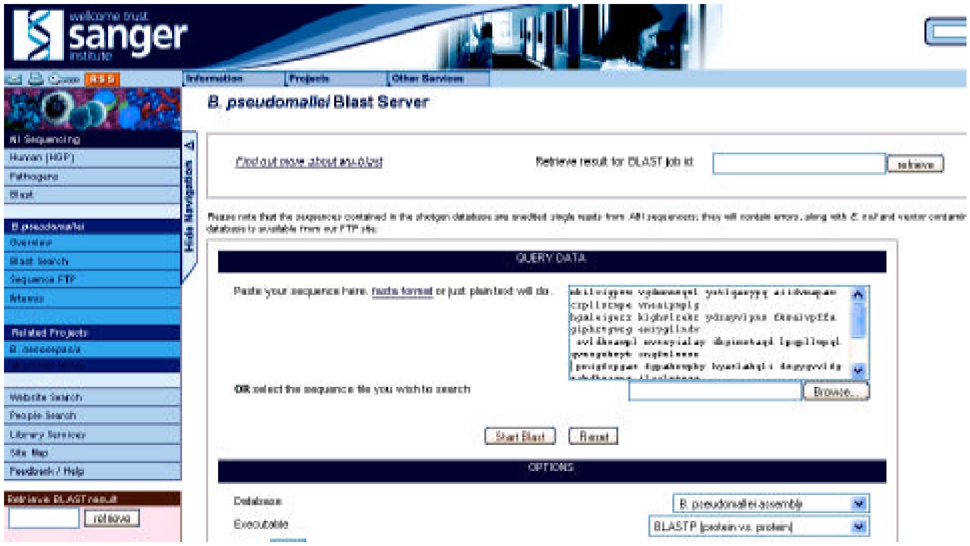

Fig. 2: *B. pseudomallei* database BLAST server

```
Query=  UNKNOWN-QUERY
        (348 letters)

Database:  BPs.pep
           5729 sequences; 1,991,322 total letters.
Searching....10....20....30....40....50....60....70....80....90....100% done

                                                                    Smallest
                                                                      Sum
                                                             High  Probability
Sequences producing High-scoring Segment Pairs:              Score  P(N)      N

BPSL0791 waaF putative ADP-heptose--LPS heptosyltransfera...  648  2.5e-65   1
BPSL1123 BPSL1123 putative LPS core biosynthesis-related ...  171  3.8e-12   1
BPSS0425 BPSS0425 putative heptosyltransferase (O-antigen...  122  6.5e-12   2
BPSS2252 BPSS2252 putative LPS biosynthesis related glyco...  160  6.2e-11   1
BPSL1120 BPSL1120 putative transferase (LPS biosynthesis-...  144  1.3e-10   2
BPSL2665 waaC1 lipopolysaccharide heptosyltransferase-1 3...  100  5.7e-08   2
BPSL2592 waaC2 lipopolysaccharide heptosyltransferase-1 3...   85  7.5e-07   2
BPSS2255 BPSS2255 glycosyltransferase family 9 protein 30...  113  2.2e-05   2
BPSS0426 BPSS0426 putative heptosyltransferase (O-antigen...   82  0.0039    2
BPSS1771 BPSS1771 putative decarboxylase 2429831:2430493 ...   64  0.83      1
BPSS2170 BPSS2170 conserved hypothetical protein 2934853:...   55  0.995     1


>BPSL0791 waaF putative ADP-heptose--LPS heptosyltransferase II 918996:920036
          reverse MW:37959
[Full Sequence]

        Length = 346

 Score = 648 (233.2 bits), Expect = 2.5e-65, P = 2.5e-65
 Identities = 147/341 (43%), Positives = 201/341 (58%)
[HSP Sequence]

Query:     2 KILVIGPSWVGDMMMSQSLYRTLQARYPQAIIDVMAPAWCRPLLSRMPEVNEAIPMPLGH 61
             + LVI P+W+GD +M+Q L+  L+  +P+ +ID +APAW  P+L RMPE+++     L H
Sbjct:     3 RALVIAPNWIGDALMAQPLFALLKKLHPRIVIDAVAPAWVAPVLERMPEIHDVHATELAH 62

Query:    62 GALEIGERRKLGHSLREKRYDRAYVLPNSFKSALVPFFAGIPHRTGWRGEMRYGLLNDVR 121
             G L++  R +L   LRE  YD AYVLPNS KSAL+P+ AGIP R G+ GE RY LLN VR
Sbjct:    63 GKLQMLHRWQLASDLRELGYDAAYVLPNSLKSALIPWLAGIPLRIGYTGEHRYALLN-VR 121

Query:   122 VLD----KEAWPLMVERYIALAYDKGIMRTAXXXXXXXXXXXXXVSEGEKSYTCNQFSLS 177
               +    +E    M + Y ALAY  G                ++E  +     +F+L
```

Fig. 3: The partial result of BLAST provided by BLAST server (http://www.sanger.ac.uk/Projects/B_pseudomallei)

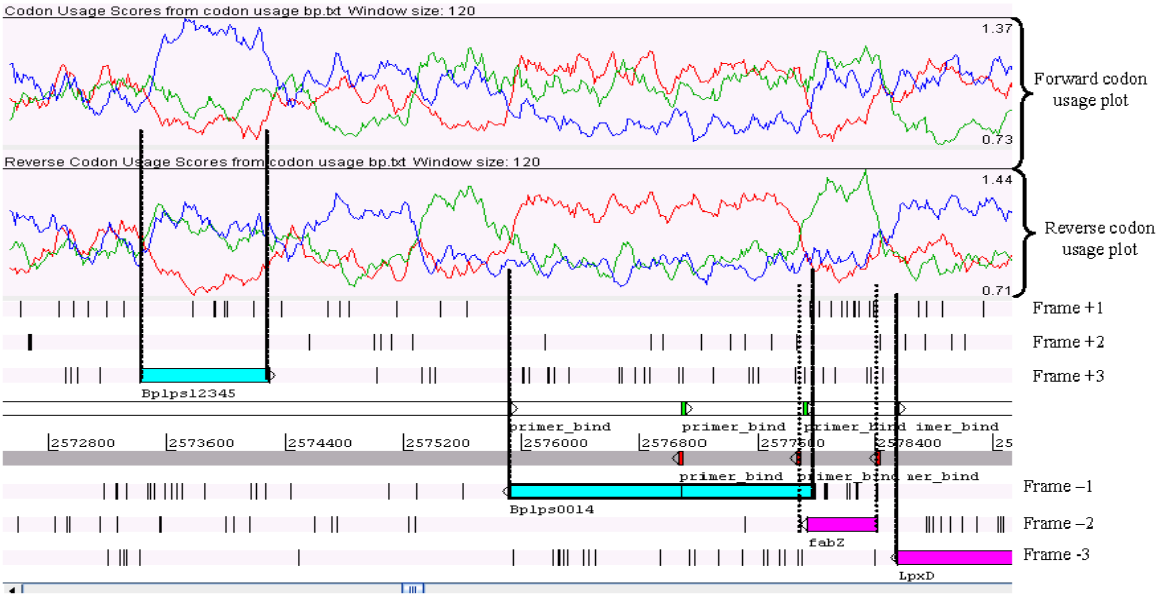

Fig. 4: Codon usage plot based on codon usage frequency for *B. pseudomallei*

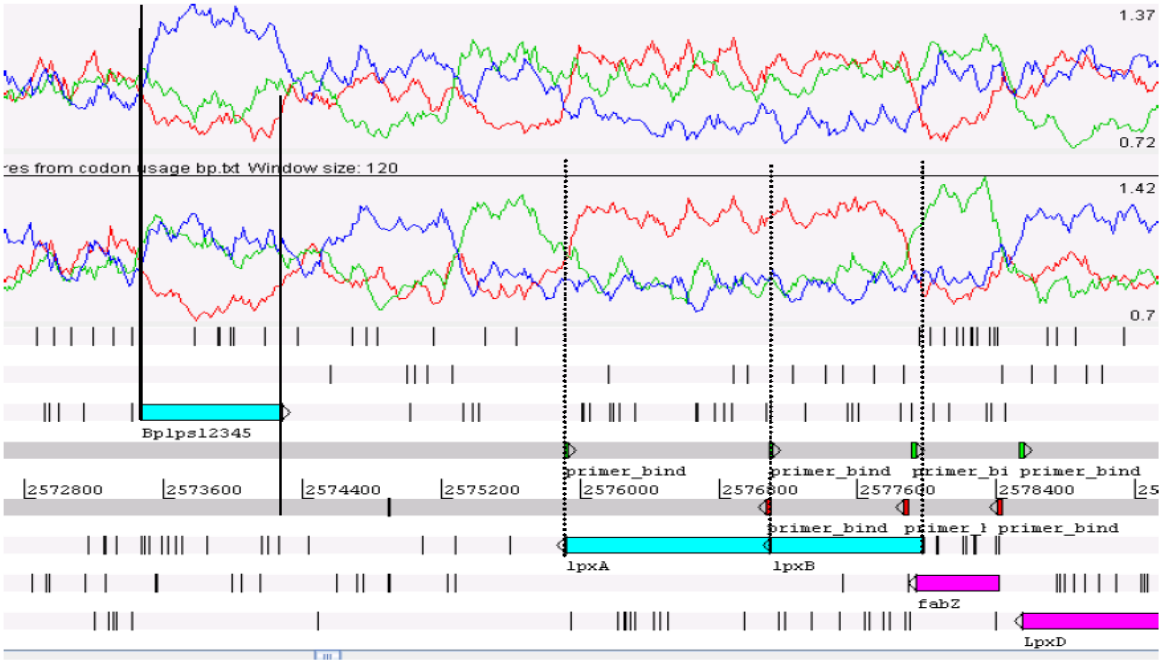

Fig. 5: Triming of annotated ORF using Artemis program ORFs were trimed according to both homology matching and codon usage plot

copied and preceded to Artemis program to identify the location and function of the ORF in *B. pseudomallei* genome.

Genetic mapping for lipopolysaccharide gene using Artemis: The *B. pseudomallei* genome sequence is visualized by using Artemis software, which is available under the GNU General Public License from http://www.sanger.ac.uk/Software/Artemis. The amino acid sequence from *B. pseudomallei* database, which hit and exhibit homologous to query sequence were fished out and the gene location in the genome was identified by using Artemis software. A few amino acid sequences, about 15-20 amino acid. Were copied from that particular

homologue subject search and proceeded to Artemis program for identifying the position and function of that ORF in *B. pseudomallei* genome using Artemis navigator tool. An ORF start with ATG, include the navigator sequence, until the stop codon was annotated as the putative ORF. In addition, *ab initio* analysis on these fishing out ORFs (open reading frame) were done using codon usage plot. Codon usage frequency for *B. pseudomallei* was used in codon usage plot. *Ab initio* gene prediction is relying on the statistical qualities of exons rather than on homologies. Then, ORFs, which were fished out were trimmed to the size referring to the codon usage scores from *B. pseudomallei* codon usage frequency (Fig. 4 and 5). Once, the pattern was recognized, the amino acid sequence of individual predicted ORFs were analyzed and the homolog matches in NCBI GenBank (PSI-BLAST) was identified. The predicted ORFs were examined individually for the similarity with well-characterized genes from other bacteria. The function and conserve domain were then identified. A few amino acid sequences of homologue subject from BLAST server were copied and pasted in Artemis navigator for identify the position of that ORF in *B. pseudomallei* genome.

Total of 6 frames can be view using Artemis software. Three forward codon usage plot, which are red, green and blue represent frame +1, +2 and +3, respectively. Same as the reverse codon usage plot, which are green blue and red represent frame -3, -2 and -1.

Confirmation of the *in silico* data using molecular based method

***Burkholderia pseudomallei* stock:** All *B. pseudomallei* culture works were done in Pathogen Laboratory, Department of Biochemical, University Kebangsaan Malaysia laboratory. *B. pseudomallei* strain D286 isolate was culture on Ashdown medium, a selected medium for *B. pseudomallei*. The picture took by research assistant (Lim Boon San) from Pathogen Laboratory, UKM. The biochemical test of *B. pseudomallei* was done by using Microbact™ 24E Gram Negative Bacteria Confirmation Kit. This test also carried out by Lim Boon San.

RESULTS AND DISCUSSION

The data mining through the gene prediction method and homology search using LPS genes from other Gram-negative bacteria successfully annotated a total of 44 putative genes involve in LPS biosynthesis pathway of *B. pseudomallei*. Each putative ORF was given an identity according to the location of genes in *B. pseudomallei* genome. Table 1 showed the properties of the annotated genes, which include the size of amino acid, similarity of protein in database, putative function, expected homology value or similarity and conserve domains. The lowest similarity of putative ORF was Bplps0008/*wzyC* whereby the protein sequence was 27% similar to lipid A core-O-antigen ligase from Magnetospirillum magnetotacticum. There were 12 ORFs, which showed 100% homology to the respective enzyme protein from *B. pseudomallei*. The ORFs are Bplps0024/*rfb(orf1)*, Bplps0025/*galE*, Bplps0027/*wbiH*, Bplps0033/*wbiB*, Bplps0035/*wzt*, Bplps0037/*rmlD*, Bplps0038/*rmlC*, Bplps0039/*rmlA*, Bplps0040/*rmlB*, Bplps0041/*apaH*, Bplps0042/ *plsC* and Bplps0043/*pyrC*. The data mining result also showed distribution of the 44 putative ORFs along the *B. pseudomallei* genome. Several ORFs were located in a cluster while, oftener were located individually. Figure 6-17 showed the distribution of ORFs along the *B. pseudomallei* genome. Total of 38 ORFs were considered located in the clusters (Fig. 8, 9, 11, 13, 15 and 16) while, 6 were located individually (Fig. 6, 7, 10, 12, 14 and, 17).

Table 1: List of the ORFs that were used in the PCR study

ORF identity	Putative gene name	Product length (bp)
Bplps0001	*lpxL*	943
Bplps0002	*waaF*	1124
Bplps0003	*fabG*	780
Bplps0004	*adk*	672
Bplps0005	*kdsB*	861
Bplps0006	*lpxK*	1072
Bplps0007	*rfaF*	1195
Bplps0008	*wzyC*	1315
Bplps0009	*waaB*	1293
Bplps0010	*rfaQ*	1108
Bplps0011	*dpm1*	1043
Bplps0012	*lpxB*	1182
Bplps0013	*lpxA*	795
Bplps0014	*fabZ*	514
Bplps0015	*lpxD*	1098
Bplps0016	*fabH*	994
Bplps0018	*waaE*	1062
Bplps0019	*udg*	1511
Bplps0020	*waaC2*	1094
Bplps0021	*waaA*	1353
Bplps0022	*waaC1*	1032
Bplps0023	*wbyC*	1173
Bplps0026	*wbiI*	1644
Bplps0034	*wbiA*	657
Bplps0035	*wzt*	1604
Bplps0036	*wzm*	868
Bplps0044	*lpxC*	968

Genetic mapping for lipopolysaccharide gene using artemis: The annotated ORFs (open reading frame) involve in LPS biosynthesis pathway were viewed through the Artemis program. Figure 6-17 showed the location of 44 annotated ORFs (pink color) in *B. pseudomallei* genome based on homologues result view by Artemis 5.0 program. The program showed 6

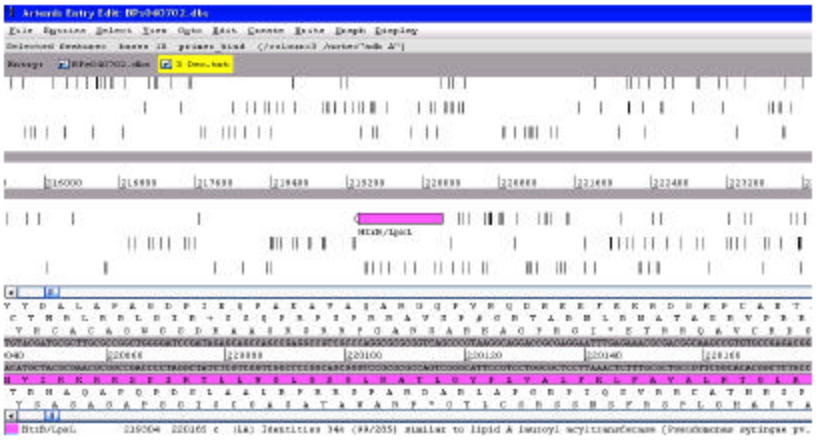

Fig. 6: The location of Bplps0001/lpxL/*htrB* gene, The location of Bplps0001/*lpxL*/*htrB* (219304..220185) gene in *B. pseudomallei* genome showed in reverse frame

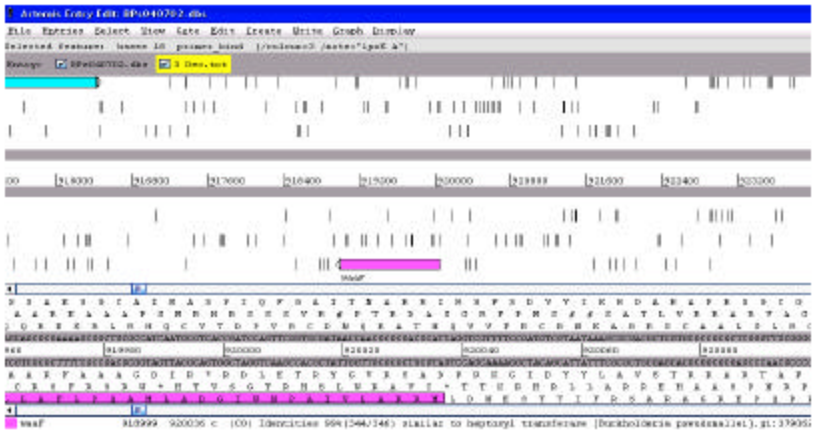

Fig. 7: The location of Bplps0002/*waaF* gene. The location of Bplps0002/*waaF* (918999..920036) gene in *B. pseudomallei* genome showed in reverse frame

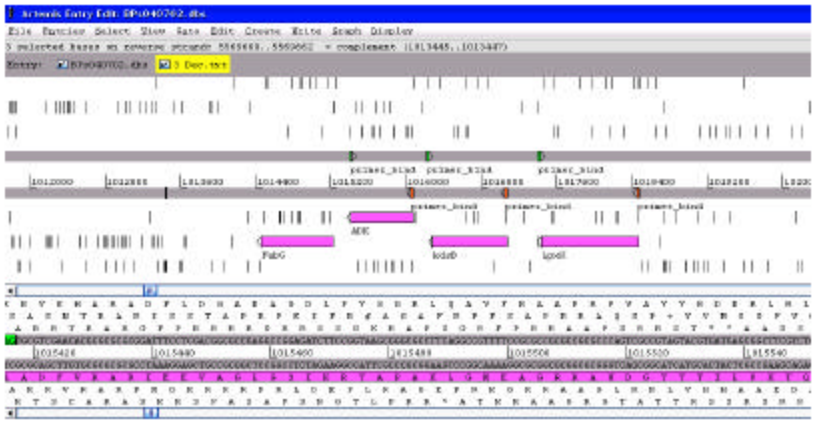

Fig. 8: The location of Bplps0003/*fabG*, Bplps0004/*adk*, Bplps0005/*kdsB* and Bplps0006/*lpxK* genes. The location of Bplps0003/*fabG* (1014455..1015210), Bplps0004/*adk* (1015411..1016070), Bplps0005/*kdsB* (1016282..1017070) and Bplps0006/*lpxK* (1017434..1018459) genes in *B. pseudomallei* genome showed in reverse frame

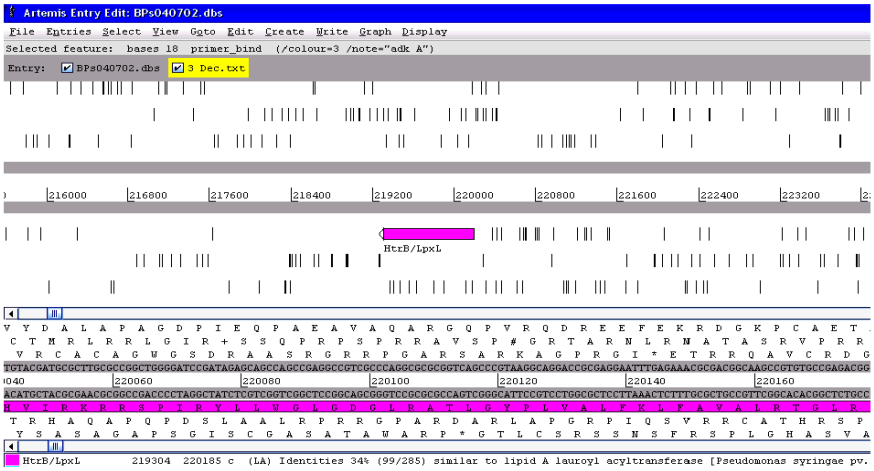

Fig. 9: The location of Bplps0007/*rfaF*, Bplps0008/*wzyC*, Bplps0009/*waaB* and Bplps0010/*rfaQ* genes. The location of Bplps0007/*rfaF* (1302531..1303673), Bplps0008/*wzyC* (1303673..1304968), Bplps0009/*waaB* (1304968..1306227) and Bplps0010/*rfaQ* (1306440..1307486) genes in *B. pseudomallei* genome showed in forward frame

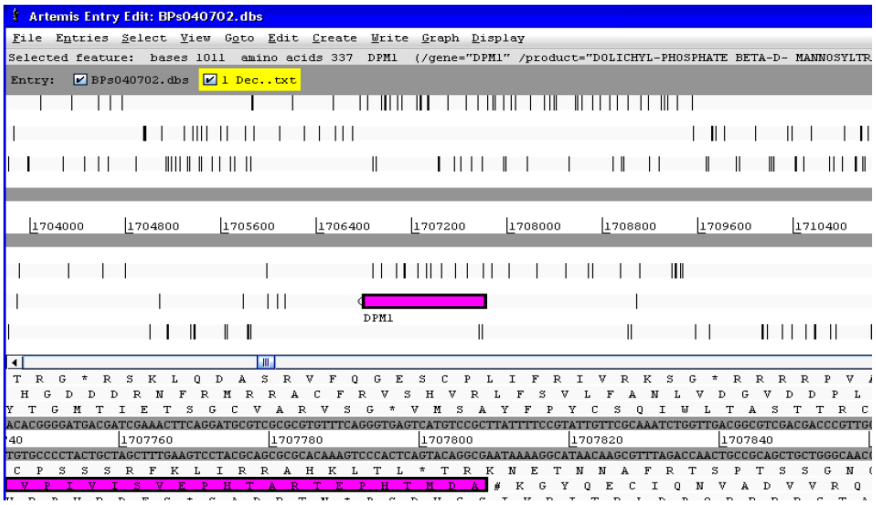

Fig. 10: The location of Bplps0011/*Dpm1* gene. The location of Bplps0011/*Dpm1* (1706798..1707808) gene in *B. pseudomallei* genome showed in reverse frame

possible reading frames: three read from 5-3' direction which showed in the upper three lines and three read from the reverse direction (5-3') of the complementary sequence, which are shown in the lower three lines. Therefore, not only the annotated ORFs locations are shown, the frames were also identified. Besides, from the main page of the program, the relative length of ORFs can be observed. The sequence of each ORF was trimmed to give the suitable start codon (ATG) and end with stop codon (TAA, TAG and TGA), thus, the identified frame called Open Reading Frame (ORF). The trimming process was based on the Psi-BLAST results from the GenBank.

Bioinformatics can speed up observation and performs analyses that impossible to perform using experimental approaches. The speed and effectiveness of computational biology has shown to provide the fastest and most comprehensive way of performing research (Irmtraud and Richard, 2002). The current study does agree that computation in biology helps in experimental research. Bioinformatics for biologist is a set of tools that help them reach the ultimate truth. The first microbial genome sequence, Haemophilus influenzae, was published in 1995. Since then, >400 microbial genome sequences have been completed or commenced (David *et al*., 2005). The massive influx of data provides the opportunity to obtain biological insights into the genomic of genomic bacteria through comparative genomics approach. The availability of complete genome

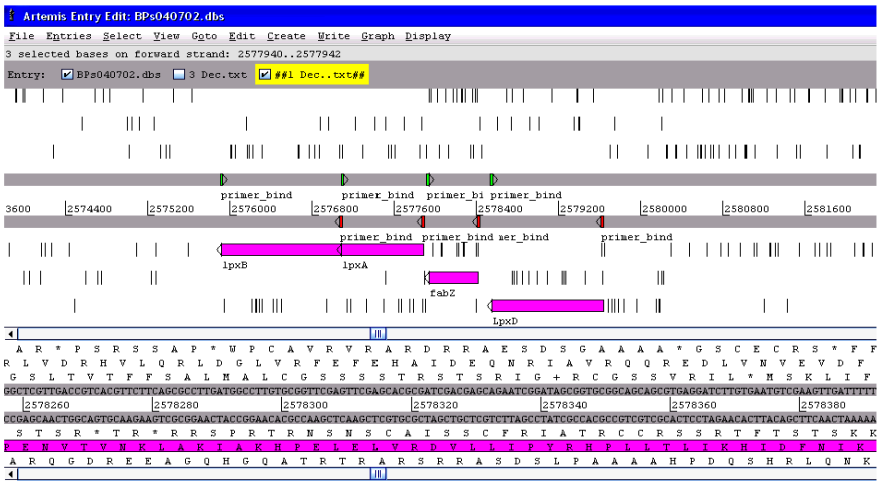

Fig. 11: The location of Bplps0012/*lpxB*, Bplps0013/*lpxA*, Bplps0014/*lpxC/fabZ* and Bplps0015/*lpxD* gene. The location of Bplps0012/*lpxB* (2575924..2577087), Bplps0013/*lpxA* (2577094..2577879), Bplps0014/*lpxC/fabZ* (2577938..2578402) and Bplps0015/*lpxD* (2578551..2579633) genes in *B. pseudomallei* genome showed in reverse frame

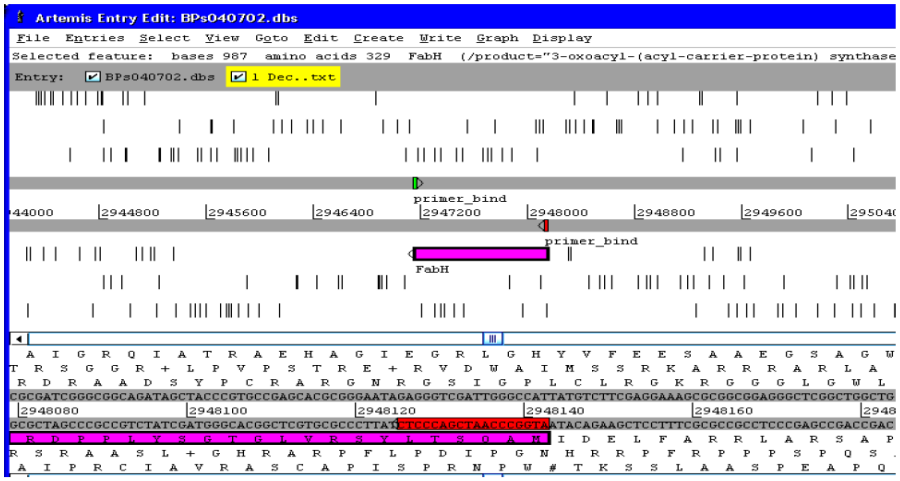

Fig. 12: The location of *Bplps0016/fabH* gene. The location of *Bplps0016/fabH* (2947156..2948142) gene in *B. pseudomallei* genome showed in reverse frame

sequence information for many pathogens and the development of sophisticated computer programs have led to a new paradigm in vaccine and drug development. Bioinformatics analysis is the first important strategy of reverse of vaccine and drug discovery (Hong-Liang *et al*., 2006). Since, the *B. pseudomallei* genome had been sequenced, the first challenge in analyzing these sequence data is the identification of hypothetical Open Reading Frames (ORFs) of unknown function. Another challenge is to define the correct genetic map position, arrangement and the function of the ORFs. It is well known that in many organisms the genes responsible for related functions are located close to each other on the chromosome. If one could accurately predict operons, the availability of a growing number of prokaryotic genomes would offer numerous significant clues relating to the function of hypothetical proteins (Ross *et al*., 1998). Furthermore, protein-coding DNA has certain periodicities and other statistical properties that are easy to detect in sequence of its length. These characteristics make prokaryotic gene finding relatively straightforward and well-designed systems are able to achieve high levels of accuracy. In referring to the current gene annotation and identification approaches (Perrier and Thioulouse, 2000; Irmtraud and Richard, 2002; Anders, 2003; Ren *et al*., 2004) a simple and optimal strategy to annotate LPS

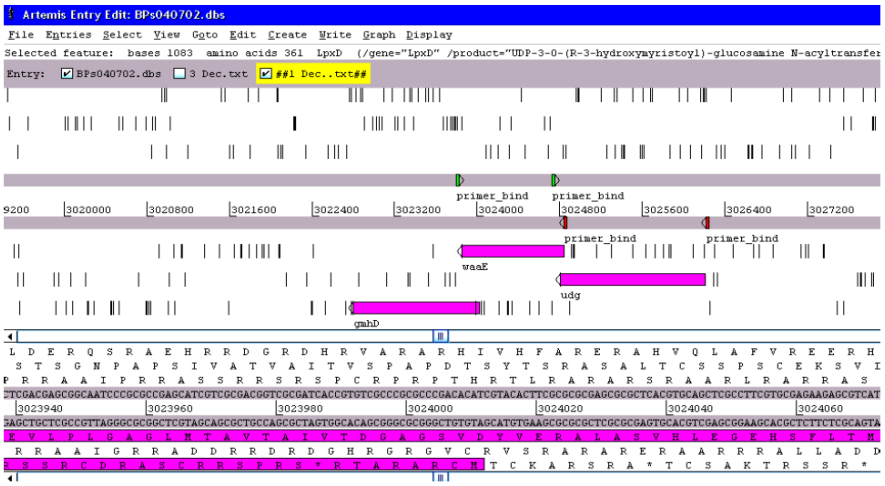

Fig. 13: The location of Bplps0017/*gmhD*, Bplps0018/*waaE* and Bplps0019/*udg* genes. The location of Bplps0017/*gmhD* (3022803..3024011), Bplps0018/*waaE* (3023854..3024837) and Bplps0019/*udg* (3024809..3026206) genes in *B. pseudomallei* genome showed in reverse frame

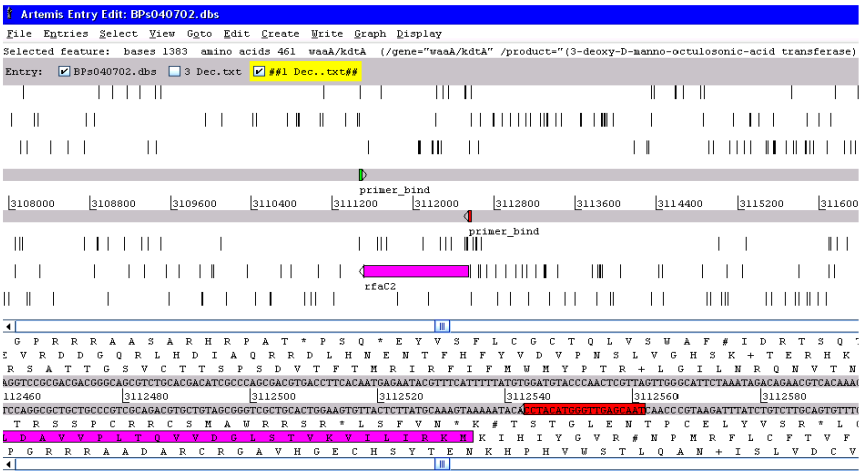

Fig. 14: The location of Bplps0020/*rfaC* gene. The location of Bplps0020/*rfaC* (3111515..3112534) gene in *B. pseudomallei* genome showed in reverse frame

biosynthesis genes in *B. pseudomallei* genomic sequences and gene function prediction was adopted. In the present study, comparison of sequence similarities was used since the comparison of nucleotide or protein sequences from the same or different organisms has been extensively compiled for systematic interrogation using bioinformatics software for molecular biology approach. Basic Local Alignment Search Tool (BLAST) is the tool located in NCBI suitable for sequence similarity determination and can contribute to the ORFs annotation. In the current study, the sequences of *B. pseudomallei* were hit 27-100% in BLASTing. Seventy to hundred precent of sequence similarity is considered as a good alignment and portrays a possible biological relationship. However, this study still accepts the hit of 27% similarity ORFs sequence of *B. pseudomallei* where, the result still in the top scoring 500 of BLAST report. Moreover, the position of the entire ORFs in gene cluster convinced the existence of a function link between the products of the co-localized genes (Overbeek *et al*., 1999). It is also, necessary to resort to ab initio gene finding, in which genomic DNA sequence alone is systematically searched for certain telltale signs of protein-coding genes. These signs can be broadly categorized as either signals, specific sequences that indicate the presence of a gene nearby. *Ab initio* gene finding might be more accurately characterized as gene prediction, since extrinsic evidence is generally required to conclusively establish that a

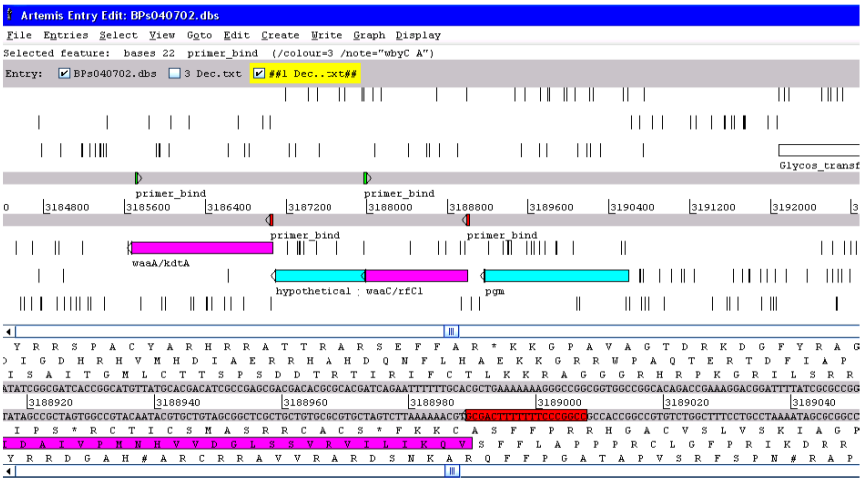

Fig. 15: The location of Bplps0021/*kdtA/waaA* and Bplps0022/*waaC* genes. The location of Bplps0021/*kdtA/waaA* (3185674..3187056) and Bplps0022/*waaC* (3187997..3188989) genes in *B. pseudomallei* genome. Blue colour bars are hypothetical protein genes showed in reverse frame

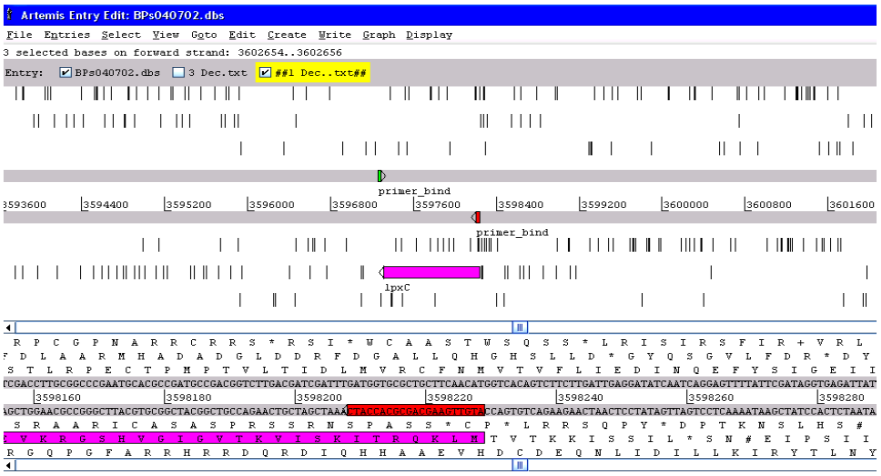

Fig. 16: The location of Bplps0044/*lpxC2* gene. The location of Bplps0044/*lpxC2* (3597314..3598228) gene in *B. pseudomallei* genome showed in reverse frame

putative gene is functional. The Artemis software was used as the bioinformatics tool to determine putative functions of genes and showed good results especially in *ab initio* gene finding. It is now the main annotation tool used for analysis of microbial genomes at the Sanger Institute. With this software, features and sequence from the study's result was read and the features on a six-frames translation of the sequence were successfully display. Additional information, such as feature colours, functional classifications were fully utilized and are added to the annotation. The purple color of annotated ORFs can be visualized in the results display as shown in Fig. 6-12. The predicted ORFs were examined individually for identity of known genes of *B. pseudomallei* with sequences and functions deposited in GenBank. Besides, the similarity with well-characterized LPS biosynthesis genes from other bacteria was successfully examined by using amino acid sequence alignment and PSI-BLAST, which were useful for searching databases, including the GenBank database and the Pfam protein motif database for possible gene functions. In each case, the literature was searched to ensure that the proteins encoded by the homologous genes were functionally characterized to avoid the perpetuation of poorly supported functional assignments. ORFs with strong homology to genes in other organisms with demonstrated functions can be fished out. All of the ORFs, which provided the latest new information about *B. pseudomallei* biology are those that could be assigned a probable function on the basis of similarity to

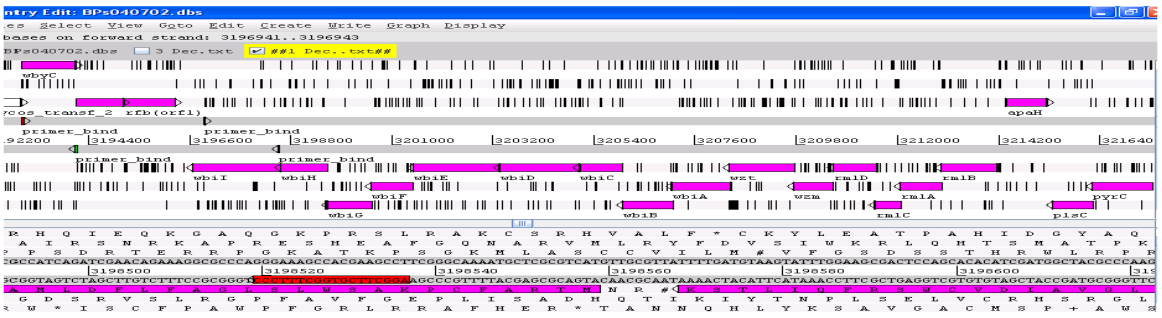

Fig. 17: The location of Bplps0023/*wbyC* (3192943..3194097), Bplps0024/*GalE* (3194115..3195134), Bplps0025/*rfb* (3195165..3196268),Bplps0026/*wbiI*(3196648..3198558),Bplps0027/*wbiH*(3198568..3199575),Bplps0028/*wbiG* (3199596..3200558),Bplps0029/*wbiF*(3200558..3201451),Bplps0030/*wbiE*(3201463..3203301),Bplps0031/*wbiD* (3203356..3205077),Bplps0032/*wbiC*(3205084..3206004),Bplps0033/*wbiB*(3206028..3207116),Bplps0034/*wbiA* (3207119..3208354), Bplps0035/*wzt* (3208354..3209760), Bplps0036/*wzm* (3209753..3210583), Bplps0037/*rmlD* (3210613..3211506),Bplps0038/*rmlC*(3211506..3212054),Bplps0039/*rmlA*(3212042..3212932),Bplps0040/*rmlB* (3212947..3214146),Bplps0041/*apaH*(3214383..3215228),Bplps0042/*plsC*(3216248..3217255),Bplps0043/*pyrC* genes in *B. pseudomallei* genome showed in forward and reverse frame

established sequence motifs, but could not be assigned a definite name. Therefore, the ORFs name was given base on the similar protein and the sequence arrangement on *B. pseudomallei* genome: for example, BPlps0001/*lpxL*. Due to the results of data mining in this study are based on *in silico*, all the annotated ORFs are considered putative or probable. Present study cannot predict exactly all gene components due to the limitation of our knowledge of complex biological processes and signals regulating gene expression.

CONCLUSION

Novel active compounds targeted at these genes will be particularly useful in overcoming the detrimental consequence of *B. pseudomallei* infection. The data presented here has identified new critical genes required for *B. pseudomallei*. The number of essential genes is sufficiently small to allow for experimental analysis, leading to a systematic strategy in designing novel antimicrobial active compounds for therapeutic intervention in melioidosis.

REFERENCES

Alistair, G.R., M. Emmanuel and B. Ewan, 2002. Genome annotation techniques: New approach and challenges. Drug Discovery Today, 7 (11): 570-576. DOI: 10.1016/S1359-6446(02)02289-4.

Anders, F., 2003. Strong associations between gene function and codon usage. APMIS, 111: 843-847. PMID: 14510641.

Christos, O., 2000. Two or three myths about bioinformatics. Bioinformatics, 16 (3): 187-189. PMID: 10869011.

Dance, D.A.B., 1991. Melioidosis: The tip of the iceberg? Clin. Microbiol. Rev., 4: 52-60.

David, A.R., S.A.M. Garry and R. Jacques, 2005. Visualization of comparative genomic analyses by BLAST score ratio. Bioinformatics, 6: 2. PMID: 2004347.

Hong-Liang, Y., Z. Yong-Zhang, Q. Jin-Hong, H. Ping, J. Xu-Cheng, Z. Guo-Ping and G. Xiao-Kui, 2006. *In silico* and microarray-based genomic approaches to identifying potential vaccine candidates against. Leptospira Interrogans BMC Genomics, 7: 293. PMID: 17109759.

Irmtraud, M.M. and D. Richard, 2002. Comparative *ab initio* prediction of gene structures using pair HMMs. Bioinformatics, 18: 1309-1318. PMID: 12376375.

Overbeek, R., M. Fonstein, M. D'Souza, G.D. Pusch and N. Maltsev, 1999. Use of contiguity on the chromosome to predict functional coupling. *In silico* Biol., 1: 93-108. PMID: 11471247.

Perrier, G. and J. Thioulouse, 2000. Use and misuse of coorespondence analysis in codon usage studies. Nucleic Acids Res., 30: 4548-4555. PMID: 12384602.

Ren, Z., O. Hong-Yu and Z. Chun-Ting, 2004. DEG: A database of essential genes. Nucleic Acids Res., 32: D271-D272. PMID: 14681410.

Ross, O., M. Fonstein, M. D'Souza, G.D. Pusch and N. Maltsev, 1998. Use of contiguity on the chromosome to predict functional coupling. *In silico* Biol., 1: 9. PMID: 11471247.

Yee, K.C., M.K. Lee, C.T. Chua and S.D. Puthucheary, 1988. Melioidosis, the great mimicker: A report of 10 cases from Malaysia. J. Trop. Med. Hyg., 91: 249-54. PMID: 3184245.