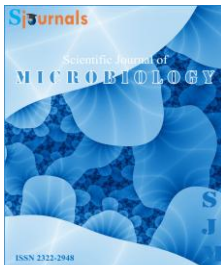
Scientific Journal of Microbiology (2016) 5(3) 108-114
ISSN 2322-2948
doi: 10.14196/sjm.v5i3.2157



Contents lists available at Sjournals

Scientific Journal of **Microbiology**

Journal homepage: www.Sjournals.com

Original article

Dynamic of IS*6110* transposase, transframe protein responsible for IS*6110* element transposition in *Mycobacterium tuberculosis* strains

Sara Thabet*, Nada Suissi

Bacteriology Laboratory, Veterinary Research Institute of Tunisia, Institute for Scientific Agricultural Research, 20 rue Jabal Lakhdhar, 1006, La Rabta,Tunis, Tunisia.

*Corresponding author; Bacteriology Laboratory, Veterinary Research Institute of Tunisia, Institute for Scientific Agricultural Research, 20 rue Jabal Lakhdhar, 1006, La Rabta,Tunis, Tunisia.

ARTICLE INFO

Article history,
Received 19 February 2016
Accepted 18 March 2016
Available online 25 March 2016
iThenticate screening 22 February 2016
English editing 15 March 2016
Quality control 21 March 2016

Keywords,
IS6110 insertion
orfAB
Evolution
Selection pressure

ABSTRACT

Transposition of IS*6110,* the *M*ycobacterium *tuberculosis* Complex specific insertion sequence, is mediated by the complex open reading frame AB (*orfAB)* transframe protein or transposase. Transposase has the crucial role to ensure IS*6110* transposition process. A single transposase proteine can control the integration of many IS*6110* copies at the same time. In this work we investigated the dynamic and the evolutionary trends of IS*6110* transposase coding gene isolated from various hosts. We had subjected these sequences to several statistical tests combined with programming package analysis. Our findings showed that IS*6110* transposase-encoding *orfAB* is subjected to purifying selection with low genetic variability rate, as demonstrated by its significantly negative Tajima's *D* statistics as well as by LRT analysis and models comparison. Even some codons tended to be positively selected, global *p-value* didn't reach reliable inference ($\omega > 1$) of significant positive selection at these positions; so statistically they were not significant. In conclusion, our findings demonstrated that IS*6110* transposase-encoding *orfAB* evolved essentially by point mutations under purifying selection acting against deleterious mutations, thus leading to an excess of low-frequency variants and purging disadvantageous non synonymous changes.

1. Introduction

Insertion sequences (ISs) are the smallest autonomously transposable mobile genetic elements widely distributed in bacterial genomes (McFadden et al., 1987; Mes and Doeleman, 2006; Preston et al., 2004; Siguier et al., 2014). ISs elements carry in their sequence the gene encoding for a transposase, thus ensuring their mobility in the bacterial genome. Transposition process of each IS is mediated by its own transposase and contributes significantly to genome diversification and plasticity (Johnson et al., 1983; Zerbib et al., 1990; Brosch et al., 1999; Fang et al., 1999; Siguier et al., 2014). Indeed, IS sequences, via their transposable activity, have been shown to induce genomic rearrangements that translates into strain-specific phenotypic variations (Siddiqi et al., 2001; Viana-Niero et al., 2006; McEvoy et al., 2007).

IS*6110* insertion sequence identified in members of the *M*ycobacterium *tuberculosis* Complex (MTBC), has been the most extensively studied of the mycobacterial mobile elements (Cole et al., 1998). IS*6110* contains two consecutive open reading frames (ORFs), *orfA* (327 bp) and *orfB* (927 bp) partially overlapping in the relative translational reading frames 0 and -1 respectively and transposes via IS*6110* transposase, a fusion protein produced following the reading shift at the overlapping region between the two ORFs (Fayet et al., 1990). In the absence of frameshifting, *orfA* and *orfB* proteins are produced from ORFs A and B respectively. *OrfA* protein exhibits a helix-turn-helix motif providing sequence-specific DNA binding to terminal IRs. *OrfB* protein contains a conserved amino acid triad at the N-terminal sequence, the DD(35)E domain, which plays an important role in the transposition process (Sekine et al., 1994). Transposase or *OrfAB* protein has two domains, a DNA binding domain at the N-terminal region and a catalytic domain at the C-terminus. It specifically recognizes the two terminal inverted repeat sequences (IRs) and ensures the excision of IS*6110* and its integration elsewhere on the genome fixing the 3'-OH end of the transposon at the 5'-phosphate end of the target site. DD(35)E motif has been involved in the catalysis of such reaction (Sekine et al., 1994).

In a previous work, genetic variability and selective forces acting on the two IS*6110 orfs* A and B have been studied separately (Thabet et al., 2015). Here, we focused on *orfAB* gene encoding IS*6110* transposase after the occurrence of the frame shift at the overlapping region in order to explore the dynamic and evolution mechanisms of IS*6110* transposase given its crucial role as key effector of transposition process. The findings were analyzed by numerous statistical tests.

2. Materials and methods

2.1. Genomic sequence data of IS*6110* transposase genes from *Mycobacterium tuberculosis* strains

A total of 327 bp full-length DNA sequence of IS*6110 orfA* and 987 bp full-length DNA sequence of IS*6110 orfB* from *Mycobacterium tuberculosis* strains isolated from various hosts were downloaded from GenBank where they were deposited under accession numbers KP844666 to KP844685 and KP844686 to KP844721 respectively.

Our data set comprised contiguous sequences without insertions or deletions. As noted before, the two ORFs A and B are partially overlapping, so we had to concatenate them at AAAG motif in the domain frameshift to have only one *orfAB* that encodes the IS*6110* transposase (figure 1).

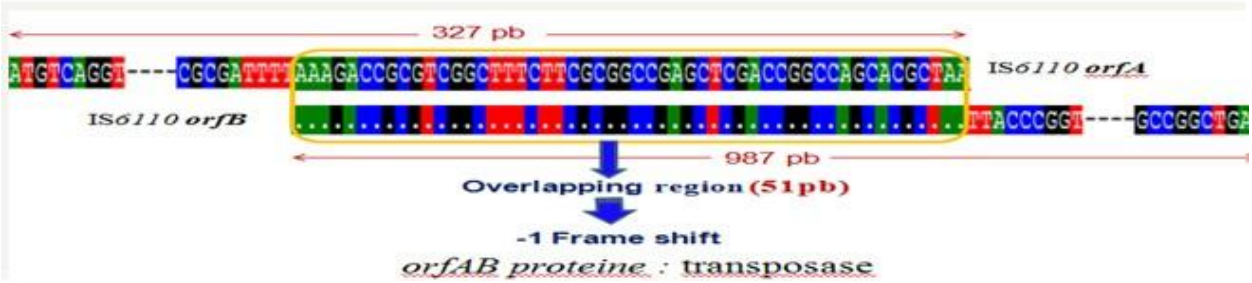


Fig. 1. *orfAB* encoding the IS*6110* transposase.

2.2. Selective neutrality testing

IS*6110* transposase encoding gene was subjected to selective neutrality testing performed with Tajima's *D* statistic. Tajima's *D* detects selective effects on the basis of molecular diversity (Tajima, 1989). It tests the relationship between two parameters, the number of polymorphic sites, S and the average number of nucleotide differences, k.

$$D = \frac{\hat{\theta}_\pi - \hat{\theta}_S}{\sqrt{\mathrm{Var}(\hat{\theta}_\pi - \hat{\theta}_S)}}$$

θπ: Estimator of the nucleotide diversity

θs: Number of polymorphic sites / sum (1 + 1/2 + 1/3 + 1/4 .. 1/(n-1))

(Denominator is a complex normalization constant C)

If $D > 0$ ⟶ Excess of mutations with imtermediate frequency

If $D < 0$ ⟶ Excess of mutations with low frequency

2.3. Selective pressure testing at codons level

The selection pressure on IS*6110* transposase-coding gene was measured by comparing nonsynonymous (dN) and synonymous (dS) substitution rates using models of the PAML package. Under neutrality (nonsynonymous changes have no associated advantage or disadvantage), the expected ratio of dN/dS (or *ω*) is 1 and significant diviation from this value can be used to identify genes that are either under purifying selection (dN < dS, nonsynonymous changes are deleterious) or under positive selection (dN > dS, nonsynonymous changes are favored because of a fitness advantage). We deleted all sequences with gaps and internal stop codons from our data set. Subsequently, we used neutral (M1 and M7) and selection (M2 and M8) models of codon evolution to establish whether positive selection was at hand and, if so, to identify the codons that are under positive selection (Mes and Doeleman, 2006). Models M1 and M7 assume a different distribution of *ω* values smaller than 1. These two models differ from the selection models M2 and M8 in the presence of a class of codons with *ω* constrained to be larger than 1 (ω_2), thereby distinguishing positive selection from purifying evolution ($\omega < 1$) and neutral evolution ($\omega = 1$).

2.3.1. Likelihood-Ratio test analysis

Likelihood-Ratio Test (LRT) analysis was performed to study the presence of positive selection at different sites of transposase amino acid sequence and to test the significance of dN/dS (*ω*) values obtained.

$$\mathrm{LRT = Log\ Ration\ Test} = \Delta = 2(\ln(LA) - \ln(LB))$$

L_A and L_B: relative likelihoods of models studied

- LRT analysis uses a parameter, the degree of freedom or df (δ). δ represents the difference between the number of models parameters; here df = 2.
- LRT value is compared to the threshold value of the Chi2 (χ^2) (threshold available in Chi2 table).
- If LRT > threshold Chi2, then MA model is significantly more likely than MB model at α risk.
- If LRT < threshold Chi2, then the above hypothesis is rejected.

2.3.2. Additional positive selection testing methods

We used an algorithmic additional tool based on selecton web server to estimate the degree of purifying selection and positive selection at each codon site. Selecton, available at: http://selecton.bioinfo.tau.ac.il, detects evolutionary forces at a single amino-acid site. In this methodology, nonsynonymous and synonymous substitutions are designed by Ka and Ks respectively and the ratio Ka/Ks is calculated for each codon site in a codon-based multiple sequence alignment (MSA). The significance of the Ka/Ks scores is also obtained by using the

LRT that compares two nested models: a null model which assumes no selection and an alternative model which does (Doron-Faigenboim et al., 2005; Stern et al., 2007).

3. Results and discussion

3.1. Genetic polymorphism in transposase gene

After concatenating the two ORFs *orfA* and *orfB* at AAAG motif for all samples, all sequences have been collected in one file in fasta format. Multiple alignments of all the sequences of *orfAB* encoding IS*6110* transposase with BioEdit5 and comparative analysis with DnaSP5 program led to recognition of 12 haplotypes. Firstly, the number of sites on this coding region was 1260 with total number of mutations, Eta: 6. Two of these mutations corresponded to synonymous substitutions at positions 63 and 72 respectively. And four were non synonymous substitutions at positions 280, 323, 919 and 1259 (Figure 2).

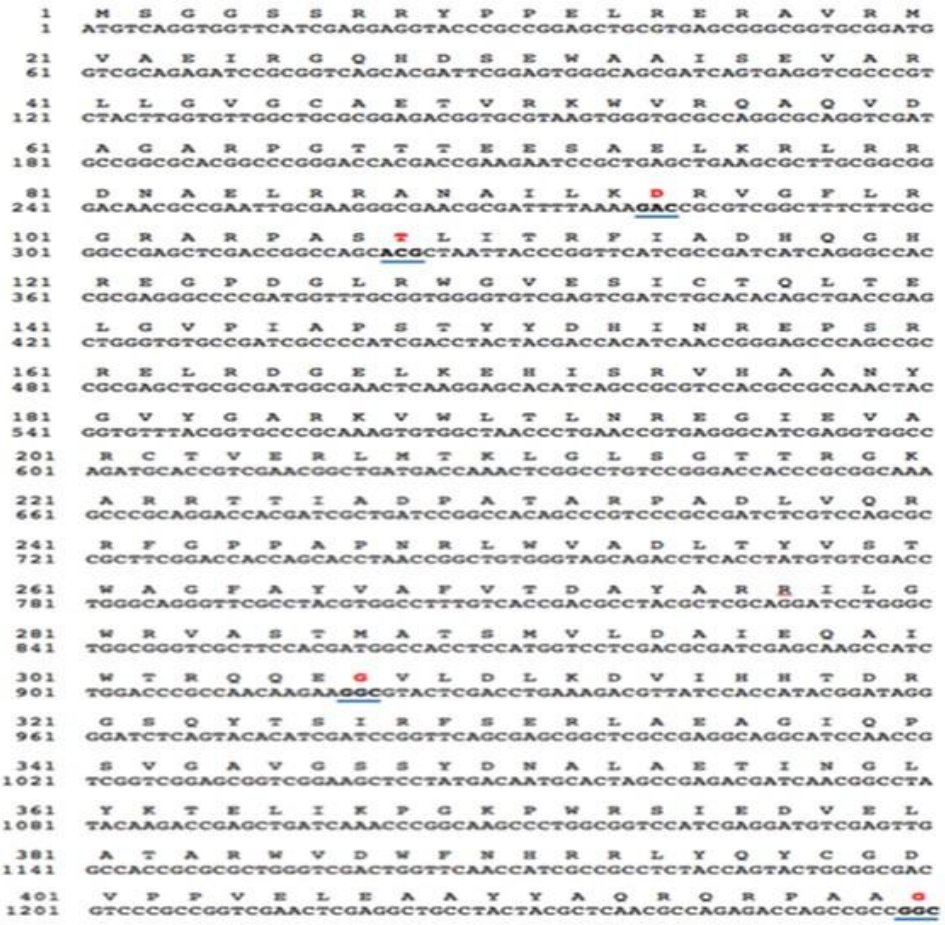


Fig. 2. Presence and distribution of four non synonymous substitutions.

The average number of nucleotide differences, k: 1.282 with nucleotide diversity, Pi of 0.00102.

3.2. Selective pressure acting on IS*6110* transposase coding *orfAB*

3.2.1. Absence of neutral selection

Statistic Tajima test proved significantly negative for IS*6110* transposase gene since *D* value determined with DnaSP program was -1.244 (<0) reflecting an excess of mutations with low frequency rejecting neutral selection theory.

3.2.2. Evidence for purifying selection

As noted before, we conducted LRT analysis to test selective pressure site by site and to compare the significance of different models using PAML package. PAML analysis of different variants obtained provided low evidence for positive selection which was not supported by Bayes Empirical Bayes (BEB) analysis. The comparison between neutral and selection models (M1 vs M2 and M7 vs M8) and Bayesian site identification showed no statistical significant difference with LRT *p-value* of ~0.9 < threshold $Chi2_{\alpha=0.05}$=5.99 (Table 1).

Table 1
Tests of positive selection and positively selected codons in the IS*6110* transposase gene of *M. tuberculosis* according to neutral models (M1 and M7) and selection models (M2 and M8) of PAML.

Model	Tree length[a]	lnL	Parameters[b, c]	LRT *p-value*
M1	0.01459	-1732.859692	ω_0 = 0.80265, ω_1 = 1.00000 p_0 = 0.99999, p_1 = 0.00001	
M2	0.01459	-1732.859694	ω_0 = 0.80260, ω_1 = 1.00000, ω_2 = 1.00000 p_0 = 0.99903, p_1 = 0.00008, p_2 = 0.00088	0.9
M7	0.01458	-1732.859736	*B* (*p* = 99.00000, *q* = 24.25459)	
M8	0.01458	-1732.859768	*B* (*p* = 99.00000, *q* = 23.49429), p_0 = 0.99999 P_1 = 0.00001, ω_2 = 2.35158	0.9

[a] Tree length is measured as the number of mutations per codon.
[b] Kappa is the transversion/transition ratio. Pi denotes the proportion of sites falling in site class ωi.
[c] Parameters p and q are the shape parameters of the beta distribution which underlies M7 and M8.
[d] The reference sequence for the amino acid designation.
[e] The probability that codons were under positive selection was determined using Bayes empirical Bayes (BEB), with the ω and its standard indicated per codon.

The estimation of purifying and positive selection in the empirical Baysian method used by selecton program and based on the calculation of the *ω* values at each codon position showed a global Ka / Ks score of 0.89 non-significant for positive selection in the protein encoded. Over the seven-color scale for representing the different type of selection, four positively selected sites were found in the amino-acid sequence 94 D, 108 T, 307 G and 420 G (Figure 3).

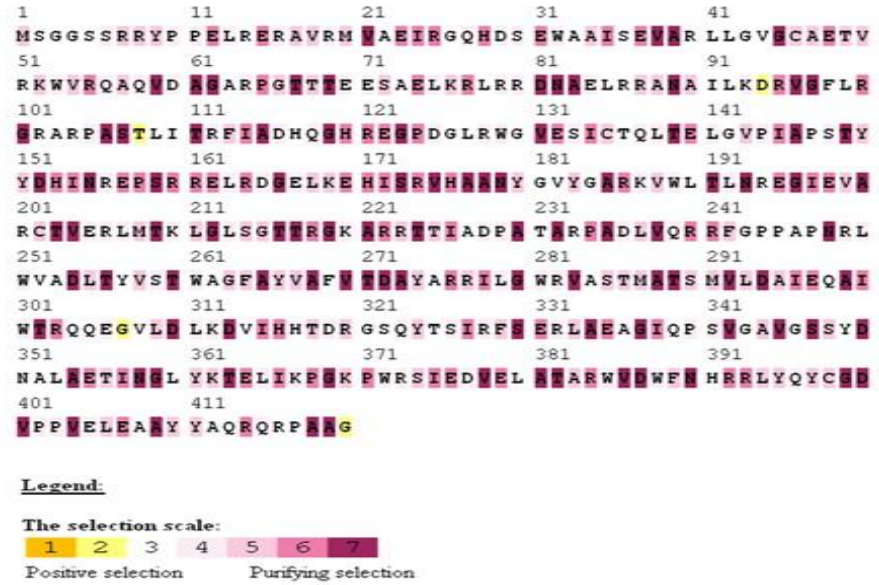


Fig. 3. Positively selected sites detected by Selecton program.

Three codons selected 94 D, 108 T and 420 G are located in non-conserved domain and then they aren't functionally important. The fourth codon selected 307 G is suggested to have important functional role as it is located in conserved domain of IS*6110* transposase protein. In fact, conserved domain database contains "building blocks" that are believed to modulate protein function and only the presence of positively selected codons in these domains suggests that these codons have important functional role. Selection pressure exerted on these four sites didn't reach reliable inference ($\omega > 1$) of significant positive selection at these positions; so statistically they were not significant. With DnaSP program, the global ω value was 0.802 and called then for purifying selection, confirming PAML and selecton results.

As an excess of nonsynonymous over synonymous substitutions at individual amino acid sites is an important indicator that positive selection has affected the evolution of a protein between extant sequences and their ancestor, we have used several methods to detect the presence and location of positively selected sites in our alignments of IS*6110* transposase coding gene. Comparing the fits of the data provided by neutral and selection models, the selection analysis with PAML method found no evidence for positive selection acting on single amino acid residues. Selecton and DnaSP results advocated neither any statistically significant positive selection in the transposase amino acid sequence. In fact, only few sites under little positive selection pressure with low effect acting on the protein structural and functional roles. Examination of transposase gene families might help to explore the functional divergence of prokaryotic genes associated with gene duplications. Because duplicated gene classes other than transposases and integrases are generally rare in bacterial genomes, their use as targets for studies of functional gene differentiation may complement similar studies in eukaryotes. Transposase genes and duplicate genes should show comparable levels of divergence. However, transposase genes are much more homogeneous than duplication genes because of the tight regulation of transposase activity in wild-type cells (Mahillon and Chandler, 1998), where transposase genes are among the lowliest expressed genes. Until recently, the role of transposase genes in adaptative evolution was thought to be minimal because surveys of particular element insertions suggested that they occur only at low frequency within species and because of their distinctiveness in terms of diversity and dynamics. However, recent studies have shown that transposon insertions play a large role in transcriptional regulation and in the evolution of regulatory and coding sequences of genomes. These findings suggest that these gene families may be suitable targets for comparative genome and bioinformatics analysis.

In conclusion, our findings demonstrated that IS*6110* transposase-encoding *orfAB* evolved essentially by point mutations under purifying selection acting against deleterious mutations, thus leading to an excess of low-frequency variants and purging disadvantageous non synonymous changes. In another component, the comparison of multiple parts of different genomes demonstrates that the *M. tuberculosis* genome is currently undergoing an active process of gene decay, analogous to the adaptation process of obligate bacterial symbionts (Cubillos-Ruiz, 2008). Such observation opens new perspectives into the evolution and the understanding of the pathogenesis of this pathogen.

Acknowledgments

This research received financial support from the Tunisian Ministry of Agriculture and Hydraulic Resources.

References

Brosch, R., Wolfgang, J.P., Stavropoulos, E., Colston, M.J., Cole, S.T., Gordon, S.V., 1999. Genomic analysis reveals variation between Mycobacterium tuberculosis H37Rv and the Attenuated M. tuberculosis H37Ra Strain. Infect. Immun., 67, 5768-5774.

Cole, S.T., Brosch, R., Parkhill, J., al, 1998. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature, 393, 537-544.

Cubillos-Ruiz, A., Morales, J., Zambrano, M.M., 2008. Analysis of the genetic variation in Mycobacerium tuberculosis strains by multiple genome alignments. BMC research Notes, 1, 110.

Doron-Faigenboim, A., Stern, A., Mayrose, I., Bacharach, E., Pupko, T., 2005. Selecton: a server for detecting evolutionary forces at a single amino-acid site. Bioinform., 21, 2101-2103.

Fang, Z., Doig, C., Kenna, D.T., Smittipat, N., Palittapongarnpim, P., Watt, B., Forbes, K.J., 1999. IS6110-mediated deletions of wild type chromosomes of Mycobacterium tuberculosis. J. Bacteriol., 181, 1014-1020.

Fayet, O., Ramond, P., Polard, P., Prère, M.F., Chandler, M., 1990. Functional similarities between retroviruses and the IS3 family of bacterial insertion sequences. Mol. Microbiol., 4, 1771-1777.

Johnson, R.C., Reznikoff, W.S., 1983. DNA sequences at the ends of transposon Tn5 required for transposition. Nature, 304, 280 - 282.

Mahillon, J., Chandler, M., 1998. Insertion sequences. Microbiol. Mol. Biol. Rev., 62, 725-774.

McEvoy, C.R.E., Falmer, A.A., Gey van Pittius, N.C., Victor, T.C., van Helden, P.D., Warren, R.M., 2007. The role of IS6110 in the evolution of Mycobacterium tuberculosis. Tuberculosis, 87, 393-404.

McFadden, J.J., Butcher, P.D., Thompson, J., Chiodini, R.J., Hermon-Taylor, J., 1987. The use of DNA probes identifying restriction-fragment-length-polymorphisms to examine the Mycobacterium avium complex. Mol. Microbiol., 1, 283 - 291.

Mes, T.H.M., Doeleman, M., 2006. Positive selection on transposase genes of insertion sequences in the Crocosphaera watsonii Genome. J. Bacteriol., 188, 7176 - 7185.

Preston, A., Parkhill, J., Maskell, D.J., 2004. The Bordetellae: lessons from genomics. Nat. Rev. Micro., 2, 379 - 390.

Sekine, Y., Eisaki, N., Ohtsubo, E., 1994. Translational control in production of transposase and in transposition of insertion sequence IS3. J. Mol. Biol., 235, 1406 - 1420.

Siddiqi, N., Shamim, M., Amin, A., Chauhan, D.S., Das, R., Srivastava, K., Singh, D., Sharma, V.D., Katoch, V.M., Sharma, S.K., Hanief, M., Hasnain, S.E., 2001. Typing of drug resistant isolates of Mycobacterium tuberculosis from India using the IS6110 element reveals substantive polymorphism. Infect. Genet. Evol., 1, 109-116.

Siguier, P., Gourbeyre, E., Chandler, M., 2014. Bacterial insertion sequences: their genomic impact and diversity. FEMS Microbiol. Rev., 38, 865 - 891.

Stern, A., Doron-Faigenboim, A., Erez, E., Martz, E., Bacharach, E., Pupko, T., 2007. Selecton 2007: advanced models for detecting positive and purifying selection using a Bayesian inference approach. Nucl. Acid. Res., 35, 506-511.

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genet., 123, 585-595.

Thabet, S., Namouchi, A., Mardassi, H., 2015. Evolutionary Trends of the Transposase-Encoding Open Reading Frames A and B (orfA and orfB) of the Mycobacterial IS6110 Insertion Sequence. PLOS ONE. 18.

Viana-Niero, C., Rodriguez, C.A.R., Bigi, F., Zanini, M.S., Ferreira-Neto, J.S., Cataldi, A., Cardoso Leao, S., 2006. Identification of an IS6110 insertion site in plcD, the unique phospholipase C gene of Mycobacterium bovis. J. Med. Microbiol., 55, 451-457.

Zerbib, D., Polard, P., Escoubas, J.M., Galas, D., Chandler, M., 1990. The regulatory role of the IS1-encoded InsA protein in transposition. Mol. Microbiol., 4, 471 – 477.

How to cite this article: Thabet, S., Suissi, N., 2016. Dynamic of IS*6110* transposase, transframe protein responsible for IS*6110* element transposition in *Mycobacterium tuberculosis* strains. Scientific Journal of Microbiology, 5(3), 108-114.