Define Spot Contamination Level $SCL[k](t)$ as the level of virus contamination of a spot k at time t. Contamination level of the spot in the certain time t is the sum of virus excretion of agents in the spot and the contamination level of the spot at time $(t-1)$.

$$SCL[k](t) = \sum_{i \in Spot[k]} AHL[i](t) + SCL[k](t-1) \times SSL[k](t) \quad (2)$$

Where Spot Sterilization Level $(0 \leq SSL[k](t) \leq 1)$ represents the effects of attenuation and sterilization on the certain spot. The smaller the SSL, the more effective the protection measure is (See TABLE III).

$$ACL[i](t) = ACL[i](t-1) \times AF\,[i](t) + SCL[k](t) \times VD[k] \quad (3)$$

Define Agent Contamination Level $ACL[i](t)$ as the amount of virus that an agent i has absorbed from the spot k where he stands at the specific time t., where Attenuation Filter $(0 \leq AF\,[i](t) \leq 1)$ represents the effect of attenuation protection on infection (e.g., hand washing) (See TABLE III) and Virtual Density $(0 \leq VD[k] \leq 1)$ represents the density of the spot k (the bigger place, the smaller VD).

When an agent i at time t absorbs a significant amount of influenza virus, he will be infected and his state will change from susceptible to infected. The probability of agent i at the time t to get infected is calculated as below.

$$p[i](t) = 1 - \exp[-PC[i](t) \times ACL[i](t)] \quad (4)$$

Where $PC\,[i](t)$ is the Physical Condition of agent i at time t $(0 \leq PC\,[i](t) \leq 1)$. Physical condition depends on vaccination status, heath condition, age, and sex. The healthier agent (smaller PC), the smaller infection probability is. If the agent is immune to the virus, PC is equal to 0, which means probability of infection is equal to 0.

Pathological transition of the disease is described in Fig. 3. Infection levels were categorized and defined into several states below [20].

- State of "0" denotes pre-infection state (susceptible to infection).
- State of "1" represents state of infection. Agents stay in this state for 3 days. This is the incubation period of the disease.
- State of "2" is designated as infection state with apparent symptoms. The probability of changing state from "1" to "2" is 0.8.
- State of "2m" represents the mild case of infection. The sequence 2m→3m→5→0i represents state transition of recovering without apparent symptoms. The sequence 2→3→5→0i describes state transition of recovering with apparent symptoms.
- State of "3s" denotes serious case of infection. The symptoms are severe influenza-like illness with cough or sore throat, plus measured fever, shortness of breath and need for hospitalization.
- State of "4c" represents critically ill infection state. For patients in this state, clinical treatment in the hospital helps them to recover. The sequence 4c→4m→5→0i describes the recovering route. Inpatients at the state of "5" are recommended to discharge from hospital.

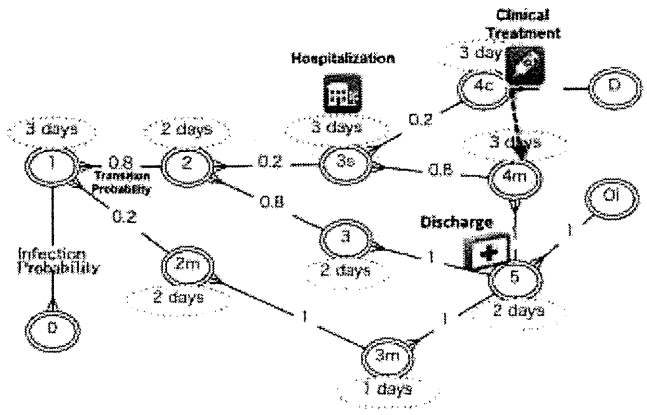

Fig. 3 State transition of influenza-like illness with high contagion, high mortality and clinical pathway for infected patient

Table I below describes the disease states and values of Agent Virus Excretion Level ($AVEL$). These values are assigned in corresponding to definition of the disease states. It is believed that influenza virus can be transmitted from infected people to others from 1 day before symptoms develop and up to 5 to 7 days after becoming sick (CDC).

DOI: 10.5963/PHF0303003

TABLE I DISEASE STATE DEFINITION AND VALUE OF AGENT VIRUS EXCRETION LEVEL (*AVEL*)

State	Definition	Fever	*AVEL*
0	Not infected	No	0.0
1	1^{st} state	Little	0.2
2	2^{nd} state	High	0.6
2m	2^{nd} mild state	Little	0.4
3	3^{rd} state	Little	0.6
3m	3^{rd} mild state	Little	0.5
3s	3^{rd} serious state	High	0.6
4c	4^{th} critical state	High	0.5
4m	4^{th} mild state	Little	0.5
5	Recovered state	No	0.0
0i	Recovered with immunity	No	0.0
D	Death	No	0.0

III. SIMULATION PARAMETERS

Simulation parameters are summarized in TABLE II. However, these parameters can be changed to adapt to any community and hospital. Simulation is executed in 30 days and repeats 30 times. The simulation program generated 30 files of each log files. The log files contain information on each agent and each spot at each hour in 30 days. The information on each agent included name, job, disease status, immunity status, influenza virus contamination level, probability of infection, etc. The information on each spot included the level of contamination. Simulation log files also included the numbers of outpatients, inpatients, number of infected HCW, number of visitors, list of infected HCW, list of contacts of each agent, etc.

Preventing transmission of influenza virus within healthcare settings is important for hospital management. Spread of influenza virus can occur among patients, HCW, and visitors; in addition, HCW may acquire influenza from persons in their household or community. The fundamental elements of nosocomial influenza infection control include influenza vaccine campaign, respiratory hygiene, monitoring HCW's health, droplet precautions, hand hygiene, environment sterilization and managing visitor access and movement within the facility [18].

Values of parameters for infection control measures are shown in TABLE III. Vaccinating children, adolescents, and young adults seems to be an appropriate vaccination strategy to reduce morbidity of the disease [21]. Based on studies of efficacy comparison of several hand hygiene products [22] and masks [23], we set values for hand hygiene and droplet precaution control measures. Biological efficacy and rate of recontamination (parameter *SSL*) is adopted from [24].

To study the impact of infection control on nosocomial infection, we vary parameters of infection control in 4 scenarios. Parameters for the four scenarios are summarized in TABLE IV. High Control and High Vaccine scenario represents for the circumstance of hospital with high resource of infection control and vaccination rate in the community is high. Scenario of Low Control and Low Vaccine represents the circumstance of hospital with low level of infection control and vaccination rate in the community is low.

Since simulation model is an abstraction of the real world, each parameter setting corresponds to the set of assumptions made by the model. The strength of simulation is that it can simulate the real world as in a variety of circumstances. Experiments can be set up and repeated many times, using a range of parameters. Those parameter changes can be made by using experimental setting function of SOARS [16].

TABLE II DESCRIPTION AND VALUE OF SIMULATION PARAMETER

Simulation parameters	
Simulation time	30 days
Simulation replication	30 times
Time step	10 min
Log time	1 hour
City population structure	
Total population	10,000 people
Age distribution	Proportion

DOI: 10.5963/PHF0303003

Child: 0- 4 y/o	8.5%
Teenager: 5- 14 y/o	16.5%
Adolescent: 15-19 y/o	10.2%
Adult: 20- 34 y/o	26.0%
Middle-aged: 35-59 y/o	29.9%
Elderly: Over 60 y/o	8.9%
Hospital structure	
Number of doctors	7
Number of nurses	18
Number of beds	28
Number of outpatient	Average of 60/day
Number of visitor	Average of 20/day

TABLE III DESCRIPTION AND VALUE OF INFECTION CONTROL PARAMETER

Vaccination Target	(Probability of vaccination)
Child	0.3
Teenager	0.5
Adolescent	0.2
Adult	0.2
Mid-aged	0.15
Elderly	0.1
Vaccinated Population	(Probability of vaccination)
High	20%
Medium	10%
Low	5%
Mask wearing	(Value of VEP)
No mask	1.0
Surgical mask	0.5
N95 mask	0.1
Hand Hygiene	(Value of AF)
Soap and water	0.62
Alcohol-based hand rubs	0.73
No treatment	1
Environmental Infection Control	(Value of SSL)
No cleaning	0.6
After cleaning	0.4
After HPV decontamination	0.03
Monitor and Manage Ill Healthcare Personnel	Not to go to work, or if at work, to stop patient-care activities, leaving work.
Patient Isolation Policy	Isolate critical influenza patients from patients of other diseases and from visitors.
Patient Isolation Policy	Isolate critical influenza patients from patients of other diseases and from visitors.
Manage Visitor Access	Limit visitors' access. Check visitors' temperature before entering the hospital.

TABLE IV INFECTION CONTROL PARAMETER OF FOUR SCENARIOS

Scenario Name	A	B	C	D
Infection Control	High Control	High Control	Medium Control	Low Control
Vaccinated Population	High Vaccine	Medium Vaccine	Low Vaccine	Low Vaccine
Hand Washing	Soap and water	Soap and water	Alcohol-based hand rubs	No
Mask	N95	N95	Surgical	No
Patient Isolation	Yes	Yes	No	No

DOI: 10.5963/PHF0303003

Cleaning	Yes	Yes	Yes	No
HPV Decontamination	Yes	Yes	No	No
Vaccinated Population	2000 (20%)	1000 (10%)	500 (5%)	500 (5%)

IV. SIMULATION RESULTS

We demonstrate simulation results by macro and micro analysis. In micro analysis, the number of infected patients and health care workers are observed in one month. Macro analysis is performed by calculating of the amount of virus in spots and in agents, respectively.

A. Macro Analysis

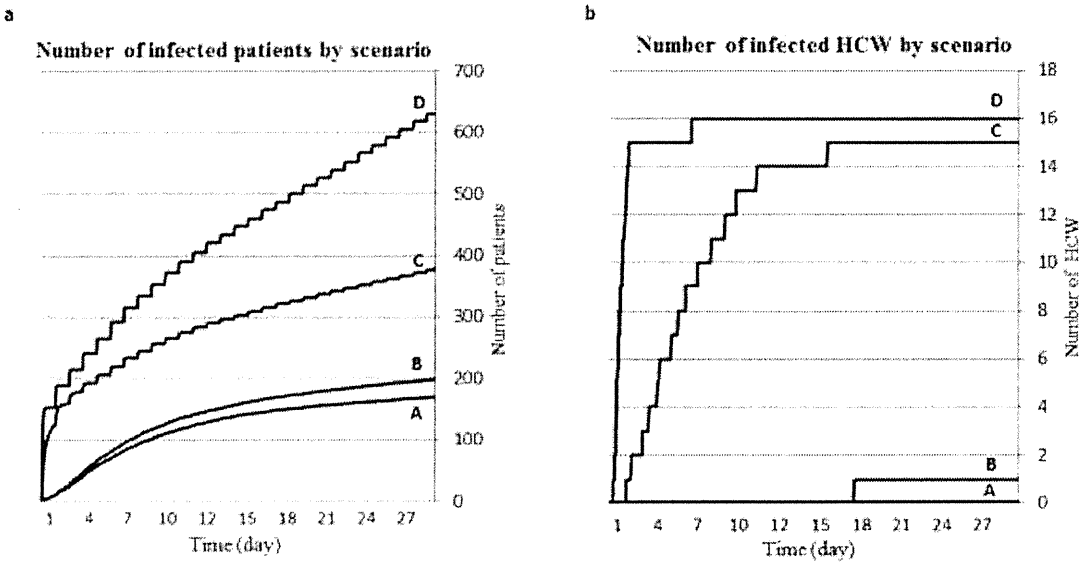

Fig. 4 Variation in average number of infected patients and health care workers (HCW) over time in the four scenarios

The aggregate number of infected patients and HCW are displayed in Figure 4a and Figure 4b, for each scenario A, B, C, D. In Figure 4a, the number of infected patients increases rapidly from 168 in scenario A and 198 in scenario B to 377 in scenario C and to 630 patients in scenario D. The relative standard deviations differ from one scenario to another, but converge around 10% in 30 days. The number of infected patients shows an increasing trend after 30 days, however with a considerably lower speed as compared to the high increasing rate at the early stage of the simulation. The infected rate among outpatients for each scenario A, B, C, D is 9%, 11%, 24% and 39%, respectively (the average number of outpatients in scenario A, B is 1800 and in scenario C, D is 1600, respectively). Note that the simulation model counts the number of people who are infected within the whole hospital, so these ratios indicate the infection risk level for every patient who is present at the hospital.

In Figure 4b, the average number of infected HCW increases dramatically from 0 and 1 in scenario A and B to 15 and 16 in scenario C and D. The relative standard deviations of number of infected HCW in scenario A and B were not calculated (since the average number is between 0 and 1). The relative standard deviation of number of infected HCW in scenario C and D converges at 14% and 17%, respectively. The number of infected HCW sees an exponential increase in scenario C and D within 2 weeks but levels off afterwards. The infection rate among HCW is 0%, 3%, 50% and 53% in scenario A, B, C and D, respectively. The simulation results imply that infection control plays a significant role in protecting HCW from nosocomial influenza infection.

To shed more light on which infection control has the most impact on preventing nosocomial influenza in HCW, we have simulated three more scenarios. In these scenarios, same low vaccination rate (3%) was set. In scenario E, high infection control measures were implemented. In scenario "E − washing hand" and "E − wearing mask", staff washing hand and wearing mask were excluded, respectively. The number of infected HCW in each scenario is shown in Figure 5. The result shows that staff washing hand combining with wearing mask could significantly reduce the number of infected HCW.

Although washing hand and wearing mask control measures were recommended worldwide, the extent to which these measures can help prevent influenza transmission has not been firmly established. Recent studies have evaluated the efficiency of those control measures [25, 26]. The authors agree with the suggestion that use of masks should always be paired with regular hand washing. In the circumstance of limited vaccine availability, using surgical mask and washing hand with soap, which are relatively inexpensive and practical, could be a good strategy in nosocomial influenza infection.

DOI: 10.5963/PHF0303003

TABLE V INFECTION CONTROL PARAMETER OF ADDITIONAL SCENARIOS

Scenario Name	E	E – washing hand	E – wearing mask
Hand Washing	Soap and water	No	Soap and water
Mask	Surgical	Surgical	No
Patient Isolation	Yes	Yes	Yes
Cleaning	Yes	Yes	Yes
HPV Decontamination	No	No	No
Vaccinated Population	300 (3%)	300 (10%)	300 (3%)

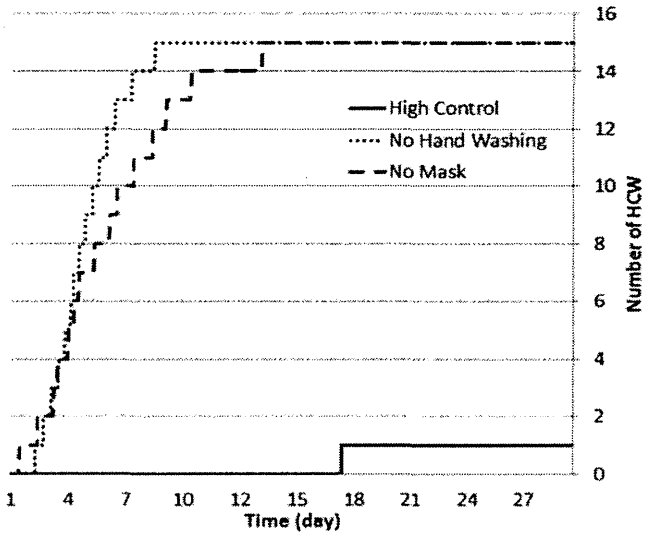

Fig. 5 Variation in average number of infected health care workers (HCW) over time in scenario E, scenario E with no staff washing hand and scenario E with no staff wearing mask.

B. Micro Analysis

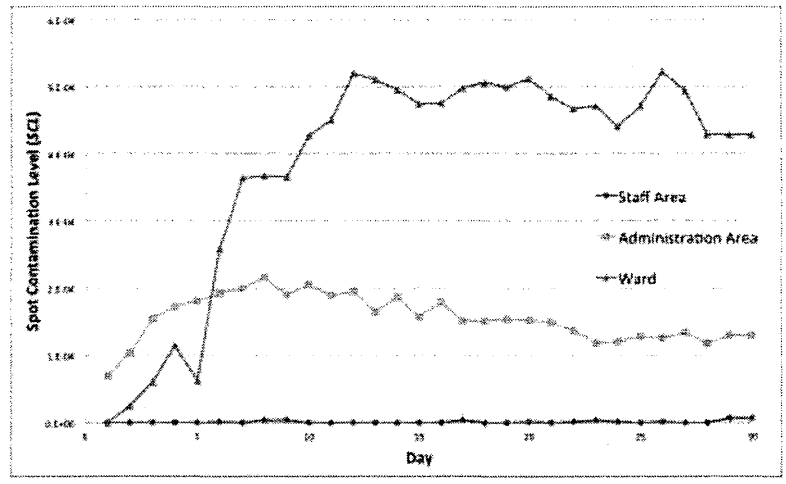

Fig. 6 Variation of Virus Contamination of areas in the hospital in scenario A

In section II.B, we have demonstrated the algorithm to calculate the amount of virtual influenza virus existing in spot and agent. Spot Contamination Level (*SCL*) at the certain time t is the sum of total amount of virus excretion of agents in the spot and the contamination level of the spot at time ($t - 1$). It depends on the number and the disease condition of infected agents existing in the spot. Figure 6 shows the average contamination level of Ward area, Administration area and Staff area in scenario A of High Control and High Vaccine. The results show that the Ward area is the most contaminated.

DOI: 10.5963/PHF0303003

The Administration area ranks the second while the Staff area is almost clean. The result implies that wards in hospital are likely to be contaminated with influenza virus when an outbreak of influenza emerges in the community.

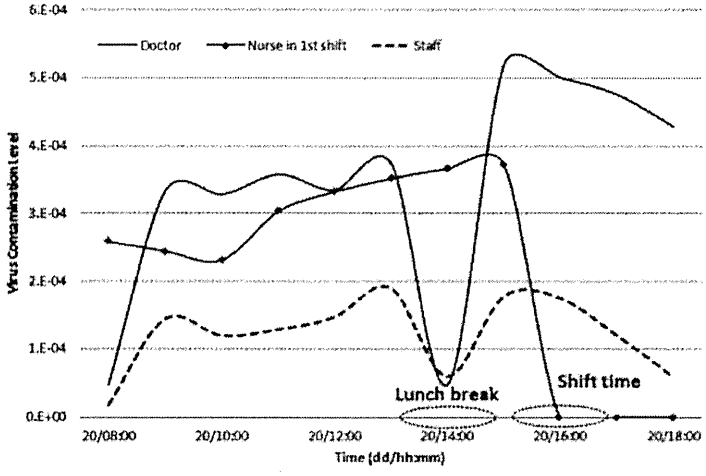

Fig. 7 Variation of virus contamination level of HCW in a working day in scenario A

Agent Contamination Level $ACL[i](t)$ is the amount of virus that an agent i has at the specific time t. Figure 7 describes the average virus contamination level of doctors, nurses and other staff in working time in 20th day when the number of inpatients reaches its peak in the scenario of A [High Control High Vaccine]. The average virus contamination level of doctors and nurses are higher than those of other staff. This could be explained by the fact that doctors and nurses work in ward area more than other staff. Sharp drops recorded in the contamination level among HCW strongly correlate with daily routines of the HCW concerned. The virus contamination level of doctors and other staff falls to their troughs at the time of lunch break (from 13:00 to 14:00). The virus contamination level of nurses also decreases rapidly when they change their shift and leave the hospital.

The conclusion of micro analysis is that doctors and nurses, who provide direct care to influenza patients have higher risk of catching influenza virus within the hospital. This conclusion supports long-standing belief in hospital infection control that annual influenza vaccination should be required for every health care worker who has direct contact with patients.

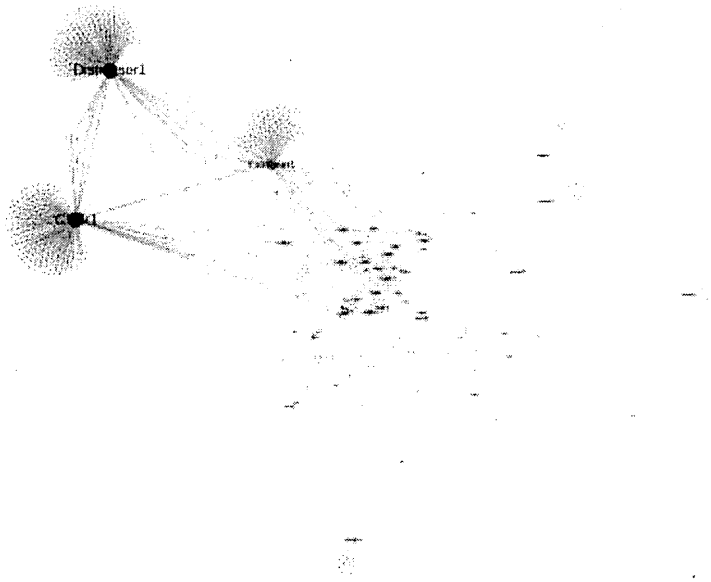

Fig. 8 Visualization of contact network in scenario D

DOI: 10.5963/PHF0303003

V. ANALYSIS ON CONTACT NETWORK

We analyze the contact network, which is generated by interactions of agents in the simulation of scenario D. We assume that once two agents come into the same spot, one contact is made between them. Each agent carries a contact list of agents who are in the same spot with him at the time t. The contact lists vary by time when agents are moving inside the hospital. The log of contact lists is converted to .csv file in order to be imported to Gephi [27], an open source graph visualization software.

Figure 8 illustrates the visualization of contact network, which we call "risk graph". Each node in the graph represents an agent in the simulation model. Lines in the graph illustrate aggregate of contact between agents the simulation of scenario D. The thicker the line, the more frequent contact between the agents has been made. The size of the node is proportional to the degree, which indicates amount of contact that he had made. The layout of the graph is Force Atlas, in which the connected nodes are attracted into the center of the graph and unconnected nodes are pushed out off the outside.

Visual conclusions of the risk graph:

- The dispenser, the clerk and the examiner (there is only one dispenser, one clerk and one examiner in the hospital) nodes are the three biggest nodes (in degree). It implies that the three heath care workers have made the most contacts with patients. However, most of the contacts were made with outpatients, so the nodes represent them are pulled out off center of the graph.
- Nodes that represent nurses, doctors and inpatients are attracted into the center of the graph (Figure 9). It implies that frequent close contacts were made between them.

Figure 9 shows the center of the risk graph, in which close contacts between doctors, nurses and inpatients are illustrated. The nodes, which are marked by blue explosion shapes, represent health care workers who have been infected during the simulation. We can see that those agents are at the center of the risk graph and have frequent contacts with inpatients. The nodes, which represent doctors, and nurses who have been infected during the simulation arose on top of the list of nodes in descending order of degree. However, the dispenser and the clerk were not infected, even though nodes representing them have a high degree. It can be explained that most of the contacts they made were with outpatients, so their risk of infection was low. The risk can also be evaluated by Spot Contamination Level of the place that the two staff were working in Figure 6. The Figure shows that staff area and administration area where the two staff work are less contaminated than ward area. It means that infection risk of those staff is lower than that of nurses who mostly work with inpatient in ward area.

Several conclusions can be drawn from the analysis on risk graph:

- Two nodes are spatially closer if they have a close and frequent contact.
- Close and frequent contacts between agents will attract them into center of the graph.
- Risk of infection can be assessed not only by the degree of the nodes but also by the amount of close contacts.

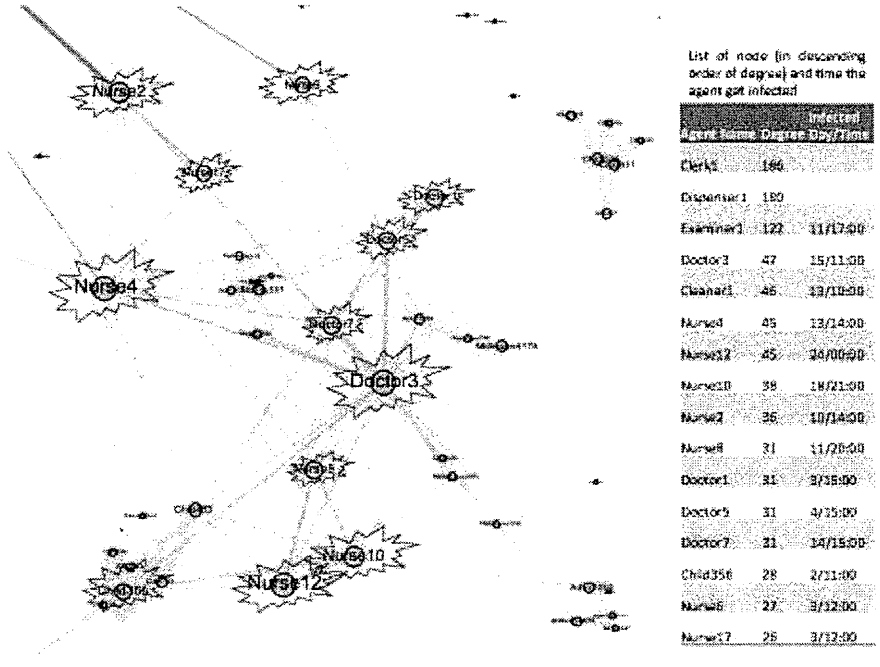

Fig. 9 Center of the risk graph and list of infected health care workers in scenario D

DOI: 10.5963/PHF0303003

VI. CONCLUSION AND DISCUSSION

We have built a simulation model for infection of an influenza-like illness in an artificial hospital and quantitatively assessed infection risk of the diseases. The simulation results have shed more light on epidemiological belief of that direct patient care HCW have high risk of catching nosocomial influenza virus and that washing hand and wearing mask are effective to prevent an outbreak of the disease in the hospital. The methodologies of quantification and visualization the infection risk have been demonstrated. The original approach has provided us with a potential methodology for risk management in infection control of nosocomial infection.

The great advantage of simulation model is that they are able to conduct experiments which are impossible or undesirable. It provides a flexibility of changing parameters to apply to other diseases rather than influenza-like illness. The computation of dynamical change of virtual influenza virus can assess the risk of infection quantitatively and visually. Even though the computational effort of the modeling method is hard, with the evolution of computing, time execution of the simulation model is constantly reduced.

The methodology of categorizing infection levels into detailed disease states can be used to apply to other pathogens like smallpox, measles and many others by changing state period and state transition probability. Compared to traditional SIR model, in which population is roughly divided in three groups of susceptible, infectious and recovered individuals, our methodology provides a better modeling of infection process.

The visualization of risk graph demonstrated above can be a valid method to assess infection risk but it is not completed. The nature of contacts that transmit virus cannot be seen from the graph. However, thanks to the development of large networks graphs visualization software, such as Gephi, we can highlight and track all contacts of agents in real time. An integration with human real time tracking systems can be potential for tracking and detecting contacts between health care workers or between health care workers and patients.

Although data and knowledge for the model have been constructed based on several field works onsite, empirical validation of the model could not be conducted due to the lack of statistics and impossibility of taking those experiments in a hospital. Although there was no observed data fitted the simulation results, outputs are qualitatively similar to observed phenomenon in the real world.

The future work is to integrate real data collecting by sensor to the simulation framework. The structure of the simulation framework is illustrated in Figure 10. We have developed and used wireless tracking systems to track real-time movement of humans in a building. The real data of movement of patients and health care workers in a real hospital can be achieved.

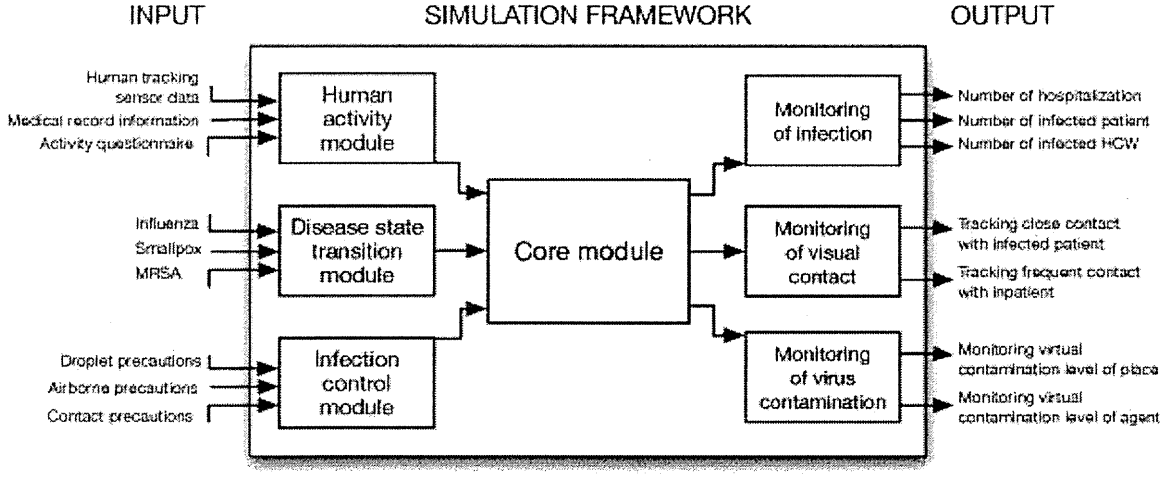

Fig. 10 Structure of the simulation framework

Activity pattern of people can also be collected via activity questionnaire. Changing parameters of the disease transition module can be applied to study other infectious diseases. Infection control measures can be changed in many scenarios depending on infection control resources of the hospital. The core module inherits from the current module but can be rebuilt to fit the structure of a new hospital. Simulation output shows real-time graph of the number of hospitalization, infected patient and HCW. By visualizing contact network, close and frequent contacts with high-risk patients can be tracked and monitored. Variation of virtual virus contamination level of places and agents can be monitored in real time. The simulation framework could be a potential decision-making support tool for hospital administrators to evaluate nosocomial infection control and it can also be used as an educational tool to study nosocomial infection.

DOI: 10.5963/PHF0303003

REFERENCES

[1] K. Klevens, R. Monina, et al. "Estimating health care-associated infections and deaths in US hospitals, 2002," Public health reports, vol. 122, no. 2, pp. 160-166, 2007.

[2] H.C.Maltezoua, M. Drancourt, "Nosocomial influenza in children." Journal of Hospital Infection, vol. 55, pp. 83-91, 2003.

[3] Bridges, Carolyn B., et al, "Prevention and control of influenza. Recommendations of the Advisory Committee on Immunization Practices (ACIP)." MMWR. Recommendations and reports: Morbidity and mortality weekly report. Recommendations and reports/Centers for Disease Control 51.RR-3, pp. 1-31, 2002.

[4] Sebille, V. & Valleron. A. J, "A computer simulation model for the spread of nosocomial infections caused by multidrug-resistant pathogens," Computers and Biomedical Research, vol. 30, no. 4, pp. 307-322, 1997.

[5] Triola, M.M., Holzman, R.S, "Agent-based simulation of nosocomial transmission in the medical intensive care unit," Computer-Based Medical Systems, Proceedings. 16th IEEE Symposium. IEEE, pp. 284-288, 2003.

[6] Boon Som Ong, Mark Chen, Vernon Lee, Joc Cing Tay, "An Individual-Based Model of Influenza in Nosocomial Environments," Lecture Notes in Computer Science, vol. 5101, pp. 590-599, 2008.

[7] Meng, Yang, et al, "An application of agent-based simulation to the management of hospital-acquired infection," Journal of Simulation, vol. 4, no. 1, pp. 60-67, 2010.

[8] Laskowski, Marek, et al, "Agent-based modeling of the spread of influenza-like illness in an emergency department: a simulation study," Information Technology in Biomedicine, IEEE Transactions , vol. 15, no. 6, pp. 877-889, 2011.

[9] N. Gilbert, "Agent-based model: Quantitative applications in the social sciences", SAGE Publications, pp. 38-46, 1999.

[10] N. Gilbert, K.Troitzsch, "Simulation for the Social Scientist", Open University Press, pp. 199-215, 2005.

[11] El-Sayed, Abdulrahman M., et al, "Social network analysis and agent-based modeling in social epidemiology," Epidemiologic Perspectives & Innovations, vol. 9, no. 1, 2002.

[12] Longini, Ira M., et al, "Containing pandemic influenza at the source," Science, vol. 309.5737, pp. 1083-1087, 2005.

[13] M.J. Keeling, M.E. Woolhouse et al, "Modelling vaccination strategies against foot-and-mouth disease," Nature, vol. 421, pp. 136-142, 2003.

[14] Meyers, Lauren Ancel, et al, "Applying network theory to epidemics: control measures for Mycoplasma pneumoniae outbreaks," Emerging infectious diseases, vol. 9, no. 2, pp. 204-210, 2003.

[15] Ichikawa, M., Koyama, Y., Deguchi, H, "Virtual City Model for Simulating Social Phenomena," Agent-Based Social Systems, vol. 7, pp. 253-264, 2010.

[16] Tanuma, H., Deguchi, H., Shimizu, T, "SOARS: Spot Oriented Agent Role Simulator - Design and Implementation," Agent-Based Simulation: From Modeling Methodologies to Real-World Applications, pp. 1-15, 2005.

[17] Bean, B., et al," Survival of influenza viruses on environmental surfaces," Journal of infectious diseases, vol. 146, no. 1, pp. 47-51, 1982.

[18] Centers for Disease Control and Prevention, "Prevention Strategies for Seasonal Influenza in Healthcare Settings," http://www.cdc.gov/flu/professionals/infectioncontrol/healthcaresettings.htm. Accessed January 25, 2011.

[19] Naffakh, Nadia, and Sylvie Van Der Werf, "April 2009: an outbreak of swine-origin influenza A (H1N1) virus with evidence for human-to-human transmission," Microbes and infection, vol. 11, no. 8, pp. 725-728, 2009.

[20] H. Deguchi, T. Saito, M. Ichikawa, H. Tanuma, "Simulated Tabletop Exercise for Risk Management - Anti Bio-Terrorism Multi Scenario Simulated Tabletop Exercise," Development in Business Simulation and Experimental Learning, vol. 38, pp.1-21, 2011.

[21] Glasser, John, et al, "Evaluation of targeted influenza vaccination strategies via population modeling," PloS one 5.9: e12777, 2010.

[22] Grayson, M. Lindsay, et al, "Efficacy of soap and water and alcohol-based hand-rub preparations against live H1N1 influenza virus on the hands of human volunteers," Clinical Infectious Diseases, vol. 48, no. 3, pp. 285-291, 2009.

[23] Loeb, Mark, et al, "Surgical mask vs N95 respirator for preventing influenza among health care workers: a randomized trial," Jama, vol. 302.17, pp. 1865-1871, 2009.

[24] Otter, J. A., et al, "Assessing the biological efficacy and rate of recontamination following hydrogen peroxide vapour decontamination," Journal of Hospital Infection, vol. 67, no. 2, pp. 182-188, 2007.

[25] Cowling, Benjamin J., et al, "Facemasks and Hand Hygiene to Prevent Influenza Transmission in HouseholdsA Cluster Randomized Trial," Annals of Internal Medicine, vol. 151, no. 7, pp. 437-446, 2009.

[26] Aiello, Allison E., et al, "Facemasks, hand hygiene, and influenza among young adults: a randomized intervention trial," PloS one 7.1: e29744, 2012.

[27] Bastian, M., Heymann, S., Jacomy, M, "Gephi: an open source software for exploring and manipulating networks." ICWSM, pp. 361-362, 2009.

DOI: 10.5963/PHF0303003

A Molecular Typing Methodology of *Mycobacterium tuberculosis* using Small Genomic Islet Patterns (TB-SGIP): A Novel Genotyping Methodology to Discriminate Clinical Strains between Beijing Family and T3-OSAKA

Tomoshige Matsumoto[1,2], Masahiro Suzuki[3], Yoshitsugu Iinuma[4], Shinji Maeda[5], Hiromi Ano[2], Yuriko Koshii[2], Tomomi Murakawa[2], Koichi Suzuki[6] and Yoshihiko Hoshino[6,*]

[1]*Department of Clinical Laboratory Medicine, Osaka Anti-Tuberculosis Association Osaka Hospital, Neyagawa-city, Osaka 572-0854, Japan;* [2]*Department of Medicine, Osaka Prefectural Medical Center for Respiratory and Allergic Diseases, Habikino-city, Osaka 583-8588, Japan;* [3]*Department of Microbiology, Aichi Prefectural Institute of Public Health, Kita-ward, Nagoya 462-8576 Japan;* [4]*Kanazawa Medical University, Kahaku, Kanazawa 920-0293 Japan;* [5]*The Research Institute of Tuberculosis, Japan Anti-Tuberculosis Association, Kiyose, Tokyo 204-8533, Japan; and* [6]*Department of Mycobacteriology, National Institute of Infectious Diseases, Higashi-Murayama, Tokyo 189-0002, Japan*

Abstract: Tuberculosis remains a leading cause of infectious disease morbidity and mortality worldwide. Therefore, the ability to discriminate between strains of the causative agent, *Mycobacterium tuberculosis* (*M. tuberculosis*), is an important issue for public health epidemiologists when tracking transmission of the disease. Several molecular typing methods are in use; however, some are labor intensive and technically demanding, while others require expensive equipment. Recently, a novel analysis of the presence or absence of genomic signatures in open reading frames (ORFs) was demonstrated to have several advantages over conventional typing methods: it uses a simple and rapid polymerase chain reaction (PCR)-based method, and the results are stable and easily comparable among institutions. We used this technology to develop a novel molecular typing method to assess the composition of small genomic islets (SGIs) derived from ORFs in the genomic sequences of *M. tuberculosis*. Analyses of the whole genome sequences of five *M. tuberculosis* complex strains led to the selection of nine SGIs. The conservation of each SGI was investigated in 40 clinical isolates from several different groups. The SGI patterns exhibited a clear difference between Beijing family members and T3-OSAKA, suggesting that the analysis of SGIs is useful for the discrimination of these strains.

Keywords: Open reading frame (ORF), Small genomic islets (SGIs), Restriction fragment length polymorphism (RFLP), Variable number of tandem repeats (VNTR), Spoligotyping, Direct repeat (DR), Multilocus sequence typing (MLST).

INTRODUCTION

Mycobacterium tuberculosis (M. tuberculosis), the etiological agent of tuberculosis, has been a significant source of morbidity and mortality since antiquity. Worldwide, approximately 8.6 million people were affected in 2012, with 1.3 million deaths [1]. The most widespread *M. tuberculosis* strains are members of the Beijing family (known as lineage 2, [2]). Beijing is the dominant family in many locations in East Asia, particularly in China [3,4], and is the primary source of the emerging multidrug resistant (MDR) strains [5,6] that are responsible for outbreaks in many countries.

Because *M. tuberculosis* is primarily transferred among humans, strain discrimination can be a vital key in tracking the course of transmission among patients. Several molecular methods are used to discriminate clinical strains of *M. tuberculosis*: restriction fragment length polymorphisms (RFLP), variable number of tandem repeats (VNTR) and spoligotyping. Insertion sequence (IS) *6110*-based RFLP is the gold standard for *M. tuberculosis* strain discrimination [7]. *M. tuberculosis* contains 0-26 copies of IS*6110* in the chromosome, which form different RFLP gel banding patterns for epidemiologically unrelated strains [8]. However, this methodology has several disadvantages: the method is labor intensive and requires a large amount of high quality bacterial genomic DNA; the analog band pattern must be analyzed digitally for inter-facility comparisons; and the polymorphisms might be of limited use in supplemental methods if there are fewer than six insertion sites.

VNTR analysis utilizes a variable number of tandem repeats for genotyping [9,10]. Bacterial genomes contain many interspersed repeat units, the number and size of which can be used to categorize microbial strains following PCR amplification [11]. The digitally acquired data are more suited for further analyses or

*Address correspondence to this author at the Department of Mycobacteriology, National Institute of Infectious Diseases, Tokyo 189-0002, Japan; Tel: +81-42-391-8211; Fax: +81-42-391-8807;
E-mail: yhoshino@nih.go.jp

E-ISSN: 2310-9386/14

comparisons. However, when VNTRs were used to analyze the same locus (target) in lineage 4 (Europe, America, Africa [2]) and lineage 2 (East Asia) mycobacteria, the method could not discriminate the Beijing strains [12,13]. A large collection of Beijing strains was not differentiated by Supply's 15-loci VNTR analysis because this assay was initially developed to analyze lineage 4, not lineage 2 (Beijing family) [14-16]. Therefore, a novel VNTR assay with better discriminatory power for East Asian strains, particularly lineage 2, is needed [17,18].

The third discrimination method, spoligotyping [19], utilizes a genomic locus that contains approximately 40 copies (the analysis of a copy number of 43 is universally available) of a short direct repeat (DR). DRs are separated by spacers whose sequence, between specific DRs, is conserved among strains. Because strains differ in terms of the presence or absence of specific spacers, the pattern of spacers in a strain can be used for genotyping. Spoligotyping has several advantages over IS*6110*-based RFLPs. Since only small amounts of DNA are required, it can be performed directly from clinical samples or from colonies shortly following inoculation. The results are expressed as positive or negative for each spacer and can be expressed in a digital format.

T3-OSAKA (also known as SIT627 [20]) was named after Osaka, Japan, where the strain was initially isolated. It was characterized with spoligotyping by the presence of spacers 1-4 and 9-12 and the absence of spacers 5-8 and 33-36 [20-23]. Previously, the T3 strain (T represents modern TB strains in the international spoligotyping database) was defined only by the absence of spacers 13 and 33-36. Thus, T3-OSAKA was expected to be a variation of the T3 strain. T3-OSAKA is found mainly in Japan, but has also been reported in Finland and Mexico, which has been confirmed in SpolDB4 (fourth international spoligotyping database) [24].

In the Osaka area, the Beijing family is the predominant strain, and here the proportion of MDR patients is higher than the national average in Japan. Because Beijing strains are expected to develop drug resistance faster than T3-OSAKA [23], the importance of discriminating between strains becomes obvious. The best way to differentiate between Beijing and T3-OSAKA strains is to detect a specific pattern of spacers by spoligotyping. But spoligotyping is expensive and requires a special apparatus. A new typing method that can discriminate T3-OSAKA and Beijing strains must be simple to perform and less expensive than spoligotyping.

A new molecular typing method, phage derived open reading frame typing (POT) is used to discriminate methicillin-resistant *Staphylococcus aureus* (*S. aureus*) (MRSA) and/or *Salmonella enterica* serovar *typhimurium* infection [25-27]. MRSA-POT was developed based on the observation that the *S. aureus* genome contains regions of open reading frames (ORFs) from bacteriophages. Phages, as well as small genomic islets (SGIs), are frequently utilized as ORF readouts. For the MRSA-POT assay, SGIs in *S. aureus* were defined as small ORFs comprised of one to five genes as well as single-nucleotide variation [28]. In a case of MRSA, the pattern of the presence or absence of such SGIs in the phages revealed high resolution and high diversity of discrimination [25,26,29,30]. The presence or absence of a particular SGI signature was associated with multilocus sequence typing (MLST) results in the discrimination of *S. aureus* [31]. Utilizing such patterns has several advantages. For example, the results are obtained rapidly and easily, are consistent among operators, and can be compared among different institutions.

The genomes of the *M. tuberculosis* complex (MTBC) have been sequenced [32,33] and are now available from websites such as the US National Library of Medicine and GenBank. The sequencing data could be used to facilitate the development of cost-effective genotyping methodologies for MTBC. Similar ORFs comprising SGIs may be present or absent in the genomic signature of different strains of *M. tuberculosis*. We have developed a novel molecular typing method for *M. tuberculosis* using SGI Patterns (TB-SGIP) and have confirmed its ability to discriminate between T3-OSAKA and members of the Beijing family.

METHODS

Selection of Candidate SGIs from Whole-Genome Sequence Data

The whole genome sequence data of five MTBC strains already published in the microbial genome database for comparative analysis web site (http://mbgd.genome.ad.jp/) were compared: CDC1551, F11, H37Ra (ATCC25177), H37Rv and *M. bovis* (AF2122_97). SGIs that are absent in at least one of the five strains were chosen using Find Differences JP (Saglasie, Yokohama), a software package that

Table 1: Loci of Small Genomic Islets in Five Sequenced *M. tuberculosis* Strains

Islet No.	Locus Tag	*M. tuberculosis* Strains					Predicted Function
		H37Rv	CDC1551	AF2122_97	H37Ra	F11	
1	MT3248	ND[a]	100[b]	100	ND	ND	PPE family protein
2	MT3428	ND	100	100	100	ND	AfsR/Dnrl/RedD family
3	MT1800	ND	100	99.833	100	100	Glycosyltransferase
4	TBFG_13356	ND	100	100	ND	100	Carbinolamine dehydrase
5	TBFG_1325	ND	100	100	ND	ND	Adenylate cyclase
6	MT1360	ND	100	ND	ND	ND	Adenylcyclase, putative
7	MT2166	ND	100	100	ND	100	PE family protein
8	MT3247	ND	100	100	ND	ND	Molybdoprotein co-factor
9	MRA_3466	100	ND	ND	ND	ND	PPE family protein

[a]ND: not detected.
[b]Percentage homology to the locus tag.

facilitates text comparison. The nucleotide sequences of the five strains were then compared again using nucleotide analysis software Genetyx Win (Genetyx, Tokyo). Among the selected SGIs, those containing a single ORF without the presence of structures resembling insertion sequences, transposases or integrases were selected (Table **1**). Non-conserved regions with larger structures such as transposons and prophages were excluded.

Detecting SGIs

SGIs were detected and confirmed with the GeneAmp PCR System 9600 (Applied Biosystems, USA, CA). Conventional PCR was performed in a 20-μl mixture containing 10 ng of template DNA, Ex Taq buffer (2 mM Mg^{2+}), dNTPs (0.2 mM each), 0.4 units of Ex Taq DNA polymerase (Takara Bio, Otsu, JAPAN) and 0.2 μmol^{-1} of primer pairs (Table **2**). The PCR consisted of 30 cycles of the following steps: 94°C for 30 s, 53°C for 30 s and 72°C for 1 min. The PCR products (5 μl) were electrophoresed on 2.5% agarose gels in 0.5x Tris-borate-EDTA at 100 V for 25 min; the bands were then visualized with ethidium bromide under UV light.

M. tuberculosis Clinical Isolates

Samples were obtained from the Osaka Prefectural Medical Center for Respiratory and Allergic Diseases and National Institute of Infectious Diseases, Japan. The *M. tuberculosis* strains were isolated from 326 pulmonary TB patients in 2009. Forty consecutive clinical isolates (12.3% = 40/326) of MTBC were collected for molecular typing, with three laboratory strains (BCG Tokyo, H37Rv, and H37Ra) used as controls. The 40 clinical isolates were drug sensitive.

Table 2: PCR Primers for the Detection of Small Genomic Islets

Islet No.	Name	Primers	Product Size
1	MT3248F MT3248R	ATTATTGAAGCCCGAGACGC GGGAACTCTGGGAACAACAA	464
2	MT3428F MT3428R	CAAAATCATCGCCCACACAG GATTGACGTTCCGAGACATAG	159
3	MT1800F MT1800R	TAGCCCAAATGTCTCGTGTG AAATCCGTCACCGTCCCACT	361
4	TBFG_13356F TBFG_13356R	TCCAACAAGTCAACCGCAC ATTTGGGTTTGGGAGCTTCG	128
5	TBFG_01325F TBFG_01325R	TCAAAAGGAACCCGCCGATCA CTATTCATCCCCGCAGGTCAT	316
6	MT1360F MT1360R	ATGACCTGCGGGATGAATAG CTTTTCCGACACCTAGCATC	292
7	MT2166F MT2166R	ATTGGGATGTTGGCGGCA CTGACTTCCTCATACGCA	250
8	MT3427F MT3427R	GTATCACCACGGAATCAAGC GCGATCACAACTAATGGCGT	211
9	MRA_3466F MRA_3466R	GCAAAACACGTCAGCAACA GGCCGAAAATATCCAAGCAA	251

Molecular Typing

IS6110-Based RFLP

*Pvu*II-digested DNA of *M. tuberculosis* was probed with the insertion element IS*6110* according to a

standardized protocol [7]. Band patterns were analyzed using the BioNumerics software package (Applied Maths). Strains with identical RFLP band patterns were categorized as a cluster.

VNTR

Among 24 mycobacterial interspersed repetitive units (MIRU) loci, including four exact tandem repeats (ETRs) and 20 other loci proposed for VNTR [9,14-16,34-37], 15 were selected for use in typing Beijing strains as previously described [38]. VNTR typing was performed using ExTaq with GC PCR buffer I (Takara Bio). The PCR mixture was prepared in a 20-μl volume with 1x GC PCR buffer I, 0.5 U Ex Taq, 200 μM each of four dNTPs, 0.5 mM of the primer set and 10 ng template DNA. All loci were amplified under the following conditions: initial denaturation at 94°C for 5 min; 35 cycles of 94°C for 30 s, 63°C for 30 s and 72°C for 3 min; followed by a final extension at 72°C for 7 min.

Spoligotyping

Spoligotyping was performed as previously described [19]. Spoligotype families were assigned as described [22,24,39].

Minimal Spanning Tree (MST) Analysis

MST analysis was performed, based on the VNTR genotypes, using BioNumerics software version 4.61 (Applied Maths) to reconstruct a hypothetical phylogenetic tree for clinical isolates. The reconstruction was performed using previously reported criteria [40].

Statistics

The Discriminatory Power (*D*) is the average probability that the typing system will assign a different type to two unrelated strains randomly sampled in the microbial population of a given taxon. *D* was calculated using Simpson's index of diversity formula [41].

RESULTS

Detecting SGIs

Non-conserved regions among the five sequenced strains were predicted using computational analysis to identify nine potential SGI targets (Table **1**). All SGIs were found in non-essential genes [32]. Conventional PCR was used to confirm the presence or absence of each genomic islet in the selected ORF using the primers listed in Table **2**. PCR analysis showed different SGIs in the ORF of the five strains, suggesting that these nine loci could be good markers for *M. tuberculosis* strain discrimination (data not shown).

Spoligotyping and RFLP Analysis

Three laboratory strains and 40 clinical isolates were evaluated by the conventional discrimination methods of spoligotyping and RFLP analysis. Spoligotyping of the clinical isolates revealed one large group clustering as T3-OSAKA (n = 18, #4-21, Table **3**) that was characterized by the presence of spacers 1-4 and 9-12 and the absence of spacers 5-8 and 33-36 [22]. Another large group clustering as Beijing (n = 8, #30-37, Table **3**) had a specific spoligotype with the presence of spacers 35-43 [22]. Other clinical isolates (#38-43, Table **3**), such as T1, T2, T3, T5, Haarlem (H) 1, and Latin America (LAM)1, also showed specific spoligotyping patterns predicted by SpolDB4 [24](Table **3**).

In the laboratory strains, spoligotyping analysis revealed the absence of spacers 3, 9, 16, and 39-43 in BCG Tokyo and the absence of spacers 20-22 and 33-36 in H37Rv and H37Ra [24]. When these spoligotyping results were compared with those of the RFLP analysis, the number of RFLP bands was variable, leading to the conclusion that they do not have the same genotype. In the T3-OSAKA group, the RFLP profile varied from one to five bands (gel fragments). In addition, four strains had spoligotyping patterns similar to those of the T3-OSAKA group. These isolates were named the Osaka-Family, although they differed in the number of RFLP bands. Another four sporadic isolates exhibited a small number of bands (1-5), but the spoligotyping patterns did not match those of either the Beijing or the T3-OSAKA strains. They were labeled non-B (Beijing), non-O (T3 OSAKA) strains. In contrast, RFLP analysis showed that the Beijing strains had multiple copy numbers of bands (9-16 bands). BCG Tokyo had two bands [42], while H37Rv and H37Ra had fourteen and fifteen bands, respectively (Table **3**) [43].

VNTR Analysis

The number of interspersed repeat units in each of the 43 isolates was analyzed by VNTR analysis (Table **4**). Based on the results in Table **4**, T3-OSAKA and OSAKA-Family strains appeared to have similar VNTR patterns, whereas members of the Beijing family shared a different pattern. We reconstructed putative phylogenetic trees of the 40 clinical isolates and three

Table 3: Comparative Results of Spoligotyping and RFLP Analysis

Isolate No.	Strain	Spoligotyping Pattern	Number of RFLP Bands
1	BCG Tokyo	absent of 3, 9, 16 and 39 to 43	2
2	H37 Rv	absent of 20 to 22 and 33 to 36	14
3	H37 Ra	absent of 20 to 22 and 33 to 36	15
4	T3 OSAKA	absent of 5 to 8 and 33 to 36	2
5	T3 OSAKA	absent of 5 to 8 and 33 to 36	1
6	T3 OSAKA	absent of 5 to 8 and 33 to 36	1
7	T3 OSAKA	absent of 5 to 8 and 33 to 36	1
8	T3 OSAKA	absent of 5 to 8 and 33 to 36	2
9	T3 OSAKA	absent of 5 to 8 and 33 to 36	5
10	T3 OSAKA	absent of 5 to 8 and 33 to 36	1
11	T3 OSAKA	absent of 5 to 8 and 33 to 36	1
12	T3 OSAKA	absent of 5 to 8 and 33 to 36	3
13	T3 OSAKA	absent of 5 to 8 and 33 to 36	1
14	T3 OSAKA	absent of 5 to 8 and 33 to 36	1
15	T3 OSAKA	absent of 5 to 8 and 33 to 36	1
16	T3 OSAKA	absent of 5 to 8 and 33 to 36	1
17	T3 OSAKA	absent of 5 to 8 and 33 to 36	1
18	T3 OSAKA	absent of 5 to 8 and 33 to 36	1
19	T3 OSAKA	absent of 5 to 8 and 33 to 36	1
20	T3 OSAKA	absent of 5 to 8 and 33 to 36	1
21	T3 OSAKA	absent of 5 to 8 and 33 to 36	1
22	OSAKA-family	absent of 1, 5 to 8, 13 and 33 to 36	2
23	OSAKA-family	absent of 5 to 8, 13 to 14 and 33 to 36	4
24	OSAKA-family	absent of 5 to 8, 13, and 33 to 36, 40	5
25	OSAKA-family	absent of 3 to 11, 13, and 33 to 36	1
26	non-B, non-O	absent of 33 to 36	1
27	non-B, non-O	absent of 13, 33 to 36	1
28	non-B, non-O	absent of 13 to 16, 29 to 32, 34 and 40 to 43	5
29	non-B, non-O	absent of 13 to 16 and 33 to 36	5
30	Beijing	absent of 35 to 43	15
31	Beijing	absent of 35 to 43	15
32	Beijing	absent of 35 to 43	14
33	Beijing	absent of 35 to 43	12
34	Beijing	absent of 35 to 43	13
35	Beijing	absent of 35 to 43	16
36	Beijing	absent of 35 to 43	16
37	Beijing	absent of 35 to 43	12
38	LAM1	absent of 3, 21, 24, and 33 to 36	11
39	T1	absent of 33 to 36	15
40	T3	absent of 13 and 33 to 36	15
41	H1	absent of 26 to 29 and 33 to 36	9
42	T5	absent of 23 and 33 to 36	14
43	T2	absent of 33 to 36 and 40	12

Table 4: Comparative Results of VNTR Analyses

No.	Strain	VNTR Pattern														
		ETR				MIRU				QUB			Mtub			
		A	C	D	E	10	16	26	40	11b	26	4156	4	21	30	39
1	BCG Tokyo	5	5	3	3	2	3	5	2	4	5	1	0	0	2	2
2	H37Rv	3	4	3	3	3	2	3	1	5	5	3	2	2	2	5
3	H37Ra	3	4	3	3	3	2	3	1	5	5	3	2	2	2	5
4	T3 OSAKA	3	3	2	3	5	11	2	3	4	11	1	0	2	3	3
5	T3 OSAKA	3	4	2	3	1	11	2	3	7	11	1	0	1	3	3
6	T3 OSAKA	3	4	2	3	3	11	2	3	9	11	1	0	2	3	3
7	T3 OSAKA	3	4	2	3	5	11	2	3	9	11	1	0	2	3	3
8	T3 OSAKA	3	4	2	3	5	11	1	3	9	12	1	0	2	3	5
9	T3 OSAKA	3	4	2	3	3	11	1	3	9	12	1	0	2	3	3
10	T3 OSAKA	3	4	2	3	5	11	2	3	5	12	1	0	3	3	4
11	T3 OSAKA	3	4	2	3	5	11	2	3	9	12	1	0	2	3	5
12	T3 OSAKA	3	4	2	3	5	11	2	3	9	12	1	0	2	3	5
13	T3 OSAKA	3	4	2	3	5	12	2	3	7	12	1	0	2	3	3
14	T3 OSAKA	3	4	2	3	5	11	2	3	10	13	1	0	2	3	3
15	T3 OSAKA	3	4	2	3	5	11	3	3	9	12	1	0	2	3	3
16	T3 OSAKA	3	4	2	3	5	12	2	3	6	12	1	0	2	3	3
17	T3 OSAKA	3	4	2	3	5	11	3	3	9	12	1	1	2	3	2
18	T3 OSAKA	3	4	2	4	5	11	2	3	9	12	1	0	2	3	4
19	T3 OSAKA	3	4	2	3	5	11	3	3	9	12	1	0	2	3	3
20	T3 OSAKA	3	4	2	3	5	11	2	3	10	13	1	0	2	3	3
21	T3 OSAKA	3	4	2	4	5	11	2	3	6	12	1	0	2	3	3
22	Osaka-Family	3	4	2	3	5	11	2	3	5	11	1	0	2	1	3
23	Osaka-Family	3	4	2	3	3	11	3	3	8	11	1	0	2	3	3
24	Osaka-Family	3	4	2	3	5	11	2	3	5	12	1	0	2	3	5
25	Osaka-Family	3	3	2	4	5	11	2	3	9	5	1	0	2	3	2
26	non-B, non-O	3	4	5	3	2	3	5	2	5	7	6	12	0	1	3
27	non-B, non-O	9	2	6	5	6	3	1	3	8	6	5	10	6	1	7
28	non-B, non-O	4	4	4	3	3	3	4	4	5	7	6	10	2	1	3
29	non-B, non-O	3	4	4	3	3	3	4	4	5	8	6	10	2	1	3
30	Beijing	4	4	3	5	3	4	7	3	7	9	5	11	4	5	7
31	Beijing	4	4	3	5	3	4	7	4	7	3	5	9	3	3	7
32	Beijing	4	4	3	5	3	4	7	3	7	3	6	11	3	5	7
33	Beijing	4	4	3	5	3	3	7	3	8	11	4	12	5	5	7
34	Beijing	4	4	3	5	3	3	7	3	8	10	4	12	3	5	7
35	Beijing	4	4	3	5	3	3	7	3	6	9	6	12	4	5	7
36	Beijing	4	4	3	5	3	3	6	3	6	9	6	12	4	5	7
37	Beijing	4	2	3	5	1	3	7	3	7	9	8	12	5	5	7
38	LAM1	1	1	4	3	3	2	3	2	4	2	6	1	5	2	1

(Table 4) contd .

No.	Strain	VNTR Pattern														
		ETR				MIRU				QUB			Mtub			
		A	C	D	E	10	16	26	40	11b	26	4156	4	21	30	39
39	T1	2	2	3	3	2	2	2	2	3	2	5	1	5	2	3
40	T3	3	1	4	3	3	2	3	2	3	2	5	1	4	2	3
41	H1	3	2	3	3	3	2	3	2	5	1	3	1	5	2	1
42	T5	3	2	4	3	3	2	3	2	3	2	6	1	5	1	1
43	T2	3	2	4	3	3	2	3	2	3	2	6	1	4	1	1

ETR: Exact tandem repeat [9].
MIRU: Mycobacterial Interspersed Repetitive Units [16].
QUB: Queen's University Belfast [37].
Mtub: Mycobacterium tuberculosis [36].

laboratory strains based on their VNTR types using the minimal spanning tree algorithm (Figure **1**). In the analysis, T3-OSAKA and Beijing isolates were clearly distinguishable, suggesting that they are genetically independent. Osaka-Family isolates were associated with T3-OSAKA, which was predicted from the spoligotyping results. Notably, although non-B, non-O isolates should differ from both T3-OSAKA and Beijing by spoligotyping, the VNTR-produced tree predicted an association between non-B, non-O and Beijing strains.

TB-SGIP Analysis

The new TB-SGIP assay was evaluated by analyzing the 40 clinical tuberculosis strains that had been examined by conventional analysis. Three laboratory strains were included as internal controls for the experiment. The results were displayed in a binary system with nine columns (1= SGI was present, 0 = SGI was absent) (Table **5**). All nine SGIs were present in the T3-OSAKA strains. All SGIs, with the exception of MT-1800, were present in the Beijing strains. These

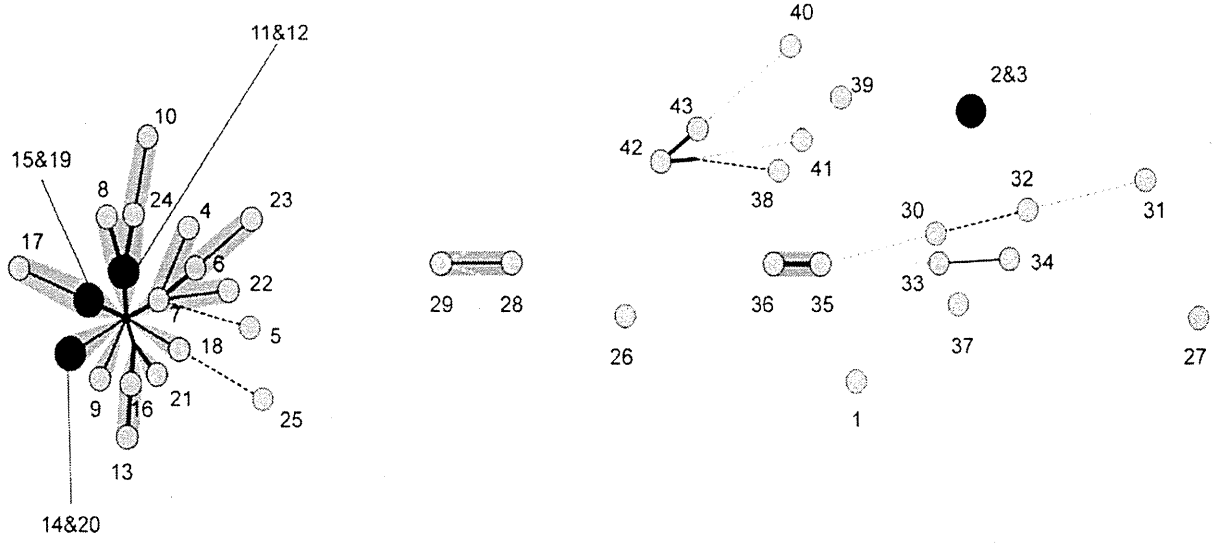

Figure 1: Minimal spanning tree of 40 clinical isolates and 3 laboratory strains based on VNTR assay.

Each circle corresponds to a different VNTR genotype. A small circle represents one strain and a large circle, two strains. Heavy lines connecting two VNTR types signify that they are single-locus variants, thin lines connect double-locus variants, and dotted lines (black) connect triple-locus variants. Dotted lines represent the most likely connection between two types differing by more than three VNTR loci. Numbers are same as those shown in Tables **3-5**. T3-OSAKA (# 4-21) and Beijing strains (# 30-37) are clearly separated in the tree.

Table 5: TB-SGIP Assays

Isolate No.	Strain	TB-SGIP								
		3248	3428	1800	13356	1325	1360	2166	3427	3466
1	BCG Tokyo	1	1	1	1	0	0	1	1	0
2	H37 Rv	1	0	0	0	0	0	0	0	1
3	H37 Ra	1	0	1	0	0	0	0	0	1
4	T3 OSAKA	1	1	1	1	1	1	1	1	1
5	T3 OSAKA	1	1	1	1	1	1	1	1	1
6	T3 OSAKA	1	1	1	1	1	1	1	1	1
7	T3 OSAKA	1	1	1	1	1	1	1	1	1
8	T3 OSAKA	1	1	1	1	1	1	1	1	1
9	T3 OSAKA	1	1	1	1	1	1	1	1	1
10	T3 OSAKA	1	1	1	1	1	1	1	1	1
11	T3 OSAKA	1	1	1	1	1	1	1	1	1
12	T3 OSAKA	1	1	1	1	1	1	1	1	1
13	T3 OSAKA	1	1	1	1	1	1	1	1	1
14	T3 OSAKA	1	1	1	1	1	1	1	1	1
15	T3 OSAKA	1	1	1	1	1	1	1	1	1
16	T3 OSAKA	1	1	1	1	1	1	1	1	1
17	T3 OSAKA	1	1	1	1	1	1	1	1	1
18	T3 OSAKA	1	1	1	1	1	1	1	1	1
19	T3 OSAKA	1	1	1	1	1	1	1	1	1
20	T3 OSAKA	1	1	1	1	1	1	1	1	1
21	T3 OSAKA	1	1	1	1	1	1	1	1	1
22	OSAKA-family	1	1	1	1	1	1	1	1	1
23	OSAKA-family	1	1	1	1	1	1	1	1	1
24	OSAKA-family	1	1	1	1	1	1	1	1	1
25	OSAKA-family	1	1	1	1	1	1	1	1	1
26	non-B non-O	0	1	1	1	1	1	1	1	1
27	non-B non-O	1	1	1	1	0	0	1	1	1
28	non-B non-O	1	1	1	1	1	1	1	1	1
29	non-B non-O	0	1	0	1	0	0	1	1	1
30	Beijing	1	1	0	1	1	1	1	1	1
31	Beijing	1	1	0	1	1	1	1	1	1
32	Beijing	1	1	0	1	1	1	1	1	1
33	Beijing	1	1	0	1	1	1	1	1	1
34	Beijing	1	1	0	1	1	1	1	1	1
35	Beijing	1	1	0	1	1	1	1	1	1
36	Beijing	1	1	0	1	1	1	1	1	1
37	Beijing	1	1	0	1	1	1	1	1	1
38	LAM1	0	1	0	1	1	1	1	1	0
39	T1	0	1	0	1	1	1	1	1	1

(Table 5) contd .

Isolate No.	Strain	TB-SGIP								
		3248	3428	1800	13356	1325	1360	2166	3427	3466
40	T3	0	1	1	1	1	1	1	1	1
41	H1	1	1	1	0	0	0	1	1	1
42	T5	1	0	1	0	1	1	1	0	1
43	T2	1	0	1	0	1	1	1	0	1

Presence or absence of SGIs = 1.
Absence of SGIs = 0.

results were more consistent with spoligo analysis rather than RFLP or VNTR. Thus, comparing spoligotyping and the TB-SGIP assay, the spoligotyping results correlated well with those of TB-SGIP, and the *D* of TB-SGIP was the same as that of spoligotyping: 0.7697 (TB-SGIP) and 0.7962 (spoligotyping). Comparisons between T3-OSAKA or OSAKA-Family strains and Beijing strains by TB-SGIP were relatively consistent with those of spoligotyping, where the *D* of the assays was 0.4049 (TB-SGIP) and 0.5839 (spoligotyping). Thus, TB-SGIP analysis and spoligotyping similarly discriminated between Beijing and T3-OSAKA strains.

DISCUSSION

In this study, we developed a novel molecular discrimination assay, TB-SGIP, for *M. tuberculosis* based on differences in patterns produced by the presence or absence of small genomic islets in different strains. TB-SGIP data were analyzed and compared with traditional methods (i.e., RFLP, VNTR and spoligotyping). TB-SGIP was not inferior to spoligotyping and is better than conventional RFLP. In addition, TB-SGIP employs conventional PCR, which is less labor intensive than VNTR and less expensive than the spoligotyping assay. TB-SGIP is also as sensitive as VNTR. Consequently, TB-SGIP would be appropriate for the typing of different lineage strains.

The *D* of TB-SGIP and spoligotyping are similar, making it suitable for global surveillance and phylogenetic analysis. Moreover, the simple combination of PCR and agarose gel electrophoresis decreases the time needed to detect Beijing or T3-OSAKA strains over conventional assays such as VNTR analysis and spoligotyping. A limitation of this analysis would be the relatively small number of non-B, non-O strains (n = 4), preventing the calculation of a false positive rate.

The TB-SGIP results were displayed with the number of small genomic islands in the strains. As all eighteen strains had the same results, T3-OSAKA appeared to be one of the ancient lineages of MTBC. A comparison of the Beijing and T3-OSAKA strains showed only one deletion of a genomic island in TB-SGIP. These data suggest that Beijing and T3-OSAKA strains might be closely related. However, other assays produced significant differences among them. Genomic data from five sequenced strains were used to establish SGIP, which could be listed as another limitation of the current analysis since there might be a sequence bias in the reference strains. This bias could be removed with future access and use of more whole genome data of *M. tuberculosis* strains.

Although H37Rv and H37Ra share a common ancestor in H37, one is virulent (H37Rv) while the other is not. The difference between the two is the presence of MT1800 in H37Ra. MT1800 might function to influence virulence attenuation between these strains. Therefore, a check for the presence of MT1800 by PCR is sufficient for discrimination of Beijing and T3-OSAKA strains. In addition, use of the TB-SGIP assay might lead to prediction of strain characteristics if further study confirms the functions of each genomic island.

In summary, TB-SGIP is as good as spoligotyping and superior to RFLP analysis for the typing of *M. tuberculosis*. In this study we compared only two strains, Beijing and T3-OSAKA; however, this methodology can be applied to the broad discrimination of *M. tuberculosis* clinical isolates.

CONCLUSION

TB-SGIP, a novel molecular typing method for *M. tuberculosis*, can differentiate between strains of *M. tuberculosis* with almost the same discriminatory power as spoligotyping. Moreover it is a simple, easy and quick method to discriminate Beijing and T3-OSAKA strains. The TB-SGIP assay is based on the presence

or absence of genomic islands, and may be able to predict strain characteristics if further study is able to clarify the functions of each island.

ACKNOWLEGEMENT

We thank Drs. Yoshiro Murase and Satoshi Mitarai for their critical review of this manuscript. This work was supported in part by a Grant-in-Aid for Scientific Research (C) from the Japan Society for the Promotion of Science for Y.H. and by a grant from the Ohyama Health Foundation for Y.H.

This study was supported by a grant-in-aid from the Ministry of Health, Labor and Welfare of Japan (H21-Shinkou-Ippan-008 and H24-Shinkou-Ippan-010).

REFERENCES

[1] World Health Organization: Global Tuberculosis Report 2012. http://apps.who.int/iris/bitstream/10665/75938/1/9789241564502_eng.pdf 2012.

[2] Gagneux S, Small PM. Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development. The Lancet Infectious Diseases 2007; 7(5): 328-337. http://dx.doi.org/10.1016/S1473-3099(07)70108-1

[3] Comas I, Chakravartti J, Small PM, Galagan J, Niemann S, Kremer K, *et al.* Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. Nature Genetics 2010; 42(6): 498-503. http://dx.doi.org/10.1038/ng.590

[4] van Crevel R, Nelwan RH, de Lenne W, Veeraragu Y, van der Zanden AG, Amin Z, *et al. Mycobacterium tuberculosis* Beijing genotype strains associated with febrile response to treatment. Emerging Infectious Diseases 2001; 7(5): 880-883. http://dx.doi.org/10.3201/eid0705.017518

[5] Kubica T, Rusch-Gerdes S, Niemann S. The Beijing genotype is emerging among multidrug-resistant *Mycobacterium tuberculosis* strains from Germany. The international journal of tuberculosis and lung disease. The Official Journal of the International Union against Tuberculosis and Lung Disease 2004; 8(9): 1107-1113.

[6] Mokrousov I, Jiao WW, Sun GZ, Liu JW, Valcheva V, Li M, *et al.* Evolution of drug resistance in different sublineages of *Mycobacterium tuberculosis* Beijing genotype. Antimicrobial Agents and Chemotherapy 2006; 50(8): 2820-2823. http://dx.doi.org/10.1128/AAC.00324-06

[7] van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, *et al.* Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. Journal of Clinical Microbiology 1993; 31(2): 406-409.

[8] Thierry D, Brisson-Noel A, Vincent-Levy-Frebault V, Nguyen S, Guesdon JL, Gicquel B. Characterization of a *Mycobacterium tuberculosis* insertion sequence, IS6110, and its application in diagnosis. Journal of Clinical Microbiology 1990; 28(12): 2668-2673.

[9] Frothingham R, Meeker-O'Connell WA. Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. Microbiology 1998; 144 (Pt 5): 1189-1196. http://dx.doi.org/10.1099/00221287-144-5-1189

[10] Mazars E, Lesjean S, Banuls AL, Gilbert M, Vincent V, Gicquel B, *et al.* High-resolution minisatellite-based typing as a portable approach to global analysis of *Mycobacterium tuberculosis* molecular epidemiology. Proceedings of the National Academy of Sciences of the United States of America 2001; 98(4): 1901-1906. http://dx.doi.org/10.1073/pnas.98.4.1901

[11] Hill V, Zozio T, Sadikalay S, Viegas S, Streit E, Kallenius G, *et al.* MLVA based classification of *Mycobacterium tuberculosis* complex lineages for a robust phylogeographic snapshot of its worldwide molecular diversity. PloS One 2012; 7(9): e41991. http://dx.doi.org/10.1371/journal.pone.0041991

[12] Kam KM, Yip CW, Tse LW, Wong KL, Lam TK, Kremer K, *et al.* Utility of mycobacterial interspersed repetitive unit typing for differentiating multidrug-resistant *Mycobacterium tuberculosis* isolates of the Beijing family. Journal of Clinical Microbiology 2005; 43(1): 306-313. http://dx.doi.org/10.1128/JCM.43.1.306-313.2005

[13] Surikova OV, Voitech DS, Kuzmicheva G, Tatkov SI, Mokrousov IV, Narvskaya OV, *et al.* Efficient differentiation of *Mycobacterium tuberculosis* strains of the W-Beijing family from Russia using highly polymorphic VNTR loci. European Journal of Epidemiology 2005; 20(11): 963-974. http://dx.doi.org/10.1007/s10654-005-3636-5

[14] Iwamoto T, Yoshida S, Suzuki K, Tomita M, Fujiyama R, Tanaka N, *et al.* Hypervariable loci that enhance the discriminatory ability of newly proposed 15-loci and 24-loci variable-number tandem repeat typing method on *Mycobacterium tuberculosis* strains predominated by the Beijing family. FEMS Microbiology Letters 2007; 270(1): 67-74. http://dx.doi.org/10.1111/j.1574-6968.2007.00658.x

[15] Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rusch-Gerdes S, Willery E, *et al.* Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. Journal of Clinical Microbiology 2006; 44(12): 4498-4510. http://dx.doi.org/10.1128/JCM.01392-06

[16] Supply P, Lesjean S, Savine E, Kremer K, van Soolingen D, Locht C. Automated high-throughput genotyping for study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial interspersed repetitive units. Journal of Clinical Microbiology 2001; 39(10): 3563-3571. http://dx.doi.org/10.1128/JCM.39.10.3563-3571.2001

[17] Maeda S, Murase Y, Mitarai S, Sugawara I, Kato S. [Rapid, simple genotyping method by the variable numbers of tandem repeats (VNTR) for *Mycobacterium tuberculosis* isolates in Japan--analytical procedure of JATA (12)-VNTR]. Kekkaku : [Tuberculosis] 2008; 83(10): 673-678.

[18] Murase Y, Mitarai S, Sugawara I, Kato S, Maeda S. Promising loci of variable numbers of tandem repeats for typing Beijing family *Mycobacterium tuberculosis*. Journal of Medical Microbiology 2008; 57(Pt 7): 873-880. http://dx.doi.org/10.1099/jmm.0.47564-0

[19] Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, *et al.* Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. Journal of Clinical Microbiology 1997; 35(4): 907-914.

[20] Millet J, Miyagi-Shiohira C, Yamane N, Sola C, Rastogi N. Assessment of mycobacterial interspersed repetitive unit-QUB markers to further discriminate the Beijing genotype in a population-based study of the genetic diversity of *Mycobacterium tuberculosis* clinical isolates from Okinawa, Ryukyu Islands, Japan. Journal of Clinical Microbiology 2007; 45(11): 3606-3615. http://dx.doi.org/10.1128/JCM.00348-07

[21] Abadia E, Zhang J, dos Vultos T, Ritacco V, Kremer K, Aktas E, *et al.* Resolving lineage assignation on *Mycobacterium tuberculosis* clinical isolates classified by spoligotyping with a new high-throughput 3R SNPs based method. Infection,

genetics and evolution. Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases 2010; 10(7): 1066-1074. http://dx.doi.org/10.1016/j.meegid.2010.07.006

[22] Brudey K, Driscoll JR, Rigouts L, Prodinger WM, Gori A, Al-Hajoj SA, *et al. Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. BMC Microbiology 2006; 6: 23. http://dx.doi.org/10.1186/1471-2180-6-23

[23] Takashima T, Iwamoto T. [New era in molecular epidemiology of tuberculosis in Japan]. Kekkaku: [Tuberculosis] 2006; 81(11): 693-707.

[24] Spol DB4. http://www.pasteur-guadeloupe.fr/tb/bd_myco.html.

[25] Suzuki M, Tawada Y, Kato M, Hori H, Mamiya N, Hayashi Y, *et al.* Development of a rapid strain differentiation method for methicillin-resistant *Staphylococcus aureus* isolated in Japan by detecting phage-derived open-reading frames. Journal of Applied Microbiology 2006; 101(4): 938-947. http://dx.doi.org/10.1111/j.1365-2672.2006.02932.x

[26] O'Sullivan MV, Kong F, Sintchenko V, Gilbert GL. Rapid identification of methicillin-resistant *Staphylococcus aureus* transmission in hospitals by use of phage-derived open reading frame typing enhanced by multiplex PCR and reverse line blot assay. Journal of Clinical Microbiology 2010; 48(8): 2741-2748. http://dx.doi.org/10.1128/JCM.02201-09

[27] Ross IL, Heuzenroeder MW. Discrimination within phenotypically closely related definitive types of *Salmonella enterica* serovar *typhimurium* by the multiple amplification of phage locus typing technique. Journal of Clinical Microbiology 2005; 43(4): 1604-1611. http://dx.doi.org/10.1128/JCM.43.4.1604-1611.2005

[28] Holden MT, Feil EJ, Lindsay JA, Peacock SJ, Day NP, Enright MC, *et al.* Complete genomes of two clinical *Staphylococcus aureus* strains: evidence for the rapid evolution of virulence and drug resistance. Proceedings of the National Academy of Sciences of the United States of America 2004; 101(26): 9786-9791. http://dx.doi.org/10.1073/pnas.0402521101

[29] Nada T, Yagi T, Ohkura T, Morishita Y, Baba H, Ohta M, *et al.* Usefulness of phage open-reading frame typing method in an epidemiological study of an outbreak of methicillin-resistant *Staphylococcus aureus* infections. Japanese Journal of Infectious Diseases 2009; 62(5): 386-389.

[30] O'Sullivan MV, Sintchenko V, Gilbert GL. Quantitative estimation of the stability of methicillin-resistant *Staphylococcus aureus* strain-typing systems by use of Kaplan-Meier survival analysis. Journal of Clinical Microbiology 2013; 51(1): 112-116. http://dx.doi.org/10.1128/JCM.01406-12

[31] Suzuki M, Matsumoto M, Takahashi M, Hayakawa Y, Minagawa H. Identification of the clonal complexes of *Staphylococcus aureus* strains by determination of the conservation patterns of small genomic islets. Journal of Applied Microbiology 2009; 107(4): 1367-1374. http://dx.doi.org/10.1111/j.1365-2672.2009.04321.x

[32] Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, *et al.* Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature 1998; 393(6685): 537-544. http://dx.doi.org/10.1038/31159

[33] Garnier T, Eiglmeier K, Camus JC, Medina N, Mansoor H, Pryor M, *et al.* The complete genome sequence of *Mycobacterium bovis*. Proceedings of the National Academy of Sciences of the United States of America 2003; 100(13): 7877-7882. http://dx.doi.org/10.1073/pnas.1130426100

[34] Roring S, Scott A, Brittain D, Walker I, Hewinson G, Neill S, *et al.* Development of variable-number tandem repeat typing of *Mycobacterium bovis*: comparison of results with those obtained by using existing exact tandem repeats and spoligotyping. Journal of Clinical Microbiology 2002; 40(6): 2126-2133. http://dx.doi.org/10.1128/JCM.40.6.2126-2133.2002

[35] Smittipat N, Billamas P, Palittapongarnpim M, Thong-On A, Temu MM, Thanakijcharoen P, *et al.* Polymorphism of variable-number tandem repeats at multiple loci in *Mycobacterium tuberculosis*. Journal of Clinical Microbiology 2005; 43(10): 5034-5043. http://dx.doi.org/10.1128/JCM.43.10.5034-5043.2005

[36] Le Fleche P, Fabre M, Denoeud F, Koeck JL, Vergnaud G. High resolution, on-line identification of strains from the *Mycobacterium tuberculosis* complex based on tandem repeat typing. BMC Microbiology 2002; 2: 37. http://dx.doi.org/10.1186/1471-2180-2-37

[37] Skuce RA, McCorry TP, McCarroll JF, Roring SM, Scott AN, Brittain D, *et al.* Discrimination of *Mycobacterium tuberculosis* complex bacteria using novel VNTR-PCR targets. Microbiology 2002; 148(Pt 2): 519-528.

[38] Matsumoto T, Koshii Y, Sakane K, Murakawa T, Hirayama Y, Yoshida H, *et al.* A novel approach to automated genotyping of *Mycobacterium tuberculosis* using a panel of 15 MIRU VNTRs. Journal of Microbiological Methods 2013; 93(3): 239-241. http://dx.doi.org/10.1016/j.mimet.2013.03.022

[39] Filliol I, Driscoll JR, Van Soolingen D, Kreiswirth BN, Kremer K, Valetudie G, *et al.* Global distribution of *Mycobacterium tuberculosis* spoligotypes. Emerging Infectious Diseases 2002; 8(11): 1347-1349. http://dx.doi.org/10.3201/eid0811.020125

[40] Maeda S, Wada T, Iwamoto T, Murase Y, Mitarai S, Sugawara I, *et al.* Beijing family *Mycobacterium tuberculosis* isolated from throughout Japan: phylogeny and genetic features. The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease 2010; 14(9): 1201-1204.

[41] Hunter PR. Reproducibility and indices of discriminatory power of microbial typing methods. Journal of Clinical Microbiology 1990; 28(9): 1903-1905.

[42] Takahashi M, Kazumi Y, Fukasawa Y, Hirano K, Mori T, Dale JW, *et al.* Restriction fragment length polymorphism analysis of epidemiologically related *Mycobacterium tuberculosis* isolates. Microbiology and Immunology 1993; 37(4): 289-294. http://dx.doi.org/10.1111/j.1348-0421.1993.tb03212.x

[43] Bifani P, Moghazeh S, Shopsin B, Driscoll J, Ravikovitch A, Kreiswirth BN. Molecular characterization of *Mycobacterium tuberculosis* H37Rv/Ra variants: distinguishing the mycobacterial laboratory strain. Journal of Clinical Microbiology 2000; 38(9): 3200-3204.

Received on 15-09-2014 Accepted on 24-09-2014 Published on 11-12-2014

DOI: http://dx.doi.org/10.14205/2310-9386.2014.02.02.2