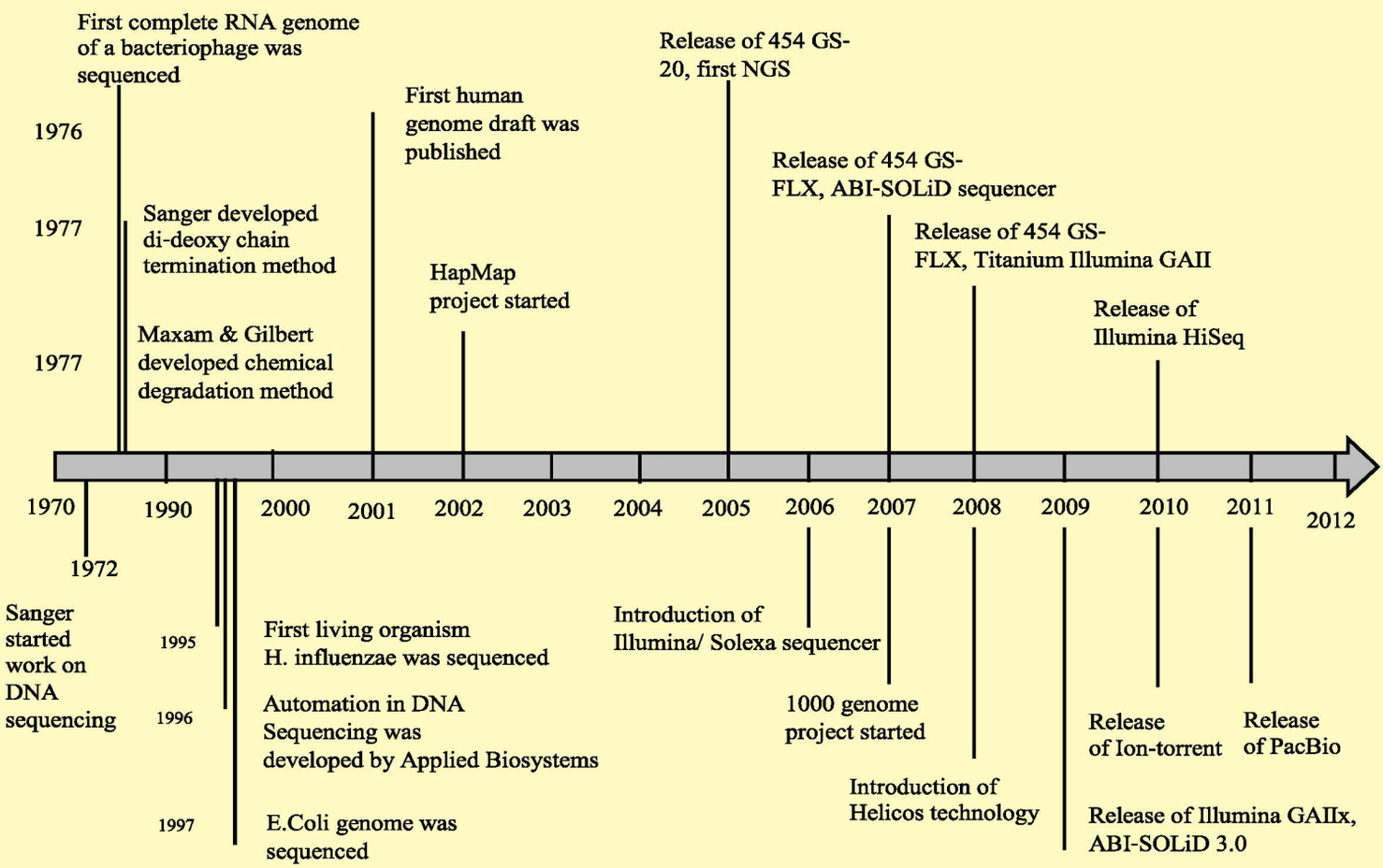
ISSN 2525-2461

Vietnam Journal of Science, Technology and Engineering

JUNE 2018 • VOL.60 NUMBER 2



The applications of massive parallel sequencing (next-generation sequencing) in research and molecular diagnosis of human genetic diseases

Vietnam Journal of Science, Technology and Engineering (VJSTE) **is a quarterly double-blind peer-reviewed academic journal. VJSTE publishes research articles and review articles in all areas of science, technology and engineering with the focus on the life sciences and environmental sciences including the following topics:**

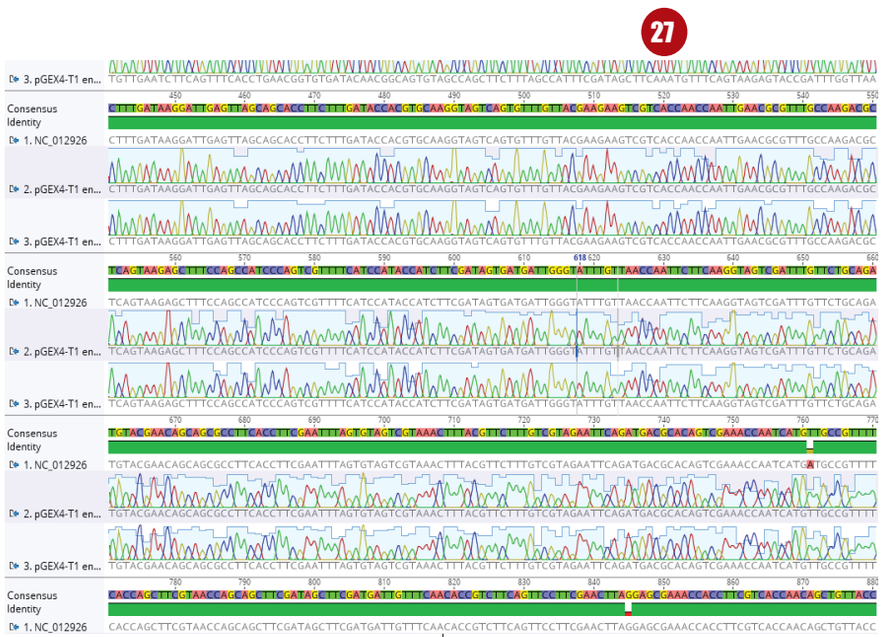
- **Agriculture**
- **Medicine**
- **Pharmacology**
- **Biology**
- **Biotechnology**
- **Ecology**
- **Climatology.**

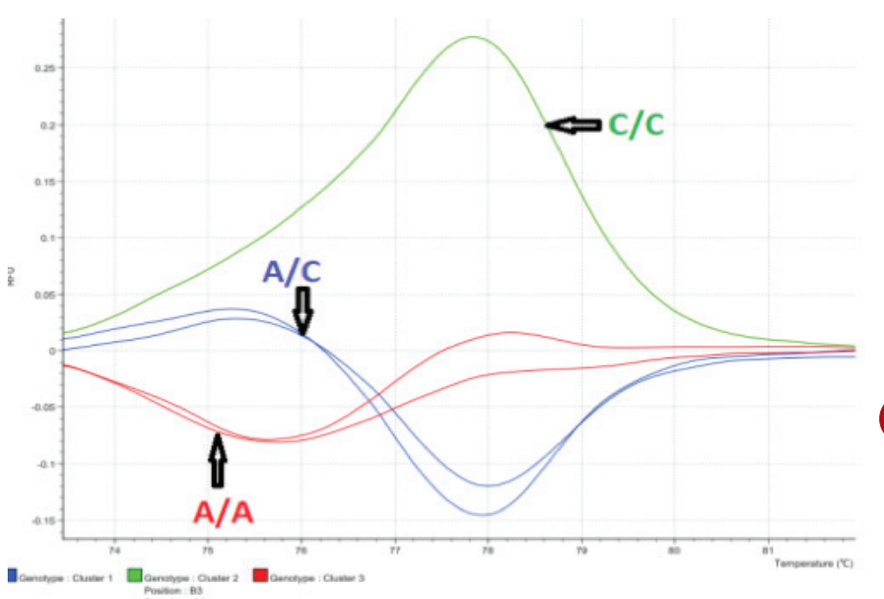
CLASSIFICATION SCHEME

1. **Mathematics and Computer Science**
   - 1.1. Mathematics
   - 1.2. Computer Science
   - 1.3. Computational Science
2. **Physical Sciences**
   - 2.1. Physics
   - 2.2. Chemistry
   - 2.3. Engineering
3. **Life Sciences**
   - 3.1. Agriculture
   - 3.2. Medicine
   - 3.3. Pharmacology
   - 3.4. Biology
   - 3.5. Biotechnology
4. **Geosciences**
   - 4.1. Geography
   - 4.2. Geology
   - 4.3. Geophysics
   - 4.4. Oceanography
5. **Nanoscience and Nanotechnology**
   - 5.1. Nanophysics
   - 5.2. Nanochemistry
   - 5.3. Nanobiotechnology
   - 5.4. Nanomedicine
   - 5.5. Nanoengineering
6. **Environmental Sciences**
   - 6.1. Ecology
   - 6.2. Climatology

CONTENTS

JUNE 2018 • VOL.60 NUMBER 2

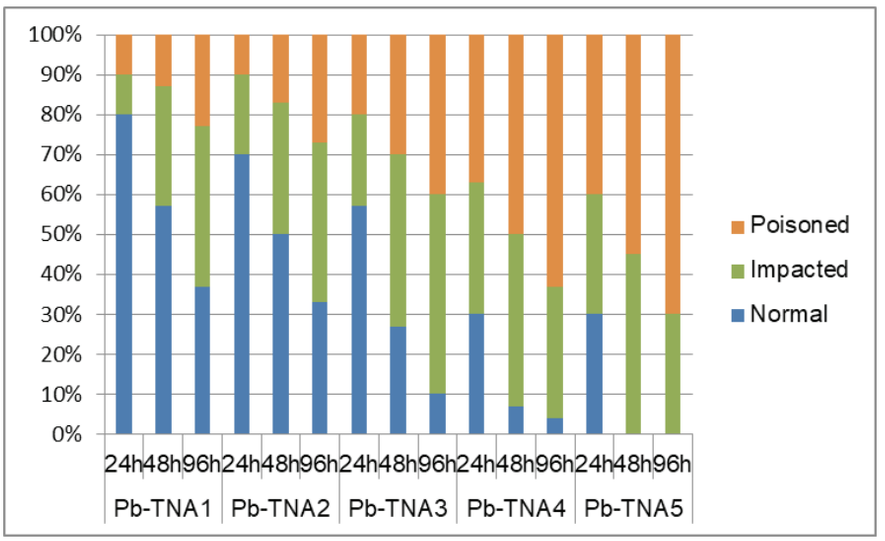




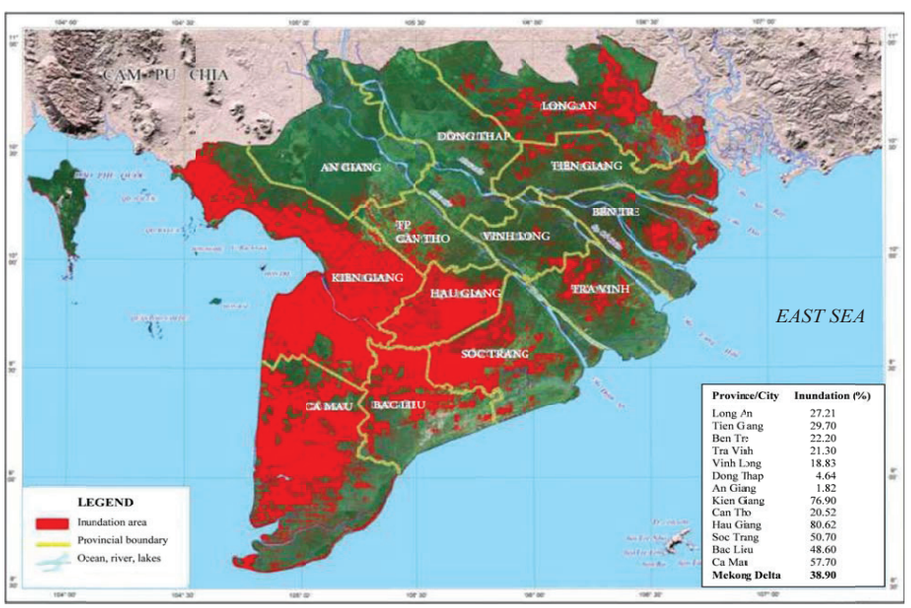
CONTENTS

JUNE 2018 • VOL.60 NUMBER 2

82



94



ENVIRONMENTAL SCIENCES

An efficient [Bmim]OH catalysed the condensations of aromatic aldehydes and diethyl malonate

Thi Thuy Duy Nguyen, Hoang Phuong Tran*

Department of Organic Chemistry, Faculty of Chemistry, University of Science, Vietnam National University, Ho Chi Minh city

Received 12 January 2018; accepted 10 April 2018

Abstract:

The rapid and green procedure for the Knoevenagel reaction between aromatic aldehydes and diethyl malonate was developed by using 1-butyl-3-methylimidazolium hydroxide, [Bmim]OH, as an efficient catalyst. The condensations showed good yields with a wide range of aromatic aldehydes in solvent-free condition and short reaction times. [Bmim]OH was easy to recover and could be reused several times without significant loss of catalytic activity.

Keywords: ***green catalyst, green chemistry, ionic liquids, Knoevenagel reaction.***

Classification number: ***2.2***

Introduction

Ionic liquids are ionic compounds that are liquid below 100°C. Their unique properties include thermal stability, biodegradability and non-volatility, and they have received special attention as environmentally benign solvents for organic synthesis. Ultimately, the combinations of organic cations and anions help to design and fine-tune their physical and chemical properties [1-4]. Due to the requirement in developing the environmentally benign processes, ionic liquids have been studied intensively as green solvents in diverse fields such as chemical reactions, electrochemistry and biochemistry [3, 5-9].

The Knoevenagel condensation reaction, which occurs between aldehydes or ketones and active methylene compounds, is a classic method for carbon-carbon bond formation [2]. The α,β-unsaturated products obtained by this method has been widely used as intermediates in organic synthesis and have been found to have increasing applications in medicine, biological science, agriculture and light-emitting materials [10]. Unfortunately, this method needs to be carried out in organic solvents and requires the Lewis base catalysts such as piperidine, pyridine, NaOH, K_2CO_3, etc., which are not recovered and reused [11-14]. Recently, the Knoevenagel reaction using ionic liquids as catalysts has been studied extensively. The ionic liquid was one of the most efficient catalytic systems that were easily recovered and reused several times [15].

In this research, we reported the Knoevenagel condensations using a basic ionic liquid, [Bmim]OH. In terms of green chemistry in organic synthesis, we synthesised [Bmim]OH via the metathesis reaction and used it as an efficient catalyst for the condensations of aromatic aldehydes and diethyl malonate under solvent-free condition. The basic catalyst could be recycled easily by the extraction and removal of water.

Experimental section

All chemicals were purchased from Sigma-Aldrich. The solvents were supplied by Xilong Chemical. The Agilent 7890A GC-MS system was equipped with a mass selective detector Agilent 5975N and a capillary column HP-5MS (length = 30 m, inner diameter = 320 µm, film thickness = 0.25 µm). The ^{1}H NMR spectra were recorded on a Bruker 500 MHz using $CDCl_3$ as solvent (Fig. 1).

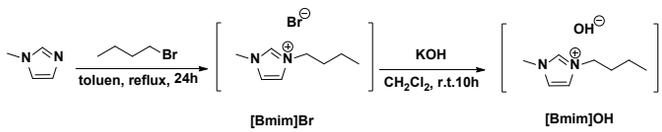


Fig. 1. Procedure for the synthesis of [Bmim]OH.

A typical procedure for the synthesis of [Bmim]Br: a round-bottomed flask (100 mL volume) was charged with a mixture of 1-methylimidazole (4.10 g, 5 mmol) and 1-bromobutane (6.85 g, 5 mmol) in toluene. Then, it was placed in an oil bath and heated at 110°C in a magnetic stirrer for 12 h. After completion the reaction, the reaction mixture was cooled down to room temperature and then washed with diethyl ether (3 x 10 mL) to obtain [Bmim]Br

*Corresponding author: Email: thphuong@hcmus.edu.vn

[16]. The purity and authenticity of the ionic liquids were confirmed by ^{1}H NMR spectroscopy.

1-butyl-3-methylimidazolium bromide [16]: a yellowish syrup. IR: ν (cm^{-1}): 3098, 2962, 2874, 1631, 1571, 1382, 1168, 1109, 760, 649. ^{1}H NMR (500 MHz, $CDCl_3$) δ 10.29 (s, 1H), 7.53 (t, *J* = 2 Hz, 1H), 7.42 (t, *J* = 2 Hz, 1H), 4.30 (t, *J* = 7.5 Hz, 2H), 4.09 (s, 3H), 1.88 (m, 2H), 1.396-1.321 (sext, *J* = 7.5 Hz, 2H), 0.93 (t, *J* = 7.5 Hz, 3H).

A typical procedure for the synthesis of [Bmim]OH: a mixture of KOH (1.2 g, 20 mmol) and [Bmim]Br (4.4 g, 20 mmol) in dry CH_2Cl_2 (20 mL) was stirred at room temperature for 10 h. After the completion of the reaction, the precipitated KBr was filtered off, and the filtrate was evaporated under vacuum to obtain the [Bmim]OH as a viscous liquid. The desired ionic liquid was washed with diethyl ether (3 × 20 mL) and dried at 90°C for 10 h to obtain the pure ionic liquid for use [17].

1-butyl-3-methylimidazolium hydroxide [17]: a honey-coloured syrup. IR: ν (cm^{-1}): 3452 (O-H), 3150, 3099, 2962, 2906 (C-H), 1651 (C=N); 1247, 1114, 860. ^{1}H NMR (500 MHz, $CDCl_3$) δ 10.28 (s, 1H), 7.51 (s, 1H), 7.39 (s, 1H), 4.31 (t, *J* = 7.5 Hz, 2H), 4.09 (s, 3H), 2.32 (bs, 1H), 1.91-1.85 (m, 2H), 1.37 (dt, *J* = 15, 7.5 Hz, 2H), 0.94 (t, *J* = 7.5 Hz, 3H).

A typical procedure for the condensation of benzaldehyde and diethyl malonate: benzaldehyde (1 mmol), diethyl malonate (1 mmol) and [Bmim]OH (0.3 mmol) were reacted at 100°C for 2 h. After cooling, the reaction mixture was extracted with diethyl ether (3 x 20 mL). Since [Bmim]OH is insoluble in diethyl ether, it is easily separated from the mixture. Then, it was washed in water, which was removed by evaporation. For the ether layer, it was decanted, washed with water and dried over Na_2SO_4. The solvent was then removed by a rotary evaporator. The desired product was purified by column chromatography. The purity and authenticity of the product was confirmed by GC-MS and ^{1}H NMR spectroscopy.

Diethyl benzenlidenemalonate [8]: GC-MS (EI, 80 eV) *m/z* 248 (M^+). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.73 (s, 1H), 7.46-7.44 (m, 2H), 7.38-7.36 (m, 3H), 4.35-4.28 (m, 4H), 1.33 (t, *J* = 7 Hz, 3H), 1.28 (t, *J* = 7 Hz, 3H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 166.7, 164.1, 142.1, 132.9, 130.5, 129.5, 128.8, 126.4, 61.7, 61.6, 14.1, 13.8.

Results and discussion

We aim to synthesise a 'green' catalyst for the Knoevenagel condensation as well as suggest an efficient and green process for the synthesis of some functionalised alkenes that are widely used as intermediates in organic synthesis.

Preparation and characterisation of catalyst: [Bmim]Br was synthesised from 1-methylimidazole and 1-bromobutane. Then, [Bmim]OH was obtained by the metathesis reaction of [Bmim]Br with KOH in dry CH_2Cl_2. [Bmim]OH was obtained in high yield and it was applied as a green basic catalyst for the Knoevenagel condensation. [Bmim]OH was characterised by FT-IR and H^1 NMR spectroscopy. In the FT-IR spectra of [Bmim]OH, the peak at 3452 cm^{-1} is a characteristic of the stretching vibration of -OH. The peaks at 2900~3000 cm^{-1} are assigned as the saturated C-H stretching vibrations. The peaks at 1651 cm^{-1} could be assigned to the stretching vibration of C=N.

Catalytic testing: the Knoevenagel condensation of benzaldehyde and dimethyl malonate is used as a model reaction to investigate reaction conditions in which the functionalised alkene was obtained in the highest yield and purity (Fig. 2). All reactions were carried out in an IKA RET BASIC (USA) heating magnetic stirrer, which is equipped with an electronic temperature controller.

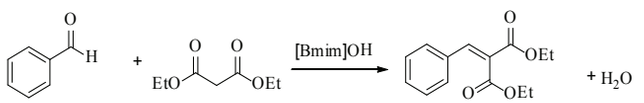


Fig. 2. Knoevenagel condensation of benzaldehyde and dimethyl malonate.

Table 1. The effect of temperature on reaction under magnetic stirring.

Entry	Temperature (°C)	Yield (%)
1	r.t	13
2	50	54
3	80	76
4	**100**	**95**
5	120	99

The condensation between aldehyde and activated methylene is preferred at high temperature (Table 1). The temperature of 100°C was chosen as the optimal temperature. When the temperature was 120°C, the yield of the reaction slightly increased (Table 1, entries 4-5).

Table 2. The effect of the reaction time.

Entry	Time (h)	GC Yield (%)
1	0.5	43
2	1.0	61
3	1.5	84
4	2.0	95

As the reaction time prolonged, there was a significant increase in the conversion of benzaldehyde. The desired product was obtained in quantity (95%) for 2 h (Table 2, entry 4).

Table 3. The effect of the amount of catalyst.

Entry	Catalyst (mol%)	Yield (%)
1	0	0
2	5	29
3	20	78
4	**30**	**91**
5	40	92
6	50	95

The effect of the catalyst was investigated on the Knoevenagel condensation under the current method. The optimal amount of the catalyst was only 0.3 equiv. (Table 3).

As a result, the optimal condition was using 0.3 equiv of [Bmim]OH and magnetic stirring at 100°C for 2 h. Additionally, we carried out the control experiments (Fig. 3). No desired product was obtained in the absence of [Bmim]OH. In addition, [Bmim]Br was not reactive in the current method because [Bmim]Br is a neutral ionic liquid.

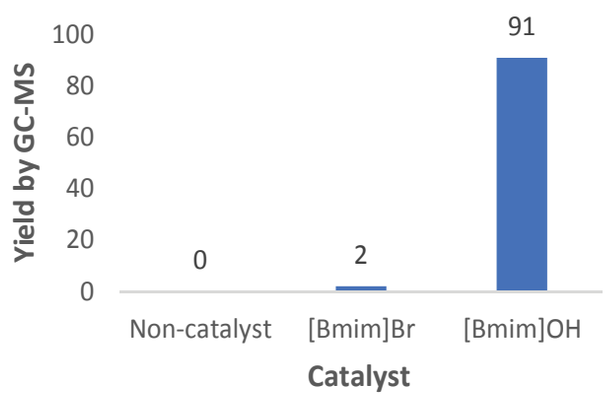
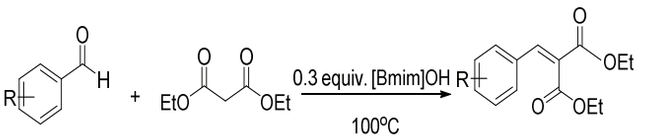
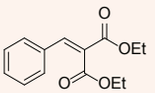
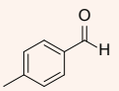
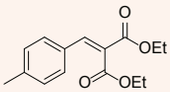
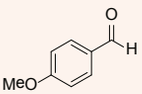
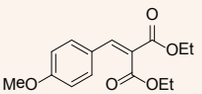
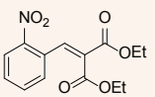
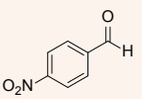
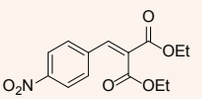


Fig. 3. Control experiments.

We applied the optimal conditions for the Knoevenagel condensation between some benzaldehyde derivatives and diethyl malonate to study the effects of the structural effect of benzaldehyde derivatives. The results are shown in Table 4.

Table 4. The Knoevenagel reaction of different substrates using [Bmim]OH under solvent-free.



Entry	Substrate	Time (h)	Product	Isolated yield (%)
1		2		90
2		3		83
3		3		87
4		3		67
5		3		75

It is evident that aromatic aldehyde with an electron-donating group such as 4-methylbenzaldehyde and 4-methoxybenzaldehyde was less reactive than benzaldehyde, by 83% and 77%, respectively (Table 4, entries 2, 3). The desired products were obtained in lower yields with nitrobenzaldehydes because these substrates were less soluble in the reaction mixture (Table 4, entries 4, 5). In addition, 2-nitrobenzaldehyde reacted at lower conversion (Table 4, entry 4) due to the increased steric hindrance of *ortho*-substitution.

The [Bmim]OH catalyst was easily recovered and reused without any loss of catalytic activity. Since [Bmim]OH is insoluble in diethyl ether, it is easily separated from the mixture. Then, it was washed and dried under vacuum.

The recovered [Bmim]OH was reactive in the Knoevenagel condensation of benzaldehyde, and dimethyl malonate produced the desired product in 89% isolated yield.

Conclusions

To conclude, the basic ionic liquid [Bmim]OH is an efficient catalyst for the Knoevenagel condensation. The ionic liquid is known as green solvent/catalyst for many organic transformations. These condensations of aromatic aldehydes and diethyl malonate were efficiently catalysed by a small amount of [Bmim]OH under a solvent-free condition with high reaction yields in short reaction times.

ACKNOWLEDGEMENTS

This research is funded by Vietnam National University, Ho Chi Minh city (VNU-HCM) under grant number 562-2018-18-03.

REFERENCES

[1] P. Moriel, E.J. García-Suárez, M. Martínez, A.B. García, M.A. Montes-Morán, V. Calvino-Casilda, M.A. Banares (2010), "Synthesis, characterization, and catalytic activity of ionic liquids based on biosources", *Tetrahedron Lett.*, **51(73)**, pp.4877-4881.

[2] M.B. Smith, J. March (2001), *March's advanced organic chemistry: reactions, mechanisms, and structure*, John Wiley & Sons.

[3] T. Welton (1999), "Room-temperature ionic liquids. Solvents for synthesis and catalysis", *Chem. Rev.*, **99(8)**, pp.2071-2083.

[4] P.H. Tran, X.T.T. Nguyen, D.K.N. Chau (2018), "Brønsted-acidic ionic liquid gel as an efficient and recyclable heterogeneous catalyst for the synthesis of bis(indolyl)methanes under solvent-free sonication", *Asian J. Org. Chem*., **7**, pp.232-239.

[5] J. Safari, Z. Zarnegar (2013), "Immobilized ionic liquid on superparamagnetic nanoparticles as an effective catalyst for the synthesis of tetrasubstituted imidazoles under solvent-free conditions and microwave irradiation", *Comptes Rendus Chim.*, **16(10),** pp.920-928.

[6] S.P. Borikar, T. Daniel (2011), "A convenient and efficient protocol for the synthesis of acylals catalyzed by Bronsted acidic ionic liquids under ultrasonic irradiation", *Ultrason. Sonochem.*, **18(5),** pp.928-931.

[7] J.R. Harjani, S.J. Nara, M.M. Salunkhe (2002), "Lewis acidic ionic liquids for the synthesis of electrophilic alkenes via the Knoevenagel condensation", *Tetrahedron Lett.*, **43(6),** pp.1127-1130.

[8] G.H. Gao, L.U. Liang, et al. (2007), "Basic ionic liquid: a reusable catalyst for knoevenagel condensation in aqueous media", *Chemical Research in Chinese Universities*, **23(2),** pp.169-172.

[9] A.R. Hajipour (2009), "Basic ionic liquids. A short review", *Rev. Lit. Arts Am.*, **6(4),** pp.647-678.

[10] M. Zahouily, M. Salah, B. Bahlaouane, A. Rayadh (2004), "Solid catalysts for the production of fine chemicals: the use of natural phosphate alone and doped base catalysts for the synthesis of unsaturated arylsulfones", *Tetrahedron*, **60(7),** pp.1631-1635.

[11] Yi-Qun Li (2000), "Potassium phosphate as a catalyst for the Knoevenagel condensation", *Journal of Chemical Research*, **2000(11)**, pp.524-525.

[12] S. Zhao, X. Wang, L. Zhang (2013), "Rapid and efficient Knoevenagel condensation catalyzed by a novel protic ionic liquid under ultrasonic irradiation", *RSC Adv.*, **3(29),** pp.11691-11696.

[13] M.R. Poor Heravi, S. Piri (2012), "Efficient Knoevenagel condensation reactions catalyzed by activated gel zirconium(IV) oxide", *J. Chem.*, **2013**, pp.1-5.

[14] A. McCluskey, P.J. Robinson, T. Hill, J.L. Scott, J.K. Edwards (2002), "Green chemistry approaches to the Knoevenagel condensation: comparison of ethanol, water and solvent free (dry grind) approaches", *Tetrahedron Lett.*, **43(17)**, pp.3117-3120.

[15] D.C. Forbes, A.M. Law, D.W. Morrison (2006), "The Knoevenagel reaction: analysis and recycling of the ionic liquid medium", *Tetrahedron Lett.*, **47(11)**, pp.1699-1703.

[16] P.N. Tshibangu, S.N. Ndwandwe, E.D. Dikio (2011), "Density, viscosity and conductivity study of 1-butyl-3-methylimidazolium bromide", *Int. J. Electrochem. Sci*, **6(6),** pp.2201-2213.

[17] B.C. Ranu, S. Banerjee (2005), "Ionic liquid as catalyst and reaction medium. The dramatic influence of a task-specific ionic liquid, [Bmim]OH, in Michael addition of active methylene compounds to conjugated ketones, carboxylic esters, and nitriles", *Org. Lett.*, **7(14),** pp.3049-3052.

Engineering properties of unfired building bricks produced using URHA-FA cement blends

Si Huy Ngo[1*], Trong Phuoc Huynh[2]

[1]*Department of Engineering and Technology, Hong Duc University*
[2]*Department of Rural Technology, College of Rural Development, Can Tho University*

Received 3 July 2017; accepted 30 November 2017

Abstract:

The production of cement and traditional fired clay bricks consumes intensive energy and inversely affects the environment. In addition, a huge quantity of solid waste materials such as rice husk ash and fly ash (FA) are generated from both industrial and agricultural activities. This study investigates the use of unground rice husk ash (URHA) and FA for manufacturing unfired building bricks. FA was used as a cement substitute (15%, 30%, and 50%), whereas URHA was used as a chippings replacement (5%, 10%, and 15%) in the brick mixtures. Test results indicate that all of the brick samples had consistent dimensions and were free of visible defects. Generally, increasing the URHA or FA replacement levels reduced the strength, bulk density, and material cost. However, it increased the water absorption capacity of the brick samples. Moreover, bricks with 10% URHA and 50% FA registered the lowest cost. Properties of all of the brick samples met the Grade M15 requirements of the TCVN 6477-2011 standard of high-quality unfired building bricks.

Keywords: ***compressive strength, cost analysis, fly ash, unfired building brick, unground rice husk ash.***

Classification number: ***2.3***

Introduction

Due to the rapid development of the construction industry in Vietnam, building bricks are consumed profusely annually. Most of them are conventional fired clay bricks or concrete bricks. Conventional fired bricks are produced from clay at high temperature, while concrete bricks are produced from ordinary Portland cement. To produce the conventional fired clay bricks, a significant energy and intensive amount of natural clay is used, leading to a negative effect on the environment due to the generation of carbon dioxide (CO_2) and the depletion of agricultural land. On the other hand, the production of cement also consumes intensive energy and releases a significant quantity of CO_2 into the air, causing greenhouse effect and contributing to climate change. Furthermore, the mass of industrial wastes is rapidly increasing and has an inverse effect on the environment. Therefore, instead of considering them as waste materials, turning such wastes into green construction materials has received much attention from researchers.

Vietnam is predominantly an agricultural country and falls within the top rice export nations in the world. In consequence, a large amount of rice husk was generated as a by-product of rice production. A part of the rice husk was used to produce animal food and fertilizer, while the rest was utilized as fuel in rural households and small businesses because of its cheap price. Rice husk ash is obtained from burning rice husk. It is worth noting that the properties of rice husk ash strongly depend on the burning conditions. When rice husk is burned at temperatures ranging between 600 and 800°C, rice husk ash consists of around 91-95% reactive silica (SiO_2) [1, 2]. Hence, it can be used for manufacturing unfired building bricks. Moreover, fly ash (FA), a by-product of coal power plants, is widely employed as a supplementary cementitious material in order to reduce the amount of cement produced. The use of FA and rice husk

*Corresponding author: Email: ngosihuy@hdu.edu.vn

ash in unfired building bricks is a visible solution to the environmental problem as well as economical effectiveness.

FA is extensively used for producing unfired building bricks. However, the properties of unfired building bricks are significantly dependent on FA content and its quality as well as forming pressure. The use of FA as the main binder in unfired building bricks was examined in some previous studies [3-7]. Under forming pressure of 10-26 MPa, unfired building bricks exhibited a compressive strength of higher than 13 MPa and water absorption of lower than 20% [3-5]. Cicek and Tanriverdi (2007) [6] investigated the use of 50-80% FA in the total amount of brick and under forming pressures varying between 0.5 and 30 MPa. Test results showed that unfired building bricks had a compressive strength of lower than 10 MPa and water absorption higher than 33%. Using 60-90% FA and forming by vibration table, unfired building bricks showed low properties with a compressive strength lower than 8 MPa and water absorption between 29-37% [7]. It was found that using the vibration table to form the sample was not as effective as using high forming pressure. Shakir, et al. (2013) studied the use of a combination of cement and FA as binder materials in unfired building bricks [8]. The compressive strength and water absorption of bricks ranged between 6.2-26.3 MPa and 12.9-19.1%, respectively. In order to increase the pozzolanic reaction of FA, the alkali-activators were added into the brick mixtures [9-11]. Kumar, et al. (2013) studied the use of 60-100% FA and 0-40% red mud in unfired building bricks [9]. These bricks exhibited good performance with a compressive strength higher than 16 MPa and water absorption lower than 7%. The combination of FA and bottom ash were investigated by Freidin (2017) [10] and Arioz, et al. (2010) [11]. Under forming pressures of 4 MPa and 30 MPa, unfired building bricks showed compressive strength up to 20 MPa and 60 MPa, respectively.

Recently, rice husk ash has been used for producing unfired building bricks [12-16]. Unfired building bricks were made from FA, rice husk ash, and sand using geopolymerization technology [12-14]. Other unfired building bricks were made from cement, FA, and rice husk ash based on the cementing reaction [15-16]. It was noted that rice husk ash was used in the following two kinds: ground rice husk ash and unground rice husk ash (URHA). In these studies, URHA was considered as fine aggregate to replace 10-40% amount of sand and ground rice husk ash was used as a supplementary cementitious material. All of the brick samples were formed under the high pressure of 35 MPa. Experimental results revealed that all unfired building bricks show a good performance with properties satisfying the requirements of TCVN 6477-2011 [17].

Compared with TCVN 6477-2011 [17], the unfired building bricks from previous studies had a water absorption much higher than 14%, over the requirements of the Vietnamese standard. All the previous studies selected FA with high quality and the loss on ignition lower than 6% as required by ASTM C618 [18]. URHA was applied to replace a part of sand, and unfired building bricks were manufactured under high forming pressure. The primary objective of this study is to investigate the use of low quality raw FA, with a high loss on ignition and URHA, in the production of unfired building bricks. Raw FA and URHA were used to replace part of the cement and chippings, respectively. The FA used herein has a 15.8% loss on ignition that is much higher than the requirement of ASTM C618 [18]. Unfired building bricks were produced under low forming pressure of around 5 MPa. The effects of FA and URHA content on the properties of the unfired building bricks such as compressive strength, water absorption, and bulk density were also investigated in accordance with TCVN 6477-2011 [17]. Moreover, cost analysis was conducted to find out the optimal brick mixture.

Materials and experimental programs

Materials

Unfired building bricks were prepared from cement, FA, chippings, and URHA, where cement and FA were used as binder materials, with properties as shown in Table 1, while chippings and URHA were used as fine aggregates. In this study, ordinary Portland cement Nghi Son PC40, with a specific gravity of 3.12, was used. FA, a raw material sourced from the Nghi Son coal power plant that was classified as class-F based on ASTM C618 [18], with a specific gravity of 2.16 and a 15.8% loss on ignition was used as a cement substitute. Chippings was a byproduct from the stone crushing process with a maximum size of 5 mm, density of 2.65 T/m^3, fineness modulus of 3.54, and moisture content of 0.5%. URHA, with a density of 2.10 T/m^3, fineness modulus of 2.58, and water absorption of 30%, was taken from the steam boiler at Nghi Son industrial zone. Gradation curves of chippings and URHA are presented in Fig. 1. Fig. 2 shows the images of chippings, URHA, and scanning electron micrograph (SEM) of URHA.

Table 1. Properties of cement and FA.

Items		Cement	FA
Physical properties	Specific gravity	3.12	2.16
	Loss on ignition (%)	1.9	15.8
Chemical composition (wt.%)	SiO_2	22.4	48.4
	Al_2O_3	5.3	20.4
	Fe_2O_3	4.0	4.8
	CaO	55.9	2.8
	MgO	2.8	1.4
	Others	4.5	4.3

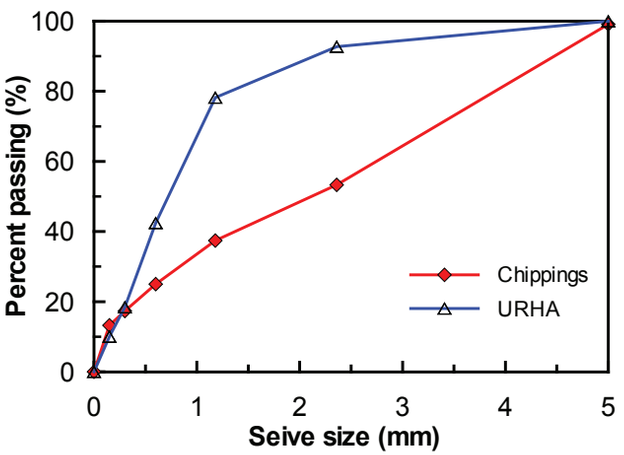


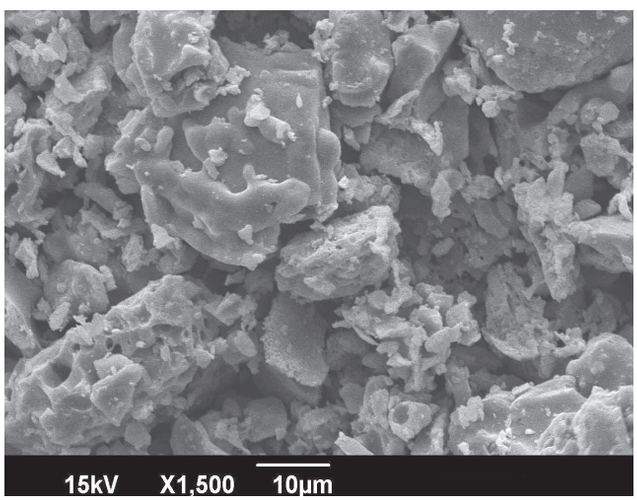
Fig. 1. Gradation curves of chippings and URHA.

Brick mixtures

The brick mixtures were divided into two groups as shown in Tables 2 and 3. Table 2 shows the first group that was designed to investigate the effect of URHA content on the properties of the unfired building bricks. In this group, URHA was used to replace 5%, 10% and 15% of chippings. Table 3 shows the second group that was designed to investigate the effect of FA content on properties of the unfired building bricks. A constant amount of 10% URHA was used for all mixtures in this group. FA was used to replace 15%, 30%, and 50% of the cement. The nomenclature of the mixtures is described as follows: M5 and M6 denote the water-to-binder ratios of 0.5 and 0.6, respectively; the numbers after them (0, 5, 10, and 15) are the percentages of URHA replacement for chippings; the numbers in front of FA (15, 30, and 50) indicate the percentages of FA replacement for cement.

(A)

(B)



(C)

Fig. 2. (A) Chippings, (B) URHA, (C) SEM image of URHA.

Table 2. First group mixture proportions.

Mixture	Ingredient proportions (kg/m^3)			
	Cement	*URHA*	*Chippings*	*Water*
M5-0	440	1693	0	220
M5-5	436	1595	84	218
M5-10	433	1499	167	216
M5-15	429	1404	248	215
M6-0	367	1756	0	220
M6-5	364	1653	87	218
M6-10	360	1553	173	216
M6-15	357	1454	257	214

Table 3. Second group mixture proportions.

Mixture	Ingredient proportions (kg/m^3)				
	Cement	*FA*	*URHA*	*Chippings*	*Water*
M5-10-15FA	364	64	1485	165	214
M5-10-30FA	297	127	1472	164	212
M5-10-50FA	210	210	1454	162	210
M6-10-15FA	304	54	1541	171	215
M6-10-30FA	248	106	1530	170	213
M6-10-50FA	176	176	1514	168	211

Samples preparation and test programs

Unfired building bricks were prepared in a steel mold, with dimensions of 220×105×65 mm, applying forming pressure of around 5 MPa that is much lower than the forming pressures used in most of the previous studies (10-35 MPa) [3-6, 11-16]. The purpose of this study is to assess the use of low forming pressure and industrial and agricultural by-products for producing unfired building bricks.

The dimensions and visible defects, compressive strength, water absorption, and bulk density of the unfired building brick samples were tested in accordance with TCVN 6477-2011 [17]. The compressive strength values were measured at the 3, 7, 14 and 28-day ages, with the presented values were the average values of the three samples. Other brick properties were measured at the 28-day ages.

Test results and discussion

Dimensions and visible defects

The measured dimensions of the unfired building bricks are shown in Table 4. All the bricks possessed a slight difference in dimensions compared with the standard size (220×105×65 mm). This is due to the deformation of the steel mold under repeated forming pressure during the production of bricks. The detected error is lower than the allowable error stipulated by TCVN 6477-2011 [17]. Table 5 shows the visible defects of the brick samples. No visible defects were observed on the surface, edge, and corner of the samples, indicating that all the unfired building bricks had consistent shape, and satisfied the TCVN 6477-2011 requirements [17].

Table 4. Dimensions of brick samples.

Dimension	Measured dimension (mm)	Allowable error (mm)
Width	105 ± 1	± 2
Length	220 ± 1	± 2
Height	65 ± 1	± 3

Table 5. Visible defects of brick samples.

Type of visible defects	Allowable level	Visible defects of brick samples
The curvature of the surface of brick (mm), no more than	3	No
The number of edges and corner cracks with the depth of 5±10 mm and the length of 10±15 mm, no more than	4	No
The number of cracks through the thickness pulling to a width that not exceeding 20 mm, no more than	1	No

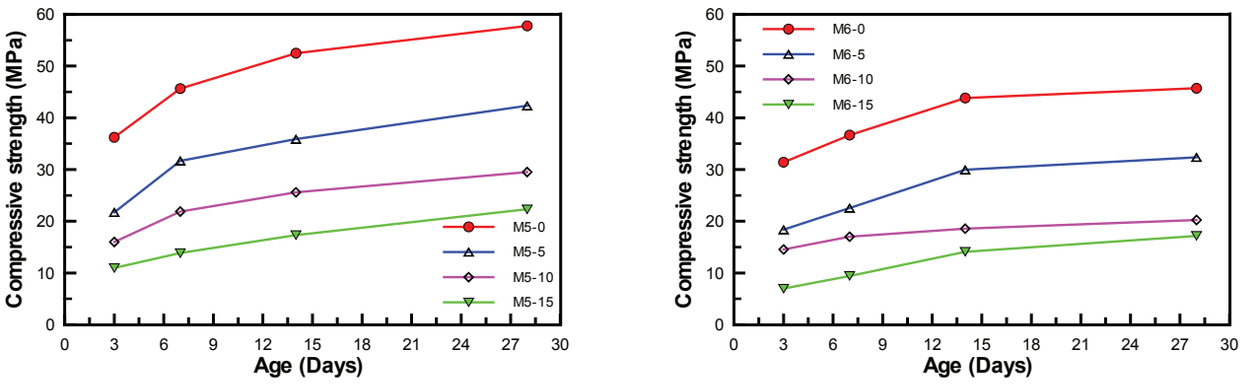


Fig. 3. Compressive strength development of the unfired building brick samples with various URHA contents.

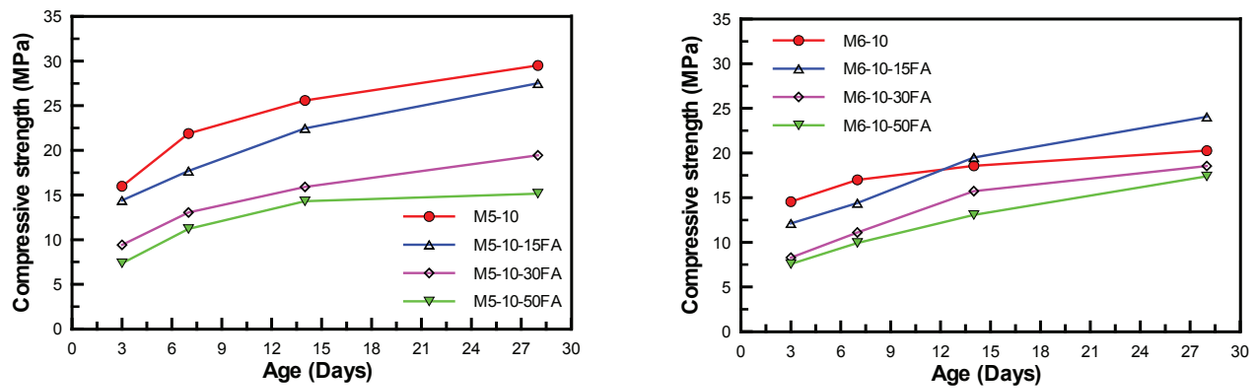


Fig. 4. Compressive strength development of the unfired building brick samples with various FA contents.

Compressive strength

The effect of URHA content on compressive strength development of brick samples is shown in Fig. 3. The compressive strength of bricks significantly reduced as the URHA content increased. For brick mixtures with water-to-binder ratios of 0.5 and 0.6, the compressive strength of the 28-day-old brick samples with 5%, 10%, and 15% URHA were, respectively, about 26.7%, 48.9%, and 61.4% and about 29.2%, 55.7%, and 62.4% lower than that of the control samples without URHA. It could be observed from Fig. 2C that URHA was made of highly porous particles that caused an inverse effect on the compressive strength of brick samples. Increasing URHA replacement levels resulted in the loss of structural compactness and, in turn, led to a lower compressive strength. However, all the brick samples incorporating URHA possessed the 28-day compressive strength that was higher than 17 MPa. Thus, these brick samples could be classified as the high-quality unfired building bricks (Grade M15) in accordance with TCVN 6477-2011 [17].

Figure 4 shows the compressive strength development of the unfired building brick samples with varied FA content. After 28 days of age, the unfired building brick samples of the M5 group with 15%, 30%, and 50% FA replacement levels had compressive strength values of 27.5, 19.4, and 15.2 MPa, respectively. These values were 6.8%, 34.1%, and 48.6% lower than the compressive strength values of the FA-free samples (M5-10), respectively. The compressive strength of the M5 mixtures decreased as the FA replacement levels increased. However, for the M6 mixtures, the brick sample with 15% FA showed the highest compressive strength at the 28-day ages (Fig. 4B). Previous studies [19-21] have proved that the use of FA at optimal

dosage enhanced the compressive strength of concrete because of the pozzolanic reaction. The optimal dosage of FA was varied, depending on its properties and mixture proportion. If FA was added over and above the optimal dosage, all of it did not participate in a chemical reaction; it acted as the fine aggregate rather than a cementitious material. In this case, the amount of FA in the M6-10-15FA mixture may be close to the optimal FA content, resulting in a higher compressive strength than control mixture M6-10. It is also worth noting that the FA used in this research was of low quality with a high loss on ignition. Thus, the compressive strength of all the brick samples reduced when FA content increased, except the M6-10-15FA mixture as mentioned above. However, similar to the first mixture group, the lowest compressive strength value among M6 brick samples was 15.2 MPa that satisfied the Grade M15 of the TCVN 6477-2011 [17].

As can be seen from to Figs. 3 and 4, the compressive strengths of mixtures with a water-to-binder ratio of 0.5 were higher than those of corresponding mixtures with a water-to-binder ratio of 0.6. This is due to the fact that the amount of binder in M5 mixtures is higher than that of M6 mixtures (Tables 2 and 3), resulting in more hydration products. Consequently, the compactness and strength capacity of bricks were enhanced because hydration products were the main carriers of strength in unfired building bricks. Therefore, the compressive strength increased since water-to-binder ratio reduced.

Water absorption

Water absorption is an important property of unfired building bricks, which significantly affects the progress and quality of construction. Bricks with high water absorption capacity will absorb a higher amount of water from mortar, affecting the bond between bricks and mortar. Therefore, the TCVN 6477-2011 [17] has limited the maximum level of water absorption of 14%. Fig. 5A shows the relationship between water absorption and URHA content. The water absorption of bricks increased with URHA content. For M5 mixtures, the unfired building bricks with the URHA replacement levels of 5%, 10%, and 15% had the water absorption levels of 52.2%, 60.7%, and 86.1%, respectively, higher than that of the control mixtures without URHA. A similar trend was observed among the M6 group bricks with 5%, 10%, and 15% URHA content that had water absorption values of 41.9%, 58.4%, and 115%, respectively, higher than that of the bricks without URHA content. This phenomenon is because of the high porosity of the URHA as mentioned above. However, all the brick samples produced in this study had water absorption values lower than 14% in accordance with TCVN 6477-2011 requirements [17].

Figure 5B shows the relationship between water absorption and FA content. As the amount of FA increased, the water absorption of bricks increased. The water absorption level of brick samples containing 50% FA was about 60% greater than that of the control mixture without FA. This finding is associated with the low quality of FA with a high loss on ignition. It is worth noting that the loss on ignition of FA is due to the loss of carbon and sulfur at high burning temperatures. The presence of unburned carbon increased the water absorption of FA [22-24], leading to an increase in the water absorption of FA bricks. However, all the FA brick samples had water absorption capacity of below 14% that satisfied the TCVN 6477-2011 requirements [17].

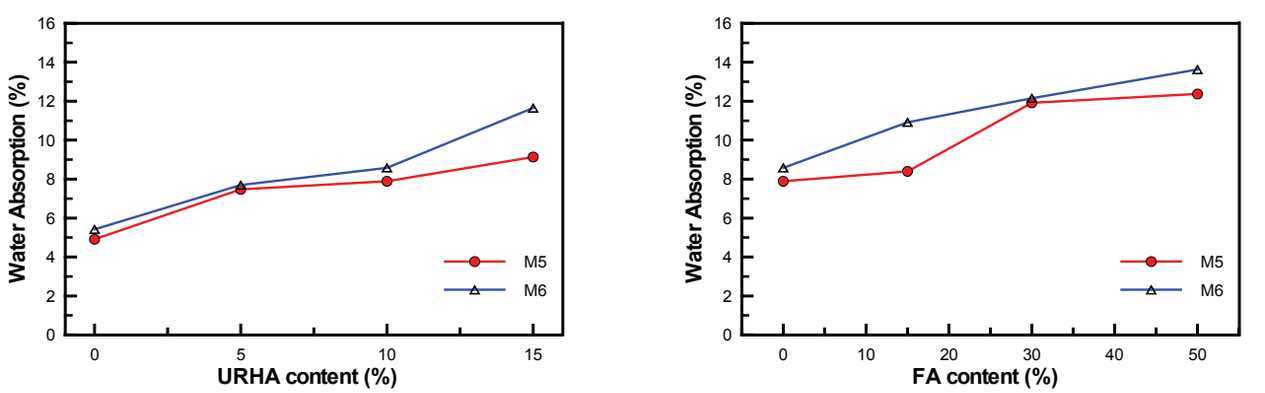


Fig. 5. Effect of (A) URHA and (B) FA contents on water absorption of the unfired building brick samples.

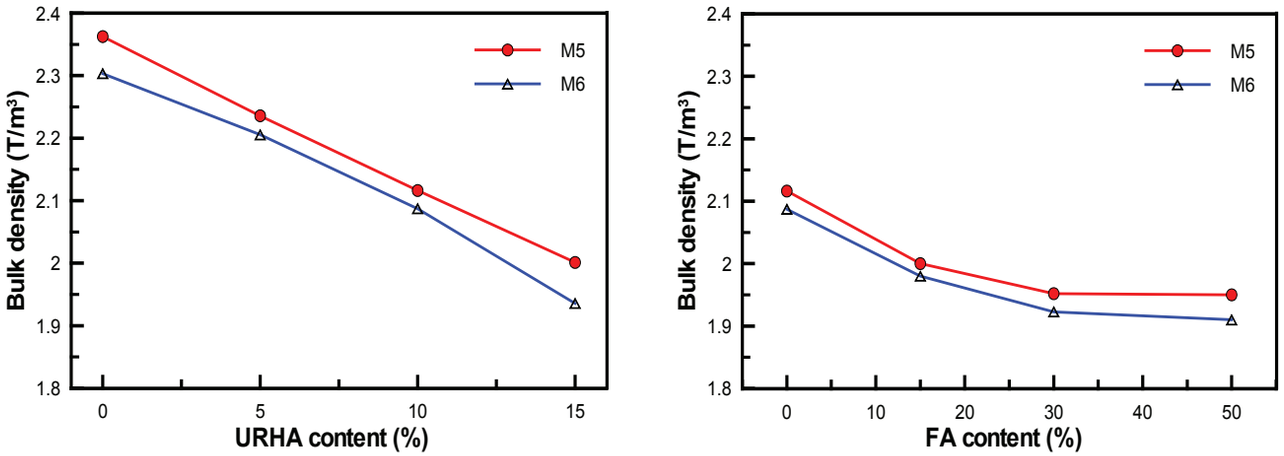


Fig. 6. Effect of (A) URHA and (B) FA contents on bulk density of the unfired building brick samples.

The water absorption of M5 mixtures was lower than that of the corresponding M6 mixtures. The lower water absorption values were mainly related to the amount of binder as mentioned previously. The water absorption of bricks has been negatively associated with its compactness and mechanical strength. In other words, brick samples with high strength and good compactness will register a low water absorption.

Bulk density

The bulk density is defined as the mass of brick divided by its volume. It is used as an indicator to classify a solid building brick. If the bulk density of bricks is high, the total mass of the building laying on the foundation is also high. Consequently, the required structure of the foundation needs to be strong enough to suffer the intensive load. Therefore, the use of light weight bricks is a good option to reduce the foundation cost. However, the bulk density is often inversely correlated with water absorption capacity. Fig. 6A shows the plot of the average bulk density of brick samples at the 28-day age against URHA content. The bulk density reduced by increasing the URHA content. The average bulk density of brick samples with 5%, 10% and 15% URHA content was around 4.8%, 9.9%, and 15.6%, respectively, lower than that of the no URHA bricks. This is mainly attributable to the lower specific density of the URHA in comparison with chippings.

Figure 6B shows the plot of the average bulk density of brick samples at the 28-day ages against FA content. Similarly, replacing cement with FA led to a reduction in bulk density of the brick samples. The average bulk density of brick samples with 15%, 30%, and 50% FA was around 5.3%, 7.8%, and 8.2%, respectively, lower than that of the FA-free bricks. This is mainly due to the lower specific density of FA compared with cement. The lowest bulk density of 1.91 T/m^3 was obtained from the M6-10-50FA mixture.

Cost analysis

The cost of bricks is a very important factor that shows the applicability of bricks in the market. Thus, the cost analysis was conducted to assess the economic efficiency of all brick mixtures. Table 6 shows the cost analysis for a brick. At the same water-to-binder ratio, the cost of each brick reduced with the use of more URHA and FA in the brick mixture. It is noted that the cost analysis was calculated based on the unit price of construction materials announced by the Department of Construction in Thanh Hoa in the first quarter of 2017. The price of water and URHA was taken as selling price in the current market. The unit price of cement, FA, chippings, water, and URHA was 1,227 VND/kg, 200 VND/kg, 1,238,000 VND/m^3, 13,860 VND/m^3, and 50,000 VND/ton, respectively. The labor cost was not included in this calculation. The result showed that the M6-10-50FA mixture had the lowest price of 500 VND, while the actual price of a Grade M15 unfired building brick available in the market was higher than 1,200 VND.

Table 6. Cost analysis for a brick.

Mixture	Material cost for a brick (VND)	Mixture	Material cost for a brick (VND)
M5-0	934	M6-0	803
M5-5	927	M6-5	797
M5-10	919	M6-10	790
M5-15	912	M6-15	784
M5-10-15FA	812	M6-10-15FA	702
M5-10-30FA	706	M6-10-30FA	614
M5-10-50FA	568	M6-10-50FA	500

Analysis for optimal mixture

As presented above, all the brick samples had compressive strength and water absorption levels that satisfied the TCVN 6477-2011 standard [17], in which the strength of bricks met the Grade M15 requirement and water absorption of bricks was below 14%. Therefore, the optimal mixture is a mixture that has the lowest bulk density and cost. Based on bulk density test and cost analysis, the M6-10-50FA brick mixture was found to be the optimal one. It had great potential to be manufactured on a large scale. This mixture provided a compressive strength value of 15.2 MPa, water absorption of 13.6%, bulk density of 1.91 T/m^3, and material cost of about 500 VND per sample.

Conclusions

In the present study, raw FA and URHA were used to produce unfired building bricks. The following conclusions may be drawn based on the above experimental results:

(1) All the brick samples made from URHA and FA had good properties that satisfied the TCVN 6477-2011 requirements. All the samples showed consistent shape without any visible defects.

(2) Using more URHA resulted in a reduction in compressive strength, bulk density, and brick cost. However, an adverse trend was observed with water absorption of brick samples.

(3) Increasing the FA content led to a reduction in compressive strength, bulk density, and brick cost, but an increase in the water absorption capacity of the brick samples, except mixture M6-10-15FA.

(4) With the compressive strength value meeting the Grade M15 requirement, a water absorption of lower than 14% and the lowest bulk density and material cost, the M6-10-50FA brick mixture was considered as the optimal mixture.

(5) The test results of this study encourage the use of raw FA and URHA in the manufacture of unfired building bricks. The recycling of such wastes is not only cost effective, but also reduces the negative impact on the environment due to the disposal of waste materials.

REFERENCES

[1] V.P. Della, I. Kuhn, D. Hotza (2002), "Rice husk ash as an alternate source for active silica production", *Materials Letters*, **57(4)**, pp.818-821, https://doi.org/10.1016/S0167-577X(02)00879-0.

[2] C.L. Hwang, L.A.T. Bui, C.T. Chen (2012), "Application of Fuller's ideal curve and error function to making high performance concrete using rice husk ash", *Computer and Concrete*, **10(6)**, pp.631-647, http://dx.doi.org/10.12989/cac.2012.10.6.631.

[3] P. Turgut (2010), "Masonry composite material made of limestone powder and FA", *Powder Technology*, **204(1)**, pp.42-47, https://doi.org/10.1016/j.powtec.2010.07.004.

[4] P. Chindaprasirt, K. Pimraksa (2008), "A study of FA-lime granule unfired brick", *Powder Technology*, **182(1)**, pp.33-41, https://doi.org/10.1016/ j.powtec.2007.05.001.

[5] Z. Zhang, J. Qian, C. You, C. Hu (2012), "Use of circulating fluidized bed combustion FA and slag in autoclaved brick", *Construction and Building Materials*, **35**, pp.109-116, https://doi.org/10.1016/j.conbuildmat.2012.03.006.

[6] T. Cicek, and M. Tanriverdi (2007), "Lime based steam autoclaved FA bricks", *Construction and Building Materials*, **21(6)**, pp.1295-1300, https://doi.org/ 10.1016/j.conbuildmat.2006.01.005.

[7] S. Kumar (2002), "A perspective study on FA-lime-gypsum bricks and hollow blocks for low cost housing development", *Construction and Building Materials*, **16(8)**, pp.519-525, https://doi.org/10.1016/S0950-0618(02)00034-X.

[8] A.A. Shakir, S. Naganathan, K.N. Mustapha (2013), "Properties of bricks made using FA, quarry dust and billet scale", *Construction and Building Materials*, **41**, pp.131-138, https://doi.org/10.1016/j.conbuildmat.2012.11.077.

[9] A. Kumar, S. Kumar (2013), "Development of paving blocks from synergistic use of red mud and FA using geopolymerization", *Construction and Building Materials*, **38**, pp.865-871, https://doi.org/10.1016/j.conbuildmat.2012.09.013.

[10] C. Freidin (2017), "Cementless pressed blocks from waste products

of coal-firing power station", *Construction and Building Materials*, **21(1)**, pp.12-18, https://doi.org/10.1016/j.conbuildmat.2005.08.002.

[11] O. Arioz, K. Kilinc, M. Tuncan, A. Tuncan, T. Kavas (2010), "Physical, mechanical and micro-structural properties of F type fly-ash based geopolymeric bricks produced by pressure forming process", *Advance in Science and Technology*, **69**, pp.69-74, doi: 10.4028/www.scientific.net/AST.69.69.

[12] C.L. Hwang, T.P. Huynh, Y. Risdianto (2016), "An application of blended FA and residual rice husk ash for producing green building bricks", *Journal of the Chinese Institute of Engineering*, **39(7)**, pp.850-858, http://dx.doi.org/10.1080/ 02533839.2016.1191376.

[13] C.L. Hwang, and T.P. Huynh (2015a), "Evaluation of the performance and microstructure of ecofriendly construction bricks made with FA and residual rice husk ash", *Advances in Materials Science and Engineering*, **2015**, 11pp, http://dx.doi.org/ 10.1155/2015/891412.

[14] C.L. Hwang, and T.P. Huynh (2015b), "Investigation into the use of URHA to produce eco-friendly construction bricks", *Construction and Building Materials*, **93**, pp.335-341, https://doi.org/10.1016/j.conbuildmat.2015.04.061.

[15] C.L. Hwang, and T.P. Huynh (2015c), "Investigation on the use of FA and residual rice husk ash for producing unfired building bricks", *Applied Mechanics and Materials*, **752-753**, pp.588-592, doi: 10.4028/www.scientific.net/AMM.752-753.588.

[16] C.L. Hwang, and T.P. Huynh (2015d), "Properties of unfired building bricks prepared from FA and residual rice husk ash", *Applied Mechanics and Materials*, **754-755**, pp.468-472, doi: 10.4028/www.scientific.net/AMM.754-755.468.

[17] Vietnamese standard TCVN 6477 (2011), *Concrete brick*, Ministry of Science and Technology, 9pp (in Vietnamese).

[18] ASTM C618 (2005), *Standard specification for coal FA and raw or calcined natural pozzolan for use in concrete*, 3pp.

[19] A. Oner, S. Akyuz, R. Yildiz (2005), "An experimental study on strength development of concrete containing FA and optimum usage of FA in concrete", *Cement and Concrete Research*, **35(6)**, pp.1165-1171, https://doi.org/10.1016/ j.cemconres.2004.09.031.

[20] A.H. Memon, S.S. Radin, M.F.M Zain, J.F. Trottier (2002), "Effects of mineral and chemical admixtures on high-strength concrete in seawater", *Cement and Concrete Research*, **32(3),** pp.373-377, https://doi.org/10.1016/S0008-8846(01)00687-1.

[21] V.G. Papadakis, and S. Tsimas (2002), "Supplementary cementing materials in concrete - Part I. Efficiency and design", *Cement and Concrete Research*, **32(10)**, pp.1525-1532, https://doi.org/10.1016/S0008-8846(02)00827-X.

[22] M. Mohebbi, F. Rajabipour, B.E. Scheetz (2015), "Reliability of loss on ignition test for determining the unburned carbon content in FA", *World of Coal Ash Conference in Nasvhille*.

[23] K. Wesch (1991), "FA in concrete: Properties and performance", *Report of Technical Committee 67-FAB use of FA in buildings*, E&FN Spon, London.

[24] J. Paya, J. Monzo, M.V. Borrachero, E. Perris, F. Amahjour (1998), "Thermo gravimetric methods for determining carbon content in FAes", *Cement and Concrete Research*, **28(5)**, pp.675-688, https://doi.org/10.1016/S0008-8846(98)00030-1.

The role of active silica and alumina in geopolymerization

Van Quang Le[1*], Quang Minh Do[2], Minh Duc Hoang[3], Hoc Thang Nguyen[4]

[1]*Vietnam Institute for Building Materials*
[2]*Ho Chi Minh city University of Technology*
[3]*Vietnam Institute for Building Science and Technology*
[4]*Ho Chi Minh city University of Food Industry*

Received 23 January 2018; accepted 18 April 2018

Abstract:

In this study, the alkaline solutions (NaOH) with concentrations from 1M to 18M, red mud (RM) and silica fume (SF) were used as reactors in geopolymer reactions. RM contains 7.40% SiO_2 and 13.65% Al_2O_3 and SF has 94.50% SiO_2, but only the active oxides can participate in the geopolymer reactions. The activity of the oxides was investigated by determining the compressive strength of the samples under different curing conditions. The characteristics of the geopolymer samples were determined by using X-ray diffraction (XRD), Fourier transform infrared spectroscopy (FTIR), thermo-gravimetric analysis (TGA)/differential thermal analysis (DTA) and nuclear magnetic resonance analysis (NMR). The experimental results indicate that active silica mainly exists in SF. In the structure of geopolymers, the silicon can bond directly with each other (Si-Si) or be linked through 'bridging' oxygen (Si-O-Si) to form independent polymer chains, while aluminium atoms can only replace the silicon atoms in Si-O-Si polymer chains to form Si-O-Al instead.

Keywords: ***active silica, alumina, geopolymer, red mud, silica fume.***

Classification number: ***2.3***

Introduction

Geopolymer is an inorganic polymer with structural units of $[SiO_4]^{4-}$ and $[AlO_4]^{5-}$ tetrahedrons [1]. The principle of the process is the formation of a polymer from the reaction of an alkaline solution (NaOH, KOH, Na_2SiO_3 and K_2SiO_3 solutions) with alumino-silicate resources [2, 3]. The structure of the geopolymer is a bonding of amorphous or semi-crystalline metal oxides with an alkaline element [4]. Therefore, raw materials for synthesising the geopolymer must contain major components of silicon dioxide, aluminium oxide and other oxides in amorphous and semi-crystalline forms. Crystal phases are inert, unreacted and not participated in geopolymer fabrication [5, 6]. The structures of the geopolymer are chains of -Si-O-Al-O- [7]. The mechanical properties of the geopolymer are influenced by the microstructure of the geopolymer.

The microstructure of the geopolymer is amorphous or semi-crystalline with three-dimensional structures based on tetrahedrons sharing oxygen atoms of the $[SiO_4]^{4-}$ and $[AlO_4]^{5-}$ molecular, which may exist in the poly-sialate form (Si:Al=2), the poly sialate disiloxo (-Si-O-Al-O) Al-O-Si-O-Si-O) (Si:Al=3) and other ratio sialate linkages (Si:Al>3). The sialate is an abbreviation for silicon-oxo-alumina [4].

The process of geopolymerization has 2 stages. The first stage is the synthesis of the geopolymer and the second stage is the polymeration of original materials with different alkaline activators. The alkali activation process of aluminosilicate is a complex process and has not been clearly explained yet [8, 9]. The major step of the geopolymer synthesis can be explained in the following stages [10, 11]:

- Extraction of active SiO_2 and Al_2O_3 in aluminosilicates by using the alkali hydroxide.

- Formation of tetrahedrons monomers.

- Formation of inorganic geopolymer structures by

*Corresponding author: Email: quanghuce83@gmail.com

monomers condensation reaction.

Geopolymerization will begin with the breakdown of the bonding Si-O-Si and then, Al atom will replace silicon atom in Si-O-Si bonding to form aluminosilicate gel with extremely large molecules [12]. This geopolymer process occurs in alkali solution. The inorganic polymer network consists of 3-dimensional aluminosilicate. In particular, the negative charge of Al in tetrahedron monomerons $[AlO_4]^{5-}$ will bond with the positive alkali ions such as Na^+ and K^+.

Geopolymers comprise the following molecular units (or chemical groups) that are presently studied and implemented in several industrial developments [13].

- Si-O-Si- siloxo, poly(siloxo).

- Si-O-Al-O- sialate, poly (sialate).

- Si-O-Al-O-Si-O- sialate siloxo, poly (sialate siloxo).

- Si-O-Al-O-Si-O-Si-O- sialate disiloxo, poly (sialate disiloxo).

- (R)-Si-O-Si-O-(R) organo siloxo, poly silicone.

- Al-O-P-O- alumino phospho, poly (alumino phospho).

- Fe-O-Si-O-Al-O-Si-O- ferro sialate, poly (ferro sialate).

Hence, any material containing amorphous oxides of silicon and aluminum such as red mud, fly ash, slag, silica fume can be used as a geopolymer material source [14].

RM is the solid waste in the manufacturing process of aluminum oxide by Bayer's technology. It contains excess sodium hydroxide (NaOH) and heavy metals that may cause many negative influences on human health and environment pollution. Thus, RM must be treated and disposed of in accordance with the regulations for hazardous waste management. The main components of RM are Fe_2O_3, SiO_2, Al_2O_3 and excess NaOH, which can be used as the material for the geopolymerization process. Furthermore, SF is also a solid waste in the metallurgical process. Silica fume has extremely fine particle size ranging from 0.1 µm to a few µm with a mean diameter of 1.5 µm. Fumed silica is mainly amorphous and hence, it is an auspicious material for geopolymerization. However, the geopolymer reactivity, physical and mechanical properties of the geopolymer products are influenced by the content of active SiO_2 and Al_2O_3 in RM and SF. The content of active SiO_2 and Al_2O_3 in RM and SF were evaluated by the amount of oxides dissolving in NaOH solutions of 1M to 15M at 80°C for 24 hours. The results showed that RM contains 4.76% active Al_2O_3 but does not contain active SiO_2 and SF contains 90.32% active SiO_2.

In this study, the geopolymer samples from RM were prepared by mixing the NaOH solution of 1M to 18M with RM in the NaOH/RM ratio of 0.4/1 (by weight). The samples were maintained at 60, 90, 120, 150, 180 and 210°C for 10 hours. Geopolymers' samples from SF were prepared by mixing the NaOH solution of 1M to 18M with SF in the NaOH/SF ratio of 0.2/1. The samples were pressed and maintained at room temperature.

The results of the microstructural analysis indicate that Si-Si and Si-O-Si bonds were formed to form independent polymer chains in the geopolymer samples. In the polymerization process, the Al atom can replace the Si atom in the polymer chain Si-O-Si to form Si-O-Al. For sufficient mechanical strength, active SiO_2 should be added to the geopolymer samples from RM; the samples from SF don't need added active oxides.

Experimental process

Materials

- RM from Tan Rai's Alumina Plant, Lam Dong Alumina Company in Vietnam.

- Silica fume (SF): Use 940U silica by Elkem Silicon Materials.

- Anhydrous NaOH: Bien Hoa Chemical Plant, Dong Nai Province in Vietnam.

Experimental process

Determination of active SiO_2 and Al_2O_3:

RM and SF were dried at 105 to 110°C to constant mass. About 2.5 g of the test sample (RM or SF) was put into a stainless-steel flask and then 25 ml of NaOH in varying concentrations (1 to 15M) were added. This was gently shaken several times, then cover with a lid and put in the oven at 80±2°C. After 24 hours, the kettle was stabilised at room temperature and the solution was filtered. The contents of silica and alumina dissolved in the solution were determined.

Experiment:

RM and SF were dried at 105 to 110°C to constant mass and sieved through a sieve of 0.08 mm.

The samples from RM were prepared by mixing RM with NaOH solution of 1M to 18M in the NaOH/RM ratio of 0.40/1 (by weight). The samples were formed in a stainless steel mold, pressed at 72 KN (10 N/mm^2) and had sizes of 90x80x40 mm. Then, the samples were removed from the mold immediately. The size of the samples conforms to TCVN 6477:2016. The samples were heated at 60°C

to 210°C for 2h, 4h, 6h, 8h and 10h. They were cured at room temperature for 28 days and then, were subjected to compressive strength testing.

The samples from SF were prepared by mixing SF with NaOH solution of 1M to 10M in the NaOH/SF ratio of 0.20 (by weight) (Table 1). Samples were prepared by semi-dry pressing at 72 KN (10 N/mm^2) in a mold with sizes of 90x80x40 mm. Then, they were removed from the mold immediately. The size of the samples conforms to TCVN 6477:2016. The samples were cured at room temperature for 28 days and then, were subjected to compressive strength and softening-coefficient. Some samples were selected to analyse the microstructure by using the methods of XRD, DTA-TG and NMR.

Table 1. Mixture proportion of geopolymer synthesis from SF and NaOH solution.

Sample	Ratio NaOH/SF	NaOH (M)
SF-Na1M	0.2	1
SF-Na2M	0.2	2
SF-Na3M	0.2	3
SF-Na4M	0.2	4
SF-Na5M	0.2	5
SF-Na6M	0.2	6
SF-Na7M	0.2	7
SF-Na8M	0.2	8
SF-Na9M	0.2	9
SF-Na10M	0.2	10

Results and discussion

Characteristics of the raw materials

The chemical compositions of RM and SF were determined by the XRF method, and the results are shown in Table 2. We can see that the silica content of SF is high, about 94.50% SiO_2. Additionally, the results in Table 2 shows that RM has a high L.O.I (loss on ignition) of about 12.50%, while SF has 2.74%.

The mineral composition of RM and SF were determined by using XRD and XRD patterns, which are shown in Fig. 1. The average particle size of RM was 9.5 μm by using the laser diffraction method.

Table 2. Composition of the materials.

Name	SiO_2	Al_2O_3	Fe_2O_3	Na_2O	L.O.I
RM (%)	7.40	13.65	56.05	3.63	12.50
SF (%)	94.5	0	0	0	2.74

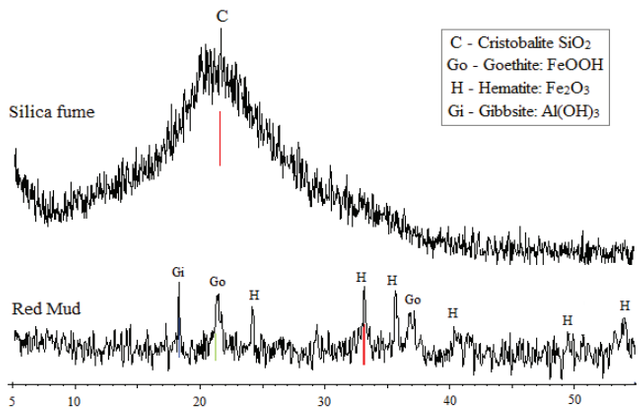


Fig. 1. XRD spectrum of the material.

Mineral compositions of RM are Goethite (FeOOH) 21%, Hematite (Fe_2O_3) 14% and Gibbsite ($Al(OH)_3$) 5%. The amorphous phase is 60%. The amorphous phase of SF is extremely high, about 99% SiO_2. The main crystal phase is cristobalite (SiO_2), and its content is extremely low.

The results of DTA of RM and SF are shown in Fig. 2 and Fig. 3. That was performed from room temperature up to 1,000°C (heating rate 5°C/min) (Fig. 3).

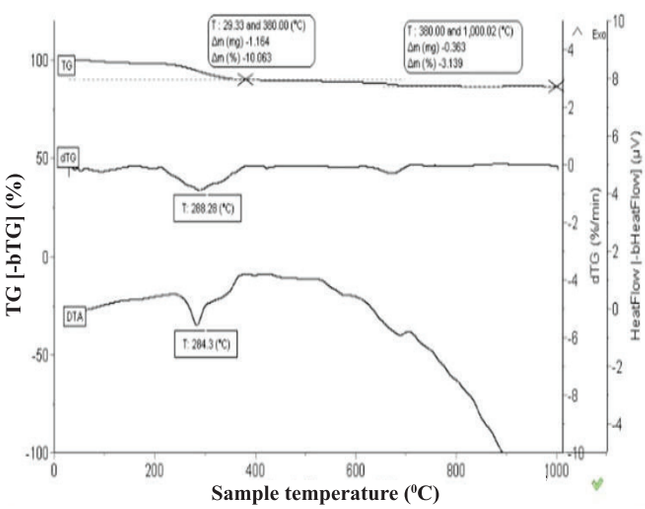


Fig. 2. The DTA curve of RM.

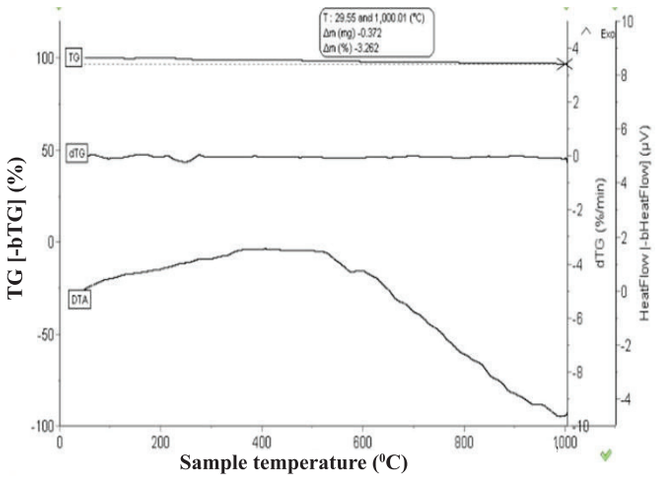


Fig. 3. Thermal analysis of fumed silica.

On the DTA curve of RM, there is an endothermic peak at 284°C, corresponding to the decomposition of $Al(OH)_3$ to Al_2O_3 and FeOOH to Fe_2O_3 [15]. The loss of ignition of RM is 10.06% after heating up to 380°C, and it continuously decreased to 3.14% from heating 380°C to 1,000°C. There is no significant heat effect on the DTA curve of SF; the loss on ignition of SF is only 3.26% during the heating.

The NMR spectrum of ^{29}Si of SF is shown in Fig. 4. The symbols Qn(mAl) are used to describe the structural monomers in aluminosilicates, where *n* represents the valence of the central silicon and *m* is the Al number around the SiO_4 monomer.

The MNR spectrum ^{29}Si of SF exhibits a narrow peak of 50.3% at 108.77 ppm. This peak is related to the number of wavelengths that may be present. The bond Q4(0Al) has a large component in the material, which is characterised by silica-rich SF.

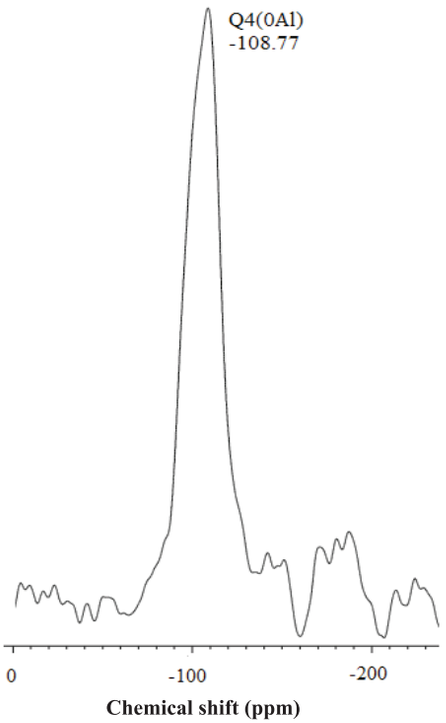


Fig. 4. NMR spectrum of ^{29}Si of SF.

The ratio of active SiO_2 and Al_2O_3 in the material

The content of active SiO_2 and Al_2O_3 in the raw materials are indicated in Table 3.

The results in Table 3 indicate that RM did not contain active SiO_2. The highest content of active alumina extracted from RM is 4.76% at the sodium hydroxide solution concentration of 5M, and the highest active silica content extracted from silica fume is 90.32% at the solution concentrations of at least 5 M.

Effects of active SiO_2 and Al_2O_3 on properties of the geopolymer

The samples from RM had not hardened. That is

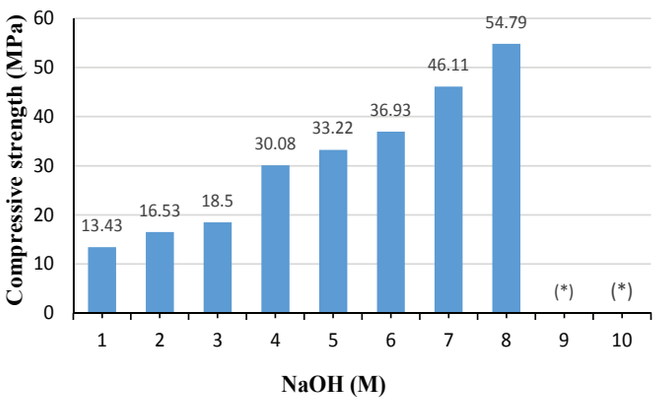
Table 3. The rates of active SiO_2 and Al_2O_3 in the raw material.

Samples		NaOH concentration (M)							
		1	*3*	*5*	*7*	*9*	*11*	*13*	*15*
RM	SiO_2 (%)	0	0	0	0	0	0	0	0
	Al_2O_3 (%)	4.13	4.74	4.76	4.76	4.76	4.76	4.76	4.76
SF	SiO_2 (%)	90.06	90.07	90.32	90.32	90.32	90.32	90.32	90.32

explained by the absence of active SiO_2 in RM. Although the content of SiO_2 in RM is 7.4% (Table 2), but they are not active SiO_2 and hence, they cannot participate in the geopolymer reaction. The content of active Al_2O_3 in the RM is 4.76% (Table 3) but they cannot polymerize because Al^{3+} is a modifier ion, and thus, they cannot form independent polymer chains.

In the presence of active SiO_2, a part of Al^{3+} having 4 oxygen coordination can replace Si^{4+} in the $[SiO_4]^{4-}$ tetrahedron to create a geopolymer network.

The lowest compressive strength of the samples (SF-Na1M) is 13.43 MPa. The highest compressive strength of the samples (SF-Na8M) is 54.79 MPa. The concentration of NaOH increased from 1M to 3M to increase the compressive strength of the samples from 13.43 MPa to 18.50 MPa. Notably, when using NaOH 4M, the compressive strength of the samples increased significantly - an increase of 67.5% compared to NaOH 3M (Fig. 5). This may explain that the concentration NaOH from 1M to 3M was insufficient to trigger the reaction. The higher the alkaline solution, the better the polymerization reaction.



(*) Sample was swollen

Fig. 5. Effect of NaOH concentration on geopolymer Compressive Strength.

Softening-coefficient is defined as the ratio of the compressive strength of a material saturated with water to that in the dry state. The lowest and the highest softening coefficient of the samples SF-Na1M and SF-Na8M were 75.28% and 99.87%, respectively. The softening-coefficient was rapidly increased from 75÷80% (of the samples SF-Na1M to SF-Na3M) to 99.19÷99.87% (of the samples SF-Na5M and SF-Na8M) (Fig. 6).

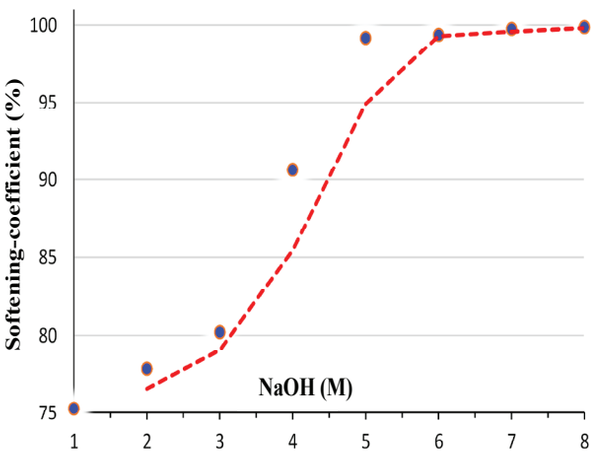


Fig. 6. Effect of NaOH concentration on the softening-coefficient.

The low softening-coefficient of the samples SF-Na1M to SF-Na4M can be explained by the low amount of alkaline solution, which is not enough to dissolve silicon and aluminum for geopolymerization. Thus, when the sample is saturated by water, many unreacted raw materials will be easily degraded to wash off, reducing the compressive strength of the sample. When the concentration of NaOH solution was increased above 5M, the geopolymer reaction increased, which led to the increase of softening-coefficient. However, when using the alkaline solution with a concentration higher than 8M, the geopolymer samples were swollen, which leads to crack and deformation. These samples did not have compressive strength.

The sample SF-Na4M was selected for structural analysis by XRD (Fig. 7), DTA-TG (Fig. 8) and NMR (Fig. 9).

On the XRD spectra of the samples SF and SF-Na4M, there is no new mineral peak. On the XRD spectrum of SF, there is only one peak corresponding to Cristobalite, which indicates that most of the silica content in SF was activated and they participated in the geopolymer reaction. Additionally, this proves that the formed phases during geopolymerization were amorphous.

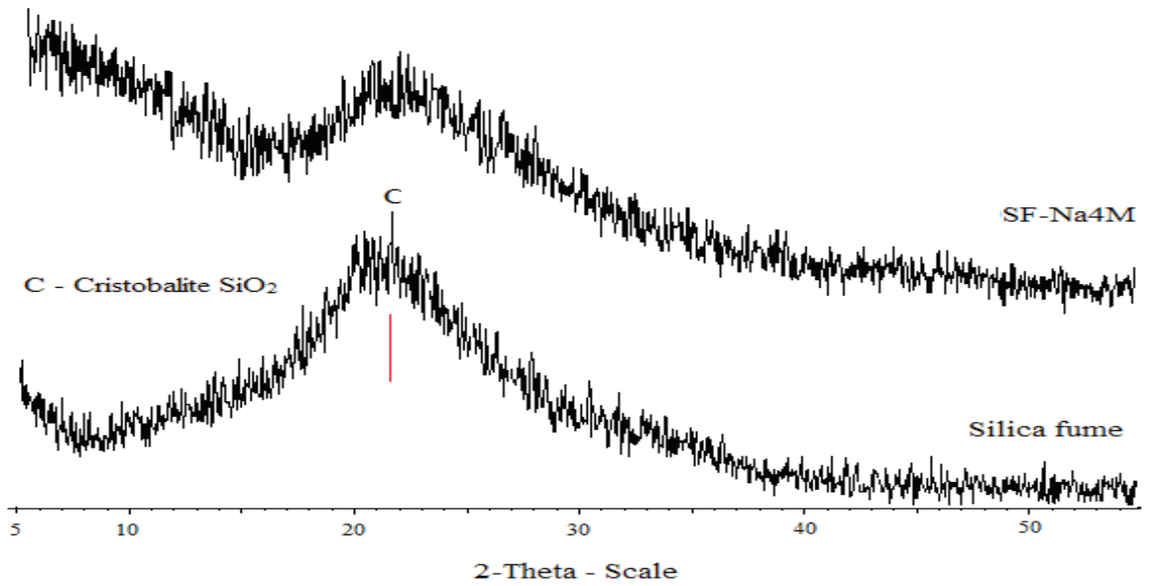


Fig. 7. XRD spectra of SF and SF-Na4M.

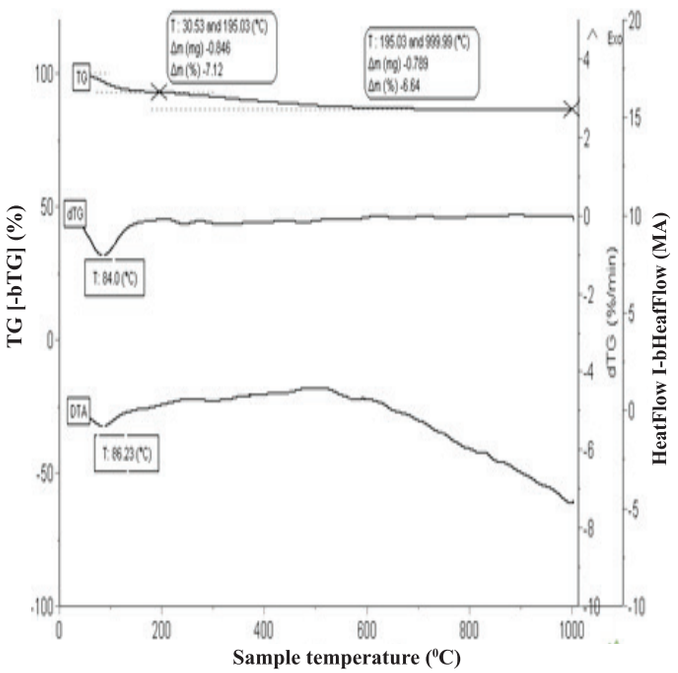


Fig. 8. The DTA curve of SF-Na4M.

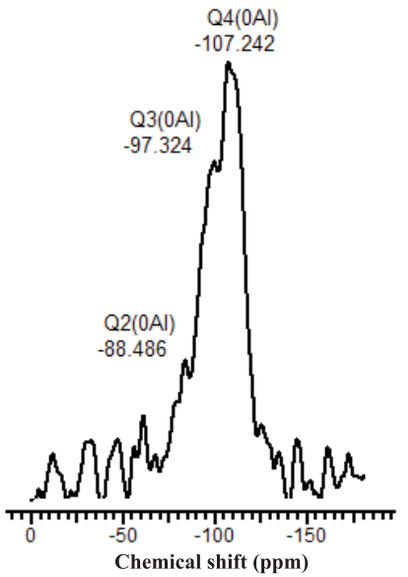


Fig. 9. NMR spectrum of ^{29}Si of the sample SF-Na4M.

The DTA-TG curves of the samples SF-Na4M are shown in Fig. 8.

Beyond only a peak of evaporation at 86°C, there is no significant heat effect on the DTA curve of the sample SF-Na4M (Fig. 8). On the TG curve, the loss on ignition is 13.76% (loss of 7.12% from room temperature to 195°C and 6.64% from 195°C to 1,000°C).

Nuclear Magnetic Resonance Analysis (NMR) of ^{29}Si

The NMR spectrum (Fig. 4) clearly shows that the 3-dimensional structure of SF was changed. The NMR spectra of SF appeared at a peak of -108.77 ppm such as Q4 (0Al) linkage (Fig. 4). After the polymerization process, two new vertices were found at -97.324 ppm and -88.486 ppm corresponding to Q3 (0Al) and Q2 (0Al) (Fig. 9). Alkaline dissolution starts with the attachment of the base OH^- to

the silicon atom, which is, thus, able to extend its valence sphere to the penta-covalent state and the new linkages are formed. Furthermore, it was found that the geopolymer reaction of the SF-Na4M sample did not occur completely. The amorphous content of SiO_2 is extremely high. This explains that the geopolymer samples still exists in Q4(0Al). SiO_2 with structure Q4(0Al) was not completely soluble and concentration of Q4(0Al) was lower than the original. In addition, the NMR intensity proportional to the number of ^{29}Si nuclei should allow the quantification of the phase. The characteristics of NMR pickups for the geopolymer samples are shown in Table 4.

Table 4. Characteristics of the ^{29}Si NMR spectrum of the SF-Na4M sample.

	Qn(mAl)	ppm	Width (ppm)	Intensity (%)
SF	Q4(0Al)	-108.707	23	50.3
GP	Q4(0Al)'	-107.242	17	37.2
	Q3(0Al)	-97.324	6	27.6
	Q2(0Al)	-88.486	5	9.3

From the data in Table 4, we have:

$$\sum(Q4(0Al)' + Q3(0Al) + Q2(0Al)) = \sum Q4(0Al) = 100\%$$

We also have the magnitude of the sum of the components in the geopolymer sample:

$$I_{(Q4(0Al)' + Q3(0Al) + Q2(0Al)} = I_{Q4(0Al)'} + I_{Q3(0Al)} + I_{Q2(0Al)} = 37.2 + 27.6 + 9.3 = 74.1\%$$

The percentage (%) of links in the geopolymer sample is calculated as follows:

Amount of phase A on phase B

$$\frac{W_A}{W_B} = \frac{I_A}{I_B} \times \frac{I_{oB}}{I_{oA}}$$

where: I_{oA}, I_{oB} are the intensity of standard diffraction beam.

The results of the linked units Qm(nAl) were shown in Table 5.

Table 5. Proportion of Qm(nAl) in the SF-Na4M.

	Qn(mAl)	ppm	Amount of phase (%)
SF	Q4(0Al)	-108.707	100
GP	Q4(0Al)'	-107.242	50.20
	Q3(0Al)	-97.324	37.25
	Q2(0Al)	-88.486	12.55

The samples from RM were not solidified although the active Al_2O_3 content was 4.76% compared to the total of 13.65%. The geopolymer samples from SF have high compressive strength, with the highest one being around 54.72 MPa of the sample SF-Na8M. This proves that active SiO_2 is indispensable and plays the most important role in the geopolymerization process. Al_2O_3 only plays a role in modifying the silicon polymer network.

Conclusions

Active silica plays the most important role in the geopolymerlyzation process because it makes the bonding and structure of the geopolymer. Silicon has the ability to bind directly to one another (Si-Si) or cross-link through silanes (Si-O-Si). When bonded via oxygen, the polymer chain can be expressed through coordinated multilane bonds, creating a three-dimensional network. The ions of the alkali oxides such as Na_2O, K_2O, CaO, MgO do not create a chain and are located in the hole coordinates.

When selecting raw materials for geopolymer materials, besides requiring materials containing the SiO_2 and Al_2O_3 components, the activity of SiO_2 must be present. In the geopolymerization process, active silica will form the bonds of monomer to achieve a geopolymer. Aluminium atom acts as a modifying ion. Al atom can only replace the Si atom in the polymer chain Si-O-Si.

It is necessary to add active SiO_2 when using RM of Tan Rai, Lam Dong to synthesize geopolymer. The active SiO_2 can be obtained from industrial waste such as fly ash, SF or glass water solution. The bonding and structure of geopolymer materials will be determined by the ratio of NaOH solution/SF and active silica. Silicon has the ability to bind directly to one another (Si-Si) or cross-link through

silanes (Si-O-Si). When bonded via oxygen, the polymer chain can be expressed through coordinated multilane bonds, creating a three-dimensional network. The ions of the alkali oxides such as Na_2O, K_2O, CaO, MgO do not create a chain and are located in the hole of structure network.

REFERENCES

[1] J. Davidovits (1989), "Geopolymers and geopolymeric materials", *Journal of Thermal Analysis and Calorimetry*, **35(2)**, pp.429-441.

[2] J. Giancaspro, P.N. Balaguru, and R.E. Lyon (2006), "Use of inorganic polymer to improve the fire response of balsa sandwich structures", *Journal of Materials in Civil Engineering,* **18**, pp.390-397.

[3] K. Goretta, J. Fuller, and E. Crawley (2006), "Geopolymers", *Air Force Office of Scientific Research Report,* Document # OSR-H-05-05.

[4] J. Davidovits (1994), "Geopolymers: man-made rocks geosynthesis and the resulting development of very early high strength cement", *Journal of Material Education,* **16,** pp.911-939.

[5] J. Davidovits (2002), "30 years of successes and failures in geopolymer applications, market trends and potential breakthroughs", *Geopolymer Conference,* Melbourne, Australia.

[6] J. Davidovits (1994), "Properties of geopolymer cements", *Proceedings 1st International Conference on Alkaline Cements and Concretes,* Scientific Research Institute on Binders and Materials (Kiev State Technical University, Ukraine), **199,** pp.131-149.

[7] J. Davidovits (2011), *Geopolymer chemistry and applications 3rd edition,* Institute Geopolymer - France.

[8] František škvára (2007), "Alkali activated materials or geopolymers?", *Ceramics - Silikáty,* **51**, pp.173-177.

[9] Joseph Davidovits (1999), "Chemistry of geopolymeric system terminology", *Géopolymère '99: Second International Conference*, pp.9-39.

[10] Van Chanh Nguyen, Dang Trung Bui, Van Tuan Dang (2008), "Recent research geopolymer concrete", *The 3rd ACF International Conference,* pp.235-241.

[11] Muhd Fadhil Nuruddi (2010), "Construction of infrastructures for sustainable futures"*, Seminar Nasional Aplikasi Teknologi Prasarana Wilayah*.

[12] R.E. Lyon, P.N. Balaguru, A. Foden, U. Sorathia, J. Davidovits, and M. Davidovics (1997), "Fire resistant aluminosilicate composites", *Fire and Materials*, **21**, pp.67-73.

[13] J. Davidovits (2015*), Geopolymer Chemistry and Applications 4th Edition,* Geopolymer Institute.

[14] H. Xu and J.S.J. Van Deventer (2000), "The geopolymerisation of alumino-silicate minerals", *International Journal of Mineral Processing*, **59**, pp.247-266.

[15] V.M. Sglavo, S. Maurina, A. Conci, A. Salviati, G. Carturan, G. Cocco (2000), "Bauxite "red mud" in the caramic industry. Part 2: Production of clay - based ceramic", *Journal of the European Society,* **20,** pp.245-252.

In vitro expression of enolase from *Streptococcus suis* serotype 2 and it's antigenicity

Hoang Bach Nguyen[1,2*], Thi Dang Khoa Nguyen[1,2], Gessica Tore[3], Van An Le[1,2], Alberto Alberti[3]

[1]*Carlo Urbani Centre, Hue University of Medicine and Pharmacy*
[2]*Department of Microbiology, Hue University of Medicine and Pharmacy*
[3]*Department of Veterinary Medicine, University of Sassari, Italy*

Received 20 November 2017; accepted 6 April 2018

Abstract:

***Streptococcus suis* serotype 2 (SS2) is one of the most important pathogens in the porcine industry and an important zoonotic agent. The absence of suitable vaccine or virulence markers makes SS2 infections more difficult to control. An immunoproteomics approach is used for identifying antigenic proteins in SS2 recognized enolase, which may represent strain-specific antigenic proteins and potential protective antigens. This study aims to clone, express enolase gene from SS2 and use western blotting to evaluate the antigenicity. Enolase gene from the SS2 strain was amplified with specific primers. The obtained PCR product was inserted into an expression vector, pGEX-4T1. The recombinant vector was then transformed into BL21 cells for protein expression. Subsequently, the immunological activity of the recombinant enolase was tested by western blotting with human sera in the convalescent phase. The glutathione S-transferase (GST)-tag fusion enolase was purified by glutathione sepharose affinity chromatography for further studies. The SS2 recombinant enolase was successfully expressed in *E. coli*. The western blotting analysis demonstrated that enolase has an antigenic property, which is recognized by patients naturally infected with SS2.**

***Keywords:* antigenicity, enolase, protein expression, *Streptococcus suis* serotype 2.**

***Classification numbers:* 3.1, 3.5**

Introduction

Streptococcus suis, commensal and opportunistic pathogens of swinea, α-hemolytic gram-positive cocci with 35 different serotypes. *Streptococcus suis* is an important pathogen in pigs and an emerging zoonotic agent in humans who are in contact with pigs or with their products [1]. *S. suis* infections in human are most often reported in countries where there is a high population of pigs. By the end of 2012, a total of 1584 cases had been reported in the literature (including 189 probable cases identified in 3 outbreaks), mainly from Thailand (36%), Vietnam (30%) and China (22%) [2]. Human infections with *S. suis* most frequently manifest as purulent meningitis, but septic shock with multiple organ failure, endocarditis, pneumonia, arthritis and peritonitis have also been reported [1]. In Thailand, China (including Hong Kong) and Vietnam, *S. suis* is an important cause of adult endocarditis, sepsis, and especially, meningitis [3]. Thirty-five serotypes (types 1-34 and 1/2) have been described based on capsular polysaccharides; serotype 2 is considered to be the most common pathogen in both humans and pigs [1, 4].

The lack of thorough knowledge about virulence markers and protective antigens can hinder the control of SS2 infections. Several approaches have been adopted to develop vaccines for *S. suis*. However, little success was achieved because the protection was either a serotype or strain dependent. More recently, interest has shifted towards the protein antigens of *S. suis* as vaccine candidates. Recently, immunoproteomics has become an effective approach for identifying immunoreactive proteins, which are essential antigens in the development of vaccine; they are also biomarkers for diagnostic and molecular therapy.

*Corresponding author: Email: nhbach@huemed-univ.edu.vn

Enolase, which is one of the remarkable antigenic proteins of SS2, was identified through immunoproteomics.

Enolase is a potential virulence factor of SS2; it was originally identified as a key glycolytic enzyme that catalyses the dehydration of 2-phosphoglycerate to phospho-enolpyruvate in the last steps of the catabolic glycolytic pathway [5, 6]. Recent studies have found that enolase plays an important role in many physiological and pathological processes, such as facilitating pathogen invasion into the host [7-9], cancer metastasis [10-14] and apoptosis [15, 16]. Recently, it was also recognized as an immunodominant antigen involved in the virulence of the Streptococcus species [6, 8, 17, 18]. It is noteworthy that enolase is an antigenic protein recognised by patient sera with SS2 infection based on the immunoproteomics approach [19]. In this study, the enolase gene from SS2 was cloned and expressed to confirm its antigenic property.

Methods and materials

Bacterial strain and patient sera collection

SS2 strains were isolated and identified in patients infected with *Streptococcus suis* serotype 2, as described in previous publication [19]. The patient sera were collected from the SS2 infected patients in convalescent phases who were hospitalized for treatment. Both the bacterial isolates and patient sera were stored in a deep freezer at -80°C [20]. One Shot™ TOP10 Chemically Competent *E. coli* (Thermo Fisher Scientific, Massachusetts, USA) and One Shot BL21 Star (DE3) Chemically Competent *E. Coli* (Thermo Fisher Scientific, Massachusetts, USA) were used as recipients of recombinant enolase, containing the vector pGEX- 4T1 (Amersham).

Amplification of enolase gene and cloning enolase gene into expression vector

A PCR assay was performed by using forward primer (5'-GGC GGA TCC ATG TCA ATT ATT ACT GAT G-3') and reverse primer (5'-ACG CTC GAG TTA TTT TTT CAA GTT GTA GAA TGA G-3'), which specifically targets 1305-bp of the enolase gene of locus taq SSUBM407_1397 of *Streptococcus suis* BM407 complete genome data (NC_012926.1). The primes were design in the Geneious v8.1 software, which contained *BamH*I and *Xho*I sites at 5'and 3' ends, respectively [21]. The PCR amplification was profiled as follows: initial denaturation at 94°C for 2 minutes, followed by 36 cycles of 95°C for 15s, 55°C for 30s and 68°C for 30s in Veriti® Thermal Cycler (Applied Biosystems, CA, USA).

The PCR products were purified using the QIA quick PCR Purification Kit (Qiagen, Hilden, Germany), following the manufacturer's instructions. Ten of the purified enolase PCR products and 0.32 µM of the primer were used for direct sequencing. To sequence both strands, two specific PCR primers were run for each enolase PCR product samples. The chromatograms were analysed by the Geneious software v8.1 and compared with the enolase sequence data available in the GenBank (http://www.ncbi.nlm.nih.gov/genbank/), using the BLASTn plugin of the Geneious software [21]. All sequences were aligned using ClustalX [22].

The 1305 bp PCR product, which corresponded to the obtained enolase, was restricted, gel eluted and inserted into the expression vector pGEX-4T1 in frame with the GST-tag sequence. The recombinant plasmid was introduced into *E. coli* TOP10 by heat shock transformation, and the clones were selected by growing the cultures on Luria-Bertani (LB) agar in the presence of 100 µg/ml ampicillin. For identifying the positives clones, miniprep was performed, and then, the putative clones were checked by restriction digestion.

Expression of recombinant S. suis enolase

The recombinant vectors were transformed into *E. coli* BL21 for expression. The transformation procedure followed the instructions of TransformAid Bacterial Transformation Kit (Thermo Fisher Scientific, USA) [23]. A single recombinant colony of BL21 was grown on a SOB (super optimal broth) medium at 37°C overnight with shaking. The protein expression was induced by the addition of 0.1 mM isopropyl b-D-1-thiogalactopyranoside (IPTG). One hour later, the cells were collected by centrigugation, and the cell pellet at were stored at -20°C; this collection process continued for a period of 4 hours with 1 hour intervals.

SDS-PAGE and Western Blotting

The recombinant enolase protein was resolved in 2 separate acrylamide gels. After electrophoresis, one gel was alternatively stained with SimplyBlue™ SafeStain (Novex, Life Technologies), and then digitalized, with the Gel Doc™ XR+ System (Bio-rad, USA).

The second gel was subjected to western blot assay. The proteins were transferred onto a nitrocellulose membrane (Hybond-C Extra, Amersham, GE) with a Mini-Trans-Blot Cell (Bio-Rad, CA, USA) at 250 mA (100 V) for one hour at 4°C. Then, the membranes were blocked with PBS-T containing 5% (w/v) skim milk. Membranes were incubated for one hour with a patient serum in 1:1500 dilutions in PBS-T, containing 2% (w/v) skim milk in a Mini-PROTEAN II Multiscreen Apparatus (Bio-Rad, CA, USA). The incubated membranes with the Goat Anti-Human IgG-HRP conjugated (Southern Biotech, Alabama, USA) as secondary antibodies diluted in the ration 1:50000 in PBS-T, containing 5% (w/v) skim milk, in one hour. The membranes were developed with Luminata™ Forte Western HRP substrate (Merck Millipore Corp., Darmstadt, Germany), and images were acquired with a ChemiDoc™ XRS+ System and were analysed by Image Lab Software (Bio-Rad, CA, USA).

Affinity purification of GST fusion enolase

The GST-tag added to the enolase allowed its isolation by affinity chromatography on glutathione column. Overnight cultures of *E. coli* BL21 cells harbouring recombinant enolase were sub-cultured in an LB medium with ampicillin 100 µg/ml and chloramphenicol 35 µg/ml and incubated at 37°C until an optical density of 0.6 at 600 nm (OD_{600}) was reached. Then, the *E. coli* were grown with 0.1 mM IPTG for 3 hours. The cells were centrifuged, and then, the pellet was re-suspended in ice-cold 1X PBS buffer with protease inhibitors. After sonication on ice, the lysates were centrifuged at 10,000×g for 10 minutes, and the recombinant protein was purified from the filtered supernatants by affinity chromatography. The procedure was performed as described in the Recombinant Protein Purification Handbook-GE. Healthcare [24].

Results and discussion

Enolase sequences analysis

Amplification of the full length coding sequences (CDS) of the enolase gene was confirmed by gel electrophoresis of the amplified fragments. The expected size of about 1305 bp was obtained, which corresponded to the SS2 enolase gene (Fig. 1).

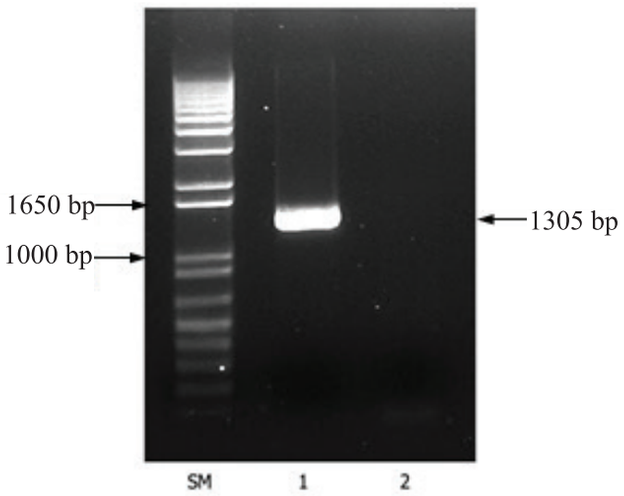


Fig. 1. PCR amplification of enolase gene from *S. suis* serotype 2. The PCR products were separated on 1.0% agarose gel. SM: 1 kb DNA Ladder (D0428, SIGMA DNA); lane 1: PCR product; lane 2: Negative control.

The raw data of enolase sequences were analysed with a multiple alignment algorithm in Geneious v8.1 with the enolase sequence from the reference genome of *S. suis* BM407 (NC_012926) in the Genbank (https://www.ncbi.nlm.nih.gov/nuccore/NC_012926.1). The sequence assembly data showed that our sequences matched with the *eno* CDS at locus_tag SSUBM407_1397 of the reference genome sequence (Fig. 2).

Cloning and expression of enolase in pGEX-4T1

The PCR products and pGEX-4T1 vector were digested by the *BamH*I/*Xho*I and purified by the ZymoClean™ Gel DNA Recovery Kit (Zymo Research Corp., USA). The enolase was cloned into the TOP10 *E. coli*. Double digestion of the enolase recombinant vector by *BamH*I/*Xho*I showed that the expected size of 1300 bp fragment on agarose gel electrophoresis corresponded to the enolase gene (Fig. 3).

Expression of recombinant enolase

The enolase gene was 1305 bp in size and encoded into a 435 amino acid protein with a predicted molecular mass of 52 kDa. The result of expression from the pGEX-4T1 is a GST-tagged fusion enolase in which the functional GST protein (26 kDa) is fused with the N-terminus of the recombinant enolase. By comparing the SDS-PAGE results of the non-induced and IPTG-induced BL21 strains (Fig. 4), it can be seen that induced BL21- pGEX 4T1-enolase lands

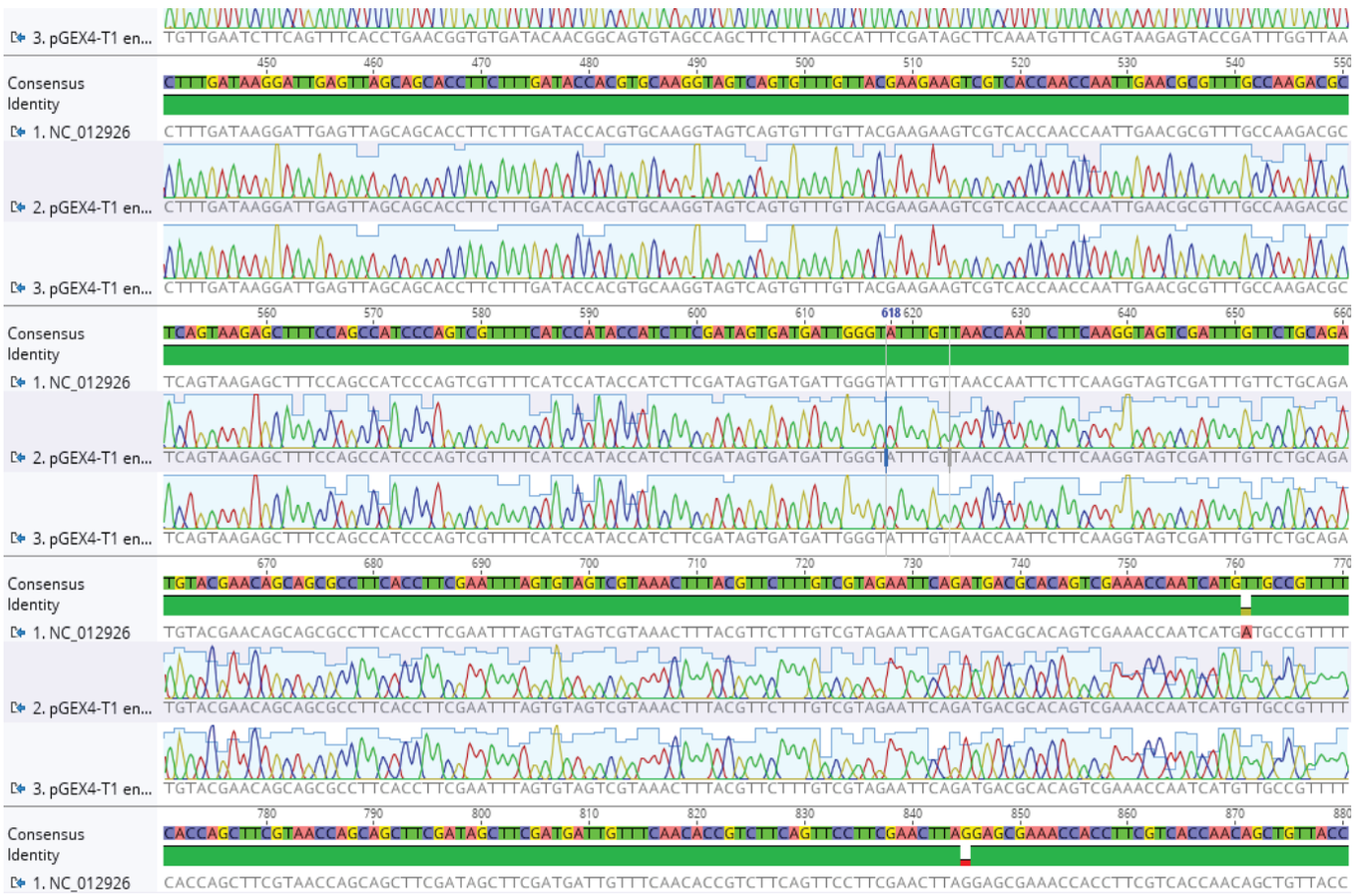


Fig.2. Analyse the chromatograms of *S. suis* enolase sequences by Align/Assembly plugin of Geneious v8.1.

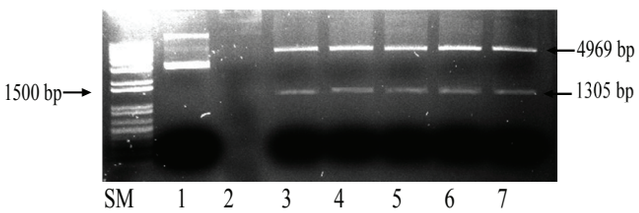


Fig. 3. Agarose gel electrophoresis of pGEX-4T1 recombinant vector digested *BamH*I/*Xho*I. SM: 1 kb DNA ladder; lane 1: pGEX-4T1-enolase undigested; lane 3-7: pGEX-4T1-enolase digested with *BamHI/XhoI.*

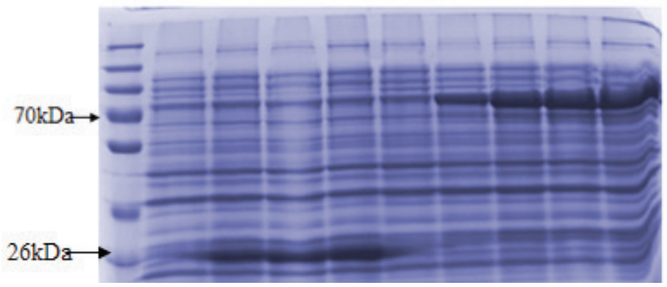


Fig. 4. SDS-PAGE analysis of recombinant enolase. Lane 1: PageRuler™ Prestained Protein Ladder Plus (Thermo Fisher Scientific Inc.); Lane 2: Non-induced BL21; Lane 3-5: Induced BL21-pGEX 4T1 empty; Lane 6-10: Induced BL21-pGEX 4T1-enolase at time zero, 1h, 2h, 3h and 4h, respectively.

have a strong band with an approximate size of 78 kDA. The other lanes (2-5) also have a light band with an approximate size of 78 kDA. The antigenic property of the recombinant enolase in the total protein of the induced BL21-pGEX 4T1-enolase was evaluated in the western blot assay.

Testing antigenic property of recombinant enolase

Western blotting was used to test the antigenic property of the recombinant enolase with sera from naturally infected patients. The primary antibody was diluted in a 1:1000 ratio, and the secondary antibody was diluted in a 1:50000 ratio. The protein reacted to the convalescent sera with a strong band with a predicted size of 78 kDa (Fig. 5).

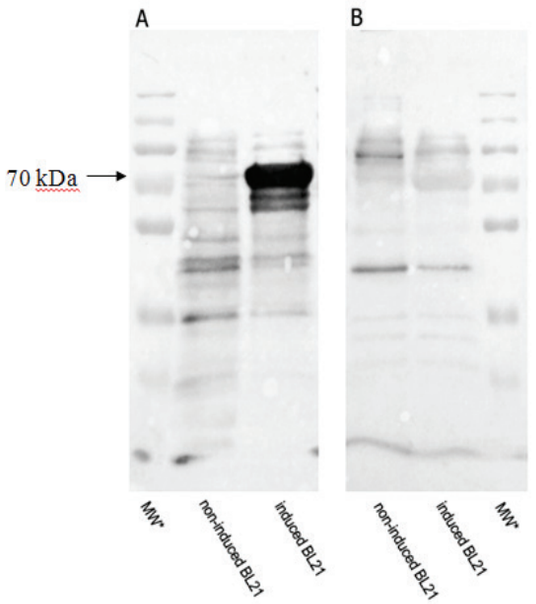


Fig. 5. Western blotting analysis of recombinant enolase. Convalescent serum **(A)** and negative serum **(B)**. MW: PageRuler™ Prestained Protein Ladder Plus (Thermo Fisher Scientific Inc.).

Purification of recombinant enolase

Protein purification was confirmed through a SDS-PAGE analysis. From the SDS-PAGE gel, it can be seen that there is a pure recombinant enolase in the elution fractions, and the fourth elution fraction contains the highest intensity band (Fig. 6).

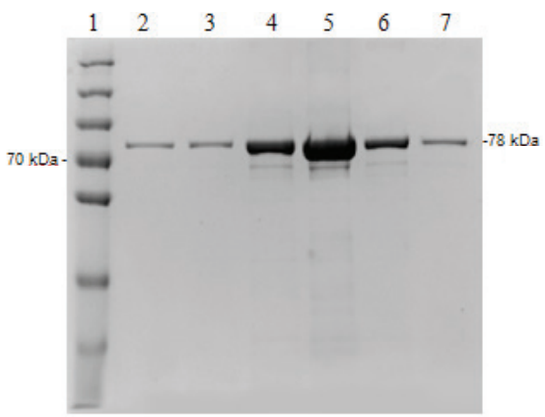


Fig. 6. SDS-PAGE analysis of affinity purification. Lane 1: Protein ladder; Lane 2-7: Elution fractions of purified enolase.

This evidence demonstrates that the recombinant enolase with antigenicity can be recognized by patient serum. It could be a candidate for developing a vaccine against SS2, which is present at the surface of all 35 *S. suis* serotypes. It contributes to *S. suis* adhesion to and invasion of host cells [8], and it is a highly conserved protein [25].

Conclusions

We have successfully cloned the SS2 enolase gene into the pGEX 4T1 vector and the recombinant enolase into BL21 cells. The recombinant enolase has antigenicity and can be recognized by patient sera infected SS2 in western blot assay.

REFERENCES

[1] Z.R. Lun, Q.P. Wang, X.G. Chen, A.X. Li, X.Q. Zhu (2007), "Streptococcus suis: an emerging zoonotic pathogen", *Lancet Infect. Dis.*, **7(3)**, pp.201-209.

[2] V.T.L. Huong, N. Ha, N.T. Huy, P. Horby, H.D.T. Nghia, V.D. Thiem, et al. (2014), "Epidemiology, clinical manifestations, and outcomes of streptococcus suis infection in humans", *Emerg. Infect. Dis.*, **20(7)**, pp.1105-1114.

[3] H.F.L. Wertheim, H.D.T. Nghia, W. Taylor, C. Schultsz (2009), "*Streptococcus suis:* An Emerging Human Pathogen", *Clin. Infect. Dis.*, **48(5)**, pp.617-625.

[4] R. Higgins, M. Gottschalk, M. Boudreau, A. Lebrun, J. Henrichsen, R.Y. Reams, et al. (1995), "Description of six new capsular types (29-34) of Streptococcus suis Fibrinohemorrhagic pneumonia in pigs naturally infected with Streptococcus suis", *J. Vet. Diagn. Invest.*, **7(3)**, pp.405-406.

[5] E. Zhang, J.M. Brewer, W. Minor, L.A. Carreira, L. Lebioda (1997), "Mechanism of enolase: the crystal structure of asymmetric dimer enolase-2-phospho-d-glycerate/enolase-phosphoenolpyruvate at 2.0 Å resolution", *Biochemistry*, **36(41)**, pp.12526-12534.

[6] V. Pancholi (2001), "Cellular and Molecular Life Sciences Multifunctional a -enolase: Its role in diseases", *Cell Mol. Life Sci.*, **58(7)**, pp.902-920.

[7] I. Veiga-Malta, M. Duarte, M. Dinis, D. Tavares, A. Videira, P. Ferreira (2004), "Enolase from Streptococcus sobrinus is an immunosuppressive protein", *Cell Microbiol.*, **6(1)**, pp.79-88.

[8] M. Esgleas, Y. Li, M.A. Hancock, J.D. Dubreuil, M. Gottschalk, M. Gottschalk (2008), "Isolation and characterization of a -enolase, a novel fibronectin-binding protein from *Streptococcus suis*", *Microbiology*, **154**, pp.2668-2679.

[9] C.A. Carter, G.S. Chhatwal (2002), "Anchorless adhesins and invasins of Gram-positive bacteria: a new class of virulence factors", *Trends in Microbiol.*, **10(5)**, pp.205-208.

[10] M. Capello, S. Ferri-borgogno, P. Cappello, F. Novelli (2011), *A -enolase: a promising therapeutic and diagnostic tumor target*, **278**, pp.1064-1074.

[11] K. Ejeskär, C. Krona, H. Carén, F. Zaibak, L. Li, T. Martinsson, et al. (2005), "Introduction of in vitro transcribed ENO1 mRNA into neuroblastoma cells induces cell death", *BMC Cancer*, **5(161)**, pp.1-14.

[12] S. Feo, D. Arcuri, E. Piddini, R. Passantino, A. Giallongo (2000), "ENO1 gene product binds to the c-myc promoter and acts as a transcriptional repressor: relationship with Myc promoter-binding protein 1 (MBP-1)", *FEBS Letters*, **473(1)**, pp.47-52.

[13] A.K. Ghosh, R. Steele, R.B. Ray (2005), "c-myc Promoter-binding protein 1 (MBP-1) regulates prostate cancer cell growth by inhibiting MAPK pathway", *J. Biol. Chem.*, **280(14)**, pp.14325-14330.

[14] G. Chang, K. Liu, C. Hsieh, T. Hu, S. Charoenfuprasert, H. Liu, et al. (2006), "Identification of a-enolase as an autoantigen in lung cancer: its overexpression is associated with clinical outcomes", *J. Immunol.*, **12(19)**, pp.5746-5755.

[15] D.S. Ucker, M.R. Jain, G. Pattabiraman, K. Palasiewicz, R.B. Birge, H. Li (2012), "Externalized glycolytic enzymes are novel, conserved, and early biomarkers of apoptosis", *J. Biol. Chem.*, **287(13)**, pp.10325-10343.

[16] H. Yang, W. Zheng, X. Zhang, F. Tang (2011), "Induction of endothelial cell apoptosis by anti-alpha-enolase antibody", *Chinese Med. Sci. J.*, **26(3)**, pp.152-157.

[17] M. Dinis, D. Tavares, I. Veiga-malta, A.J.M.M. Fonseca, E.B. Andrade, G. Trigo, et al. (2009), "Oral therapeutic vaccination with Streptococcus sobrinus recombinant enolase confers protection against dental caries in rats", *J. Infect. Dis.*, **199(1)**, pp.116-123.

[18] J. Kolberg, A. Aase, S. Bergmann, T.K. Herstad, G. Rødal, R. Frank, et al. (2006), "Streptococcus pneumoniae enolase is important for plasminogen binding despite low abundance of enolase protein on the bacterial cell surface", *Microbiology,* **152(5)**, pp.1307-1317.

[19] B.H. Nguyen, D.H.N. Phan, H.X. Nguyen, Van A Le, A. Alberti (2015), "Molecular diagnostics and ITS-based phylogenic analysis of streptococcus suis serotype 2 in central Vietnam", *J. Infect. Dev. Ctries.,* **9(6)**, pp.624-630.

[20] B.H. Nguyen, Van A Le, A. Alberti (2016), "Identification antigenic proteins of Streptococcus suis serotype 2 by immunoproteomics approach", *J. Prat. Med.*, Vietnamese, **1005**, pp.500-504.

[21] M. Kearse, R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, et al. (2012), "Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data", *Bioinformatics*, **28(12)**, pp.1647-1649.

[22] M.A. Larkin, G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, et al. (2007), "Clustal W and Clustal X version 2.0", *Bioinformatics*, **23(21)**, pp.2947-2948.

[23] S. Harper, D.W. Speicher (2011), "Purification of proteins fused to glutathione S-tranferase", *Methods Mol. Biol.*, **681**, pp.259-280.

[24] G.E. Healthcare (2009), *Recombinant Protein Purification Handbook*, Protein Purification.

[25] S. Ehinger, W.D. Schubert, S. Bergmann, S. Hammerschmidt, D.W. Heinz (2004), "Plasmin(ogen)-binding alpha-enolase from streptococcus pneumoniae: crystal structure and evaluation of plasmin(ogen)-binding sites", *J. Mol. Biol.*, **343(4)**, pp.997-1005.

The applications of massive parallel sequencing (next-generation sequencing) in research and molecular diagnosis of human genetic diseases

Hieu T. Nguyen[1*], Huong T.T. Le[1], Liem T. Nguyen[1], Hua Lou[2], Thomas LaFramboise[2]

[1]Vinmec Research Institute of Stem Cell and Gene Technology, Vinmec International Hospitals, Hanoi, Vietnam
[2]Department of Genetics and Genome Sciences, School of Medicine, Case Western Reserve University, Cleveland, Ohio, USA

Received 2 February 2018; accepted 22 May 2018

Abstract:

Next-generation sequencing (NGS) is a high throughput sequencing technology, which has revolutionized both basic and clinical research of the human genetic disorders. This technology is also called massively parallel sequencing (MPS) due to its ability to generate a huge amount of output data in a cost- and time-effective manner. NGS is widely utilized for different sequencing applications such as targeted sequencing (a group of candidate genes), exome sequencing (all coding regions), and whole genome sequencing (the entire human genome). With NGS, a variety of genomic aberrations can be screened simultaneously such as common and rare variants, structural variations (amplifications and deletion), copy-number variation, and fusion transcripts. NGS technologies combined with advanced bioinformatic analysis have tremendously expanded our knowledge. On the one hand, the basic research area involves direct use of NGS to identify novel variations and determine human disease mechanisms. On the other hand, clinical research is being advanced by high-throughput genetic tests with high resolution and clinically relevant genetic information for molecular diagnoses of human disorders. In this communication, we introduce NGS technologies and review a few key areas where NGS has made a significant impact, with an emphasis on the application of NGS to the identification of the molecular bases of human genetic diseases.

Keywords: ***human genetic diseases, massively parallel sequencing, molecular diagnosis, next-generation sequencing.***

Classification numbers: ***3.2, 3.5***

Introduction

For more than four decades, Sanger sequencing based on the dideoxy chain termination principle has been considered the gold standard method for determining a DNA sequence and the identification of genomic variations to support the diagnosis of genetic diseases [1]. For monogenic diseases with clear clinical and biochemical presentations, and well characterized mutation landscapes, sequencing the target regions by the Sanger method is an accurate and cost-effective way to obtain a conclusive molecular diagnosis. Nevertheless, as most inherited diseases are often genetically and clinically heterogeneous, the selection of candidate gene(s) and/or gene region(s) for sequence analysis is costly, laborious, and time consuming, which often delays diagnosis and treatment, causing anxiety for patients and their families. Many neurological disorders such as ataxias, epilepsy, and migraines are caused by mutations in one of many genes. For example, 65 genes were shown to be responsible for Retinitis Pigmentosa, one common form of hereditary retinal degeneration, demonstrating its high heterogeneity and diversity of inheritance patterns. The diagnosis of mitochondrial diseases is another demonstration for an extreme situation, for which clinical phenotypes significantly overlap and heterogeneous mutations span more than 1,300 genes [2-5].

The number of recognized polygenic conditions has greatly increased due to the rapid discovery of new genes, genetic conditions, and phenotypic ranges. Thus, the traditional step-wise molecular diagnostic approach of single genes or candidate genes is no longer adequate to identify the molecular etiologies of diseases. Additionally, amplicon-based Sanger sequencing is notorious for allele

*Corresponding author: Email:htn13@case.edu

dropout due to either Single Nucleotide Polymorphism (SNPs) at the PCR primer sites and large deletions including one or both of the primer sites [6]. Moreover, the complexity of genomic research and application including diagnosis of genetic diseases demands a depth of information beyond the capacity of traditional DNA sequencing technologies. The need to address these drawbacks has spurred the birth of a new NGS approach, which is more comprehensive, accurate, and effective. Massive parallel sequencing technologies (MPS or NGS) have enabled sequencing of many, usually short fragments of nucleic acid at the same time to provide deep sequencing coverage of individual samples or indexing of multiple samples.

During the past decade, NGS has revolutionized nearly every area of biological sciences by generating the enormous genetic information for the identification of genomic variations, disease mechanisms, and disease-associated markers, which has led to the development of better diagnostics tools and treatment therapies. High-throughput sequencing including (1) targeted sequencing (genes of interest), (2) whole exome sequencing (protein-coding portions), and (3) whole genome sequencing (the entire human genome) allows the detection of mutations in multiple genes in a cost-time effective fashion [7-13].

A number of research studies have successfully utilized the NGS technology to identify genes related to diseases [14], causative mutations [15] and epigenetic modulations correlated with particular disorders [16, 17]. NGS approaches have been also applied to the molecular diagnosis of genetic diseases, particularly complex disorders with heterogeneous clinical phenotypes and various underlying genetic causes [18-21]. Generally, sequencing more than 3 billion base pairs of the whole human genome is economically unfeasible, computationally challenging and technically demanding. Thus, in clinical setting, it is frequently desirable to capture or enrich genes demonstrated to be important for a particular clinical phenotype, followed by NGS. Here, we review NGS technologies and summarize the recent applications of NGS in both basic and clinical research with a focus on the molecular diagnosis of human genetic diseases.

Overview of NGS technologies

To date, there have been several generations of sequencing technologies, which are different regarding sequencing principles, sequencing chemistries, and instrumentation (Fig. 1).

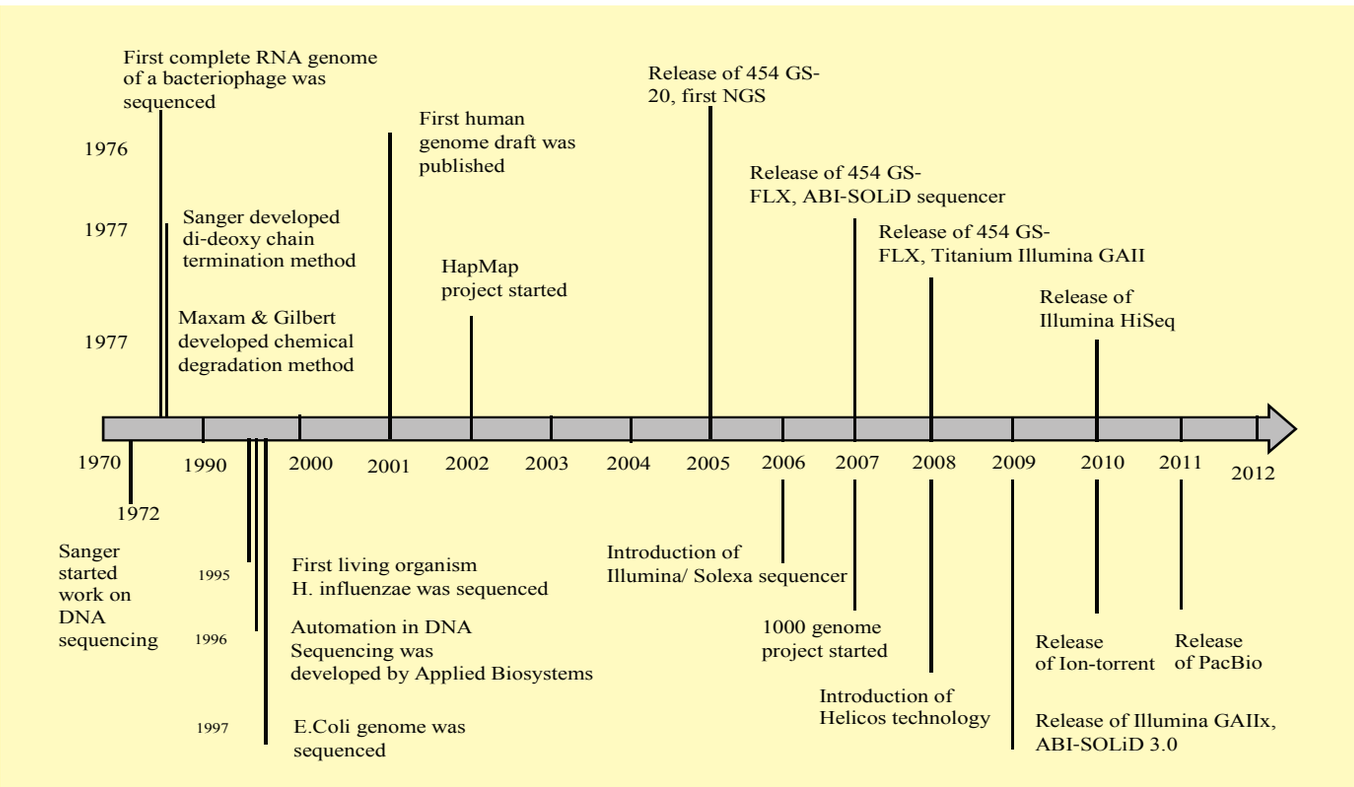


Fig. 1. Timeline of introduction of DNA sequencing technologies and platforms [22].

The first generation of sequencing was defined by the Sanger and Maxam-Gilbert techniques, which are capable of sequencing a few hundred base pairs at a time, and could be used for single gene sequencing [1].

NGS technologies, also called second-generation sequencers, was first introduced to the scientific community in 2005, over 30 years after Sanger sequencing was introduced. The major advancement of second-generation sequencing is its capability to produce sequencing data in a massively parallel manner, thus generating huge amounts of data in a cost- and time-effective fashion. Next generation sequencers are featured by several platforms that produce the large scale of sequencing data (output data size up to gigabases), including Roche 454, Illumina Solexa, and ABI-SOLiD technologies. These technologies differ in their sequencing principles; specifically, Illumina's sequencing by DNA synthesis (Sequencing By Synthesize - SBS), Roche 454's sequencing by pyrosequencing, and ABI SOLiD's sequencing by oligonucleotide ligation (Sequencing by ligation - SBL) [23-26]. Notably, although Roche's 454 was the first commercial NGS platform appeared on the market, it is no longer available, which is indicative of rapid advancement of the field.

Third generation sequencing (or next-NGS) was developed with the purpose of making sequencing cheaper than second-generation sequencing. Third generation sequencers utilize technologies that interrogate single molecules of DNA without amplifying them through PCR, thereby overcoming problems of PCR amplification biases and de-phasing. These sequencers include Helicos Helioscope (Helicos) based on single molecule sequencing [27], which went bankrupt in 2012, and Pacific Bioscience (PacBio), a single molecular real-time (SMRT) instrument [28, 29].

The Ion Torrent, a smaller scale sequencer, could be placed between the second and third generation as it is not a single-molecule sequencing technique and the sequencing detection is not based on fluorescence signal. Semiconductor is the sequencing basis for Ion Torrent, allowing the detection of protons (H^+) generated by enzymatic reactions [30].

Complete Genomics, Oxford Nanopore, and Plonator use different sequencing principles and chemistries, which do not belong to second or third generation and could be placed under fourth generation. No matter what the sequencing chemistries and different company platforms are, these technologies share the same principle, which is to simultaneously sequence an enormous amount of separated genomic regions [7, 29, 31] (Table 1).

Table 1. Comparison of important NGS platforms.

	Roche 454 GS FLX Plus	Illumina Solexa HiSeq200	ABI SOLiD 5500xl	Ion Torrent	Pacific Bio	Helicos
Sequencing methods	Pyrosequencing	Reversible Dye Terminators	Sequencing by ligation	H^+ Detection	ZMW-Single molecule	Heliscope-Single molecule
Read lengths	700-1,000 bp	2 x100 bp	75 bp	200 bp	Average 1,200, up to 3 kb	25-55 (average 35 bp)
Sequencing runtime	23 hrs	11-14 days	6-7 days	2 hrs	Less than a day	Less than a day
Data generated	700 MB/run	600 GB/run	10-15 GB/day	10-1,000 MB (depends upon chip used)	70-140 MB/ cell	21-35 GB/run
Advantages	- Longer read length - Small data files	Huge data	High-quality data	- Low cost - Very fast	- Longer read than 454 - Fast	- Big data among single molecule synthesis
Concerns	- Less data - Homopolymer	- Short reads - Dephasing - Long run time	- Short reads - Long run time	- Less data - Small read	- Random indel errors	- Small reads - Higher raw error rate

Regardless of sequencing platforms and principles, the application of NGS to research and clinical diagnosis comes in several different scales, which are based on coverage depth. In most NGS experiments, the genome (either the whole genome or targeted "panel" of genes) is fragmented into short fragments of a few hundred base pairs. These fragments are individually read and aligned to generate longer contiguous sequences computationally. In order to get significant redundancy, each individual nucleotide needs to be read several times. The number of times that a given nucleotide in the genome has been read in an experiment is indicated as sequencing depth (also known as read depth). Regarding coverage, there are two concepts that need to be clarified. First, the "breath of coverage" concept is often understood as a measure of what proportion of the total intended genome is represented in the data set. Second, the "depth of coverage" concept can be used to describe the average raw or aligned read depth. The coverage depth varies, depending on the size of the targeted region and the application goal. Shifting the focus from a single large gene to a group of genes, to the whole exome (~20,000 genes), and ultimately, to the whole genome, increases the complexity but decrease the read depth coverage and the ability to call copy number variations (CNVs). When designing an NGS experiment to investigate a clinical question or questions, understanding of depth and coverage concepts can help in tailoring the experimental design and bioinformatics tools to obtain the most meaningful data.

Currently, Illumina NGS platforms are the most commonly used tools for NGS-based basic and clinical research of genetic diseases. Therefore, in this review, we will focus on Illumina NGS system for our discussion on the applications NGS in research and the molecular diagnosis of human genetic diseases. The typical Illumina sequencing workflow from sample collection to NGS analysis contains several steps (Fig. 2): DNA extraction, DNA fragmentation, target sequence enrichment, library construction and sample indexing, loading onto the sequencer for cluster generation and sequencing. The sequence images are subsequently converted to base calls followed by filtering for high-quality base calls, sequence alignment, data analysis and variant calling, and finally interpretation and reporting.

In clinical NGS, quality control procedures must be incorporated to monitor the performance of each step and to ensure that the final results are accurately and appropriately interpreted according to each patient's clinical presentation. The sequence analyses consist of three major steps. The

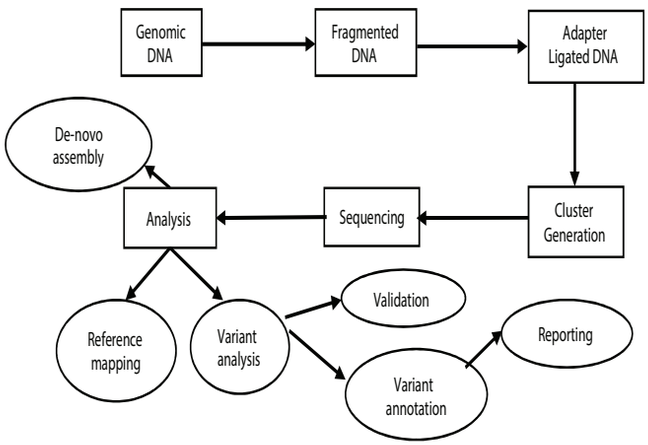


Fig. 2. Basic scheme of a next generation experiment using an Illumina sequencing platform.

primary analysis involves the image capture, the conversion of the image to base calls, and the assignment of quality scores to base calls. The secondary analysis is the filtering of reads based on quality followed by alignment and/or assembly of the reads. Finally, the tertiary analysis involves variant calls based on a reference sequence, variant annotation, data interpretation, and result reporting. Quality control at each step is required because an NGS experiment often involves a large number of samples, a complex workflow and bioinformatic pipeline, and a high reagent cost.

NGS is being applied to identify (causal) genetic variants associated with a genetic disease or phenotype under many different methods such as whole-genome sequencing (WGS), whole-exome sequencing (WES), methylome sequencing, transcriptome sequencing, and targeted sequencing (Fig. 3). While WGS allows sequencing of the entire patients' genomes, WES focuses on the coding regions (exons) of a genome, which take up about 2% of the human genome. However, WES is not suitable to identify most CNVs and other structural modifications. Besides DNA, NGS can also be applied to determine levels of gene expression (transcriptome sequencing or RNA-Seq), splice variants, gene fusions, genomic rearrangements, allele-specific expression, posttranscriptional modifications, microRNAs, small and long noncoding RNAs. Methylome sequencing focuses on DNA methylation. Finally, targeted sequencing, focusing on a selection of genes of interest for a specific disease, is a great choice regarding time and cost for clinical applications of NGS.

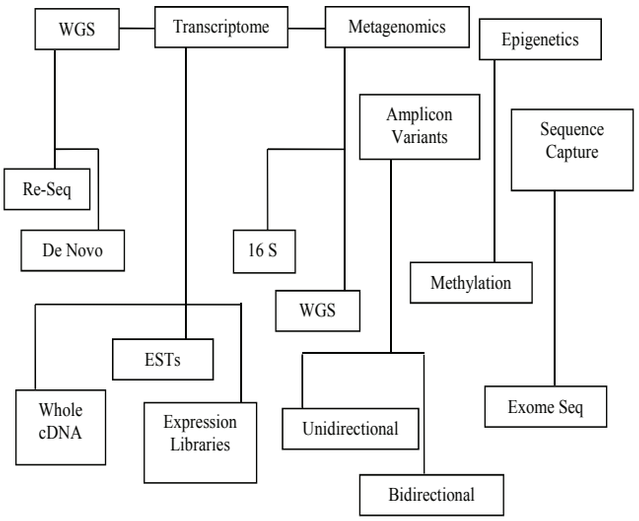


Fig. 3. Schematic diagram of different NGS applications and sequencing methods [22].

Advantages of NGS technologies

Comparing with the first generation sequencing method, NGS technology has the most apparent advantage, which is the capability to massively parallel sequence the genome to obtain high throughput output (up to millions of DNA fragments) on each run. The capability of MPS of NGS has constantly been improved by the development of both sequencing technologies and wet bench. Specifically, the revolutionary development of clonal DNA fragment amplification techniques and the sequence reading technologies together with the improvements in the wet bench portion such as target capture methods enable NGS to sequence the whole genome or particular areas of interest with deep coverage (Table 2).

In the context of clinical applications, NGS technologies offer a number of advantages over the traditional sequencing methods as shown in Table 2. Specifically,

(1) MPS allows sequencing of a group of biomarkers from multiple samples in each run. It is often desirable to simultaneously process many samples to minimize waiting time for results for patients.

(2) Each patient can be simultaneously screened for various genomic aberrations such as single nucleotide and multi-nucleotide polymorphisms (SNPs), insertions and deletions (indels), copy number variations (CNVs), and gene/transcript fusions. Simultaneous screening enables consolidating multiple tests into one MPS run; therefore, lowering the overall healthcare costs and patient sample requirement as compared to the low- and medium-throughput tests.

(3) NGS provides high sequencing depth and coverage for DNA fragments of interest (over 100X), offering sensitive detection (limit-of-detection) and a high level of confidence.

(4) In particular, it is possible to relatively quantitate the allelic fraction of a mutation by estimating the number of DNA strands harboring the genetic alterations and abnormalities among the total sequencing reads, leading to better understanding of the pathogenicity of the tested sample.

Application of NGS

Cancer

As cancer is a genetic disease caused by heritable or somatic mutations, application of NGS has revolutionized

Table 2. Comparison of clinical sequencing technologies [32].

Sequencing technology	DNA input	Throughput of Sequencing output	Multiplexing ability	Types of detected mutations	Workflow and interpretation complexity	Quantitative
Sanger	High	Low	None	Point mutations, Insertions and deletions	Low/ Intermediate	No
Pyrosequencing	Intermediate	Low/ Intermediate	None	Point mutations	Intermediate	Yes
NGS	Low	High	Yes (high)	Point mutations, insertions, deletions, gene expression, fusion, and copy number variations	High	Yes

the field. Cancer research has traditionally been complicated by the fact that there is no clear-cut mechanism for all types of cancers; therefore, there is an urgent need to analyze a large number of genetic variations in the human genome that could be responsible for cancer phenotypes. A large number of cancer cases need to be compared to healthy individuals regarding their genetic make-up, particularly focusing on several genetic targets. This area has been substantially aided by the application of NGS, as many genomes can be sequenced in a short amount of time. In the context of bench-to-bedside applications, NGS has contributed greatly to commercially available gene panels for cancer screening, diagnosis, prognosis and pharmacogenesis. For instances, Extended RAS Panel, an FDA-approved NGS kit, helps clinicians identify colorectal cancer patients eligible for Vectibix treatment [33]. Vectibix is the first FDA-approved monoclonal anti-epidermal growth factor receptor antibody for first-line treatment for patients with wild-type RAS metastatic colorectal cancer (mCRC). NGS targeted panel approach enables simultaneous interrogation of 56 variants across the K-RAS and N-RAS genes to determine the mutant status of RAS genes in a single test. Data generated by the NGS RAS Panel help identify mCRC patients with wild type RAS genes who will be treated with Vectibix. The Extended RAS Panel highlights the importance of NGS-based biomarker screening in therapeutic decision-making in cancer treatment planning.

NGS is empowering the worldwide collaborations for cataloging the mutations and genomic landscapes in multiple cancer types such as The Cancer Genome Atlas (TCGA) [34] and the International Cancer Genome Consortium (ICGC) [35]. These large-scale projects aimed at generating high-quality genomic sequences for a large number of tumors from various types and subtypes of cancer. The massive amount of data generated by TCGA and the ICGC will help us refine cancer classification systems and interrogate the interplay between DNA alternations, RNA expression changes, and epigenomic landscapes in order to gain a comprehensive overall picture of cancer genomics, thereby assisting in discoveries related to diagnostics, prognostics, and therapeutics. For instance, WES and WGS studies have identified new high- and moderate-risk genes in different types of cancers, such as the pancreatic cancer susceptibility genes PALB2 and ATM [36], and the hereditary colorectal cancer moderate-risk genes POLD1 and POLE [37].

In addition to DNA sequencing, RNA sequencing is also used to sequence non-coding RNAs like microRNAs (miRNAs) and long non-coding RNAs (lncRNAs), which have significant functions in cancer pathogenesis and have been demonstrated to be ubiquitously dysregulated in tumorigenesis [38]. Also, epigenetic modifications, particularly DNA methylation, are well-documented and well-studied in some cancers [31]. For example, using DNA methylation and miRNA profiles, a recent study reports that DNA methylation contributes to deregulation of 12 cancer-associated miRNAs and breast cancer progression [39]. The authors also found a strong association between hypermethylation of MIR-127 and MIR-125b-1 and breast cancer progression, particularly metastasis, and concluded that MIR-127 and MIR-125b-1 hypermethylation could be potential biomarkers of breast cancer metastasis [39].

Many NGS-based studies have been conducted to identify novel genetic alterations leading to oncogenesis, metastasis and cancer progression and to survey tumor complexity and heterogeneity [34]. These efforts have provided significant achievements for many diseases such as melanoma, acute myeloid leukemia, breast, lung, liver, kidney, ovarian, colorectal, head and neck cancers [34].

In the past few years, NGS technology has been applied to provide a comprehensive molecular diagnosis of cancers [40]. NGS technology enables the simultaneous sequencing of a large number of target genes and provides early detection and diagnostic markers to develop NGS-based cancer molecular diagnosis [41-43].

WES is currently the most commonly used in clinical diagnostics because it covers more than 95% of the exons, which contain 85% of disease-causing mutations [44]. Moreover, WES has also been applied for determining somatic mutations in tumors [44]. WGS can be utilized to monitor cancer progression, treatment efficacy, and the molecular mechanisms underlying resistance development. However, WGS is expensive and computationally burdensome because of the enormous amount of output data. Indeed, targeted cancer panels are currently most commonly used as diagnostic and prognostic tools in clinical oncology due to their advantages such as low cost and relatively simple interpretation [45, 46].

Breast cancer (BC) is a good example of the application of NGS as an effective method to increase the detection rate of high-risk cases [47]. Previous studies have shown that BRCA1 and BRCA2 mutations cause about 30% of BC cases. A genetic test using BRCA1 and BRCA2 mutations has been recommended; however, mutations in other genes

such as ATM, CHEK2, PALB2, and TP53, have also been shown to confer high BC risk [43]. Therefore, a multiple-gene sequencing panel was developed using NGS, which contained 68 genes including BRCA1, BRCA2, ATM, and TP53. The genes in this panel had cancer risk association for patients with early-onset or familial breast cancer. Currently, the approach of targeted sequencing holds great potential for the rapid diagnosis of not only breast cancer but also other kinds of cancers.

Mendelian and rare diseases

The mendelian or monogenic disease is caused by a mutation at a single gene locus. The location of a single gene could be on an autosome or a sex chromosome, and its inheritance could be either in a dominant or a recessive or an X-linked fashion. There are a number of reports for the use of NGS in identifying the causal variants and in the diagnosis of genetic disorders.

Miller syndrome is the first rare Mendelian disorder whose causal mutations were identified by WES. This syndrome mainly affects the development of the face and limbs. The authors described dihydroorotate dehydrogenase (*DHODH*) mutations in 3 affected families following filtering against public single nucleotide polymorphism (SNP) databases and eight haplotype map (HapMap) exomes [48] U. To date, WES and WGS have identified over 100 genes responsible for various Mendelian diseases; some examples are listed in Table 3.

In the last few years, technological advancements in NGS, especially target enrichment methods, have led to the identification of genetic variations responsible for more than 40 rare disorders. NGS facilitates researchers with the required capacity to analyze large panels of genes for suspected genetic diseases. These diseases vary from single gene disorders such as Neurofibromatosis Type 1 (NF1), Marfan syndrome (MFS), and spastic paraplegia [50, 53, 54] to diseases caused by a group of related genes such as hypertrophic cardiomyopathy and congenital disorders of glycosylation (CDG) [19, 20, 51]. NGS has also been applied to multi-gene disorders including X-linked intellectual disability (XLID) [18] and retinitis pigmentosa [52], as well as defined disorders without identified genetic causes [55-57] (Table 4).

Cystic fibrosis (CF) was the first disease for which the FDA approved an NGS assay for in vitro diagnostic use [62]. It is a Mendelian autosomal recessive disorder that affects the lungs and digestive system of about 70,000 people worldwide. There is no way to prevent CF; therefore, the best defense against this disease is early diagnosis. NGS offers a complete, accurate, and comprehensive interrogation into the whole cystic fibrosis gene for increased clarity in molecular CF testing. NGS-based CF molecular tests enable earlier detection in affected individuals and selection of optimized therapies. Besides diagnosis, NGS-based CF molecular tests can be applied for population screening to determine CF carrier status, newborn screening for CF, and

Table 3. Several publications on the application of WES and WGS to clinical practice [49].

Gene/disease	Enrichment method	Sequencing platform/ chemistry	Sample	Average coverage	References
Miller syndrome (WES)	Agilent array-based capture	Genomic Analyzer (GAII)/76 base read, Single-End (SBS)	Three kindreds	40X	[50]
Kabuki syndrome (WES)	Agilent array-based capture	GAII/Single End or Pair End (SBS)	Ten unknown	40X	[51]
Inflammatory Bowel disease (WES)	NimbleGen exome array-based capture	GS-FLX (SBS, pyrosequencing)	1 patient	34X	[52]
Charcot-Marie-Tooth (WGS)	Direct genomic DNA	SOLiD (SBL)	Family members	30X	[9]
Dopa-responsive dystonia (WGS)	Direct genomic DNA	SOLiD4 (SBL)	Twins and family members	30X	[11]

Table 4. Examples of publications on the application of NGS targeted sequencing to clinical practice [49].

Disease	Gene	Enrichment method	Sequencing platform/ chemistry	Sample	Average coverage	References
Neurofibromatosis Type I (autosomal dominant)	NF1 is a large gene with many exons	NimbleGen oligo array capture	GS-FLX (SBS, pyrosequencing)	2 known	>30X	[53]
Marfan Syndrome (autosomal dominant)	FBN1 is a large gene with many exons	Multiplex PCR	GS-FLX (SBS, pyrosequencing)	5 known 87 unknown	~174X	[54]
Hereditary Spastic Paralegias (HSP: A group of inherited neurodegenerative disorders)	SPG5 and SPG7 genes are involved in the autosomal recessive form of HSP	Fluidigm	GS-FLX (SBS, pyrosequencing)	187 patients	72X for run 1 25X for run 2	[50]
Dilated Cardiomyopathy (DCM) (a group of genetically heterogeneous disorders)	Panel of 19 genes known to cause DCM	Pooled PCR amplicons	GAII (SBS)	5 known	~50X	[19]
Congenital Disorder of Glycosylation (CDG) (a group of diseases caused by over 30 genes involved in the N-linked glycosylation)	Panel of 24 genes known to cause CDG	Fluidigm Raindance	SOLiD version 3/50 base read, SE (SBL)	12 known	616X (FD) 455X (RD)	[36]
Retinitis Pigmentosa (RP) (a group of diseases caused by over 40 known genes)	Panel of 45 genes known to cause RP	NimbleGen oligo array capture	GAII/32 base read, SE (SBS)	2 known 3 unknown	486X (1 sample per lane) 98X (4 samples per lane)	[58]
X-Linked Intellectual Disability (XLMR) (a group of genetically heterogeneous disorders)	Panel of 86 genes known to cause XLMR	Raindance	GAII (SBS)	3 known 21 unknown	Coverage per base ranging from 92X to 445X	[18]
Mitochondrial diseases	Mitochondrial DNA (mtDNA)	2 overlapping PCR fragments	GAII (SBS)	2 known	~1,785x	[59]
Mitochondrial diseases	Panel of 362 nuclear genes are known to involve in mitochondrial diseases.	Agilent array based capture	GAII/36 base read, SE (SBS)	2 patients 1 normal	37X-51X for nuclear genes, 3,000-5,000X for mtDNA	[21]
Human Leukocyte Antigen (HLA) genotyping	HLA genes	HLA gene amplification	MiSeq/250 base read, PE (SBS)	211 known 79 unknown	>67X	[60]
Ataxias	A panel of 58 genes known to cause human ataxia	Agilent SureSelect targeted capture	Illumina/51 base read, PE (SBS)	50 patients	94% of regions of interest > 5X	[61]

genetic counseling regarding couples' reproductive risks and family planning options.

Pre-natal diagnosis

Traditionally, molecular prenatal diagnosis requires invasive methods to draw an amniocentesis or chorionic villus sample and detect chromosomal abnormalities. Besides cost, these procedures pose a miscarriage risk at an approximate rate of 0.5%. Therefore, it is highly desirable to develop a non-invasive method for prenatal diagnosis to avoid the risk of fetal loss.

One of the most valuable applications of NGS technology is molecular genetic testing in pre-diagnostics. The pioneering work of Denis Lo and his coworkers at The Chinese University of Hong Kong [63] demonstrated that more than 10% of a mother's cell-free DNA is from the fetal genome at the end of the first trimester. Recently, there has been rapid progress in applying NGS technology to the detection of fetal chromosomal abnormality in fetal DNA from cell-free DNA fragments in maternal plasma.

In 2011, three large-scale studies involving multiple centers established non-invasive prenatal tests (NIPTs) that have had a significant impact on prenatal care [64-66]. These studies showed that the detection of fetal trisomy 21 could be performed at nearly 100% sensitivity and 98% specificity by multiplexed MPS of maternal plasma DNA. Since its introduction in 2011, NIPT has been standardized as a recommended test for high-risk pregnancies [1]. NIPT has also evolved from exclusively trisomy 21 testing to include trisomy 18, trisomy 13, sex chromosome aneuploidies, and microdeletions. In 2016, one clinical validation study demonstrated that genome-wide NIPT could provide high resolution, sensitive, and specific detection of a wide range of fetal subchromosomal and whole chromosomal abnormalities that were previously only pinpointed by invasive karyotyping testing [67]. In some cases, this NIPT also provided further information about the origin of genetic material that had not been identified by the invasive karyotype method. Therefore, the implementation of the NGS-based prenatal screening of fetal chromosome abnormalities using circulating cell-free nucleic acid in the maternal blood is one of the great advancements in providing effective and safe prenatal diagnostics.

Besides screening for chromosomal abnormalities, NGS technology can also be applied to the prenatal mutation detection of genetic disorders. A proof-of-concept NGS-based study to detect fetal β-thalassemia mutations using maternal blood demonstrated the possibility of investigating specific genetic disease loci [68]. In this study, NGS enables sequencing of fetal DNA fragments that could subsequently be assembled into a complete fetal genomic map with the parental genomes as guides. Then, the fetal genome could then be screened for mutations prenatally and noninvasively. This approach was applied to identify whether the fetus carries β-thalassemia mutations in the case study of a family where the pregnant mother carried one gene mutation, and the father carried a different mutation for the blood disease β-thalassemia. From the maternal plasma DNA sequencing data, they found that the fetus inherited the paternal mutation. Then, they used relative haplotype dosage analysis to test if the fetus had inherited the genomic region that contained the maternal mutation. The authors found that the fetus had not inherited the maternal mutation; therefore, the fetus was a heterozygous carrier for β-thalassemia. This is one of the pioneering studies showing that sequencing of maternal plasma cell-free DNA provides noninvasive prenatal genome-wide scanning for genetic disorders [68].

The current global status of using NGS in disease management

The rapid development of MPS has opened up the opportunities to turn scientific discoveries about DNA and the way it works into a promising life-saving reality for patients worldwide. A clear example of this global influence of NGS in disease management is the launching of several massively sequencing projects for precision medicine in developed countries. These projects are 100K Genome project in the UK, Precision Medicine Initiative in USA, Japan, India, and Middle East, all aiming at bringing the benefits of genomics to patients. Precision medicine is an emerging approach for disease treatment and prevention, in which health care and medical decisions, practices, and products should be individually tailored to each patient's variability in genes, environment, and lifestyle.

Currently, cancer and rare diseases, which are strongly linked to changes in the genome, are the primary focus for precision medicine. In case of cancer, DNA from the tumor and DNA from the patient's healthy cells, thanks to NGS, can be sequenced and compared; the precise gene changes

are detected. Understanding these genomic alterations is crucial to predict how well a person will respond to a particular treatment such as radiotherapy, or indicate which treatment will be the best for individual patients. An excellent example in use already is to prescribe the medicine tailored to a woman's breast cancer genotyping. Herceptin will be effective for a woman with HER2 positive but not for the one who is HER2 negative. Additionally, genomics coupled with NGS can also be used to track infectious disease, precisely pinpointing the origin and nature of the outbreak through examining the whole genomes of infectious agents.

Future perspectives

The advent of NGS has opened up many new frontiers and it, in the future, will continue to play a crucial role in the research and molecular diagnostics of genetic diseases. NGS will keep providing novel insights into disease mechanisms, metabolics, and signaling pathways at a resolution never previously possible. Information obtained from NGS is being used to improve diagnostics, and to develop more effective and more personalized treatments for disease and patient care. Furthermore, targeted NGS will still hold great potential for speed and cost-effectiveness of sequencing by focusing on the portions of the genome that are relevant to the question of the study. It is also beneficial in identifying and developing panels for biomarkers associated with a particular type of health condition.

NGS technologies are capable of helping scientists and clinicians study genomes of individuals faster than ever before, opening up the new era of Personalized Medicine. Every individual is different in their genetic make-up and is susceptible to different diseases, infections, and disorders. Therefore, knowing one's genomic sequence will help determine accurate and proper care and will elucidate increased risks for hereditary diseases. With decreasing costs and rapid developments of NGS technologies, it is easy to envision that all patients will soon have their genomes sequenced when they visit their doctors. The information generated by NGS can provide information on the different types of disease-causing alterations in individuals in the short turn-around time required to screen patients either for clinical trials, or for diagnostics in clinical settings.

In the future, sequencing of individual genomes of interest under different living, nutritional, or treatment conditions will benefit the medical communities by guiding disease control, progression, and prevention, and rational usage of molecularly guided treatments. These discoveries will ultimately bring a better understanding of disease pathogenesis, contributing to a new era of molecular pathology and personalized medicine. With the knowledge of precision medicine, we can increase treatment efficacy, reduce toxicity, and therefore decrease disease burden for both patients and society.

Finally, to better understand the genetic etiology of diseases, to improve effective molecular diagnosis, and to apply genetic information in precision medicine and personalized medicine, it will be critical, in the long run, to combine the NGS results with genome-wide association studies (GWASs) as well as gene-gene and gene-environment interactions.

Challenges

In pursuing NGS-based research and implementing NGS technologies in clinical applications, many hurdles may be encountered that need to be resolved.

First of all, enormous amounts of data that must be properly managed, stored, and analyzed are the obvious challenge posed by NGS. As the reagent costs of sequencing decrease with the development of better reagents and improved protocols, the number of sequencing projects is continuously increasing, the complexity of data analysis and management appear to be the primary limiting factor among researchers. Specifically, computing skills, hardware, storage, and network capabilities are necessary and critical to managing the massive data sets generated by large-scale NGS studies. Also, rapidly developing technology constantly require upgrades of analytic software and bioinformatics pipelines, which is costly and warrants revalidation before implementation.

Second, the complexity of NGS caused by the large size of the genome tested in multiple barcoded samples leads to the challenge of data validation. Thorough validation of the tests must be performed to implement NGS as a routine diagnostic test, as the majority of the NGS assays is intended for research only. NGS is an iterative process, which is the major problem in validating the performance characteristics of a clinical test for accuracy and reproducibility. Validation of a NGS tests involves optimizing simultaneously the

sequencing platform, the specific test/panel of genes, and the bioinformatics pipeline. Therefore, after altering one small part of the workflow such as changing the input DNA quantity, adding or changing even a single primer, or adjusting a parameter in the bioinformatics pipeline, the entire assay must be revalidated. Given that NGS assays require days to weeks to complete, validating these tests is time-consuming and costly. Moreover, the addition of a new marker or a new gene in the implemented sequencing panels will lead to the revalidation of existing ones or complete redesign and validation.

Finally, the clinical laboratories also have to face the challenges of clinical reporting when performing clinical NGS tests. It is necessary to establish standardized mutation annotation and classification such as germline, somatic, variant of unknown significance, that interface with literature and publicly available mutation databases (COSMIC, dbSNP) to generate a comprehensive report for a patient with disease information. In addition, a large number of mutations can be incidental findings and of unknown significance, posing a challenge for the laboratories to report due to the medicolegal implications. Therefore, laboratories should communicate their policy regarding the reporting of incidental findings and follow them through reporting.

Drawbacks

Although the potential of genomics is enormous, certain disease scenarios will be helped by NGS more than others. As aforementioned above, cancer and rare diseases are the areas that NGS has been widely applied to detect *de novo* mutations or mosaicism in sporadic patients without a prior hypothesis about the causal gene. NGS has also been used to elucidate the genetic causes for Mendelian and heterogeneous disorders. However, there are many diseases that do not very much benefit from NGS such as noninherited metabolic syndrome and nutritional deficiency. Indeed, a phenotype is the holistic outcome of many different layers of molecular regulation and is also influenced by environmental interactions. DNA is only one regulatory layer; many other complex molecular processes occurred at RNA and protein levels also contribute to the disease phenotypes. Therefore, it is an unrealistic hope of trying to relate all human traits with DNA level information, and it is not justified to say that NGS is helpful to all diseases.

Conclusions

NGS is a powerful technology generating enormous amounts of data, which is revolutionizing almost all aspects of research in human health and diseases. Several key areas reviewed here demonstrate that NGS is a practical, attractive, and economically feasible technology for clinical applications. The implementation of NGS approaches to clinical diagnostics has still been an on-going pursuit and requires a thorough understanding of the principles, utilities, challenges, limitations, and interpretation of the results as well as the improvement of the technologies. Therefore, continuous improvement of the accuracy of sequencing chemistries, computational algorithms, bioinformatics analytical tools, and the improved methods for sample processing and variant interpretation will make NGS a new "gold standard" method for clinical laboratories.

ACKNOWLEDGEMENTS

We thank Vinmec Research Institute of Stem Cell and Gene Technology, Vinmec International Hospital, and particularly, the Laboratory of Gene Technology for their support and careful review of the manuscript.

REFERENCES

[1] F. Sanger, S. Nicklen, A.R. Coulson (1977), "DNA sequencing with chain-terminating inhibitors", *Proceedings of the National Academy of Sciences of the United States of America*, **74,** pp.5463-5467.

[2] D.J. Pagliarini, S.E. Calvo, B. Chang, S.A. Sheth, S.B. Vafai, S.E. Ong, G.A. Walford, C. Sugiana, A. Boneh, W.K. Chen, D.E. Hill, M. Vidal, J.G. Evans, D.R. Thorburn, S.A. Carr, V.K. Mootha (2008), "A mitochondrial protein compendium elucidates complex I disease biology", *Cell*, **134**, pp.112-123.

[3] S. Calvo, M. Jain, X.H. Xie, S.A. Sheth, B. Chang, O.A. Goldberger, A. Spinazzola, M. Zeviani, S.A. Carr, V.K. Mootha (2006), "Systematic identification of human mitochondrial disease genes through integrative genomics", *Nature Genetics*, **38**, pp.576-582.

[4] D.R. Thorburn (2004), "Mitochondrial disorders: Prevalence, myths and advances", *Journal of Inherited Metabolic Disease*, **27**, pp.349-362.

[5] C. Scharfe, H.H.S. Lu, J.K. Neuenburg, E.A. Allen, G.C. Li, T. Klopstock, T.M. Cowan, G.M. Enns, R.W. Davis (2009), "Mapping Gene Associations in Human Mitochondria using Clinical Disease Phenotypes", *PLOS Computational Biology*, **5**, e1000374.

[6] J. Wang, H. Yu, V.W. Zhang, X. Tian, Y. Feng, G. Wang, E. Gorman, H. Wang, R.E. Lutz, E.S. Schmitt, S. Peacock, L.J. Wong (2015), "Capture-based high-coverage NGS: a powerful tool to uncover a wide spectrum of mutation types", *Genetics in Medicine*, **18**, p.513.

[7] E.R. Mardis (2008), "Next-generation DNA sequencing methods", *Annual Review of Genomics and Human Genetics*, **9**, pp.387-402.

[8] M.L. Metzker (2010), "Applications of NGS sequencing technologies - the next generation", *Nature Reviews Genetics*, **11**, pp.31-46.

[9] J. Shendure, H. Ji (2008), "Next-generation DNA sequencing", *Nature Biotechnology*, **26**, pp.1135-1145.

[10] R.A. Gibbs (2011), "Bringing genomics and genetics back together", *Science*, **331**, pp.548-548.

[11] F.S. Collins (2011), "Faces of the Genome", *Science*, **331**, pp.546-546.

[12] E.D. Green, M.S. Guyer, I. Natl. Human Genome Res. (2011), "Charting a course for genomic medicine from base pairs to bedside", *Nature*, **470**, pp.204-213.

[13] E.S. Lander (2011), "Initial impact of the sequencing of the human genome", *Nature*, **470**, pp.187-197.

[14] S.B. Ng, D.A. Nickerson, M.J. Bamshad, J. Shendure (2010), "MPS and rare disease", *Human Molecular Genetics*, **19**, pp.R119-R124.

[15] R.D. Morin, M.D. O'Connor, M. Griffith, F. Kuchenbauer, A. Delaney, A.L. Prabhu, Y. Zhao, H. McDonald, T. Zeng, M. Hirst, C.J. Eaves, M.A. Marra (2008), "Application of MPS to microRNA profiling and discovery in human embryonic stem cells", *Genome Research*, **18**, pp.610-621.

[16] R.D. Hawkins, G.C. Hon, B. Ren (2010), "Next-generation genomics: an integrative approach", *Nature Reviews Genetics*, **11**, pp.476-486.

[17] R. Lister, M. Pelizzola, R.H. Dowen, R.D. Hawkins, G. Hon, J. Tonti-Filippini, J.R. Nery, L. Lee, Z. Ye, Q.M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A.H. Millar, J.A. Thomson, B. Ren, J.R. Ecker (2009), "Human DNA methylomes at base resolution show widespread epigenomic differences", *Nature*, **462**, pp.315-322.

[18] H. Hu, K. Wrogemann, V. Kalscheuer, A. Tzschach, H. Richard, S.A. Haas, C. Menzel, M. Bienek, G. Froyen, M. Raynaud, H. Van Bokhoven, J. Chelly, H. Ropers, W. Chen (2009), "Mutation screening in 86 known X-linked mental retardation genes by droplet-based multiplex PCR and massive parallel sequencing", *The HUGO Journal*, **3**, pp.41-49.

[19] S. Gowrisankar, J.P. Lerner-Ellis, S. Cox, E.T. White, M. Manion, K. LeVan, J. Liu, L.M. Farwell, O. Lartchouk, H.L. Rehm, B.H. Funke (2010), "Evaluation of Second-Generation Sequencing of 19 Dilated Cardiomyopathy Genes for Clinical Applications", *Journal of Molecular Diagnostics*, **12**, pp.818-827.

[20] K.V. Voelkerding, S. Dames, J.D. Durtschi (2010), "NGS for Clinical Diagnostics-Principles and Application to Targeted Resequencing for Hypertrophic Cardiomyopathy A Paper from the 2009 William Beaumont Hospital Symposium on Molecular Pathology", *Journal of Molecular Diagnostics*, **12**, pp.539-551.

[21] V. Vasta, S.B. Ng, E.H. Turner, J. Shendure, S.H. Hahn (2009), "Next generation sequence analysis for mitochondrial disorders", *Genome Medicine*, **1**, doi: 10.1186/gm100.

[22] A.K. Gupta, U.D. Gupta (2014), "NGS and its applications A2 - Verma, Ashish S", chapter 19, in A. Singh (Ed.) *Animal Biotechnology*, Academic Press, San Diego, pp.345-367.

[23] M. Margulies, M. Egholm, W.E. Altman, S. Attiya, J.S. Bader, L.A. Bemben, J. Berka, M.S. Braverman, Y.J. Chen, Z.T. Chen, S.B. Dewell, L. Du, J.M. Fierro, X.V. Gomes, B.C. Godwin, W. He, S. Helgesen, C.H. Ho, G.P. Irzyk, S.C. Jando, M.L.I. Alenquer, T.P. Jarvie, K.B. Jirage, J.B. Kim, J.R. Knight, J.R. Lanza, J.H. Leamon, S.M. Lefkowitz, M. Lei, J. Li, K.L. Lohman, H. Lu, V.B. Makhijani, K.E. McDade, M.P. McKenna, E.W. Myers, E. Nickerson, J.R. Nobile, R. Plant, B.P. Puc, M.T. Ronan, G.T. Roth, G.J. Sarkis, J.F. Simons, J.W. Simpson, M. Srinivasan, K.R. Tartaro, A. Tomasz, K.A. Vogt, G.A. Volkmer, S.H. Wang, Y. Wang, M.P. Weiner, P.G. Yu, R.F. Begley, J.M. Rothberg (2005), "Genome sequencing in microfabricated high-density picolitre reactors", *Nature*, **437**, pp.376-380.

[24] S. Bennett (2004), "Solexa Ltd", *Pharmacogenomics*, **5**, pp.433-438.

[25] S.T. Bennett, C. Barnes, A. Cox, L. Davies, C. Brown (2005), "Toward the $1000 human genome", *Pharmacogenomics*, **6**, pp.373-382.

[26] J. Shendure, G.J. Porreca, N.B. Reppas, X.X. Lin, J.P. McCutcheon, A.M. Rosenbaum, M.D. Wang, K. Zhang, R.D. Mitra, G.M. Church (2005), "Accurate multiplex polony sequencing of an evolved bacterial genome", *Science*, **309**, pp.1728-1732.

[27] T.D. Harris, P.R. Buzby, H. Babcock, E. Beer, J. Bowers, I. Braslavsky, M. Causey, J. Colonell, J. DiMeo, J.W. Efcavitch, E. Giladi, J. Gill, J. Healy, M. Jarosz, D. Lapen, K. Moulton, S.R. Quake, K. Steinmann, E. Thayer, A. Tyurina, R. Ward, H. Weiss, Z. Xie (2008), "Single-molecule DNA sequencing of a viral genome", *Science*, **320**, pp.106-109.

[28] J. Korlach, K.P. Bjornson, B.P. Chaudhuri, R.L. Cicero, B.A. Flusberg, J.J. Gray, D. Holden, R. Saxena, J. Wegener, S.W. Turner (2010), "Real-time DNA sequencing from single polymerase molecules", in: N.G. Walter (Ed.) *Methods in Enzymology*, **Vol.472**, Single Molecule Tools, Pt A: Fluorescence Based Approaches, pp.431-455.

[29] E.H. Turner, S.B. Ng, D.A. Nickerson, J. Shendure (2009), "Methods for Genomic Partitioning", *Annual Review of Genomics and Human Genetics*, **10**, pp.263-284.

[30] J.M. Rothberg, W. Hinz, T.M. Rearick, J. Schultz, W. Mileski, M. Davey, J.H. Leamon, K. Johnson, M.J. Milgrew, M. Edwards, J. Hoon, J.F. Simons, D. Marran, J.W. Myers, J.F. Davidson, A. Branting, J.R. Nobile, B.P. Puc, D. Light, T.A. Clark, M. Huber, J.T. Branciforte, I.B. Stoner, S.E. Cawley, M. Lyons, Y. Fu, N. Homer, M. Sedova, X. Miao, B. Reed, J. Sabina, E. Feierstein, M. Schorn, M. Alanjary, E. Dimalanta, D. Dressman, R. Kasinskas, T. Sokolsky, J.A. Fidanza, E. Namsaraev, K.J. McKernan, A. Williams, G.T. Roth, J. Bustillo (2011), "An integrated semiconductor device enabling non-optical genome sequencing", *Nature*, **475**, pp.348-352.

[31] K.V. Voelkerding, S.A. Dames, J.D. Durtschi (2009), "NGS: From Basic Research to Diagnostics", *Clinical Chemistry*, **55**, pp.641-658.

[32] R.R. Singh, R. Luthra, M.J. Routbort, K.P. Patel, L.J. Medeiros (2016), "Implementation of NGS in clinical molecular diagnostic

laboratories: advantages, challenges and potential", *Expert Review of Precision Medicine and Drug Development*, **1**, pp.109-120.

[33] I. Company (2016), *Extended RAS Panel*.

[34] NIH (2017), *The Cancer genome ATLAS*.

[35] International cancer genome consortium (2017).

[36] S. Jones, R.H. Hruban, M. Kamiyama, M. Borges, X. Zhang, D.W. Parsons, J.C.H. Lin, E. Palmisano, K. Brune, E.M. Jaffee, C.A. Iacobuzio-Donahue, A. Maitra, G. Parmigiani, S.E. Kern, V.E. Velculescu, K.W. Kinzler, B. Vogelstein, J.R. Eshleman, M. Goggins, A.P. Klein (2009), "Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene", *Science*, **324**, p.217.

[37] C. Palles, J.B. Cazier, K.M. Howarth, E. Domingo, A.M. Jones, P. Broderick, Z. Kemp, S.L. Spain, E. Guarino, I. Salguero, A. Sherborne, D. Chubb, L.G. Carvajal-Carmona, Y. Ma, K. Kaur, S. Dobbins, E. Barclay, M. Gorman, L. Martin, M.B. Kovac, S. Humphray, C.C. The, W.G.S.C. The, A. Lucassen, C.C. Holmes, D. Bentley, P. Donnelly, J. Taylor, C. Petridis, R. Roylance, E.J. Sawyer, D.J. Kerr, S. Clark, J. Grimes, S.E. Kearsey, H.J.W. Thomas, G. McVean, R.S. Houlston, I. Tomlinson (2012), "Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas", *Nature Genetics*, **45**, p.136.

[38] M. Esteller (2011), "Non-coding RNAs in human disease", *Nature Reviews Genetics*, **12**, p.861.

[39] I.V. Pronina, V.I. Loginov, A.M. Burdennyy, M.V. Fridman, V.N. Senchenko, T.P. Kazubskaya, N.E. Kushlinskii, A.A. Dmitriev, E.A. Braga (2017), "DNA methylation contributes to deregulation of 12 cancer-associated microRNAs and breast cancer progression", *Gene*, **604**, pp.1-8.

[40] J. Gagan, E.M. Van Allen (2015), "NGS to guide cancer therapy", *Genome Medicine*, **7**, p.80.

[41] R. Luthra, H. Chen, S. Roy-Chowdhuri, R.R. Singh (2015), "NGS in clinical molecular diagnostics of cancer: advantages and challenges", *Cancers*, **7**, pp.2023-2036.

[42] V.G. LeBlanc, M.A. Marra (2015), "NGS Approaches in Cancer: Where have they brought us and where will they take us?", *Cancers*, **7**, pp.1925-1958.

[43] R. Kamps, D.R. Brandão, J.B. Bosch, D.A. Paulussen, S. Xanthoulea, J.M. Blok, A. Romano (2017), "NGS in oncology: genetic diagnosis, risk prediction and cancer classification", *International Journal of Molecular Sciences*, **18**, doi: 10.3390/ijms18020308.

[44] B. Rabbani, M. Tekin, N. Mahdieh (2013), "The promise of whole-exome sequencing in medical genetics", *Journal of Human Genetics*, **59**, p.5.

[45] K.A. Johansen Taber, B.D. Dickinson, M. Wilson (2014), "The promise and challenges of next-generation genome sequencing for clinical care", *JAMA Internal Medicine*, **174**, pp.275-280.

[46] T.A. Manolio, R.L. Chisholm, B. Ozenberger, D.M. Roden, M.S. Williams, R. Wilson, D. Bick, E.P. Bottinger, M.H. Brilliant, C. Eng, K.A. Frazer, B. Korf, D.H. Ledbetter, J.R. Lupski, C. Marsh, D. Mrazek, M.F. Murray, P.H. O'Donnell, D.J. Rader, M.V. Relling, A.R. Shuldiner, D. Valle, R. Weinshilboum, E.D. Green, G.S. Ginsburg (2013), "Implementing genomic medicine in the clinic: the future is here", *Genetics in Medicine*, **15**, p.258.

[47] G.H.J.M. Leyten, D. Hessels, F.P. Smit, S.A. Jannink, H. de Jong, W.J.G. Melchers, E.B. Cornel, T.M. de Reijke, H. Vergunst, P. Kil, B.C. Knipscheer, C.A. Hulsbergen-van de Kaa, P.F.A. Mulders, I.M. van Oort, J.A. Schalken (2015), "Identification of a Candidate Gene Panel for the Early Diagnosis of Prostate Cancer", *Clinical Cancer Research*, **21**, p.3061.

[48] B. Rabbani, N. Mahdieh, K. Hosomichi, H. Nakaoka, I. Inoue (2012), "NGS: impact of exome sequencing in characterizing Mendelian disorders", *Journal of Human Genetics*, **57**, pp.621-632.

[49] M. Seifi, A. Ghasemi, S. Raeisi, S. Heidarzadeh (2017), "Application of NGS in Clinical Molecular Diagnostics", *Brazilian Archives of Biology and Technology*, ***60***, http://dx.doi.org/10.1590/1678-4324-2017160414.

[50] N.A. Schlipf, R. Schuele, S. Klimpe, K.N. Karle, M. Synofzik, J. Schicks, O. Riess, L. Schoels, P. Bauer (2011), "Amplicon-based high-throughput pooled sequencing identifies mutations in CYP7B1 and SPG7 in sporadic spastic paraplegia patients", *Clinical Genetics*, **80**, pp.148-160.

[51] M.A. Jones, S. Bhide, E. Chin, B.G. Ng, D. Rhodenizer, V.W. Zhang, J.J. Sun, A. Tanner, H.H. Freeze, M.R. Hegde (2011), "Targeted polymerase chain reaction-based enrichment and NGS for diagnostic testing of congenital disorders of glycosylation", *Genetics in Medicine*, **13**, pp.921-932.

[52] D.A. Simpson, G.R. Clark, S. Alexander, G. Silvestri, C.E. Willoughby (2011), "Molecular diagnosis for heterogeneous genetic diseases with targeted high-throughput DNA sequencing applied to retinitis pigmentosa", *Journal of Medical Genetics*, **48**, p.145.

[53] L.S. Chou, C.S.J. Liu, B. Boese, X. Zhang, R. Mao (2010), "DNA Sequence Capture and Enrichment by Microarray Followed by NGS for Targeted Resequencing: Neurofibromatosis Type 1 Gene as a Model", *Clinical Chemistry*, **56**, pp.62-72.

[54] M. Baetens, L. Van Laer, K. De Leeneer, J. Hellemans, J. De Schrijver, H. Van De Voorde, M. Renard, H. Dietz, R.V. Lacro, B. Menten, W. Van Criekinge, J. De Backer, A. De Paepe, B. Loeys, P.J. Coucke (2011), "Applying Massive Parallel Sequencing to Molecular Diagnosis of Marfan and Loeys-Dietz Syndromes", *Human Mutation*, **32**, pp.1053-1062.

[55] J.R. Lupski, J.G. Reid, C. Gonzaga-Jauregui, D. Rio Deiros, D.C.Y. Chen, L. Nazareth, M. Bainbridge, H. Dinh, C. Jing, D.A. Wheeler, A.L. McGuire, F. Zhang, P. Stankiewicz, J.J. Halperin, C. Yang, C. Gehman, D. Guo, R.K. Irikat, W. Tom, N.J. Fantin, D.M. Muzny, R.A. Gibbs (2010), "Whole-Genome Sequencing in a Patient with Charcot-Marie-Tooth Neuropathy", *New England Journal of Medicine*, **362**, pp.1181-1191.

[56] S.B. Ng, K.J. Buckingham, C. Lee, A.W. Bigham, H.K. Tabor, K.M. Dent, C.D. Huff, P.T. Shannon, E.W. Jabs, D.A. Nickerson, J. Shendure, M.J. Bamshad (2009), "Exome sequencing identifies the cause of a mendelian disorder", *Nature Genetics*, **42**, p.30.

[57] S.B. Ng, A.W. Bigham, K.J. Buckingham, M.C. Hannibal, M.J. McMillin, H.I. Gildersleeve, A.E. Beck, H.K. Tabor, G.M. Cooper, H.C. Mefford, C. Lee, E.H. Turner, J.D. Smith, M.J. Rieder, K.I. Yoshiura, N. Matsumoto, T. Ohta, N. Niikawa, D.A. Nickerson, M.J. Bamshad, J. Shendure (2010), "Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome", *Nature Genetics*, **42**, p.790.

[58] D.A. Simpson, G.R. Clark, S. Alexander, G. Silvestri, C.E. Willoughby (2011), "Molecular diagnosis for heterogeneous genetic diseases with targeted high-throughput DNA sequencing applied to retinitis pigmentosa", *Journal of Medical Genetics*, **48**, pp.145-151.

[59] S. Tang, T. Huang (2010), "Characterization of mitochondrial DNA heteroplasmy using a parallel sequencing system", *Biotechniques*, **48**, pp.287-292.

[60] E.T. Weimer, M. Montgomery, R. Petraroia, J. Crawford, J.L. Schmitz "Performance Characteristics and Validation of NGS for Human Leucocyte Antigen Typing", *The Journal of Molecular Diagnostics*, **18**, pp.668-675.

[61] A.H. Németh, A.C. Kwasniewska, S. Lise, R. Parolin Schnekenberg, E.B.E. Becker, K.D. Bera, M.E. Shanks, L. Gregory, D. Buck, M. Zameel Cader, K. Talbot, R. de Silva, N. Fletcher, R. Hastings, S. Jayawant, P.J. Morrison, P. Worth, M. Taylor, J. Tolmie, M. O'Regan, R. Valentine, E. Packham, J. Evans, A. Seller, J. Ragoussis (2013), "NGS for molecular diagnosis of neurological disorders using ataxias as a model", *Brain*, **136**, pp.3106-3118.

[62] I. Company (2012), *Cystic Fibrosis*.

[63] Y.M.D. Lo, N. Corbetta, P.F. Chamberlain, V. Rai, I.L. Sargent, C.W.G. Redman, J.S. Wainscoat (1997), "Presence of fetal DNA in maternal plasma and serum", *Lancet*, **350**, pp.485-487.

[64] R.W.K. Chiu, H. Sun, R. Akolekar, C. Clouser, C. Lee, K. McKernan, D. Zhou, K.H. Nicolaides, Y.M.D. Lo (2010), "Maternal Plasma DNA Analysis with MPS by Ligation for Noninvasive Prenatal Diagnosis of Trisomy 21", *Clinical Chemistry*, **56**, pp.459-463.

[65] M. Ehrich, C. Deciu, T. Zwiefelhofer, J.A. Tynan, L. Cagasan, R. Tim, V. Lu, R. McCullough, E. McCarthy, A.O.H. Nygren, J. Dean, L. Tang, D. Hutchison, T. Lu, H. Wang, V. Angkachatchai, P. Oeth, C.R. Cantor, A. Bombard, D. van den Boom (2011), "Noninvasive detection of fetal trisomy 21 by sequencing of DNA in maternal blood: a study in a clinical setting", *American Journal of Obstetrics and Gynecology*, **204**, pp.205e1-205e11.

[66] G.E. Palomaki, E.M. Kloza, G.M. Lambert-Messerlian, J.E. Haddow, L.M. Neveux, M. Ehrich, D. van den Boom, A.T. Bombard, C. Deciu, W.W. Grody, S.F. Nelson, J.A. Canick (2011), "DNA sequencing of maternal plasma to detect Down syndrome: An international clinical validation study", *Genetics in Medicine*, **13**, pp.913-920.

[67] R.B. Lefkowitz, J.A. Tynan, T. Liu, Y. Wu, A.R. Mazloom, E. Almasri, G. Hogg, V. Angkachatchai, C. Zhao, D.S. Grosu, G. McLennan, M. Ehrich (2016), "Clinical validation of a noninvasive prenatal test for genomewide detection of fetal copy number variants", *American Journal of Obstetrics & Gynecology*, **215**, pp.227.e1-227.e16.

[68] Y.M.D. Lo, K.C.A. Chan, H. Sun, E.Z. Chen, P. Jiang, F.M.F. Lun, Y.W. Zheng, T.Y. Leung, T.K. Lau, C.R. Cantor, R.W.K. Chiu (2010), "Maternal Plasma DNA Sequencing Reveals the Genome-Wide Genetic and Mutational Profile of the Fetus", *Science Translational Medicine*, **2**, doi: 10.1126/scitranslmed.3001720.

Constructing a molecular genotyping assay for rs11077 based on real-time polymerase chain reaction high resolution melting (PCR HRM) technique for the prognosis of hepatocellular carcinoma (HCC) patients

Thi Hao Pham[1], Thanh Huy Nguyen[1], Minh Phung Truong[1], Thi Thuy Hang Le[2], Quoc Dang Quan[3], Quang Tri Le[4], Hoang Chuong Nguyen[5*]

[1]Center for Research and Application in Bioscience
[2]Ho Chi Minh city University of Food Industry
[3]Center of Science and Technology Development, Ho Chi Minh Communist Youth Union
[4]7A Military Hospital
[5]University of Science, Vietnam National University Ho Chi Minh city

Received 5 March 2018; accepted 19 May 2018

Abstract:

***XPO5* codes for the nuclear transport factor exportin-5, which is a membrane-bound protein. This gene is responsible for the transport of pre-miRNA from the nucleus to the cytoplasmic compartments, thereby adjusting the whole miRNA expression level. The reduction of the miRNA levels was recorded when *XPO5* was knocked down. rs11077 is found in the 3′UTR region of *XPO5*, and this SNP might affect mRNA stability and be associated with the altered expression of *XPO5*. This leads to the universal suppression of miRNA expression profiles, thereby mediating the HCC survival. HCC patients bearing C/C and A/C genotypes of rs11077 had a survival rate of 60% after 3 years; and this rate was reduced to 24.7% with HCC patients bearing the A/A genotype. In this study, we constructed a molecular assay based on a real-time PCR HRM technique for rs11077 genotyping. We successfully designed the primer pair for the real-time PCR HRM of rs11077. We also found the optimal concentration of $MgCl_2$ to arrive at a clear differentiation of the three genotypes of rs11077. Thereafter, we characterised the analytical specificity and the precisions of the molecular assay. The SNP genotyping results were compared between the real-time PCR HRM and nucleotide sequencing. Finally, we evaluated the molecular assay on 123 human blood samples. The rs11077 genotyping assay in this study could be used for the prognosis of HCC patients.**

Keywords: ***hepatocellular carcinoma, real-time PCR HRM, rs11077, SNP.***

Classification numbers: ***3.2, 3.5***

Introduction

Hepatocellular carcinoma (HCC) is the most common type of primary liver cancer, and is also the 5th most common type of cancer in the world. Worldwide, more than 700,000 cases are diagnosed each year; and the high mortality rates make HCC the third leading cause of cancer deaths in the world, following lung cancer and stomach cancer. The incidence of HCC varies across the geographical regions of the world, as it relates to the frequency difference of risk factors, the most pertinent ones being chronic hepatitis B and C infection [1, 2]. Vietnam is among the countries with the highest rates of HCC in the world [3].

rs11077 in the 3'UTR region of *XPO5* has been shown to be related to HCC. This SNP consists of two alleles - C and A - that produce three different genotypes, namely A/A, C/C, and A/C. It was found that the 3-year survival rate for HCC patients with genotype A/C and C/C was 60%, while in HCC patients with genotype A/A, it was 24.7%. Therefore, making the distinction between the three genotypes of rs11077 in the *XP05* gene is necessary to be predictive for the treatment of HCC [4, 5].

There are many molecular methods for SNP genotyping such as real-time PCR, PCR-RFLP, ARMS-PCR, PCR-sequencing, and real-time PCR HRM. Among them, real-time PCR HRM has many advantages over other methods such as ease of operation and shorter process duration (compared to PCR-RFLP, PCR-sequencing), which does not require a complicated primer design and PCR optimisation (compared to ARMS-PCR). It also does not require

*Corresponding author: Email: hoangchuongnguyen@gmail.com

complicated analytical devices (automatic nucleotide sequencing machine). Finally, it also provides accurate genotyping result without requiring expensive chemical reagents (Taqman probe). For these reasons, we performed this research aiming at developing a molecular assay for rs11077 genotyping that could be used to genotype rs11077 in clinical practice.

Materials and methods

Reagents

The human blood samples and bacterial strains were supplied by Center for Research and Application in Bioscience (Ho Chi Minh city, Vietnam). The blood samples were collected from healthy people. All the chemical reagents for the DNA extraction, real-time PCR HRM, and agarose gel electrophoresis were purchased from Bioline, Merck and Sigma. The nucleotide sequencing kit was supplied by Applied Biosystems. The primers were synthesised and supplied by Phu Sa Biochem.

Primer design

A DNA fragment containing rs11077 was obtained from GenBank to be used as the template for the primer design. Two primer pairs were designed using the AnnHyb software, in which one pair was designated for genotyping rs1801133 using the real-time PCR HRM method and the other was designated for the nucleotide sequencing of this SNP. The oligo characteristics of these primers in terms of Tm, percentage of GC, free energy of the secondary structures (hairpin, homodimer, heterodimer) were examined using the OligoAnalyzer software to ensure good performance during PCR. Finally, selective binding to the target region containing rs11077 of these primers was checked using the Primer-Blast software.

DNA extraction

Whole blood was first treated with a red blood cell lysis buffer to remove the erythrocytes and obtain the lymphocytes. The lymphocytes were lysed with the guanidine thiocyanate-containing solution in the presence of silica particles. Proteins and other impurities were removed, and the genomic DNA was absorbed by the silica particles. The DNA-containing silica were washed with washing buffer and ethanol 70% to completely remove the remaining impurities and salts. Finally, the genomic DNA was eluted from the silica particles in nuclease-free water and kept at -20^{o}C until used.

Real-time PCR HRM

The 20 µl-volume real-time PCR reaction was set up in 0.1 ml tube with the following components: 10 µl of SensiFast^{TM} HRM master mix 2X (Bioline), 2 µl of the 10 µM-concentration CN5-CN6 primer pair, 5 µl of the genomic DNA template, and 3 µl of water. The reaction program was initiated at 95^{o}C for 120 seconds followed by 40 cycles at 95^{o}C for 10 seconds, 60^{o}C for 10 seconds, and 72^{o}C for 10 seconds. The HRM analysis on PCR product was started at 60^{o}C to 97^{o}C with 0.1^{o}C increment. The results were analysed using MyGo-Pro PCR software based on the melting curve shape on the normalised melting curves and melting point (Tm) of the amplified products.

Nucleotide sequencing

The PCR products containing rs11077 were obtained using PCR with the CN15-CN16 primer pair. These PCR products were purified before being labelled with the appropriate fluorescents. The fluorescent-labelled PCR products were analysed on the ABI 3500 genetic analyser. The nucleotide sequence of the target PCR product was analysed based on the fluorescence signals and was then compared to the original sequence containing rs11077 on GenBank nucleotide database.

Analytical specificity

The selective amplification of the CN13-CN14 primer pair was checked using real-time PCR with the genetic materials from bacteria such as *Escherichia coli, Staphylococcus aureeus, Pseudomonas aeruginosa, Shigella dysenteria, Vibrio cholera,* and *Klebsiella pneumoniae* that may co-exist in human body. In addition, the PCR with the 27F-1495R was performed on these genetic materials to prove that the negative results in the above real-time PCR were not a result of the PCR inhibition.

Precisions

The real-time PCR HRM for genotyping rs11077 was repeated five times in the same test conditions on the same day on the samples containing known C/C, C/A, and A/A genotypes to check the repeatability of the method. Similarly, the rs11077 genotyping protocol was repeated five times in various test conditions on the samples containing known C/C, C/A, and A/A genotypes to check reproducibility. The degree of deviation in the rs11077 genotyping result on the samples was assessed using the value of the coefficient of variation (CV) and the unit of calculation was expressed in

percentage.

Data analysis

All the data in this study was analysed with Excel (Microsoft Office 2013).

Results

Primer design

We designed the primer pair for genotyping rs11077 using a real-time HRM PCR assay and the primer pair for nucleotide sequencing of this SNP. The nucleotide sequences of the primers were shown as follows:

CN13: AGTACCTCCAAGGACCAGG

CN14: AAAGGGGATGTTAGCACTAAAGAC

CN15: CCTTTTGCTGCTGGGCTGG

CN16: TGAGTGGACCTTGAGGCTG

The CN13-CN14 primer pair was designed for genotyping rs11077, and the PCR product with this primer pair was 51 bp in size. In contrast, the CN15-CN16 primer pair was designed for sequencing rs11077 using the Sanger technique, and the PCR product with this primer pair was 300 bp in size.

In the next step, we checked the technical parameters of the primers such as Tm, the GC component, and the free energy of the secondary structures using the OligoAnalyzer software. The results are shown in Table 1.

Table 1. Technical parameters of the designed primers.

Parameters	Primer			
	CN13	*CN14*	*CN15*	*CN16*
Nucleotide	19	24	19	19
GC content (%)	57.9	41.7	63.2	57.9
Tm (°C)	55.8	54.9	59.8	56.9
Hairpin (kcal/mole)	-1.58	0.49	-0.6	-0.65
Self-dimer (kcal/mole)	-4.67	-4.9	-3.14	-4.67
Hetero-dimer (kcal/mole)	-4.5		-4.67	

The results in Table 1 showed that the four primers met the specific requirements that allowed them to work well in the PCR. Finally, we tested the theoretical specificity of these primers using the Blast software. The results showed that the primers matched only the human DNA in the target gene *XPO5* (data not shown). In conclusion, the primers that we designed were suitable for the subsequent experiments.

Building the real-time PCR HRM for rs11077 genotyping

With the CN13-CN14 primer pair, we set up a real-time PCR HRM reaction on five human DNA samples. The results are illustrated in Fig. 1.

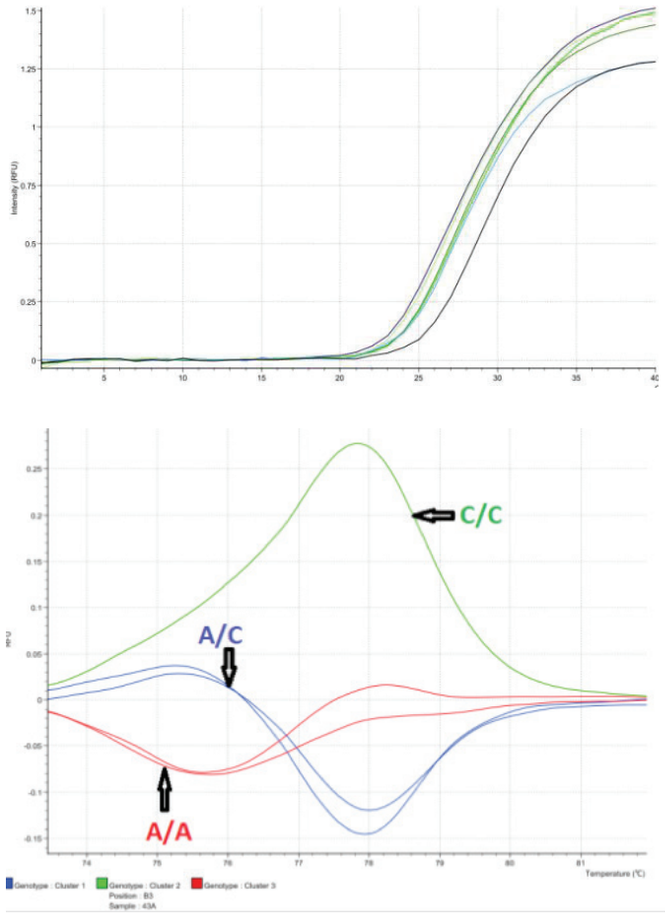


Fig. 1. The real-time PCR HRM results on five human DNA samples.

The results in Fig. 1 show that the real-time PCR results were positive for five DNA samples. While analysing the melting curve using HRM software, three different melting curve patterns that corresponded to the three genotypes A/A, A/C, and C/C of rs11077 were observed. The predicted A/A, A/C, and C/C genotypes based on the 3 melting curve models were confirmed by Sanger's nucleotide sequencing.

For rs11077 nucleotide sequencing, we used the CN15-CN16 primer pair, as the PCR product from the CN13-

CN14 primer pair was not suitable for Sanger nucleotide sequencing because of its small size (51 bp). The results of the nucleotide sequences are shown in Fig. 2.

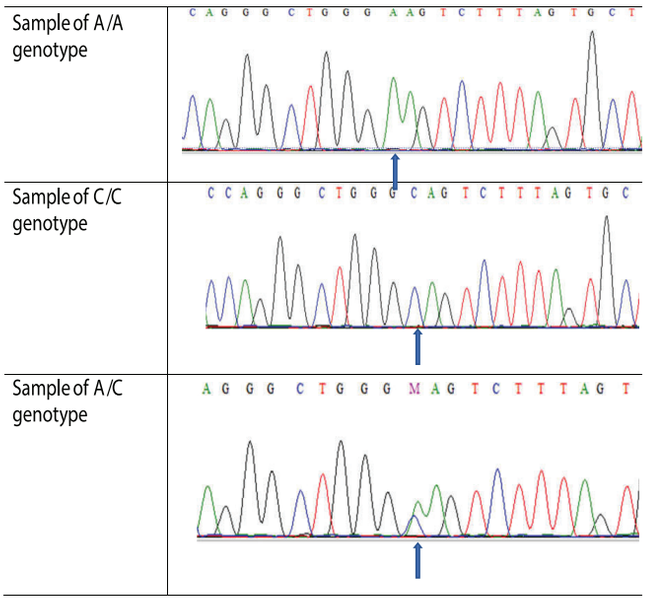


Fig. 2. Results of Sanger's nucleotide sequencing of the A/A, A/C, and C/C genotypes.

The results in Fig. 2 show the occurrence of the peaks corresponding to the nucleotides of rs11077. Specifically, the sample containing the genotype A/A contained a peak of Adenine, and the sample containing the genotype C/C contained a peak representing Cytosine. Samples containing genotype A/C contain the presence of a double peak, representing Adenine and Cytosine. Thus, the result of Sanger's nucleotide sequencing confirmed that the rs11077 genotyping results obtained by using real-time PCR HRM were completely accurate.

Optimisation of $MgCl_2$ concentration

$MgCl_2$ is the major component that influences the melting temperature of PCR products when analysed by HRM. Therefore, we investigated the optimal $MgCl_2$ concentration for distinguishing between the three melting curve patterns corresponding to three genotypes A/A, A/C, and C/C of rs11077. In the PCR master mix for real-time PCR HRM with unknown concentration of $MgCl_2$, we investigated added $MgCl_2$ concentrations in the real-time PCR HRM as follows: 0 mM, 0.5 mM, 1 mM, 1.5 mM, 2 mM, and 2.5 mM. The results of the optimisation of $MgCl_2$ concentration are shown in Fig. 3.

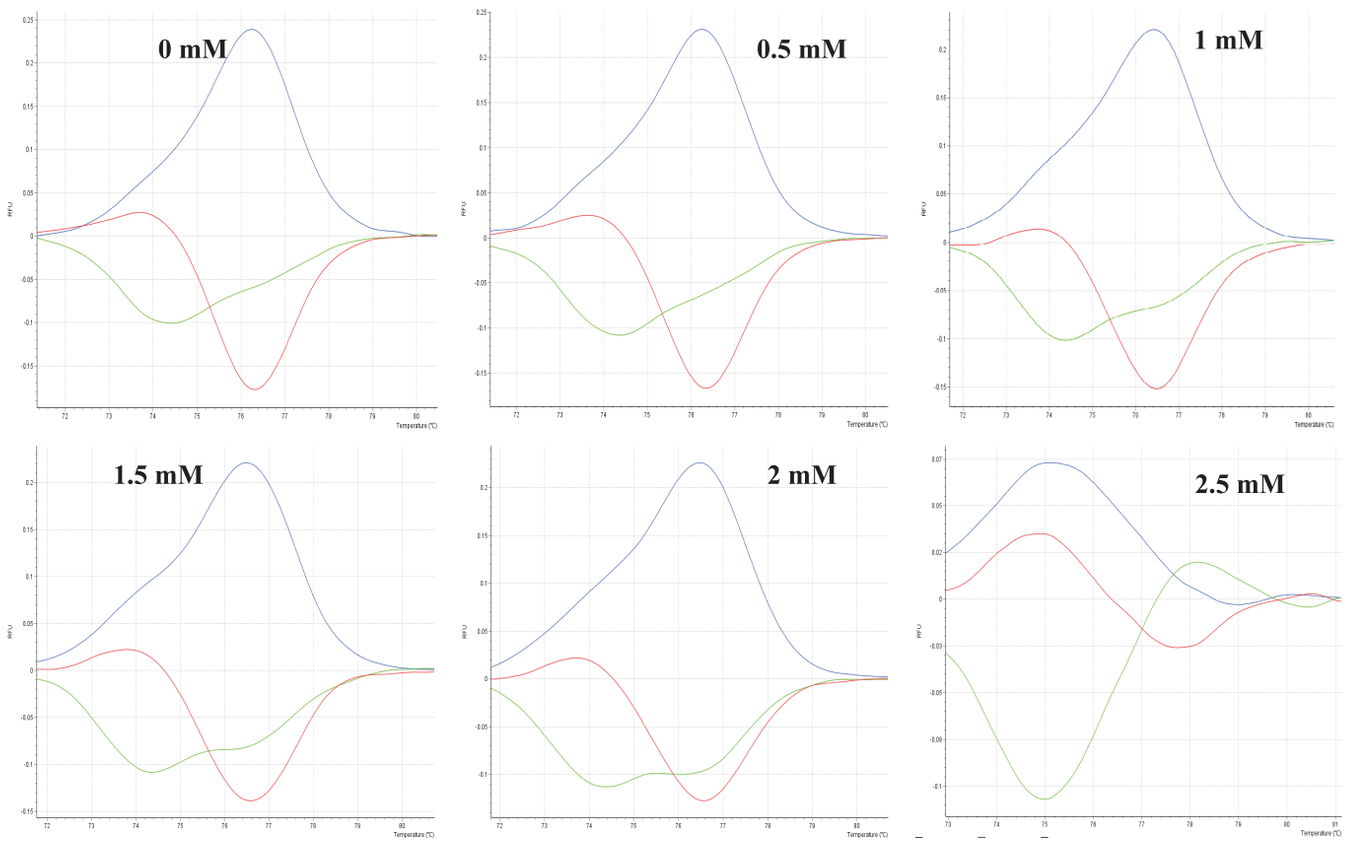


Fig. 3. Optimisation of $MgCl_2$ concentrations for the real-time PCR HRM for genotyping rs11077.

The results in Fig. 3 show that the three melting curves that correspond to the three genotypes A/A, A/C, and C/C of rs11077 were most clearly distinguished at the added $MgCl_2$ concentrations of 0 mM, 0.5 mM, and 1 mM. The other $MgCl_2$ concentrations presented unspecific curves. With this result, we chose 0 mM as the added concentration of $MgCl_2$ for subsequent studies.

Analytical specificity of the real-time PCR HRM protocol

Analytical specificity of the real-time PCR HRM assay was demonstrated by the selective amplification of the human DNA region containing rs11077 by the target primer pair. In this experiment, we investigated the selective amplification of the CN13-CN14 primer pair on the genetic material of various agents, including human and bacteria (*Escherichia coli, Staphylococcus aureus, Pseudomonas aeruginosa, Shigella dysenteriae, Vibrio cholerae,* and *Klebsiella pneumoniae*) in the real-time PCR. The results of the selective amplification of the CN13-CN14 primer pair are presented in Fig. 4.

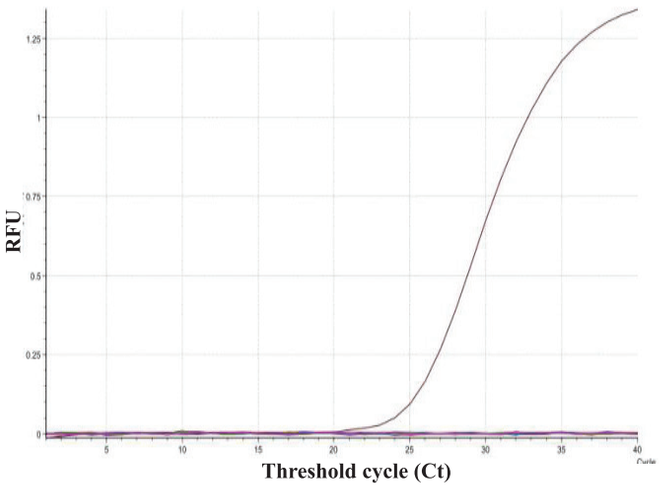


Fig. 4. Selective amplification of the CN13-CN14 primer pair on various genetic materials from human and bacteria.

The results in Fig. 4 show that only the human DNA sample gave a positive result in the real-time PCR reaction with the CN13-CN14 primer pair, whereas DNA samples from the bacteria produced negative results when reacting with the same primer pair. Next, to confirm that the negative results in the real-time PCR reaction with the CN13-CN14 primer pair on the bacterial DNA samples were truly negative, we performed the PCR reaction with the 27F-1495R primer pair on these DNA samples. 27F-1495R is the primer pair specific for the 16S rRNA gene of all eubacteria with the sequences as follows: 27F (GAGAGTTTGATCCTGGCTCAG) and 1495R (CTACGGCTACCTTGTTACGA). The 27F-1495R primer pair gives the PCR product of ~ 1.4 kb. The PCR results are shown in Fig. 5.

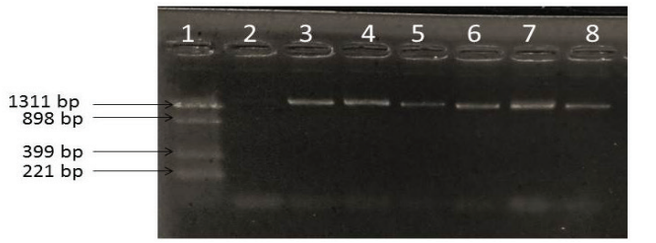


Fig. 5. The PCR results of the bacterial DNA with the 27F-1495R primer pair. Lane 1: DNA ladder, lane 2: negative control; lane 3-8: DNA from *Escherichia coli, Staphylococcus aureus, Pseudomonas aeruginosa, Shigella dysenteriae, Vibrio cholerae*, and *Klebsiella pneumoniae* respectively.

The results in Fig. 5 show that the DNA samples from *Escherichia coli, Staphylococcus aureus, Pseudomonas aeruginosa, Shigella dysenteriae, Vibrio cholerae,* and *Klebsiella pneumoniae* were positive for PCR with the 27F-1495R primer pair. This result confirmed that the negative results in the real-time PCR reaction with the CN13-CN14 primer pair on the bacterial DNA samples were truly negative. Thus, the real-time PCR HRM assay was specifically designed to genotype rs11077 in human.

Precisions

We genotyped rs11077 using the real-time PCR HRM assay on three samples of A/A, A/C, and C/C genotypes five times in the same experiment batch to measure the repeatability of the assay. Additionally, we genotyped rs11077 using the real-time PCR HRM assay on three samples of A/A, A/C, and C/C genotypes five times in the five experiment batches to measure the reproducibility of the assay. The repeatability and reproducibility were expressed by the coefficient of variation (CV), and the CV values were calculated as percentages. The CV value of the repeatability test of the real-time PCR HRM assay was 0.083%, and the reproducibility test results produced a CV value of 0.353%. These CV values proved the high precision of the real-time PCR HRM protocol for genotyping rs11077.

Evaluating the real-time PCR HRM protocol on 123 human blood samples

We evaluated the performance of the real-time PCR

HRM protocol for genotyping rs11077 on the 123 human blood samples that were supplied by Center for Research and Application in Bioscience. The real-time PCR HRM results are shown in Fig. 6.

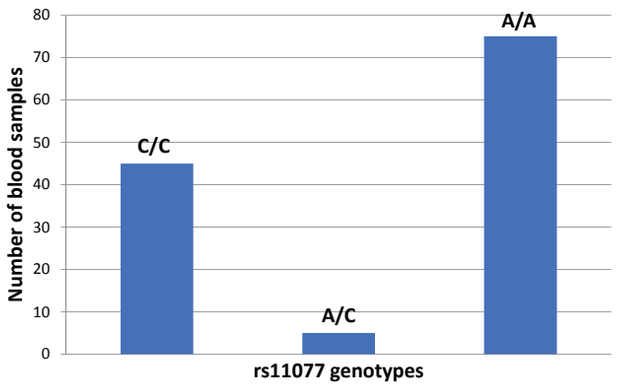


Fig. 6. rs11077 genotyping results on 123 human blood samples using the real-time PCR HRM assay.

Based on the real-time PCR HRM results on 123 human blood samples, 75 samples were found to contain genotype A/A, 5 samples contained genotype A/C, and 43 sample contained genotype C/C. We also performed nucleotide sequencing on 10 random samples out of the 123 samples to confirm the genotyping results by the real-time PCR HRM assay (supplementary data). The comparison showed that the genotyping results on rs11077 between the real-time PCR HRM assay and the Sanger's nucleotide sequencing were identical.

Discussion

There have been many studies on the association of SNPs with human diseases. In HCC, it has been found that a number of SNPs have been associated with this disease, particularly during the progression of the disease. The different genotypes of an SNP can affect the disease differently, which means a certain genotype of an SNP that can make a person more susceptible to develop a disease than another. Among the SNPs related to HCC, such as rs3783553, rs16405, rs99985, rs2910164, and rs11614913, we chose rs11077 to construct a molecular assay for genotyping using real-time PCR HRM. This is because this SNP is heavily involved in the prognostic of HCC patients.

Sanger's nucleotide sequencing is the gold standard for genotyping SNPs, but it is seemed impractical to employ the same in this method in the clinical practice for genotyping rs11077 owing to its complicated process and the required expensive apparatus. In research, other molecular methods such as PCR-RFLP, real-time PCR, molecular hybridisation were used for SNP genotyping. In this study, we chose the real-time PCR HRM method to genotype rs11077, because it has a simple operation procedure but still gives accurate results. In order to genotype rs11077 by distinguishing the melting curve based on HRM analysis, the size of the target PCR product must be so small that a single nucleotide change in the homologous nucleotide sequences can change the melting curve of these nucleotide sequences. Typically, the target PCR size for HRM analysis is from 100-300 bp; however, in this study, we designed the primer pair amplifying the target PCR of only 51 nucleotides. The smaller the target PCR, the more distinct will be the genotype of a SNP. In addition, the primer pair designed for the small target PCR product will avoid other SNPs that are adjacent to the target SNP, as these SNPs will interfere with the melting curves of the target SNP. Through an *in silico* analysis, the size of PCR from CN13-CN14 primer pair was found to be 51 bp, excluding neighbouring SNPs of rs11077. Moreover, the CN15-CN16 primer pair occupies most of the nucleotide sequence of the PCR product except for the position of rs11077, indicating that different human DNAs containing rs11077 will be distinguished merely by rs111077. However, the size of the target PCR product was 51 bp with the pair of CN13-CN14 primers, which was not long enough to be analysed by Sanger's nucleotide sequencing. Thus, we designed the CN15-CN16 primer pair to target the Sanger nucleotide sequence to confirm that the rs11077 subtype is real-time PCR HRM. The PCR product size by the CN15-CN16 is 300 bp, which is sufficient for sequencing by the Sanger technique. The results of sequencing by the Sanger technique confirmed that the genotyping results by real-time PCR HRM on rs11077 were accurate.

Finally, we evaluated the performance of the real-time PCR HRM assay for genotyping rs11077 on the 123 clinical blood samples. Results showed that there were 75 samples bearing genotype A/A, 5 samples carrying genotype A/C, and 43 samples carrying genotype C/C. According to the work of Liu, et al. (2014), people carrying genotype A/A will have a poor treatment prognosis with a 3-year survival rate of 24.7% [4]. In this study, the proportion of people

carrying the genotype A/A accounted for 60.09%, which might be one of the important factors contributing to the high mortality of HCC patients in the Vietnamese population. Knowledge of the genetic information that influences the cancer phenotype is essential for health authorities to control primary liver cancer in the community.

Conclusions

We successfully constructed a molecular assay based on the real-time PCR HRM technique for genotyping rs11077, an SNP involved in the prognosis of HCC patients. The real-time PCR HRM for genotyping rs11077 exhibited good performance in terms of analytical specificity, repeatability, and reproducibility. When evaluated on 123 clinical human blood samples, the real-time PCR HRM assay delivered accurate genotyping results. The assay can be developed to be a prognosis tool for the treatment of HCC patients.

ACKNOWLEDGEMENTS

We thank the Center for Research and Application in Bioscience for the human blood samples and bacterial strains used in this study.

REFERENCES

[1] Y.A. Ghouri, I. Mian, J.H. Rowe (2017), "Review of HCC: Epidemiology, etiology, and carcinogenesis", *Journal of Carcinogenesis*, **16(1)**, doi: 10.4103/jcar.JCar_9_16.

[2] L.R. Roberts, G.J. Gores (2005), "HCC: molecular pathways and new therapeutic targets", *Seminars in Liver Disease,* **25(2)**, pp.212-225.

[3] R.X. Zhu, W.K. Seto, C.L. Lai, M.F. Yuen (2016), "Epidemiology of HCC in the Asia-Pacific region", *Gut and Liver*, **10(3)**, pp.332-339.

[4] S. Liu, J. An, J. Lin, Y. Liu, L. Bao, W. Zhang, J.J. Zhao (2014), "Single nucleotide polymorphisms of microRNA processing machinery genes and outcome of HCC", *PLoS One*, **9(3)**, p.e92791.

[5] R. Yi, Y. Qin, I.G. Macara, B.R. Cullen (2003), "Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs", *Genes & Development*, **17(24)**, pp.3011-3016.

Establishment of multiplex PCR for detection of genes related to Quinolone resistance

Ngo Tat Trung[1*], Dao Thanh Quyen[1], Phan Quoc Hoan[1], Thirumalaisamy P.P. Velavan[2], Le Huu Song[1,3*]

[1]Dept. of Molecular Biology, 108 Military Central Hospital, Hanoi, Vietnam
[2]Institute of Tropical Medicine, Tübingen University, Germany
[3]Faculty of Infectious and Tropical Diseases, 108 Military Central Hospital, Hanoi, Vietnam

Received 29 January 2018; accepted 10 May 2018

Abstract:

Background/aims: It has been shown that quinolone resistance arises due to mutations in the quinolone resistance-determining regions of the drug targets. This study aimed to optimise a multiplex PCR assay to track plasmid-mediated low-level quinolone resistance profiles.

Subjects and methods: A multiplex PCR-based-method in which the primers were already established by our team. About 44 samples were collected from 44 patients who enrolled in this study by using surgical site infection (SSI).

Results: By targeting the conserved domains of *qnrA, qnrB, qnrS and the qnrVC* gene families, the primer number was reduced significantly to only four pairs in one multiplex PCR. Using multiplex PCR, 3/44 SSI samples were found to be carrying fluoroquinolone-resistance genes (*qnrA, qnrB, qnrS, qnrVC*).

Conclusion: A multiplex PCR for detecting pathogens as well as identifying quinolone resistance genes all in one reaction was successfully established.

***Keywords:* plasmid-mediated quinolone resistance genes, plasmids, SSI, qnr.**

***Classification numbers:* 3.3, 3.5**

Introduction

Hospital-acquired bacterial infections and drugs resistance of bacteria are big healthcare problems not only in Vietnam but also worldwide. It not only causes severe consequences for patients but also causes economic losses and global diseases. Surgical site infection (SSI) is one of the most common nosocomial infection that occurrs in operated patients. The diagnosis of SSI seems to be easy but, sometimes, the identified pathogens and their profile of antibiotics resistance are inconcludable by culture. Bacterial culture is the classical method for identifying pathogens. However, there are many disadvantages such as, i) long waiting for the result and ii) false negative rate is still high.

Recent advances in nucleic acid testing have opened up doors for molecular detection of microbes. One of the latest methodologies employs multiplexing assays, thus, offering a chance to not only investigate various microbial pathogens causing SSI but also to quickly detect the antibiotic resistance genes.

In bacteria, the type II topoisomerases DNA gyrase and topoisomerase IV change the topology of DNA. These enzymes cleave both strands of DNA and allow one double-stranded DNA molecule to pass through another and then reseal the break. DNA gyrase wraps DNA around itself, which is responsible for introducing negative superhelical twists to plasmid DNA molecules that topoisomerase IV does not, and DNA becomes relaxed. Quinolones inhibit

*Corresponding author: Email: tatrungngo@gmail.com, lehuusong@108-icid.com

these enzymes by stabilising the DNA-DNA gyrase complex or topoisomerase IV and thus, block the progression of polymerase and DNA replication. The widespread use of these agents has contributed to the rise of bacterial quinolone resistance [1]. Previous studies have shown that the mechanisms of quinolone resistance arise by mutations in chromosomal genes [2] and plasmid-mediated low-level resistance mechanisms (*qnrA, qnrB, qnrS, qnrVC*) [3]. These genes have a large geographic distribution (mainly in *Enterobacteriaceae*) [4-6]. Since these genes appear in various bacteria species, the prevalence of detecting this type of resistance phenotypically is only partially known. Although these mechanisms cause low-level resistance, but they favour and complement the selection of other resistance mechanisms [4].

In this study, we aim to develop and optimise a multiplex PCR approach in order to track the plasmid-mediated low-level quinolone resistance mechanisms. By applying our assays to a collection of SSI samples, the first prevalence of fluoroquinolone resistance genes (*qnrA, qnrB, qnrS, qnrVC*) in Vietnam can be described.

Subjects and methods

Subjects

For this study, we included patients who were hospitalised at 108 Military Central Hospital, Hanoi, Vietnam from February 2012 to December 2012 with at least one of these conditions: (1) purulent incisional drainage from deep layers of soft tissue within 30 days after the operation or within 1 year after the operation if a prosthesis was inserted and (2) local signs and symptoms of pain or tenderness, swelling and erythema with the incision opened by the surgeon or confirmed by the attending surgeon or physician. Patients provided their written informed consent prior to participation, and the study was approved by the local ethics committee of the 108 Military Central Hospital.

Exudates from deep incisional surgical infection sites (aspirate beneath the incision area) were collected using sterile syringes and, in parallel, subjected to bacterial culture or stored at minus 80 degrees Celsius for molecular diagnostic.

In total, ninety-one patient samples (n=91) were enrolled in this study procedure. About 44/91 SSI samples were proved to be bacterial culture positive.

Primer design

In order to screen genes encoding important quinolone resistance plasmids *(qnrA, qnrB, qnrS and qnrVC),* the gene bank accession numbers of the *qnrA, qnrB, qnrS* and *qnrVC* gene families were acquired from http://blast.ncbi.nlm.nih.gov/Blast.cgi and imputed into Vector NTI software 11.5 (Invitrogen, California).

DNA extraction and multiplex PCR

Aliquots of exudates from SSIs were allocated for total bacterial culture and the remnants were immersed into a universal lysis solution (200 mM NaOH, 1% SDS) and heated for 5 minutes at 95^{o}C. An equal volume of 1 M Tris-HCl was also added to neutralise the pH down to about 7.5. The resulting solution was subjected to standard phenol/chloroform/isoamyl alcohol extraction. The precipitated DNA was reconstituted in 150 µl TE (25 mM Tris-base, pH 8.0, 1 mM EDTA). About 5 µl of the reconstituted DNA was used as a template for multiplex PCR in 25 µl reactions containing 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM Mg^{2+} and 250 µM of each deoxynucleotide triphosphate and optimised amount of individual primers. Thermal amplification cycles were carried out as follows: initial denaturation at 95^{o}C for 4 minutes, 35 cycles of 94^{o}C for 25 seconds, 58^{o}C for 45 seconds and 72^{o}C for one minute.

Confirmatory sequencing of amplicons

To avoid false positive results due to cross-reactivity of the selected primer pairs with human DNA, amplified products were cleaned up and subjected to bidirectional PCR sequencing. The outcome was compared to an annotated gene-bank database (Blast search http://blast.

ncbi.nlm.nih.gov/Blast.cgi) in order to identify human genomic information.

Results

Obtained primers

By using the described assay for designing primers, which ones were selected by vector NTI software and targeted for PCR primer binding. The multiplex PCR assay SHPT@QNR (*qnrA, qnrB, qnrS, qnrVC*) was developed and optimised in a way that neighbouring PCR amplicons differed from each other by 60 to 100 bp, which helps to resolve the PCR fragments into visible bands in agarose gel electrophoresis. The primer details can be found in Table 1.

Table 1. Primers used for amplifying conserved domains of quinolone resistance gene families (*qnrA, qnrB, qnrS, qnrVC*).

Primer name	Primer sequencing (5' to 3')	Amplicon size (bp)
Tr-qnrB-F	TGARTTTATYGGCTGYCARTTYTATGATCG	216
Tr-qnrB-R	CAGGTGCGMGTRGTGATCATATTCAT	
Tr-qnrS-F	AACTGCAAGTTCATTGAACAGGGTGATATC	357
Tr-qnrS-R	AACACCTCGACTTAAGTCTGACTCTTTCA	
Tr-qnrVC-F	GGGTGYGATTTTTCTTAYKCKGATCTT	447
Tr-qnrVC-R	CTGYTGCCACGARCAKATTTTTACACC	
Tr-qnrA-F	CAAGAGGATTTCTCACGCCAGGATTT	562
Tr-qnrA-R	CCCCMTCGAGGTTGACCCG	

SHPT@QNR multiplex assay reliably differentiates between qnrA, qnrB, qnrS, and qnrVC positivity

As shown in Fig. 1, individual amplicons derived from multiplexing differ from one another by 60 to 100 bp and do not generate any unspecific bands. All the PCR fragments were resolved into visible bands in agarose gel electrophoresis.

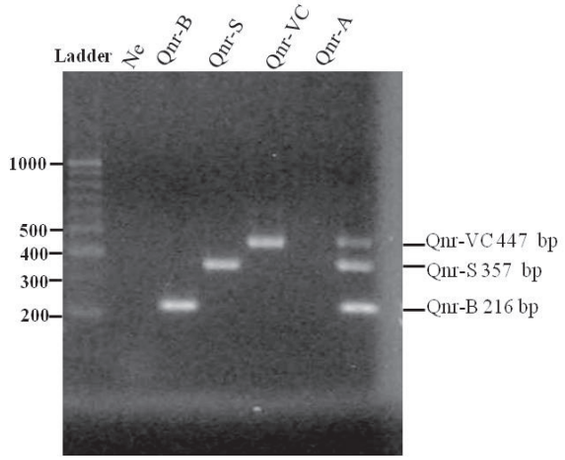


Fig. 1. Multiplex PCR assays (SHPT@QNR) for screening quinolone resistance genes (qnrB, qnrS, qnrVC and qnrA).

Screening patient samples for quinolone resistance utilising the SHPT@QNR assay

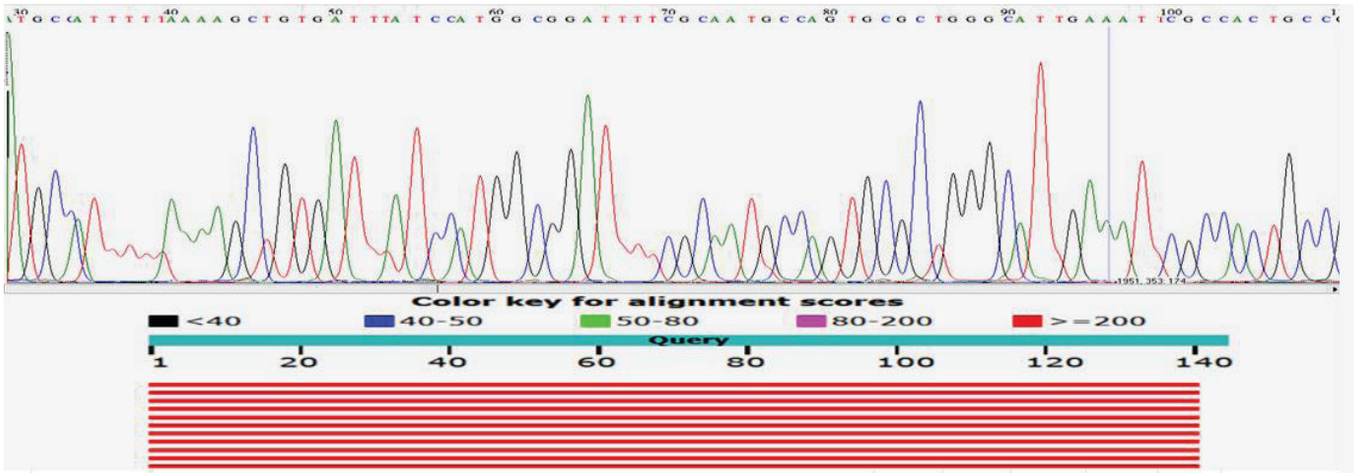
After optimising the amplifying abilities of the SHPT@QNR assay (Fig. 1), it was used to screen the potential quinolone resistance genes (*qnrA, qnrB, qnrS, qnrVC*) from all 44 bacterial culture positive SSI samples (out of 91 suspected SSI biopsies). As seen in Table 2, 3 out of 44 SSI samples harbour the following genetic material: NKSM 33 (*qnrB, qnrVC), NKSM 40 (qnrB)*, NKSM55 (*qnrS),* which encode *qnrB, qnrS, qnrVC,* respectively.

Table 2. Quinolone resistance genes (*qnrA, qnrB, qnrVC)* were found in collected SSI samples. Noted: SSI-ID = code of sample.

SSI-ID	Bacterial strain	Detected Quinolone resitance gene		
		QnrA	QnrB	QnrVC
NKSM33	E. coli, P. aeruginosa	Positive	-	Positive
NKSM40	K. pneumoniae	Positive	-	-
NKSM55	P. aeruginosa	-	Positive	-

The sequencing results

The positive PCR was confirmed by direct sequencing. The obtained sequences were analysised by using blast stool (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and then, they were compared with the sequence on the GeneBank. We showed that the identity of the nucleotide sequences from qnrB, qnrVC and qnrS PCR products were 99%, 99% and 100%, respectively. The result is shown in Fig. 2.



Description	Max score	Total score	Query cover	E value	Ident	Accession
Klebsiella aerogenes strain AR_0062 plasmid tig00000053_pilon, complete sequence	279	279	97%	2e-71	99%	CP026757.1
Klebsiella pneumoniae strain FDAARGOS_447 plasmid unnamed3, complete sequence	279	279	97%	2e-71	99%	CP023950.1
Enterobacter hormaechei plasmid pMTY11043_IncHI2 DNA, complete genome, strain: TUM11043	279	279	97%	2e-71	99%	AP018352.1
Klebsiella pneumoniae strain QH38-1 QnrB (qnrB) gene, complete cds (Qnr B)	279	279	97%	2e-71	99%	KY769120.1
Klebsiella pneumoniae strain QH4-1 QnrB (qnrB) gene, complete cds	279	279	97%	2e-71	99%	KY769118.1
Pseudomonas aeruginosa strain Pa25 class I integron ANT(2")-Ia, VIM-2 (blaVIM-2), ANT(2")-Ia,	700	700	81%	0.0	99%	JX861889.1
Vibrio cholerae non-O1/non-O139 strain DRV181 class 1 integron, partial sequence	700	700	81%	0.0	99%	HM015625.1
Vibrio cholerae strain VC627 class I integron integrase (intI1) gene, partial cds, and streptomycin	700	700	81%	0.0	99%	EU436855.2
Vibrio cholerae O1 biovar El Tor strain MCV09 transposon SXT QnrVC3 (qnrVC3) and dihydrofol (Qnr VC)	695	695	81%	0.0	99%	HM015627.1
Escherichia coli strain J53 transposon SXT, partial sequence	695	695	81%	0.0	99%	FJ968160.1
Vibrio parahaemolyticus strain B3-5-1 QnrVC1 gene, partial cds	684	684	78%	0.0	99%	KU363013.1
Shigella flexneri strain M11560 quinolone resistance pentapeptide repeat protein QnrS1 (qnrS) ge (Qnr S)	544	544	97%	8e-151	100%	KU848190.1
Uncultured bacterium clone RS_qnrs_OTU3 quinolone resistance protein (qnrS) gene, partial cds	544	544	97%	8e-151	100%	KU604991.1
Uncultured bacterium clone RS_qnrs_OTU1 quinolone resistance protein (qnrS) gene, partial cds	544	544	97%	8e-151	100%	KU604989.1

Fig. 2. The sequencing results qnrB, qnrVC, qnrS PCR amplicons.

Discussion

Hospital-acquired bacterial infections paired with antibiotic resistance is a major contemporary public health threat, especially in developing countries like Vietnam, where the medical infrastructure is sub-standard and utilisation of antibiotics is barely regulated. So far, local specific epidemiological data on hospital infection and acquired bacterial antibiotic resistant property were only sporadically reported [7, 8]. In this situation, relevant diagnostic methodologies for monitoring bacterial infections are needed. Together with bacterial infectivity, nosocomial infections related antibiotic resistance is also of public concern. It raises a quest to look for suitable tracking tools. From the experience revealed during the clinical performance in our institution, we suspect the existence of such microorganisms carrying the extended spectrum betalactamese or carbapenemese capability as well as quinolone resistance. However, once the patients have received prophylaxis, broad range antibiotic pressure might inhibit the colonial growth in bacterial culture testing. This is why culture-based antibiotic resistance assays, such as minimum inhibitory concentration tests, often can't be carried out in a clinical setting. That leads to emerging

antibiotic resistance, which is covered up until it has been deeply entrenched in the bacterial population.

The recent emergence of the so-called nucleic acid-based detection methodology has opened up possibilities of not only quickly tracking the presence of different microbial pathogens, which is only present in limited numbers, but also the genes encoding for antibiotic resistance. A challenge establish PCR in the clinic is the diversity of genetics materials coding for antibiotic resistance. Since a single PCR reaction can only detect one genotype, the number of PCR reactions proportionally increases with the diversity level of antibiotic resistance, putting a strain on resources such as time, manpower and money. The multi-primer format is an option here. However, the increase in the number of primers would force them to compete with each other and/or create a risk of getting unspecific signals. In this study, we designed primers complementary to the conserved regions of relevant antibiotic resistance gene families, therefore, reducing the primer number and ending up with only four primer pairs, which still possess the capacity of detecting 51 different genotypes of quinolone resistance gene families.

Conclusions

In short, we reported the establishment of a four-primer-pairs optimised multiplex PCR assay that is significant for screening common gene families that cause quinolone resistance for bacteria.

ACKNOWLEDGEMENTS

This study was funded by the Vietnam National Foundation for Science and Technology Development (NAFOSTED) under the Grant number 108.06-2017.21. The funding agency had no role in the study design, data collection and analysis, decision to publish, and/or preparation of the manuscript.

Conflict of interests

The authors declare no conflict of interests in this study.

REFERENCES

[1] E.S. Kim, and D.C. Hooper (2014), "Clinical importance and epidemiology of quinolone resistance", *Infect. Chemother.*, **46(4)**, pp.226-238.

[2] P.M. Hawkey, and A.M. Jones (2009), "The changing epidemiology of resistance", *J. Antimicrob. Chemother*., **64,** Suppl. 1, pp.i3-10.

[3] PP. Nordmann, and L. Poirel (2005), "Emergence of plasmid-mediated resistance to quinolones in Enterobacteriaceae", *J. Antimicrob. Chemother.*, **56(3)**, pp.463-469.

[4] A. Robicsek, G.A. Jacoby, and D.C. Hooper (2006), "The worldwide emergence of plasmid-mediated quinolone resistance", *Lancet Infect. Dis.*, **6(10)**, pp.629-640.

[5] A. Paauw, et al. (2007), "Failure to control an outbreak of qnrA1-positive multidrug-resistant Enterobacter cloacae infection despite adequate implementation of recommended infection control measures", *J. Clin. Microbiol.*, **45(5)**, pp.1420-1425.

[6] J. Liu, et al. (2015), "The dissemination mode of drug-resistant genes in Enterobacter cloacae", *Indian J. Med. Microbiol.*, **33**, pp.S87-92.

[7] D. Nguyen, et al. (2001), "Incidence and predictors of surgical-site infections in Vietnam", *Infect. Control Hosp. Epidemiol.*, **22(8)**, pp.485-492.

[8] K.V. Nguyen, et al. (2013), "Antibiotic use and resistance in emerging economies: a situation analysis for Vietnam", *BMC Public Health*, **13**, p.1158.

DNA barcoding, an approach for molecular identification of Huyen-sam (*Scrophularia* L.) samples collected in Northern Vietnam

Manh Minh Bui[1], Anh Tuan Vu[2], Phuong Nhung Vu[1], Quang Cu Pham[2], Dang Ton Nguyen[1,3], Thi Thu Hue Huynh[1,3*]

[1]*Institute of Genome Research, Vietnam Academy of Science and Technology*
[2]*General Department of Logistics - Techniques*
[3]*Graduate University of Science and Technology, Vietnam Academy of Science and Technology*

Received 4 December 2017; accepted 26 March 2018

Abstract:

Huyen-sam (Vietnamese name) which belongs to *Scrophularia* L. genus is a valuable herb. This medicinal plant is classified as *Scrophularia ningpoensis* Hemsl. Huyen-sam roots, which contain a large amount of bioactive compounds, have a similar morphology to its relatives. DNA barcodes promise to be a precise and reliable tool for distinguishing the processed Huyen-sam materials from their counterfeits. However, studies about using DNA barcodes for classification of *Scrophularia* L. in Vietnam are not available. Here, we conducted a taxonomic analysis of eight *Scrophularia* L. samples collected from the mountain areas of Northern Vietnam. Based on the combined sequence data of ribosomal nuclear *ITS*, a part of chloroplast *rbcL* gene and *trnL-trnF* intergenic spacer, phylograms of *Scrophularia* L. were generated by both Bayesian inference and maximum likelihood bootstrap method. The phylogenetic analysis showed that the tested samples have a sister relationship to *S. ningpoensis*. Hopefully, the analysis strategy that we used would contribute to further phylogenetic analyses of medicinal plants of Vietnam in the future.

Keywords: ***Bayesian inference, DNA barcodes, ITS, maximum likelihood, medicinal plants, phylogenetic, rbcL, Scrophularia L., trnL-trnF.***

Classification numbers: ***3.3, 3.5***

Introduction

Scrophularia L., which is commonly called "figwort" is a plant genus belonging to family *Scrophulariaceae*. The genus comprises about 200-300 species distributed in Central Asia, Europe (Mediterranean), North America and China [1-3]. Huyen-sam (Vietnamese name) or *Scrophularia ningpoensis* Hemsl., whose root is a valuable natural herb, is usually used for the treatment of inflammation, constipation and fever [4-6]. The main bioactive compounds present in *S. ningpoensis*'s root are harpagoside, angroside C, acteoside and cinnamic acid, which have anti-inflammatory, antimicrobial and antioxidant effects [3, 6, 7]. Sourced from Southeast China, this herb is also domestically grown in some northern districts of Vietnam, such as Lao Cai, Ha Giang and Cao Bang [6]. Due to the similarity in the morphology, *S. ningpoensis*'s root can be mistaken for its close relatives, such as *S. buergeriana* Miq. or *S. kakudensis* Franch. Consequently, the demand for new molecular markers that support the identification of processed *S. ningpoensis*'s samples has become increasingly necessary. However, phylogenetic studies based on molecular markers of *Scrophularia* L. are very limited in Vietnam. These discoveries would play an important role in assuring the quality of processed herb in Vietnam market.

DNA barcoding is a conventional method for the identification of unknown living organism specimens. This approach can be applied to a wide range of species from microbes to higher animals. By analysing the evolutionary rate of small genome fragments as substitutes for morphology aspects, the method provides a quick and cost-effective species identification, especially for higher plant taxons [8, 9]. The searching for universal DNA

*Corresponding author: Email: hthue@igr.ac.vn

barcodes for plants is still ongoing; however, there is a common agreement that more than one region is needed for performing the taxonomy consensus analysis [10, 11].

The selected plant DNA barcodes are usually the genome regions which have a suitable evolutionary rate for generating enough changes in various nucleotide sites during generations. The majority of plant DNA barcoding studies have utilized the DNA regions located on the plastid genome, e.g. *ribulose 1,5-bisphosphate carboxylase large subunit (rbcL)*, *maturase K (matK)* and multiple intergenic regions such as tRNA Leucine - tRNA Phenylanaline (*trnL-trnF*), tRNA Histidine - photosystem binding protein A (*trnH-psbA*), tRNA Glutamine - ribosomal protein S16 (*trnQ-rps16*), tRNA Cysteine - tRNA Asparagine (*trnC-trnD*), and tRNA Alanine - tRNA Histidine (*trnA-trnH*) [11-15]. In addition, the nuclear internal transcribed spacer (ITS) and 18s RNA can also perform as useful barcodes for classifying flower plants [11, 16-18].

During the last decade, taxonomy studies of *Scrophularia* L. based on the DNA barcodes have been gradually conducted. The phylogenetic relationships among *Scrophularia* L. taxa collected from different parts of American continent were analysed from the combined data from the sequence of *ITS*, the chloroplast *trnQ-rps16* and *psbA-trnH* intergenic spacers [2]. Another study on the evolutionary relationships of *Scrophularia* L. species in Western Mediterranean and Macaronesia was completed by the data of ITS and *trnQ-rps16* by Bayesian binary MCMC (BBM) analysis [17].

In this paper, we perform a phylogenetic analysis of *Scrophularia* L. samples in Northern Vietnam. More particularly, the analysis is based on the sequence data from nuclear *ITS*, chloroplast *rbcL,* and *trnL-trnF*. The combined data set from these DNA regions are objects for generating the phylogenetic tree by Bayesian inference (BI) and maximum likelihood (ML) analysis. The main aim of this project is to contribute to the molecular classification study of genus *Scrophularia* L. in Vietnam.

Materials and methods

Plant materials

Eight leaf samples of *Scrophularia* L. were collected from the cultivated gardens located in different mountain districts in Northern Vietnam namely HSa-1A, HSa-1B, HSa-2A, HSa-3A, HSa-3B, HSa-4B, HSa-5A, HSa-8A. All the samples were preserved in silica gel for a completed desiccation.

DNA extraction, amplification and sequencing

Total DNA was extracted from about 100 mg of the dried leaf following the CTAB extraction method [19]. The extracted DNA was resuspended in 50 µl miliQ water, and standard 50 ng of the DNA was used for amplification. The primers for amplification of target regions were designed based on the reference sequence on Genbank (Table 1).

Table 1. List of primers used in the study.

Primer	DNA regions	Primer sequences (5'→3')	Amplicon size (bp)
ITS-AB-101	*ITS*	ACGAATTCATGGTCCGGTGAAGTGTTCG	800
ITS-AB-102		TAGAATTCCCCGGTTCGCTCGCCGTTAC	
rbcL-F	*rbcL*	ATTTGAACTGGTGACACGAG	600
rbcL-R		CGAAATCGGTAGACGCTACG	
TrnL-PF	*TrnL-trnF*	AGTGTTGGATTCAAGCTGGTG	1100
TrnL-PR		TGGTTGTGAGTTCACGTTCT	

The condition of amplification was optimized for 20 µl of PCR, including 50 ng extracted DNA, 2.5 µM of each primer, 0.75 unit of Phusion polymerase (Thermo Scientific), 1 mM of each dNTP and Phusion PCR buffer. The amplification thermocycles were performed as follows: 1 cycle of denaturing at 94°C for 4 minutes, 35 cycles of amplification including 94°C/30s followed by annealing at 52°C/30s (*trnL-trnF* and *rbcL*) or 54°C/30s (*ITS)* and extension 72°C/1 min 30s; ending with a final extension step of 72°C/7 mins. The PCR products were checked by electrophoresis on 0.8% agarose gel. Successful PCR products were purified by Thermofisher Scientific DNA

purification kit (K0512). Sequencing was carried out using the BigDyeTM terminator v3.1 cycle sequencing kit (Applied Biosystems) in a final volume of 20 µl. Sequence runs were performed on an ABI 3500 genetic analyser following Sanger's principle.

Alignment and phylogenetic construction

All the DNA sequences generated from this study were assembled, edited and aligned manually using Bioedit v7.0.5.9 which embedded the ClustalW v1.8 [20]. To access the closeness of the relationships between tested plant samples and the species of *Scrophularia* L., the DNA sequences of 3 examined regions namely *ITS, rbcL* and trnL-*trnF* of species involving in genus *Scrophularia* L. and some other genera of *Lamiales* as outgroup were downloaded from Genbank (www.ncbi.nlm.nih.gov) and aligned (Table 2).

Table 2. Taxons included in this study, with Genbank accession numbers.

Species	GenBank accession numbers		
	ITS	***rbcL***	***trnL-trnF***
Scrophularia ningpoensis	FJ609731.1	GQ436721.1	AY695886.1
Scrophularia buergeriana	JQ065663.1	NC031437.1	KP718626.1
Scrophularia takesimensis	JQ065681.1	KP718628.1	AY695886.1
Scrophularia kakudensis	JQ065674.1	-	KM979600.1
Scrophularia zvartiana	KY067618.1	-	-
Scrophularia arguta	-	-	AJ430936.1
Scrophularia californica	-	-	HQ412946.1
Orobanche gracilis	JX193303.1	AY582198.1	-
Orobanche californica	KC480368.1	AY582178.1	-
Phelipanche ramosa	AY209315.1	AY582252.1	-
Conopholis americana	AY209289.1	-	-
Epifagus virginiana	AY209290.1	-	-
Lathraea squamaria	KC480353.1	-	-
Plantago maritima	AY101879.1	KR297244.1	AY101934.1
Plantago media	AJ548964.1	KF602241.1	AY101920.1
Coffea canephora	MF417755.1	NC030053.1	AF102405.2
Coffea arabica	MF417758.1	EF044213.1	DQ153829.1
Jasminum nudiflorum	AF534817.1	DQ673255.1	EU281146.1
Olea europaea	KF805102.1	DQ673304.1	AF231866.1
Olea woodiana	JX862658.1	NC015608.1	LN515476.1
Utricularia macrorhiza	-	NC025653.1	AF482657.1
Utricularia gibba	-	NC021449.1	AF482657.1
Ajura reptans	EF508061.1	Z37385.1	GU381470.1
Salvia miltiorrhiza	DQ132863.1	KC473307.1	DQ667523.1
Pogostemon stellatus	KP718621.1	NC031434.1	NC031434.1
Phyllostegia velutina	KF529547.1	NC029820.1	KU724134.1
Stachys sylvatica	JN680361.1	AF502022.1	NC029824.1

-: the sequence is unavailable.

The phylogenetic taxonomy analysis was conducted with a BI and an ML approach from 3 datasets (ITS, chloroplast and combined data). The BI was calculated by MrBayes 3.2.6 using a Metropolis-coupled Markov Chain Monte Carlo (MCMCMC) algorithm [21, 22]. The certainty of the node generated by BI were supported by the posterior probability (PP) value, which ranged from 0 to 1. The Combined chloroplast partition evolution was assumed to follow the general time reversible model with a proportion of the sites invariable and the rate for the remaining sites drawn from a gamma distribution (GRT+Γ+I model), while ITS region follows a SYM model (SYM+Γ model) [23]. The MrBayes was executed with 2 runs and four chains (3 hot - 1 cold) with the default temperature of hot chain t = 0.2 for 10 million generations, sampling every 2000th generation to generate 10,000 trees. A burn-in ratio of 10% of sampled trees was discarded and the BI consensus tree were generated from 80% of the remainders. Besides, the alignment of the combined dataset (*ITS, rbcL* and *trnL-trnF*) and ML was performed by the software MEGA 6.06 with the bootstrap method of 1,000 replications [24]. The consensus tree was drawn using Figtree v.1.2.3.

For controlling the incongruence between phylogenetic trees generated by BI and ML bootstrap method, the phylograms received from the chloroplast and nuclear datasets were analysed separately before combining. The incongruence taxons and nodes with a high level of BS (70%) or surpassing the Bayesian support of 85% were discarded [25, 26].

Results

DNA extraction and amplification

The extracted DNA from examined samples were used as templates for amplification of *ITS, rbcL* and *trnL-trnF* regions using designed primers. The size of PCR products was checked by electrophoresis on Agarose gel 0.8%. The correct PCR products were purified and sequenced for generating the DNA sequences afterward (Fig. 1).

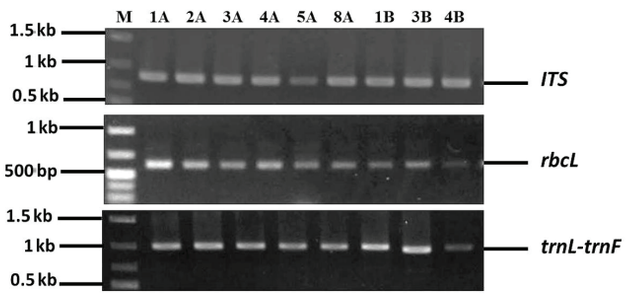


Fig. 1. Gel electrophoresis image of PCR products from *Scrophularia* samples on 0.8% agarose gel. 1A, 2A, 3A, 5A, 8A, 1B, 3B, and 4B are correspondent to HSa-1A, HSa-2A, HSa-3A, HSa-5A, HSa-8A, HSa-1B, HSa-3B and HSa-4B, respectively. The sample *ITS, rbcL* and *trnL-trnF* fragments have the size 800, 600 and 1,100 bp, respectively. The PCR products clearly showed the DNA bands with the correct size of desired fragments.

DNA sequence alignment

In this study, a total of 24 sequences were generated from tested *Scrophularia* samples including 8 each for the nuclear *ITS* regions, chloroplast *rbcL* genes and *trnL-trnF* intergenic spaces. The sequences obtained from examined samples were aligned with correspondent references for creating 3 separated alignment datasets namely ITS, Chloroplast combination (*rbcL+trnL-trnF*) and Mixed combination (*ITS+rbcL+trnL-trnF*). The Chloroplast combination is a merger of 2 plastid sequence, while the Mixed combination was generated by adding the ITS sequence to the Chloroplast combination. The mixed combined data matrix contained 2,065 aligned characters with the average sequence length of 1,826.1 bp. The average sequence length was 544.9 bp for ITS alignment data, 551.7 bp for *rbcL* and 729.8 bp for *trnL-trnF.* We also estimated the mean of evolution distance between the taxons included in each dataset using the Maximum composite likelihood model with Gamma distribution and assuming rate variation and pattern heterogeneity among sites. The *trnL-trnF* showed the highest overall mean of evolution distance (0.414) followed by the ITS regions (0.374), which indicates that these regions have relatively high evolution rates on *Scrophularia* L. genus. The detailed statistic information about the aligned dataset, including mean G+C content, number of conserved nucleotides and parsimony-informative sites, were provided in Table 3.

Table 3. Alignment characteristics and statistics for ITS, *rbcL* region, *trnL-trnF* intergenic space, combined chloroplast, and combined dataset.

	ITS	rbcL	trnL-trnF	Comb. Chloroplast	Combined
Number of taxa	32	31	33	26	23
Average sequence length (bp)	544.9	551.7	729.8	1277.6	1826.1
Aligned sequence length	608	580	1019	1457	2065
Conserved characters	228	376	223	788	1109
Parsimony-informative characters	298	112	538	383	601
Overall mean evolution distance	0.347	0.08	0.414	0.111	0.09
%G+C content	61.5	45.2	35.2	38.7	45.6

The table contains the number of conserved characters, parsimony - informative characters and mean % GC content of aligned sequences. Overall mean evolution distance was estimated through maximum composite likelihood model with Gamma distribution and assuming rate variation and pattern heterogeneity among sites.

Phylogenetic tree construction

In this study, we generated phylogenetic trees for three separately aligned datasets. For all the BI consensus trees, the average standard deviations of split frequencies when 2-run coverage at stationary distribution remained at lower than 0.002. The low split frequency indicated an increasing similarity of runs, and the results were adequate for the next analysis. The phylogenetic trees were obtained from 9,002 sampled trees after the Bayesian runs. The BI and ML analyses of each dataset showed high congruence on topologies, especially on the *Scrophulariaceae* family. Therefore, we plotted the BS value of ML analysis onto the respective BI consensus tree.

Individual phylogenetic trees generated from the nuclear barcode *ITS* and Chloroplast combination dataset (*rbcL+trnL-trnF*) were shown in Figs. 2 and 3, respectively. The clades of the examined sample in both consensus trees are highly similar in branch length with each other and grouped with *Scrophularia ningpoensis,* suggesting a close relationship. However, the node between *S. ningpoensis* and HSa-8A was not well-supported by the BI analysis (PP: 0.53, BS: 97). The BI and ML analyses also did not support the node of *S. takesimensis* and *S. zvartiana* (PP: 0.68, BS: 43) and the node of outgroup and tested samples (PP: 0.5, BS: 7), indicating the uncertainty of the ITS trees (Fig. 2). On the outgroup clades, the *Lathraea squamaria* was incorrectly ordered into the clade of *Orobanchaceae* instead

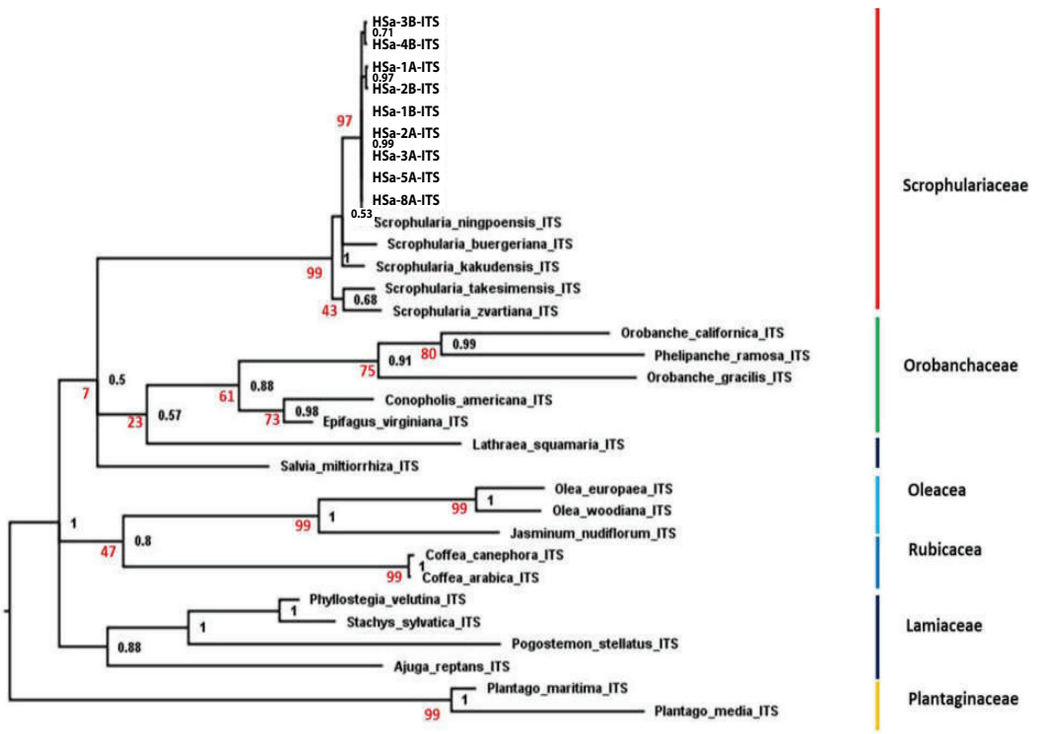


Fig. 2. Bayesian consensus tree of *Scrophularia* L., generated from the ITS dataset and reference sequences from 5 other genera of *Lamiales*. PP values are given in black number next to each node, while the corresponding BS values are given in red numbers. PP values are obtained from 9,002 trees. The scale bar indicates the average expected changes per site of sequences in the study.

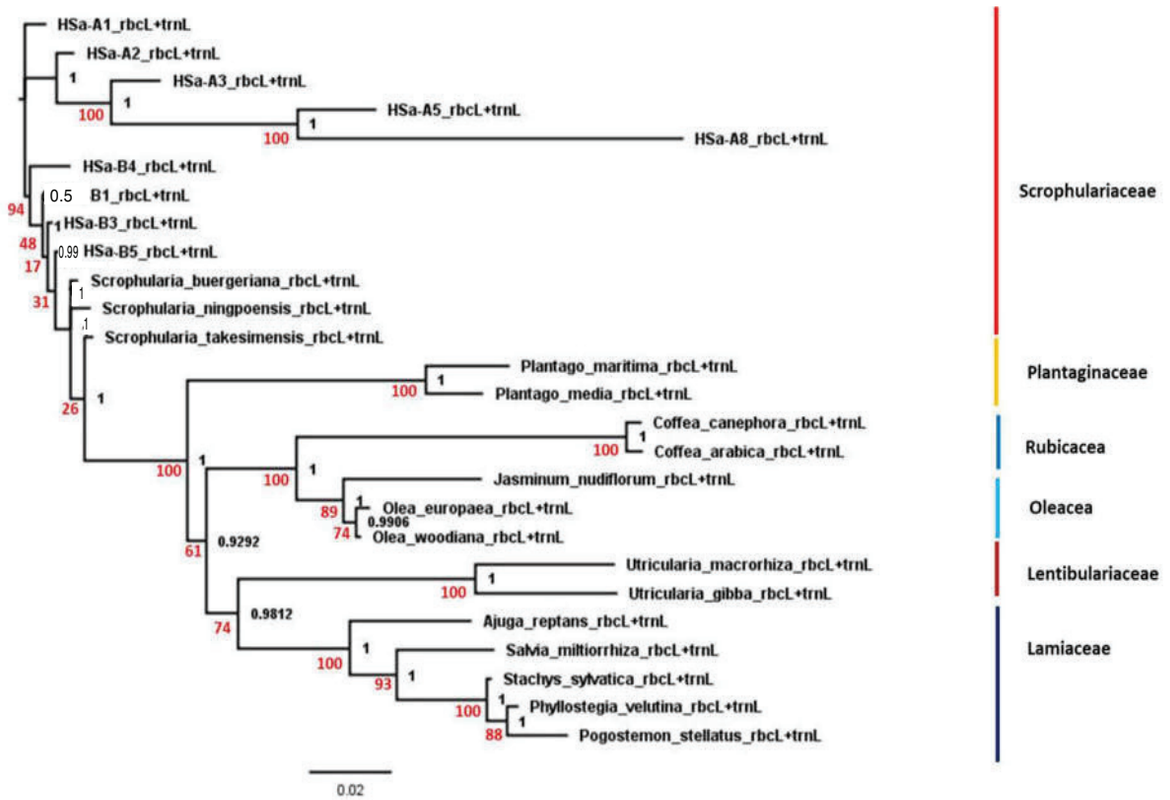


Fig. 3. Bayesian consensus tree *Scrophularia* L., generated from the Chloroplast combination dataset (*rbcL* and *trnL-trnF* intergenic spacers) of tested samples and reference sequences from 5 other genera of *Lamiales*. PP are given in black number next to each node, while the corresponding BS values are given in red numbers. PP values are obtained from 9,002 sampled trees. The scale bar indicates the average expected changes per site of sequences in the study.

of Lamiaceae; however, the PP and BS values for this classification were quite low (PP: 0.57, BS: 23), indicating uncertainty.

Turning to Combined chloroplast consensus tree (Fig. 3), all the nodes were well supported by the BI analysis (i.e. PP values are larger than 0.9) and the clade of outgroups had high values of both BS and PP. The branch length of taxons HSa-A1, HSa-AB1, HSa-A2, HSa-B3, HSa-B5 were similar and grouped together on a clade of the consensus tree. This clade also had a close relationship with *S. buergeriana* and *S. ningpoensis*. Nevertheless, the node between the tested group and *S. buergeriana* received a weak BS value from ML bootstrap analysis (BS: 31). A similar circumstance occurred with the node between the outgroups and tested samples (BS: 26).

The topologies involving the phylogram generated from the mixed combination dataset were relatively congruent with the Chloroplast combination tree (Fig. 4). All the nodes received a good support from BI analysis. The clades including examined samples (HSa-3A, HSa-1B, HSa-1A, HSa-4B, HSa-2A) were highly supported to be the sister of *S.ningpoensis* (PP: 1, BS: 91), while it is quite weak on ITS tree (PP: 0.53, BS: 97). The HSa-5A and HSa-8A were classified into a separated clade with high PP and BS value (PP: 1, BS: 100), suggesting that they belong to a completely different group involved in the tested samples. In addition, all the outgroup clades showed a consistency with the APGII classification (Angiosperm Phylogeny Group, 2003) with a high support from BI and ML bootstrap analysis. This result indicated the high efficiency of using Combined dataset for building phylogenetic trees not only for *Scrophulariaceae,* but also other families of *Lamiales*.

Discussion and conclusions

Scrophularia L. is one small genus which belongs to *Scrophulariaceae* family. The genus includes a relatively small number of species (about 200-300) in comparison

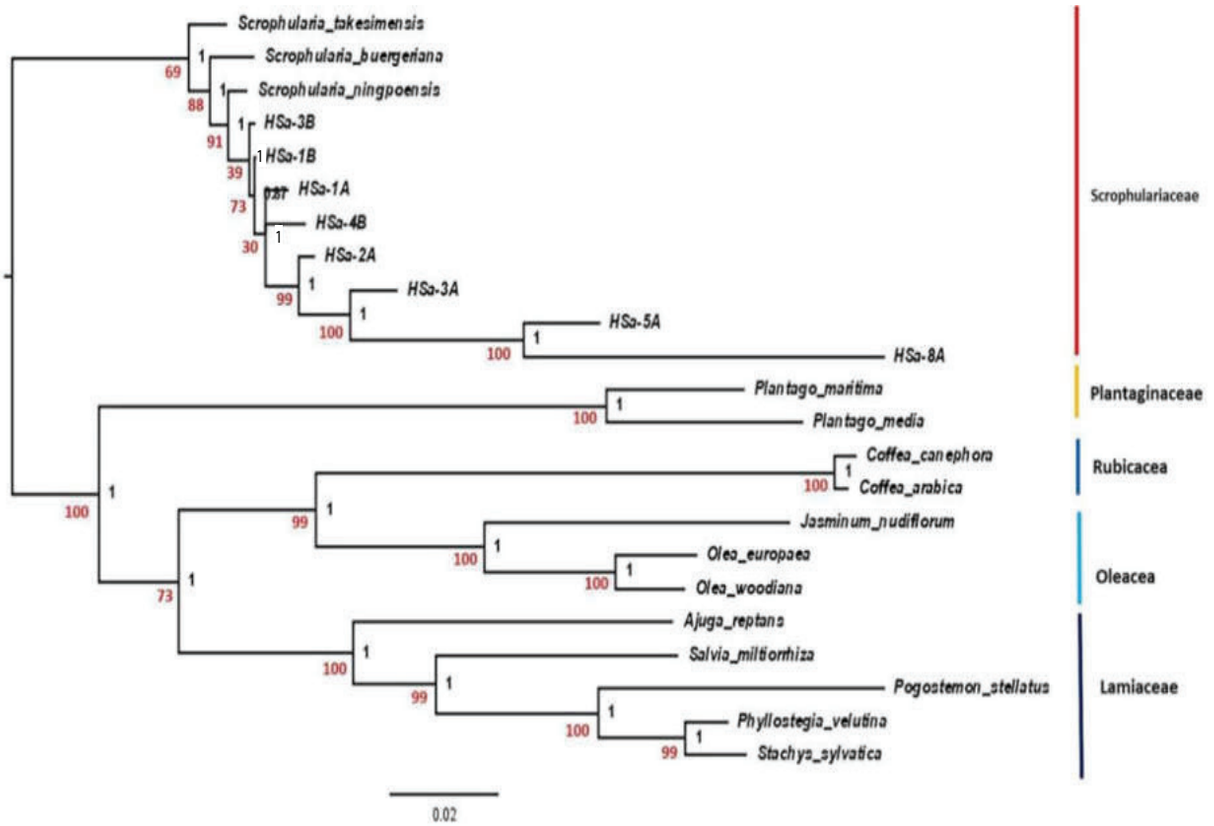


Fig. 4. Bayesian consensus tree *Scrophularia* L., generated from a mixed combination dataset (*ITS, rbcL* and *trnL-trnF* intergenic spacers) of tested samples and reference sequences from 4 other genera of *Lamiales*. PP are given in black number next to each node, while the corresponding BS values are given in red numbers. PP values are obtained from 9,002 sampled trees. The scale bar indicates the average expected changes per site of sequences in the study.

with the total number species of *Scrophulariaceae* family; however, a lot of them are utilized as traditional medicine. In Vietnam, the first study on *Scrophularia* L. was instituted by Do Tat Loi in 1962 under a catalogue of medical plants in Vietnam. The study classified Huyen-sam as *S. buergeriana* Miq. or *S. oldhami* with numerous benefits for human health, such as pulse-quickening, anti-inflammation, and antibiotic [27]. In the Vietnam Plant data Centre, the classification of *Scrophulariaceae* is largely based on the morphology of reproductive trait; e.g. stamen exsertion, corolla shape or leaf organization structure (www.botanyvn.com). However, the morphology knowledge for distinguishing *Scrophularia* L's processed roots which are valuable herbs is unavailable. In this study, we aim to develop a strategy using DNA barcoding, a molecular approach, to support the classification of *Scrophularia* L. samples in Vietnam.

DNA barcode is versatile and cost-effective for the plant taxonomist. The most remarkable advantage of using DNA barcoding is the wide range of applicable plant samples. The method could be applied for DNA samples obtained from different parts of the plant (e.g. leaf, root, flower...) in various kinds of preservation conditions (fresh, dry…). Upon DNA barcode analysis, the taxon identification can be processed without a detailed description of morphology [28]. In addition, DNA barcode information could support the finding of new species from a collection or confirmation of preserved plant materials [29].

Here, we followed a general pipeline of taxonomic analysis of *Scrophularia* L. genus on three controversial barcode regions *(ITS, rbcL* and *trnL-trnF*). For more details, *ITS* is well-known as one of the most importance barcodes for plant classification. This region includes 2 more variable partitions namely ITS1, ITS2 and a conserved 5.8S sequence. Due to the convenience of amplification, the ITS regions are widely used for performing taxonomy analysis of the fungi, monocot and dicot [10, 11]. However, the

ITS has a quite complex evolution pattern that correlates with nuclear genome and causes difficulties for analysis [8]. In this study, we applied a SYM+Γ substitution model which was suggested by Scheunert and colleagues [23] for generating the phylogram from the ITS data. The utilization of ITS dataset was also integrated into a large number of Scrophulariaceae family classification studies [2, 17, 30].

We also did the taxonomy analysis with the two-locus barcode located on chloroplast (*rbcL* and *trnL-trnF*) which are also widely used as plant DNA barcodes for *Scrophulariaceae* [28, 31, 32]. By merging the Chloroplast dataset with the ITS data, we have improved the reliability level of the analysis with a higher value of BI and ML bootstrap analysis. The combination of multi-loci for generating phylogenetic trees is a controversial method for reducing the inconsistency from different single locus analyses and creating a 'total evidence' approach [25]. In addition, the utilization of ITS locus combined with two plastid loci is proposed as the silver standard method for land plant classification [10]. The adding of a locus which has lower rates of evolution in plant plastids such as spacer regions (*trnH-psbA, trnL-trnF*…) and *rbcL* has shown more effective and precise results in separating closely related plants [33].

In conclusion, we have identified a close relationship between the Huyen-sam samples HSa-3A, HSa-1B, HSa-1A, HSa-4B, HSa-2A with the *S. ningpoensis* using combined DNA barcodes generated from 3 loci namely *ITS, rbcL* and *trnL-trnF.* The sequence data generated from this project could enhance the further studies about the diversity of medicinal plant in Vietnam or confirmation of plant material in Vietnam herb market. The phylogenetic analysis followed a general pipeline with BI and ML bootstrap analysis. This strategy could be promisingly applied for not only *Scrophularia* L., but also other valuable herbs of the other genera in Vietnam.

ACKNOWLEDGEMENTS

This study was supported by the Preservation of Vietnamese Herbarium genetic resources for medicine development project. All the experiments and analysis in the research were performed at Institute of Genome Research, Vietnam Academy of Science and Technology, Hanoi, Vietnam.

REFERENCES

[1] E. Fischer (2004), *Scrophulariaceae (in The families and genera of vascular plants)*, Springer, **7**, pp.333-432.

[2] A. Scheunert, G. Heubl (2011), "Phylogenetic relationships among new world *Scrophularia* L. (Scrophulariaceae): New insights inferred from DNA sequence data", *Plant Systematics and Evolution,* **291(1-2)**, pp.69-89.

[3] J. Tian, X. Ye, Y. Shang, Y. Deng, K. He, X. Li (2012), "Preparative isolation and purification of harpagoside and angroside C from the root of *Scrophularia ningpoensis* Hemsley by high-speed counter-current chromatography", *Journal Separation Science*, **35(19)**, pp.2659-2664.

[4] T. Kajimoto, M. Hidaka, K. Shoyama, T. Nohara (1989), "Iridoids from *Scrophularia ningpoensis*", *Phytochemistry,* **28(10)**, pp.2701-2704.

[5] A.P. Sagare, C.L. Kuo, F.S. Chueh, H.S. Tsay (2001), "De novo regeneration of scrophularia yoshimurae Yamazaki (*Scrophulariaceae*) and quantitative analysis of harpagoside, an iridoid glucoside, formed in aerial and underground parts of in vitro propagated and wild plants by HPLC", *Biological and Pharmaceutical Bulletin*, **24(11)**, pp.1311-1315.

[6] Nguyen Tien Ban (2005), *Scrophulariaceae (in Plant category of Vietnam)*, Agricultural Publisher, **3**, p.233.

[7] S. Yang, J. Li, Y. Zhao, B. Chen, C. Fu (2011), "Harpagoside variation is positively correlated with temperature in *Scrophularia ningpoensis* Hemsl", *J. Agricultural Food Chemistry,* **59(5)**, pp.1612-1621.

[8] O. Seberg, G. Petersen (2009), "How many loci does it take to DNA barcode a crocus?", *PLOS ONE,* **4(2)**, p.e4598, doi: 10.1371/journal.pone.0004598.

[9] L.T.T. Hien, H.D. Boer, N.V. Hai, L.T. Huong, N.M. Huong, L. Bjork (2012), "DNA molecule bar code and bar code data system", *Journal of Biotechnology (VN),* **10**, pp.393-405.

[10] M.W. Chase, et al. (2005), "Land plants and DNA barcodes: short-term and long-term goals", *Philosophical Transactions of the Royal Society, Biological Sciences,* **360(1462)**, pp.1889-1895.

[11] W.J. Kress, et al. (2005), "Use of DNA barcodes to identify flowering plants", in *Proceedings of the National Academy of Sciences of the United States of America,* **102**, pp.8369-8374.

[12] D.Y. Wang, et al. (2017), "Evaluation of DNA barcodes in Codonopsis (Campanulaceae) and in some large angiosperm plant genera", *PLOS ONE,* **12(2)**, p.e0170286, doi.org/10.1371/journal.pone.0170286.

[13] J.M. Saarela, et al. (2013), "DNA barcoding the Canadian Arctic flora: core plastid barcodes (*rbcL* + *matK*) for 490 vascular plant species", *PLOS ONE,* **8(10)**, p.e77982, doi.org/10.1371/journal.pone.0077982.

[14] X. Guo, X. Wang, W. Su, G. Zhang, R. Zhou (2011), "DNA barcodes for discriminating the medicinal plant Scutellaria baicalensis (Lamiaceae) and its adulterants", *Biological and Pharmaceutical Bulletin,* **34(8)**, pp.1198-1203.

[15] C. Garcia Robledo, et al. (2013), "Tropical plant-herbivore networks: reconstructing species interactions using DNA barcodes", *PLOS ONE,* **8(1)**, p.e52967, doi.org/10.1371/journal.pone.0052967.

[16] L.T. Hương, N.N. Linh, B.M. Minh, H.H. Hạnh, H.T.T. Hue, N.V. Hai, H.V. Huan, L.T.T. Hien (2017), "Applied DNA barcode assists

in the identification of some ginseng species in the genus ginseng (*Panax L.*)", *Journal of Biotechnology (VN)* **15**, pp.63-72.

[17] A. Scheunert, G. Heubl (2014), "Diversification of Scrophularia (Scrophulariaceae) in the western mediterranean and macaronesia-phylogenetic relationships, reticulate evolution and biogeographic patterns", *Molecular Phylogenetics and Evolution,* **70**, pp.296-313.

[18] G.N. Feliner, J.A. Rosselló (2007), "Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants", *Molecular Phylogenetics and Evolution,* **44(2)**, pp.911-919.

[19] J.D. Clarke (2009), "Cetyltrimethyl ammonium bromide (CTAB) DNA miniprep for plant DNA isolation", *Cold Spring Harbor Protocols*, **2009(3)**, doi: 10.1101/pdb.prot5177.

[20] T.A. Hall (1999), "BioEdit (version 7.0.5.1): a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT", *Nucl. Acids Symp. Ser.,* **41**, pp.95-98.

[21] F. Ronquist, J.P. Huelsenbeck (2003), "MrBayes 3: Bayesian phylogenetic inference under mixed models", *Bioinformatics*, **19(12)**, pp.1572-1574.

[22] F. Ronquist, et al. (2012), "MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space", *Systematic Biology*, **61(3)**, pp.539-542.

[23] A. Scheunert, et al. (2012), "Phylogeny of tribe Rhinantheae (Orobanchaceae) with a focus on biogeography, cytology and re-examination of generic concepts", *Taxon*, **61(6)**, pp.1269-1285.

[24] K. Tamura, et al. (2013), "MEGA6: molecular evolutionary genetics analysis version 6.0", *Molecular Biology and Evolution,* **30(12)**, pp.2725-2729.

[25] J.P. Huelsenbeck, J.J. Bull, C.W. Cunningham (1996), "Combining data in phylogenetic analysis", *Trends in Ecology & Evolution*, **11(4)**, pp.152-158.

[26] M.A. Perez Gutierrez, et al. (2012), "Phylogeny of the tribe Fumarieae (Papaveraceae s.l.) based on chloroplast and nuclear DNA sequences: evolutionary and biogeographic implications", *American Journal of Botany*, **99(3)**, pp.517-528.

[27] D.T. Loi (2004), *Vietnamese medicinal plants and herbs*, Medical Publishing House.

[28] D. Do, L.Z. Drábková (2017), "Herbarium tale: The utility of dry specimens for DNA barcoding Juncaceae", *Plant Systematics and Evolution*, **304(2)**, pp.1-14.

[29] D.E. Schindel, S.E. Miller (2005), "DNA barcoding a useful tool for taxonomists", *Nature*, **435**, p.17, doi:10.1038/435017b.

[30] A. Scheunert, G. Heubl (2017), "Against all odds: Reconstructing the evolutionary history of Scrophularia (Scrophulariaceae) despite high levels of incongruence and reticulate evolution", *Organisms Diversity & Evolution*, **17(2)**, pp.323-349.

[31] K.L. Bell, V.M. Loeffler, B.J. Brosi (2017), "An *rbcL* reference library to aid in the identification of plant species mixtures by DNA metabarcoding", *Applications in Plant Sciences*, **5(3)**, doi: 10.3732/apps.1600110.

[32] J.H. Chau, N. O'Leary, W.B. Sun, R.G. Olmstead (2017), "Phylogenetic relationships in tribe Buddlejeae (Scrophulariaceae) based on multiple nuclear and plastid markers", *Botanical Journal of the Linnean Society*, **184(2)**, pp.137-166.

[33] E. Pennisi (2007), "Wanted: a barcode for plants", *Science*, **318(5848)**, pp.190-191, doi: 10.1126/science.318.5848.190.

Study on soil biology in Vietnam - achievements and challenges

Quang Manh Vu*

Center for Biodiversity Resources Education & Development (CEBRED), Hanoi National University of Education (HNUE)

Received 4 July 2017; accepted 5 January 2018

Abstract:

The article introduces the achievements and challenges in the research on the soil biology of Vietnam. It is focussed on microfauna, mesofauna and macrofauna, including families of arachnids (Arachnida), chilopods (Chilopoda), diplopods (Diplopoda), insects (Insecta), oligochaetes (Oligochaetes), and nine orders. Until present, the soil fauna diversity of Vietnam is known to have 1,809 species and subspecies, belonging to 687 genera and subgenera, and 195 families and subfamilies. The number of soil animal species identified have decreased in the following order: (1) Araneida: 491 > (2) Oribatida: 320 > (3) Hymenoptera: Formicidae: 307 > (4) Oligochaeta: 212 > (5) Diplopoda: 136 > (6) Collembola: 132 > (7) Isoptera: 101 > (8) Chilopoda: 71 > (9) Scorpionida: 39.

Basing on the study of the oribatid mites (Oribatida) fauna, and the study results obtained during the period from 1977 until now, it also proposes further research directions on the soil biology of Vietnam as followings: (i) Study the biodiversity of soil organisms, (ii) Study ecology and function of soil organisms, (iii) Study of soil organisms contributes to the conservation and sustainable management of the environment and soil ecosystems, and (iv) Study soil organisms as indicators of environmental climate change in Vietnam.

Keywords: ***achievements, challenges, further research, soil fauna, Vietnam.***

Classification number: *3.4*

Introduction

Vietnam has been regarded as one of the 16 countries containing the highest biodiversity values in the World. The discovery of a large animal, the Sao la (*Pseudoryx nghetinhensis*) in 1993 from Vietnam, has received great attention not only from scientists but also from public media, globally (*World Conservation Monitoring Centre, 1992*). This forest-dwelling bovine, Saola, Vu Quang ox, or Asian bicorn, also, infrequently, is one of the world's rarest large mammals, found only in the Annamite Range of Vietnam and Laos.

The above-mentioned example of Saola is about a large mammal; and how about the animal communities that live in the soil of Vietnam? Soil biology is the science of the study of microbial and faunal activity and ecology in soil. Soil life, soil biota, soil fauna or edaphon is a collective term that encompasses all organisms that spend a significant portion of their life cycle within a soil profile or at the soil-litter interface. Generally, soil animals are ranged according to their size as microfauna, mesofauna, macrofauna and megafauna, with 20-200 µm, >0,2-2,0 mm, >2,0-20,0 mm and >20,0 mm, respectively *(wikipedia.org/wiki/Soil_biology*). The studies of soil animals in Vietnam were firstly carried out by Thai Tran Bai and his colleagues in the 1970-1980s of the last century [1]. As a result, it is found that Vietnam's soil fauna is extremely diverse and has a great biodiversity. In general, our understanding about the soil animals of Vietnam are not enough and still insufficient [2-6].

This paper introduces the achievements and challenges in research on the soil fauna of Vietnam. It is focussed on soil microfauna, mesofauna and macrofauna such as arachnids (Arachnida), chilopods (Chilopoda), diplopods (Diplopoda),

*Email: vqmanh@hnue.edu.vn

insects (Insecta), and oligochaetes (Oligochaetes) [7, 8]. Basing on the case study of the oribatid mites (Oribatida) fauna, it also proposes further research directions for soil biology to address climate environment change in Vietnam [9].

Soil fauna diversity in Vietnam

Spiders (Arachnida: Araneida)

Vietnam's spider (Arachnida: Araneida) fauna was first undertaken by Simon (1886, 1903, 1904 & 1909) and Hogg (1922). After that, there has been a break for over 60 years until Zabka (1985) reported his results on taxonomy of Salticidae species, in which 51 species and 8 genera were described as new for science. Hereafter and Ono (1999, 2002, 2003, 2004a & b) described 4 new species for science of the family Liphistiidae, and 11 new species for science of the family Zodariidae. Peng and Li (2003) reported new localities of 13 species of jumping spiders, including one that was described as a new species for science. Tu and Li (2004 & 2006) did a primary report on the family Linyphiidae from Vietnam with 7 new species for science and 4 new records. Grismado and Ramizez (2004) described one new Zodariidae species. Jager and Vedel (2005) also reported one Sparassidae species from Vietnam. In the time between 2009 and 2015, 53 new species of spiders (Lin, Pham & Li, 2009; Liu, Li & Pham, 2010; Zhang, Li & Pham, 2013; Yao, Pham & Li, 2015; Pham, 2015) were recorded. A total of 205 spider species are recorded only in Vietnam so far, of which 78 species were reported by Simon (1886, 1903, 1904 & 1909), 2 species by Walckenaer (Simon, 1909), 1 species by Gunther (Simon, 1909), and 10 species by Hogg (1922); and the rest of the species were reported in recent years (after Pham Dinh Sac 2015 [10]).

At present, 491 spider species (Arachnida: Araneida) belong to 219 genera and 43 families are known from Vietnam (Table). It is probably that many of them are endemic, and further study on spider fauna, not only from Vietnam but also from other Asian Southeast countries, particularly Laos and Cambodia, is needed to be carried out.

Scorpions (Arachnida: Scorpionida)

After Vu Quang Manh (2008), Vu Quang Manh & Nguyen Dang Thin (2009) firstly presented a list of 17 species and sub-species of scorpions belonging to five families and seven genera, known in Vietnam at that time. They were as follows: I. Family BUTHIDAE C.L. Koch, 1837: (1) *Isometrus maculates* (DeGeer, 1778), (2) *Isometrus vittalus* Pocock, 1900, (3) *Lichas mucrotus* (Fabricius, 1798); II. Family CHAERILIDAE Pocock, 1893: (4) *Chaerilus celebensis* Pocock, 1894, (5) *Chaerilus truncatus* Karsch, 1879, (6) *Chaerilus variegatus* Simon, 1877, (7) *Chaerilus variegatus variegatus* Simon, 1877; III. Family ISCHNURIDAE Simon, 1879: (8) *Liocheles austrasiae* (Fabricius, 1775); IV. Family SCORPIONIDAE Latreille, 1802: (9) *Heterometrus laoticus* Couzigin, 1981, (10) *Heterometrus patersii* (Thorell, 1876), (11) *Heterometrus patersii patersii* (Thorell, 1876), (12) *Heterometrus spinifer* (Ehrenberg, 1828), (13) *Heterometrus spinifer spinifer* (Ehrenberg, 1828), and (14) *Heterometrus liurus* (Pocock, 1897); and V. Family SCORPIOPIDAE Kraepelin, 1905: (15) *Euscorpiops kaftani* (Kovarík, 1993), (16) *Scorpiops montanus* Karsch, 1879, and (17) *Scorpiops oligotrichus* Fage, 1933. Among them, *Euscorpiops kaftani*, recorded in the Cuc Phuong National Park by Kovarik, in 1993, is an endemic one. Vu Hong Quang (1996), Vu Hong Quang, Vu Quang Manh, Le Xuan Hue (1996), have made an important contribution to scorpion control and management (after Vu Hong Quang 1996 [11]; Vu Quang Manh, Nguyen Dang Thin 2009 [12]).

In recent years, the Vietnam's fauna of scorpions (Arachnida: Scorpionida) have being been investigated actively by Phan Dinh Sac and his collaborators, with descriptions of many species that are new for science (After Lourenço, Dinh-Sac Pham, 2010; Lourenco, Pham Dinh Sac, 2015; Dinh Sac Pham, Thi Hang Tran, Wilson R. Lourenço, 2017) [13]. At present, 39 scorpion species belong to 12 genera and 8 families have been recorded in Vietnam (Table).

Oribatid mites (Arachnida: Oribatida)

The investigation on the oribatid mites (Acari: Oribatida) of Vietnam started in 1967, with identifications of 33 oribatid species recovered from Vietnam. All were new for the fauna of Vietnam, and included 29 species and 4 genera that were new for science (Balogh & Mahunka, 1967). In addition to the first research done by Hungarian Balogh and Mahunka (1967), very important contributions to the knowledge of the oribatid mites of Vietnam were conducted by other foreign authors, including those of Rajski, Szudrowicz (1974), Golosova (1983, 1984), Golosova, et al. (1985), Jeleva, Vu (1987), Zonev, Vu (1987), Mahunka (1987, 1988, 1989), Behan-Pelletier (1989), Pavlichenko (1991, 1994), Stary (1993), Krivolutsky, Vu and Phan (1998) (after Vu Quang Manh, 2015b [9]; Krivolutzkij, Vu, Phan, 1998 [14]).

In 1986, the research work of Vu Quang Manh titled '*Faunal - Ecological Studies on oribatid mites (Acarina:*

Oribatei) community in Northern Vietnam' defended successfully at the Sofia University St. Kliment Ohridski (Bulgaria), for a PhD title. In this work, the author has identified 73 oribatid species from northern Vietnam, including 39 that were recorded as new for the fauna of Vietnam and 7 species described as new for science [15]. In the monograph entitled '*Fauna of Vietnam T. 21. Acari: Oribatida*', the author has included 150 Oribatid species and subspecies known to be found in Vietnam, including 44 species, which represent new records for the fauna of Vietnam [16].

In the base of the studies carried out in Vietnam during the period of 1979 until 2014, and based on the oribatid materials obtained throughout the country, it is found that the oribatid fauna of Vietnam is represented by 320 species (including four subspecies) belonging to 163 genera, 64 families (including two subfamilies) and 30 super families. It is highly diversified, with a high number of species with limited distribution; 34.68% of the total number are probably endemic species. One hundred and fifty-five species, representing 48.44% of the total oribatid fauna, were recorded for the first time as part of the fauna of Vietnam (Table). The oribatid fauna of Vietnam occupies 3.09% (320 vs. 10,342 species), 13.05% (163 vs. 1,249 genera) and 38.0% (62 vs. 163 families) of the World oribatid fauna [9, 16-20].

Recently, the Russian colleague Sergey G. Ermilov and his foreign collaborators have made an important contribution to the knowledge of the oribatid mite fauna of the Southeast and Southwest of Vietnam. Ermilov (2015) introduces a list of 535 oribatid mites (Acari: Oribatida) species of Vietnam, from 222 genera and 81 families [21]. However, some of their oribatid specimens have been obtained by unofficial lines, and Ermilov rarely collected materials by himself in the field. Some of their data on the geographical locations, names of locations and natural conditions of Vietnam have been obtained from illegal sources [22]. Particularly, the administrative maps of the socialist republic of Vietnam in their publications were presented wrongly, lacking islands and the island regions [21, 23, 24]. Therefore, the data on Vietnam's oribatid mites given by the Russian colleague Sergey G. Ermilov must be checked and revised carefully before consulting them [21]. Even so, although lacking knowledge of Vietnam's nature, this is an encouraging attempt by the Russian colleague on the study of oribatid mites of Vietnam, and it confirms the importance of studying about soil animals in Vietnam.

Centipedes (Myriapoda: Chilopoda)

In 2013, on the base mainly of the literature, the centipede (Chilopoda) fauna of Vietnam was introduced by Tran, Le, Nguyen [25]. As a result, a total of 71 species belonging to 26 genera and 13 families of four orders, namely Scolopendromorpha, Geophilomorpha, Lithobiomorpha and Scutigeromorpha, has been registered from Vietnam (Table). Four genera, *Tonkinodentus, Alluropus, Anopsobiella* and *Megalacrus*, are monotypic, and twenty-two species as well as subspecies are known only in Vietnam.

Millipedes (Myriapoda: Diplopoda)

In 2004, the millipede (Diplopoda) fauna of Vietnam was firstly reviewed by Enghoff, Henrik, Golovatch and Nguyen Duc Anh (2004) (after Enghoff, et al., 2004 [26]). As a result, a list of 136 millipede species belonging to 74 genera and 28 families were introduced (Table).

Springtails (Insecta: Collembola)

Being one of the main components of soil microarthropods (Microarthropoda), springtails or collembolans as well as oribatid mites are subject of numerous research works in almost every region all over the world. In Vietnam, the study on the collembolans fauna are insufficient [2, 3, 17].

Faunal-ecological studies on collembolans fauna have been carried out continuously by Vietnamese specialists [27-29]. Untill the present, the collembolan fauna (Apterygota: Collembola) of Vietnam are represented by 132 species belonging to 62 genera and 18 families (Table).

Ants (Insecta: Hymenoptera)

The ant fauna (Hymenoptera: Formicoidea: Formicidae) of Vietnam was firstly investigated by Bingham (1903), Santchi (1920) and Karawajew (1935). Based on their study results, the ant fauna of Vietnam was recorded to have about 160 species. The recent studies of Dlussky and Radchenko (1988), Radchenko (1993) and Bui Tuan Viet (2002a, 2002b, 2005) have introduced the ant fauna of Vietnam with 307 species recorded, belonging to 74 genera and 9 subfamilies (Table) (after Viet Bui Tuan, 2005 [30]).

Termites (Insecta: Isoptera)

Bathellier, in 1927, firstly studied Indochina's termite fauna (Blattodea: Isoptera), including Cambodia, Laos, and Vietnam. After this study, the termite fauna of North and South Vietnam have been investigated by Nguyen Duc

Kham (1969) and Nguyen Tan Vuong (1997). Vu Van Tuyen (1982) is one of the Vietnamese leading specialists on termites and has contributed great achievements in termite control (1994, 2004, 2010) (after Nguyen Duc Kham, et al., 2007 [31]).

In the study by Q.M. Vu, H.H. Nguyen, R. Smith (2007) on the termites (Isoptera) of Xuan Son National Park, a lowland and lower mountain evergreen and limestone forest in northern Vietnam, 15 species in 8 genera and 2 families were recorded. Termitidae was the dominant family with 6 genera and 12 species. The genus *Odontotermes* with five, contained the largest number of species. Five species were new records for northern Vietnam: *Odontotermes maesodensis* Ahmad, 1965, *Nasutitermes ovatus* Fan, 1983, *Pericaptitermes latignathus* (Holmgren, 1913), *Pericaptitermes nitobei* Shiraki, 1909 and *Bulbitermes laticephalus* Ahmad, 1965. The inventory included eight fungus-growing species: *Macrotermes barneyi* Light, 1924, *Ma. annandalei* (Silvestri, 1914), *O. yunnanensis* Tsai et Chen, 1963, *O. hainanensis* Light, 1924, *O. formosanus* Sharaki, 1909, *O. maesodensis* Ahmad, 1965, *O. graveli* (Silvestri, 1914) and *Microtermes pakistanicus* Ahmad, 1965. Five species, *M. barneyi*, *O. yunnanensis*, *O. hainanensis*, *O. formosanus* and *M. pakistanicus*, occurred in all habitat types. Six species identified are considered special pests because their activities weaken earthen structures. They are as follows: *M. pakistanicus, M. barneyi*, *M. annandalei*, *O. yunnanensis*, *O. hainanensis* and *O. formosanus* (after Vu, Nguyen, Smith, 2007 [32]).

Up to 2007, Nguyen Duc Kham, et al. have presented a list of 101 termite species known from Vietnam, belonging to 33 genera and four families [31] (Table).

Earthworms (Annelida: Oligochaeta)

Vietnam's earthworm fauna was firstly investigated by Edmond Perrier (1872 & 1875) with a description of *Perionyx excavatus* Perrier, 1872, and then with the second species *Amynthas juliani* (Perrier, 1875). Additionally, earlier there was a record of 2 species, namely *Metaphire posthuma* (Vaillant, 1868) (as *Perichaeta affinis* Perrier, 1872) and *Pheretima houlleti*.

Recently, from studies of Thai (1985) through today (T. Tung Nguyen, D. Nguyen Anh, T.T. Nguyen Binh, Robert J. Blackemore, 2016), Vietnam's earthworm fauna has been recorded with 212 species, belonging to 24 genera and 8 families [33, 34].

The oribatid mites (Acari: Oribatida) fauna of Vietnam

Introduction

The case study of the oribatid mites (Acari: Oribatida) fauna of Vietnam has been carried out during the period of 1977 until now [9]. The investigation on the oribatid mites (Acari: Oribatida) of Vietnam started in 1967, with identifications of 33 oribatid species recovered from Vietnam. All were new for the fauna of Vietnam, and included 29 species and 4 genera that were new for science (Balogh & Mahunka, 1967). The study results obtained indicated that oribatid mites of Vietnam are very diverse, and they have a bioindicator potential (Vu, 1990, 2000; Vu, et al., 1985, 1987, 1995, 2002; Vu, et al., 2010, 2011, 2012, 2013; Dao, Vu, 2010, 2013; Nguyen, Vu, 2011, 2012, 2013, [2, 5, 14-16, 18-20]).

Ecological studies based on the oribatid mite community structure have also been conducted according to landscape, altitudinal zonations, soil and habitat type, and season. They were carried out in a number of national parks (NP) Xuan Nha (province of Son La), NP Xuan Son (Phu Tho), NP Tam Dao (Vinh Phuc), NP Cuc Phuong (Ninh Binh), NP Ba Vi (Ha Noi), NP Cat Ba (Hai Phong), in uplands and the Delta of the Hong river, NP Ben En (Thanh Hoa), NP Phong Nha - Ke Bang (Quang Binh), as well as in some sites in Central and Southern Vietnam (Vu, 2004, 2006, 2008; Vu, et al., 2002, 2003, 2009; Vu, Nguyen, 2000; Dao, Vu, 2011; Ermilov, Anichkin, 2013; Minor, Ermilov, 2015 [2, 7, 8, 15, 17-20, 23, 24]).

In general, the studies on the Oribatid mites of Vietnam can be divided into 3 main periods as follows:

1. 1967-1986: This is the period when a total of 73 species have been recorded from 33 known species. In 1967, 33 oribatid species of Vietnam were identified (Balogh & Mahunka, 1967) (after Vu Quang Manh, 2015b [9]). All these species were recorded as new for Vietnam, including 29 species and 4 genera new for science. In 1980, Vu Quang Manh found that, the soil arthropods, Oribatida and Collembola, are much diversified and still poorly known in Vietnam. This author has identified 73 oribatid species of northern Vietnam, including 39 species recorded new for the Vietnam's fauna, and 7 species new for science.

2. 1987-2007: This is the period when a total of 150 species have been recorded from 73 known species. In 2007, in his study titled '*Fauna of Vietnam T. 21. Acari: Oribatida*' Vu Quang Manh presented 150 oribatid species known in Vietnam, with 44 species recorded new for the

fauna.

3. 2008-2014: This is the period when a total of 320 species have been recorded from 150 known species. The oribatid mite fauna of Vietnam has been recorded with 320 species and sub-species, belonging to 60 families and 163 genera. Among them, 111 species (34.68% of the total) have been described as new for science, and recorded only in Vietnam.

Results and discussions

Based on the studies carried out in Vietnam during the period of 1979 until now, and based on the oribatid materials obtained throughout the country [9], the most important results obtained are as follows:

1. Up to December 2014, the oribatid fauna of Vietnam is represented by 320 species (including four subspecies) belonging to 163 genera, 64 families (including two subfamilies) and 30 superfamilies. It is highly diversified, with a high number of species with limited distribution; 34.68% of the total number are probably endemic species. One hundred and fifty-five (155) species, representing 48.44% of the total oribatid fauna, were recorded for the first time for the fauna of Vietnam. The oribatid fauna of Vietnam occupies 3.09% (320 vs. 10,342 species), 13.05% (163 vs. 1,249 genera) and 38.0% (62 vs. 163 families) of the World oribatid fauna.

2. According to the number of families, genera, species and subspecies recorded, the systematic structure of the oribatid fauna of Vietnam is diverse. However, the number of genera per family, as well as the number of species and subspecies per genus, is not high. Almost all the families consist of one or 2-3 genera (42.2% and 39.1%, respectively, of the 64 families and subfamilies). A majority of the genera are represented by one species (68.10% of the 163 genera). Only one family is represented by more than 10 genera, i.e. the family Oppiidae Grandjean, 1954, with 23 genera. Only two genera are represented by more than 10 species: *Galumna* Heyden, 1826 and *Pergalumna* Grandjean, 1936, represented by 13 and 11 species, respectively. It is shown that *P. arboriseta* Jeleva et Vu is clearly a species distinct from *P. hirsutus* (Aoki 1961) (after Vu Quang Manh, 2015 [9]).

3. The zoogeographical structure of the oribatid fauna of Vietnam is highly diversified. It consists mostly of Oriental species represented by 60.3% (193 species out of 320 species recorded). Other zoogeographical elements represented are Palaearctic-Oriental species (12.2%), cosmopolitan species (10.6%), Afrotropical (Ethiopical) species (6.9%), Australian-Oriental species (5.0%), Neotropical-Oriental species (3.8%), Nearctic-Oriental species (0.9%) and Pacific-Oriental species (0.3%). A substantial part of the oribatid fauna of Vietnam - 111 species, representing 34.68% - consists of species with distribution restricted only to the country. These are probably endemic species.

4. The oribatid fauna in Vietnam is grouped into three main zoogeographical regions. There are differences between these regions of the country, and even between different sub-regions of these three parts. The three main zoogeographical regions of Vietnam are the following: (A) The region between (I) Northwest, (II) Northeast, (IV) Red River Delta, (V) Red River Delta: NP Cat Ba Island, (VI) North Central; (B) The region between (III) Red River Delta: Uplands and (VII) Central North: NP Phong Nha - Ke Bang, and (C) The region between (VIII) Southern - Mekong River Delta: NP Bu Gia Map, (IX) Southern - Mekong River Delta: NP Cat Tien.

5. From the North to South of Vietnam, the distribution of the oribatid fauna can be divided into six zoogeographical zonations, as follows:

(i) Region between (I) Northwest and (II) Northeast (North Vietnam) of Vietnam, with eight characteristic oribatid species including (1) *Papilacarus arboriseta* (Jeleva et Vu, 1987), (2) *Nothrus baviensis* (Krivolutsky, 1998), (3) *Nothrus montanus* (Krivolutsky, 1998), (4) *Gibbicepheus baccanensis* (Jeleva et Vu, 1987), (5) *Leobodes monstruosus* (Jeleva et Vu, 1987), (6) *Perxylobates brevisetus* (Mahunka, 1988), (7) *Xylobates monodactylus* (Haller, 1884), and (8) *Scheloribates cruciseta* (Jeleva et Vu, 1987).

(ii) Region of Red River Delta (IV) (North Vietnam) of Vietnam, with four characteristic oribatid species including (1) *Kokoppia dendricola* (Jeleva et Vu, 1987), (2) *Perxylobates vietnamensis* (Jeleva et Vu, 1987), (3) *Scheloribates praeincisus* (Berlese, 1916), and (4) *Lamellobates ocularis* (Jeleva et Vu, 1987).

(iii) Region of NP Cat Ba Island (V) of the Red River Delta (North Vietnam) of Vietnam, with two characteristic oribatid species including (1) *Scheloribates laevigatus*, and (2) *Fissicepheus elegans* (Balogh et Mahunka, 1967).

(iv) Region between (III) the Uplands of the Red River Delta and (VII) NP Phong Nha - Ke Bang: Central North (North Vietnam and Central North Vietnam) of Vietnam, with four characteristic oribatid species including (1) *Tectocepheus cuspidentatus* (Knulle, 1954),

(2) *Austrachipteria phongnhae* (Ermilov et Vu, 2012), (3) *Scheloribates praeincisus* (Berlese, 1916), and (4) *Galumna kebangica* (Ermilov et Vu, 2012).

(v) Region of (VI) NP Ben En: North Central (North Central Vietnam) of Vietnam, with six characteristic species including (1) *Papilacarus benenensis* (Vu, Ermilov et Dao, 2010), (2) *Setoxylobates foveolatus* (Balogh et Mahunka, 1967), (3) *Perxylobates thanhhoaensis* (Ermilov, Vu, Trinh et Dao, 2010), (4) *Xylobates lophotricus* (Belese, 1904), (5) *Galumna tenensis* (Ermilov, Vu et Nguyen, 2011) and (6) *Pergalumna granulatus* (Balogh et Mahunka, 1967).

(vi) Region between (VIII) Southern - Mekong River Delta: NP Bu Gia Map and (IX) Southern - Mekong River Delta: NP Cat Tien (South Vietnam) of Vietnam, with eight characteristic oribatid species including (1) *Arthrodamaeus vietnamicus* (Ermilov et Anichkin, 2011), (2) *Acrotocepheus (Otocepheus) vietnamicus* (Ermilov et Anichkin, 2011), (3) *Unguizetes cattienensis* (Ermilov et Anichkin, 2011), (4) *Galumna levisensilla* (Ermilov et Anichkin, 2010), (5) *Galumna pseudokhoii* (Ermilov et Anichkin, 2010), (6) *Neogalumna seniczaki* (Ermilov et Anichkin, 2010), (7) *Pergalumna indistincta* Ermilov et Anickin, 2011, and (8) *Pergalumna yurtaevi* (Ermilov et Anickin, 2011).

6. The diversity analysis of oribatid communities in various habitat types shows that the most particular and distinctive communities are formed in natural forests; under agricultural intensification and in human-disturbed habitats these communities are changed, with communities in grasslands and shrubs occupying an intermediate position between natural forests and human-affected habitats. The habitat of the grasslands and scrubs can play the role of a transformational ecosystem for reestablishment of the soil oribatid mite community. A tendency for formation of two distinct oribatid communities in the two main soil types studied (alluvial and ferralitic) is detected. It is possible that the soil type plays a major role in determining the species composition of the oribatid communities. Three oribatid species, namely *Tectocepheus velatus* (Michael, 1880), *Scheloribates praeincisus* (Berlese, 1916) and *Lamellobates ocularis* (Jeleva et Vu, 1987), are the most widespread and most persistent species of northern Vietnam. They can be considered as bioindicators of disturbed soil ecosystems.

Future research direction for soil fauna in Vietnam

Recently, the major soil animal groups that have been studied in Vietnam include nine orders (1) spiders (Araneida), (2) scorpions (Scorpionida), (3) oribatid mites

Table. Soil fauna diversity by number of species, genera and families.

Taxa animal classes and orders	Number of species & subspecies	Number of genera & subgenera	Number of families & subfamilies	Total
i.1. Spiders (Arachnida: Araneida)	491	219	43	491/219/43
i.2. Scorpions (Arachnida: Scorpionida)	39	12	8	39/12/8
i.3. Oribatid mites (Acari: Oribatida)	320	163	64	320/163/64
ii.4. Centipedes (Myriapoda: Chilopoda)	71	26	13	71/26/13
iii.5. Millipedes (Myriapoda: Diplopoda)	136	74	28	136/74/28
iv.6. Springtails (Insecta: Apterygota: Collembola)	132	62	18	132/62/18
iv.7. Ants (Insecta: Hymenoptera: Formicidae)	307	74	9	307/74/9
iv.8. Termites (Insecta: Isoptera)	101	33	4	101/54/4
v.9. Earthworms (Annelida: Oligochaeta)	212	24	8	212/24/8
Total: 5 families and 9 orders	**1809**	**697**	**195**	1809/687/195

(Oribatida), (4) centipedes (Chilopoda), (5) millipedes (Diplopoda), (6) springtails (Collembola), (7) ants (Hymenoptera: Formicidae), (8) termites (Isoptera), and (9) earthworms (Oligochaeta). They belong to five animal classes, namely (I) Archnids (Arachnida), (II) Centipedes (Chilopoda), (III) Millipedes (Diplopoda), (IV) Insects (Insecta), and (V) Earthworms (Oligochaeta). Until this moment, the soil fauna diversity of Vietnam is known with 1,809 species and subspecies, belonging to 687 genera and subgenera, and 195 families and subfamilies (Table).

The number of soil animal species identified decreases in the following order: (1) Araneida: 491 > (2) Oribatida: 320 > (3) Hymenoptera: Formicidae: 307 > (4) Oligochaeta: 212 > (5) Diplopoda: 136 > (6) Collembola: 132 > (7) Isoptera: 101 > (8) Chilopoda: 71 > (9) Scorpionida: 39 (Table). The order mentioned above does not mean that the species diversity of those soil animal is different in soil ecosystems, but it is only in relation to the level of research of each soil animal group in Vietnam.

In general, our understandings on soil fauna of Vietnam are insufficient [2, 7, 8, 35, 36]. Therefore, further research on soil biology in Vietnam will include the following five main directions:

1. Study the biodiversity of soil organisms.

2. Study the ecology and function of soil organisms.

3. Study of soil organisms contributes to the conservation and sustainable management of the soil environment.

4. Study soil organisms as indicators of environmental climate change in Vietnam.

5. Study of soil organisms contributes to the training of soil biology specialists.

ACKNOWLEDGEMENTS

This report is funded partially by the Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 106.14-2012.46.

REFERENCES

[1] Thai Tran Bai, Vu Quang Manh, et al. (1986), "Results of fundamental and practical studies on the soil invertebrate are community in Vietnam", *Research study contributing to educational and socio-economical development*, Hanoi University of Education, **1**, pp.68-74.

[2] Vu Quang Manh (1993), *Final report of the scientific-fundamental project Soil faunal community structures and their role in typical ecosystem of Vietnam*, code: KT/04.05.01.13.

[3] Vu Quang Manh & Vu Van Tuyen (Eds.) (1993), "Studies on soil ecology in Vietnam", *Journal of Biology,* special issue, **15(4)**, pp.1-78.

[4] Vu Quang Manh (Editor-in-Chief) (1995), *Soil biodiversity*, Science and Technology Publishing House, pp.1-136.

[5] Vu Quang Manh (Ed.) (2000), *Soil organisms resources and sustainable development of soil ecosystem*, Agr. Publishing House, pp.1-324.

[6] Vu Quang Manh (2003), *Soil ecology*, HNUE Publishing House, pp.1-265.

[7] Vu Quang Manh (2011), *Final report of the NAFOSTED project Arthropod community structures (Arthropoda) and their role in sustainable management of soil ecosystem in Vietnam*, code: 106.15.13.09, 2009-2011.

[8] Vu Quang Manh (2015a), *Final report of the NAFOSTED project Soil oribatid mite (Acari: Oribatida) and animal community, their bioindicator's role of climate environmental change in Vietnam*, code: 106.14-2012.4, 2013-2015.

[9] Vu Quang Manh (2015b), *The oribatid mite fauna (Acari: Oribatida) of Vietnam - systematics, zoogeograpph and formation*, PENSOFT, Sofia-Moscow, pp.1-212.

[10] Pham Dinh Sac (2015), *Spiders in the forests of Northern Vietnam*, Vietn. Acad. Publishing, pp.1-166.

[11] Vu Hong Quang (1996), *Study on biology-ecological characteristics of Scorpions (Buthidae) at Lab conditions*, PhD. thesis, pp.1-145.

[12] Vu Quang Manh, Nguyen Dang Thin (2009), "Scorpions (Chelicrata : Scorpionida) A venomous arthropod in Vietnam", *Journal of Science of HNUE*, **54(1)**, pp.90-97.

[13] Dinh Sac Pham, Thi Hang Tran, Wilson R. Lourenço (2017), "Diversity and endemicity in the scorpion fauna of Vietnam. A preliminary synopsis", *Comptes Rendus Biologies*, **340(2)**, pp.65-137.

[14] D. Krivolutzkij, Q.M. Vu, T.V. Phan (1998), "The oribatid mites of Vietnam", *The Severtsov Institute of Problems of Ecology and Evolution, Russian Academy Sciences, Moscow,* The Russian-Vietnamese Tropical Centre, Moscow-Hanoi, pp.130-145.

[15] Vu Quang Mạnh (1986), *A study of soil oribatid mite (Acari, Oribatei) fauna of Northern Vietnam*, PhD. Thesis, Sofia University, Bulgaria, pp.1-175.

[16] Vu Quang Mạnh (2007), *Fauna of Vietnam T. 21: Acari: Oribatida*, Techniques Publishing House, 355 pp.

[17] Q.M. Vu, T.T. Nguyen (2000), "Microarthropod community structures (Oribatei and Collembola) in Tam Dao national park", *Journal of Biosciences*, **25(4)**, pp.379-386.

[18] Dao Duy Trinh (2011), *Species diversity and structures of oribatid mites community (Acari: Oribatida) of Xuan Son national park, Phu Tho province*, PhD. thesis, HNUE Vietnam, pp.1-145.

[19] Nguyen Hai Tien (2012), *Species diversity and structures of oribatid mites community (Acari: Oribatida) of Phong Nha - Ke Bang national park, province of Quang Binh*, PhD. thesis, HNUE, pp.1-150.

[20] Q.M. Vu (2012), "Oribatid soil mite (Acari: Oribatida) of Northern Vietnam: Species distribution and densities according to soil and habitat type", *The Pan-Pacific Entomologist*, **87(4)**, pp.209-222.

[21] S. Ermilov (2015), "A list of oribatid mites (Acari, Oribatida) of Vietnam", *Zookeys*, **546,** pp.61-85.

[22] Socialist Republic of Vietnam (2008), *The national assembly: Law on biodiversity*, No. 20/2008/QH12.

[23] S. Ermilov, A. Anichkin (2014), "Taxonomic study of oribatid mites (Acari, Oribatida) of Bi Dup - Nui Ba National Park (southern Vietnam)", *Zootaxa*, **3834**, pp.1-86.

[24] A. Minor, S. Ermilov (2015), "Effect of topography on soil litter mites (Acari: Oribatida, Mesostigmata) in a tropical monsoon forest in Southern Vietnam", *Exp. Appl. Acarol*., doi: 10.1007/s10493-015-9955-7.

[25] B.T.T. Tran, X.S. Le, D.A. Nguyen (2013), "An annotated checklist of centipedes (Chilopoda) of Vietnam", *Zootaxa*, **3722(2)**, pp.219-244.

[26] H. Enghoff, S. Golovatch, D.A Nguyen (2004), "A review of the millipede fauna of Vietnam (Diplopoda)", *Arthropoda Selecta*, **13(1/2)**, pp. 29-43.

[27] Vu Quang Manh, Nguyen Tri Tien (1993), "A list of Collembolan (Insecta: Collembola) known in Vietnam", *Journal of Biology,* Vietnamese, special issue **15(4)**, pp.20-25.

[28] Nguyen Tri Tien (1994), *Collembolan (Insecta: Collembola) fauna of Northern Vietnam*, PhD. thesis, pp.1-180.

[29] Nguyen Thi Thu Anh (2009), *Collembolan (Insecta: Collembola) fauna in agricultural soil of Northern Vietnam*, PhD. thesis, pp.1-145.

[30] Viet Bui Tuan (2005), "Contribution to study biodiversity and environmental conservation in case of study on ant biodiversity in Vietnam", *Report on insect inventory project in tropic Asia* (TAIIV), pp.103-115.

[31] Nguyen Duc Kham, Nguyen Tan Vuong, Trinh Van Hanh, Nguyen Van Quang, Le Van Trien, Nguyen Thuy Hien, Vu Van Nghien, Ngo Truong Son, Vo Thu Hien (2007), *Fauna of Vietnam. 15. Isoptera*, Science and Technics Publishing House, Hanoi, pp.1-303.

[32] Q.M. Vu, H.H. Nguyen, R. Smith (2007), "The termites (Isoptera) of Xuan Son national park, Northern Vietnam", *The Pan-Pacific Entomologist*, **83(2)**, pp.85-94.

[33] Thai Tran Bai (1985), *The Earthworm fauna (Oligochaeta) of Vietnam*, DSc. Thesis, Lomonosov University, Moscow, USSR, pp.1-285.

[34] T.T. Nguyen, A.D. Nguyen, B.T. Tran, Robert J. Blackemore (2016), "A comprehensive checklist of earthworm species and subspecies from Vietnam (Annelida: Clitellata: Oligochaeta: Almidae, Eudrilidae, Glossoscolecidae, Lumbricidae, Megascolecidae, Moniligastridae, Ocnerodrilidae, Octochaetidae)", *Zootaxa,* **4140(1),** pp.1-92.

[35] V.V. Lien, Timothy C.T. Bonebrake, Q.V. Manh, T.N. Nha (2015), "Butterfly diversity and habitat variation in a disturbed forest in Northern Vietnam", *The Pan-Pacific Entomologists*, **91(1)**, pp.29-38.

[36] V.V. Lien, Timothy C. Bonebrake, Q.V. Manh, et al. (2016), "Warming threat compounds habitat degradation impacts on a tropical butterfly community in Vietnam", *Global Ecology and Conservation*, **8**, pp.203-211.

BWTaligner: a genome short-read aligner

Lam Nguyen[1], Xuan Thi Trinh[2], Hien Trinh[3], Dang Hung Tran[4], Cuong Nguyen[1*]

[1]*Vinmec Research Institute of Stem Cell and Gene Technology*
[2]*Faculty of Information Technology, Hanoi Open University*
[3]*Laboratory of Genetic Engineering, Institute of Biotechnology, Vietnam Academy of Science and Technology*
[4]*Hanoi National University of Education*

Received 6 April 2018; accepted 15 May 2018

Abstract:

The development of next-generation sequencing technologies has helped sequence large genomes easily, producing a huge number of short-reads - small fragments of DNA. Despite the existence of many developed alignment tools, mapping short-read datasets to the reference genome, a crucial step of genome analysis, still remains a challenge. In this study, we develop a short-read alignment program, BWTaligner, based on the Burrows-Wheeler transform compression - exact and inexact matching. We tested it on the paired-end read data simulated from chromosome 9 of the rice genome to compare the alignment and single-nucleotide polymorphism (SNP) calling between our aligner and BWA - the preferred alignment program. The results showed that the BWA delivers higher recall and F-score, while BWTaligner has better precision in high coverage depth.

Keywords: ***Burrows-Wheeler transform, high-throughput sequencing, paired-end short reads, sequence alignment.***

Classification number: ***3.5***

Introduction

The development of massive parallel sequencing technologies has stimulated the production of a vast number of short-reads, which are small fragments of DNA genomes. As the mapping of short-read datasets to large genomes presents a huge challenge to the existing sequencing programs, more and more algorithms are being improved in order to reduce the execution time and increase the mapping accuracy. At the outset, hash table-based methods either hash the short-read sequences or the reference genome and many alignment tools have been developed to resolve this. The aligners based on hashing short reads are typically MAQ [1], ZOOM [2], and SHRiMP [3]. MAQ is one of the old programs that supports ungapped sequence alignments and shown quality scores, while ZOOM limits a number of mismatches. SHRiMP indexes both the short-reads and the genome. These aligners have a flexible memory footprint, which have been capable of overhead when a small number of reads are mapped. The tools hashing the genome, such as SOAP 1 [4], PASS [5], MOM [6], mrFAST/mrsFAST [7], and BFAST [8] can be parallelized using numerous threads; however, they need a large memory to build an index for the reference genome. Interestingly, mrFAST/mrsFAST employs a seed-and-extend strategy that initially identifies candidate positions for a short-read and then uses different alignment algorithms, such as the Smith-Waterman algorithm [9], for mapping. In addition to initial hash table-based methods, the other alignment algorithm is a slider that merges and sorts the reference subsequences

*Corresponding author: Email: v.cuongn@vinmec.com

and short-reads.

As the alignment algorithms using a hash table often require a large amount of memory, new alignment programs based on suffix/prefix tries were generated to reduce the memory requirements. The suffix/prefix tries perform backward searched and the Burrows-Wheeler Transform (BWT) [10] for exact matching, which has led to the development of several aligners, including Bowtie [11], SOAP 2 [12], and BWA [13]. Furthermore, they also provide support for paired-end alignment. Bowtie, including Bowtie 1 and 2 [14], is one of the first programs to use FM-index [15, 16], which is built on the BWT and mimics backward search. For reads shorter than about 50 bp, Bowtie 1 is sometimes more sensitive, while Bowtie 2 supports gapped alignment and works better for longer short-reads. SOAP 2 combines the hashing and FM-index to speed up but uses more memory than BWA and Bowtie. The efficiency of the BWA aligner for inexact matching is widely known, and it is still used by researchers. All of these aligners are fast and have been optimized for multi-core Central Processing Units (CPUs). However, increasing the speed of alignment process provides time saving, especially with regard to processing large-scale data; hence, a multiple-core Graphics Processing Units (GPUs)-based method is a powerful choice. There are several alignment tools based on GPU, including SOAP3 [17] and BarraCUDA [18].

In this research, the introduction of BWTaligner based on the BWT algorithm, exact and inexact matching, has been made. Moreover, we have evaluated the performance of BWTaligner on simulated data by comparing it with BWA in single-nucleotide polymorphism (SNP) calling - the finding corresponding to the variations occurring in the genome.

Materials and methods

Burrow-Wheeler transform

The BWT construction reduces the execution speed and memory in the running process. Let G be a reference genome sequence that is constructed by four nucleotides (A, C, G, T). The symbol $ is lexicographically smaller than all the characters in G and only appears at the end to form a new sequence, G$. The matrix M, which is built from the rotations of G$, is sorted by lexicographical order, and each column is a permutation of G$. The transformed B can be attained by taking the last column of matrix M. A suffix array (SA) is defined as an array of integers with the starting position of the i-th smallest suffix of G. This algorithm is illustrated in Fig. 1.

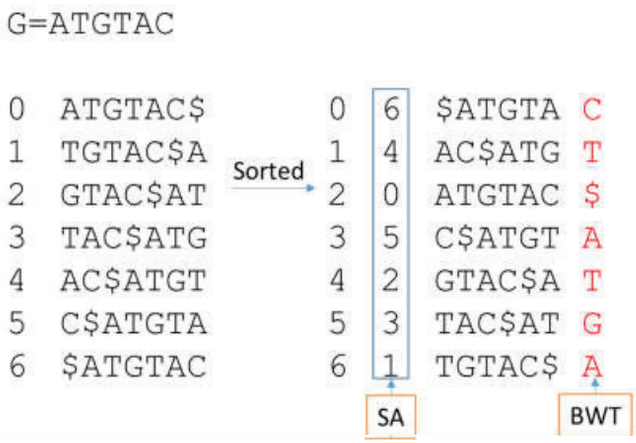


Fig. 1. The BWT construction and suffix array of reference genome G = ATGTAC. BWT matrix includes seven rows that are in lexicographical order. Forward BWT is defined as CT$ATGA and SA is (6, 4, 0, 5, 2, 3, 1).

Exact matching

Let Q be a query sequence that is a substring of reference genome G. A backward search based on FM index [15] was used to find each occurrence of Q (Fig. 2), which is actually the search for an SA interval. Oc(α,i) is the number of occurrences of α in B[0,i]. C(α) is the number of symbols in Q[0,n-2] that are lexicographically smaller than α ϵ G. Ferragina and Manzini showed the SA interval for searching for all occurrences of Q in G using the forward BWT as follows:

$$Rs = C(Q[i]) + O\ (Q[i], Rs(i+1) - 1) + 1, \quad i \in [0, |Q|]$$
$$Re = C(Q[i]) + Oc\big(Q[i], Re(i+1)\big), \quad i \in [0, |Q|]$$

where: Rs and Re indicate the start and end of the SA interval, respectively. Q is a substring of G if and only if Rs(i) ≤ Re(i). However, as the total array of Oc is sorted, more memory and execution time are required. For reducing

the memory footprint of the Oc array, only a part of the Oc is stored and calculated using the length of Oc (L).

```
Backward_Search(Q[1,q])
{
     i = q, c = Q[q], First = C[c]+1, Last = C[c+1];
     while ( ( First ≦ Last) and i ≧ 2 )
     {
          c = Q[i-1];
          First = C[c] + Oc(c, First-1)+1;
          Last  = C[c] + Oc(c, Last); i = i-1;
     }
     if (Last < First) then
          return "no occurrence" ;
     else
          return ( First, Last );
}
```

Fig. 2. The pseudocode of backward search.

Inexact matching

Sequencing errors and the differences between the sequence and reference genome are inexact matches. The inexact matches can be attained by comparing input sequences with the reference genome in order to identify variations, such as substitutions, insertions, and deletions. However, the exact matching only provides ungapped alignment; therefore, insertions and deletions were not allowed. Hence, inexact match searching could be converted to exact matches based on all the permutations of short reads. A total of permutations can be performed by the 4-ary tree, where each permutation represents different routes (Fig. 3).

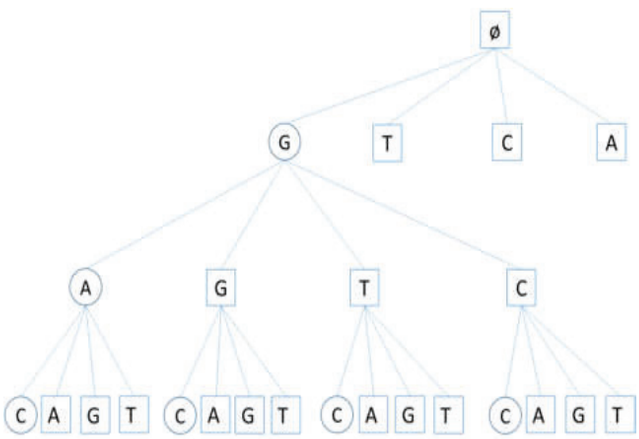


Fig. 3. A 4-ary tree example for searching the inexact matches of sequence "GAC" using BWT. The circles are defined as the original bases and rectangles as the mutated bases.

Each node denotes a base in the query with the same position. Currently, there are two approaches to searching for all inexact matches - depth-first search (DFS) and breadth-first search (BFS). The BFS approach requires a large memory capacity for storing all the results, so this approach is impractical for GPU computing. Our aligner implemented the DFS approach, where the memory expenditure is small and equivalent to the size of the tree. Nevertheless, recursive functions are still supported for the Fermi architecture [19]. Fig. 4 illustrates the pseudocode of inexact matching. The number of inexact matching of sequences can be estimated through calculating the number of bases that do not exactly match the genome. z(*) is defined as the full length of the query sequence (Q), where z(i) is defined as the number of inexact matches in the query Q[i + 1, |Q|-1] ($0 \leq i \leq |Q| -1$). For seed alignment, zw(*) is calculated, where zw(i) represents the number of

```
# G: genome, Q: query sequence
# bwt and rbwt are forward and reverse BWT of G
1. Searching for inexact matches
Inexact_matches_search_full_length(Q,G,f,l,z)
{
   z=0;f=0;l=|G|
   for i = |Q|-1 to 0;
   {
      f = bwt.C(Q[i], f-1)+1;l = bwt.C(Q[i]) + bwt.Oc(Q[i],l); z[i]=z
   }
      if f > l then
         f = 0; l = |G|; ++z;
         return "inexact matches";
}
Inexact_matches_search_seed_length(W, G, f, l, z)
# W: seed length (a part of Q)
{
   zw=0, f=0; l = |G|
   for i = W-1 to 0
   {
      f = bwt.C(Q[i], f-1)+1; l = bwt.C(Q[i]) + bwt.Oc(Q[i],l); zw[i]=z
      if f > l then
         f = 0; l = |G|; ++z;
         return "inexact matches";
   }
}
2. The initialization of the stack from the first base of Q()
Initialization_stack_first_base()
{
   while (stack is not empty)
   {
      access to current node (top node)
      if there is not mutation on current node then
         continue;
      fi
      there is mutation in the current node
      calculate SA interval (f,l)
      if f <= l then
         if kepping to the end of Q then
            return "a hit is found";
         else
            push next base in Q into the stack;
         fi
      fi
   }
}
```

Fig. 4. The pseudocode for inexact matching.

substitutions mismatching correctly to the substring Q[i+1, W-1] ($0 \leq i \leq W-1$).

Results and discussion

We evaluated the performance of the BWTaligner by comparing it to BWA version 0.6.2, the alignment tool using simulated reads that is most widely used. The paired-end datasets were simulated from chromosome 9 in the reference rice genome, Nipponbare version 7.0 (Genbank accession number PRJDB1747, 23,012,720 bp) using *wgsim*, a short-read simulator (version 0.3.1) with a different depth of coverage, including 5X, 10X, and 30X 100 bp-paired-end reads; 0.085% mutation rates (19,560 SNPs); and 0.02% base error rates. First, we evaluated the alignment quality of each tool and subsequently called SNPs from aligned short reads using *mpileup* in SAMtools and VarScan. Finally, the identified SNPs were compared with the simulated SNPs. All the tests were carried out on a workstation with a X5650 @ 2.67 GHz 24-core processor and 198 GB RAM running the Ubuntu 14.10.

Table 1 showed that over 99.0% of simulated paired-end reads aligned with the reference genome. With regard to the number of aligned reads, BWA was slightly higher at three depths of coverage. The SNP calling performance of the aligners was evaluated using the precision, recall, and F-score (Table 2). Precision is defined as TP/(TP+FP), recall as TP/(TP+FN), and F-score as 2*precision*recall/(precision+recall), where TP is a true positive, that is, the number of correct SNPs. FP is a false positive, performing the mismatch, and FN is a false negative, representing that the simulated SNPs were not determined along with the missed SNP. From Table 3, it can be seen that at the lower coverage (5X and 10X), BWA has better precision, while at 30X depth, the precision of BWTaligner is higher (99.16% as compared to 99.03% of BWA). While the precision is a positive predictive value and based on the number of positive SNPs, the recall, i.e., the sensitivity, is considered as the number of negative SNPs and F-score value, which is the harmonic mean of the precision and recall. The results of our study showed that BWA always leads to a higher recall and F-score than BWTaligner at all coverages. Furthermore, F-scores tend to increase as the depth of coverage gets higher. This denotes that the depth of coverage plays an important role in the accuracy of alignment and SNP calling.

Table 1. Alignment of simulated reads.

Depth of coverage	Stimulated paired-end reads	Aligned reads using BWA (%)	Aligned reads using BWTaligner (%)
5X	575,318	99.57	99.38
10X	1,150,636	99.58	99.41
30X	3,451,908	99.58	99.41

Table 2. Performance of SNP calling under different coverage.

		BWA		BWTaligner	
Call SNP at 5X	TP	1,182	6.01%	891	4.55%
	FP	3	0.02%	9	0.05%
	FN	18,468	93.97%	18,669	95.40%
Call SNP at 10X	TP	9,439	47.98%	8,223	41.92%
	FP	21	0.11%	58	0.30%
	FN	10,211	51.91%	11,337	57.79%
Call SNP at 30X	TP	19,155	96.56%	18,951	96.10%
	FP	187	0.94%	161	0.82%
	FN	495	2.50%	609	3.09%

Table 3. Precision, recall and F-measure between BWA and BWTaligner.

	BWA			BWTaligner		
	5X	*10X*	*30X*	*5X*	*10X*	*30X*
Precision	0.9974	0.9978	0.9903	0.9900	0.9930	0.9916
Recall	0.0601	0.4804	0.9748	0.0456	0.4204	0.9689
F-score	0.1134	0.6485	0.9825	0.0871	0.5907	0.9801

Conclusions

We preliminarily present the BWTaligner based on the basic algorithms, including the Burrows-Wheeler Transform, backward search for exact and inexact matching. This tool will be further developed and empirically studied so as to address the short-read alignment challenge with regard to time and accuracy.

REFERENCES

[1] H. Li, J. Ruan, R. Durbin (2008), "Mapping short DNA sequencing reads and calling variants using mapping quality scores", *Genome Research*, **18**(**11**), pp.1851-1858.

[2] H. Lin, Z. Zhang, M.Q. Zhang, B. Ma, M. Li (2008), "ZOOM! Zillions of oligos mapped", *Bioinformatics*, **24**(**21**), pp.2431-2437.

[3] S.M. Rumble, P. Lacroute, A.V. Dalca, M. Fiume, A. Sidow, M. Brudno (2009), "SHRiMP: accurate mapping of short color-space reads", *PLoS Computational Biology*, **5**(**5**), pp.e1000386.

[4] R. Li, Y. Li, K. Kristiansen, J. Wang (2008), "SOAP: short oligonucleotide alignment program", *Bioinformatics*, **24**(**5**), pp.713-714.

[5] D. Campagna, A. Albiero, A. Bilardi, E. Caniato, C. Forcato, S. Manavski, G. Valle (2009), "PASS: a program to align short sequences", *Bioinformatics*, **25**(**7**), pp.967-968.

[6] H.L. Eaves, Y. Gao (2009), "MOM: maximum oligonucleotide mapping", *Bioinformatics*, **25**(**7**), pp.969-970.

[7] F. Hach, F. Hormozdiari, C. Alkan, F. Hormozdiari, I. Birol, E.E. Eichler, S.C. Sahinalp (2010), "mrsFAST: a cache-oblivious algorithm for short-read mapping", *Nature Methods*, **7**(**8**), pp.576-577.

[8] N. Homer, B. Merriman, S.F. Nelson (2009), "BFAST: an alignment tool for large scale genome resequencing", *PLOS ONE*, **4**(**11**), pp.e7767.

[9] R. Mott (2005), "Smith-Waterman Algorithm", *Encyclopedia of Life Sciences*, Chichester: John Wiley & Sons, Ltd.

[10] M. Burows, D.J. Wheeler (1994), *A block sorting lossless data compression algorithm*, CA, Digital Equipment Corporation.

[11] B. Langmead, C. Trapnell, M. Pop, S.L. Salzberg (2009), "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome", *Genome Biology*, **10**(**3**), pp.R25.

[12] R. Li, C. Yu, Y. Li, T.W. Lam, S.M. Yiu, K. Kristiansen, J. Wang (2009), "SOAP2: an improved ultrafast tool for short read alignment", *Bioinformatics*, **25**(**15**), pp.1966-1967.

[13] H. Li, R. Durbin (2009), "Fast and accurate short read alignment with Burrows-Wheeler transform", *Bioinformatics*, **25**(**14**), pp.1754-1760.

[14] B. Langmead, S.L. Salzberg (2012), "Fast gapped-read alignment with Bowtie 2", *Nature Methods*, **9**(**4**), pp.357-359.

[15] P. Ferragina, G. Manzini (2005), "Indexing compressed text", *Journal of the ACM*, **52**(**4**), pp.552-581.

[16] P. Ferragina, G. Manzini (2000), "Opportunistic data structures with applications", *Proceedings of the 41st annual symposium on foundations of computer science*, pp.390-398, IEEE.

[17] C.M. Liu, T. Wong, E. Wu, R. Luo, S.M. Yiu, Y. Li, T.W. Lam (2012), "SOAP3: ultra-fast GPU-based parallel alignment tool for short reads", *Bioinformatics*, **28**(**6**), pp.878-879.

[18] P. Klus, S. Lam, D. Lyberg, M. Cheung, G. Pullan, I. McFarlane, B.Y. Lam (2012), "BarraCUDA - a fast short read sequence aligner using graphics processing units", *BMC Research Notes*, **5**(**1**), p.27.

[19] E. Lindholm, J. Nickolls, S. Oberman, J. Montrym (2008), "NVIDIA Tesla: a unified graphics and computing architecture", *IEEE Micro*, **28**(**2**), pp.39-55.

Risk assessment of lead and cadmium on Juveniles of *Cyprinus carpio* in laboratory scale

Thi Yen Ta*, Thi Trinh Le, Thi Thuy Trinh, Thi Tham Trinh, Thi Mai Thao Pham, Thi Hong Phuong Pham

Ha Noi University of Natural Resources and Environment

Received 1 March 2018; accepted 28 May 2018

Abstract:

This paper assesses the risk of lead and cadmium heavy metals on *Cyprinus carpio* in laboratory conditions. The research determined the 96 hours LC50 value of lead nitrate and cadmium nitrate in the fish *Cyprinus carpio*. This study combined the ecological toxicology test with the ecological hazard description to determine the acute toxicity effects of lead and cadmium on *Cyprinus carpio*. Ecological toxicology took place in 96 hours with test concentration of 0.25; 0.5; 1.0; 1.5; 2.0 mg/l on lead and 0.5; 1.0; 1.5; 2.0; 3.0 mg/l on cadmium. Through probit analysis, the LC50 after 96 hours of lead and cadmium in *Cyprinus carpio* was found to be 0.987 mg/l and 1.171 mg/l, respectively. Through monitoring the biological behaviour of *Cyprinus carpio* when exposed to lead and cadmium, it was observed that the number of deaths is proportional to the concentration of chemical exposure time. The abnormal morphology and behaviour of the fish also increased with testing time and lead and cadmium concentration. *Cyprinus carpio* also absorbs lead and cadmium in its body; the cumulative content is similar as above.

Keywords: ***acute toxicity, cadmium nitrate, Cyprinus carpiro, lead nitrate, 96 hours LC50.***

Classification number: ***6.2***

Introduction

Toxic chemicals released into the environment, either from point sources such as industrial and municipal discharges or from non-point sources such as agricultural runoff and atmospheric deposition, are capable of contaminating surface waters and sediments [1]. Heavy metals are a group of toxic chemicals persistent in the environment, which are bio-accumulative and non-biodegradable in the food chain [2]. Heavy metals also disrupt and result in the contamination of ecosystems; they can be both carcinogenic and non-carcinogenic for human health. Heavy metals in the human body do not degrade, which accounts for their chronic toxicities. Air contaminated by heavy metals may pollute soil and water, resulting in contaminated crops and consumables. Erosion of natural deposits of rock minerals and atmospheric deposition of gaseous emissions from tailpipes of industrial engine allow the mobility of heavy metals into the aquatic environment. Heavy metals persist in the aquatic environment and, based on their available concentrations, bioaccumulate in the tissues of aquatic plants and animals.

Examples of heavy metals that have been released into the environment include cadmium (Cd), lead (Pb), nickel (Ni), arsenic (As), mercury (Hg) and chromium (Cr) among others; they are probable carcinogens in humans. Lead reduces and increases, while cadmium accumulates [3]. Lead and cadmium can cause damage to the nervous, cardiovascular, and human skeletal systems.

Fishes are organisms that survive mainly in water bodies. Fish is food to humans, as it remains a relatively cheap source of protein. The nutritional composition of fish encompasses both macro and trace nutrients beneficial to the human biological system. The major nutritional constituents of fish are water, proteins, lipids, minerals and vitamin B2 [4]. However, aquatic ecosystems polluted with cadmium and lead threatens the suitability of fish as an important food source for humans. Fish being the final chain in the aquatic food web is able to bioaccumulate heavy metals in the aquatic environment. The accumulated metals in fishes are transferable to humans through the food chain. Fish safety, just as food safety, is an important public health issue because humans can develop numerous diseases from the consumption of contaminated fish [5].

There have been many studies on the effects of toxins on the growth, development and reproduction of fish species [6]. According to EPA's Ecotoxicological Testing Guidelines, fishes are considered to be highly susceptible and can be easily observed during the test, so it is chosen as the ecotoxicological test organism [1]. LC50 is used to

*Corresponding author: Email: tayen87@gmail.com

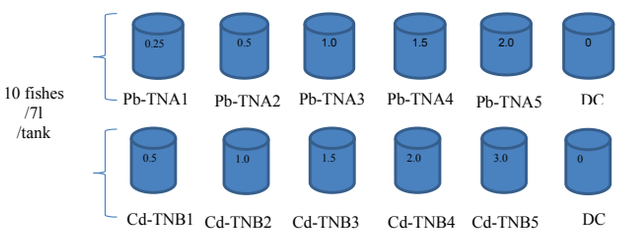
evaluate the effects of toxins on the test organisms through the lethal levels [6, 7].

Waste water from many industrial production activities in Vietnam, especially waste water from the recycling of metals, contains lead and cadmium. When theses concentration exceeds the allowed standard, they affect humans through biological amplification in the food chain. *Cyprinus carpio* is a common food source in Vietnam so, if it is living in a polluted environment, it poses a high risk to health.

Methods

Juveniles of *Cyprinus carpio* with a mean body weight 8-10 g and standard length of 8-9 cm were used for the study. They were collected from the fish research laboratory of the Research Institute for Aquaculture No 1, Vietnam. The fishes were acclimatised in the laboratory for 5 days in plastic tanks of 20 litres capacity before the experiment. During the acclimatisation, the fishes were fed daily with regular feed stock, i.e., Durhante fish pellet, which was equivalent to 5% mean body weight of a fish. Natural groundwater was used to feed the fishes. Water samples were taken at three different sampling times and analysed for cadmium and lead. The acceptable feeding water must be free of heavy metals. Commercial lead nitrate and cadmium nitrate were used as Cd and Pb sources, respectively. Only individuals who are still healthy will be used for ecological toxicology.

Each experiment was conducted with 5 different concentrations of lead and cadmium and a control sample. After conducting two exploratory experiments, the experimental concentration range was determined as follows:



The concentrations of Pb and Cd in the fishes were determined after 6h, 12h, 24h… 96h of the experiment. The meat of the fishes was homogenised and added with nitric acid-peroxide. Then, the samples were digested using a microwave digester (MWS-2). The completely digested samples were allowed to cool to room temperature; then, they were filtered (glass wool) and made up to 50 ml. All the digested samples were analysed using an atomic absorption spectrophotometer (Thermo Scientific) and an air-acetylene flame.

Homogenised samples were spiked with three different concentrations of heavy metals for determination recovery. Each running in triplicate and blanks were carried through the whole procedure described above. The recovery of result analysis ranged from 70.7% to 118.5%.

Ecotoxicity testing was performed by using the APHA method and EPA guidelines. During the experiment, oxygen was continually pumped to maintain the required DO level for the fish, and the temperature and pH parameters were measured daily.

All research results were evaluated and analysed in Excel. The LC50 of lead and cadmium in *Cyprinus carpio* was determined through probit analysis [8].

Results and discussion

Physical-chemical characteristics of water samples

The physical-chemical characteristics of sample waters were analysed. Its main parameters were investigated, including temperature (28-29ºC), pH (7.1-7.6), dissolved oxygen (6.7-8.4 mg/l), hardness (43.1-52.5 mg $CaCO_3$/l) and total alkalinity (112-120 mg $CaCO_3$/l). Cadmium and lead were not detected in the samples. The results were compared with the Food and Agriculture Organisation's (FAO) [9a, 9b] guidelines for fish pond water quality, as shown in Table 1.

Table 1. Physical-chemical characteristics of water samples.

Parameter	Range	FAO guideline
Temperature (ºC)	28-29	25-30
pH	7.1-7.6	6.5-8.5
Dissolved Oxygen (mg/l)	6.7-8.4	>3
Hardness (mg $CaCO_3$/l)	43.1-52.5	>25*
Total alkalinity (mg $CaCO_3$/l)	112-120	>25*
Cadmium (mg/l)	ND**	
Lead (mg/l)	ND**	

Source: FAO [9a]; *FAO [9b], ** not detected.

Effects of lead on Cyprinus carpio

Through confirmed experiments and results in 96 hours, the following was observed: The higher the dose and longer the exposure time, the greater the impact of chemicals. Specifically, at the lowest Pb^{2+} concentration of 0.25 mg/l, the number of dead fish was the lowest, and at the highest concentration of 2.0 mg/l, the number of dead fish was the highest. During the first 3 hours, the fishes were not

Table 2. Probit for 96 hours exposure to lead.

Concentration (ppm)	Log_{10} Concentration	Total no of test fish	Replicate 1			Replicate 2			Replicate 3		
			No of death	*% mortality*	*probit*	*No of death*	*% mortality*	*probit*	*No of death*	*% mortality*	*probit*
0	0	10	-	-	-	-	-	-	-	-	-
0.25	-0.602	10	2	20.00	4.16	2	20.00	4.16	2	20.00	4.16
0.5	-0.301	10	4	40.00	4.75	4	40.00	4.75	3	30.00	4.48
1.0	0	10	4	40.00	4.75	4	40.00	4. 75	4	40.00	4.75
1.5	0.176	10	6	60.00	5.25	6	60.00	5.25	7	70.00	5.52
2	0.301	10	7	70.00	5.52	7	70.00	5.52	7	70.00	5.52

affected but, after 96 hours, the lowest and highest number of dead fish increased. It is evident that at the same time, the mortality variation was very large for the selected concentration range.

Probit analysis was applied to the mean results calculated in Table 2 to determine the LC50 of the carp.

Effect of cadmium on Cyprinus carpio

Through confirmed experiments and monitoring results in 96 hours, we see that the effect of toxicity is greater when the dose is higher and exposure time is longer. Specifically, at the lowest Cd^{2+} concentration of 0.5 mg/l, the lowest number of dead fish was recorded, and at the highest concentration of 3.0 mg/l, the highest number of dead fish was recorded. During the first 3 hours, the fish was not affected but, after 96 hours, the lowest and highest number of dead fish increased. It is evident that at the same time, the mortality variation was very large for the selected concentration range.

The 96-hour 50% lethal concentration (LC50-96 hours) of both cadmium and lead was calculated using the regression method. The number of deaths of test fish observed at each concentration after 96 hours of exposure for each of the three replicates is shown in the probit Tables 2 and 3. LC50-96 hours of cadmium and lead for *Cyprinus carpio* is shown in Table 4. Similar to the result presented in Table 1, the number of dead fishes was proportional to the lead test concentration in the three replicates.

Table 3. Probit for 96 hours exposure to cadmium.

Concentration (ppm)	Log_{10} Concentration	Total no of test fish	Replicate 1			Replicate 2			Replicate 3		
			No of death	*% mortality*	*Probit*	*No of death*	*% mortality*	*probit*	*No of death*	*% mortality*	*Probit*
0	0	10	-	-	-	-	-	-	-	-	-
0.5	-0,301	10	2	20.00	4.16	3	30.00	4.48	2	20.00	4.16
1.0	0	10	4	40.00	4.75	4	40.00	4.75	4	40.00	4.75
1.5	0.176	10	6	60.00	5.25	5	50.00	5.00	5	50.00	5.00
2.0	0.301	10	8	80.00	5.84	7	70.00	5.52	8	80.00	5.84
3.0	0.477	10	8	80.00	5.84	8	80.00	5.84	8	80.00	5.84

Table 4. LC50-96 hours of cadmium (Cd) and lead (Pb) for *Cyprinus carpio.*

Heavy metal	Replicate 1 LC50 (mg/l)	Replicate 2 LC50 (mg/l)	Replicate 3 LC50 (mg/l)	Mean LC50 (mg/l)
Cadmium	1.147	1.161	1.202	1.171
Lead	0.996	0.996	0.968	0.987

LC50-96 hours results show that there was no significant difference in the three replicates for both Cd and Pb. Although no death was recorded in the control groups, the mortality percentage of the test organism increased by increasing the test concentration. The increase in mortality with increase in toxicant concentration may be due to the increase of toxicant solubility and species' susceptibility. It accompanies the high toxicant concentration in the aquatic medium (LC50) of lead in the three replicates, which was significantly lower than those of cadmium. This result suggests that lead is more toxic to *Cyprinus carpio* than cadmium. Higher LC50 connotes less toxicity. Higher concentration is required to achieve a 50% mortality of test organisms.

Some authors in the world have also conducted toxicological studies to determine the LC50-96 hours in some organisms and obtained different results. Specifically, according to the result of Brraich Onkar Singh and Kaur Manjeet, the concentration of lead nitrate (LC50-96 hours) in Labeo rohita is 34.20 mg/l [10]. Zeynab Abedi1, et al. identified the LC50-96 hours of $CdCl_2$, $CrCl_3$ and Pb $(NO_3)_2$ for P. hypophthalmus as 64.89, 7.46 and 48.06 mg/l, respectively [11]. This suggests that the toxicological effects of lead and cadmium on different species and in different experimental conditions will yield different results.

Effect of lead and cadmium on Cyprinus carpio

This experiment was conducted according to the concentrations selected in the previous exploratory experiment. The confirmed experiment was repeated twice. The results have been averaged as follows:

Average number of dead fish = (1^{st} death + 2^{nd} death+ 3^{rd} death)/3

The results are shown in Fig. 1 and Fig. 2 below:

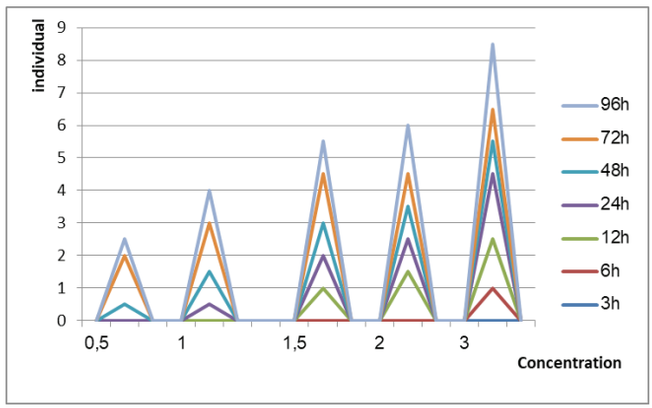


Fig. 1. Effect of lead on *Cyprinus carpio.*

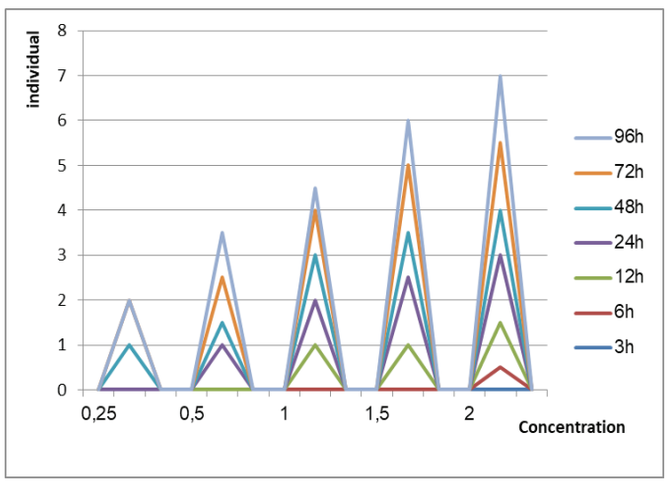


Fig. 2. Effect of cadmium on *Cyprinus carpio.*

The confirmed and monitored results after 96 hours (Fig. 3) showed that the effect of chemicals was greater when the dose was higher and exposure time was longer. Specifically, when Pb^{2+}, Cd^{2+} concentrations were the lowest at 0.25 ppm and 0.5 ppm, respectively, the number of dead fishes was minimum; at the highest concentration of lead at 2.0 ppm and cadmium at 3.0 ppm, the number of dead fishes was the highest. During the first 3 hours, the tested fishes were not affected but, after 96 hours, any amount of concentration increased the number of dead fish. At the same time, variation in lethal effects was significant for the range of concentrations chosen.

Effect of concentration and exposure time of lead and cadmium to Cyprinus carpio

The effect of lead and cadmium exposure on the *Cyprinus carpio's* bioassay is shown in Table 5.

Table 5. Expression of *Cyprinus carpio* in the experiments with lead.

Expression	Normal	Impacted	Poisoned
Swimming at the tank bottom	x		
Normal breathing	x		
Fast swimming		x	
Losing the swimming direction, rushing into the tank		x	x
Change eye colour (light yellow to grey brown)		x	
Scabbing, red marks on the body		x	x
Bleeding bring		x	x
Mucus secretion		x	x
Breathing by oral and bring		x	
Loss of balance, sluggishness, abdominal swimming		x	x
Die			x

The impacts of lead and cadmium exposure on *Cyprinus carpio* are shown in Figs. 3 and 4. The number of fishes is correlated to the level of concentration as well as exposure time. The percentage of normal fish diminishes. From the starting point up to 48 hours of observation, the percentage of affected fishes increases slightly due to prolonged exposure. After 48 hours, the adversely affected fishes increase suddenly with the exposure time being prolonged to 72 hours and 96 hours.

The test concentration of cadmium is two-fold with that of lead, but the effects appear almost simultaneously. It is suggested that the toxic levels of lead are two times greater than that of cadmium.

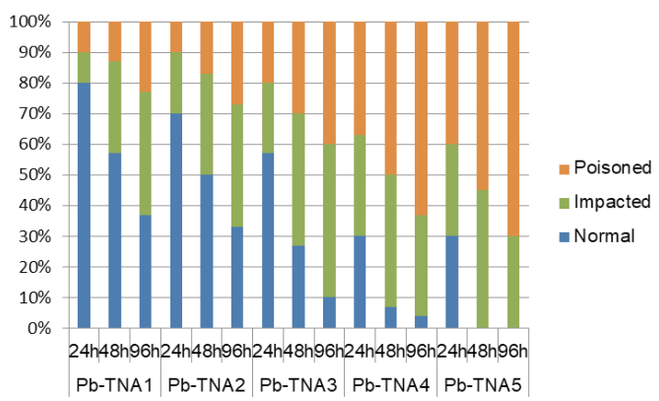


Fig. 3. Indication of the effect of lead on *Cyprinus carpio*.

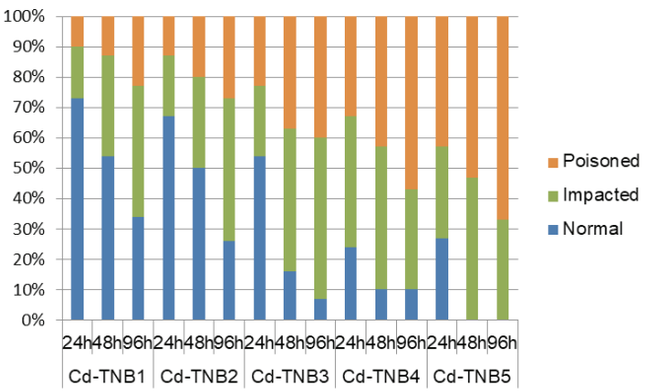


Fig. 4. Indication of the effect of cadmium on *Cyprinus carpio*.

Concentration of lead and cadmium in Cyprinus carpio

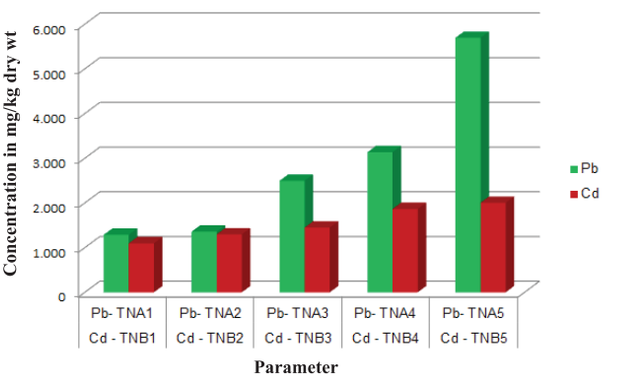


Fig. 5. Concentration of lead and cadmium in *Cyprinus carpio*.

Figure 5 shows that the concentration of lead and cadmium in *Cyprinus carpio* increases with the experimental concentration range. The lowest cumulative concentrations are found in formula Pb-TNA1, Cd-TNB1 and the highest in Pb-TNA5, Cd-TNB5. *Cyprinus carpio* accumulates lead content higher than cadmium.

The result shows that the concentration of Pb and Cd causes a significant effect on the *Cyprinus carpio* in 96 hours. It directly affects the physiological health, and it indirectly affects populations and ecosystems, which may affect human health if humans consume fishes from the affected ecosystem.

Conclusions

The effect of chemicals is greater when the dose is higher and exposure time is longer. The result showed that the LC50 of Cd and Pb were 1.171 mg/l and 0.987 mg/l, respectively. Our study provides good information about LC50 of two heavy metals on the juveniles of *Cyprinus*

carpio, which is a useful basis for risk assessment.

Fish abnormalities (effects, poisoning) increase with an increase in the concentration of lead and cadmium and longer exposure time to them. The test concentration of cadmium is two-fold with that of lead, but the effects appear almost simultaneously. It has been suggested that the toxic levels of lead are two times greater than that of cadmium.

The concentration of lead and cadmium in *Cyprinus carpio* increases with the experimental concentration range. *Cyprinus carpio* accumulates lead higher than cadmium.

REFERENCES

[1] USEPA (2000), *Guidance for Assessing Chemical Contaminant Data for Use in Fish Advisories; Risk Assessment and Fish Consumption Limits Third Edition,* Office of Science and Technology Office of Water (4305), U.S. Environmental Protection Agency Washington, EPA 823-B-00 008.

[2] K. Uysal, E. Yılmaz, K. Esengül (2008), "The determination of heavy metal accumulation ratios in muscle, skin and gills of some migratory fish species by inductively coupled plasma-optical emission spectrometry (ICP-OES) in Beymelek Lagoon (Antalya/Turkey)", *Microchemical Journal,* **90,** pp.67-70.

[3] European Commission (2001), "Commission Regulation (EC) No. 466/2001 of 8th March 2001 setting maximum levels for certain contaminants in foodstuffs", *Off. J. Eur. Commun.,* **77**, pp.1-13.

[4] Jasper Abowei, C.C. Tawari (2011), "Some basic principles of fish processing in Nigeria", *Asian Journal of Agricultural Sciences,* **3(6)**, pp.437-452.

[5] H. Kassa, S.G. Silverman, K. Baroudi (2010), "Effect of a manager training and certification program on food safety and hygiene in food service operations", *Environmental Health Insights,* **4**, pp.13-20.

[6] C.D. Nwani, S.N. Nagpure, R. Kumar, B. Kushwaha (2010), "Lethal concentration and toxicity stress of Carbosulfan, Glyphosate and Atrazine to freshwater air breathing fish Channa punctatus (Bloch)", *International Aquatic Research,***2**, pp.105-111.

[7] F.E. Olaifa, A.K. Olaifa, O.O. Lewis (2003), "Toxic stress of lead on Clarias Gariepinus (African Catfish) Fingerlings", *African Journal of Biomedical Research,* **6**, pp.101-104.

[8] D.J. Finney (1978), "Probit Analysis", *Cambridge University Press,* London.

[9a] FAO (2013a), *Improving pond water quality: Training Series for Aquaculture,* available:ftp://ftp.fao.org/fi/cdrom/fao_training/FAO_Training/General/x6709e/x6709e0 2.htm Accessed 24th June, 2013.

[9b] FAO (2013b), *Pond conditioning through Liming. Training Series for Aquaculture,* available:ftp://ftp.fao.org/fi/cdrom/fao_training/FAO_Training/General/x6709e/x6709e0 5.htm Accessed 24th June.

[10] Brraich Onkar Singh and Kaur Manjeet (2015), "Determination of LC50 of Lead Nitrate for a fish, Labeo rohita (Hamilton - Buchanan)", *International Research Journal of Biological Sciences*, **4(8)**, pp.23-26.

[11] Zeynab Abedi, Mohammadkazem, Sohraby Kohestan Eskandari, Hossein Rahmani (2012), "Comparison of Lethal concentrations (LC_{50}-96h) of $CdCl_2$, $CrCl_3$ and $Pb(NO_3)_2$ in common Carp (Cyprinus carpio) and Sutchi Catfish (Pangasius Hypophthalmus)", *Iranian Journal of Toxicology*, **6(18)**, pp.672-680.

Assessment procedure for sea level rise economic damage due to climate change in agricultural land use: case study in Nam Dinh province

Van Doanh Vu[1*], Ha Phong Doan[2], Thi Thu Trang Bui[1]

[1]*Hanoi University of Natural Resources and Environment*
[2]*Vietnam Institute of Meteorology, Hydrology and Climate Change*

Received 15 January 2018; accepted 24 April 2018

Abstract:

Nam Dinh province is located in the Red river delta with a relatively flat terrain. The North East of Nam Dinh is bordered by the Red river and the southwest by the Day river and has a coastline of approximately 80 km. Given the geographical location, Nam Dinh has favourable conditions for agricultural and aquaculture production. Climate change creates a large number of potential risks for Nam Dinh. This includes sea level rise, typhoons, flood tide, etc. According to the climate change and sea level rise scenario in 2016, if the sea level rises by 100 cm, the flood risk area in coastal districts such as Hai Hau would be 67.34%, Giao Thuy 64.6%, Nghia Hung 81.61% and Xuan Truong 59.3% [1]. According to the monitoring data from Vietnam Institute of Meteorology, Hydrology and Climate Change (IMHEN), the salinity intrusion area due to sea level rise is likely to widen, which degrades and affects the area of the mangrove ecosystem [2]. In addition, the sea level rise may reduce the resilience of irrigation system due to the changes in estuarine and coastal dynamics such as water levels, flows, waves, etc. Through synthesising the local and international guidelines and in combination with expert consultation and field surveys, this paper develops an assessment procedure for sea level rise that causes economic damage to agricultural land use in order to quantify the impact of sea level rise in 4 coastal districts of Xuan Truong, Giao Thuy, Hai Hau and Nghia Hung in Nam Dinh province.

Keywords: ***agricultural land, climate change, coastal district, damage assessment procedure, Nam Dinh province, sea level rise.***

Classification number: ***6.2***

Introduction

The Climate change and sea level rise scenarios 2016 of the Ministry of Natural Resources and Environment (MONRE, Vietnam) estimated that if the sea level rises by 100 cm, approximately 16.8% of the Red river delta area would be in danger of flooding. Accordingly, Nam Dinh is one of the two provinces (together with Thai Binh province) to have the highest flood risk with approximately 58.0% (in area) of the province facing flood risk.

Nam Dinh is located in the Red River delta with a relatively flat terrain. The North East of the province is bordered by the Red river and the southwest by the Day river and has a coastline of nearly 80 km. Nam Dinh has favourable conditions for agricultural and aquaculture production. Climate change creates a great number of potential risks including sea level rise, typhoons, flood tide, etc.

Previous studies on assessing socio-economic damages caused by climate change, especially sea level rise, have been carried out at the regional and global scale, or for national group, sector or specific countries such as the United States, Singapore, Indonesia, etc. Specially, the study by Susmita Dasgupta, et al. (2007) [3] on the impact of sea level rise on developing countries have listed 10 countries where sea level may rise by 1 m. According to this study, the affected land area of Vietnam is estimated to be 5.17%, which equates to 10.79% of the population being affected and a loss of 10.21% GDP. In Vietnam, MONRE has developed and updated three climate change and sea level rise scenarios for Vietnam from 2009 to 2016 [1]. In addition, there exists a great number of researches on the impact of climate change on different sectors, fields and regions such as the impact of climate change on the economy, or research on the impact of climate change - sea level rise in specific coastal provinces such as Ca Mau and Thua Thien - Hue. However, these researches only performed qualitative impact assessment. Only a few studies have quantified the extent of damage,

*Corresponding author: doanh2002vn@gmail.com

yet the scale is too wide for the sea level rise scenarios to be suitable for Vietnam conditions; moreover, the land use changes are not included in these studies.

In fact, there are different methods of calculating economic losses due to climate change, in general, and sea level rise, in particular. These methods have been widely applied in the world. However, the application of such methods is still limited to the case of Vietnam (such as the extent of damages for different target groups and areas due to the impact of sea level rise). Therefore, identifying the objects affected by sea level rise, selecting and using appropriate methods of economic damage calculation to assess the economic damage of the sea level rise because of climate change is important and should be realistic to improve the scientific basis of research assessment of economic damage to agricultural land due to the sea level rise caused by climate change.

This study seeks to establish assessment procedures for economic damage caused to agricultural land use in Nam Dinh. The results of this study form an important basis for the selection of methods to assess the economic damage caused by the sea level rise to the agriculture land use in the Nam Dinh province and can be applied to other areas with similar conditions.

Materials and methods

In order to build an economic damage assessment procedure of sea level rise, the development and standardisation of the database is necessary. Three main methods of research were used:

(i) Collect document and data of the world, in general, and of Vietnam, in particular.

(ii) Use the Delphi method to collect opinions from expert groups on the selection of evaluation scenarios and the impacts of sea level rise on agricultural production.

(iii) Conduct field surveys and community consultation to identify the objects and areas likely to be impacted by sea level rise.

The databases for developing the assessment procedure for damage caused by sea level rise are described in detail below:

a) Documents on climate change and sea level rise:

This document includes: The 1st and 3rd Report of the Intergovernmental Panel on Climate Change (IPCC) [4]; IPCC Technical Guidelines on assessing the impact of climate change and adaptation [5]; Scenario of Climate change and sea level rise of MONRE in 2012, 2016 [1, 6]; Sea level data from the Hydrographic Data and Information Center be long to Viet Nam Meteorological and Hydrological Administration; Guidebook for Assessing the Impacts of Climate Change and Identifying Adaptation Measures of the IMHEN [7]; Climate Change Action Plan for Nam Dinh provincial People's Committee [8]; Irrigation planning to 2020 and orientation to 2030, etc [9].

b) Data on agricultural land in the Nam Dinh province:

Statistical yearbook of Nam Dinh province from 2010 to 2015 [10]; Land use data in 2010, 2015; Land use planning map to 2020 and vision to 2030 (General Department of Land Administration be long to MONRE); Report on socio-economic development of four research districts: Report on development planning for agriculture, fisheries and salt production in 2010-2020, vision to 2030 in the Nam Dinh province; Research and dissertation related to land use in the area, etc.

c) Documents and data on economic damage assessment:

Documents and data on economic damage assessment includes the theory of total economic value (TEV) and the theory of methods used in estimating the economic value commonly used; The researches in the world, areas and in Vietnam on the impact of climate change, sea level rise on land resources, especially on agricultural land; Document on the valuation of economic value of related subjects have been implemented in Vietnam and the Nam Dinh province.

The paper utilises the results from previous researches such as doctoral dissertation, national-level research projects, ministerial-level research project and organisational-level research project that have been conducted on the Nam Dinh area.

Results and discussion

The database supports the framework of the procedure of assessing the economic damage caused by sea level rise

Delphi consultation results:

From two independent rounds of the consultative panel, the synthesised results of the consultation are as follows:

a) Selection of climate change and sea level rise scenarios:

Climate change and sea level rise scenario for Nam Dinh with greenhouse emission level equivalent to RCP 6.0 is used to assess the impact of sea level rise on Nam Dinh land use. On the one hand, the rationale behind the selection of the scenarios from local scientists and officials is due to the potentially high level of impacts of climate change and sea level rise on the Nam Dinh province. On the other hand, the selection of the RCP 6.0 scenarios is also a conservative selection to improve the ability to cope with climate change and sea level rise of the locals in the future.

b) Identification of areas and objects affected by sea level rise:

The study identified four districts and four groups of agricultural land in Nam Dinh with a high risk. The four groups of agriculture land include: paddy land, aquaculture land, salt land and special-use forest land, including protection forest and production forest (since the coastal area of Nam Dinh province only has mangrove forest, thus, mangrove forest the identify name is mangroves forest replaces all other protection and production forest category).

c) Determine the damage level (K):

To assess the economic damage, under the guidance of Inter-ministerial Circular No. 43/2015/TTLT-BNNPTNT-BKHDT [11] on guidelines for statistics and assessment of damages caused by natural disasters and in consultation with local officials and economic experts, the study determined the damage level of agriculture land due to sea level rise for two areas namely the inside and the outside of the dyke protection area (as shown in Table 1).

Table 1. Damage level in two areas inside and outside the dyke protection area.

Areas inside the dyke protection area	Areas outside the dyke protection area
Seriously damaged K = 0.5	Completely damaged K = 1
Partially damaged K = 0.3	Very serious damaged K = 0.7
Mangrove forest ecosystems damaged in the period of 2020-2030 K = 0.2	
Mangrove forest ecosystems damaged in the period of 2020-2030 K = 0.4	

+ The area outside of the dyke protection area is likely to be completely damaged or seriously damaged.

+ The area inside the dyke protection area is less likely to be flooded and the level of impact of the sea level rise is not as direct (due to the dyke system); the level of damage is categorised as seriously damaged and partially damaged.

+ For mangrove forest: based on expert consultation at the Mangrove Forest Research Center - Hanoi Pedagogic University, officers at Xuan Thuy National Park, Nghia Hung Biosphere Reserve and from reference research projects such as Thi Kim Cuc Nguyen (2012) [12], Gilman Eric, et al. (2007) [13], the authors have selected the following damage level: Damage level of 0.2 for from 2020-2030 (corresponding to the sea level rise between 12 and 18 cm) and damage level of 0.4 for from 2040 to 2050 (corresponding to the sea level rise between 24 and 32 cm).

Survey results and community consultation:

The results of survey and community consultation were fairly similar to the applied Delphi method with specific results as follows:

a) Identification of four districts with high risk of sea level rise in Nam Dinh:

Results from consultation and survey of four districts affected by sea level rise include: Nghia Hung, Hai Hau, Giao Thuy and Xuan Truong, of which there are three districts adjacent to the sea. Xuan Truong, although not a coastal district, yet due to terrain conditions and the similarity of irrigation systems with Giao Thuy (the shared irrigation system with Xuan Thuy district) faces the impact of climate change and sea level rise, which could be seen in serious salinity intrusion.

b) Identification of agricultural land groups affected by sea level rise:

The study identified four groups of affected agricultural lands from a total of eight types: paddy land, aquaculture land, salt land and special-use forest land including protection forest and production forest in the four research districts directly affected by sea level rise. In addition, consulted opinions suggested that the dyke systems and drainage sluices and drainage ditches in these four districts will be affected by sea level rise. Summary of affected objects in four districts divided into two areas inside and outside the dyke protection area is as follows:

Table 2. Affected objects located inside and outside of the dyke protection area in four districts.

Area	Affected objects outside the dyke protection area	Affected objects inside the dyke protection area
Nghia Hung, Hai Hau, Giao Thuy	- Area of aquaculture - Area of mangrove forest - Area of salt production	- The sea dyke system - Salinity warning system and Salinity prevented system (not yet invested) - Area of paddy land - Area of aquaculture - Area of salt production
Xuan Truong	- Not affected	- Salinity warning system (investment enhancement) - Area of paddy land

The assessment procedure of economic damage caused by sea level rise

The procedure of determining the economic damage caused by sea level rise has been created in a scientific manner based on domestic and international guidelines from renowned agencies and organisations. The applicability of the guideline is also tested during consultations with experts, administrators and communities in conjunction with the field survey.

Scientific foundation to create the procedure:

To develop a procedure for assessing economic losses due to sea level rise, the study utilised collected documents on climate change, sea level rise, information on the assessment of economic value globally and in Vietnam, which includes 'Technical Guidelines on Impact Assessment of Climate Change and Adaptation' by IPCC (1994) [5] and 'Guidelines for Impact Assessment of Climate change and Identification of Adaptive Measures' by the IMHEN [14].

Foundation to create the procedure:

In addition to using documents in the world and in Vietnam, the research is also based on expert opinion through the application of the Delphi method, the field survey and community consultation. The study has developed a procedure for the assessment of damage caused by sea level rise on the Nam Dinh province. The detailed procedure of assessing economic damage are presented in Fig. 1 with seven main steps:

(1) Select sea level rise scenario under RCP 6.0 by the Delphi method and community consultation.

According to the Delphi Expert Consultation method and community consultation, the selected scenarios are RCP 6.0, representing a high average GHG emission. The reason is for enhanced prevention. Sea level rise of 12, 18, 24 and 32 cm were selected from the scenarios in the period 2020 to 2050.

(2) Create a flood risk map according to the sea level rise according to RCP 6.0 scenario.

- Use the DEM model, topographic data, cross section, slope... and spatial interpolation analysis from the 'decision tree' approach to map the flood risk and the impact of sea level rise to agriculture land.

- Estimate the level of flooding due to sea level rise according to the climate change scenario. Digitise attributed data (sea level rise inundation level) into the basic map based on the 'water surface elevation' method at a selected value.

- Use topographic data, cross-section and slope as inputs for the DEMs model to construct the digital elevation model for four coastal districts of the Nam Dinh province, including:

+ Field survey data: coastal topography, sea level rise, saline intrusion within the last 10 years in four districts Xuan Truong, Giao Thuy, Hai Hau and Nghia Hung; Community surveys on changes in agriculture land use, focusing on the conversion of paddy land, salt land and mangrove forest into aquaculture land (by local people).

+ Map data and statistics: input data for spatial analysis to establish a flooding risk map, including the impact of natural-human relationships: background map topography of four districts with a 1:10,000 scale; Thematic maps on land use status in 2010, 2015 and Nam Dinh land use planning map in 2020 with a scale of 1: 50,000 [15-17]; Area of natural land of four districts according to the statistics of Nam Dinh province by General Statistics Office of Vietnam (GSO) in 2010, 2015.

Identification, description of different wetlands (in dykes, outside dykes) as a basis for map overlays (spatial analysis, floodplain interpolation by decision tree method).

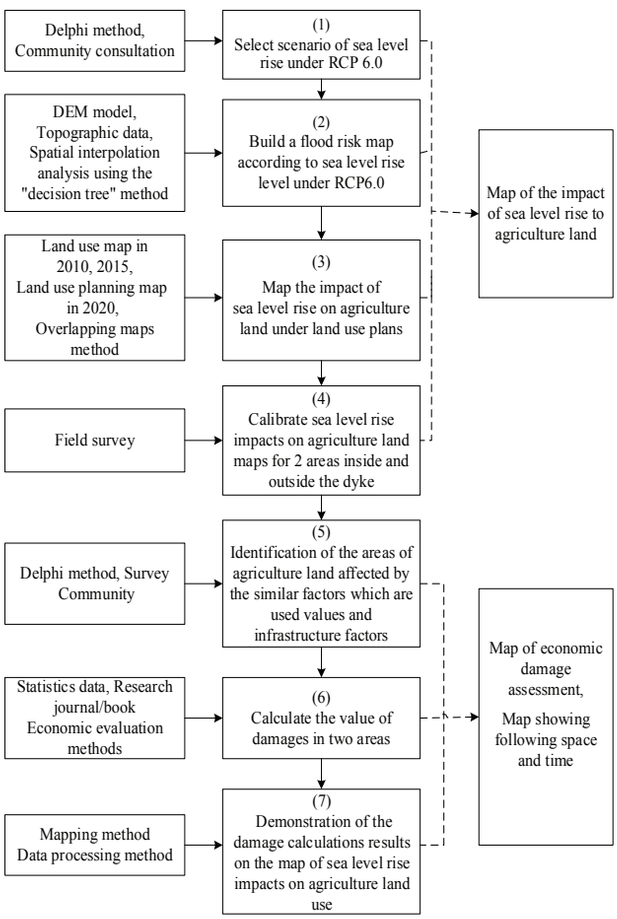


Fig. 1. Economic damage assessment procedure.

(3) Build the map of the impact of sea level rise on agriculture land under land use plans such as land use status map in 2010, 2015 and planning map of 2020.

- Using the method of overlapping maps from the flood risk map in step 2 with three land use maps including land use status map of 2010, 2015 and land use planning map to 2020 vision to 2030 (The use of three maps as three land use scenarios may occur from 2020 to 2050 for more options for local review when the impact of the sea level rise occurs).

- Map of flood risk caused by sea level rise to agriculture

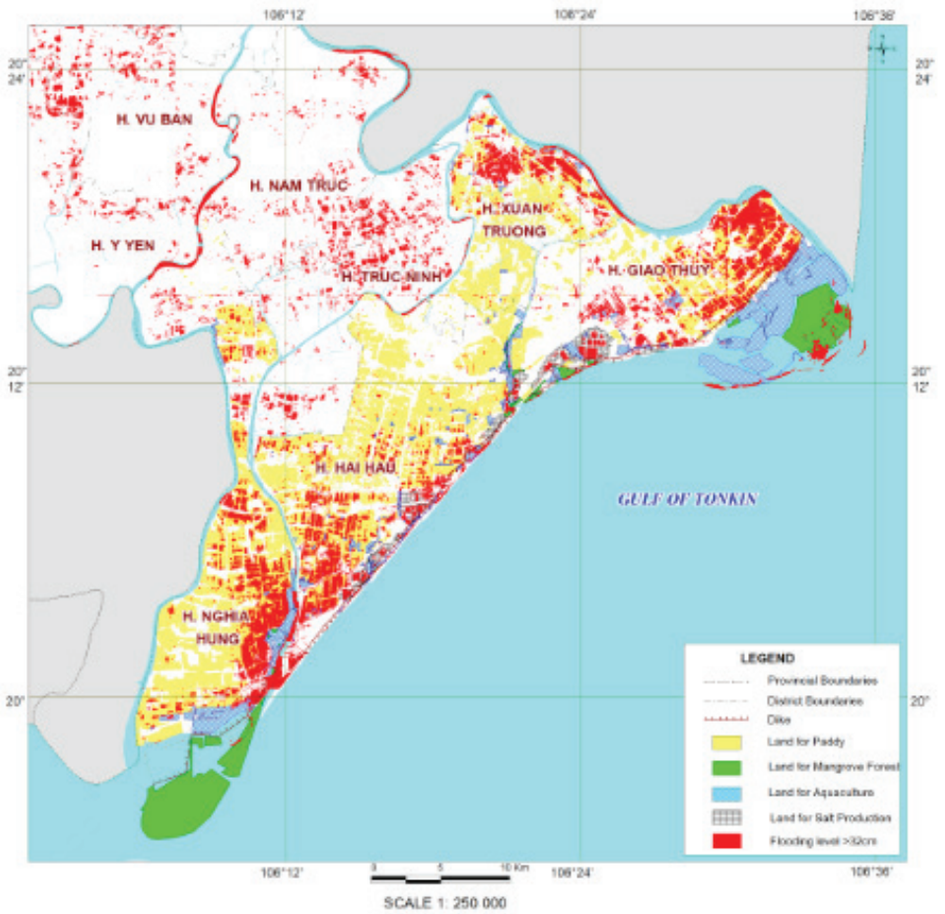


Fig. 2. Map of flood risk caused by sea level rise to agriculture land use for four districts by 2050 according to the land use planning map of 2020.

land use was established by combining floodplains in sea level rise inlays (interpolation on ArcGIS software based on DEMs) and submerged areas according to current land use map (Fig. 2). Overlapped floodplains will be pooled into flooded areas. Calculation of the flood risk area (Si) was adjusted according to the survey data in four districts of Xuan Truong, Giao Thuy, Hai Hau and Nghia Hung.

The map of flood risk caused by sea level rise to agriculture land use for four districts shows that the total area of the agriculture land use was impacted by sea level rise, the proportion of agriculture land use that was most affected by the sea level rise and the object most affected by the sea level rise.

(4) Calibrate maps of sea level rise impacts on agricultural land for two areas inside and outside the dyke protection area.

- Calibrate flood impact maps due to sea level rise in order to correct the area and the flood risk area to eliminate errors of the DEMs model. Calibrate the map according to the actual survey data from suspected points due to spontaneous conversion of land use purpose of local people or errors in interpolation.

- Comparing the methods of mapping the impact of flood risk caused by sea level rise to agriculture land use in the Nam Dinh province and the method of developing flood risk maps from the MONRE scenario includes the additional step of investigating the reality date to develop flood risk maps that are most appropriate. This is a new contribution in the method of mapping the impacts of flood risk caused by sea level rise to agriculture land use in the Nam Dinh province in order to accurately determine the vulnerable area of land types in the area protected by the dyke system.

(5) Identification of the agricultural area (including paddy area, aquaculture area, salt area, special-use forest land and mangroves forest) affected by the used values and infrastructure factors such as dykes and system to prevent salty.

The use of supporting tools in Arcgis and Mapinfo software to calculate the area affected by sea level rise in the four districts of the agriculture land group include:

Table 3. Area of agricultural land affected by sea level rise in both areas inside and outside the dyke from 2020-2050 according to the planning map of 2020.

District	Type of land	Area of land use master plan 2020 (ha)	Flooded area outside the dyke affects to each type of land (ha)				Flooded area inside the dyke affects to each type of land (ha)			
			2020 12 cm	*2030 18 cm*	*2040 24 cm*	*2050 32 cm*	*2020 12 cm*	*2030 18 cm*	*2040 24 cm*	*2050 32 cm*
Nghia Hung	Area of paddy land	8599.4	0.0	0.0	0.0	0.0	599.0	1031.5	1483.3	2160.0
	Area of aquaculture	4639.3	6.2	6.5	7.1	7.5	97.9	129.5	163.4	211.5
	Area of salt production	31.0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
	Area of mangrove forest	2213.7	51.1	57.7	64.3	72.8	7.8	8.5	9.1	9.8
Hai Hau	Area of paddy land	8014.4	0.0	0.0	0.0	0.0	494.6	792.3	1126.6	1633.2
	Area of aquaculture	3090.6	7.9	18.7	31.0	49.5	14.7	29.9	51.7	78.9
	Area of salt production	213.7	2.7	5.2	7.8	13.1	20.2	40.1	67.6	111.2
	Area of mangrove forest	84.5	0.2	0.3	0.3	0.6	0.4	2.9	4.8	7.0
Giao Thuy	Area of paddy land	6561.0	0.0	0.0	0.0	0.0	290.5	694.4	1054.6	1508.4
	Area of aquaculture	5647.7	41.6	90.7	124.4	161.1	3.3	17.6	37.3	69.6
	Area of salt production	305.3	2.2	8.9	13.4	23.0	1.4	14.6	48.9	103.7
	Area of mangrove forest	2178.4	21.4	59.8	100.8	169.1	0.7	1.7	3.3	7.0
Xuan Truong	Area of paddy land	4608.8	0.0	0.0	0.0	0.0	138.7	277.2	412.3	577.4
	Area of aquaculture	1196.2	0.0	0.0	0.0	0.0	7.5	11.0	14.0	16.1
	Area of salt production	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Area of mangrove forest	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

mangrove forest, aquaculture land, salt land and paddy land. This supports the identification of the risk area affected in the two areas inside and outside of the protection area of the dyke system of the four districts. In addition, the study considered the system of sea dykes and the system of salty water drainage for two areas inside and outside the sea dykes.

Table 3 is a typical spreadsheet of the area of agricultural land affected by the sea level rise in both areas inside and outside the dyke. It shows the details of the agricultural land affected by sea level rise. In Hai Hau and Giao Thuy districts, the proportion of the area of salt production affected is extremely high, equivalent to 58.2% and 41.5% of total area of salt production in the two districts by 2050.

(6) Calculate the value of damages in two areas inside and outside the dyke.

The procedure to determine the value of damage caused by sea level rise to agriculture land in 4 districts of Nghia Hung, Hai Hau, Giao Thuy and Xuan Truong consists of 3 main components:

- Determine the average economic value (Gj) of the objects affected by conducting field surveys, collecting data from Nam Dinh Statistical Yearbook and reviewing the literature (journal/book). The affected objects include agriculture land, the dyke systems and the salinity

prevention system. The TEV of Bolt, Pearce and the system of valuation according to Babie are used [18-20].

- Select the discount coefficient (r) to calculate the value in 2010 (the comparative price being applied in statistics in Vietnam).

- Calculate the value of economic damage to agriculture land, dykes and salinity prevention system in 4 districts according to formula (1):

$$TH_{DNN} = \sum(Si \times Gj).K \quad (1)$$

where: TH_{DNN}: damage value of agriculture land caused by sea level rise in one area; Si: area of agriculture land type i affected by sea level rise. In this study, the agriculture land consists of four main categories: aquaculture land, mangroves, salt land and paddy land; Gj: average economic value j of 1 unit of agricultural land; K: damage level.

(7) Map the damage calculations result on the map of impact of sea level rise on agriculture land use.

Through mapping of the results, it could be observed that there are changes in the degree of impacted level to agriculture land caused by the sea level rise in space (study area) and time (from 2020 to 2050). The mapping of the calculated results proved to be significant in rapidly assessing the objects, locations and the difference between the affected objects and the affected area.

Conclusions

This study has created the procedure of evaluating the economic damage of sea level rise due to climate change to agriculture land use in the coastal area of Nam Dinh province with seven main steps. The process of assessing the economic damage of sea level rise due to climate change is scientifically supported by the guidelines of prestigious agencies in the world and in Vietnam in the field of climate change and sea level rise. Besides that, the procedure also ensures the practicality and the possibility of replication when consulting experts and councils and combined with the actual survey.

When the economic damage assessment procedure for the coastal area of Nam Dinh is completed, the data set for assessing the economic damage of sea level rise due to climate change in the coastal area of Nam Dinh province will include:

+ Database on climate change scenarios, sea level rise, data on agricultural land use and related documents on economic valuation of Nam Dinh province.

+ The flood risk map system and maps of the impact of sea level rise on agricultural land use in 4 districts of Nghia Hung, Hai Hau, Giao Thuy and Xuan Truong at four time points from 2020 to 2050 for three other land use plans.

+ The dataset on the flood risk area affected by sea level rise for four districts was separated into two areas inside and outside the dyke, which showed that the area of flood risk was affected by sea level rise in four districts.

This dataset has not only significant scientific value but also practical value for local government and scientists in the context of climate change and sea level rise, which have an impact on the local conditions that are becoming increasingly serious.

The procedure of assessing the economic damage of the sea level rise due to climate change to the agricultural land use in the coastal area of Nam Dinh province can be applied to areas with similar terrain and database.

Recommendations and outlook from the results

This study hypothesised that the flood risk area will be economically affected by four different levels of damage to the two areas inside and outside the dyke.

In order to assess the impact of sea level rise on more accurate and efficient use of agricultural land, there should be an agreement on the databases collected for the calculation from the local statistical yearbook, data from land use status maps and land use planning maps, local research works.

In the future, the application of the economic damage assessment method will be widely used, as, in addition to determining the value of ecosystem management services, it can also be applied to determine compensation for damage in conflicts, environmental conflicts, etc. Therefore, there will be more research on assessing the economic value of ecosystem in the field of natural resources and environment to continue improving the methodology, unify the process of assessment for specific objects and moving towards institutionalising legal documents that regulate ecological economic valuation.

REFERENCES

[1] Ministry of Natural Resources and Environment (2016), *Climate change and sea level rise scenarios for Vietnam*.

[2] Vietnam Institute of Meteorology, Hydrology and Climate Change (2013), *Report on Water Resources Planning of Nam Dinh Province to 2020*.

[3] Susmita Dasgupta, Benoit Laplante, Craig Meisner, David Wheeler & Yan Jianping (2007), "The impact of sea level rise on developing countries: a comparative analysis", *Policy Research Working Paper*, **4136,** The World Bank.

[4] The Intergovernmental Panel on Climate Change (2007), *Climate*

Change 2007 Impacts, Adaptation and Vulnerability, Working Group II, Contribution to the Fourth Assessment, Cambridge University Press.

[5] IPCC (1994), *Technical Guidelines on Assessing the Impact of Climate Change and Adaptation,* University College London and Center for Global Environmental Research.

[6] Ministry of Natural Resources and Environment (2012), *Climate change and sea level rise scenarios for Vietnam*.

[7] Vietnam Institute of Meteorology, Hydrology and Climate Change (2011), *Guidance document on Assessment the impact of climate change and identification of adaptation measures,* Final report on cooperation projects with Denmark, Hanoi.

[8] The People's Committee of Nam Dinh province (2011), *Report of Action Plan for Responding to Climate Change of Nam Dinh Provincial People's Committee.*

[9] The People's Committee of Nam Dinh province (2013), *Report of Irrigation planning to 2020 and orientation to 2030.*

[10] General Statistics Office (2010-2015), *Nam Dinh Statistical Yearbook*, Statistical Publishing House.

[11] *Circular 43/2015/TTLT-BNNPTNT-BKHTT Associate circular guiding statistics and assessment of damage caused by natural disasters*, issued on November 23th 2015.

[12] Thi Kim Cuc Nguyen, Van Dat Tran (2012), "Study on adaptability of coastal mangrove ecosystems under the impact of sea level rise in the Red River Delta", *Journal of Science and Technology Irrigation and Environment*, **37**, pp. 45-52.

[13] Gilman Eric, et al. (2007), "Assessment of Mangrove Response to Projected Relative Sea level Rise and Recent Historical Reconstruction of Shoreline Position", *Environmental Monitoring Assessment*, **124**, pp.105-130.

[14] Vietnam Institute of Meteorology, Hydrology and Climate Change (2011), *Guidance document on Assessment the impact of climate change and identification of adaptation measures,* Final report on cooperation projects with Denmark, Hanoi.

[15] General Department of Land Administration - MONRE (2011), *Thematic Map of land use in 2010*, Hanoi.

[16] General Department of Land Administration - MONRE (2011), *Thematic Map of land use in 2015*, Hanoi.

[17] General Department of Land Administration - MONRE (2012) *Nam Dinh Land Use Planning Map in 2020*, Hanoi.

[18] Bolt Katherine, Giovanni Ruta, Maria Sarraf (2005), *Estimating the Cost of Environmental Degradation: A Training Manual in English, French and Arabic*, Report Nº 106 Environmental Department Papers, Environmental Economic Series, World Bank, Washington.

[19] E.B Barbier, M. Acreman and D. Knowler (1997), *Economic valuation of wetlands: aguide for policy market and planners*.

[20] United Nations (1994), *United Nations Framework Convention on Climate Change,* New York.

Basic trends and strategic viewpoints for sustainable development in adaptation to climate change in the Mekong delta

Dinh Hoa Tran*

Vietnam Academy for Water Resources

Received 25 December 2017; accepted 3 April 2018

Abstract:

The Mekong delta is the southernmost region of Vietnam, with a population of approximately 20 million and a total area (excluding islands) of about 4 million hectares, of which about 2.60 million hectares are used for developing agriculture and aquaculture. The Mekong delta is particularly important for the overall development of the country. However, in recent years, under the impact of global warming, the Intergovernmental Panel on Climate Change has defined the Mekong delta as one of the three most vulnerable plains, due to climate change, sea level rise, and land subsidence. Especially, the impact of water resources exploitation of upper Mekong countries has been posing great challenges to the sustainable development of the region. Recognizing the nature and development trend of the Mekong delta, new threats and challenges that help to form scientific bases for development models and solutions for the region is very important. This article presents assessments on the challenges and strategic viewpoints in regard to sustainable development and adaptation to climate change of the Mekong delta.

Keywords: *climate change, sustainable development, water resources, water resources security.*

Classification number: *6.2*

Introduction

Located in the southern-most part of the country, right next to Ho Chi Minh city (the second centre of political and lead economic development of the country), Mekong delta has a population of nearly 20 million (approximately one quarter of the country population); its agricultural production accounts for over 50% of the country's total production while food export and fruit and fishery export amount to over 90% and 70% of the country respectively. The development of the Mekong delta in the last decades (where agriculture is the mainstream occupation) is a miracle that has been recognized by international communities. This is also a brave base-region, its resilience marked by the nation's historic victory on April 30th. It can be said that the Mekong delta has confirmed its extremely important position in all aspects of economy, politics, society and national security and has received the trust of all the people of the country.

The Mekong delta has experienced the adverse impact of intense human and natural activities. It is a new delta with diverse ecosystems and fertile soils. However, much of the soil suffers from saline contamination and is heavily affected by aluminum. In addition, this area has been influenced adversely by climate change and the over-exploitation of water resources by upstream countries.

However, no long-term investment strategy has been proposed for the Mekong delta, in its development planning; there has been no regional planning, with consideration of technical, economic, social and ecological factors. Meanwhile, we have planned for short-term periods or individual sectors. We have not yet evaluated the full of challenges and opportunities of the region, nor created a breakthrough and positive change. Therefore, it is urgent to have a sustainable development strategy for the delta, which

*Email: tranhoa08@gmail.com

will not only substantially contribute to the development of Vietnam, but also receive the special attention of international community.

Basic trends and viewpoints of challenges in the Mekong delta

In recent years, the Mekong delta has been adversely affected by climate change and the over-exploitation of water resources by the upstream countries. This has considerably impacted the local people, as also development and production activities in the region. This poses great challenges to the sustainable development of not only the Mekong delta, but the country as a whole.

A large number of studies have been carried out on the Mekong delta by researchers and scientific institutes, for identifying an optimal solution to the delta's problems and challenges. However, there is still not the same point of view between the solutions.

Scientists and research organizations from both Vietnam and foreign contries have put forth quite a few ideas and viewpoints on the trends and impacts of these issues on the life and development of the Mekong delta. Research results and viewpoints are very positive, with the desire to help detect the true nature of the impact, in order to find the optimal solution for the region. However, there are still some issues on which there has been no consensus, in terms of assessment viewpoints. So it requires a comprehensive understanding of the mechanism of existing issues and challenges in the region.

Land subsidence in the Mekong delta

Mekong delta was formed about 8,000 years from alluvial deposition. Morphological evolution of the Mekong delta includes two primary processes, viz., alluvial deposition (during flood seasons) and natural subsidence (a continuous slow process lasting centuries) [1]. Currently, a decrease in fluvial sediment supply and widespread over-exploitation of the groundwater and sand have resulted in enhancing subsidence through aquifer compaction (Fig. 1). It is important to note that people are not aware of the land subsidence, as the process is normally slow and hidden from morphological changes, due to deposition or erosion.

In fact, as many major river deltas in the world such as Mississippi (USA), Colorado (USA), Yellow river (China)… [3], due to the combined effect of natural subsidence, global sea-level rise and reduced alluvial deposition, the Mekong delta has become vulnerable to inundation. Recently, in the Mississippi river delta, where tens of hydropower dams have been constructed along the 3,734 km of Mississippi river bank, through 10 states. From 1932 to 2010, in the coastal areas of Louisiana and Mississippi, river deltas have been flooded, with a total affected area of 5,000 km^2 (Fig. 2), which is equivalent to half the area of the southern Ca Mau peninsula.

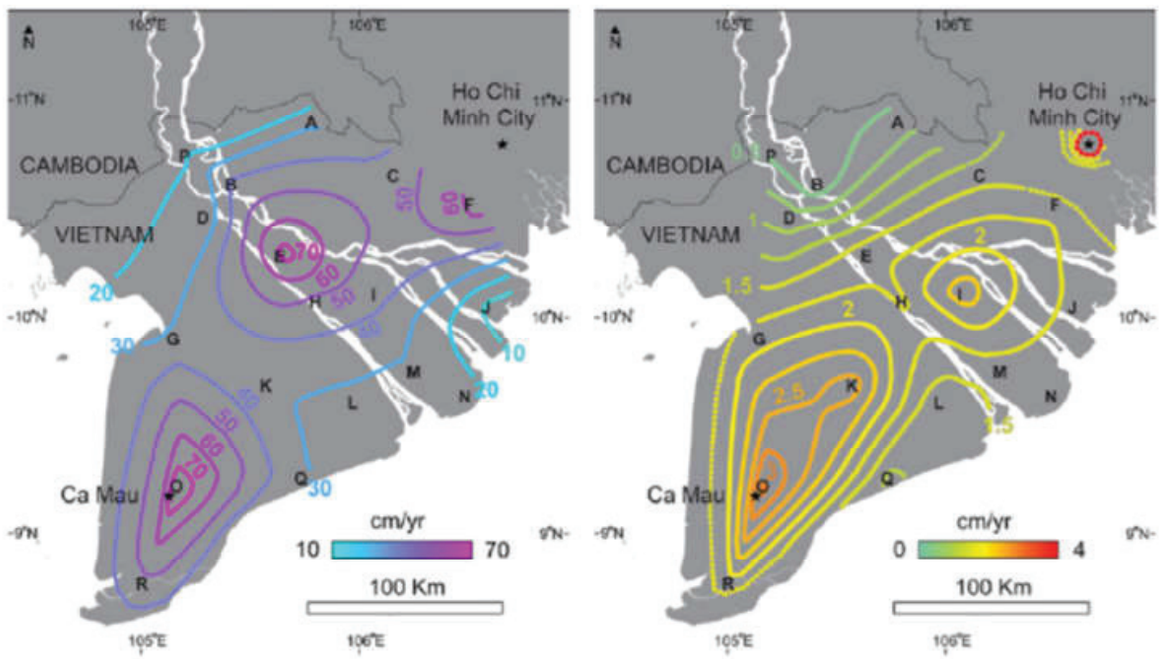


Fig. 1. The rate of lowering of the groundwater table (left) and surface subsidence of the Mekong delta (right) [2].

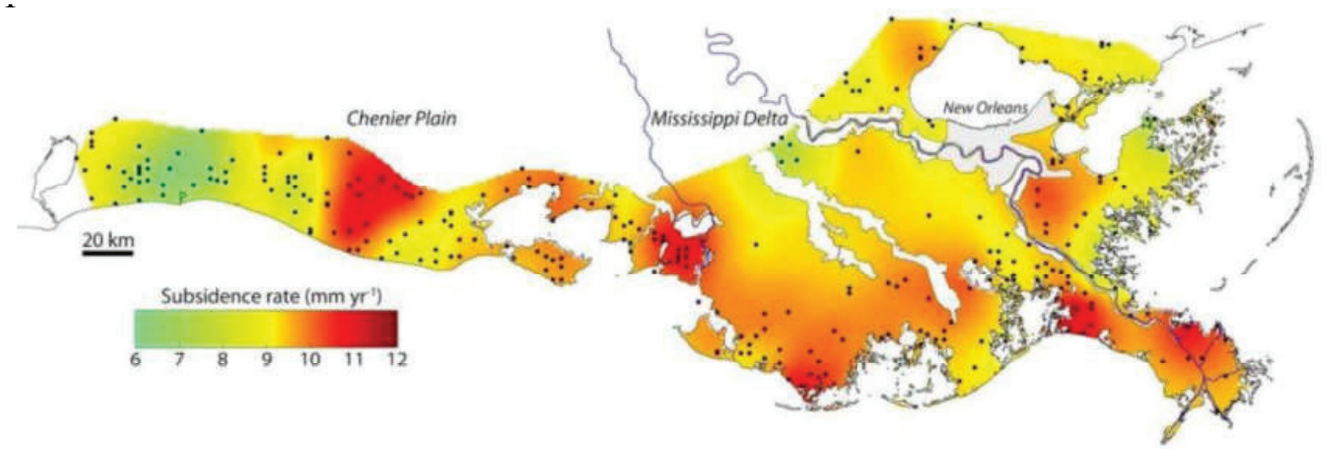


Fig. 2. Natural subsidence in coastal areas in Louisiana and Mississippi river deltas (black dots are observation locations) [4].

From the realities facing us in the major deltas of the world with conditions similar to those of the Mekong delta, it has been found that natural land subsidence in the deltas is a clear and irreversible trend. Groundwater over-exploitation is one of the primary factors responsible for the increasing subsidence in the Mekong river delta. Therefore, it is necessary to have properly observation, management, and analysis. In other words, a comprehensive survey should be conducted for formulating proper solutions.

River and coastal erosion

Beside the natural changes, overexploitation and operations from human activities in the upstream countries and the Mekong delta have resulted in raising the challenges to the Mekong delta. Climate change will have considerable impact on the delta such as flow regime changes, sea-

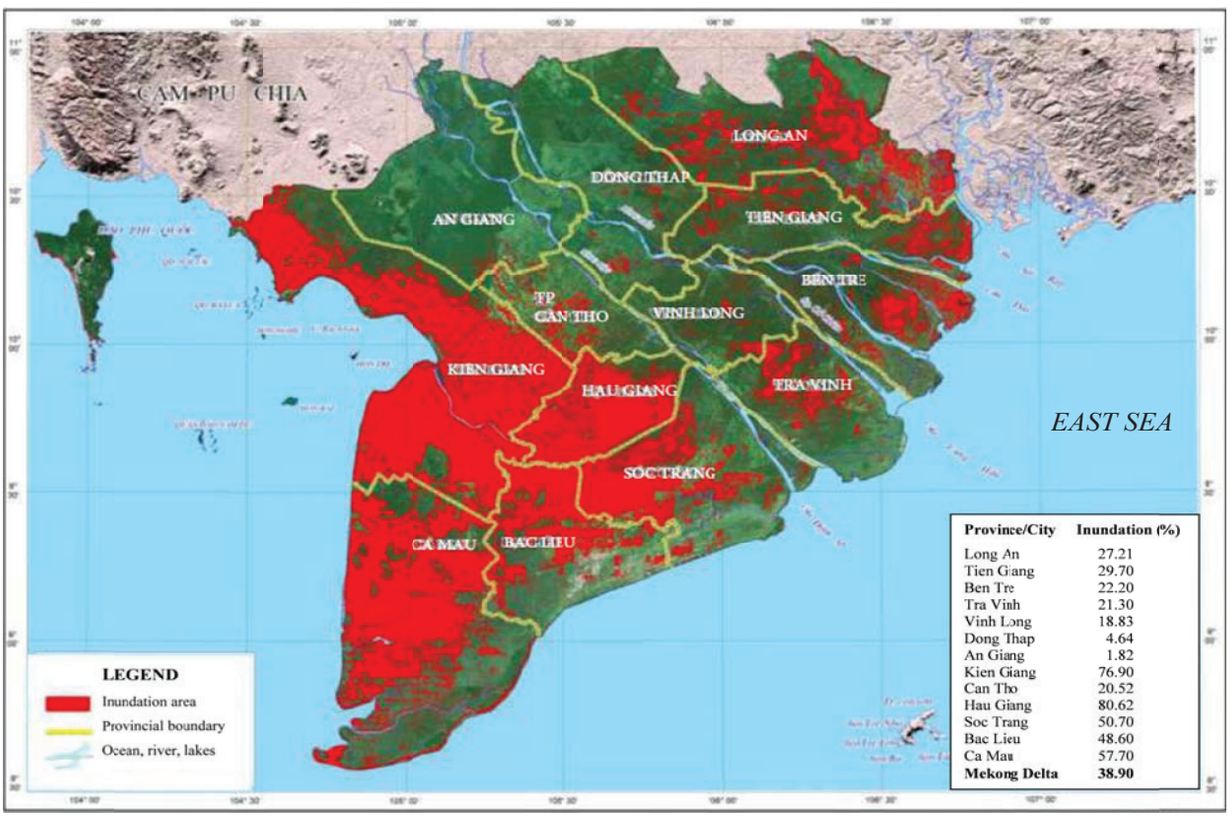


Fig. 3. Inundation map of the Mekong delta projected according to climate change scenario and sea level rise [5].

level rise which causes increased inundation hazard (Fig. 3) and severe erosion in coastal areas [5]. In addition, the operation and over-exploitation of water resources in the reservoirs have resulted in changing of bankfull discharge and decreasing of alluvial deposition, which causes increased erosion in rivers, channels, and coastal areas. Thus, downstream rivers and coastal deposition locations are generally unstable, as per inherent nature.

The recovery of eroded coastal areas due to natural cycles is infeasible; maintaining the sustainability and reduction of eroded locations appears unrealistic, as the capability for controlling these locations has been eroded. The trend of river and coastal erosion is increasing and is unavoidable, in view of the natural condition and exploitation of water, both by domestic sources and by other upstream countries of the Mekong delta. For example, the volume of sand exploitated is in exceess of the fluvial sediment supply. Therefore, using non-structural or structural strategies to restore the eroded bank to its original condition is infeasible. In the Red River delta, the river bed has been declining gradually after hydroelectric power dams were built upstream. For example, in Hai Hau beach - Nam Dinh province, one of the famous Churches has collapsed due to beach erosion. These phenomena are also found in coastal areas of large-scale river basins around the world. America has recently been studying solutions for partial sediment restoration in the coastal areas of Louisiana and Mississippi river deltas (Fig. 4).

Therefore, in order to protect the land, we need to find solutions and accept the lowering of the river bed and elevation of coastal plains. It is easy to recognize that the Mekong delta is unified and has inter-connected canal systems, due to which any changes in a local area will affect the whole system. Any solutions proposed to protect the river and coastal banks should be considered carefully, to avoid other unwanted effects.

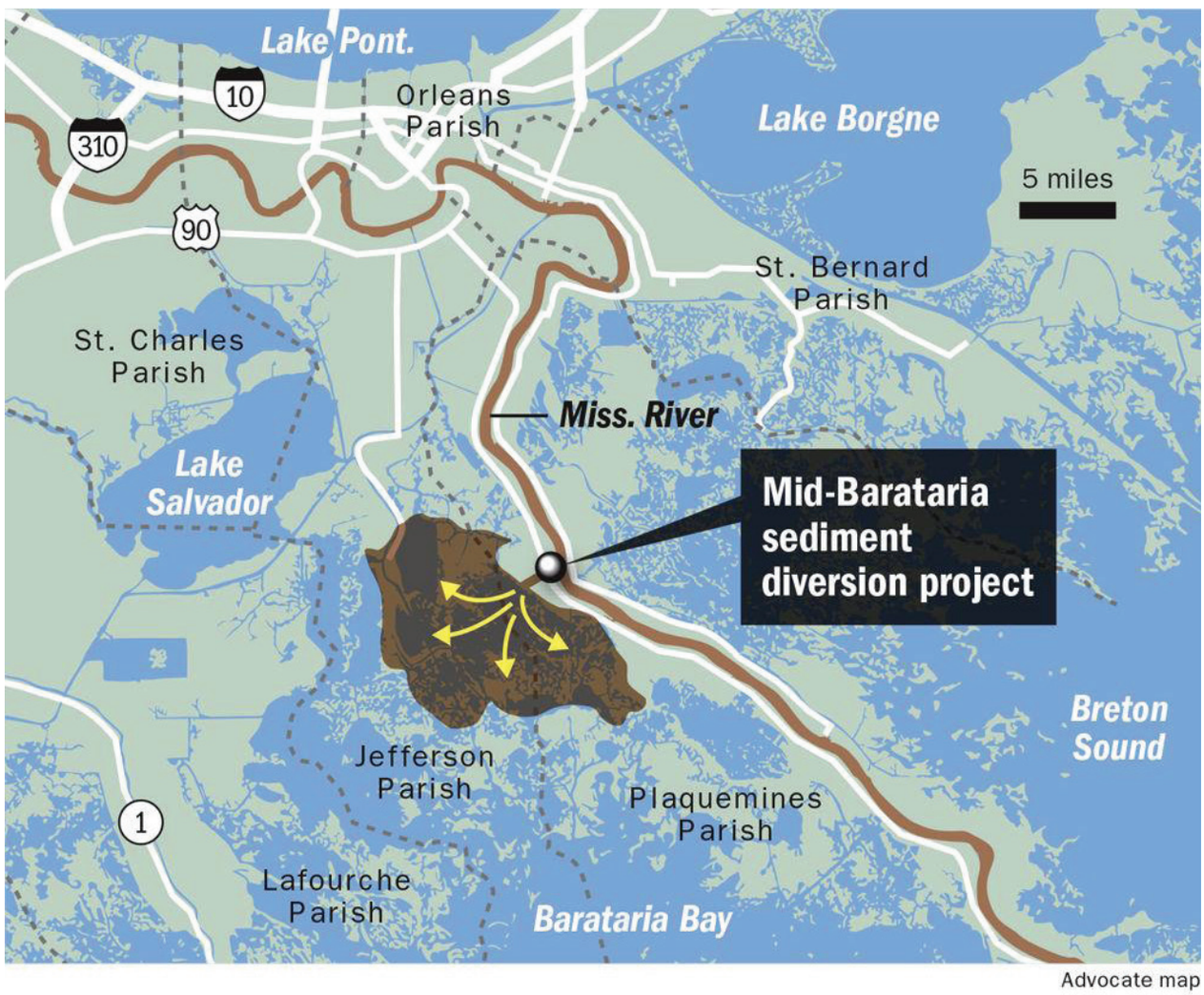


Fig. 4. Pilot project to transfer sediment from the river to plains [6].

Degradation of water resources

Till the time of this report, the total volume of the flows to the Mekong delta have not decreased significantly. The reduction of the total flow volume will become severe, when all the proposed upstream works have been completed. In particular, if the regulation works on the Tonle Sap river for flood protection and rising water level in the Tonle Sap purposes, as well as oil and gas exploitation, and projects of Thailand, Laos, Cambodia to lift water from the Mekong river for agricultural development (covering an estimated total area of more than 1.8 hectares) are implemented, it will lead to extreme conditions of flood and drought flows [7].

The impacts from exploitation of water resources projects upstream are not only reducing the quantity of water, but also its quality, causing reduction of organic elements in water. Further, the pollutant water from pesticides, posing a grave threat to biodiversity, yield, and quality of fisheries, will be a great challenge for the Mekong delta. Along with that, the pollution from waste water and sewage from aquaculture are getting worse. Although the coastal zone in the Mekong delta is conducive for aquaculture and biodiversity, production is unstable due to the frequent occurrence of epidemics. These effects are more severe than the lack of water.

Viewpoint on renovation and sustainable development in the Mekong delta

In the context of international integration, local and regional levels across the country having strategic development based on the strength of each region, there is essential to recognise advantages and minimise shortcomings in order to maximise the available resources. By analyzing and identifying the challenges for the Mekong delta as described above, sustainable development of the delta must be harmonised with three core elements: "*Land - Water - People*". The development history of Vietnam is always associated with "*Water*". Especially the Mekong delta, the development of the region over the past 300 years was always associated with the way which people behaved, vis-a-vis the relationship between "*Land*" and "*Water*". With relatively flat terrain characteristics, an inter-connected canal system and similarity in culture pervading the whole region, it can be said that the Mekong delta is a unified form and has a great interaction with each other. Therefore, Mekong delta needs to have regional linkages, highly systematic activities that bring more efficiency in economic development as well as in adaptation to climate change and other factors.

Placing the Mekong delta in the context of water degradation in terms of quantity, quality, and suspended particles, and increasing environmental pollution, it is very likely that development orientation for the Mekong delta will have to "*Concentrate on the large-scale agricultural production of high quality and productivity; combining services, ecotourism and appropriate industry*". It is time to change the "*traditional food security*" strategy of the recent decades into "*high-quality food security*", by promoting high-value and export-oriented products and looking into large-scale and professional production models.

Conclusions

It can be concluded that a key issue for strategic orientation transformation of the Mekong Delta is the drastic restructuring of its agricultural sector, including: (i) Reorganizing and redistributing land use and farming practices in respect of the decreasing rice cultivation and the number of crops; (ii) Increasing the area of aquaculture; (iii) Increasing the area of fruit trees; (iv) Promoting and using land for ecotourism; (v) Increasing the area of mangroves; and (vi) Re-planning the whole system of irrigation works with respect to ecological areas: fresh water, salt water, and brackish water.

All these strategies will not only enhance local standards of living, but also ensure harmony and conformity with the laws of nature. In other words, "*challenges*" can be turned into "*opportunities*" for development.

REFERENCES

[1] Ta Thi Kim Oanh, Nguyen Van Lap, M. Tateishi, I. Kobayashi, Y. Saito (2005), *Sediment facies changes and delta evolution during Holocene in the Mekong river delta, Vietnam*, in: Mega-deltas of Asia, China Ocean Press.

[2] Laura E. Erban, Steven M. Gorelick, Howard A. Zebker (2014), "Groundwater extraction, land subsidence, and sea-level rise in the Mekong delta, Vietnam", *Environ. Res. Lett.*, **9(8)**, doi:10.1088/1748-9326/9/8/084010.

[3] B.R. Couvillion, J.A. Barras, G.D. Steyer, W. Sleavin, M. Fischer, H. Beck, N. Trahan, B. Griffin, D. Heckman (2011), *Land area change in coastal Louisiana from 1932 to 2010*, USGS - Scientific Investigations Map 3164, scale 1:265,000, 12pp.

[4] Jaap H. Nienhuis, Torbjörn E. Törnqvist, Krista L. Jankowski, Anjali M. Fernandes, Molly E. Keogh (2017), "A new subsidence map for coastal Louisiana", *GSA Today*, **27**, pp.58-59, doi: 10.1130/GSATG337GW.1.

[5] Ministry of Natural Resources and Environment (2016), *Climate change, sea level rise scenarios for Viet Nam*, 170pp.

[6] Louisiana's 2012 Coastal Master Plan.

[7] To Quang Toan (2016), *A study on the impacts of cascaded hydropower dams on the Mekong mainstream to flow, environment, socio-economic situation of the Mekong delta and propose of mitigation measures*, State level research report (KC08.13/11-15).

Vietnam Journal of Science, Technology and Engineering (VJSTE): Manuscript Template

General Instructions on using this template and submitting a manuscript to VJSTE: Thank you for preparing a manuscript for submission to VJSTE. Using this template, or following the guidelines below, will help us in processing your paper. Our goal is to be able to identify each section of your manuscript so that we can accurately record the title, authors, abstract, etc. and also be able to enrich it by including reference links and an accurate layout.

You can submit your paper at http://vietnamscience.vn

So that we can identify the parts of your paper, and even if you do not use our template, please begin each section with the specific key words listed below. Please do not use paragraph breaks in the title, author list, or abstract.

Title

Authors

Affiliations

Abstract

One Sentence Summary

Main Text

Acknowledgements

References

Several of these headings are optional, for example, not all papers will include a one-sentence summary, figures or tables.

Please use the .docx format (all version after Word 2007).

More specific formatting instructions are provided in the actual template, which follows.

Title: How to Format a Science Paper (replace with your real title)

Authors: The author list should be one single paragraph (no breaks). Authors should be listed by first name or initial followed by last name and separated by commas. Use superscript numbers to link affiliations, and symbols *†‡ for author notes. For example, X. Jones[1]*, P. Smith[1,2]

Affiliations:

[1]Affiliations should be preceded by superscript numbers corresponding to the author list, and each affiliation should end with a period.

[2]Each affiliation should be a separate paragraph.

*Correspondence to: include the email addresses or telephone number of the corresponding author(s). Please use the asterisk (*) symbol for the corresponding author information.

Abstract: The abstract should be about 100-150 words, and organized in this structure: An opening sentence that sets the question that you address and is comprehensible to the general reader, background content specific to this study, results, and a concluding sentence. It should be one paragraph only.

One Sentence Summary: A brief summary of the main result of your paper, without excessive jargon and less than 150 characters.

Main Text: In general, this should include a brief (1-2 paragraph) introduction, followed by a statement of the specific scope of the study, followed by results and then interpretations. Please avoid statements of future work or claims of priority, and avoid repeating the conclusions at the end.

All of the Figures and Tables should be cited in order (which should be cited as, for example, figure S1, and table S1). You can include page breaks if you would like to place the figures within the text.

Subheads: These can be included in Research Articles or Reviews and should be brief.

Equations can be included. We do not recommend using the native Word 2007, 2008, 2010 or 2011 equation editor. This can in some cases produce less reliable MathML, the online markup language we use, which may result in display errors. Instead, use the legacy equation editor in word (Insert menu; select insert object; select word equation) or use Mathtype (recommended). If you enter equations in simple LaTeX, check that they will convert accurately (Word 2007 and higher can convert simple LaTeX equations).

Fig. #: (Begin each figure caption with a label, "**Fig. 1.**" for example, as a new paragraph.).

Fig. 1. The figure caption should begin with an overall descriptive statement of the figure followed by additional text. They should be immediately after each figure. Figure parts are indicated with capital letters (**A**). If you prefer, you can place both the actual figures and captions logically through the text near where they are cited rather than at the end of the file (but not both). If a paragraph in the main text begins with the name of a figure, write out "Figure" in full (e.g., <para>"Figure 1 shows….").

Fig. 2. You can place graphics above each equation as part of this file.

Table #: (Begin each table caption with a label "**Table 1.**", etc.) as a new paragraph.

Table 1. Start this caption with a short description of your table. Format tables using the Word Table commands and structures. Do not create tables using spaces or tabs characters.

Acknowledgments: The acknowledgments should include a statement about where the data reported in the paper are presented, archived, or available. If in an archive, include the accession number or a placeholder for it. Please also include relevant funding information such as grant numbers and funding agencies. You can also include a statement of author contributions here. The acknowledgements can be multiple paragraphs.

References:

References should be cited in bracket with a number [1]. Multiple reference citations are separated by commas [2, 3] or if a series, dashes [4-6].

References are listed in numerical order, and in the same order in which they are cited in text. The reference list should include all and only those references you have cited in the text. The reference list appears at the end of the paper.

Please include the above heading "References".

Please do not mix in references with explanatory notes.

The reference consist of: Family name(s) and initials of all authors (year of publication), "Titles of journal articles" (optional), *Title of journal, book or other publication*, **volume(number)**, page numbers.

Example of a reference list:

[1] T. Binzoni, T.S. Leung, D. Boggett, D.T. Delpy (2002), "A new near infrared laser-Doppler flowmeter for deep tissue perfusion monitoring", *MAGMA*, **14,** pp.74-75.

[2] T. Binzoni, T.S. Leung, D. Boggett, D.T. Delpy (2003), "Non-invasive laser Doppler perfusion measurements of large tissue volumes and human skeletal muscle blood RMS velocity", *Phys. Med. Biol.*, **48(1)**, pp.2527-2549.

ISSN 2525-2461

Vietnam Journal of Science, Technology and Engineering

JUNE 2018 • VOL.60 NUMBER 2

EDITORIAL OFFICE

113 Tran Duy Hung Str., Cau Giay Dist., Hanoi

Tel: (84.24) 39436793; Fax: (84.24) 39436794

Email: khcnvn@most.gov.vn

Website: vietnamscience.vn

PUBLICATION LICENCE

No.1153/GP-BTTTT 26th July 2011
No.2528/GP-BTTTT 26th December 2012
No.592/GP-BTTTT 28th December 2016