BACKGROUND

Vibrio vulnificus (*V. vulnificus*), a gram-negative bacterium of the family *Vibrionaceae*, is a worldwide inhabitant of salt water [1,2]. These bacteria tend to be more common in warmer waters (17–20°C) [3,4]. *V. vulnificus* causes serious illness including necrotizing fasciitis and septicemia, and death in persons with preexisting liver disease or compromised immune systems [5–7]. People with chronic liver disease, particularly liver cirrhosis, are more prone to developing infection, and are at greatest risk for an adverse outcome [8,9]. Other predisposing factors are iron overload and hemochromatosis, and immunosuppression caused by steroid treatment, malignancy, human immunodeficiency virus (HIV) infection, renal failure and organ transplantation [10,11].

V. vulnificus infection was first reported by Roland in 1970 in a case of endotoxic shock with leg gangrene [12]. In Japan, Matsuo et al. reported the first case of *V. vulnificus* infection in 1978 [13]. There have since been case reports of approximately 200 patients over a period of about 30 years [14]. However, because the 200 cases represent only those that were published, the actual number of *V. vulnificus* infections is considered to be higher [14]. The annual number of *V. vulnificus* septicaemia cases in Japan has been estimated at 425 (95% CI 238–752) [15]. The prevalence of *V. vulnificus* septicaemia is estimated at 3.3 per million in Japan. The annual number of *V. vulnificus* infection in Japan is notably higher than in other countries, such as Korea and the USA [15]. The prevalence of *V. vulnificus* septicaemia is low in the general population, and estimated at 0.6 per million in USA [8]. A study of the epidemiological and clinical characteristics of *V. vulnificus* infections reported in Japan from 1975 to 2005 [14] found that about 90% of Japanese patients with *V. vulnificus* infection had liver disease such as liver cirrhosis, hepatocellular carcinoma (HCC), and chronic hepatitis.

It is estimated that approximately 2 million Japanese people are chronically infected with hepatitis C virus (HCV) [16]. Approximately 35,000 patients died due to HCC in Japan, and the number of deaths in Japan from HCC continues to increase. In Japan, approximately 80% of HCCs are caused by HCV and about 10% by hepatitis B virus (HBV). The increase in the number of HCC patients due to HCV in turn contributes to the increase in the number of deaths in Japan from HCC.

In Japan, patients with liver disease are not provided adequate educational opportunities. Therefore, in this study, we assessed knowledge about *V. vulnificus* infection in patients with chronic liver disease.

MATERIAL AND METHODS

Subjects

Between August 1, 2008 and October 31, 2008, anonymous questionnaires relating to general knowledge of *V. vulnificus* infections were given to all patients with chronic liver diseases who had been treated at 14 geographically-distinct institutions in Japan, as well as to their attending physicians. A physician at each participating institution completed a questionnaire with the patient's medical information and handed the questionnaire to the patient. Next the patient was interviewed about *V. vulnificus* infection. The questionnaire was conducted in one-to-one interview style by patient and physician. A physician at each medical institution returned the completed questionnaires to Kurume University of Medicine; 1,336 completed questionnaires were recovered, and the collection rate was 97.3% (1,336/1,373). The 14 medical organizations were those where many liver specialists authorized by the Japan Association for the Study of the Liver work full-time.

We mailed questionnaires directly to these 14 medical institutions through a collaborative study. A database for the results of our investigation was compiled at the Department of Digestive Disease Information & Research, Kurume University School of Medicine.

Items of investigation

Anonymous questionnaires asked patients and their attending physicians to respond to the following items; patient background (age, gender, diagnosis of liver diseases, comorbidities, and steroid use), patient awareness and understanding of *V. vulnificus* infection, frequency of eating raw fish and shellfish, raw shrimp and sushi, the season in which raw fish was eaten, and frequency of bathing in the sea and shellfish gathering. After the patients answered the questionnaires, we provided them with literature containing basic information about *V. vulnificus* infection.

The investigation was conducted in accordance with the "ethical guidelines on epidemiological studies" of the Ministry of Education and Science and the Ministry of Health, Labour and Welfare, and observed the spirit of the Helsinki Declaration. Physicians at study facilities explained to patients the content and significance of the study and obtained consent in accordance with each facility's regulations.

Statistical analysis

All data are expressed as mean ± standard error. Differences between the 2 groups were analyzed using the Welch's test and the Mann-Whitney U test. Differences were judged significant for $p<0.05$ (2-tailed). All statistical analyses were conducted using JMP Version 6 (SAS Institute, Cary, NC, USA).

RESULTS

Patient's background

We analyzed 1,336 questionnaires in which 656 indicated they were males, 670 females, and 10 did not specify gender. Mean age was 61.4±12.3, as shown in Table 1.

Among the 1,336 patients, the distribution of diagnoses of liver disease was as follows: HCV-related liver diseases 760 (56.9%), HBV-related liver diseases 266 (19.9%), HCV & HBV-related liver diseases (simultaneous infection) 4 (0.3%), non-B non-C-related liver diseases 19 (1.4%), other liver diseases 273 (20.4%), and no answer 14 (1.0%). Some institutions differed significantly in patients' age, gender distribution, or liver diseases, compared to the overall averages (Table 1).

Table 1. Clinical information for 1,336 patients from whom questionnaires returned.

Prefecture	Medical institution	n	Collection rate of questionnaire	Age			Sex				Liver diseases												
				Mean	SD	P value	Male	Female	No answer	P value	HCV-related liver disease		HBV-related liver disease		HCV & HBV-related liver disease		NBNC-related liver disease		The other		No answer		P value
			(%)	year			n	n	n		n	(%)	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)	
Hokkaido	Sapporo Medical University School of Medicine	100	(100.0)	59.4	13.4	NS	44	55	1	NS	48	(48.0)	27	(27.0)	0	(0.0)	4	(4.0)	19	(19.0)	2	(2.0)	NS
Fukushima	Fukushima Medical University School of Medicine	97	(97.0)	63.6	12.2	NS	38	52	7	NS	42	(43.3)	12	(12.4)	0	(0.0)	2	(2.1)	34	(35.1)	7	(7.2)	<0.05
Chiba	Chiba University Graduate School of Medicine	97	(97.0)	58.8	13.5	NS	47	50	0	NS	63	(65.0)	15	(15.5)	0	(0.0)	1	(1.0)	18	(18.6)	0	(0.0)	NS
Shizuoka	ELM Medical Clinic	100	(100.0)	57.2	12.2	0.001	71	29	0	<0.0001	38	(38.0)	36	(36.0)	0	(0.0)	0	(0.0)	26	(26.0)	0	(0.0)	<0.001
	Miyazaki Clinic	100	(100.0)	51.0	15.3	<0.000000001	53	47	0	NS	40	(40.0)	37	(37.0)	0	(0.0)	0	(0.0)	23	(23.0)	0	(0.0)	<0.001
Aichi	Social Insurance Chukyo Hospital	100	(100.0)	61.4	14.1	NS	44	55	1	NS	59	(59.0)	14	(14.0)	0	(0.0)	1	(1.0)	25	(25.0)	1	(1.0)	NS
Fukuoka	Kurume University School of Medicine	213	(100.0)	60.6	11.6	NS	86	127	0	0.01	135	(63.4)	38	(17.8)	0	(0.0)	1	(0.5)	39	(18.3)	0	(0.0)	NS
Saga	Inuzuka Hospital	100	(100.0)	64.4	11.0	<0.05	47	52	1	NS	85	(85.0)	6	(6.0)	1	(1.0)	0	(0.0)	8	(8.0)	0	(0.0)	<0.00001
Nagasaki	Narao Hospital	122	(81.3)	66.5	10.8	<0.00001	68	54	0	NS	71	(58.2)	42	(34.4)	0	(0.0)	1	(0.8)	7	(5.7)	1	(0.8)	<0.0001
	National Nagasaki Medical Center	59	(98.3)	64.5	10.5	NS	29	30	0	NS	47	(79.7)	6	(10.2)	1	(1.7)	0	(0.0)	5	(8.5)	0	(0.0)	<0.01
Oita	Oita University	100	(100.0)	59.6	13.4	NS	41	59	0	NS	53	(53.0)	16	(16.0)	2	(2.0)	3	(3.0)	25	(25.0)	1	(1.00	<0.05
	National Hospital Organization Oita Medical Center	48	(96.0)	64.9	12.5	<0.05	23	25	0	NS	31	(64.6)	8	(16.7)	0	(0.0)	4	(8.3)	4	(8.3)	1	(2.1)	0,001
	Oita Cardiovascular Hospital	50	(100.0)	67.0	10.9	<0.001	29	21	0	NS	36	(72.0)	8	(16.0)	0	(0.0)	2	(4.0)	4	(8.0)	0	(0.0)	NS
	Abe Diabetes Clinic	50	(100.0)	62.0	10.6	NS	36	14	0	0.001	12	(24.0)	1	(2.0)	0	(0.0)	0	(0.0)	36	(72.0)	1	(2.0)	<0.00000000000001
Total		1336	(97.3)	61.4	12.3		656	670	10		760	(56.9)	266	(19.9)	4	(0.3)	19	(1.4)	273	(20.4)	14	(1.0)	

Liver cirrhosis was observed in 304 (22.8%) patients, including those with HCV-related liver cirrhosis (177 cases), HBV-related liver cirrhosis (66), HCV & HBV-related liver cirrhosis (1), non-B non-C-related liver cirrhosis (11), and other liver diseases such as primary biliary cirrhosis (PBC) and autoimmune hepatitis (AIH) (49).

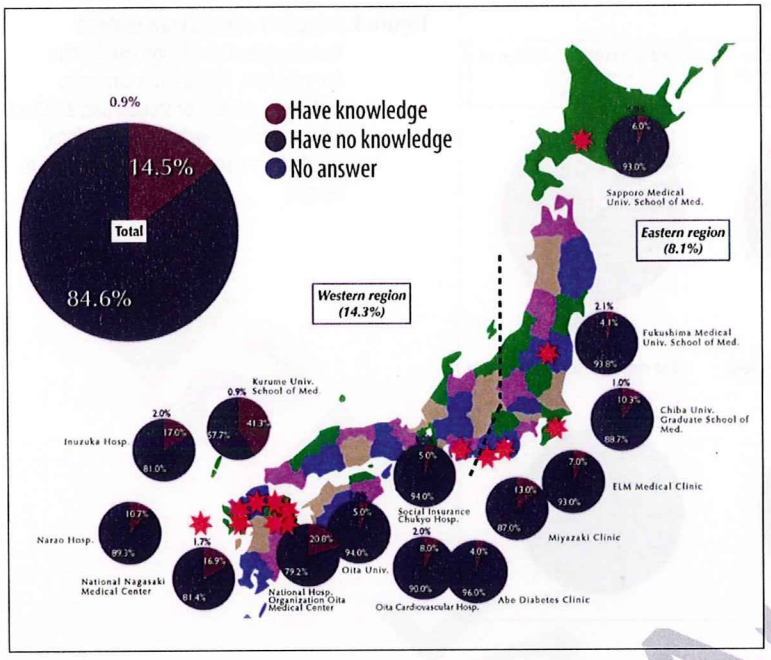

Figure 1. Knowledge of *V. vulnificus* infections among all patients with liver diseases. Only 14.5% of such patients had knowledge of this infection. Fourteen red stars indicate the location of each medical institution. Japan consists of 47 prefectures. Half of east of Japan, including Tokyo, where Japan is metropolitan, is called eastern Japan, and the western half of Japan is called western Japan. The broken line indicates the boundary between the 2 areas.

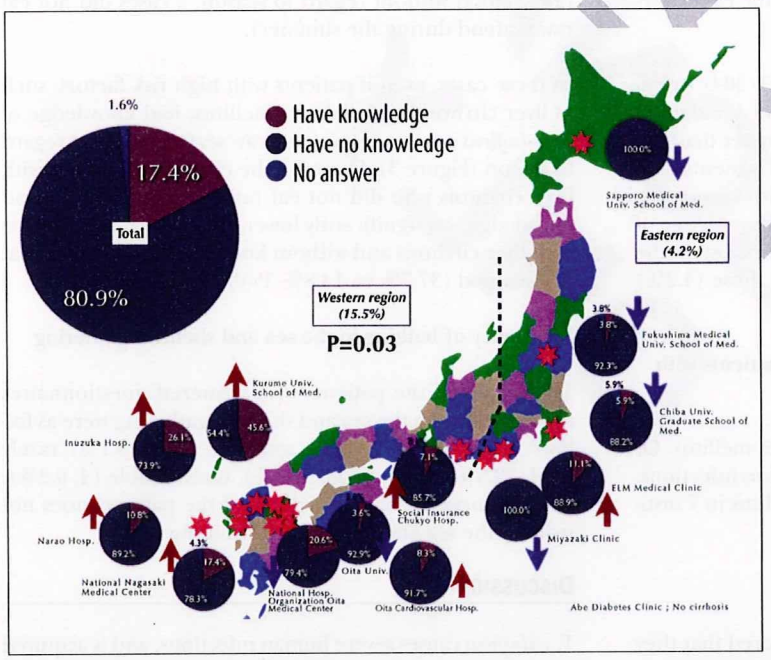

Figure 2. Knowledge of *V. vulnificus* infections in patients with liver cirrhosis. The rate of patient knowledge in the western region of Japan was significantly higher than in the eastern region. The upward pointing arrow indicates an increase in the rate of *V. vulnificus* infections in a given institution compared to Figure 1. A down-pointing arrow indicates a decrease compared to Figure 1.

There were associated comorbidities in 732 (54.8%) of all patients with liver disease. These were classified using International Classification of Diseases (ICD) criteria: diseases of the circulatory system (372 cases), endocrine, nutritional and metabolic diseases (316), diseases of the digestive system (73), malignant neoplasms (54), diseases of the genitourinary system (33), diseases of the nervous system (23), diseases of the musculoskeletal system and connective tissue (18), diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism (17), diseases of the respiratory system (16), mental and behavioral disorders (7), diseases of the skin and subcutaneous tissue (4), diseases of the eye and adnexa (4), certain infectious and parasitic diseases (2), and other diseases (6).

There were 563 patients (42.1%) with no comorbidities, 40 patients provided no answer about comorbidities, and 1 patient was unassessable. There were 60 patients who took oral or topical steroids for their liver disease or comorbidities.

Knowledge of *V. vulnificus* infection in patients with liver diseases

Only 14.5% (194/1,336) of patients with liver disease had general knowledge regarding *V. vulnificus* infections. The level of patient knowledge varied widely among medical institutes, ranging from 4.0% to 41.3%. The mean rate (14.3%) of knowledge among patients who resided in the western re-

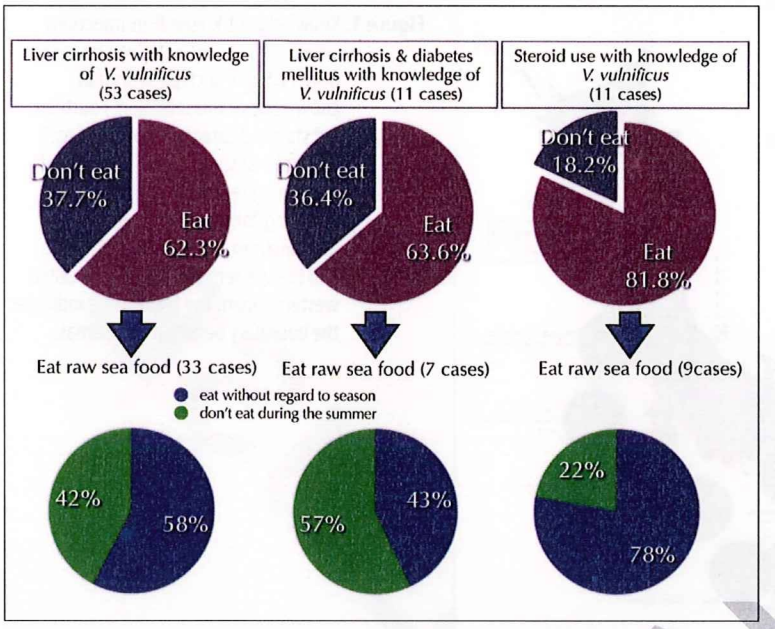

Figure 3. Frequency of eating raw seafood. Even if patients had high risk factors for infection, such as liver cirrhosis, diabetes mellitus, or steroid use, and had knowledge of *V. vulnificus* infections, most ate raw seafood without regard to season.

gion of Japan was higher than that (8.1%) in the eastern region (Figure 1).

Of 304 patients with liver cirrhosis, 17.4% (53/304) (minimum 0%, maximum 45.6%) had knowledge of *V. vulnificus* infection (Figure 2). This rate (17.4%) was higher than the mean rate (14.5%) of knowledge among all patients with liver diseases, but the proportion of those with knowledge was lower in 6 institutes. The rate (15.5%) of knowledge of *V. vulnificus* infection among those with liver cirrhosis in the western region was significantly higher than those (4.2%) in the eastern region (P=0.03).

Knowledge of *V. vulnificus* infection among patients with liver cirrhosis and diabetes mellitus

Sixty patients had liver cirrhosis and diabetes mellitus. Of these, 11 (18.3%) had knowledge of *V. vunificus* infections. Patients with liver cirrhosis and diabetes mellitus in 7 institutes had no knowledge of the infection.

Frequency of intake of raw seafood

A total 1,170 (87.6%) of 1,336 patients answered that they often eat raw seafood. Most (1,002 cases, 85.6%) of the patients answered that they eat raw seafood without regard to season. There was significant difference between patients with knowledge and without knowledge who eat raw seafood (P<0.00001).

Thirty-three of 53 patients who suffered from liver cirrhosis and who had knowledge of *V. vulnificus* infection ate raw seafood (19 cases ate raw seafood without regard to season; 14 did not eat raw seafood during the summer). Seven of 11 patients, who suffered from liver cirrhosis and diabetes mellitus and with knowledge of *V. vulnificus* infection, ate raw seafood (3 cases ate raw seafood without regard to season; 4 cases did not eat raw seafood during the summer). Nine of 11 patients who took steroids and who had knowledge of *V. vulnificus* infection ate raw seafood (7 cases ate raw seafood without regard to season, 2 cases did not eat raw seafood during the summer).

In these cases, even if patients with high risk factors, such as liver cirrhosis and diabetes mellitus, had knowledge of *V. vulnificus* infections, most ate raw seafood without regard to season (Figure 3). However, the rate of the patients with liver cirrhosis who did not eat raw seafood and who had knowledge was significantly lower than that of the patients with liver cirrhosis and without knowledge who did not eat raw seafood (37.7% vs. 14.8%, P=0.0001).

Frequency of bathing in the sea and shellfish gathering

The results of the patients who answered questionnaires about bathing in the sea and shellfish gathering were as follows: often (18 cases, 1.3%), sometimes (122, 9.1%), rarely (394, 29.5%), never (768, 57.5%), unassessable (4, 0.3%), and no answer (30, 2.2%). Most of the patients does not swim in the sea and did not go clamming.

Discussion

V. vulnificus causes severe human infections, and is acquired through wounds or contaminated seafood. In Japan, many cases of *V. vulnificus* infection have been reported to occur in the western region and more than half of the infections were reported to occur in Kyusyu [14,17]. Inoue et al. did a retrospective survey in which 1,693 hospitals from across Japan were surveyed, including advanced life saving emergency centers and dermatology institutions [17]. Ninety-four cases were confirmed as *V. vulnificus* infections over 5 years. The authors reported that many *V. vulnificus* infections occurred in Kyusyu, especially in the coastal areas of the Ariake and Yatsushiro Seas.

One reason for the high incidence of *V. vulnificus* infection in the western region in Japan is thought to be higher seawater temperature. *V. vulnificus* proliferates in areas where, or during months when, the water temperature exceeds

17–20°C [3,4]. The other reason is the greater number of HCV carriers in Kyusyu. Geographically, HCC is more frequent in western than eastern Japan [16].

The awareness of *V. vulnificus* infections among Japanese physicians is reported to be low [15]. Only 15.7% of emergency-physicians were reported to have a basic knowledge of *V. vulnificus* infections. In 2004, Osaka et al. reported that emergency-room physicians who work in the western region of Japan had more knowledge of *V. vulnificus* infections [15]. The Ministry of Health, Labour and Welfare warned of the risk of *V. vulnificus* infection on their website in 2006.

Our study demonstrates that awareness of *vulnificus* infections among patients with chronic liver diseases is low. Medical institutions in Japan, except for Kurume University of Medicine, did not provide educational opportunities for learning about *V. vulnificus* infections. Although the 15.5% rate of knowledge among patients with liver cirrhosis in the western region was significantly higher than that in the eastern region (P=0.03), this rate is far from adequate.

The most significant host factor contributing to virulence is chronic liver disease [8,9]. This may act in several ways including: portal hypertension, causing shunting of the bacteria around reticuloendothelial cells in the liver [18,19]; decreased clearance of bacteria from the portal circulation by Kupffer's cells in the diseased liver [19]; increased iron in the serum, as seen in patients with cirrhosis and hemochromatosis, which promotes growth of *V. vunificus* [7,20]; and achlorhydria occurring naturally or induced by medications [8,19,21].

Factors conferring high risk include: liver disease and other diseases with possible hepatic involvement or elevated serum iron levels (including cirrhosis, alcoholism, malignancy, hemochromatosis, or thalassemia major) [8,9,19,20]; therapeutically induced or naturally low gastric acid (achlorhydria or antacid or H2 blocker use) [8,19,21]; and conditions that compromise the immune system (HIV infection, diabetes mellitus, renal disease, or steroid dependency) [10,11,19].

Primary liver cancer, 95% of which is HCC, is ranked third among men and fifth among women as a cause of death from malignant neoplasms in Japan [22,23]. The number of deaths and death rate of HCC has been increasing. Geographically, HCC is more frequent in western than eastern Japan. Meanwhile, according to the Ministry of Internal Affairs and Communications, yearly per capita fish consumption in Japan was 63.2 kilograms on average for 2003–2005, about 4 times higher than the world average. The Japanese custom of eating raw fish and shellfish such as sashimi or sushi has become widely known throughout the world. Their traditional eating habits are attributed to the fact that patients with knowledge about *V. vulnificus* infections still ate raw seafood.

Therefore, it is important for physicians in Japan to expand their knowledge of *V. vulnificus* infections and become familiar with prevention methods. It is also important for patients with liver diseases to acquire the necessary knowledge of *V. vulnificus* infections and prevention methods, such as avoidance of eating raw seafood during the summer. Because of rapid aggravation and high mortality, patients should also keep an emergency contact number handy.

Conclusions

In conclusion, standardized guidelines for prevention of *V. vunificus* infections and education of patients with liver diseases should be required.

References:

1 Blake PA, Merson MH, Weaver RE et al: Disease caused by a marine Vibrio. Clinical characteristics and epidemiology. N Engl J Med, 1979; 300: 1–5

2 Morris JG Jr, Black RE: Cholera and other vibrioses in the United States. N Engl J Med, 1985; 312: 343–50

3 Wright AC, Hill RT, Johnson JA et al: Distribution of Vibrio vulnificus in the Chesapeake Bay. Appl Environ Microbiol, 1996; 62: 717–24

4 Heidelberg JF, Heidelberg KB, Colwell RR: Seasonality of Chesapeake Bay bacterioplankton species. Appl Environ Microbiol, 2002; 68: 5488–97

5 Centers for Disease Control and Prevention (CDC). Vibrio vulnificus infections associated with raw oyster consumption – Florida, 1981–1992. MMWR Morb Mortal Wkly Rep, 1993; 42: 405–7

6 Hlady WG, Klontz KC: The epidemiology of Vibrio infections in Florida, 1981–1993. J Infect Dis, 1996; 173: 1176–83

7 Wright AC, Simpson LM, Oliver JD: Role of iron in the pathogenesis of Vibrio vulnificus infections. Infect Immun, 1981; 34: 503–7

8 Haq SM, Dayal HH: Chronic liver disease and consumption of raw oysters: a potentially lethal combination – a review of Vibrio vulnificus septicemia. Am J Gastroenterol, 2005; 100: 1195–99

9 Shapiro RL, Altekruse S, Hutwagner L et al: The role of Gulf Coast oysters harvested in warmer months in Vibrio vulnificus infections in the United States, 1988–1996. Vibrio Working Group. J Infect Dis, 1998; 178: 752–59

10 Strom MS, Paranjpye RN: Epidemiology and pathogenesis of Vibrio vulnificus. Microbes Infect, 2000; 2: 177–88

11 Morris JG Jr: Cholera and other types of vibriosis: a story of human pandemics and oysters on the half shell. Clin Infect Dis, 2003; 37: 272–80

12 Roland FP: Leg gangrene and endotoxin shock due to vibrio parahaemolyticus – an infection acquired in New England coastal waters. N Engl J Med, 1970; 282: 1306

13 Matsuo T, Kohno S, Ikeda T et al: Fulminating lactose-positive Vibrio septicemia. Acta Pathol Jpn, 1978; 28: 937–48

14 Oishi H, Ura Y, Mitsumizo S, Nakashima M: A collective review of Vibrio vulnificus infection in Japan. Kansenshogaku Zasshi, 2006; 80: 680–89

15 Osaka K, Komatsuzaki M, Takahashi H et al: Vibrio vulnificus septicaemia in Japan: an estimated number of infections and physicians' knowledge of the syndrome. Epidemiol Infect, 2004; 132: 993–96

16 Yoshizawa H: Hepatocellular carcinoma associated with hepatitis C virus infection in Japan: projection to other countries in the foreseeable future. Oncology, 2002; 62(Suppl.1): 8–17

17 Inoue Y, Ono T, Matsui T et al: Epidemiological survey of Vibrio vulnificus infection in Japan between 1999 and 2003. J Dermatol, 2008; 35: 129–39

18 Blake PA, Merson MH, Weaver RE et al: Disease caused by a marine Vibrio. Clinical characteristics and epidemiology. N Engl J Med, 1979; 300; 1–5

19 Koenig KL, Mueller J, Rose T: Vibrio vulnificus. Hazard on the half shell. West J Med, 1991; 155: 400–3

20 Hor LI, Chang TT, Wang ST: Survival of Vibrio vulnificus in whole blood from patients with chronic liver diseases: association with phagocytosis by neutrophils and serum ferritin levels. J Infect Dis, 1999; 179: 275–78

21 Johnston JM, Becker SF, McFarland LM: Gastroenteritis in patients with stool isolates of *Vibrio vulnificus*. Am J Med, 1986; 80: 336–38

22 Kiyosawa K, Uemura T, Ichijo T et al: Hepatocellular Carcinoma: Recent Trends in Japan. Gastroenterology, 2004; 127(5 Suppl.1): 17–26

23 Umemura T, Ichijo T, Yoshizawa K et al: Epidemiology of hepatocellular carcinoma in Japan. J Gastroenterol, 2009; 44(Suppl.19): 102–7

© Med Sci Monit, 2009; 15(9): CR453-459
PMID: 19721396

Received: 2009.03.14
Accepted: 2009.05.21
Published: 2009.09.01

High incidence of multiple primary carcinomas in HCV--infected patients with oral squamous cell carcinoma

Yumiko Nagao[1ABCDEFG], **Michio Sata**[1,2ADEG]

[1] Department of Digestive Disease Information & Research, Kurume University School of Medicine, Kurume, Japan
[2] Division of Gastroenterology, Department of Medicine, Kurume University School of Medicine, Kurume, Japan

Source of support: This study was supported in part by a Grant-in-Aid for Scientific Research (C) (No. 18592213) from the Ministry of Education, Culture, Sports, Science and Technology of Japan, and was supported in part by Health and Labour Sciences Research Grants for Research on Hepatitis from the Ministry of Health, Labour and Welfare of Japan

Authors' Contribution:
A Study Design
B Data Collection
C Statistical Analysis
D Data Interpretation
E Manuscript Preparation
F Literature Search
G Funds Collection

Summary

Background: Hepatitis C virus (HCV) infection has been associated with several extrahepatic manifestations. Oral cancer is one of them. We investigated the association among oral squamous cell carcinoma (OSCC), multiple primary cancers (MPCs), insulin resistance and HCV infection.

Material/Methods: Upper gastrointestinal tract examination and determination of the presence of HCV infection were routinely done for 60 primary OSCC patients. Occurrence of MPCs was evaluated between 1992 and 2008.

Results: Of the 60 patients, 21 (35%: 15 males and 6 females; mean age 67.3±11.9 years) developed MPCs. Antibodies to HCV were found in 26.7% (16/60) of cases. The incidence of MPCs in HCV-infected OSCC cases was 62.5% (10/16 cases, $P<0.01$ vs the non-HCV-infected OSCC group); for cases without HCV infection it was 25% (11/44 cases). In HCV-infected cases, 10 MPCs with patients, hepatocellular carcinoma (HCC) was the most common outcome (5 cases), whereas gastric cancer was the most common outcome (6 cases) in non-HCV-infected 11 MPCs. In logistic regression analysis, the adjusted odds ratios on staging IV, anti-HCV positive, and over 70 years old were 15.50, 13.45, and 4.46, respectively, indicating that there were significant differences. Furthermore, the patients with HCV-infected MPCs had hyperinsulinemia.

Conclusions: HCV infection was strongly associated with the occurrence of MPCs as well as primary OSCC. HCV-infected OSCC patients in Japan should receive medical treatment to inhibit development of HCC. In patients with HCV infection, it is important to clinically examine organs other than the liver.

key words: **multiple primary cancers (MPCs) • oral squamous cell carcinoma (OSCC) • hepatitis C virus (HCV) • hepatocellular carcinoma (HCC) • lichen planus • insulin resistance • extrahepatic manifestations**

Abbreviations: **anti-HCV** – anti-bodies to HCV; **anti-HBc** – antibody to hepatitis B core antigen; **CLEIA** – chemiluminescent enzyme immunoassay; **HBsAg** – hepatitis B surface antigen; **HCC** – hepatocellular carcinoma; **HCV** – hepatitis C virus; **HOMA-IR** – homeostasis model assessment; **IFN** – interferon; **MPCs** – multiple primary cancers; **OSCC** – oral squamous cell carcinoma

Full-text PDF: http://www.medscimonit.com/fulltxt.php?l,5CID=878175
Word count: 2612
Tables: 4
Figures: 1
References: 30

Author's address: Yumiko Nagao, Department of Digestive Disease Information & Research, Kurume University School of Medicine, 67 Asahi-machi, Kurume 830-0011, Japan, e-mail: nagao@med.kurume-u.ac.jp

BACKGROUND

The development of multiple primary cancers (MPCs) is frequently detected in patients with oral squamous cell carcinoma (OSCC). Patients with OSCC are at risk of developing second cancers or MPCs, particularly at sites within the upper digestive tract and airway [1,2]. Routine upper gastrointestinal panendoscopy identifies synchronous MPCs in 9–14% of patients [3].

In recent years in Japan, there has been an upward trend in MPCs in patients with head and neck cancer [4]. The reasons behind this are increases in carcinoma itself, progress in diagnostic techniques, improvements in treatment outcomes, and increased mean survival time.

Since 1981, malignant neoplasms have been the leading cause of death in Japan. During the past 20 years, primary liver cancer, 95% of which is hepatocellular carcinoma (HCC), has ranked third in men and fifth in women in Japan as the cause of death from malignant neoplasms [5]. The number of deaths from HCC is expected to increase by 2010–15 [6]. Of the HCC cases in Japan, ~16% are caused by hepatitis B virus (HBV) infection and~80% by hepatitis C virus (HCV) infection. The increase in incidence of HCC in Japan has largely been attributable to HCV infection. Geographically, HCC is more frequent in western than eastern Japan.

HCV infection has also been associated with extrahepatic manifestations and immune-mediated phenomena [7]. For example, HCV is associated with the development of OSCC. We reported for the first time an association between HCV and OSCC [8], and provided evidence, at the national level in Japan, for the high prevalence of HCV infection in patients with OSCC [9]. The subjects included 305 patients with OSCC and 276 patients with non-malignant disease (the control group) from five geographically-distinct institutions. The incidence of HCV infection in Japanese OSCC patients has been reported as 16.7–24.0% [8,9]. We also investigated the prevalence of HCV infection in oral cancer patients with MPCs [10]. Of 327 patients with OSCC, 59 (18.0%) exhibited MPCs. In the OSCC patients with MPCs, serum HCV antibodies (anti-HCV) and HCV RNA were detected in 36.7% and 28.6%, respectively [10].

Meanwhile, insulin resistance emerges as a very important host factor in patients with chronic hepatitis C. Hyperinsulinaemia is associated with accelerated HCC growth [11]. We concluded that HCV infection induces insulin resistance, which causes an increase in the incidence of extrahepatic manifestations such as lichen planus in HCV-infected individuals [12,13]. Lichen planus is an inflammatory disease of the skin and oral mucosa. The HCV infection rates in lichen planus patients are high especially in Japan [14]. Oral lichen planus should be considered as a precancerous lesion, particularly in patients presenting HCV infection [15]. Prevalence of smoking history, presence of hypertension, extrahepatic malignant tumor, and insulin resistance were significantly higher in 17 patients with lichen planus than in 70 patients without lichen planus [13].

In the current study, we surveyed the incidence of MPCs in OSCC patients with or without HCV infection and investigated the relationship between OSCC and insulin resistance.

MATERIAL AND METHODS

Subjects

This retrospective study included 60 primary OSCC patients who had visited our clinic at the Kurume University Hospital in Japan for the first time between November 1992 and December 1994. The 60 patients with OSCC included 39 males and 21 females. Their ages ranged from 32 to 85 years, with an average age of 64.8±13.7 years. These patients resided in the northern Kyushu region of Japan where the prevalence of HCV infection is the highest in the country [5,16]. The stages of OSCC were as follows; stage I (15 cases), II (24), III (6), and IV (15).

MPCs were identified according to the definition proposed by Warren and Gates: there must be histological evidence of malignancy in each tumor, they must be separated from each other by normal tissue, and one tumor must not be a metastasis of another [17]. Patients with multiple OSCCs were excluded from the study. MPCs detected <6 months after OSCC diagnosis were defined as synchronous; those detected >6 months after diagnosis were defined as metachronous [17].

Methods

Upper gastrointestinal tract examinations were routinely performed in all OSCC patients using an endoscope. This was done on the first visit or first day of medical treatment in order to confirm the presence of MPCs such as carcinomas of the larynx, pharynx, esophagus, and stomach regardless of whether symptoms were present.

Sera from all 60 OSCC patients were used for the following liver function tests at the time of the first visit to our hospital: serum aspartate aminotransferase (AST), alanine aminotransferase (ALT), gammaglutamyl transpeptidase (γ-GTP), lactate dehydrogenase (LDH), total protein (TP), and albumin (Alb). Sera were also examined for the presence or absence of HCV or HBV infection. Anti-HCV antibodies and hepatitis B virus surface antigen (HBsAg) were measured by a chemiluminescent enzyme immunoassay (CLEIA) kit and a chemiluminescent immunoassay (CLIA), respectively. In 59 of 60 patients, HCV RNA in serum was detected using the Amplicore HCV test. In 58 of 60 patients, antibody to hepatitis B core antigen (anti-HBc) was found using a CLEIA kit. Ultrasonographic examination for all subjects was performed in order to examine the shape of the liver and lesions occupying the liver. Computed tomography and liver biopsy were performed in some patients.

Plasma glucose levels were measured by a glucose oxidase method for all subjects and serum insulin levels were measured using a sandwich enzyme immuno assay kit (EIKEN CHEMICAL, Tokyo, Japan). Insulin resistance was calculated on the basis of fasting levels of plasma glucose and insulin, according to the homeostasis model assessment (HOMA-IR) method [18]. The formula for the HOMA-IR is: HOMA-IR = fasting glucose (mg/dL) × fasting insulin (μU/mL)/405.

Their district, a history of liver dysfunction, blood transfusion, alcohol consumption, and smoking at the time of the first medical examination were collected as background information; OSCC was based upon their medical record cards.

Table 1. Incidence difference of MPCs depend on the presence or absence of HCV infection.

		Anti-HCV negative n=44 (%)		Anti-HCV positive n=16 (%)		P value A versus B
Age	Mean (year) ±SD	64.3±14.5		66.1±11.0		NS
Sex	Male	30	(68.2)	9	(56.2)	NS
	Female	14	(31.8)	7	(43.8)	
MPCs	Number	11	(25.0)	10	(62.5)	p<0.01
	Primary oral SCC					
	Tongue	2	(18.2)	6	(60.0)	
	Gingiva	5	(45.5)	3	(30.0)	
	Buccal mucosa	2	(18.2)	0	(0.0)	
	Sinus	1	(9.1)	0	(0.0)	
	Oropharynx	1	(9.1)	1	(10.0)	
	Number of MPCs					
	Double		(81.8)	10	(100.0)	
	Triple		(9.1)	0	(0.0)	
	Quadruple		(9.1)	0	(0.0)	
	Organ of MPCs	Stomach	6	Liver	5	
		Esophagus	2	Colon	2	
		Skin	2	Lung	1	
		Thyroid	1	Throid	1	
		Pharynx	1	Bone marrow*	1	
		Kidney	1			
		Liver	1			
		Total	14	Total	10	
	Occurrence time					
	Synchronus		6		5	
	Metachronus		6**		5	

* Acute myeloid leukemia (AML); ** One patient with quadruple cancer had cancer of the gingiva-esophagus (synchronous)-skin (synchronous)-hypopharynx (metachronous). SD – standard deviation; NS – no significance.

We observed the occurrence of MPCs from the first medical examination day to the last check-up day or nearest day preceding October 17, 2008. MPCs were diagnosed based on histopathology by the pathology laboratory which collected samples from all other medical departments of our hospital; or the diagnosis was made at other medical institutions.

Furthermore, the 60 patients whom we followed were divided into four groups: (i) MPCs with HCV infection, (ii) MPCs without HCV infection, (iii) non-MPCs with HCV infection, (iv) non-MPCs without HCV infection. We examined insulin resistance in these four groups.

Statistical analysis

All data are expressed as mean ± standard error. Differences between two groups were analyzed using the Mann-Whitney U test and the Chi-square test. Differences were judged significant for $p<0.05$ (two-tailed). Adjusted odds ratios were calculated using logistic regression analysis. All statistical analyses were conducted using JMP Version 6 (SAS Institute, Cary, NC, USA). The level of statistical significance was defined as 0.05.

Results

Incidence of MPCs

The details of the 60 patients studied are shown in Table 1. The mean period of follow-up was 2914.6±1536.7 days. Of the 60 patients with OSCC, 21 (35%: 15 males and 6 females; mean age 67.3±11.9 years) developed MPCs. Among the 21 patients, there were a total of 24 affected organs. The affected organs were: 6 liver cases (25%), 6 stomach (25%), 2 esophagus (8.3%), 2 colon (8.3%), 2 thyroid (8.3%), 2

Table 2. Background factors of 60 patients in onset of OSCC.

		Total n=60 (%)	Group A MPCs n=21 (%)	Group B Non-MPCs n=39 (%)	P value A versus B
Age	Mean (year) ±SD	64.8 ± 13.7	67.3 ± 11.9	63.4 ± 14.4	NS
Age group	20-69 years old	35 (58.3)	10 (47.6)	25 (64.1)	NS
	70 years or older	25 (41.7)	11 (52.4)	14 (35.9)	
Sex	Male	39 (65.0)	15 (71.4)	24 (61.5)	NS
	Female	21 (35.0)	6 (28.6)	15 (38.5)	
Stage	I	15 (25.0)	4 (19.0)	11 (28.2)	NS
	II	24 (40.0)	6 (28.6)	18 (46.2)	
	III	6 (10.0)	2 (9.5)	4 (10.3)	
	IV	15 (25.0)	9 (42.9)	6 (15.4)	
Period of follow-up	Mean (days) ±SD	2914.6±1536.7	3512.3±1355.0	2675.5±1457.9	NS
History of liver dysfunction	Yes	16 (26.7)	10 (47.6)	6 (15.4)	p<0.01
	No	41 (68.3)	9 (42.9)	32 (82.1)	
	Unknown	3 (5.0)	2 (9.5)	1 (2.6)	
History of blood transfusion	Yes	7 (11.7)	5 (23.8)	2 (5.1)	p<0.05
	No	48 (80.0)	13 (61.9)	35 (89.7)	
	Unknown	5 (8.3)	3 (14.3)	2 (5.1)	
Alcohol consumption	Yes	29 (48.3)	11 (52.4)	18 (46.2)	NS
	No	29 (48.3)	10 (47.6)	19 (48.7)	
	Unknown	2 (3.3)	0 (0.0)	2 (5.1)	
Smoking history	Yes	24 (40.0)	10 (47.6)	14 (35.9)	NS
	No	34 (56.7)	11 (52.4)	23 (59.0)	
	Unknown	2 (3.3)	0 (0.0)	2 (5.1)	

OSCC – oral squamous cell carcinoma; MPCs – multiple primary cancers; SD – standard deviation, NS: no significance.

skin (8.3%), 1 pharynx (4.2%), 1 kidney (4.2%), 1 lung (4.2%), and 1 bone marrow (leukemia, 4.2%). Nineteen patients had second primary cancers: one patient had three, and one patient had four primary cancers.

Incidence of HCV infection

Anti-HCV were detected in sera from 16 of the 60 patients with oral cancer (26.7%). The diagnosis of liver disease following the development of primary OSCC included: asymptomatic HCV carrier 6.3% (1/16), past HCV infection 6.3% (1/16), chronic hepatitis C 25% (4/16), liver cirrhosis 37.5% (6/16), HCC with liver cirrhosis 18.8% (3/16), and HCC post interferon (IFN) treatment for chronic hepatitis C 6.3% (1/16). Just after we succeeded in eliminating HCV by IFN treatment, a 38-year-old man developed simultaneous HCC and OSCC. The incidence of MPCs in an HCV-infected OSCC or in a non-HCV-infected OSCC patient was 62.5% (10/16 cases, P<0.01 vs the non-HCV-infected OSCC group) and 25% (11/44), respectively. In 10 MPC patients who were HCV-infected, HCC was the most common carcinoma (5 cases); In 11 MPC patients who were not HCV-infected, gastric cancer was the most common (6 cases).

Risk factors by univariate analysis

We compared characteristics of 21 subjects who had MPCs (group A) and 70 subjects who did not have MPCs (group B). The average age in group A was 67.3±11.9 years; there were 15 males and 6 females. The average age in group B was 63.4±14.4 years; there were 24 males and 15 females. Table 2 shows clinical features of groups A and B. A history of liver dysfunction in group A was found in 10 (47.6%, p<0.01 vs group B); a history of blood transfusion in group A was found in 5 (23.8%, p<0.05 vs group B).

We analyzed for differences between these two groups in AST, ALT, ALP, γGTP, LDH, TP, Alb, insulin, blood glucose level, and HOMA-IR. The laboratory data of both groups are shown in Table 3. Prevalence of anti-HCV antibodies was significantly higher in group A than in group B (p<0.01).

Significant differences in the development of MPCs included a history of liver dysfunction, blood transfusion, and anti-HCV positivity.

Table 3. Laboratory data of 60 patients in onset of OSCC.

		Total n=60	Group A MPCs n=21	Group B Non-MPCs n=39	P value A versus B
AST (IU/L)	(Mean ± SD)	31.1±23.5	34.6±22.4	29.1±24.1	NS
ALT (IU/L)	(Mean ± SD)	19.5±18.5	22.7±15.4	17.7±19.9	NS
ALP (IU/L)	(Mean ± SD)	15.6±2.0	33.2±2.1	7.1±1.9	NS
γ-GTP (IU/L)	(Mean ± SD)	23.4±20.5	25.5±18.7	22.3±21.3	NS
LDH (IU/L)	(Mean ± SD)	337.1±66.8	351.1±56.5	330.4±70.8	NS
TP (g/dL)	(Mean ± SD)	7.6±0.5	7.7±0.5	7.6±0.5	NS
Alb (g/dL)	(Mean ± SD)	4.0±0.4	3.9±0.3	4.0±0.4	NS
Insulin (μU/L)	(Mean ± SD)	11.9±9.4	14.1±9.0	10.8±9.5	NS
Blood glucose level (mg/dL)	(Mean ± SD)	90.9±40.6	89.8±19.3	91.5±47.7	NS
HOMA-IR	(Mean ± SD)	3.0±3.7	3.3±2.3	2.9±4.2	NS
Anti-HCV	Positive	16 (26.7%)	10 (47.6%)	6 (15.4%)	p<0.01
	Negative	44 (73.3%)	11 (52.4%)	33 (84.6%)	
HCV RNA	Positive	13 (21.7%)	7 (33.3%)	6 (15.4%)	NS
	Negative	46 (76.7%)	13 (61.9%)	33 (84.6%)	
	Uncertain	1 (1.7%)	1 (4.8%)	0 (0.0%)	
HBsAg	Positive	1 (1.7%)	0 (0.0%)	1 (2.6%)	NS
	Negative	59 (98.3%)	21 (100.0%)	38 (97.4%)	
Anti-HBc	Positive	39 (65.0%)	14 (66.7%)	25 (64.1%)	NS
	Negative	19 (31.7%)	5 (23.8%)	14 (35.9%)	
	Uncertain	2 (3.3%)	2 (9.5%)	0 (0.0%)	

SD – standard deviation; NS – no significance; AST – serum aspartate aminotransferase; ALT – alanine aminotransferase; γ-GTP – gammaglutamyl transpeptidase; LDH – lactate dehydrogenase; TP – total protein; Alb – albumin; HOMA IR – homeostasis model assessment.

Multivariate analysis

According to multivariate analysis, three factors – stage IV, anti-HCV positivity, and over 70 years old – were identified as factors associated with OSCC patients having an increased chance of developing MPCs. The adjusted odds ratios for these three factors were 15.50, 13.45, and 4.46, respectively, and each was statistically significant (Table 4).

Table 4. Results of multivariate analysis.

	Adjusted odds ratio (95% confidence interval)		P value
Stage IV	15.50	(0.39–2.58)	P=0.0124
Anti-HCV positive	13.45	(0.50–2.30)	P=0.0039
70 years or older	4.46	(0.04–1.56)	P=0.0480

Insulin resistance for the four groups

Of the 60 subjects (16 anti-HCV antibody positive and 44 anti-HCV negative), 10 had MPCs with HCV infection (group 1), 11 had MPCs without HCV infection (group 2), 6 lacked MPCs but had HCV infection (group 3), and 33 lacked MPCs and HCV infection (group 4). Fasting insulin levels at the time of the first visit to our hospital were: 16.3±7.9, 12.1±9.5, 13.5±12.6, or 10.3±8.7, in groups 1, 2, 3 and 4, respectively. Fasting insulin levels for group 1 was significantly higher than for group 4 (p=0.01, Figure 1A). HOMA-IR values seven years prior in groups 1, 2, 3, and 4 were, respectively, 3.5±1.6, 3.0±2.7, 3.1±3.0, and 2.9±4.4. A HOMA-IR value for group 1 was significantly higher than for group 4 (p=0.01, Figure 1B).

Discussion

We have already reported a high incidence of HCV among patients with OSCC [8,9]. Furthermore, we investigated the characteristics and incidence of MPCs in patients with OSCC treated between 1974 and 1995, suggesting that HCV infection increases the risk of developing MPCs [10].

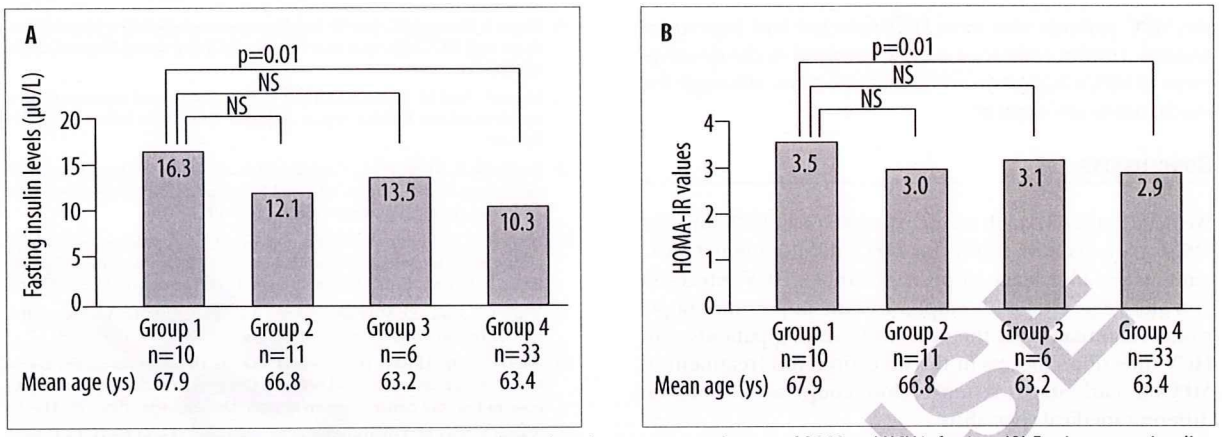

Figure 1. Association of carcinomas with insulin resistance depend on the presence or absence of OSCC and HCV infection. (**A**) Fasting serum insulin levels and (**B**) HOMA-IR values.

In the present study, the incidence of MPCs in patients with OSCC was 35% (21/60 patients) during 2914.6±1536.7 days of follow-up. The incidence of anti-HCV positivity was 26.7% (16/60 patients). The incidence of MPCs in an OSCC patient that was HCV-infected was significantly higher than in one that was not infected (62.5% vs 25%, $p<0.01$). HCC was the most common form of HCV-infected MPCs, and gastric cancer was the most common form of non-HCV-infected MPCs. These findings suggest a strong association between HCV infection and OSCC. The incidence of MPSc with the exclusion of 5 HCC in an OSCC patient that was HCV-infected was also higher than in one that was not infected (45.1% vs 25%). The affected carcinomas in extrahepatic organs of OSCC patients with HCV infection were: 2 colons, 1 lung, 1 thyroid, and 1 bone marrow (leukemia). Even excluding HCC, HCV-infected patients were at a high risk of developing extrahepatic MPCs.

Multivariate analysis demonstrated that stages of OSCC, being anti-HCV positive, and being over 70 years old increased the risk that patients with OSCC would develop MPCs. In OSCC patients who are HCV-infected, it is important to clinically examine the liver other than the oral cavity and gastrointestinal regions.

HCV infection induces not only chronic liver disease but also extrahepatic manifestations. Indeed, we experienced and reported five head and neck SCC among HCV-infected patients: (i) the patient who developed buccal mucosa cancer after IFN therapy for chronic hepatitis C [19], (ii) the patient who had simultaneous double primary cancers, including tongue cancer and HCV-related HCC [20], (iii) the patient who developed tongue cancer during the treatment of HCV-related liver disease [20], (iv) the patient with chronic hepatitis C, who developed worsening of lichen planus lesions during treatment with IFN plus ribavirin [21] and subsequently developed larynx cancer, and (v) the patient who developed tongue cancer during treatment for chronic hepatitis C [22].

It is presumed that between 1 and 2 million Japanese people are chronically infected with HCV. Because many such people are unaware that they are infected, carriers may develop liver cirrhosis and HCC, and this poses a serious problem. HCV-related HCC has increased and is now the cause of a majority of cases in Japan. Thus, the increased rates of death due to primary liver cancer in Japan appear to reflect the increase in numbers of HCV-related HCC [5]. IFN therapy, an antiviral agent, contributes to the prevention of occurrence of HCC and to improvement in long-term prognosis [23,24]. HCV-infected OSCC patients should also receive medical treatment to inhibit development of HCC, especially in Japan where the average life expectancy has increased year after year. In 2006, the life expectancies at birth were 79.0 years for males and 85.8 for females (Abridged Life Table, Ministry of Health, Labour and Welfare). Meanwhile, in patients with HCV infection, it is important to clinically examine organs other than the liver.

Satoh et al reported autopsy cases collected from the Annual of the Pathological Autopsy Cases in Japan, which is issued by the Japanese Society of Pathology for the past five years 1997–2001 [4]. A total of 134,997 cases had autopsies in Japan over five years. Of these, 321 were tongue cancer. The incidence of MPCs, affecting both the tongue and other organs, was reported to be 35.2% (113/321). In cases of double cancers including tongue cancer, commonly occurring cancers were reported to be lung, liver, esophagus, and thyroid. We think that there is a strong relation between OSCC and HCV infection, as can be seen from the fact that the second most common MPCs with tongue cancer, according to the results of autopsies, is liver cancer (reported by Satoh et al).

Several studies and our previous reports suggest that HCV infection antedates insulin resistance [25,26]. We showed molecular mechanisms for HCV core-induced insulin resistance [26]. Meanwhile, in a large population-based cohort study, Park et al. reported that among male cancer survivors, prediagnosis smoking, alcohol consumption, obesity, and insulin resistance (all risk factors for cancer development) affected cancer prognosis [27]. Previous studies in breast, prostate, and colorectal cancers demonstrated that insulin resistance can influence outcomes through systemic consequences of hyperinsulinemia [28–30]. Insulin receptors are overexpressed in those cancer tissues, so high insulin levels could promote the selective growth advantage of cancer cells [28–30]. We conclude that HCV infection induces insulin resistance and may cause lichen planus, a precancerous lesion [12,13]. In the present study,

the MPC patients who were HCV-infected had hyperinsulinemia. Insulin resistance may be involved in the development of MPCs in patients with HCV infection, although the mechanisms are unclear.

Conclusions

We demonstrated a high incidence of MPCs in HCV-infected OSCC patients. Risk factors for MPCs developing in OSCC patients are high stage of primary cancer, HCV infection, and older age. Our study emphasizes the importance of periodic examination of the oral cavity among patients with HCV infection. Success in the detection and treatment of MPCs at early stages requires close cooperation between different medical specialists.

References:

1. Day GL, Blot WJ: Second primary tumors in patients with oral cancer. Cancer, 1992; 70: 14–19
2. Crosher R, McIlroy R: The incidence of other primary tumours in patients with oral cancer in Scotland. Br J Oral Maxillofac Surg, 1998; 36: 58–62
3. Licciardello JT, Spitz MR, Hong WK: Multiple primary cancer in patients with cancer of the head and neck: second cancer of the head and neck, esophagus, and lung. Int J Radiat Oncol Biol Phys, 1989; 17: 467–76
4. Satoh M, Oikawa Y, Furuya I: A statistical study of autopsy cases of tongue cancer in Japan (Part VI) (in Japanese). Dent J Iwate Med Univ, 2005; 30: 53–64
5. Kiyosawa K, Uemura T, Ichijo T et al: Hepatocellular Carcinoma: Recent Trends in Japan. Gastroenterology, 2004; 127 (5 Suppl.1): 17–26
6. Shibuya K, Yano E: Regression analysis of trends in mortality from hepatocellular carcinoma in Japan, 1972–2001. Int J Epidemiol, 2005; 34: 397–402
7. Pawlotsky JM, Ben Yahia M, Andre C et al: Immunological disorders in C virus chronic active hepatitis: A prospective case-control study. Hepatology, 1994; 19: 841–48
8. Nagao Y, Sata M, Tanikawa K et al: High prevalence of hepatitis C virus antibody and RNA in patients with oral cancer. J Oral Pathol Med, 1995; 24: 354–60
9. Nagao Y, Sata M, Itoh K et al: High prevalence of hepatitis C virus antibody and RNA in patients with head and neck squamous cell carcinoma. Hepatol Res, 1997; 7: 206–12
10. Yoshida M, Nagao Y, Sata M et al: Multiple primary neoplasms and hepatitis C virus infection in oral cancer patients. Heparol Res, 1997; 9: 75–81
11. Saito K, Inoue S, Saito T et al: Augmentation effect of postprandial hyperinsulinaemia on growth of human hepatocellular carcinoma. Gut, 2002; 51: 100–4
12. Nagao Y, Kawaguchi T, Tanaka K et al: Extrahepatic manifestations and insulin resistance in an HCV hyperendemic area. Int J Mol Med, 2005; 16: 291–96
13. Nagao Y, Kawasaki K, Sata M: Insulin resistance and lichen planus in patients with HCV-infectious liver diseases. J Gastroenterol Hepatol, 2008; 23: 580–85
14. Nagao Y, Sata M, Tanikawa K et al: Lichen planus and hepatitis C virus in the northern Kyushu region of Japan. Eur J Clin Invest, 1995; 25: 910–14
15. Gandolfo S, Richiardi L, Carrozzo M et al: Risk of oral squamous cell carcinoma in 402 patients with oral lichen planus: a follow-up study in an Italian population. Oral Oncol, 2002; 40: 77–83
16. Hayashi J, Yoshimura E, Nabeshima A et al: Seroepidemiology of hepatitis C virus infection in hemodialysis patients and the general population in Fukuoka and Okinawa, Japan. J Gastroentol, 1994; 29: 276–81
17. Warren S, Gates O: Multiple primary malignant tumors. A survey of the literature and statistical study. Am J Cancer, 1932; 16: 1358–414
18. Matthews DR, Hosker JP, Rudenski AS et al: Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. Diabetologia, 1985; 28: 412–19
19. Nagao Y, Sata M, Fukuizumi K et al: Oral cancer and hepatitis C virus (HCV): can HCV alone cause oral cancer? – a case report. Kurume Med J, 1996; 43: 97–100
20. Nagao Y, Sata M, Noguchi S et al: Various extrahepatic manifestations caused by hepatitis C virus infection. Int J Mol Med, 1999; 4: 621–25
21. Nagao Y, Kawaguchi T, Ide T et al: Exacerbation of oral erosive lichen planus by combination of interferon and ribavirin therapy for chronic hepatitis C. Int J Mol Med, 2005; 15: 237–41
22. Nagao Y, Hiromatsu Y, Nakashima T, Sata M: Graves' ophthalmopathy and tongue cancer complicated by peg-interferon α-2b and ribavirin therapy for chronic hepatitis C: A case report and review of the literature. Molecular Medicine Reports, 2008; 1: 625–31
23. Yoshida H, Shiratori Y, Moriyama M et al: Interferon therapy reduces the risk for hepatocellular carcinoma: national surveillance program of cirrhotic and noncirrhotic patients with chronic hepatitis C in Japan. IHIT Study Group. Inhibition of Hepatocarcinogenesis by Interferon Therapy. Ann Intern Med, 1999; 131: 174–81
24. Yoshida H, Arakawa Y, Sata M et al: Interferon therapy prolonged life expectancy among chronic hepatitis C patients. Gastroenterology, 2002; 123: 483–91
25. Hui JM, Sud A, Farrell GC et al: Insulin resistance is associated with chronic hepatitis C and virus infection fibrosis progression. Gastroenterology, 2003; 125: 1695–704
26. Kawaguchi T, Yoshida T, Harada M et al: Hepatitis C virus down-regulates insulin receptor substrates 1 and 2 through up-regulation of suppressor of cytokine signaling 3. Am J Pathol, 2004; 165: 1499–508
27. Park SM, Lim MK, Shin SA, Yun YH: Impact of prediagnosis smoking, alcohol, obesity, and insulin resistance on survival in male cancer patients: National Health Insurance Corporation Study. J Clin Oncol, 2006; 24: 5017–24
28. Jee SH, Ohrr H, Sull JW et al: Fasting serum glucose level and cancer risk in Korean men and women. JAMA, 2005; 293: 194–202
29. Dawson SI: Long-term risk of malignant neoplasm associated with gestational glucose intolerance. Cancer, 2004; 100: 149–55
30. Hsing AW, Gao YT, Chua S Jr et al: Insulin resistance and prostate cancer risk. J Natl Cancer Inst, 2003; 95: 67–71

LETTERS

Endogenous non-retroviral RNA virus elements in mammalian genomes

Masayuki Horie[1]*, Tomoyuki Honda[1,2]*, Yoshiyuki Suzuki[3], Yuki Kobayashi[3], Takuji Daito[1], Tatsuo Oshida[4], Kazuyoshi Ikuta[1], Patric Jern[5], Takashi Gojobori[3], John M. Coffin[5] & Keizo Tomonaga[1,6]

Retroviruses are the only group of viruses known to have left a fossil record, in the form of endogenous proviruses, and approximately 8% of the human genome is made up of these elements[1,2]. Although many other viruses, including non-retroviral RNA viruses, are known to generate DNA forms of their own genomes during replication[3–5], none has been found as DNA in the germline of animals. Bornaviruses, a genus of non-segmented, negative-sense RNA virus, are unique among RNA viruses in that they establish persistent infection in the cell nucleus[6–8]. Here we show that elements homologous to the nucleoprotein (N) gene of bornavirus exist in the genomes of several mammalian species, including humans, non-human primates, rodents and elephants. These sequences have been designated endogenous Borna-like N (EBLN) elements. Some of the primate EBLNs contain an intact open reading frame (ORF) and are expressed as mRNA. Phylogenetic analyses showed that EBLNs seem to have been generated by different insertional events in each specific animal family. Furthermore, the EBLN of a ground squirrel was formed by a recent integration event, whereas those in primates must have been formed more than 40 million years ago. We also show that the *N* mRNA of a current mammalian bornavirus, Borna disease virus (BDV), can form EBLN-like elements in the genomes of persistently infected cultured cells. Our results provide the first evidence for endogenization of non-retroviral virus-derived elements in mammalian genomes and give novel insights not only into generation of endogenous elements, but also into a role of bornavirus as a source of genetic novelty in its host.

Bornaviruses are the only animal RNA viruses that achieve a highly cell-associated life cycle within the nuclear envelope[6–9], and can therefore provide not only new models of RNA virus replication, but also insight into dynamics of RNA molecules in eukaryote cells. In an effort to understand whether bornaviruses mimic host factors to maintain persistent infection in the nucleus, we searched human protein databases for sequences with similarity to BDV proteins. This search identified two hypothetical human proteins (GeneID LOC340900 and LOC55096), each of which has significant sequence similarity to BDV N (Fig. 1a and Supplementary Table 1). BDV N is a major structural protein, which tightly encapsidates the viral RNA to form the nucleocapsid. The LOC340900 sequence encodes a protein of comparable length (366 residues) to BDV N (370 residues), whereas LOC55096 seems to contain several frameshift mutations relative to BDV N, resulting in a shorter ORF length (Fig. 1a). Both LOC340900 and LOC55096 showed an overall 41% sequence identity and 58% similarity to BDV N and 72% identity to each other. The close relationship between BDV N and the homologous genes was further demonstrated by the alignment of transcription regulatory sequences on either side of BDV N (Fig. 1b). The S and T motifs in flanking sequences of both putative human proteins were well conserved with those of BDV (Fig. 1b). In addition, a poly-A sequence appears after the T1-like motif in the 3′ flanking region of LOC55096 (Fig. 1b). The homology of the human genes to BDV N was also confirmed by a permutation test (Supplementary Fig. 1). These findings indicated that both human genes may be endogenous elements related to BDV N gene, and therefore we designated them EBLNs (LOC340900, EBLN-1 and LOC55096, EBLN-2).

To investigate the presence of EBLN sequences in other animal species, we conducted tblastn searches using BDV N as a query in eukaryote and whole-genome shotgun databases at NCBI. Sequences with blast *E*-values of 10^{-10} or lower were identified as EBLNs. We found two additional human elements (EBLN-3 and -4) as well as a number of related sequences in various mammalian species, including marsupials (Supplementary Table 2). Orthologous genes to human EBLNs were identified in the genomes of non-human anthropoid primates, including chimpanzee, gorilla, orang-utan, and macaque (Supplementary Table 2). We also detected primate EBLNs in the genomes of the suborder Strepsirrhini, including the mouse lemur and Garnett's galago. Furthermore, two species of the Afrotheria, African elephant and cape hyrax, and four rodents were found to have EBLNs with *E*-values of less than 10^{-20} (Supplementary Table 2). An EBLN locus with a high level of similarity to BDV N was also identified in the thirteen-lined ground squirrel (TLS) genome (Supplementary Fig. 2a). Like the human EBLNs, the TLS EBLN contained a 3′ poly-A sequence, as well as S and T signal motifs, in its 3′ flanking region (Supplementary Fig. 2b). Almost all EBLN fragments, except for EBLN-1 and the TLS gene, contained several stop codons in the predicted coding sequences, or lacked the identifiable flanking sequences. In addition, we found that all anthropoid EBLNs, except for EBLN-4, are expressed as mRNAs in some human and monkey-derived cell lines (Supplementary Fig. 3). A previous study reported the interaction of human EBLN-2 with other cellular proteins, such as AP1S1, TUSC2/FUS1 and FANCC (ref. 10) (Supplementary Table 1), indicating that anthropoid EBLNs may encode functional proteins.

To investigate whether other mammalian species contain EBLN-related sequences in their genomes further, we conducted Southern blot hybridization under low-stringency conditions using human, murine and TLS EBLN as probes (Fig. 1c and d). Along with the clear signals in primate genomes, we detected reproducible faint positive bands in murine and shrew genomes when using a human EBLN probe (Fig. 1c, dots). The signals were also observed using a mouse

[1]Department of Virology, Research Institute for Microbial Diseases (BIKEN), Osaka University, Osaka 565-0871, Japan. [2]Japan Society for the Promotion of Science (JSPS), Chiyoda-ku, Tokyo 102-8472, Japan. [3]Center for Information Biology and DNA Data Bank of Japan, National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan. [4]Department of Life Science and Agriculture, Obihiro University of Agriculture and Veterinary Medicine, Obihiro, Hokkaido 080-8555, Japan. [5]Department of Molecular Biology and Microbiology, Tufts University School of Medicine, Boston, Massachusetts 02111, USA. [6]PRESTO, Japan Science and Technology Agency (JST), Chiyoda-ku, Tokyo 102-0075, Japan.
*These authors contributed equally to this work.

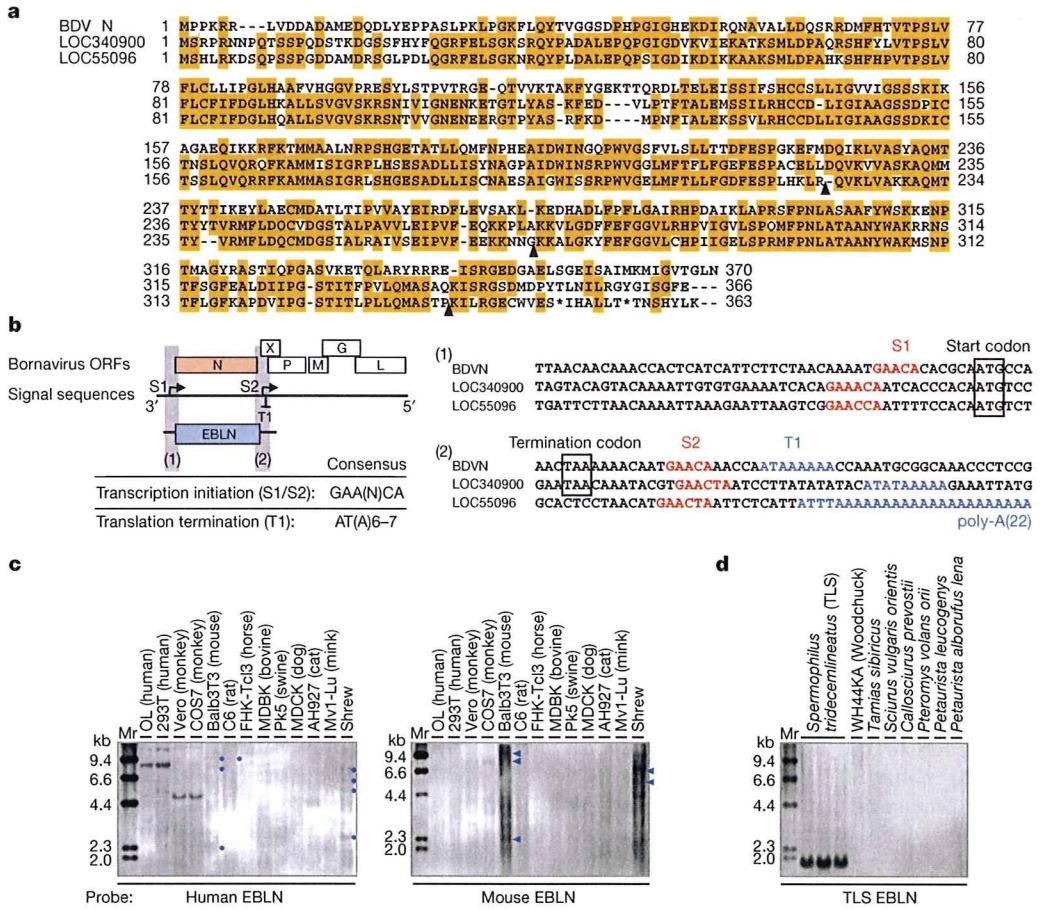

Figure 1 | Bornavirus N-like elements in mammalian genomes. **a**, Alignment between predicted amino acid sequences of BDV N and two human bornavirus N-like elements. Black arrowheads indicate predicted frameshift sites in LOC55096. **b**, Sequence alignments of transcription signal sites (S1/S2 and T1) at both the 5′ and 3′ ends of the bornavirus N ORF. A schematic representation of bornavirus genome structure is shown. **c**, **d**, Low-stringency Southern blot hybridizations of DNA from various mammalian species using human EBLN-1 and mouse EBLN chr.11 (**c**) and TLS EBLN (**d**) as probes. Dots and arrowheads on the right side of the murine and shrew lanes in panel **c** indicate the positions of reproducible positive signals. Mr, Molecular marker.

EBLN probe (Fig. 1c, arrowheads), indicating that the faint bands are most likely to be EBLN-related sequences. In fact, EBLN-like sequences, albeit with *E*-values greater than 10^{-10}, were found in the Eurasian shrew genome in our tblastn searches. On the other hand, except for TLS, no positive band was detected by the TLS probe in the genomes of several different squirrel species, such as woodchuck (*Marmota* spp.), the closest species to the TLS (*Spermophilus* spp.) (Fig. 1d)[11], indicating that the ground squirrels are likely to be the only host species of EBLN within the squirrel family. The BDV N probe detected many faint and smear bands that include the signals detected by EBLN-specific probes in both selected mammalian species and the squirrel families (Supplementary Fig. 4), indicating that EBLN-related fragments are more widely distributed in the mammalian genome.

We next performed a comprehensive phylogenetic analysis using nucleotide sequences of all EBLNs with *E*-values less than 10^{-20} (Fig. 2 and Supplementary Fig. 5). In addition to EBLNs, we included avian bornaviruses (ABVs)[12] and an exogenous reptile bornavirus (RBV) sequence, which was detected in a cDNA library from a *Bitis gabonica* (Gaboon viper) venom gland[13] (Supplementary Fig. 6). As shown in Fig. 2, the anthropoid and murine EBLNs are clustered phylogenetically within each host order. By contrast, EBLNs from other species, including African elephant, cape hyrax and guinea pig, form branches independent from the evolutionary lineage of their hosts, indicating that these EBLNs had most likely invaded each species via independent integration events. Interestingly, the TLS EBLNs form a tight cluster more closely related to modern exogenous bornaviruses than to those of other animals. Considering that a closely related species does not contain EBLNs, the integration of squirrel EBLN could have been a very recent event. A phylogenetic analysis using all primate EBLNs, including marmoset (Supplementary Fig. 7), showed that the integration events leading to the primate EBLNs occurred in the Haplorrhini at least before the split between rhesus macaque and marmoset.

To investigate whether current bornaviruses are able to be copied into DNA to produce EBLN-like elements, we first performed PCR analyses using DNA of persistently BDV-infected cells. As shown in Fig. 3a and Supplementary Table 3, BDV DNA was clearly detected in some cell lines by a primer set targeted to the BDV N region. To understand which viral RNA species serve as template for the DNA form of BDV, we used several primers within the BDV genome for amplification. The results showed that primer sets straddling the boundaries of BDV transcription units could not amplify BDV-specific DNA (Fig. 3b and c), indicating that the DNA is transcribed from mRNAs of BDV. We detected BDV-specific DNA in the brains of persistently BDV-infected mice (Supplementary Fig. 8), indicating that BDV can produce DNA forms *in vitro* and *in vivo*. We next performed Alu-PCR to investigate whether BDV DNA detected in the infected cells exists as integrated or extrachromosomal DNA. As shown in Fig. 3d and Supplementary Fig. 9, an Alu-specific PCR product was detected in BDV-infected cells only when using an N-specific forward primer about 30 days post-infection. This observation indicated that although BDV DNA in infected cells may be mainly extrachromosomal, the N gene is integrated into the host genome during persistent infection.

We further characterized the BDV DNA insertions and flanking cellular sequences by using Alu-PCR and inverse PCR (Supplementary Fig. 10)[14]. Integration sites were present on various chromosomes

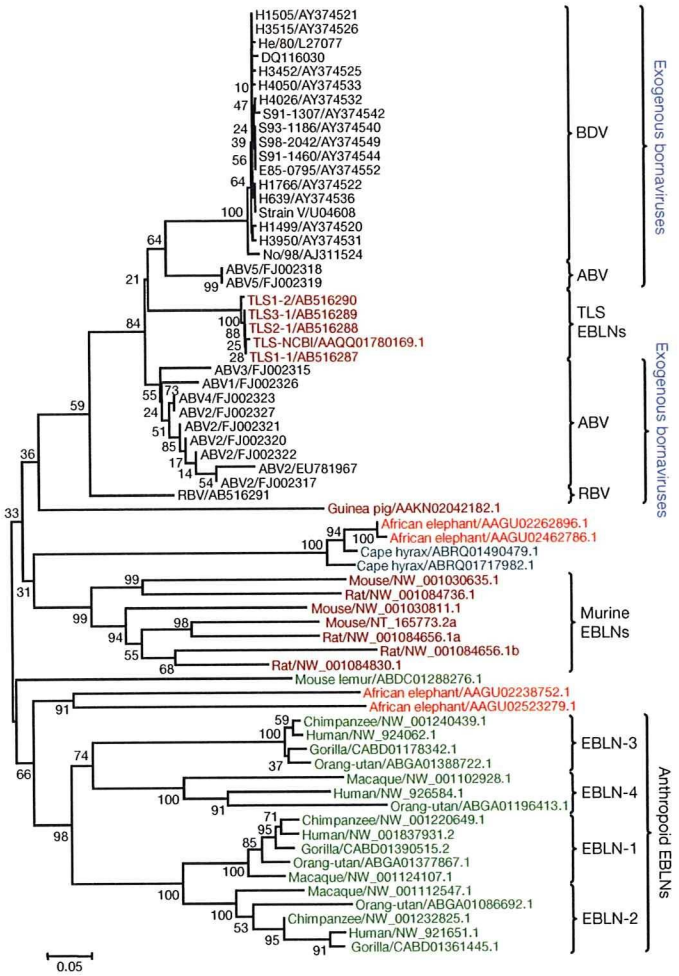

Figure 2 | Phylogenetic tree of exogenous bornaviruses and mammalian EBLNs. The bootstrap probability is indicated for each interior branch. The scale bar indicates the number of amino acid substitutions per site. Animals belonging to the same order are indicated by the same colour. Strain and sequence accession numbers are given for each sequence.

(Fig. 4). Similar to some mammalian EBLNs, many BDV DNA insertions contained a 3′ poly-A sequence (Fig. 4b and c). In addition, integrations of truncated BDV N DNA were also found in some clones. No apparent consensus sequences were found at the sites, although target site duplications (TSDs) were detected in some clones from the inverse PCR (Fig. 4c). We also found deletions, as well as sequence rearrangement, of host genome adjacent to BDV DNA insertions (Fig. 4c). These results indicate that modern BDV is able to produce DNA forms leading to insertion of EBLN-like elements into its host's genome.

This report is the first to provide evidence of endogenous sequences derived from a non-retroviral RNA virus in mammalian species. Phylogenetic analyses demonstrate that the oldest primate EBLN observed must have appeared in an ancestor of primates after the separation between Strepsirrhini and Haplorrhini, implying that bornaviruses have coexisted with primates for an evolutionary history stretching at least 40 million years. Thus, bornaviruses are the first non-retroviral RNA virus whose existence in prehistoric times has been confirmed. To date, the evolution/origin of RNA viruses is a major puzzle in the relationship between viruses and mammalian hosts, because simple molecular clock calculations using an average rate of nucleotide substitutions estimate the origin of RNA viruses to be a very recent event[15–17]. Despite replication during tens of millions of years as exogenous viruses, the amino acid sequences of current BDV N seem surprisingly conserved relative to EBLNs. This conservation demonstrates the inapplicability of simple molecular clocks to RNA virus evolution. Discovery of EBLNs in several

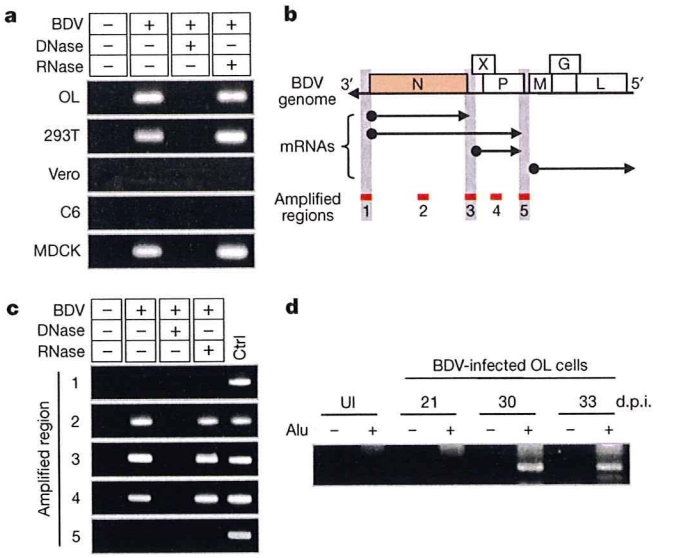

Figure 3 | Reverse transcription and integration of BDV RNA in mammalian cells. **a**, PCR amplification of BDV N-specific cDNA in BDV-infected cells. OL and 293T, human; Vero, monkey; C6, rat; MDCK, dog. **b**, Schematic representation of the bornavirus genome and mRNAs. Regions for the PCR amplification are indicated by red bars. **c**, Region-dependent amplification of BDV cDNA in infected OL cells. The numbers on the left side of the panels correspond to the amplification regions in panel **b**. Ctrl indicates the results of RT–PCR using RNA from BDV-infected OL cells. **d**, Integration of BDV DNA. Genomic DNA was isolated from BDV-infected OL cells at the indicated days after infection, and Alu-PCR was performed with (+) or without (–) the Alu primer. UI, genomic DNA from uninfected cells; d.p.i., days post-infection.

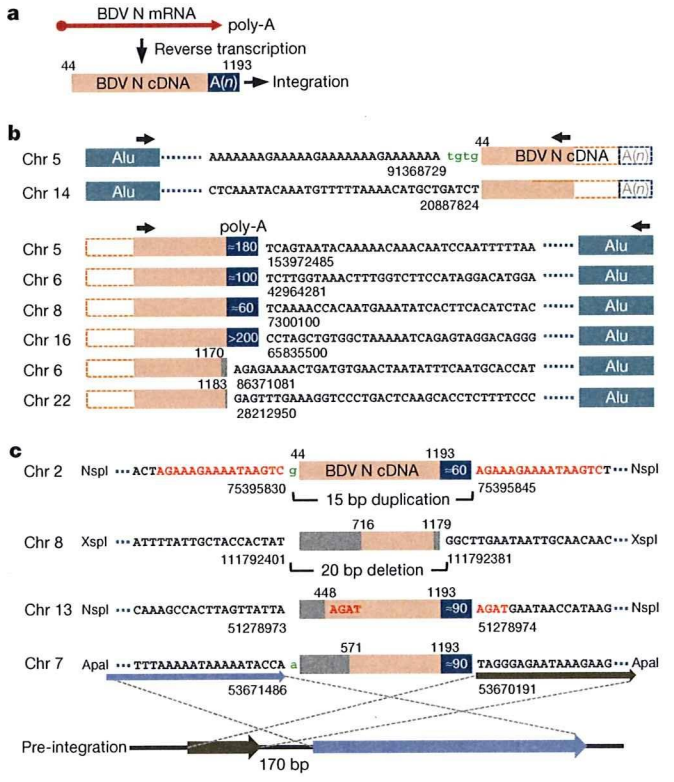

Figure 4 | Structures of BDV N integration events in OL cells. **a**, Structure of BDV N cDNA. The numbering corresponds to nucleotide positions in the BDV genome. The BDV N transcript runs from nucleotide positions 44 to 1193. **b, c**, Structures of BDV N integrations detected by Alu-PCR (**b**) and inverse PCR (**c**). Grey rectangles in the N cDNA indicate truncated regions. Black lettering, host genome sequences; green lettering, inserted nucleotides; red lettering, predicted TSDs. The blue box indicates the position and length of the poly-A sequence. The pre-integration form of chromosome 7 is shown in panel **c**.

mammalian species will help shed light on the evolutional history of RNA viruses and their hosts.

The sequence characteristics of both EBLNs and BDV DNA insertions in host genomes indicate that the reverse transcriptase activity encoded by retrotransposons, such as long interspersed nucleotide elements (LINEs), is likely to be involved in the reverse transcription and integration of bornavirus mRNAs, although some clones showed no apparent TSDs (ref. 18). LINE-1s (L1) are abundant retrotransposons, whose enzymes are able to sometimes target cellular mRNAs and produce processed pseudogenes in mammalian genomes[19–21]. The organization of sequences flanking EBLN-2 is consistent with the action of L1. The sequence shows the presence of an AluSx element immediately downstream of the 3′ poly-A tail of EBLN-2 (Supplementary Fig. 11). The key observation is that the EBLN-2/AluSx element is flanked by a perfect 9-bp TSD. Because the AluSx itself is not flanked by TSDs and the 3′ end of Alu is known to be recognized by L1 during target-primed reverse transcription, the presumed EBLN-2/AluSx chimaera element was most likely created and integrated by the L1 machinery. Thus, it is likely that EBLNs are processed pseudogenes derived from ancient bornavirus infections. At present, the reasons why bornaviruses but not other non-retroviral RNA viruses, and why only N and not other genes, have been preserved in mammalian genomes as endogenous elements are not clear. There are several possibilities. First, bornaviruses may have greater access to the germline. Second, the BDV N mRNA, like some cellular RNAs, may have features that, by chance, make it a favourable template for L1-mediated reverse transcription[22,23]. Third, the predominant transcription of BDV N mRNA in infected cells may also favour its association with the L1 replication machinery. The selectivity for BDV N mRNA implies a role for specific structural features, perhaps in conjunction with one or more of the other possibilities. Our data also raise the possibility that, like some endogenous retroviruses, EBLNs may have some function in their host species. An analysis of the non-synonymous to synonymous substitution ratios among anthropoid EBLNs indicates functional, albeit weak, evolutionary conservation. This finding implicates bornaviruses as a new source of genetic innovation in their hosts. Further studies will be needed to explore this possibility.

METHODS SUMMARY

Homology searches (blastp, tblastn) were conducted using the amino acid sequence of BDV N H1499 (International Nucleotide Sequence Database accession number AY374520) as a query and the genomic sequences of 234 eukaryotes as a database at the genomic blast server at the National Center for Biotechnology and Information, NCBI. Sequence hits with *E*-values less than 10^{-10} were collected together with neighbouring hits, if any, with higher *E*-values and combined according to their alignment pattern with BDV N. The resulting amino acid sequence was examined for the presence of a BDV_P40 domain (Pfam accession number PF06407.3) using HMMPFAM. The sequence was identified as a putative EBLN when the domain was detected with the *E*-value of less than 10^{-10}.

The putative EBLN amino acid sequences that were identified with *E*-value of less than 10^{-20} in both tblastn and HMMPFAM were used for the phylogenetic analysis with N sequences of various exogenous bornaviruses. The multiple alignments of EBLN and BDV N amino acid sequences were made according to the alignment pattern of EBLN sequences to BDV N in the tblastn results. The phylogenetic tree was constructed using the neighbour-joining method[24] and the evolutionary distance measured as the proportion of difference (*p* distance) with the pairwise deletion option in MEGA (version 4.0)[25]. The reliability of interior branches in the phylogenetic tree was assessed by the bootstrap method with 1,000 resamplings.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 2 September; accepted 17 November 2009.

1. Jern, P. & Coffin, J. M. Effects of retroviruses on host genome function. *Annu. Rev. Genet.* **42**, 709–732 (2008).
2. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
3. Zhdanov, V. M. Integration of viral genomes. *Nature* **256**, 471–473 (1975).
4. Klenerman, P., Hengartner, H. & Zinkernagel, R. M. A non-retroviral RNA virus persists in DNA form. *Nature* **390**, 298–301 (1997).
5. Geuking, M. B. *et al.* Recombination of retrotransposon and exogenous RNA virus results in nonretroviral cDNA integration. *Science* **323**, 393–396 (2009).
6. Tomonaga, K., Kobayashi, T. & Ikuta, K. Molecular and cellular biology of Borna disease virus infection. *Microbes Infect.* **4**, 491–500 (2002).
7. de la Torre, J. C. Molecular biology of Borna disease virus and persistence. *Front. Biosci.* **7**, d569–d579 (2002).
8. Lipkin, W. I. & Briese, T. in *Fields Virology* 5th edn (eds Knipe, D. M. & Howley, P. M.) 1829–1851 (Lippincott Williams & Wilkins, 2007).
9. Chase, G. *et al.* Borna disease virus matrix protein is an integral component of the viral ribonucleoprotein complex that does not interfere with polymerase activity. *J. Virol.* **81**, 743–749 (2007).
10. Ewing, R. M. *et al.* Large-scale mapping of human protein–protein interactions by mass spectrometry. *Mol. Syst. Biol.* **3**, 89 (2007).
11. Mercer, J. M. & Roth, V. L. The effects of Cenozoic global change on squirrel phylogeny. *Science* **299**, 1568–1572 (2003).
12. Kistler, A. L. *et al.* Recovery of divergent avian bornaviruses from cases of proventricular dilatation disease: identification of a candidate etiologic agent. *Virol. J.* **5**, 88 (2008).
13. Francischetti, I. M., My-Pham, V., Harrison, J., Garfield, M. K. & Ribeiro, J. M. Bitis gabonica (Gaboon viper) snake venom gland: toward a catalog for the full-length transcripts (cDNA) and proteins. *Gene* **337**, 55–69 (2004).
14. Hui, E. K., Wang, P. C. & Lo, S. J. Strategies for cloning unknown cellular flanking DNA sequences from foreign integrants. *Cell. Mol. Life Sci.* **54**, 1403–1411 (1998).
15. Holmes, E. C. Molecular clocks and the puzzle of RNA virus origins. *J. Virol.* **77**, 3893–3897 (2003).
16. Duffy, S., Shackelton, L. A. & Holmes, E. C. Rates of evolutionary change in viruses: patterns and determinants. *Nature Rev. Genet.* **9**, 267–276 (2008).
17. Korber, B., Theiler, J. & Wolinsky, S. Limitations of a molecular clock applied to considerations of the origin of HIV-1. *Science* **280**, 1868–1871 (1998).
18. Morrish, T. A. *et al.* DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nature Genet.* **31**, 159–165 (2002).
19. Maestre, J., Tchenio, T., Dhellin, O. & Heidmann, T. mRNA retroposition in human cells: processed pseudogene formation. *EMBO J.* **14**, 6333–6338 (1995).
20. Esnault, C., Maestre, J. & Heidmann, T. Human LINE retrotransposons generate processed pseudogenes. *Nature Genet.* **24**, 363–367 (2000).
21. Kazazian, H. H. Jr. Mobile elements: drivers of genome evolution. *Science* **303**, 1626–1632 (2004).
22. Zhang, Z., Carriero, N. & Gerstein, M. Comparative analysis of processed pseudogenes in the mouse and human genomes. *Trends Genet.* **20**, 62–67 (2004).
23. Pavlicek, A. & Jurka, J. in *Genomic disorders* (eds Lupski, J. R. & Stankiewicz, P.) 57–72 (Humana Press, 2006).
24. Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).
25. Tamura, K., Dudley, J., Nei, M. & Kumar, S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**, 1596–1599 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Kawahara for helping the capture of the wild shrews (*Sorex unguiculatus* and *Sorex gracillimus*) at Kiritappu wetland, Hokkaido, Japan. We thank I. Francischetti for provision of Gaboon viper (*Bitis gabonica*) venom gland tissue and a cDNA library, D. Vaughan for thirteen-lined ground squirrel (*Spermophilus tridecemlineatus*) brain and liver tissues, and K. Maeda, T. Miyazawa and N. Ohtaki for providing culture cell lines from several mammalian species. This work was supported by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) Grants-in-aid for Scientific Research on Priority Areas (Infection and Host Responses; Matrix of Infection Phenomena) (K.T.), PRESTO (RNA and Biofunctions) from Japan Science and Technology Agency (JST) (K.T.), a Health Labour Sciences Research Grants for Research on Measures for Intractable Diseases (H20 nanchi ippan 035) from the Ministry of Health, Labor and Welfare of Japan (K.T.), research grant R37 CA 089441 from the National Cancer Institute (J.M.C.) and a fellowship from the Wenner-Gren Foundation (P.J.). J.M.C. was a Research Professor of the American Cancer Society with support from the George Kirby Foundation.

Author Contributions K.T. designed research; M.H., T.H., T.D. and K.T. conducted experiments using virus and culture systems; T.O. collected samples; Y.S., Y.K. and T.G. performed phylogenetic analysis; M.H., T.H., Y.S., K.I., P.J., T.G., J.M.C. and K.T. analysed data; and M.H., Y.S., P.J., J.M.C. and K.T. wrote the manuscript. All authors discussed the results.

Author Information The TLS EBLN and RBV sequences reported here have been deposited in the DDBJ/EMBL/GenBank and the accession numbers are shown in Figure 2. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.T. (tomonaga@biken.osaka-u.ac.jp).

METHODS

Permutation test. A permutation test was conducted to examine the homology of human EBLNs to the N gene of BDV, taking into account their base composition. The nucleotide sequence of each EBLN was aligned with that of the BDV N gene (strain CRP3A: accession number AY114161) using CLUSTAL W. Gaps were eliminated from the alignment, and the proportion of identical sites (q) was computed. Nucleotide sequences of both the EBLN and the BDV N gene were randomly permuted using pseudorandom numbers, and the q value was computed as indicated above. The permutation process was repeated 10,000 times, and the distribution of the q value between two unrelated sequences of the same base composition as the original EBLN and the N gene was obtained. The probability (p) of observing the q value equal to or greater than the original value in the comparison of unrelated sequences was obtained from the distribution.

Tissue samples. Tissues from three weanling thirteen-lined ground squirrel (*Spermophilus tridecemlineatus*) born in May 2008 (four generations from wild stock) were provided from the Ground Squirrel Captive Breeding Colony at the University of Wisconsin Oshkosh, USA. Immediately after decapitation, brain and liver were rapidly dissected, cut into 5 mm cubes, immersed in chilled methanol, and stored frozen in liquid nitrogen until use. Shrew tissues (brain and liver) were isolated from wild-captured long-clawed shrews (*Sorex unguiculatus*) in Hokkaido, Japan. The shrews were captured under sampling permission of the government of Hokkaido. Immediately after capture, tissue samples were fixed in RNAlater (Ambion) and stored frozen until use. Gaboon viper (*Bitis gabonica*) venom gland tissue was obtained as frozen samples from the Laboratory of Malaria and Vector Research at National Institute of Allergy and Infectious Diseases, National Institutes of Health, USA. Ethanol-fixed tissues from Siberian flying squirrels (*Pteromys volans orii*) and Eurasian red squirrels (*Sciurus vulgaris orientis*) were obtained from the Department of Life Science and Agriculture, Obihiro University of Agriculture and Veterinary Medicine, Obihiro, Hokkaido, Japan.

DNA isolation. Total DNA from cultured cells was isolated using QIAamp DNA Blood Mini kit (Qiagen). Two monkey cell lines, Vero and COS7, used in this study are derived from African green monkey. High molecular mass DNA was extracted by using a Blood and Cell Culture DNA Mini kit (Qiagen). Genomic DNAs of shrews, ground squirrels and the Gaboon viper were prepared from tissue samples using a phenol/chloroform extraction method or the Blood and Cell Culture DNA Mini kit. To minimize the risks of contamination, DNA extraction was performed in UV-irradiated safety cabinet with UV-irradiated pipettes, tubes and filter tips.

DNA samples. Genomic DNAs from chipmunks (*Tamias sibiricus*), Japanese giant flying squirrels (*Petaurista leucogenys*) and red and white giant flying squirrels (*Petaurista alborufus lena*) were obtained from the Department of Life Science and Agriculture, Obihiro University of Agriculture and Veterinary Medicine, Obihiro, Hokkaido, Japan.

Southern blot hybridization. Genomic DNA (5 μg) was digested with appropriate restriction endonucleases (TaKaRa). After electrophoresis in a 0.9% agarose gel, DNA was transferred onto positively charged Nylon membranes (Roche) and baked at 120 °C for 30 min. The membrane was prehybridized in DIG Easy Hyb (Roche) at 32 °C for 30 min. Human and TLS EBLN and BDV N probes were labelled by DIG-High Prime (Roche). Hybridization was performed in DIG Easy Hyb containing 25 ng ml^{-1} probe at 32 °C overnight. The membrane was washed twice with 2× SSC, 0.1% SDS at room temperature for 5 min, and then washed twice with 0.5× SSC, 0.1% SDS at 50 °C for 15 min. For chemiluminescence detection, Anti-DIG-alkaline phosphatase, Fab (Roche) and CDP-Star (Roche) were used according to the manufacturer's instructions. The low-stringency condition can theoretically detect sequences having at least 75% identity with each probe.

F-PERT assay. F-PERT (fluorescent product-enhanced reverse transcriptase) assay was performed as described previously[26]. Briefly, cells were lysed in disruption buffer (40 mM Tris-HCl, pH 8.1; 50 mM KCl; 20 mM dithiothreitol; 0.2% NP-40) and the protein concentration was measured. For the reverse transcription reaction, 1 μg of the cellular protein in 10 μl disruption buffer and an equal volume of 2× RT mix (100 mM KCl; 20 mM Tris-HCl pH 8.3; 11 mM $MgCl_2$; 1 mM dATP, dCTP, dGTP and dTTP; 0.4 μM reverse primer: 5′-CACAGGTCAAACCTCCTAG GAATG-3′, 0.2% NP-40; 20 mM dithiothreitol; 0.8 U μl^{-1} RNasin (Promega); 314 ng μl^{-1} calf thymus DNA (Sigma) and 1.5 ng MS2 RNA (Roche)) were mixed and incubated at 48 °C for 30 min. cDNA was mixed with forward primer: 5′-TCCTGCTCAACTTCCTGTCGAG-3′, reverse primer, probe: 5′-(FAM)-TC TTTAGCGAGACGCTACCATGGCTA-(TAMRA)-3′ and 2× TaqMan Universal PCR Master Mix (Applied Biosystems). Real-time PCR was carried out in an ABI 7900HT Fast Real-Time PCR System using the following parameters: 95 °C 10 min, then 50 cycles consisting of 94 °C for 30 s and 64 °C for 1 min. SuperScript III reverse transcriptase (Invitrogen) was used as standard control.

Virus infection. The BDV strains, huP2br, He/80 and recombinant BDV expressing GFP (rBDV-5′ GFP), were used in this study. Virus stock was prepared from the supernatants of BDV-infected cells. Confluent BDV-infected cells were washed with 20 mM HEPES, pH 7.5 and incubated with 5 ml of 20 mM HEPES (pH 7.5) containing 250 mM $MgCl_2$ and 1% FCS for 1.5 h at 37 °C. Supernatants were harvested and centrifuged at 2,500g for 5 min. The resulting supernatants were used for virus stock. The infectious titre was determined by focus forming assay as described previously[27]. The cell lines used in this study were cultured in Dulbecco's modified Eagle's medium (DMEM)-containing 10% fetal bovine serum (FBS). Newborn Balb/c mice (Oriental kobo) were inoculated intracranially with 200 focus forming units of BDV stock per animal within 24 h after birth. Infected animals were sacrificed at 21 days post-infection. The brains were collected for further analyses. All animal experiments conformed to the guide for the care and use of laboratory animals in the Research Institute for Microbial Diseases, Osaka University, Japan.

Alu-PCR analysis. Integration of BDV sequences into host genomes was detected by using primers specific to human Alu repeats and to BDV N region. First round amplification was performed in a final volume of 25 μl containing 0.5 U Ex Taq (TAKARA), 1× Ex Taq buffer, 0.2 mM dNTP, BDV N-specific primer, Alu primer and 100 ng of high molecular mass genome DNA. As control, PCR without the Alu primer was also performed. The condition of first PCR was as follows: denature for 5 min, 20 cycles of 94 °C for 30 s, 53 °C for 30 s, 72 °C for 4 min, followed by an extended elongation at 72 °C for 10 min. The second round PCR reaction was carried out with 1 μl of the first reaction using BDV N-specific nested primers. The reaction was run as follows: denature for 5 min, 40 cycles of 94 °C for 30 s, 60 °C for 30 s, 72 °C for 20 s with the final extension at 72 °C for 3 min. The sequence information for primers used in Alu-PCR is available on request.

Amplification of virus–host junction. Virus–host junctions were amplified by using Alu-PCR and inverse PCR methods. Alu-PCR analysis was performed as described previously[28]. Briefly, the first round PCR reaction was carried out with 100 ng of high molecular mass genome DNA in a final volume of 25 μl containing 0.5 U Ex Taq, 0.2 mM dNTP, 2 μM BDV-specific primer and 0.2 μM Alu primer under the following conditions: denaturing at 94 °C for 1 min, 10 cycles of 94 °C for 30 s, 59 °C for 30 s, 70 °C for 3 min, followed by an extended elongation at 70 °C for 10 min. After amplification, 0.5 U of uracil DNA glycosylase (New England Biolabs) was added into the tubes and incubated at 37 °C for 30 min. After heating at 94 °C for 10 min to break DNA strands at apurinic dUTP sites, the next amplification primers, Tag- and BDV-specific primers, were added. Second round PCR was performed as follows: after denaturing at 94 °C for 2 min, 20 cycles of touchdown PCR in which the annealing temperature was decreased one degree every other cycle from 65 °C to 56 °C. The remaining 20 cycles were run with the annealing temperature at 55 °C, followed by an extended elongation at 72 °C for 3 min. One microlitre of the second round PCR products was further amplified with Tag- and BDV-specific primers as follows: after denaturing for 2 min, 25 cycles of 94 °C for 30 s, 60 °C for 30 s, 72 °C for 3 min with the final extension at 72 °C for 3 min. Amplified DNA was electrophoresed, extracted and then sequenced.

Inverse PCR was described elsewhere[29]. Briefly, 1 μg genomic DNA was digested with an appropriate restriction enzyme, including ApaI, BamHI, EcoRI, NspI, PstI or XspI, for 3 h. Digested DNA was purified with QIAquick PCR Purification kit (Qiagen) and diluted with T4 DNA ligase buffer to a final DNA concentration of 1 ng μl^{-1}, and then T4 DNA ligase (New England Biolabs) was added to a final concentration of 4 U μl^{-1}. After ligation at 16 °C for 16 h, ligated DNA was isolated using a QIAquick PCR Purification kit. Five microlitres of the eluate were used for nested PCR. First round PCR was conducted in a 50 μl final volume containing 1 U TaKaRa Ex Taq, 0.2 mM dNTP and 0.2 μM BDV-specific primer set with the following program: after denaturing at 94 °C for 2 min, 20 cycles of 94 °C for 30 s, 70 °C for 30 s (temperature was decreased one degree every other cycle), 72 °C for 4 min and 20 cycles of 94 °C for 30 s, 60 °C for 30 s, 72 °C for 4 min with the final extension at 72 °C for 3 min. Second round PCR was performed with 1 μl of the first reaction. The reaction condition was 94 °C for 2 min, 25 cycles of 94 °C for 30 s, 58 °C for 30 s, 72 °C for 4 min with the final extension at 72 °C for 3 min. PCR products were electrophoresed and DNA was extracted from the desired bands and sequenced. Sequence information for primers used in this study is available on request.

26. Lovatt, A. *et al.* High throughput detection of retrovirus-associated reverse transcriptase using an improved fluorescent product enhanced reverse transcriptase assay and its comparison to conventional detection methods. *J. Virol. Methods* **82**, 185–200 (1999).
27. Ohtaki, N. *et al.* Downregulation of an astrocyte-derived inflammatory protein, S100B, reduces vascular inflammatory responses in brains persistently infected with Borna disease virus. *J. Virol.* **81**, 5940–5948 (2007).
28. Minami, M., Poussin, K., Brechot, C. & Paterlini, P. A novel PCR technique using Alu-specific primers to identify unknown flanking sequences from the human genome. *Genomics* **29**, 403–408 (1995).
29. Wo, Y. Y., Peng, S. H. & Pan, F. M. Enrichment of circularized target DNA by inverse polymerase chain reaction. *Anal. Biochem.* **358**, 149–151 (2006).

Published online 22 October 2009 *Nucleic Acids Research, 2010, Vol. 38, Database issue* ***D33–D38***
doi:10.1093/nar/gkp847

DDBJ launches a new archive database with analytical tools for next-generation sequence data

Eli Kaminuma, Jun Mashima, Yuichi Kodama, Takashi Gojobori, Osamu Ogasawara, Kousaku Okubo, Toshihisa Takagi and Yasukazu Nakamura*

Center for Information Biology and DNA Data Bank of Japan, National Institute of Genetics, Research Organization for Information and Systems, Yata, Mishima 411-8510, Japan

Received September 15, 2009; Accepted September 22, 2009

ABSTRACT

The DNA Data Bank of Japan (DDBJ) (http://www.ddbj.nig.ac.jp) has collected and released 1 701 110 entries/1 116 138 614 bases between July 2008 and June 2009. A few highlighted data releases from DDBJ were the complete genome sequence of an endosymbiont within protist cells in the termite gut and Cap Analysis Gene Expression tags for human and mouse deposited from the Functional Annotation of the Mammalian cDNA consortium. In this period, we started a novel user announcement service using Really Simple Syndication (RSS) to deliver a list of data released from DDBJ on a daily basis. Comprehensive visualization of a DDBJ release data was attempted by using a word cloud program. Moreover, a new archive for sequencing data from next-generation sequencers, the 'DDBJ Read Archive' (DRA), was launched. Concurrently, for read data registered in DRA, a semi-automatic annotation tool called the 'DDBJ Read Annotation Pipeline' was released as a preliminary step. The pipeline consists of two parts: basic analysis for reference genome mapping and *de novo* assembly and high-level analysis of structural and functional annotations. These new services will aid users' research and provide easier access to DDBJ databases.

INTRODUCTION

The DNA Data Bank of Japan (DDBJ) is one of three databanks that constitute the DDBJ/EMBL-Bank/GenBank International Nucleotide Sequence Database (INSD), which was established through close collaboration with the European Bioinformatics Institute (EBI) in Europe and the National Center for Biotechnology Information (NCBI) in the USA. DDBJ is administered by the Center for Information Biology and DDBJ (CIB-DDBJ) of the National Institute of Genetics (http://www.nig.ac.jp/index-e.html) with funding endorsement from the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT). All researchers can submit their data to one of the three summit databanks to register it with INSD. The data that are enrolled are exchanged on every day, so that the three collaborating databanks share virtually the same data at any given time. The syntax for the INSD entries is discussed among the three databanks at an INSD collaborative meeting held once every year. The agreed rules are reflected in feature tables that define the common syntax (http://www.ddbj.nig.ac.jp/FT/full_index.html).

In the last year, we started novel web services that focus on daily announcements using Really Simple Syndication (RSS) technology and visualization of DDBJ content with high readability. Furthermore, a new data archive database for massive amounts of raw sequencing reads from next-generation sequencers was officially launched. The expert annotators of the DDBJ Read Archive (DRA) issue original accession numbers for submitted data. Concurrently, there was a preliminary release of a raw read annotation pipeline tool. This analytical pipeline tool supports reference genome mapping, *de novo* assembly and further annotation analyses, such as single nucleotide polymorphism (SNP) detection. The following sections describe three major advancements of the DDBJ databases, the novel announcement web services and the new archiving database with analytical tools for raw sequencing reads.

DEVELOPMENT OF DDBJ DATABASES

We have introduced newly released DDBJ databases, databases within the framework of INSD and other individual databases that have been appended from last year's report (1).

*To whom correspondence should be addressed. Tel: +81 55 981 6859; Fax: +81 55 981 6889; Email: yanakamu@genes.nig.ac.jp

Datasets contributing to INSD through DDBJ

In the period from July 2008 to June 2009, DDBJ collected and released original data on 1 701 110 entries/ 1 116 138 614 bases. More than 90% of the data came from Japanese researchers and the Japan Patent Office (JPO) and the rest was mainly from researchers in China, Korea and Taiwan. We call this dataset the 'INSD-core data'. It consists of INSD data in traditional format and includes general sequence data, complete genomes, expressed sequence tags (ESTs), etc., but excludes whole-genome shotgun (WGS), mass sequence for genome annotation (MGA) and third party annotation (TPA). Large sets of contigs (i.e. overlapping reads) and finished sequences without annotation from an ongoing genome project can be submitted to INSD as WGS data. DDBJ has released one WGS entry (2 878 428 bp) on *Staphylococcus aureus* ssp. *aureus* Mu50-omega and 23 675 009 MGA entries (80 069 915 counts). All of these INSD-core, WGS and MGA data were collected, reviewed and accessioned by DDBJ. Another portion of the INSD-core data contains the complete genome sequence of an endosymbiont within protist cells in the gut of the termite (Candidatus *Azobacteroides pseudotrichonymphae* genomovar. CFP2) submitted by Institute of Physical and Chemical Research (RIKEN) and National Institute of Genetics; full-length cDNA (HTC) and EST entries for the tomato (*Solanum lycopersicum*) submitted by Kazusa DNA Research Institute; Genome Survey Sequences entries for the rat (*Rattus norvegicus* LE/Stm) submitted by Kyoto University; EST entries for rhizomes of Chinese liquorice (*Glycyrrhiza uralensis*) submitted by Chiba University, MGA entries for the human and mouse submitted by RIKEN Omics Science Center; and MGA entries for small RNAs of the silkworm (*Bombyx mori*) submitted by the University of Tokyo. These data can be obtained at the DDBJ ftp site (http://www.ddbj.nig.ac.jp/ftp_soap-e.html). The reader may find it worthwhile to refer to the two sets of data on the complete genome sequence of an endosymbiont within protist cells in the termite gut and the MGA datasets used in Functional annotation of the mammalian cDNA (FANTOM; http://fantom.gsc.riken.jp/). This bacterial endosymbiont is widely known as a model organism for the study of cellulolysis. With regard to the endosymbiont, functional annotation of the bacterial genome has revealed that nitrogen fixation and cellulolysis are coupled within the protist's cells (2). An MGA dataset from the FANTOM consortium identified a large-scale gene network that controls the differentiation of the human myeloid leukaemia cell line THP-1 from monoblast to monocyte by applying next-generation sequencing technology and the Cap Analysis Gene Expression (CAGE) method (3).

Datasets released from DDBJ

In Table 1, we summarize numbers of published records collected and released from DDBJ. A primary database is a database as originally constructed and a secondary database is based on a primary database. An MGA is defined as a sequence that is produced in large quantity for the purpose of genome annotation, such as CAGE and 5'SAGE. A TPA (4) is a nucleotide sequence data collection in which each entry is obtained by assembling primary entries publicized from INSD and/or the Trace Archive with additional feature annotations determined by experimental or inferential methods by the TPA submitter. The DDBJ Trace Archive (DTA) is a permanent repository of DNA sequence chromatograms (traces), base calls and quality estimates for single-pass reads from various large-scale sequencing projects. The DTA has operated since 2008. In 2009, a simple metadata search system and a viewer of trace data for DDBJ-accepted data were added. Gene Trek in Prokaryote Space (GTPS) (5) is a database of prokaryotic genome data that have been reannotated by analyzing the original data in various ways. Genome Information Broker (GIB) (6) is a comprehensive data repository of complete microbial genomes in the public domain. GIB distributes genome sequence data and annotation 1 day after the data are submitted to INSD. The DDBJ Amino Acid Sequence Database (DAD) is produced by extracting all translated sequences from the DDBJ periodical release, including all INSD (DDBJ/EMBL-Bank/ GenBank) entries. We also support two other databases by providing maintenance service: Center for Information Biology gene Expression database (CIBEX) (7) is a public database for microarray data and stores MIAME-compliant data in accordance with MGED Society recommendations; Genomes TO Protein structures and function (GTOP) (8) is a database consisting of data

Table 1. Datasets released from DDBJ

Type	Database name	No. of records	Released date
Primary DB	INSD-core (processed by DDBJ)	17 440 910 entries (1 701 110 entries)	29 May 2009
	WGS	1 246 513 entries	10 September 2009
	MGA	34 740 058 entries	1 June 2009
	TPA	593 entries	10 September 2009
	DTA	2 submissions	7 July 2008
	DRA	12 submissions	11 September 2009
Secondary DB	DAD	14 710 673 entries	29 May 2009
	GTPS	690 genomes	25 May 2009
	GIB	982 genomes	10 September 2009

The number of records represents only published data.

analyses of proteins identified by various genome projects. The GTOP database mainly uses sequence homology analyses and features extensive use of information on 3D structures.

DAILY RELEASE ANNOUNCEMENT AND COMPREHENSIVE VISUALIZATION OF DDBJ DATABASES

To deliver up-to-date information from DDBJ to researchers every day, we started the daily publication of newly released data from DDBJ by implementing the following two new functions into the DDBJ web services. The first function is the announcement of RSS feeds of contents of data released from DDBJ databases each day (Figure 1). The second function is the visualization of DDBJ entries as word cloud figures. The following sections explain these in detail.

Frequent announcement of daily data releases by RSS

The first new function is the RSS publication of daily data releases by DDBJ. The RSS is a family of web-feed formats used to publish frequently updated items such as blog entries and news headlines (9). RSS feeds are also used by biological databases such as PubMed Central (http://www.pubmedcentral.nih.gov/) and ArrayExpress (10). A list of new enhancements in FLATFILE/WGS/CON/TPA is generated as RSS feeds every day. The contents of the RSS feeds are generated in connection with the respective VERSION, ACCESSION ID, DEFINITION if these are present in REFERENCE tags. The unit of published content is set by PROJECT of DBLINK; however, if there are no XML tags in PROJECT, the TITLE of the REFERENCE tag and the AUTHOR are substituted for the PROJECT.

Comprehensive visualization of DDBJ entries by word cloud images

In addition to daily publication of database updates, information on classified statistics in DDBJ databases such as species and features is worthwhile for users. DDBJ already provides several statistics on its web site, such as the gross numbers of registered entries and of bases in registered databases, with numerical values and graphs. However, with conventional media it is difficult to provide an overview of the features of DDBJ databases at a glance. Therefore, we apply the word cloud image program Wordle (http://www.wordle.net/) to statistics on the frequency of DDBJ database. This program generates a word cloud image based on the frequency of keywords appearing in a text document or webpage.

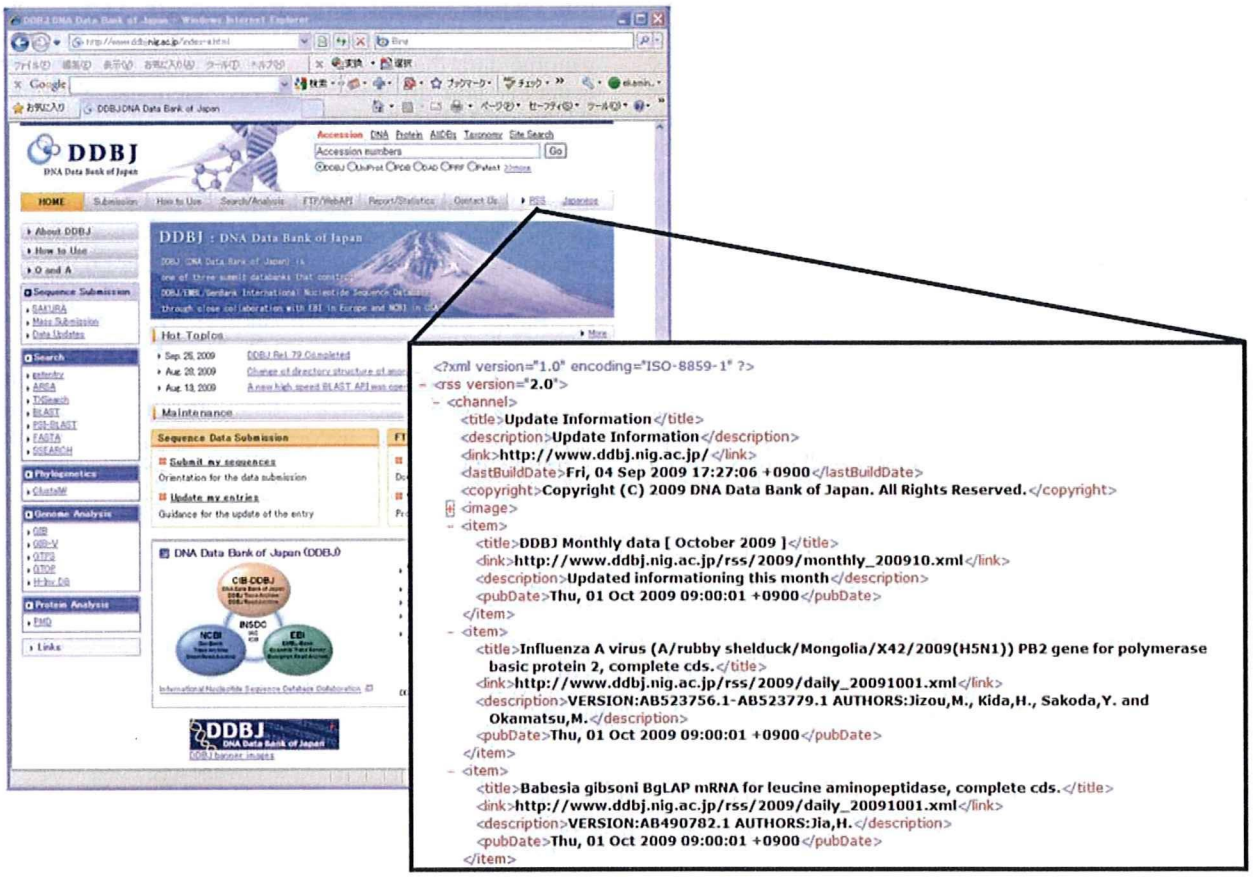

Figure 1. A feed file for RSS 2.0 is published from the DDBJ homepage every day (http://www.ddbj.nig.ac.jp/rss/update_information.xml). Daily released contents of DDBJ databases can be confirmed via RSS reader programs.