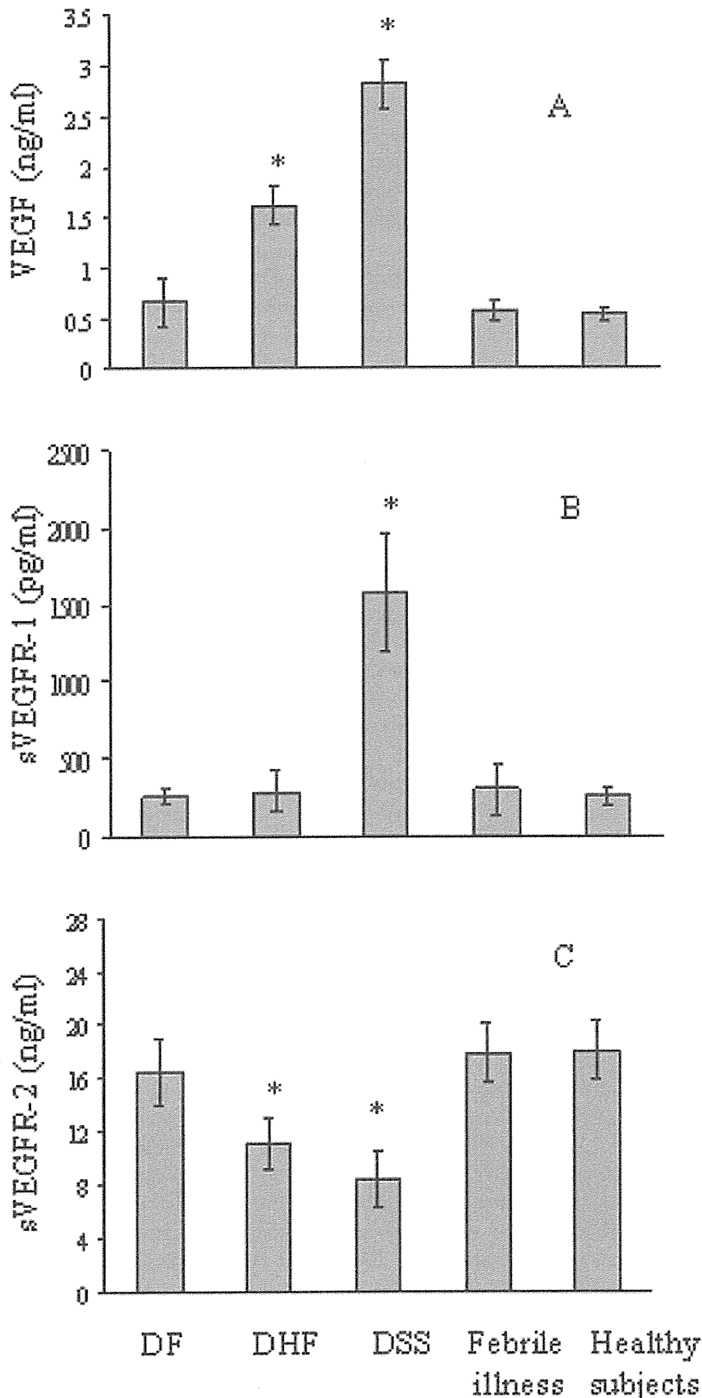

Figure 1. Plasma levels of VEGF, and sVEGFR-1 and -2 in Dengue patients and control groups. Plasma levels of VEGF (A), and sVEGFR-1 (B) and -2 (C) were examined for Dengue patients (DF, DHF, and DSS) on admission and for control subjects (febrile illness and healthy subjects) by ELISA. VEGF *$p<0.01$ (DHF and DSS versus DF, febrile illness or healthy subjects), sVEGFR-1 *$p<0.01$ (DSS versus DF, DHF, febrile illness or healthy subjects), sVEGFR-2 *$p<0.01$ (DHF and DSS versus DF, febrile illness or healthy subjects). Representative results from three independent experiments are shown.
doi:10.1371/journal.pntd.0001505.g001

level was observed when Dengue virus was inoculated with normal human serum (1:1,000 final dilution) or when UV-inactivated Dengue virus was inoculated with human Dengue virus immune or normal human serum. In addition, no VEGF production by KU812 and HMC-1 cells was observed after mock-infection with human Dengue immune or normal human serum. These results suggested the importance of antibody to Dengue virus for mast cell secretion of VEGF *in vitro*.

As it is known that KU812 and HMC-1 cells are permissive to Dengue virus infection when the virus is inoculated together with human Dengue immune serum [20], the antibody-dependent infection of KU812 cells with Dengue virus was examined by immunofluorescence analysis in the presence and absence of human Dengue immune serum 24 h after the inoculation. Positive immunofluorescence was only observed in cells infected in the presence of human Dengue virus-immune serum, suggesting the

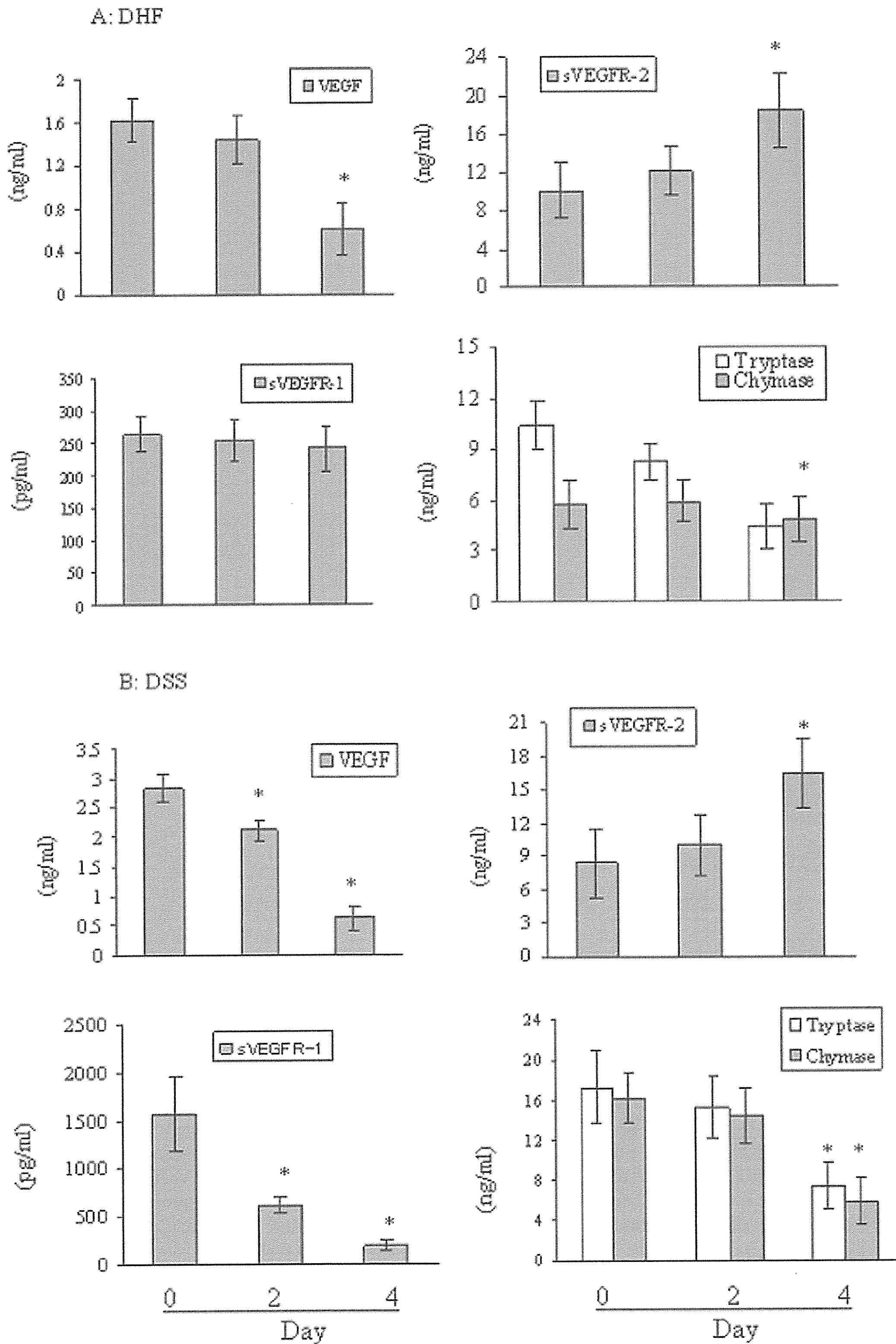

Figure 2. Plasma levels of VEGF, sVEGFRs, tryptase, and chymase in Dengue patients during the admission period. Plasma levels of VEGF, sVEGFR-1 and -2, tryptase, and chymase were examined for Dengue patients with DHF (A) and DSS (B) by ELISA. VEGF in DHF *$p<0.01$ (day 4 versus day 0 and 2) and DSS *$p<0.01$ (day 2 and 4 versus day 0), sVEGFR-1 in DSS *$p<0.01$ (day 2 and 4 versus day 0), sVEGFR-2 in DHF *$p<0.01$ (day 4 versus day 0 and 2) and DSS *$p<0.01$ (day 4 versus day 0 and 2), tryptase in DHF *$p<0.01$ (day 4 versus day 0 and 2) and DSS *$p<0.01$ (day 4 versus day 0 and 2), chymase in DSS *$p<0.01$ (day 4 versus day 0 and 2). Results are representative of two independent experiments.
doi:10.1371/journal.pntd.0001505.g002

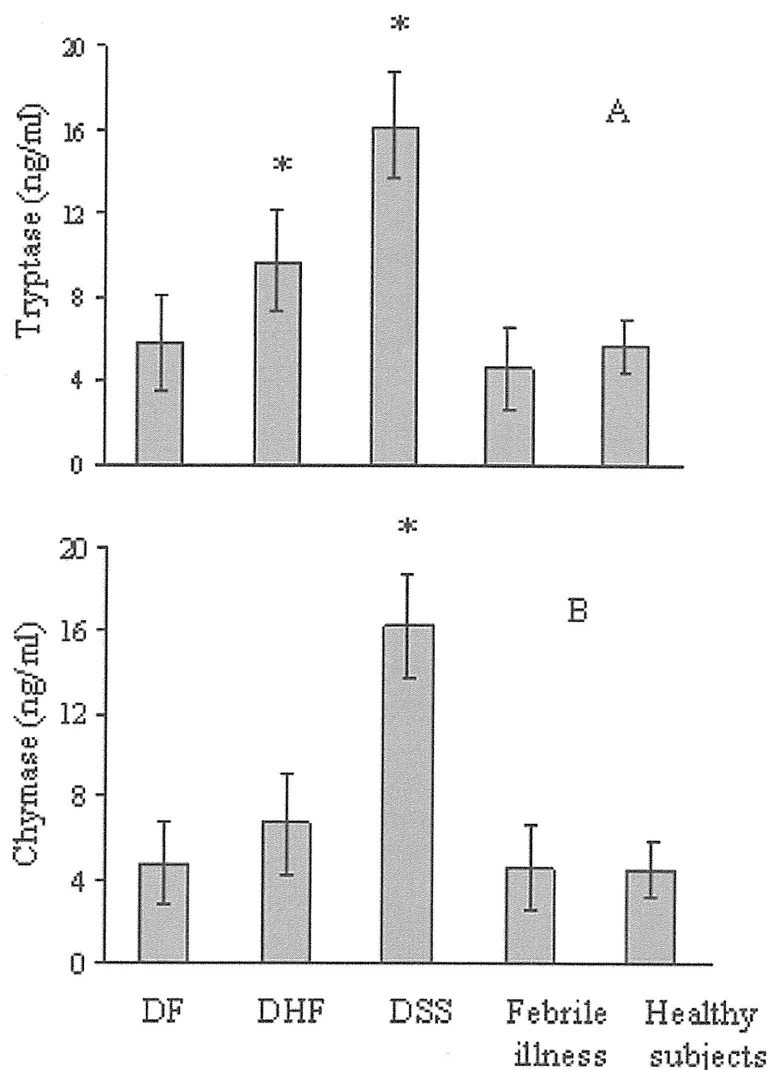

Figure 3. Plasma levels of tryptase and chymase in Dengue patients and control groups. Plasma levels of tryptase (A) and chymase (B) were examined for Dengue patients (DF, DHF, and DSS) on admission and for control groups (febrile illness and healthy subjects) by ELISA. Tryptase *$p<0.01$ (DHF and DSS versus DF, febrile illness or healthy subjects), and Chymase *$p<0.01$ (DSS versus DF, DHF, febrile illness or healthy subjects). Representative results from three independent experiments are shown.
doi:10.1371/journal.pntd.0001505.g003

occurrence of permissive infection of Dengue virus (Fig. 5). To determine the role of IL-9 in VEGF production by mast cells, KU-812 and HMC-1 cells were inoculated with Dengue virus and human Dengue virus-immune serum (1:1,000 final dilution) in the presence and absence of IL-9. Although a low level of VEGF production by KU-812 and HMC-1 cells was observed without IL-9, VEGF levels were significantly increased in the presence of IL-9 (Table 3). The effect of IL-9 on VEGF production by KU812 and HMC-1 cells was not observed in the presence of normal human serum (data not shown). Taken together, these findings suggested the possibility that Dengue virus induces VEGF secretion from human mast cells during infection, and that IL-9 supports the production of VEGF in mast cells.

Discussion

Recently, Srikiatkhachorn et al. [48] compared the plasma levels of VEGF-A and sVEFGR-1 and -2 between DHF and DF patients, and found a rise of VEGF-A and decline of sVEGFR-2 levels in DHF patients, with the severity of plasma leakage inversely correlating with sVEGFR-2 levels. These findings seemed to be consistent with our present results that VEGF and sVEGFR-2 were significantly increased and reduced, respectively, in DHF and DSS patients. Although the reason why sVEGFR-2 levels are decreased in DHF and DSS patients is not clear, as VEGF binding to VEGFR-2 on endothelial cells results in receptor phosphorylation, changes in endothelial cell morphology and proliferation, and maintenance of physiological condition of blood vessels, decreased sVEGFR-2 levels in severe Dengue patients might represent the dysfunction of homeostasis in vascular endothelial cells and correlate with increased plasma leakage [49]. We additionally observed a significant increase of sVEGFR-1 levels in DSS patients, which suggests that activation of monocytes/macrophages by Dengue virus leads to increased expression of soluble and surface VEGFR-1 on cells during severe Dengue infection, as was previously reported [49].

Regarding the relationship between VEGF level and severity of Dengue virus infection, Tseng et al. [3] observed the elevation of

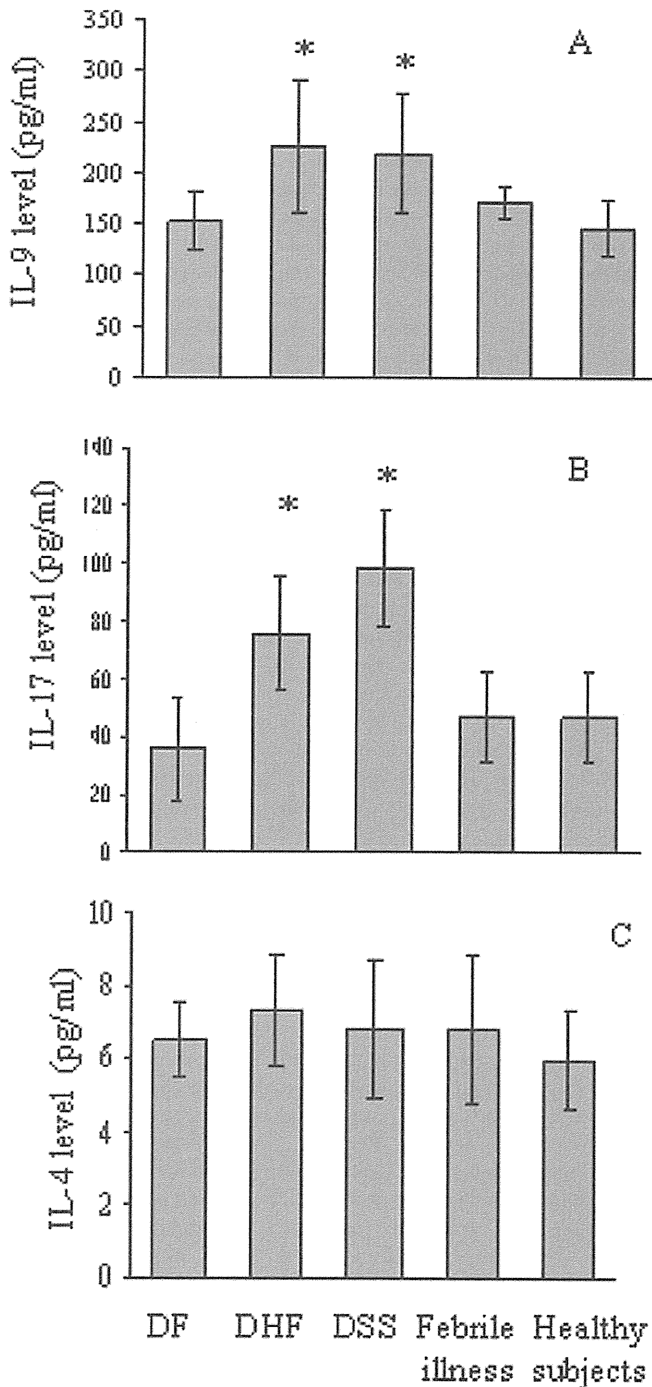

Figure 4. Plasma levels of IL-9, -17, and -4 in Dengue patients and control groups. Plasma levels of IL-9 (A), -17 (B), and -4 (C) were examined for Dengue patients (DF, DHF and DSS) and control groups (febrile illness and healthy subjects) by ELISA. IL-9, *$p<0.01$ (DHF and DSS versus DF, febrile illness and healthy subjects), IL-17, *$p<0.01$ (DHF and DSS versus DF, febrile illness and healthy subjects). Representative results from three independent experiments are shown.
doi:10.1371/journal.pntd.0001505.g004

circulating VEGF levels in adult DHF patients during the early phases of Dengue infection, as compared to DF patients and healthy controls. In a study of a pediatric population with DHF, Srikiatkhachorn et al. [6] also detected a rise in circulating VEGF in the early febrile and defervescent stages of Dengue infection, but not during the later convalescent stage. However, subsequent studies reported contradictory findings, as increased circulating VEGF concentrations were not observed during the early febrile and toxic stages in DHF, but lower VEGF concentrations were detected in patients with more severe Dengue infection [50–52]. Several underlying reasons may explain these differences, such as poor study design, small sample size, and the lack of a standardized collection methodology and storage of blood samples used for the measurement of VEGF. In addition, VEGF is also expressed at low levels in a wide variety of normal adult human and animal tissues, and at higher levels in a few selected sites,

Table 2. VEGF production by KU812 and HMC-1 cells exposed to Dengue virus.

Cell line	Serum	DV	UDV	C3/36[a]	C48/80[b] (ng/ml)
KU812	HDIS	4.2±0.9*	0.4±0.3	0.4±0.1	7.5±0.8*
KU812	NHS	0.5±0.2	0.4±0.2	0.4±0.1	7.3±0.4*
HMC-1	HDIS	2.3±0.4*	0.5±0.2	0.4±0.1	4.6±0.4*
HMC-1	NHS	0.5±0.1	0.4±0.1	0.3±0.2	4.2±0.7*

KU812 and HMC-1 cells were inoculated with Dengue virus-2 (DV) or UV-irradiated Dengue virus-2 (UDV) in the presence of human Dengue-immune serum (HDIS, 1:1000 final dilution) or normal human serum (NHS, 1:1000 final dilution), and VEGF levels in culture supernatants were then examined 24 h later. Significant VEGF production was not observed when KU812 and HMC-1 cells were infected with DV alone (data not shown).
HDIS in KU812*p$<$0.01 (HDIS and C48/80 versus UDV or C3/36), NHS in KU812 *p$<$0.01 (C48/80 versus DV, UDV or C3/36), HDIS in HMC-1 *p$<$0.01 (DV versus UDV, C3/36 or C48/80), NHS in HMC-1 *p$<$0.01 (C48/80 versus DV, UDV, or C3/36).
[a]: C3/36 medium alone served as a negative control.
[b]: Activation of mast cells with C48/80 (300 μg/ml) was used as a positive control.
doi:10.1371/journal.pntd.0001505.t002

namely, podocytes of the renal glomerulus, cardiac myocytes, prostatic epithelium and semen, and certain epithelial cells of the adrenal cortex and lung [53]. Dovrak et al. [54] reported that VEGF is substantially overexpressed at both the mRNA and protein levels in a high percentage of malignant animal and human tumors, as well as in many transformed cell lines. Thus, studies of VEGF production by mast cells during Dengue virus infection are complicated by these alternate sources of VEGF in human and animals, and may affect circulating VEGF levels.

Incubation of KU812 and HMC-1 cells with Dengue virus in the presence of human Dengue virus-immune serum resulted in enhanced VEGF production, which was not observed when UV-inactivated Dengue virus was incubated with human Dengue virus-immune serum or when Dengue virus was used alone to infect KU812 cells (Table 2). As the permissive infection of Dengue virus was observed in KU812 cells (Fig. 5), these findings suggest the critical importance of antibodies to Dengue virus for VEGF production by highly infected mast cells and indicate that infected mast cells can secrete VEGF without stimulation through FcεRI. Our results appear consistent with the findings that Dengue virus infection induces the production of chemokines by human mast cells without stimulation of FcεRI in the presence of human Dengue immune serum [22]. Brown et al. [55] reported that FcγRII plays a dominant role in antibody-enhanced Dengue virus infection of the human mast cell lines HMC-1 and KU812, and in the associated CCL5 release. In studies of DHF epidemics, Halstead et al. [56] and Guzmán et al. [57] demonstrated that secondary infection is the most important host risk factor for DHF.

Boesiger et al. and Grützkau et al. [46,47] reported that mouse and human mast cells produce and secrete VPF/VEGF, and release VEGF upon stimulation through FcεRI or after challenge with chemical mast cell activators. Notably, the FcεRI-dependent secretion of VEGF by either mouse or human mast cells is significantly increased in cells that have undergone upregulation of FcεRI surface expression by preincubation with IgE. As Koraka [58] reported that Dengue virus-specific IgE levels were significantly higher in DHF and DSS patients compared to those in DF and non-Dengue patients, FcεRI may be important for mast cell activation via IgE antibody in Dengue virus infection. However, we did not determine whether the patient sera collected in the present contained IgE antibody against Dengue virus. To clarify the importance of FcεRI for VEGF production by mast cells in Dengue virus infection, further studies are needed.

High levels of VEGF in culture supernatants were detected when KU812 and HMC-1 cells were cultured in the presence of IL-9 (Table 3). As IL-9 enhances the survival of mast cells and induces their production of proinflammatory cytokines, including Th1 and Th2 cytokines [59], it is possible that IL-9 primes HMC-1 and KU812 cells *in vitro* to respond to Dengue virus infection by promoting VEGF production. To evaluate the contribution of IL-9 and IL-17 to Dengue virus infection, we measured the plasma levels of these two cytokines in Dengue patients and found that both IL-9 and IL-17 were significantly increased in DHF and DSS compared with DF, febrile illness and healthy subjects. These findings suggest that Th9 and Th17 cells contribute to the inflammatory response to severe Dengue virus infection. It is possible that IL-9 may act additively or synergistically with other factors, such as additional Th2 cytokines, to induce the mast cell response observed in this study. However, as the level of IL-4 was not increased in the plasma of Dengue patients, our findings suggest the independent involvement of IL-9 secreted by Th2 cells in Dengue virus infection. Recently, IL-9-producing cells have been described as a new subset of the T helper cell population separate from Th2 that produces IL-9 in large quantities and contributes uniquely to immune responses [60,61]. This cell population has been named 'Th9', and IL-9 secreted by T cells, particularly Th9 cells, may regulate chronic allergic inflammation [62]. Moreover, IL-9 has been recently proposed to function as a Th17-derived cytokine that contributes to inflammatory diseases [36].

Tryptase and chymase levels were significantly increased in DHF and DSS, and DSS, respectively, on admission compared with DF, febrile illness, and healthy subjects (Fig. 3). However, 2–4 days after admission, the levels of these proteases had returned to similar levels with the other patient groups (Fig. 2). These findings support the concept that mast cells and mast cell degranulation play important roles in the pathogenesis of DHF/DSS and might be suitable targets for new therapies and prevention of Dengue infection. However, it is presently unclear whether Dengue virus infection in mast cells directly induces chymase and tryptase production and secretion. Recently, Kitamura-Inenaga et al. [63] reported that encephalomyocarditis virus infection results in mast cell chymase and tryptase production *in vivo*, and additionally, viral infections have been shown to cause the accumulation of mast cells in the nasal mucosa during the first days of a symptomatic, naturally acquired respiratory infection [64]. However, the relevance and underlying mechanisms of mast cell infection and activation in the setting of viral infections remain to be characterized in detail.

Immunocytohistochemical studies in human tissues have identified two mast cell phenotypes distinguishable by their neutral protease content, namely the 'mast cell-tryptase' (MCT) phenotype and the 'mast cell-tryptase-chymase' (MCTC) phenotype [65]. MCT appears to be associated with immune system-related mast cells that play a primary role in host defenses and are preferentially located at mucosal surfaces. MCT mast cells are increased in number in areas of T lymphocyte infiltration and in allergic disease, and are reduced in number in acquired and chronic immunodeficiency syndromes [65]. In contrast, the MCTC phenotype appears to be associated with non-immune system-related mast cells that primarily function in angiogenesis and tissue remodeling, rather than immunologic protection, and are found predominantly in submucosal and connective tissues. In addition, MCTC mast cells are not increased in numbers in areas

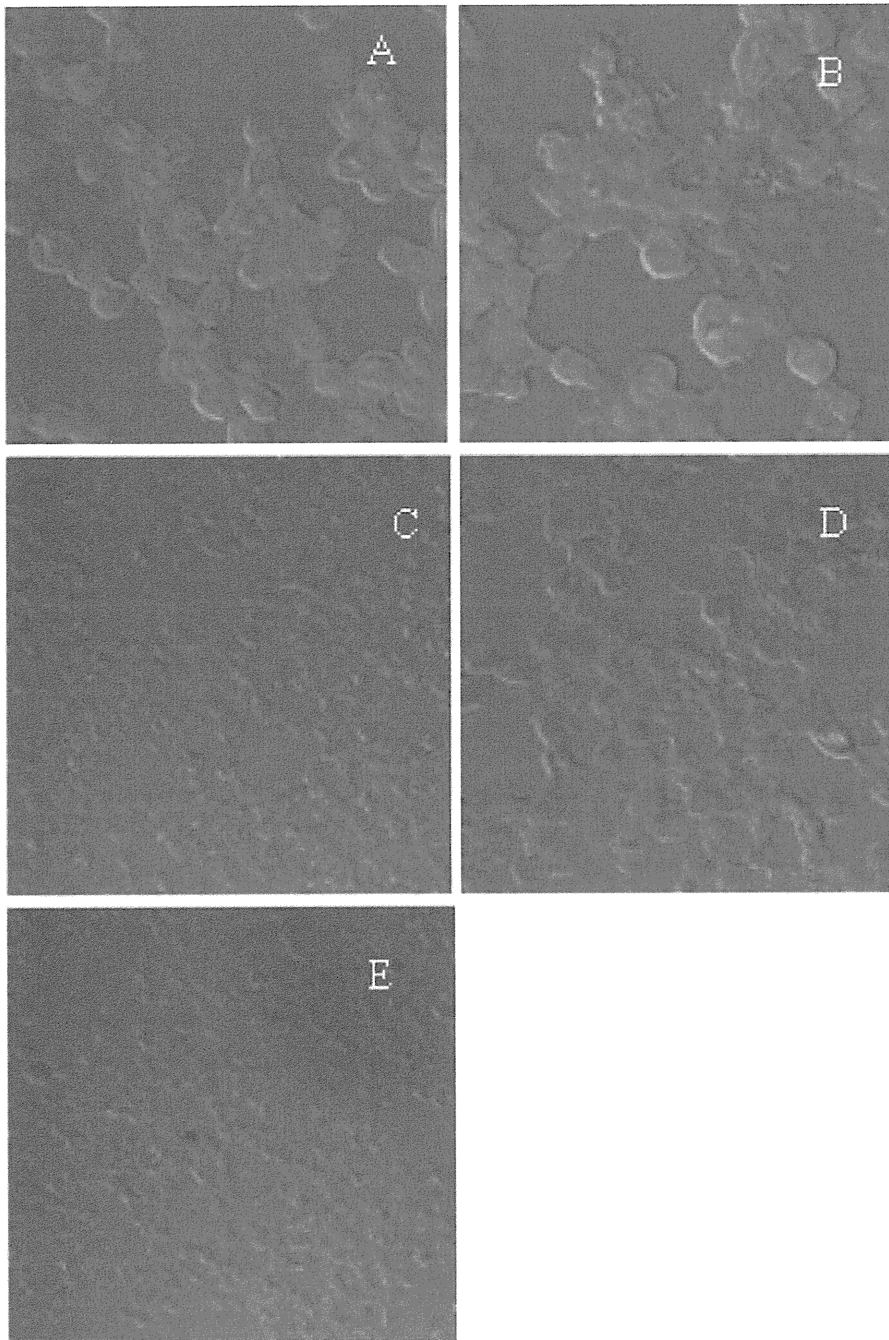

Figure 5. Immunofluorescence staining of Dengue virus-infected KU812 cells. KU812 cells inoculated with a combination of Dengue virus-2 and human Dengue virus-immune or normal human serum were harvested 24 h post-infection. The harvested cells were incubated with mouse anti-Dengue virus monoclonal antibody 1B7 (30) or isotype-matched mouse IgG2a antibody (negative control) as a primary antibody. Anti-mouse IgG labeled with FITC was used as a secondary antibody. (A)–(C) KU812 cells inoculated with a combination of Dengue virus and human Dengue virus-immune serum (A 1:1000, ×40; B 1:10000 final dilution, ×40) or normal human serum (C 1:1000 final dilution, ×20). (D)–(E) KU812 cells inoculated with a combination of UV-inactivated Dengue virus and human Dengue virus-immune serum (D 1:1000, ×40) or human Dengue immune serum alone (E 1:1000, ×20). Similar results were obtained from two additional experiments.
doi:10.1371/journal.pntd.0001505.g005

of heavy lymphocytic infiltration and are not decreased in number in immunodeficiency syndromes [65]. In the present study, a significant increase of chymase was observed in the plasma of DSS patients as compared with those of DF, DHF, and the control group, suggesting the possibility that MCTC mast cells contribute to the pathogenesis of severe forms of Dengue virus infection. However, further study is needed to clarify the roles of tryptase and chymase in severe Dengue virus infection.

Concerning the ability of mediators produced by mast cells other than VEGF to activate endothelial cells, King et al. [22] reported that Dengue virus plus Dengue virus-specific antibody treatment results in selective production of the T-cell chemoattractants RANTES, MIP-1α, and MIP-1β by KU812 and HMC-1 human mast cell-basophil lines. In addition, Brown et al. [66] demonstrated that antibody-enhanced Dengue virus infection of primary human cord blood-derived mast cells (CBMCs) and

Table 3. Effect of IL-9 on VEGF production in KU812 and HMC-1 cells.

Cell line	IL-9	DV	UDV	C3/36[a]	C48/80[b] (ng/ml)
KU812	+	6.4±1.4*	0.5±0.3	0.6±0.3	8.4±1. 2*
KU812	−	4.1±1.5*	0.4±0.1	0.4±0.2	7.4±1.8*
HMC-1	+	3.5±0.4*	0.6±0.1	0.5±0.2	4.9±0.4*
HMC-1	−	2.3±0.2*	0.4±0.2	0.3±0.1	4.2±0.2*

KU812 and HMC-1 cells were inoculated with Dengue virus-2 (DV) or UV-irradiated Dengue virus-2 (UDV) in the presence or absence of IL-9 (200 ng/ml), and VEGF levels were then examined. VEGF production by KU812 and HMC-1 cells was significantly increased in the presence of IL-9 when the cells were cultured with DV and human Dengue virus-immune serum (1:1000 final dilution). However, significant VEGF production was not observed in cells when KU812 and HMC-1 cells were infected with DV or UDV in the presence of normal human serum (1:1000 final dilution) and IL-9 (data not shown). Significant VEGF production was not observed when KU812 and HMC-1 cells were infected with DV alone (data not shown).
IL-9+ in KU812 *p<0.01 (DV and C48/80 versus UDV or C3/36), IL-9− in KU812 *p<0.01 (DV and C48/80 versus UDV or C3/36), IL-9+ in HMC-1 *p<0.01 (DV and C48/80 versus UDV or C3/36), IL-9− in HMC-1 *p<0.01 (DV and C48/80 versus UDV or C3/36).
[a]: C3/36 medium alone (mock infection) served as a negative control.
[b]: Activation of mast cells with C48/80 (300 μg/ml) was used as a positive control.
*p<0.01.
doi:10.1371/journal.pntd.0001505.t003

HMC-1 cells results in the release of ICAM-1 and VCAM-1, which subsequently activate human endothelial cells. St. John et al. [67] reported that the response to mast cell activation involves the *de novo* transcription of cytokines, including TNF-α and IFN-α, and chemokines, such as CCL5, CXCL12, and CX3CL1, which are well characterized to recruit immune effector cells, including cytotoxic lymphocytes, to sites of peripheral inflammation.

In conclusion, we found that mast cells and mast cell-derived mediators, namely VEGF, and the mast cell-specific proteases tryptase and chymase participate in the development of severe forms of Dengue virus infection, which is accompanied by elevated circulating levels of IL-9 and -17. As tryptase and chymase are known as selective markers of non-immune system-related activation of mast cells in submucosal and connective tissues, these two proteases, particularly chymase, might serve as good predictive markers of Dengue disease severity.

Author Contributions

Conceived and designed the experiments: TF. Performed the experiments: TF LAM NTH. Analyzed the data: LAM NTH MK KM MY. Contributed reagents/materials/analysis tools: NTPL VTQH TTT CTPN VDT TTNH YO MK KM MY. Wrote the paper: TF KH NW.

References

1. Rothman AL, Ennis FA (1994) Immunopathogenesis of dengue hemorrhagic fever. Virology 257(1): 1–6.
2. Burke DS, Monath TP (2001) Flaviviruses. *In* DM. Knipe, PM. Howley, eds. Fields virology, 4th ed., vol. 1. Philadelphia: Lippincott Williams and Wilkins. pp 1043–1126.
3. Tseng CS, Lo HW, Teng HC, Lo WC, Ker CG (2005) Elevated levels of plasma VEGF in patients with dengue hemorrhagic fever. FEMS Immunol Med Microbiol 43: 99–102.
4. Senger DR, Galli SJ, Dvorak AM, Perruzzi CA, Harvey VS, et al. (1983) Tumor cells secrete a vascular permeability factor that promotes accumulation of ascites fluid. Science 219: 983–985.
5. Zebrowski BK, Yano S, Liu W, Shaheen RM, Hicklin DJ, et al. (1999) Vascular endothelial growth factor levels and induction of permeability in malignant pleural effusions. Clin Cancer Res 5: 3364–3368.
6. Leung DW, Cachianes G, Kuang WJ, Goeddel DV, Ferrara N (1989) Vascular endothelial growth factor is a secreted angiogenic mitogen. Science 246: 1306–1309.
7. Claesson-Welsh L (2003) Signal transduction by vascular endothelial growth factor receptors. Biochem Soc Trans 31: 20–24.
8. Shibuya M, Ito N, Claesson-Welsh L (1999) Structure and function of vascular endothelial growth factor receptor-1 and -2. Curr Top Microbiol Immunol 237: 59–83.
9. Quinn TP, Peters KG, De Vries C, Ferrara N, Williams LT (1993) Fetal liver kinase 1 is a receptor for vascular endothelial growth factor and is selectively expressed in vascular endothelium. Proc Natl Acad Soci USA 90: 7533–7537.
10. Sawano A, Iwai S, Sakurai Y, Ito M, Shitara K, et al. (2001) Flt-1, vascular endothelial growth factor receptor 1, is a novel cell surface marker for the lineage of monocyte-macrophages in humans. Blood 97: 785–791.
11. Kim I, Moon SO, Kim SH, Kim HJ, Koh YS, et al. (2001) Vascular endothelial growth factor expression of intercellular adhesion molecule 1 (ICAM-1), vascular cell adhesion molecule 1 (VCAM-1), and E-selectin through nuclear factor-kappa B activation in endothelial cells. J Biol Chem 276: 7614–7620.
12. Reinders ME, Sho M, Izawa A, Wang P, Mukhopadhyay D, et al. (2003) Proinflammatory functions of vascular endothelial growth factor in alloimmunity. J Clin Invest 112: 1655–1665.
13. Lucerna M, Mechtcheriakova D, Kadl A, Schabbauer G, Schafer R, et al. (2003) NAB2, a corepressor of EGR-1, inhibits vascular endothelial growth factor-mediated gene induction and angiogenic responses of endothelial cells. J Biol Chem 278: 11433–11440.
14. Kuenen BC, Levi M, Meijers JC, Kakkar AK, van Hinsbergh VW, et al. (2002) Analysis of coagulation cascade and endothelial cell activation during inhibition of vascular endothelial growth factor/vascular endothelial growth factor receptor pathway in cancer patients. Arterioscler Thromb Vasc Biol 22: 1500–1505.
15. Harada M, Mitsuyama K, Yoshida H, Sakisaka S, Taniguchi E, et al. (1998) Vascular endothelial growth factor in patients with rheumatoid arthritis. Scand J Rheumatol 27: 377–380.
16. Taha Y, Raab Y, Larsson A, Carlson M, Loof L, et al. (2004) Vascular endothelial growth factor (VEGF)–a possible mediator of inflammation and mucosal permeability in patients with collagenous colitis. Dig Dis Sci 49: 109–115.
17. Tuchinda M, Dhorranintra, B, Tuchinda P (1977) Histamine content in 24-hour urine in patients with dengue haemorrhagic fever. Southeast Asian J Trop Med Public Health 8: 80–83.
18. Bhamarapravati H, Tuchinda P, Boonyapaknavik V (1967) Pathology of Thailand hemorrhagic fever: a study of 100 autopsy cases. Ann Trop Med Parasitol 61: 500–510.
19. King CA, Marshall JS, Alshurafa H, Anderson R (2000) Release of vasoactive cytokines by antibody-enhanced Dengue virus infection of a human mast cell/basophil line. J Virol 74: 7146–7150.
20. King CA, Anderson R, Marshall JS (2002) Dengue Virus Selectively Induces Human Mast Cell Chemokine Production. J Virol 76: 8408–8419.
21. Metcalfe DD, Baram D, Mekori YA (1997) Mast cells. Physiol Rev 114: 1–9.
22. Galli SJ, Kalesnikoff J, Grimbalderson MA, Piliponsky AM, Williams CT, et al. (2005) Mast cells as "tunable" effector and immunoregulatory cells: recent advances. Annu Rev Immunol 23: 749–786.
23. Galli SJ, Maurer M, Lantz CS (1999) Mast cells as sentinels of innate immunity. Curr Opin Immunol 11: 53–59.
24. Marshall JS, Bienenstock J (1994) The role of mast cells in inflammatory reactions of the airways, skin, and intestine. Curr Opin Immunol 6: 853–859.
25. Echtenacher B, Mannel DN, Hultner L (1996) Critical protective role of mast cells in a model of acute septic peritonitis. Nature 381: 75–77.
26. Church MK, Holgate ST, Shute JK, Walls AF, Sampson AP (1998) Mast cell-derived mediators. Allergy: Principles and practice, by Elliot Meddleton, 5th ed. St. Louis: Mosby.
27. Féger F, Varadaradjalou S, Gao Z, Abraham SN, Arock M (2002) The role of mast cells in host defense and their subversion by bacterial pathogens. Trends Immunol 23: 151–158.
28. Schwartz LB, Lewis RA, Austen KF (1981) Tryptase from human pulmonary mast cells. Purification and characterization. J Biol Chem 256: 11939–11943.
29. Schwartz LB, Sakai K, Bradford TR, Ren S, Zweiman B, et al. (1995) The alpha form of human tryptase is the predominant type present in blood at baseline in normal subjects and is elevated in those with systemic mastocytosis. J Clin Invest 96: 2702–2710.
30. Schwartz LB, Irani AM, Roller K, Castells MC, Schechter NM (1987) Quantitation of histamine, tryptase, and chymase in dispersed human T and TC mast cells. J Immunol 138: 2611–2615.

31. He S, Walls AF (1998) Human mast cell chymase induces the accumulation of neutrophils, eosinophils and other inflammatory cells *in vivo*. Br J Pharmacol 125: 1491–1500.
32. Hultner L, Druez C, Moeller J, Uyttenhove C, Schmitt E, et al. (1990) Mast cell growth-enhancing activity (MEA) is structurally related and functionally identical to the mouse T cell growth factor P40/TCGFIII (interleukin 9). Eur J Immunol 20: 1413–1416.
33. Renauld JC, Vink A, Louahed J, Van Snick J (1995) Interleukin-9 is a major anti-apoptotic factor for thymic lymphomas. Blood 85: 1300–1305.
34. Nowak EC, Weaver CT, Turner H, Begum-Haque S, Becher B, et al. (2009) IL-9 as a mediator of Th17-driven inflammatory disease. J Exp Med 206: 1653–1660.
35. World Health Organization (1997) Dengue hemorrhagic fever: diagnosis, treatment, prevention, and control. Geneva. pp 1–11.
36. Rigau-Pérez JG (2006) Severe dengue: the need for new case definitions. Lancet Infect Dis 6: 297–302.
37. World Health Organization (2005) Dengue, dengue haemorrhagic fever and dengue shock syndrome in the context of the integrated management of childhood illness. WHO/FCH/CAH/05.13.
38. Lan NTP, Kikuchi M, Huong VTQ, Ha DQ, Thuy TT, et al. (2008) Protective and Enhancing H2LA Alleles, HLA-DRB1*0901 and HLA-A*24, for Severe Forms of Dengue Virus Infection, Dengue Hemorrhagic Fever and Dengue Shock Syndrome. PLoS Negl Trop Dis 2(10): e304. doi:10.1371/journal.pntd.0000304.
39. Lanciotti RS, Calisher CH, Gubler DJ, Chang GJ, Vorndam AV (1992) Rapid detection and typing of dengue viruses from clinical samples by using reverse transcriptase-polymerase chain reaction. J Clin Microbiol 30: 545–551.
40. Innis BL, Nisalak A, Nimmannitva S, Kusalerdchariya S, Chongswasdi V, et al. (1989) An enzyme-linked immunosorbent assay to characterize dengue infections where dengue and Japanese encephalitis co-circulate. Am J Trop Med Hyg 40: 418–427.
41. Gubler DJ, Kuno G, Sather GE, Valez M, Oliver A (1984) Mosquito cell cultures and specific monoclonal antibodies in surveillance for dengue viruses. Am J Trop Med Hyg 33: 158–165.
42. Kishi K (1985) A new leukemia cell line with Philadelphia chromosome characterized as basophil precursors. Leuk Res 9: 381–390.
43. Butterfield JH, Weiler D, Dewald G, Gleich GJ (1988) Establishment of an immature mast cell line from a patient with mast cell leukemia. Leuk Res 12: 345–355.
44. Igarashi A (1978) Isolation of a Singh's *Aedes albopictus* cell clone sensitive to dengue and chikungunya viruses. J Gen Virol 40: 531–544.
45. Kinoshita H, Mathenge EGM, Hung NT, Huong VTQ, Kumatori A, et al. (2009) Isolation and characterization of two phenotypically distinct Dengue type-2 virus isolate from the same Dengue hemorrhagic fever patients. Jpn J Infect Dis 62: 343–350.
46. Boesiger J, Tsai M, Maurer M, Yamaguchi M, Brown LF, et al. (1998) Mast cells can secrete vascular permeability factor/vascular endothelial cell growth factor and exhibit enhanced release after immunoglobulin E–dependent upregulation of Fcεreceptor I expression. J Exp Med 21: 1135–1145.
47. Grützkau A, Krüger-Krasagakes S, Baumeister H, Schwarz C, Kögel H, et al. (1998) Synthesis, storage, and release of vascular endothelial growth factor/vascular permeability factor (VEGF/VPF) by human mast cells: Implications for the biological significance of VEGF206. Mol Biol Cell 9: 875–884.
48. Srikiatkhachorn A, Ajariyakhajorn C, Endy TP, Kalayanarooj S, Libraty DH, et al. (2007) Virus-induced decline in soluble vascular endothelial growth receptor 2 is associated with plasma leakage in Dengue hemorrhagic fever. J Virol 81: 1592–1600.
49. Watenberger J, Claessen-Weish L, Siegbahn A, Shibuya M, Heldin CH (1994) Different signal transduction properties of KDR and Flt-1, two receptors for vascular endothelial growth factor. J Biol Chem 269: 26988–26995.
50. Sathupan P, Khongphattanayothin A, Srisai J, Srikaew K, Poovorawan Y (2007) The role of vascular endothelial growth factor leading to vascular leakage in children with dengue virus infection. Ann Trop Paediatr 27: 179–84.
51. Seet RCS, Chow AWL, Quek AML, Chen Y-H, Lim ECH (2009) Relationship between circulating vascular endothelial growth factor and its soluble receptors in adults with dengue virus infection: a case—control study. Int J Infect Dis 13: e248–e253.
52. Becquart P, Wauquier N, Nkoghe D, Ndjoyi-Mbiguino A, Padilla C, et al. (2010) Acute dengue virus 2 infection in Gabonese patients is associated with an early innate immune response, including strong interferon alpha production. BMC Infect Dis 10: 356–366.
53. Ferrara N, Davis-Smyth T (1997) The biology of vascular endothelial growth factor. Endocr Rev 18: 4–25.
54. Dvorak HF, Brown LF, Detmar M, Dvorak AM (1995) Vascular permeability factor/vascular endothelial growth factor, microvascular hyperpermeability, and angiogenesis. Am J Pathol 146: 1029–1039.
55. Brown M, King CA, Sherren C, Marshall JS, Anderson R (2006) A dominant role for FcγRII in antibody-enhanced dengue virus infection of human mast cells and associated CCL5 release. J Leukoc Biol 80: 1242–1250.
56. Halstead SB, Mimmannitya S, Cohen SN (1970) Observations related to pathogenesis of dengue hemorrhagic fever. IV. Relation of disease severity to antibody response and virus recovered. Yale J Biol Med 42: 311–328.
57. Guzman MG, Kouri G (2008) Dengue haemorrhagic fever integral hypothesis: confirming observations, 1987—2007. Trans Roy Soc Trop Med Hyg 102: 522–523.
58. Koraka P, Murgue B, Deparis X, Setiati TE, Suharti C, et al. (2003) Elevated levels of total and Dengue virus-specific immunoglobulin E in patients with varying disease severity. J Med Virol 70: 91–98.
59. Louahed J, Kermouni A, Van Snick J, Renauld JC (1995) IL-9 induces expression of granzymes and high-affinity IgE receptor in murine T helper clones. J Immunol 154: 5061–5070.
60. Veldhoen M, Uyttenhove C, van Snick J, Helmby H, Westendorf A, et al. (2008) Transforming growth factor-beta 'reprograms' the differentiation of T helper 2cells and promotes an interleukin 9-producing subset. Nat Immunol 9: 1341–1346.
61. Dardalhon V, Awasthi A, Kwon H, Galileos G, Gao W, et al. (2008) IL-4 inhibits TGF-beta induced Foxp3^{+} T cells and, together with TGF-beta, generates IL-9^{+} IL-10^{+} Foxp3^{-} effector T cells. Nat Immunol 9: 1347–1355.
62. Soroosh P, Doherty TA (2008) Th 9 and allergic disease. Immunology 127: 450–458.
63. Kitaura-Inenaga K, Hara M, Higuchi K, Yamamoto K, Yamaki A, et al. (2003) Gene expression of cardiac mast cell chymase and tryptase in a murine model of heart failure caused by viral myocarditis. Circ J 67: 881–884.
64. Alho OP, Karttunen TJ, Karttunen R, Tuokko H, Koskela M, Uhari M (2003) Lymphocyte and mast cell counts are increased in the nasal mucosa in symptomatic natural colds. Clin Exp Immunol 131: 138–142.
65. Pejler G, Abrink M, Ringvall M, Wernersson S (2007) Mast Cell Proteases, Advances in Immunology, Volume 95:169–229. Elsevier Inc. ISSN 0065-2776, DOI: 10.1016/S0065-277.
66. Brown MG, Hermann LL, Issekutz AC, Marshall JS, Rowter D, Al-Afif A, Anderson1 R (2011) Dengue virus infection of mast cells triggers endothelial cell activation. J Virol 85: 1145–1115.
67. St. John AL, Rathorea APS, Yap H, Ng M-L, Metcalfe DD, et al. (2011) Immune surveillance by mast cells during dengue infection promotes natural killer (NK) and NKT-cell recruitment and viral clearance. Proc Nat Acad Soc USA 108: 9190–9195.

PLoS NEGLECTED TROPICAL DISEASES

Protective Human Leucocyte Antigen Haplotype, HLA-DRB1*01-B*14, against Chronic Chagas Disease in Bolivia

Florencia del Puerto[1], Juan Eiki Nishizawa[2], Mihoko Kikuchi[1,3], Yelin Roca[4], Cinthia Avilas[4], Alberto Gianella[4], Javier Lora[4], Freddy Udalrico Gutierrez Velarde[5], Sachio Miura[6], Norihiro Komiya[7], Koji Maemura[7], Kenji Hirayama[1]*

1 Department of Immunogenetics, Institute of Tropical Medicine (NEKKEN), and Global COE Program, Nagasaki University, Nagasaki, Japan, **2** Clinica Siraní, Santa Cruz, Bolivia, **3** Center for International Collaboration Research (CICORN), Nagasaki University, Nagasaki, Japan, **4** Centro Nacional de Enfermedades Tropicales, Santa Cruz, Bolivia, **5** Hospital Universitario Japonés, Santa Cruz, Bolivia, **6** Department of Tropical Medicine and Parasitology, School of Medicine, Keio University, Tokyo, Japan, **7** Department of Cardiovascular Medicine, Nagasaki University Graduate School of Biomedical Science, Nagasaki, Japan

Abstract

Background: Chagas disease, caused by the flagellate parasite *Trypanosoma cruzi* affects 8–10 million people in Latin America. The mechanisms that underlie the development of complications of chronic Chagas disease, characterized primarily by pathology of the heart and digestive system, are not currently understood. To identify possible host genetic factors that may influence the clinical course of Chagas disease, Human Leucocyte Antigen (HLA) regional gene polymorphism was analyzed in patients presenting with differing clinical symptoms.

Methodology: Two hundred and twenty nine chronic Chagas disease patients in Santa Cruz, Bolivia, were examined by serological tests, electrocardiogram (ECG), and Barium enema colon X-ray. 31.4% of the examinees showed ECG alterations, 15.7% megacolon and 58.1% showed neither of them. A further 62 seropositive megacolon patients who had undergone colonectomy due to acute abdomen were recruited. We analyzed their HLA genetic polymorphisms (HLA-A, HLA-B, MICA, MICB, DRB1 and TNF-alpha promoter region) mainly through Sequence based and LABType SSO typing test using LUMINEX Technology.

Principal Findings: The frequencies of HLA-DRB1*01 and HLA-B*14:02 were significantly lower in patients suffering from megacolon as well as in those with ECG alteration and/or megacolon compared with a group of patients with indeterminate symptoms. The DRB1*0102, B*1402 and MICA*011 alleles were in strong Linkage Disequilibrium (LD), and the HLA-DRB1*01-B*14-MICA*011haplotype was associated with resistance against chronic Chagas disease.

Conclusions: This is the first report of HLA haplotype association with resistance to chronic Chagas disease.

Citation: del Puerto F, Nishizawa JE, Kikuchi M, Roca Y, Avilas C, et al. (2012) Protective Human Leucocyte Antigen Haplotype, HLA-DRB1*01-B*14, against Chronic Chagas Disease in Bolivia. PLoS Negl Trop Dis 6(3): e1587. doi:10.1371/journal.pntd.0001587

Editor: Helton da Costa Santiago, National Institutes of Health, United States of America

Received September 2, 2011; **Accepted** February 11, 2012; **Published** March 20, 2012

Copyright: © 2012 del Puerto et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This study was supported in part by the Grant-in-Aid for 21c COE program (2003–2008), the Global Center of Excellence (GCOE) Program (2008–2011), Nagasaki University and Grant-in-Aid for Research A (23256003) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: hiraken@nagasaki-u.ac.jp

Introduction

Following an extensive control program consisting of vector control, serological screening in blood banks and identification and treatment of congenital transmission, the estimated number of people infected with *Trypanosoma cruzi*, the causal agent of Chagas disease in Latin America, has fallen from approximately 20 million in 1981 to around 8–10 million in 2009 [1,2]. Most of the seropositive patients are chronically infected and more than 10,000 deaths are estimated to occur annually from the disease [1,3,4].

Cardiac myositis and autonomous neuroplexus degeneration of the digestive tract are major histopathological alterations that can arise during Chagas disease, and may lead to cardiac failure, digestive abnormalities, megacolon or megaesophagus. Based on these pathologies, there are often considered to be three major clinical forms of Chagas disease; cardiac, digestive and indeterminate [5,6]. This variation in pathological manifestation has been reported to relate to differences in host immune response, such as the ability to control parasitaemia, the strength of inflammatory reactions, and the induction of autoimmune like responses [7–11]. Indeterminate phase T-cells have been reported to correspond with modulatory responses such as increased IL-10 production by $CD4^{+}CD28^{-}$ T cells and the expression of CTLA-4, a ligand that is involved in modulation of T-cell responses by $CD8^{+}$ T-cells [12,13]. Additionally, the unregulated production of IFN-gamma by $CD8^{+}$ T-cells in cardiac Chagas disease patients has been reported, which might result in the destruction of heart tissue due to its cytotoxic effect [12].

The highly polymorphic HLA Class I and II molecules determine the efficiency of *T. cruzi* epitope presentation to T lymphocytes that could affect the clinical course of Chagas disease

Author Summary

Chronic Chagas disease consists of four different forms categorized on the basis of their clinical manifestations, namely; cardiac, digestive, cardiodigestive and indeterminate. In Latin America, there are 8–10 million seropositive persons who are at risk of, or have already developed serious clinical complications and who have limited access to effective treatment. The cardiac and digestive forms are characterized by tissue damage caused by persistent infection of *Trypanosoma cruzi* and are thought to be modulated by host immunity. In our large scale screening for chronic Chagas disease in Santa Cruz, Bolivia, hearts and colons of 229 seropositive patients were examined. We found 31.4% of patients had abnormal electrocardiograms (ECGs), 15.7% presented with megacolon, 5.2% had a combination of abnormal ECG and megacolon, and 58.1% were of indeterminate status. Previously, we attempted to ascertain whether parasite genetic polymorphism might account for the differences in clinical manefestations, by analyzing parasite DNA taken from the same study group (with the addition of a further 62 megacolon post-operational patients). We found no relationships between parasite lineages and clinical disease form. The present study reveals that host HLA polymorphisms associate with clinical manifestations of Chagas.

[14,15]. Several HLA alleles and haplotypes have been reported to be associated with Cardiac Chagas disease in Chile, Venezuela, Brazil and Guatemala [16–18]. The HLA region contains not only classical HLA genes but a wide variety of immunologically relevant genes, such as nonclassical class I genes (MICA, MICB) [19,20], and class III genes (TNF-alpha, beta), that may be involved in pathogenesis [21,22].

In the present study, we investigated HLA class I (A, B, MICA, MICB), Class II (DRB1) and Class III (TNF-alpha) gene polymorphism in seropositive chronic Chagas patients in Bolivia, characterized by electrocardiogram (ECG), barium enema colon X-ray examinations and/or surgical operation.

Materials and Methods

Patients

The study subjects were described previously [23]. Two hundred and ninety one patients with chronic Chagas disease (136 men and 155 women, mean age 45 years) were recruited from the Centro Nacional de Enfermedades Tropicales (CENETROP) (91 men and 119 women), Hospital Primero de Mayo (12 men and 7 women) and from post-operative patients at the Hospital Universitario Japonés (HUJ) (33 men and 29 women) in Santa Cruz, Bolivia.

Upon medical examination of patients other than the HUJ patients, if serological tests (Indirect Haemaglutination assay (IHA) and Indirect Immunofluorescence test (IIF) [3]) were positive, they were asked to participate in the study and signed informed consent was obtained. Informed consent was also obtained for the HUJ post-operative patients, using the same protocol.

ECG abnormalities were diagnosed based on the Minnesota Code Criteria and were confirmed independently by two cardiologists. Colon X-ray with barium enema examination was performed for the detection of megacolon. To exclude the possibility of including individuals who were asymptomatic upon examination, but who may not have had adequate time post-infection for symptoms to become apparent, those under 30 years of age were excluded.

Finally, 229 seropositive Chagas outpatients in Santa Cruz, Bolivia, were examined by ECG and/or barium enema colon X-ray as shown in Table 1. The 62 post-operational patients from HUJ were confirmed to be suffering from Chagas megacolon during the admission period.

The experimental protocol was approved by the Institutional Ethical Review Committee of the Institute of Tropical Medicine, Nagasaki University (No. 0210170018) and the Centro Nacional de Enfermedades Tropicales (CENETROP).

DNA extraction

Genomic DNA was extracted from 10 mL of whole blood containing 10 mM EDTA using a DNA extraction kit (QIAGEN GmbH, FRG) and was stored at −20°C until use.

Typing of MICA, MICB, TNF alpha promoter region and GCT repeat polymorphism in the MICA gene transmembrane region (MICA-TM)

MICA, MICB, TNF-alpha promoter region and MICA-TM typing was performed as previously described [19–21]. Sequences were obtained from a 3730 DNA Analyzer (Applied Biosystems, USA) and submitted to the Assign- software ATF (Conexio Genomics ATF, Australia) for allele identification. The TNFP alleles were determined according to Higuchi T *et al* (1998) and Ubalee R *et al* (2001) [21,22].

MICA-TM alleles were typed by sequencing on a 3730 DNA analyzer (Applied Biosstems, USA) and analyzed with the GeneMapper Software Version 3.7 (Applied Biosystems, USA). The fluorescent primers, 5′ F (GCC CAG TGT ATA ACA AGT CCC-6-FAM) and 3′ R (CCT TAC CAT CTC CAG AAA CTG C) were used.

HLA-A, HLA-B, HLA-DRB1 typing

Typing was carried out according to the manufacturer's specification for LABType SSO Typing, testing for each locus using LUMINEX Technology (ONE LAMBDA, INC, USA) and the retrieved output was analyzed by HLA Fusion software (ONE LAMBDA, INC, USA) for allele identification.

Statistical analysis

Statistical analysis was performed at the two and four digits levels. Allele frequencies less than 5% were removed from the analysis. The statistical significance and odds ratio (OR) of allele frequencies between each group was determined by Chi square and Fisher's exact tests using the StatsDirec software (StatsDirect Ltd, UK). *P*-values were considered significant when <0.05 following Bonferoni correction (Pc). Hardy-Weinberg Equilibrium, linkage disequilibrium (LD) and Haplotype analyses were calculated with PyPopWin32.0.7.0 software [24].

Results

Frequencies of HLA-B*14 and HLA-DRB1*01 significantly decreased in megacolon patients compared with indeterminate patients

In the two digits analysis (Table S1, S3, S5), we observed a significant decreased frequency of alleles DRB1*01 and B*14 in the megacolon patients compared with indeterminate patients (Table 2). The frequency of HLA-B*14 was also significantly lower in the patients with ECG alteration compared to the indeterminate patients. In the four digits analysis (Table S2, S4, S6), DRB1*01:01,

Table 1. Subjects.

Grouping Criteria	IIF[ζ]	HAA[φ]	Mega colon*	ECG alteration**	N = 291	Male/Female	Age ±SD
Indeterminate	+	+	−	−	70	31/39	43±8.5
Megacolon	+	+	+	−	36	17/19	47±12.0
Megacolon	+	+	+	NE	45	26/19	56±16.6
ECG alteration	+	+	−	+	40	20/20	39±9.8
ECG alteration	+	+	NE	+	20	10/10	41±10.6
Megacolon and ECG alteration	+	+	+	+	17	6/11	48±14.1
Not determined	+	+	NE	−	36	15/21	42±8.8
Not determined	+	+	−	NE	27	11/16	43±7.5

*Examination of Colon by Barium enema and X-ray: +; Megacolon, −; no abnormality, NE; not examined.
**Examination of Electrocardiogram. +; typical alterations, −; no alteration, NE; not examined.
[ζ]Indirect Immunofluorescence Test.
[φ]Haemaglutination Assay.
doi:10.1371/journal.pntd.0001587.t001

*01:02 and B*14:02 had the same tendency towards lower allele frequencies in the megacolon patients compared with the indeterminate patients. The frequency of the HLA-B*14:02 allele was significantly lower in the ECG alteration and/or megacolon patients compared with those of indeterminate symptoms (Table 2). The frequency of MICA*011 was also significantly lower in the complication positive groups (Table 2, S7).

TNF-alpha promoter region polymorphism and MICB alleles

None of the TNF-*alpha* promoter region and MICB (Table S11 and S8) alleles tested here were significantly associated with the clinical groups compared with the indeterminate group.

Linkage disequilibrium (LD) between the alleles of HLA-DRB1, A, B, and MICA, B loci

Linkage disequilibrium (LD) was calculated for all combinations of alleles (Table S1, Table S2, Table S3, Table S4, Table S5, Table S6, Table S7, Table S8, Table S9), and significantly strong LDs was observed between the alleles, HLA-DRB1*01, A*33:01, B*14:02 and MICA*011 as shown in Table 3. Except for HLA-A*33:01, all the linkage group alleles showed the same tendency of being associated with protection against megacolon (Table 2). HLA-A*01:01 showed strong LD with HLA-B*08:01, DRB1*0301 and MICB*008. Although HLA-A*01:01 showed a non-significant tendency of association with non-megacolon patients, none of the LD group alleles showed any similar tendency (Table 2). No LD was observed between any HLA alleles and TNF-alpha promoter haplotypes.

Discussion

We have previously shown that there was no association between *T. cruzi* lineage or sub lineage and the clinical manifestation of Chagas disease in samples from Bolivia [23]. Here, we have analysed the same samples in order to determine if there were any associations between clinical disease and host genetic variation. We analysed polymorphic genes located in the MHC region, consisting of three sub regions; class I, II and III. Human MHC, HLA class I and II molecules play a crucial role in determining individual acquired immune responsiveness through the presentation of pathogen-derived peptides to $CD8^+$ and $CD4^+$ T-cells. In the class III sub region, there are a variety of genes related to immunity such as complement, TNF-alpha, Lymphotoxin, *etc.* [24].

Our study revealed that the HLA-DRB1*01-B*14-MICA*011 haplotype was significantly associated with protection from Chagasic megacolon. As the linkage disequilibrium between the HLA-B*14 and DRB1*01 was strong (Table 3), it was difficult to determine the primary associated locus within the haplotype.

HLA-class I is the antigen-presenting molecule on the cell-surface of infected host cells which stimulates microbe-derived antigen specific $CD8^+$ T-cells. Therefore, HLA-B*14 itself could be directly related to protective T-cell immunity as was suggested by the Chagas disease mouse model [25].

If HLA-B*14 is more efficient at stimulating protective T-cells through the binding of antigenic peptides, other HLA-class I molecules that share the same antigen binding motif should show the same protective association. HLA-B*14 consists of three 4-digit alleles, HLAB*14:01, 14:02 and 14:06 that share the same antigen binding motifs but the latter two alleles are so rare that it was impossible to analyse their effect. As the HLA-B*14 alleles belong to the B27 supertype group that share the same anchoring residues in the peptide binding groove, we analyzed the member alleles B*38:01, 39:04, 39:05. 39:06, 39:14, 48:01, 48:03 for their total effect on protection against complication, and found no association (Tables S6 and S10) [26]. This finding did not, therefore, support the hypothesis that the association between HLA-B*14 and protection against clinical Chagas disease is driven by the ability of the gene product to more effectively stimulate protective T-cells than other alleles of this gene.

The peripheral human and mice $CD8^+$T-cells reactive to *Trypanosome* antigens have been identified [27]. Interestingly, HLA-A*02:01 restricted epitopes from cruzipain and FL-160 were frequently recognized by PBMC of patients with Cardiopathy [28]. Moreover, an HLA-A2 tetramer experiment showed that the number of IFN-gamma producing amastigote-specific $CD8^+$ T-cells inversely correlated with the severity of the disease [29]. It will be interesting to see if the same phenomenon occurred in the HLA-B*14 patients.

Immuno-regulatory mechanisms have been reported to be associated with clinical forms including Treg cells [30], NKT cells and NK cells [31]. Whereas the HLA involvement in the induction of Treg cells is not yet clear, HLA-class I can interact with NK cells to suppress their activity through various inhibitory receptors such as KIR. HLA-B*14 belongs to the Bw6 family, the members of which preferentially stimulate specific members of the KIR family [32,33], which could regulate NK cell activity during inflammation. HLA-non classical class I, MICA*011, which was closely linked to HLA-B*14 and DRB1*01 might also be functional as MICA is

Table 2. Two-digit and four-digit alleles association with Clinical Manifestations of Chagas disease.

HLA locus	Indeterminate (N = 70)		Megacolon (N = 98)		ECG Alteration (N = 77)		ECG alteration and/or Megacolon (N = 158)				
HLA-DRB1*	**n**	**(%)**	**n**	**(%)**	**n**	**(%)**	**n**	**(%)**	**Pv**	**Pc**	**OR [95%CI]**
01	12	(17.1)	**1**	**(1.0)**	9	(11.7)	10	(6.3)	0.0001*	0.001*	0.049 [0.001–0.358]*
01:01	4	(5.7)	0	(0.0)	2	(2.6)	2	(1.3)			
01:02	8	(11.4)	**1**	**(1.0)**	3	(3.9)	**4**	**(2.5)**	0.011* 0.009ς	NS* NSς	0.089 [0.002–0.687] * 0.200 [0.043–0.790]
01:03	1	(1.4)	0	(0.0)	4	(5.2)	4	(2.5)			
03:01φ	4	(5.7)	10	(10.2)	3	(3.9)	13	(8.2)			
HLA-A*	**n**	**(%)**	**n**	**(%)**	**n**	**(%)**	**n**	**(%)**	**Pv**	**Pc**	**OR [95%CI]**
01	10	(14.3)	**4**	**(4.1)**	16	(20.8)	20	(12.7)	0.024*	NS*	0.250 [0.056–0.943]*
01:01φ	4	(5.7)	3	(3.1)	**14**	**(18.2)**	17	(10.8)	0.024**	NS**	3.667 [1.067–15.993]**
01:06	4	(5.7)	1	(1.0)	1	(1.3)	1	(0.6)			
01:07	0	(0.0)	0	(0.0)	0	(0.0)	1	(0.6)			
01:14	2	(2.9)	0	(0.0)	1	(1.3)	0	(0.0)			
33:01	1	(1.4)	5	(5.1)	1	(1.3)	6	(3.8)			
HLA-B*	**n**	**(%)**	**n**	**(%)**	**n**	**(%)**	**n**	**(%)**	**Pv**	**Pc**	**OR [95%CI]**
14	10	(14.3)	**2**	**(2.0)**	**3**	**(3.9)**	**5**	**(3.2)**	0.004* 0.040** 0.003ς	0.049* NS** 0.040ς	0.125 [0.013–0.623]* 0.245 [0.042–1.010]** 0.196 [0.051–0.666]ς
14:01	1	(1.4)	0	(0.0)	2	(2.6)	2	(1.3)			
14:02	8	(11.4)	**2**	**(2.0)**	**1**	**(1.3)**	**3**	**(1.9)**	0.018* 0.014** 0.004ς	NS* NS** 0.049ς	0.161 [0.016–0.854]* 0.103 [0.002–0.805]** 0.149 [0.025–0.675]ς
14:06	1	(1.4)	0	(0.0)	0	(0.0)	0	(0.0)			
08:01φ	4	(5.7)	6	(6.1)	5	(6.5)	11	(7.0)			
MICA*	**n**	**(%)**	**n**	**(%)**	**n**	**(%)**	**n**	**(%)**	**Pv**	**Pc**	**OR [95%CI]**
011	6	(8.6)	**0**	**(0.0)**	1	(1.3)	**1**	**(0.6)**	0.005* 0.004ς	NS* 0.048ς	0.050 [0.000–0.832]* 0.069 [0.001–0.584]ς
034	1	(1.4)	4	(4.1)	2	(2.6)	6	(3.8)			
MICB*	**n**	**(%)**	**n**	**(%)**	**n**	**(%)**	**n**	**(%)**	**Pv**	**Pc**	**OR [95%CI]**
008φ	9	(12.9)	5	(5.1)	9	(11.7)	14	(8.9)			

Footnote for Table 2.
*Comparison between Megacolon vs Indeterminate.
**Comparison between ECG alteration vs Indeterminate.
ςComparison between (ECG+ &/or Megacolon+) vs Indeterminate.
φLinkage Disequilibrium group of HLA-DRB1*03:01-MICB*008-B*08:01-A*01:01 as shown in Table 3.
doi:10.1371/journal.pntd.0001587.t002

known to stimulate gamma-delta T-cells in the gut mucosa; a phenomenon that could relate to megacolon [34].

HLA-class II can present antigen to $CD4^{+}$ T-cells so HLA-DRB1*01 may also be directly involved in the pathogenesis as well as HLA-B*14. Many autoimmune diseases are reported to be associated with specific HLA-class II alleles [35]. Auto-reactive processes that involve the activation of cytokine producing T-cells may occur during infection. As was previously suggested, autoimmune mechanisms in the pathogenesis of chronic Chagas heart and colon may be regulated by the HLA-class II.

We analyzed 4-digit HLA-DRB1 alleles for association with Chagas disease clinical manifestations. As shown in Table 2, the HLA-DRB1*01 group included three alleles, DRB1*01:01, 01:02, 01:03 and all of them showed the same protective tendency when compared between the megacolon and indeterminate symptom groups. As two of them, 01:01 and 01:02 shared the same peptide-binding motif [36], we considered that the HLA-DR molecule itself was functionally related to resistance to megacolon. It was previously reported that the DRB1*01 allele was associated with susceptibility to Chagas cardiomyopathy in Venezuela [37]. This opposite association to the megacolon resistance observed in the present work requires further clarification.

Despite the strong LD shown within the HLA-DRB1*01-B*14 haplotype, there was no strong linkage between TNF-alpha promoter alleles that may influence the levels of its production by immune cells [21]. However, between HLA-DRB1 and HLA-B loci, 1270 kb of class III sub region containing more than 60 genes such as complements, heat shock proteins, 21-hydroxylase, are present that might be relevant to pathogenesis. The HLA-B*14:02-DRB1*01:02 haplotype was reported to be associated with V281L polymorphism in 21-hydroxylase in African-American and Caucasian populations [38]. The same kind of

Table 3. Linkage analysis of the alleles with positive association to the complication of Chagas.

Allele 1	Allele 2	observed	expected	diseq	chisq	Degree of freedom	P value
HLA-DRB1*							
01:02	A*33:01	4	0.27	0.006	53.81	20	<0.0001
01:02	B*14:02	8	1.71	0.011	26.81	16	<0.05
01:02	MICA*011	6	0.25	0.009	124.92	28	<0.0001
HLA-A*							
01:01	DRB1*03:01	8	1.01	0.012	51.79	20	<0.0005
01:01	B*08:01	10	0.82	0.016	111.62	20	<0.0001
01:01	MICB*008	8	1.23	0.011	40.29	25	<0.05
HLA-B*							
14:02	DRB1*01:02	8	1.71	0.011	26.81	16	<0.05
14:02	MICA*034	4	0.21	0.007	69.09	28	<0.0001
14:02	MICA*011	8	0.24	0.013	239.21	28	<0.0001
MICA*							
011	B*14:02	8	0.24	0.013	239.21	28	<0.0001
011	DRB1*01:02	6	0.25	0.009	124.92	28	<0.0001
MICB*							
008	A*01:01	8	1.23	0.011	40.29	25	<0.05
008	B*08:01	10	1.11	0.015	73.57	20	<0.0001
008	DRB1*03:01	8	1.25	0.011	36.12	20	<0.05

doi:10.1371/journal.pntd.0001587.t003

abnormality was also reported in the HLA-A*01:01-B*08-DRB1*03 haplotype that is associated with several diseases such as allergy and viral infectious diseases [39]. It was not significantly associated with Chagas. However, HLA-A*01:01 also showed a decreased frequency in the megacolon patients. Whole sequencing of the class III region of the associated haplotypes would be the next target for clarification of any genetic resistance.

We have no information relating to the immunological characteristics of the individuals who poscessed those haplotypes that might be associated with lymphocyte activation during a chronic infection. We did, however, analyze the relationship between individuals' specific antibody titers and their HLA alleles (data not shown) but there was no clear correlation. About 7.1% of the seropositive indeterminate individuals were estimated to carry this haplotype; therefore further to identify its characteristic immunological function is feasible. To our knowledge, this is the first report of resistant HLA haplotype association with chronic Chagas diagnosed by the active examination of silent colon lesion.

Supporting Information

Table S1 The frequency of the Alleles of HLA-DRB1 locus. Two digits analysis.
(DOC)

Table S2 The frequency of the Alleles of HLA-DRB1 locus. Four digits analysis.
(DOC)

Table S3 The frequency of the Alleles of HLA-A locus. Two digits analysis.
(DOC)

Table S4 The frequency of the Alleles of HLA-A locus. Four digits analysis.
(DOC)

Table S5 The frequency of the Alleles of HLA-B locus. Two digits analysis.
(DOC)

Table S6 The frequency of the Alleles of HLA-B locus. Four digits analysis.
(DOC)

Table S7 The frequency of the Alleles of MICA locus.
(DOC)

Table S8 The frequency of the Alleles of MICB locus.
(DOC)

Table S9 The frequency of GCT triplet polymorphism in the MICA-transmembrane region.
(DOC)

Table S10 The frequency of B Supertype.
(DOC)

Table S11 The frequency of TNF-alpha promoter region polymorphism.
(DOC)

Acknowledgments

We are grateful to Dr. Richard Culleton, Dr. Tetsuo Yanagi at NEKKEN Nagasaki University, and Dr. Graciela Russomando at IICS University of Asuncion Paraguay, for their critical discussion and suggestion.

Author Contributions

Conceived and designed the experiments: JEN FUGV SM KH. Performed the experiments: FdP JEN MK YR CA AG JL FUGV. Analyzed the data: MK YR NK KM KH. Contributed reagents/materials/analysis tools: JEN YR FUGV KH. Wrote the paper: FdP JEN MK FUGV KM KH. Diagnostic of Chagas: JEN YR CA AG JL FUGV SM NK KM. Genetic analysis: FdP MK KH.

References

1. Quantative estimation of Chagas disease in the Americas. Montevideo, Pan American Health Organization (2006) (OPS/HDM/CD/425-06).
2. Chagas disease control and prevention in Europe. Report of a WHO Informal Consultation (2009) Geneva, Switzerland, 17–18 December. Geneva: World Health Organization, 2010 (WHO/HTM/NTD/IDM/2010.1).
3. WHO (2002) WHO Control of Chagas disease: second report of a WHO expert committee. WHO technical report series. 905 p.
4. First WHO report on neglected tropical diseases: working to overcome the global impact of neglected tropical diseases (2010) (WHO/HTM/NTD/2010.1).
5. Marin-Neto JA, Rassi A, Jr. (2009) Actualizacion sobre la cardiopatia de la enfermedad de Chagas en el primer centenario de su descubrimiento. Rev Esp Cardiol 62(11): 1211–1216.
6. Rassi Jr. A, Rassi A, Marin-Neto JA (2010) Chagas Disease. Lancet 375: 1388–1402.
7. Marin-Neto JA, Cunha-Neto E, Maciel BC, Simoes MV (2007) Pathogenesis of chronic Chagas heart disease. Circulation 115: 1109–1123.
8. Arce-Fonseca M, Ballinas-Verdugo MA, Reyes PA, Aranda-Fraustro A, Monteon VM (2005) Autoantibodies to human heart conduction system in Chagas' disease. Vector Borne Zoonotic Dis 5: 233–236.
9. Manoel-Caetano Fda S, Silva AE (2007) Implications of genetic variability of *Trypanosoma cruzi* for the pathogenesis of Chagas disease. Cad Saude Publica 23: 2263–2274.
10. Souza PE, Rocha MO, Rocha-Vieira E, Menezes CA, Chaves AC, et al. (2004) Monocytes from patients with indeterminate and cardiac forms of Chagas disease display distinct phenotypic and functional characteristics associated with morbidity. Infect Immun 72: 5283–5291.
11. Cunha-Neto E, Nogueira LG, Teixeira PC, Ramasawmy R, Drigo SA, et al. (2009) Immunological and non-immunological effects of cytokines and chemokines in the pathogenesis of chronic Chagas disease cardiomyopathy. Mem Inst Oswaldo Cruz Jul 104 (l)1: 252–258.
12. Menezes CA, Rocha MO, Souza PE, Chaves AC, Gollob KJ, et al. (2004) Phenotypic and functional characteristics of CD28C and CD28K cells from chagasic patients: distinct repertoire and cytokine expression. Clin Exp Immunol 137: 129–138.
13. Souza PE, Rocha MO, Menezes CA, Coelho JS, Chaves AC, et al. (2007) *T cruzi* infection induces differential modulation of costimulatory molecules and cytokines by monocytes and T cells from patients with indeterminate and cardiac Chagas disease. Infect Immun 75: 1886–1894.
14. Rao X, Costa AI, van Baarle D, Kesmir C (2009) A comparative study of HLA binding affinity and ligand diversity: implications for generating immunodominant CD8+ T cells responses. J Immunol 182(3): 1526–1532.
15. Alvarez MG, Postan M, Weatherly DB, Albareda MC, Sidney J, et al. (2008) HLA class I-T cell epitopes from trans-Sialidase proteins reveal functionally distinct subsets of CD8+ T cells in chronic chagas disease. PLoS Negl Trop Dis 2(9): e288.
16. Campbell DA, Westenberger SJ, Sturm NR (2004) The determinants of Chagas disease: connecting parasite and host genetics. Curr Mol Med 4(6): 549–562.
17. Fae KC, Drigo SA, Cunha-Neto E, Ianni B, Mady C, et al. (2000) HLA and ß-myosin heavy chain do not influence susceptibility to Chagas' disease cardiomyopathy. Microbe and Infection 2: 745–751.
18. Aida K, Juarez S, Kikuchi M, Gil M, Ayau O, et al. (2000) HLA-B35 and MICA-A5 synergistically enhanced susceptibility to Chagas Heart disease. MHC 7: 63–70.
19. Field SF, Nejentsev S, Walker NM, Howson JM, Godfrey LM, et al. (2008) Sequencing-based genotyping and association analysis of the MICA and MICB genes in type 1 diabetes. Diabetes 57(6): 1753–1756.
20. Mizuki N, Ota M, Kimura M, Ohno S, Ando H, et al. (1997) Triplet repeat polymorphism in the transmembrane region of the MICA gene: a strong association of six GCT repetitions with Behçet disease. Proc Natl Acad Sci U S A 94(4): 1298–1303.
21. Higuchi T, Seki N, Kamizono S, Yamada A, Kimura A, et al. (1998) Polymorphism of the 5′-flanking region of the human tumor necrosis factor (TNF-alpha) gene in Japanese. Tissue Antigens 51: 605–612.
22. Ubalee R, Suzuki F, Kikuchi M, Tasanor O, Wattanagoon Y, et al. (2001) Strong association of a tumor necrosis factor-alpha promoter allele with cerebral malaria in Myanmar. Tissue Antigens 58(6): 407–410.
23. del Puerto R, Nishizawa JE, Kikuchi M, Iihoshi N, Roca Y, et al. (2010) Lineage analysis of circulating *Trypanosoma cruzi* parasites and their association with clinical forms of Chagas disease in Bolivia. PLoS Negl Trop Dis 4(5): e687.
24. Lancaster A, Nelson MP, Single RM, Meyer D, Thomson G (2003) "PyPop: a software framework for population genomics: analyzing large-scale multi-locus genotype data". In: Pacific Symposium on Biocomputing RB. Altman, et al., editor. Singapore vol. 8: 514–525.
25. Freitas JM, Andrade LO, Pires SF, Lima R, Chiari E, et al. (2009) The MHC gene region of murine hosts influences the differential tissue tropism of infecting Trypanosoma cruzi strains. PLoS One 4(4): e5113.
26. Sidney J, Peters B, Frahm N, Brander C, Sette A (2008) HLA class I supertypes: a revised and updated classification. BMC Immunology 9: 1.
27. Martin DL, Weatherly DB, Laucella SA, Cabinian MA, Crim MT, Sullivan S, et al. (2006) CD8+ T-Cell responses to Trypanosoma cruzi are highly focused on strain-variant trans-sialidase epitopes. PLoS Pathog 2(8): e77.
28. Fonseca SG, Moins-Teisserenc H, Clave E, Ianni B, Nunes VL, et al. (2005) Identification of multiple HLA-A*0201-restricted cruzipain and FL-160 CD8+ epitopes recognized by T cells from chronically Trypanosoma cruzi-infected patients. Microbes Infect 7(4): 688–697.
29. Laucella SA, Postan M, Martin D, Hubby Fralish B, Albareda MC, et al. (2004) Frequency of interferon- gamma -producing T cells specific for Trypanosoma cruzi inversely correlates with disease severity in chronic human Chagas disease. J Infect Dis 189(5): 909–918.
30. de Araujo FF, Vitelli-Avelar DM, Teixeira-Carvalho A, RenatoZuquim, Antas P, Assis Silva Gomes J, et al. (2011) Regulatory T Cells Phenotype in Different Clinical Forms of Chagas' Disease. PLoS Negl Trop Dis 5(5): e992.
31. Vitelli-Avelar DM, Sathler-Avelar R, Dias JC, Pascoal VP, Teixeira-Carvalho A, et al. (2005) Chagasic patients with indeterminate clinical form of the disease have high frequencies of circulating CD3+CD16-CD56+ natural killer T cells and CD4+CD25High regulatory T lymphocytes. Scand J Immunol 62(3): 297–308.
32. Lanier LL (2005) NK cell recognition. Annu Rev Immunol 23: 225–274.
33. Vyas Y, Selvakumar A, Steffens U, Dupont B (1998) Multiple transcripts of the killer cell immunoglobulin-like receptor family, KIR3DL1 (NKB1), are expressed by natural killer cells of a single individual. Tissue Antigens 52(6): 510–519.
34. Groh V, Steinle A, Bauer S, Spies T (1998) Recognition of stress-induced MHC molecules by intestinal epithelial gammadelta T cells. Science 279(5357): 1737–1740.
35. Thorsby E (2011) On the future of HLA. Tissue Antigens 78(4): 229–240.
36. Fu XT, Bono CP, Woulfe SL, Swearingen C, Summers NL, et al. (1995) Pocket 4 of the HLA-DR (alpha, beta 1*0401) molecule is a major determinant of T cells recognition of peptide. J Exp Med 81(3): 915–26.
37. Colorado IA, Acquatella H, Catalioti F, Fernandez MT, Layrisse Z (2000) HLA class II DRB1, DQB1, DPB1 polymorphism and cardiomyopathy due to Trypanosoma cruzi chronic infection. Hum Immunol 61(3): 320–325.
38. Dorak MT, Yee LJ, Tang J, Shao W, Lobashevsky ES, et al. (2005) HLA-B, -DRB1/3/4/5, and -DQB1 gene polymorphisms in human immunodeficiency virus-related Kaposi's sarcoma. J Med Virol 76(3): 302–310.
39. Price P, Witt C, Allcock R, Sayer D, Garlepp M, et al. (1999) The genetic basis for the association of the 8.1 ancestral haplotype (A1, B8, DR3) with multiple immunopathological diseases. Immunol Rev 167: 257–274.

Mbanefo et al. BMC Genomics 2012, 13:260
http://www.biomedcentral.com/1471-2164/13/260

BMC Genomics

RESEARCH ARTICLE Open Access

Origin of a novel protein-coding gene family with similar signal sequence in *Schistosoma japonicum*

Evaristus Chibunna Mbanefo[1,6], Yu Chuanxin[2], Mihoko Kikuchi[1], Mohammed Nasir Shuaibu[1], Daniel Boamah[1], Masashi Kirinoki[3], Naoko Hayashi[3], Yuichi Chigusa[3], Yoshio Osada[4], Shinjiro Hamano[5] and Kenji Hirayama[1*]

Abstract

Background: Evolution of novel protein-coding genes is the bedrock of adaptive evolution. Recently, we identified six protein-coding genes with similar signal sequence from *Schistosoma japonicum* egg stage mRNA using signal sequence trap (SST). To find the mechanism underlying the origination of these genes with similar core promoter regions and signal sequence, we adopted an integrated approach utilizing whole genome, transcriptome and proteome database BLAST queries, other bioinformatics tools, and molecular analyses.

Results: Our data, in combination with database analyses showed evidences of expression of these genes both at the mRNA and protein levels exclusively in all developmental stages of *S. japonicum*. The signal sequence motif was identified in 27 distinct *S. japonicum* UniGene entries with multiple mRNA transcripts, and in 34 genome contigs distributed within 18 scaffolds with evidence of genome-wide dispersion. No homolog of these genes or similar domain was found in deposited data from any other organism. We observed preponderance of flanking repetitive elements (REs), albeit partial copies, especially of the *RTE*-like and *Perere* class at either side of the duplication source locus. The role of REs as major mediators of DNA-level recombination leading to dispersive duplication is discussed with evidence from our analyses. We also identified a stepwise pathway towards functional selection in evolving genes by alternative splicing. Equally, the possible transcription models of some protein-coding representatives of the duplicons are presented with evidence of expression *in vitro*.

Conclusion: Our findings contribute to the accumulating evidence of the role of REs in the generation of evolutionary novelties in organisms' genomes.

Keywords: Signal sequence trap, *Schistosoma japonicum*, Repetitive elements, Gene duplication, Secreted proteins, Non-allelic homologous recombination

Background

Evolutionary novelties generated as an upshot of the "nascence" of new protein-coding genes are the bedrock of adaptive evolution and acquisition of novel molecular functions. The ever-growing vast and diverse protein repertoire in organisms can be ascribed to these events, and may explain the increasing heterogeneity among organisms of otherwise common ancestry [1-5]. Since the pioneering definitive treatise on gene duplication by Ohno about four decades ago [6], geneticists and evolutionary biologists have advanced this traditional notion; creating remarkable insights into the composite patterns and underlying mechanisms of genetic innovations. Some of these mechanisms are illustrated in a supplementary figure (Additional file 1). The advent of the genomics era has most importantly armed scientists with a valuable tool to enhance discovery of the rather intriguing mechanisms underlying the "birth" of new genes [5].

Apart from the canonical gene duplication model as proposed by Ohno [6]; extensive studies in various organisms have not only elucidated other models of gene duplication, including "dispersed" duplication in addition to the more definitive "tandem" duplication [7-13]; but has also revealed multiple mechanisms leading to the emergence of new functional genes. These include but not limited to: recombination by exon shuffling or exon "scrambling"

* Correspondence: hiraken@nagasaki-u.ac.jp
[1]Department of Immunogenetics, Institute of Tropical Medicine (NEKKEN), and Global COE Program, Nagasaki University, 1-12-4 Sakamoto, 852-8523, Nagasaki, Japan
Full list of author information is available at the end of the article

BioMed Central
© 2012 Mbanefo et al.; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

[4,14-18]; retrotransposition by retrotransposons yielding intronless chimeric genes [18-25]; transduction of genomic segments by transposable elements by skipping the characteristic weak polyadenylation signal in retrotransposons leading to the mobilization of adjacent genomic sequence; or may involve a repetitive element (RE) mediated DNA level recombination (DLR) by a non-allelic homologous recombination (NAHR) mechanism, in which the REs provide the requisite homologous sequences for the recombination of genomic sequences in a non-allelic manner [7,20,26-30]. Horizontal gene transfer between organisms although infrequent, can give rise to new genes in the recipient organism [31-33]. *De novo* origination of protein coding genes from previously non-coding genomic sequences is a very important mechanism previously underrated, but accumulating data in many organism show that this event occur more often than previously thought [2,3,34-40]. Equally, a new gene can arise from the fusion of two genes [1,3,22] or fission of a "parent" gene [41]. These mechanisms seldom operate singly as they frequently overlap, collaborating in the creation of nascent genes as depicted in the famous origins of *Jingwei* and *Sphinx* in *Drosophila* species [14,19].

Schistosoma japonicum along with *S. mansoni* and *S. haematobium* are the principal schistosome species causing human schistosomiasis. Uncharacteristic of other human invading schistosomes, *S. japonicum* is also able to infect several non-human mammalian hosts. While *S. japonicum* and *S. mansoni* inhabit the periportal veins and cause an intestinal form of the disease, characterized by liver granulomatous fibrosis as a consequence of host immune response to the eggs lodged in the hepatic sinusoids [42,43]; *S. haematobium* causes urinary schistosomiasis at the vesical bladder plexus. Although *S. japonicum* produces similar lesions like *S. mansoni*, the fibrotic lesions and hepatosplenomegaly, the most severe outcome of schistosomiasis, is relatively more frequent and severe in *S. japonicum* [44]. Also, in contrast to *S. mansoni* and *S. haematobium,* acute disease due to *S. japonicum* is common in endemic foci and is associated with severe and persistent manifestations that may rapidly progress the host mediated immunopathogenesis, terminating in a network of fibrotic lesions [45]. Secreted proteins from the parasite ova embolized in the liver of the host are accessible to the host immune cells being located at the host-parasite interface and thus constantly exposed to the host liver tissues. Such interactions play critical role in the initiation and progression of granuloma and fibrosis formation by mediating inflammation [42-45]. Secreted protein candidates thus, possess great potentials for application in interventions aimed at preventing severe hepatic pathogenesis [46,47] among other applications.

Nascent genes confer extra functional capacities for the organisms to confront the challenges of the ever dynamic environment, and may equally, albeit rarely, inflict some functional constraints. In any case, recently evolved characteristics could best be attributed to either: protein family or domains expansion, gene loss events [48], or more likely, evolution of new genes. *S. japonicum* relatively exhibit a higher degree of parasitism and dependence on host derived molecules and signals as inferred from genomic and transcriptomic studies [49-51]; it is able to infect a wide range of hosts, and produces relatively more severe pathogenesis [45]. While these could be attributed to a number of other factors including: selective pressure of parasite-host interactions, the extensive gene loss and protein domain elimination or expansion events observed in its genome and transcriptome [49]; the evolution of novel functional protein coding genes before and after the divergence from other members of the genus *Schistosoma* could account for these extra characteristics.

Here, we report putative evolutionary novel gene family of Asian schistosomes, *S. japonicum* on the premise that no homologs of the genes were found in the genome of its evolutionary close relatives in the genus *Schistosoma,* or in any other organism with a complete sequenced genome. The genes first caught our attention as genes bearing similar or same signal sequence from our previous work that identified some secreted protein coding genes from the eggs of *S. japonicum* using a signal sequence trap (SST) [47]. Given the available tools prior to the publication of the *S. japonicum* genome sequence, we had attributed this observation to some alternative or trans-splicing models. The present analysis was inspired by the availability of the invaluable tool presented by the recently published partially assembled genome of this parasite [49]. We adopted an integrated approach utilizing extensive BLAST queries and other bioinformatics tools, transcription and expression analyses, southern hybridization of genomic DNA and evolutionary analyses. We describe evidence of "genome-wide" dispersed duplication of a protein coding gene locus, which may have arisen recently from previously non-coding genomic sequence. The role of repetitive elements as major mediators of the dispersive duplication is analyzed and discussed. Detailed evidence of the potential transcription models of some protein-coding representatives of the duplicons with similar signal sequence is presented and supported by our observations. Finally, based on the identification of non-coding mRNA transcripts as alternatively spliced variants of protein coding mRNAs, we propose that the new genes could be under significant functional selection.

Results and discussion

Sequence characteristics of a novel protein-coding gene family with similar signal sequences in *S. japonicum*

To identify secreted proteins from the eggs of *S. japonicum*, we previously utilized a signal sequence trap (SST) and isolated at least 15 full length *S. japonicum* egg stage cDNAs encoding secreted or membrane binding proteins [47]. In addition, we observed that six of these genes have same or similar signal sequences (Table 1) from our analyses in [47]. Multiple alignment of the initial SST isolated messenger RNAs (mRNAs) is presented as a supplementary information (Additional file 2), while the multiple alignment of the corresponding protein sequences showing the similar signal peptides is presented in Figure 1 with the phylogenetic tree of the SST identified family members. Given the available tools at the time we made this observation, we had attributed this trend to some alternative splicing or trans-splicing models. Here, we took advantage of the recently characterized and published partial assembly of the genome sequence and transcriptome of *S. japonicum* to unravel the possible underlying mechanisms of signal sequence similarity among SST identified genes. BLASTN search on the whole non-redundant (nr) nucleotide collections and all expressed sequence tags (ESTs) in GenBank including the *S. japonicum* transcriptome using the similar signal sequence as query showed that a total of 181 mRNA sequences and 14 ESTs all belonging to *S. japonicum* bear the similar signal sequence. Based on information in the UniGene database that provides sets of transcript sequences that appear to come from the same transcription locus, these mRNA sequences with similar signal sequence were placed in 27 distinct UniGene entries (Table 2). By further sequence alignments of the returned mRNA sequences and information from UniGene, we grouped the mRNA transcripts according to their gene products and identified at least 7 distinct egg proteins, somula protein, 53 other hypothetical protein sequences and 10 non-coding mRNAs, all bearing the similar signal motif. All protein products of the mRNAs in the public database bearing the similar signal sequence were characteristically short, with one of them containing only 54 amino acids residues. A genome wide BLAST search using the similar signal sequence as query against all whole genome shotgun (WGS) reads also produced hits on 34 *S. japonicum* genome contigs (Table 3) distributed within 18 genome scaffolds (Table 4), thus confirming the existence of such sequences in the genome at multiple loci. These loci were non-redundant and non-overlapping as confirmed from the partially mapped scaffolds of this parasite's genome accessible in GeneDB [52]. For clarity, we restricted further analyses to the initial cDNAs we had identified from our previous study using the SST.

To assess whether some homologs or at least some similar domains exist in other species, BLASTN and BLASTP searches using both the signal sequence and the entire coding sequences of the mRNAs and protein sequences as queries showed that these genes have no homologs in any other organism, but their expression in *S. japonicum* is supported by evidence from transcriptome and proteomic data. A search on several protein domain databases showed that although our candidates were classified in the same protein family with similar domains and assigned to a domain ID (ProDom:PD884968), no related domain or protein family was found in any other organism. The absence of these genes in the genome of *S. mansoni, S. haematobium* and other published genomes cannot possibly be attributed to sequencing gaps or annotation errors since the WGS sequencing approach is considerably reliable [49], and the fact that we adopted a multiple species approach covering the entire available sequenced genomes of all species makes this even more improbable [37,40]. Given the accumulating evidences of *de novo* origin of new genes from previously non-coding DNA sequences [2,34-40], we propose that the coding sequence of these genes may have recently originated *de novo* from previously non-coding DNA sequences in the ancestral forms, and subsequently duplicated and dispersed in the genome. This represents a more plausible interpretation than the improbable alternative hypothesis of concurrent gene deletion or inactivation in multiple ancestral lineages.

Table 1 *SST isolated *S. japonicum* egg cDNAs with similar signal peptide

Gene Products	GenBank cDNA Accession	GenBank Protein Accession	Signal Peptide
SjCP1084	AY570737 (1027 bp)	AAS68242 (271aa)	MRIINLVIISTALLLINLLQTKSQ
SjCP3611	AY570744 (983 bp)	AAS68249 (260aa)	MRIIILGIISTVLLLINLLQTKSQ
SjCP501	AY570753 (1038 bp)	AAS68258 (174aa)	MRIINLVNISTVLLLINLLQTKSQ
SjCP3842	AY570748 (854 bp)	AAS68253 (203aa)	MFKMRIINLVNISTVLLLINLLQTKSR
SjCP400	AY570756 (848 bp)	AAS68261 (124aa)	MFKMRIINLVNISTVLLLINLLQTKSQ
SjCP1531	AY570742 (1037 bp)	AAS68247 (274aa)	MFKVRIINLVNISTVLLLINLLQTKSQ

*SST: Signal Sequence Trap.

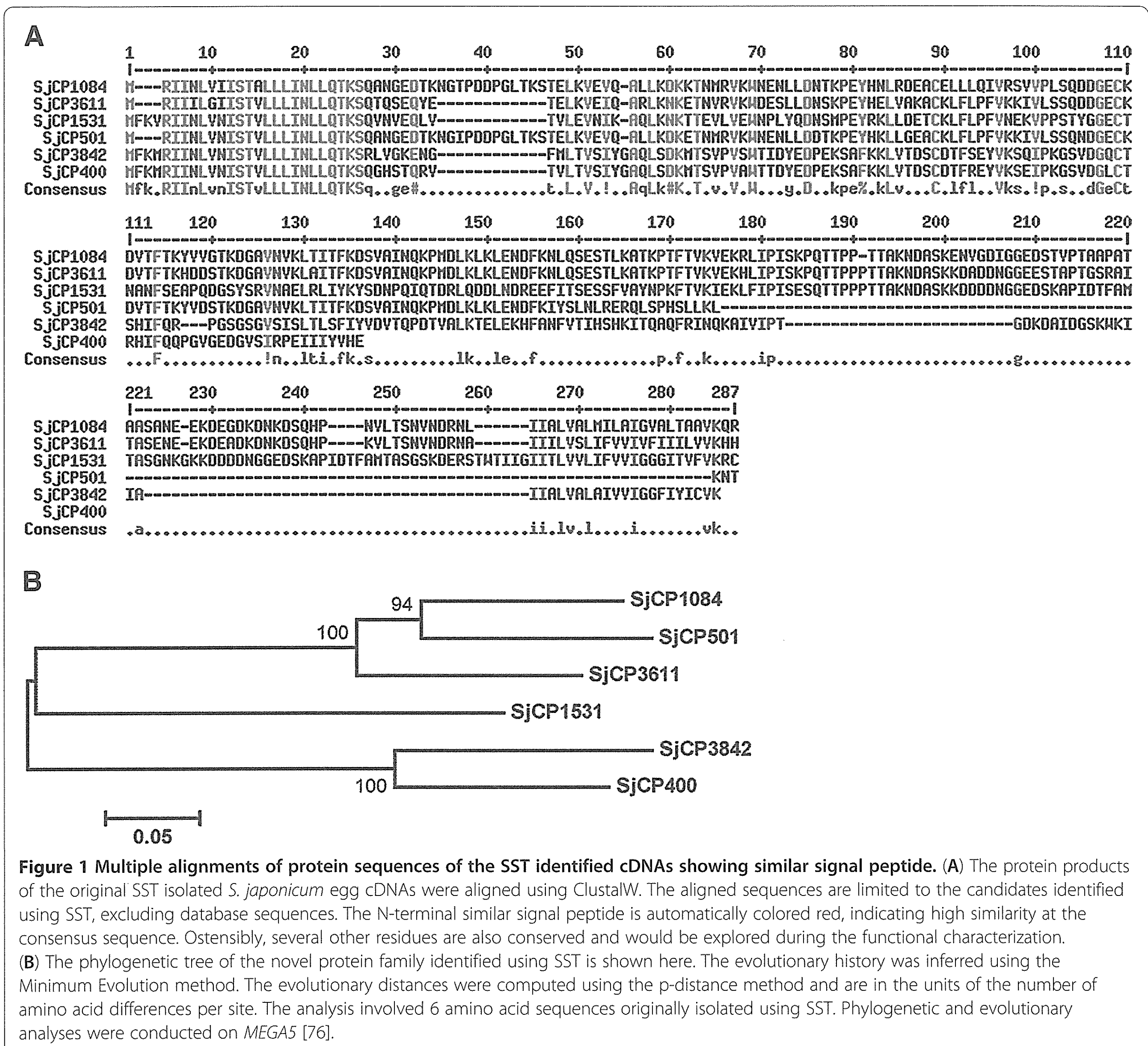

Figure 1 Multiple alignments of protein sequences of the SST identified cDNAs showing similar signal peptide. (**A**) The protein products of the original SST isolated *S. japonicum* egg cDNAs were aligned using ClustalW. The aligned sequences are limited to the candidates identified using SST, excluding database sequences. The N-terminal similar signal peptide is automatically colored red, indicating high similarity at the consensus sequence. Ostensibly, several other residues are also conserved and would be explored during the functional characterization. (**B**) The phylogenetic tree of the novel protein family identified using SST is shown here. The evolutionary history was inferred using the Minimum Evolution method. The evolutionary distances were computed using the p-distance method and are in the units of the number of amino acid differences per site. The analysis involved 6 amino acid sequences originally isolated using SST. Phylogenetic and evolutionary analyses were conducted on *MEGA5* [76].

Species and strain specific expression

To further exclude the possibility of false negative observations, we assessed the presence of the gene loci among different species and strains of *Schistosoma in vitro* using southern blots. This genomic locus and its duplicons was found to be exclusively present in all the strains of *S. japonicum* using southern hybridization experiment utilizing genomic DNA samples of different strains of *S. japonicum* (Japanese, Chinese and Philippines), and other species of *Schistosoma* including *S. mansoni*, *S. haematobium* and *S. mekongi* (Figure 2), covering all the major clades in the genus. The result of southern hybridization using 462 base-pair digoxigenin labeled hybridization probe containing the similar signal sequence and designed to be specific to the gene loci under consideration showed that this genomic sequence was not found in any other species of *Schistosoma* except in all strains of *S. japonicum* analyzed. Several bands representing the duplicated loci are apparent in the hybridized blots (Figure 2). The analyzed samples is composed of representatives of the species complexes of this genus and further provide insight into the inter-species, intra-species and intra-strain variations that may exist among the members of the genus *Schistosoma*. In line with the widely accepted Asian origin hypothesis deduced from the evolutionary biogeography of this genus as inferred from evidences at the morphological, karyotype and molecular levels [53], it is highly plausible that this genomic sequence has recently evolved exclusively within the *S. japonicum* complex long after the divergence of the ancestors of the African species and other re-invading Asian species with origin from Africa [53,54]. The fact

that this gene locus was completely lacking in other Asian species like *S. mekongi* of common ancestry with *S. japonicum* even throws more light on the most probable evolution of these other Asian species which are thought to have evolved from same ancestor or as descendants of *S. japonicum* based on mitochondrial gene arrangement [55]. Either *S. japonicum* and other Asian species in the *S. japonicum* group evolved independently

Table 2 *UniGene entries for *S. japonicum* mRNAs and ESTs bearing the similar signal sequence (n = 195)

UniGene Name	UniGene ID (UID)	Set of likely mRNA transcripts (GenBank)	Gene Products (Database annotation)
Sja.1526	1476162	AY814448, BU780442[est]	Egg protein SjCP3611
Sja.1611	1476247	FN317637, BU772954[est]	Hypothetical protein
Sja.1628	1476264	AY570742, FN320556, FN320555, FN320553, FN320552, FN320551, FN320550, FN320549	Egg protein SjCP1531
Sja.1676	1476312	AY570748[SST], AY223245, AY222916, AY813542, EF127834, EF140742, FN323799, FN323800, FN323801, FN323803, FN323793, FN323792, FN323791, FN323790, FN323788, FN323785, FN323782, FN323781, FN323779, FN323778, FN323777, FN323776, FN323773, FN323772, FN323771, FN323770, FN323769, FN323768, FN323767, FN323766, FN323765, FN323764, FN323763, FN323762, BU772060[est], BU766145[est], CX862012[est]	Egg protein SjCP3842
Sja.2063	1476798	FN321064, FN321061	Egg protein SjCP1084
Sja.2065	1476800	AY570753[SST], AY570744[SST], AY814685, FN327232, FN327137, FN318042, FN321065, FN321060, FN321059, FN321058, FN321057, FN321056, FN321055, FN329815[nc], BU768978[est], BU780021[est]	Egg protein SjCP3611, Egg protein SjCP501, Hypothetical proteins
Sja.2070	1476805	AY599749[SST]	Egg protein SjCP1731
Sja.5326	2034920	FN326953, FN330298[nc]	Hypothetical protein
Sja.9771	2493712	AY570756[SST], FN327121, FN327254, FN327253, FN327241, FN327233, FN327229, FN327224, FN327222, FN327216, FN327196, FN327185, FN327163, FN327158, FN327154, FN327129, FN327125, FN327115, FN327089, FN327083, FN327073, FN327057, FN327050, FN327049, FN327045, FN327042, FN327035, FN327022, FN327018, FN327014, FN327000, FN326998, FN326978, FN326973, FN326961, FN326960, FN326959, FN326930, FN326905, FN326883, FN326882, FN326881, FN326859, FN326857, FN326852, FN326851, FN326841, FN326831, FN326829, FN326822, FN326808, FN326801, FN326790, FN326770, FN326740, FN330540[nc]	Egg protein SjCP400, Somula protein
Sja.11083	2671933	AY915467, FN327219, FN327063, FN326828, FN326826, FN323794, FN323797, FN323798, FN323802, FN323789, FN323787, FN323786, FN323784, FN323783, FN323780, FN323774, FN323761, FN323760, FN323759, FN323758, FN323757, FN320521, FN320520, FN320519, FN320518, FN320517, FN320516, FN320515, FN320513	Egg protein SjCP3842, Hypothetical protein
Sja.11325	2672175	AY813755, FN320057, FN320056, FN320514, FN329566[nc], BU768160[est], BU774105[est], BU770186[est], BU779051[est]	Egg protein SjCP3842, Hypothetical protein
Sja.11840	2895838	FN327242, FN327131, FN327087, FN326854, BU776301[est]	Hypothetical protein
Sja.11891	2895889	AY813975, FN329814[nc], BU769048[est]	Egg protein SjCP1084
Sja.13298	3987026	FN320059	Hypothetical protein
Sja.13324	3987052	AY570737[SST], FN328299[nc]	Egg protein SjCP1084
Sja.13882	3987610	FN330716[nc]	None
Sja.13956	3987684	FN330422[nc]	None
Sja.14071	3987799	FN329677[nc]	None
Sja.14095	3987823	FN329269[nc]	None
Sja.14561	3988289	FN327139, FN323795, FN323796, FN323775	Egg protein SjCP3842
Sja.14562	3988290	FN327130, FN326955, FN326901	Egg protein SjCP1084
Sja.14565	3988293	FN327099	Egg protein SjCP1084
Sja.14614	3988342	FN320058	Hypothetical protein
Sja.14627	3988355	FN319007	Hypothetical protein
Sja.14941	3988669	FN320554	Hypothetical protein
Sja.15036	5233761	FN326786, FN318043, CX861530[est]	Hypothetical protein
Sja.15108	5233833	AY810465, FN321062	Hypothetical protein

* UniGene is a database of sets of transcript sequences that appear to come from the same transcription locus.
The original set of cDNAs we earlier identified using signal sequence trap bear the superscript tag ([SST]).
Transcripts with tags ([est]) and ([nc]) are expressed sequence tags (ESTs) and non-coding mRNAs respectively.

from a common ancestor, or the evolution of this locus and the subsequent dispersed duplication occurred recently after the other Asian forms have diverged (see phylogenetic relationship in Additional file 3). An alternative explanation is that the gene was not fixed or was deleted from the genome of the other Asian and African species. Since the last hypothesis is highly unlikely, we concluded that our observation was a product of a newly evolved gene locus possibly from mutations or modifications on a previously non-coding sequence in the ancestral forms, which was subsequently severally amplified and dispersed in the genome of *S. japonicum* after all other species of the genus had diverged. Furthermore, a close look at the banding pattern of the restriction digested genomic DNA of different strains of *S. japonicum* as observed in the southern blotting result revealed that possible intra-species and intra-strain genetic variations could exist among the members of the species complex (Figure 2). Whether the *S. japonicum* complex (Japanese, Chinese, Philippines and animal infecting Formosa strains) is made up of four geographical strains, four subspecies or four independent biological species remains contentious. Be that as it may, this presents an interesting subject for further research and could be further explored using a wider array of isolates from different regions.

Nevertheless, while it is completely normal to verify this exclusive evolution and dispersed duplication hypothesis by confirming the physical localization of the gene loci in the genome and chromosomes by performing synteny analysis, we are unable to achieve this because we do not have access to a fully mapped chromosome information of the genome of *S. japonicum*. However, the distribution of the contigs and scaffolds bearing the similar signal sequence apparently suggests a dispersed distribution. To confirm this hypothesis and to exclude the possibility of overlapping among the loci, we generated the restriction map of six of the genome scaffolds bearing duplicated loci based on the information on the genome map, performed southern hybridization using restriction endonuclease digested genomic DNA from *S. japonicum* species and strains; and were able to match the expected probe binding fragment sizes with the observed bands on the hybridization blots (Additional file 4). Also an ancestral homolog is required for synteny analysis, however, we could not find a homolog in *S. mansoni*, another member of the genus with sequenced genome; and the genome of other more closely related species like *S. mekongi* and *S. malayensis* are not yet sequenced. Unless new evidences emerge from future updates in the sequenced genomes, we hold true that these genes have newly evolved, probably from modifications on previously non-coding ancestral DNA sequences and subsequently disperse duplicated. As opined in previous studies, the short length of our identified genes is an expected property of nascent genes because of improbability of evolution of long open reading frames (ORFs) and the complexity of intron splicing signal [38]. We expect these novel genes to be of functional significance since new genes tend to display accelerated sequence and structural changes towards neo-functionalization [1], and most newly characterized genes from other species have been shown to be characteristically functional [35,56]. Other workers showed that the common pathway for *de novo* protein-coding gene evolution involves a piece of DNA sequence to be transcribed via recruitment of all transcription core promoters, other elements and machines; followed by the acquisition of a translatable ORF through mutations or other sequence alteration mechanisms [2,35]. Together, our findings support the presence of these intrinsic features of novel genes in the identified candidates, including the gradual model of novel protein coding gene origination.

Evidence of dispersed duplication from a source gene locus

The mechanisms behind dispersed duplication could be hidden within the DNA sequences of the duplicates or the adjacent flanking genomic sequences. In line with this, we explored the DNA sequences of the gene loci and the surrounding genomic sequences to identify possible mechanisms underlying dispersed duplication proposed in our hypothesis. A genome-wide BLAST search against WGS reads using the similar signal sequence as query returned 34 contigs of varying lengths and degrees of degradation (Table 3). By manually tracing these 34 contigs to the genome scaffolds, we found that they were distributed within 18 scaffolds (Table 4), apparently widely dispersed in the genome of *S. japonicum* as inferred from the genome map. To explore the mechanism of such dispersed duplication of a genomic sequence, a comparative analysis involving a parent gene in an ancestral species is often required. However, since we were unable to find any parental homolog in the available genome data and proteomes, and because gene duplication produces a diverse set of progeny loci with varied degrees of homology to an ancestral source locus when it exists [9], we performed a comparative sequence analysis on the 34 contigs as representatives of the gene loci. The result revealed a particular prominent contig in the *S. japonicum* WGS reads [GenBank: CABF01020060], the longest of the set of "duplicons" (43.7 kb), which significantly encompassed the length of the other contigs (Figure 3). CABF01020060 was therefore putatively selected as the duplication 'source locus' and utilized as such for most of the analyses performed in this study.