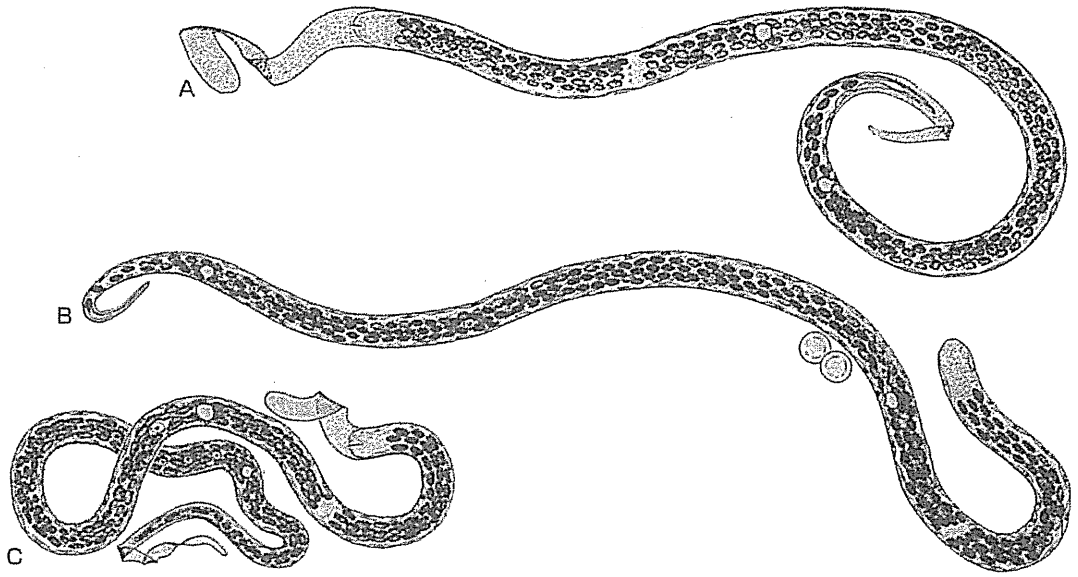

図 56-10 A：血液中のバンクロフト糸状虫のミクロフィラリア．尾端には核がないことに着目(225-300×8-10 μm)．B：皮膚内の回旋糸状虫のミクロフィラリア(血液中にはほとんどいない)(300-350×5-9 μm)(円は赤血球の大きさを代表)．C：血中のロア糸状虫 *Loa loa* のミクロフィラリア．末端まで核が存在することに着目(250-300×6-9 μm)．

病原性と臨床所見

リンパ節にいる成虫は炎症を惹起し，それが最終的にはリンパ管を遮り浮腫を引き起こす．著名な下肢の浮腫は**象皮病** elephantiasis とよばれる．ミクロフィラリアは何ら症状を引き起こさない．

感染初期は無症状である．後に，発熱，リンパ炎や蜂巣炎が続く．次第にリンパ流の通過障害が下肢や生殖器，特に陰嚢の浮腫や線維化に繋がる．象皮病は主として長い間繰り返し感染した患者にみられる．ただ一度だけ感染した旅行者が典型的な象皮病になることはない．

ボルバキア属 *Wolbachia* はリケッチアのような細菌であり，バンクロフト糸状虫や回旋糸状虫などのフィラリア細胞内に見出される．ボルバキアはエンドドキシン様分子を放出し，それがバンクロフト糸状虫や回旋糸状虫の病原性に寄与しているらしい．ドキシサイクリンはボルバキアを死滅させる抗生物質であるが，この薬剤が線虫感染によって惹起される炎症を減弱させるという事実も，ボルバキアが病原性に関係しているという説を支持している．

熱帯性肺好酸球症は夜間の顕著な咳嗽や息切れとして特徴づけられる．これらの症状は肺内ミクロフィラリアに反応して惹起される高 IgE 血症と好酸球増多症で特徴付けられる即時型過敏反応による．

疫 学

この病気はアフリカやアジア，南米といった熱帯地方に蔓延する．ベクターであるカの種は場所によって異なる．2～3億人が感染している．

検査診断

夜間に患者から得られた厚層塗沫標本の中にミクロフィラリアを見出すことによる(図 56-11)．血清学的検査は有用ではない．

治 療

ジエチルカルバマジンはミクロフィラリアに対してのみ有効である．成虫に有効な薬剤はない．

予 防

殺虫剤の使用によるカのコントロール，カの吸血を防ぐ衣類の着用，ベットネットや忌避剤の使用などが予防法となる．

回旋糸状虫

疾 患

回旋糸状虫 *Onchocerca volvulus* がオンコセルカ症(河川盲目症)の原因である．

重要な特徴

雌のブユ female blackfly ブユ属 *Simulium* が吸血し，刺し傷の上に幼虫を残したときにヒトは感染する．幼虫は傷口から皮膚に侵入し，皮下に移行し，通常は**真皮内結節** dermal nodule の中で成虫へと分化する．雌はミクロフィラリアを産み出す(図 56-10B)．ミクロフィラリアはブユの吸血とともにブユに取り込まれ，感染型幼虫へと分化し生活環を完了させる．ヒトは唯一の固有宿主である．

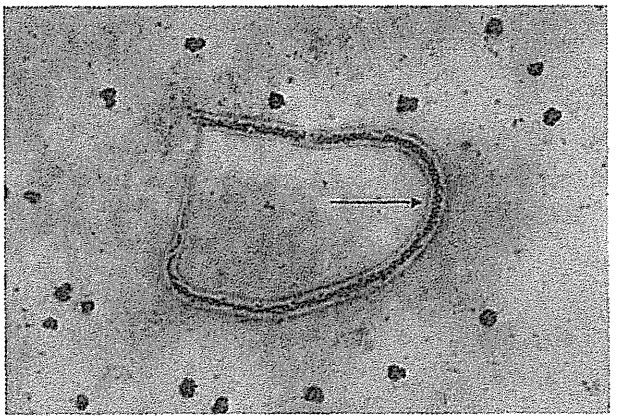

図 56-11 バンクロフト糸状虫の血中ミクロフィラリア．矢印は血中のミクロフィラリアを指す．(Dr. M. Melvin, Public Health Image Library, CDC. より許可を得て転載)

病原性と臨床所見

皮下組織では成虫のタンパクに反応して炎症が惹起され，掻痒感のある皮疹や結節ができる．ミクロフィラリアは皮下組織を移行し，究極的には眼に集中する[*1]．そこで失明に繋がる病態が形成される．皮下の弾性線維の消失がしわの多い皮膚の原因となり，それが鼠径部にみられる場合には「**鼠径部皮膚下垂** hanging groin」とよばれる．皮膚が厚くなり，剥がれ，乾燥して激しい痒みを伴うのは「**トカゲ皮膚** lizard skin」とよばれる皮膚炎の特徴である．

オンコセルカ症の病原性発現におけるボルバキアの役割はバンクロフト糸状虫の項に記載済みである．

疫 学

アフリカと中米で数百万人のヒトが罹患している．この病気は失明の主要な原因の1つである．ブユが河川で成長し，そのような河川の側に住むヒトが罹患するので**河川盲目症** river blindness とよばれる．感染の浸淫地における感染率はしばしば80%を超える．

検査診断

生検皮膚中にミクロフィラリアを同定することによる(図 56-10B)．本ミクロフィラリアは血液内を循環しないので，血液を調べてもミクロフィラリアは検出できない．好酸球増多症はしばしばみられる．血清学的検査は有用ではない．

治 療

イベルメクチンはミクロフィラリアに対しては有効であるが，成虫には無効である．スラミンは成虫を殺すことができるが，きわめて毒性が高く，眼病変で患う人に特に使用される．皮下結節は外科的に除去できるが，新しい皮下結節ができるので，浸淫地における外科的処置による寛解はあまりない．

予 防

殺虫剤の使用などブユのコントロールが予防法となる．イベルメクチンも発症予防に繋がる．

ロア糸状虫

疾 患

ロア糸状虫 *Loa loa* が**ロア糸状虫症** loiasis の原因である．

重要な特徴

メクラアブ deer fly メクラアブ属 *Chrysops* が吸血し，刺し傷の上に幼虫を残したときにヒトは感染する．幼虫は傷口から皮膚に侵入し，体内を彷徨い，成虫へと分化する．雌はミクロフィラリアを産み出す(図 56-10C)．ミクロフィラリアは特に日中に血液中に入り，メクラアブの吸血時に取り込まれ，感染型幼虫へと分化し，次のヒトに感染して生活環を維持する．

病原性と臨床所見

ミクロフィラリアや成虫に対しての炎症反応は惹起されないが，過敏性反応は一過性で限局した非紅斑性の皮下浮腫の原因となる(カラバル腫張 Calabar swellings)．成虫が**眼球結膜を泳ぎ渡る**様子を発見することはもっとも衝撃的であり，無害ではあるもののひどく当惑させられる．

疫 学

この疾患はベクターであるメクラアブの分布する熱帯の中央および西アフリカにのみ存在する．

検査診断

診断は血液塗抹標本中にミクロフィラリアを同定することによる(図 56-10C)．有用な血清学的検査はない．

治 療

ジエチルカルバマジンはミクロフィラリアを殺し，成虫も殺すと考えられる．眼内の成虫は外科的に取り出す必要がある．

*1(訳注)：眼に集中するのではなく「全身の皮膚に分布するが，眼にも分布する」．

予 防

殺虫剤によるメクラアブのコントロールが予防法となる.

メジナ虫

疾 患

メジナ虫 *Dracunculus medinensis* がメジナ虫症 dracunculiasis の原因である.

重要な特徴

ヒトは**飲料水中**に混入し体内に感染型幼虫を有する小さな**ケンミジンコ** crustaceans **を飲み込むと**感染する. 幼虫は小腸で放出され, 体内を移行し, 成虫へと成長する. 1メートルにもなる成虫の雌は皮膚に潰瘍をつくり, 動く幼虫を水中に産み落とす. ケンミジンコは幼虫を食べ, 幼虫はケンミジンコの中で脱皮して感染型幼虫となる. 水の中に生きるこれらケンミジンコがヒトに飲み込まれると生活環は一周する.

病原性と臨床所見

雌成虫は炎症を惹起する物質を産生し, 通常は下腿に焼け付くような皮膚潰瘍を生じる. 丘疹は**焼けるように痛痒く**, 潰瘍は2次感染に繋がる. 通常診断は臨床的に**皮膚の潰瘍部位に虫**を検出することによる.

疫 学

この疾患は熱帯のアフリカ, 中東, インドの広範な地域に分布し, 何百万人もの人が感染している[*2].

検査診断

検査室での診断は必要ない.

治 療

昔ながらの治療法は数日かけて棒状のものに成虫を巻き付けて徐々に引き抜くことである. チアベンダゾールやメトロニダゾールの服用によって虫が巻き取りやすくなる.

予 防

飲料水の濾過や煮沸が予防法となる.

[*2](訳注):国際社会と地域社会の努力により, 1986年に350万人と推定された患者数は2009年には3200人にまで減じた.

幼虫移行症の原因となる線虫

イヌ回虫, ネコ回虫

疾 患

イヌ回虫 *Toxocara canis* は内臓幼虫移行症のおもな原因の1つである. ネコ回虫 *Toxocara cati* やいくつかの関連線虫もまたこの病態の原因となる.

重要な特徴

イヌ回虫の固有宿主はイヌである. イヌ回虫の雌成虫はイヌの腸の中で産卵し, 虫卵を含む糞便は土壌の上に落とされる. ヒトが虫卵が混入した土を口に運ぶと, 虫卵は小腸で孵化し, 幼虫はいくつもの臓器, 特に肝臓, 脳, 眼へ移行する. 幼虫は最終的には被嚢されて死滅する. ヒトの中で生活環が完結することはない. ヒトは偶発的に感染するデッド・エンド(行き止まり)宿主である.

病原性と臨床所見

幼虫に対する遅延型過敏症の結果として死んだ幼虫の周囲に肉芽腫形成が認められる. もっとも重篤な臨床像は網膜を巻き込んで起こる失明である. 熱発, 肝腫大や好酸球増多症がよくみられる.

疫 学

幼い子供は虫卵を含んだ土を口に入れやすくもっとも感染しやすい. 米国ではイヌ回虫はイヌによくみられる寄生虫である.

検査診断

血清学的な検査がよく行われるが, 確定診断は組織中に幼虫を検出することによる. 高 IgE 血症や好酸球増多症は補助診断となる.

治 療

アルベンダゾールやメベンダゾールが使用されるが, きちんと効果が証明された治療法はない. 多くの患者は治療することなく回復する.

予 防

イヌの駆虫や子供が土を口に運ぶことを防ぐことなどが予防法となる.

イヌ鉤虫，ネコ鉤虫

皮膚幼虫移行症 cutaneous larva migrans はイヌ鉤虫 *Ancylostoma caninum* やネコ鉤虫 *Ancylostoma braziliensis* の感染型幼虫によっても引き起こされる．それら線虫はヒト体内で生活環を完遂することはできない．幼虫は皮膚に浸入し，**皮下組織を移行し**，炎症反応を引き起こす．「**皮膚爬行症いん線発疹** creeping eruption」の部位は特に痒みが強い．本症は米国では南部に分布し，感染性の土に接する機会の多い子供や建築関係者に多い．診断に検査はあまり機能せず，臨床的になされることが多い．チアベンダゾールの経口もしくは局所的な使用が有効である．

広東住血線虫

ラットの肺に寄生する線虫である広東住血線虫 *Angiostrongylus cantonensis* は好酸球性髄膜炎を引き起こす．この髄膜炎は脊髄液や血液中に多くの好酸球が認められることを特徴とする．通常，少なくとも白血球の 10%以上に好酸球を認める．幼虫は不完全にしか調理していないカニやエビや貝などとともに摂取される．この感染はアジアの国々でもっともよくみられる．診断は臨床の場でなされることが多いが，ときに検査室で脊髄液の中に幼虫を検出することがある．治療法はない．ほとんどの患者は後遺症もなく自然治癒する．

好酸球性髄膜炎はそのほか 2 つの線虫によっても引き起こされる．ネコやイヌの胃壁に寄生する線虫である有棘顎口虫 *Gnathostoma spingerum* は不完全調理されたサカナを食べて感染し，アライグマの回虫である *Baylisascaris procyonis* はたまたまアライグマの糞便が口から入ったときに感染する．これらの病原体は住血線虫より重篤な症状を引き起こし，ときに致死的である．有棘顎口虫にはアルベンダゾールが有効であるが，アライグマ回虫に有効な治療法はない．

アニサキス

アニサキス症は線虫アニサキスシンプレクス *Anisakis simplex* の幼虫によって引き起こされる．成虫はクジラやイルカ，アザラシなどの海産哺乳類の腸管に寄生する．成虫が産卵した卵はオキアミ類に食べられ，感染オキアミはサケ，サバ，ニシンなどに食べられる．ヒトはこれら**海産物を生食する際に幼虫を摂取し**，幼虫は胃や腸の粘膜に侵入する．胃腸炎，腹痛，好酸球増多症，便潜血などが典型的な症状である．急性期の症状は虫垂炎に酷似しており，慢性期の症状は消化管腫瘍に似ている．

米国でのほとんどのケースは日本食料理店で寿司や刺身(特にサケやフエダイ)の摂食に起因する．診断は典型的には内視鏡や腹腔鏡でなされる．微生物学的もしくは血清学的検査は診断には役に立たない．有効な薬はない．外科的に虫体を除去することが不可欠である．海産物を適切に調理すること，食べる前に 24 時間以上冷凍することが予防に繋がる．

アニサキス科に属するほかの線虫に，シュードテラノバデセピエンス *Pseudoterranova decepiens* があり，この幼虫は非侵襲性のアニサキス症を引き起こす．この幼虫は不完全調理のサカナを食べて感染し，嘔吐や腹痛を引き起こす．診断は幼虫を消化管や吐瀉物の中に見出すことによる．薬はない．幼虫は内視鏡の間に取り除くことができる．

まとめ

本章で解説されているまとめは，529 ページを参照．必要最低限の事項が簡潔に要約されている．

練習問題：USMLE(米国医師資格試験)と課程試験

本章で論じられた内容に関する問題は，第 12 部の米国医師資格試験練習問題を参照(578 ページ)．もしくは第 13 部の米国医師資格試験実践的練習問題を参照(603 ページ)．

MAJOR ARTICLE

Entamoeba moshkovskii Is Associated With Diarrhea in Infants and Causes Diarrhea and Colitis in Mice

Chikako Shimokawa,[1] Mamun Kabir,[3] Mami Taniuchi,[4] Dinesh Mondal,[3] Seiki Kobayashi,[5] Ibne Karim M. Ali,[4] Shihab U. Sobuz,[3] Masachika Senba,[2] Eric Houpt,[4] Rashidul Haque,[3] William A. Petri Jr,[4] and Shinjiro Hamano[1,6]

[1]Department of Parasitology, [2]Department of Pathology, Institute of Tropical Medicine (NEKKEN), Nagasaki University, Nagasaki, Japan; [3]Parasitology Laboratory, Laboratory Sciences Division, ICDDR, B, Dhaka, Bangladesh; [4]Division of Infectious Diseases and International Health, University of Virginia, Charlottesville; [5]Department of Tropical Medicine and Parasitology, Keio University School of Medicine, Shinjuku, Tokyo, Japan; and [6]Global COE program, Nagasaki University, Japan

Background. *Entamoeba moshkovskii* is prevalent in developing countries and morphologically indistinguishable from pathogenic *Entamoeba histolytica* and nonpathogenic *Entamoeba dispar*. It is not known if *E. moshkovskii* is pathogenic.

Methods. Mice were intracecally challenged with the trophozoites of each *Entamoeba* spp. to test the ability to cause diarrhea, and infants in Bangladesh were prospectively observed to see if newly acquired *E. moshkovskii* infection was associated with diarrhea.

Results. *E. moshkovskii* and *E. histolytica* caused diarrhea and weight loss in susceptible mice. *E. dispar* infected none of the mouse strains tested. In Mirpur, Dhaka, Bangladesh, *E. moshkovskii*, *E. histolytica*, and *E. dispar* were identified in 42 (2.95%), 66 (4.63%), and 5 (0.35%), respectively, of 1426 diarrheal episodes in 385 children followed prospectively from birth to one year of age. Diarrhea occurred temporally with acquisition of a new *E. moshkovskii* infection: in the 2 months preceding *E. moshkvskii*-associated diarrhea, 86% (36 of 42) of monthly surveillance stool samples were negative for *E. moshkovskii*.

Conclusions. *E. moshkovskii* was found to be pathogenic in mice. In children, the acquisition of *E. moshkovskii* infection was associated with diarrhea. These data are consistent with *E. moshkovskii* causing disease, indicating that it is important to reexamine its pathogenicity.

Entamoeba histolytica causes extensive mortality and morbidity worldwide through diarrheal disease and abscess formation in parenchymal tissues such as liver, lung, and brain. In contrast, other amoebae that infect humans include *Entamoeba dispar, Entamoeba moshkovskii, Entamoeba coli, Entamoeba hartmanni, and Endolimax nana*, which have been considered nonpathogenic commensals of the human gut [1–3]. *Dientamoeba fragilis* and *Entamoeba polecki* have been associated with diarrhea and *Entamoeba gingivalis* with periodontal disease [4, 5].

E. moshkovskii is genetically related to *E. histolytica* and *E. dispar* and is microscopically indistinguishable from them in its cyst and trophozoite forms [6]. This species of *Entamoeba* was first identified in sewage in Moscow by Tshalaria in 1941 [7] and was initially thought to be a free-living common protozoan species in anoxic sediments and in environments such as brackish coastal pools. The first human isolate was obtained from a resident of Laredo, Texas, who suffered from diarrhea, weight loss, and epigastric pain in 1961 [8]. This finding would seem to suggest and/or support that *E. moshkovskii* can be pathogenic. At first, this isolate was named *E. histolytica* Laredo strain and shared biological features with *E. moshkovskii*.

Received 22 October 2011; accepted 20 February 2012; electronically published 21 June 2012.

Presented in part: Part of the information has been presented at the ASTMH 60th Annual Meeting, December 2011, Philadelphia, at the 45th annual Japan-US joint conference on parasitic diseases, January 2011, Tokyo, Japan, and at the 2nd international conference on climate change and neglected tropical diseases, September 2010, Dhaka, Bangladesh.

Correspondence: Shinjiro Hamano, MD, PhD, Department of Parasitology, Institute of Tropical Medicine, Nagasaki University, Sakamoto 1-12-4, Nagasaki 852-8523, Japan (shinjiro@nagasaki-u.ac.jp).

The Journal of Infectious Diseases 2012;206:744–51

DOI: 10.1093/infdis/jis414

Both the Laredo strain and *E. moshkovskii* grow at room temperature and were resistant to osmotic shock and to drugs used in the chemotherapy of amoebiasis such as emetine [9]. Subsequent molecular studies revealed that *E. histolytica* Laredo is identical with *E. moshkovskii* [10].

E. moshkovskii is a common *Entamoeba* infection in humans in some settings. It is composed of anywhere from as little as 1% to as high as 50% of the *E. histolytica/E. dispar/E. moshkovskii* complex parasites detected in fecal samples in limited studies from Australia, Bangladesh, India, Iran, Tanzania, and Turkey [6, 11–16]. These studies for the most part tested stool samples submitted to clinical microbiology laboratories from patients with gastrointestinal symptoms, suggesting that *E. moshkovskii* could cause disease. However, in HIV-1–infected individuals in northern Tanzania, *E. moshkovskii* was not associated with enteric symptoms nor immune status [17]. Thus, the ability of *E. moshkovskii* to cause disease in humans remains unclear.

Here we tested the ability of *E. moshkovskii* to cause colitis and diarrhea in a murine model system, in which intracaecal inoculation with *E. histolytica* trophozoites into CBA/J, C3H/HeN, and C3H/HeJ mice leads to amebic colitis [18–20]. In addition, we tested in a longitudinal study of children in Bangladesh not only if *E. moshkovskii* was present in stool samples from infants with diarrhea, but whether the *E. moshkovskii* infection was newly acquired at the time of the diarrheal illness.

MATERIALS AND METHODS

Mice

Male CBA/J, C57BL6/J, BALB/c, C3H/HeN, and C3H/HeJ mice were purchased from the Jackson Laboratory. Animals were maintained under specific pathogen-free conditions at Animal Research Center for Tropical Infectious Diseases, Nagasaki University, and were challenged when they were 5–8 weeks old.

Cultivation of *Entamoeba* spp.

Trophozoites of the *E. moshkovskii* Laredo strain were a gift from Dr Seiki Kobayashi, Keio University, School of Medicine (originally from the late Professor Louis S. Diamond, National Institutes of Health, Bethesda, Maryland). Trophozoites of *E. histolytica,* originally laboratory strain HM1:IMSS (American Type Culture Collection, Manassas, Virginia), were from Professor Eric Houpt, University of Virginia, which were sequentially passaged in vivo through the mouse cecum [18]. Cecal contents were cultured at 25° and 37°C, respectively, in BIS-33 medium supplemented with heat-inactivated 10% adult bovine serum, 25 U/mL penicillin, and 25 mg/mL streptomycin [21]. Trophozoites of *E. dispar* AS16IR were also provided by Dr Seiki Kobayashi and cultured in YIMDHA-S media at 37°C. Trophozoites under log phase of growth were used in the experiments.

Intracecal Inoculation of *Entamoeba* spp.

Trophozoites were harvested from culture tubes of *E. histolytica* HM1:IMSS, *E. moshkovskii* Laredo and *E. dispar* AS16IR strains by incubating the tubes on ice for 5–10 minutes. Then, the trophozites were collected, and the number of trophozoites was determined. We anesthetized mice with domitor (medetomidine hydrochloride: 0.1 mg/kg) and dormicum (midazolam: 0.1 mg/kg), shaved their abdomens to incise the skin and exteriorized each cecum from the peritoneum, and injected 150 μL of 1×10^6 each trophozoites into the proximal, middle, and apical sites of cecum. Then the cecum was blotted and the peritoneum and the skin were sutured. Mice were kept on warming blankets at 37°C throughout. Survival rates were ≥85% in all strains. The study was approved by the animal ethical review board of Nagasaki University.

PCR Amplification for Diagnosis of *Entamoeba* spp. Infection in Mice

For isolation of *Entamoeba* DNA from mouse stools, QIAamp DNA Stool kits (QIAGEN, Valencia, California) were used according to manufacturer's instructions. The primer sequences used for polymerase chain reaction (PCR) were described elsewhere [22].

Pathology of Murine Amoebic Colitis

At the indicated days after intracecal challenge, mice were killed, the ceca fixed in phosphate buffered 10% formalin, and then cut into 4–6 equal cross-sections and embedded in paraffin, and 4 μm slides were stained with H&E.

Child Study Area and Population

The study was conducted in Mirpur, an urban slum in Dhaka. Infants were enrolled in the first week after birth and followed until one year of age, beginning in January 2008. Field research assistants (FRAs) visited each study house every other day and collected information related to child morbidity, especially for diarrheal illness, through a structured questionnaire. If the FRA found any child with an acute illness, then she referred the child to the study clinic for further management by the medical officer. Parents or guardians were also encouraged to visit the study clinic for medical assistance if the study child became sick. FRAs collected nondiarrheal monthly stool specimens as well as diarrheal stool specimens from the home or in the study field clinic. All stool specimens were transported from the field to the clinic using a cold box. In the field clinic an aliquot of the diarrheal stool specimens was placed into Carry-Blair medium. All specimens were transported from the field clinic to the ICDDR,B Parasitology laboratory within 3 hours of collection, with a cold chain maintained. Diarrhea was defined as having ≥3 unformed or abnormal stools (as per the mother's perception) in a 24-hour period. A diarrheal episode was defined as being separated from another episode by at least 3 diarrhea-free days.

The study was approved by the Institutional Review Board of the University of Virginia, and the Ethical Review Committee of the International Centre for Diarrhoeal Disease Research, Bangladesh. Informed written consent was obtained from the parents or guardians for the participation of their child in the study.

Detection of Enteropathogens

Stool samples were cultured for enteric pathogens including *Vibrio cholerae* O1/O139, *Salmonella* spp., *Shigella* spp., and *Campylobacter jejuni*. Enzyme-linked immunosorbent assay (ELISA) methods were used to detect LT and ST producing enterotoxigenic *E. coli* (ETEC) [23]. *Entamoeba histolytica, Cryptosporidium,* and *Giardia* were identified by real-time PCR as described elsewhere [24]. Rotavirus, astrovirus, and adenovirus were detected by ELISA using commercial kits (ProSpectT Rotavirus Catalog R240396, ProSpectT Astrovirus Catalog R240196, and ProSpectT Adenovirus Catalog R240096, respectively). Multiplex (RT-)PCR and probe-based detection with Luminex beads for conceivable diarrhea-causative microbes was performed as described elsewhere in the literature [25–27].

The DNA was extracted using a slightly modified QIAamp DNA Stool Mini Kit protocol (Qiagen Inc, Valencia, California) [24]. The RNA was extracted using the QuickGene RNA tissue kit SII [25, 26]. For the (RT-)PCR-Luminex assay, either the forward or the reverse primer per target was labeled with biotin-TEG at 5′ ends. After (RT-)PCR was performed with the conditions described elsewhere, samples were analyzed on the BioPlex-200 system using bead on which coupling and hybridization were performed according to published protocols [28].

Amplification of Arg^{TCT} Gene Fragment and Sequencing

The *E. moshkovskii*–specific primer pair, EmR-1 and EmR-2, was used to specifically amplify the *E. moshkovskii* Arg^{TCT} gene fragment [13]. Amplification was performed using the high-fidelity Sahara DNA polymerase (Bio-Line, US). Sequencing was performed on an Applied Biosystems 377 Prism DNA Sequencer, using the BigDye terminator chemistry and EmR-1 or EmR-2 primer.

Statistical Analysis

The χ^2 test and Mann–Whitney *U* test were used where they were applicable.

RESULTS

E. moshkovskii Established the Infection in Mice

We previously showed that C3H/HeN, C3H/HeJ, and CBA/J mice allowed the establishment of *E. histolytica* infection, whereas many strains of mice including C57BL/6 and BALB/c mice did not, indicating that susceptibility to *E. histolytica* infection depended on the genetic background of the host [18–20]. Trophozoites of either *E. histolytica, E. moshkovskii,* or *E. dispar* were intracecally injected into congenic strains of mice. *E. histolytica* successfully infected the ceca of C3H/HeN, C3H/HeJ, and CBA/J mice. *E. moshkovskii* infected the ceca of CBA/J mice in approximately 68% (51 of 75) of mice at 4 days after challenge, as determined by both culture and PCR of intracecal contents. Likewise, C3H/HeN and C3H/HeJ mice were infected with *E. moshkovskii* in 60% and 40% of cases at 4 days, respectively, whereas infection rates of C57BL/6 and BALB/c mice were 5.6% and 0.0% at 4 days, respectively (Table 1). Nonpathogenic *E. dispar* did not infect any mouse strain tested. These data demonstrated that in contrast to nonpathogenic *E. dispar* that did not infect, *E. moshkovskii* had a similar host genetic susceptibility to infection in the murine model as did pathogenic *E. histolytica*.

Table 1. Susceptibility of Congenic Strains of Mice to *Entamoeba histolytica, Entamoeba moshkovskii,* or *Entamoeba dispar* Infection

	E. histolytica (%)	*E. moshkovskii* (%)	*E. dispar* (%)
BALB/c	1/15 (6)	0/10 (0)	0/10 (0)
C57BL/6	2/20 (10)	1/18 (6)	0/15 (0)
C3H/HeJ	8/15 (53)	4/10 (40)	0/10 (0)[a]
C3H/HeN	7/15 (47)	6/10 (60)	0/10 (0)[a]
CBA/J	61/90 (68)	51/75 (68)	0/20 (0)[a]

[a] $P < .05$ compared to *E. histolytica* or *E. moshkovskii* (χ^2 test).

E. moshkovskii Induced Intestinal Symptoms in Mice

Intestinal symptoms and body weight were monitored after challenging CBA/J mice with *E. moshkovskii*. A total of 71% (51/72) of CBA/J mice inoculated with *E. moshkovskii* were infected by 3 days after challenge. Diarrhea was observed in 39% (20/51) and dysentery in 6% (3/51) (Figure 1*A–C*). In successfully infected mice, amoebae were observed in the lumen of the ceca (Figure 1*D*). Mice with bloody diarrhea exhibited a thickened and contracted ceca (Figure 1*E*). Histopathological examination of ceca from these mice revealed epithelial ulceration, hemorrhagic changes, and tissue destruction (Figure 1*F*). Furthermore, obvious weight loss was observed during the course of *E. moshkovskii* infection in CBA/J mice, which was more severe than that observed during pathogenic *E. histolytica* infection, both of which were significant compared to control sham-operated mice (Figure 1*H*). Together these data indicated that *E. moshkovskii* was virulent in mice.

E. moshkovskii Was Expelled Within 14 Days After Challenge, Whereas *E. histolytica* Chronically Infected in the Ceca of Mice

The time course of each *Entamoeba* spp. infection in susceptible strains of mice was observed. As was reported [20], *E. histolytica* established chronic infection in not only CBA/J

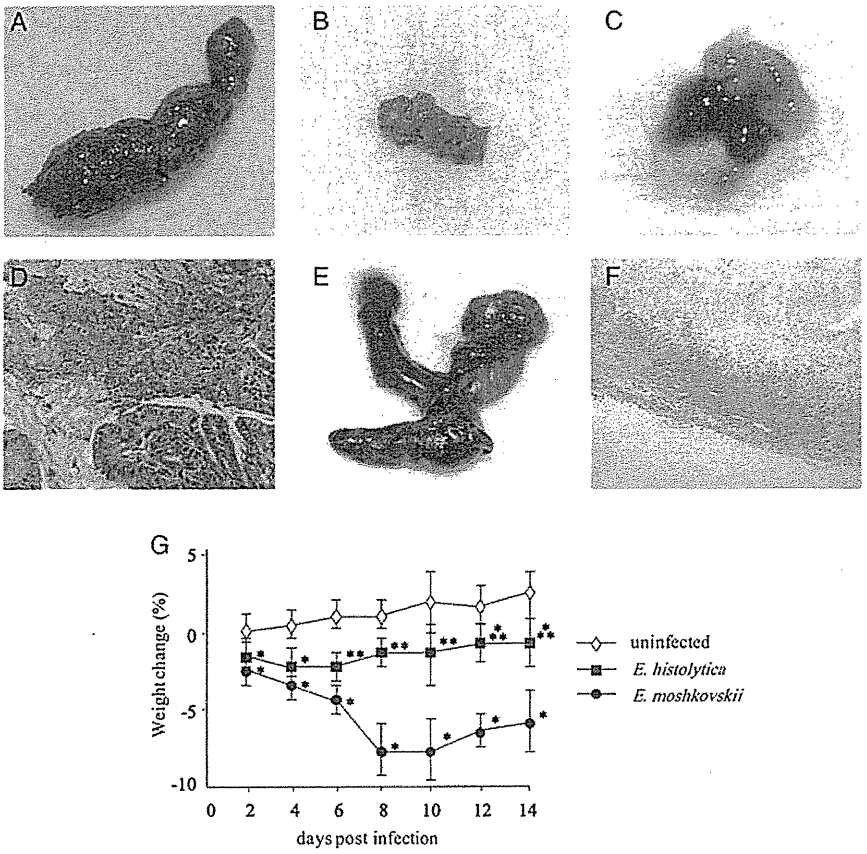

Figure 1. *Entamoeba moshkovskii* induced intestinal symptoms and weight loss in CBA/J mice. CBA/J mice were intracecally inoculated with 1×10^6 trophozoites of *E. moshkovskii*. After infection, diarrhea, colitis, and weight loss were monitored. Normal (*A*), loose (*B*), and bloody feces (*C*) were observed as was indicated in the results. Amoebae were observed in the lumen of the ceca in successfully infected mice (*D*). Macroscopic and histopathological observations of ceca in mice exhibited bloody diarrhea were shown in panels *E* and *F*. Changes in body weight were monitored in successfully infected 15 mice per group (*G*), in which CBA/J mice were intracecally inoculated with 1×10^6 trophozoites of *Entamoeba histolytica* (*solid squares*), *E. moshkovskii* (*solid circles*), or medium alone (*open diamonds*). The study was repeated 3 times with similar results. $^*P < 1.0 \times 10^{-6}$, $^{**}P < 1.0 \times 10^{-5}$ and $^{***}P < 1.0 \times 10^{-4}$ compared with sham-operated mice (Mann–Whitney *U* test).

but also C3H/HeJ and C3H/HeN mice, whereas neither C57BL/6 nor BALB/c allowed establishment of *E. histolytica* infection (Figure 2). In contrast, *E. moshkovskii* did not cause chronic infection, being expelled by approximately 2 weeks after challenge in CBA/J mice (Figure 2). A similar time to clearance was seen in C3H/HeN and C3H/HeJ mice.

In Infants in Bangladesh, *E. moshkovskii* Was Detected in Diarrheal Samples With Similar Frequency to *E. histolytica*

The association between diarrheal episodes and infection with each *Entamoeba* spp. was tested in children in Mirpur, Dhaka, Bangladesh. These studies were part of a prospective cohort study on diarrheal diseases [29]. Newborn children were enrolled in the Mirpur community of Dhaka, Bangladesh, and prospectively followed for diarrheal illness by every other day home visits. A total of 1426 diarrheal episodes were recorded during the first 12 months of life in 385 children. PCR analyses of the diarrheal samples revealed that 66 episodes were positive for *E. histolytica* (4.63%), 42 were positive for *E. moshkovskii* (2.95%), and 5 episodes were positive for *E. dispar* (0.35%). As such, in diarrheal samples, the detection rates of either *E. histolytica* or *E. moshkovskii* were 13.2 and 8.4 times higher than that of nonpathogenic *E. dispar*. Two episodes were found to be mixed infections with *E. histolytica* and *E. moshkovskii*, but no other mixed infections of *Entamoeba* spp. were found.

E. moshkovskii Infection Was Newly Acquired in Children With Diarrhea

In order to attempt to discern if the *E. moshkovskii* detected in the diarrheal stool sample could be the cause of diarrhea, we tested if it was newly acquired at the time of diarrhea. The preceding 2 months of surveillance stool samples collected when the child did not have diarrhea were tested for the

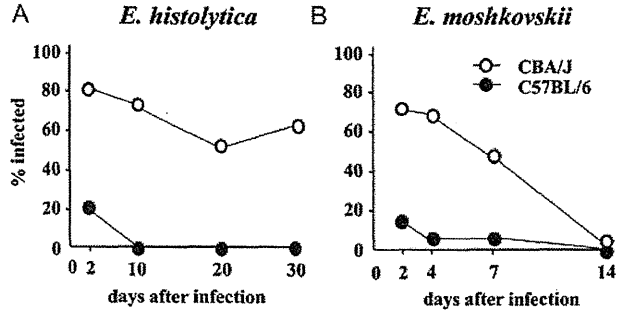

Figure 2. *Entamoeba moshkovskii* was expelled within 2 weeks in CBA/J Mice. CBA/J (open circles) and C57BL/6 (solid circles) mice were intracecally inoculated with 1×10^6 trophozoites of *Entamoeba histolytica* (*A*) or *Entamoeba moshkovskii* (*B*). Time course of each *Entamoeba* spp. infection was then monitored by detection of the parasites in stool by culture and polymerase chain reaction (PCR).

presence of *E. moshkovskii*. This study design therefore temporally controlled for *E. moshkovskii* infection in the 42 infants with diarrhea attributed to this parasite. In the 1 and 2 months preceding *E. moshkvskii*-associated diarrhea, 93% (39/42) and 86% (36/42) of monthly surveillance stool samples, respectively, were negative for *E. moshkovskii* (Table 2). This supported the hypothesis that temporal acquisition of a new *E. moshkovskii* infection led to diarrheal episodes in some proportion of these children.

E. moshkovskii-Associated Diarrhea Was of Similar Severity to Other Causes of Diarrhea

The diarrheal severity score was comparable among episodes associated with *E. histolytica, E. moshkovskii,* and other causes: 4.89 ± 0.22, 4.71 ± 0.24, and 4.84 ± 0.05, respectively. The duration of diarrhea was also comparable among these episodes positive for *E. histolytica, E. moshkovskii,* and others: 4.44 ± 0.44, 4.74 ± 0.49, and 4.84 ± 0.10 days, respectively (Table 3). The mean age of the onset of diarrheal episodes associated with *E. histolytica, E. moshkovskii,* and others was found to be 7.72 ± 0.75, 9.12 ± 0.73, and 9.09 ± 0.18 months, respectively, without any significant differences. Thus, the diarrhea related to *E. moshkovskii* was indistinguishable from diarrhea related to *E. histolytica* in severity, duration, and age of onset (Table 3).

Table 2. Prevalence of *Entamoeba moshkovskii* Asymptomatic Infection Preceding *E. moshkovskii*-Associated Diarrhea in 42 Children

Category	Preceding 1-Month Surveillance Stool	Preceding 2-Month Surveillance Stool
E. moshkovskii (+)	3	6
E. moshkovskii (–)	39	36
Total	42	42

Table 3. Severity and Duration of Diarrhea Associated With *Entamoeba histolytica* or *Entamoeba moshkovskii*

Pathogens	Severity Score (mean ± SE)	Duration (days) (mean ± SE)	Age of Onset in Months (mean ± SE)
E. histolytica	4.89 ± 0.22	4.44 ± 0.44	7.72 ± 0.75
E. moshkovskii	4.71 ± 0.24[a]	4.74 ± 0.49[a]	9.12 ± 0.73[a]
Others	4.84 ± 0.05	4.84 ± 0.10	9.09 ± 0.18

[a] No significant difference in diarrheal severity score or duration for episodes associated with *E. histolytica, E. moshkovskii,* or other enteropathogens infection.

Additional Enteropathogens Were Identified in Stool Samples From *E. moshkovskii* Infected Children

As there are many microbes that can potentially induce diarrhea, the presence of other diarrheagenic microbes was tested in the 42 diarrheal samples that were associated with *E. moshkovskii*. The 42 samples were examined for other conceivable diarrhea-causative microbes infection using standard bacterial culture techniques, fecal antigen detection, and multiplex PCR combined with probe-based detection with Luminex beads (Table 4) [25, 26]. In the 42 diarrheal stool samples with *E. moshkovskii,* 12 samples (28.6%) contained >4 other pathogens, 13 (31.0%) had 3 pathogens, 14 (33.3%) had 2 pathogens, 1 (2.3%) had 1 pathogen, and 2 samples were positive solely for *E. moshkovskii* (Table 4). The application of these state-of-the-art diagnostic techniques in this cohort has on average identified a minimum of 2 different enteropathogens in every diarrheal stool sample (E. Houpt and M. Taniuchi, personal communication, 2011). It was therefore not surprising that the diarrheal episodes associated with *E. moshkovskii* were commonly coinfected.

E. moshkovskii Isolates Were Genetically Diverse in the Infants

In order to investigate the genetic diversity in *E. moshkovskii* strains detected in the infected children's stools, we used a tRNA-gene linked locus (R-R), which previously showed PCR size differences among *E. moshkovskii* strains from Bangladesh [12, 30]. Twenty-six *E. moshkovskii*-positive stool DNAs (6 from asymptomatic children and 20 from diarrheal children) were amplified using the *E. moshkovskii* specific nested PCR primers described elsewhere[12]. However, PCR did not reveal any obvious product size differences among these samples (data not shown). Because same size PCR products do not necessarily mean identical DNA sequences, we sequenced PCR products directly without cloning them into any vectors (in order to minimize the chances of any sequence selection) to detect sequence variation. Sequencing did reveal that the

Table 4. Other Enteropathogens Detected in *Entamoeba moshkovskii* (+) Diarrheal Stool Samples

Name of Organism	No. of Samples
Encephalitozoon intestinalis	1
Cyclospora cayetanensis	1
Cystoisospora belli	1
Enterocytozoon bieneusi	6
Adenovirus	1
Astrovirus	4
Sapovirus	2
Norovirus G1	0
Norovirus G2	4
Rotavirus	1
E. histolytica	2
Giardia intestinalis	10
Cryptosporidium spp.	1
Vibrio cholera/parahaemolyticus	3
EAEC	15
ETEC	3
EPEC	5
EHEC	0
EIEC/Shigella spp.	23
Salmonella (pan)	2
Aeromonas (pathogenic)	14
Yersinia (pan)	0
Campylobacter jejuni/coli	23

E. moshkovskii strains detected in this study were polymorphic in locus R-R; although unlike the *E. histolytica* and *E. dispar* sequences[31], no short tandem repeats could be detected in *E. moshkovskii*. Single-nucleotide polymorphisms (SNPs) were detected in 2 of the 6 asymptomatic children-derived sequences and in 6 of the 20 diarrheal children-derived sequences (Supplementary Figure 1). These SNPs could be used to divide them into 9 different genotypes—18 strains with identical locus R-R sequences and the remaining 8 strains containing ≥1 distinct SNPs (Supplementary Figure 1 and Table 5). Because we used a high-fidelity DNA polymerase (Bio-Line, US) during PCR amplification, it was unlikely that these SNPs were erroneously introduced by the DNA polymerase. The sequence alignment at locus R-R revealed that the *E. moshkovskii* strains of this study were comparatively more diverse than the reference *E. moshkovskii* Laredo strain, but closer to the only Bangladeshi strain (ID:MS15-3646) sequenced previously (labeled as Em-Laredo and Em-BANGLA, respectively, in Supplementary Figure 1). The SNPs detected in this study were distributed randomly across the locus R-R sequences, and as a result, these SNPs could not be used to differentiate asymptomatic and diarrheal strains of *E. moshkovskii*. However, we noticed from the sequence traces that the 2 asymptomatic strains (IDs:8056-CMS15 and 7086-CMS15)

Table 5. Single-Nucleotide Polymorphisms (SNPs) in the *E. moshkovskii* Strains from Bangladesh at Locus R-R

ID	Clinical Status	No. of SNPs	Position and SNP Type
7040-CDS05	Diarrhea	6	T204A, C205T, T206C, C208del, T209del, T210del
7063-CDS02	Diarrhea	1	T71C
7161-CDS05	Diarrhea	1	T141G
7146-CDS02	Diarrhea	1	A235T
8119-CDS02	Diarrhea	2	T83C, T137C
8113-CDS04	Diarrhea	3	T81C, C120T, G221A
7086-CMS15	Asymptomatic	2	T203W, T204Y
8056-CMS15	Asymptomatic	1	A135R

All positions are based on the consensus sequence in the alignment. W = A/T; Y = C/T; R = A/G.

showed allelic variation in all 3 SNPs (T203W, T204Y, and A135R), whereas none of the 6 diarrheal strains showed any allelic variations in their respective SNPs (Table 5 and Supplementary Figure 2). The significance of this remains unknown at present.

DISCUSSION

This work draws into question the paradigm that *E. moshkovskii* is avirulent. In the murine model of intestinal amebiasis, *E. moshkovskii* caused diarrhea, weight loss, and colitis. In this way, *E. moshkovskii* shared with *E. histolytica,* but not the nonpathogen *E. dispar,* the ability to cause disease. In children in Bangladesh, the new acquisition of *E. moshkovskii* infection was associated with diarrhea.

E. moshkovskii infected the ceca of C3H/HeN, C3H/HeJ, and CBA/J mice, but not C57BL/6 or BALB/c mice, which was consistent with the host range of pathogenic *E. histolytica.* In contrast, the nonpathogenic parasite *E. dispar* was unable to infect the intestine of any strains of mice tested. The finding that *E. moshkovskii* shared with *E. histolytica* the ability to infect mice indicates that they share virulence mechanisms, which are not present in *E. dispar.*

Mouse strain-dependent resistance to *E. histolytica* infection was mediated by nonhematopoietic cells [19]. Relatively few loci on C57BL/6 chromosomes 1 and 2 correlated with resistance to intestinal amebiasis [32]. In humans, one important means of innate resistance of intestinal epithelial cells to amebiasis is leptin, which acts via STAT3 signaling to protect intestinal epithelial cells from parasite killing [33, 34]. In this context, it will be interesting to examine whether this observation is also true in *E. moshkovskii* infection, because of the similar host range as *E. histolytica.* If the mechanism of resistance to *E. histolytica* and *E. moshkovskii* observed in many

inbred strains of mice is shared with humans, identification of regional candidate genes in mice has implications for further understanding the human variability to amebic infection.

E. moshkovskii induced intestinal symptoms including diarrhea and bloody stool, typical symptoms of amebiasis, indicating that *E. moshkovskii* was pathogenically similar to *E. histolytica* at least in mice. Weight loss was also observed during the course of infection, which was more severe in mice infected with *E. moshkovskii* than with *E. histolytica*. The observation that *E. moshkovskii* induced severe intestinal symptoms accompanied by weight loss reemphasizes that it is potentially pathogenic.

However, it is unclear what kinds of differences among *Entamoeba* spp. result in the different outcomes of infection in the murine model. *Entamoeba histolytica* possesses molecules such as pathogen-associated molecular patterns (PAMPs) on its surface that stimulate proinflammatory cytokines production from antigen-presenting cells [35]. We are investigating whether parasite PAMPs, host MyD88 signaling, and the pattern of proinflammatory cytokines produced in response qualitatively differ between *Entamoeba* species, may provide clues to the different severity between CBA/J mice infected with *E. histolytica*, *E. moshkovskii*, and *E. dispar*.

Entamoeba moshkovskii isolates infecting children were genetically heterogeneous, as evidenced by PCR typing of tRNA locus R-R. It will be important in future studies of the potential pathogenicity of *E. moshkovskii* to take into account this heterogeneity; as for the case of *E. histolytica*, not every genotype is equally capable of causing disease [36].

The study subjects reported here differ from those of the previous study examining *E. moshkovskii* infection in preschool children in Dhaka, Bangladesh [13], which was not focused solely on diarrheal stool samples, but also included monthly stool samples from asymptomatic children. In addition, the current study reports on a novel birth cohort longitudinally followed from birth to 1 year of age. Therefore, it is important to discuss the association between diarrhea in infants and *E. moshkovskii* infection in the context of the cohort.

In conclusion, we found that *E. moshkovskii* caused diarrhea, colitis, and weight loss in mice and that in Bangladeshi children acquisition of a new *E. moshkovskii* infection occurred temporally with diarrhea. These data are consistent with *E. moshkovskii* causing diarrhea and indicate that it is important to reexamine its pathogenicity.

Supplementary Data

Supplementary materials are available at *The Journal of Infectious Diseases* online (http://jid.oxfordjournals.org/). Supplementary materials consist of data provided by the author that are published to benefit the reader. The posted materials are not copyedited. The contents of all supplementary data are the sole responsibility of the authors. Questions or messages regarding errors should be addressed to the author.

Notes

Acknowledgments. We thank M. Iwami-Sano and C. Matsuzaki-Moriya for technical assistance, F. Hara and M. Hayashida for animal husbandry, H. Hisaeda for comments, members of Department of Parasitology, Kyushu University and Institute of Tropical Medicine, Nagasaki University for helpful discussion, and members of Parasitology Laboratory, ICDDR, B., and all of field research assistants for helpful assistance in the field.

Financial support. This work was supported by NIH grant 5RO1 AI043596 (to W. A. P.), a Grant-in-Aid for Scientific Research on Priority Areas from MEXT (21022037 to S. H.), Grants-in-Aid for International Scientific Research (B) from JSPS (20406008, 23406009 to S. H.), a Health Labour Sciences Research Grant (H20-Shinkoh-Ippan-016, H23-Shinkoh-Ippan-014 to S. H.), the Takeda Foundation, the Uehara Foundation (to S. H.), and the Global COE Program, Nagasaki University, supported by MEXT (to S. H.).

Potential conflicts of interest. Dr Petri receives royalties from a licensing agreement with TechLab, Inc., for amebiasis diagnostics. These royalties are donated in their entirety to the American Society of Tropical Medicine and Hygiene without benefit to Dr Petri. All other authors report no potential conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Irusen EM, Jackson TF, Simjee AE. Asymptomatic intestinal colonization by pathogenic *Entamoeba histolytica* in amebic liver abscess: prevalence, response to therapy, and pathogenic potential. Clin Infect Dis **1992**; 14:889–93.
2. Blessmann J, Ali IK, Nu PA, et al. Longitudinal study of intestinal *Entamoeba histolytica* infections in asymptomatic adult carriers. J Clin Microbiol **2003**; 41:4745–50.
3. Gathiram V, Jackson TF. A longitudinal study of asymptomatic carriers of pathogenic zymodemes of *Entamoeba histolytica*. S Afr Med J **1987**; 72:669–72.
4. Vandenberg O, Peek R, Souayah H, et al. Clinical and microbiological features of dientamoebiasis in patients suspected of suffering from a parasitic gastrointestinal illness: a comparison of *Dientamoeba fragilis* and *Giardia lamblia* infections. Int J Infect Dis **2006**; 10:255–61.
5. Stark D, Beebe N, Marriott D, Ellis J, Harkness J. Prospective study of the prevalence, genotyping, and clinical relevance of *Dientamoeba fragilis* infections in an Australian population. J Clin Microbiol **2005**; 43:2718–23.
6. Fotedar R, Stark D, Marriott D, Ellis J, Harkness J. *Entamoeba moshkovskii* infections in Sydney, Australia. Eur J Clin Microbiol Infect Dis **2008**; 27:133–7.
7. Scaglia M, Gatti S, Strosselli M, Grazioli V, Villa MR. *Entamoeba moshkovskii* (Tshalaia, 1941): morpho-biological characterization of new strains isolated from the environment, and a review of the literature. Ann Parasitol Hum Comp **1983**; 58:413–22.
8. Dreyer DA. Growth of a strain of *Entamoeba histolytica* at room temperature. Tex Rep Biol Med **1961**; 19:393–6.
9. Fotedar R, Stark D, Beebe N, Marriott D, Ellis J, Harkness J. Laboratory diagnostic techniques for *Entamoeba* species. Clin Microbiol Rev **2007**; 20:511–32, table of contents.
10. Clark CG, Diamond LS. The Laredo strain and other '*Entamoeba histolytica*-like' amoebae are *Entamoeba moshkovskii*. Mol Biochem Parasitol **1991**; 46:11–8.
11. Tanyuksel M, Ulukanligil M, Guclu Z, Araz E, Koru O, Petri WA Jr. Two cases of rarely recognized infection with *Entamoeba moshkovskii*. Am J Trop Med Hyg **2007**; 76:723–4.

12. Fotedar R, Stark D, Beebe N, Marriott D, Ellis J, Harkness J. PCR detection of *Entamoeba histolytica*, *Entamoeba dispar*, and *Entamoeba moshkovskii* in stool samples from Sydney, Australia. J Clin Microbiol **2007**; 45:1035–7.
13. Ali IK, Hossain MB, Roy S, et al. *Entamoeba moshkovskii* infections in children, Bangladesh. Emerg Infect Dis **2003**; 9:580–4.
14. Khairnar K, Parija SC, Palaniappan R. Diagnosis of intestinal amoebiasis by using nested polymerase chain reaction-restriction fragment length polymorphism assay. J Gastroenterol **2007**; 42: 631–40.
15. Nazemalhosseini Mojarad E, Nochi Z, Sahebekhtiari N, et al. Discrimination of *Entamoeba moshkovskii* in patients with gastrointestinal disorders by single-round PCR. Jpn J Infect Dis **2010**; 63:136–8.
16. Parija SC, Khairnar K. Mutation detection analysis of a region of 16S-like ribosomal RNA gene of *Entamoeba histolytica*, *Entamoeba dispar* and *Entamoeba moshkovskii*. BMC Infect Dis **2008**; 8:131.
17. Beck DL, Dogan N, Maro V, Sam NE, Shao J, Houpt ER. High prevalence of *Entamoeba moshkovskii* in a Tanzanian HIV population. Acta Trop **2008**; 107:48–9.
18. Hamano S, Asgharpour A, Stroup SE, Wynn TA, Leiter EH, Houpt E. Resistance of C57BL/6 mice to amoebiasis is mediated by nonhemopoietic cells but requires hemopoietic IL-10 production. J Immunol **2006**; 177:1208–13.
19. Houpt ER, Glembocki DJ, Obrig TG, et al. The mouse model of amebic colitis reveals mouse strain susceptibility to infection and exacerbation of disease by CD4+ T cells. J Immunol **2002**; 169:4496–503.
20. Asgharpour A, Gilchrist C, Baba D, Hamano S, Houpt E. Resistance to intestinal *Entamoeba histolytica* infection is conferred by innate immunity and Gr-1+ cells. Infect Immun **2005**; 73:4522–9.
21. Diamond LS, Harlow DR, Cunnick CC. A new medium for the axenic cultivation of *Entamoeba histolytica* and other *Entamoeba*. Trans R Soc Trop Med Hyg **1978**; 72:431–2.
22. Hamzah Z, Petmitr S, Mungthin M, Leelayoova S, Chavalitshewinkoon-Petmitr P. Differential detection of *Entamoeba histolytica*, *Entamoeba dispar*, and *Entamoeba moshkovskii* by a single-round PCR assay. J Clin Microbiol **2006**; 44:3196–200.
23. Qadri F, Saha A, Ahmed T, Al Tarique A, Begum YA, Svennerholm AM. Disease burden due to enterotoxigenic *Escherichia coli* in the first 2 years of life in an urban community in Bangladesh. Infect Immun **2007**; 75:3961–8.
24. Haque R, Roy S, Siddique A, et al. Multiplex real-time PCR assay for detection of *Entamoeba histolytica*, *Giardia* intestinalis, and *Cryptosporidium* spp. Am J Trop Med Hyg **2007**; 76:713–7.
25. Liu J, Kibiki G, Maro V, et al. Multiplex reverse transcription PCR Luminex assay for detection and quantitation of viral agents of gastroenteritis. J Clin Virol **2011**; 50:308–13.
26. Taniuchi M, Verweij JJ, Noor Z, et al. High throughput multiplex PCR and probe-based detection with Luminex beads for seven intestinal parasites. Am J Trop Med Hyg **2011**; 84:332–7.
27. Taniuchi M, Verweij JJ, Sethabutr O, et al. Multiplex PCR method to detect *Cyclospora*, *Cystoisospora*, and *Microsporidia* in stool samples. Diagn Microbiol Infect Dis **2011**; 71:386–90.
28. Baums IB, Goodwin KD, Kiesling T, Wanless D, Diaz MR, Fell JW. Luminex detection of fecal indicators in river samples, marine recreational water, and beach sand. Mar Pollut Bull **2007**; 54:521–36.
29. Mondal D, Minak J, Alam M, et al. Contribution of enteric infection, altered intestinal barrier function and maternal malnutrition to infant malnutrition in Bangladesh. Clin Infect Dis **2011**; 54:185–92.
30. Ali IK, Mondal U, Roy S, Haque R, Petri WA Jr, Clark CG. Evidence for a link between parasite genotype and outcome of infection with *Entamoeba histolytica*. J Clin Microbiol **2007**; 45:285–9.
31. Ali IK, Zaki M, Clark CG. Use of PCR amplification of tRNA gene-linked short tandem repeats for genotyping *Entamoeba histolytica*. J Clin Microbiol **2005**; 43:5842–7.
32. Hamano S, Becker S, Asgharpour A, et al. Gender and genetic control of resistance to intestinal amebiasis in inbred mice. Genes Immun **2008**; 9:452–61.
33. Duggal P, Guo X, Haque R, et al. A mutation in the leptin receptor is associated with *Entamoeba histolytica* infection in children. J Clin Invest **2011**; 121:1191–8.
34. Guo X, Roberts MR, Becker SM, et al. Leptin signaling in intestinal epithelium mediates resistance to enteric infection by *Entamoeba histolytica*. Mucosal Immunol **2011**; 4:294–303.
35. Guo X, Houpt E, Petri WA Jr. Crosstalk at the initial encounter: interplay between host defense and ameba survival strategies. Curr Opin Immunol **2007**; 19:376–84.
36. Ali IK, Solaymani-Mohammadi S, Akhter J, et al. Tissue invasion by *Entamoeba histolytica*: evidence of genetic selection and/or DNA reorganization events in organ tropism. PLoS Negl Trop Dis **2008**; 2: e219.

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

[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.

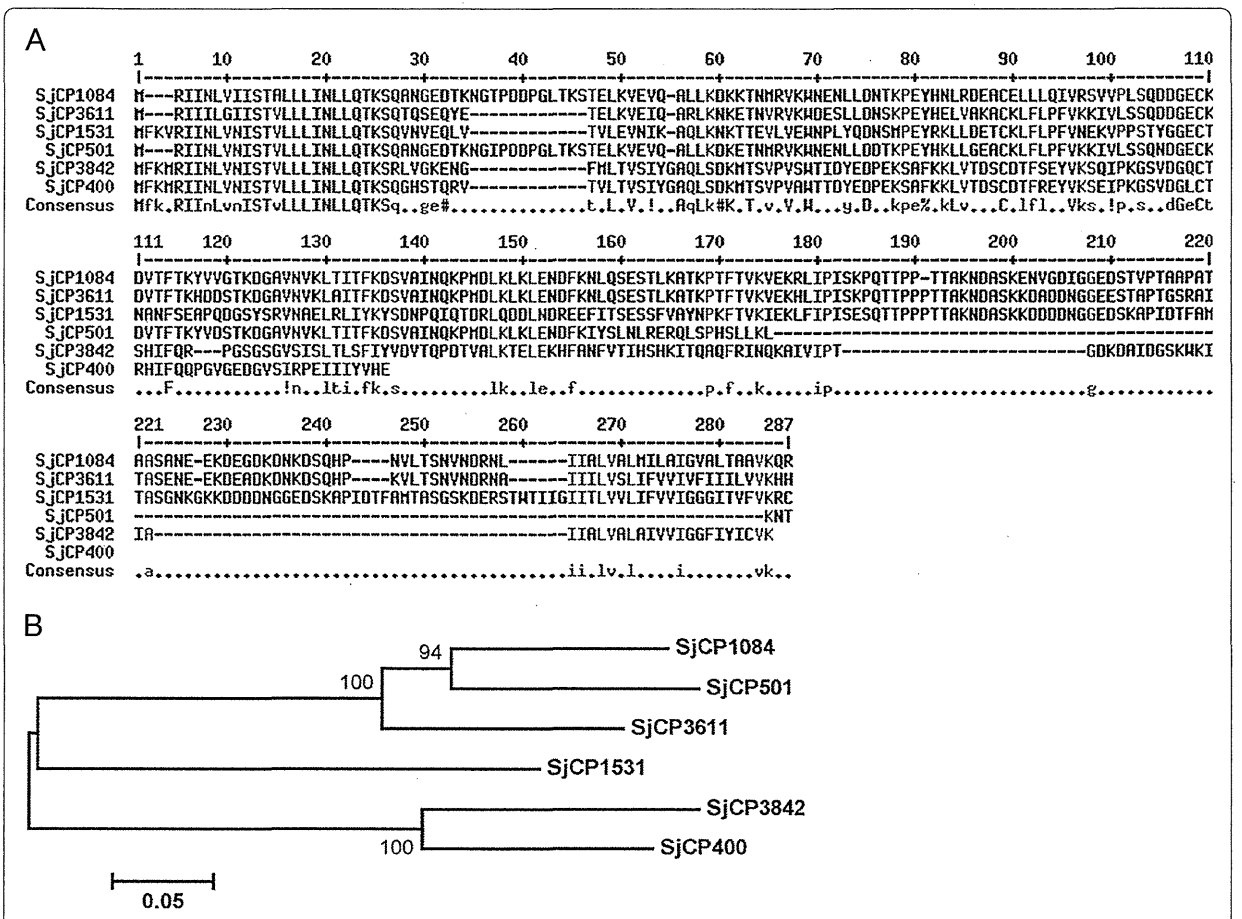

Figure 1 Multiple alignments of protein sequences of the SST identified cDNAs showing similar signal peptide. (**A**) The protein products of the original SST isolated *S. japonicum* egg cDNAs were aligned using ClustalW. The aligned sequences are limited to the candidates identified using SST, excluding 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.

Table 3 *Schistosoma japonicum* genome contigs containing the similar signal sequence (n = 34)

*Contig [GenBank Accession No.]	Transcription Strand	Size, kb	Signal Sequence coordinates
SJC_C002611 [CABF01002612]	-	69.7	4848 – 4779
SJC_C002621 [CABF01002622]	-	0.6	276 – 205
SJC_C002622 [CABF01002623]	-	1.6	999 – 928
SJC_C002627 [CABF01002628]	-	3.0	2413 – 2342
SJC_C002629 [CABF01002630]	-	14.5	3856 – 3785
SJC_C013669 [CABF01013761]	-	6.8	3284 – 3217
SJC_C019814 [CABF01020047]	+	29.2	19023 – 19094
SJC_C019817 [CABF01020050]	-	10.1	6502 – 6431
SJC_C019827 [CABF01020060]	-	43.7	42511 – 42440
SJC_C022876 [CABF01022876]	-	12.4	9335 – 9264
SJC_C022884 [CABF01022884]	+	12.9	493 – 564
SJC_C023364 [CABF01023364]	-	12.4	10498 – 10427
SJC_C025268 [CABF01025296]	-	12.1	7860 – 7789
SJC_C027826 [CABF01027854]	+	4.3	433 – 504
SJC_C027833 [CABF01027861]	-	11.9	5768 – 5697
SJC_C027838 [CABF01027866]	+	19.0	12367 – 12438
SJC_C032855 [CABF01032892]	-	22.3	383 – 322
SJC_C032859 [CABF01032896]	-	9.6	4663 – 4602
SJC_C043165 [CABF01043187]	+	4.9	2484 – 2544
SJC_C057153 [CABF01057161]	+	2.8	337 – 408
SJC_C061392 [CABF01061395]	-	7.4	4411 – 4342
SJC_C067189 [CABF01067176]	+	4.9	388 – 459
SJC_C067567 [CABF01067411]	-	3.2	925 – 854
SJC_C070280 [CABF01070230]	-	6.8	271 – 200
SJC_C072631 [CABF01072590]	+	1.9	1646 – 1717
SJC_C072632 [CABF01072591]	+	2.6	446 – 517
SJC_C073741 [CABF01073691]	-	2.4	2389 – 2319
SJC_C075160 [CABF01075030]	-	6.7	5231 – 5160
SJC_C076469 [CABF01076032]	+	3.1	1295 – 1366
SJC_C077101 [CABF01078976]	-	1.1	656 – 585
SJC_C080985 [CABF01080674]	+	2.3	1094 – 1165
SJC_C081391 [CABF01080757]	-	2.1	1918 – 1847
SJC_C081246 [CABF01080893]	-	1.3	1131 – 1060
SJC_C097686 [CABF01092393]	+	5.6	4249 – 4320

(-) are contigs with 'signal sequence' on the negative strand (anti-sense) of the genome while
(+) are contigs with 'signal sequence' on the positive strand (sense) in the genome
*Contigs are representative of disperse duplicated gene loci. We indicated the ranges for the signal sequence motif.

To investigate possible role by repetitive elements (REs) in mediating such dispersed duplication with a clue from previous studies [20,26-29], we performed repeat masking on the putative duplication source locus and the other 33 duplicons and observed a preponderance of flanking REs, especially of non-LTR class prominent of which were the *S. japonicum RTE (*retrotransposable element*)*-like retrotransposon (*SjR2*) and the *Perere* class of retrotransposons *(SjR1)* (Additional file 5). An almost full copy of *SjR2* was found upstream of the coding region of the putative source locus in addition to other six albeit partial copies of *SjR2*. Alignment of the other contigs to the putative duplication source locus revealed that both the dispersed similar signal sequence and the repeat elements are considerably aligned at very similar positions, further showing that they were likely

Table 4 *Schistosoma japonicum* Scaffolds containing the similar signal sequence (n = 18)

Scaffolds [GenBank Accession]	Contigs within the scaffolds
SJC_S000013 [FN330988]	CABF01002611, CABF01002612, CABF01002622, CABF01002623, CABF01002628, CABF01002630
SJC_S000219 [FN331192]	CABF01020047, CABF01020050
SJC_S000220 [FN331193]	CABF01020060
SJC_S000273 [FN331245]	CABF01022876, CABF01022884
SJC_S000284 [FN331256]	CABF01023364
SJC_S000329 [FN331301]	CABF01025296
SJC_S000394 [FN331366]	CABF01027854, CABF01027861, CABF01027866
SJC_S005820 [FN336777]	CABF01067176
SJC_S007785 [FN338731]	CABF01070230
SJC_S008639 [FN339578]	CABF01072590, CABF01072591
SJC_S009177 [FN340103]	CABF01073691
SJC_S010134 [FN341037]	CABF01075030
SJC_S011206 [FN342077]	CABF01076032
SJC_S011724 [FN342573]	CABF01078976
SJC_S014521 [FN345237]	CABF01080674
SJC_S014753 [FN345459]	CABF01080893
SJC_S014868 [FN345568]	CABF01080757
SJC_S026182 [FN354050]	CABF01092393

duplicated from a single source locus. The fact that the duplicons are not absolutely homologous and the degenerative nature of the RE sequences suggests variation within members, typical of evolving genes (Figure 4). Because homology with the other duplicates did not terminate 3` of this putative source locus, we recruited and adjoined two contigs [GenBank:CABF01020061 and GenBank:CABF01020062] downstream of the putative source locus according to the genome assembly information, thereby creating flanking sequences of at least 5 kilobasepairs on each side of the gene duplication source locus. This sequence was then aligned with the genome contigs and scaffolds to identify the exact point at which homology was lost, which could arguably represent the breakpoint of duplication. Further attempt to identify the exact breakpoints was not successful due to unfilled sequencing gaps in the scaffolds but examination of the downstream flanking sequence from the point where homology was terminated showed a prominent retrotransposon of the *Perere* class flanking the duplicated loci 3` of the locus (see short movie in Additional file 5). Taken together, our data show that the duplication source locus was flanked on either side by *RTE*-like and *Perere* class retrotransposons. These two classes of non-LTR retrotransposons have significantly high copy number, making up 12.63 % of the *S. japonicum*

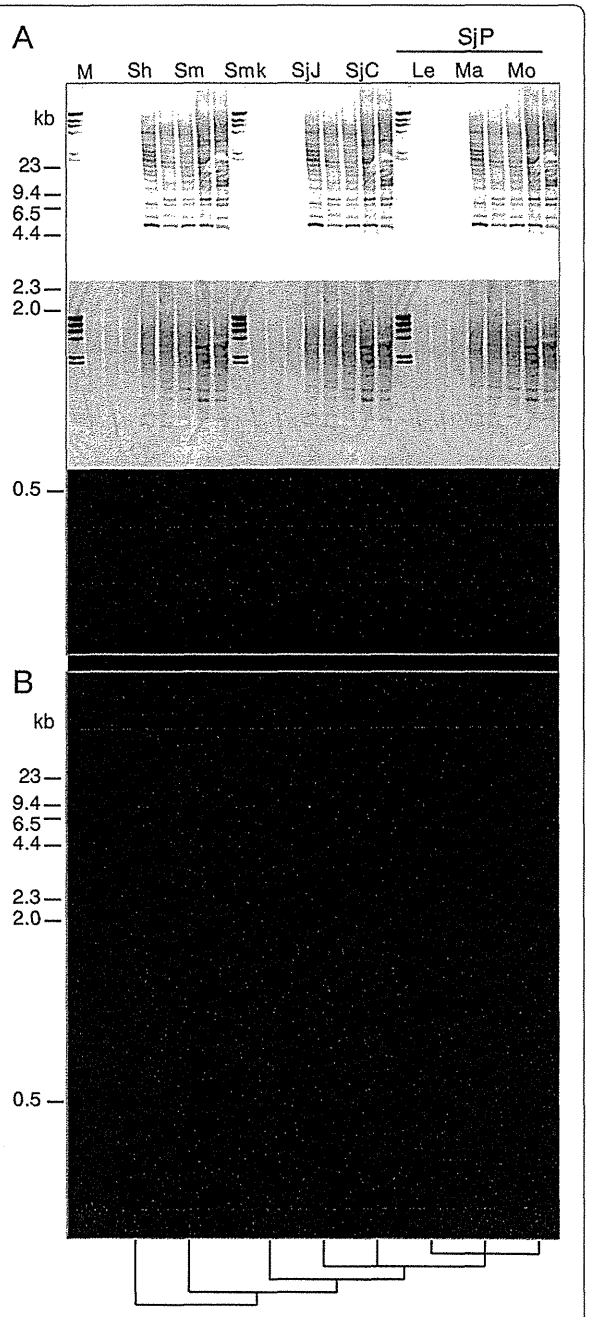

Figure 2 Southern blotting confirms duplicated loci exclusively in *S. japonicum*. (**A**) Southern hybridization with digoxigenin-labeled probes showing the presence of duplicated loci with several bands due to copies of the duplicated source locus. Lanes 2-9 corresponds to *Eco*RI + *Eco*RV double digested genomic DNA of different species and strains of Schistosoma (*S. haematobium*, *S. mansoni*, *S. mekongi*, *S. japonicum* (Japanese, Chinese, Philippines' Leyte, Mindanao and Mindoro isolates). 'M' is Digoxigenin-labeled DNA molecular weight marker. Notice the differential banding pattern among different strains and between isolates of the same strain. (**B**) Same experiment as in (A) was replicated using a different pair of restriction enzymes (EcoRI + *Hind*III). Also inter-strain and intra-strain variation in the banding pattern is apparent.