nature genetics

Plasmodium cynomolgi genome sequences provide insight into *Plasmodium vivax* and the monkey malaria clade

Shin-Ichiro Tachibana[1,13], Steven A Sullivan[2], Satoru Kawai[3], Shota Nakamura[4], Hyunjae R Kim[2], Naohisa Goto[4], Nobuko Arisue[5], Nirianne M Q Palacpac[5], Hajime Honma[1,5], Masanori Yagi[5], Takahiro Tougan[5], Yuko Katakai[6], Osamu Kaneko[7], Toshihiro Mita[8], Kiyoshi Kita[9], Yasuhiro Yasutomi[10], Patrick L Sutton[2], Rimma Shakhbatyan[2], Toshihiro Horii[5], Teruo Yasunaga[4], John W Barnwell[11], Ananias A Escalante[12], Jane M Carlton[2,14] & Kazuyuki Tanabe[1,5,14]

***P. cynomolgi*, a malaria-causing parasite of Asian Old World monkeys, is the sister taxon of *P. vivax*, the most prevalent malaria-causing species in humans outside of Africa. Because *P. cynomolgi* shares many phenotypic, biological and genetic characteristics with *P. vivax*, we generated draft genome sequences for three *P. cynomolgi* strains and performed genomic analysis comparing them with the *P. vivax* genome, as well as with the genome of a third previously sequenced simian parasite, *Plasmodium knowlesi*. Here, we show that genomes of the monkey malaria clade can be characterized by copy-number variants (CNVs) in multigene families involved in evasion of the human immune system and invasion of host erythrocytes. We identify genome-wide SNPs, microsatellites and CNVs in the *P. cynomolgi* genome, providing a map of genetic variation that can be used to map parasite traits and study parasite populations. The sequencing of the *P. cynomolgi* genome is a critical step in developing a model system for *P. vivax* research and in counteracting the neglect of *P. vivax*.**

Human malaria is transmitted by anopheline mosquitoes and is caused by four species in the genus *Plasmodium*. Of these, *P. vivax* is the major malaria agent outside of Africa, annually causing 80–100 million cases[1]. Although *P. vivax* infection is often mistakenly regarded as benign and self-limiting, *P. vivax* treatment and control present challenges distinct from those of the more virulent *Plasmodium falciparum*. Biological traits, including a dormant (hypnozoite) liver stage responsible for recurrent infections (relapses), early infective sexual stages (gametocytes) and transmission from low parasite densities in the blood[2], coupled with emerging antimalarial drug resistance[3], render *P. vivax* resilient to modern control strategies. Recent evidence indicates that *P. falciparum* derives from parasites of great apes in Africa[4], whereas *P. vivax* is more closely related to parasites of Asian Old World monkeys[5–7], although not itself infective of these monkeys.

P. vivax cannot be cultured *in vitro*, and the small New World monkeys capable of hosting it are rare and do not provide an ideal model system. *P. knowlesi*, an Asian Old World monkey parasite recently recognized as a zoonosis for humans[8], has had its genome sequenced[9], but the species is distantly related to *P. vivax* and is phenotypically dissimilar. In contrast, *P. cynomolgi*, a simian parasite that can infect humans experimentally[10], is the closest living relative (a sister taxon) to *P. vivax* and possesses most of the same genetic, phenotypic and biological characteristics—notably, periodic relapses caused by dormant hypnozoites, early infectious gametocyte formation and invasion of Duffy blood group–positive reticulocytes. *P. cynomolgi* thus offers a robust model for *P. vivax* in a readily available laboratory host, the Rhesus monkey, whose genome was recently sequenced[11]. Here, we report draft genome sequences of three *P. cynomolgi* strains and comparative genomic analyses of *P. cynomolgi*, *P. vivax*[12] and *P. knowlesi*[9], three members of the monkey malaria clade.

We sequenced the genome of *P. cynomolgi* strain B, isolated from a monkey in Malaysia and grown in splenectomized monkeys (Online Methods). A combination of Sanger, Roche 454 and Illumina chemistries was employed to generate a high-quality reference assembly at 161-fold coverage, consisting of 14 supercontigs (corresponding to the 14 parasite chromosomes) and ~1,649 unassigned contigs, comprising

[1]Laboratory of Malariology, Research Institute for Microbial Diseases, Osaka University, Suita, Japan. [2]Department of Biology, Center for Genomics and Systems Biology, New York University, New York, New York, USA. [3]Laboratory of Tropical Medicine and Parasitology, Institute of International Education and Research, Dokkyo Medical University, Shimotsuga, Japan. [4]Genome Information Research Center, Research Institute for Microbial Diseases, Osaka University, Suita, Japan. [5]Department of Molecular Protozoology, Research Institute for Microbial Diseases, Osaka University, Suita, Japan. [6]The Corporation for Production and Research of Laboratory Primates, Tsukuba, Japan. [7]Department of Protozoology, Institute of Tropical Medicine (NEKKEN) and Global COE (Centers of Excellence) Program, Nagasaki University, Nagasaki, Japan. [8]Department of Molecular and Cellular Parasitology, Graduate School of Medicine, Juntendo University, Tokyo, Japan. [9]Department of Biomedical Chemistry, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. [10]Tsukuba Primate Research Center, National Institute of Biomedical Innovation, Tsukuba, Japan. [11]Center for Global Health, Centers for Disease Control and Prevention, Divison of Parasitic Diseases and Malaria, Atlanta, Georgia, USA. [12]Center for Evolutionary Medicine and Informatics, The Biodesign Institute, Arizona State University, Tempe, Arizona, USA. [13]Present address: Career-Path Promotion Unit for Young Life Scientists, Kyoto University, Kyoto, Japan. [14]These authors jointly directed this work. Correspondence should be addressed to K.T. (kztanabe@biken.osaka-u.ac.jp) or J.M.C. (jane.carlton@nyu.edu).

Received 25 January; accepted 9 July; published online 5 August 2012; doi:10.1038/ng.2375

a total length of ~26.2 Mb (**Supplementary Table 1**). Comparing genomic features of *P. cynomolgi*, *P. knowlesi* and *P. vivax* reveals many similarities, including GC content (mean GC content of 40.5%), 14 positionally conserved centromeres and the presence of intrachromosomal telomeric sequences (ITSs; GGGTT(T/C)A), which were discovered in the *P. knowlesi* genome[9] but are absent in *P. vivax* (**Fig. 1**, **Table 1** and **Supplementary Table 2**).

We annotated the *P. cynomolgi* strain B genome using a combination of *ab initio* gene prediction programs trained on high-quality data sets and sequence similarity searches against the annotated *P. vivax* and *P. knowlesi* genomes. Not unexpectedly for species from the same monkey malaria clade, gene synteny along the 14 chromosomes is highly conserved, although numerous microsyntenic breaks are present in regions containing multigene families (**Fig. 2** and **Table 2**). This genome-wide view of synteny in six species of *Plasmodium* also identified two apparent errors in existing public sequence databases: an inversion in chromosome 3 of *P. knowlesi* and an inversion in chromosome 6 of *P. vivax*. The *P. cynomolgi* genome contains 5,722 genes, of which approximately half encode conserved hypothetical proteins of unknown function, as is the case in all the *Plasmodium* genomes sequenced to date. A maximum-likelihood phylogenetic tree constructed using 192 conserved ribosomal and translation- and transcription-related genes (**Supplementary Fig. 1**) confirms the close relationship of *P. cynomolgi* to *P. vivax* compared to five other *Plasmodium* species. Approximately 90% of genes (4,613) have reciprocal best-match orthologs in all three species (**Fig. 3**), enabling refinement of the existing *P. vivax* and *P. knowlesi* annotations (**Supplementary Table 3**). The high degree of gene orthology enabled us to identify specific examples of gene duplication (an important vehicle for genome evolution), including a duplicated homolog of *P. vivax Pvs28*—which encodes a sexual stage surface antigen that is a transmission-blocking vaccine candidate[13]—in *P. cynomolgi* (**Supplementary Table 4**). Genes common only to *P. cynomolgi* and *P. vivax* ($n = 214$) outnumber those that are restricted to *P. cynomolgi* and *P. knowlesi* ($n = 100$) or *P. vivax* and *P. knowlesi* ($n = 17$). Such figures establish the usefulness of *P. cynomolgi* as a model species for studying the more intractable *P. vivax*.

Notably, most of the genes specific to a particular species belong to multigene families (excluding hypothetical genes; **Table 2** and **Supplementary Table 5**). This suggests repeated lineage-specific gene duplication and/or gene deletion in multigene families within the three monkey malaria clade species. Moreover, copy numbers of the genes composing multigene families were generally greater in the *P. cynomolgi–P. vivax* lineage than in *P. knowlesi*, suggesting repeated gene duplication in the ancestral lineage of *P. cynomolgi* and *P. vivax* (or repeated gene deletion in the *P. knowlesi* lineage). Thus, the genomes of *P. cynomolgi*, *P. vivax* and *P. knowlesi* can largely be distinguished by variations in the copy number of multigene family members. Examples of such families include those that encode proteins involved in evasion of the human immune system (*vir*, *kir* and *SICAvar*) and invasion of host red blood cells (*dbp* and *rbp*).

In malaria-causing parasites, invasion of host erythrocytes, mediated by specific interactions between parasite ligands and erythrocyte receptors, is a crucial component of the parasite lifecycle. Of great interest are the *ebl* and *rbl* gene families, which encode parasite ligands required for the recognition of host erythrocytes. The *ebl* genes encode erythrocyte binding–like (EBL) ligands such as the Duffy-binding proteins (DBPs) that bind to Duffy antigen receptor for chemokines (DARC) on human and monkey erythrocytes. The *rbl* genes encode the reticulocyte binding–like (RBL) protein family, including reticulocyte-binding proteins (RBPs) in *P. cynomolgi* and *P. vivax*, and normocyte-binding proteins (NBPs) in *P. knowlesi*, which bind to unknown erythrocyte receptors[14]. We confirmed the presence of two *dbp* genes in *P. cynomolgi*[15] (**Supplementary Table 6**), in contrast to the one *dbp* and three *dbp* genes identified in *P. vivax* and *P. knowlesi*, respectively. This raises an intriguing hypothesis that *P. vivax* lost one *dbp* gene, and thus its infectivity of Old World monkey erythrocytes, after divergence from a common *P. vivax–P. cynomolgi* ancestor. This hypothesis is also supported by our identification of single-copy *dbp* genes in two other closely related Old World monkey malaria-causing parasites, *Plasmodium fieldi* and *Plasmodium simiovale*, which are incapable of infecting humans[16]. These two Old World monkey species lost one or more *dbp* genes during divergence that confer infectivity to humans, whereas *P. cynomolgi* and *P. knowlesi* retained *dbp* genes that allow invasion of human erythrocytes (**Supplementary Fig. 2**).

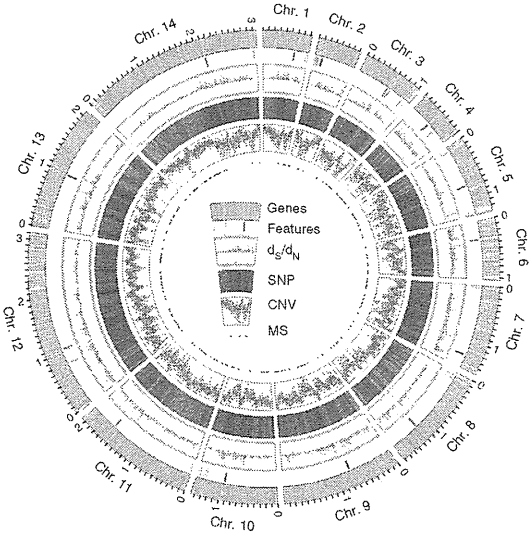

Figure 1 Architecture of the *P. cynomolgi* genome and associated genome-wide variation data. Data are shown for each of the 14 *P. cynomolgi* chromosomes. The six concentric rings, from outermost to innermost, represent (i) the location of 5,049 *P. cynomolgi* genes, excluding those on small contigs (cyan lines); (ii) genome features, including 14 centromeres (thick black lines), 43 telomeric sequence repeats (short red lines), 43 tRNA genes (red lines), 10 rRNAs (dark blue lines) and several gene family members, including 53 *cyir* (dark green lines), 8 *rbp* (brown lines), 13 *sera* (serine-rich antigen; pink lines), 25 *trag* (tryptophan-rich antigen; purple lines), 12 *msp3* (merozoite surface protein 3; light gray lines), 13 *msp7* (merozoite surface protein 7; gray lines), 25 *rad* (silver lines), 8 *etramp* (orange lines), 16 *Pf-fam-b* (light blue lines) and 7 *Pv-fam-d* (light green lines); (iii) plot of d_S/d_N for 4,605 orthologs depicting genome-wide polymorphism within *P. cynomolgi* strains B and Berok (black line) and divergence between *P. cynomolgi* strains B and Berok and *P. vivax* Salvador I (red line); a track above the plot indicates *P. cynomolgi* genes under positive selection (red) and purifying selection (blue), and a track below the plot indicates *P. cynomolgi–P. vivax* orthologs under positive selection (red) and purifying selection (blue); (iv) heatmap indicating SNP density of 3 *P. cynomolgi* strains plotted per 10-kb windows: red, 0–83 SNPs per 10 kb (regions of lowest SNP density); blue, 84–166 SNPs per 10 kb; green, 167–250 SNPs per 10 kb; purple, 251–333 SNPs per 10 kb; orange, 334–416 SNPs per10 kb; yellow, 417–500 SNPs per10 kb (regions of highest SNP density); (v) $\log_2$ ratio plot of CNVs identified from a comparison of *P. cynomolgi* strains B and Berok; and (vi) map of 182 polymorphic intergenic microsatellites (MS, black dots). The figure was generated using Circos software (see URLs).

Table 1 Comparison of genome features between *P. cynomolgi*, *P. vivax* and *P. knowlesi*, three species of the monkey malaria clade

Feature	*P. cynomolgi*	*P. vivax*[12]	*P. knowlesi*[9]
Assembly			
Size (Mb)	26.2	26.9	23.7
Number of scaffolds[a]	14 (1,649)	14 (2,547)	14 (67)
Coverage (fold)	161	10	8
GC content (%)	40.4	42.3	38.8
Genes			
Number of genes	5,722	5,432	5,197
Mean gene length (bp)	2,240	2,164	2,180
Gene density (bp per gene)[b]	4,428.2	4,950.5	4,416.1
Percentage coding[b]	51.0	47.1	49.0
Structural RNAs			
Number of tRNA genes	43	44	41
Number of 5S rRNA genes	3	3	0[c]
Number of 5.8S, 18S and 28S rRNA units	7	7	5
Nuclear genome			
Number of chromosomes	14	14	14
Number of centromeres	14	14	14
Isochore structure[d]	+	+	–
Mitochondrial genome			
Size (bp)[e]	5,986 (AB444123)	5,990 (AY598140)	5,958 (AB444108)
GC content (%)	30.3	30.5	30.5
Apicoplast genome			
Size (bp)	29,297[f]	5,064[g]	N/A
GC content (%)	13.0	17.1	N/A

N/A, not available.
[a]Small unassigned contigs indicated in parentheses. [b]Sequence gaps excluded. [c]Not present in *P. knowlesi* assembly version 4.0. [d]Regions of the genome that differ in density and are separable by CsCl centrifugation; isochores correspond to domains differing in GC content. [e]Identified in other studies (GenBank accessions given in parentheses). [f]Partial sequence (~86% complete) generated during this project. [g]Partial sequence of reference genome only published[12]; actual size is ~35 kb.

We found multiple *rbp* genes, some truncated or present as pseudogenes, in the *P. cynomolgi* genome (**Fig. 1** and **Table 2**). Phylogenetic analysis showed that *rbl* genes from *P. cynomolgi*, *P. vivax* and *P. knowlesi* can be classified into three distinct groups, RBP/NBP-1, RBP/NBP-2 and RBP/NBP-3 (**Supplementary Fig. 3**), and suggests that these groups existed before the three species diverged. All three groups of RBP/NBP are represented in *P. cynomolgi*, whereas *P. vivax* and *P. knowlesi* lack functional genes from the RBP/NBP-3 and RBP/NBP-1 groups, respectively. Thus, *rbl* gene family expansion seems to have occurred after speciation, indicating that the three species have multiple species-specific erythrocyte invasion mechanisms. Notably, we found an ortholog of *P. vivax rbp1b* in some strains of *P. cynomolgi* but not in others (**Supplementary Table 6**). To our knowledge, this is the first example of a CNV for a *rbp* gene between strains of a single *Plasmodium* species, highlighting how repeated creation and destruction of *rbl* genes, a signature of adaptive evolution, may have enabled species of the monkey malaria clade to expand or switch between monkey and human hosts.

The largest gene family in *P. cynomolgi*, consisting of 256 *cyir* (cynomolgi-interspersed repeat) genes, is part of the *pir* (plasmodium-interspersed repeat) superfamily that includes *P. vivax vir* genes (n = 319) and *P. knowlesi kir* genes (n = 70) (**Table 2**). *Pir*-encoded proteins reside on the surface of infected erythrocytes and have an important role in immune evasion[17]. Most *cyir* genes have sequence similarity to *P. vivax vir* genes (n = 254; **Supplementary Table 7**) and are found in subtelomeric regions (**Fig. 1**), but, notably, 11 *cyir* genes have sequence similarity to *P. knowlesi kir* genes (**Supplementary Table 7**) and occur more internally in the chromosomes, as do the *kir* genes in *P. knowlesi*. As with 'molecular mimicry' in *P. knowlesi* (mimicry of host sequences by pathogen sequences)[9], one CYIR protein (encoded by PCYB_032250) has a region of 56 amino acids that is highly similar to the extracellular domain of primate CD99 (**Supplementary Fig. 4**), a molecule involved in the regulation of T-cell function. A new finding is that *P. cynomolgi* has two genes whose sequences are similar to *P. knowlesi SICAvar* genes (**Supplementary Table 7**) that are expressed on the surfaces of schizont-infected macaque erythrocytes and are involved in antigenic variation[18].

The ability to form a dormant hypnozoite stage is common to both *P. cynomolgi* and *P. vivax* and was first shown in laboratory infections of monkeys by mosquito-transmitted *P. cynomolgi*[19]. In a search for candidate genes involved in the hypnozoite stage, we identified nine coding for 'dormancy-related' proteins that had the upstream ApiAP2 motifs[20] necessary for stage-specific transcriptional regulation at the sporozoite (pre-hypnozoite) stage (**Supplementary Table 8**). The candidates include kinases that are involved in cell cycle transition; hypnozoite formation may be regulated by phosphorylation of proteins specifically expressed at the pre-hypnozoite stage. Our list of *P. cynomolgi* candidate genes represents an informed starting point for experimental studies of this elusive stage.

We sequenced *P. cynomolgi* strains Berok (from Malaysia) and Cambodian (from Cambodia) to 26× and 17× coverage, respectively, to characterize *P. cynomolgi* genome-wide diversity through analysis of SNPs, CNVs and microsatellites. A comparison of the three *P. cynomolgi* strains identified 178,732 SNPs (**Supplementary Table 9**) at a frequency of 1 SNP per 151 bp, a polymorphism level somewhat

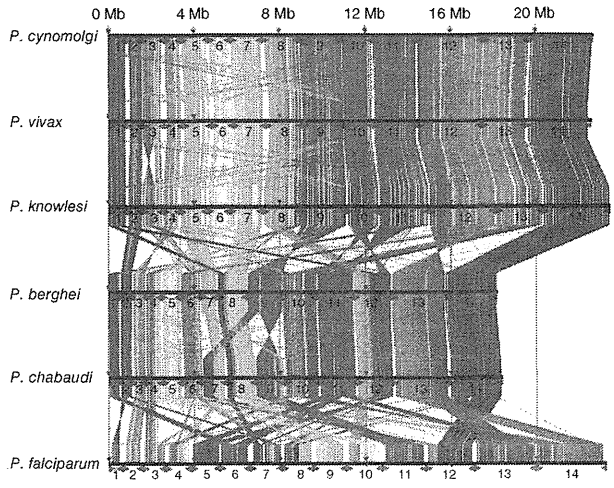

Figure 2 Genome synteny between six species of *Plasmodium* parasite. Protein-coding genes of *P. cynomolgi* are shown aligned with those of five other *Plasmodium* genomes: two species belonging to the monkey malaria clade, *P. vivax* and *P. knowlesi*; two species of rodent malaria, *P. berghei* and *P. chabaudi*; and *P. falciparum*. Highly conserved protein-coding regions between the genomes are colored in order from red (5′ end of chromosome 1) to blue (3′ end of chromosome 14) with respect to genomic position of *P. cynomolgi*.

Table 2 Components of multigene families of *P. cynomolgi*, *P. vivax* and *P. knowlesi* differ in copy number

Family	Multigene family	Localization	Arrangement	*P. cynomolgi*	*P. vivax*	*P. knowlesi*	Putative function and other information
1	*pir* (*vir*-like)	Subtelomeric	Scattered and clustered	254	319[a]	4	Immune evasion
2	*pir* (*kir*-like)	Subtelomeric and central	Scattered and clustered	11	2	66[a]	Immune evasion
3	*SICAvar*	Subtelomeric and central	Scattered and clustered	2	1	242[a]	Antigenic variation, immune evasion
4	*msp3*	Central	Clustered	12	12	3	Merozoite surface protein
5	*msp7*	Central	Clustered	13	13	5	Merozoite surface protein
6	*dbl* (*dbp/ebl*)	Subtelomeric	Scattered	2	1	3	Host cell recognition
7	*rbl* (*rbp/nbp/rh*)	Subtelomeric	Scattered	8[a]	10[a]	3[a]	Host cell recognition
8	*Pv-fam-a* (*trag*)	Subtelomeric	Scattered and clustered	36	36	26[a]	Tryptophan-rich
9	*Pv-fam-b*	Central	Clustered	3	6	1	Unknown
10	*Pv-fam-c*	Subtelomeric	Unknown[b]	1	7	0	Unknown
11	*Pv-fam-d* (*hypb*)	Subtelomeric	Scattered	18	16	2	Unknown
12	*Pv-fam-e* (*rad*)	Subtelomeric	Clustered	27	44	16	Unknown
13	*Pv-fam-g*	Central	Clustered	3	3	3	Unknown
14	*Pv-fam-h* (*hyp16*)	Central	Clustered	6	4	2	Unknown
15	*Pv-fam-i* (*hyp11*)	Subtelomeric	Scattered	6	6	5	Unknown
16	*Pk-fam-a*	Central	Scattered	0	0	12[a]	Unknown
17	*Pk-fam-b*	Subtelomeric	Scattered	0	0	9	Unknown
18	*Pk-fam-c*	Subtelomeric	Scattered	0	0	6[a]	Unknown
19	*Pk-fam-d*	Central	Scattered	0	0	3[a]	Unknown
20	*Pk-fam-e*	Subtelomeric	Scattered	0	0	3[a]	Unknown
21	*PST-A*	Subtelomeric and central	Scattered	9[a]	11[a]	7	αβ hydrolase
22	*ETRAMP*	Subtelomeric	Scattered	9	9	9	Parasitophorous vacuole membrane
23	*CLAG* (*RhopH-1*)	Subtelomeric	Scattered	2	3	2	High-molecular-wieght rhoptry antigen complex
24	*PvSTP1*	Subtelomeric	Unknown[b]	3	10[a]	0	Unknown
25	*PHIST* (Pf-*fam-b*)	Subtelomeric	Scattered and clustered	21	20	15	Unknown
26	*SERA*	Central	Clustered	13[a]	13[a]	8[a]	Cysteine protease

[a]Pseudogenes, truncated genes and gene fragments included. [b]Gene arrangement could not be determined due to localization on unassigned contigs.

similar to that found when *P. falciparum* genomes are compared[21,22]. We calculated the pairwise nucleotide diversity (π) as 5.41×10^{-3} across the genome, which varies little between the chromosomes. We assessed genome-wide CNVs between the *P. cynomolgi* B and Berok strains, using a robust statistical model in the CNV-seq program[23], by which we identified 1,570 CNVs (1 per 17 kb), including 1 containing the *rbp1b* gene on chromosome 7 (**Supplementary Fig. 5**). Finally, mining of the *P. cynomolgi* B and Berok strains identified 182 polymorphic intergenic microsatellites (**Supplementary Table 10**), the first set of genetic markers developed for this species. These provide a toolkit for studies of genetic diversity and population structure of laboratory stocks or natural infections of *P. cynomolgi*, many of which have recently been isolated from screening hundreds of wild monkeys for the zoonosis *P. knowlesi*[24].

We estimated the difference between the number of synonymous changes per synonymous site (d_S) and the number of nonsynonymous changes per nonsynonymous site (d_N) over 4,563 pairs of orthologs within *P. cynomolgi* strains B and Berok and 4,601 pairs of orthologs between these two *P. cynomolgi* strains and *P. vivax* Salvador I, using a simple Nei-Gojobori model[25]. We found 63 genes with $d_N > d_S$ within the two *P. cynomolgi* strains and 3,265 genes with $d_S > d_N$ (**Supplementary Table 11**). Genes with relatively high d_N/d_S ratios include those encoding transmembrane proteins, such as antigens and transporters, among which is a transmission-blocking target antigen, Pcyn230 (encoded by PCYB_042090). Notably, the *P. vivax* ortholog (PVX_003905) does not show evidence for positive selection[26], suggesting species-specific positive selection. We explored the degree to which evolution of orthologs has been constrained between *P. cynomolgi* and *P. vivax* and found 83 genes under possible accelerated evolution but 3,739 genes under possible purifying selection (**Supplementary Table 12**). This conservative estimate indicates that at least 81% of loci have diverged under strong constraint, compared with 1.8% of loci under less constraint or positive selection (**Fig. 1**), indicating that, overall, the genome of *P. cynomolgi* is highly conserved in single-locus genes compared to *P. vivax* and emphasizing the value of *P. cynomolgi* as a biomedical and evolutionary model for studying *P. vivax*.

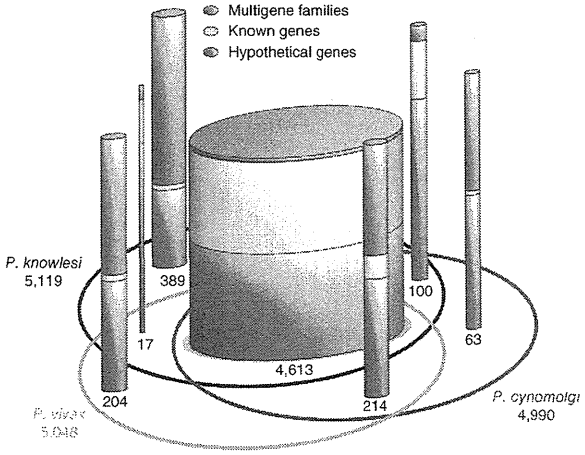

Figure 3 Comparison of the genes of *P. cynomolgi*, *P. vivax* and *P. knowlesi*. The Venn ellipses represent the three genomes, with the total number of genes assigned to the chromosomes indicated under the species name. Cylinders depict orthologous and non-orthologous genes between the three genomes, with the number of genes in each indicated and represented graphically by cylinder relative width. In each cylinder, genes are divided into three categories whose thickness is represented by colored bands proportional to category percentage.

Our generation of the first *P. cynomolgi* genome sequences is a critical step in the development of a robust model system for the intractable and neglected *P. vivax* species[27]. Comparative genome analysis of *P. vivax* and the Old World monkey malaria-causing parasites *P. cynomolgi* and *P. knowlesi* presented here provides the foundation for further insights into traits such as host specificity that will enhance prospects for the eventual elimination of vivax-caused malaria and global malaria eradication.

URLs. PlasmoDB, http://plasmodb.org/; Circos, http://circos.ca/; MIcroSAtelite Identification tool (MISA), http://pgrc.ipk-gatersleben.de/misa/; dbSNP, http://www.ncbi.nlm.nih.gov/projects/SNP/snp_viewBatch.cgi?sbid=1056645.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Sequence data for the *P. cynomolgi* B, Cambodian and Berok strains have been deposited in the DNA Data Bank of Japan (DDBJ), the European Molecular Biology Laboratory (EMBL) and the GenBank databases under the following accessions: B strain sequence reads DRA000196, genome assembly BAEJ01000001–BAEJ01003341 and annotation DF157093–DF158755; Cambodian strain sequence reads DRA000197; and Berok strain sequence reads SRA047950. SNP calls have been submitted to dbSNP (NYU_CGSB_BIO; 1056645) and may also be downloaded from the dbSNP website (see URLs). Sequences of the *dbp* genes from *P. cynomolgi* (Cambodian strain), *P. fieldi* (A.b.i. strain) and *P. simiovale* (AB617788–AB617791) and the *P. cynomolgi* Berok strain (JQ422035–JQ422036) and *rbp* gene sequences from the *P. cynomolgi* Berok and Cambodian strains (JQ422037–JQ422050) have been deposited. A partial apicoplast genome of the *P. cynomolgi* Berok strain has been deposited (JQ522954). The *P. cynomolgi* B reference genome is also available through PlasmoDB (see URLs).

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

We thank H. Sawai for suggestions on genome analysis, D. Fisher for help with genome-wide evolutionary analyses and the NYU Langone Medical Center Genome Technology Core for access to Roche 454 sequencing equipment (funded by grant S10 RR026950 to J.M.C. from the US National Institutes of Health (NIH)). Genome and phylogenetic analyses used the Genome Information Research Center in the Research Institute of Microbial Diseases at Osaka University. This work was supported by grants from the Ministry of Education, Culture, Sports, Science and Technology of Japan (18073013, 18GS03140013, 20390120 and 22406012) to K.T., an NIH grant (R01 GM080586) to A.A.E. and a Burroughs Wellcome Fund grant (1007398) and an NIH International Centers of Excellence for Malaria Research grant (U19 AI089676-01) to J.M.C. The content is soley the responsibility of the authors and does not necessarily represent the official views of the NIH.

AUTHOR CONTRIBUTIONS

K.T., J.M.C., A.A.E. and J.W.B. conceived and conducted the study. S.K., Y.K., Y.Y., S.-I.T. and J.W.B. provided *P. cynomolgi* material. S.N., N.G., T.Y. and H.R.K. constructed a computing system for data processing, and S.-I.T., H.H., P.L.S., S.A.S. and H.R.K. performed scaffolding of contigs and manual annotation of the predicted genes. S.N. performed sequence correction of supercontigs and gene prediction. S.-I.T., S.N., N.G., N.A., M.Y., O.K., K.T., H.R.K., R.S., S.A.S. and J.M.C. analyzed data. S.-I.T., N.M.Q.P., T.T., T.M., K.K., J.M.C., T.H., A.A.E., J.W.B. and K.T. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/ng.2375.

1. Mendis, K., Sina, B.J., Marchesini, P. & Carter, R. The neglected burden of *Plasmodium vivax* malaria. *Am. J. Trop. Med. Hyg.* **64**, 97–106 (2001).
2. Mueller, I. *et al.* Key gaps in the knowledge of *Plasmodium vivax*, a neglected human malaria parasite. *Lancet Infect. Dis.* **9**, 555–566 (2009).
3. Baird, J.K. Resistance to chloroquine unhinges vivax malaria therapeutics. *Antimicrob. Agents Chemother.* **55**, 1827–1830 (2011).
4. Rayner, J.C., Liu, W., Peeters, M., Sharp, P.M. & Hahn, B.H. A plethora of *Plasmodium* species in wild apes: a source of human infection? *Trends Parasitol.* **27**, 222–229 (2011).
5. Cornejo, O.E. & Escalante, A.A. The origin and age of *Plasmodium vivax*. *Trends Parasitol.* **22**, 558–563 (2006).
6. Escalante, A.A. *et al.* A monkey's tale: the origin of *Plasmodium vivax* as a human malaria parasite. *Proc. Natl. Acad. Sci. USA* **102**, 1980–1985 (2005).
7. Mu, J. *et al.* Host switch leads to emergence of *Plasmodium vivax* malaria in humans. *Mol. Biol. Evol.* **22**, 1686–1693 (2005).
8. Singh, B. *et al.* A large focus of naturally acquired *Plasmodium knowlesi* infections in human beings. *Lancet* **363**, 1017–1024 (2004).
9. Pain, A. *et al.* The genome of the simian and human malaria parasite *Plasmodium knowlesi*. *Nature* **455**, 799–803 (2008).
10. Eyles, D.E., Coatney, G.R. & Getz, M.E. Vivax-type malaria parasite of macaques transmissible to man. *Science* **131**, 1812–1813 (1960).
11. Gibbs, R.A. *et al.* Evolutionary and biomedical insights from the rhesus macaque genome. *Science* **316**, 222–234 (2007).
12. Carlton, J.M. *et al.* Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. *Nature* **455**, 757–763 (2008).
13. Saxena, A.K., Wu, Y. & Garboczi, D.N. *Plasmodium* p25 and p28 surface proteins: potential transmission-blocking vaccines. *Eukaryot. Cell* **6**, 1260–1265 (2007).
14. Iyer, J., Gruner, A.C., Renia, L., Snounou, G. & Preiser, P.R. Invasion of host cells by malaria parasites: a tale of two protein families. *Mol. Microbiol.* **65**, 231–249 (2007).
15. Okenu, D.M., Malhotra, P., Lalitha, P.V., Chitnis, C.E. & Chauhan, V.S. Cloning and sequence analysis of a gene encoding an erythrocyte binding protein from *Plasmodium cynomolgi*. *Mol. Biochem. Parasitol.* **89**, 301–306 (1997).
16. Coatney, G.R., Collins, W.E., Warren, M. & Contacos, P.G. *The Primate Malarias* (US Department of Health, Education and Welfare, Washington, DC, 1971).
17. Cunningham, D., Lawton, J., Jarra, W., Preiser, P. & Langhorne, J. The *pir* multigene family of *Plasmodium*: antigenic variation and beyond. *Mol. Biochem. Parasitol.* **170**, 65–73 (2010).
18. al-Khedery, B., Barnwell, J.W. & Galinski, M.R. Antigenic variation in malaria: a 3′ genomic alteration associated with the expression of a *P. knowlesi* variant antigen. *Mol. Cell* **3**, 131–141 (1999).
19. Krotoski, W.A. The hypnozoite and malarial relapse. *Prog. Clin. Parasitol.* **1**, 1–19 (1989).
20. Campbell, T.L., De Silva, E.K., Olszewski, K.L., Elemento, O. & Llinas, M. Identification and genome-wide prediction of DNA binding specificities for the ApiAP2 family of regulators from the malaria parasite. *PLoS Pathog.* **6**, e1001165 (2010).
21. Mu, J. *et al.* Genome-wide variation and identification of vaccine targets in the *Plasmodium falciparum* genome. *Nat. Genet.* **39**, 126–130 (2007).
22. Volkman, S.K. *et al.* A genome-wide map of diversity in *Plasmodium falciparum*. *Nat. Genet.* **39**, 113–119 (2007).
23. Xie, C. & Tammi, M.T. CNV-seq, a new method to detect copy number variation using high-throughput sequencing. *BMC Bioinformatics* **10**, 80 (2009).
24. Lee, K.S. *et al. Plasmodium knowlesi*: reservoir hosts and tracking the emergence in humans and macaques. *PLoS Pathog.* **7**, e1002015 (2011).
25. Nei, M. & Gojobori, T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**, 418–426 (1986).
26. Doi, M. *et al.* Worldwide sequence conservation of transmission-blocking vaccine candidate Pvs230 in *Plasmodium vivax*. *Vaccine* **29**, 4308–4315 (2011).
27. Carlton, J.M., Sina, B.J. & Adams, J.H. Why is *Plasmodium vivax* a neglected tropical disease? *PLoS Negl. Trop. Dis.* **5**, e1160 (2011).

ONLINE METHODS

Parasite material. Details of the origin of the *P. cynomolgi* B, Berok and Cambodian strains, their growth in macaques and isolation of parasite material are given in the **Supplementary Note.**

Genome sequencing and assembly. *P. cynomolgi* B strain was sequenced using the Roche 454 GS FLX (Titanium) and Illumina/Solexa Genome Analyzer IIx platforms to 161× coverage. In addition, 2,784 clones (6.8 Mb) of a ~40-kb insert fosmid library in pCC1FOS (EpiCentre Biotechnologies) was sequenced by the Sanger method. A draft assembly of strain B was constructed using a combination of automated assembly and manual gap closure. We first generated *de novo* contigs by assembling Roche 454 reads using GS *De novo* Assembler version 2.0 with default parameters. Contigs of >500 bp were mapped to the *P. vivax* Salvador I reference assembly[12] (PlasmoDB; see URLs). *P. cynomolgi* contigs were iteratively arrayed through alignment to *P. vivax*–assembled sequences with manual corrections. A total of 1,264 aligned contigs were validated by mapping paired-end reads from fosmid clones using blastn ($e < 1 \times 10^{-15}$; identity > 90%; coverage > 200 bp) implemented in GenomeMatcher software version 1.65 (ref. 28). Additional linkages (699 regions) were made using PCR across the intervening sequence gaps with primers designed from neighboring contigs. The length of sequence gaps was estimated from insert lengths of the fosmid paired-end reads, the size of PCR products and homologous sequences of the *P. vivax* genome. Supercontigs were then manually constructed from the aligned contigs. Eventually, we obtained 14 supercontigs corresponding to the 14 chromosomes of the parasite, with a total length of ~22.73 Mb, encompassing ~80% of the predicted *P. cynomolgi* genome. A total of 1,651 contigs (>1 kb) with a total length of 3.45 Mb was identified as unassigned subtelomeric sequences by searching against the *P. vivax* genome using blastn. Additionally, to improve sequence accuracy, we constructed a mapping assembly of Illumina paired-end reads and the 14 supercontigs and unassigned contigs as reference sequences using CLC Genomics Workbench version 3.0 with default settings (CLC Bio). Comparison of the draft *P. cynomolgi* B sequence with 23 *P. cynomolgi* protein-coding genes (64 kb) obtained by Sanger sequencing showed 99.8% sequence identity (**Supplementary Table 13**). The *P. cynomolgi* Berok and Cambodian strains were sequenced to 26× and 17× coverage, respectively, using the Roche 454 GS FLX platform, with single-end and 3-kb paired-end libraries made for the former and a single-end library only made for the latter. For phylogenetic analyses of specific genes, sequences were independently verified by Sanger sequencing (**Supplementary Table 14** and **Supplementary Note**).

Prediction and annotation of genes. Gene prediction for the 14 supercontigs and 1,651 unassigned contigs was performed using the MAKER genome annotation pipeline[29] with *ab initio* gene prediction programs trained on proteins and ESTs from PlasmoDB Build 7.1. For gene annotation, blastn ($e < 1 \times 10^{-15}$; identity > 70%; coverage > 100 bp) searches of *P. vivax* (PvivaxAnnotated Transcripts_PlasmoDB-7.1.fasta) and *P. knowlesi* (PknowlesiAnnotatedTranscripts_PlasmoDB-7.1.fasta) predicted proteomes were run, and the best hits were identified. All predicted genes were manually inspected at least twice for gene structure and functional annotation, and orthologous relationships between *P. cynomolgi*, *P. vivax* and *P. knowlesi* were determined on synteny. A unique identifier, PCYB_######, was assigned to *P. cynomolgi* genes, where the first two of the six numbers indicate chromosome number. Paralogs of genes that seemed to be specific to either *P. cynomolgi*, *P. vivax* or *P. knowlesi* were searched using blastp with default parameters, using a cutoff *e* value of 1×10^{-16}.

Multiple genome sequence alignment. Predicted proteins of *P. cynomolgi* B strain were concatenated and aligned with those from the 14 chromosomes of 5 other *Plasmodium* genomes: *P. vivax*, *P. knowlesi*, *P. falciparum*, *P. berghei* and *P. chabaudi*, using Murasaki software version 1.68.6 (ref. 30).

Search for sequence showing high similarity to host proteins. Eleven *P. cynomolgi* CYIR proteins (with sequence similarity to *P. knowlesi* KIR) were subjected to blastp search for regions having high similarity to host *Macacca mulatta* CD99 protein, with cutoff *e* value of 1×10^{-12} and compositional adjustment (no adjustment) against the nonredundant protein sequence data set of the *M. mulatta* proteome in NCBI.

Phylogenetic analyses. Genes were aligned using ClustalW version 2.0.10 (ref. 31) with manual corrections, and unambiguously aligned sites were selected for phylogenetic analyses. Maximum-likelihood phylogenetic trees were constructed using PROML programs in PHYLIP version 3.69 (ref. 32) under the Jones-Taylor-Thornton (JTT) amino-acid substitution model. To take the evolutionary rate heterogeneity across sites into consideration, the R (hidden Markov model rates) option was set for discrete γ distribution, with eight categories for approximating the site-rate distribution. CODEML programs in PAML 4.4 (ref. 33) were used for estimating the γ shape parameter, α values. For bootstrap analyses, SEQBOOT and CONSENSE programs in PHYLIP were applied.

Candidate genes for hypnozoite formation. We undertook two approaches. First, genes unique to *P. vivax* and *P. cynomolgi* (hypnozoite-forming parasites) and not found in other non-hypnozoite–forming *Plasmodium* species were identified. We used the 147 unique genes identified in the *P. vivax* genome[12] to search the *P. cynomolgi* B sequence. For the orthologs identified in both species, ~1 kb of sequence 5′ to the coding sequence was searched for four specific ApiAP2 motifs[20]—PF14 0633, GCATGC; PF13_0235_D1, GCCCCG; PFF0670w_D1, TAAGCC; and PFD0985w_D2, TGTTAC—which are involved in sporozoite stage–specific regulation and expression (corresponding to the pre-hypnozoite stage). Second, dormancy-related proteins were retrieved from GenBank and used to search for *P. vivax* homologs. Candidate genes ($n = 128$) and orthologs of *P. cynomolgi* and five other parasite species were searched in the region ~1 kb upstream of the coding sequence for the presence of the four ApiAP2 motifs. Data for *P. vivax*, *P. knowlesi*, *P. falciparum*, *P. berghei*, *Plasmodium chabaudi* and *Plasmodium yoelii* were retrieved from PlasmoDB Build 7.1.

Genome-wide screen for polymorphisms. For SNP identification, alignment of Roche 454 data from strains B, Berok and Cambodian was performed using SSAHA2 (ref. 34), with 0.1 mismatch rate and only unique matches reported. Potential duplicate reads generated during PCR amplification were removed, so that when multiple reads mapped at identical coordinates, only the reads with the highest mapping quality were retained. We used a statistical method[35] implemented in SAMtools version 0.1.18 to call SNPs simultaneously in the case of duplicate runs of the same strain. SNPs with high read depth (>100) were filtered out, as were SNPs in poor alignment regions at the ends of chromosomes (**Supplementary Note**).

Nucleotide diversity (π) was calculated as follows. For each site being compared, we calculated allele frequency by counting the two alleles and measured the proportion of nucleotide differences. Letting π be the genetic distance between allele *i* and allele *j*, then the nucleotide diversity within the population is

$$\pi = \sum_{i,j} P_i P_j \pi_{ij}$$

where P_i and P_j are the overall allele frequencies of *i* and *j*, respectively. Mean π was calculated by averaging over sites, weighting each by $\sum_{i=1}^{n-1} \frac{1}{i}$, where *n* is the number of aligned sites. Average d_N/d_S ratios were estimated using the modified Nei-Gojobori/Jukes-Cantor method in MEGA 4 (ref. 36).

CNV-seq[23] was used to identify potential CNVs in *P. cynomolgi*. Briefly, this method is based on a statistical model that allows confidence assessment of observed copy-number ratios from next-generation sequencing data. Roche 454 sequences from *P. cynomolgi* strain B assembly were used as the reference genome, and the *P. cynomolgi* Berok strain was used as a test genome; the sequence coverage of the Cambodian strain was considered too low for analysis. The test reads were mapped to the reference genome, and CNVs were detected by computing the number of reads for each test strain in a sliding window. The validity of the observed ratios was assessed by the computation of a probability of a random occurrence, given no copy-number variation.

Polymorphic microsatellites (defined as repeat units of 1–6 nucleotides) between *P. cynomolgi* strains B and Berok were identified by aligning contigs

from a *de novo* assembly of Berok (generated using Roche GS Assembler version 2.6, with 40-bp minimum overlap, 90% identity) to the B strain using the Burrows-Wheeler Aligner (BWA)[37] and allowing for gaps. Using the Phred-scaled probability of the base being misaligned by SAMtools[35], indel candidates were called from the alignment. In-house Python scripts were used to then cross-reference with the microsatellites found in the reference strain B assembly identified by MISA (see URLs). All homopolymer microsatellites were discarded to account for potential sequence errors introduced by 454 sequencing.

Selective constraint analysis of 4,563 orthologs between *P. cynomolgi* strains B and Berok and 4,601 orthologs between these strains and *P. vivax* Salvador I used MUSCLE[38] alignments with stringent removal of gaps and missing data (*P. cynomolgi* Berok orthologs were identified through a reciprocal best-hit BLAST search against strain B genes). Analyses were conducted using the Nei-Gojobori model[25]. To detect values that could not be explained by chance, we estimated the standard error by a bootstrap procedure with 200 pseudoreplicates for each gene. The expected value for d_S/d_N is 0 if a given pair of sequences is diverging without obvious effects on fitness. In the case of the comparison within *P. cynomolgi*, values with a difference of ± 2 s.e.m. from 0 were considered indicative of an excess of synonymous ($d_S/d_N > 0$) or nonsynonymous ($d_S/d_N < 0$) changes. In the case of the comparison between *P. cynomolgi* and *P. vivax*, we used a more stringent criterion of ± 3 s.e.m. from 0.

28. Ohtsubo, Y., Ikeda-Ohtsubo, W., Nagata, Y. & Tsuda, M. GenomeMatcher: a graphical user interface for DNA sequence comparison. *BMC Bioinformatics* **9**, 376 (2008).
29. Cantarel, B.L. *et al.* MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* **18**, 188–196 (2008).
30. Popendorf, K., Tsuyoshi, H., Osana, Y. & Sakakibara, Y. Murasaki: a fast, parallelizable algorithm to find anchors from multiple genomes. *PLoS ONE* **5**, e12651 (2010).
31. Thompson, J.D., Higgins, D.G. & Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
32. Felsenstein, J. *PHYLIP, Phylogeny Inference Package*, 3.6a3 edn (University of Washington, Seattle, 2005).
33. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
34. Ning, Z., Cox, A.J. & Mullikin, J.C. SSAHA: a fast search method for large DNA databases. *Genome Res.* **11**, 1725–1729 (2001).
35. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
36. Tamura, K., Dudley, J., Nei, M. & Kumar, S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**, 1596–1599 (2007).
37. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
38. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

Virology 432 (2012) 29–38

Contents lists available at SciVerse ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

Trans-complemented hepatitis C virus particles as a versatile tool for study of virus assembly and infection

Ryosuke Suzuki [a,*], Kenji Saito [a], Takanobu Kato [a], Masayuki Shirakura [b], Daisuke Akazawa [a], Koji Ishii [a], Hideki Aizaki [a], Yumi Kanegae [c], Yoshiharu Matsuura [d], Izumu Saito [c], Takaji Wakita [a], Tetsuro Suzuki [e,**]

[a] *Department of Virology II, National Institute of Infectious Diseases, 1-23-1 Toyama, Shinjuku-ku, Tokyo 162-8640, Japan*
[b] *Influenza Virus Research Center, National Institute of Infectious Diseases, Tokyo 208-0011, Japan*
[c] *Institute of Medical Science, University of Tokyo, Tokyo 108-8639, Japan*
[d] *Research Institute for Microbial Diseases, Osaka University, Osaka 565-0871, Japan*
[e] *Department of Infectious Diseases, Hamamatsu University School of Medicine, 1-20-1 Handayama, Higashi-ku, Hamamatsu, Shizuoka 431-3192, Japan*

ARTICLE INFO

Article history:
Received 30 March 2012
Returned to author for revisions
23 April 2012
Accepted 25 May 2012
Available online 22 June 2012

Keywords:
HCV
HCVtcp
Trans-packaging
Single-round infection

ABSTRACT

In this study, we compared the entry processes of *trans*-complemented hepatitis C virus particles (HCVtcp), cell culture-produced HCV (HCVcc) and HCV pseudoparticles (HCVpp). Anti-CD81 antibody reduced the entry of HCVtcp and HCVcc to almost background levels, and that of HCVpp by approximately 50%. Apolipoprotein E-dependent infection was observed with HCVtcp and HCVcc, but not with HCVpp, suggesting that the HCVtcp system is more relevant as a model of HCV infection than HCVpp. We improved the productivity of HCVtcp by introducing adapted mutations and by deleting sequences not required for replication from the subgenomic replicon construct. Furthermore, blind passage of the HCVtcp in packaging cells resulted in a novel mutation in the NS3 region, N1586D, which contributed to assembly of infectious virus. These results demonstrate that our plasmid-based system for efficient production of HCVtcp is beneficial for studying HCV life cycles, particularly in viral assembly and infection.

Introduction

Over 170 million people worldwide are chronically infected with hepatitis C virus (HCV), and are at risk of developing chronic liver diseases (Hoofnagle, 2002). HCV is an enveloped virus of the family *Flaviviridae*, and its genome is a positive-strand RNA consisting of the 5′-untranslated region (UTR), an open reading frame encoding viral proteins (core, E1, E2, p7, NS2, NS3, NS4A, NS4B, NS5A, and NS5B) and the 3′-UTR (Suzuki et al., 2007).

Host–virus interactions are required during the initial steps of viral infection. It was previously reported that CD81 (Bartosch et al., 2003a, b; McKeating et al., 2004; Pileri et al., 1998), scavenger receptor class B type I (Bartosch et al., 2003a, b; Scarselli et al., 2002), claudin-1 (Evans et al., 2007; Liu et al., 2009) and occludin (Benedicto et al., 2009; Evans et al., 2007; Liu et al., 2009; Ploss et al., 2009) are critical molecules for HCV entry into cells. CD81 interacts with HCV E2 via a second extracellular loop (Bartosch et al., 2003a, b; Hsu et al., 2003) and its role in the internalization process was confirmed (Cormier et al., 2004; Flint et al., 2006). It has also been shown that infectious HCV particles produced in cell cultures (HCVcc) exist as apolipoprotein E (ApoE)-enriched lipoprotein particles (Chang et al., 2007) and that ApoE is important for HCV infectivity (Owen et al., 2009).

Investigation of HCV had been hampered by difficulties in amplifying the virus in vitro before development of robust cell culture systems based on JFH-1 isolates (Lindenbach et al., 2005; Wakita et al., 2005; Zhong et al., 2005). Retrovirus-based HCV pseudoparticles (HCVpp), in which cell entry is dependent on HCV glycoproteins, have been used to study virus entry (Bartosch et al., 2003a; Hsu et al., 2003). Vesicular stomatitis virus (VSV)-based pseudotypic viruses bearing HCV E1 and E2 and replication-competent recombinant VSV encoding HCV envelopes have also been available as surrogate models for studies of HCV infection (Mazumdar et al., 2011; Tani et al., 2007).

It was recently shown that HCV subgenomic replicons can be packaged when structural proteins are supplied in *trans* (Adair et al., 2009; Ishii et al., 2008; Masaki et al., 2010; Steinmann et al., 2008). These *trans*-complemented HCV particles (HCVtcp) are infectious, but support only single-round infection and are unable to spread. Establishment of flexible systems to efficiently produce HCVtcp should contribute to studying HCV assembly, in particular encapsidation of the viral genome, and entry to cells with less stringent biosafety and biosecurity measures. Although single-round infection can be achieved by using the HCVcc system with receptor knock-out

* Corresponding author. Fax: +81 3 5285 1161.
** Corresponding author. Fax: +81 53 435 2338.
E-mail addresses: ryosuke@nih.go.jp (R. Suzuki), tesuzuki@hama-med.ac.jp (T. Suzuki).

0042-6822/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.virol.2012.05.033

cells, the single-round HCVcc system is not suitable for studying virus entry. We previously described plasmid-based production of HCVcc and HCVtcp (Masaki et al., 2010). Here, we demonstrated that HCVtcp production can be enhanced by introducing the previously reported cell-culture adaptive mutations and by deleting sequences not essential for replication in the subgenomic replicon construct. By providing genotype 1b-derived core-to-p7 in addition to intragenotypic viral proteins, chimeric HCVtcp were generated. Furthermore, blind passage of HCVtcp in the packaging cells resulted in the identification of a novel cell culture-adaptive mutation in NS3 that enables us to establish the efficient production of HCVtcp with structural proteins from various strains. Taken together, our system for producing single-cycle infectious HCV particles should be useful in the study of entry and assembly steps of the HCV life cycles. This technology may also have potential to be the basis for the safer vaccine development.

Results

Enhancement of HCVtcp production by adaptive mutations in E2, p7 and NS2 and by deleting sequences not essential for replication from replicon construct

In our HCVtcp system, the RNA polymerase I (Pol I)-driven replicon plasmid, which carries a dicistronic subgenomic luciferase reporter replicon of JFH-1 strain with a Pol I promoter and terminator (pHH/SGR-Luc), as well as a plasmid containing core-NS2 cDNA under the CAG promoter (pCAGC-NS2) were used (Masaki et al., 2010). In an effort to improve the yield of HCVtcp production, cell culture-adaptive mutations in E2 (N417S), p7 (N765D) and NS2 (Q1012R) which were previously selected from serial passage of HCVcc (Russell et al., 2008) were introduced into the core-NS2 expression plasmid (Fig. 1A) (residues are numbered

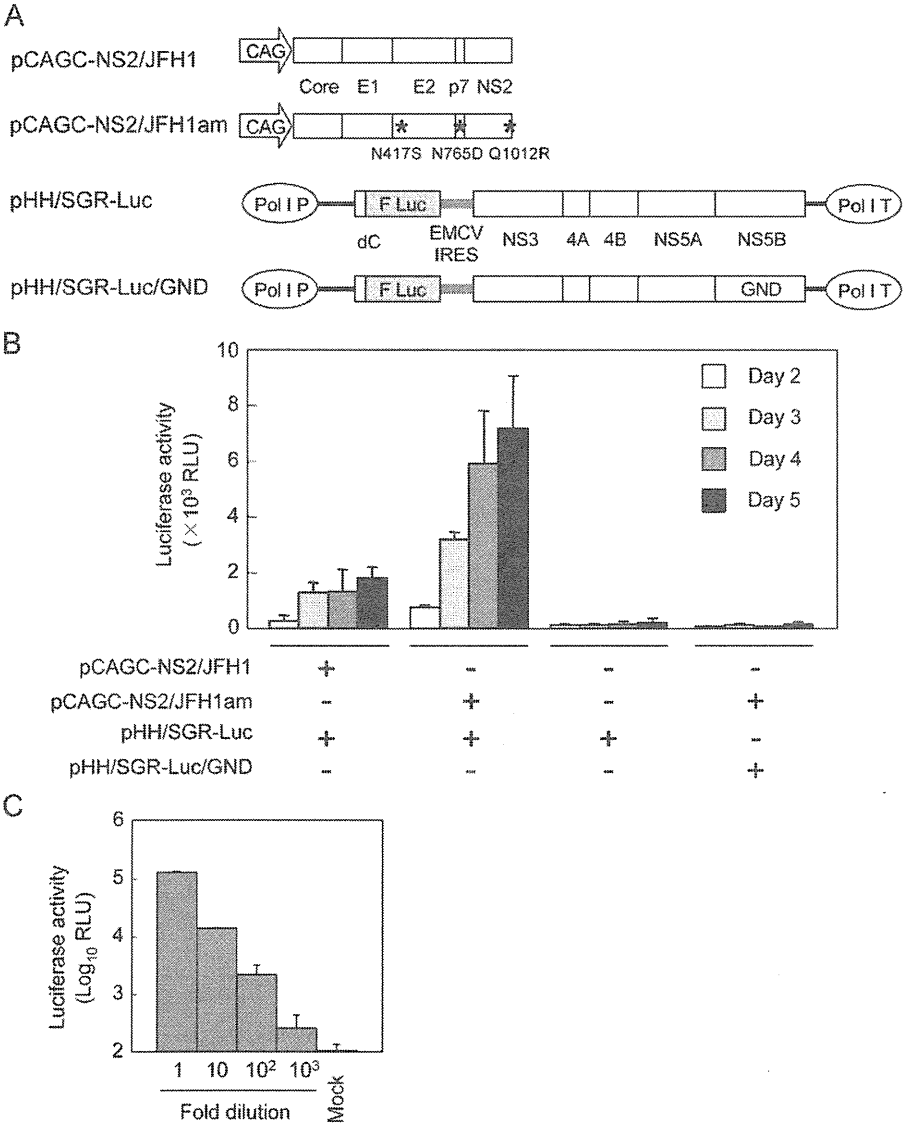

Fig. 1. HCVtcp production by two-plasmid transfection. (A) Schematic representation of plasmids is shown. HCV polyproteins derived from JFH-1 are indicated by white boxes. HCV UTRs are indicated by bold lines. The internal ribosomal entry site from encephalomyocarditis virus (EMCV IRES) is denoted as gray lines. Adaptive mutations are indicated as asterisks. F Luc: firefly luciferase gene; CAG: CAG promoter; Pol I P: RNA polymerase I promoter; Pol I T: RNA polymerase I terminator; GND: replication-deficient GND mutation. (B) Luciferase activity in Huh7.5.1 cells inoculated with supernatant from cells transfected with indicated plasmids at the indicated time points. Data are averages of triplicate values with error bars showing standard deviations. (C) Luciferase activity in cells inoculated with serially diluted HCVtcp.

according to positions within the JFH-1 polyprotein). Supernatants of cells transfected with plasmids (Fig. 1A) were collected and were used to infect Huh7.5.1 cells, which were analyzed by luciferase assay. Introduction of adaptive mutations (pCAGC-NS2/JFH1am) resulted in more than 4-fold higher production of HCVtcp at 5 day post-transfection, as compared to wild-type (WT) (pCAGC-NS2/JFH1) (Fig. 1B), indicating that the adaptive mutations contribute to enhancing HCVtcp production. To confirm that luciferase activity levels in HCVtcp-infected cells are correlated with the number of infectious particles, Huh7.5.1 cells were inoculated with serial dilutions of HCVtcp. Luciferase activity was well correlated with viral load (Fig. 1C), indicating that luciferase assay in HCVtcp-infected cells can be used to quantify HCV infection.

In order to further explore the efficient production of HCVtcp, we generated replicon constructs that lack the luciferase gene or include the partial coding sequences for structural proteins instead of reporter (Fig. 2A). Replication of each replicon in plasmid-transfected cells was then assessed by Western blotting (Fig. 2B). Among the constructs tested, NS5B levels were lowest in cells expressing pHH/SGR-Luc. NS5B levels in cells replicating other replicons appeared to be comparable. Cells were infected with supernatants of cells transfected with each replicon plasmid, along with pCAGC-NS2/JFH1am, followed by infectious unit assay (Fig. 2C). The highest production of HCVtcp was obtained from cells transfected with pHH/SGR, where the luciferase sequence was deleted from pHH/SGR-Luc, thus suggesting that deletion of the sequence not essential for RNA replication in the replicon may contribute to enhancing HCVtcp production.

Production of chimeric HCVtcp by providing heterologous core-p7

In order to elucidate whether *trans*-encapsidation of JFH-1 replicon can be achieved by providing core-p7 from other HCV strains, core-NS2 plasmids were constructed (Fig. 3A). In these plasmids, core through the N-terminal 33 aa of NS2, which contains transmembrane domain 1 of NS2, was derived from either H77c (genotype 1a), THpa (genotype 1b), Con1 (genotype 1b) or J6 (genotype 2a) strain. Residual NS2 was derived from JFH-1, as described previously (Pietschmann et al., 2006). HCVtcp was efficiently produced by core-p7 of J6 and THpa strains, but its production was less efficient in the case of Con1 strain. *Trans*-packaging was not detectable when core-p7 of H77c strain was used (Fig. 3C). Among HCV strains tested, difference in luciferase activity levels in HCVtcp-infected cells (Fig. 3C) were in agreement with that in the viral RNA levels in the culture supernatants of the transfected cells (Fig. 3B). Although the efficacy of *trans*-complementation was variable among strains, chimeric HCVtcp can be generated by providing genotype 1b-derived core-p7 in addition to intragenotypic viral proteins, and was used in subsequent studies.

ApoE- and CD81-dependent infection by HCVtcp

There is accumulating evidence that apolipoproteins, particularly ApoE, contribute to HCV production and infectivity (Chang et al., 2007; Owen et al., 2009). To determine whether ApoE is involved in infection of target cells by HCVtcp, we infected cells in the presence of increasing concentrations of anti-ApoE antibody.

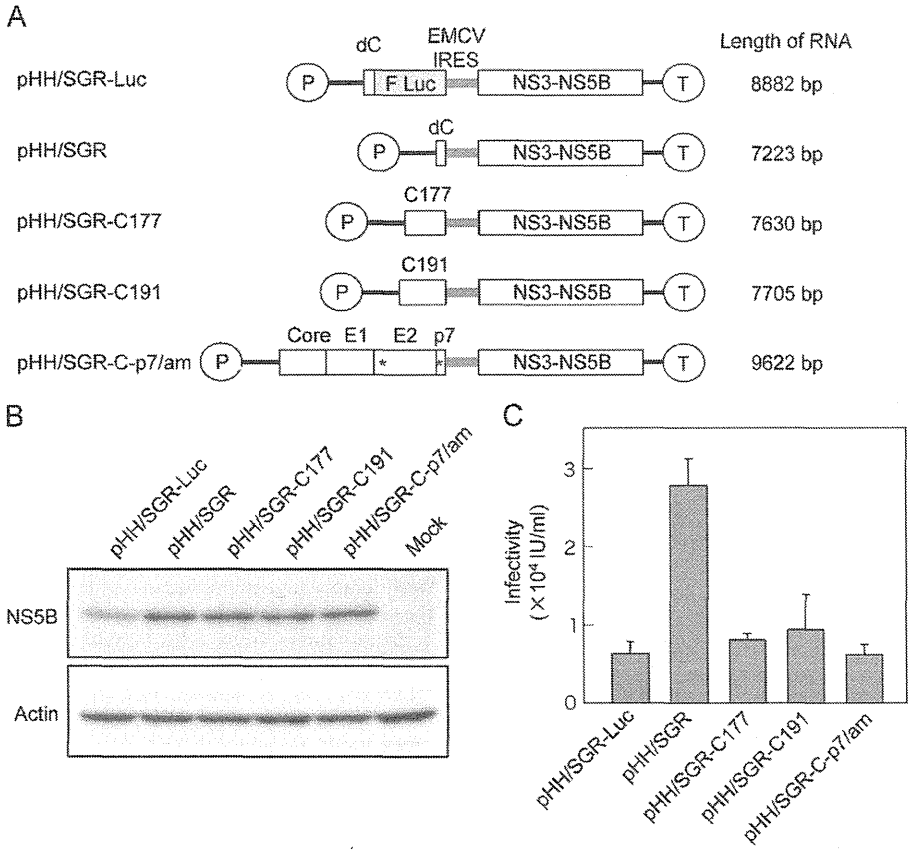

Fig. 2. Production of HCVtcp with different replicon constructs. (A) Schematic representation of plasmids used for production of HCVtcp. Deduced length of transcribed RNA from each construct is shown on the right. HCV polyproteins from JFH-1 strain are indicated by open boxes. HCV UTRs are indicated by bold lines. The EMCV IRES is denoted by gray bars. Adaptive mutations are indicated by asterisks. F Luc: firefly luciferase gene; P: RNA polymerase I promoter; T: RNA polymerase I terminator. (B) Detection of NS5B and actin in Huh7.5.1 cells transfected with indicated plasmids at 4 day post-transfection. (C) Infectivity of culture supernatants from cells transfected with indicated replicon plasmids along with pCAGC-NS2/JFH1am at 4 day post-transfection.

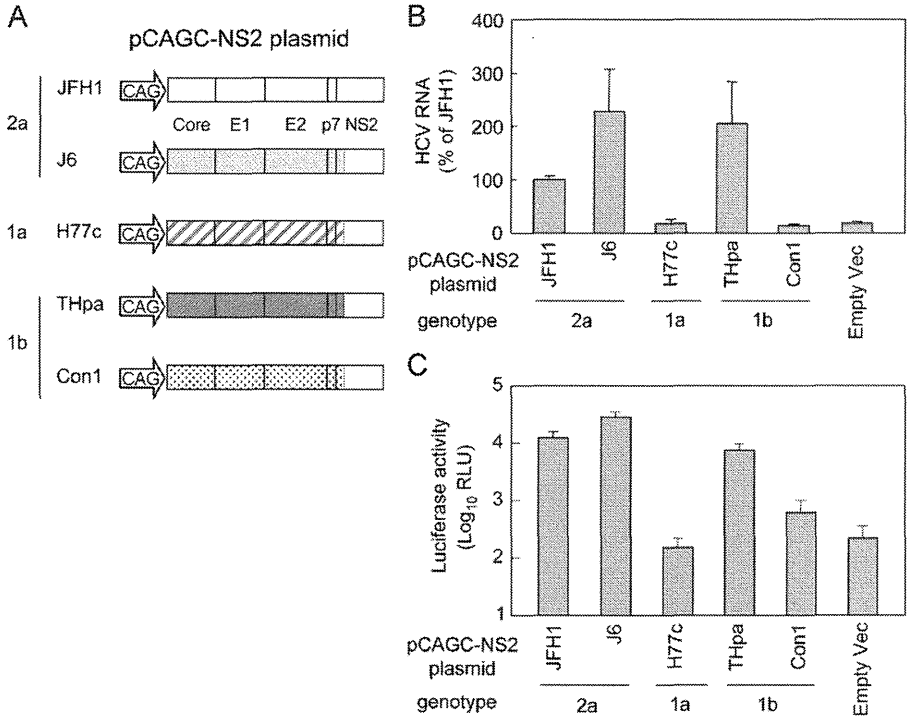

Fig. 3. HCVtcp production with structural proteins from various strains. (A) Schematic representation of plasmids used. HCV polyproteins of JFH-1, J6, H77c, THpa and Con1 strain are shown in the open box, bright gray box, box with diagonal lines, dark gray box and dotted box, respectively. (B) Relative levels of HCV RNA in the supernatant from cells transfected with indicated plasmids along with pHH/SGR-Luc. (C) Luciferase activity in cells inoculated with supernatant from cells transfected with indicated plasmids along with pHH/SGR-Luc at 4 day post-transfection.

pCAGC-NS2/THpa and pCAGC-NS2/JFH1am were used as core-NS2 plasmids for HCVtcp production carrying core-p7 derived from genotypes 1b and 2a (HCVtcp-1b and HCVtcp-2a, respectively). HCVpp derived from JFH-1 and VSVpp were generated and used for comparison. Infection with HCVtcp-1b or HCVtcp-2a was blocked by anti-ApoE antibody in a dose-dependent manner. In contrast, anti-ApoE antibody did not affect infection with HCVpp and VSVpp (Fig. 4A).

The CD81 dependence of infection was also compared between HCVtcp and HCVpp (Fig. 4B). Anti-CD81 antibody inhibited the entry of HCVtcp-1b, HCVtcp-2a, and HCVpp in a dose-dependent manner. The antibody had no effect on VSVpp infection. HCVtcp infection appears to be more sensitive to anti-CD81 antibody when compared with HCVpp infection; more than 60% inhibition was observed at 0.08 μg/mL anti-CD81 antibody for HCVtcp-1b and HCVtcp-2a, whereas approximately 50% inhibition was observed for HCVpp at 2 μg/mL antibody. Neutralization of HCVcc by anti-ApoE and anti-CD81 antibodies was also determined. Antibodies blocked HCVcc infection (Fig. 4C and D), as observed with HCVtcp. These results suggest that ApoE, as well as CD81, play an important role in HCVtcp infection. Thus, HCVtcp may be more useful for evaluating the HCV entry process than HCVpp.

Identification of novel culture-adaptive mutation in NS3 by serial passage of HCVtcp in packaging cells

The HCVtcp system was further applied to analyses of genetic changes during serial passages in target cells. As an initial attempt, supernatants of cells co-transfected with pCAGC-NS2/JFH1am and pHH/SGR were inoculated into Huh7.5.1 cells transiently transfected with pCAGC-NS2/JFH1am. However, infectious titer was lost after repeated inoculation, likely due to low HCVtcp titers and low efficiency of plasmid transduction (data not shown). To overcome this, we utilized recombinant adenovirus vectors (rAdVs) to provide core-NS2. As we were not able to obtain rAdV directly expressing core-NS2, conditional transgene expression based on a Cre-loxP strategy was employed (Kanegae et al., 1995). We constructed an rAdV containing core-NS2 gene downstream of a stuffer DNA flanked by a pair of loxP sites (AxCALNLH-CNS2). When cells were doubly infected with AxCALNLH-CNS2 and the Cre-expressing rAdV, AxCANCre (Kanegae et al., 1995), the Cre-mediated excisional deletion removed the stuffer DNA, resulting in core-NS2 expression under control of the CAG promoter (Fig. 5A). As expected, tightly regulated production of HCVtcp was observed. The cells infected with AxCANCre and AxCALNLH-CNS2 along with transduction of pHH/SGR-Luc produced HCVtcp at high levels. Production of HCVtcp was undetectable when either AxCANCre or AxCALNLH-CNS2 was not infected (Fig. 5B). The Cre-mediated rAdV expression system appears to have yielded considerably higher production of HCVtcp when compared with the settings for plasmid co-transfection.

Supernatants from cells in which core-NS2 was expressed using rAdVs and the subgenomic RNA derived from pHH/SGR replicated were inoculated into cells infected with AxCALNLH-CNS2 and AxCANCre (Fig. 6A). Blind passage was performed by sequentially transferring culture supernatants to cells infected with the above rAdVs. The two independent 10 blind passages (p10) showed virus titers of $>1\times10^6$ IU/mL, which were markedly higher than those of the passage 0 (p0) stock cultures (4×10^4 IU/mL). Side-by-side infection analysis revealed that the HCVtcp p10 #1 achieved a virus titer approximately 36 times higher than that of HCVtcp p0 on the packaging cells at 6 day post-infection (Fig. 6B). Sequencing of the entire replicon in the supernatants at p10 in two independent experiments revealed

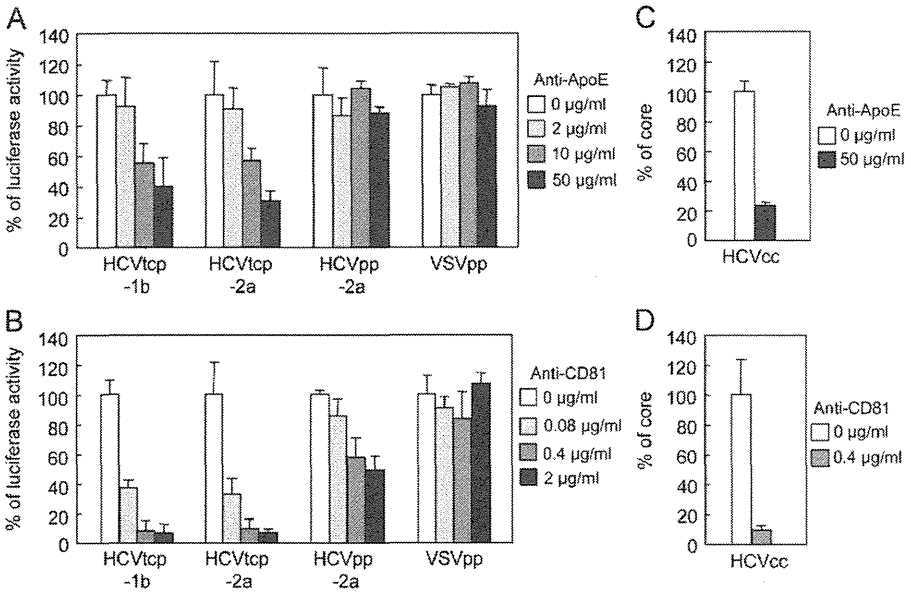

Fig. 4. Effects of anti-ApoE and anti-CD81 antibodies on HCV entry. (A) Aliquots of virus sample were incubated with increasing concentrations of anti-ApoE antibodies for 1 h and were then added to Huh7.5.1 cells. Luciferase activity was determined at 72 h post-infection and is expressed relative to activity without antibodies (white bar). (B) Huh7.5.1 cells were preincubated for 1 h with increasing concentrations of anti-CD81 antibodies, followed by inoculating virus samples. Luciferase activity was determined and expressed as shown in (A). (C) Aliquots of HCVcc were incubated with anti-ApoE antibodies for 1 h and were then added to Huh7.5.1 cells at an MOI of 0.05. Intracellular core levels were quantitated at 24 h post-infection and are expressed relative to levels without antibodies (white bar). (D) Huh7.5.1 cells were preincubated for 1 h with anti-CD81 antibodies. HCVcc infection and measurement of core proteins were performed as indicated in (C).

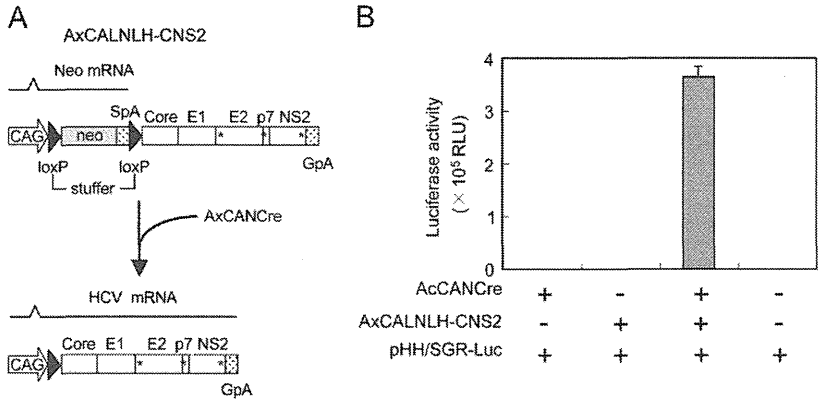

Fig. 5. Transgene activation mediated by rAdVs expressing Cre recombinase under control of CAG promoter. (A) Cre recobminase expressed by AxCANCre recognizes a pair of its target sequences loxP in AxCALNLH-CNS2, and removes the stuffer region resulting in expression of HCV core-NS2 polyprotein by CAG promoter. CAG: CAG promoter; SpA: SV40 early polyA signal; GpA: rabbit b-globin poly(A) signal. (B) Luciferase activity in Huh7.5.1 cells inoculated with 4-day post-transfection culture supernatant from cells transfected with pHH/SGR-Luc, and then infected with indicated rAdVs.

that both passaged HCVtcp had an identical nonsynonymous mutation in the NS3 region (N1586D) (Fig. 6C).

In order to examine the role of NS3 mutation identified on HCV RNA replication and on HCVtcp production, the N1586D mutation was introduced into pHH/SGR-Luc. Luciferase activities of the N1586D-mutated replicon were apparently lower than those of the WT-replicon, thus suggesting that the NS3 mutation reduced viral RNA replication (Fig. 7A). HCV RNA levels in the supernatants of cells transfected with WT- or mutant replicon plasmid along with pCAGC-NS2/JFH1am and luciferase activity in cells inoculated with supernatants from the transfected cells were then determined (Fig. 7B). The viral RNA level secreted from cells replicating the N1586D-mutated replicon was lower than that from cells replicating WT replicon (Fig. 7B, left). By contrast, a significantly higher infectivity of HCVtcp produced from the mutant replicon-cells was observed, as compared to WT replicon-cells (Fig. 7B, right), suggesting that the adaptive mutation increased the specific infectivity (almost 9-fold) of the virus particles. To further determine whether the N1586D mutation affects infectious viral assembly and/or virus release, we used the CD81-negative Huh-7 subclone, Huh7-25 (Akazawa et al., 2007), which may produce infectious particles, but is not susceptible to HCV entry due to a lack of CD81 expression, therefore allowing us to examine viral assembly and release without the influence of reinfection by produced HCVtcp. Measurement of intracellular and extracellular HCVtcp indicated that Huh7-25 cells replicating the N1586D-mutated replicon produced more infectious virus than WT in both supernatants and cell lysates (Fig. 7C). Thus, it can be concluded that the N1586D mutation contributes to enhanced infectious viral assembly, not RNA replication. We could not exclude the possibility that N1586D mutation affects virus release, since the mutation enhanced extracellular virus titers more than did the intracellular titer.

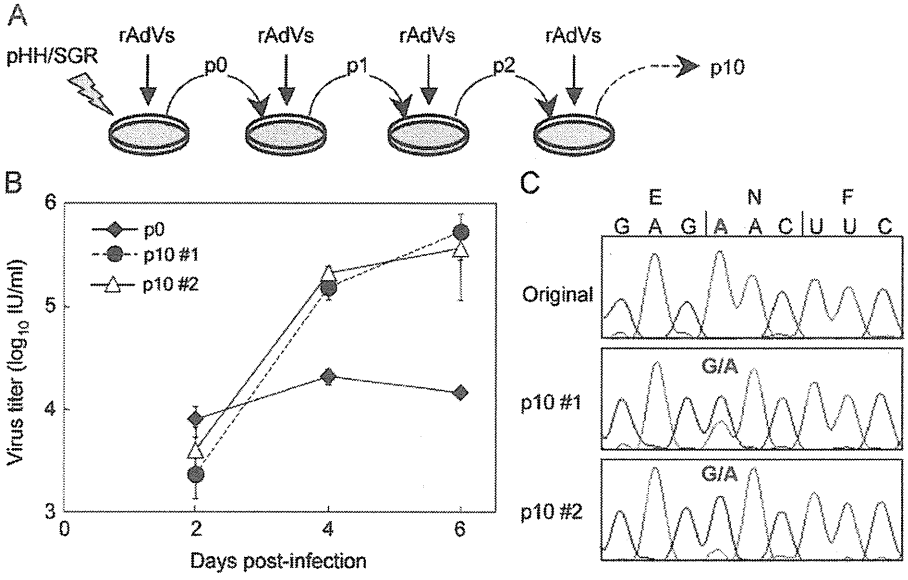

Fig. 6. Genotypic changes in HCVtcp following blind passage. (A) Experimental procedure for blind passage of HCVtcp. Huh7.5.1 cells were transfected with pHH/SGR and were doubly infected with AxCANCre and AxCALNLH-CNS2. Culture fluids were collected and were inoculated into cells infected with AxCANCre and AxCALNLH-CNS2. These procedures were repeated 10 times with two independent samples (#1 and #2). (B) Growth curves of HCVtcp p0 and p10 on Huh7.5.1 cells expressing core-NS2. Cells were infected with HCVtcp at an MOI of 0.05, and medium was collected at the indicated time points and subjected to titration. (C) Nucleotide sequences of original and blind-passaged replicons from HCVtcp. Nucleotides of mutated position are shown in red and bold.

The impact of the N1586D mutation on production of intra- and intergenotypic HCVtcp chimeras was also investigated. The N1586D mutation in the replicon enhanced the production of chimeric HCVtcp by providing core-p7 from all strains examined, although not statistically significant in THpa, and Con1 strains (Fig. 7D). Finally, to determine whether the N1586D mutation was responsible for enhancing HCVcc production, this mutation was introduced into pHHJFH1, which carries the full-length wild-type JFH-1 cDNA (Masaki et al., 2010), yielding pHHJFH1N1586D. The virus titer obtained from cells transfected with the pHHJFH1N1586D was significantly higher than that of WT (Fig. 7E), thus demonstrating that the N1586D mutation enhances yields of HCVcc, in addition to HCVtcp.

Discussion

Single-round infectious viral particles generated by *trans*-packaging systems are considered to be valuable tools for studying virus life cycles, particularly the steps related to entry into target cells, assembly and release of infectious particles. However, limited HCV strains have been applied for the efficient production of HCVtcp to date. In this study, we improved the HCVtcp system in order to enhance the productivity of infectious particles. Production of chimeric HCVtcp by providing genotype 1b-derived core-p7, in addition to intragenotypic viral proteins, was also confirmed. Furthermore, we exploited the system to investigate genetic changes during serial passage of target cells and identified a novel cell culture-adaptive mutation in NS3, which also contributes to enhance the productivity of HCVtcp.

HCVpp (Bartosch et al., 2003a; Hsu et al., 2003) has proven to be a valuable surrogate system by which the study of viral and cellular determinants of the viral entry pathway is possible. Early steps of HCV infection, including the role of HCV glycoprotein heterodimers, receptor binding, internalization and pH-dependent endosomal fusion, have been at least in part mimicked by HCVpp (Lavie et al., 2007). However, as HCVpp is generated in non-hepatic cells such as the human embryo kidney cells 293T, it is likely that the cell-derived component(s) of HCVpp differ from those of HCVcc. Hepatocytes play a role in maintaining lipid homeostasis in the body by assembling and secreting lipoproteins, including VLDL. It is highly likely that HCV exploits lipid synthesis pathways, as there is a tight link between virion formation and VLDL synthesis. Down-regulation of ApoE considerably reduces HCV production (Benga et al., 2010; Chang et al., 2007; Hishiki et al., 2010; Jiang and Luo, 2009; Owen et al., 2009). Infectivity of HCVcc is also neutralized by anti-ApoE antibodies (Chang et al., 2007). These data suggest that ApoE is important for HCV infectivity. Furthermore, Niemann-Pick C1-like 1 (NPC1L1), involving cholesterol uptake receptor, was recently identified as a host factor for HCV entry (Sainz et al., 2012). Knockdown of NPC1L1 had no effect on the entry of HCVpp whereas HCVcc entry was impaired, possibly due to different cholesterol content of these particles. Here, we found that the anti-ApoE antibody neutralized infection by HCVtcp and HCVcc, but not by HCVpp (Fig. 4A and C), thus suggesting that biogenesis and/or secretion pathways of VLDL are involved in HCVtcp similarly to HCVcc, but not in HCVpp.

We also observed that infectivity of HCVtcp and HCVcc is more efficiently neutralized by the anti-CD81 antibody, as compared to that of HCVpp (Fig. 4B and D). It has recently been reported that E2 of HCVcc contained both high-mannose-type and complex-type glycans, whereas most of the glycans on HCVpp-associated E2 were complex-type, which is matured by Golgi enzymes (Vieyres et al., 2010). Mutational analysis of the N-linked glycosylation sites in E1/E2 demonstrated that several glycans on E2 may affect the sensitivity of HCVpp against antibody neutralization, as well as access of CD81 to its binding site on E2 (Helle et al., 2010). The differences in sensitivity between HCVtcp and HCVpp to neutralization by anti-CD81 antibody observed here may be due to differences in carbohydrate composition of HCV glycoproteins during expression and processing of E1/E2 in cells and morphogenesis of HCVtcp and HCVpp.

By analyzing the various replicons for *trans*-packaging, we observed the highest production of HCVtcp with replicons from pHH/SGR, which lacked sequences not essential for RNA

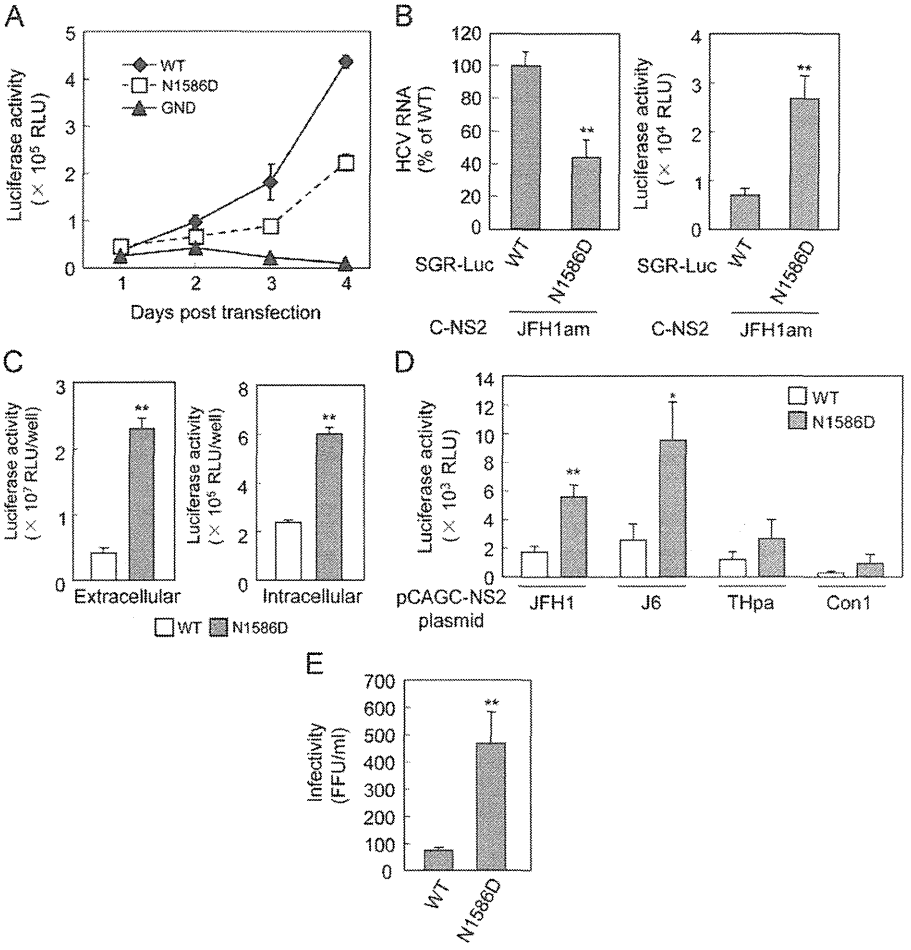

Fig. 7. Effects of N1586D mutation on RNA replication and production of HCVtcp or HCVcc. (A) RNA replication of replicons in cells transfected with pHH/SGR-Luc (WT) or N1586D mutant. Luciferase activities at 1 to 4 day post-transfection were determined. (B) Relative levels of HCV RNA in the supernatants from cells transfected with pHH/SGR-Luc (WT) or N1586D mutant plasmid along with pCAGC-NS2/JFH1am were shown in the left panel. Luciferase activities in cells inoculated with supernatants from cells transfected with indicated plasmids at 4 day post-transfection were shown in the right panel. (C) Luciferase activity in cells inoculated with supernatant and cell lysates from Huh7-25 cells transfected with pHH/SGR-Luc (WT) or N1586D mutant plasmid along with pCAGC-NS2/JFH1am at 5 day post-transfection. (D) Luciferase activity in cells inoculated with culture supernatant from cells transfected with pHH/SGR-Luc (WT) or N1586D mutant plasmid along with indicated core-NS2 plasmids at 4 day post-transfection. (E) Infectivity of supernatant from cells transfected with pHH/JFH1 (WT) or its derivative plasmid containing N1586D mutation at 6 day post-transfection. Statistical differences between WT and N1586D were evaluated using Student's *t*-test. $^{*}p < 0.05$, $^{**}p < 0.005$ vs. WT.

replication, while less efficient productivity was observed from pHH/SGR-Luc, pHH/SGR-C177, pHH/SGR-C191 and pHH/SGR-C-p7/am (Fig. 2C). Differences in the replication efficiency of the replicon do not appear to be a major determinant for HCVtcp productivity, at least in the present settings, as all replicon constructs except pHH/SGR-Luc replicated at similar levels, as confirmed by Western blotting (Fig. 2B). Although the shorter viral genome sequence may offer advantages over the longer sequence, further investigation is required in order to understand the molecular mechanisms underlying viral genome packaging. By comparing pHH/SGR vs. pHH/SGR-C177, pHH/SGR-C191 and pHH/SGR-C-p7/am, it is likely that the expression of the structural protein in *cis* does not increase HCVtcp production when sufficient amounts of structural proteins are supplied in *trans*.

Blind passage of HCVtcp in packaging cells infected with rAdVs providing core-NS2 enabled us to identify a novel culture-adaptive mutation in NS3. The N-terminal third of NS3 forms a serine protease, together with NS4A, and its C-terminal two-thirds exhibits RNA helicase and RNA-stimulated NTPase activities. In addition, similarly to flaviviruses (Kummerer and Rice, 2002; Liu et al., 2002), it is now apparent that HCV NS3 is also involved in viral morphogenesis (Han et al., 2009; Ma et al., 2008), although its precise role and underlying molecular mechanism(s) have not fully been elucidated. Two cell-culture adaptive NS3 mutations which are involved in HCV assembly have been identified. The Q1251L mutation in helicase subdomain 1 resulted in approximately 30-fold higher production of HCV without affecting NS3 enzymatic activities (Ma et al., 2008). The M1290K adaptive mutation was also located in subdomain 1 of the NS3 helicase (Han et al., 2009). The N1586D mutation identified here was located in subdomain 3 of helicase. Analogous to Q1251L and M1290K, the N1586D mutation enhanced the infectious viral assembly by increasing specific infectivity without affecting the efficiency of viral RNA replication. Considering the possibility that NS3 plays a role in linking between the viral replicase and assembly sites (Jones et al., 2011), it is likely that NS3 helicase is one of the determinants for interaction with the structural proteins. Our results, together with earlier studies, suggest that chimeric and defective mutations as well as supplying the viral components in *trans*, function as selective pressures in virion assembly.

In summary, we have established a plasmid-based reverse genetics for efficient production of HCVtcp with structural

proteins from various strains. Single-round infectious HCVtcp can complement the HCVcc and HCVpp systems as a valuable tool for the study of HCV life cycles.

Materials and methods

Cells

Huh7 derivative cell line Huh7.5.1 and Huh7-25 were maintained in Dulbecco modified Eagle medium (DMEM) supplemented with nonessential amino acids, 100 U of penicillin/mL, 100 μg of streptomycin/mL, and 10% fetal bovine serum at 37 °C in a 5% CO_2 incubator.

Plasmids

Plasmids pHHJFH1, pHH/SGR-Luc, pHH/SGR-Luc/GND and pCAG/C-NS2 were as described previously (Masaki et al., 2010). In this study, plasmid pCAG/C-NS2 was designated as pCAGC-NS2/JFH. The plasmid pCAGC-NS2/JFHam having adaptive mutations in E2 (N417S), p7 (N765D), and NS2 (Q1012R) in pCAGC-NS2/JFH was constructed by oligonucleotide-directed mutagenesis. These mutations were also introduced in pHHJFH1, resulting in pHHJFH1am. To generate core-NS2 expression plasmids with different strains of HCV, the cDNA coding core to the first transmembrane region of NS2 (33 amino acids) in pCAGC-NS2/JFH was replaced with the corresponding sequence of the J6 (Lindenbach et al., 2005), H77c (Yanagi et al., 1997), THpa (Shirakura et al., personal communication) and Con1 (Koch and Bartenschlager, 1999) strains. The THpa sequence contained the P to A mutation at 328 aa at E1 in the original TH strain. To generate pHH/SGR, pHH/SGR-Luc was digested with MluI and PmeI, followed by Klenow enzyme treatment and self-ligation to delete the luciferase coding sequence. To generate pHH/SGR-C177, pHH/SGR-C191 and pHH/SGR-C-p7/am, cDNA coding the partial core and luciferase in pHH/SGR-Luc were replaced with coding sequences for mature core (177aa), full-length core (191aa) or core-p7 polyprotein containing adaptive mutations in E2 and p7, respectively. The selected NS3 mutation (N1586D) was introduced into pHH/SGR-Luc and pHHJFH1 by oligonucleotide-directed mutagenesis.

Generation of viruses

HCVcc and HCVtcp were generated as described previously (Masaki et al., 2010). For the production of HCVpp-2a, plasmid pcDNAdeltaC-E1-E2(JFH1)am having adaptive mutations in E2 (N417S) in pcDNAdeltaC-E1-E2(JFH1) (Akazawa et al., 2007) was constructed by oligonucleotide-directed mutagenesis. Murine leukemia virus pseudotypes with VSV G glycoprotein expressing luciferase reporter (VSVpp) were generated in accordance with previously described methods (Akazawa et al., 2007; Bartosch et al., 2003a).

Luciferase assay

Huh7.5.1 cells were seeded onto a 24-well plate at a density of 3×10^4 cells/well 24 h prior to inoculation with reporter viruses. Cells were incubated for 72 h, followed by lysis with 100 μL of lysis buffer. Luciferase activity of the cells was determined using a luciferase assay system (Promega, Madison, WI). All luciferase assays were performed in triplicate.

Quantification of HCV infectivity and HCV RNA

To determine the titers of HCVtcp and HCVcc, Huh7.5.1 cell monolayers prepared in multi-well plates were incubated with dilutions of samples and then replaced with media containing 10% FBS and 0.8% carboxymethyl cellulose. Following incubation for 72 h, monolayers were fixed and immunostained with rabbit polyclonal anti-NS5A antibody, followed by Alexa Fluor 488-conjugated anti-rabbit secondary antibody (Invitrogen), and stained foci or individual cells were counted and used to calculate a titer of focus-forming units (FFU)/mL for spreading infections or infectious units (IU)/mL for non-spreading infections. For intracellular infectivity, the cell pellet was resuspended in culture media, and cells were lysed by four freeze–thaw cycles. Cell debris was pelleted by centrifugation for 5 min at 4000 rpm. Supernatant was collected and used for titration. To determine the amount of HCV RNA in culture supernatants, RNA was extracted from 140 μL of culture medium by QIAamp Viral RNA Mini Kit (QIAGEN, Valencia, CA) and treated with DNase (TURBO DNase; Ambion, Austin, TX) at 37 °C for 1 h. Extracted RNA was further purified by using an RNeasy Mini Kit, which includes RNase-free DNase digestion (QIAGEN). Copy numbers of HCV RNA were determined by real-time quantitative reverse transcription-PCR as described previously (Wakita et al., 2005).

Antibodies

Mouse monoclonal antibodies against actin (AC-15) and CD81 (JS-81) were obtained from Sigma (St. Louis, MO) and BD Biosciences (Franklin Lakes, NJ), respectively. Goat polyclonal antibody to ApoE (LV1479433) was obtained from Millipore (Tokyo, Japan). Anti-NS5A and anti-NS5B antibodies were rabbit polyclonal antibody against synthetic peptides.

Neutralization assay

For neutralization experiments with anti-CD81 antibody, Huh7.5.1 cells were incubated with dilutions of anti-CD81 antibody for 1 h at 37 °C. Cells were then infected with viruses for 5 h at 37 °C. For neutralization experiments with anti-ApoE antibody, viruses were incubated with various concentrations of anti-ApoE antibody at room temperature for 1 h and cells were infected with viruses for 5 h at 37 °C. Following infection, supernatant was removed and cells were incubated with culture medium, and luciferase activity was determined at 3 day post-infection for HCVtcp and pseudotyped viruses. For neutralization experiments with HCVcc generated with pHHJFH1am, a multiplicity of infection (MOI) of 0.05 was used for inoculation, and intracellular core protein levels were monitored by ELISA (Ortho Clinical Diagnostics) at 24 h post-infection.

Immunoblotting

Transfected cells were washed with PBS and incubated with lysis buffer (50 mM Tris–HCl, pH 7.4, 300 mM NaCl, 1% triton X-100). Lysates were then sonicated for 5 min and were added to the same volume of SDS sample buffer. Protein samples were boiled for 10 min, separated by SDS-PAGE, and transferred to PVDF membrane. After blocking, membranes were probed with first antibodies, followed by incubation with peroxidase-conjugated secondary antibody. Antigen–antibody complexes were visualized using an enhanced chemiluminescence detection system (Super Signal West Pico Chemiluminescent Substrate; PIERCE, Rockford, IL), in accordance with the manufacturer's protocols.

Generation of recombinant adenoviruses

rAdV, AxCANCre, expressing Cre recombinase tagged with nuclear localization signal under CAG promoter was prepared as described previously (Baba et al., 2005). The target rAdV AxCALNLH-CNS2 expressing HCV core-NS2 polyprotein with adaptive mutations in E2, p7 and NS2 was generated as follows. Cosmid pAxCALNLwit2 is identical to pAxCALNLw (Sato et al., 1998), except that both the terminal sequences of the rAdV genome are derived from pAxCAwit2 (Fukuda et al., 2006). The core-NS2 fragment obtained from pCAGC-NS2/JFH1am by StuI-EcoRI digestion and subsequent Klenow treatment was inserted into the SwaI site of pAxCALNLwit2. The resultant cosmid pAxCALNLH-CN2it2 was digested with PacI and transfected into 293 cells to generate rAdV AxCALNLH-CNS2.

Preparation of packaging cells for HCVtcp

Huh7.5.1 cells were coinfected with AxCANCre at an MOI of 1 and AxCALNLH-CNS2 at an MOI of 3 for expression of JFH-1 core-NS2 polyprotein containing the adaptive mutations in E2, p7 and NS2.

RNA preparation, RT-PCR and sequencing

Total cellular RNA was extracted with TRIzol reagent (Invitrogen, Carlsbad, CA), and subjected to reverse transcription with random hexamer and Superscript III reverse transcriptase (Invitrogen). Three fragments of HCV cDNAs that cover the entire HCV subgenomic replicon genome, were amplified by nested PCR with TaKaRa Ex Taq polymerase (Takara, Shiga, Japan). Amplified products were separated by agarose gel electrophoresis, and were used for direct DNA sequencing.

Acknowledgments

We are grateful to Francis V. Chisari (The Scripps Research Institute) for providing Huh7.5.1 cells. We thank M. Sasaki, M. Matsuda, and T. Date for their technical assistance, and T. Mizoguchi for the secretarial work. We also thank T. Masaki for their helpful discussions. This work was supported in part by grants-in-aid from the Ministry of Health, Labor, and Welfare and the Ministry of Education, Culture, Sports, Science, and Technology, Japan.

References

Adair, R., Patel, A.H., Corless, L., Griffin, S., Rowlands, D.J., McCormick, C.J., 2009. Expression of hepatitis C virus (HCV) structural proteins in trans facilitates encapsidation and transmission of HCV subgenomic RNA. J. Gen. Virol. 90 (Part 4), 833–842.

Akazawa, D., Date, T., Morikawa, K., Murayama, A., Miyamoto, M., Kaga, M., Barth, H., Baumert, T.F., Dubuisson, J., Wakita, T., 2007. CD81 expression is important for the permissiveness of Huh7 cell clones for heterogeneous hepatitis C virus infection. J. Virol. 81 (10), 5036–5045.

Baba, Y., Nakano, M., Yamada, Y., Saito, I., Kanegae, Y., 2005. Practical range of effective dose for Cre recombinase-expressing recombinant adenovirus without cell toxicity in mammalian cells. Microbiol. Immunol. 49 (6), 559–570.

Bartosch, B., Dubuisson, J., Cosset, F.L., 2003a. Infectious hepatitis C virus pseudoparticles containing functional E1-E2 envelope protein complexes. J. Exp. Med. 197 (5), 633–642.

Bartosch, B., Vitelli, A., Granier, C., Goujon, C., Dubuisson, J., Pascale, S., Scarselli, E., Cortese, R., Nicosia, A., Cosset, F.L., 2003b. Cell entry of hepatitis C virus requires a set of co-receptors that include the CD81 tetraspanin and the SR-B1 scavenger receptor. J. Biol. Chem. 278 (43), 41624–41630.

Benedicto, I., Molina-Jimenez, F., Bartosch, B., Cosset, F.L., Lavillette, D., Prieto, J., Moreno-Otero, R., Valenzuela-Fernandez, A., Aldabe, R., Lopez-Cabrera, M., Majano, P.L., 2009. The tight junction-associated protein occludin is required for a postbinding step in hepatitis C virus entry and infection. J. Virol. 83 (16), 8012–8020.

Benga, W.J., Krieger, S.E., Dimitrova, M., Zeisel, M.B., Parnot, M., Lupberger, J., Hildt, E., Luo, G., McLauchlan, J., Baumert, T.F., Schuster, C., 2010. Apolipoprotein E interacts with hepatitis C virus nonstructural protein 5A and determines assembly of infectious particles. Hepatology 51 (1), 43–53.

Chang, K.S., Jiang, J., Cai, Z., Luo, G., 2007. Human apolipoprotein E is required for infectivity and production of hepatitis C virus in cell culture. J. Virol. 81 (24), 13783–13793.

Cormier, E.G., Tsamis, F., Kajumo, F., Durso, R.J., Gardner, J.P., Dragic, T., 2004. CD81 is an entry coreceptor for hepatitis C virus. Proc. Natl. Acad. Sci. USA 101 (19), 7270–7274.

Evans, M.J., von Hahn, T., Tscherne, D.M., Syder, A.J., Panis, M., Wolk, B., Hatziioannou, T., McKeating, J.A., Bieniasz, P.D., Rice, C.M., 2007. Claudin-1 is a hepatitis C virus co-receptor required for a late step in entry. Nature 446 (7137), 801–805.

Flint, M., von Hahn, T., Zhang, J., Farquhar, M., Jones, C.T., Balfe, P., Rice, C.M., McKeating, J.A., 2006. Diverse CD81 proteins support hepatitis C virus infection. J. Virol. 80 (22), 11331–11342.

Fukuda, H., Terashima, M., Koshikawa, M., Kanegae, Y., Saito, I., 2006. Possible mechanism of adenovirus generation from a cloned viral genome tagged with nucleotides at its ends. Microbiol. Immunol. 50 (8), 643–654.

Han, Q., Xu, C., Wu, C., Zhu, W., Yang, R., Chen, X., 2009. Compensatory mutations in NS3 and NS5A proteins enhance the virus production capability of hepatitis C reporter virus. Virus Res. 145 (1), 63–73.

Helle, F., Vieyres, G., Elkrief, L., Popescu, C.I., Wychowski, C., Descamps, V., Castelain, S., Roingeard, P., Duverlie, G., Dubuisson, J., 2010. Role of N-linked glycans in the functions of hepatitis C virus envelope proteins incorporated into infectious virions. J. Virol. 84 (22), 11905–11915.

Hishiki, T., Shimizu, Y., Tobita, R., Sugiyama, K., Ogawa, K., Funami, K., Ohsaki, Y., Fujimoto, T., Takaku, H., Wakita, T., Baumert, T.F., Miyanari, Y., Shimotohno, K., 2010. Infectivity of hepatitis C virus is influenced by association with apolipoprotein E isoforms. J. Virol. 84 (22), 12048–12057.

Hoofnagle, J.H., 2002. Course and outcome of hepatitis C. Hepatology 36 (5 Suppl. 1), S21–9.

Hsu, M., Zhang, J., Flint, M., Logvinoff, C., Cheng-Mayer, C., Rice, C.M., McKeating, J.A., 2003. Hepatitis C virus glycoproteins mediate pH-dependent cell entry of pseudotyped retroviral particles. Proc. Natl. Acad. Sci. USA 100 (12), 7271–7276.

Ishii, K., Murakami, K., Hmwe, S.S., Zhang, B., Li, J., Shirakura, M., Morikawa, K., Suzuki, R., Miyamura, T., Wakita, T., Suzuki, T., 2008. Trans-encapsidation of hepatitis C virus subgenomic replicon RNA with viral structure proteins. Biochem. Biophys. Res. Commun. 371 (3), 446–450.

Jiang, J., Luo, G., 2009. Apolipoprotein E but not B is required for the formation of infectious hepatitis C virus particles. J. Virol. 83 (24), 12680–12691.

Jones, D.M., Atoom, A.M., Zhang, X., Kottilil, S., Russell, R.S., 2011. A genetic interaction between the core and NS3 proteins of hepatitis C virus is essential for production of infectious virus. J. Virol. 85 (23), 12351–12361.

Kanegae, Y., Lee, G., Sato, Y., Tanaka, M., Nakai, M., Sakaki, T., Sugano, S., Saito, I., 1995. Efficient gene activation in mammalian cells by using recombinant adenovirus expressing site-specific Cre recombinase. Nucl. Acids Res. 23 (19), 3816–3821.

Koch, J.O., Bartenschlager, R., 1999. Modulation of hepatitis C virus NS5A hyperphosphorylation by nonstructural proteins NS3, NS4A, and NS4B. J. Virol. 73 (9), 7138–7146.

Kummerer, B.M., Rice, C.M., 2002. Mutations in the yellow fever virus nonstructural protein NS2A selectively block production of infectious particles. J. Virol. 76 (10), 4773–4784.

Lavie, M., Goffard, A., Dubuisson, J., 2007. Assembly of a functional HCV glycoprotein heterodimer. Curr. Issues Mol. Biol. 9 (2), 71–86.

Lindenbach, B.D., Evans, M.J., Syder, A.J., Wolk, B., Tellinghuisen, T.L., Liu, C.C., Maruyama, T., Hynes, R.O., Burton, D.R., McKeating, J.A., Rice, C.M., 2005. Complete replication of hepatitis C virus in cell culture. Science 309 (5734), 623–626.

Liu, S., Yang, W., Shen, L., Turner, J.R., Coyne, C.B., Wang, T., 2009. Tight junction proteins claudin-1 and occludin control hepatitis C virus entry and are downregulated during infection to prevent superinfection. J. Virol. 83 (4), 2011–2014.

Liu, W.J., Sedlak, P.L., Kondratieva, N., Khromykh, A.A., 2002. Complementation analysis of the flavivirus Kunjin NS3 and NS5 proteins defines the minimal regions essential for formation of a replication complex and shows a requirement of NS3 in cis for virus assembly. J. Virol. 76 (21), 10766–10775.

Ma, Y., Yates, J., Liang, Y., Lemon, S.M., Yi, M., 2008. NS3 helicase domains involved in infectious intracellular hepatitis C virus particle assembly. J. Virol. 82 (15), 7624–7639.

Masaki, T., Suzuki, R., Saeed, M., Mori, K., Matsuda, M., Aizaki, H., Ishii, K., Maki, N., Miyamura, T., Matsuura, Y., Wakita, T., Suzuki, T., 2010. Production of infectious hepatitis C virus by using RNA polymerase I-mediated transcription. J. Virol. 84 (11), 5824–5835.

Mazumdar, B., Banerjee, A., Meyer, K., Ray, R., 2011. Hepatitis C virus E1 envelope glycoprotein interacts with apolipoproteins in facilitating entry into hepatocytes. Hepatology 54 (4), 1149–1156.

McKeating, J.A., Zhang, L.Q., Logvinoff, C., Flint, M., Zhang, J., Yu, J., Butera, D., Ho, D.D., Dustin, L.B., Rice, C.M., Balfe, P., 2004. Diverse hepatitis C virus glycoproteins mediate viral infection in a CD81-dependent manner. J. Virol. 78 (16), 8496–8505.

Owen, D.M., Huang, H., Ye, J., Gale Jr., M., 2009. Apolipoprotein E on hepatitis C virion facilitates infection through interaction with low-density lipoprotein receptor. Virology 394 (1), 99–108.

Pietschmann, T., Kaul, A., Koutsoudakis, G., Shavinskaya, A., Kallis, S., Steinmann, E., Abid, K., Negro, F., Dreux, M., Cosset, F.L., Bartenschlager, R., 2006. Construction and characterization of infectious intragenotypic and intergenotypic hepatitis C virus chimeras. Proc. Natl. Acad. Sci. USA 103 (19), 7408–7413.

Pileri, P., Uematsu, Y., Campagnoli, S., Galli, G., Falugi, F., Petracca, R., Weiner, A.J., Houghton, M., Rosa, D., Grandi, G., Abrignani, S., 1998. Binding of hepatitis C virus to CD81. Science 282 (5390), 938–941.

Ploss, A., Evans, M.J., Gaysinskaya, V.A., Panis, M., You, H., de Jong, Y.P., Rice, C.M., 2009. Human occludin is a hepatitis C virus entry factor required for infection of mouse cells. Nature 457 (7231), 882–886.

Russell, R.S., Meunier, J.C., Takikawa, S., Faulk, K., Engle, R.E., Bukh, J., Purcell, R.H., Emerson, S.U., 2008. Advantages of a single-cycle production assay to study cell culture-adaptive mutations of hepatitis C virus. Proc. Natl. Acad. Sci. USA 105 (11), 4370–4375.

Sainz Jr., B., Barretto, N., Martin, D.N., Hiraga, N., Imamura, M., Hussain, S., Marsh, K.A., Yu, X., Chayama, K., Alrefai, W.A., Uprichard, S.L., 2012. Identification of the Niemann-Pick C1-like 1 cholesterol absorption receptor as a new hepatitis C virus entry factor. Nat. Med. 18 (2), 281–285.

Sato, Y., Tanaka, K., Lee, G., Kanegae, Y., Sakai, Y., Kaneko, S., Nakabayashi, H., Tamaoki, T., Saito, I., 1998. Enhanced and specific gene expression via tissue-specific production of Cre recombinase using adenovirus vector. Biochem. Biophys. Res. Commun. 244 (2), 455–462.

Scarselli, E., Ansuini, H., Cerino, R., Roccasecca, R.M., Acali, S., Filocamo, G., Traboni, C., Nicosia, A., Cortese, R., Vitelli, A., 2002. The human scavenger receptor class B type I is a novel candidate receptor for the hepatitis C virus. EMBO J. 21 (19), 5017–5025.

Steinmann, E., Brohm, C., Kallis, S., Bartenschlager, R., Pietschmann, T., 2008. Efficient trans-encapsidation of hepatitis C virus RNAs into infectious virus-like particles. J. Virol. 82 (14), 7034–7046.

Suzuki, T., Ishii, K., Aizaki, H., Wakita, T., 2007. Hepatitis C viral life cycle. Adv. Drug Deliv. Rev. 59 (12), 1200–1212.

Tani, H., Komoda, Y., Matsuo, E., Suzuki, K., Hamamoto, I., Yamashita, T., Moriishi, K., Fujiyama, K., Kanto, T., Hayashi, N., Owsianka, A., Patel, A.H., Whitt, M.A., Matsuura, Y., 2007. Replication-competent recombinant vesicular stomatitis virus encoding hepatitis C virus envelope proteins. J. Virol. 81 (16), 8601–8612.

Vieyres, G., Thomas, X., Descamps, V., Duverlie, G., Patel, A.H., Dubuisson, J., 2010. Characterization of the envelope glycoproteins associated with infectious hepatitis C virus. J. Virol. 84 (19), 10159–10168.

Wakita, T., Pietschmann, T., Kato, T., Date, T., Miyamoto, M., Zhao, Z., Murthy, K., Habermann, A., Krausslich, H.G., Mizokami, M., Bartenschlager, R., Liang, T.J., 2005. Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. Nat. Med. 11 (7), 791–796.

Yanagi, M., Purcell, R.H., Emerson, S.U., Bukh, J., 1997. Transcripts from a single full-length cDNA clone of hepatitis C virus are infectious when directly transfected into the liver of a chimpanzee. Proc. Natl. Acad. Sci. USA 94 (16), 8738–8743.

Zhong, J., Gastaminza, P., Cheng, G., Kapadia, S., Kato, T., Burton, D.R., Wieland, S.F., Uprichard, S.L., Wakita, T., Chisari, F.V., 2005. Robust hepatitis C virus infection in vitro. Proc. Natl. Acad. Sci. USA 102 (26), 9294–9299.

JCM
Journals.ASM.org

Japanese Reference Panel of Blood Specimens for Evaluation of Hepatitis C Virus RNA and Core Antigen Quantitative Assays

Asako Murayama,[a] Nao Sugiyama,[a] Koichi Watashi,[a] Takahiro Masaki,[a] Ryosuke Suzuki,[a] Hideki Aizaki,[a] Toshiaki Mizuochi,[b] Takaji Wakita,[a] and Takanobu Kato[a]

Department of Virology II, National Institute of Infectious Diseases, Tokyo, Japan,[a] and Department of Research on Blood and Biological Products, National Institute of Infectious Diseases, Tokyo, Japan[b]

An accurate and reliable quantitative assay for hepatitis C virus (HCV) is essential for measuring viral propagation and the efficacy of antiviral therapy. There is a growing need for domestic reference panels for evaluation of clinical assay kits because the performance of these kits may vary with region-specific genotypes or polymorphisms. In this study, we established a reference panel by selecting 80 donated blood specimens in Japan that tested positive for HCV. Using this panel, we quantified HCV viral loads using two HCV RNA kits and five core antigen (Ag) kits currently available in Japan. The data from the two HCV RNA assay kits showed excellent correlation. All RNA titers were distributed evenly across a range from 3 to 7 log IU/ml. Although the data from the five core Ag kits also correlated with RNA titers, the sensitivities of individual kits were not sufficient to quantify viral load in all samples. As calculated by the correlation with RNA titers, the theoretical lower limits of detection by these core Ag assays were higher than those for the detection of RNA. Moreover, in several samples in our panel, core Ag levels were underestimated compared to RNA titers. Sequence analysis in the HCV core region suggested that polymorphisms at amino acids 47 to 49 of the core Ag were responsible for this underestimation. The panel established in this study will be useful for estimating the quality of currently available and upcoming HCV assay kits; such quality control is essential for clinical usage of these kits.

Hepatitis C virus (HCV) is a major cause of chronic liver disease worldwide (15). There is no protective vaccine against this virus, and once an individual is infected, HCV often establishes persistent infection and leads to chronic hepatitis, cirrhosis, and hepatocellular carcinoma (9). The most widely used therapy for HCV infection is the combined administration of pegylated alpha interferon and ribavirin (29). However, this treatment is problematic, as it has limited efficacy, high cost, and severe adverse effects (8, 25). To estimate the outcome of antiviral therapy, and to understand the state of viral propagation, it is important to determine the HCV viral load in chronic hepatitis C patients by the use of accurate and reliable HCV quantitative assays (9, 14). For this purpose, several commercial assay kits for HCV RNA and core antigen (Ag) quantification are currently used in Japan. For quantification of HCV RNA levels, two real-time quantitative reverse transcription-PCR (qRT-PCR)-based assay kits are available, including the COBAS AmpliPrep/COBAS TaqMan HCV test (CAP/CTM-RNA; Roche Diagnostics, Tokyo, Japan) and the Abbott RealTime HCV test (ART-RNA; Abbott Japan, Tokyo, Japan). These assays are known to have high sensitivity and a wide dynamic range, but they require technical skill and attention to maintaining the specified conditions (4–6, 16, 24, 33–35). Alternatively, HCV viremia can be quantified by assessment of HCV core Ag level (1–3, 7, 10, 12, 13, 17–22, 27, 30–32). Five HCV core Ag assay kits are commercially available in Japan, including Architect HCV Ag (Architect-Ag; Abbott Japan), Lumipulse Ortho HCV Ag (Lumipulse-Ag; Fujirebio, Tokyo, Japan), Lumispot Eiken HCV Ag (Lumispot-Ag; Eiken Chemical, Tokyo, Japan), the Ortho HCV Ag ELISA test (ELISA-Ag; Ortho Clinical Diagnostics, Tokyo, Japan), and the Ortho HCV Ag IRMA test (IRMA-Ag; Ortho Clinical Diagnostics, Tokyo, Japan). These assays have some disadvantages compared to those measuring HCV RNA (notably, low sensitivity and narrow range of quantification) but also have some advantages (including ease of use, reduced risk of contamination, reduced cost, and reliability even with samples stored at room temperature for extended periods of time [1, 32]). Although core Ag levels are thought to be related closely to HCV RNA titers, the correlation and linearity of core Ag levels have not yet been fully evaluated. In addition, these quantitative parameters are known to be affected by nucleotide and amino acid sequences at the target regions of the assays (5, 6, 28, 34), and this sequence variation depends on genotypes or predominant strains in specific geographical regions.

In this study, we established a Japanese reference panel of samples for evaluation of HCV RNA and core Ag levels by collecting donated blood specimens that tested positive for HCV RNA and anti-HCV antibodies. Using this reference panel, we evaluated the HCV loads in these specimens with two HCV RNA assay kits and five core Ag assay kits and assessed correlations among the data generated by these kits.

MATERIALS AND METHODS

Preparation of reference panel. To establish a reference panel for HCV quantitative assays, a total of 80 donated plasma samples were selected. All of these specimens, supplied by the Japanese Red Cross Blood Centers, tested positive for the presence of HCV RNA and anti-HCV antibodies. These samples, collected in Japan from May to September of 2007, were obtained from Japanese blood donor volunteers in various regions of

Received 20 February 2012 Returned for modification 21 March 2012
Accepted 30 March 2012
Published ahead of print 11 April 2012
Address correspondence to Takanobu Kato, takato@nih.go.jp.
Supplemental material for this article may be found at http://jcm.asm.org/.

doi:10.1128/JCM.00487-12

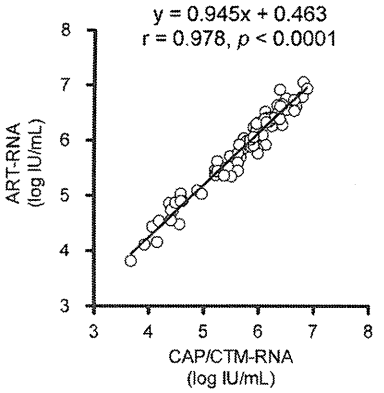

FIG 1 Correlation of HCV RNA titers as quantified by two commercial kits.

Japan. The samples were divided into 1-ml aliquots and stored at −80°C until use.

Quantification of HCV RNA and core Ag. The HCV RNA titer was measured with two real-time qRT-PCR kits, CAP/CTM-RNA (detection range, 1.5×10^1 to 6.9×10^7 IU/ml) and ART-RNA (detection range, 1.2×10^1 to 1.0×10^8 IU/ml). Additionally, samples were assessed using five HCV core Ag assay kits, including Architect-Ag (detection range, 3 to 20,000 fmol/liter), Lumipulse-Ag (detection range, 50 to 50,000 fmol/liter), Lumispot-Ag (detection range, 20 to 400,000 fmol/liter), ELISA-Ag (detection range, 44.4 to 3,600 fmol/liter), and IRMA-Ag (detection range, 20 to 20,000 fmol/liter). All assays were performed by the respective manufacturers at their research laboratories.

Sequencing and genotyping of HCV in reference panel samples. Viral RNA was extracted with the QIAamp viral RNA kit (Qiagen, Valencia, CA) from 140 μl of each plasma sample. HCV RNA was amplified by RT-PCR with primers corresponding to the 5′ untranslated region (UTR) (43S-IH, 5′-CCTGTGAGGAACTACTGTCTTC-3′; c/s17-ssp, 5′-CCGG GAGAGCCATAGTGGTCTGCG-3′) and the E1 region (1323R-IH, 5′-G GCGACCAGTTCATCATCAT-3′); the amplified products were sequenced directly. HCV genotypes of the isolated strains were assigned by phylogenetic analysis using an alignment with a representative strain of each genotype.

Statistical analysis. The correlations of obtained quantitative data were assessed by Pearson's correlation coefficient analysis, and values for r and P were calculated. A P value of <0.05 was considered to indicate statistical significance. Analysis was performed using Prism 5 software (GraphPad Software, Inc., La Jolla, CA).

Nucleotide sequence accession numbers. The accession numbers of C-01 to C-80 are AB705312 to AB705391, respectively.

RESULTS

Quantification of HCV RNA levels. The reference panel established in this work was used to measure HCV RNA levels with the CAP/CTM-RNA and ART-RNA kits. The correlation of the data obtained with the two kits is shown in Fig. 1. The RNA titers of these samples were distributed evenly, and all values were within the dynamic ranges of both assays. The HCV titers ranged from 3.68 to 6.88 and 3.82 to 7.08 log IU/ml in CAP/CTM-RNA and ART-RNA, respectively, and the correlation was significant ($r = 0.978$; $P < 0.0001$).

Quantification of HCV core Ag levels. HCV core Ag levels were measured using Architect-Ag, Lumipulse-Ag, Lumispot-Ag, ELISA-Ag, and IRMA-Ag kits. Among the 80 specimens in the reference panel, core Ag levels could be measured in all samples using Architect-Ag and ELISA-Ag kits, whereas core Ag levels were below the detection limit in 4, 2, and 1 samples using Lumipulse-Ag, Lumispot-Ag, and IRMA-Ag kits, respectively (Fig. 2; also, see Fig. S1 in the supplemental material). Significant correlations were observed between assays of HCV core Ag and HCV RNA ($r = 0.9065$ to 0.9666 and $P < 0.0001$ compared with CAP/CTM-RNA data [Fig. 2]); $r = 0.8877$ to 0.9552 and $P < 0.0001$ compared with ART-RNA data [see Fig. S1 in the supplemental material]). The theoretical lower limits of detection of these assays were calculated by use of these correlation formulas and were 3.2 and 3.4 log IU/ml for Architect-Ag, 4.2 and 4.2 log IU/ml for Lumipulse-Ag, 3.7 and 3.9 log IU/ml for Lumispot-Ag, 3.6 and 3.8 log IU/ml for ELISA-Ag, and 3.6 and 3.8 log IU/ml for IRMA-Ag (compared to CAP/CTM-RNA and ART-RNA, respectively). These calculated detection limits were substantially higher than those for the RNA quantitative assays (1.18 and 1.08 log IU/ml for CAP/CTM-RNA and ART-RNA, respectively).

In addition, we found that several samples showed considerable deviation from the linear regression (Fig. 2; also, see Fig. S1 in the supplemental material). To identify the deviating samples, we used Bland-Altman plot analysis (Fig. 3; also, see Fig. S2 in the supplemental material). This plot shows the difference between the titer values of HCV RNA and core Ag as a function of the average of these two values. Several samples demonstrated discordance between the measured HCV RNA and core Ag levels. Among these samples, we focused on samples with discordant results in multiple core Ag assays compared to both RNA quantitative assays. For sample C-01, core Ag levels were underestimated when measured with Architect-Ag, Lumipulse-Ag, and Lumispot-Ag in comparison with CAP/CTM-RNA (Fig. 3) and when measured with Architect-Ag, Lumipulse-Ag, Lumispot-Ag, and IRMA-Ag in comparison with ART-RNA (see Fig. S2 in the supplemental material). Likewise, for sample C-73, core Ag levels were underestimated when measured with Architect-Ag, Lumipulse-Ag, and IRMA-Ag in comparison with CAP/CTM-RNA (Fig. 3) and when measured with Architect-Ag and Lumipulse-Ag in comparison with ART-RNA (see Fig. S2 in the supplemental material). Thus, sample-specific underestimation was observed in several HCV core Ag kits.

Nucleotide sequences in core region of reference panel samples. To clarify the sources of these underestimates of HCV core Ag levels, HCV RNA was extracted from each of the samples in the reference panel, and the nucleotide sequences of core regions were determined. Phylogenetic analysis with these sequences permitted classification of the individual strains by genotype. Of 80 samples in the reference panel, 1 (1.3%) was genotype 1a, 35 (43.8%) were genotype 1b, 26 (32.5%) were genotype 2a, and 18 (22.5%) were genotype 2b (Table 1; also, see Fig. S3 in the supplemental material). These strains were distributed evenly among reference strains of each genotype and cover the sequence diversity of strains isolated in Japan (see Fig. S3 in the supplemental material). The genotypes of samples associated with underestimated core Ag values (samples C-01 and C-73) were both classified as genotype 2a.

Predicted amino acid sequences of HCV core protein were aligned with the consensus core protein sequence for the genotype 1b strains obtained in this study (see Fig. S4 in the supplemental material). Excluding the genotype-specific sequence variations, a specific amino acid polymorphism was identified at amino acid (aa) residue 48 (Ala to Thr) in samples C-01 and C-73. Sample C-01, which yielded underestimated values in most core Ag assays, also possessed an additional polymorphism in the same region,

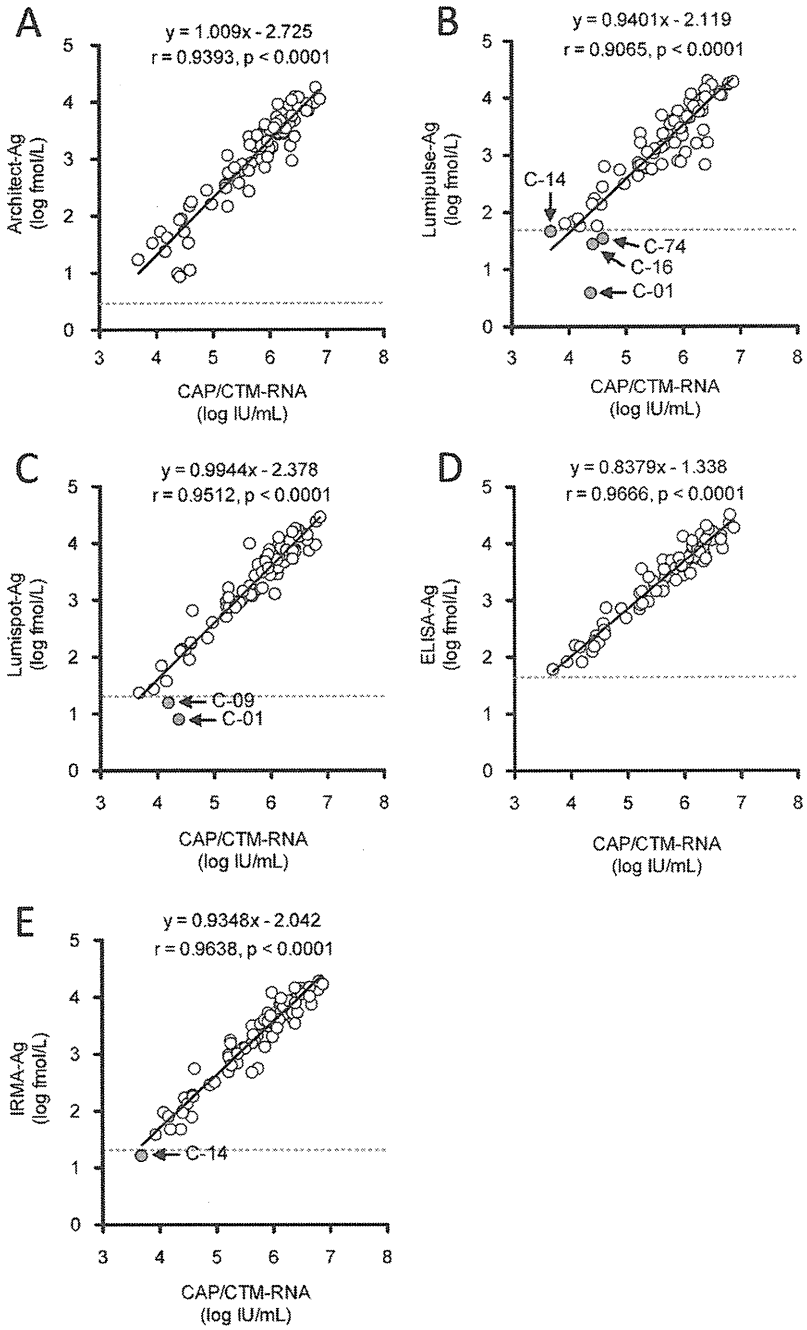

FIG 2 Correlation between CAP/CTM-RNA and core Ag levels as quantified by five commercial kits. Data for core Ag levels were converted to log fmol/liter prior to analysis. In each plot, the lower limit of detection of the respective core Ag assay is indicated by a dotted line. Data for samples below the lower detection limit of each assay are indicated by shaded circles labeled with the respective sample designations.

specifically an Arg-to-Gly substitution at aa 47. We suspected that these polymorphisms altered the antigenicity of the core protein, thereby reducing detected core Ag levels and leading to underestimation of values by the core Ag quantification kits. To assess the correlation of these polymorphisms with the underestimation of core Ag values, strains containing polymorphisms in this region (at aa 47 to 49 [Fig. 4]) were identified in Bland-Altman plots of HCV RNA and core Ag (Fig. 3; also, see Fig. S2 in the supplemental material). A total of 12 strains exhibited polymorphisms at these positions, including 2 strains of genotype 1b, 8 of genotype 2a, and 2 of genotype 2b (Table 1). In the Bland-Altman plot of CAP/CTM-RNA and Architect-Ag, 4 of 12 values (for samples C-01, C-16, C-73, and C-74) were located under the line of the lower 95% limit of agreement (Fig. 3A). Likewise, in the plot of CAP/CTM-RNA and Lumipulse-Ag, 3 of 12 values (those for samples C-01, C-67, and C-73) were located under the line of the lower