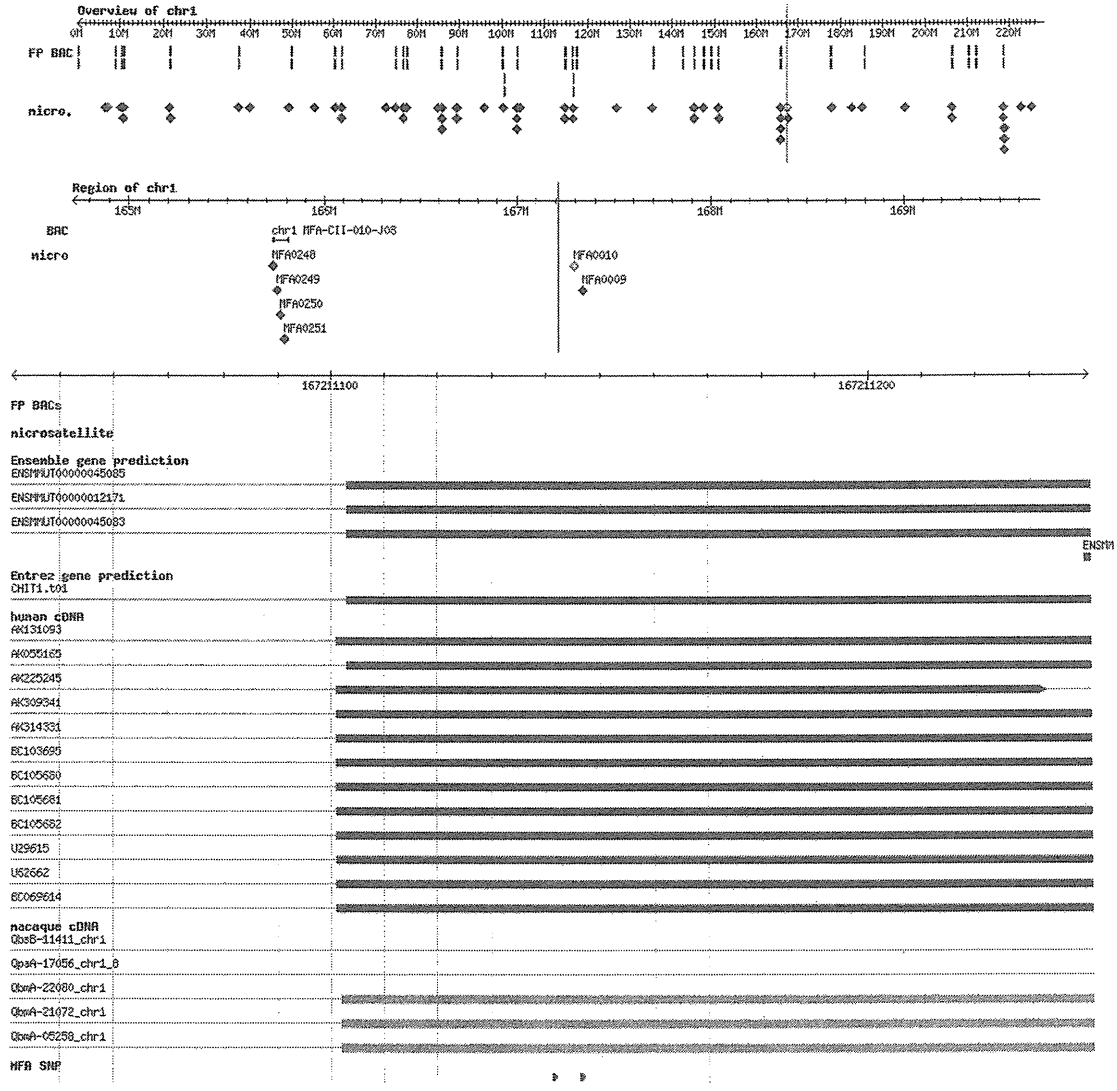
(a)



(b)

***Macaca fascicularis* MFA SNP**

chr1 167211141 [167211141 .. 167211141]

Position	Reference allele	Alternative allele	Reference count	Alternative count	Zygosity	Transcript	AA change	Rhesus QV
167211141	T	C	0	30	Homozygous	CHIT1.t01:447->;	LT;	60

Figure 5 Screenshots of the macaque genome database. (a) cDNA clones, BAC clones, microsatellite markers, gene predictions, and SNVs on chromosome 1 are shown in the genome browser. **(b)** Detailed information on each SNV is linked from the browser.

size is possibly associated with admixture with the rhesus macaque, since their habitats were largely connected by the formation of Sundaland. However, it should be noted that the time estimation largely depends on the generation time parameter of macaques. If we adopt a longer generation time parameter - for example, 10 to 12 years

as the median age of females giving offspring - the most recent bottleneck event would shift earlier, 33,000 to 40,000 years ago.

Conclusions

We identified 9.7 million high-quality SNVs between the Malaysian cynomolgus and the reference (Indian rhesus) macaque genomes. The list of whole-genome SNVs will be useful for many future applications, such as an array-based genotyping system of macaque individuals. In contrast to humans, the genetic variation of experimental animals, especially of monkeys, is largely unexplored. The whole-genome sequence of a Malaysian cynomolgus macaque has unveiled hidden genetic variations among these widely used experimental animals and will benefit future evolutionary and biomedical studies.

Materials and methods

Animal and blood sampling

Whole blood cells for genomic DNA were obtained from a 25-year-old male cynomolgus macaque (Malaysian), housed at the Tsukuba Primate Research Center (TPRC), National Institute of Biomedical Innovation (NIBIO), Tsukuba, Ibaraki, Japan, in accordance with the TPRC guidelines. The sampled macaque was an F1 progeny of unrelated wild individuals captured in the south of Kuala Lumpur. These macaques were cared for and handled according to the guidelines established by the Institutional Animal Care and Use Committee of NIBIO and the standard operating procedures for macaques at the TPRC. Blood collection was conducted at the TPRC in accordance with the guidelines of the Laboratory Biosafety Manual, World Health Organization. Genomic DNA was isolated from 10 ml of peripheral blood with EDTA using a Qiagen Genomic DNA purification kit (Qiagen K. K., Tokyo, Japan). The isolated DNA samples were kept at -80°C until use.

Genome sequencing

Genome sequencing was performed using the SOLiD 3 Plus System (Life Technologies, Gaithersburg, MD, USA). Fragment (50 bp) and mate-pair (25 bp × 2) libraries were generated using the macaque genomic DNA. Mate-pair libraries of 600 to 800 bp and 800 to 1,000 bp insert sizes were prepared, and each library was run in two slides. Library preparations and all SOLiD runs were performed as per the standard manufacturer's protocols.

Mapping sequence data on the Indian rhesus macaque genome

SOLiD sequence data were mapped on the rhesus macaque draft genome sequence (GenBank accession numbers NC_007858 to NC_007878). The assembly QV of the genome was retrieved from the UCSC website [37]. The reads were mapped using the BioScope (Life Technologies) local alignment algorithm with parameters of 25 bp seed length, 2 mismatches in a seed, and mismatch penalty score -2.0 (default threshold). The algorithm finds genomic regions that match to the first 25 bp of each read, allowing at most 2 mismatches, and extending the regions until the score exceeds the threshold. 'PCR and optical duplicates' reads (defined by BioScope; mapped to more than 100 loci, duplicates) and mate-pair reads incongruently mapped on the reference genome (unpaired reads) were filtered out. All mapped sequence reads were deposited to public databases (DNA data bank of Japan (DDBJ) Sequence Read Archive: DRA000430). Chinese rhesus macaque and Vietnamese cynomolgus macaque genome sequences were downloaded from the public database (accession numbers SRA023855 and SRA023856) and aligned to the rhesus macaque genome sequence using the Bowtie 2 program [38] with a local alignment algorithm. A pre-aligned African genome sequence (NA19239) was retrieved from the 1000 Genomes project website [39]. In all resequenced genomes, SNVs were called using SAMtools with a default parameter setting, except for a mismatch tuning parameter (option -C) of 50.

Indel detection

The detection and calling of small and large indels were performed using the software implemented in BioScope software v1.3.1. Briefly, small indels were identified using sequence reads mapped with alignment gaps, and large indels were identified using incongruent distances between mate-pair reads. The small indel-finding algorithm could detect deletions shorter than 12 bp and insertions shorter than 4 bp. In both analyses, a default setting of parameters was applied.

Gene annotation

Entrez Gene annotations in the National Center for Biotechnology Information database were used for classifying SNVs into annotations [40]. Genes assigned to multiple genomic loci were excluded from the analysis. Among 27,424 annotated transcripts in the Indian rhesus macaque genome, 944 showed inconsistencies with the draft genome sequence and were removed from further analyses. When we counted the number of variants at a site with overlapping annotations, we assigned an order of priority as follows: coding exon > non-coding exon > intron > intergenic. For example, when a site was annotated as a coding exon of some transcripts and as an intron of the others, the site was classified as a coding exon. In total, 19,574 protein-coding genes, consisting of 26,480 transcripts, were analyzed. Orthologous genes between human and macaque were determined using the

annotations of the Ensembl database [41]. Only one-to-one orthologs were used for subsequent analyses.

Estimation of demographic parameters

We used PSMC (pairwise sequentially Markovian coalescent) software to infer the demographic history of the Malaysian cynomolgus macaque [31]. Briefly, the program estimates the distribution of coalescent time between two haploid genomes, deduced from the rate of heterozygous SNVs across the genome sequence, with ancestral recombination events inferred by the hidden Markov model. The following parameters were used: time interval = 6 + 29*2, generation time = 6, mutation rate per generation = 2.5×10^{-8}, and the number of iterations = 25. The 95% confidence intervals were estimated using 200 times bootstrap resampling of 5 Mb genome blocks.

Additional material

Additional file 1: Figures S1 to S5 and Tables S1 to S3. Figure S1: chromosomal distribution of fold coverage of quality controlled mapped reads (duplicates and unpaired mate-pair reads were filtered out) on the reference rhesus macaque genome are shown. All chromosomes exceed 37-fold. Figure S2: minimum coverage of quality controlled mapped reads (duplicates and unpaired mate-pair reads were filtered out) on the reference rhesus macaque genome is shown. Genomic regions with at least five-fold coverage were used in the SNV analysis. Figure S3: SNV density along each chromosome. The red and blue lines represent the number of heterozygous and homozygous SNVs in 1 Mb windows, respectively. The step size of window sliding was 100 kb. Figure S4: small indel discovery rate and rhesus macaque genome quality. The red and blue lines represent the rate of small deletions and insertions, respectively, with given rhesus macaque genome sequence quality values (QVs). Small indels at sites having QV < 45 in the rhesus macaque genome sequence were filtered out. Figure S5: distribution of large-indel lengths identified in the cynomolgus macaque genome. Indels were identified using the distance information from the mate-pair libraries. Indel regions containing ambiguous genome sequences were excluded. Table S1: summary of SOLiD libraries and sequence reads (mapped to the Vietnamese cynomolgus macaque genome sequence). Table S2: pattern of nucleotide changes. Table S3: immune- and drug-response genes with completely segregating nonsynonymous SNVs between cynomolgus and rhesus macaques.

Abbreviations

BAC: bacterial artificial chromosome; QV: quality value; SNV: single nucleotide variant; UTR: untranslated region.

Acknowledgements

This study was conducted through the Cooperative Research Program at the Tsukuba Primate Research Center, National Institute of Biomedical Innovation (supported by the Ministry of Health, Labour and Welfare, Japan). This work was partially supported by a Grant-in-Aid for Young Scientists (B) KAKENHI 22700460 and 24700428.

Author details

[1]Laboratory of Rare Disease Biospecimen, Department of Disease Bioresources Research, National Institute of Biomedical Innovation, 7-6-8 Saito-asagi, Ibaraki, Osaka 567-0085, Japan. [2]Center for Human Evolution Modeling Research, Primate Research Institute, Kyoto University, Inuyama, Aichi 484-8506, Japan. [3]Tsukuba Primate Research Center, National Institute of Biomedical Innovation, 1-1 Hachimandai, Tsukuba, Ibaraki 305-0843, Japan. [4]Division of Evolutionary Genetics, Department of Population Genetics, National Institute of Genetics, 1111 Yata, Mishima, Shizuoka 411-8540, Japan. [5]Department of Genetics, The Graduate University for Advanced Studies (SOKENDAI), 1111 Yata, Mishima, Shizuoka 411-8540, Japan.

Authors' contributions

AH, RS, TM, YY and NO contributed to the design of this research. AH, YK, IT, RT and NO performed the experiments. AH, RS, MH and NO contributed to data analysis. AH, RS and NO wrote the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 9 December 2011 Revised: 20 June 2012
Accepted: 2 July 2012 Published: 2 July 2012

References

1. Carlsson HE, Schapiro SJ, Farah I, Hau J: **Use of primates in research: a global overview.** *Am J Primatol* 2004, **63**:225-237.
2. Fooden J: **Provisional classifications and key to living species of macaques (primates: *Macaca*).** *Folia Primatol (Basel)* 1976, **25**:225-236.
3. Uno Y, Iwasaki K, Yamazaki H, Nelson DR: **Macaque cytochromes P450: nomenclature, transcript, gene, genomic structure, and function.** *Drug Metab Rev* 2011, **43**:346-361.
4. Ebeling M, Kung E, See A, Broger C, Steiner G, Berrera M, Heckel T, Iniguez L, Albert T, Schmucki R, Biller H, Singer T, Certa U: **Genome-based analysis of the nonhuman primate *Macaca fascicularis* as a model for drug safety assessment.** *Genome Res* 2011, **21**:1746-1756.
5. Gibbs RA, Rogers J, Katze MG, Bumgarner R, Weinstock GM, Mardis ER, Remington KA, Strausberg RL, Venter JC, Wilson RK, Batzer MA, Bustamante CD, Eichler EE, Hahn MW, Hardison RC, Makova KD, Miller W, Milosavljevic A, Palermo RE, Siepel A, Sikela JM, Attaway T, Bell S, Bernard KE, Buhay CJ, Chandrabose MN, Dao M, Davis C, Delehaunty KD, Ding Y, *et al*: **Evolutionary and biomedical insights from the rhesus macaque genome.** *Science* 2007, **316**:222-234.
6. Fooden J: **Rhesus and crab-eating macaques: intergradation in Thailand.** *Science* 1964, **143**:363-364.
7. Street SL, Kyes RC, Grant R, Ferguson B: **Single nucleotide polymorphisms (SNPs) are highly conserved in rhesus (*Macaca mulatta*) and cynomolgus (*Macaca fascicularis*) macaques.** *BMC Genomics* 2007, **8**:480.
8. Osada N, Uno Y, Mineta K, Kameoka Y, Takahashi I, Terao K: **Ancient genome-wide admixture extends beyond the current hybrid zone between *Macaca fascicularis* and *M. mulatta*.** *Mol Ecol* 2010, **19**:2884-2895.
9. Osada N, Hashimoto K, Kameoka Y, Hirata M, Tanuma R, Uno Y, Inoue I, Hida M, Suzuki Y, Sugano S, Terao K, Kusuda J, Takahashi I: **Large-scale analysis of *Macaca fascicularis* transcripts and inference of genetic divergence between *M. fascicularis* and *M. mulatta*.** *BMC Genomics* 2008, **9**:90.
10. Yan G, Zhang G, Fang X, Zhang Y, Li C, Ling F, Cooper DN, Li Q, Li Y, van Gool AJ, Du H, Chen J, Chen R, Zhang P, Huang Z, Thompson JR, Meng Y, Bai Y, Wang J, Zhuo M, Wang T, Huang Y, Wei L, Li J, Wang Z, Hu H, Yang P, Le L, Stenson PD, Li B, *et al*: **Genome sequencing and comparison of two nonhuman primate animal models, the cynomolgus and Chinese rhesus macaques.** *Nat Biotechnol* 2011, **29**:1019-1023.
11. Fang X, Zhang Y, Zhang R, Yang L, Li M, Ye K, Guo X, Wang J, Su B: **Genome sequence and global sequence variation map with 5.5 million SNPs in Chinese rhesus macaque.** *Genome Biol* 2011, **12**:R63.
12. Fawcett GL, Raveendran M, Deiros DR, Chen D, Yu F, Harris RA, Ren Y, Muzny DM, Reid JG, Wheeler DA, Worley KC, Shelton SE, Kalin NH, Milosavljevic A, Gibbs R, Rogers J: **Characterization of single-nucleotide variation in Indian-origin rhesus macaques (*Macaca mulatta*).** *BMC Genomics* 2011, **12**:311.
13. Kanthaswamy S, Satkoski J, George D, Kou A, Erickson BJ, Smith DG: **Interspecies hybridization and the stratification of nuclear genetic variation of rhesus (*Macaca mulatta*) and long-tailed macaques (*Macaca fascicularis*).** *Int J Primatol* 2008, **29**:1295-1311.

14. Smith DG, McDonough JW, George DA: **Mitochondrial DNA variation within and among regional populations of longtail macaques (*Macaca fascicularis*) in relation to other species of the *fascicularis* group of macaques.** *Am J Primatol* 2007, 69:182-198.
15. Stevison LS, Kohn MH: **Determining genetic background in captive stocks of cynomolgus macaques (*Macaca fascicularis*).** *J Med Primatol* 2008, 37:311-317.
16. Bonhomme M, Cuartero S, Blancher A, Crouau-Roy B: **Assessing natural introgression in 2 biomedical model species, the rhesus macaque (*Macaca mulatta*) and the long-tailed macaque (*Macaca fascicularis*).** *J Hered* 2009, 100:158-169.
17. Delson E: **Fossil macaques, phyletic relationships and a scenario of deployment.** In *The Macaques: Studies in Ecology, Behavior, and Evolution.* Edited by: Lindburg DG. New York: Van Nostrand Reinhold Co; 1980:10-30.
18. Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF, Denisov G, Lin Y, MacDonald JR, Pang AW, Shago M, Stockwell TB, Tsiamouri A, Bafna V, Bansal V, Kravitz SA, Busam DA, Beeson KY, McIntosh TC, Remington KA, Abril JF, Gill J, Borman J, Rogers YH, Frazier ME, Scherer SW, Strausberg RL, *et al*: **The diploid genome sequence of an individual human.** *PLoS Biol* 2007, 5:e254.
19. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, Carter RJ, Keira Cheetham R, Cox AJ, Ellis DJ, Flatbush MR, Gormley NA, Humphray SJ, Irving LJ, Karbelashvili MS, Kirk SM, Li H, Liu X, Maisinger KS, Murray LJ, Obradovic B, Ost T, Parkinson ML, Pratt MR, *et al*: **Accurate whole human genome sequencing using reversible terminator chemistry.** *Nature* 2008, 456:53-59.
20. Wang J, Wang W, Li R, Li Y, Tian G, Goodman L, Fan W, Zhang J, Li J, Zhang J, Guo Y, Feng B, Li H, Lu Y, Fang X, Liang H, Du Z, Li D, Zhao Y, Hu Y, Yang Z, Zheng H, Hellmann I, Inouye M, Pool J, Yi X, Zhao J, Duan J, Zhou Y, Qin J, *et al*: **The diploid genome sequence of an Asian individual.** *Nature* 2008, 456:60-65.
21. Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen Y-J, Makhijani V, Roth GT, Gomes X, Tartaro K, Niazi F, Turcotte CL, Irzyk GP, Lupski JR, Chinault C, Song X-z, Liu Y, Yuan Y, Nazareth L, Qin X, Muzny DM, Margulies M, Weinstock GM, Gibbs RA, Rothberg JM: **The complete genome of an individual by massively parallel DNA sequencing.** *Nature* 2008, 452:872-876.
22. Ahn S-M, Kim T-H, Lee S, Kim D, Ghang H, Kim D-S, Kim B-C, Kim S-Y, Kim W-Y, Kim C, Park D, Lee YS, Kim S, Reja R, Jho S, Kim CG, Cha J-Y, Kim K-H, Lee B, Bhak J, Kim S-J: **The first Korean genome sequence and analysis: Full genome sequencing for a socio-ethnic group.** *Genome Res* 2009, 19:1622-1629.
23. Kim J-I, Ju YS, Park H, Kim S, Lee S, Yi J-H, Mudge J, Miller NA, Hong D, Bell CJ, Kim H-S, Chung I-S, Lee W-C, Lee J-S, Seo S-H, Yun J-Y, Woo HN, Lee H, Suh D, Lee S, Kim H-J, Yavartanoo M, Kwak M, Zheng Y, Lee MK, Park H, Kim JY, Gokcumen O, Mills RE, Zaranek AW, *et al*: **A highly annotated whole-genome sequence of a Korean individual.** *Nature* 2009, 460:1011-1015.
24. Fujimoto A, Nakagawa H, Hosono N, Nakano K, Abe T, Boroevich KA, Nagasaki M, Yamaguchi R, Shibuya T, Kubo M, Miyano S, Nakamura Y, Tsunoda T: **Whole-genome sequencing and comprehensive variant analysis of a Japanese individual using massively parallel sequencing.** *Nat Genet* 2010, 42:931-936.
25. **QFbase.**. [http://genebank.nibio.go.jp/cgi-bin/gbrowse/rheMac2/].
26. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup GPDP: **The Sequence Alignment/Map format and SAMtools.** *Bioinformatics* 2009, 25:2078-2079.
27. McKernan KJ, Peckham HE, Costa GL, McLaughlin SF, Fu Y, Tsung EF, Clouser CR, Duncan C, Ichikawa JK, Lee CC, Zhang Z, Ranade SS, Dimalanta ET, Hyland FC, Sokolsky TD, Zhang L, Sheridan A, Fu H, Hendrickson CL, Li B, Kotler L, Stuart JR, Malek JA, Manning JM, Antipova AA, Perez DS, Moore MP, Hayashibara KC, Lyons MR, Beaudoin RE, *et al*: **Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding.** *Genome Res* 2009, 19:1527-1541.
28. Nei M: *Molecular Evolutionary Genetics* Columbia University Press; 1987.
29. Ohta T: **The nearly neutral theory of molecular evolution.** *Annu Rev Ecol Systematics* 1992, 23:263-286.
30. Dutrillaux B, Biemont MC, Viegas Pequignot E, Laurent C: **Comparison of the karyotypes of four Cercopithecoidae: *Papio papio, P. anubis, Macaca mulatta,* and *M. fascicularis*.** *Cytogenet Cell Genet* 1979, 23:77-83.
31. Li H, Durbin R: **Inference of human population history from individual whole-genome sequences.** *Nature* 2011, 475:493-496.
32. Hernandez RD, Hubisz MJ, Wheeler DA, Smith DG, Ferguson B, Rogers J, Nazareth L, Indap A, Bourquin T, McPherson J, Muzny D, Gibbs R, Nielsen R, Bustamante CD: **Demographic histories and patterns of linkage disequilibrium in Chinese and Indian rhesus macaques.** *Science* 2007, 316:240-243.
33. Higashino A, Osada N, Suto Y, Hirata M, Kameoka Y, Takahashi I, Terao K: **Development of an integrative database with 499 novel microsatellite markers for *Macaca fascicularis*.** *BMC Genet* 2009, 10:24.
34. Matsumoto J, Kawai S, Terao K, Kirinoki M, Yasutomi Y, Aikawa M, Matsuda H: **Malaria infection induces rapid elevation of the soluble Fas ligand level in serum and subsequent T lymphocytopenia: possible factors responsible for the differences in susceptibility of two species of *Macaca* monkeys to *Plasmodium coatneyi* infection.** *Infect Immun* 2000, 68:1183-1188.
35. Hamada Y, Urasopon N, Hadi I, Malaivijitnond S: **Body size and proportions and pelage color of free-ranging *Macaca mulatta* from a zone of hybridization in Northeastern Thailand.** *Int J Primatol* 2006, 27:497-513.
36. Heaney LR: **A synopsis of climatic and vegetational change in Southeast-Asia.** *Climatic Change* 1991, 19:53-61.
37. **UCSC Genome Browser.**. [http://ucsc.genome.edu/].
38. **Bowtie 2.**. [http://bowtie-bio.sourceforge.net/bowtie2/index.shtml].
39. **1000 Genomes.**. [http://www.1000genomes.org/].
40. Maglott D, Ostell J, Pruitt KD, Tatusova T: **Entrez Gene: gene-centered information at NCBI.** *Nucleic Acids Res* 2011, 39:D52-D57.
41. Flicek P, Amode MR, Barrell D, Beal K, Brent S, Chen Y, Clapham P, Coates G, Fairley S, Fitzgerald S, Gordon L, Hendrix M, Hourlier T, Johnson N, Kähäri A, Keefe D, Keenan S, Kinsella R, Kokocinski F, Kulesha E, Larsson P, Longden I, McLaren W, Overduin B, Pritchard B, Riat HS, Rios D, Ritchie GRS, Ruffier M, Schuster M, *et al*: **Ensembl 2011.** *Nucleic Acids Res* 2011, 39:D800-D806.

doi:10.1186/gb-2012-13-7-r58
Cite this article as: Higashino *et al.*: Whole-genome sequencing and analysis of the Malaysian cynomolgus macaque (*Macaca fascicularis*) genome. *Genome Biology* 2012 13:R58.

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**).

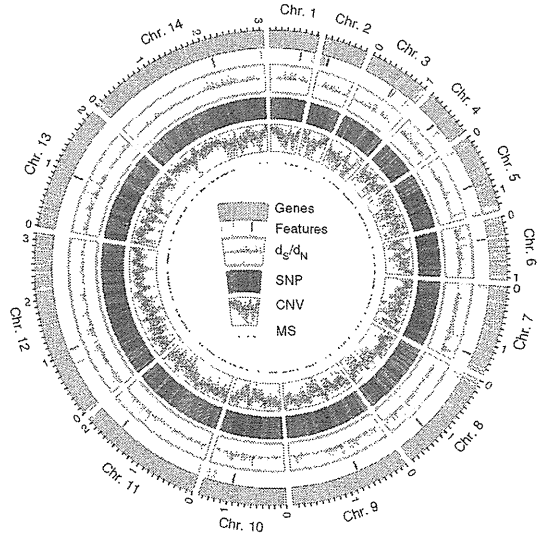


Figure 1 Architecture of the *P. cynomolgi* genome and associated genome-wide variation data. Data are shown for each of the 14 *P. cynomolgi* chromosomes. The six concentric rings, from outermost to innermost, represent (i) the location of 5,049 *P. cynomolgi* genes, excluding those on small contigs (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

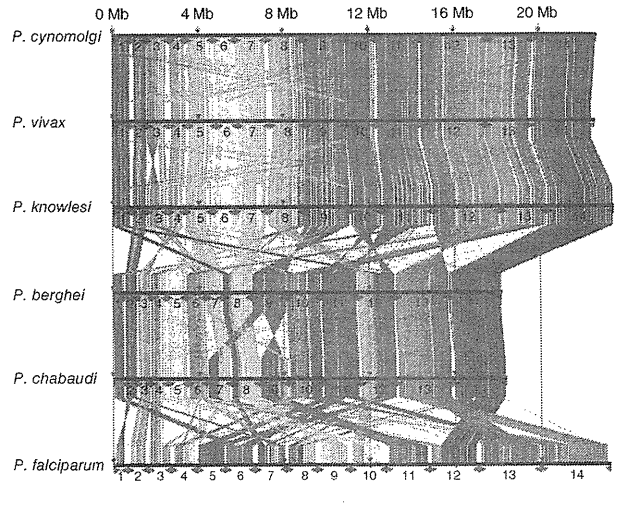


Figure 2 Genome synteny between six species of *Plasmodium* parasite. Protein-coding genes of *P. cynomolgi* are shown aligned with those of five other *Plasmodium* genomes: two species belonging to the monkey malaria clade, *P. vivax* and *P. knowlesi*; two species of rodent malaria, *P. berghei* and *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*.

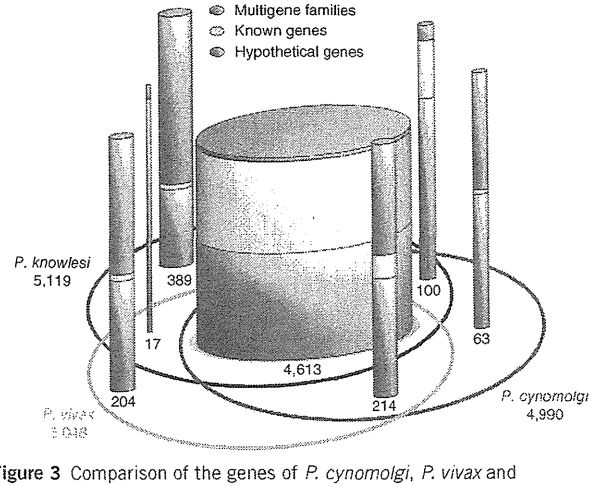


Figure 3 Comparison of the genes of *P. cynomolgi*, *P. vivax* and *P. knowlesi*. The Venn ellipses represent the three genomes, with the total number of genes assigned to the chromosomes indicated under the species name. Cylinders depict orthologous and non-orthologous genes between the three genomes, 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).

	Journal	MSP No.	Dispatch: February 5, 2013	CE: Dinesh Kumar
	MIM	12025	No. of Pages: 6	PE: Gisella Huang

Microbiol Immunol 2013; xx: 1–6
doi: 10.1111/1348-0421.12025

ORIGINAL ARTICLE

Genomic polymorphisms in 3β-hydroxysterol Δ24-reductase promoter sequences

Nagla Elwy Salem[1,2,3], Makoto Saito[1], Yuri Kasama[1], Makoto Ozawa[4,5], Toshiko Kawabata[4,5], Shinji Harada[2], Hiroko Suda[6], Katsuhiro Asonuma[6], Ahmed El-Gohary[3] and Kyoko Tsukiyama-Kohara[1,4,5]

[1]Department of Experimental Phylaxiology, Kumamoto University, Kumamoto, Japan, [2]Department of Medical Virology, Faculty of Life Sciences, Kumamoto University, Kumamoto, Japan, [3]Department of Clinical Pathology, Faculty of Medicine Suez Canal University, Ismailia, Egypt, [4]Transboundary Animal Diseases Center, Joint Faculty of Veterinary Medicine, Kagoshima University, Kagoshima, Japan, [5]Laboratory of Animal Hygiene, Joint Faculty of Veterinary Medicine, Kagoshima University, Kagoshima, Japan and [6]Department of Transplantation and Pediatric Surgery, Postgraduate School of Medical Science, Kumamoto University, Kumamoto, Japan

ABSTRACT

It was recently reported by the present team that 3β-hydroxysterol Δ24-reductase (DHCR24) is induced by hepatitis C virus (HCV) infection. In addition, upregulation of DHCR24 impairs p53 activity. In human hepatoma HuH-7 cells, the degree of DHCR24 expression is higher than in normal hepatic cell lines (WRL68) at the transcriptional level. The genomic promoter sequence of DHCR24 was characterized and nucleotide substitutions were observed in HuH-7 cells at nucleotide numbers −1453 (G to A), −1420 (G to T), −488 (A to C) and −200 (G to C). The mutations of these sequences from HuH-7 cell types to WRL68 cell types suppressed DHCR24 gene promoter activity. The sequences were further characterized in hepatocytes from patient tissues. Four tissues from HCV-positive patients with cirrhosis or hepatocellular carcinoma (#1, 2, 3, 5) possessed HuH-7 cell type sequences. Interestingly, one patient with liver cirrhosis (#4) possessed WRL68 cell-type sequences; this patient had been infected with HCV and was HCV negative for 17 years after interferon therapy. Next, the effect of HCV infection on these polymorphisms was examined in humanized chimeric mouse liver and HuH-7 cells. The human hepatocytes possess WRL68 cell type and did not show the nucleotide substitution after HCV infection. The HCV-replicon was removed by interferon treatment and established the cured K4 cells. These cells possess HuH-7 cell type sequences. Thus, this study showed the genomic polymorphism in DHCR24 promoter is not directly influenced by HCV infection.

Key words 3β-hydroxysterol Δ24-reductase, hepatitis C virus, promoter.

Liver cancer is one of the most prevalent forms of cancer (1). More than 80% of cases occur in developing countries; however, Japan also has a remarkably high incidence (2). Among the primary liver cancers, HCC is the most common (3). Its incidence is increasing: between 1975 and 2005, age-adjusted HCC rates tripled (4).

One crucial cause of HCC is HCV infection (5). DHCR24, which functions as an oxidoreductase during cholesterol biosynthesis (6, 7), is linked to HCV-associated hepatocarcinogenesis and development of HCC (8–10). Infection of hepatocytes with HCV results in over-expression of DHCR24. This enzyme protects cells from oxidative stress and inhibits p53 activity (8), thus

Correspondence
Kyoko Tsukiyama-Kohara, Transboundary Animal Diseases Center, Joint Faculty of Veterinary Medicine Kagoshima University, 1-21-24 Korimoto, Kagoshima 890-0065, Japan.
Tel: +81 99 285 3589; fax: +81 99 285 3589; email: kkohara@agri.kagoshima-u.ac.jp

Received 24 October 2012; revised 10 December 2012; accepted 21 December 2012.

List of Abbreviations: DHCR24, 3β-hydroxysterol Δ24-reductase; DMEM, Dulbecco's modified Eagle's medium; HCC, hepatocellular carcinoma; HCV, hepatitis C virus; IFN, interferon; SVR, sustained viral response.

contributing to the development of HCC (5). These facts prompted us to investigate whether the molecular features of *DHCR24* are linked to HCC development. To this end, we characterized the promoter region of *DHCR24* in HCC cell lines and clinical samples.

MATERIALS AND METHODS

Cell lines and growth conditions

HuH-7 and HepG2 cells were cultured in (DMEM; Sigma-Aldrich, St. Louis, MO, USA) supplemented with 10% FCS (Sigma-Aldrich). WRL68 cells were cultured in DMEM supplemented with 1 mM sodium pyruvate (Invitrogen, Carlsbad, CA, USA), 0.1 mM non-essential amino acids (Invitrogen) and 10% FCS. HuH-7 cell-based HCV replicon harboring cell lines (R6FLR-N) (11) were cured off HCV by interferon treatment (12) and designated as K4 cells.

Northern and western blotting

Northern and western blotting were performed as previously described (8).

Sequencing of genomic DNA and reporter plasmid construction

Genomic DNA was extracted from HuH-7 and WRL68 cells using standard methods. DNA from the promoter region of *DHCR24* (~5 kb) was amplified using PCR (sense primer: 5′-CACTCCTGCTCACCACTGAT-3′; antisense primer: 5′-GTAGTAGATATCGAAGATAAGCGAGAGCGG-3′). These fragments were individually cloned into the upstream region of the firefly luciferase gene in the pGL3-Basic vector (Promega, Madison, WI, USA) at the *Xho*I and *Nco*I sites (as we had done previously for the HepG2 cell line) (6). DNA sequences were determined using standard methods. Reporter plasmids that possessed chimeric promoters were constructed using restriction enzyme sites for *Tth111*I (position −2160) and *Bss*HII (position −1030).

Dual luciferase reporter assay

Using Lipofectamine LTX (Invitrogen), HepG2 cells (1×10^4 cells/well in a 96-well plate) were transfected with a reporter plasmid (0.25 μg/well) together with an internal control plasmid (phRL-TK; 0.025 μg/well) encoding *Renilla* luciferase (Promega). Forty-eight hours after transfection, the cells were assayed with the Dual-Glo Luciferase Assay System (Promega). Luminescence was measured using a TriStar LB941 microplate reader (Berthold Technologies GmbH, Bad Wildbad, Germany).

Liver tissue samples from chimeric mice or patients infected with hepatitis C virus

Severely combined immunodeficient mice carrying human primary hepatocytes were purchased from BD BioSciences (Franklin Lakes, NJ, USA) and African American, male, 5-year-old, HCV negative mice from PhoenixBio (Hiroshima, Japan) (13). These "human liver chimeric" mice were inoculated or mock-inoculated with plasma collected from an HCV-positive (HCR6 strain (14), GenBank accession #AY045702) patient in accordance with the requirements of the Declaration of Helsinki. HCV infection in the mice thus infected was confirmed by using quantitative PCR for HCV mRNA as previously described (9). The protocols for the animal experiments were pre-approved by the local Ethics Committee, and the animals were maintained in accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals.

Informed consent for this clinical study was obtained from five patients with HCV (Table 1) at the Kumamoto University Hospital (Kumamoto, Japan), in accordance with the Helsinki Declaration prior to 2003, and the protocol was approved by the Regional Ethics Committee. LiverPool 20-donor pooled cryopreserved human hepatocytes (Celsis IVT, Baltimore, MD, USA) were purchased and used as the normal human liver tissue control. HCV RNA was detected by the COBAS TaqMan HCV test (Hoffman–La Roche, Basel, Switzerland). Liver

Table 1. Summary of patients with HCC

Patient ID	Sex	Age (years)	Diagnosis	ALT (IU/mL)	Outcome of IFN treatment	HCV RNA detection[†]
#1	F	60	LC	22	NR	+
#2	M	65	LC	31	NT	+
#3	F	57	LC	24	NR	+
#4	M	61	LC, HCC	12	SVR	−[‡]
#5	M	51	LC, HCC	91	NR	+

[†]Serum was tested for HCV RNA using quantitative PCR; [‡]In 1995, #4 was diagnosed with HCV-associated LC and HCC and HCV RNA was detected in his serum. As a result, #4 was treated with IFN. Since then, no HCV RNA has been detected in this patient's serum (>17 years). ALT, alanine aminotransferase; F, female; LC, liver cirrhosis; M, male; NR, no response; NT, not treated.

tissue was obtained from either mice or patients and processed for DNA sequencing. Two DNA fragments (corresponding to positions −1600 to −1292 and −631 to −86) were amplified using PCR with Tth-Bss forward and reverse primers (5′-ATTTCAACATGTCATTAACA-3′ and 5′-TTCTAGCACGGTGCTTTGTG-3′) and Bss-Nco forward and reverse primers (5′-CCAGCCA-TAGCCTTCCATG-3′ and 5′-AATGGCGAGCCGCGC CGG-3′), respectively. The amplified fragments were directly sequenced using the same set of primers.

Statistical analysis

Student's *t*-test was used to test the statistical significance of the results. P values of < 0.05 were considered statistically significant.

RESULTS

First, we measured DHCR24 expression in cell lines of noncancerous hepatocytes (WRL68) and hepatoma cells (HuH-7 and HepG2). Compared with noncancerous hepatocytes, DHCR24 expression in the two hepatoma cell lines was considerably increased with respect to both mRNA and protein (Fig. 1a, b). In addition, the different culture media used for the WRL68 and HuH-7 cells did not significantly influence the degree of expression of DHCR24 protein (Fig. 1c).

To identify the genetic characteristic(s) that govern DHCR24 upregulation, we isolated genomic DNA from these three cell lines and sequenced the *DHCR24* promoter region (nucleotide positions −4976 to +113, where +1 indicates the transcription start site). For this analysis, we sequenced three molecular clones from each cell line. Alignments of WRL68 and HuH-7 sequences showed different nucleotides at four positions: (i) an A to G switch at −1453 (i.e., A in WRL68 and G in HuH-7); (ii) a T to G switch at −1420; (iii) a C to A switch at −488; and (iv) a C to G switch at −200 (Fig. 2). The two hepatoma cell lines (HuH-7 and HepG2) had no nucleotide differences within these regions.

Next, we investigated whether these small changes in the promoter sequence affect gene expression in a heterologous context. We constructed reporter plasmids that placed the firefly luciferase gene under the control of *DHCR24* promoter sequences (either from HuH-7 or WRL68 cells) (Fig. 3a). We measured the promoter activity of each construct in HepG2 cells with dual-luciferase assays. The *DHCR24* promoter derived from HuH-7 cells showed significantly greater activity (i.e., induced greater expression) than the WRL68 promoter (Fig. 3b). We also constructed two reporter plasmids that contained chimeric promoters. In each of these chimeras,

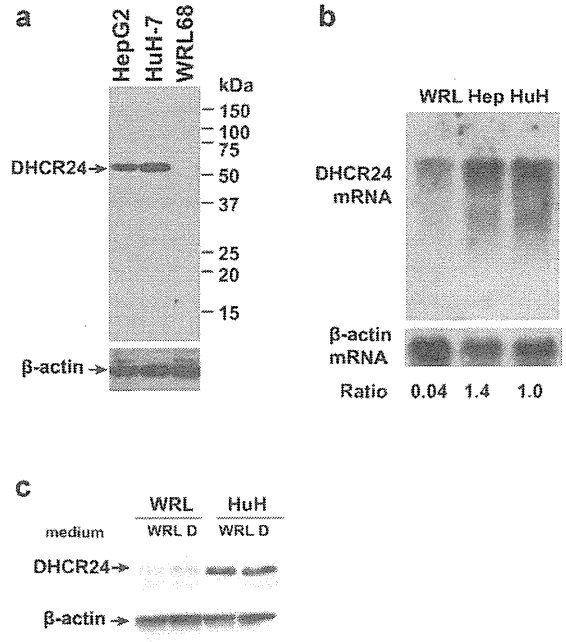


Fig. 1. Expression of DHCR24 in hepatoma cell lines. (a) Lysates from WRL68, HuH-7 and HepG2 cells were subjected to western blot analysis using antibodies directed against DHCR24 (upper panel) and β-actin (lower panel). (b) RNA was extracted from WRL68, HepG2 and HuH-7 cells and subjected to northern blot analysis using probes specific for *DHCR24* (upper panel) and β-actin (lower panel). Band intensities were quantified with a densitometer. Relative band intensity ratios (*DHCR24*/β-actin) are indicated below the gel images (the ratio for HuH-7 cells was set at 1). (c) DHCR24 protein (upper panel) and β-actin (lower panel) were detected in WRL 68 or HuH-7 cells with culture media for WRL68 cells (DMEM, 1 mM sodium pyruvate and 1 mM nonessential amino acids) or HuH-7 cells (DMEM alone).

we replaced HuH-7 fragments containing two polymorphisms with wild-type WRL68 sequences (Fig. 3a). These chimeric promoters had less activity than did intact promoters from both HuH-7 and WRL68 cells (Fig. 3b). These results indicate that the *DHCR24* promoter from HuH-7 cells contributes to the strong degree of *DHCR24* expression. In addition, all four nucleotide sequences of HuH-7 cell type in promoter fragments might be important for strong promoter activity.

Thereafter, we examined whether polymorphisms within the *DHCR24* promoter could be detected in clinical samples. We collected samples of liver tissue from five patients infected with HCV (Table 1) and sequenced the *DHCR24* promoter region (Table 2). Of the five samples tested, four (#1–3 and #5) showed all four of the polymorphisms associated with strong promoter activity (i.e., G, G, A and G nucleotides at positions −1453, −1420, −488, and −200). In contrast, promoter

COLOUR ONLINE, B&W IN PRINT

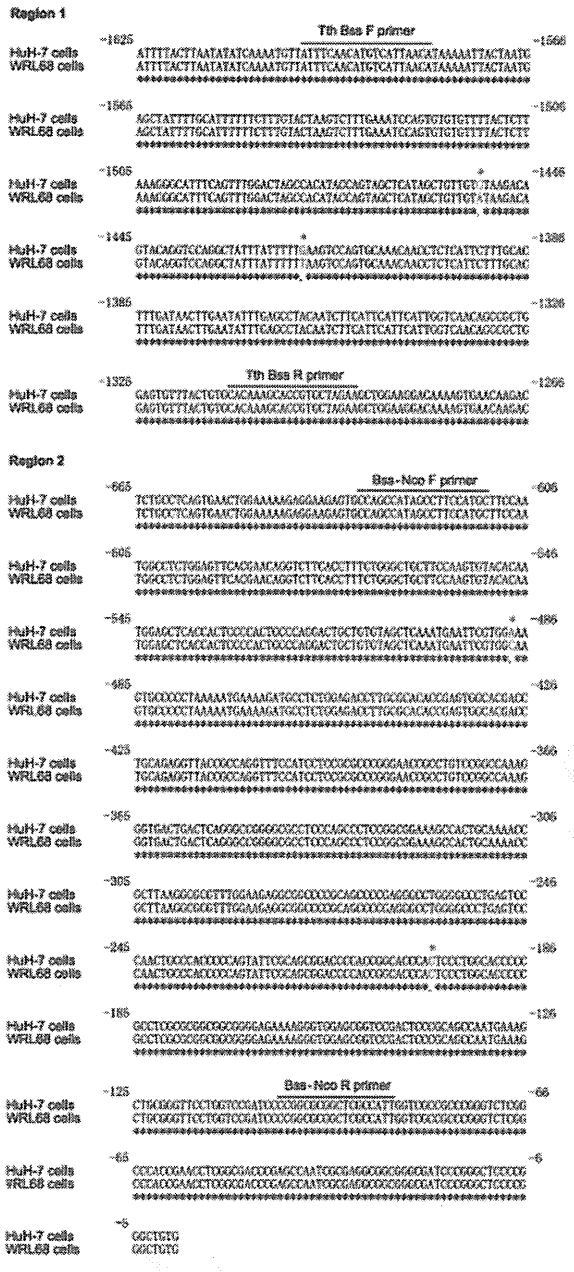


Fig. 2. Alignment of *DHCR24* promoter sequences. Nucleotide sequences from *DHCR24* promoter regions obtained from HuH-7 and WRL68 cells are shown. Cell-type-specific differences between these sequences (at positions −1453, −1420, −488, and −200) are indicated by asterisks and colors (red and blue represent nucleotides in HuH-7 and WRL68 cells, respectively). Positions of primer sequences are indicated.

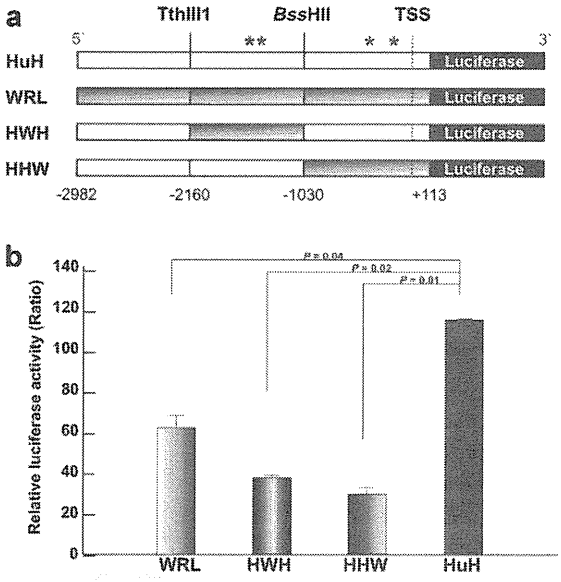


Fig. 3. Effect of nucleotide changes on *DHCR24* promoter activity. (a) Schematic diagrams of reporter constructs. Intact or chimeric *DHCR24* promoter sequences were used to drive the expression of firefly luciferase. Fragments of DNA derived from HuH-7 and WRL68 cells are colored white and grey, respectively. The asterisks indicate the position of each nucleotide polymorphism. Restriction enzyme sites (*Tth111*I and *Bss*HII) and transcription start sites are indicated. (b) Promoter activity of reporter constructs. HepG2 cells were transfected with the indicated reporter construct together with a control plasmid encoding *Renilla* luciferase. The relative ratio of firefly/*Renilla* luciferase activity is shown. Error bars indicate the standard deviation of two independent experiments. Each experiment was performed in triplicate. TSS, transcription start sites.

Table 2. Summary of nucleotide substitutions within the *DHCR24* promoter region

	Nucleotide position			
Origin of DNA sample	−1453	−1420	−488	−200
HuH-7 cells (high)†	G	G	A	G
WRL68 cells (low)‡	A	T	C	C
Patient #1	G	G	A	G
Patient #2	G	G	A	G
Patient #3	G	G	A	G
Patient #4	A	T	C	C
Patient #5 (NC)	G	G	A	G
Patient #5 (C)	G	G	A	G
20-donor pool§	A	T	C	C

†DHCR24 was expressed strongly in HuH-7 cells (Fig. 1), ‡DHCR24 was expressed weakly in WRL68 cells (Fig. 1), §Pooled normal human hepatocytes from a 20-donor pool. C, cancerous region; NC, non-cancerous region.

Table 3. Summary of nucleotide substitutions within the *DHCR24* promoter region with or without HCV infection

		Nucleotide position			
Origin of DNA sample	HCV	−1453	−1420	−488	−200
HuH-7 cells (high)	−	G	G	A	G
WRL68 cells (low)	−	A	T	C	C
Chimeric mouse liver	−	A	T	C	C
HCV infected chimeric mouse liver	+[†]	A	T	C	C
HCV replicon cells (R6FLR-N)	+	G	G	A	G
Cured K4 cells	+	G	G	A	G

[†]7.5×10^6 copies/mL of HCV in patient plasma was inoculated.

sequences from patient 4 (#4) had nucleotides associated with weak activity at these positions (i.e., A, T, C, and C). Intriguingly, only #4 exhibited an SVR, which is characterized by the absence of detectable HCV RNA in serum for >24 weeks following IFN treatment. The SVR status of #4 has persisted since 1995. In #5, promoter sequences were the same in cancerous and non-cancerous regions of the liver. These results suggest that the four polymorphisms within the *DHCR24* promoter region may influence the susceptibility to malignancy and IFN responsiveness of hepatoma cells and thus influence the fate of patients with HCC.

To assess the impact of HCV infection on genomic polymorphism in DHCR24 promoter sequences, we determined the sequences in human hepatocytes that had been transplanted into severely combined immunodeficient mice that we infected or mock-infected with HCV We detected markedly high titers of HCV only in the infected mice (Table 3). Sequencing revealed that all four polymorphic nucleotide positions were of the weak activity type. Notably, we detected no nucleotide differences between HCV- and mock-infected mice in the targeted regions (Table 3). We also established cured K4 cells by treating HCV replicon cells R6FLR-N with IFN. Analysis of the genomic sequence of these cell lines showed no nucleotide differences in R6-FLR-N and K4 cells (Table 3). These results suggest that the differences in the DHCR24 promoter sequence are ingenerate rather than induced by HCV infection.

DISCUSSION

In this study, we analyzed the promoter sequences associated with *DHCR24* in hepatocytes and identified polymorphisms that regulate the degree of expression of downstream genes (Figs. 1–3). #4 had an SVR in response to IFN treatment; thus these *DHCR24* promoter sequence polymorphisms are potential biomarkers for predicting patients' responsiveness to IFN treatment.

Genomic polymorphisms within the *DHCR24* promoter region may influence binding of transcription factors (Supplementary Fig. S1). In fact, a T-to-G nucleotide substitution at position −1420 generates a potential binding site for the protein encoded by the caudal homeobox gene (*CdxA*), a homeobox transcription factor responsible for gastrointestinal tract development and epithelial differentiation (15). A C-to-A substitution at position −488 generates potential binding sites for nuclear factor kappa-light-chain enhancer of activated B cells and STATx (16), as well as a low-affinity binding site for Nkx-2 (17). Finally, a C-to-G substitution at position −200 potentially abolishes a p300 binding site (18). These changes in transcription factor binding affinities could upregulate *DHCR24* expression, thereby promote carcinogenesis.

Previously, we discovered that DHCR24 is a host factor involved in HCV-associated development of HCC (8, 9). This protein is upregulated by HCV infection (8), and reduced degrees of expression (via siRNA knockdown) inhibit HCV replication (9). These findings are consistent with the role of DHCR24 in cholesterol biosynthesis (6, 7), which is important for HCV replication (19). Also, because the efficiency of HCV replication might have been lower in #4 than in other patients with strongly active DHCR24 promoter, the weak DHCR24 expression in this patient (Supplementary Fig. S2) might have contributed to the efficacy of IFN treatment.

In conclusion, we have discovered polymorphisms in the promoter region of DHCR24 gene that have not been induced by HCV infection. Future study will clarify their biological significance.

ACKNOWLEDGMENTS

The authors thank Dr Michinori Kohara, Tokyo Metropolitan Institute of Medical Science, Tokyo, Japan for his generous support, which included supplying reagents. This work was supported by grants from the Ministry of Health, Science and Welfare and the Ministry of Education, Science and Culture, Japan.

DISCLOSURE

The authors have no financial relationships to disclosure.

REFERENCES

1. Wong C.M., Ng I.O. (2008) Molecular pathogenesis of hepatocellular carcinoma. *Liver Int* **28**: 160–74.
2. Center M.M., Jemal A. (2011) International trends in liver cancer incidence rates. *Cancer Epidemiol Biomarkers Prev* **20**: 2362–8.
3. Perz J.F., Armstrong G.L., Farrington L.A., Hutin Y.J., Bell B.P. (2006) The contributions of hepatitis B virus and hepatitis C

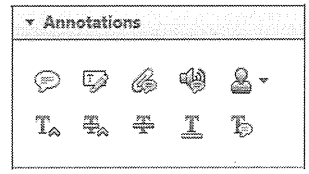
virus infections to cirrhosis and primary liver cancer worldwide. *J Hepatol* **45**: 529–38.
4. Altekruse S.F., McGlynn K.A., Reichman M.E. (2009) Hepatocellular carcinoma incidence, mortality, and survival trends in the United States from 1975 to 2005. *J Clin Oncol* **27**: 1485–91.
5. Farazi P.A., DePinho R.A. (2006) Hepatocellular carcinoma pathogenesis: from genes to environment. *Nat Rev Cancer* **6**: 674–87.
6. Greeve I., Hermans-Borgmeyer I., Brellinger C., Kasper D., Gomez-Isla T., Behl C., Levkau B., Nitsch R.M. (2000) The human DIMINUTO/DWARF1 homolog seladin-1 confers resistance to Alzheimer's disease-associated neurodegeneration and oxidative stress. *J Neurosci* **20**: 7345–52.
7. Wu C., Miloslavskaya I., Demontis S., Maestro R., Galaktionov K. (2004) Regulation of cellular response to oncogenic and oxidative stress by Seladin-1. *Nature* **432**: 640–5.
8. Nishimura T., Kohara M., Izumi K., Kasama Y., Hirata Y., Huang Y., Shuda M., Mukaidani C., Takano T., Tokunaga Y., Nuriya H., Satoh M., Saito M., Kai C., Tsukiyama-Kohara K. (2009) Hepatitis C virus impairs p53 via persistent overexpression of 3beta-hydroxysterol delta24-reductase. *J Biol Chem* **284**: 36442–52.
9. Takano T., Tsukiyama-Kohara K., Hayashi M., Hirata Y., Satoh M., Tokunaga Y., Tateno C., Hayashi Y., Hishima T., Funata N., Sudoh M., Kohara M. (2011) Augmentation of DHCR24 expression by hepatitis C virus infection facilitates viral replication in hepatocytes. *J Hepatol* **55**: 512–21.
10. Saito M., Kohara M., Tsukiyama-Kohara K. (2012) Hepatitis C virus promotes expression of the 3beta-hydroxysterol delta24-reductase through Sp1. *J Med Virol* **84**: 733–46.
11. Watanabe T., Sudoh M., Miyagishi M., Akashi H., Arai M., Inoue K., Taira K., Yoshiba M., Kohara M. (2006) Intracellular-diced dsRNA has enhanced efficacy for silencing HCV RNA and overcomes variation in the viral genotype. *Gene Ther* **13**: 883–92.
12. Blight K.J., McKeating J.A., Rice C.M. (2002) Highly permissive cell lines for subgenomic and genomic hepatitis C virus RNA replication. *J Virol* **76**: 13001–14.
13. Mercer D.F., Schiller D.E., Elliott J.F., Douglas D.N., Hao C., Rinfret A., Addison W.R., Fischer K.P., Churchill T.A., Lakey J.R., Tyrrell D.L., Kneteman N.M. (2001) Hepatitis C virus replication in mice with chimeric human livers. *Nat Med* **7**: 927–33.
14. Inoue K., Umehara T., Ruegg U.T., Yasui F., Watanabe T., Yasuda H., Dumont J.M., Scalfaro P., Yoshiba M., Kohara M. (2007) Evaluation of a cyclophilin inhibitor in hepatitis C virus-infected chimeric mice *in vivo*. *Hepatology* **45**: 921–8.
15. Hecht N.B. (1995) The making of a spermatozoon: a molecular perspective. *Dev Genet* **16**: 95–103.
16. Kim J., Sharma S., Li Y., Cobos E., Palvimo J.J., Williams S.C. (2005) Repression and coactivation of CCAAT/enhancer-binding protein epsilon by sumoylation and protein inhibitor of activated STATx proteins. *J Biol Chem* **280**: 12,246–54.
17. Shiojima I., Komuro I., Mizuno T., Aikawa R., Akazawa H., Oka T., Yamazaki T., Yazaki Y. (1996) Molecular cloning and characterization of human cardiac homeobox gene CSX1. *Circ Res* **79**: 920–9.
18. Rikitake Y., Moran E. (1992) DNA-binding properties of the E1A-associated 300-kilodalton protein. *Mol Cell Biol* **12**: 2826–36.
19. Aizaki H., Lee K.J., Sung V.M., Ishiko H., Lai M.M. (2004) Characterization of the hepatitis C virus RNA replication complex associated with lipid rafts. *Virology* **324**: 450–61.

USING e-ANNOTATION TOOLS FOR ELECTRONIC PROOF CORRECTION

Required software to e-Annotate PDFs: Adobe Acrobat Professional or Adobe Reader (version 8.0 or above). (Note that this document uses screenshots from Adobe Reader X)
The latest version of Acrobat Reader can be downloaded for free at: http://get.adobe.com/reader/

Once you have Acrobat Reader open on your computer, click on the Comment tab at the right of the toolbar:

This will open up a panel down the right side of the document. The majority of tools you will use for annotating your proof will be in the Annotations section, pictured opposite. We've picked out some of these tools below:

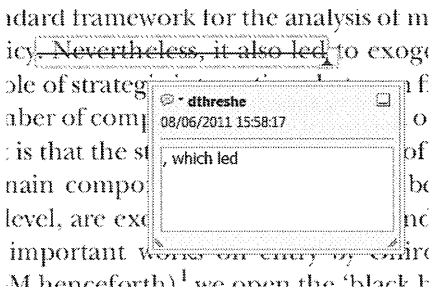


1. Replace (Ins) Tool – for replacing text.

Strikes a line through text and opens up a text box where replacement text can be entered.

How to use it

- Highlight a word or sentence.
- Click on the Replace (Ins) icon in the Annotations section.
- Type the replacement text into the blue box that appears.

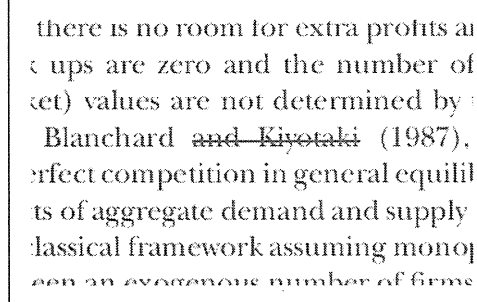


2. Strikethrough (Del) Tool – for deleting text.

Strikes a red line through text that is to be deleted.

How to use it

- Highlight a word or sentence.
- Click on the Strikethrough (Del) icon in the Annotations section.

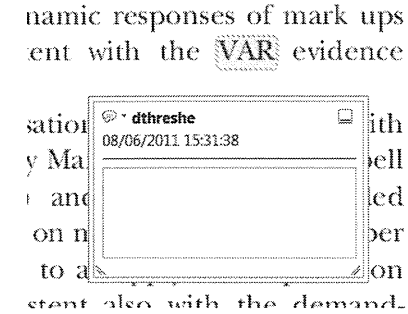


3. Add note to text Tool – for highlighting a section to be changed to bold or italic.

Highlights text in yellow and opens up a text box where comments can be entered.

How to use it

- Highlight the relevant section of text.
- Click on the Add note to text icon in the Annotations section.
- Type instruction on what should be changed regarding the text into the yellow box that appears.

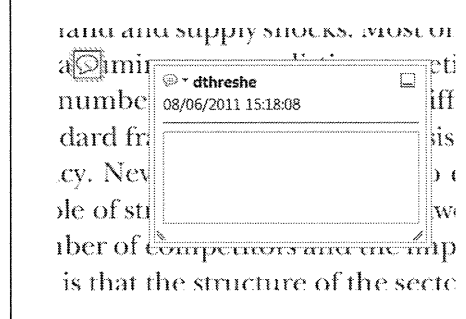


4. Add sticky note Tool – for making notes at specific points in the text.

Marks a point in the proof where a comment needs to be highlighted.

How to use it

- Click on the Add sticky note icon in the Annotations section.
- Click at the point in the proof where the comment should be inserted.
- Type the comment into the yellow box that appears.

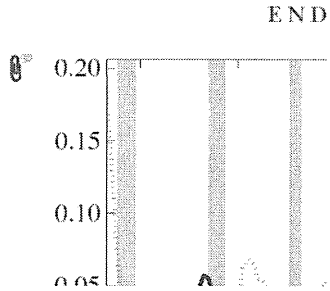


5. Attach File Tool – for inserting large amounts of text or replacement figures.

Inserts an icon linking to the attached file in the appropriate pace in the text.

How to use it

- Click on the Attach File icon in the Annotations section.
- Click on the proof to where you'd like the attached file to be linked.
- Select the file to be attached from your computer or network.
- Select the colour and type of icon that will appear in the proof. Click OK.

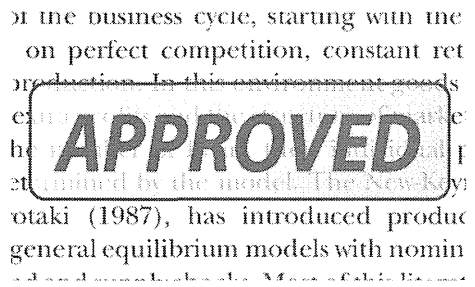


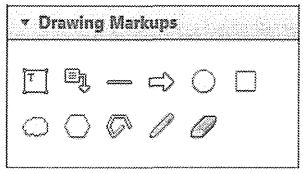
6. Add stamp Tool – for approving a proof if no corrections are required.

Inserts a selected stamp onto an appropriate place in the proof.

How to use it

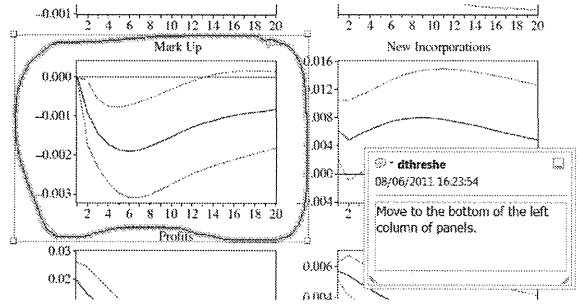
- Click on the Add stamp icon in the Annotations section.
- Select the stamp you want to use. (The Approved stamp is usually available directly in the menu that appears).
- Click on the proof where you'd like the stamp to appear. (Where a proof is to be approved as it is, this would normally be on the first page).





7. Drawing Markups Tools – for drawing shapes, lines and freeform annotations on proofs and commenting on these marks.

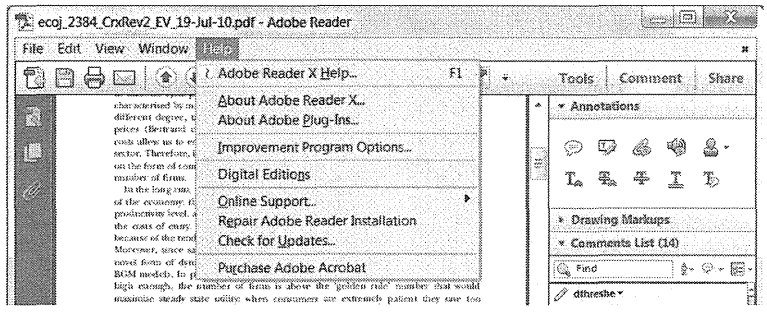
Allows shapes, lines and freeform annotations to be drawn on proofs and for comment to be made on these marks..



How to use it

- Click on one of the shapes in the Drawing Markups section.
- Click on the proof at the relevant point and draw the selected shape with the cursor.
- To add a comment to the drawn shape, move the cursor over the shape until an arrowhead appears.
- Double click on the shape and type any text in the red box that appears.

For further information on how to annotate proofs, click on the Help menu to reveal a list of further options:



OPEN ACCESS Freely available online

PLOS | ONE

Immunization with a Recombinant Vaccinia Virus That Encodes Nonstructural Proteins of the Hepatitis C Virus Suppresses Viral Protein Levels in Mouse Liver

Satoshi Sekiguchi[1], Kiminori Kimura[2], Tomoko Chiyo[1], Takahiro Ohtsuki[1], Yoshimi Tobita[1], Yuko Tokunaga[1], Fumihiko Yasui[1], Kyoko Tsukiyama-Kohara[3], Takaji Wakita[4], Toshiyuki Tanaka[5], Masayuki Miyasaka[6], Kyosuke Mizuno[7], Yukiko Hayashi[8], Tsunekazu Hishima[8], Kouji Matsushima[9], Michinori Kohara[1]*

1 Department of Microbiology and Cell Biology, Tokyo Metropolitan Institute of Medical Science, Setagaya-ku, Tokyo, Japan, **2** Division of Hepatology, Tokyo Metropolitan Komagome Hospital, Bunkyo-ku, Tokyo, Japan, **3** Transboundary Animal Diseases Center, Joint Faculty of Veterinary Medicine, Kagoshima University, Korimoto, Kagoshima, Japan, **4** Department of Virology II, National Institute of Infectious Diseases, Shinjuku-ku, Tokyo, Japan, **5** Laboratory of Immunobiology, Department of Pharmacy, School of Pharmacy, Hyogo University of Health Sciences, Chuo-ku, Kobe, Japan, **6** Laboratory of Immunodynamics, Department of Microbiology and Immunology, Osaka University Graduate School of Medicine, Suita, Osaka, Japan, **7** Chemo-Sero-Therapeutic Research Institute, Okubo, Kumamoto, Japan, **8** Department of Pathology, Tokyo Metropolitan Komagome Hospital, Bunkyo-ku, Tokyo, Japan, **9** Department of Molecular Preventive Medicine, School of Medicine, University of Tokyo, Bunkyo-ku, Tokyo, Japan

Abstract

Chronic hepatitis C, which is caused by infection with the hepatitis C virus (HCV), is a global health problem. Using a mouse model of hepatitis C, we examined the therapeutic effects of a recombinant vaccinia virus (rVV) that encodes an HCV protein. We generated immunocompetent mice that each expressed multiple HCV proteins via a Cre/*loxP* switching system and established several distinct attenuated rVV strains. The HCV core protein was expressed consistently in the liver after polyinosinic acid–polycytidylic acid injection, and these mice showed chronic hepatitis C-related pathological findings (hepatocyte abnormalities, accumulation of glycogen, steatosis), liver fibrosis, and hepatocellular carcinoma. Immunization with one rVV strain (rVV-N25), which encoded nonstructural HCV proteins, suppressed serum inflammatory cytokine levels and alleviated the symptoms of pathological chronic hepatitis C within 7 days after injection. Furthermore, HCV protein levels in liver tissue also decreased in a CD4 and CD8 T-cell-dependent manner. Consistent with these results, we showed that rVV-N25 immunization induced a robust CD8 T-cell immune response that was specific to the HCV nonstructural protein 2. We also demonstrated that the onset of chronic hepatitis in CN2-29$^{(+/-)}$/MxCre$^{(+/-)}$ mice was mainly attributable to inflammatory cytokines, (tumor necrosis factor) TNF-α and (interleukin) IL-6. Thus, our generated mice model should be useful for further investigation of the immunological processes associated with persistent expression of HCV proteins because these mice had not developed immune tolerance to the HCV antigen. In addition, we propose that rVV-N25 could be developed as an effective therapeutic vaccine.

Citation: Sekiguchi S, Kimura K, Chiyo T, Ohtsuki T, Tobita Y, et al. (2012) Immunization with a Recombinant Vaccinia Virus That Encodes Nonstructural Proteins of the Hepatitis C Virus Suppresses Viral Protein Levels in Mouse Liver. PLoS ONE 7(12): e51656. doi:10.1371/journal.pone.0051656

Editor: Naglaa H. Shoukry, University of Montreal, Canada

Received March 13, 2012; **Accepted** November 5, 2012; **Published** December 17, 2012

Funding: This study was supported by grants from the Ministry of Education, Culture, Sports, Science, and Technology of Japan; the Program for Promotion of Fundamental Studies in Health Sciences of the Pharmaceuticals and Medical Devices Agency of Japan; and the Ministry of Health, Labor, and Welfare of Japan. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

* E-mail: kohara-mc@igakuken.or.jp

Introduction

Hepatitis C virus (HCV) is a major public health problem; approximately 170 million people are infected with HCV worldwide [1]. HCV causes persistent infections that can lead to chronic liver diseases such as chronic hepatitis, liver cirrhosis, and hepatocellular carcinoma (HCC) [2]. Antiviral drugs are not highly effective in individuals with a chronic infection; furthermore, an effective vaccine against HVC has not been developed. A convenient animal model of HCV infection will greatly facilitate the development of an effective HCV vaccine.

Transgenic mice that express HCV proteins have been generated to study HCV expression [3,4]; however, in each of these cases, the relevant transgenes is expressed during embryonic development; therefore, the transgenic mice become immunotolerant to the transgenic products, and consequently, the adult mice are not useful for investigations of the pathogenesis of chronic hepatitis C. To address this problem, we developed a system that can drive conditional expression of an HCV transgene; our system involves the Cre/*loxP* system and a recombinant adenovirus capable of expressing Cre recombinase [5,6]. Concerns have been expressed that an adenovirus and transient expression of HCV proteins could induce immune responses [5] and, therefore, obscure any evidence of the effect of the host immune responses on chronic liver pathology. Therefore, here, we used a Cre/*loxP* switching system to generate an immunocompetent mouse model