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**).

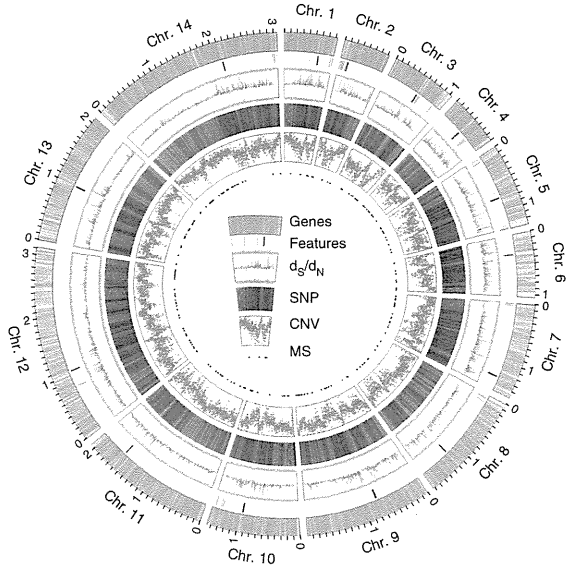

Figure 1 Architecture of the *P. cynomolgi* genome and associated genome-wide variation data. Data are shown for each of the 14 *P. cynomolgi* chromosomes. The six concentric rings, from outermost to innermost, represent (i) the location of 5,049 *P. cynomolgi* genes, excluding those on small contigs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

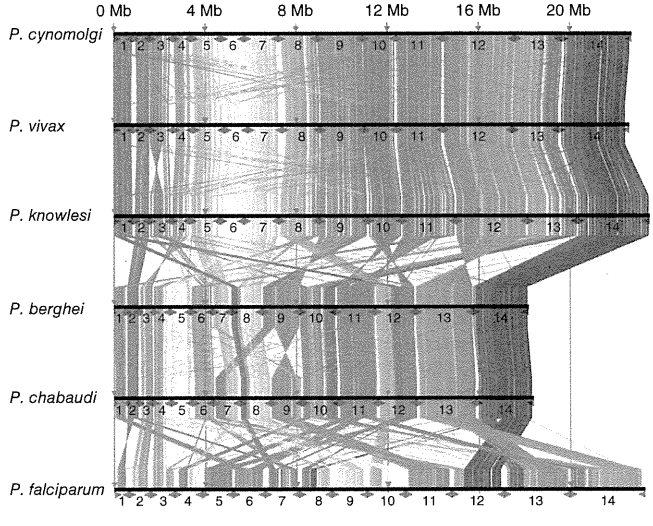

Figure 2 Genome synteny between six species of *Plasmodium* parasite. Protein-coding genes of *P. cynomolgi* are shown aligned with those of five other *Plasmodium* genomes: two species belonging to the monkey malaria clade, *P. vivax* and *P. knowlesi*; two species of rodent malaria, *P. berghei* and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

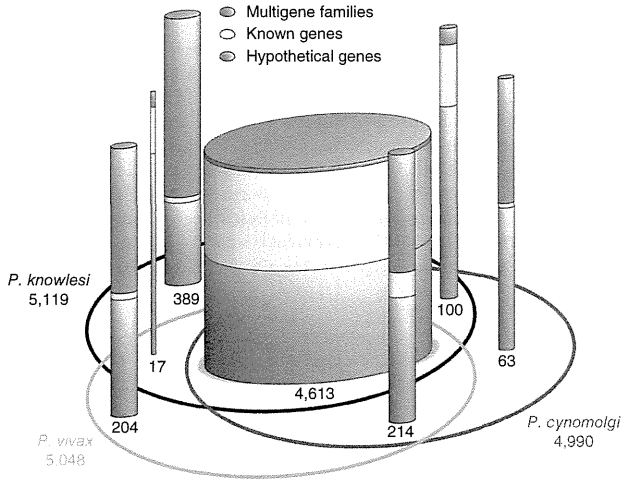

Figure 3 Comparison of the genes of *P. cynomolgi*, *P. vivax* and *P. knowlesi*. The Venn ellipses represent the three genomes, with the total number of genes assigned to the chromosomes indicated under the species name. Cylinders depict orthologous and non-orthologous genes between the three genomes,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

Hum. Reprod. Advance Access published December 6, 2011
Human Reproduction, Vol.0, No.0 pp. 1–8, 2011
doi:10.1093/humrep/der408

human reproduction ORIGINAL ARTICLE *Embryology*

Symmetrical division of mouse oocytes during meiotic maturation can lead to the development of twin embryos that amalgamate to form a chimeric hermaphrodite

Junko Otsuki[1,*], Yasushi Nagai[1], Alexander Lopata[2], Kazuyoshi Chiba[3], Lubna Yasmin[4], and Tadashi Sankai[4]

[1]Department of Obstetrics and Gynecology, Nagai Clinic, 607-1 Kamihikona Misato, Saitama 341-0004, Japan [2]Department of Obstetrics and Gynecology, University of Melbourne, Royal Women's Hospital, Parkville, Victoria, Australia [3]Department of Biological Sciences, Ochanomizu University, Tokyo, Japan [4]Tsukuba Primate Research Center, National Institute of Biomedical Innovation, Ibaraki, Japan

*Correspondence address. Tel: +81-48-959-1311; Fax: +81-48-959-1125; E-mail: otsuki.midori.junko@gmail.com

Submitted on May 31, 2011; resubmitted on October 19, 2011; accepted on November 1, 2011

BACKGROUND: Gentle compression of mouse oocytes during meiosis-I prevented the usual extrusion of a small polar body and resulted in the symmetrical division of the ooplasm into two cells of similar size within the zona pellucida. The purpose of our study was to determine whether such cells, equivalent to two small oocytes, were capable of embryonic development and would result in birth following transfer to the uterus.

METHODS: IVF of the 2-celled oocytes was performed and the twin intra-zonal embryos were observed. In each case, the two embryos that originated from fertilized cells with two pronuclei were observed to amalgamate and form a single morula and subsequent blastocyst that was transferred to the uterus of a recipient of a different mouse strain. FISH analysis was performed on sectioned paraffin-embedded tissue of the offspring.

RESULTS: In symmetrically divided oocytes each cell contained a metaphase II spindle. Both cells were fertilizable and cleaved to form twin embryos within the same zona pellucida. Most twin embryos amalgamated to form a single compacted morula, which progressed to hatched blastocysts that contained a single inner cell mass. In total, 104 of these blastocysts were transferred to 19 mice, two of which became pregnant, resulting in the birth of three offspring. FISH analysis showed that one newborn contained both XX and XY cells.

CONCLUSIONS: We found that two small oocytes fertilized within the same zona pellucida to form twin embryos that amalgamate to establish a single chimeric embryo. This may be one mechanism that leads to the formation of a chimeric hermaphrodite when an embryo containing XX cells mixes with its intra-zonal twin containing XY cells.

Key words: 2-celled oocytes / meiotic spindle / embryo amalgamation / chimera / hermaphrodite origin

Introduction

Oocytes aspirated from their follicles are occasionally found to contain two cells, of similar size, within their zona pellucida. This condition has been described as 'immediate cleavage' in some papers (Kaufman, 1973; Van de Leur and Zeilmaker, 1990). Some investigators have reported that in mouse oocytes, immediate cleavage can be induced by exposing tubal ova to high or low temperatures, to hypertonic or hypotonic solutions or to stimulation with an electric current, or by ether anesthesia (Komar, 1982). In the human, a presumed 'embryo' consisting of two cells, each containing two pronuclei, was found in an IVF laboratory and the authors suggested that this resulted from parthenogenetic activation of the oocyte leading to immediate cleavage, followed by separate fertilization of each cell (Van de Leur and Zeilmaker, 1990). However, the mechanism that induces immediate cleavage has not been clearly explained for more than two decades. It should be noted that parthenogenetic activation of maturing oocytes does not always induce their cleavage into two equal cells. In most cases, parthenogenetically activated mature oocytes extrude second polar bodies and only

a small proportion of oocytes (Kaufman, 1973) cleave immediately into two equal cells.

In the present study, we found that the gentle compression of mouse oocytes during meiotic maturation prevented the extrusion of a small polar body, but instead resulted in the formation of two cells of similar size within the zona pellucida. One of these cells was considered to be an extruded large first polar body. In such 'double' oocytes each cell attained meiotic metaphase II (MII). Our aim was to determine whether each cell could be fertilized by a separate sperm, and whether the two intra-zonal embryos retained developmental competence. We therefore transferred such embryos to the uterus of recipient mice, to determine whether they were capable of developing to birth.

Materials and Methods

Mouse oocytes

Mice were exposed to a 12 h light/dark cycle with food and water provided ad libitum. Female BDF1 mice, aged 7–12 weeks and weighing 27–34 g, were injected with 7.5 IU of equine chorionic gonadotrophin (Teikoku-Zoki Pharmaceuticals, Tokyo, Japan) followed by an injection of 5 IU of hCG (Mochida Pharmaceutical, Japan). Oocytes undergoing meiotic maturation were retrieved from the ovaries 7–8 h after hCG injection. The cumulus cells were removed by gentle pipetting in HEPES-buffered human tubal fluid (HTF; *In Vitro* Care, Ferderick, MD, USA) medium containing 1 mg/ml human serum albumin (HSA; *In Vitro* Care). The cumulus-free oocytes containing a meiotic-1 (M1) spindle were washed three times in HTF medium containing 1 mg/ml HSA and they were then cultured at 37°C in a humidified atmosphere of 5% CO_2, 5% O_2 and 90% N_2. All experimental mice in this study were cared for using procedures approved by the Animal Care Committees of both the Nagai Clinic and the National Institute of Biomedical Innovation, Tsukuba Primate Research Center. Both institutions approved all animal handling and experimental procedures used in the present studies.

Time-lapse observations

Using a CO_2 chamber (SKHC-303, Sankei, Japan) equipped with a DIC inverted microscope (Nikon ECLIPSE TE2000-U), time-lapse recordings were performed on the cultured mouse oocytes. The inverted microscope was equipped with a digital camera (Nikon D200) connected to a computer and a display using Nikon Capture (Nikon, Japan) software. Digital images of the oocytes in culture were recorded every 1–2 min with an exposure time of 2 s. The chamber was maintained at 37°C with a humidified atmosphere of 5% CO_2, 5% O_2 and 90% N_2.

Compression of the ooplasm near the meiotic spindle

During the meiotic maturation in culture, mouse oocytes were pressed down to a thickness of 20–30 μm between a cover glass and the bottom of a glass dish using autoclaved silicon grease as support between them. This procedure was applied to compress the cortical cytoplasm when the M1 spindle was near the surface of the oocyte in order to impede the spindle's movement and promote symmetrical division of the oocyte.

Disruption of the zona pellucida by laser for IVF

After an oocyte divided into two cells of similar size, fertilization of each of the cells was achieved by *in vitro* insemination after partial disruption of the oocyte's zona pellucida using laser (OCTAX Laser Shot System for ART, from OCTAX Microscience). The use of the laser system for making an opening in the zona pellucida has been described in detail (Germond *et al.*, 1995). Briefly, a 1.48-μm diode laser aiming beam and a collimated 1.48 μm continuous-wave laser beam are passed into an inverted microscope (Olympus IX-70; Tokyo, Japan), redirected by several mirrors and focused on the microscopic field. The power routinely available at the image plane of the objective is 47 mW, corresponding to a maximum power density of 94 kW/cm^2. During the zona opening procedure, oocytes were suspended in 10 μl droplets of HTF medium containing 1 mg/ml HSA under mineral oil in a 35-mm culture dish.

IVF and embryo culture

Mouse spermatozoa were collected from the cauda epididymis and suspended in HTF medium that contained 1 mg/ml HSA. After pre-incubation for 1 h, an aliquot of sperm suspension was added to the medium containing oocytes to provide a final sperm concentration of 200 sperm/μl. After 3–4 h of insemination, oocytes were examined to confirm the presence of extruded second polar bodies. These fertilized oocytes were transferred into LGGG medium (Life Grobal, Japan) that contained 0.1 mg/ml rHSA (rHA; Vitrolife, Sweden). At 16 h after insemination, the oocytes were examined for the presence of two pronuclei.

Embryo culture experiments were performed using double small oocytes, each containing two pronuclei. Embryos arising from polyspermic fertilization were excluded. In addition, a study was performed to compare embryo development following fertilization of the two small oocytes in the one zona pellucida, with embryo development originating from the fertilization of a single large oocyte. The formation of blastocysts was the endpoint of this study, comparing embryo development from 104 double small oocytes with 106 single large oocytes. Statistical analysis was performed using Welch's *t*-test employing the statistical software StatMateIII (ATMS, Inc., Tokyo, Japan). A difference was considered to be significant when its *P*-value was <0.05.

Embryo transfer

The recipients for embryo transfer were pseudo-pregnant day 3 mice of an ICR strain. These recipients were selected after detecting the presence of copulatory plugs following mating with vasectomized ICR male mice. Day 4 cultured BDF1 mouse blastocysts were transferred to the recipients using a glass transfer pipette.

FISH analysis

Dual-color FISH analysis was performed on sections of paraffin-embedded tissues obtained from a newborn mouse, presumed to be a chimera. Briefly, 5 μm sections were washed in phosphate-buffered saline for 5 min, digested in pepsin solution (0.02% in 0.1 N HCl) at 37°C for 5 min and then dehydrated. A digoxigenin-labeled mouse X probe and a biotin-labeled mouse Y probe were applied to the pretreated sections, covered with cover slips and simultaneously denatured at 90°C for 10 min. Hybridization was carried out at 37°C overnight. Sections were then washed with 50% formamide/2 × saline sodium citrate (SSC) at 37°C for 20 min and 1 × SSC for 15 min at room temperature and blocking solution (5% skim milk, 0.1% non-idet P-40, 0.1 M phosphate buffer, pH7.5) was applied at 37°C for 30 min. The mouse X and Y probe signals were detected with Cy5-labeled anti-digoxigenin and avidin-Cy3, respectively. The sections were treated with antibodies at 37°C for

30 min, washed three times with 0.1% non-idet P-40/2 × SSC, counter stained with 4,6-diamidino-2-phenylindole and mounted on slides. The same procedures were applied to control tissues obtained from a normal male. The FISH images were captured with a CW4000 FISH system.

Results

Spindle behavior and oocyte division when the ooplasm was compressed

Meiotic maturation of mouse oocytes usually results in an asymmetric cell division. Polar body extrusion occurs after the spindle migrates to the oocyte's cortical region along its long axis and becomes positioned perpendicular to the surface. When gentle compression was applied to the oocyte, the MI spindle adopted a position in which its axis was observed to be parallel to the surface. A cleavage furrow was observed to form adjacent to the spindle and this surface indentation appeared to push its way toward the spindle's midzone. Then, as the entire midzone structure appeared to shrink, the spindle together with the cleavage furrow moved across the ooplasm and divided the oocyte into two approximately equal cells. Thus, instead of extruding a typical small polar body, the cytoplasm was bisected into two similar oval cells. Subsequently, a structure resembling a meiotic MII spindle formed in each of the daughter cells that now resembled two small oocytes within the same zona pellucida (Fig. 1).

In vitro embryonic development after fertilization of the 2-celled oocytes

After partial disruption of their zona with laser, *in vitro* insemination of the 2-celled oocytes was performed to evaluate the developmental potential of each daughter cell. When such 2-celled MII mouse oocytes were fertilized, in the presence of small openings in their zona pellucida, a typical second polar body was extruded from each cell and early male and female pronuclei were seen in each cell by 6 h after insemination (Fig. 2A). Division of each cell initially produced distinct twin 2-celled and then twin 4-celled embryos (Fig. 2B and C). However, beyond these stages, we could not determine whether embryos remained separated or whether mixing of their blastomeres was occurring. Moreover, compaction at about the 8-celled stage made it difficult to distinguish separate embryos (Fig. 2C). In all cases, amalgamation of the embryos produced a single morula (Fig. 2D and E). In these studies, we observed more than 100 blastocysts that originated from the amalgamated twin embryos, and each blastocyst was found to contain a single inner cell mass (Fig. 2F). Such blastocysts were transferred to recipient mice.

Comparison of embryo development after fertilization of small double oocytes and single large oocytes

After IVF, syngamy usually occurred in each small oocyte by 16–18 h after insemination. In these small double oocytes, the first cleavage division was observed in 72.1% of the zygotes by 24 h after insemination (Fig. 3). As a rule one cell divided before the other, and such asynchrony was seen regularly. Each fertilized cell divided into two regular blastomeres to form twin 2-celled embryos (Fig. 2B). The separate twin embryos, each containing two blastomeres, were quite distinct inside their zona. By 48 h after insemination, two definitive 4-cell embryos were observed in 44.1% of the fertilized double oocytes and in 88.5% of single normal oocytes ($P < 0.001$). At earlier times, 3-celled embryos were occasionally noticed in both groups. More advanced twin embryo stages, including approximately 8-cell embryos, could not be clearly distinguished, possibly due to the onset of compaction. However, by 60 h after insemination, it was clear that a significantly lower percentage of fertilized double oocytes attained this stage (Fig. 3). Similarly, a significantly lower percentage of fertilized double oocytes reached the morula and blastocyst stages by 72 and 96 h after insemination.

The birth of mouse pups after transfer of amalgamated embryos and FISH analysis

In total, 104 chimeric BDF1 mouse blastocysts were transferred to the uterus of 19 ICR mouse recipients. Two out of 19 mice became pregnant and three pups were born (Fig. 4). Two of these were partly eaten by the mother and one died on Day 5 after birth.

FISH analysis on the head and neck tissues of the partly eaten offspring showed that it contained both XX and XY cells (Fig. 5A). This indicated that an XX/XY chimera was established by the amalgamation of the twin embryos, inside the same zona pellucida, as a result of our experimental procedures.

Discussion

The human oocyte in the dominant follicle is arrested at the germinal vesicle stage until it is stimulated to resume meiotic maturation. After germinal vesicle breakdown, the first meiotic spindle is formed and homologous chromosomes become aligned along its midzone. The subsequent first meiotic division is asymmetrical and produces two cells: the secondary oocyte and the first polar body. Meiosis then remains arrested until the oocyte is fertilized or activated by another mechanism. Following activation meiosis resumes and the oocyte undergoes a second asymmetric division, again producing two cells: a zygote and the second polar body. Both first and second polar bodies are small and usually degenerate shortly after their extrusion. On rare occasions, the first meiotic division may be symmetrical and produces a large polar body. This large polar body can be regarded as a second intra-zonal oocyte, which can be fertilized by a second sperm. In the present study we have demonstrated that a controlled compression of mouse oocytes during meiosis-1, resulted in the formation of two fertilizable oocytes inside the same zona pellucida. Symmetric division, in which the oocyte and the released polar body are similar in size, can probably also occur during the extrusion of the second polar body.

In mouse oocytes, during meiosis-1, the established spindle migrates along its long axis, reaches the cortical region in a vertical position and remains in this position during its asymmetric meiotic division (Azoury *et al.*, 2008; Brunet and Verlhac, 2011; Schuh and Ellenberg, 2008) to produce a small first polar body. The question that needs to be answered from the present study is how, following gentle compression, oocytes undergo symmetrical meiotic division to yield two cells that are approximately equal in size. During this symmetrical division, the long axis of the meiotic

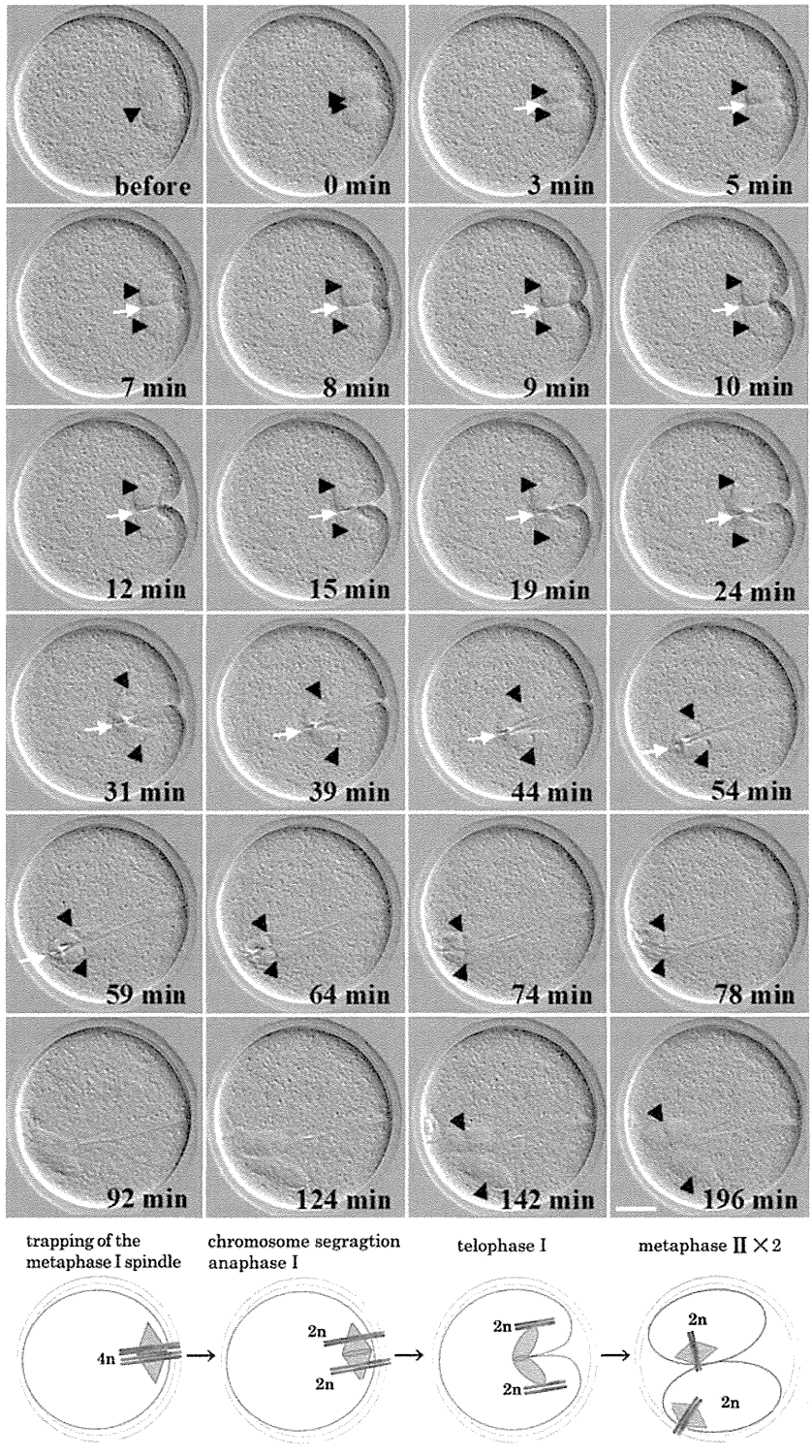

Figure 1 The time-course of symmetric oocyte division when the cortical ooplasm near the MI spindle was compressed. The sequence of oocyte changes, obtained from time-lapse photography, shows the time-course of a symmetrical meiotic division when the MI spindle was experimentally compressed in the cortical region of the oocyte. A cleavage furrow formed adjacent to the spindle and appeared to move toward the midzone (8–78 min). Subsequently, the spindle and cleavage furrow traversed the ooplasm, so that the oocyte became divided into two approximately equal cells (92–124 min), instead of extruding a typical small polar body. After the two cells were established, a MII spindle formed in each cell (142–196 min). An arrow shows the midzone. Scale bars = 20 μm.

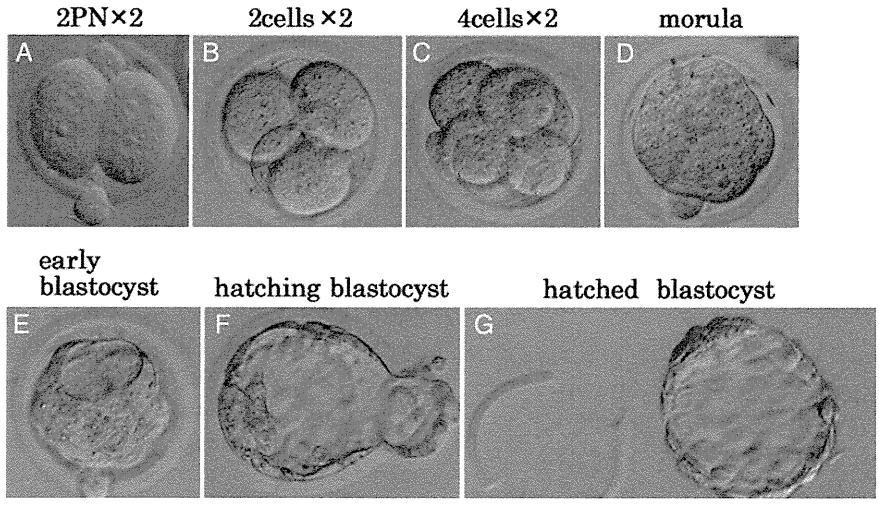

Figure 2 (**A**) Each of the two small oocytes was fertilized, revealed by the presence of female and male pronuclei and an extruded second polar body from each of the cells. (**B**) Each fertilized cell divided to form twin 2-celled embryos. (**C**) Two 4-celled embryos formed twin embryos within the same zona pellucida. (**D**) The twin embryos amalgamated to form a single compacted morula. (**E**) A single cavitating morula has formed. (**F**) A hatching blastocyst contained a single inner cell mass. (**G**) The chimeric blastocys is fully hatched.

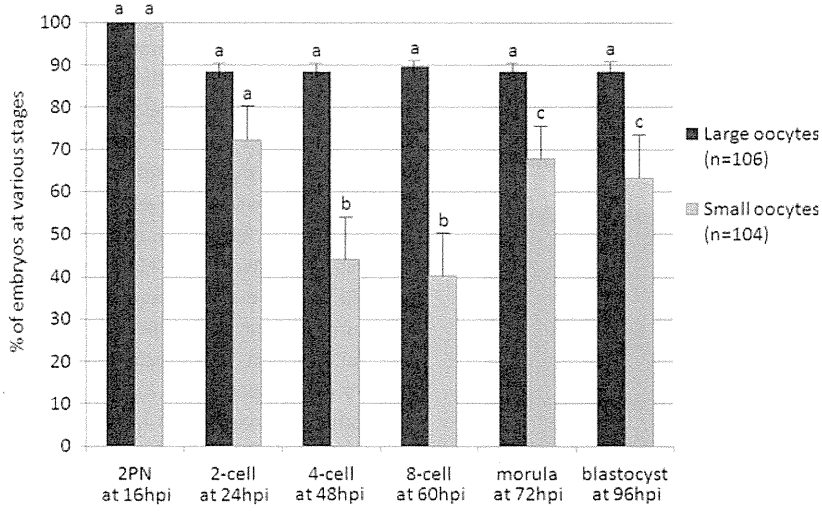

Figure 3 In the histogram, the timing of embryo development is compared after fertilization of small double oocytes and single normal large oocytes. At 24 h post insemination (hpi), 72.1% of small oocytes and 88.5% large oocytes formed 2-cell embryos. The formation of 4-cell embryos at 48 hpi in the two groups was 44.1 and 88.5%, respectively. The establishment of 8-cell embryos was more difficult to judge because of possible cell mixing and onset of compaction, but it was judged that 40.4% of the small oocytes and 89.6% of the large oocytes reached this stage. At 72 hpi, the establishment of a single morula was observed in 67.8% of the small oocyte group and 88.5% of the large oocyte group. By 96 hpi, blastocysts were observed in 63.3% of the small oocyte group, compared with 88.5% in the large oocyte group. The differences in embryo development were statistically significant at each time interval. a,a; NS, a,b; $P < 0.001$; a,c; $P < 0.05$.

spindle has been observed to be parallel to the surface of the oocyte. It would appear that the compression of meiotic oocytes may induce the spindle to take up a position that is parallel, instead of perpendicular, to the plasma membrane. It has been suggested that the compression of the oocytes may delay the onset of anaphase-I and this could lead to the spindle adopting a position parallel to the surface (Melina Schuh, personal communication, 2011). If a cleavage furrow begins while the spindle is parallel to the surface, symmetrical oocyte division will occur as illustrated in Fig. 1. However, the process that induces the spindle to become

Figure 4 A newborn chimera on Day 4 after its birth and an inset figure that shows the pup with its mother.

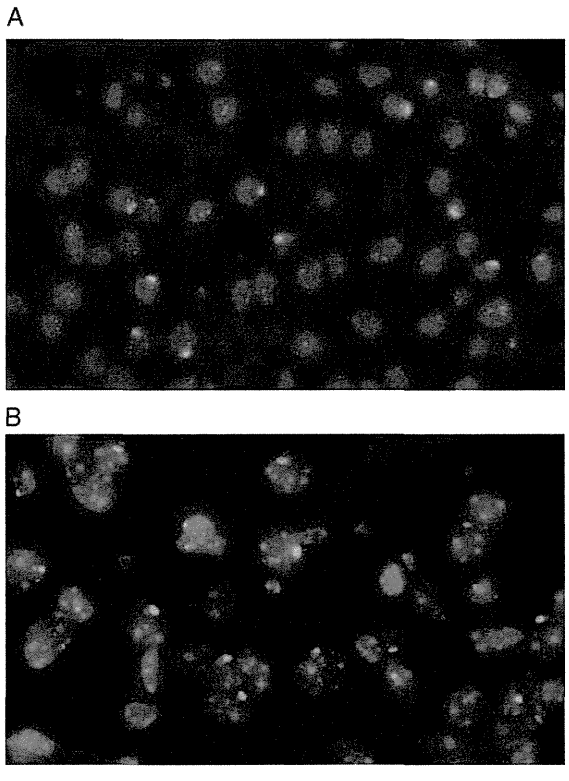

Figure 5 (**A**) Metaphase FISH analysis of the head and neck tissues of the chimera newborn mouse showing the X chromosomes stained in red and the Y chromosomes in green. (**B**) A control FISH analysis of head tissues of a normal male mouse showing the X chromosomes stained in red and the Y chromosomes in green.

parallel to the surface (Zhu *et al.*, 2003) and bring about symmetric oocyte division is not understood.

Our findings prompted us to compare 'immediate cleavage', described in human oocytes (Giltay *et al.*, 1998; Souter *et al.*, 2007) with the experimental model developed in the present studies. It has been proposed that 'immediate cleavage' involves the parthenogenetic activation of a mature oocyte, followed by symmetric division of the oocyte. One of the cells inside the zona would represent an activated small oocyte, while the other cell would represent a large second polar body. The female chromatin in such cells would form early pronuclei. If such cells can be fertilized, after the cortical granule reaction, the penetrated sperm would need to undergo decondensation to establish pronuclei that can undergo syngamy and further development. Currently there is no experimental information to support these proposals. The findings obtained in the present studies provide insight into the possibility that 'immediate cleavage' does not require parthenogenetic activation of an oocyte, but instead the 'cleavage' occurs in the course of meiotic maturation that involves symmetrical division of the oocyte. The semi-identical twins observed by Souter *et al.* (2007) could arise from the separate fertilization of the double oocytes. In such cases, the fertilization of each cell by separate sperm could produce non-identical twins when the inner cell mass of an amalgamated chimeric embryo becomes separated into two embryonic masses.

In this study, when an MII spindle formed in each daughter cell, each cell would have contained a different DNA pattern due to chromosome crossover during prophase-I of meiosis. Chromosomal crossover is an exchange of genetic material between homologous maternal and parternal chromosomes. At the first meiotic division, the paired homologs separate, but the chromatids remain attached to one another through their centromeres. Consequently, the secondary oocyte and the first polar body receive different versions of each homologous chromosome. In the present study, following symmetric division, each cell contained an MII spindle near the surface of each small oocyte. We found that each of these small oocytes could be fertilized and retained developmental competence. When a separate sperm fertilized each of the intra-zonal cells, two embryos were formed within the same zona pellucida, and these twin embryos eventually amalgamated to develop into a single chimeric blastocyst. In the present study, fertilization and development of the two small oocytes within the same zona pellucida was found to be one mechanism that could lead to the formation of a genetic intersex. It is likely that this occurred in our study when one embryo, developed after the entry of an X sperm, amalgamated with its twin embryo, formed by the entry of a Y sperm (Fig. 6). Our results suggest that a low incidence of other embryos of this kind could develop to term to give rise to chimeric organisms composed of a mixture of XX and XY cells.

A true chimera is an organism containing organs or body components consisting of two or more tissues containing cells of different genetic composition. Blood chimeras are completely different because they arise from the mingling of blood cells during multiple pregnancies conceived naturally or following IVF/ICSI treatment cycles (Miura and Niikawa, 2005; Walker *et al.*, 2007). Such blood chimeras are derived from two or more separately implanted embryos that originated from two or more different oocytes. In contrast to blood chimeras, the organisms that we produced in the present study are chimeras originating from a single oocyte that divided into two fertilizable cells within the same zona pellucida. We have demonstrated that each of these cells produced an embryo that amalgamated into a single chimeric embryo but only a small proportion of such chimeras were capable of implanting and of developing to term.

A number of factors may account for the low incidence of implantation and development to term of the amalgamated embryos following

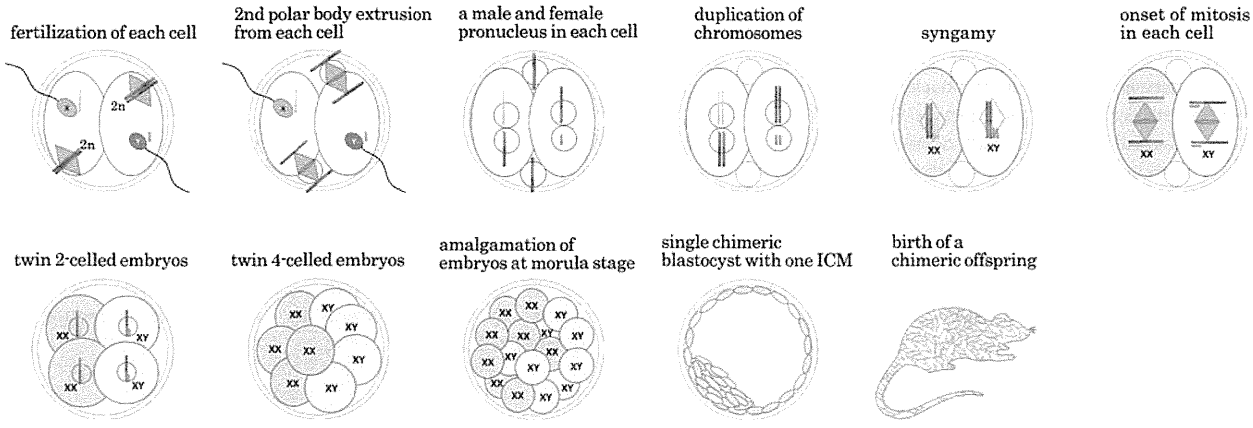

Figure 6 A diagram of fertilization and development of the two small oocytes showing initially the formation of twin embryos that eventually amalgamate within the same zona pellucida. If one embryo contains XX cells and the other XY cells, then this diagram depicts a proposed mechanism for the formation of a chimeric hermaphrodite organism.

their transfer to recipients. All of the oocytes used in the present studies were matured *in vitro*. Moreover, during their meiotic maturation in culture, the oocytes may have been stressed by the compression that was applied to induce symmetrical division. Also, the resulting two small oocytes acquired a markedly different nucleocytoplasmic ratio, compared with single large oocytes. It has been reported that such differences in the cytoplasmic to nuclear proportions could influence the cell cycle, cleavage rates, the onset of DNA and RNA synthesis as well as differentiation at more advanced embryonic stages (Satoh, 1985; Kominami and Takata, 2003).

In the present study, the FISH results indicated that the chimera contained both XX and XY cells in the newborn mouse head and neck tissues. We showed that such chimera could arise from the symmetrical division of oocytes during meiosis-1, followed by the fertilization of each cell by separate sperm inside the same zona pellucida. Although the size of each cell was about half that of a normal oocyte, each developed as a single embryo until mixing of their blastomeres followed by amalgamation of the twin embryos produced a single chimera. We propose that this mechanism could lead to the formation of a true hermaphrodite containing both male and female tissues.

Authors' roles

J.O. designed the main study and chimera embryo development and wrote the main part of the paper. She substantially contributed to the conception and design, acquisition of data, and the analysis and interpretation of data; she drafted the article and revised it critically for important intellectual content, and contributed to the final approval of the version to be published. Y.N. performed FISH analysis and assisted in plans for the experiments. He substantially contributed to analysis and interpretation of data. He contributed to drafting and revision of the paper for important intellectual content, and to the final approval of the version to be published. A.L. wrote parts of the paper and helped to interpret the results. He substantially contributed to analysis and interpretation of data. He contributed to drafting and revision of the paper for important intellectual content, and to the final approval of the version to be published. K.C. provided advice on experimental protocols and procedures. He substantially contributed to analysis and interpretation of data. He contributed to drafting and revision of the paper for important intellectual content, to the final approval of the version to be published. L.Y. assisted in the embryo transfer. She substantially contributed to analysis and interpretation of data. She contributed to drafting and revision of the paper for important intellectual content, and to the final approval of the version to be published. T.S. performed the embryo transfers, assisted in the experimental design and provided advice on experimental protocols. He substantially contributed to analysis and interpretation of data. He contributed to drafting and revision of the paper for important intellectual content, and to the final approval of the version to be published.

Funding

The funding for the project was provided by the Nagai clinic.

Conflict of interest

None declared.

References

Azoury J, Lee KW, Georget V, Rassinier P, Verlhac MH. Spindle positioning in mouse oocytes relies on a dynamic meshwork of actin filaments. *Current Biol* 2008;**18**:1514–1519.

Burnet S, Verlhac MH. Positioning to get out of meiosis: the asymmetry of division. *Hum Reprod Update* 2011;**17**:68–75.

Germond M, Nocera D, Senn A, Rink K, Delacrétaz G, Fakan S. Microdissection of mouse and human zona pellucida using a 1.48-microns diode laser beam: efficacy and safety of the procedure. *Fertil Steril* 1995;**64**:604–611.

Giltay JC, Brunt T, Beemer FA, Wit JM, van Amstel HKP, Pearson PL, Wijmenga C. Polymorphic detection of a parthenogenetic maternal

and double paternal contribution to a 46,XX/46,XY hermaphrodite. *Am J Hum Genet* 1998;**62**:937–940.

Kaufman MH. Parthenogenesis in the mouse. *Nature* 1973;**242**:475–476.

Komar A. Fertilization of parthenogenetically activate mouse eggs. I. Behaviour of sperm nuclei in the cytoplasm of parthenogenetically activated eggs. *Exp Cell Res* 1982;**139**:361–367.

Kominami T, Takata H. Timing of early developmental events in embryos of a tropical sea urchin. *Echinometra mathaei*. *Zool Sci* 2003;**20**:617–626.

Miura K, Niikawa N. Do monochorionic dizygotic twins increase after pregnancy by assisted reproductive technology? *J Hum Genet* 2005; **50**:1–6.

Satoh N. Recent advances in our understanding of the temporal control of early embryonic development in amphibians. *J Embryol Exp Morph* 1985; **89**:257–270.

Schuh M, Ellenberg J. A new model for asymmetric spindle positioning in mouse oocytes. *Current Biol* 2008;**18**:1986–1992.

Souter VL, Parisi MA, Nyholt DR, Kapur RP, Henders AK, Opheim KE, Gunther DF, Mitchell ME, Glass IA, Montgomery GW. A case of true hermaphroditism reveals an unusual mechanism of twinning. *Hum Genet* 2007;**121**:179–185.

Van de Leur SJ, Zeilmaker GH. Double fertilization *in vitro* and the origin of human chimerism. *Fertil Steril* 1990;**54**:539–540.

Walker SP, Meagher S, White SM. Confined blood chimerism in monochorionic dizygous (MCDZ) twins. *Prenat Diagn* 2007; **27**:369–372.

Zhu Z-Y, Chen D-Y, Li J-S, Lian L, Lei L, Han Z-M, Sun Q-y. Rotation of meiotic spindle is controlled by microfilaments in mouse oocytes. *Reprod Biol* 2003;**68**:943–946.

Comparative Medicine
Copyright 2012
by the American Association for Laboratory Animal Science

Vol 62, No 1
February 2012
Pages 1–4

Case Report

Triplet Pregnancy in a Cynomolgus Monkey (*Macaca fascicularis*) after Double Embryo Transfer

Juri Morichika,[1] Chizuru Iwatani,[1] Hideaki Tsuchiya,[1] Shinichiro Nakamura,[1] Tadashi Sankai,[1,2] and Ryuzo Torii[1,*]

At our research center, cynomolgus monkeys (*Macaca fascicularis*) are bred by mating or intracytoplasmic sperm injection (ICSI) and embryo transfer. We typically transfer 2 embryos, because the pregnancy rate is better than that for single embryo transfer. In the case we present here, 2 embryos that had been frozen and thawed after ICSI were transplanted into a recipient female macaque, and a multiple pregnancy (3 fetuses) was confirmed. All 3 fetuses were miscarried between days 81 and 85 of pregnancy. One fetus, which was wrapped in the amnion, was expelled along with its own placenta and one other. Because the other placenta had 2 umbilical arteries, 2 fetuses may have shared it. Therefore, we believe this pregnancy was a case of triplets, including a set of twins from an embryo that divided after transfer.

Abbreviation: ICSI, intracytoplasmic sperm injection.

The birth of twins to rhesus[2] and Japanese macaque[7] monkeys has been reported, but the probability of twins is extremely low (0.027% to 0.21%). At our institution, we carry out indoor artificial breeding of cynomolgus monkeys (*Macaca fascicularis*) by using a variety of techniques, including ovarian stimulation, oocyte collection, intracytoplasmic sperm injection (ICSI), and embryo transfer. Double embryo transfer typically is performed, because the pregnancy rate is better than that for single embryo transfer.

Here we describe a case of triplets generated from double embryo transfer. We believe that 1 of the 2 transplanted embryos divided after transfer.

Case Report

Experimental procedures were approved by the animal care and use committee of Shiga University of Medical Science (Shiga, Japan). Oocytes were collected from one sexually mature female cynomolgus monkey (age, 6 y; weight, 2.3 kg), and another (age, 5 y; weight, 2.5 kg) was the recipient of embryo transfer. The animal rooms were maintained at 25 ± 2 °C, with a relative humidity of 50% ± 5% and 12:12-h light:dark cycle (lights on, 0800 to 2000). Cynomolgus monkeys were housed individually (cage dimensions, 500 mm × 800 mm × 800 mm). In the morning, each animal was fed commercial pellet monkey chow (20 g per kg of body weight; CMK1, CLEA Japan, Tokyo), supplemented with 20 to 50 g of sweet potato in the afternoon. Water was given ad libitum by an automatic supplier.

Received: 09 Aug 2011. Revision requested: No revisions requested. Accepted: 22 Sep 2011.
[1]Research Center for Animal Life Science, Shiga University of Medical Science, Shiga, Japan, and [2]Tsukuba Primate Research Center, National Institute of Biomedical Innovation, Tsukuba, Ibaraki, Japan.
**Corresponding author. Email: torii@belle.shiga-med.ac.jp*

Oocytes were collected by laparoscopy (LA6500, Machida Endoscope, Tokyo, Japan) from adult female monkeys that had been treated with gonadotropin.[6] Insemination was performed by ICSI with oocytes and thawed sperm, and embryos were cultured in vitro.[8] Three days after fertilization, 12-cell and 8-cell embryos were vitrified (Cryotop and Vitrification Kit; Kitazato BioPharma, Shizuoka, Japan).[10] On day 547 after freezing, the embryos were thawed, and the 8-cell and 12-cell embryos were cultured for 16 h. The 8-cell embryo was confirmed to have a normal form (that is, showing neither death nor degeneration of the blastomere), and the 12-cell embryo had developed to the 16- to 32-cell stage (Figure 1). Both embryos were transplanted by laparoscopy through the fimbria into the oviduct of the recipient female macaque[8,9] at 0 to 4 d after ovulation. Pregnancy was determined by ultrasonography (SSD-620, Aloka, Tokyo, Japan) 37 d after embryo transfer. The presence of at least 2 heartbeats confirmed a multiple pregnancy (Figure 2).

At our research center, the number of days (mean ± 1 SD; $n = 14$) from ICSI to birth is 157 ± 8. The first miscarriage (fetus A; Figure 3) in the macaque we present here occurred on day 81 after ICSI (not counting the freezing time). Ultrasonography after miscarriage confirmed the presence of a heartbeat. Two days later (day 83), a second miscarriage (fetus B; Figure 3) occurred, with the third miscarriage (fetus C) on day 85. Fetus C was expelled with its placenta. Another placenta was expelled on the same day. Because this second placenta had 2 umbilical arteries (Figure 3), we believe that fetuses A and B had shared it, suggesting that they were monozygotic twins; fetus C, which had a separate placenta, likely originated from the second embryo. However, because we did not perform DNA analysis, we cannot exclude the possibility that the 3 fetuses were monozygotic triplets.

At our research center, double embryo transfer of cynomolgus macaques is performed to improve the pregnancy rate, and twin pregnancies do occur. Typically, however, one or both of the twins

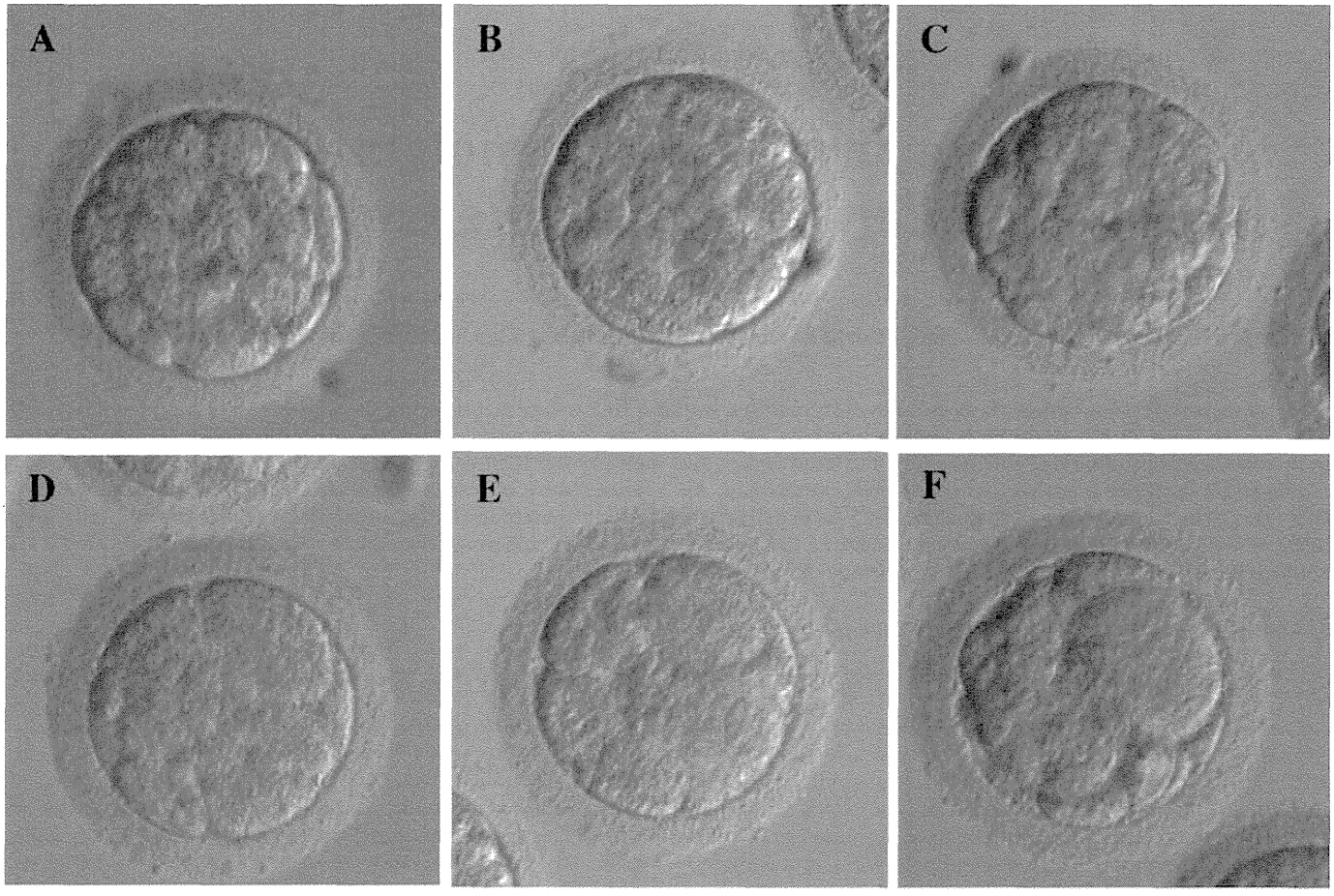

Figure 1. Cynomolgus monkey embryos (A, D) before freezing, (B, E) after thawing, and (C, F) at transfer. (A through C) Embryo that was frozen–thawed at 12 cells and transferred to the recipient at 16 to 32 cells after 16 h of culture. (D through F) Embryo frozen–thawed at 8 cells and transferred to the recipient at 8 cells after 16 h of culture. Bar, 50 μm.

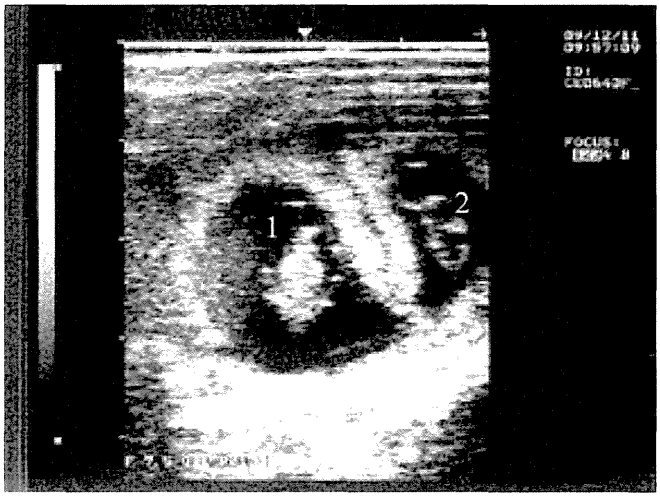

Figure 2. Ultrasonographic image of the pregnancy on day 37.

are miscarried. In addition, multiple pregnancies from single embryo transfer had not occurred previously at our institution, nor had triplets resulted from double embryo transfer. In the current case, natural separation after the 8-cell stage led to monozygotic multiple fetuses; to the best of our knowledge, we are the first to report such a result after assisted reproductive methods in cynomolgus monkeys. Embryo splitting similarly led to a twin pregnancy in a rhesus monkey, but one twin was miscarried.[5] Although the 2 heartbeats confirmed on day 37 led us to diagnose a twin pregnancy, we now believe that it actually was a triple pregnancy. We were unable to perform chromosomal and DNA analysis on any of the 3 fetuses described herein.

Monozygotic multiples and dizygotic quadruplets and quintuplets are well known to occur in humans.[3,4] The incidence of human monozygotic twins after assisted reproductive technologies varies depending on the number of days in embryo culture and the fertilization method, such as in vitro fertilization and ICSI. Ovary stimulation and expanded blastocyst transfer are other possible causes of monozygotic twins in humans.[1]

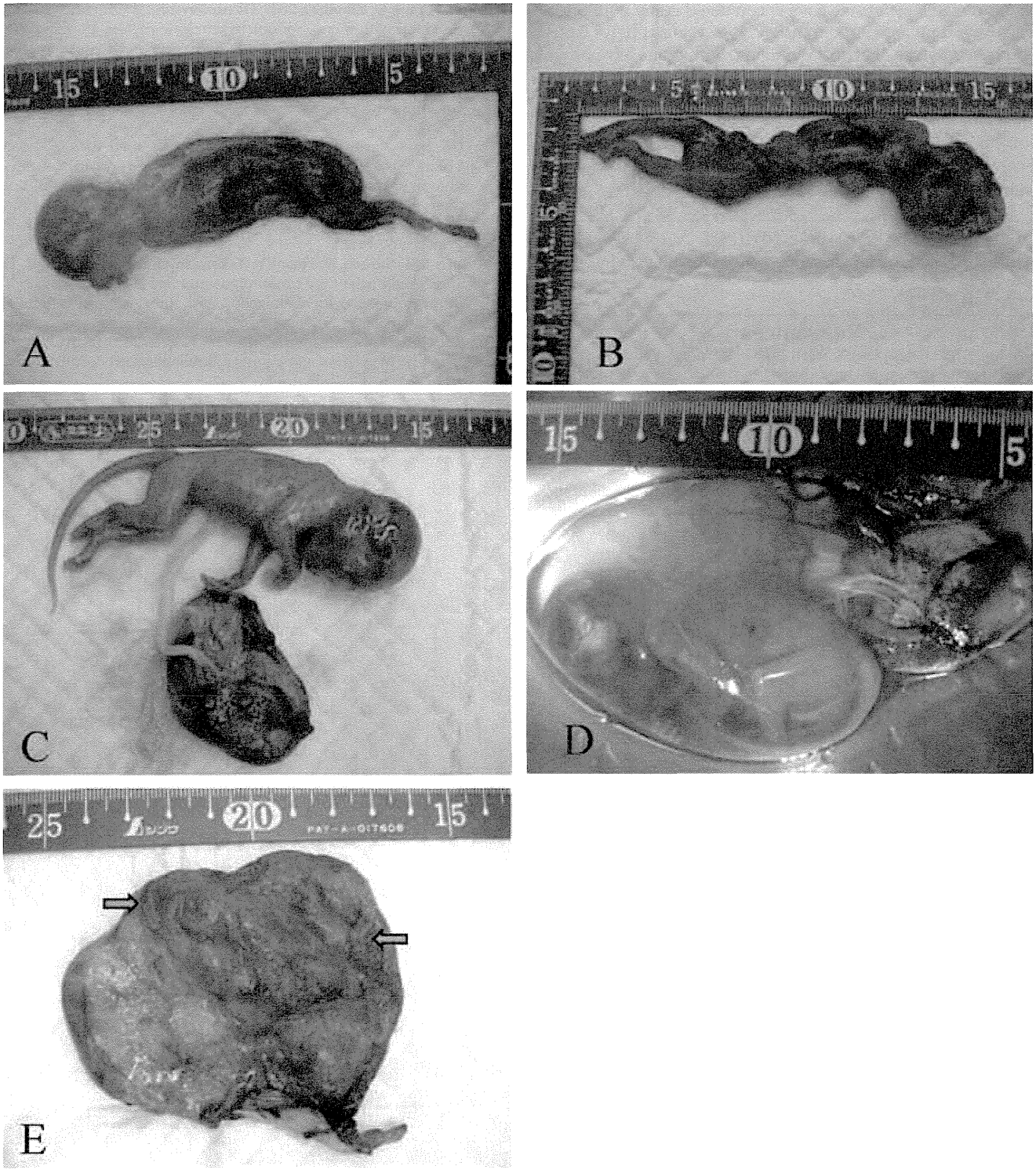

Figure 3. Dissection view of 3 miscarried fetuses. (A) Fetus (fetus A; male; length, 14 cm; weight, 35.0 g) miscarried on day 81 of pregnancy (25 January 2010). The tips of both feet were lost. The positions and shapes of the internal organs were normal, and the complexion was poor. (B) Fetus (fetus B; male; length, 15 cm; weight, 30.1 g) miscarried on day 83 of pregnancy (27 January 2010). Autolysis was pronounced, and portions of various internal organs had been resorbed. (C) Fetus (fetus C; male; length, 15 cm; weight, 51.1 g) miscarried on day 85 of pregnancy (29 January 2010). Fetus C was expelled while still wrapped in the amnion. The positions and shapes of the internal organs were normal, and it had its own placenta (weight, 19.2 g). Development of the internal organs was more advanced than in fetus A. (D) Appearance of fetus C when it was found. (E) Placenta (weight, 37.4 g) that was expelled with that of fetus C. We believe that fetuses A and B shared this placenta, because it had 2 umbilical arteries (arrows).