

Restriction-Modification Systems as Mobile Genetic Elements in the Evolution of an Intracellular Symbiont

Hao Zheng,^{†,1} Carsten Dietrich,^{†,1} Yuichi Hongoh,² and Andreas Brune^{*,1}

¹Department of Biogeochemistry, Max Planck Institute for Terrestrial Microbiology, Marburg, Germany

²Department of Biological Sciences, Tokyo Institute of Technology, Tokyo, Japan

[†]These authors contributed equally to this work.

*Corresponding author: E-mail: brune@mpi-marburg.mpg.de.

Associate editor: Eduardo Rocha

Abstract

Long-term vertical transmission of intracellular bacteria causes massive genomic erosion and results in extremely small genomes, particularly in ancient symbionts. Genome reduction is typically preceded by the accumulation of pseudogenes and proliferation of mobile genetic elements, which are responsible for chromosome rearrangements during the initial stage of endosymbiosis. We compared the genomes of an endosymbiont of termite gut flagellates, “*Candidatus Endomicrobium trichonymphae*,” and its free-living relative *Endomicrobium proavitum* and discovered many remnants of restriction-modification (R-M) systems that are consistently associated with genome rearrangements in the endosymbiont genome. The rearrangements include apparent insertions, transpositions, and the duplication of a genomic region; there was no evidence of transposon structures or other mobile elements. Our study reveals a so far unrecognized mechanism for genome rearrangements in intracellular symbionts and sheds new light on the general role of R-M systems in genome evolution.

Key words: restriction-modification system, CRISPR-Cas system, mobile genetic elements, endosymbiont, genome reduction.

Mutualistic symbioses between bacteria and eukaryotes are ubiquitous in nature, and they can greatly affect the genomes of bacteria (Bennett and Moran 2015). Intracellular bacterial symbionts are notable among cellular life forms because of their extremely reduced genomes, which allow exploration of the evolutionary processes that occur during the transition from free-living to endosymbiotic lifestyles (Moran et al. 2008). The highly reduced genomes of many ancient endosymbionts have been sequenced, but only few younger endosymbionts have been studied, mostly because of the lack of suitable models (Moran and Bennett 2014). This leaves enormous gaps in our knowledge of the forces shaping genome evolution in the early stages of an endosymbiotic association. Nevertheless, it is generally accepted that the early stage of genome erosion in endosymbionts is characterized by proliferation of mobile genetic elements (MGEs), specifically transposons, which promote genome rearrangements that lead to the accumulation of disrupted genes and the first deletions of larger gene regions (McCutcheon and Moran 2012).

Endomicrobia are a class of bacteria (phylum *Elusimicrobia*) that comprise numerous intracellular symbionts of termite gut flagellates (Stingl et al. 2005; Ohkuma et al. 2007). One of them is “*Candidatus Endomicrobium trichonymphae*,” which was acquired from free-living ancestors about 40–70 Ma and since then has been vertically transmitted, leading to cospeciation with its flagellate host (Ikeda-Ohtsubo and Brune 2009; Zheng, Dietrich, Thompson, et al. 2015). The small genome size of *C. Endomicrobium trichonymphae* strain Rs-D17 (1.13 Mbp) and the presence of many pseudogenes

indicate that it is still in an early stage of genome reduction (Hongoh et al. 2008).

The recent isolation and genome sequencing of *Endomicrobium proavitum*, a close but free-living relative of *C. Endomicrobium trichonymphae* (Zheng, Dietrich, Radek, et al. 2015; Zheng and Brune 2015), provides a unique model for studying the early stages of evolution of an intracellular symbiont. Here, we report evidence that the genome rearrangements in strain Rs-D17 are associated with restriction-modification (R-M) systems that apparently act as MGEs—a finding that adds an entirely new facet to the mechanisms involved in genome evolution in endosymbiotic associations.

Results and Discussion

Comparative genome analysis revealed that 82% of the protein-coding genes of strain Rs-D17 have homologs in *E. proavitum*. However, despite a moderate level of average amino acid identity ($61.3 \pm 0.6\%$) encoded by the two genomes, a genome-wide alignment revealed that the genome of the endosymbiont underwent a large number of rearrangements (fig. 1); *Mauve* software identified 166 synteny breaks (supplementary fig. S1, Supplementary Material online). It was therefore quite unexpected that we could not find any conventional MGEs, such as prophage DNA or transposons. Instead, the genome of the endosymbiont contains many R-M systems (fig. 1 and supplementary table S1, Supplementary Material online), which are scarce in *E. proavitum* (supplementary table S2, Supplementary Material online).

© The Author 2015. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Open Access

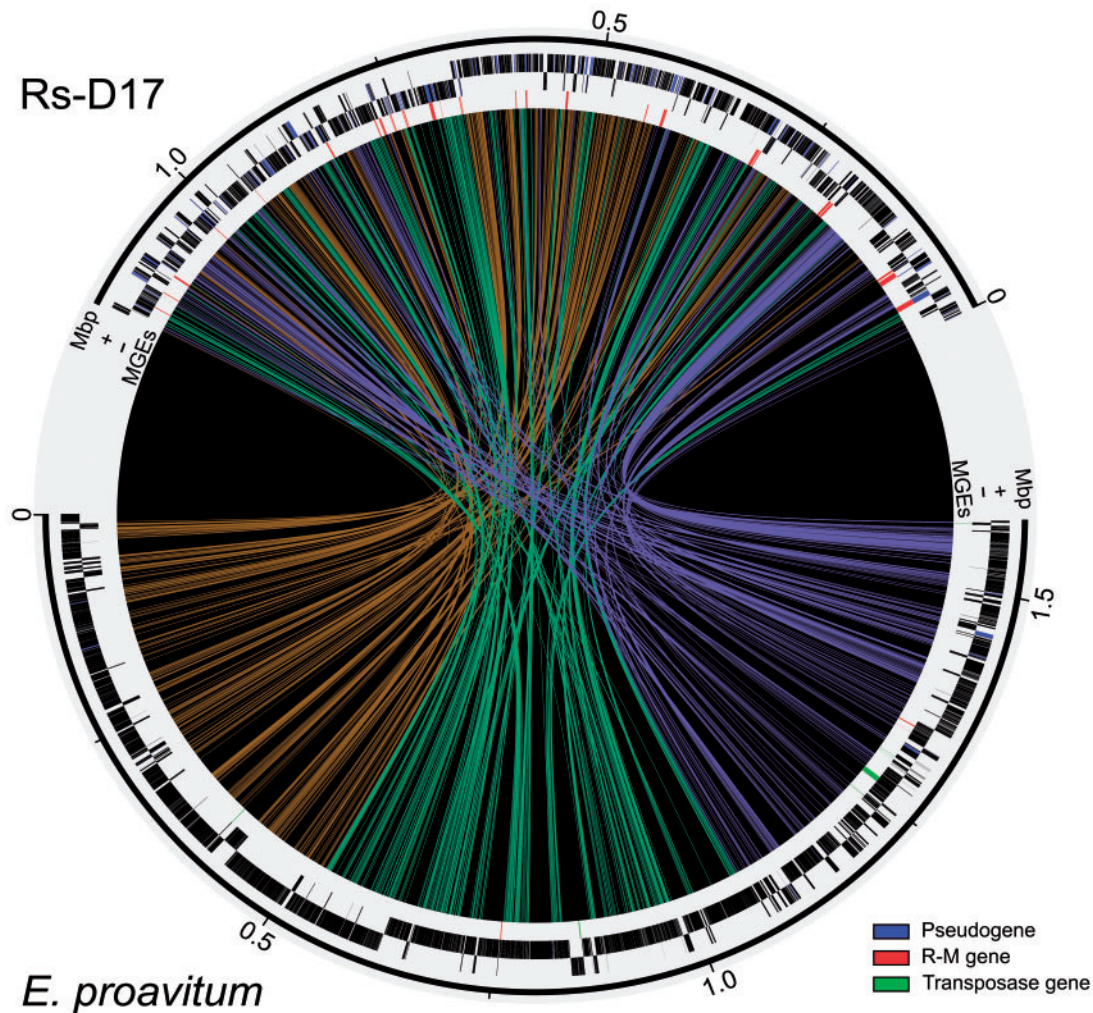


Fig. 1. Comparison of genomes of *Candidatus Endomicrobium trichonymphae* strain Rs-D17 (upper half) and its closest free-living relative *Endomicrobium proavatum* (lower half). The concentric rings denote the following features (from outside): Number of base pairs; Open reading frames on the forward (+) and reverse (–) strands; and MGEs. In the innermost graph, the homologous genes in the two genomes are connected by lines. A guanine–cytosine skew diagram of the two genomes is shown in [supplementary figure S2, Supplementary Material](#) online.

Table 1. R-M Systems and Solitary Methyltransferase Genes in *Candidatus Endomicrobium trichonymphae* Strain Rs-D17.

R-M Systems	Number					Rearrangement Patterns		
	R	M	S	Genes (pseudo)	R-M Systems (inactive)	Insertions	Duplications	Transpositions
Type I	2	2	2	6 (2)	2 (1)	0	0	2
Type II ^a	11	17	n.a.	28 (25)	13 (13)	7	1	5
Type III	4	3	n.a.	7 (5)	3 (2)	1	0	2
Solitary M genes ^b	n.a.	5	n.a.	5 (5)	n.a.	4	0	1
Total	17	27	2	46 (37)	18 (16)	12	1	10

NOTE.—See [supplementary table S1, Supplementary Material](#) online, for a full list of R-M genes in the genome. S, DNA sequence specificity; pseudo, pseudogenes; inactive, inactive R-M systems with R or M pseudogenes; n.a., not applicable.

^aOne system corresponds to Type IIC ([supplementary table S1, Supplementary Material](#) online).

^bAll solitary M genes were classified as Type II methyltransferases. They were considered solitary if they were at least 10 genes away from the next R-M system (Oliveira et al. 2014) or if they recognized different sequences (i.e., the neighboring M genes B3 and B4 in [supplementary table S1, Supplementary Material](#) online).

R-M systems are composed of genes encoding restriction endonucleases (R) and methyltransferases (M) and serve to cleave any DNA that has not been modified appropriately (Pingoud 2004). It is widely believed that R-M systems are maintained by bacteria to defend the cells from viral infection

and the entry of other foreign DNA (Kobayashi 2001). Strain Rs-D17 contains 18 R-M systems and 5 solitary M genes ([table 1](#)), which exclusively belong to Type I–III. Most of the R-M genes belong to Type II, the most common type in prokaryotes (Oliveira et al. 2014). Most R-M systems in strain Rs-D17

have been inactivated by pseudogenization of one or more genes (table 1). A truncation of R-M genes has also been reported for other bacteria (Furuta et al. 2014).

Although none of the R-M genes of strain Rs-D17 have homologs in the genome of the ancestral *E. proavatum*, most of them were also found in a metagenomic library of closely related *Endomicrobia* from a different flagellate species (supplementary table S1, Supplementary Material online; Zheng et al., unpublished data), which suggests that the R-M genes were acquired after the separation of free-living and endosymbiotic lines. The top BLASTP hits of the R-M genes are associated with diverse and only distantly related bacterial taxa (supplementary table S1, Supplementary Material online), which agrees with the concept that R-M systems have undergone frequent horizontal gene transfer (Rocha et al. 1999; Furuta and Kobayashi 2013). A comparison of pentanucleotide frequencies revealed that the Euclidean distances between certain R-M genes and other protein-coding genes in strain Rs-D17 are significantly larger than the average intragenomic distance (supplementary fig. S3 and Materials and Methods, Supplementary Material online); this also indicates horizontal gene transfers. However, movement of target recognition domains among R-M systems (Furuta and Kobayashi 2012) cannot be evaluated because most of the genes are pseudogenized.

R-M systems often co-occur with CRISPR-Cas systems, which sometimes act synergistically in the defense against foreign DNA (Dupuis et al. 2013). The genome of strain Rs-D17 contains two CRISPR-Cas systems (Type IC and Type IIC), which comprise the typical arrays of direct repeats separated by short spacer sequences close to the *cas* genes (supplementary fig. S4A, Supplementary Material online). They are present also in *E. proavatum*, and in the case of the Type IIC system, also in *Elusimicrobium minutum* (Herlemann et al. 2009), which indicated that the latter system may be conserved in this phylum. In addition, the genome of strain Rs-D17 contains a third *cas* gene cassette (Type IIC) without a CRISPR locus (supplementary fig. S4B, Supplementary Material online). Because R-M and CRISPR-Cas systems do not neighbor each other, it is unlikely that their co-occurrence in the genome of strain Rs-D17 is due to a cotransfer in “defense islands” (Oliveira et al. 2014). Although the *cas* genes and the repeat sequences are conserved between strain Rs-D17 and *E. proavatum* (supplementary fig. S5, Supplementary Material online), they do not share the same spacer sequences (supplementary table S3, Supplementary Material online), which indicates that the two organisms have experienced entirely different invasion situations. Because the spacer sequences of the two CRISPR loci do not match other regions of the genome, the CRISPR-Cas systems are unlikely to cleave endogenous DNA (Cong et al. 2013).

When we compared the genome regions flanking R-M genes in strain Rs-D17 with the syntenic regions in *E. proavatum*, we found that they were consistently associated with sites of obvious genome rearrangements. The most common rearrangements were insertions and transpositions; a duplication of a genomic region was observed only once (table 1). A typical example of an insertion is shown in figure 2A, where

a pseudogenized R-M gene set, together with two heterologous genes encoding hypothetical proteins, interrupts the space between two DNA mismatch repair (MMR) genes (*mutS* and *mutL*), which are contiguous in *E. proavatum*. This insertion is of particular interest in an intracellular symbiont because it might suppress the MMR pathway and thereby reduce the stringency of homologous recombination (Li 2008), causing gene deletions and, subsequently, a decreased genome size. The second example (fig. 2B) shows a genomic region that is contiguous in *E. proavatum* but interrupted in strain Rs-D17 by a pseudogenized M gene. During its insertion between the *aroF* and *aroA* genes, the pseudogene apparently replaced the *tyrA* gene. Interestingly, one of the flanking regions, including *tyrA*, was duplicated in this process, which possibly rescued the capacity of the endosymbiont for biosynthesis of aromatic amino acids and their provision to the host (Hongoh et al. 2008). In contrast to the situation in *Helicobacter pylori* (Furuta et al. 2011), this duplication was not associated with inversions. The third example shows a complex series of transpositions (fig. 2C) that led to a major rearrangement of a larger genome region: Several genes (*rpmE-prmC*; *murA*) contiguous in *E. proavatum* have been separated and moved to different regions in strain Rs-D17 (the *murA* gene is now flanked by an M gene). In addition, a larger stretch of DNA (ca. 2 kb) located distally from that region in *E. proavatum* has been inserted proximally to the original site of transposition and is now associated with a disrupted R-M system. Although the consequences of this rearrangement remain unclear, other rearrangements should directly affect the metabolic capacities of the endosymbiont. For example, the pseudogenization of the *glnA* gene encoding glutamine synthetase, which limits the ability of strain Rs-D17 to incorporate ammonia and requires the import of glutamine (Hongoh et al. 2008), may be related to the transposition event responsible for the presence of the flanking R-M system (A17 in supplementary fig. S6, Supplementary Material online).

Previous studies comparing the genomes of closely related prokaryotes have already documented that R-M systems are associated with different types of genome rearrangements (Tsuru et al. 2006; Furuta and Kobayashi 2013). This study is the first case where R-M systems acting as MGEs are responsible for the genome rearrangements characteristic of the early stage of genome reduction in an intracellular symbiont. It has been shown that an apparent mobility of R-M systems can be caused by their carriage on other mobile elements or by forming a composite transposon with other insertion sequences (ISs; Furuta et al. 2010; Takahashi et al. 2011). However, this is unlikely in the case of strain Rs-D17. Despite intensive searches (supplementary materials and methods, Supplementary Material online), we found no evidence for the presence of other MGEs. Moreover, the R-M genes of strain Rs-D17 were never flanked by direct repeats, which are involved in the multiplication of R-M systems in other bacteria (Nobusato et al. 2000; Sadykov et al. 2003). In view of the relatively young age of the symbionts, it is unlikely that mobile elements were originally present but eventually

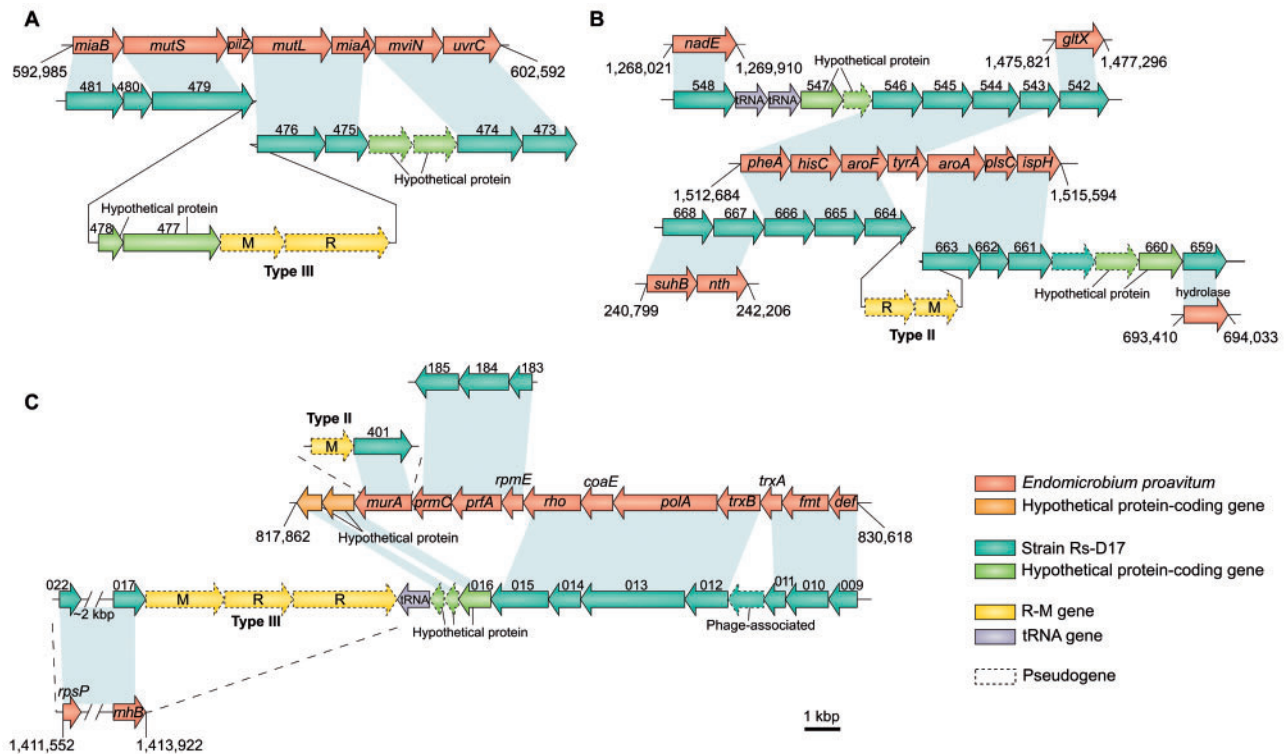


FIG. 2. Examples of genome rearrangement sites flanked by R-M systems in *Candidatus Endomicrobium trichonymphae* strain Rs-D17. (A) Simple insertion, here accompanied by foreign genes. (B) Region duplication. (C) Complex rearrangement, involving several transpositions and inversions. The number above each gene of strain Rs-D17 is the locus tag used in the annotation list of Hongoh et al. (2008). The positions of start and end points of the gene clusters in *Endomicrobium proavitum* are indicated (numbers). A detailed analysis of all rearrangement sites in the genome of strain Rs-D17 that are associated with R-M systems is shown in [supplementary fig. S6, Supplementary Material](#) online. The frequencies of these events are shown in [table 1](#).

lost from the genomes. However, it is possible that R-M systems themselves act as MGEs (Furuta and Kobayashi 2013).

Genome reduction in endosymbionts depends on the age of the association and is considered to occur in several stages (McCutcheon and Moran 2012). During the initial transition to an intracellular lifestyle, many genes of the endosymbionts are no longer required and the reduced strength of purifying selection leads to rapid accumulation of pseudogenes and rearrangements throughout the genome. In older endosymbionts, the ongoing deletion removes pseudogenes and mobile elements, resulting in highly compact genomes (Moran and Bennett 2014). The relatively recent acquisition of intracellular *Endomicrobia* (Ikeda-Ohtsubo and Brune 2009) matches both the strong accumulation of pseudogenes and massive genome rearrangements. The intermediate genome size of strain Rs-D17 is 29% smaller than that of its free-living relative (1.59 Mbp; Zheng and Brune 2015), probably due to its strictly vertical transmission, but still substantially larger than the genomes of the more ancient endosymbionts, which can be extremely reduced (Moran and Bennett 2014).

So far, the only model for a nascent symbiosis that allows a comparison of the genomes of both intracellular symbionts and their free-living relative is the *Sodalis*-allied clade of insect endosymbionts and the closely related isolate, strain HS (Clayton et al. 2012). In this case, the primary endosymbionts of insects show numerous genomic rearrangements that are

attributed mostly to IS elements that are abundant in the genome (Oakeson et al. 2014). Although the massive genome rearrangements in *Endomicrobia* endosymbionts seem to be caused by a different mechanism (i.e., involving R-M systems as MGEs), they led to the same result. Apparently, the relaxed selection in endosymbionts generates many dispensable genes (“cryptic” pseudogenes; Clayton et al. 2012) that become a landing place for any MGEs.

The reasons for the presence of R-M systems but not IS elements in the genomes of endosymbiotic *Endomicrobia* remain unclear. Because the gut flagellates phagocytose wood particles and extracellular bacteria (Yamaoka and Nagatani 1977), it is possible that phages from the environment can also invade the intracellular bacteria located in the cytoplasm of the host cell. In that case, the presence of R-M systems would provide a competitive advantage in the establishment of a stable symbiosis. Alternatively, it is possible that the R-M systems accelerate the evolution of endosymbionts. It has been shown that DNA methylation can drive adaptive genome evolution: Changes in the methylome lead to changes in the transcriptome and eventually in phenotype, thereby providing targets for selection (Furuta et al. 2014).

Our evidence for the involvement of R-M systems in genome rearrangement in intracellular *Endomicrobia* provides a new perspective for future studies targeting the genome evolution of endosymbionts, the evolutionary roles of R-M systems (Rocha et al. 2001), and their potential

application as an evolutionary genome engineering tool (Asakura et al. 2011).

Supplementary Material

Supplementary materials and methods, figures S1–S7, and tables S1–S3 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

This work was supported by the Max Planck Society. H.Z. is a doctoral student in the International Max Planck Research School for Environmental, Cellular and Molecular Microbiology (IMPRS-MIC), Marburg. We thank Lennart Randau and Srivatsa Dwarakanath (MPI, Marburg) for discussions on CRISPR-Cas systems.

References

- Asakura Y, Kojima H, Kobayashi I. 2011. Evolutionary genome engineering using a restriction-modification system. *Nucleic Acids Res.* 39:9034–9046.
- Bennett GM, Moran NA. 2015. Heritable symbiosis: the advantages and perils of an evolutionary rabbit hole. *Proc Natl Acad Sci U S A.* 112:10169–10176.
- Clayton AL, Oakeson KF, Gutin M, Pontes A, Dunn DM, von Niederhausern AC, Weiss RB, Fisher M, Dale C. 2012. A novel human-infection-derived bacterium provides insights into the evolutionary origins of mutualistic insect-bacterial symbioses. *PLoS Genet.* 8:e1002990.
- Cong L, Ran FA, Cox D, Lin SL, Barretto R, Habib N, Hsu PD, Wu XB, Jiang WY, Marraffini LA, et al. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science* 339:819–823.
- Dupuis ME, Villion M, Magadan AH, Moineau S. 2013. CRISPR-Cas and restriction-modification systems are compatible and increase phage resistance. *Nat Commun.* 4:2087.
- Furuta Y, Abe K, Kobayashi I. 2010. Genome comparison and context analysis reveals putative mobile forms of restriction-modification systems and related rearrangements. *Nucleic Acids Res.* 38:2428–2443.
- Furuta Y, Kawai M, Yahara K, Takahashi N, Handa N, Tsuru T, Oshima K, Yoshida M, Azuma T, Hattori M, et al. 2011. Birth and death of genes linked to chromosomal inversion. *Proc Natl Acad Sci U S A.* 108:1501–1506.
- Furuta Y, Kobayashi I. 2012. Movement of DNA sequence recognition domains between non-orthologous proteins. *Nucleic Acids Res.* 40:9218–9232.
- Furuta Y, Kobayashi I. 2013. Restriction-modification systems as mobile genetic elements. In: Roberts AP, Mullany P, editors. *Bacterial integrative mobile genetic elements*. Austin (TX): Landes Bioscience. p. 85–103.
- Furuta Y, Namba-Fukuyo H, Shibata TF, Nishiyama T, Shigenobu S, Suzuki Y, Sugano S, Hasebe M, Kobayashi I. 2014. Methylome diversification through changes in DNA methyltransferase sequence specificity. *PLoS Genet.* 10:e1004272.
- Herlemann DPR, Geissinger O, Ikeda-Ohtsubo W, Kunin V, Sun H, Lapidus A, Hugenholtz P, Brune A. 2009. Genomic analysis of “*Elusimicrobium minutum*,” the first cultivated representative of the phylum “*Elusimicrobia*” (formerly Termite Group 1). *Appl Environ Microbiol.* 75:2841–2849.
- Hongoh Y, Sharma VK, Prakash T, Noda S, Taylor TD, Kudo T, Sakaki Y, Toyoda A, Hattori M, Ohkuma M. 2008. Complete genome of the uncultured Termite Group 1 bacteria in a single host protist cell. *Proc Natl Acad Sci U S A.* 105:5555–5560.
- Ikeda-Ohtsubo W, Brune A. 2009. Cospeciation of termite gut flagellates and their bacterial endosymbionts: *Trichonympha* species and “*Candidatus Endomicrobium trichonymphae*”. *Mol Ecol.* 18:332–342.
- Kobayashi I. 2001. Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Res.* 29:3742–3756.
- Li GM. 2008. Mechanisms and functions of DNA mismatch repair. *Cell Res.* 18:85–98.
- McCutcheon JP, Moran NA. 2012. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol.* 10:13–26.
- Moran NA, Bennett GM. 2014. The tiniest tiny genomes. *Annu Rev Microbiol.* 68:195–215.
- Moran NA, McCutcheon JP, Nakabachi A. 2008. Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet.* 42:165–190.
- Nobusato A, Uchiyama I, Ohashi S, Kobayashi I. 2000. Insertion with long target duplication: a mechanism for gene mobility suggested from comparison of two related bacterial genomes. *Gene.* 259:99–108.
- Oakeson KF, Gil R, Clayton AL, Dunn DM, von Niederhausern AC, Hamil C, Aoyagi A, Duval B, Baca A, Silva FJ, et al. 2014. Genome degeneration and adaptation in a nascent stage of symbiosis. *Genome Biol Evol.* 6:76–93.
- Ohkuma M, Sato T, Noda S, Ui S, Kudo T, Hongoh Y. 2007. The candidate phylum “Termite Group 1” of bacteria: phylogenetic diversity, distribution, and endosymbiont members of various gut flagellated protists. *FEMS Microbiol Ecol.* 60:467–476.
- Oliveira PH, Touchon M, Rocha EP. 2014. The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Res.* 42:10618–10631.
- Pingoud A. 2004. *Restriction endonucleases*. Berlin, Heidelberg: Springer.
- Rocha EP, Danchin A, Viari A. 1999. Analysis of long repeats in bacterial genomes reveals alternative evolutionary mechanisms in *Bacillus subtilis* and other competent prokaryotes. *Mol Biol Evol.* 16:1219–1230.
- Rocha EP, Danchin A, Viari A. 2001. Evolutionary role of restriction/modification systems as revealed by comparative genome analysis. *Genome Res.* 11:946–958.
- Sadykov M, Asami Y, Niki H, Handa N, Itaya M, Tanokura M, Kobayashi I. 2003. Multiplication of a restriction-modification gene complex. *Mol Microbiol.* 48:417–427.
- Stingl U, Radek R, Yang H, Brune A. 2005. “*Endomicrobia*”: cytoplasmic symbionts of termite gut protozoa form a separate phylum of prokaryotes. *Appl Environ Microbiol.* 71:1473–1479.
- Takahashi N, Ohashi S, Sadykov MR, Mizutani-Ui Y, Kobayashi I. 2011. IS-linked movement of a restriction-modification system. *PLoS One* 6:e16554.
- Tsuru T, Kawai M, Mizutani-Ui Y, Uchiyama I, Kobayashi I. 2006. Evolution of paralogous genes: reconstruction of genome rearrangements through comparison of multiple genomes within *Staphylococcus aureus*. *Mol Biol Evol.* 23:1269–1285.
- Yamaoka I, Nagatani Y. 1977. Cellulose digestion system in the termite, *Reticulitermes speratus* (Kolbe). II. Ultra-structural changes related to the ingestion and digestion of cellulose by the flagellate, *Trichonympha agilis*. *Zool Mag.* 86:34–42.
- Zheng H, Brune A. 2015. Complete genome sequence of *Endomicrobium proavitum*, a free-living relative of the intracellular symbionts of termite gut flagellates (phylum *Elusimicrobia*). *Genome Announc.* 3:e00679–15.
- Zheng H, Dietrich C, Radek R, Brune A. 2015. *Endomicrobium proavitum*, the first isolate of *Endomicrobia* class. nov. (phylum *Elusimicrobia*)—an ultramicrobacterium with an unusual cell cycle that fixes nitrogen with a Group IV nitrogenase. *Environ Microbiol.* doi: 10.1111/1462-2920.12960.
- Zheng H, Dietrich C, Thompson CL, Meuser K, Brune A. 2015. Population structure of *Endomicrobia* in single host cells of termite gut flagellates (*Trichonympha* spp.). *Microbes Environ.* 30:92–98.