論文内容の要旨

論文題目

Genome rearrangements and restriction-modification systems: analyses by genome comparison

（ゲノム再編と制限修飾系：ゲノム比較による解析）

氏名　　古田　芳一

Abstract

Genome rearrangements such as indels, recombinations and inversions are one of the most important resources of adaptive evolution of life. These rearrangements can be induced by double strand breaks and the change in DNA modification status, which is the result of the activity of restriction-modification (RM) systems. To reveal the relationship between these two factors, genome rearrangements and RM systems, I adopted the genome comparison analysis, especially for the comparison of the *Helicobacter pylori* genome. This species are famous for variable and plastic genomes due to high rate of mutation and recombination, and also famous for possessing many RM systems. First, I classified genome inversions in 10 *H. pylori* genomes into four mechanisms, which include the novel mechanism of genome rearrangement called DNA duplication associated with inversion (DDAI). Second, RM systems in all prokaryote genome sequences are investigated and revealed the mobility of them through flanking repeat sequences in various configuration. Third, RM systems in *H. pylori* are systematically investigated and revealed the mobility of target recognition domains of RM genes in addition to the mobility of RM system itself. By these results, I revealed the generality of RM systems' mobility and their frequent involvement in genome rearrangements, which sometimes result in change of gene composition in genomes.

Results

Birth and death of genes linked to chromosomal inversion

Birth and death of genes by mechanism such as gene duplication are central to evolution, and yet the underlying mechanisms remain elusive. Availability of closely-related complete genome sequences now helps to trace gene copy number in relation to genome organization. The Gram-negative human stomach

pathogen *Helicobacter pylori* is known for its plastic genome and geographical differentiation. We sequenced 4 Japanese *H. pylori* strains and analyzed with other 6 *H. pylori* genome sequences which were already available on database. Whole genome sequence comparison of *H. pylori* revealed a copy number change specific to East Asian strains in genes of outer membrane protein family. When the position of these genes are compared with the genome context, some of them were positioned at the endpoint of large genome inversions. I named this novel mechanism of genome rearrangements as DNA Duplication Associated with Inversion (DDAI, **Figure 1**). All the long genome inversions in 10 *H. pylori* genomes were analyzed and their mechanism were classified into 4 types: (i) DDAI, (ii) homologous recombination at long inverted repeat, (iii) recombination at short inverted repeat, and (iv) inversion adjacent to a mobile element. These analysis of long inversion events allowed reconstruction of synteny evolution in this species. These results may serve as a paradigm in analyzing long and short-term genome evolution in various organisms and in cancer cells thorough extensive DNA sequencing.

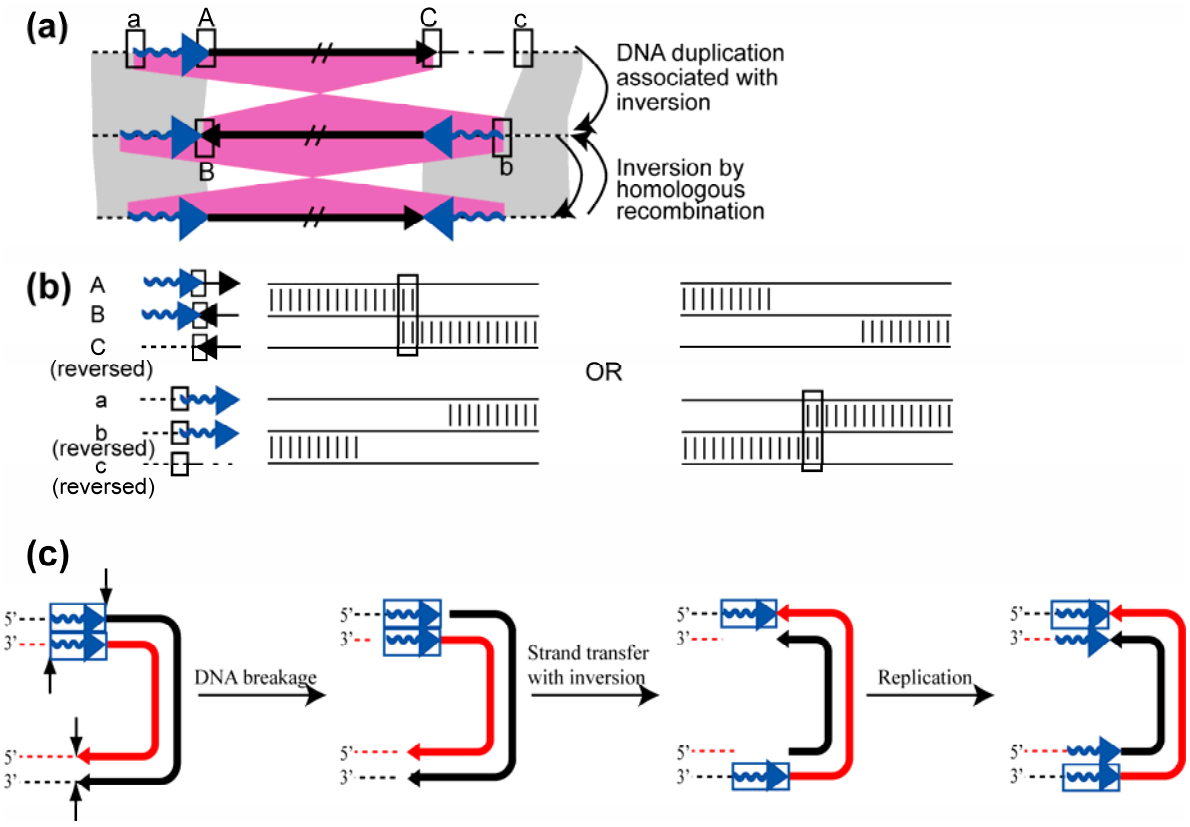

Figure 1 DDAI mechanism. (a) A genomic region represented as a wavy blue arrow is duplicated in inverted orientation at the opposite end of a chromosomal inversion. This is followed by inversion through homologous recombination involving the duplicated regions. (b) Nucleotide sequence alignments around the inversion break points. Only one of the two sequence alignment sets shows an overlap after DDAI. (c) Hypothetical mechanism of DDAI. DNA breakages are inserted at four positions, leading to generation of nick and double strand break. Strand exchange occurs with inversion, followed by closure of gaps by replication.

Putative mobile forms of restriction-modification systems and related rearrangements

The mobility of restriction-modification (RM) gene complexes and their association with genome rearrangements is a subject of active investigation. Here I conducted systematic genome comparisons and genome context analysis on fully sequenced prokaryotic genomes to detect RM-linked genome rearrangements. RM genes were frequently found to be linked to mobility-related genes such as

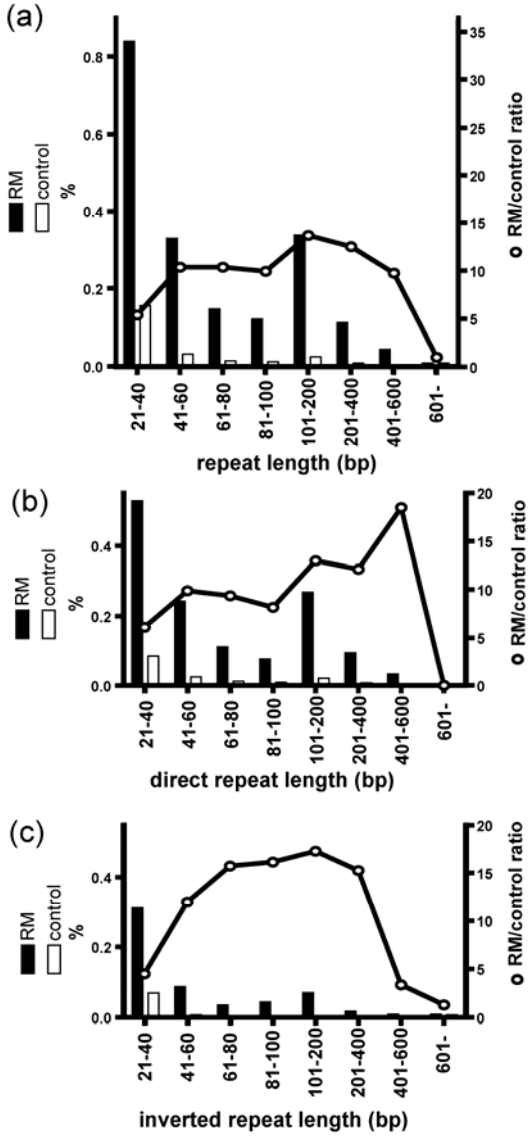

Figure 2 Frequency of genes flanked by (a) direct or inverted repeats, (b) direct repeats, and (c) inverted repeats. The vertical axis indicates % of the 11,554 compared RM-system-flanking sequence pairs. Black and white bars represent frequencies of flanking repeats and control genes, respectively. White circles indicate the ratio of RM systems to control genes for repeat frequency.

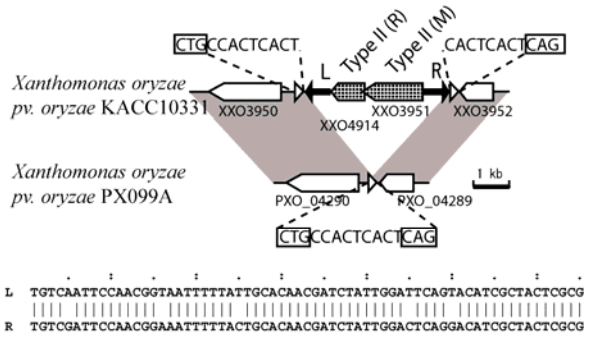

Figure 3 Transposon-like structure of RM systems flanked by repeat sequences. Triangles and arrows represent different sets of repeat sequences. (a) Type II RM genes in *X. oryzae* pv. oryzae KAC10331 are flanked by 65-bp inverted repeats (aligned below). The resulting unit is further flanked by 8-bp direct repeats (underlined), which are identical to the 8-bp sequence at the empty locus in *X. oryzae* pv. oryzae PX099A. The short direct repeat sequences are flanked by part of the predicted recognition sequence of the RM system (boxed) in the other genome. (b) Type II genes in *N. gonorrhoeae* NCCP 11945 are flanked by 26-bp inverted repeats (aligned below). The resulting unit is further flanked by 8-bp direct repeats, which are identical to the 8-bp sequence at the empty locus in *N. meningitides* MC58.

integrase and transposase homologs. They were flanked by direct and inverted repeats at significantly high frequency (**Figure 2**). Insertion by long target duplication was observed for I, II, III, and IV restriction types. I found several RM genes flanked by long inverted repeats, some of which had apparently inserted into a genome with a short target duplication. In some cases, only a portion of an apparently complete RM system was flanked by inverted repeats (**Figure 3**).

Domain movement within a gene

A function of a protein is carried out by a specific domain localized in its specific position. In the present work, I report that a specific amino acid sequence can move between a domain at one position and a domain at another position within a same gene. This discovery was made during sequence comparison of restriction enzyme genes within a bacterial species called *Helicobacter pylori*. In the specificity subunit of Type I restriction enzymes, DNA sequence recognition is mediated by Target Recognition Domain 1 (TRD1) and TRD2. To our surprise, several sequences are shared by TRD1 and TRD2 of genes (alleles) at the same locus (chromosomal location): these sequences appear to have moved between these two domains (**Figure 4**). The gene/protein organization can be represented as x-(TRD1)-y-x-(TRD2)-y, where x and y stand for the repeat sequences. The movement is likely realized by recombination at these flanking DNA repeats. In accord with this hypothesis, recombination at these repeats also appears to have decreased the two TRDs into one TRD, or increased into three (TRD1-TRD2-TRD2) and to have allowed TRD sequence movement between genes at different loci. Similar movements of a sequence between TRD1 and TRD2 were observed for the specificity subunit of a Type IIG restriction enzyme. The lateral domain movements within a protein, designated here as DOMO (*do*main *mo*vement), represent novel routes for diversification of proteins.

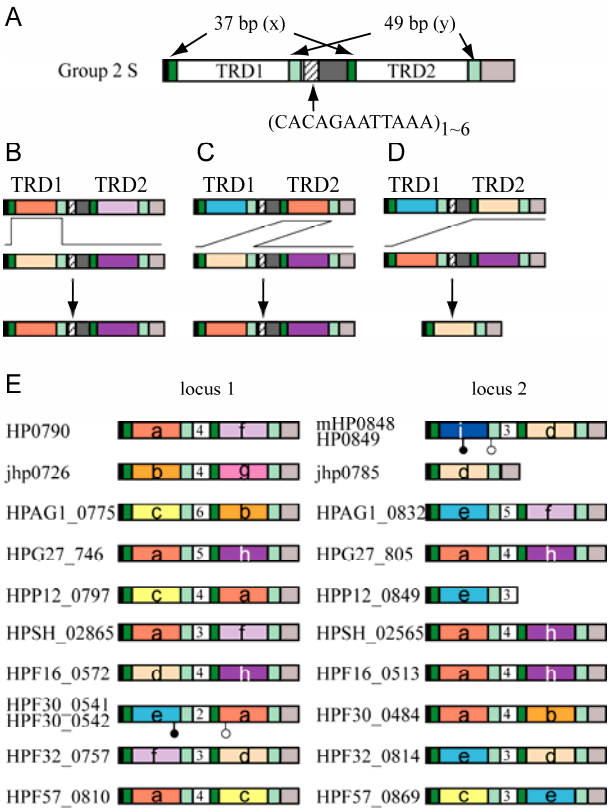

Figure 4 Diversity at target recognition domains of Group 2 specificity subunit. (A) Gene context of Group 2 specificity subunit. (B, C) Mechanisms of TRD exchange by homologous recombination. (D)Mechanism of TRD loss by homologous recombination. (E) Alleles of Group 2 specificity subunit paralogs at locus 1 and 2.