Birth and death of genes linked to chromosomal inversion

Yoshikazu Furutaa,b,1, Mikihiko Kawai,b,c,1, Koji Yahara,1,2, Noriko Takahashia,b, Naofumi Handia,b, Takeshi Tsurub,f, Kenshiro Oshimab, Masaru Yoshidab, Takeshi Azumab, Masahira Hattorig, Ikuo Uchiyamac, and Ichizo Kobayashia,b,f,2

*Department of Medical Genome Sciences, Graduate School of Frontier Sciences, University of Tokyo, Minato-ku, Tokyo 108-8639, Japan; †Institute of Medical Science, University of Tokyo, Minato-ku, Tokyo 108-8639, Japan; ‡Laboratory of Genome Informatics, National Institute for Basic Biology, National Institutes of Natural Sciences, Okazaki, Aichi 444-8585, Japan; ¶Graduate School of Medicine, Kurume University, Kurume, Fukuoka 830-0011, Japan; **Fujitsu Kyushu Systems Ltd., Fukuoka, Fukuoka 814-8589, Japan; ††Department of Biophysics and Biochemistry, Graduate School of Science, University of Tokyo, Minato-ku, Tokyo 108-8639, Japan; ‡‡Department of Computational Biology, Graduate School of Frontier Sciences, University of Tokyo, Kashiwa, Chiba 277-8561, Japan; and †††Department of Gastroenterology, Graduate School of Medicine, Kobe University, Chuo-ku, Kobe, Hyogo 650-0017, Japan.

Edited by Patrick Sung, Yale University, New Haven, CT, and accepted by the Editorial Board December 7, 2010 (received for review August 24, 2010)

The birth and death of genes is central to adaptive evolution, yet the underlying genome dynamics remain elusive. The availability of closely related complete genome sequences helps to follow changes in gene contents and clarify their relationship to overall genome organization. Helicobacter pylori, bacteria in our stomach, are known for their extreme genome plasticity through mutation and recombination and will make a good target for such an analysis. In comparing their complete genome sequences, we found that gain and loss of genes (loci) for outer membrane proteins, which mediate host interaction, occurred at breakpoints of chromosomal inversions. Sequence comparison there revealed a unique mechanism of DNA duplication: DNA duplication associated with inversion. In this process, a DNA segment at one chromosomal locus is copied and inserted, in an inverted orientation, into a distant locus on the same chromosome, while the entire region between these two loci is also inverted. Recognition of this and three more inversion modes, which occur through reciprocal recombination between long or short sequence similarity or adjacent to a mobile element, allowed reconstruction of synteny evolution through inversion events in this species. These results will guide the interpretation of extensive DNA sequencing results for understanding long- and short-term genome evolution in various organisms and in cancer cells.

C hanges in the gene repertoire of a genome are central to adaptive evolution. For example, microorganisms change host-interacting proteins on their surface, evading the immune response and presenting alternative modes of adhesion, and human cancer cells show amplification of oncogenes. In addition to horizontal transfer (1), various mechanisms for generation and loss of genes have been recognized (2). A gene may move to a distant locus through a variety of recombination mechanisms. Replication error and illegitimate recombination may generate a tandem duplication, which may lead to further amplification through homologous interaction (3). These mechanistic analyses, however, are not usually connected to genome-wide dynamics.

Recent innovations in DNA sequencing are providing a complementary approach to the study of genome-wide changes. Comparing complete genome sequences between multiple closely related organisms may reveal the dynamics of genome evolution in detail (4). We can assess copy number variation by measuring the exact number of homologous genes in a genome and trace both subtle mutations and global genome rearrangements. In bacteria, this approach has turned out to be especially powerful (5, 6).

In this study, we compared 10 genome sequences of Helicobacter pylori, a bacterial species that is present in half the human population and is responsible for stomach cancer and other diseases (7). H. pylori genomes show high plasticity through homologous recombination and have experienced geographical differentiation (8). Detailed comparison of these 10 complete genome sequences led to discovery of a unique mechanism of DNA duplication that is linked to chromosomal inversion. We inferred the mechanisms of all of the large inversions seen and reconstructed genome synteny evolution through inversion.

Results

Linkage of Changes in Number of Paralogous Gene Loci and Chromosomal Inversion. In this work, we compared genome organization of 10 H. pylori strains: five Western strains [four European strains (26695, HPAG1, G27, and P12) and one West African strain (399)] and five Eastern strains [four East Asian strains (F16, F30, F32, and F57) and one Amerind strain (Shi470)]. H. pylori genes involved in host interaction, such as those coding for outer membrane proteins (OMPs), show large divergence between the Western strains and the Eastern strains (9). By comparing the complete genome sequences of Eastern strains with those of Western strains, we found that differences in the number of loci for two OMP families, oipA (one locus in the West vs. two loci in the East) and hopMN (two loci in the West vs. one locus in the East), are both linked to a large chromosomal inversion (Fig. L4; for details, see Genome Comparison in Materials and Methods and SI Materials and Methods). These changes may have contributed to host adaptation because OMPs in general, and these OMP families in particular, are important in host interaction and pathogenicity (9–11).

DNA Duplication Associated with Inversion. Further comparative sequence analysis of the inversion breakpoints led to our proposal for a unique process of DNA duplication, which we designated DNA duplication associated with inversion (DDAI), illustrated schematically in Fig. 1B. The DDAI concept explains the observed linkage of the copy number variations to an inversion as follows: a DNA region in one locus (the blue wave arrow at a-A in Fig. 1B) generates a copy at a distant locus (C-C) on the same chromosome in a process that also results in the inversion of the...
Fig. 1. Gene duplication/decay through the DDAI process. (A) Linkage of oipA gene duplication and hopH gene decay with chromosomal inversion between schematic Western and East Asian genomes. The upper side corresponds to P12 genome and the lower side to one of the two possible ancestral structures of the four Japanese genomes (Fig. S5). (B) DDAI. (B1) Genomic region in strain 1, represented as a blue wavy arrow, is duplicated in inverted orientation in an event associated with a chromosomal inversion in strain 2. The blue wavy segment originally present is inserted in inverted orientation in an event associated with a chromosomal inversion (inversions B2 and D2; Fig. 2). The DDAI event was followed by inversion through homologous recombination between the duplicated regions (Fig. 1C, third and fourth lines) in two genomes (inversions B2 and D2; Fig. 2). The loss of one of the hopN/hopM loci (Fig. 1A) is explained by a DDAI event that resulted in a gene fusion as follows: A DNA region covering the 3′ end of the dpnA gene (encoding a DNA methyltransferase; blue wavy arrow at a-A in Fig. 1D1) was duplicated into a site within the hopN gene (at C-c in Fig. 1D1), generating a hopN'-dpnA fusion gene (at b in Fig. 1D1). This was likely inactive and seems to have acquired a secondary deletion within hopN'. Here the alignment a-b-c shows an overlap, whereas a-B-C contains a gap. The DDAI event was again followed by inversion through homologous recombination interpretation of large-scale orientation, arrowheads on the DNA schematically indicate global orientation, not 3′ ends. A pair of staggered single-strand breaks is made flanking a region (blue wavy arrows) to be duplicated. A double-strand DNA break is made at the target site. Single-strand ends of the region to be duplicated are ligated to the target. This strand transfer results in inversion between the two sites when the gaps are closed by replication, which may be primed from the unligated termini that remain. (C) Concrete example of gene duplication resulting from an inferred DDAI event. (C1) Hypothetical step for hopN duplication (inversion D1) followed by inversion through homologous recombination (inversions B2 and D2; the Amerind strain Shi470 has a secondary deletion in the right homologous segment, not illustrated here). (Cii) Sequence alignments of the labeled boxes. A nonhomologous substitution, not accounted for by the above simple model (Biii), is found near the junction when the first and second lines in Ci were compared. (Ciii) A second concrete example of DDAI, which was followed by gene decay. (Di) Hypothetical steps for loss of a hopN gene. A region covering the dpnA C terminus was duplicated and fused to a part of hopN (inversion C2). The fusion gene underwent further deletion by a mechanism independent of this model, followed by further inversion through homologous recombination (inversion D2). (Dii) Sequence alignments. Substitution of nonhomologous material (138 bp at box B added, and part of the C terminus of hopN lost, comparing the top line with the hypothetical intermediate) is not accounted for by the above simple model (Biii). P12, J99, and Shi470 are not illustrated here because P12 and J99 had a different hopM allele at the hopM locus, and Shi470 had lost the gene at the hopN locus by simple deletion. In all of the panels, gray and pink shading between rows indicates similarity at the DNA level. Gray shading indicates syntenic alignment, whereby two chromosomes carry the same sequences in the same order; pink shading indicates inverted alignment, whereby two chromosomes carry the same sequences but in inverted orientation. Blue wavy arrows represent the region subject to duplication. Straight arrows represent global orientation of a DNA region to be inverted. Boxed regions are the breakpoints of the indicated alignments; capital letters indicate one set containing similar sequences; lowercase letters indicate a second set. In C and D, arrowed boxes (with wings) represent ORFs, with homologous segments carrying the same pattern; a pentagon (arrow box with no wing) indicates a relevant intergenic region, with homologous regions labeled by the same pattern.
More DDAIs. We identified a third example of the DDAI type inversion (inversion J in Fig. 2 and Table 1), this time in a European strain. As illustrated in Fig. S2i, part of a hypothetical gene (HPAG1_1222) was duplicated with inversion (14). An overlap is seen in alignment A-B-C (Fig. S2ii). Occurrence of this inversion was noted earlier (14).

A DDAI event may have also occurred in an inversion involving a smaller region. A region of approximately 1 kb in the cag pathogenicity island that includes cagQ was reported to be inverted in some Japanese strains (15). Our detailed sequence comparison confirmed a similar inversion in Shi470 and in three out of the four Japanese genomes and, furthermore, strongly suggested that these inversion events likely occurred by two steps of DDAI with generation of inverted repeat pairs of 87 bp and 82 bp (Fig. S3).

**Homologous recombination.** The second mechanism of inversion, homologous recombination (Table 1 and Fig. S4ii), has long been recognized in various organisms (16). Notably, most events (five of six) seem to have taken place between duplicated regions generated in the first place by DDAI (inversions B2, D2, E, K, and Figs. S2 and S3). This association can be explained by the dependence of homologous recombination on near-identity of substrate sequences (17). Immediately after the duplication, the regions share perfect identity and are prone to engage in homologous

---

**Table 1. Inversions between complete genomes**

<table>
<thead>
<tr>
<th>Label</th>
<th>Figure</th>
<th>Inverted repeats*</th>
<th>Effect on gene</th>
<th>Inserted element</th>
<th>RM linkage†</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA duplication associated with inversion (DDAI)</td>
<td>A</td>
<td>1C</td>
<td>1007/1035 (oipA)</td>
<td>Dilation (oipA)</td>
<td>+</td>
</tr>
<tr>
<td>C2</td>
<td>1D</td>
<td>325/330 (dpnA)</td>
<td>Decay (hopN); partial duplication (dpnA)</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>J</td>
<td>S2</td>
<td>244/255 (hypothetical)</td>
<td>Partial duplication (hypothetical)</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Homologous recombination</td>
<td>B1</td>
<td>56</td>
<td>(Repeat 8, deleted)</td>
<td>TnPZ</td>
<td>+</td>
</tr>
<tr>
<td>B2</td>
<td>1C</td>
<td>462/522 (oipA)</td>
<td>TnPZ</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>D2</td>
<td>1C</td>
<td>984/1020 (oipA)</td>
<td>TnPZ</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>1D</td>
<td>331/331 (dpnA)</td>
<td>TnPZ</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td></td>
<td>1399/1407 (repeat 8)</td>
<td>TnPZ</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>S2</td>
<td>244/255 (hypothetical)</td>
<td>TnPZ</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Recombination at a short sequence similarity</td>
<td>C1</td>
<td>3</td>
<td>8/10 (endonuclease)</td>
<td>Splitting (endonuclease)</td>
<td>+</td>
</tr>
<tr>
<td>D1</td>
<td>57</td>
<td>5/5 (hypothetical, TnPZ)</td>
<td>Splitting (hypothetical, TnPZ)</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Inversion adjacent to a mobile element insertion</td>
<td>G</td>
<td>S8A</td>
<td>IS605</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>H1</td>
<td>4</td>
<td>Type II RM</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H2</td>
<td>S8B</td>
<td>Type II RM</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>S8C</td>
<td>Type I RM</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>S8D</td>
<td>TnPZ</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M1</td>
<td>S8F</td>
<td>IS605</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M2</td>
<td>S8F</td>
<td>IS605</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M3</td>
<td>S8F</td>
<td>IS605</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M4</td>
<td>S8G</td>
<td>IS605</td>
<td>+</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Large (>5 kb) inversions were analyzed (Materials and Methods). For details, see Table S1. RM, restriction–modification genes; TnPZ, conjugative transposable element (20).

*Sequence present in multiple strains and present as an inverted repeat in at least one strain.
†Linkage to RM genes and endonuclease genes not further than five ORFs.
‡Repeat 8, a repetitive element (19).
§Previously reported (14, 18).
Inversion through recombination, resulting in another inversion event (Fig. S4 i and ii). Nonreciprocal homologous interaction such as gene conversion would further tend to homogenize their sequences (Fig. S4iv). Indeed the duplicated oipA alleles in the East Asian strains are very similar to each other in their sequence and are more similar within, than between, genomes (Fig. S5). When sequence divergence reaches a critical point, homologous recombination and, therefore, further inversion will occur less readily (Fig. S4 iii–v).

Among other pairs of known inverted repeats in the H. pylori genomes (18, 19), the repeat 8 pair showed inversion [inversions B1 (Shi470) and F (F16 and F32); Table S1]. Repeat 8 seems to have been lost after inversion B1 through one of the conjugative transposable element TnPZ (20) and secondary deletion events in Shi470 (Fig. S6).

Recombination at short sequence similarity. The third type of inversion (Table 1) occurred through reciprocal recombination involving a short (5–10 bp) region of sequence similarity. The example in Fig. 3 (inversion C1) involved a sequence within a gene for a putative endonuclease and a distant intergenic sequence and resulted in the splitting of this gene into two parts at two very distant loci (Fig. 2).

In another example (inversion D1; Fig. 2 and Table 1), recombination between a 5-bp sequence within a gene on a copy of TnPZ and a 5-bp sequence in a distant gene resulted in splitting of this conjugative transposable element (Fig. S7).

Inversion adjacent to a mobile element. The remaining examples of inversion (Table 1) all carried a mobile element. The second line of evidence is from the discovery of an intermediate inversion adjacent to a mobile element (Fig. S7).

The first line of evidence is our discovery of an intermediate form for inversion H1 with a restriction–modification system insertion (Fig. 4i). The restriction–modification system seems to have inserted into the genome of HPG1 with an ~400-bp target duplication, a mode already seen with several groups of restriction–modification systems (22, 23). A single copy of the repeat was observed at the empty site in strain G27. A segment adjacent to the restriction–modification system has been inverted in strain J99. This inversion involved a site approximately 100 bp away from the duplicated copy of the repeat in HPG1 (Fig. 4e).

The second line of evidence is from inversion G (Fig. S8i). A TnPZ seems to have inserted with duplication of a target sequence, and IS605 seems to have inserted into this TnPZ copy. This seems to have been followed by inversion at one end of the copy of the insertion sequence (IS). The IS end and the inversion break point are very close (Fig. S8ii).

The third line of evidence is that the overlap of the end of the insert (either a restriction–modification system or an IS) and the inversion break point was observed for inversion I, L, M1, M2, M3, and M4 (Figs. S8 C, D, E, and G).

Reconstruction of Inversion History in H. pylori. On the basis of these findings and interpretations, we reconstructed the evolutionary history of genome synteny involving large inversion events in H. pylori (Fig. 5).

The phylogenetic tree simply based on the number of inversion events (Fig. S4) shows the likely order of the events, but its shape is different from a tree based on gene sequences (8). One reason likely comes from the assumption that all of the inversion events occur at a constant rate during evolution (24). According to our observations, this assumption cannot be justified at least for this species, because of the variety in the molecular mechanisms of inversion and their mutual relationships. Inversion by homologous recombination would occur more frequently than DDAI or inversion by short-homology recombination. Homology generated by a DDAI event would promote later frequent homologous recombination. The last hypothesis is supported by the observation that there are multiple inversions at oipA and hopN′-dpnA by homologous recombination in the Japanese lineage after the DDAI event in the ancestor (Fig. 5 A and B).

The difference between the two trees may become smaller when we learn the frequency of the four types of inversion and assign a unique branch length to each of the four.

Discussion

DDAI is similar to the replicative inversion process used by specific DNA transposons (12, 25, 26) but differs in two important aspects: the nature of DNA duplicated and the sequences at
the breakpoints. Only one of the two sequence alignment sets showed an overlap (Fig. 1 Bii, Cii, and Dii and Fig. S2i), which implies deletion/insertion at one of the breakpoints in the framework of the model in Fig. 1Bii. DNA degradation from the initial double-strand break may have resulted in the deletion. The induction of natural competence by DNA damage recently reported (27) could be related to the deletion/insertion. Instead, the two distinct processes, inversion and deletion/insertion, may have been evolutionarily related. The short nucleotide overlap at an inversion breakpoint also reminds us of the other replicative models, such as the microhomology-mediated break-induced replication model (28). Although the possibility of secondary rearrangements can never be excluded in the genome comparison approach, we prefer the DDAI concept of a single event generating duplication and inversion at the same time, or at almost the same time, as an attractive and productive working hypothesis for further genome comparison and experimental analyses.

We do not yet know what catalyzes DDAI reaction. Enzymes proposed to be involved in transposons’ replicative inversion, such as transposases, DNA polymerases, DNA ligases, and topoisomerases (12), are all available in H. pylori according to genome annotation (19). In addition, the initiating incisions for DDAI may have been catalyzed by restriction–modification systems in place of the transposases in replicative inversion. The following considerations favor this possibility: (i) restriction-modification genes are abundant in H. pylori genomes (19); (ii) some restriction–modification enzymes show a nicking activity (29); (iii) some restriction enzymes are similar to transposases in structure and function (30); and (iv) some restriction–modification units are similar to DNA transposons in organization (23).

A related issue is the mechanism of inversion adjacent to mobile element insertion (type iv). The four restriction–modification systems as well as the IS605 and Tn17Z copies near the inversions may have caused them. Indeed, in one of the inversions adjacent to a restriction–modification system, several recognition sites of the linked restriction–modification system were found at the breakpoints (inversion I; Fig. S8C).

The tight association of inversion and gene births and deaths demonstrated in this work will contribute to our understanding of the dynamics of genome evolution in various organisms, both in germ lines and in somatic lines, and in normal and cancerous cells. Reconstruction of genome synteny evolution must account for the variety in inversion mechanisms and the wealth of biological information embedded in the genomes.

Materials and Methods

**Genome Sequences.** Sequence data were obtained from National Center for Biotechnology Information Genome for the circularized chromosomes of the following H. pylori strains: European strains, 26695 (NC_000915.1), HPAG1 (NC_000886.1), G27 (NC_011333.1), P12 (NC_011498.1); West African strain, J99 (NC_000921.1); Amerind strain, Shi470 (NC_010698.2); and East Asian strains, F16 (AP011194.0), F30 (AP011194.1), F32 (AP0111943.1), and F57 (AP011194.5).**

**Genome Comparison.** Genome alignments and dotplots were obtained with Comparative Genome Analysis Tool (31) and used for evolutionary inference (6, 25, 32, 33). Detection of inversion breakpoints and extraction of genome blocks were helped by Mauve (34). Genome blocks smaller than 5 kb were ignored. History of genome inversion events was reconstructed using the Multiple Genome Rearrangement program (35) with manual curation (SI Materials and Methods). In Fig. 1A, the other inversions were ignored for clarity. The two schematic structures correspond to that of strain P12 and one of the two equally possible hypothetical ancestral structures of the four Japanese strains (the intermediate structure labeled with inversions C1 and C2 in Fig. S8B).

**ACKNOWLEDGMENTS.** We thank Shigeru Iida for comments on an early version and Elisabeth Raleigh for expert help in improving the manuscript during revision. This work was supported by the Institute for Bioinformatics Research and Development, the Japan Science and Technology Agency; by a grant (to I.K.) from the global COE (Center of Excellence) project “Genome Information Big Bang” from Ministry of Education, Culture, Sports, Science and Technology-Japan (MEXT), Suzuken Memorial Foundation, Urakami Foundation; by Grant-in-Aid for Scientific Research 20310125 from the Japan Society for the Promotion of Science (to I.U.); by a grant (to N.H.) from MEXT, Takeda Foundation, Sumitomo Foundation, Kato Memorial Bioscience Foundation, and Naito Foundation; and by Grants-in-Aid for Scientific Research on Priority Areas “Comprehensive Genomics” from MEXT (to M.H.).
Manual Refinement of Phylogenetic Trees Based on Inversion Events.

In brief, Multiple Genome Rearrangement (MGR) constructs an inversion-based phylogenetic tree by minimizing the number of inversion events between genomes (1), based only on comparison of the order and orientation of parts in each genome. The method does not take into account, for example, the high frequency of homologous recombination between long identical inverted repeats. We analyzed sequences at the breakpoints of each inversion in detail by sequence alignments and dotplots and revealed that some of the inversions are flanked by inverted repeats at the breakpoints. We modified the MGR tree to include the duplication associated with inversion (DDAI) events and the ensuing inversion events by homologous recombination, which were sometimes ignored for the following reason. We hypothesized the DDAI-related inversion events as genome A → (DDAI) → genome B → (homologous recombination) → genome C. We can distinguish genome A and genome C with respect to the presence/absence of the repeats flanking the inverted region, whereas MGR does not distinguish genome A and genome C because it considers only the orientation of the inverted region and neglects the repeats. Therefore, MGR often considers the above two-step process as a one-step event, that is, genome B → genome A (= genome C), trying to minimize the total number of inversions.

Author Contributions in Detail

Y.F. was the first to present DDAI idea, contributed to informatics analyses including sequence alignments and examination of every inversion in sequence level, and wrote the manuscript; M.K. contributed to informatics analyses, including discovery of linkage of copy number variation and genome inversion, and carried out initial survey of genome rearrangements; K.Y., T.T., and I.U. contributed to informatics analyses; N.T. and N.H. contributed to preparation of genomic DNA; K.O. and M.H. contributed to sequencing and assembly; M.Y. and T.A. provided the strains; and I.K. contributed to design and writing.

Fig. S1. (Continued)
Fig. S1. (Continued)
**Fig. S1.** Inversion breakpoints indicated on genome-wide dotplots. Both termini of an inversion labeled as in Table 1 (main text) are indicated by a red circle. 
(A) Inversions A, B1, B2, D1, D2, F, and G. (B) Inversions C1, C2, D1, and E. (C) Inversions H1 and H2. (D) Inversion I. (E) Inversions J, K, and L. (F) Inversions M1, M2, M3, and M4.

(i) 
(ii) 

**Fig. S2.** Partial gene duplication resulting from a DDAI event in a European strain. (i) Hypothetical steps duplicating part of a hypothetical gene (HPAG1_1222, inversion J), followed by further inversion through homologous recombination (inversion K). The DDAI is again accompanied by non-homologous substitution near the target site (box b and the stippled ORF on the first line). (ii) Sequence alignments.
Reconstruction of inversion events around cagQ on cag pathogenicity island. Approximately 1-kb region including cagQ in a hypothetical ancestor (second line) was inverted in Shi470 (third line) by a DDAI event, generating 87-bp inverted repeats (black triangle). A part of this 1-kb region was simply deleted (first line). Another DDAI event in Shi470 resulted in generation of 82-bp inverted repeats (white triangles) (fourth line). Finally, inversion by homologous recombination between the 82-bp inverted repeats (white triangles) occurred (fifth line).

Fig. S4. Expected consequences of DDAI. The regions duplicated by DDAI (i) are involved in homologous recombination leading to inversion (ii). Sequence divergence (iii) is prevented by gene conversion (iv). Further sequence divergence (v) inhibits their homologous recombination.

Similarity between oipA alleles. Pairwise nucleotide sequence identity is shown.
Fig. S6. Inversion by homologous recombination at repeat 8 (inversion B1). (i) Reconstructed steps. After the inversion, the left repeat 8 copy was deleted, and the right repeat 8 copy experienced insertion of TnPZ.

(ii) Genome context comparison.

(iii) Sequence alignments at the inversion break points. Boxed sequences are the 9-bp repeat sequences generated by TnPZ insertion.
Fig. S7. Inversion by reciprocal recombination at a short sequence similarity, which resulted in splitting of a genomic island, TnPZ (inversion D1). (i) Reconstructed steps. Intact type 1b TnPZ carries two copies of 5′ CTCTT at A1 and A2. Recombination involving them may have led to deletion of the sequence between them. The resulting single copy experienced reciprocal recombination with another copy within a gene (HPF32_0137) outside of TnPZ, causing inversion. (Alternatively, the deletion within TnPZ may have taken place after the inversion or coupled to it.) (ii) Sequence alignments at the inversion breakpoints. The sequence of type 1b TnPZ in P12 is used for A1 and A2. The sequence of HPF32_0137 is used for C.
Fig. S8. (Continued)
Fig. S8. (Continued)
Fig. S8. (Continued)
Fig. S8. Inversions associated with insertion of a mobile element. (A) Inversion G. (Ai) Reconstructed steps. Insertion of intact type 2 TnPZ was followed by an internal deletion. (Alternatively, insertion of a TnPZ copy with deletion occurred in the first place.) Then IS605 inserted into this defective TnPZ. Inversion took place adjacent to this IS605 copy. (Aii) Genome comparison. (Aiii) Sequence alignments at the inversion break points. Boxed sequences are the 7-bp target duplication generated by TnPZ insertion. (B) Inversion H. (Bi) Reconstructed steps for G27. A restriction-modification (RM) system was hypothesized to have inserted into a hypothetical ancestor (first line) with tandem duplication of a long target region (white triangle) (second line). Inversion occurred adjacent to this RM including a copy of the duplicated target sequence (third line), followed by deletion of the RM (fourth line). (Bii) Reconstructed steps for J99. The RM was lost from the hypothetical intermediate (second line) in Bi. (Biii) Sequence alignments. (C) Inversion I. (Ci) Genome comparison. (Cii) Sequence alignments. The R (restriction) gene is depicted by an arrow. The recognition sequences of the RM system (5'TGCA 3') underlined are near the break points. (D) Inversion L. (Di) Genome comparison. The region adjacent to type I restriction system was inverted. (Dii) Sequence alignments. The S (specificity) gene is depicted by an arrow. (E) Inversion M1 and M2. (Ei) Reconstructed steps. The ancestor (first line) carries a restriction system. Insertion of IS605 (second line) is followed by inversion of its neighbor region (inversion M1) (third line). Next, a region adjacent to the restriction system was inverted (inversion M2) (fourth line), resulting in splitting of its specificity subunit gene. (Eii) Genome comparison. (Eiii) Sequence alignments. (F) Inversion M3. (Fi) Reconstructed steps. Inversion took place near TnPZ and IS605. (Fii) Genome comparison. (Fiii) Sequence alignments. (G) Inversion M4. (Gi) Reconstructed steps. IS605 insertion is followed by adjacent inversion. (Gii) Genome comparison. (Giii) Sequence alignments.

Other Supporting Information Files

Dataset S1 (XLS)