Synopsis of Original Research Paper

A mechanism to generate variation in pathogenicity-related tandem paralogues of *Staphylococcus aureus*

Ichizo Kobayashi

Graduate School of Frontier Sciences, University of Tokyo & Institute of Medical Science, University of Tokyo

Staphylococci are normal inhabitants on skin and mucous membrane of humans and can become pathogenic. Especially, *Staphylococcus aureus* can cause severe inflammation in various tissues including skin which sometimes results in serious disorder. In addition, this bacterium has developed resistance to practically all types of antibiotics. Due to its clinical importance, whole genome sequences of multiple *S. aureus* strains have been decoded. To date, complete genome was sequenced for more than ten strains and has become publicly available. In their genome, several tandem clusters of paralogous genes, likely pathogenicity-related, have been found in genomic islands. Intergenomic comparison with respect to these clusters revealed polymorphisms in them. In the case of *lpl* gene cluster, encoding lipoprotein homologues, the variation was very extensive. Our multiple sequence alignment revealed presence of a region highly conserved not only at the amino acid sequence level but also at the nucleotide sequence level and regions to its 5' and 3' sides, which are more variable. The highly conserved nucleotide sequences are likely to have provided a site for homologous recombination generating the variation of this region. Comparison of phylogenies of the 5'-variable region and the 3'-variable region revealed significant incongruence within the same ORF. By contrast, pairs of 3'-variable region of an ORF and 5'-variable region of its downstream ORF gave more congruent phyogenies with groups of conserved pairs, which suggested their linkage. An intergenic region sandwiched by such a pair of variable regions seemed to have co-evolved with it. These lines of observations supports our hypothesis that homologous recombination at the central conserved region have played a major role in generating variations in this cluster. This model explains not only formation of various types of rearrangements through multiple crossing-over events but also generation of a novel ORF with different sets of two variable regions. The crossing-over events caused extensive shuffling of the two variable regions in one ORF, but maintained a conserved unit comprising 3'-variable region, intergenic region, and 5'-variable region spanning adjacent ORFs. This characteristic mode of tandem paralogue diversification, maintaining 3'-part of a gene, intergenic region, and 5'-part of its downstream gene as a unit of evolution, is unique among previously studied paralogue rearrangements, in which an ORF tended to have been considered as the unit of evolution.