論文の内容の要旨

# Evolutionary Study on Deletion Polymorphism of the *GSTM1* Gene in Humans

（ヒト *GSTM1* 欠失多型の進化学的研究）

齊藤真理恵

The glutathione S-transferase (GST) gene superfamily comprises phase II detoxification enzymes that catalyze conjugation of glutathione (GSH) to xenobiotics (Sheehan et al., 2001). GSTs play key roles in cellular protection against xenobiotics (McIlwain et al., 2006). Genetic variants of GSTs have been studied with respect to disease susceptibility and drug resistance. The GST-μ (GSTM) family is encoded by a tandem gene cluster on chromosome one; a whole gene deletion of the *GSTM1* has been found (Xue et al., 1998). The *GSTM1* enzyme impairment is thought to result in inefficient detoxification, which leads to genetic damages and increased disease risks (Sheehan et al., 2001; McIlwain et al., 2006) and response rates to some chemotherapy (Hayes and Pulford, 1995).

The *GSTM1* deletion homozygous genotypes have extensively been studied in various human populations from the viewpoint of epidemiology. The *GSTM1* deletion homozygous genotypes can be observed in various human population commonly (Garte et al., 2001; Gaspar et al., 2002; Buchard et al., 2007; Saadat, 2007; Fujihara et al., 2009; Piacentini et al., 2011). For

example, the prevalence of the *GSTM1* deletion homozygous genotype in Europeans, Asians and Africans was 47~57%, 42~54% and 16~36%, respectively (Garte et al., 2001). Such high frequency and these differences in the frequencies of the *GSTM1* deletion homozygous genotypes among human populations may have been because of an evolutionary advantage; however, the reasons (1) why has this deletion been maintained in the human populations at very high frequencies and (2) why does frequency of this deletion vary among populations remain unknown. The aim of this study is to reveal evolutionary force which shaped the distribution of the human *GSTM1* deletion from the viewpoint of biological anthropology.

In chapter 1, I investigated geographical distribution of the *GSTM1* deletion homozygous genotype distribution. The *GSTM1* deletion homozygous individuals showed higher sensitivity to UVB than in individuals with *GSTM1* wild-type allele. It is thus speculated that UV light irradiation was the selective pressure which facilitated relatively low frequencies of the *GSTM1* deletion homozygous genotype in Africans (Dandara et al., 2002). Dandara et al (2002) postulated that tropical populations adaptively maintained the *GSTM1* gene to protect their cells against oxidative stresses caused by strong UV irradiation. Meanwhile, little was known about the *GSTM1* genotype frequency among other populations residing in the tropics such as Southeast Asian and Oceanic populations. I collected the *GSTM1* deletion homozygous genotype data from 19 populations in Southeast Asia and Oceania, which were lacking in the previous studies and incorporated the data from experiments with published *GSTM1* deletion genotype data from 81 human populations. Comprehensive analysis of frequency of the *GSTM1* deletion genotype revealed the geographic distribution of this polymorphism of 81 populations in the previous studies and 19 populations in Southeast Asia and Oceania by the present study. It revealed that most Southeast Asian and Oceanic populations showed high frequencies of the *GSTM1* deletion homozygous genotype. There was no correlation between the *GSTM1* null

allele frequency and the absolute latitude in the worldwide populations contrary to the previously raised expectation. This non-latitudinal geographical pattern of the *GSTM1* deletion is thought to be attributed to human migration, genetic drift or adaptations, but is not due to adaptation to UV irradiation.

It has been technically difficult to test neutrality of whole gene deletion polymorphisms by comparing sequences of the wild-type allele with those of mutant alleles. In chapter 2, I calculated statistics for neutrality tests and analyzed haplotypes using the flanking regions of the *GSTM1* deletion, following the methods of Eaaswarkhanth et al. (2016), using the 1000 genome datasets. Using 1000 Genome datasets, I observed that LD between the *GSTM1* deletion and SNVs was moderately conserved only in CHB. The decay of LD beyond the *GSTM1* deletion suggests that gene conversions and recombinations broke the LD between the flanking SNVs and the deletion. However, this result does not disprove the possibility of recurrent deletion. For the Tajima's D, no difference were observed between the target regions which locate on the *GSTM1* deletion flanking regions and control regions. It suggests that the *GSTM1* deletion is neutral or LD between the deletion and target regions were too weak to catch signature of natural selection on the deletion. Haplotypes of the target regions were highly differentiated between East Asia and Africa. The East Asian-dominant SNVs on target regions significantly change other *GSTM* genes according to the GTEx-portal data. It is thus possible that in East Asia, the *GSTM1* deletion hitchhiked the East-Asian dominant haplotypes which experienced non-neutral evolution.

The *GSTM1* deletion allele is thought to have been generated by homologous recombination of two SDs (segmental duplications) (Xu et al. 1998). SDs have been defined as long (≥1–5 kb) and highly similar (≥90% similarity) sequences appear to have arisen by duplication (Bailey and Eichler, 2006). SDs account for 5% of the human genome (Bailey et al.,

2002). Since SDs can generate gene copy number variations by initiating non-allelic homologous recombination, they have been contributed gene family evolution in the primate genomes, in particular, in the ape genomes (Bailey and Eichler, 2006). SDs and gene copy number variations thus have had considerable impact on the primate genomic evolution. Recently developed genomic datasets make it possible to conduct comparative genomic analysis of various species to reveal genomic evolution. In chapter 3, I conducted *in silico* comparative genomic analyses among primates to reveal evolutionary history of the *GSTM* genes and SDs.

Comparative genomic analyses revealed the evolutionary history of the primate *GSTM* genes. The chimpanzee *GSTM5*, gorilla *GSTM4* and rhesus macaque *GSTM1* were pseudogenized. The nonhuman primate *GSTM1* genes were flanked by the two SD sequences as well as humans. It is thus suggested that the SDs are thought to be generated in the early stage of primate evolution. The nonhuman primate SDs may cause the *GSTM1* gene deletion.

In chapter 4, experimental comparative studies of the *GSTM1* region among primates were conducted. Chimpanzees have the *GSTM1* deletion allele as well as humans at polymorphic state. Sliding window analyses and phylogenetic analyses demonstrated that the human *GSTM1* deletion allele and chimpanzee *GSTM1* deletion allele were generated independently. Sliding window analyses also revealed break points of SD fusion and gene conversions in the human and the chimpanzee. A linage-specific sequence-swapping in chimpanzee lineage was also detected.

Studies on this thesis revealed the complexity of the evolutionary history of a gene deletion polymorphism which has biomedical importance: recurrent deletions in humans and chimpanzee, frequent gene conversions, and recombinations in the flanking SDs. The framework of this study can be expanded to studies on other copy number variations.