Molecular evolutionary and population genetic analysis of mental activity-related genes in humans

Hielim Kim

DOCTOR OF PHILOSOPHY

Department of Biosystems Science
School of Advanced Sciences
The Graduate University for Advanced Studies

2007

ACKNOWLEDGEMENTS

My first, and most earnest, acknowledgment must go to my supervisor, Prof. Yoko Satta, for providing me the opportunity to study at the Sokendai and her expert guidance and support in all respects. It is my best fortune to study with her in my life.

I would like to express my sincere appreciation and respect to Prof. Naoyuki Takahata for his critical comments and willing support for my research. His works and thoughts have made a great impact on my research.

I would like to thank the members of my thesis committee Prof. Tatsuya Ota, Mariko Hasegawa and Yoko Satta at the Graduate University for Advanced Studies (Sokendai) and Prof. Katsushi Tokunaga at the University of Tokyo for valuable comments and excellent guidance on my thesis.

My special thanks go to Prof. Andrew G. Clark at the Cornell University for his hospitality during my stay at the Cornell University. During the stay, I gained the great experiences in his laboratory and received many detailed and insightful comments on my published paper from him and his colleagues.

My deep gratitude goes to Prof. Yeongbin Song at Ewha womans University for his mental support and helpful advice, and Prof. Kyuyeong Song at the University of Ulsan whose introduction to genetics has led me where I am standing now.

I am grateful to many colleagues in Sokendai, particularly Drs. Hiromi Sawai, Mineyo Iwase, Satoko Kaneko, Tasuku Nishioka, Toshiyuki Hayakawa and Yasuhiro Go for their encouragements and considerations for my academic life. I also would like to thank all the members of Sokendai.

My sincerest appreciation goes to the Iwatani Naoji foundation for financial support. Without the support, this thesis and my Ph.D. course could never have been accomplished.

Many thanks to my precious friends, in particular Jinyoung and her little love, who will be born in May, for emotional support and love. I also thank Hitomi, who is my best friend in Japan, for her great friendship.

Finally, my most heartfelt acknowledgment must go to my family for their unconditional support at my every step along the way. I deeply thank my parents, pretty sister, and lovely aunt, Songi imo. Their encouragement and everlasting love make this thesis as well as me.

ABSTRACT

The aim of this thesis is to understand the human evolution, in particular mental activity of humans, and I have focused on genes related to sphingolipid (SL) metabolism. SL regulates neuronal developments by involving signal transduction. Some genetic disorders of SL metabolism (lipid storage disease) show typical symptoms of mental retardation and dysfunction of nervous system. In the evolutionary process of the acquisition of human specific mental activity, genes related to SL metabolism are likely to play important roles and to be candidate genes on which positive Darwinian selection operated.

To identify genes selected positively, the long range-haplotype test was applied to eight genes associated with lipid storage diseases using the HapMap data. The test shows that a particular haplotype of the *N-acylsphingosine amidohydrolase* (*ASAH1*; Acid ceramidase) gene has maintained stronger and longer linkage disequilibrium (LD) than haplotypes of simulated neutral genes. Positive selection has resulted in the spread of a selected variant in an ancestral population so rapidly that not enough time is there for recombination to decay the LD of the variant. Thus the result suggests that positive selection might have operated on the evolution of *ASAH1*.

To examine the evolution of *ASAH1* in the human population, I determined nucleotide sequences (~ 11 kb) of *ASAH1* from a world-wide sample of 60 chromosomes. In the strong LD region (SL region; ~ 4.4 kb) of the sequenced region, I found that two allelic lineages (V and M) have been maintained for 2.4 ± 0.4 million years (my) in the human population. Computer simulations suggest that the long persistence of the allelic lineages is likely to be attributed to population structure of humans in Africa before the Pleistocene period. The genetic diversity and the time to

the most recent common ancestor (TMRCA) of the other loci are compatible with the demographic history revealed by *ASAH1*. Therefore, it is speculated that each of the allelic lineages has persisted in each subpopulations in Africa and an admixture of two lineages has occurred by time of the dispersal of modern humans from Africa.

In addition, signatures of positive Darwinian selection for haplotypes belonging to the V lineage have been detected from the pattern and level of polymorphism of the two lineages. The haplotypes of the V lineage are predominant (62%) but have exhibited small nucleotide diversity (π = 0.05%), recent TMRCA (200 ~ 340 thousand years) and strong LD in the SL region. The diversity is significantly smaller in the SL region than the other regions but this reduction of diversity is not seen in the haplotypes of the M lineage. These observations are consistent with the rapid expansion of the haplotypes of the V lineage by positive selection. For the V lineage, I found that the Val residue at the 72nd amino acid residue, characteristic of the V lineage, is human specific among primates, suggesting that this Val residue could be a target of positive selection. No variation at the 659 bp region surrounding this Val residue in the V lineage is also consistent with this. Computer simulations with assuming various ancestral population-structures have confirmed that the observed small nucleotide diversity of the V lineage is not accounted for only by neutral evolution. From the above observations, it has been argued that positive selection has operated on the V lineage against to the M lineage since the “out of Africa” of modern humans. This is consistent with the archeological evidence supports the emergence of behavioral modernity of humans 70 ~ 80 kya in Africa.

Moreover, the subject of the study has been expanded to four genes possessing the domain of ceramidase activity besides the *ASAH1* gene. Phylogenetic analyses show that the origin of three kinds of ceramidase (acid, neutral, and alkaline) is prior to the

split of vertebrates and invertebrates. The amino acid sequences of five groups of genes have been highly conserved in each group, which is consistent with the reported functional differentiation among ceramidase in the pathway of ceramide metabolism. Further, I found that two *N-acylsphingosine amidohydrolase 2* (*ASAH2*) paralogs, *ASAH2B* and *ASAH2C* are expressed in humans only and was born by duplication in the human lineage. Interestingly, a previous study showed significant decrease of *ASAH2B* transcripts in the brain of Alzheimer's disease patients. This report has suggested a role of *ASAH2B* in brain and the gene should be an attractive target of further study with respect to the human evolution.

In this thesis, I have revealed the demographic history of human populations, the recent positive selection on *ASAH1*, and human specific genes, *ASAH2B* and *ASAH2C*. It is crucial for elucidatin of the human evolution that the comprehension of demographic history leading to modern humans and the human specific evolution of mental activity associated genes.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER
1. INTRODUCTION

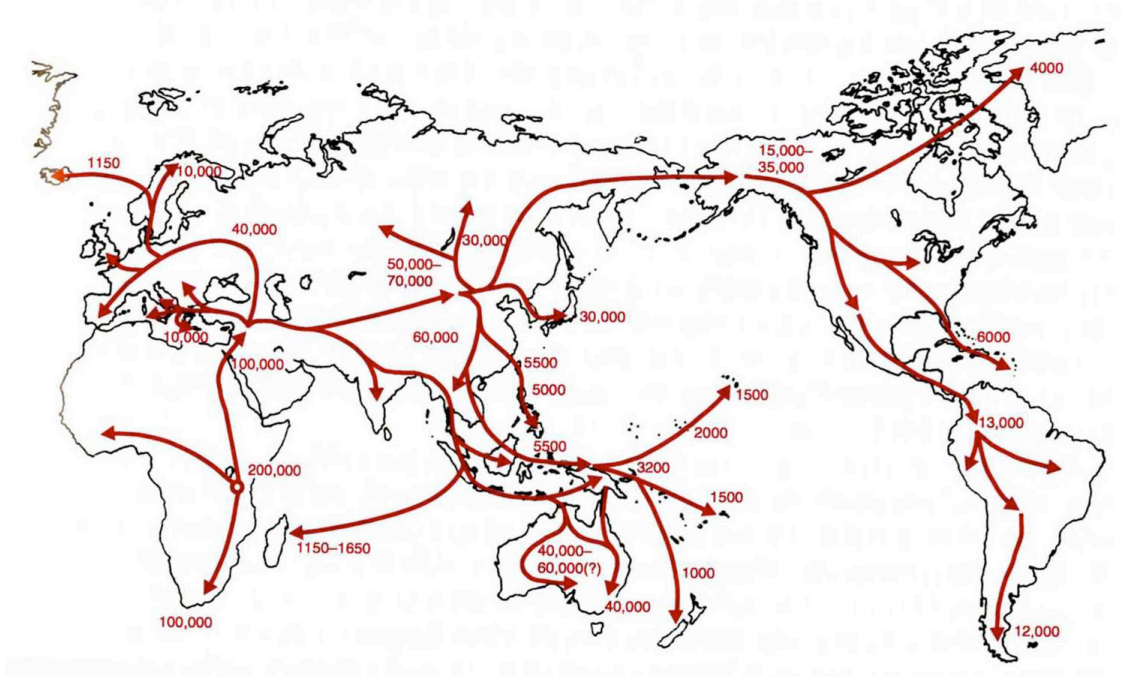
1.1. Mental activity of modern humans

H*omo sapiens* has evolved to adapt to new and diverse environments, showing rapid population expansion since the exodus from Africa, 60 ~ 80 thousand years ago (kya) (Figure 1-1) (Watson *et al*. 1997; Macaulay *et al*. 2005; Kivisild *et al*. 2006; Mellars 2006b). The explanation to the question, why and how did modern humans migrate from Africa to whole world is important to understand the evolution of human mental activity. The period of the "cultural explosion" of humans, the Middle/Upper Palaeolithic transition, is concordant with the time of "out of Africa". The archeological records reveal the emergence of technical and social advances such as symbolic behavior, cultural division, burials and complex technology in this period. At the Blombos cave in South Africa, the ornaments of shell beads, pieces of red ochre, and other crafted tools with dating ~ 75 kya were found (Henshilwood *et al*. 2002). Similar ornament of shell beads was found also at Morocco, North Africa and was dated to ~ 82 kya (Bouzouggar *et al*. 2007) (Figure 1-2). Personal ornamentation is sign of symbolic material cultures (d'Errico 2005; Mellars 2006b), and thus the shell beads have been the first evidence of human's symbolic behavior to date. These findings imply a wide-spread distribution of bead-making of symbolic behaviors in Africa before 80 kya and may suggest existence of exchanging systems or long-distance social networks at this early period (Bouzouggar *et al*. 2007). Besides, the Upper Paleolithic technology of tool-making was dramatically changed so that not only new and specialized tools requiring a high level of skill was invented but also new materials, such as bone, ivory,

and antler was used (Klein 1999). Further, the earliest cave art and sculptures have been excavated from in Europe and were dated about 30 kya. Therefore, humans attained modern behavior recently, just before the dispersal to whole world in Africa.

Several anthropologists have noticed the coproduct of population spreading and cultural transition, suggesting the acquisition of human specific mental activity is a possible driving force for subsequent dispersal around the world (Klein 1999; Mellars 2006b). In the process of modern human evolution, it is likely that some genes, especially those related to mental activity of modern human specificity, have evolved under natural selection (Nei 1983; Klein 2000). Recent studies have reported positively selected genes for mental activity and/or brain development in the human lineage: *Abnormal spindle-like microcephaly associated* and *Microcephalin* in relation to brain size (*ASPM*: Zhang 2003; *MCPH1*: Evans *et al*. 2004; Wang and Su 2004, but see Currat *et al.* 2006; Yu *et al.* 2007), *Dopamine receptor D4* and *Monoamine oxidase A* in relation to emotional activity (*DRD4*: Ding *et al*. 2002; *MAOA*: Gilad *et al*. 2002), *Forkhead box P2* in relation to language (*FOXP2*: Enard *et al*. 2002; Zhang *et al*. 2002), and *Spinocerebellar ataxia type 2* and *Pituitary adenylate cyclase-activating polypeptide* in relation to neurodegenerative disorders (*SCA2*: Yu *et al*. 2005; *PACAP*: Wang *et al*. 2005). In addition, many causal genes for several types of mental retardation, possibly related to brain and cognitive development, have recently been reported (Inlow and Restifo 2004; Schumacher *et al*. 2007; Mervis and Becerra 2007). These genes are further thought to influence mental activity. However, no evidence supports that emergence of human modernity is derived from human specific evolution of genes related to mental activity, not yet.

Figure 1-1. The out of Africa of humans



The figure is cited from Klein and Takahata 2002 Fig. 10.10. The number is the estimated dates of arrival of modern humans at the sites and all dates are before present.

Figure 1-2. Geographic distribution of shell beads in Africa

The map was retrieved from the google map (http://maps.google.co.jp/). Red circles indicate sites where shell beads were found in Africa. The pictures were downloaded from web sites (http://www.svf.uib.no/sfu/blombos/; http://www.nhm.ac.uk/).

1.2. Lipid storage diseases

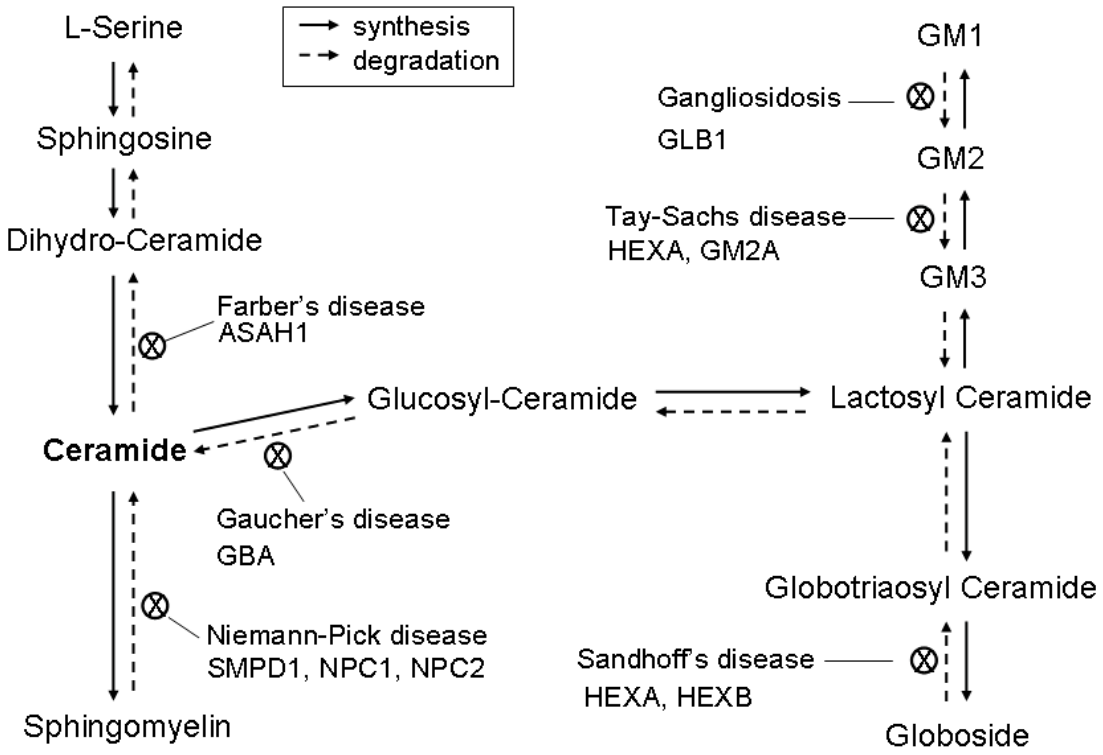
Among genes related to human mental activity, interest lie in sphingolipid (SL) metabolism associated genes. The inherited deficiency of these genes result in the lipid storage diseases (also known lysosomal storage diseases; LSDs) such as Gaucher, Tay-Sachs, Farber and Niemann-Pick diseases and Sandhoff's disease (Figure 1-3). The deficiency causes intra-lysosomal accumulation of unmetabolized SLs, and common symptoms are mental retardation and dysfunction of nervous systems (Futerman and Meer 2004).

LSDs are known well as highly frequent genetic disease among the Ashkenazi Jews, even though many LSDs are lethal. For example, the frequency of Gaucher disease is 0.028 and it is most common LSD in the Ashkenazim (Beutler and Grabowski 2001). Because of this unusual high frequency and the wide geographic range of disease-associated alleles, several studies have suggested that natural selection has favored heterozygous carriers of alleles affecting LSD (Rotter and Diamond 1987; Diamond 1994; Motulsky 1995; Cochran *et al.* 2006). Besides, Cochran *et al.* claimed that the heterozygotes at the LSD loci associated with high IQ since SL accumulations in neuron promote axonal growth and branching (Schwartz et al. 1995; Walkley et al. 2000; Walkley 2003), which appear to be a necessary step in learning (Holloway 1966; Leuner et al. 2003). In the other hand, the founder effect has been suggested to account for the frequent presence of LSD in the Ashkenazi Jews (Risch *et al.* 1995; Goldstein *et al.* 1999; Niell *et al*. 2003; Risch *et al.* 2003; Frisch *et al.* 2004). Risch *et al.* (2003) compared the frequency and geographic distribution of disease-associated alleles between LSD and non-LSD in the Ashkenazi Jewish population. There is no significant difference between two groups of diseases. The observations do not support natural selection for LSD-associated alleles and suggest that the high frequency of LSD in the

population results from genetic drift, such as founder effect. This is also examined by a statistical test (Slatkin 2004).

Though natural selection for LSD-associated genes remains unclear, it could be subjected to associations of LSD with neuronal functions and mental activity in humans.

Figure 1-3. The pathway of sphingolipid metabolism and lipid storage disease



The figure illustrates the pathway of sphingolipid metabolism. Circled X means a metabolism deficiency due to mutations of gene related to the metabolism. A disease name and disease-associated genes are indicated with the circled X.

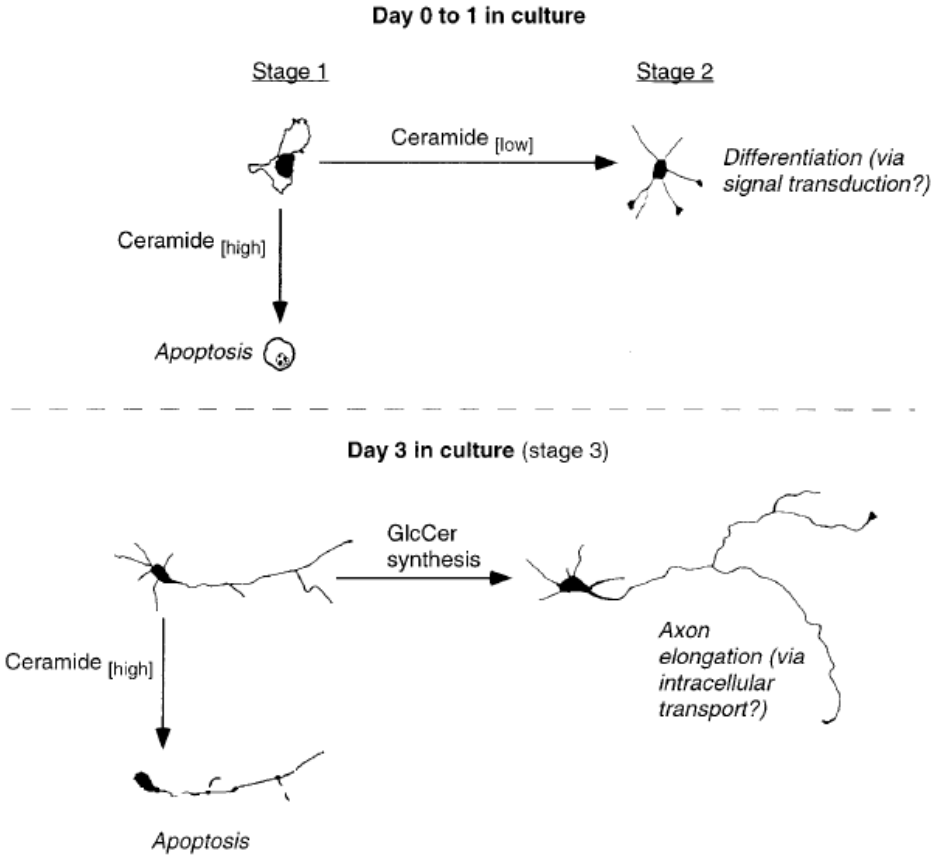
1.3. Sphingolipids and ceramides

Sphingolipids (SLs) are ubiquitous components of eukaryotic cell membranes and have important roles in cell membranes. The molecular structure of SLs is relatively stiff compared to other membrane lipids and then concentrated SLs in membranes result in a lateral organization of the cell membranes, named lipid rafts (Simons and Ikonen 1997). This membrane domain serves the clustering of intracellular signaling molecules and receptors, thus, facilitate, signal transduction (Bollinger *et al.* 2005). SL is extensive complex molecule, and around 500 different carbohydrate structures have been described in glycolsphingolipids. This diversity favors a variety of signaling pathway, including SLs as first and second messengers. The pathway mediates cell apoptosis, differentiation, and growth (Futerman and Hannun 2004; Futerman and Riezman 2005). The biochemical pathway of synthesis and degeneration of SLs have been well determined (Hannun and Luberto 2004), and this metabolism modulates various cellular responses (Futerman and Riezman 2005). Further, SLs are enriched in neuronal plasma membranes and highly relates to neuronal development such as neurite outgrowth (Schwarz and Futerman 1997; Buccoliero *et al*. 2002; Buccoliero and Futerman 2003; Ruvolo 2003).

In particular, ceramide is the simplest SL and is the hub of SL metabolism and serves as the first point of significant SL accumulation in the *de novo* pathway (Hannun and Obeid 2002; Merrill 2002). This pathway involves acid ceramidase (ASAH1; also known as *N*-acylsphingosine amidohydrolase; ASAH; AC, MIM# 228000, EC 3.5.1.23) (Rother *et al*. 1992), which hydrolyzes ceramides into sphingosines and free fatty acids (Gatt 1963). Ceramides and sphingosines of the product of acid ceramidase are both important cell signaling molecules (Spiegel *et al.* 1998; Pyne 2002; Ruvolo 2003).

In neuronal cells, ceramides play some roles at different stages of neuronal development (Figure 1-4; Schwarz and Futerman 1997) and are involved in the regulation of neuronal proliferation, differentiation, survival and apoptosis (Buccoliero and Futerman 2003; Ruvolo 2003). An inherited deficiency leads to accumulation of ceramides in various tissues, resulting in Farber disease, also known as Farber lipogranulomatosis (Sugita *et al*. 1972). Farber disease is a rare disorder with an autosomal recessive mode of inheritance. Typical symptoms include painful swelling of the joints, hoarseness, and premature death, and, depending on the tissues affected by the storage of ceramides, severe nervous system dysfunction (Moser 1995). Several mutations for Farber disease have been reported; one single nucleotide deletion (V96del: Muramatsu *et al*. 2002) and nine single nonsynonymous mutations (Y36C, V97E, E138V, L182V, T222K, G235R, R254G, N320D, and P362R: Bar *et al*. 2001; Koch *et al.* 1996; Li *et al.* 1999; Muramatsu *et al*. 2002, Devi *et al.* 2006). Besides, Huang et al. (2004) reported that ASAH1 encodes acid ceramidase associated with Alzheimer's disease. They showed that the level and activity of ASAH1 elevates in Alzheimer's disease brain and suggested that ASAH1 might play a role in controlling neuronal apoptosis mediated by signaling pathway and be involved in the molecular mechanism of Alzheimer's disease.

Figure 1-4. Roles of ceramides in neuronal cells



The figure is taken from Schwarz and Futerman (1997) Figure 10.

Demonstrated are the roles of ceramide and Glycosylceramide (GlcCer) at different stages of neuronal development.

1.4. Ceramidase

A ceramidase gene family consists of five members, *ASAH1*, *ASAHL*, *ASAH2*, *ASAH3*, and *ASAH3L*. Among them, there are three types of ceramidases (EC 3.5.1.23) described to date. These are classified as acid, neutral, and alkaline ceramidases according to the pH range that supports their optimum enzymatic activity (Nikolova-Karakashian and Merrill Jr. 2000). They are encoded by *ASAH1*, *ASAH2*, and *ASAH3*, respectively. An acid ceramidase is encoded by *ASAH1* gene and pH for the optimal activity is about 5. The catalysis of acid ceramidase has been detected in lysosomes which contain hydrolyzing enzymes that are active at the acidic pH of this organelle. ASAH1 was known to be the main hydrolyzing enzyme of the biosynthetic pathway of ceramides. It was known that the amino acid sequence of *ASAH1* is most similar to *N-acylsphingosine amidohydrolase-like* (*ASAHL*) gene on chromosome 4 (Hong *et al* 1999). However, the roles of *ASAHL* have not been known.

N-acylsphingosine amidohydrolase 2 (*ASAH2*) encodes neutral ceramidase and the catalysis by this enzyme was detected in plasma membrane and mitochondria (Tani *et al.* 2000). Neutral ceramidase is an integral membrane protein and the mammalian neutral ceramidase is highly expressed in the small intestine. Null mutants of for *ASAH2* in mice have a normal life span but show limited abnormalities in the intestinal degradation of ceramides (Kono *et al.* 2006). This suggests that *ASAH2*-encoded neutral ceramidase is a key enzyme for the catabolism of dietary ceramides (Kono *et al.* 2006; Nilsson and Duan 2006). An alkaline ceramidase is encoded by the *N-acylsphingosine amidohydrolase 3* (*ASAH3*) gene. The amino acid sequence of the *N-acylsphingosine amidohydrolase 3-like* (*ASAH3L*) gene is highly similar to *ASAH3.* It has been unknown about the functions of two genes. *ASAH3L* ESTs are detected in plasma membrane.

CHAPTER
2. MATERIALS AND METHODS

2.1. The long-range haplotype test

The long-range haplotype (LRH) test (Sabeti *et al*. 2002) for eight LSD-associated genes was conducted using the HapMap Project data, which was released in June 2006 (http://www.hapmap.org) (The International HapMap Consortium 2005; The International HapMap Consortium 2007). The eight genes are *ASAH1, Glucosidase acid beta* (*GBA*), *ß-Galactosidase 1* (*GLB1*), *Ganglioside GM2-activator* (*GM2A*), *Hexosaminidase A* (*HEXA*), *Hexosaminidase B* (*HEXB*), *Niemann-Pick disease type C1* (*NPC1*), and *Niemann-Pick disease type C2* (*NPC2*) (Table 2-1). For each of Yoruba in Ibadan from Nigeria (YRI) and Northern and Western Europeans in Utah (CEU), the HapMap data of 60 unrelated individuals were analyzed. For Han Chinese from Beijing (CHB) and Japanese from Tokyo (JPT), similar data of 45 and 44 unrelated individuals were used, respectively. In order to estimate haplotype phases accurately, chromosomes with ≥ 10% missing genotypes (undetermined genotypes) were excluded. The haplotype phase of these trimmed HapMap data was estimated using the program PHASE v2.1 (Stephens *et al*. 2001; Stephens and Donnelly 2003). After the phase estimation, haplotypes with a low probability were further excluded. Thus the total number of single nucleotide polymorphisms (SNPs) and chromosomes used in the LRH test ranges from 83 to 100 and 76 to 120, respectively, as given in Table 2-2. The extended haplotype homozygosity (EHH) and relative-EHH (REHH) in ~ 200 kb surrounding specified genomic regions of interest were measured using the software Sweep.1.0 (Sabeti *et al*. 2002).

The significance of the LRH test results was examined with the simulation program *ms* (Hudson 2002). In the simulation, neutral polymorphism data in a sample of 120 DNA sequences, each of which ~ 200 kb length, were generated without recombination to make the test conservative. One hundred segregating sites were randomly sampled so as to imitate the actual data for the eight LSD-associated genes (Table 2-2). One thousand replications were carried out for each of the eight genes. For each gene in each population, the observed and simulated EHH and REHH were compared within the bin that contained haplotypes of the same frequency. The standard deviations of the observed values from the mean in their bin were calculated using the EHH significance calculator option of Sweep 1.0.

Table 2-1. List of eight genes applied for the LRH test

Gene Symbol	Description	Chr. arm	Molecular function	Clinical disorder
ASAH1	*N*-Acylsphingosine amidohydrolase	8p	Hydrolase	Farber disease
GLB1	β-Galactosidase 1	3p	Hydrolase	GM1-gangliosidosis
GBA	Glucocerebrosidase	1q	Hydrolase	Gaucher disease
GM2A	Ganglioside GM2-activator	5q	Enzyme regulator	Tay-Sachs disease, AB variant
HEXA	Hexosaminidase A	15q	Hydrolase	Tay-Sachs disease, Sandhoff disease
HEXB	Hexosaminidase B	5q	Hydrolase	Sandhoff disease
NPC1	Niemann-Pick disease type C1	18q	Transporter	Niemann-Pick disease
NPC2	Niemann-Pick disease type C2	14q	Transporter	Niemann-Pick disease

Table 2-2. The number of chromosomes and SNPs used in the LRH test of eight LSD-associated genes

	Gene symbol	Populations				Gene symbol	Populations			
		YRI	CEU	CHB	JPT		YRI	CEU	CHB	JPT
	ASAH1					***HEXA***				
n_{ch}[a]		118	116	78	82		120	120	90	88
n_{snp}[b]		96	96	96	99		100	100	100	100
	GBA					***HEXB***				
n_{ch}		116	120	90	88		118	116	-	82
n_{snp}		83	93	97	95		95	88	-	90
	GLB1					***NPC1***				
n_{ch}		-	110	78	-		118	114	90	88
n_{snp}		-	93	97	-		99	100	100	100
	GM2A					***NPC2***				
n_{ch}		112	120	90	82		114	114	-	76
n_{snp}		100	99	94	95		98	100	-	100

[a]: Number of chromosomes

[b]: Number of SNPs

YRI: Yoruba in Ibadan from Nigeria

CEU: Northern and Western Europeans in Utah

CHB: Han Chinese from Beijing

JPT: Japanese from Tokyo

2.2. DNA samples, PCR and sequencing

The 30 human genomic DNA samples used in this study come from 15 Africans (10 Pygmies, 2 African Americans, and 3 Yoruba) and 15 non-Africans (4 Amerinds, 5 Europeans, and 6 Asians). The repository numbers of these samples in the Coriell Cell Repositories are NA10470 – 10473, 10492-10496, 10469, 10965, 10970, 10975, 11197, 11322, 11324, 11373, 11521, 11587, 13597, 13607, 13617 – 13618, 13820, 13838, 14537, 14661, 18523, 18853, and 19208.

ASAH1 gene is located on the short arm of chromosome 8 (8p22 - p21.3), is ~ 28.5 kb long, appears to be a single-copy gene, and encodes 14 exons. Depending on the splicing pattern, *ASAH1* is translated into either 395 or 411 amino acids (NP_808592 and NP_004306). PCR was used to amplify the part of *ASAH1* (~ 12.5 kb), ranging from chromosome position 17969623 to 17982155 (NCBI build 36.2). The primers were designed using the program Primer3 (Rozen and Skaletsky 2000) and are given in Table 2-3. PCR was performed with 4 pmol of each primer, 150 ng of human genomic DNA, 0.2 mM dNTPs, 0.7 µl of Elongase@Enzyme Mix (Invitrogen, Carlsbad, CA) and 4 µl of PCR buffer containing 1.9 mM $MgCl_2$ in a total volume of 20 µl. A RoboCycler Gradient 96 (Stratagene, La Jolla, CA) and TGradient (Whatman Biometra, Goettingen) were used under the following conditions depending on primer pairs: denaturation at 94° for 2 min followed by 40 amplification cycles of 94° for 30 sec, 55 – 59° for 30 sec and 68° for 10 min, and ending with extension at 68° for 20 min. The amplified products were purified using ExoSAP-IT (United States Biochemical, Cleveland, CA) and sequenced directly. Except for repeated sequences and nucleotides with low quality peaks, the ~ 11 kb region was used for subsequent analyses. Sequencing reactions were performed using BigDye® Terminator v1.1 or v3.1 Cycle Sequencing Kits (Applied Biosystems, Foster City, CA) and analyzed on an ABI

PRISM 377, 3100, or 3730 DNA sequencer (Applied Biosystems). To avoid sequencing errors, the PCR products were read at least twice in both directions.

Table 2-3. Nucleotide sequences of PCR primers

Forward		Reverse	
1F	5’-ttcagcatttctttcccactg-3’	1R	5’-atttgctataaaacacacagtctgc-3’
2F	5’-tgagattcgtgaattttaagca-3’	2R	5’-tgctttgaaatctagttcttgga-3’
3F	5’-gcctgcgtatggctataaatactaa-3’	3R	5’-ctctcatactctgctgattgtcaac-3’
4F	5’-acttcattgtaatcccttcattgtc-3’	4R	5’-tgcctggtctgtactaccattt-3’
5F	5’-gaaaactgtttaacgagaggtaaaa-3’		

2.3. Population genetic analyses

From the DNA sequence data of the ~ 11 kb region at *ASAH1*, haplotype phases were inferred using the program PHASE v.2.1 (Stephens *et al*. 2001; Stephens and Donnelly 2003) and fastPHASE1.0.1 (Scheet and Stephens 2006). All estimated haplotypes were used for further analyses. The nucleotide diversity (π) (Nei and Li 1979) and Tajima's *D* (Tajima 1989) were computed, and the HKA test (Hudson *et al.* 1987) was applied using the program DnaSP4.10 (Rozas *et al*. 2003). A gene tree for the strong LD region (after excluding two possible recombinants) was constructed and the time to the most recent common ancestor (TMRCA) was estimated using the software Genetree (Griffiths and Tavaré 1995), assuming the effective population size (N_e) of 10^4 and the generation time (g) of 20 years (Takahata 1993; Klein and Takahata 2002). To estimate the TMRCA in an alternative way, the average p distance between the V and M lineages (p_B) was used. The chimpanzee sequence was downloaded from the Ensemble database (ENSPTRT00000037110). The average per-site pairwise distance (d = 0.0174 ± 0.0003) between 60 human and one chimpanzee chromosomes was calculated in the strong LD region of ~ 4.4 kb using the program MEGA 3.1 (Kumar *et al*. 2004). Under the assumptions that humans and chimpanzees diverged 5 – 7 million years (my) ago, the between-lineage TMRCA was estimated as $(p_B/d) \times 5 - 7$ my. The average nucleotide difference between the root haplotype and every individual in the sample was calculated in each of the V and M lineage. The ratio of these differences to d was then used to estimate the within-lineage TMRCA.

2.4. Computer simulations

In the computer simulations, various models of a panmictic population with bottleneck or expansion as well as those of both recent and ancient structured

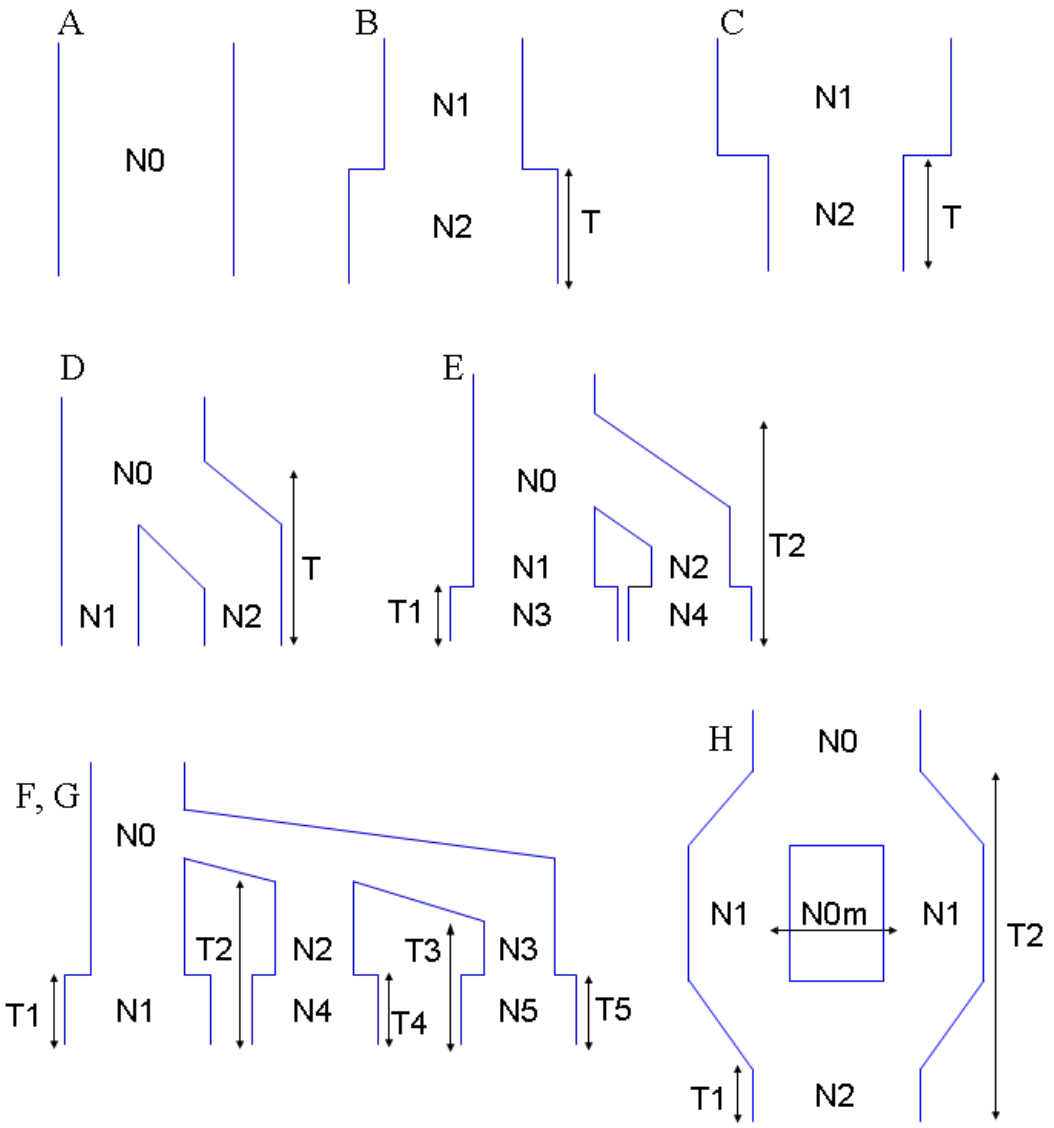
populations were assumed. Under the assumption of selective neutrality, the software *ms* (Hudson 2002) efficiently evaluated the extent of neutral variation such as the π value. 22 different sets of demographic parameters were examined with 50,000 replications for each (Figure 2-1 and Table 2-4) by commonly specifying the following parameters: the number of chromosomes, 60; the number of segregating sites, 49 (the observed value in the region of strong LD); and the generation time (g), 20 years. The other parameters for models followed the previous studies (Takahata 1995; Marth *et al.* 2004; Voight *et al.* 2005; Williamson *et al.* 2005).

To evaluate the effect of an advantageous mutation on the pattern of polymorphism in comparison with neutral cases, a forward simulation was also carried out (Figure 2-2). An ancestral sequence of L = 1000 base pair length was generated at random. Two hundred copies ($2N_0$ = 200) of the sequences evolve according to a finite site model of neutral mutations and random sampling, and this process was repeated till equilibrium ($6N_0$ generations). Each of the two subpopulations (popS and popN) was composed of these 100 sequences ($N_0 = 2N_1 = 100$, where N_1 is the size of subpopulation). To trace a particular lineage (N lineage) in a subpopulation, a single mutation for labeling is introduced into a single gene in the popN. It was assumed that neutral mutations scaled by N_0 occur at a rate of $N_0L\mu = 0.4$ per generation, where μ stands for the neutral mutation rate per site per generation, followed by migration (at a rate of N_0m = 0.1, 0.01, and 0.001 where m is the migration rate per gene) and random sampling of N_0 sequences for the next generation. No recombination was assumed. Repeating this process for additional $10N_0$ generations, two subpopulations admix with each other and the two become a single panmictic one. At $0.25N_0$ generations before this unification, a single advantageous mutation of $N_0s = 50$ was introduced into a non-N lineage in the popS and a new lineage (S lineage) is generated. The time of $0.25N_0$

generations is long enough for the mutation to fix in the popS. The average fixation time of a single advantageous mutation with $N_0s = 50$ within a subpopulation is ~ $0.2N_0$ generations (Takahata 1991). Simulation was terminated when the frequency of the S lineage in the entire population reached ~ 0.7. Then the π values within the S and N lineage and in the entire population were measured in each replication and ~ 100 such π values were collected.

For a neutral case, the population size of N_0 was set to be 20, because the simulation of the same scheme takes enormously long time to collect sufficient number of data. The simulation was ceased at $0.1N_0$ generations after the unification of two subpopulations. At any segregating site, the lineage with a nucleotide of its frequency ~ 0.7 is defined as the S lineage, while if the frequency is ~ 0.3 the lineage with the nucleotide is defined as the N lineage. The π values within the S or N lineage and in the entire population were calculated in the same way for the case of selection, and ~ 1000 π values within each lineage were collected. Although the population size appears too small, the result of simulations shows a good agreement with the theoretical expectations.

Figure 2-1. The demographic models used in the simulations

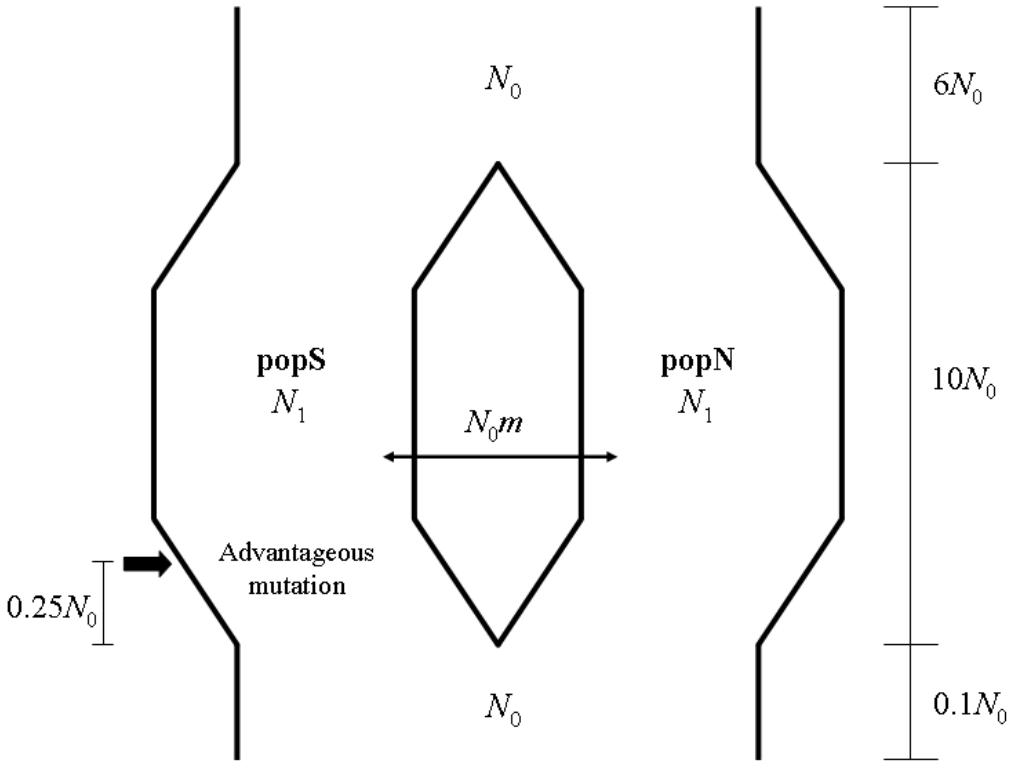


A. Constant population size, B. Population expansion, C. Ancient bottleneck, D. Recent population-structure, E. Recent population-structure and expansion, F. Recent population-structure and bottleneck, G. Recent population-structure and bottleneck/expansion, H. Ancient population-structure. Parameters applied the models are shown in the Table 2-4.

Table 2-4. The parameters of the demographic models

Model	Parameters
A	$N0=10^4$
B	N1=8211, N2=51340, T=18160 (Williamson *et al.* 2005)
C	$N2=10^4$, T=2my
1	$N1=10^5$
2	$N1=10^6$
D	$N0=10^4$, T=60ky, N1+N2=N0
1	N1/N2=1/9
2	N1/N2=2/8
3	N1/N2=3/7
4	N1/N2=4/6
5	N1/N2=5/5
E	$N0=10^4$, N1+N2=N0, N1/N3=N2/N4=5, T1=20ky, T2=60ky
1	N1/N2=1/9
2	N1/N2=2/8
3	N1/N2=3/7
4	N1/N2=4/6
5	N1/N2=5/5
F	N0=10659, N1=22565, T1=20ky, T2=80ky, T3=40ky
1	N2=N3=N0*0.1, N2/N4=N3/N5=0.1, T4=60ky, T5=32ky
2	N2=N3=N0*0.2, N2/N4=N3/N5=0.2, T4=40ky, T5=22ky
3	N2=N3=N0*0.3, N2/N4=N3/N5=0.3, T4=28ky, T5=14ky(Voight *et al.* 2005)
G	$N0=10^4$, N1=18000, N2=3000, N3=2000, N4=25000, N5=20000
	T1=150ky, T2=76ky, T3=70ky, T4=64ky, T5=60ky (Marth *et al.* 2004)
H	$N0=10^4$, N1=N0 / 2, N2=N0, T1=20ky, T2=2my
1	N0m=0.1
2	N0m=0.01
3	N0m=0.001
4	N0m=0 (Takahata 1995)

Figure 2-2. The ancient population-structure model for the forward simulations



N_0 is the population size of panmictic state, and m is rate of migration per gene in a generation. N_1 is the population size of subdivided population. In the present model, $N_1 = N_0/2$. $10N_0$ and $0.1N_0$ generations represent the time of random mating and that of partial isolation, respectively. An advantageous mutation is introduced in to popS at $0.25N_0$ generations before the admixture. Without selection, the simulation is terminated at $0.1N_0$ generations after the admixture. With introduction of advantageous mutation, when the frequency of the mutation (= the frequency of the S lineage) reaches to 0.65 ~ 0.75, the simulation is terminated.

2.5. Phylogenetic and dot-matrix analyses

Ceramidase DNA sequences in various species were retrieved from the NCBI and Ensembl Databases. Some sequences were found by using the BLAST program (Altschul *et al.* 1990) in the database of all assemblies of genome. The accession numbers of the sequences used are shown in Table 2-5. These nucleotide sequences were aligned by Clustal X (Thompson *et al*. 1997) and the alignments were further checked manually. All neighbor-joining (NJ) trees (Saitou and Nei 1987) were constructed by MEGA 3.1 (Kumar *et al*. 2004).

For the dot-matrix analysis, humans and chimpanzees genomic sequences were downloaded from the NCBI databases. The chromosome positions of human nucleotide sequences are chr.10: 46-46.3Mb, 51.4-51.7Mb and 52.1-52.3Mb, and those of chimpanzee sequences are chr.10: 45.9-46.3Mb, 48.9M-49.2Mb, 49.5-49.6Mb. The analysis is performed by software Dotter with the default conditions (Sonnhammer and Durbin 1995).

Table 2-5. The accession numbers of nucleotide sequences

	ASAH1	*ASAHL*	*ASAH2*	*ASAH3*	*ASAH3L*
Human	NM_177924	ENST00000286733	ENST00000374028	ENST00000301452	NM_001010887/NT_005403
Chimpanzee	ENSPTRT00000037110	ENSPTRT00000030115	ENSPTRT00000004678	ENSPTRT00000019038	XM_520508
Macaque	ENSMMUT00000021738	ENSMMUT00000024604	ENSMMUT00000009882	ENSMMUT00000008001	ENSMMUT00000024981
Mouse	NM_019734	ENSMUST00000031361	ENSMUST00000096119	ENSMUST00000056113	NM_139306
Rat	NM_053407	ENSRNOT00000003102	ENSRNOT00000016688	−	XM_233138
Cow	ENSBTAT00000014960	−	−	−	−
Dog	ENSCAFT00000039154	−	−	−	−
Opossum	ENSMODT00000024178	ENSMODT00000024415	ENSMODT00000020951	ENSMODT00000002389	ENSMODT00000005127
Platypus	ENSORLT00000011543	−	ENSOANT00000007514		−
Chicken	NM_001006453	ENSGALT00000018799	ENSGALT00000005958	ENSGALT00000002417	XM_424820
X.tropicalis	ENSXETT00000047334	ENSXETT00000057513	ENSXETT00000014491	ENSXETT00000047334	NM_001017116
Zebra fish	ENSDART00000040434/ ENSDART00000017122	−	ENSDART00000027793	ENSDART00000023797	XM_001339790
Fugu	ENSSTOT00000002346	NEWSINFRUT00000151462	SINFRUT00000148873	−	SINFRUT00000141610
Tetraodon	−	GSTENT00014088001	−	GSTENT00019254001	−
C.savigny	−	−	ENSCSAVT00000007875/ ENSCSAVT00000009623/ ENSCSAVT00000014795	−	SINCSAVESTT00000008790
C.intestinalis	ENSCINT00000004990	ENSCINT00000009949	ENSCINT00000006321 ENSCINT00000006462 ENSCINT00000011274	ENSCINT00000022025	ENSCINT00000000094
C.briggsae	−	contig cb25.fpc0010	−	−	−
C.elegans	F27E5.1 / K11D2.2	Y55D5A.3	−	W02F12.2:W02F12.2	−
Aedes	−	−	AAEL007028-RA/ AAEL007030-RA	−	AAEL001645-RA
Anopheles	−	−	AGAP000973-RA/ AGAP003730-RA	−	AGAP008729-RA
Fruit fly	−	−	CG1471-RE	−	CG13969-RA
Yeast		−	−	YBR183W / YPL087W	−
D.discoideum	DDB0190948/DDB0184555	−	dcd2A / dcd2B	−	−
A.thaliana		−	AT5G5898 / AT1G07380 / AT2G38010	−	−

CHAPTER
3. THE LONG-RANGE HAPLOTYPE TEST FOR THE LSD-ASSOCIATED GENES

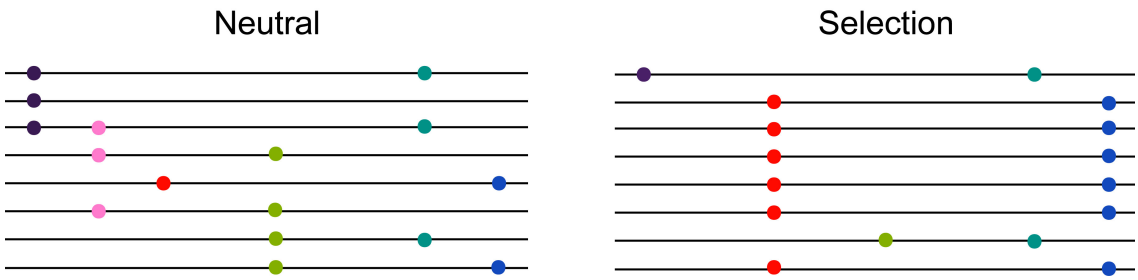
3.1. The purpose of the long-range haplotype test

To identify genes selected positively during human evolution, we focused on eight LSD-associated genes whose deficiency clearly results in symptoms of mental retardation. The LRH test (Sabeti *et al*. 2002) was applied to these genes using the HapMap data of four populations YRI, CEU, CHB, and JPT (The International HapMap Consortium 2005).

The test detects signatures of positive selection, *i.e.* selective sweep. Under neutral evolution, new variants require a long time to reach high frequency in a population, and linkage disequilibrium (LD) associated with the variants will decay substantially during the time due to recombination (Kimura 1983). Positive selection, however, results in rapid spread of a selected variant within a population such that recombination does not have enough time to break down the haplotype with the variant (Figure 3-1). Thus a haplotype under positive selection maintained unusually long-range LD with its frequency in a population (Aquadro *et al.* 1994; Sabeti *et al*. 2006). To detect the signature, the LRH test measures the extended haplotype homozygosity (EHH) and relative-EHH (REHH) using the SNP genotyping data. The test first selects a 'core region' is a zone with strong LD and then calculates the EHH at a distance x from the core region. The EHH is defined as the probability that two randomly chosen chromosomes sharing the haplotype in the core region are identical by descent (Sabeti *et al.* 2002). However, local variation of recombination rates is known in the human genome (Myers *et al.* 2005) and then the EHH does not reflect this variation. To correct

for the local variation, the REHH is calculated as the ratio of the EHH of tested core haplotype to the combined EHH of all other core haplotypes. The EHH of neutral haplotype is lower than that of haplotype possessing advantageous allele. Thus high EHH and REHH represent the transmission of a haplotype with long-range LD and high frequency compared to haplotypes under neutral evolution. Consequently, the LRH test detects signatures of recent positive selection.

Figure 3-1. The pattern of polymorphism with and without selection in a population

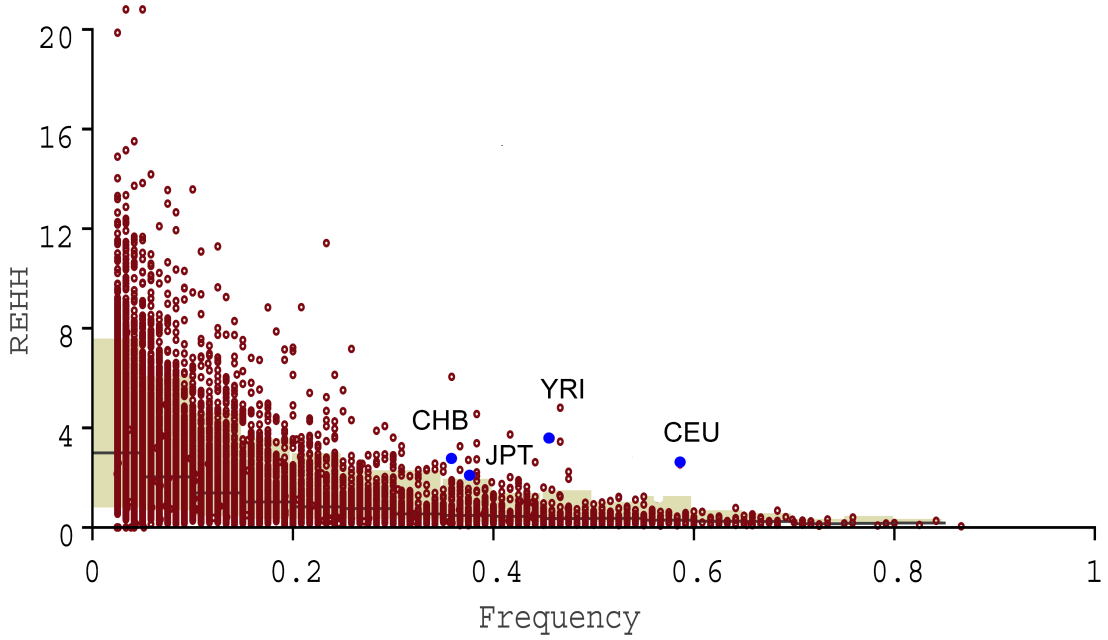


This is an example of the pattern of polymorphism. A line is a chromosome and a colored spot is polymorphism as a SNP. Left figure represents variation of a population under neutrality, whereas right figure expresses that under positive selection. In the selection, the red allele is advantageous to other alleles and subsequently spreads rapidly in a population. The blue allele is linked with the red one by chance and spreads simultaneously. Thus we could observe strong LD in a region operated positive selection, even in a long range.

3.2. The result of the long-range haplotype test

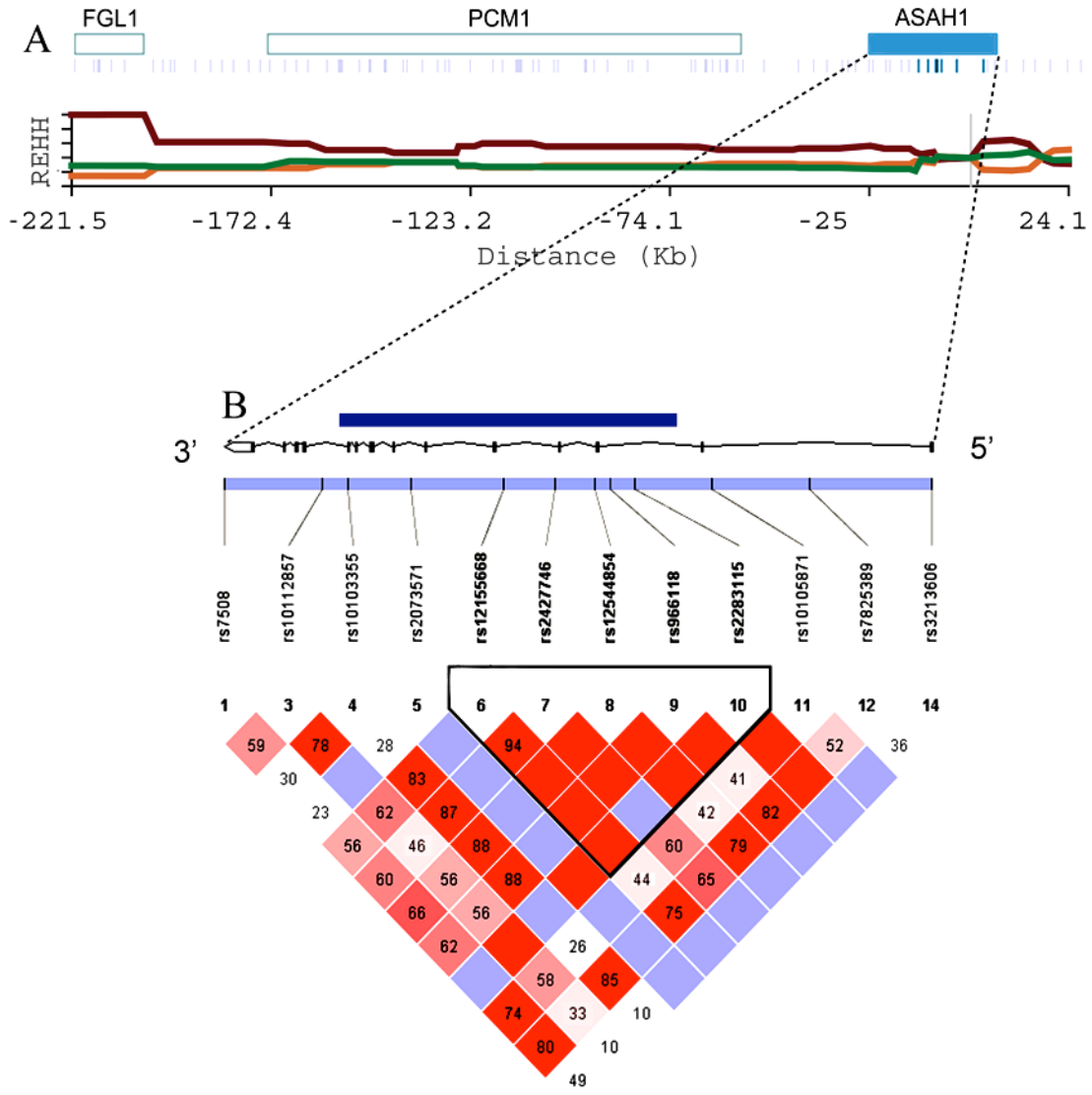
For applying the LRH test, the core region of each of the eight LSD-associated genes is selected separately in YRI, CEU, CHB, and JPT. For each gene, the core region turns out to be almost the same in the four populations, or the pattern of LD does not differ among populations. Both the EHH and the REHH of each haplotype in the core region (core haplotype) were measured in either the ~ 200 kb region upstream or downstream from the core region and compared with data simulated under a neutral model without recombination (Materials and Methods). Among these 16 surrounding regions of the eight genes, a particular core haplotype including *ASAH1* shows a significantly high EHH and REHH in the ~ 200 kb downstream region for all four populations ($P < 0.05$: Figure 3-2, Figure 3-3A, and Table 3-1). The EHH and REHH of core haplotypes in *GBA* are also significantly high in three populations, but the haplotype sequences are not consistent to each other across populations. The observation might be understood through that bottleneck effect or local selection for European and East Asian. However, the sequences of the core haplotype with significance of *ASAH1* are same and show strong LD across the four populations (Figure 3-3B and 3-4). The most parsimonious explanation for the sharing of the results of *ASAH1* across African and non-African is that selective sweep occurred in the region prior to the dispersal of modern humans from Africa. Further, it is likely that the selective variants are associated with the particular haplotype of *ASAH1*.

Figure 3-2. The REHH of *ASAH1* gene



The REHH values (Y axis) are plotted against the core haplotype frequency (X axis). Observed REHHs were compared with simulated data with 1,000 replications. Brown dots indicate simulation results, and blue dots represent the observations of each four populations.

Figure 3-3. A map of genes and SNPs, REHH by distance, and a LD plot of *ASAH1*

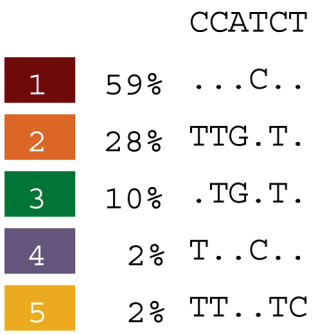


A. Blue rectangles indicate genes in the ~ 200 kb downstream region of *ASAH1*, with the official symbol of the gene shown on the rectangles. Under the genes, blue bars indicate the SNPs used for the LRH test and dark blue bars represent SNPs in the core region. The X axis indicates the distance from the core region and the Y axis the REHH values of each haplotype in the region in the figure below. Core haplotype sequences are shown in the Figure 3-4A. The data presented is based on the CEU panel from the HapMap.

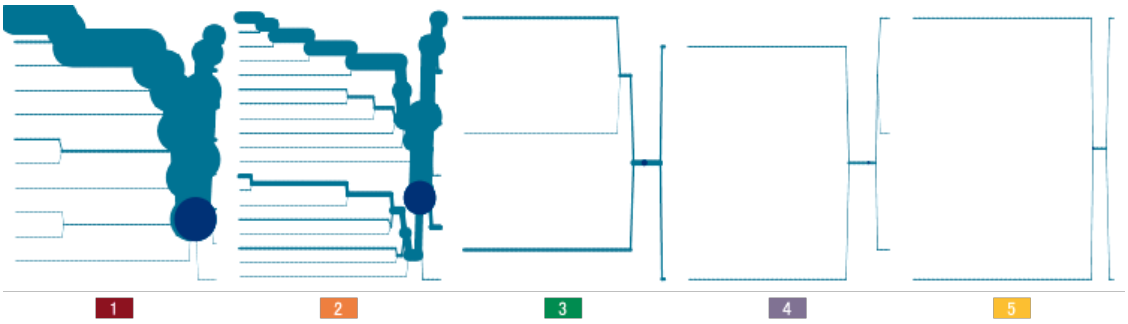
B. A LD plot was drawn using the software Haploview 3.32 (Barrett *et al.* 2005) and analyzed based on HapMap genotype data in which SNPs were shared in all four populations. The black line expresses the gene structure of *ASAH1*, and the dark blue rectangle above the line represents the region sequenced in this study. The black bar in the light blue rectangle shows the location of the SNPs used for the LD analysis, with the reference SNP ID indicated under each SNP. The triangle represents the LD plot calculated from these SNPs. Numbers in squares represent *D'* values (×100: Lewontin 1964) and the color of each square expresses the extent of LD; bright red: *LOD* (Morton 1955) ≥ 2 and *D'* = 1, red and shades of pink: *LOD* ≥ 2 and *D'* < 1, blue: *LOD* < 2 and *D'* = 1, and white: *LOD* < 2 and *D'* < 1. The triangle surrounded by a bold black line shows the LD block according to the definition of Gabriel *et al.* (2002).

Figure 3-4. The LRH test of *ASAH1* for the European populations

A



B



A. Haplotype sequences of the core region and their frequency. Dots indicate nucleotides with an identical sequence to the top sequence. The color of each rectangle represents each core haplotype as indicated in the Figure 3-3A. The haplotype of the number 1 shows the significantly high EHH and REHH in the Figure 3-2.

B. Haplotype bifurcation diagrams of each haplotype in the core region. The root of each diagram is a haplotype in the core region, identified by a dark blue circle. Blue lines indicate haplotypes in the ~ 200 kb region and the thickness of each corresponds to the number of chromosomes. The numbers and colors of rectangles correspond to the core haplotypes shown in (A).

Table 3-1. The significance of the LRH test results for eight LSD associated genes

		ASAH1	*GBA*	*GLB1*	*GM2A*	*HEXA*	*HEXB*	*NPC1*	*NPC2*
YRI	EHH	0.37	0.59	– [a]	1.00	0.61	1.00	0.71	1.00
	REHH	0.00*	0.10	–	0.35	0.08	0.74	0.48	0.62
CEU	EHH	0.00*	0.00*	0.35	0.51	NaN [b]	0.14	0.14	0.91
	REHH	0.00*	0.01*	0.04*	0.05	NaN	0.05	0.28	0.13
CHB	EHH	0.03*	0.00*	1.00	0.23	0.14	–	0.38	–
	REHH	0.01*	0.02*	0.07	0.00*	0.10	–	0.38	–
JPT	EHH	0.01*	0.02*	–	0.90	0.12	NaN	0.79	0.99
	REHH	0.04*	0.04*	–	0.24	0.58	NaN	0.87	0.29

[a]: Haplotype was not determined. [b]: EHH (REHH) was not measured.

Red letters*: P-value < 0.05

CHAPTER
4. COALESCENCE ANALYSES OF *ASAH1* IN THE HUMAN

4.1. Two distinct lineages at *ASAH1*

As aforementioned in the chapter 3, the small genetic variation and strong LD at *ASAH1* suggest signatures of selective sweep in the human population. To disclose the *ASAH1* evolution in the human population, an about 11 kb region of the *ASAH1* gene was sequenced for 15 Africans and 15 non-Africans. The sequenced region includes exons 3-10, which are part of the functional domain of the enzyme, and covers the strongest LD block including the haplotypes with significantly high EHH and REHH (Figure 3-2B). There are 106 segregating sites and 37 haplotypes within this region (Figure 4-1). In random recombination, the larger number of segregating sites generates the larger number of haplotypes. Therefore, the small number of haplotypes relative to the large number of segregating sites directly indicates fairly strong LD in this region. Nonetheless, the region was classified into two in terms of LD, a ~ 4.4 kb region under strong LD (SL) and the remaining ~ 6.6 kb region under moderately strong LD (ML).

For the SL region, 17 haplotypes with 49 segregating sites were estimated. With an ancestor nucleotide sequence of those haplotypes being inferred from a chimpanzee sequence, the gene tree constructed for the SL region (Figure 4-2) reveals the presence of two distinct lineages at the locus. They are distinguished at two nonsynonymous polymorphic sites, M72V and I93V. One lineage possesses Val (derived type) at amino acid site 72 and Ile (ancestral type) at amino acid site 93, whereas the other possesses Met (ancestral type) and Val (derived type) at these sites, respectively. The two lineages are named M and V with respect to this M72V dimorphism. These lineages are found in

both the African and the non-African samples (Table 4-1). The TMRCA between the V and M lineages is estimated as 2.4 ± 0.4 my from the average pairwise nucleotide differences between these clusters (p_B) (Materials and Methods).

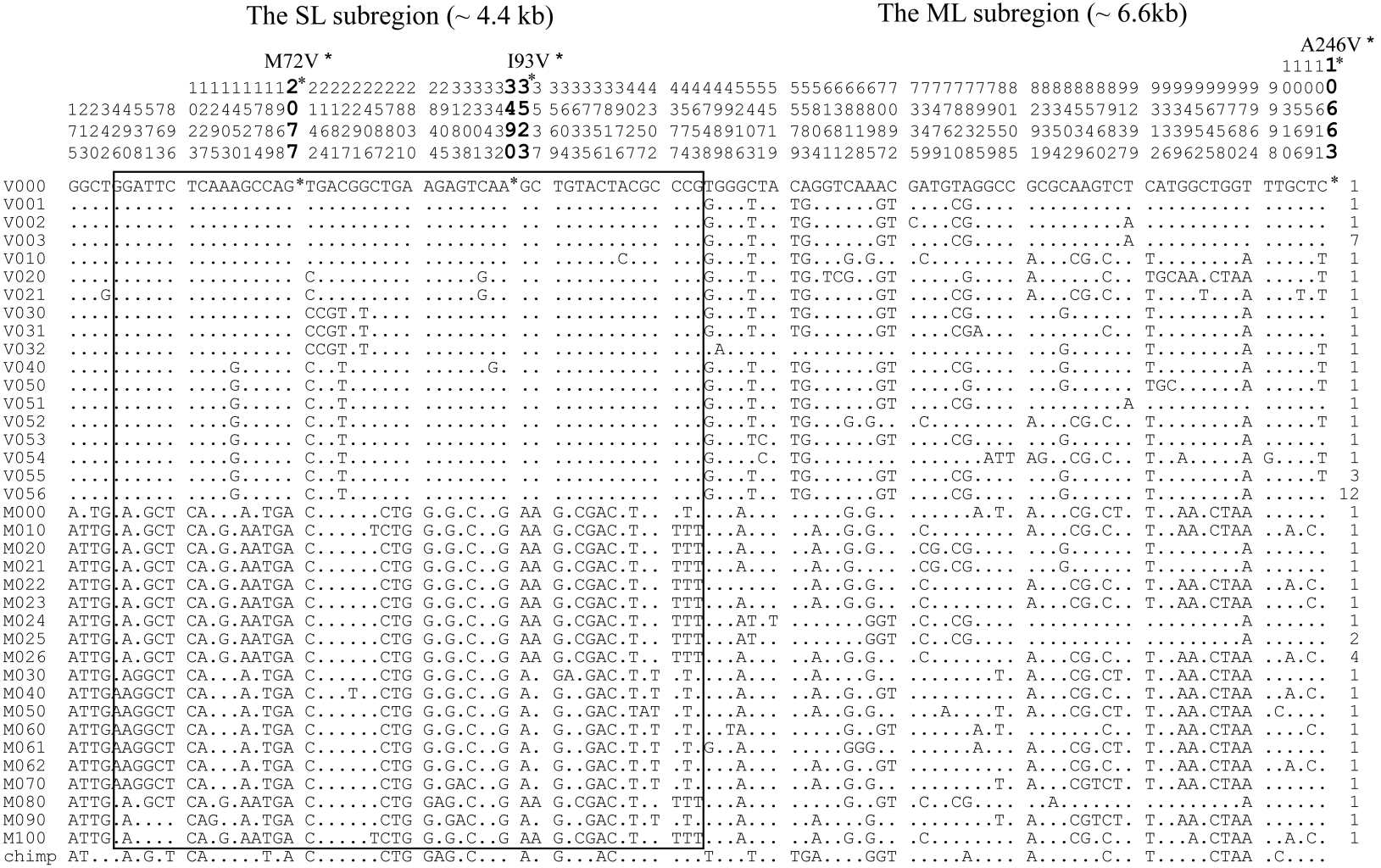
This TMRCA is relatively old compared to the average TMRCA of 0.8 my, under the assumption of $N_e = 10^4$, $g = 20$ years, and a panmictic population (Takahata 1993; Klein and Takahata 2002). However, there are several reports about genes with ancient TMRCA. Figure 4-3A is the distribution of estimated TMRCA from several studies based on the reported data of 39 loci (Figure 4-3B), 17 loci of autosomes including *ASAH1* and 22 loci of X chromosome. In the distribution, the TMRCA of 6 loci out of 16 loci of autosomes excluding *ASAH1* are older than 2 my.

To examine whether the TMRCA at *ASAH1* is unusual, I compared the nucleotide diversity and divergence of *ASAH1* with those of other ten neutral loci (Satta and Takahata 2004) by using the HKA test (Hudson *et al.* 1987). Null hypothesis of the HKA test is that the level of genetic variations would be the same within and among species between loci under neutral evolution. For applying the test, data of genetic variations within and between species at several neutral loci are needed, and I could use the data from Satta and Takahata 2004. Results of the test showed no significant difference between *ASAH1* and the other loci. These suggest that the genetic variation of *ASAH1* is not exceptional under neutrality, meaning that the ancient TMRCA at *ASAH1* is not unusual (Table 4-2).

Among the 10 loci, the TMRCA at the CMP-*N*-acetylneuraminic acid hydroxylase (*CMAH*) locus is 2.9 my and older than that of *ASAH1* (Hayakawa *et al.* 2006). The ancient TMRCA at *CMAH* is also attributed to two allelic lineages maintained for a long time in Africa as *ASAH1*. It is interesting that DNA sample used in Hayakawa *et al*. is partly the same as that used in this study. The observation of

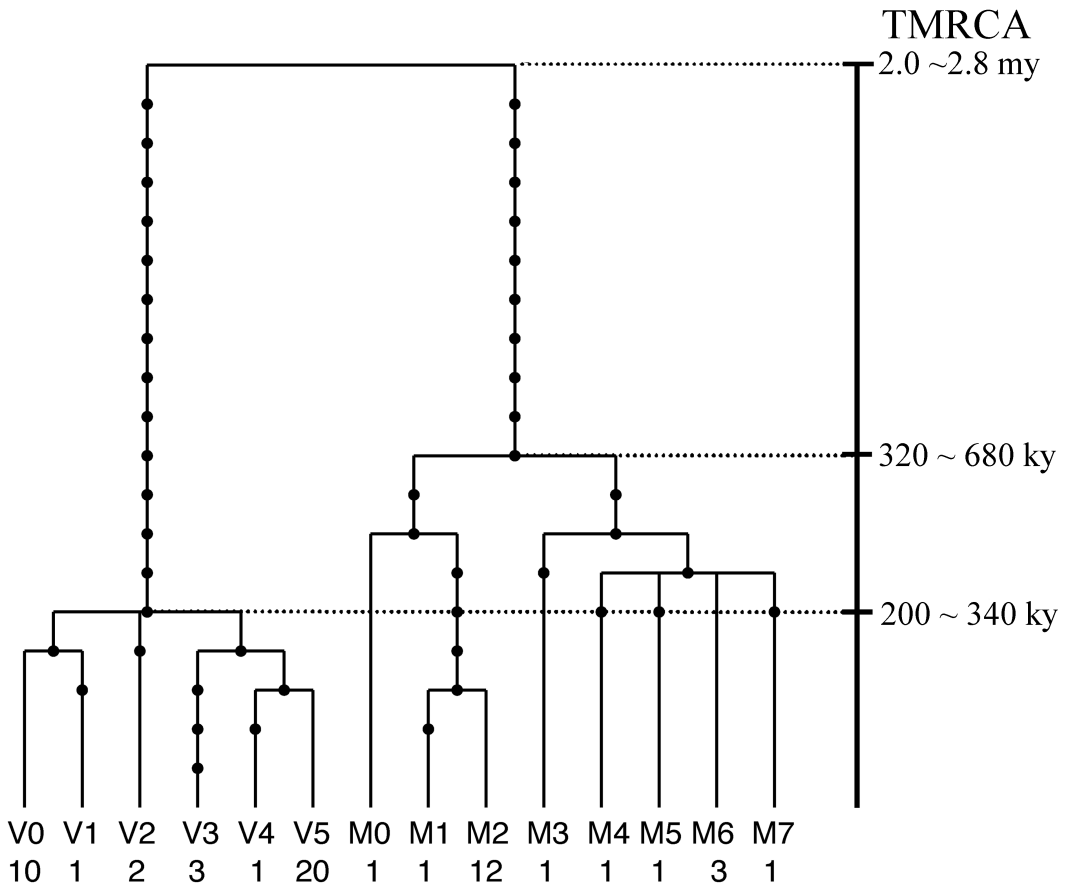
allelic lineages maintained for a long time at independent loci and/or the same sample could suggest particular demographic history of humans in the Pleistocene period in Africa.

Figure 4-1. Segregating sites in 37 haplotypes and their frequencies in a world-wide sample of 60 chromosomes



Dots indicate nucleotides that are identical to those in the top line sequence. The number above the sequence indicates the bp position of each site, and sites in the coding region are represented by the larger font size than the others. Nucleotides with an asterisk (∗) indicate a nonsynonymous polymorphic site; the amino acid substitutions at these sites are shown above the top line sequence. Sequences in the rectangle are in complete linkage disequilibrium except for two haplotypes (M090 and M100). The numbers in the far right-hand column indicate the frequency of each haplotype. This nucleotide sequence data appears in the DDBJ/EMBL/GenBank nucleotide sequence databases with the accession numbers AB371370 – AB371406.

Figure 4-2. A gene tree of 13 human *ASAH1* haplotypes in the SL region



The chimpanzee sequence was used to root the tree. Closed circles represent nucleotide substitutions. The haplotype names at the tip of the tree are provided in Table 4-1. The numbers below the haplotype names represent the frequency of each haplotype on the 58 examined chromosomes, excluding two recombinant sequences (M090 and M100). Since the nucleotide at the 32nd site (Figure 4-1) in M080 is the same as that in the chimpanzee, parallel substitution was thought to have occurred at this site; the site is incompatible with others and therefore excluded from the tree. After exclusion of this site, M080 in the SL region is identical to M2. For the method of TMRCA estimation, see the Materials and Methods.

Table 4-1. The distribution of the SL region haplotypes in a world-wide sample of 60 chromosomes

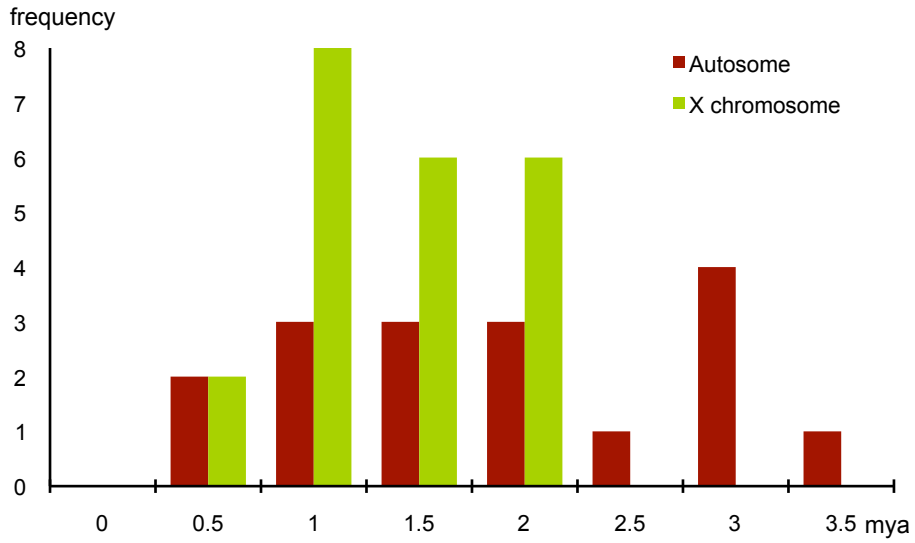
Haplotype	African	Non-African				Total
		European	Asian	Amerind	Total[a]	
V0	4	3	2	1	6	10
V1			1		1	1
V2	2					2
V3	3					3
V4	1					1
V5	10		8	2	10	20
M0	1					1
M1				1	1	1
M2	3	6		2	8	11
M3		1			1	1
M4	1					1
M5	1					1
M6	3					3
M7				1	1	1
M8	1					1
M9			1		1	1
M10				1	1	1
Total	30	10	12	8	30	60

[a]: Total of chromosomes in the non-African samples.

The first column shows the haplotype name in the SL region corresponding to those in Figure 4-1. In the figure, the haplotype name consists of one letter and three numbers. The first two numbers in Figure 4-1 indicate the haplotype number in the SL region indicated in this table. From the second to last column, the number of chromosomes in the samples is shown.

Figure 4-3. The distribution of TMRCA

A



B

Autosome			X chromosome		
Locus	TMRCA	Reference	Locus	TMRCA	Reference
12q.tel	190	Baird *et al.* 2000	*X21.1RRM2P4*	190	Garrigan *et al.* 2005a
CD209	280	Barreiro 2005	*RRM2P4*	200	Garrigan *et al.* 2005b
MCPH1	170	Evans *et al.* 2006	*ALAS2*	65.6	Hammer *et al.* 2004
β-globin	163	Harding 1997	*MSN*	60.5	Hammer *et al.* 2004
CMAH	290	Hayakawa *et al.* 2006	*APXL*	84	Hammer *et al.* 2004
ASAH1	240	Kim and Satta 2008	*AMELX*	117.8	Hammer *et al.* 2004
FUT2	300	Koda 2001	*TNFSF5*	134	Hammer *et al.* 2004
psGBA	27	Martinez-Arias 2001	*Dmd44*	193.8	Hammer *et al.* 2004
Mc1r	85	Satta and Takahata 2004	*Dmd7*	57.6	Hammer *et al.* 2004
ECP	51	Satta and Takahata 2004	*LICAM*	76.2	Hammer *et al.* 2004
EDN	303	Satta and Takahata 2004	*G6PD*	69.2	Hammer *et al.* 2004
HFE	108	Satta and Takahata 2004	*PDHA1*	191	Harris 1999
α-globin	143	Satta and Takahata 2004	*FIX*	28.2	Harris 2001
MART	300	Stefansson 2005	*P2Y10*	89.8	Kaessmann *et al.* 1999
APOE	31.1	Stephanie 2000	*Zfx*	121	Satta and Takahata 2004
Ace	111	Takahata *et al.* 2001	*MAOA*	143	Satta and Takahata 2004
Lpl	91	Takahata *et al.* 2001	*Gk*	49	Satta and Takahata 2004
			Plp	154	Satta and Takahata 2004
			Xp11.22	144	Shimada 2006
			Hprt	53	Takahata *et al.* 2001
			DACH2	120	Yu *et al.* 2002
			Dys44	162	Zietkiewicz *et al.* 2003

A. The X axis is the estimated TMRCA with the unit of my and the Y axis is frequency of the TMRCA. B. The list of reported data applied to (A). TMRCA is indicated by the unit of ky.

Table 4-2. The result of the HKA test

Locus1	Locus2[a]	P-value
ASAH1	*CMAH*	0.74
	EDN	0.65
	ECP	0.40
	Mclr	0.15
	α-globin	0.26
	β-globin	0.26
	HFE	0.69
	ZFX	0.72
	Pdha1	0.68
	Xq13.3	0.41

[a]: Satta and Takahata 2004

4.2. Coalescence simulations under various demographic models

The TMRCA of *ASAH1* was estimated from π value of the SL region (Nei and Li 1979). The value is 0.37 ± 0.02%, (Table 4-3), more than four times higher than the average value in the human genome (0.08%; Sachidanandam *et al*. 2001). For inference of the demographic history in the human evolution, coalescence simulations were performed. The simulation examined whether this large π is compatible with demographic models which have been proposed for the human evolution so far (see the Materials and Methods, and Figure 2-1). Then probability of $\pi > 0.37\%$ was estimated from the simulation. Then results of 50,000 replications showed that the probability was lower than 0.03 except for the ancient population-structure model (Table 4-4). The model assumes two subpopulations with limited gene flow maintained for 2 my (Figure 2-1H).

It should be noted that this ancient population-structure model is consistent with π values at other, 10 loci used for the HKA test (Table 4-2). The probability of $\pi <$ observed values of each loci was estimated under the model and then more than half of locus is consistent with their observed values (Table 4-5). In the result, however, the observed π value of *CMAH* rejects the ancient population-structure model even though the ancient TMRCA has been observed for the gene. The π value of *CMAH* (0.05%; Hayakawa *et al.* 2006) is smaller than the average of human genome (0.08%; Sachidanandam *et al*. 2001) and results from a low level of sequence differences ($\pi = 0.02\%$) in the non-African sample. This π value of *CMAH* is too small to test the model by the simulation. The results suggest that the human diversity could be explained by the ancient population-structure demographic history.

Table 4-3. Genetic diversity at *ASAH1* in a world-wide sample of 60 chromosomes

	Lineage[a]	N^b	S^c	H^d	π^e
World-wide	All	60	49	17	0.37 (0.024)
	V	37	9	6	0.05 (0.006)
	M	23	19	11	0.13 (0.017)
African	All	30	44	11	0.36 (0.043)
	V	20	8	5	0.06 (0.010)
	M	10	12	6	0.13 (0.016)
Non-African	All	30	41	9	0.39 (0.026)
	V	17	4	3	0.04 (0.005)
	M	13	16	6	0.11 (0.034)

[a]: Define V and M [b] : Number of sequences

[c]: Number of segregating sites [d]: Number of haplotypes

[e]: Percentage nucleotide diversity per site (standard deviation)

Table 4-4. The average of the nucleotide diversity (%) in each demographic model

Model	A	B	C	
			1	2
Average	0.233	0.195	0.233	0.235
s.d.[a]	0.6	0.6	0.7	0.7
Probability	0.023	0.005	0.025	0.030

Model	D				
	1	2	3	4	5
Average	0.220	0.220	0.215	0.206	0.196
s.d.	0.6	0.6	0.6	0.6	0.5
Probability	0.019	0.016	0.012	0.008	0.004

Model	E				
	1	2	3	4	5
Average	0.189	0.185	0.178	0.168	0.158
s.d.	0.6	0.6	0.5	0.5	0.5
Probability	0.005	0.003	0.001	0.001	0.000

Model	F			G
	1	2	3	
Average	0.188	0.190	0.192	0.201
s.d.	0.6	0.6	0.6	0.5
Probability	0.003	0.004	0.004	0.004

Model	H			
	1	2	3	4
Average	0.276	0.329	0.340	0.341
s.d.	0.7	0.5	0.5	0.5
Probability	**0.072**	**0.198**	**0.230**	**0.237**

Details of the models are described in the Figure 2-1 and Table 2-4. Probability represents the proportion of replications with $\pi > 0.37\%$ in 50,000 replications under each parameter combinations. Red letter means that the probability is > 0.05.

Table 4-5. The probability of observed π under the ancient population-structure

Loci	Observed π (%)	*Nm*[a]		
		0.001	0.01	0.1
CMAH	0.05	0.00	0.00	0.00
EDN	0.09	0.02	0.03	**0.12**
ECP	0.09	**0.11**	**0.15**	**0.22**
Mclr	0.11	**0.21**	**0.25**	**0.46**
α-globin	0.11	0.04	**0.05**	**0.06**
β-globin	0.18	**0.50**	**0.53**	**0.76**
HFE	0.08	0.03	**0.07**	**0.34**
ZFX	0.08	0.01	0.03	**0.10**
Pdha1	0.16	**0.11**	**0.18**	**0.61**
Xq13.3	0.04	0.00	0.01	0.03

[a]: Migration rate per population. **Red letter**: probability > 0.05

4.3. The human demographic history

Simulation studies support the ancient population-structure model is consistent with the large TMRCA at *ASAH1* and other loci. The distribution of the V and M lineages in 60 chromosomes is compatible with the demography. The frequency of the V and M lineage in the total sample is 0.62 and 0.38, respectively (Table 4-1). Both lineages are detected in African and non-African and the frequencies of two lineages are similar in African and non-African. This indicates that two lineages should be in Africa before the out of Africa. Further, the distribution of two lineages in an individual supports the demography also. Under the assumption of Hardy-Weinberg equilibrium, the expected heterozygosity with V and M is 0.47. However, the observed value in the sample is 0.23, which is significantly lower than the expectation ($P < 0.005$, Chi-Square test). This heterozygosity deficiency is also observed in both the African and non-African. Under overdominance selection, it is expected to observe an excess of heterozygotes over Hardy-Weinberg equilibrium, although the extent is not necessarily large if mating occurs at random every generation. This deficiency is inconsistent with the possibility of overdominance that might maintain two allelic lineages for a long time (Takahata 1990). Alternatively, this deficiency suggests that the V and M lineages have been maintained in a partially isolated subpopulation until the exodus of Africa. When migration is limited, it is likely that genes within each subpopulation coalesce to a common ancestor and form a single cluster in a tree (Takahata 1991).

On the other hand, some studies claimed that the ancient TMRCA could be explained by regionally interbreeding with the archaic *Homo* after the out of Africa of *Homo sapiens* (Templeton 2002; Evans *et al.* 2006; Garrigan and Hammer 2006; Shimada *et al.* 2007; Cox *et al.* 2008). In particular, Evans *et al.* (2006) argued the detection of introgression into *Homo sapiens* from archaic *Homo* lineage representing at

locus of *MCPH*. As the *ASAH1* genealogy, two distinct lineages have been detected at the *MCPH* locus. Two loci have shown shared features in their genealogy, not only the ancient TMRCA of two distinct lineages (1.7 my at *MCPH*) but also less frequent recombination between two lineages. Further, it is similar that the genetic diversity and the coalescence age are very different between two lineages (details about *ASAH1* are discussed in the chapter 5). In the study of Evans *et al.* (2006), simulation studies showed that the observation of the maintaining of allelic lineages for a long time is incompatible with the panmictic population model and instead suggests population-structure and introgression model. However, in contrast to *ASAH1*, the population-structure model was rejected in their study. They suggested an admixture of the human and Neanderthal based on the coexistence of two populations in Europe. The reason of the rejection of the population-structure model is that other loci in the human genome have not shown similar pattern of genealogy as observed for *MCPH*. About this point, I have already described in the chapter 4-2 that we can find the ancient TMRCA at several loci in the human genome (Figure 4-3). In addition, for the conclusion of introgression, it is required the interbreeding the Neanderthal and the human at the out of Africa and then the positive selection for an introgressed allele of *MCPH* since the introgressed allele is likely to be extinct by genetic drift because of rare frequency of the allele. Evans *et al* (2005) suggested that the introgressed allele of *MCPH* associated gene with brain size is advantageous to modern humans and spread rapidly in the human population but some studies have not agreed with this positive selection (Currat *et al.* 2006; Yu *et al.* 2007). Besides, they have to explain the common presence of the introgressed allele in African (Evans *et al.* 2005).

Simulation studies by Nordvorg (1998 and 2001) pointed out that the power to detect admixed ancestry of polymorphism at a locus is very low, thus even recent

studies have failed to detect the evidence of ancient admixture (Serre *et al.* 2004; Green *et al.* 2006). Moreover, Takahata *et al.* (2001) examined 13 loci available for polymorphic nucleotide sequence data and have shown that the ancestral lineage is traced back to Africa at all but one locus. Further, in the observation of *ASAH1*, all observed haplotypes are coalesced to an ancestral haplotype in African, which provides no evidence of ancient large-scale admixture after the exodus of Africa. Consequently, in this chapter it is revealed that the pattern of genetic diversity of *ASAH1* and other loci are compatible with the proposal that the human population was once geographically structured and genetically differentiated in Africa before the period of Pleistocene (Takahata 1995, Satta and Takahata 2004).

CHAPTER 5. *ASAH1* POLYMORPHISM OF THE HUMAN POPULATION

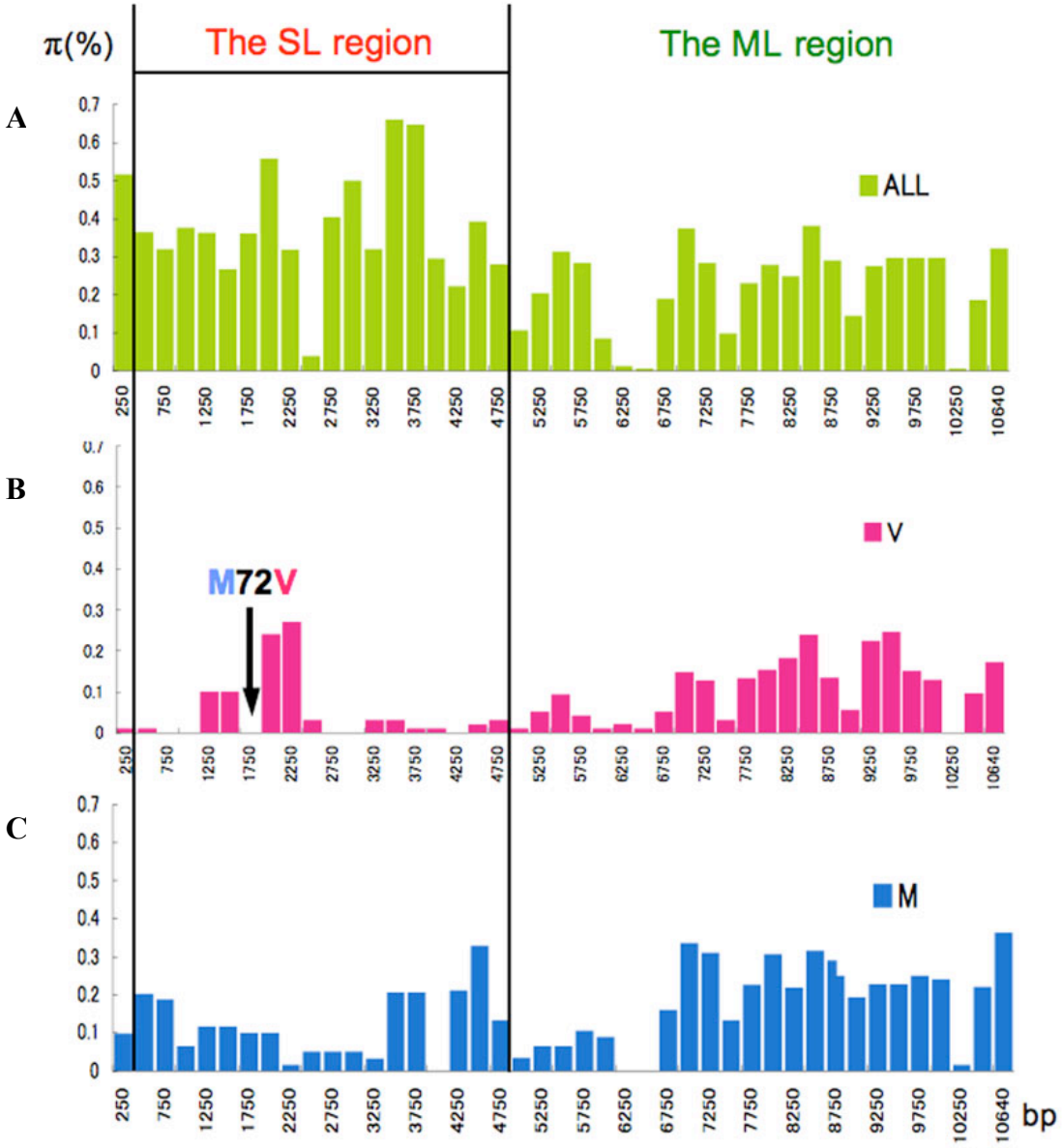
5.1. Signatures of Darwinian positive selection

This chapter describes the *ASAH1* genetic variation within each lineage in the human population. The π value within the V lineage (π_{VSL} = 0.05 ± 0.01%; Table 4-3) is smaller than the overall π value on chromosome 8 (0.08%) (Sachidanandam *et al*. 2001) on which *ASAH1* is located. More importantly, the π_{VSL} value is significantly smaller than the nucleotide diversity within the M lineage (π_{MSL} = 0.13 ± 0.02%). Furthermore, the number of haplotypes in the V lineage is only six, yet it is 11 in the M lineage. These hold true in both African and non-African samples (Table 4-1). I can therefore expect that the TMRCA within the V lineage is younger than that within the M lineage as far as the SL region is concerned (Figure 4-2). Indeed, the TMRCA of the V lineage is estimated as 200 ± 50 ky from the Genetree analysis (Griffiths and Tavaré 1995) and 340 ± 80 ky based on the average nucleotide diversity (Materials and Methods). On the other hand, the TMRCA of the M lineage is 320 ± 70 ky from the Genetree analysis and 680 ± 180 ky from the nucleotide diversity. Compared with the M lineage, the relatively recent origin of the predominant V lineage implies that it has been rapidly increasing in frequency. In accord with this, Tajima's *D* value is negative (-0.11) for the V lineage and positive (0.27) for the M lineage, although both are not statistically different from zero ($P > 0.1$) (Tajima 1989). Moreover, the rapid expansion of the V lineage is also consistent with the result of the LRH test that the haplotype showing the significantly high EHH and REHH is included in the haplotypes of the V lineage (Figure 3-4).

To see if the reduced level of polymorphism within the V lineage is restricted to the SL region, the π value is compared with that of the ML region defined as moderately strong LD (Figure 5-1). Possible recombination in the ML region does not allow us to make phylogenetic analysis and the concept of lineage *per se* becomes equivocal. For this reason, I simply defined the V and M lineages in the ML region as those that are linked with V and M in the SL region. I then calculated the nucleotide diversity as $\pi_{VML} = 0.10 \pm 0.02\%$ and $\pi_{MML} = 0.18 \pm 0.02\%$. They are not significantly different from each other ($P > 0.05$, Z-test), but π_{VML} is significantly larger than π_{VSL} ($P < 0.01$, Z-test; Figure 5-1).

As mentioned earlier, the frequency of the V lineage in the total sample is higher (0.62) than that of the M lineage (0.38). The predominance of the V lineage is observed in both Africans and non-Africans (Table 4-1). The HapMap data also gives similar results: 0.83 in YRI, 0.56 in CEU, and 0.67 in both CHB and JPT. Despite the relatively high frequency of the V lineage irrespective of data and populations, the π value is smaller and the within-lineage TMRCA is shorter than the corresponding value in the M lineage. All these features consistently indicate positive Darwinian selection operating for the V lineage. In addition, it should be noted that this small genetic diversity was limited to the SL region, suggesting that the target of the selection is located within the region.

Figure 5-1. Window analysis of π in the sequenced region of ~11 kb length



The π values were calculated in a window size of 500 bp with a step size of 250 bp. The X axis indicates the distance in the sequenced region and the Y axis the π value of each window.

A. The π values obtained from a sample of 60 sequences.

B. The π values within the V lineage. The black arrow indicates the location of the M72V polymorphism.

C. The π values within the M lineage.

5.2. Lineage-specific amino acid changes

In the SL region of *ASAH1*, there are 46 segregating sites in introns, one synonymous and two nonsynonymous segregating sites. The synonymous mutation is observed in only one chromosome (V040), whereas two nonsynonymous mutations (M72V and I93V) are "fixed" within each lineage. Regarding M72V, the Val is a derived and human-specific amino acid, because in chimpanzees, orangutans, and rhesus monkeys the amino acid at this site is exclusively occupied by Met. Regarding I93V, the Val is shared by rhesus monkeys, but the site is occupied by Ile in chimpanzees and orangutans (Figure 5-2). It is likely that site 93 is subjected to recurrent substitution to Val.

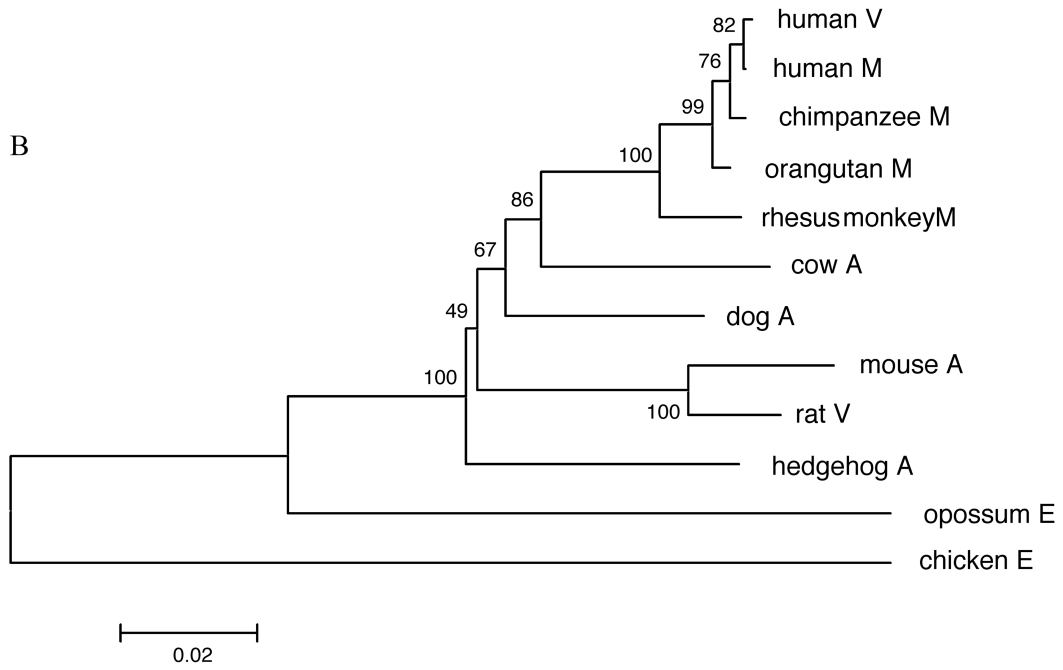
The SL region contains exon 3 (31 amino acids) and exon 4 (29 amino acids). Although these exons do not encode the active center of ASAH1, mutations in these exons are associated with Faber disease. Three such nonsynonymous mutations are Y36C, V96del, and V97E (Bar *et al*. 2001; Muramatsu *et al*. 2002). It thus appears that these exons encode important functional parts of ASAH1. It is interesting to note that the 659 bp region encompassing M72V did not accumulate any mutations within the V lineage (Figure 5-1). This is again a signature of selective sweep on the region encompassing Val at site 72, suggesting the amino acid is a likely target of positive selection for the V lineage.

Figure 5-2. Alignment of avian and mammalian *ASAH1* orthologs in the SL region (A) and the NJ tree obtained based on nonsynonymous substitutions in the entire *ASAH1* of these orthologs (B)

A

```
                A46P                                M72V                                I93V
human V     TYRGAVPWYT INLDLPPYKR WHELMLDKAP VLKVIVNSLK NMINTFVPSG KIMQVVDEKL
human M     .......... .......... .......... M......... .......... .V........
chimpanzee  ....P..... .......... .......... M......... .......... ..V.......
orangutan   ....P..... .......... .....V.... M......... .......... ..V.......
rhesus      ....PA.... .......... .....A.... M......... ..V....... .V..I.....
mouse       ....P...H. .......... ....LAQ... A.RIL...IT SLV....... .L.KM..Q..
rat         ....P..... .......... ....LAH... ..RTL...IS .LV.A..... ....M.....
cow         ....P..... .......... .....VV... A.......M. .IV.A..... .-LYL..Q..
dog         ....S..... .......... .....A.... A..I...... DLVTA..... ....I..Q..
hedgehog    ....P..... .......... ..K..T.... A.RI...... .VV.A..... .V..M..Q..
opossum     ....P..... .......... ...VIT.LT. EVRH..KD.L D.V.A.F... .VTKM...R.
chicken     .FK.N..T.I .......S.. .D..IRA.KT E..AVIQNI. DIA...F... ..VDI..H.I
```

B



A. Dots indicate amino acids that are identical to those in the top line sequence. A dash represents a deletion. Bold letters indicate amino acid substitutions in the human lineage compared with birds and mammals.

B. Numbers near the node indicate the bootstrap probability in 500 replications. The scale bar below the tree represents the number of nonsynonymous substitutions per site. The nonsynonymous sites of 751 bp were compared. The chicken sequence was used as an outgroup. Capital letters next to the species name indicate the amino acid at position 72.

5.3. Theoretical considerations of positive selection operating on *ASAH1*

The above results show that positive selection has favored the V lineage commonly sharing Val at amino acid site 72 and reduction of genetic diversity in the lineage despite its predominance. To examine the relationship between π within a lineage and population size with and without selection, I carried out computer simulations under the ancient population-structure model (Materials and Methods; Figure 2-2). To check the validity of the computer program for the model, I compare simulation results with the theoretical expectation for neutral cases.

To obtain the theoretical formulas for the coalescence time of two neutral genes at a locus, I use formula 4 for a finite island model in Takahata (1995). The model assumes that the ancestral panmictic population of effective size M has been subdivided into l island populations with effective size N_1 for t_i generations. Our model is different in that the subpopulations have further admixed to form a panmictic population of effective size N_0 for subsequent t_m generations to the present. I sample two genes from this panmictic population and derive the formula of the mean coalescence time. Although the formula is complicated for any values of l, M, N_1, and N_0, it is much simplified for the case of $l = 2$, $M = N_0 = 2N_1$, and this simplified formula is sufficient to our present purpose.

The coalescence can occur in the admixed population during the period of t_m generation. This conditional TMRCA, T_0, is ready given by

$$T_0 = \int_0^{t_m} t \frac{1}{2N_0} e^{-\frac{t}{2N_0}} dt = 2N_0 \left[1 - (1 + \frac{t_m}{2N_0}) a \right] \qquad (1)$$

where $a = e^{-\frac{t_m}{2N_0}}$. On the other hand, if two genes do not coalesce during t_m, the two ancestral gene lineages reside in the same subpopulations with probability $P(0)$ and in two different populations with probability $Q(0)$ in Takahata's designations. Obviously, $P(0) = Q(0) = 1/2$ in the present model with $l = 2$. Formula 4 in Takahata (1995) then reduces to the following T_1

$$T_1 = 4N_1 + \frac{1}{4m} - \frac{Q(t_i)}{2m} \tag{2}$$

where

$$Q(t_i) = \frac{1}{2R_2}\left[\frac{R_1 + R_2}{2} e^{(R_2 - R_1)t_i} + \frac{R_2 - R_1}{2} e^{-(R_1 + R_2)t_i}\right],$$

$$R_1 = 2m + \frac{1}{4N_1} \text{ and } R_2 = \sqrt{R_1^{\,2} - \frac{m}{N_1}}.$$

Noting that this coalescence is conditional with probability a and from formulas (1) and (2), I obtain the formula of the unconditional coalescence time $T = T_0 + aT_1$:

$$T = 2N_0 - a(2N_0 + t_m - T_1). \tag{3}$$

For $4N_0m > 1$, T approaches to $2N_0$ generations, whereas for small $4N_0m$, T becomes large in proportion to the reciprocal of m.

Finally, to define the effective population size and the expected nucleotide diversity π for the present non-equilibrium demographic model, I set $N_e = T/2$ as in Nei and Takahata (1993) and obtain

$$N_e = N_0 - \frac{a}{2}(2N_0 + t_m - T_1) \quad (4a)$$

$$\pi = 4N_e\mu. \quad (4b)$$

Under neutrality, simulation assumes $N_0L\mu = 0.4$, $L = 1000$, and $N_0 = 20$. In a wide range of N_0m, π in formulas (4) agrees well with that in simulation (Table 5-1). The same simulation also demonstrates that the π value within predominant lineage (S lineage) is significantly larger than that within the other lineage (N lineage), even though with limited gene flow ($4N_0m = 0.004 \sim 0.4$) (Table 5-1; t-test, $P <$ 1E-130). Thus neutrality cannot explain the observed reduction of nucleotide diversity in the V lineage in the SL region. On the other hand, in the presence of selection, the mean and variance of nucleotide diversity in S lineage are significantly smaller than those in N lineage with any tested migration rate (Table 5-1; F-test: $P <$ 1E-10, t-test: $P <$ 1E-6). Selective sweep can thus result in reduced nucleotide diversity within a lineage that carries an advantageous mutation. This pattern is exactly what I observed at *ASAH1*.

Table 5-1. The distribution of nucleotide diversity under ancient population-structure model defined in text

	N_0m	Total		S lineage		N lineage	
		mean[a]	s.d.[b]	mean	s.d.	mean	s.d.
Neutral	0.1	0.33 (0.29)	0.17	0.18	0.15	0.05	0.08
	0.01	0.46 (0.47)	0.17	0.25	0.19	0.05	0.06
	0.001	0.51 (0.51)	0.16	0.28	0.21	0.05	0.07
Selection	0.1	0.37	0.18	0.01	0.02	0.1	0.07
	0.01	0.39	0.12	0.01	0.02	0.08	0.06
	0.001	0.38	0.15	0.01	0.02	0.07	0.06

[a]: Percentage mean of nucleotide diversity (theoretical expectation)

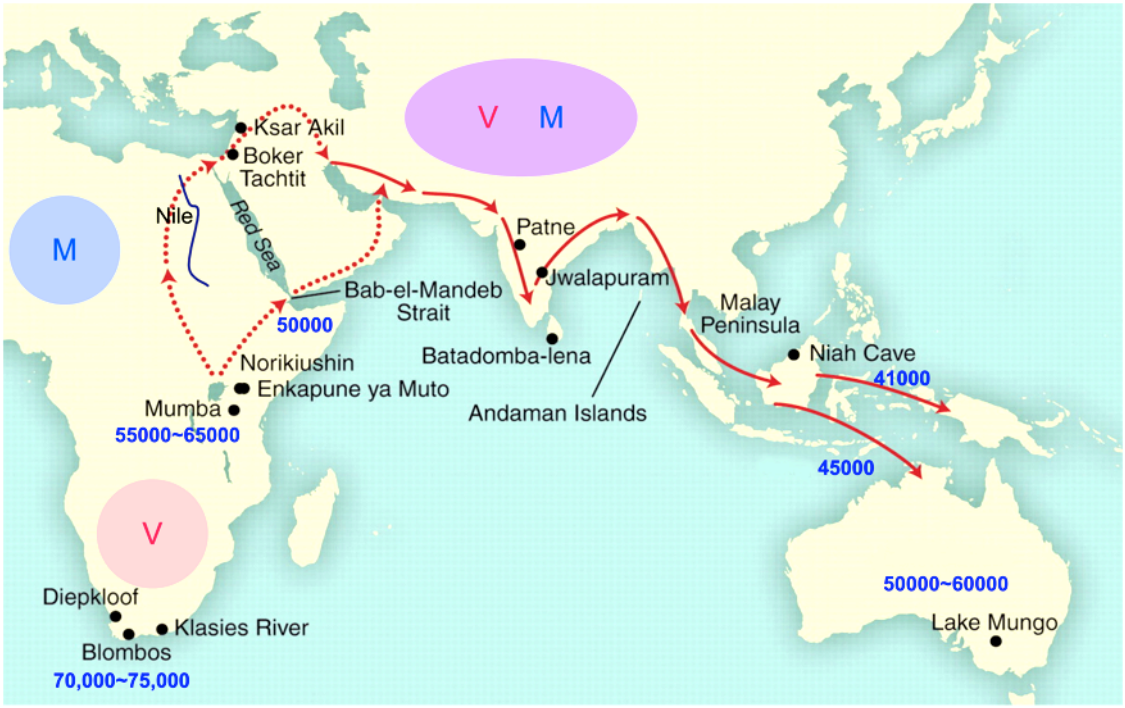
[b]: Percentage standard deviation of nucleotide diversity

5.4. Biological significances of *ASAH1* in the evolution of human mental activity

The pattern and level of genetic variability at *ASAH1* cannot be explained by demographic causes alone and/or overdominance selection. Rather, they strongly suggest positive Darwinian selection operating on the predominant V lineage of a relatively recent origin. It is likely that the human-specific Val at amino acid site 72 is responsible for positive Darwinian selection and may have occurred before the most recent common ancestor of all V lineage haplotypes, 200 ~ 340 ky. When modern humans dispersed from Africa, admixture of the distinct V and M lineages occurred and the advantageous V lineage has since spread in the entire population (Figure 5-3). Although there is no functional assay of the Val substitution at present, it is conceivable

that the Val influences the enzyme activity of ASAH1, thereby affecting catalysis of ceramides and possibly neuronal development as well. Further comprehension of *ASAH1* under positive Darwinian selection should shed light on evolution of the acquisition of behavioral modernity in human evolution.

Figure 5-3. The dispersal of two lineages from Africa.



The map is retrieved from Mellars 2006a Fig.1. A red arrow is the inferred route of migration of modern humans from Africa to Eurasia. A blue number indicates the estimated dates of archeological records at sites and all dates are before present (Mellars 2006a; 2006b). This figure represents an imagination of distributions of subpopulation in Africa in the period of the modern human migrations from Africa. There were two geographically distant subpopulations in Africa and each allele (V and M) was shared in each subpopulation, severally. It is speculated that gene flow between subpopulations occurred by occasion of the dispersal of humans to non-Africa.

CHAPTER
6. PHYLOGENETIC ANALYSES OF CERAMIDASE GENES

6.1. The phylogeny of ceramidase family

In the previous chapters, the evolution of *ASAH1*, which encodes acid ceramidase, and the human evolution were described. In this chapter, I tried to disclose not only the origin of *ASAH1* but also the evolution of the other ceramidase, neutral and alkaline ceramidase.

First, I constructed a phylogenic tree of five ceramidase genes with using amino acid sequences, *ASAH1*, *ASAHL*, *ASAH2*, *ASAH2L*, *ASAH3*, and *ASAH3L* (Figure 6-1). The homologous amino acid sequences of vertebrates, invertebrates, and plants were detected in public databases. These five genes are so conserved that the sequences are easily aligned. In the case of neutral ceramidase, bacteria have retained the enzyme activity but their sequences were excluded from the analysis owing to their large divergences. Homologous genes of the five genes are clearly clustered in separate groups. It reveals not only that the divergence between five groups of genes is large but also each of these genes is conserved within a group. The conservations could suggest five groups of genes attained specific roles in the evolution of ceramidase. In mammals, the functional difference between ASAH1 and ASAH2 has been claimed that the catalysis of ASAH1 is for biosynthetic ceramides in lysosome but that of ASAH2 is for dietary ceramides in intestine (Kono *et al.* 2006; Nilsson and Duan 2006). Only two sequences of the *Dictyostelium discoideum* are not clustered with any other sequences and located between clusters of the acid and alkaline ceramidase. It is likely that

functional differentiations among ceramidase are not distinct in the *Dictyostelium discoideum.*

For the detailed molecular evolution of five groups of genes, the phylogenic tree of each groups of ceramidase genes was also constructed (Figure 6-2, 6-3 and 6-4). The homologous amino acid sequences of *ASAH1* (acid ceramidase) and *ASAHL* could not be detected in invertebrates by blast analysis. As seen in the phylogenetic tree constructed by homologs of two genes, each group of genes makes a cluster including all homologs without the *Dictyostelium discoideum* (Figure 6-2). Therefore, although ASAHL amino acid sequences are possessing characteristics of acid ceramidase amino acid sequences, the function of ASAHL is expected to be diverged from that of ASAH1.

The phylogenic tree of *ASAH3* and *ASAH3L* suggests two points (Figure 6-4). First, *ASAH3* gene should be derived from *ASAH3L* gene because *ASAH3* is vertebrate-specific gene while *ASAH3L* is detected in vertebrates and invertebrates. The tree suggests that *ASAH3L* duplicated to *ASAH3* before the split of vertebrates and invertebrates, and then generated *ASAH3* was lost in the invertebrate lineage. Second is that the evolutionary pattern of two genes are different from each other. *ASAH3L* amino acid sequences are extremely conserved in birds and mammals, relative to *ASAH3*. While the amino acid homology of human and mouse *ASAH3* is 80% in 264 sites, the corresponding homology of *ASAH3L* is 93% in 275 sites. This higher homology suggests that stronger purifying selection has operated on *ASAH3L* in mammals, and then two groups of genes have attained specific functions separately for ASAH3 and ASAH3L after the duplication of *ASAH3L* to *ASAH3*.

In contrast to the absence of *ASAH1* and *ASAHL* in invertebrates, the genes belonging to *ASAH2* and *ASAH3L* groups are found in the fruit fly (*CG1471* and *CG13969*) and mosquito. Therefore, they may have both homologs of mammal neutral

and alkaline ceramidase but may not have that of acid ceramidase. Yoshimura *et al.* (2002) reported that RNAi for the neutral ceramidase encoded by *CG1471* in *Drosophila* remarkably decreased the activity at not only neutral but also acidic pH. This result may indicate that the catabolism for ceramides is mainly performed by the *CG1471*-encoded ceramidase in the fruit fly at both neutral and acidic conditions. With regard to the fruit fly *ASAH3L* homologous gene (*CG13969*) of mammals, it was reported that the *bwa* mutant, possessing null allele of *CG13969,* shows defects of brain structure formation (Boquet *et al.* 2000). This indicates that the gene may be involved in brain or central nervous systems in the fruit fly. The central nervous system has been highly conserved even between the fruit fly and human. Symptoms similar to human neurodegenerated diseases are shown in the fruit fly, and the fruit fly provides models for human diseases, for example, Huntington's disease (Jackson *et al.* 1998), Parkinson's disease (Feany and Bender 2000), and *etc.*. This conservation is compatible with that 77% and 87% of human disease and mental retardation associated-genes have been found their orthologs in *Drosophila*, respectively (Reither *et al.* 2001; Inlow and Restifo 2004). Consequently, if the functions of alkaline ceramidase in the fruit fly have been conserved to the human, we can expect the importance of the alkaline ceramidase in the central nervous system.

Figure 6-1. The phylogenetic tree of ceramidase gene

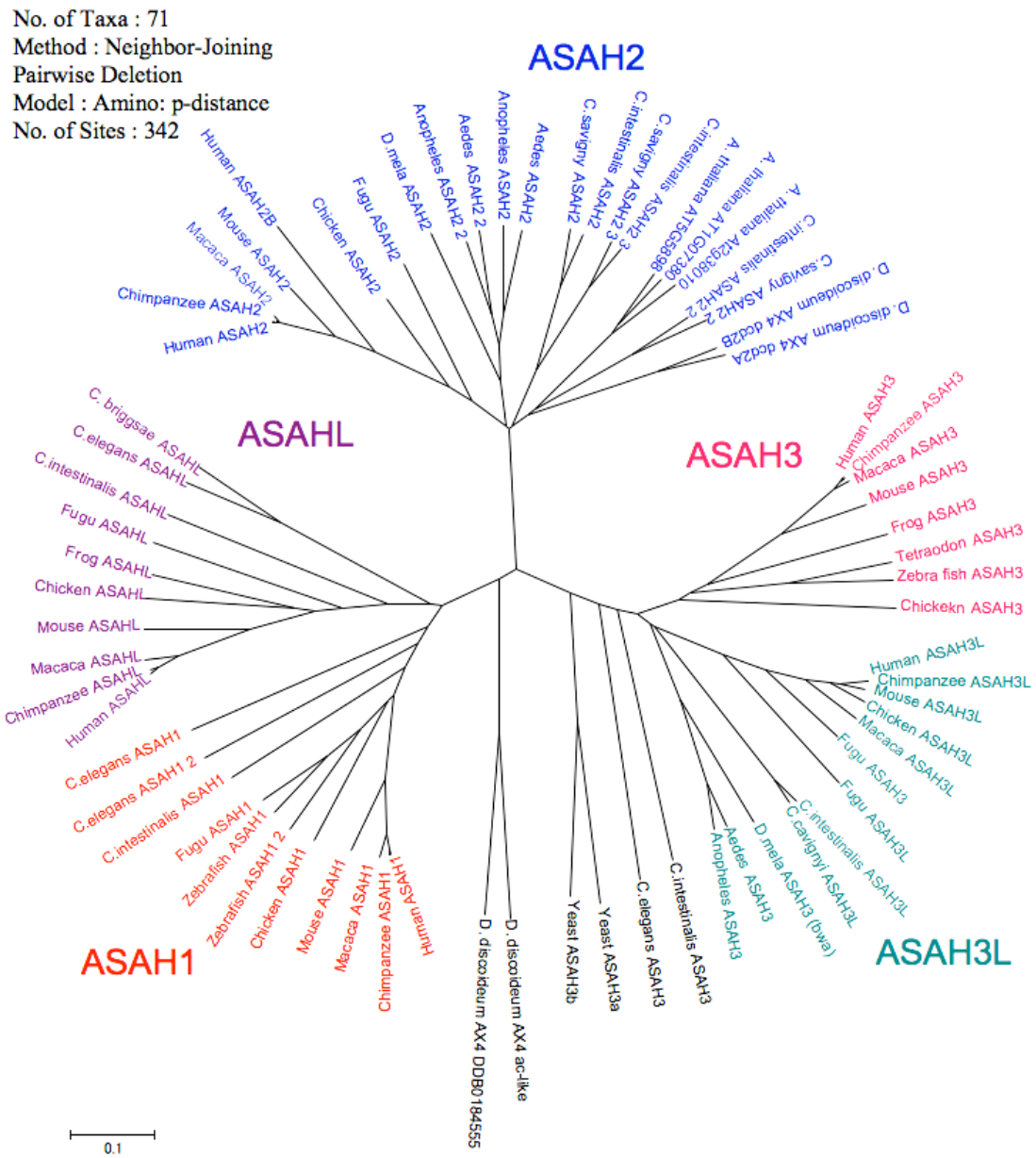


Figure 6-2. The phylogenetic tree of *ASAH1* and *ASAHL*

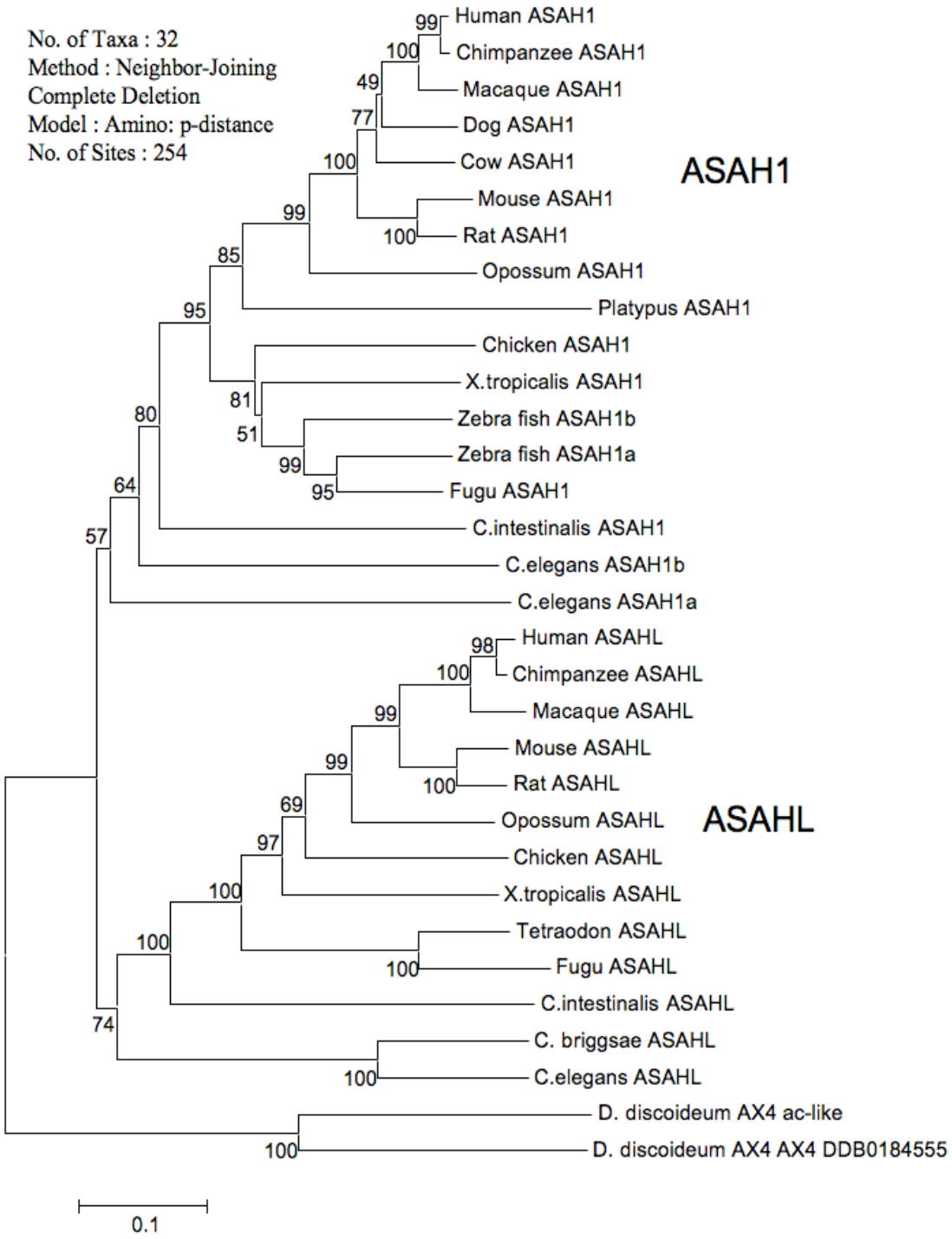


Figure 6-3. The phylogenetic tree of *ASAH2*

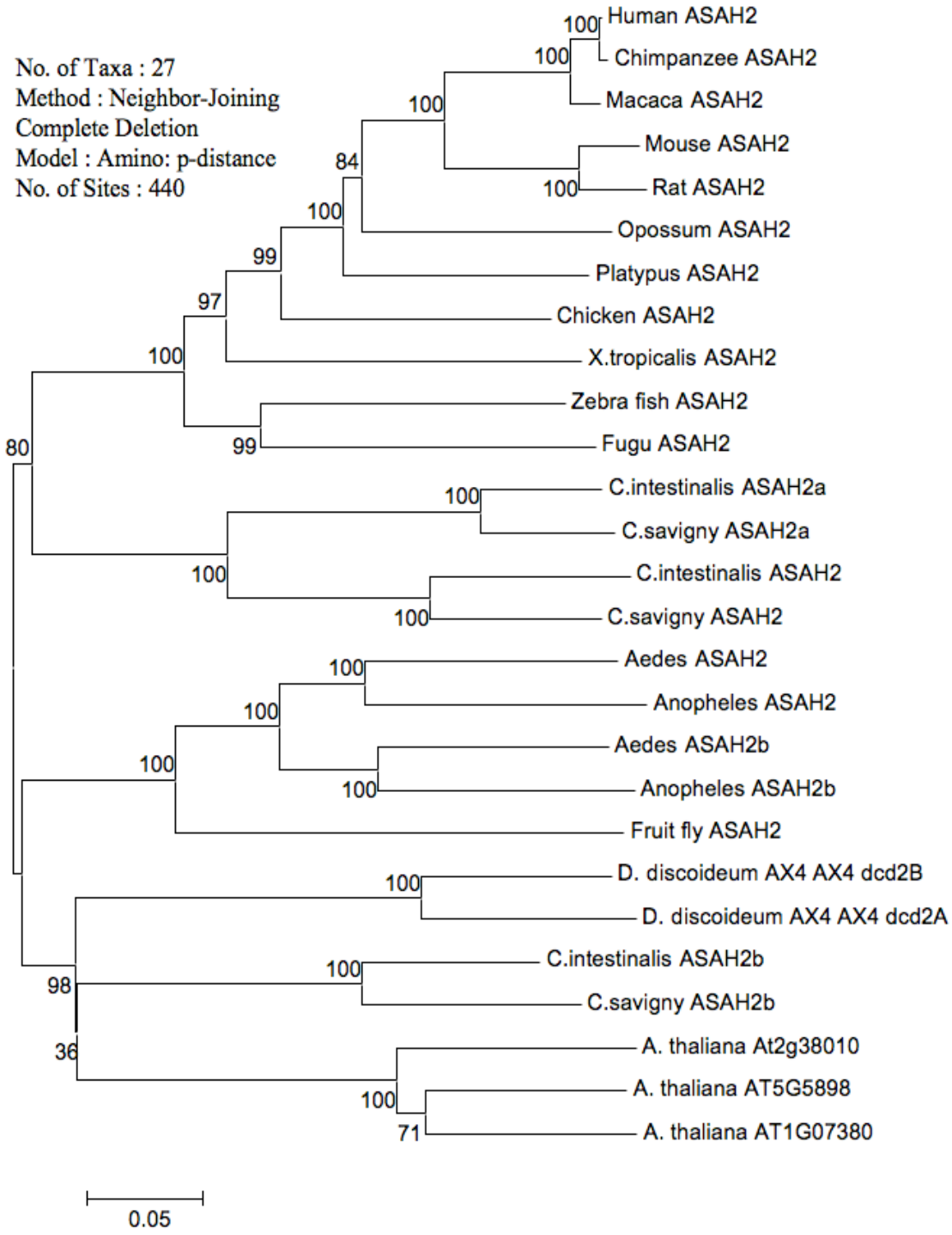
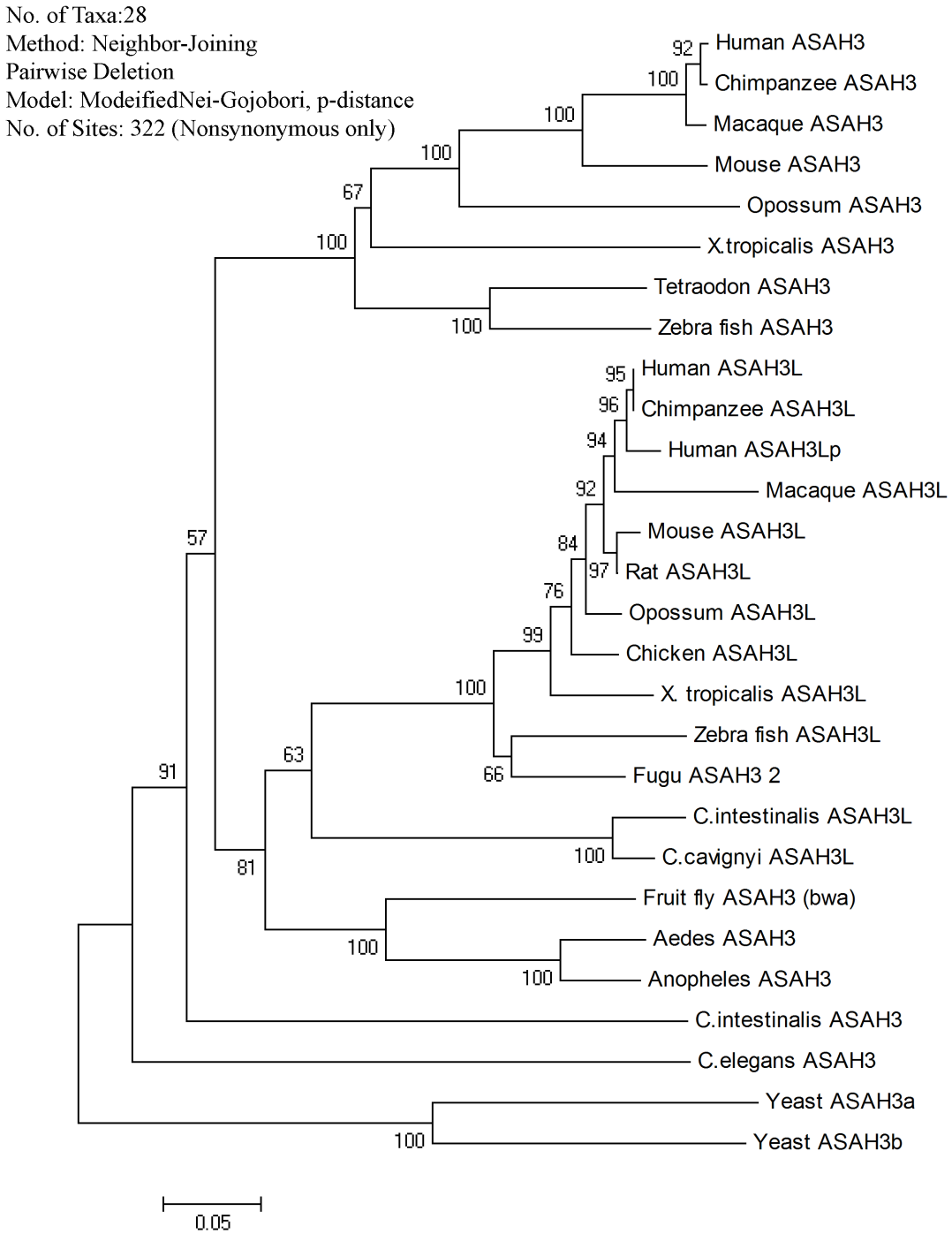


Figure 6-4. The phylogenetic tree of *ASAH3* and *ASAH3L*



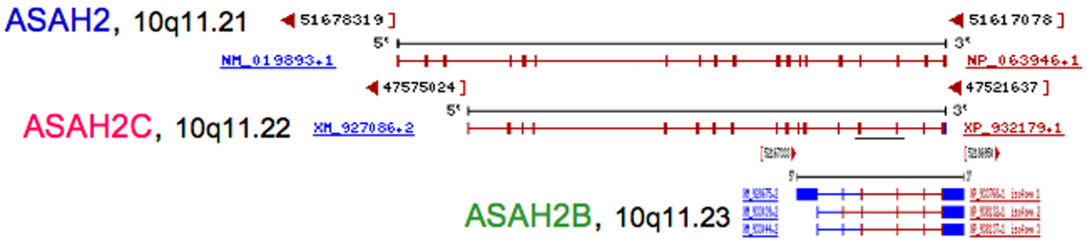
6.2. The blast and dot-matrix analyses for *ASAH2*

Two partial sequences of *ASAH2* duplicates were found in the human and chimpanzee genome by the blast analysis using *ASAH2* nucleotide sequences as queries. In humans, there are *ASAH2* and its two homologs on chromosome 10, and those sequences are named as *ASAH2*, *ASAH2B*, and *ASAH2C* in NCBI. Three sequences are located at 10q.11.21, 10q11.23, and 10q.11.22, and their lengths for the entire coding region are about 61.2, 19.6 and 53.4 kb, respectively (Figure 6-5). In the chimpanzee genome, *ASAH2B* is found but *ASAH2C* is not detected. One premature stop codon is found in *ASAH2B*, thus this chimpanzee sequence is designated as *ASAH2Bp*, a psuedogene of *ASAH2B*, in this study. Both *ASAH2B* and *ASAH2C* sequences were not detected in the rhesus monkey and other primates by blast searches, suggesting that these duplications should be occurred after the split of the human and rhesus monkey.

To confirm the range of duplicated region encompassing *ASAH2* on chromosome 10, I carried out dot-matrix analyses. Figure 6-6 is the result of the analysis for two separate 1 Mb regions including human *ASAH2* (10q11.21~22) and *ASAH2C* (10q11.23). In the result of dot-matrix analysis, two diagonal lines were noteworthy. The long one in the figure shows the homology between the ~ 160 kb *ASAH2*- and *ASAH2C*- containing regions, and the short one shows the homologous ~ 50 kb regions between *ASAH2B* and partial *ASAH2* and their downstream regions. There are long contig gaps at down and upstream of a contig (NT_077571), including *ASAH2C* (Figure 6-6). *ASAH2C* seems to be segmental duplication but we could not determine the length of duplicated region because of the contig gaps. Figure 6-7 is the magnified figure of the ~ 160 kb homologous region between *ASAH2* and *ASAH2C* in humans. Comparison between the region including ASAH2C and its syntenic region in the chimpanzee shows that a diagonal line is only ~ 80 kb long. This region corresponds to *FAM21B* gene

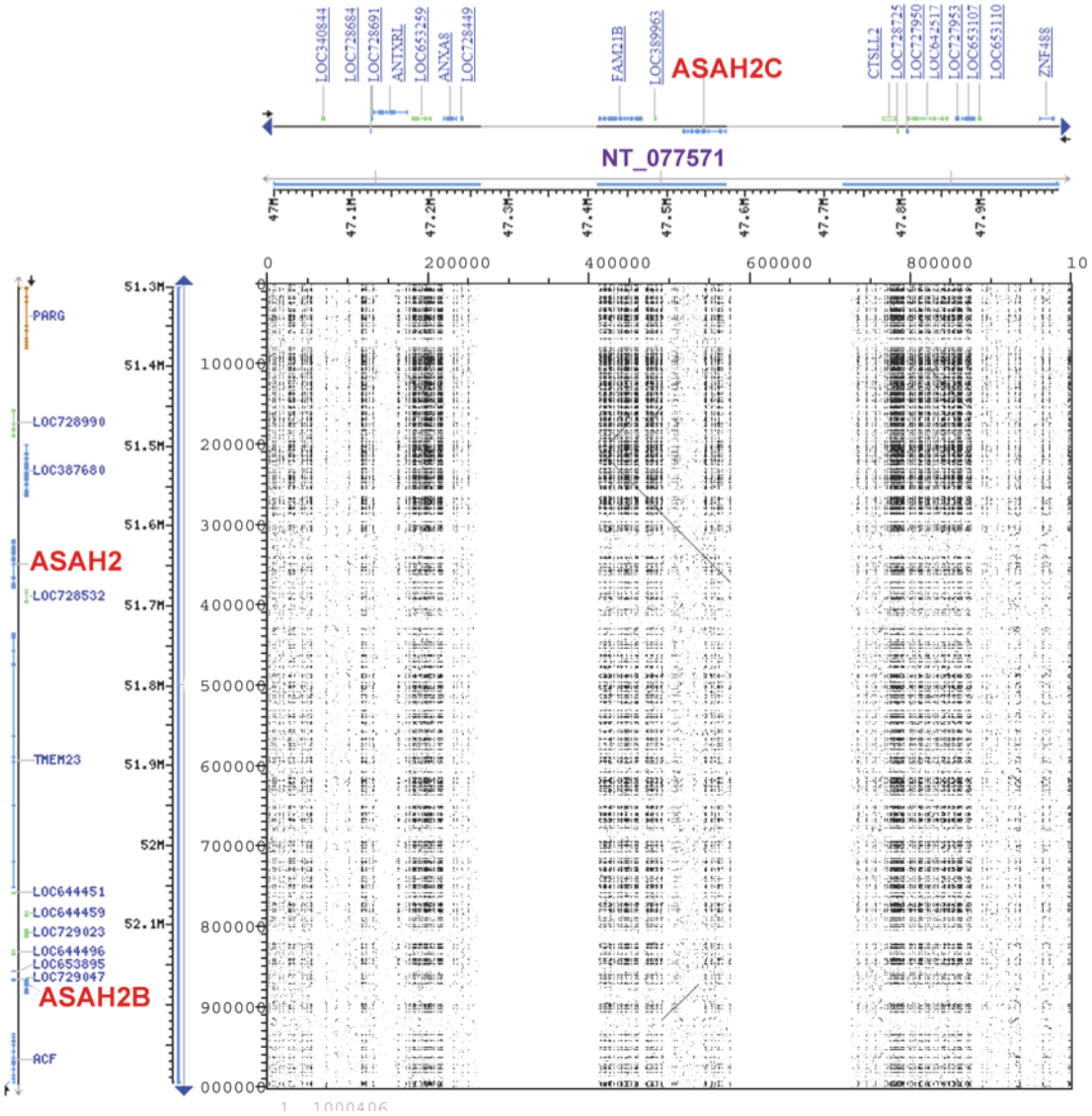
region in humans and *LOC466067* gene in chimpanzees (Figure 6-8). The divergence between the human *FAM21B* and chimpanzee *LOC466067* is 1.6%, which is consistent with the divergence of the human and chimpanzee, 1.2% (The chimpanzee sequencing and analysis consortium 2005). In comparison of the region including chimpanzee ASAH2 with the ~ 80 kb region of chimpanzee (Figure 6-9), the ~ 50 kb homologous region was detected. The homologous region corresponds to the *LOC466067* gene region and the upstream (*LOC743843*) of *ASAH2*, indicating the absence of *ASAH2* homolog and that the *ASAH2C* sequence was duplicated in the human lineage specifically.

Figure 6-5. *ASAH2* and two paralogs



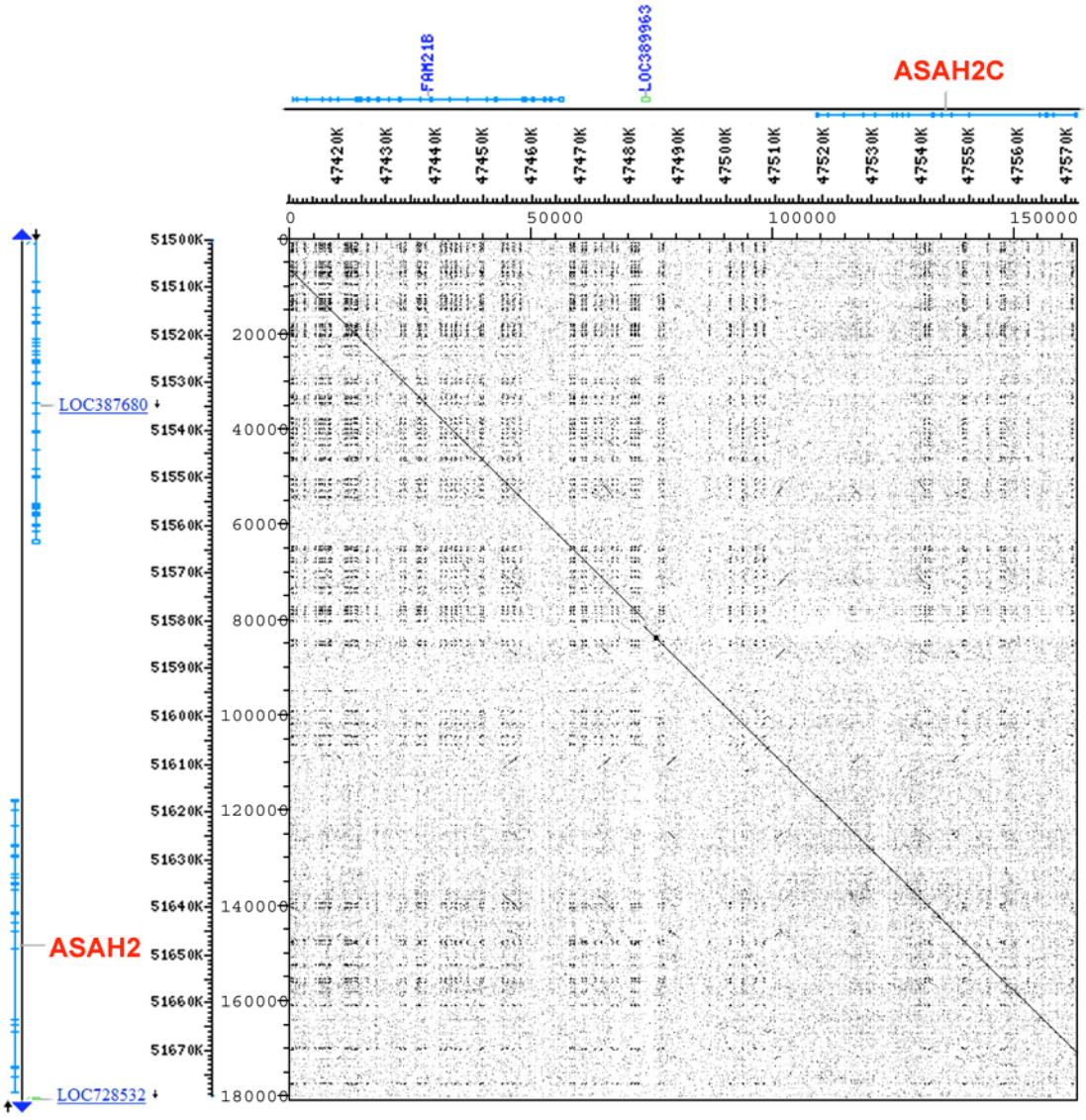
The figure was retrieved from the database of NCBI Entrez Gene (http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene). A red and blue bar represent coding and untranslated region, respectively.

Figure 6-6. Comparison of two loci, 47 – 48 Mb region and 51.3 – 52.3 Mb region on the human chromosome 10



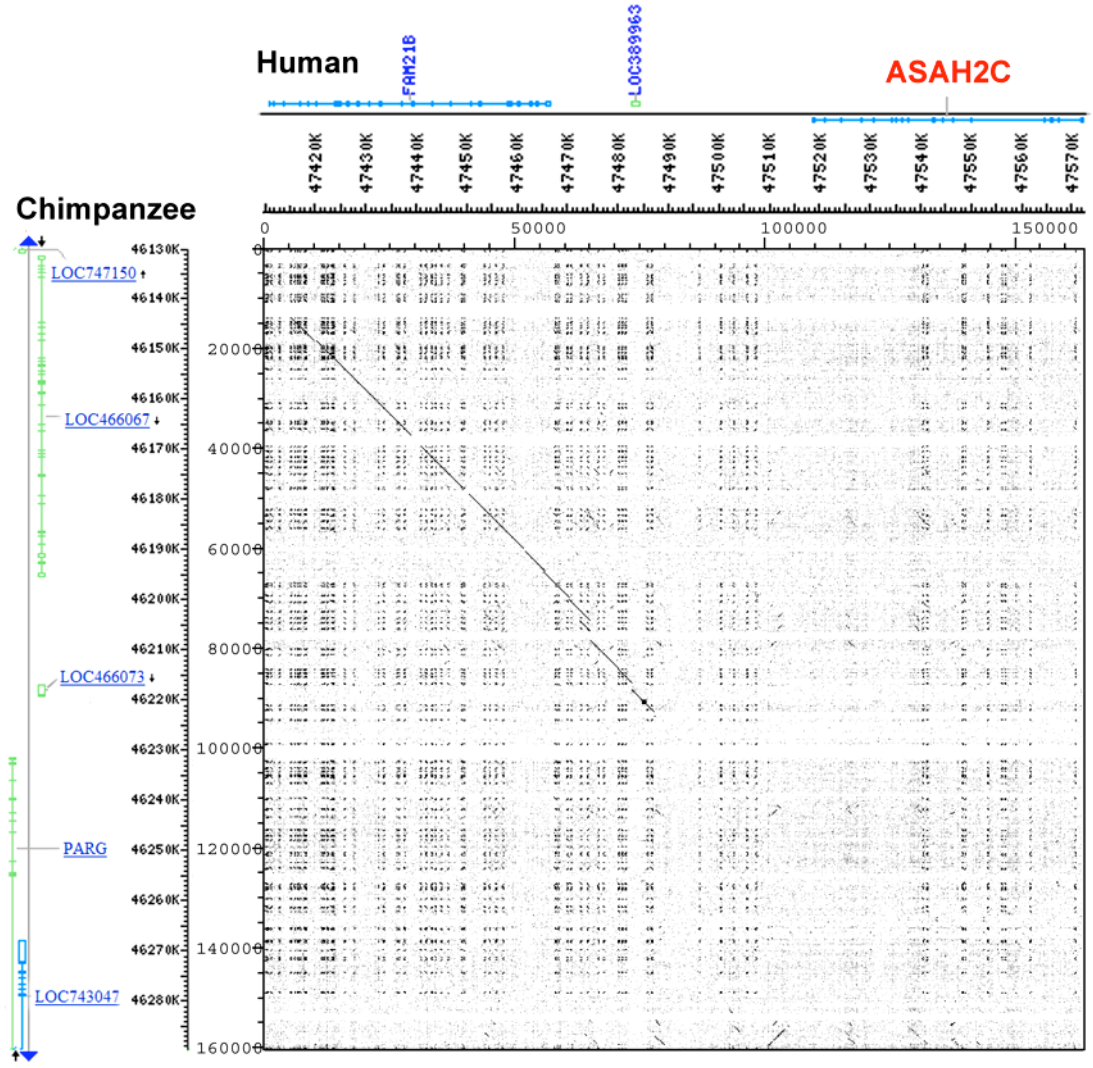
Two 1 Mb loci of genomic sequences were compared with each other based on a sliding window of 35 bp and a dot in a box represents identity larger than 50%. (This window size is same for the following Figures, 6-7 to 6-9.) Along two axes, a colored line marks a coding region with gene symbol, and a blue line on a chromosome position map indicates a position and length of contig. In the locus of 47 – 48 Mb on chromosome 10, there are two long sequence gaps within contigs. Then the white area in the box was resulted due to the gap.

Figure 6-7. Comparison of two loci, 47.41 – 47.57 Mb region and 51.5 – 51.68 Mb region on the human chromosome 10



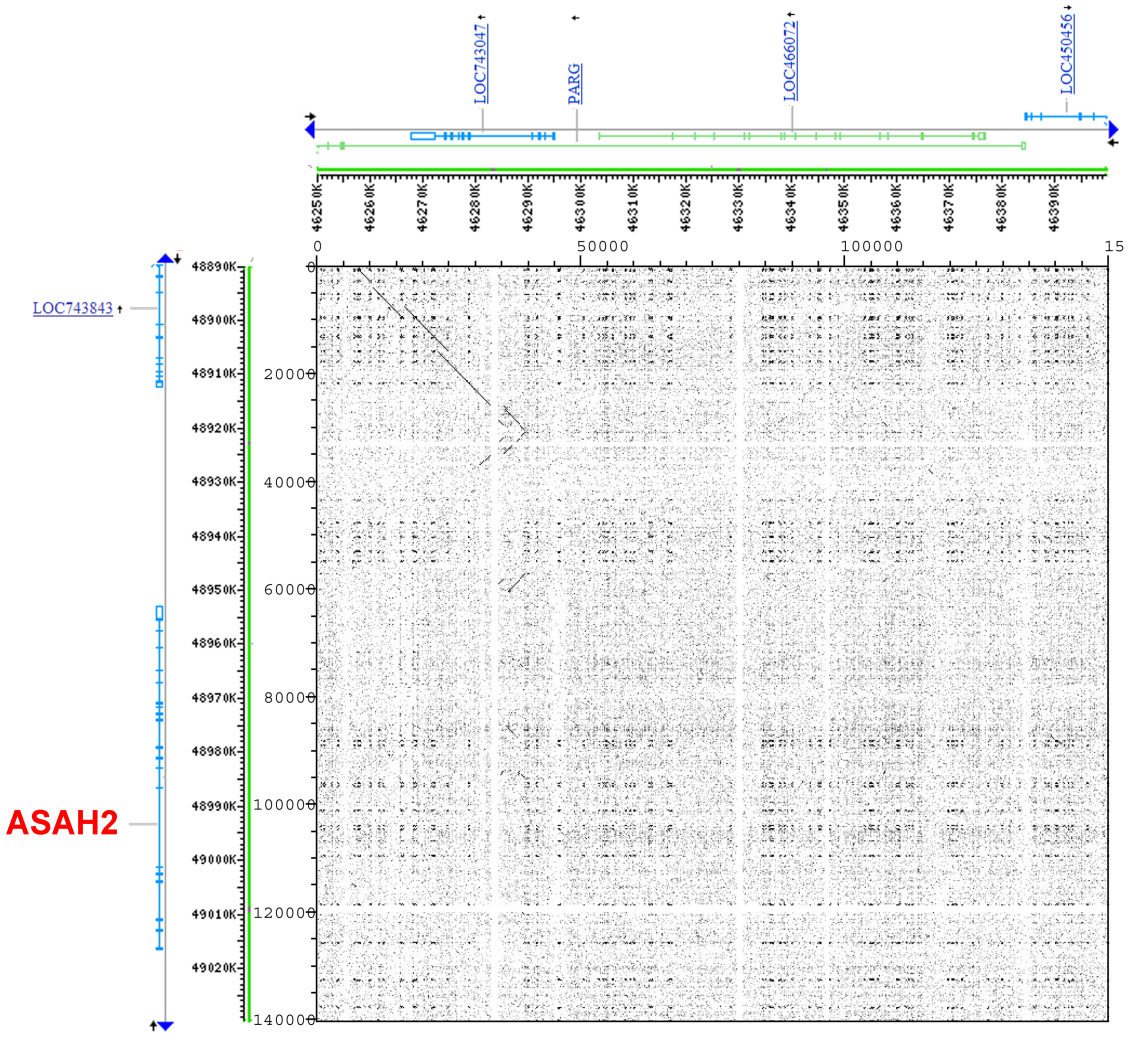
The dot-plot was constructed based on the comparison of two independent loci on the human chromosome 10. The diagonal line in the box indicates conserved nucleotide sequences between two regions of around 160 kb-length.

Figure 6-8. Comparison of 47.41 – 47.57 Mb region on the human chromosome 10 with 46.13 – 46.29 Mb region on the chimpanzee chromosome 10



The dot-plot shows conserved genomic sequences of about 80 kb region in the humans and chimpanzee syntenic region.

Figure 6-9. Comparison of 47.41 – 47.57 Mb region with 46.13 – 46.29 Mb region on the chimpanzee chromosome 10



In the dot-plot, diagonal lines show similarity between two loci on the chimpanzee chromosome 10, around 50 kb regions.

6.3. Human specific *ASAH2* paralogous genes

From the above, I detected two paralogous genes of *ASAH2* in the human genome. To understand molecular evolution of two genes, NJ tree was constructed and divergence (p-distance) of all pairwise sequences were calculated by using primate genomic sequences (Figure 6-10A and Table 6-1A). The tree and divergence reveal that *ASAH2C* is closest to human *ASAH2*. However, the tree is contradictory to the expectation that human *ASAH2B* is clustered with chimpanzee *ASAH2Bp* which is orthologous gene. The nucleotide divergence of human *ASAH2-ASAH2B* is smaller than that of *ASAH2B-ASAH2Bp* (Table 6-1A). Figure 6-11 shows that distances between human *ASAH2* and human *ASAH2B*, and between human *ASAH2* and *ASAH2C* are not constant along the aligned region. The region could be separated into two regions, R1 and R2. The divergence between *ASAH2* and *ASAH2B* is larger in R1 than R2, and the divergence between *ASAH2* and *ASAH2C* is smaller in the R1 than R2. This inconsistence depending on the region suggests different evolutionary history. The tree and divergence in the R1 region reveal that the human *ASAH2* gene was duplicated to the *ASAH2B* gene in an ancestor of the human and the chimpanzee, and to *ASAH2C* after the divergence of humans and chimpanzees. (Figure 6-10B and Table 6-1B). For the R2 region however, divergence of *ASAH2-ASAH2B* (0.2%) is lower than divergence of the human-chimpanzee, 1.2% (The chimpanzee sequencing and analysis consortium 2005), and divergence of *ASAH2-ASAH2C* (0.5%) is larger than the divergence of *ASAH2-ASAH2B*. (Figure 6-10C and Table 6-1C). For the ~ 20 kb region, the divergence between human 3'-*ASAH2* and 3'-*ASAH2C* nucleotide sequences is 0.3%, and one between human 3'-*ASAH2* and 3'-*ASAH2B* is 1.1%. The divergence of these intergeneic regions is consistent with the history of the R1 (Figure 6-12). The divergence of the R2 region is incompatible with the history of the R1 and intergenic

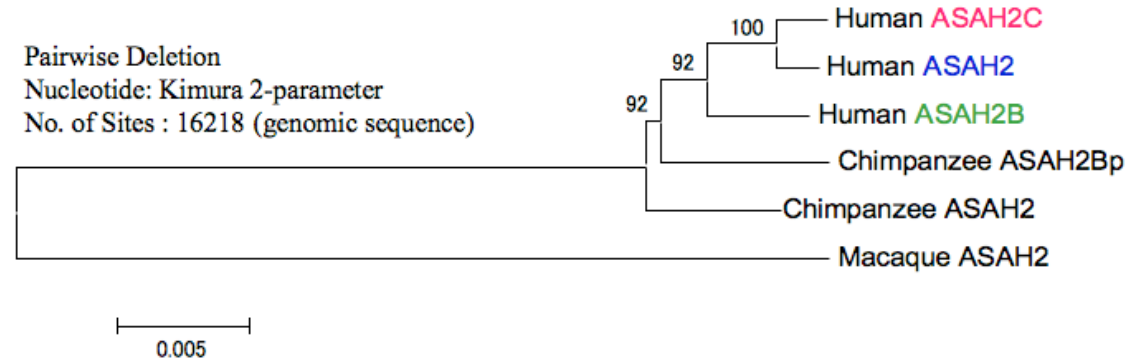
region, suggesting some sequence interactions among paralogs in the region. It is likely that human *ASAH2B* acquired new nucleotide sequences in the human lineage, and *ASAH2B* is human specific. Regarding human *ASAH2* and *ASAH2C*, divergence in the R1 is 0.1% (Table 6-1B). Assuming the divergence time of the human and chimpanzee is 6 mya, *ASAH2C* was duplicated at approximately 500 kya.

ASAH2B sequences are found in both humans and chimpanzees but the chimpanzee gene is likely a pseudogene. The sequence (*ASAH2Bp*) has not been annotated in the NCBI database, and transcriptional information of *ASAH2Bp* is unknown. In contrast, the transcription of human *ASAH2B* has been examined by Avramopoulos *et al.* (2007), and they reported possible function of human *ASAH2B*. *ASAH2B* expression was detected clearly in brain, whereas *ASAH2* expression was strong in the colon and small intestine and very weak in other tissues (Avramopoulos *et al.* 2007, Figure 3). In addition, they found that *ASAH2B* expression reduced in late onset Alzheimer's disease patients. These results might suggest that *ASAH2B* gene acquired new functions in brain after the emergence. Regarding *ASAH2C* transcription, any premature stop codon is not present in the sequence and two or three EST sequences corresponding to some exons of *ASAH2C* have been registered in the NCBI database. Therefore, *ASAH2C* may actually express and function in some tissues.

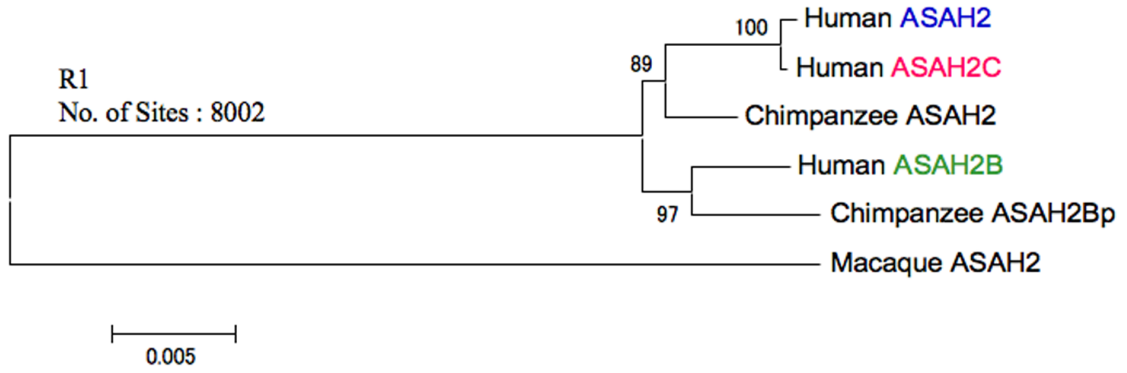
In this chapter, two *ASAH2* paralogs, *ASAH2B* and *ASAH2C* are identified in the human genome. Human *ASAH2B* has likely acquired human-specific nucleotide sequences by gene conversion-like events and has been alive in contrast to chimpanzee one. In addition, it is suggested that *ASAH2B* expression probably has a role in brain by the recent study. The birth of *ASAH2C* is quite recent and certainly human specific gene. These genes should be attractive target for studying humanization.

Figure 6-10. Neighbor-Joining tree of *ASAH2* paralogs of primates

A



B



C

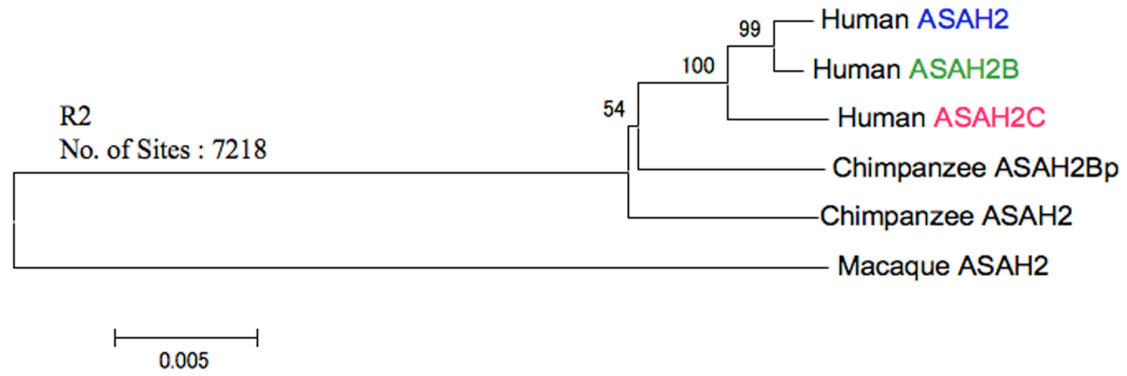
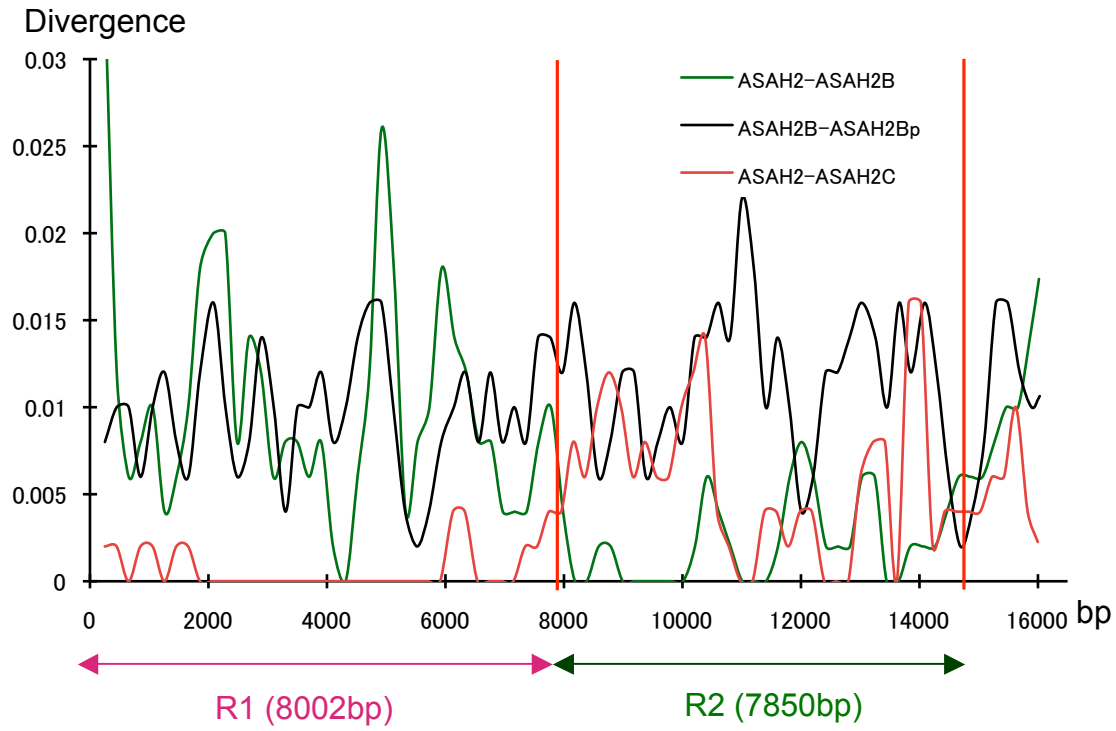


Figure 6-11. Window analysis of divergence and the alignments of ASAH2 paralogs

A



B

R1

```
Human_ASAH2     AGCTTGTTCA AATACGGCAG CGAGCCAATG AAGAACGGGG GATTATGGCA GCCCACCTCC
Human_ASAH2C    .......... ....T....A ...C...... .......... .......... ..........
Human_ASAH2B    TAAGAACATG GGCGTTAAGA TAG.ATTGCC CCCGGTAAAA AGCCCAATTG ATTTTATCTA
Chimp_ASAH2Bp   TAAGAACATG G.CGT.AAGA .AG.ATT... CCC...A.AA AGCCCAA..G A.T..AT.TA
Chimp_ASAH2     .........G G.CCT..AGA .AGCA.T... C.CG..A.AA A.CCC..... A.T...T.T.
Macaca_ASAH2    .C.......G A.CGG.AAGA .AGCATT..A C.....A.AA AGCCC.A... A.T..AT.T.
```

R2

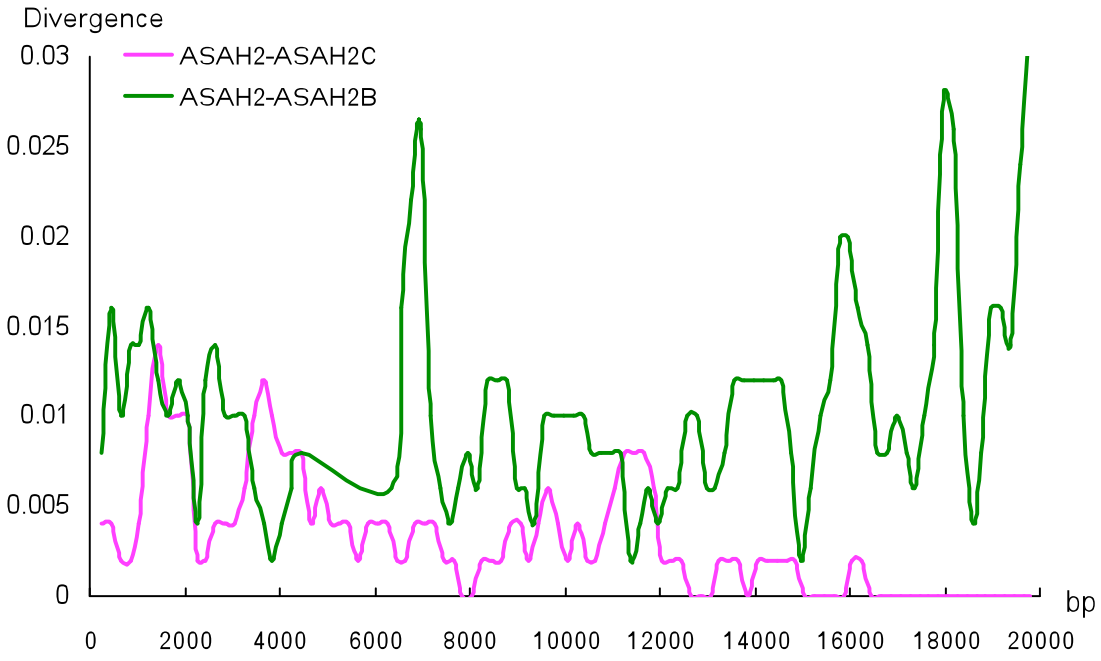
```
Human_ASAH2     CTCCCTAAGC TAGGCGCCGA CAAGTGGTGA CTAGCAATTG AGCCCCTCCG AGCCAAGTGA
Human_ASAH2C    .......... .....A.A.. ....C.CCTG TCGATGGCCC GATTTTCTT. GAAT..AC.G
Human_ASAH2B    AATTTCGGAT CCAAT.TATG TGCTCTC... ..G....... .........C GA..GGACAG
Chimp_ASAH2Bp   AATTTCGGAT .CA.....T. .GC.CTC.TG ..GA..G.C. ..T...C... .A...GAC..
Chimp_ASAH2     .....C..AT ........T. ..C.CTC.TG T.GAT.GCC. ..T...C... .A...G.C.G
Macaca_ASAH2    .....CGGAT .CA..A..T. .GT.CTC.TG -..A..G.CC .AT...CT.- .A.T.GACA-
```

```
Human_ASAH2     TCGTTTTTTT ----ACTCTA GCTAGTTAGT AGAATTA
Human_ASAH2C    CTAAAAAAAA AGAAG..GCG .G..ACCT.C .......
Human_ASAH2B    CT........ TATT.TC.CG A.CC.CCTA. GAGGCGG
Chimp_ASAH2Bp   .-A..----- ----.T..CG AGC.A.CTA. G.GGC.G
Chimp_ASAH2     .TA------- ----.T..CG AGC.A.CT.. G.GGC..
Macaca_ASAH2    --A.....-- ----.T...G AGC.ACCT.. G.GG...
```

A. The Y axis represents pair-wise divergence. A window size of the divergence and step size is 500bp and 200bp, respectively. The X axis indicates distance in the nucleotide sequences of an alignment of paralogous region of human *ASAH2*, *ASAH2B*, and *ASAH2C*. The region is classified into two regions, R1 and R2, and a red line indicates the boundary of the regions.

B. The alignment of *ASAH2* paralogs in primates where only segregation sites are shown. In the first line of the human *ASAH2* sequence, pink letters indicate nucleotides same with that of *ASAH2C*, green letters indicate those same with *ASAH2B,* and, blue letters indicate *ASAH2*-specific nucleotides among three sequences. Dots mean nucleotides same to that of the human *ASAH2*.

Figure 6-12. Window analysis of divergence at the 3' region of *ASAH2*, *ASAH2B*, and *ASAH2C*



The Y axis represents pair-wise divergence of paralogous 3' region of human *ASAH2*, *ASAH2B* and *ASAH2C*. A window size of the divergence and step size is 500bp and 200bp, respectively. The X axis indicates position in the ~ 20 kb region. In the 17 ~ 20 kb region, the divergence between 3' region of human *ASAH2* and *ASAH2C* is zero.

Table 6-1. p-distance of *ASAH2* paralogous genes in primates

A. Number of sites: 15797 (Complete deletion)

		1	2	3	4	5	6
1	Human_ASAH2						
2	Human_ASAH2C	0.003					
3	Human_ASAH2B	0.007	0.008				
4	Chimpanzee_ASAH2Bp	0.013	0.014	0.010			
5	Chimpanzee_ASAH2	0.010	0.010	0.011	0.011		
6	Macaque_ASAH2	0.060	0.060	0.060	0.061	0.059	

B. Number of sites: 7828 (Complete deletion)

R1		1	2	3	4	5	6
1	Human_ASAH2						
2	Human_ASAH2C	0.001					
3	Human_ASAH2B	0.011	0.011				
4	Chimpanzee_ASAH2Bp	0.008	0.008	0.010			
5	Chimpanzee_ASAH2	0.014	0.014	0.009	0.011		
6	Macaque_ASAH2	0.064	0.064	0.065	0.062	0.065	

C. Number of sites: 6983 (Complete deletion)

R2		1	2	3	4	5	6
1	Human_ASAH2						
2	Human_ASAH2C	0.005					
3	Human_ASAH2B	0.002	0.005				
4	Chimpanzee_ASAH2Bp	0.012	0.012	0.011			
5	Chimpanzee_ASAH2	0.012	0.013	0.012	0.012		
6	Macaque_ASAH2	0.056	0.056	0.056	0.056	0.057	

CHAPTER

7. SUMMARY AND SIGNIFICANCE

It has been remained to a lot of questions regarding the human evolution, despite of a lot of studies. The purpose of this thesis is to understand the human evolution, in particular specificity of human mental activity, from a view of molecular and population genetic analyses. I have felt a great interest in 'the out of Africa', because the archeological records show that behavioral modernity of humans was revealed in the period of "the out of Africa". Modern humans subsequently have adapted to new and diverse environments in the whole world and the population of modern humans has expanded rapidly. If the attaining the advanced mental activity of humans is not independent to the dispersal of modern humans, we could expect that natural selection has operated on genes related to mental activity recently, before the out of Africa and after the split of *Homo sapiens* and Neanderthals. However, it has not been reported to genes concerned to acquire the human modernity.

In this thesis, I have focused on the evolution of genes related to sphingolipid (SL) metabolism as candidate genes on which positive Darwinian selection operated in the human lineage. The genes regulate to neuronal developments and are associated with mental retardations. The chapter 1 mentioned the aim of this thesis and introduced background of human modernity and SL, and details of materials and methods for the thesis are described in the chapter 2.

To identify target genes, the long range-haplotype test was carried out and it is described the purpose and results of the test, in the chapter 3. The test applied for eight genes associated with the lipid storage disease which is the deficiency of genes related

to SL metabolism. The test showed a significant result that extended LD of a particular haplotype of *ASAH1* is longer and stronger than the expectation under neutral evolution, suggesting rapid expansion of the haplotype in the human population.

In the chapter 4, the inference of human demography is discussed. To examine the pattern and mode of *ASAH1* polymorphism in human population, I performed population genetic analyses by using the nucleotide sequences (~ 11 kb) at *ASAH1* from the 60 world-wide samples. For the sequenced region, two distinct lineages (V and M) were detected and are coalesced ~ 2.4 mya. Computer simulations and the ancient TMRCA of other several loci suggest the human demography that the human population was structured geographically and differentiated genetically in Africa before the Pleistocene period.

In the chapter 5, I detected signatures of positive Darwinian selection for the V lineage. All haplotypes of the V lineage show small genetic diversity despite of a high frequency of the haplotypes in a population. Human specific amino acid of Val at M72V shared in the V lineage, and could be a possible target of positive selection. The result of computer simulations is consistent with positive selection for an advantageous allele of the V lineage against to the M lineage. The selection might have operated since the out of Africa of modern humans, likely corresponding to the admixture of two lineages.

While in the chapter 6, I performed analyses of molecular evolution for *ASAH1* and four related genes (*ASAH2, ASAH2L, ASAH3,* and *ASAH3L*) encoding kinds of ceramidase. Phylogenetic analyses revealed that the origin of five genes is before the split of vertebrates and invertebrates. Further, I examined two *ASAH2* paralogs of *ASAH2B* and *ASAH2C,* and found that *ASAH2C* was born in the human lineage, and *ASAH2B* is expressed genes in only humans. It was reported that the transcription of

ASAH2B in the brain of Alzheimer's disease patients significantly decrease. These reveal possible functions of *ASAH2B* gene in the human brain.

Finally, this thesis has suggested the human specific evolution of some ceramidase, and the demographic history of humans. I expect that findings in the thesis are certainly worthy for comprehensions regarding the human evolution in particular, the evolution of modern humans.

REFERENCES

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman, 1990 Basic local alignment search tool. J. Mol. Biol. **215:** 403-410.

Aquadro, C. F., D. J. Begun and E. C. Kindahl, 1994 Non-neutral evolution, pp. 46-56 in *Selection, recombination and DNA polymorphism in Drosophila*, edited by B. Golding. Chapman & Hall, London.

Avramopoulos, D., R. Wang, D. Valle, M. D. Fallin and S. S. Bassett, 2007 A novel gene derived from a segmental duplication show perturbed expression in Alzheimer's disease. Neurogenetics **8:** 111-120.

Baird, D. M., J. Coleman, Z. H. Rosser, and N. J. Royle, 2000 High levels of sequence polymorphism and linkage disequilibrium at the telomere of 12q: implications for telomere biology and human evolution. Am. J. Hum. Genet. **66:** 235-250.

Bar, J., T. Linke, K. Ferlinz, U. Neumann, E. H. Schuchman *et al*, 2001 Molecular analysis of acid ceramidase deficiency in patients with Farber disease. Hum. Mutat. **17:** 199-209.

Barreiro, L. B., E. Patin, O. Neyrolles, H. M. Cann, B. Gicquel *et al*, 2005 The heritage of pathogen pressures and ancient demography in the human innate-immunity CD209/CD209L region. Am. J. Hum. Genet. **77:** 869-886.

Barrett, J. C., B. Fry, J. Maller and M. J. Daly, 2005 Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. **15:** 263-265.

Beutler, E. and G. A. Grabowski, 2001 Gaucher disease, pp. 3635-3668 in *The Metabolic and Molecular Bases of Inherited Disease*, edited by C. R. Scriver, A. L. Beaudet, W. S. Sly and D. Vialle. McGraw-Hill, New York.

Bollinger, C. R., V. Teichgräber and E. Gulbins, 2005 Ceramide-enriched membrane domains. Biochimica et Biophysica Acta **1746:** 284-294.

Boquet, I., R. Hitier, M. Dumas, M. Chaminade and T. Préat, 2000 Central brain postembryonic development in *Drosophila*: Implication of genes expressed at the interhemispheric junction. J. Neurobiol. **42:** 33-48.

Bouzouggar, A., N. Barton, M. Vanhaeren, F. d'Errico, S. Collcutt *et al*, 2007 82,000-year-old shell beads from North Africa and implications for the origins of modern human behavior. Proc. Natl. Acad. Sci. U. S. A. **104:** 9964-9969.

Buccoliero, R., and A. H. Futerman, 2003 The roles of ceramide and complex sphingolipids in neuronal cell function. Pharmacol. Res. **47:** 409-419.

Buccoliero, R., J. Bodennec and A. H. Futerman, 2002 The role of sphingolipids in neuronal development: Lessons from models of sphingolipid storage diseases. Neurochem. Res. **27:** 565-574.

Cochran, G. J., Hardy and H. Harpending, 2006 Natural history of Ashkenazi intelligence. J. Biosoc. Sci. 38: 659-693.

Cox, M. P., F. L. Mendez, R. M. Karafet, M. M. Pilkington, S. B. Kingan *et al*, 2008 Testing for archaic Hominin admixture on the X-chromosome: Model likelihoods for the modern human RRM2P4 region from summaries of genealogical topology under the structured coalescent. Genetics.

Currat M., L. Excoffier, W. Maddison, S. P. Otto, N. Ray *et al*, 2006 Comment on "Ongoing adaptive evolution of ASPM, a brain size determinant in Homo sapiens" and "Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans". Science **313:** 172.

d'Errico, F., C. Henshilwood, M. Vanhaeren and K. van Niekerk, 2005 Nassarius kraussianus shell beeds from Blombos cave: evidence for symbolic behaviour in the Middle Stone Age. J. Hum. Evol. **48:** 3-24.

Devi, A. R. R., M. Gopikrishna, R. Ratheesh, G. Savithri, G. Swarnalata *et al*, 2006 Farber lipogranulomatosis: clinical and molecular genetic analysis reveals a novel mutation in an Indian family. J. Hum. Genet. **51:** 811-814.

Diamond, J. M., 1994 Human genetics: Jewish lysosomes. Nature **368:** 291-292.

Ding, Y. C., H. C. Chi, D. L. Grady, A. Morishima, J. R. Kidd *et al*, 2002 Evidence of positive selection acting at the human dopamine receptor D4 gene locus. Proc. Natl. Acad. Sci. U. S. A. **99:** 309-314.

Enard, W., M. Przeworski, S. E. Fisher, C. S. Lai, V. Wiebe *et al*, 2002 Molecular evolution of FOXP2, a gene involved in speech and language. Nature **418:** 869-872.

Evans, P. D., J. R. Anderson, E. J. Vallender, S. S. Choi and B. T. Lahn, 2004 Reconstructing the evolutionary history of microcephalin, a gene controlling human brain size. Hum. Mol. Genet. **13:** 1139-1145.

Evans, P. D., S. L. Gilbert, N. Mckel-Bobrow, E. J. Vallender, J. R. Anderson *et al,* 2005 Microcephalin, a gene regulating brain size, contributes to evolve adaptively in humans. **309:** 1717-1720.

Evans, P. D., N. Mekel-Bobrov, E. J. Vallender, R. R. Hudson and B. T. Lahn, 2006 Evidence that the adaptive allele of the brain size gene *microcephalin* introgressed into *Homo Sapiens* from an archaic *Homo* lineage. Proc. Natl. Acad. Sci. U. S. A. **103:** 18178-18183.

Feany, M. B., and W. W. Bender, 2000 A Drosophila model of Parkinson's disease. Nature **404:** 394-398.

Frisch, A., R. Colombo, E. Michaelovsky, M. Karpati, B. Goldman and L. Peleg 2004 Origin and spread of the 1278insTATC mutation causing Tay-Sachs disease in Ashkenazi Jews: genetic drift as a robust and parsimonious hypothesis. Hum. Genet. **114:** 366-376.

Futerman, A. H., and G. Meer, 2004 The cell biology of lysosomal storage disorder. Nature **5:** 554-565.

Futerman, A. H., and H. Riezman, 2005 The ins and outs of sphingolipid synthesis. Trends Cell Biol. **15:** 312-318.

Futerman, A. H., and Y. A. Hannun, 2004 The complex life of simple sphingolipids. EMBO Rep. **5:** 777-782.

Gabriel, S. B., S. F. Schaffner, H. Nguyen, J. M. Moore, J. Roy *et al,* 2002 The structure of haplotype blocks in the human genome. Science **296:** 2225-2229.

Garrigan, D., and M. F. Hammer, 2006 Reconstructing human origins in the genomic era. Nat. Rev. Genet. **7:** 669-680.

Garrigan, D., Z. Mobasher, S. B. Kingan, J. A. Wilder and M. F. Hammer, 2005 Deep haplotype divergence and long-range linkage disequilibrium at xp21.1 provide

evidence that humans descend from a structured ancestral population. Genetics **170:** 1849-1856.

Garrigan, D., Z. Mobasher, T. Severson, J. A. Wilder and M. F. Hammer, 2005 Evidence for archaic Asian ancestry on the human X chromosome. Mol. Biol. Evol. **22:** 189-192.

Gatt, S., 1963 Enzymic hydrolysis and synthesis of ceramides. J. Biol. Chem. **238:** 3131-3133.

Gilad, Y., S. Rosenberg, M. Przeworski, D. Lancet and K. Skorecki, 2002 Evidence for positive selection and population structure at the human MAO-A gene. Proc. Natl. Acad. Sci. U. S. A. **99:** 862-867.

Goldstein, D. B., D. E. Reich, N. Bradman, S. Usher, U. Seligsohn and H. Peretz, 1999 Age estimates of two common mutations causing factor XI deficiency: recent genetic drift is not necessary for elevated disease incidence among Ashkenazi Jews. Am. J. Hum. Genet. **64:** 1071-1075.

Griffiths, R. C., and S. Tavaré, 1995 Unrooted genealogical tree probabilities in the infinitely-many-sites model. Math. Biosci. **127:** 77-98.

Hannun, Y. A., and C. Luberto, 2004 Lipid metabolism: ceramides transfer protein adds a anew dimension. Curr. Biol. **14:** R163-R165.

Hannun, Y. A., and L. M. Obeid, 2002 The ceramide-centric universe of lipid-mediated cell regulation: Stress encounters of the lipid kind. J. Biol. Chem. **277:** 25847-25850.

Harding, R. M. S. M. Fullerton, R. C. Griffiths, J. Bond, M. J. Cox *et al,* 1997 Archaic African and Asian lineages in the genetic ancestry of modern humans. Am. J. Hum. Genet. **60:** 772-789.

Harris, E. E., and J. Hey, 1999 X chromosome evidence for ancient human histories. Proc. Natl. Acad. Sci. U. S. A. **96:** 3320-3324.

Harris, E. E., and J. Hey, 2001 Human populations show reduced DNA sequence variation at the factor IX locus. Curr. Biol. **11:** 774-778.

Hayakawa, T., I. Aki, A. Varki, Y. Satta and N. Takahata, 2006 Fixation of the human-specific CMP-N-acetylneuraminic acid hydroxylase pseudogene and implications of haplotype diversity for human evolution. Genetics **172:** 1139-1146.

Henshilwood, C. S., F. d'Errico, R. Yates, Z. Jacobs, C. Tribolo *et al,* 2002 Emergence of modern human behavior: Middle Stone Age engravings from South Africa. Science **295:** 1278-1280.

Holloway, R. L., 1966 Dendritic branching: some preliminary results of training and complexity in rat visual cortex. Brain Research **2:** 393–396.

Huang, Y., H. Tanimukai, F. Liu, K. Iqbal, I. Grundke-Iqbal and C. X. Gong, 2004 Elevation of the level and activity of acid ceramidase in Alzheimer's disease brain. Eur. J. Neurosci. **20:** 3489-3497.

Hudson, R. R., M. Kreitman and M. Aguadé, 1987 A test of neutral molecular evolution based on nucleotide data. Genetics **116:** 153-159.

Hudson, R. R., 2002 Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics **18:** 337-338.

Ina, Y., 1995 New methods for estimating the numbers of synonymous and nonsynonymous substitutions. J. Mol. Evol. **40:** 190-226.

Inlow, J. K., and L. L. Restifo, 2004 Molecular and comparative genetics of mental retardation. Genetics **166:** 835-881.

International HapMap Consortium, 2005 A haplotype map of the human genome. Nature **437:** 1299-1320.

International HapMap Consortium, 2007 A second generation human haplotype map of over 3.1 million SNPs. Nature **449:** 852-861.

Jackson, G. T., I. Salecker, X. Dong, X. Yao, N. Arnheim et al., 1998 Polyglutamine-expanded human *huntingtin* transegenes induce degeneration of Drosophila photoreceptor neurons. Neuron **21:** 633-642.

Kimura, M., 1983 The Neutral Theory of Molecular Evolution. Cambridge Univ. press, New york.

Kivisild, T., P. Shen, D. P. Wall, B. Do, R. Sung *et al*, 2006 The role of selection in the evolution of human mitochondrial genomes. Genetics **172:** 373-387.

Klein, J., and N. Takahata, 2002 Where Do We Come From? The Molecular Evidence for Human Descent. Springer-Verlag, Berlin.

Klein, R. G., 1999 The human career: Human biological and cultural origin 2nd, pp. 495-573 in *Anatomically Modern Humans*, The University of Chicago Press. Ltd., London.

Klein, R. G., 2000 Archeology and the Evolution of Human Behavior. Evolutionary Anthropology **9:** 17-36.

Koch, J., S. Gärtner , C. M. Li, L. E. Quintern, L. Bernardo *et al,* 1996 Molecular cloning and characterization of a full-length complementary DNA encoding human acid ceramidase. Identification of the first molecular lesion causing Farber disease. J. Biol. Chem. **271:** 33110-33115.

Kono, M., J. L. Dreier, J. M. Ellis, M. L. Allende, D. N. Kalkofen *et al*, 2006 Neutral ceramidase encoded by the Asah2 gene is essential for the intestinal degradation of sphingolipids. J. Biol. Chem. **281:** 7324-7331.

Kumar, S., K. Tamura and M. Nei, 2004 MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform **5:** 150-163.

Lewontin R. C., 1964 The interaction of selection and linkage. Ⅰ. General considerations; heterotic models. Genetics **49;** 49-67.

Li, C. M., J. H. Park, X. He, B. Levy, F. Chen *et al,* 1999 The human acid ceramidase gene (ASAH): structure, chromosomal location, mutation analysis, and expression. Genomics **62:** 223-231.

Macaulay, V., C. Hill, A. Achilli, C. Rengo, D. Clarke *et al*, 2005 Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. Science **308:** 1034-1036.

Martínez-Arias, R., F. Calafell, E. Mateu, D. Comas, A. Andrés and J. Bertranpeti, 2001 Sequence variability of a human pseudogene. Genome Res. **11:** 1071-1085.

Marth, G. T., E. Czabarka, J. Murvai and S. T. Sherry, 2004 The allele frequency spectrum in genome-wide human variation data reveals signals of differential demographic history in three large world populations. Genetics. **166:** 351-372.

Mellars, P., 2006 Going East: New genetic and archaeological perspectives on the modern human colonization of Eurasia. Science **313:** 796-800.

Mellars, P., 2006 Why did modern human populations disperse from Africa ca. 60,000 years ago? A new model. Proc. Natl. Acad. Sci. U. S. A. **103:** 9381-9386.

Merrill, A. H. Jr, 2002 De novo sphingolipid biosynthesis: A necessary, but dangerous, pathway. J. Biol. Chem. **277:** 25843-25846.

Mervis, C. B., and A. M. Becerra, 2007 Language and communicative development in Williams syndrome. MRDD Res. Rev. **13:** 3-15.

Morton, N. E., 1955 Sequential tests for the detection of linkage. Am. J. Hum. Genet. **7:** 277-318.

Moser, H.W., 1995 Ceramidase deficiency: Farber lipogranulomatosis, pp. 2589-2599 in *The Metabolic Basis of Inherited Disease 7th ed,* edited by C. R. Scriver, A. L. Beaudet, W. S. Sly and D. Valle, McGraw-Hill, New York.

Motulsky, A. G. 1995 Jewish disease and origins. Nat. Genet. **9:** 99-101.

Muramatsu, T., N. Sakai, I. Yanagihara, M. Yamada, T. Nishigaki *et al*, 2002 Mutation analysis of the acid ceramidase gene in Japanese patients with Farber disease. J. Inherit. Metab. Dis. **25:** 585-592.

Myers, S., L. Bottolo, C. Freeman, G. McVean and P. Donnelly, 2005 A fine-scale map of recombination rates and hotspots across the human genome. Science **310:** 321-324.

Nei, M., 1983 Genetic polymorphism and the role of mutation in evolution, pp. 165-190 in *Evolution of Genes and Protein*, edited by M. Nei and R. K. Koehn. Sinauer Assoc. Inc., Massachusetts.

Nei, M., and N. Takahata, 1993 Effective population size, genetic diversity, and coalescence time in subdivided populations. J. Mol. Evol. **37:** 240-244.

Nei, M., and T. Gojobori, 1986 Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3:** 418-426.

Nei, M., and W. H. Li, 1979 Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. U. S. A. **76:** 5269-5273.

Niell, B.L., J. C. Long, G. Rennert, S. B. Gruber, 2003 Genetic anthropology of the colorectal cancer-susceptibility allele APC I1307K: evidence of genetic drift within the Ashkenazim. Am. J. Hum. Genet. **73:** 1250-1260.

Nikolova-Karakashian, M. and A. H. Merrill Jr, 2000 Ceramidases. Methods Enzymol. **311:** 194-201.

Nilsson. A. and R. D. Duan, 2006 Absorption and lipoprotein transport of sphingomyelin. J. Lipid Res. **47:** 154-171.

Nordborg, M. 1998 On the probability of Neanderthal ancestry. Am. J. Hum. Genet. **63:** 1237-1240.

Nordborg, M. 2001 On detecting ancient admixture, pp. 123-136 in *Genes, Fossils and Behavior* edited by P. Donnelly and R. A. Foley. IOS Press, Amsterdam, the Netherlands.

Pyne, S., 2002 Cellular signaling by sphingosine and sphingosine-1-phosphate. Their opposing roles in apoptosis. Subcell. Biochem. **36:** 245-268.

Reiter, L. T., L. Potocki, S. Chien, M. Gribskov and E. Bier, 2001 A systematic analysis of human disease-associated gene sequences in *Drosophila melanogaster*. Genome Res. **11:** 1114-1125.

Risch, N., D. de Leon, L. Ozelius, P. Kramer, L. Almasy, B. Singer, S. Fahn, X. Breakefield and S. Bressman, 1995 Genetic analysis of idiopathic torsion dystonia in Ashkenazi Jews and their recent descent from a small founder population. Nat. Genet. **9:** 152-159.

Risch, N., H. Tang, H. Katzenstein and J. Ekstein, 2003 Geographic distribution of disease mutations in the Ashkenazi Jewish population supports genetic drift over selection. Am. J. Hum. Genet. **72:** 812-822.

Rother, J., G., Van Echten, G. Schwarzmann and K. Sandhoff, 1992 Biosynthesis of sphingolipids: Dihydroceramide and not sphinganine is desaturated by cultured cells. Biochem. Biophys. Res. Commun. **189:** 14-20.

Rotter, J. I. and J. M. Diamond, 1987 What maintains the frequencies of human genetic diseases? Nature **329:** 289-290.

Rozas, J., J. C. Sanchez-DelBarrio, X. Messeguer and R. Rozas, 2003 DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics **19:** 2496-2497.

Rozen, S., and H. Skaletsky, 2000 Primer3 on the WWW for general users and for biologist programmers, pp. 365-386 in *Bioinformatics Methods and Protocols:*

Methods in Molecular Biology, edited by S. Krawetz and S. Misener. Humana Press, Totowa.

Ruvolo, P. P., 2003 Intracellular signal transduction pathways activated by ceramide and its metabolites. Pharmacol. Res. **47:** 383-392.

Sabeti, P. C., S. F. Schaffner, B. Fry, J. Lohmueller, P. Varilly *et al*, 2006 Positive natural selection in the human lineage. Science **312:** 1614-1620.

Sabeti, P. C., D. E. Reich, J. M. Higgins, H. Z. Levine, D. J. Richter *et al*, 2002 Detecting recent positive selection in the human genome from haplotype structure. Nature **419:** 832-837.

Sachidanandam, R., D. Weissman, S. C. Schmidt, J. M. Kakol, L. D. Stein *et al*, 2001 A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature **409:** 928-933.

Saitou, N., and M. Nei, 1987 The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406-425.

Satta, Y., and N. Takahata, 2004 The distribution of the ancestral haplotype in finite stepping-stone models with population expansion. Mol. Ecol. **13:** 877-886.

Scheet, P., and M. Stephens, 2006 A fast and flexible statistical model for large-scale population genotype data: Applications to inferring missing genotypes and haplotypic phase. Am. J. Hum. Genet. **78:** 629-644.

Schumacher, J., P. Hoffmann, C. Schmäl, G. Schulte-Körne and M. M. Nöthen, 2007 Genetics of dyslexia: the evolving landscape. J. Med. Genet. **44:** 289-297.

Schwarz, A., and A. H. Futerman, 1997 Distinct roles for ceramide and glucosylceramide at different stages of neuronal growth. J. Neurosci. **17:** 2929-2938.

Schwartz, A., E. Rapaport, K. Hirschberg and A. H. Futerman, 1995 A regulatory role for sphingolipids in neuronal growth: inhibition of sphingolipid synthesis and degradation have opposite effects on axonal branching. J. Biol. Chem. **270:** 10990-10998.

Serre, D., A. Langaney, M. Chech, M. Teschler-Nicola, M. Paunovic *et al,* 2004. No evidence of Neanderthal mtDNA contribution to early modern humans. PloS Biol. **2:** E57.

Shimada M. K., K. Panchapakesan, S. A. Tishkoff, A. Q. Nato Jr. and J. Hey, 2007 Divergent haplotypes and human history as reveled in a worldwide survey of X-linked DNA sequence variation. Mol. Biol. Evol. **24:** 687-698.

Simons, K. and E. Ikonen, 1997 Functional rafts in cell membranes. Nature. **387:** 569-572.

Slatkin, M., 2004 A population-genetic test of founder effects and implications for Ashkenazi Jewish diseases. Am. J. Hum. Genet. **75:** 282-293.

Sonnhammer, E. L. and R. Durbin, 1995 A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. Gene 167: GC1-10.

Spiegel, S., O. Cuvillier, L. Edsall, T. Kohama, R. Menzeleev *et al*, 1998 Roles of sphingosine-1-phosphate in cell growth, differentiation, and death. Biochemistry **63:** 69-73.

Stefansson, H., A. Helgason, G. Thorleifsson, V. Steinthorsdottir, G. Masson *et al*, 2005 A common inversion under selection in Europeans. Nat. Genet. **37:** 129-137.

Stephens, M., and P. Donnelly, 2003 A comparison of Bayesian methods for haplotype reconstruction from population genotype data. Am. J. Hum. Genet. **73:** 1162-1169.

Stephens, M., N. J. Smith and P. Donnelly, 2001 A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet. **68:** 978-989.

Sugita, M., J. T. Dulaney and H. W. Moser, 1972 Ceramidase deficiency in Farber's disease (lipogranulomatosis). Science **178:** 1100-1102.

Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585-595.

Takahata, N. and S. R. Palumbi, 1985 Extranuclear differentiation and gene flow in the finite island model. Genetics **109:** 441-457.

Takahata, N., 1990 A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism. Proc. Natl. Acad. Sci. U. S. A. **87:** 2419-2423.

Takahata, N., 1991 Genealogy of neutral genes and spreading of selected mutations in a geographically structured population. Genetics. **129:** 585-595.

Takahata, N., 1993 Allelic genealogy and human evolution. Mol. Biol. Evol. **10:** 2-22.

Takahata, N., 1994 Repeated failures that led to the eventual success in human evolution. Mol. Biol. Evol. **11:** 803-805.

Takahata, N., 1995 A genetic perspective on the origin and history of humans. Annu. Rev. Ecol. Syst. **26:** 343-372.

Tani, M., N. Okino, K. Mori, T. Tanigawa, H. Izu *et al*, 2000 Molecular cloning of the full-length cDNA encoding mouse neutral ceramidase. A novel but highly conserved gene family of neutral/alkaline ceramidases. J. Biol. Chem. **275:** 11229-11234.

Templeton A. R., 2002 Out of Africa again and again Nature **416:** 45-51.

The Chimpanzee Sequencing and Analysis Consortium, 2005 Initial sequence of the chimpanzee genome and comparison with the human genome. Nature **437:** 69-87.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin and D. G. Higgins, 1997 The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25:** 4876-4882.

Voight, B. F., A. M. Adams, L. A. Frisse, Y. Qian, R. R. Hudson *et al*, 2005 Interrogation multiple aspects of variation in a full resequencing data set to infer human population size changes. Proc. Natl. Acad. Sci. U. S. A. 102: 18508-18513.

Walkley, S. U., 2003 Neurobiology and cellular pathogenesis of glycolipid storage diseases. Philos. Trans. of the R. Soc. Lond. B Biol. Sci. **358:** 893-904.

Walkley, S. U., M. Zervas and S. Wiseman, 2000 Gangliosides as modulators of dendritogenesis in normal and storage disease-affected pyramidal neurons. Cereb. Cortex **10:** 1028-1037.

Wang, Y, Q. and B. Su, 2004 Molecular evolution of microcephalin, a gene determining human brain size. Hum. Mol. Genet. **13:** 1131-1137.

Wang, Y, Q., Y.P. Qian, S. Yang, H. Shi, C. H. Liao *et al*, 2005 Accelerated evolution of the pituitary adenylate cyclase-activation polypeptide precursor gene during human origin. Genetics **170:** 801-806.

Watson, E., P. Forster, M. Richards and H. J. Bandelt, 1997 Mitochondrial footprints of human expansions in Africa. Am. J. Hum. Genet. **61:** 691-704.

Williamson, S. H., R. Hernandez, A. Fledel-Alon, L. Zhu, R. Nielsen *et al,* 2005 Simultaneous inference of selection and population growth from patterns of variation in the human genome. Proc. Natl. Acad. Sci. U. S. A. **102:** 7882-7887.

Yoshimura, Y., N. Okino, M. Tani and M. Ito, 2002 Molecular cloning and characterization of a secretory neutral ceramidase of *Drosophila melanogaster*. J. Biochem. (Tokyo) **132:** 229-236.

Yu, F., P. C. Sabeti, P. Hardenbol, Q. Fu, B. Fry *et al*, 2005 Positive selection of a pre-expansion CAG repeat of the human SCA2 gene. PLoS Genet. **1:** e41.

Yu, F., R. S. Hill, S. F. Schaffner, P. C. Sabeti, E. T. Wang *et al,* 2007 Comment on "Ongoing adaptive evolution of ASPM, a brain size determinant in Homo sapiens". Science **316:** 370.

Zhang, J., 2003 Evolution of the human ASPM gene, a major determinant of brain size. Genetics **165:** 2063-2070.

Zhang, J., D. M. Webb and O. Podlaha, 2002 Accelerated protein evolution and origins of human-specific features: Foxp2 as an example. Genetics **162:** 1825-1835.

Zhang, J., H. F. Rosenberg and M. Nei, 1998 Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc. Natl. Acad. Sci. U. S. A. **95:** 3708-1713.

Zietkiewicz, E. V. Yotova, D. Gehl, T. Wambach, I. Arrieta *et al.* 2003 Haplotypes in the dystrophin DNA segment point to a mosaic origin of modern human diversity. Am. J. Hum. Genet. **73:** 994-1015.