A Thesis

for the Degree of Master of Science

Accelerated evolution of genome sequences among phylogenetically diverged species with dN/dS ratio analysis

dN/dS 분석을 통한 계통발생학적 유전자 진화 가속에 대한 고찰

2013 년 8 월

서울대학교 대학원

농생명공학부 동물생명공학전공

안현주

Abstract

Accelerated evolution of genome sequences among phylogenetically diverged species with dN/dS ratio analysis

Hyeonju Ahn

Department of Agricultural Biotechnology

The Graduate School

Seoul National University

dN/dS ratio has been widely used to estimate evolutionary acceleration of genome sequence. dN/dS is the ratio of non-synonymous substitution number in non-synonymous sites (dN) to synonymous substitution number in synonymous sites (dS). Synonymous substitution means mutations that make no difference in coding protein of the gene, and non-synonymous substitution means mutations making a difference in the coding protein. This estimator is used to measure accelerated evolution of orthologous genes of related species, which aimed to reveal evolutionary trend among phylogenetically diverged species.

I

With comparing orthologs of six species, which were human, mouse, horse, dog, cow, and pig, I tried to find genetic evolutionary evidences supporting differences of monotocous and polytocous traits. These species were grouped into three monotocous and three polytocous species: human, horse, and cow as a monotocous group and mouse, dog, and pig as polytocous. In this study, I suggested some candidate genes supposed to correlate with evolutionary difference of the reproductive trait. Genes evolutionally accelerated in each group showed functional differences from functional annotation analysis, and dN/dS values were higher in orthologs of monotocous species than polytocous in general.

I also performed dN/dS ratio analysis on ten primates including human to demonstrate genetic evolution related to brain functions. 41 genes were suggested from the analysis to go through accelerated evolution more in human than other nine primates. Functions of the genes represented possibilities of human evolution directed to improvement of memory and cognitive ability. And two of ten genes highly accelerated in human were associated neuronal disorder diseases, which also indicated evolution in human neuronal functions specifically.

Contents

List of Tables

List of Figures

Chapter 1. Literature review

After human genomic sequencing project, lots of researches based on genomic sequence database have been performed. Analysis of genome is meaningful that genome has kept a record of transition of evolutionary history. Although mutations of genome sequence often occurred in chance, some of them made no significant change in phenotype, which are called synonymous substitutions. Thus such mutations didn't take negative or positive pressure which was reaction to phenotypic changes in natural adaptation. Other mutations leading to changes in proteins, namely non-synonymous substitutions, were negatively critical for survival of the creature in general and such mutations therefore disappeared in population immediately after occurrences. Only a few of the non-synonymous mutations have fixed in genome sequence because the fixations brought functional changes that usually have benefits to individuals or populations.

Various estimators were designed to measure evolutionary alteration of genome sequence, and dN/dS ratio has been used widely as one of them. dN/dS ratio is the ratio of non-synonymous substitution number in non-synonymous sites (dN) to synonymous substitution number in synonymous sites (dS) (Sergey Kryazhimskiy *et al.* (2008)). As described above, synonymous substitution is regarded as neutral mutation and non-

synonymous substitution is presumed experience selection. Therefore, dN/dS ratio represents selective power by comparing natural selection with neutral mutation. If genes have dN/dS ratio values over one, it means that non-synonymous substitution occurred frequently than neutral mutation and natural selection enhanced change of the genes in protein level; while dN/dS ratio values calculated less than one suggests that the genes underwent purifying selection which is suppression power on protein change by natural selection.

Many researches are shown to use dN/dS ratio to estimate gene evolution rate in phylogenetic lineage divergence. For example, in study comparing two parasites which were *Theileria annulata* and *T. parva*, researchers estimated selective pressure on orthologs between the parasites by calculating dN/dS values of genes (Arnab Pain *et al.* (2005)). They matched these dN/dS values with gene expression data and investigated evolutionary selective pressure acting on macroschizont polypeptides positively and on merozoites negatively. Based on the relation between two data set, aspect of protein expression and their corresponding dN/dS values, the researchers presumed that regulatory functions were diversified after speciation of two parasites.

Conservation of gene sequences coding the postsynaptic density from human neocortex (hPSD) was examined between humans, primates and rodents by calculating dN/dS ratio (Àlex Bayés *et al.* (2010)). Comparing human and mouse, the median dN/dS values for hPSD genes were significant to be less than those of whole protein coding genome. Significant conservation in hPSD coding region was also shown in comparing dN/dS ratio values of other pairs, which were human/chimp, human/macaque and mouse/rat. The researchers concluded that lower dN/dS values of the hPSD genes than other genome indicated purifying selection in the hPSD genes not to unique to the human lineage. Likewise, dN/dS ratio estimation has been used as an estimator to presume orthologous gene evolution between related species.

PAML is a package program that performs phylogenetic analysis of DNA and protein sequences using the maximum likelihood method (Ziheng Yang (2007)). This package provides YN00 program for estimation of dN/dS ratio with various methods including NG86, LWL85, LPB, and LWL85m (Masatoshi Nei *et al.* (1986), Wen-Hsiung Li *et al.* (1985), Wen-Hsiung Li (1993), P Pamilo *et al.* (1993), Ziheng Yang (2006)). And CODEML is another program supported by PAML package, which

implements the maximum likelihood estimation for a new distance method defined by Yang (2006) (Ziheng Yang (2006)).

Functional annotation tools are often used to figure out resultant gene lists which are under positive selection power supported by dN/dS values. DAVID bioinformatics resources website offers analytic tools designed to obtain biological meanings from input gene lists (W. Huang da *et al.* (2009)). DAVID functional annotation tool extracts representative features of gene lists and categorized the features such as biological process, cellular components, molecular functions, and KEGG pathways. The results of DAVID analysis are organized by forms of chart or table, and summarized information of the results is also provided.

Chapter 2. Key genes differentially evolved between monotocous species and polytocous species

2.1. ABSTRACT

This study aimed to reveal differentially evolved genes and their distinct functions resulting in the number of offspring at a birth and to find distinct functions with comparative analysis between 3 monotocous mammals and 3 polytocous mammals using a maximum likelihood estimator of dN/dS.

Total significantly accelerated genes and significantly accelerated genes related in reproduction in monotocous species were much than that in polytocous species. Functions of male gamete generation, muscle, neuron and signal transduction are accelerated in monotocous species in contrast to oocyte meiosis and transcription process accelerated in polytocous species. *IL-2, CGA* and *ADAM32* have relations to the reproductive traits with highly accelerated evolution in monotocous species. *CGA* gene could affect multiple ovulations leading to regulation of litter size. The rate of acceleration of highly accelerated genes related to reproductive functions showing that 'monotocous' trait is more likely to be a result of adapted evolution than 'polytocous'.

2.2. INTRODUCTION

Mammals can be divided into two groups of monotocous or polytocous by their progeny number per birth. The mechanism determining this reproductive trait of each species has not been identified clearly. However, significantly involved genes of the mechanism can be suggested based on genome-wide comparison. No study has directly focused on the evolutional genetic differences between monotocous and polytocous species yet, but lots of researchers studied factors affecting litter size in diverse species. For example, in commercial pig breeds which are Chinese Meishan and Large White, it is shown that the estrogen receptor (ER) locus is associated with increased litter size(M Rothschild *et al.* (1996)). The prolactin receptor gene was also identified its association with total number born (TNB) and number born alive (NBA) by least squares method in five PIC lines(C. K. Tuggle A. L. Vincent, Max F. Rothschild, G. Evans, T. H. Short, O. I. Southwood, G. S. Plastow (1997)). There is another report about genes of monotocous species to prove that natural mutations in an ovary-derived factor, such as $FecX^I$ gene on Inverdale sheep, can lead to increase of ovulation rate and infertility phenotypes in a dosage-sensitive manner(S M Galloway *et al.* (2000)). Retinol-binding protein 4 (*RBP4*), estrogen

receptor, and prolactin receptor genes were demonstrated their connection with litter size and the number of piglets born alive in German pig lines as well(C Drogemuller *et al.* (2001)).

In this study, I identify significantly evolved genes causing reproductive difference between monotocous and polytocous species by comparing rate of accelerated evolution between one to one orthologs among 3 monotocous and 3 polytocous species, which are human, horse and cow as monotocous species and mouse, dog and pig as polytocous species using a maximum likelihood estimator. And, I classify those genes into functional classes focusing on the reproductive functions. Also I show difference in rate of acceleration between the polytocous specific evolved genes and the monotocous specific evolved genes.
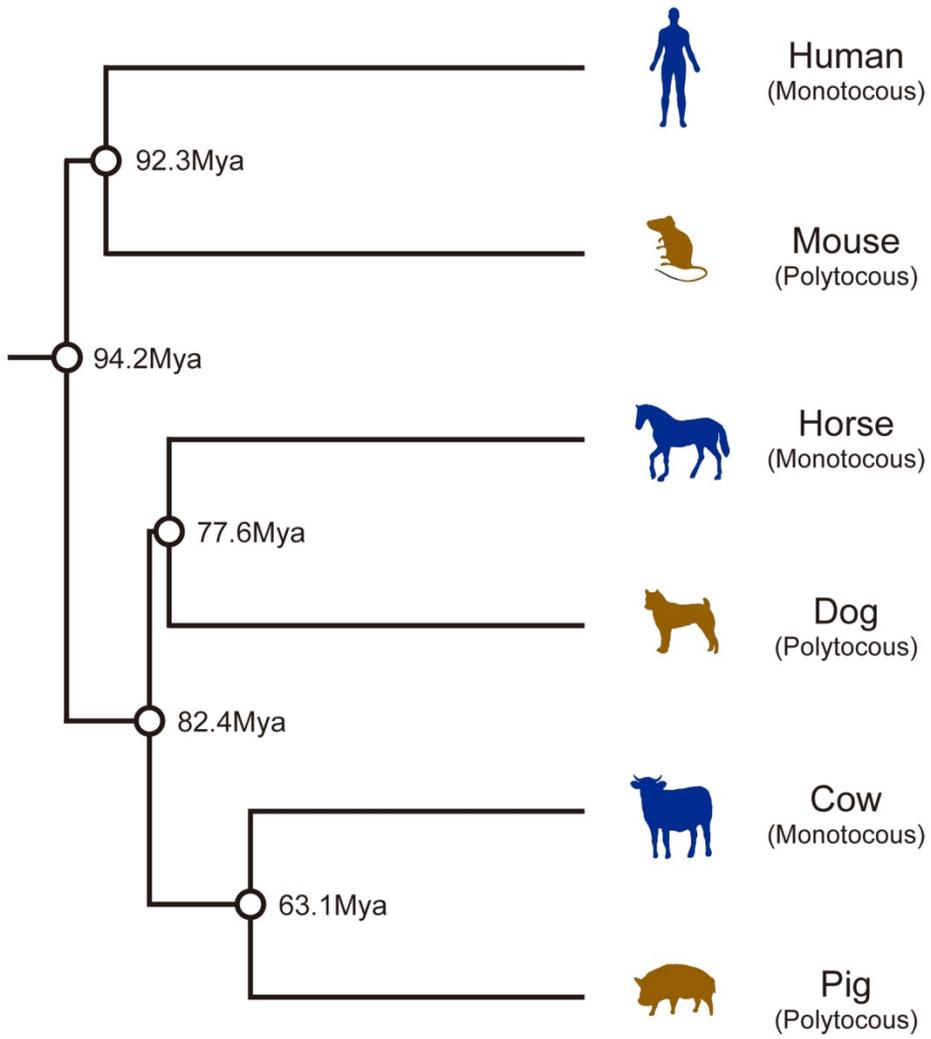
Figure 2.1. Phylogenetic tree of 6 mammal species.

2.3. MATERIALS AND METHODS

mRNA reference and protein sequence data of 6 species were downloaded from ENSEMBL(T. Hubbard (2002)). One to one orthologs of 6 species were defined by Mestortho(Kyung Mo Kim *et al.* (2008)). As a result, 9000 1:1 orthologs for the 6 species were collected. Phylogenetic tree of 6 species was supported by Timetree(S Blair Hedges *et al.* (2006)). Prank aligned the orthologous gene sets of 6 species and its option was default(Ari Löytynoja *et al.* (2005)). Poorly aligned sites of orthologs were filtered by Gblocks(J. Castresana (2000)). Using codeml of PAML 4(Ziheng Yang (2007)) (F3X4 codon frequency under branch model), dN/dS values (ω) of each orthologous genes were calculated twice. First calculation was performed with H0 (null hypothesis) that genes have evolved with the same rate across 6 species (model=0, NSsites=0) and H1 (alternative hypothesis) that genes have got specific positive selection pressure in monotocous species (model=2, NSsites=0). On codeml of H1, human, horse and cow, monotocous species, were set by foreground group and others were background group. The second was done with same H0 but different H1 that genes have been selected positively in polytocous species. Orthologs were filtered out with its dS>3 or its ω>5(Cristian I Castillo-Davis *et al.* (2004),

Christopher S Peacock *et al.* (2007)) . So far, a set of 8419 orthologs remained in monotocous group. Polytocous group had a gene set of 8417 orthologs. Significant genes were selected by FDR<0.05(Derek Y Chiang *et al.* (2003)). Also, genes having foreground ω values bigger than background ω values were picked. After all, 735 orthologs remained in monotocous species and 203 orthologs in polytocous species.

2.4. RESULTS AND DISCUSSION

I analyzed 9000 orthologs of 6 mammal species to detect genes differently evolved between monotocous and polytocous traits. As the results, 735 genes showed evidence of monotocous specific accelerated evolution and 203 genes showed evidence of polytocous specific accelerated evolution with FDR <0.05 (Table 2.2, 2.3).

Functions related in reproduction among 20 highly accelerated genes

I took each top 20 accelerated genes of polytocous and monotocous species in order of great differences between foreground ω and background ω as highly accelerated genes and focused on them(Table 2.2, 2.3). Three out of the 20 genes in polytocous species are related to meiosis in biological process (*CDK8*, *SET* and *TERF1*) by COREMINE analysis (Figure 2.4). Figure 2.2 shows eight genes of the 20 genes in monotocous species were clustered with ovulation term as biological process and two ovulation related hormones. T-cell activation and induction of interleukin-2 (IL-2) directly stimulate the release of luteinizing hormone (LH) and follicle-stimulating hormone (FSH) in a specific dose range(M Umeuchi *et al.*

(1994)) which hormones take part in regulating ovulation(A Christensen *et al.* (2012)). *CGA*, chorionic gonadotropin alpha, codes the alpha subunit of glycoprotein hormones that contain FSH, LH, TSH, and HCG(J G Pierce *et al.* (1976)). *ADAM32* has a potential role in sperm development or fertilization(Inchul Choi *et al.* (2003)) (Figure 2.5).
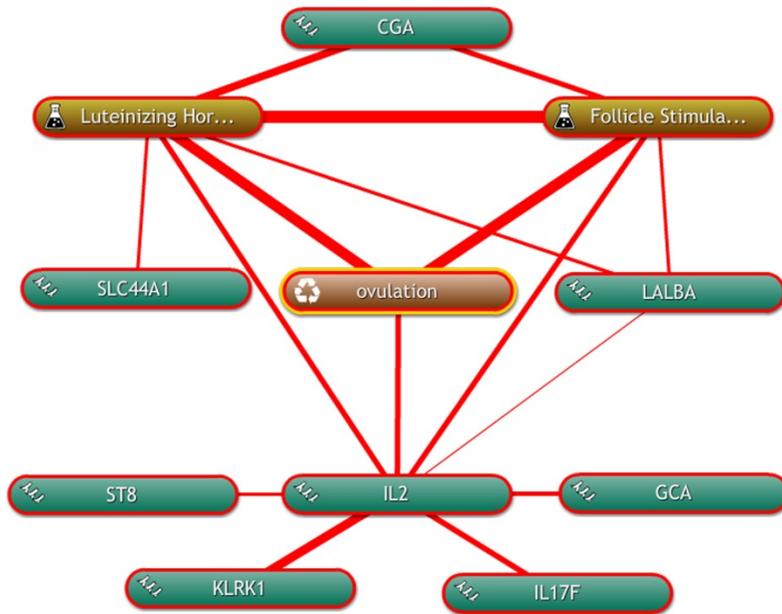
Figure 2.2. Connections among highly accelerated genes in monotocous species, ovulation term as biological process and two ovulation related hormones of luteinizing hormone(LH) and follicle stimulating hormone(FSH) using COREMINE. The thickness of the lines between two terms displays the number of articles having the both terms together. The length of the lines doesn't have meaning of distance.
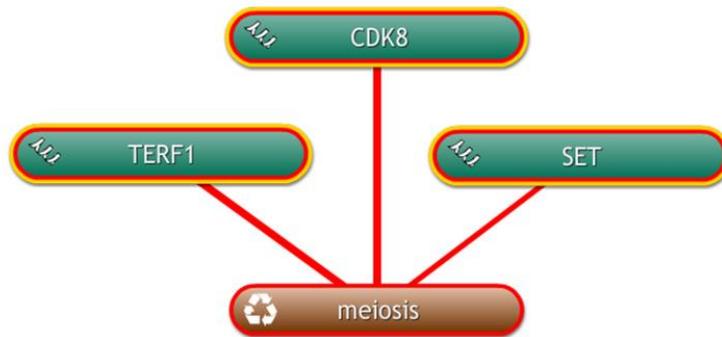
Figure 2.4. Connections among highly accelerated genes in polytocous species, meiosis as a biological process and three related genes using COREMINE. The thickness of the lines between two terms indicates the number of articles with both terms. Line length does not represent distance.
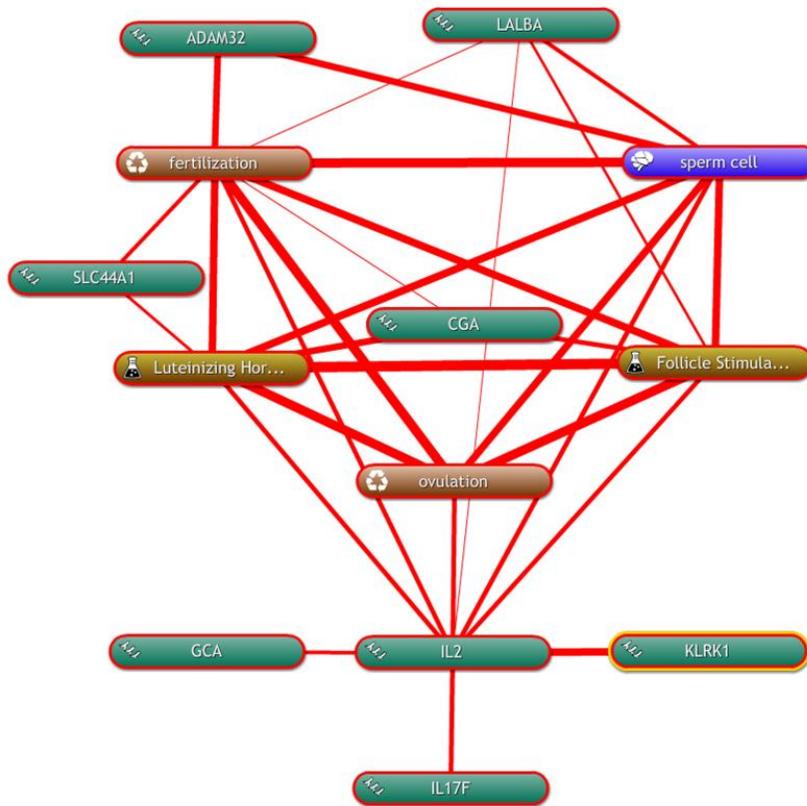
Figure 2.5. Connections among highly accelerated genes in monotocous species, ovulation and fertilization as biological processes, sperm cell as a feature of anatomy, and two hormones—luteinizing hormone (LH) and follicle-stimulating hormone (FSH)—using COREMINE. The thickness of the lines between two terms indicates the number of articles with both terms. Line length does not represent distance.

Table 2.2. List of top 20 genes differentially accelerated in monotocous species

Gene	ΔlnL	Foreground dN/dS	Background dN/dS	FDR
DEFB123	12.53324	1.9523	0.2718	0.006921
ZNF75D	15.96063	1.61039	0.53613	0.001717
C1orf54	8.441912	1.50078	0.43017	0.036163
IL2	10.31047	1.47713	0.49661	0.016811
SCIMP	10.48672	1.21329	0.35809	0.015771
LALBA	9.618706	0.93535	0.33244	0.022301
GCA	30.24602	0.70658	0.10568	2.81E-06
IL17F	33.16577	0.61593	0.10733	7.04E-07
FABP2	17.58106	0.65233	0.14851	0.000855
SLC44A1	112.815	0.52451	0.06547	3.04E-23
CGA	28.81538	0.49265	0.05822	5.26E-06
KLRK1	8.875466	0.75588	0.32371	0.030733
ADAM32	25.11975	0.81571	0.38603	2.85E-05
C6orf163	27.98434	0.60259	0.17487	7.56E-06
NXT2	10.07988	0.48901	0.07362	0.018475
CPA4	27.33101	0.63241	0.22308	1.04E-05
CERS3	18.49539	0.64199	0.23441	0.000576
COMMD8	15.12557	0.52197	0.11711	0.002404
ZBP1	11.35457	0.7843	0.38035	0.010942
TMPRSS11D	14.03167	0.78968	0.38886	0.003688

Twenty genes were sorted in descending order according to differences between foreground dN/dS and background dN/dS. Genes were filtered according to their acceleration in monotocous species by FDR < 0.05.

18

Table 2.3. List of top 20 genes differentially accelerated in polytocous species

Gene	ΔlnL	Foreground dN/dS	Background dN/dS	FDR
SAT1	18.12234	0.29578	0.04037	0.000571
SSB	53.79014	0.29519	0.04338	5.57E-11
DZIP1L	16.4915	0.55396	0.30466	0.001116
C20orf72	12.42532	0.51103	0.26529	0.006387
CCDC58	7.869252	0.30042	0.06234	0.03963
KIAA1958	68.37548	0.29801	0.06022	6.74E-14
TCTEX1D2	8.451344	0.29317	0.06054	0.031464
ACCN2	93.07348	0.22427	0.00654	9.06E-19
SET	26.76556	0.22615	0.00955	1.32E-05
CDK8	41.2362	0.22404	0.01	1.97E-08
FBXO15	9.981984	0.4604	0.24906	0.017077
TERF1	9.60053	0.4599	0.25416	0.019622
NRSN2	11.23157	0.32469	0.12124	0.010409
GPBP1L1	20.33151	0.29432	0.10077	0.000222
HIPK1	82.62362	0.21309	0.03467	1.12E-16
CTDSPL2	39.01921	0.18591	0.00856	5.72E-08
BTF3	18.0742	0.19109	0.01491	0.000583
USP9X	184.8863	0.19466	0.02345	3.74E-38
AC004381.6	7.68336	0.51151	0.34211	0.042644
WIPF3	8.95985	0.35286	0.18621	0.025375

Twenty genes are sorted in descending order according to differences between foreground dN/dS and background dN/dS. Genes were filtered according to their acceleration in polytocous species by FDR < 0.05.

A role of *CGA* gene on multiple ovulations

I could find a scenario for multiple ovulations based on the gonadotropins focusing on the putative key gene, *CGA*. Two gonadotropins of FSH and LH secreted by pituitary gland are regarded as regulators of ovarian function(Aurélie Vinet *et al.* (2012)). In estrus cycle, the number of follicles that will be ovulated is affected by a threshold of FSH level allowing mature of follicles and the size of pool of gonadotropin-dependent follicles. The duration for plasma FSH level to stay above the threshold determines the number of mature gonadotropin-dependent follicles. And basically, large increase in the amount of gonadotropin-dependent follicles on normal FSH level can induce multiple ovulations among lots of mature follicles. These two mechanisms are not mutually exclusive and capable of occurring simultaneously. Our finding that *CGA* gene has accelerated evolved more in monotocous species can be an evidence for the scenario. Focusing on the role of plasma FSH level in the hypothesis, although *CGA* is a common constituent not only on FSH but among all glycoproteins, it is assumable that this gene takes a crucial role deciding the single or multiple ovulations.

Difference of acceleration rate between polytocous and monotocous species

Figure 2.3 shows that histograms of difference in ω between foreground and background species of significantly accelerated genes of polytocous and monotocous. It also shows that the number of significantly accelerated genes in monotocous species is much than that of polytocous species. Overall rate of acceleration in monotocous species is higher than that in polytocous species. Also, the positions of candidate genes to be related in determining number of offspring in a single birth with highly accelerated evolution are different between monotocous and polytocous species. Rate of acceleration of the candidate genes of monotocous species are higher than that of polytocous species. This may show that 'monotocous' as trait is more likely to be acquired trait than 'polytocous' with fitness in their environment.
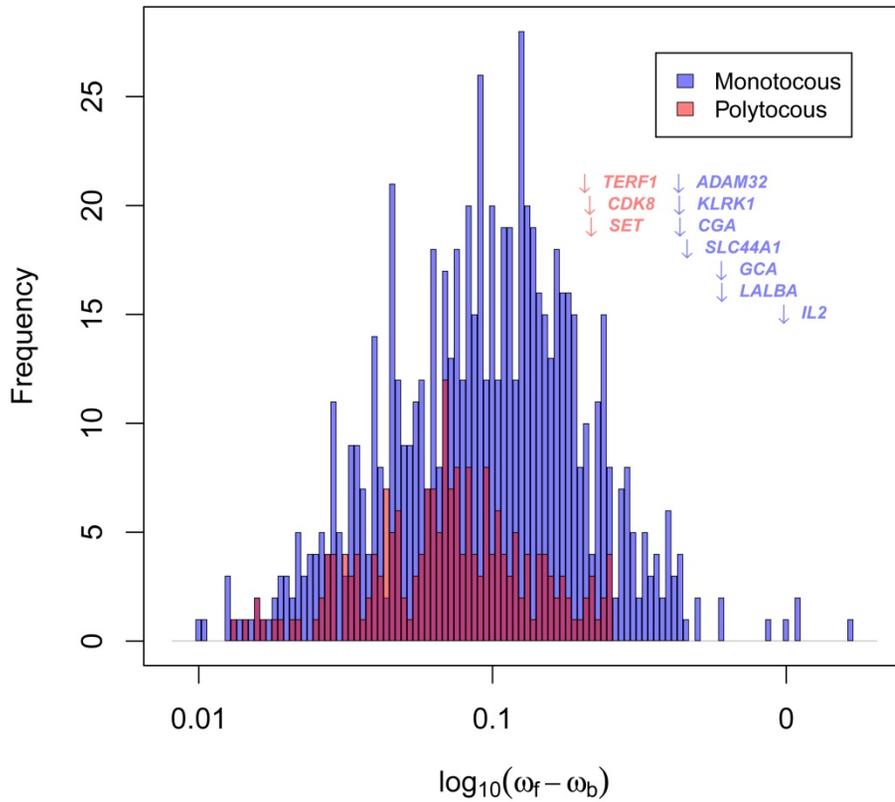
Figure 2.3. Histogram of difference in ω between foreground and background species using significantly accelerated genes of polytocous and monotocous. Red colored genes are highly significantly accelerated in polytocous species and blue colored genes are highly significantly accelerated in monotocous species. Horizontal positions of the arrows are

22

position of the class mark of their genes. There is no point in vertical positions of the arrows. ωf: ω ratio in foreground branches, ωb: ω ratio in background branches.

Functional enrichment of differentially accelerated genes between monotocous and polytocous species

I performed gene enrichment analysis among differentially accelerated genes between monotocous and polytocous species by their functions using DAVID(W. Huang da *et al.* (2009)), and arranged GO terms by biological processes(BP), cellular components(CC), molecular functions(MF), and KEGG pathways in Table 2.1. The table showed some distinctive features of each group in a little different aspect. A few of terms related to muscle scored significant p-values in monotocous species: muscle organ development, muscle system process, cell motion, cytoskeleton, and arrhythmogenic right ventricular cardiomyopathy (ARVC). Two terms for immune system, leukocyte activation and Fc gamma R-mediated phagocytosis, were shown to be enriched in monotocous species as well. It is possible to assume relation between terms to signal transduction system in monotocous species. A term of cell surface receptor linked signal transduction had the most enriched genes among BP terms and genes for ion transport can be assumed to have a role in signal transduction(H M Lander (1997)). Some CC terms, cell surface, integral to membrane, and plasma membrane, were also connected to signal transduction despite their covering

broad sense(Kai Simons *et al.* (2000)). In addition, nervous system seems to be accelerated in monotocous species: axon guidance and neuron differentiation. Genes for cytoskeleton is supposed to attribute nervous system since it is involved in axonal pathfinding(P. C. Letourneau (1996)). Likewise, axon guidance receptors exhibit isoform-specific homophilic binding trait (or identical protein binding) by alternative splicing of drosophila immunoglobulin(Woj M. Wojtowicz *et al.* (2004)). Accelerated evolution of male gamete generation in monotocous species is meaningful in contrast to oocyte meiosis in polytocous species. Transcription process is also a notable feature of GO terms in polytocous species.

Table 2.1. Analyzing enrichment of genes differentially accelerated in monotocous and polytocous species using DAVID. GO terms were arranged by biological processes, cellular components, molecular functions and KEGG pathways.

	Biological processes	Cellular components	Molecular functions	KEGG pathways
Monotocous species	Leukocyte activation (17) Proteolysis (55) Regulation of growth (21) Muscle organ development (15) Neuron differentiation (28) Cell adhesion (43) Cell surface receptor linked signal transduction (90) Cell motion (29) Male gamete generation (24) Lipid biosynthetic process (24) Muscle system process (13) Nitrogen compound biosynthetic process (25) Ion transport (64)	Extracellular region (98) Cell projection (41) Cytoskeleton (76) Cell surface (26) Integral to membrane (246) Plasma membrane (199) Cell fraction (75)	Peptidase activity (40) Passive transmembrane transporter activity (35) Alkali metal ion binding (28) Identical protein binding (40) Cytoskeletal protein binding (33) Nucleotide binding (121)	Axon guidance (11) Arrhythmogenic right ventricular cardiomyopathy (ARVC) (8) ABC transporters (7) Fc gamma R−mediated phagocytosis (10)
Polytocous species	Protein catabolic process (14) Positive regulation of macromolecule metabolic process (23) Regulation of transcription (47)	Membrane-enclosed lumen (38) Intracellular non-membrane-bounded organelle (38)	Passive transmembrane transporter activity (10) Protein kinase activity (14) Transcription regulator activity (28)	MAPK signaling pathway (9) Oocyte meiosis (6) Cell cycle (5)

Numbers in the brackets represents the numbers of genes related to the terms.

Functional differences of differentially accelerated genes between monotocous and polytocous species

Also, I searched the functions of genes differentially accelerated in each group of species from Entrez Gene database at NCBI(Donna Maglott *et al.* (2005)). Genes were categorized by their functions with representative terms. Genes having accelerated evolution in monotocous species were sorted into divergent functions: estrogen, steroid, sperm, immune, transcription, signal transduction, solute carrier, ATP·AMP·GTP, myosin, eye, olfactory, ear, nervous system, oncogene and cell adhesion. Polytocous species had acceleration in evolution of functions which are ubiquitin ligase, translation, transcription, cell cycle regulation, cell growth, mitochondrial, actin, autism and Alzheimer's disease.

The list shows some differences through two groups clearly. First of all, accelerated evolution of sensory organ including nervous system in monotocous species is noticeable. This is meaningful to assume evolutionary history between monotocous and polytocous species. Evolutionary trend of nervous system and brain, for instance, was lightened by contrasting primates with rodents(Steve Dorus *et al.* (2004)). In the study, protein evolution rates are shown to be higher in primates than rodents, and

especially in human out of primates, protein evolution is obviously accelerated depending on the lineage from ancestral primates to human. In addition, lots of candidate genes that are assumed to have important roles in the evolution of human brain were also identified, and two of the candidates, *SHH* and *GRIK4*, were listed in our accelerated genes in monotocous species as well. I observed another evident point that two muscular fibers (actin and myosin) are separately accelerated within two groups. Genes for myosin (*MYH1, MYO1E, MYO1F,* and *MYO5C*) were shown to have higher ω value in monotocous species than polytocous species, but genes for actin (*ABLIM3, CAPZB, OXSR1, SMARCC1,* and *SMARCD1*) were shown to be accelerated in polytocous species. I could find a clue to explain higher ω values of myosin genes which might induce morphological changes or functional alteration of myosin related organs by mutation to its inactivation. *MYH16* gene coding myosin heavy chain 16 is able to be an example to show the correlation of accelerated evolution of myosin genes with changes in morphological features of species(Hansell H Stedman *et al.* (2004)). In primates, a protein coded by *MYH16* is only found in muscles on the jaw. This gene has gone through a frameshift since the humans and chimpanzees diverged into each lineage, and *MYH16* gene lost its function by the

mutation. Furthermore, higher value of ω of *MYH16* shown especially in human lineage means that the gene has undergone purifying selection in all ancestral primate lineages except for human. These evolution processes also led to changes of physical appearance, which are expressed as size reduction of masticatory muscle fibers in human. The case of *MYH16* can indicate that accelerated evolution of four myosin genes in monotocous species may represent a possibility to induce certain changes in terms of their functions or physical structure where they are mainly expressed. However, in spite of an example of *MYH16*, it is difficult to confirm completely the consequence of accelerated evolution of four myosin genes due to their roles on varied functions. Genes coding proteins for actin filaments in polytocous species are also hard to see the effect of its accelerated evolution as there is no research studying actin evolution in mammal. I can only deduce the consequence based on the case of higher value of ω ratio on *MYH16* gene, described above.

Table 2.4. Classification of genes differentially accelerated in monotocous species

Function	Genes
Estrogen	CGA, ESRRG
Steroid	CYP46A1, HSDL1, NR4A3, SRRG
Sperm	ADAM32, ADCY10, CCDC39, CCIN, DNAH8, MNS1, SEPT14
Immune	GCA, IGSF3, IL17F, MASP1, MLLT10
Transcription	BHLHB9, GFI1, GTF2E1, GTF3C1, HMX2, LMX1B, MLLT10, NR4A3, OLIG3, OTP, PAX9, RFX1, RUNX1, SIM2, SMARCB1, TBX2, THOC3
Signal transduction	DCAF8L1, DLG1, GNB5, LPHN1, MPP2, PLCB4, PLXNA2, SCN9A, SHH, SULF1, WNT2
Solute carrier	SLC1A1, SLC1A6, SLC22A12, SLC35C2, SLC38A4, SLC44A1, SLC5A1, SLC6A19, SLC7A2
ATP, AMP, GTP	ABCG4, ADCY10, AFG3L2, AMPD3, ATP10A, ATP8B1, DNM1, RASGRP2, XRCC5
Myosin	MYH1, MYO1E, MYO1F, MYO5C
Eye	EML1, MIP, PLCB4, RPE65
Olfactory	OR5M11, OR8U1
Ear	HMX2, TECTA, TECTB
Nervous system	ADAM22, ADAM9, AFG3L2, B3GAT2, CDH22, DLG1, GABRE, KCNA7, NEURL, SIM2, SLC1A1, SPM6B
Oncogene	FYN, GFI1, SRC, WNT2, WNT5A
Cell adhesion	B3GAT2, CDH23, DSG4, LPHN1

Genes differentially accelerated in monotocous species are arranged according to their functions with reference to information in the NCBI database.

Table 2.5. Classification of genes differentially accelerated in polytocous species

Function	Genes
Ubiquitin ligase	NEDD4L, TRIM2
Translation	CPEB1, DDX23, DHX36, HNRNPA2B1, RPL37A, SSB
Transcription	ETV5, PAX3, RTF1, SMARCC1, SMARCD1
Cell cycle regulation	CDK8, ZYG11B
Cell growth	IGFBP5, LTBP1, PPP2R2B, TGFB2
Mitochondrial	NNT, SDHA
Actin	ABLIM3, CAPZB, OXSR1, SMARCC1, SMARCD1
Autism	MARK1, SEMA5A
Alzheimer's disease	APBA1

Genes differentially accelerated in polytocous species are arranged according to their functions with reference to information in the NCBI database .

2.5. CONCLUSIONS

In this research, not only total significantly accelerated genes but also the significantly accelerated genes related in reproduction in monotocous species were much than that in polytocous species. I suggested 3 of significantly highly accelerated evolved genes (*IL-2*, *CGA* and *ADAM32*) in monotocous species, which have a role in reproductive function. Especially, *CGA* is the gene differentiating litter size of species by controlling the number of ovulated follicles. By reproduction related terms in all significantly accelerated genes, genes related to male gamete generation were accelerated in monotocous species in contrast that genes related to oocyte meiosis were accelerated in polytocous species. With rate of acceleration of reproduction related genes, 'monotocous' trait is more likely to be a result of adapted evolution than 'polytocous'.

Besides, terms about muscle, neuron and signal transduction were enriched among the accelerated genes in monotocous species whereas transcription process was enriched in polytocous species with enrichment analysis of those significantly accelerated genes using DAVID(W. Huang da *et al.* (2009)). Also monotocous species had lots of accelerated genes on

terms of estrogen, steroid, sperm, immune response, myosin, solute carrier, other sensory organs, and so on. In contrast, genes of polytocous species showed rapid evolution in functions such as translation, cell growth, actin, and psychopathy symptoms. This study has a limitation that it is rare to find preceding studies supporting our results; however, it is meaningful that this study revealed definite genomic differences between monotocous species and polytocous species and could suggest the evolutionary flow.

Chapter 3. Evolutionary advancement of memory and cognitive ability in human in contrast to non-human primates using dN/dS analysis

3.1. ABSTRACT

It is well known that some of human brain functions have been more developed than other non-human primates' brain functions. Long memory capacity is shown to be developed in human specifically in contrast to memory capability of other primate species. Language faculty, high cognitive ability and learning are the other traits shown to be developed in human lineage. Because primates have been diverged from same lineage, it is meaningful to analyze accelerated evolution of genes of primates including human and to compare the results between the species.

I performed a dN/dS ratio analysis on orthologous genome sequences of 10 primate species which were human, bushbaby, chimpanzee, gorilla, gibbon, macaque, marmoset, mouse lemur, orangutan and tarsier. Genes were picked among the orthologous gene lists based on differences of dN/dS values between human and non-human primates which represented accelerated evolution of the genes in human. In the selection, 41 genes remained. From DAVID analysis, functions of the genes were mainly related to actin filament and MAPK signaling pathway, which were associated with neuronal mechanisms. Also, BCHE and RIT2 were noticeable among ten

genes which had high differences of dN/dS values because the genes related to neuronal disorder diseases like Alzheimer's disease and Parkinson's disease.

3.2. INTRODUCTION

There have been a lot of researches that studied distinctive features of human brain function, which have been expressed as long memory, language faculty, high cognitive ability, learning, and much more. In an evolutionary point of view, many researchers have performed their works on identifying factors which have affected human-specific brain faculties by studying human and non-human primates. For example, metabolic modifications of human brain were suggested as a factor evolutionally inducing rapid expansion of brain size and extreme increase in cognitive capabilities, by investigating human, chimpanzee, and rhesus macaque (Xing Fu *et al.* (2011)). In comparison between human, chimpanzee, rhesus macaque, and pigtail macaque, gene expression analysis of two thrombospondins (THBS4 and THBS2) showed that expression of two genes had been increased evolutionally, which had resulted in changes of synaptic organization and plasticity, distinctive cognitive abilities, and the vulnerability to neurodegenerative disease in human (Mario Cáceres *et al.* (2007)). In addition, a research which studied molecular evolution of microcephalin gene in human and non-human primates indicated that the

gene had gone through positive selection and brain of human and primates was enlarged concurrently in evolution history (Yin-qiu Wang *et al.* (2004)).

I also performed comparative analysis of orthologous genes of human and non-human primates using a maximum likelihood method. I used 10 primates including human for this study, and analyzed orthologous genes which were more evolutionally accelerated in human than in non-human primates around brain function.

3.3. MATERIALS AND METHODS

I downloaded sequences of nucleotide and amino acid of 10 primates, which were human, bushbaby, chimpanzee, gorilla, gibbon, macaque, marmoset, mouse lemur, orangutan and tarsier, from ENSEMBL (T. Hubbard (2002)). Orthologous gene pairs of human and one of non-human primates were defined by ENSEMBL (T. Hubbard (2002)), and I picked up a set of 8694 orthologs linked across all of 10 species based on one to one correspondence. I referred Timetree (S Blair Hedges *et al.* (2006)) for phylogenetic tree of 10 primates. Orthologous genes of 10 primates were aligned by Prank on default options (Ari Löytynoja *et al.* (2005)). Poorly aligned positions of orthologous gene sequences were eliminated by Gblocks (J. Castresana (2000)), and 8313 orthologs remained. I used codeml of PAML 4 (Ziheng Yang (2007)) with branch model to calculate dN/dS values of each orthologous gene. I set null hypothesis (H0) on same rate evolution of genes throughout 10 primates (model=0, NSsites=0) and alternative hypothesis (H1) on positive selection pressure to human genes rather than other nine primates (model=2, NSsites=0). In H1, human was therefore set as foreground species while other nine primates were set as

41

background species. I filtered orthologous genes by their values resulting from codeml with dS>3 and dN/dS>5 (Cristian I Castillo-Davis *et al.* (2004), C. S. Peacock *et al.* (2007)). After filtering, 6505 orthologs remained.

3.4. RESULTS AND DISCUSSION

2910 orthologs accelerated more in human were identified by selecting orthologs that dN/dS values of foreground (human) were greater than those of background (other primates). Analyzing their functions focusing on brain-related points, I used 41 orthologs that remained from more accelerated orthologs filtered by FDR<0.2 (B. Angulo *et al.* (2008)). (Table 3.2)

Functional annotation of genes highly accelerated in human than other primates

I performed a functional annotation analysis using DAVID (W. Huang da *et al.* (2009)), and the result is shown in Table 3.1. It is noticeable that terms related to cytoskeleton or actin filament were enriched: actin filament capping, cytoskeleton organization, cytoskeleton, and actin binding. Some previous researches showed that cytoskeleton or actin filament play important roles in neuronal mechanisms. Actin, brain tropomyosin, and myosin is shown to be major constituents of rat synaptosomes and to be associated with neurotransmitter release (AL Blitz *et al.* (1974)).

Cytoskeletal machinery that is responsible for growth cone migration induces axonal pathfinding (P. C. Letourneau (1996)). Dendritic spines contain high concentrations of actin, which support a possibility that the postsynaptic element is primarily associated with neuronal plasticity (Maria Fischer *et al.* (1998)). As neuronal (synaptic) plasticity has been regarded as a strong candidate that plays an essential role in learning and memory (S. F. Cooke *et al.* (2006)), and based on other actin functions related to brain, I can assume that accelerated evolution of actin genes more in human than in other primates supports improvement of cognitive ability of human brain in contrast to non-human primates.

Terms related to MAPK signaling pathway and Ras protein is also notable in terms of brain functions. MAPK signaling pathway is connected to signals from cell surface receptors to regulatory targets in cells and controls cell survival (Lufen Chang *et al.* (2001)). During acquisition of fear memory, MAPK signaling pathway was activated (Kiyofumi Yamada *et al.* (2003)). As well as actin filament is supposed to be involved in memory and learning by its participation in neuronal plasticity, MAPK signaling pathway is also seemed to have a role in passive memory process in neuronal mechanisms. Furthermore, activation of MAPK signaling pathway

coordinated with Ras-phosphatidylinositol 3-kinase (PI3K)-Akt-mammalian target of rapamycin (mTOR) signaling pathway induced dendritic complexity (V. Kumar *et al.* (2005)).

Table 3.1. Functional annotation of 41 genes more evolutionally accelerated

in human than in other primates

Biological process	Cellular component	Molecular function	KEGG pathway
Negative regulation of macromolecule metabolic process (5) Cell migration (4) Actin filament capping (2) Regulation of Ras protein signal transduction (3) Positive regulation of molecular function (5) Cytoskeleton organization (4)	Membrane fraction (5) Cytoskeleton (7)	Sequence-specific DNA binding (5) Zinc ion binding (10) Actin binding (4) Transcription regulator activity (8) Beta-amyloid binding (2)	MAPK signaling pathway (3)

Top 10 genes among 41 genes highly accelerated in human than other primates

I selected 10 genes from 41 highly accelerated genes in descending order of dN/dS differences between foreground group and background group, which represent human and other nine primates, for looking into pivotal genetic evolutional flow along with human lineage: BCHE, PSMA7, RIT2, RNF170, PPIC, OR4K1, NR2F2, GIT2, ADAM12, and RCL1 (Table 3.2). All dN/dS differences of 10 genes were over 1.

It was found that mutations of BCHE gene affected prolonged apnea during anesthesia using neuromuscular blocking agents and BCHE was related to Alzheimer's disease and Parkinson's disease through interacting with DCP1 gene and apolipoprotein E ε4 allele (Csaba Barta *et al.* (2001), Kari M Mattila *et al.* (2000)). PSMA7, proteasome subunit alpha type 7, had a role in regulation of HIF-1α and HCV IRES activity; HIF-1α is an important transcription factor for cellular responses and HCV IRES has a critical role in HCV replication (Sayeon Cho *et al.* (2001), Martin Krüger *et al.* (2001)). RIT2 was shown to associate with Parkinson's disease as well as BCHE gene (Nathan Pankratz *et al.* (2012)). RNF170 is related to IP3 receptor mechanism and autosomal dominant sensory ataxia (Justine P Lu *et*

al. (2011), Paul N Valdmanis *et al.* (2011)). PPIC encodes peptidyl-prolyl isomerase C that its family catalyzes proline and accelerate the folding of protein (Donna Maglott *et al.* (2005)), but its own functions have not been identified in detail so far. Encoding olfactory receptor protein, OR4K1 is seemed to associate with neuronal responses related to olfactory sense (Donna Maglott *et al.* (2005)). NR2F2 was responsible for a regulatory circuitry which is critical for regulation of pluripotency and differentiation in human embryonic stem cells and had crucial role in the activation of neural genes during early differentiation in particular (Alessandro Rosa *et al.* (2010)). It was demonstrated that GIT2 had many distinct forms by alternative splicing and its short isoform was associated with localization of paxillin and actin cytoskeletal organization leading to intracellular signaling (Richard T Premont *et al.* (2000), Yuichi Mazaki *et al.* (2001)). GIT2 is also related to both induction and regulation of cell motility (Scott R Frank *et al.* (2006)). ADAM12 was related to tumor aggressiveness and progression in liver cancers and induced shedding of heparin-binding epidermal growth factor which led to cardiac hypertrophy (Hélène Le Pabic *et al.* (2003), Masanori Asakura *et al.* (2002)). In addition, ADAM12 showed up-regulation directly after myoblasts differentiated into myotubes and the gene

48

was expressed in brain ubiquitously (Marie-Florence Galliano *et al.* (2000), Ulrike Novak (2004)). Finally, RCL1 was shown to cleave pre-rRNA into the small and large subunit as a nuclease (Darryl M Horn *et al.* (2011)).

The functions of 10 genes demonstrate the notable evolutionary point in human. BCHE gene which has the highest difference of dN/dS between human and other primates is related to Alzheimer's disease and Parkinson's disease. RIT2 also has an association with Parkinson's disease. Alzheimer's disease is demonstrated as progressive impairment of memory and other cognitive functions (Guy McKhann *et al.* (1984)). Patents of Parkinson's disease mostly had Lewy bodies which were also found in the cerebral cortex and other patent subjects showed some symptoms which were progressive supranuclear palsy, multiple system atrophy, Alzheimer's disease, Alzheimer-type pathology, and basal ganglia vascular disease (Andrew J Hughes *et al.* (1992)). Thus, I think that accelerated evolution of BCHE and RIT2 more in human than in other primates can be critical evidence supporting advancement of memory and cognitive ability of human in evolution.

Table 3.2. List of 41 orthologous genes showing accelerated evolution in human sorted in descending order of difference of dN/dS values between human as foreground and nine primates as background

Gene	$\varDelta$ lnL	Foreground dN/dS	Background dN/dS	FDR
BCHE	10.79141	2.64688	0.1869	0.184272
PSMA7	16.34442	2.34684	0.01837	0.022136
RIT2	10.34844	2.01571	0.10032	0.194664
RNF170	17.22142	1.84355	0.02627	0.016992
PPIC	11.31662	1.68767	0.05757	0.154886
OR4K1	12.43369	1.66977	0.15198	0.107692
NR2F2	23.52511	1.27531	0.02496	0.001181
GIT2	16.93649	1.23297	0.04498	0.018509
ADAM12	10.19711	1.30684	0.20176	0.199399
RCL1	10.31276	1.00963	0.11277	0.194664
HPR	15.7973	1.03654	0.17585	0.026671
IKBKB	22.3497	0.84893	0.01026	0.001935
MEF2C	29.52484	0.83538	0.07608	0.000125
MAX	10.16497	0.78426	0.03435	0.199399
SPTA1	23.96572	0.98931	0.29721	0.001073
COL25A1	44.81715	0.7843	0.10744	1.66E-07
FGD4	32.65615	0.73008	0.0608	4.21E-05
LHX6	10.71817	0.69201	0.02609	0.184272
CAP1	10.64062	0.70698	0.10551	0.184272
FAM126B	16.63022	0.73245	0.13385	0.020472
PLCE1	10.97118	0.72059	0.1261	0.17726
IFT20	15.38743	0.74579	0.17111	0.029172
CCDC90A	10.67711	0.69146	0.11835	0.184272
LRRFIP1	15.49628	0.93565	0.40017	0.028791
DHX35	14.29158	0.53568	0.05982	0.049983

SMG5	21.72787	0.47276	0.04404	0.002189
FYCO1	10.83898	0.63509	0.21073	0.184272
MLTK	11.64394	0.50182	0.09035	0.144608
LRRC8C	10.22593	0.38884	1.00E-04	0.199399
NRSN1	21.86161	0.45653	0.07584	0.002189
SH2D4B	11.65586	0.48963	0.12782	0.144608
SETD3	11.54466	0.36955	0.04643	0.144608
CA6	13.62816	0.38141	0.09841	0.068289
PPP2R2D	15.72888	0.25734	0.02433	0.026671
PHF1	20.91239	0.31023	0.07777	0.002908
AIMP2	16.27118	0.38826	0.1559	0.022136
RC3H2	20.86295	0.21577	0.00886	0.002908
PAX3	24.18252	0.20795	0.00545	0.001073
CCDC132	11.0139	0.19873	0.0039	0.17726
SPTBN1	20.5581	0.19568	0.01572	0.003166
ERC2	28.83517	0.08512	0.01826	0.000125

3.5. CONCLUSIONS

In this study, dN/dS comparative analysis resulted in 41 genes which have gone through accelerated evolution more in human than other nine primates. These genes were mainly associated with cytoskeleton or actin filament, and MAPK signaling pathway. Many studies suggested that actin or cytoskeletal machinery in brain correlated with neurotransmitter release, axonal pathfinding and neuronal plasticity. MAPK signaling pathway was connected to passive memory mechanism and dendritic complexity. These associations indicated that human genome had evolved in the direction of improvement in memory and cognitive ability. In addition, it is noticeable that two of ten genes with highly accelerated evolution in human, BCHE and RIT2, had a relation to neuronal disorder diseases which were Alzheimer's disease and Parkinson's disease. It is meaningful that this study suggests evidences of accelerated evolution in human genome for specific brain functions.

References

1 . Sergey Kryazhimskiy, and Joshua B Plotkin (2008). *The population genetics of dN/dS*. PLoS genetics **4**(12): e1000304

2 . Arnab Pain, Hubert Renauld, Matthew Berriman, Lee Murphy, Corin A Yeats*, et al.* (2005). *Genome of the host-cell transforming parasite Theileria annulata compared with T. parva*. Science **309**(5731): 131-133

3 . lex Bayés, Louie N Van De Lagemaat, Mark O Collins, Mike Dr Croning, Ian R Whittle*, et al.* (2010). *Characterization of the proteome, diseases and evolution of the human postsynaptic density*. Nature neuroscience **14**(1): 19-21

4 . Ziheng Yang (2007). *PAML 4: phylogenetic analysis by maximum likelihood.* Molecular biology and evolution **24**: 1586-1591

5 . Masatoshi Nei, and Takashi Gojobori (1986). *Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions*. Molecular biology and evolution **3**(5): 418-426

6 . Wen-Hsiung Li, Chung-I Wu, and Chi-Cheng Luo (1985). *A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes*. Molecular biology and evolution **2**(2): 150-174

7 . Wen-Hsiung Li (1993). *Unbiased estimation of the rates of synonymous and nonsynonymous substitution*. Journal of molecular evolution **36**(1): 96-

8 . P Pamilo, and No Bianchi (1993). *Evolution of the Zfx and Zfy genes: rates and interdependence between the genes*. Molecular biology and evolution **10**(2): 271-281

9 . Ziheng Yang (2006) *Computational molecular evolution*. Oxford University Press Oxford

10 . W. Huang Da, B. T. Sherman, and R. A. Lempicki (2009). *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources*. Nat Protoc **4**(1): 44-57

11 . M Rothschild, C Jacobson, D Vaske, C Tuggle, L Wang*, et al.* (1996). *The estrogen receptor locus is associated with a major gene influencing litter size in pigs.* Proceedings of the National Academy of Sciences of the United States of America **93**: 201-205

12 . C. K. Tuggle A. L. Vincent, Max F. Rothschild, G. Evans, T. H. Short, O. I. Southwood, G. S. Plastow (1997). *The Prolactin Receptor Gene is Associated with Increased Litter Size In Pigs*. Swine Research Report

13 . S M Galloway, K P Mcnatty, L M Cambridge, M P Laitinen, J L Juengel*, et al.* (2000). *Mutations in an oocyte-derived growth factor gene (BMP15) cause increased ovulation rate and infertility in a dosage-sensitive manner.* Nat Genet **25**: 279-283

14 . C Drogemuller, H Hamann, and O Distl (2001). *Candidate gene markers for litter size in different German pig lines*. Journal of animal

science **79**: 2565-2570

15 . T. Hubbard (2002). *The Ensembl genome database project*. Nucleic acids research **30**: 38-41

16 . Kyung Mo Kim, Samsun Sung, Gustavo Caetano-Anollés, Jae Yong Han, and Heebal Kim (2008). *An approach of orthology detection from homologous sequences under minimum evolution*. Nucleic acids research **36**: e110

17 . S Blair Hedges, Joel Dudley, and Sudhir Kumar (2006). *TimeTree: a public knowledge-base of divergence times among organisms*. Bioinformatics (Oxford, England) **22**: 2971-2972

18 . Ari Löytynoja, and Nick Goldman (2005). *An algorithm for progressive multiple alignment of sequences with insertions*. Proceedings of the National Academy of Sciences of the United States of America **102**: 10557-10562

19 . J. Castresana (2000). *Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis*. Molecular biology and evolution **17**: 540-552

20 . Cristian I Castillo-Davis, Fyodor A Kondrashov, Daniel L Hartl, and Rob J Kulathinal (2004). *The functional genomic distribution of protein divergence in two animal phyla: coevolution, genomic conflict, and constraint*. Genome research **14**: 802-811

21 . Christopher S Peacock, Kathy Seeger, David Harris, Lee Murphy, Jeronimo C Ruiz*, et al.* (2007). *Comparative genomic analysis of three*

Leishmania species that cause diverse human disease. Nature genetics **39**(7): 839-847

22 . Derek Y Chiang, Alan M Moses, Manolis Kellis, Eric S Lander, and Michael B Eisen (2003). *Phylogenetically and spatially conserved word pairs associated with gene-expression changes in yeasts.* Genome biology **4**: R43

23 . *COREMINE medical.*   http://www.coremine.com/medical

24 . M Umeuchi, T Makino, M Arisawa, S Izumi, S Saito*, et al.* (1994). *The effect of interleukin-2 on the release of gonadotropin and prolactin in vivo and in vitro*. Endocrine journal **41**(5): 547

25 . A Christensen, Ge Bentley, R Cabrera, Hh Ortega, N Perfito*, et al.* (2012). *Hormonal regulation of female reproduction*. Hormone and Metabolic Research **44**(08): 587-591

26 . J G Pierce, M R Faith, L C Giudice, and J R Reeve (1976). *Structure and structure-function relationships in glycoprotein hormones.* Ciba Foundation symposium **41**: 225-250

27 . Inchul Choi, Jong-Min Woo, Sunghee Hong, Yong-Keun Jung, Do Han Kim*, et al.* (2003). *Identification and characterization of ADAM32 with testis-predominant gene expression*. Gene **304**: 151-162

28 . Aurélie Vinet, Laurence Drouilhet, Loys Bodin, Philippe Mulsant, Stéphane Fabre*, et al.* (2012). *Genetic control of multiple births in low ovulating mammalian species.* Mammalian genome : official journal of the

International Mammalian Genome Society

29 . H M Lander (1997). *An essential role for free radicals and derived species in signal transduction.* FASEB journal : official publication of the Federation of American Societies for Experimental Biology **11**: 118-124

30 . Kai Simons, and Derek Toomre (2000). *Lipid rafts and signal transduction*. Nat Rev Mol Cell Biol **1**(1): 31-39

31 . P. C. Letourneau (1996). *The cytoskeleton in nerve growth cone motility and axonal pathfinding*. Perspect Dev Neurobiol **4**(2-3): 111-123

32 . Woj M. Wojtowicz, John J. Flanagan, S. Sean Millard, S. Lawrence Zipursky, and James C. Clemens (2004). *Alternative Splicing of Drosophila Dscam Generates Axon Guidance Receptors that Exhibit Isoform-Specific Homophilic Binding*. Cell **118**(5): 619-633

33 . Donna Maglott, Jim Ostell, Kim D Pruitt, and Tatiana Tatusova (2005). *Entrez Gene: gene-centered information at NCBI*. Nucleic acids research **33**: D54-58

34 . Steve Dorus, Eric J Vallender, Patrick D Evans, Jeffrey R Anderson, Sandra L Gilbert*, et al.* (2004). *Accelerated evolution of nervous system genes in the origin of Homo sapiens.* Cell **119**: 1027-1040

35 . Hansell H Stedman, Benjamin W Kozyak, Anthony Nelson, Danielle M Thesier, Leonard T Su*, et al.* (2004). *Myosin gene mutation correlates with anatomical changes in the human lineage.* Nature **428**: 415-418

36 . Xing Fu, Patrick Giavalisco, Xiling Liu, Gareth Catchpole, Ning Fu, *et al.* (2011). *Rapid metabolic evolution in human prefrontal cortex*. Proceedings of the National Academy of Sciences **108**(15): 6181-6186

37 . Mario Cáceres, Carolyn Suwyn, Marcelia Maddox, James W. Thomas, and Todd M. Preuss (2007). *Increased Cortical Expression of Two Synaptogenic Thrombospondins in Human Brain Evolution*. Cerebral Cortex **17**(10): 2312-2321

38 . Yin-Qiu Wang, and Bing Su (2004). *Molecular evolution of microcephalin, a gene determining human brain size*. Human Molecular Genetics **13**(11): 1131-1137

39 . C. S. Peacock, K. Seeger, D. Harris, L. Murphy, J. C. Ruiz, *et al.* (2007). *Comparative genomic analysis of three Leishmania species that cause diverse human disease*. Nat Genet **39**(7): 839-847

40 . B. Angulo, A. Suarez-Gauthier, F. Lopez-Rios, P. P. Medina, E. Conde, *et al.* (2008). *Expression signatures in lung cancer reveal a profile for EGFR-mutant tumours and identify selective PIK3CA overexpression by gene amplification*. J Pathol **214**(3): 347-356

41 . Al Blitz, and Re Fine (1974). *Muscle-like contractile proteins and tubulin in synaptosomes*. Proceedings of the National Academy of Sciences **71**(11): 4472-4476

42 . Maria Fischer, Stefanie Kaech, Darko Knutti, and Andrew Matus (1998). *Rapid Actin-Based Plasticity in Dendritic Spines*. Neuron **20**(5): 847-854

43 . S. F. Cooke, and T. V. P. Bliss (2006). *Plasticity in the human central nervous system*. Brain **129**(7): 1659-1673

44 . Lufen Chang, and Michael Karin (2001). *Mammalian MAP kinase signalling cascades*. Nature **410**(6824): 37-40

45 . Kiyofumi Yamada, and Toshitaka Nabeshima (2003). *Brain-Derived Neurotrophic Factor/TrkB Signaling in Memory Processes*. Journal of Pharmacological Sciences **91**(4): 267-270

46 . V. Kumar, M. X. Zhang, M. W. Swank, J. Kunz, and G. Y. Wu (2005). *Regulation of dendritic morphogenesis by Ras-PI3K-Akt-mTOR and Ras-MAPK signaling pathways*. J Neurosci **25**(49): 11288-11299

47 . Csaba Barta, Maria Sasvari-Szekely, Adrien Devai, Erika Kovacs, Maria Staub*, et al.* (2001). *Analysis of Mutations in the Plasma Cholinesterase Gene of Patients with a History of Prolonged Neuromuscular Block during Anesthesia*. Molecular Genetics and Metabolism **74**(4): 484-488

48 . Kari M Mattila, Juha O Rinne, Matias Röyttä, Pekka Laippala, Timo Pietilä*, et al.* (2000). *Dipeptidyl carboxypeptidase 1 (DCP1) and butyrylcholinesterase (BCHE) gene interactions with the apolipoprotein E ε4 allele as risk factors in Alzheimer's disease and in Parkinson's disease with coexisting Alzheimer pathology*. Journal of medical genetics **37**(10): 766-770

49 . Sayeon Cho, Youn-Jeong Choi, Jeong-Min Kim, Seong-Tae Jeong, Jae-

Hoon Kim, *et al.* (2001). *Binding and regulation of HIF-1α by a subunit of the proteasome complex, PSMA7*. FEBS letters **498**(1): 62-66

50 . Martin Krüger, Carmela Beger, Peter J Welch, Jack R Barber, Michael P Manns, *et al.* (2001). *Involvement of proteasome α-subunit PSMA7 in hepatitis C virus internal ribosome entry site-mediated translation.* Molecular and cellular biology **21**(24): 8357-8364

51 . Nathan Pankratz, Gary W Beecham, Anita L Destefano, Ted M Dawson, Kimberly F Doheny, *et al.* (2012). *Meta‑analysis of Parkinson's Disease: Identification of a novel locus, RIT2.* Annals of neurology **71**(3): 370-384

52 . Justine P Lu, Yuan Wang, Danielle A Sliter, Margaret Mp Pearce, and Richard Jh Wojcikiewicz (2011). *RNF170 protein, an endoplasmic reticulum membrane ubiquitin ligase, mediates inositol 1, 4, 5-trisphosphate receptor ubiquitination and degradation.* Journal of Biological Chemistry **286**(27): 24426-24433

53 . Paul N Valdmanis, Nicolas Dupré, Mathieu Lachance, Shawn J Stochmanski, Veronique V Belzil, *et al.* (2011). *A mutation in the RNF170 gene causes autosomal dominant sensory ataxia.* Brain **134**(2): 602-607

54 . Alessandro Rosa, and Ali H Brivanlou (2010). *A regulatory circuitry comprised of miR-302 and the transcription factors OCT4 and NR2F2 regulates human embryonic stem cell differentiation.* The EMBO journal **30**(2): 237-248

55 . Richard T Premont, Audrey Claing, Nicolas Vitale, Stephen J Perry, and

Robert J Lefkowitz (2000). *The GIT family of ADP-ribosylation factor GTPase-activating proteins functional diversity of GIT2 through alternative splicing*. Journal of Biological Chemistry **275**(29): 22373-22380

56 . Yuichi Mazaki, Shigeru Hashimoto, Katsuya Okawa, Asako Tsubouchi, Kuniaki Nakamura*, et al.* (2001). *An ADP-ribosylation factor GTPase-activating protein Git2-short/KIAA0148 is involved in subcellular localization of paxillin and actin cytoskeletal organization*. Molecular biology of the cell **12**(3): 645-662

57 . Scott R Frank, Molly R Adelstein, and Steen H Hansen (2006). *GIT2 represses Crk-and Rac1-regulated cell spreading and Cdc42-mediated focal adhesion turnover*. The EMBO journal **25**(9): 1848-1859

58 . Hélène Le Pabic, Dominique Bonnier, Ulla M Wewer, Alexandre Coutand, Orlando Musso*, et al.* (2003). *ADAM12 in human liver cancers: TGF-β-regulated expression in stellate cells is associated with matrix remodeling*. Hepatology **37**(5): 1056-1066

59 . Masanori Asakura, Masafumi Kitakaze, Seiji Takashima, Yulin Liao, Fuminobu Ishikura*, et al.* (2002). *Cardiac hypertrophy is inhibited by antagonism of ADAM12 processing of HB-EGF: metalloproteinase inhibitors as a new therapy*. Nature medicine **8**(1): 35-40

60 . Marie-Florence Galliano, Clotilde Huet, Jessica Frygelius, Anna Polgren, Ulla M Wewer*, et al.* (2000). *Binding of ADAM12, a marker of skeletal muscle regeneration, to the muscle-specific actin-binding protein, α-actinin-2, is required for myoblast fusion*. Journal of Biological Chemistry

275(18): 13933-13939

61 . Ulrike Novak (2004). *ADAM proteins in the brain*. Journal of Clinical Neuroscience **11**(3): 227-235

62 . Darryl M Horn, Saundra L Mason, and Katrin Karbstein (2011). *Rcl1 protein, a novel nuclease for 18 S ribosomal RNA production*. Journal of Biological Chemistry **286**(39): 34082-34087

63 . Guy Mckhann, David Drachman, Marshall Folstein, Robert Katzman, Donald Price*, et al.* (1984). *Clinical diagnosis of Alzheimer's disease Report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease*. Neurology **34**(7): 939-939

64 . Andrew J Hughes, Susan E Daniel, Linda Kilford, and Andrew J Lees (1992). *Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases*. Journal of Neurology, Neurosurgery & Psychiatry **55**(3): 181-184

국문초록

dN/dS는 유전자 염기서열의 진화 가속에 대한 분석에 널리 쓰이고 있다. dN/dS는 코딩 단백질에 변화가 생기는 염기 치환인 비동의 치환(non-synonymous substitution)이 일어난 비율(dN)과, 코딩 단백질에 변화를 주지 않는 염기 치환인 동의 치환(synonymous substituion)이 일어난 비율(dS)을 비교한 것이다. 이 추정치는 계통분류학적으로 연관되어 있는 종들 사이의 유전적 흐름을 밝히기 위해 사용되며, 그 종들 간의 이종상동성 유전자(orthologous gene)들의 진화 가속 정도를 추정한다.

첫번째 연구에서는, 포유류 6종(인간, 쥐, 말, 개, 소, 돼지)을 비교하여 단태동물과 다태동물 사이의 차이를 뒷받침하는 유전적 진화 근거를 찾기 위해 노력했다. 이 여섯 종들 중 인간, 말, 소는 단태동물로 분류되고, 쥐와 개, 돼지는 다태동물로 분류된다. 이 연구에서, 나는 번식 특성의 진화적 차이에 관련되는 것으로 추정되는 후보 유전자들을 제시했다. 또한 단태동물 그룹과 다태동물 그룹 각각에서 진화 가속을 겪은 유전자들이 기능적인 차이점을 보인다는 것을 알 수 있었고, 이종상동성 유전자들의 dN/dS 값이 다태동물보다 단태동물에서 더 높게 추정되는 것을 확인했다.

두번째 연구에서는 사람의 뇌 기능과 관련된 유전적 진화를 설명하기 위해 인간을 포함한 유인원 10종에 대해 dN/dS 비율 분석을 수행했다. 유인원 9종에 비해 인간에서 더 많이 진화적 가속을 겪은 유전자는 분석 결과 41개를 뽑을 수 있었다. 이 유전자들의 기능을 분석한 것을 토대로, 나는 유인원들과 비교하여 인간 종 진화가 기억력과 인식 능력이 향상되는 방향으로 이루어졌을 가능성을 제시했다. 그리고 인간에서 진화 가속이 가장 많이 된 열 개 유전자들 중 두 유전자가 뉴런 장애 질병(neuronal disorder disease)에 관련된 것을 밝혀 인간의 신경 특이적 진화의 근거를 뒷받침했다.