Comparative Analysis of Large Genome in Human-Chimpanzee

인간-침팬지간 대량의 지놈서열 비교분석

Tae-Hyung Kim[1,2], Dae-Soo Kim[1,2], Yeo-Jin Jeon[1,2], Hwan-Gue Cho[1,3] and Heui-Soo Kim[1,2*]

1 Interdisciplinary program of bioinformatics, Pusan National University

2 Dept of Biology, College of Natural Sciences, Pusan National University

3 ALGORIGENE, Bioinformatics Lab. Dept. Computer Science, Pusan National University

*To whom correspondence should be addressed. E-mail: khs307@pusan.ac.kr

Abstract

With the availability of complete whole-genomes such as the human, mouse, fugu and chimpanzee chromosome 22, comparative analysis of large genomes from cross-species at varying evolutionary distances is considered one of a powerful approach for identifying coding and functional non-coding sequences. Here we describe a fast and efficient global alignment method especially for large genomic regions over mega bases pair. We used an approach for identifying all similarity regions by HSP (Highest Segment Pair) regions using local alignments and then large syntenic genome based on the both extension of anchors at HSP regions in two species and global conservation map. Using this alignment approach, we examined rearrangement *loci* in human chromosome 21 and chimpanzee chromosome 22. Finally, we extracted syntenic genome 30 Mb of human chromosome 21 with chimpanzee chromosome 22, and then identified genomic rearrangements (deletions and insertions ranging in size from 0.3 to 200 kb). Our experiment shows that all insertion/deletion (indel) events in excess of 300 bp within chimpanzee chromosome 22 and human chromosome 21 alignments in order to identify new insertions that had occurred over the last 7 million years of evolution. Finally we also discussed evolutionary features throughout comparative analyses of Ka/Ks (non-synonymous / synonymous substitutions) rate in orthologous 119 genes of chromosome 21 and 53 genes of MHC-I class in human and chimpanzee genome.

Introduction

지난 10년간 많은 생물학적 연구 노력으로 여러 생물종의 전체 지놈을 얻게 되었다. 현재까지 완벽하게 지놈이 결정된 생물은 고세균류 16종, 박테리아 96종, 진핵생물에 있어서는 효모(Saccharomyces cerevisiae), 말라리아 균(Plasmodium falciparum), 모기

(Anopheles gambiae), 꼬마선충(Caenorhabditis elegans), 애기장대(Arabidopsis thaliana), 쌀(Oryza sativa), 초파리(Drosophila melanogaster), 복어(Fugu ribripes), 인간(Homo sapiens)은 완벽한 지놈서열이 밝혀진 상태이며 마우스(Mus musculus), 쥐(Rattus norvegicus,), 어류의 모델 생물인 제브라피쉬(Zebrafish)의 경우는 곧 완벽한 전체 지놈을 얻게 될 것이다. 그리고 몇 년 후에는 인간의 특이적 질병 연구를 위하여 휴먼과 제일 가까운 영장류에 있어서도 침팬지와 마카카 원숭이의 지놈서열이 우선적으로 선택되어 결정 될 것이다.

밝혀진 지놈 서열들은 종간의 차이 및 진화의 고리를 찾기 위해 지놈을 비교하기 시작하였으며 앞서 수행되어 보고된 연구로써 복어와 마우스 지놈의 휴먼과 전체 지놈 비교 분석을 통해 종간의 잘 보존된 영역을 발견하여 이로써 새로운 유전자 발굴에 이바지 하였으며 더불어 단백질 코딩영역은 아니지만 유전자를 조절 하는 보존된 비번역 서열들을 찾아 내는데 이용되었다[1], [2], [3]. 이러한 중요한 비교 분석결과를 도출할 수 있음에도 불구하고 척추동물의 지놈 비교분석에 있어 가장 문제가 되고 있는 것은 지놈에 넓게 차지하여 흩어져있는 반복 서열[4]인 2bp(simple repeat)에서 크게는 10kb (interspersed repeat or retrotransposon)정도의 지놈서열이 연속적으로 또는 무작위로 반복되어 흩어져 있다. 이러한 메커니즘으로 인해 척추동물의 지놈의 크기를 50%이상 확장 시켜 오면서 지놈의 대부분을 차지하고 있다[5]. 그러므로 반복 서열들은 paralogous genome을 형성하게 되며 더욱 더 orthologous genome을 동정하기에 어려워 지게 된다. 또 한가지 문제로써 대부분 BAC clone 레벨에서 지놈 서열결정을 하게 됨으로 인해 BAC 서열을 이용해 다른 생물종에 있어 orghologous한 지놈을 추출하고 동시에 이들 100~200kb정도에 해당하는 BAC서열들을 assembly해야 한다. 본 작업에서는 이러한 문제를 해결하기 위해 local alignment와 global alignment의 장점을 취합하기로 하였다. 즉 우리가 잘 알고 있는 local alignment(BLAST, FASTA, cross_match, 등)의 경우는 비교적 작은 서열의 데이터를 query로 거대 서열 데이터베이스에 적용하여 두 서열 유사성 검색에 있어 발생된 가능한 서열영역 들을 찾아냄으로써 연구자들이 이들 데이터들을 비교 분석할 수 있는 이점이 있었지만 서열이 지놈레벨로 확대 되게 되면 paralogous한 지놈영역에 의해 유사서열 영역은 지놈 길이에 비례해서 많이 얻어지게 될 것이며 이로 인해 두 생물종간에 존재하는 실제 orthologous 영역과 paralogous한 영역의 구별이 더욱 더 어려워지게 된다.

Global alignment는 두 서열 전체에 걸친 서열 정열을 통해 비교적 긴 지놈을 얻어낼 수가 있다. 하지만 지놈레벨로 확대되게 되면 대부분 Needlemann-Wunsch 알고리즘으로 구현된 global alignment(ALIGN, NAP, GAP0, 등) 의 경우에 있어 지놈의 길이에 커게 영향을 받아 상당히 많은 메모리 공간과 계산 시간을 필요하게 되며 Mb단위의 지놈서열을 비교할 때 어려움이 있다.

그래서 본 연구에서 채택한 방법으로 비교적 긴 지놈 서열의 해싱을 통해 빠른 local alignment를 수행하는 Megablast 프로그램을 이용하여 두 지놈내 서열 유사성 검색에 의한 가능한 HSP(높은 점수를 가진 영역)를 모두 결정하고 각각의 지놈에서 발생한 HSP들이 방향과 순서가 같으며 연속적

인 것을 표지지점으로 하여 점진적으로 시작과 끝을 이어나가는 global alignment map구성의 방법을 활용함으로써 비교적 효과적이고 빠르게 두 지놈 전체에 걸쳐 orthologous genomes을 얻어낼 수가 있었다. 이렇게 얻어진 orthologous 지놈은 지놈의 기능 및 진화를 연구하는데 있어 매우 훌륭하게 활용될 수 있다6], [7]. 각 지놈에 포함되어 있는 단백질 번역지역 (protein coding region)에 있어서 차이 및 지놈 재배열 (genomic rearrangement) 그리고 유전자, 조절영역, 반복서열 등의 진화적 거리 및 돌연변이 속도를 계산함으로써 그들의 생성 시기를 알 수가 있으며 종 특이적인 영역을 발견함으로써 표현형 연구에 있어 큰 도움이 되고 있다.

본 연구에서는 global alingnment mapping방법을 이용해 인간과 가장 가까운 진화적 거리를 가진 Chimpanzee (Pan troglodytes)와 인간의 지놈 서열을 비교 분석하는 작업을 수행하였다.

Materials and Methods

휴먼과 침팬지 서열 데이터

침팬지 지놈 프로젝트에 의해 밝혀진 22번 염색체에 해당하는 297개의 BAC 서열을 휴먼과 cross-species global alignment map구성을 위한 데이터 셋으로 활용하였다. 이 영역에 대한 상대 지놈으로는 휴먼 지놈 프로젝트 수행 후 완성된 휴먼지놈 (build 33, 14-Apr-2003) 데이터를 이용하였다. 그리고 휴먼과 침팬지의 MHC-1 class영역 지놈 서열 데이터도 추가적으로 분석하였다.

Global Alignment Map 구성

35Mb정도의 지놈 서열 정렬을 위해 본 작업에서는 4단계의 순차적 작업을 계획하였다. 첫 번째로 지놈내 반복서열을 제거(masking)하는 전처리 과정을 거쳤으며, 두 번째로 비교적 빠르고 정확하게 찾기 위한 local alignment (Megablast)툴을 이용 모든 가능한 표지 지점(anchor)을 찾았다. 세 번째로 global alignment map을 구성하기 위한 최적의 표지 지점을 선택하였으며, 마지막으로 어셈블리를 수행하였다.

1. 전처리 작업

휴먼과 침팬지를 포함한 포유류 지놈의 상당히 많은 부분이 paralogous (반복서열, RNA 유전자, 패밀리 유전자)영역들에 의해 실제 지놈을 비교 분석함에 있어 orthologous 한 영역을 찾는 것을 어렵게 하고 있다. 그러므로 RepeatMasker (http://ftp.genome.Washington.edu/RM/RepeatMasker.html, Smit and Green) 프로그램을 이용하여 휴먼 지놈내 존재하는 반복서열 및 과거 바이러스 유래 내재 반복 서열 (retrotransposon) 그리고 RNA 유전자 등과 같은 실제 orthologous 한 영역을 찾는데 방해가 될만한 이들 반복서열 영역을 masking 하였다.

2. 빠른 local alignment 전략

local alignment방법을 사용 지놈내 높은 상동성에 의해 검출되며 두 지놈의 전체 orthologous 지도작성을 가능케 할 수 있는 지놈내 흩어져 있는 표지 지점을 빠르고 정확하게 모두 찾고자 하였다. 이때 사용한 툴은 hash 및 여러heuristic기법을 적용한 blast 알고리즘에 greedy algorithm을 새로 적용하여 Mega단위의 지놈서열을 local

alignment를 하는데 강력한 프로그램인 Megablast툴을 활용하였으며 이 프로그램만 가지고는 지놈 내부의 모든 orthologous한 영역을 추출하지는 못하나, 본 작업에서 목표로 하는 최적의 global alignment map을 구성하기 위하여 모든 가능한 표지지역을 찾는데 있어서는 효과적인 툴이다.

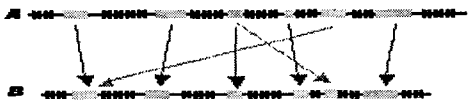


그림1. 두 지놈 내 존재하는 반복서열을 모두 masking (Ns) 한 뒤 표지지점으로 가능한 모든 HSP를 찾아냄.

3. *Global alignment 를 위한 지놈내 표지지역 선택*

local alignment에 의해 결정된 global alignment map의 표지 지역이 될 후보 HSP들을 선택하기 위해 휴먼 지놈을 reference 서열로 하여 침팬지의 HSP결과를 순서대로 재배치 하였으며 침팬지 각각의 BAC 내부에서도 순서대로 재배치하였다. 이들 BAC 서열들은 다양한 길이와 quality를 가지고 있으며 서로 다른 BAC들이 휴먼 지놈의 같은 영역에 여러 중복된 HSP를 발생하게 된다. 그러므로 어떠한 BAC 서열의 HSP를 global alignment map을 구성할 표지로 선택할 것인지를 결정하여야 한다. 이때 제안된 한가지 방법으로 reference 지놈에 맵핑된 BAC들을 서로 연결했을 때 gap을 전혀 발생시키지 않거나 가능하면 작게 하는 BAC들의 HSP그룹을 우선순위로 선택하였으며, 만약 비슷한 길이의 gap을 발생시키게 된다면 이때는 가장 길이가 길어 서열의 완성도가 높은 BAC clone의 HSP그룹을 선택 하였다.

그림 2. 여러 BAC clone들이 중복되어져 있음을 보여준다. (파란선: 중복되어 서열 결정된 BAC clone, 검은선: 높은 quality와 어셈블리시 갭을 발생시키지 않은 BAC clone)

4. *표지지점을 연결하는 global alignment 맵 구성 및 BAC 서열의 지놈 어셈블리*

두 지놈상의 전체 global alignment map을 구성하기 위한 HSP 표지 지점들이 선택되었으며 선택된 표지지점을 근거하여 이어나가되 중복된 서열을 잘라내고 붙여 하나의 지놈 contig (30Mb)를 어셈블리 하였다. 이렇게 BAC clone과 같이 비교적 긴 길이(100~267kb)와 많은(297개)개수의 BAC서열들을 assembly하기 위해서는 global alignment map을 활용하여 reference 휴먼지놈과 비교함으로 해서 침팬지의 BAC을 assembly와 동시에 휴먼과 지놈 비교에 사용될 수 있는 orthologous한 침팬지 30Mb정도의 지놈 영역을 얻어낼 수가 있었다.

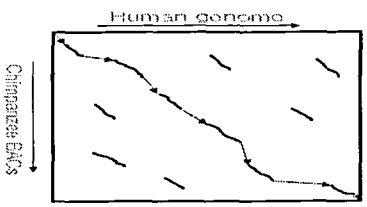


그림 3. HSP를 이용한 global alignment map구성.

지놈 분석

global alignment map에 의해 얻어진 orthologous한 침팬지 지놈 서열과 휴먼 서

열을 비교하기 위해 DotPlot과 PipMaker를 활용하였으며 지놈내 유전자 위치정보는 genbank에서 휴먼의 코딩영역을 추출하였으며 반복서열에 대한 위치 정보는 RepeatMasker를 이용하여 지놈서열에 존재하는 반복서열(LINE, HERV, ALU, simple repeat등)을 동정하였다. 그리고 30MB정도 되는 orthologous지놈을 DotPlot이나 PipMaker로 분석 가능한 적당한 크기인 1.45MB로 잘랐다. 최종적으로 23개의 orthologous 지놈 contig서열을 만들어 휴먼과 침팬지 서열을 비교하는데 활용하였다.

번역서열의 Ka와 Ks 계산

휴먼지놈에 이미 알려진 단백질 번역위치를 이용하여 침팬지에서도 같은 번역서열을 추출하였으며 각 유전자 쌍을 ClustalW를 사용하여 서열정열을 하였다. 서열정열 결과에서 삽입/결실이 발생하거나, 종결코돈이 예상한 것보다 앞서 나타난 유전자 쌍은 모두 제거 하였다. Nei-Gojobori Method의 P_S= Sd/S 와 P_N = Nd/N의 수식을 이용 각 번역서열의 Ka (Non-Synonymous)와 Ks (Synonymous)를 계산하였다.

Results

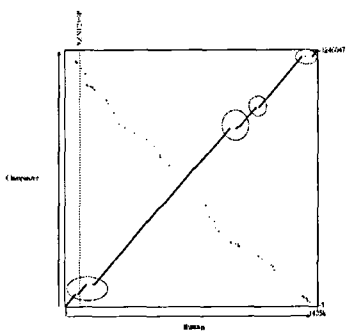
Global alignment map구성과 어셈블리

기존에 잘 알려진 지놈 서열을 global alignment하는 알고리즘인 Needleman-Wunsch(1970)는 비교할 서열의 길이가 길어질수록 비례해서 수행하는 시간이 많이 소요되며 대량의 서열을 다루기 위해서는 무리가 있다. 현재 이러한 문제점을 극복하기 위해 MUMmer [8][9], GLASS [10], VISTA [11] 그리고 AVID [12] 와 같은 global alignment프로그램이 개발되었으며 지놈레벨의 서열들을 효과적으로 정렬할 수가 있었다. 하지만 이들은 휴먼과 같은 매우 큰 reference지놈 서열과 비교할 대상의 서열이 다수의 BAC 서열에 기초할 경우 global alignment와 동시에 어셈블리도 또한 수행되어져야 하나 그렇지 못하다. 그러므로 본 작업에서는 이를 가능케 하기위해 긴 지놈의 서열을 global alignment 해주는 프로그램들이 공통적으로 수행하고 있는 알고리즘인 지놈내 서열상의 유사성이 높은 anchor를 찾아 같은 방향과 순서를 고려하여 연결하여 나감으로 해서 하나의 커다란 지놈을 구성하는 방식을 그대로 활용하였다. 이들 BAC clone contig을 global alignment와 동시에 어셈블리하기 위해 4단계(1.반복서열 제거, 2.Megablast (blastz) 프로그램을 활용하여 가능한 HSP생성, 3.global alignment map을 생성키 위한 후보 BAC및 anchor들 선택, 4.이들 BAC과 anchor를 바탕으로 중복을 제거 하고 하나의 길다란 othologous 지놈 추출)와 같이 단계별 작업을 거쳐 휴먼 지놈의 21번 염색체의 contig 5개와 침팬지 22번 염색체의 BAC 서열 297개를 이용해 긴 30Mb정도의 orthologous 지놈을 assembly하여 얻어낼 수가 있게 되었다.

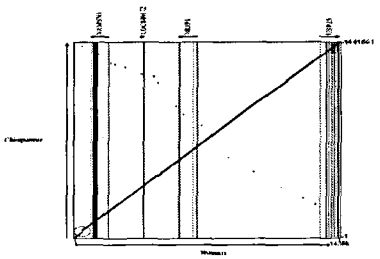
Large-scale의 지놈 rearrangement 분석

앞서 휴먼과 침팬지 지놈에서 얻어낸 30MB정도의 orthologous genome을 통해 우리는 여러 흥미로운 생물학적 의미를 가진 발견을 할 수가 있었다. 이들 두 생물의 지놈은 공통조상으로부터 최근에 분기

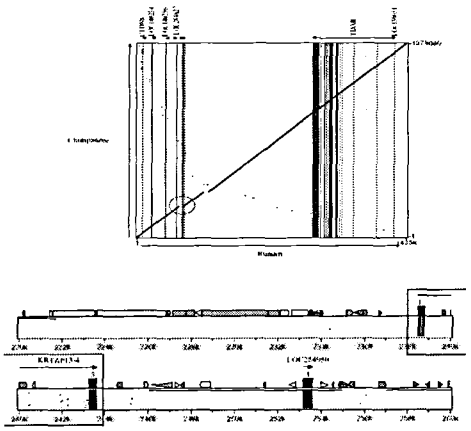
(5Myr~6Myr)되어 짧은 진화 역사를 가지고 있음으로 인해 코딩영역뿐만 아니라 인트론 영역 그리고 인트제닉한 영역까지도 대체적으로 매우 높게 잘 보존되어 있음을 알 수가 있었다. 하지만 가까운 진화적 거리를 가지고 있음에도 불구하고 지놈 서열 사이의 변이는 매우 미흡한 반면에 빈번한 지놈 재배치 (rearrangement)이 발생함을 확인하였다. 작게는 0.3kb에서 크게는 200kb 정도의 지놈 재배열이 발생하였으며 한 예로 인트제닉 지놈에 있어 삽입 및 결실(그림 4. A, B)이 나타난 곳과 유전자 코딩영역의 결실(그림 4. C) 그리고 지놈 서열의 반복에 의한 지놈의 확장(그림 4. D)과 같은 여러 다양한 지놈 재배열을 보여주고 있었다. 이러한 다양한 지놈 재배열중 KRTAP13-4라는 유전자를 포함한 5kb정도의 지놈이 침팬지에서는 결실이 되거나 인간에게 삽입이 발생한 것으로 확인되었으며 털 형성에 관여하는 표현형의 변화에 관계가 있을 것으로 보아진다. 하지만 단 하나의 유전자영역의 지놈 재배열만 보였으며 전체 21번 염색체 상에서 하나만 차이를 보이는 것으로 보아 유전자 수적인 차이에 의한 침팬지 휴먼사이의 표현형 변화는 크게 기대되지는 않는다.



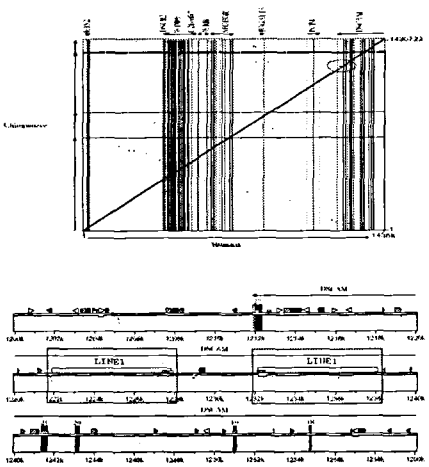
A. 휴먼 지놈에서는 존재하나 침팬지 지놈에서는 존재하지 않는 영역.



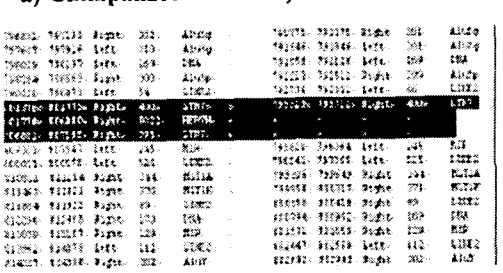
B. 침팬지 지놈에서는 존재하나 휴먼 지놈에서는 존재하지 않는 영역.



C. 휴먼지놈 내부에 keratin 단백질을 코딩하는 유전자 KRTAP13-4 (5kb) 를 포함한 영역 (24kb) 가 침팬지 지놈에서는 결실되어 나타나고 있음을 보여줌.



D. 휴먼 지놈의 지놈 조각중 일부 영역인 LINE1 element가 역 반복서열로 중복되어 palindrome 구조를 형성.



E. LTR 상동성 재배열에 의한 HERV (Human Endogenous Retrovirus) 지놈의 결실

그림 4. 휴먼과 침팬지 사이에 지놈 rearrangement발생에 대한 여러 분석 결과.

plot들의 색깔들은 다음과 같은 요소를 지정하고 있다. 붉은색: 정방향 엑손, 푸른색: 역방향 엑손, 오렌지색: 정방향 UTR(untranslation region), 초록색: 역방향 UTR, 노란색: 인트론, 흰색: Intergenic서열 지놈 재배열이 발생한 영역을 붉은색 동그라미로 표시를 하였으며 C, D에서는 더 자세한 지놈 영역을 확인하기 위해 그 아래에 Pip(percent identity plot)을 보여주고 있다. E는 Retrotransposon 영역의 삽입/결실 결과를 텍스트 파일로 보여줌.

휴먼 21 번 염색체와 침팬지 22 번 염색체의 orthologous 유전자들의 Positive 선택압 측정

휴먼과 침팬지에 있어서 orthologous 유전차 147 개 추출한 뒤 번역서열 내부에 삽입/결실이 일어나거나 종결코돈이 나오는 28 개 데이터를 제외하면 총 119 개의 유전자를 대상으로 선택압 측정한 결과 $K_a/K_s > 1$ 인 유전자를 147 개 유전자 중에 12 개를 찾았으며 LOC254950 유전자의 경우는 K_a/K_s = 1.66 으로 비교적 높은 positive 선택압을 받았으며 *KAP15.1* 로 명명되어 있는 유전자였다[13]. 이 유전자의 기능 또한 앞서 설명된 KRTAP13-4 와 비슷한 패밀리 유전자로써 모세포 형성에 관여하고 있었다.

휴먼 21 번 염색체의 orthologous 유전자 그룹과 MHC-1 class 유전자 그룹의 positive 선택압 비교 분석

휴먼 21 번 염색체 유전자와 침팬지 22 번 염색체의 orthologous 전체 유전자들의 진화적인 특징을 알기 위해서 MHC-1 class orthologous 유전자 54 개를 ka, ks 계산을 하고 비교를 하였다. 이들 각 유전자 그룹(human_chr21-chimpanzee_chr22, MHC-1)을 한 직선에 적합하기 위하여 회귀분석을 하였으며, 이때 직선의 기울기는 ka/ks 와 거의 일치한다는 가정을 하였으며, 유전자 그룹들간의 진화적 패턴을 잘 설명할 수가 있다고 보아진다. 이를 비교한 결과에 따르면 MHC-1 class 의 경우는 기울기 (ka/ks)가 0.587917 의 값이 나왔으며 human_chr21-chimpanzee_chr22 의 경우는 MHC-1 class 와 비교해서 1/3 배 정도의 기울기 0.201868 (ka/ks)를 보였다(그림 5).

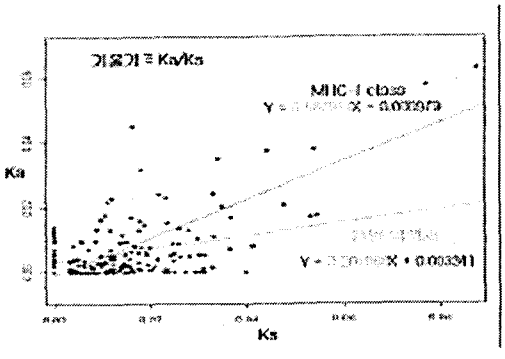


그림 5. MHC-1 class와 human_21-chimpanzee_22 유전자 그룹으로써의 진화적 선택압 비교.

Discussion and Further study

생물 종 간(cross-species)의 전체 지놈을 비교하고 해석하는 방법은 지놈의 기능 및 진화적인 관계를 이해하는데 중요하다.

본 논문에서는 이러한 전체 지놈 서열 비교의 문제에 대한 시도로써 수십~수백개의 BAC *contig*서열들을 reference 지놈에 맵핑하여 순서와 방향이 잘 보존된 영역을 연결함으로 해서 global alignment map을 형성한 뒤 대량의 orthologous한 지놈을 얻어 내는 방법을 고안하였다.

이렇게 얻어진 지놈은 두 종간의 유전학적 차이와 표현형과 연관해서 해석 가능하며 또한 유전자들의 기능과 진화적인 관계를 연관시켜 설명할 수 있는 고리를 찾을 수 있을 것이다. 본 논문에서는 짧은 진화적 거리를 가지고 있으며 비교적 정확한 분기연대를 동위원소 추정법에 의해서 잘 알고 있는 침팬지와 휴먼의 지놈을 선택하여 비교하게 되었다. 그 결과 전체적으로 번역서열영역과 비 번역서열 사이의 돌연변이율은 크게 차이를 보이지 않았으나 짧은 분기시간대에도 불구하고 0.3kb ~ 200kb에서 해당하는 지놈 재배열이 다수 일어나는 것을 확인하였다. 그리고 대부분의 지놈 재배열이 반복서열 영역에서 발생하였으며 이들 반복 서열들이 지놈 재배열을 매개할 수 있는 가능을 확인하였다. 유전자 영역에 있어서 지놈 재배열은 200여 개 가까운 유전자 중 실제 표현형 단계에서 발현된다고 확인된 단 하나의 유전자 *KRTAP13-4*만이 확인되었으며 휴먼의 삽입 또는 침팬지 유전자 결실로 보아진다. 이 유전자는 모세포 형성의 중요한 기능을 하고 있으며 인간에게만 존재하며 침팬지에서는 존재하지 않는 것으로 보아 두 종간에 표현형에서 뚜렷한 차이를 보이는 머리 및 몸에서 형성되어지는 털과 어떤 관계가 있을 것으로 보아진다. 그리고 더불어 positive selection의 선택압을 받는 유전자 분석에서도 *KRTAP13-4*와 유사한 패밀리 유전자인 *KAP15.1*가 높은 ka/ks 값으로 확인됨으로 인해 이들 패밀리 유전자들에 특히 집중하여 지놈 재배열과 단백질 영역의 급속한 돌연변이가 일어남으로 인해 휴먼의 모세포 형성이 침팬지와 큰 차이를 보임을 지놈 분석을 통해 예측할 수가 있었다. 이와 더불어 human 21번과 침팬지 22번 염색체에 존재하는 orthologous유전자와 MHC-1 class의 orhtologous유전자의 positive selection 선택압의 통계적 분석을 통해 이들 두 그룹이 매우 짧은 종간 진화적 거리에도 불구하고 매우 다른 선택압을 받아왔음을 알 수가 있었다. 특히 MHC-1 class의 경우는 항상 존재하는 생체 내외부의 다양한 이물질의 공격을 방어하는 생체 면역에 관여함으로 인해 다양성을 크게 요구 받게 되었으며 이들 유전자의 단백질 레벨에서의 돌연변이가 이득이 되었을 것으로 보인다.

휴먼의 21 염색체의 유전자 그룹의 경우는 수소의 유전자만이 positive selection 유전자의 성질을 보이고 있으며 21번 염색체에 있는 유전자들은 대체적으로 다양성을 요구하지 않는 유전자로 구성되어 있음이 MHC-1 class와 비교를 통해 확인할 수가 있었다.

유전자 기능과 진화적인 관계를 연관시켜 확인한 결과 ka/ks >1에 해당하는 유전자들은 대체적으로 뇌와 신경발달, 면역에 관여하고 있으며 단백질의 세포학적 구성물질로써의 기능 분류에서는 유동성 및 다양성을 요구하는 막 단백질 형성에 관여하고 있음을 확인하였다. 반대로 ka/ks < 1에 해당하는

negative selection 선택압을 받는 유전자는 포유류로부터 미생물까지 잘 보존되는 세포 생장주기(cell cycle) 및 에너지 대사(energy metabolic) 기능으로 분류되었다.

차후 이러한 global alignment map 방법을 개선하여 다른 영장류 지놈과 마우스 및 초파리와 같은 진화적으로 먼 생물종에 있어서도 또한 대량의 지놈 분석을 통하여 다종간 지놈 비교를 위한 orthologous 맵작성 및 분석을 가능케 하고자 한다.

References

[1] Hardison, R. and Miller, W. 1993. Use of long sequence alignments to study the evolution and regulation of mammalian globin gene clusters. *Mol. Biol. Evol.* 10: 73-102

[2] Gumucio, D., Shelton, D., Zhu, W., Millinoff, D., Gray, T., Bock, J., Slightom, J., and Goodman, M. 1996. Evolutionary strategies for the elucidation of cis and trans factors that regulate the developmental switching programs of the β-like globin genes. *Mol. Phylog. Evol.* 5: 18-32

[3] Pennacchio, L.A. and Rubin, E.M. 2001. Genomic strategies to identify mammalian regulatory sequences. *Nat. Rev. Genet.* 2: 100-109

[4] Chen, R., Bouck, J.B., Weinstock, G.M., and Gibbs, R.A. 2001. Comparing vertebrate whole-genome shotgun reads to the human genome. *Genome Res.* 11: 1807-1816

[5] Liu, Ge, Program, NISC Comparative Sequencing, Zhao, Shaying, Bailey, Jeffrey A., Sahinalp, S. Cenk, Alkan, Can, Tuzun, Eray, Green, Eric D., Eichler, Evan E. 2003. Analysis of Primate Genomic Variation Reveals a Repeat-Driven Expansion of the Human Genome. *Genome Res.* 13: 358-368

[6] Batzoglou, S., Pachter, L., Mesirov, J.P., Berger, B., and Lander, E.S. 2000. Human and mouse gene structure: Comparative analysis and application to exon prediction. *Genome Res.* 10: 950-958

[7] Kent, J. 2002. BLAT — The BLAST-like alignment tool. *Genome Res.* 12: 656-664

[8] Delcher, A.L., Kasif, S., Fleischman, R., Peterson, J., White, O., and Salzberg, S.L. 1999. Alignment of whole genomes. *Nucleic Acids Res.* 27: 2369-2376

[9] Delcher, A.L., Phillippy, A., Carlton, J., and Salzberg, S.L. 2002. Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res.* 30: 2478-2483

[10] Batzoglou, S., Pachter, L., Mesirov, J., Berger, B., and Lander, E.S. 2000. Human and mouse gene structure: Comparative analysis and application to exon prediction. *Genome Res.* 10: 950-958

[11] Mayor, C., Brudno, M., Schwartz, J.R., Poliakov, A., Rubin, E.M., Frazer, K.A., Pachter, L., and Dubchak, I. 2000. VISTA:

Visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics.* 16: 1046-1047

[12] Bray, N., Dubchak, I., and Pachter, L. 2003. AVID: A global alignment program. *Genome Res.* 13: 97-102

[13] Michael, A.R., Lutz, L., Hermelita, W., Claudia, E., Silke, P. and J? gen S. 2002. Characterization of a First Domain of Human High Glycine-Tyrosine and High Sulfur Keratin-associated Protein (KAP) Genes on Chromosome 21q22.1. *J. Biol. Chem.* 277: 48993-49002