

## Comparative Analysis of Large Genome in Human - Chimpanzee

-

Tae-Hyung Kim<sup>1,2</sup>, Dae-Soo Kim<sup>1,2</sup>, Yeo-Jin Jeon<sup>1,2</sup>, Hwan-Gue Cho<sup>1,3</sup> and Heui-Soo Kim<sup>1,2\*</sup>

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

10

16 , 96 ,  
(*Saccharomyces cerevisiae*),  
(*Plasmodium falciparum*),  
(*Anopheles gambiae*), (Caenorhabditis

elegans), (Arabidopsis thaliana), orthologous  
(Oryza sativa), (Drosophila melanogaster), 100~200kb BAC  
(Fugu ribripes), (Homo sapiens) assembly .  
local alignment  
(Mus musculus), (Rattus norvegicus.), global alignment  
(Zebrafish) 가 local  
alignment(BLAST, FASTA, cross\_match, )  
query  
가 가

paralogous

orthologous paralogous

Global alignment

[1], [2],

[3].

가 .

Needleman-Wunsch

가 가

global alignment(ALIGN, NAP, GAP0,

)

[4] 2bp(simple repeat) 10kb

(interspersed repeat or retrotransposon)

Mb

50%

local alignment

Megablast

[5].

paralogous

genome

가

HSP(

가

orthologous genome

)

가

BAC

HSP

가

clone

BAC

가

global alignment map

orthologous genomes 가 . (masking) ,

orthologous local alignment (Megablast) 가

(anchor) .

6], [7]. global alignment map

(protein coding ,

region) .

(genomic rearrangement) ,

, I.

가 paralogous ( ,

RNA , )

global alignment mapping orthologous

가 가 RepeatMasker

가 Chimpanzee (*Pan troglodytes*) (<http://ftp.genome.Washington.edu/RM/RepeatMasker.html>, Smit and Green)

## Materials and Methods

(retrotransposon) RNA

orthologous masking

가

22

297 BAC

cross-species global alignment map

(build 33, 14-

Apr-2003)

MHC-1 class

가

Global Alignment Map

blast hash heuristic

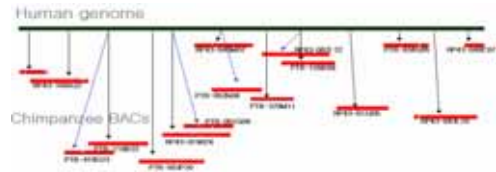
Mega greedy algorithm local

alignment

35Mb Megablast

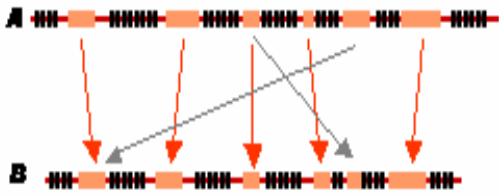
4

가  
map  
orthologous  
global alignment  
가



2. BAC clone

( :  
BAC clone, : quality  
BAC clone)



1.  
masking (Ns) 가  
HSP

4. global alignment  
BAC  
global alignment map  
HSP

3. Global alignment

local alignment  
alignment map  
global  
HSP

가  
contig (30Mb)  
BAC clone  
(100~267kb) (297 ) BAC  
assembly  
global alignment  
map reference  
BAC assembly

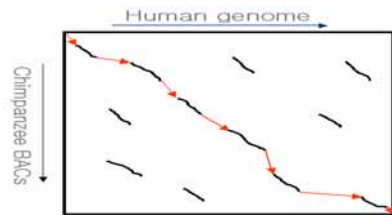
HSP  
BAC  
BAC  
quality 가

BAC  
HSP  
BAC  
HSP  
global alignment map

가  
BAC  
reference  
gap  
가  
BAC  
HSP

HSP  
gap  
가  
가  
가  
BAC clone HSP

orthologous 30Mb  
가



3. HSP  
global alignment map

global alignment map  
orthologous

DotPlot PipMaker MUMmer [8][9], GLASS [10], VISTA [11]  
 AVID [12] global alignment  
 genbank  
 RepeatMasker reference  
 (LINE, HERV, ALU, simple BAC  
 repeat ) 30MB global alignment  
 orthologous DotPlot  
 PipMaker 가  
 1.45MB 23 가 global  
 orthologous contig alignment

anchor

**Ka Ks**

BAC clone contig  
 global alignment  
 ClustalW 4 (1. , 2.Megablast  
 (blastz) 가 HSP  
 / , , 3.global alignment map  
 BAC anchor , 4. BAC  
 anchor  
 . Nei-Gojobori Method  $P_S =$   
 Sd/S  $P_N = N_d/N$  orthologous )  
 Ka (Non-Synonymous) Ks 21  
 (Synonymous) . contig 5 22 BAC  
 297 30Mb

## Results

### Global alignment map

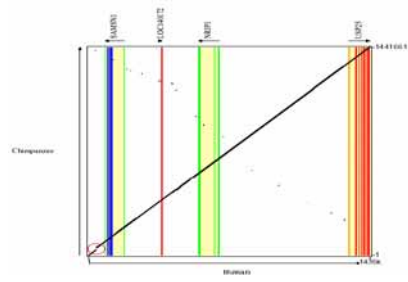
### Large-scale rearrangement

global  
 alignment Needleman-  
 Wunsch(1970) 가 30MB orthologous genome 가  
 가 .  
 가 .

(5Myr~6Myr)

가

가 . 가  
가



B.

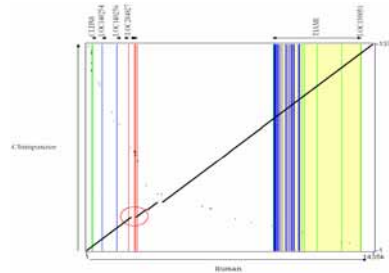
(rearrangement)

200kb

( 4. A, B)

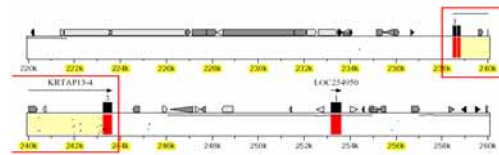
( 4. C)

( 4. D)



KRTAP13-4

5kb



C.

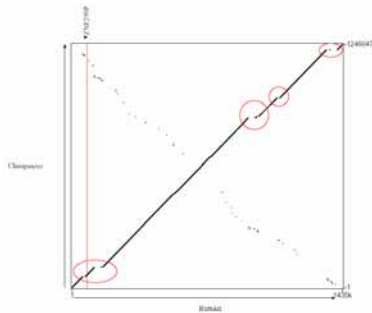
keratin

KRTAP13-4 (5kb)

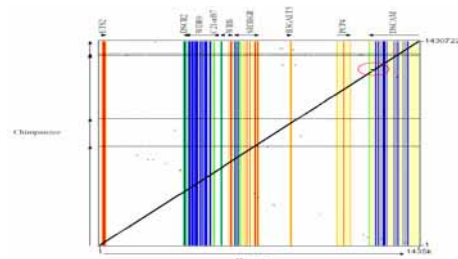
가

(24kb) 가

21



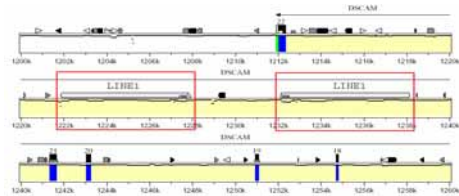
A.



D.

LINE1 element가

palindrome



a) Chimpanzee

b) Human

[13].

|        |        |       |      |       |        |        |       |     |       |
|--------|--------|-------|------|-------|--------|--------|-------|-----|-------|
| 798932 | 797133 | Right | 302  | AluSp | 790775 | 791075 | Right | 301 | AluSp |
| 797687 | 797916 | Left  | 210  | AluSp | 791156 | 791046 | Left  | 201 | AluSp |
| 798026 | 798197 | Left  | 169  | IRa   | 791958 | 792126 | Left  | 168 | IRa   |
| 798204 | 798583 | Right | 300  | AluSp | 792213 | 792111 | Right | 299 | AluSp |
| 798820 | 798873 | Left  | 54   | LINE2 | 792726 | 792801 | Left  | 66  | LINE2 |
| 801376 | 801775 | Right | 400  | LTR   | 795313 | 795713 | Right | 400 | LTR   |
| 801776 | 808906 | Right | 5022 | HERV  |        |        |       |     |       |
| 806681 | 807185 | Right | 795  | LTR   |        |        |       |     |       |
| 807353 | 807847 | Left  | 245  | IRP   | 795820 | 796064 | Left  | 245 | IRP   |
| 808033 | 808178 | Left  | 126  | LINE2 | 796541 | 797061 | Left  | 521 | LINE2 |
| 810811 | 811154 | Right | 344  | MLT1A | 799308 | 799649 | Right | 344 | MLT1A |
| 811483 | 811821 | Right | 370  | MLT1K | 799958 | 800317 | Right | 371 | MLT1K |
| 811954 | 811922 | Right | 89   | LINE2 | 800350 | 800418 | Right | 89  | LINE2 |
| 812396 | 812460 | Right | 173  | IRa   | 800794 | 800962 | Right | 189 | IRa   |
| 813029 | 813167 | Right | 129  | IRP   | 801531 | 801659 | Right | 129 | IRP   |
| 812962 | 814073 | Left  | 112  | LINE2 | 802447 | 802508 | Left  | 112 | LINE2 |
| 814307 | 814508 | Right | 302  | AluT  | 802692 | 802993 | Right | 302 | AluT  |

*KRTAP13-4*

21 orthologous MHC-1 class Positive

E. LTR  
Endogenous Retrovirus)

4.  
rearrangement  
plot

21 orthologous MHC-1 class  
22  
orthologous  
54 ka, ks

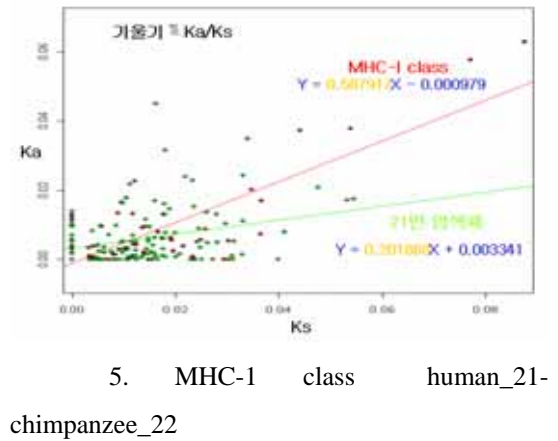
UTR(untranslation region) ,  
UTR, : Intergenic  
C, D

(human\_chr21-chimpanzee\_chr22, MHC-1)  
가 , 가

Pip(percent identity plot)  
Retrotransposon /

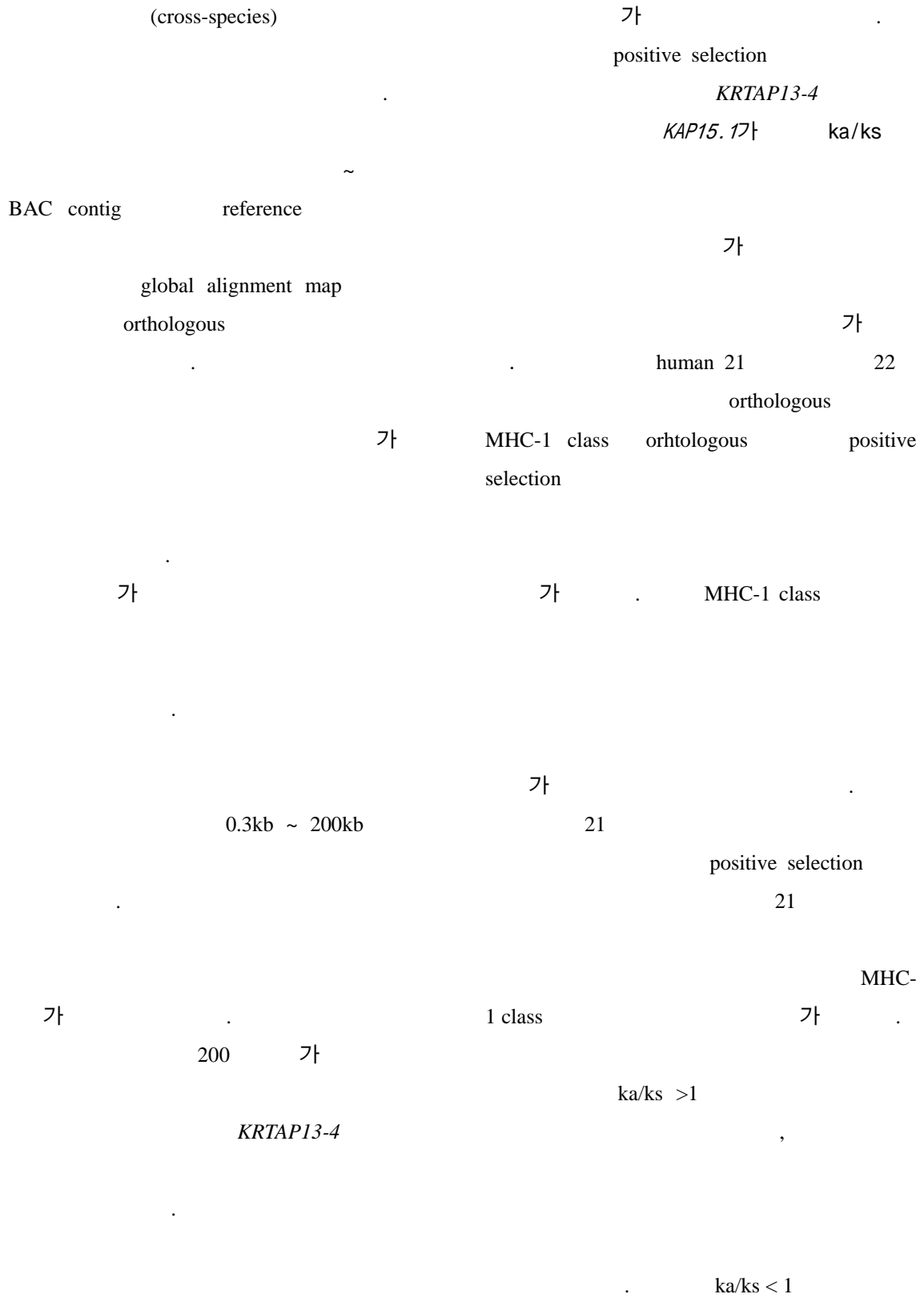
MHC-1 class  
(ka/ks)가 0.587917  
human\_chr21-chimpanzee\_chr22  
MHC-1 class 1/3  
0.201868 (ka/ks) ( 5).

21 orthologous Positive  
22  
orthologous  
147  
/  
28 119  
 $K_a/K_s > 1$  147  
12 LOC254950  
 $K_a/K_s = 1.66$   
positive *KAP15.1*



5. MHC-1 class human\_21-chimpanzee\_22

## Discussion and Further study





negative selection  
 (cell cycle) (energy  
 metabolic) .  
 global alignment map  
 orthologous  
 가 .

## 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  $\beta$ -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. Cen, 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 Jürgen 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

---