



Short tandem repeats in the human, cow, mouse, chicken, and lizard genomes are concentrated in the terminal regions of chromosomes



Kazuharu Misawa*

Advanced Center for Computing and Communication, Riken, Hirosawa 2-1, Wako, Saitama 351-0198, Japan

ARTICLE INFO

Article history:

Received 31 December 2015
Received in revised form 17 January 2016
Accepted 17 January 2016
Available online 23 February 2016

Keywords:

Short tandem repeat
Evolution
Chromosome
Mammalian genome
Next generation sequencer

ABSTRACT

Repetitive sequences in a genome cause mapping errors, especially in the case of short reads, because of the presence of similar or identical sequences. Distribution of repetitive sequences in a genome must be studied to distinguish between mappable and unmappable regions. Previous studies showed that short tandem repeats (STRs) are clustered in the terminal regions of chromosomes in the human genome. It is an open question whether formation of STRs in the terminal regions of chromosomes occurs only in humans. The present study investigated the distribution of STRs in the human, cow, mouse, chicken, and lizard genomes. In this study, it was shown that STRs were concentrated in the terminal regions of chromosomes not only in the human genome, but also in the mouse, cow, chicken, and lizard genomes. The results suggested the mechanism through which STRs are shared by amniotes in which mammals, birds and lizards are included. Thus, we must be careful with the genomic sequences at chromosome ends of amniotes by using the next generation sequencers.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

The human genome contains various repetitive sequences that account for approximately 50% of the complete genome (Lander et al., 2001). These sequences contain perfect or slightly imperfect copies of DNA motifs of variable lengths (Charlesworth et al., 1994; Wicker et al., 2007). Repetitive elements in a genome sometimes cause mapping errors, especially in the case of short reads, during next-generation sequencing because of the presence of similar or identical sequences (Trapnell & Salzberg, 2009; Treangen & Salzberg, 2012; Misawa, 2013; Li & Freudenberg, 2014; Li et al., 2014). A quantitative understanding of these repeat sequences in a genome is necessary to distinguish between mappable and unmappable regions.

Previous studies showed that short tandem repeats (STRs) are mainly distributed in the terminal regions of human chromosomes (Cross et al., 1990; Amarger et al., 1998; Giraudeau et al., 1999; Vergnaud & Denoeud, 2000; Näslund et al., 2005; Ames et al., 2008). On the contrary, STRs seem to be distributed almost uniformly in the mouse (Bois et al., 1998) and rat genomes (Amarger et al., 1998). Two hypotheses explain the difference in STR distribution between human and rodent genomes. The first hypothesis is “concentration early” hypothesis, which suggests that STRs were concentrated in the terminal regions of chromosomes in a common ancestor of humans and rodents and that the STRs in the terminal regions of chromosomes were lost in

the rodent lineage. The second hypothesis is the “concentration late” hypothesis, which suggests that concentration of STRs in the terminal regions of chromosomes occurred only in the human lineage. To test the 2 hypotheses, it is necessary to examine the state of terminal STRs in an ancestor of mammals.

In this study the distribution of STRs in the chromosomes of humans, mice, cows, chickens, and lizards was investigated. It must be noted that the centromeric DNA sequences of most eukaryotes contain large arrays of repetitive DNA (Mehta et al., 2010). STRs on chromosomes must be differentiated from centromeric repeats. Firstly, centromeres of chromosomes must be carefully checked in this study, since centromeric positions depend on chromosomes. Next, repeat units that were longer than centromeric repeats were excluded in this study. Telomeric repeats were also excluded in the analyses. By comparing the distributions of STRs in the human, cow, mouse, chicken and lizard genomes, the ancestral state of terminal repeats in the mammalian genome was discussed.

2. Materials and methods

2.1. Sequences of the human, cow, mouse, chicken, and lizard genomes

In this study, the sequences of the NCBI human genome (Build 37) were investigated. The chimpanzee genome assembly started with the alignment of chimpanzee reads to the human build (The Chimpanzee Sequencing and Analysis Consortium, 2005). Such sequences might be “humanized” so that their structures might be similar to the human genome than they really are. For this reason, genomes of non-human primates were not investigated in the present study.

* Tohoku Medical Megabank Organization, Tohoku University, 2-1 Seiryō-machi, Aoba-ku, Sendai, Miyagi 980-8573, Japan. Tel.: +81 22 273 6051, +81 48 462 1111x3876; fax: +81 22 274 5938, +81 48 462 4634.

E-mail address: misawa@megabank.tohoku.ac.jp.

Table 1
The proportion of human minisatellites in terminal regions.

Chromosomes		Whole chromosome		p-Terminal region ^a		q-Terminal region ^a	
Number	Size	Sum of length	Odds ratio ^b	Sum of length	Odds ratio ^b	Sum of length	Odds ratio ^b
Human							
chr1	226,934,550	2,189,788	0.95*	30,397	2.98*	11,981	1.17*
chr2	238,357,373	2,391,392	0.98*	35,765	3.51*	31,249	3.06*
chr3	194,797,140	1,825,646	0.92*	9412	0.92*	32,102	3.15*
chr4	188,683,360	1,906,581	0.99*	27,554	2.70*	10,174	1.00
chr5	177,745,260	1,724,484	0.95*	58,343	5.72*	16,574	1.62*
chr6	200,830,336	1,975,336	0.96*	8736	0.86*	10,220	1.00
chr7	156,036,559	1,660,995	1.04*	35,340	3.46*	8562	0.84*
chr8	143,015,011	1,444,158	0.99*	33,711	3.31*	8754	0.86*
chr9	121,526,573	1,221,053	0.99*	7018	0.69*	18,944	1.86*
chr10	131,714,747	1,345,856	1.00	28,453	2.79*	54,676	5.36*
chr11	131,369,619	1,274,493	0.95*	31,563	3.09*	26,702	2.62*
chr12	130,531,895	1,276,500	0.96*	9926	0.97	37,187	3.65*
chr13	95,589,878	1,056,592	1.08*	15,689	1.54*	54,938	5.39*
chr14	88,289,540	858,851	0.95*	24,389	2.39*	17,383	1.70*
chr15	82,221,392	726,475	0.87*	13,731	1.35*	11,098	1.09*
chr16	78,934,753	891,847	1.11*	37,072	3.63*	25,121	2.46*
chr17	79,910,435	825,724	1.01*	54,061	5.30*	7778	0.76*
chr18	74,761,510	767,649	1.01*	9085	0.89*	41,741	4.09*
chr19	56,110,841	884,158	1.54*	37,279	3.65*	18,440	1.81*
chr20	59,605,520	596,085	0.98*	8246	0.81*	41,390	4.06*
chr21	35,186,384	427,080	1.19*	24,525	2.40*	28,593	2.80*
chr22	34,894,566	417,971	1.17*	26,469	2.60*	27,155	2.66*
chrX	151,350,560	1,548,112	1.00	40,966	4.02*	16,365	1.60*
chrY	25,753,566	384,935	1.47*	42,284	4.15*	9287	0.91*
Total	2,904,151,368	29,621,761	1.00	650,014	2.66*	566,414	2.31*
Mouse							
chr1	198,927,608	2,152,035	0.94*	8394	0.73*	9343	0.81*
chr2	178,793,830	2,027,669	0.98*	13,009	1.12*	9906	0.86*
chr3	160,365,011	1,773,578	0.96*	42,283	3.66*	9727	0.84*
chr4	163,852,048	1,937,372	1.02*	51,106	4.42*	14,517	1.26*
chr5	150,395,578	1,816,298	1.04*	14,961	1.29*	10,770	0.93*
chr6	155,394,424	1,712,082	0.95*	20,386	1.76*	13,782	1.19*
chr7	143,002,394	1,640,823	0.99*	12,670	1.10*	13,444	1.16*
chr8	126,561,704	1,468,776	1.00	11,782	1.02	9408	0.81*
chr9	121,200,903	1,446,949	1.03*	44,812	3.87*	13,346	1.15*
chr10	127,154,958	1,477,836	1.00*	17,091	1.48*	12,072	1.04*
chr11	125,610,892	1,513,043	1.04*	17,853	1.54*	9769	0.84*
chr12	119,064,322	1,373,578	1.00	16,163	1.40*	12,741	1.10*
chr13	117,640,084	1,327,852	0.98*	12,864	1.11*	11,341	0.98
chr14	123,524,422	1,357,421	0.95*	6314	0.55*	9218	0.80*
chr15	100,884,676	1,170,894	1.00	9557	0.83*	11,621	1.00
chr16	98,776,437	1,152,784	1.01*	36,436	3.15*	13,946	1.21*
chr17	106,827,956	1,454,958	1.18*	53,114	4.59*	9621	0.83*
chr18	88,172,611	1,018,206	1.00	20,199	1.75*	16,537	1.43*
chr19	58,524,654	706,939	1.04*	15,743	1.36*	12,686	1.10*
chrX	166,886,737	1,939,805	1.01*	12,337	1.07*	11,712	1.01
chrY	92,210,412	1,032,377	0.97*	11,327	0.98	10,378	0.90*
Total	2,723,771,661	31,501,275	1.00	448,401	1.85*	245,885	1.01*
Cow							
chr1	161,428,367	1,285,251	1.01*	6924	0.88*	13,773	1.75*
chr2	141,965,563	1,109,070	1.00*	18,997	2.42*	11,288	1.44*
chr3	126,844,711	986,330	0.99*	16,601	2.12*	14,039	1.79*
chr4	123,809,850	963,355	0.99*	18,474	2.35*	13,997	1.78*
chr5	125,249,322	1,009,561	1.03*	15,492	1.97*	12,682	1.62*
chr6	122,519,025	953,964	0.99*	9765	1.24*	9138	1.16*
chr7	113,029,157	886,037	1.00	8675	1.11*	8540	1.09*
chr8	116,846,264	868,995	0.95*	12773	1.63*	13,184	1.68*
chr9	108,503,706	864,202	1.01*	8154	1.04	15,672	2.00*
chr10	105,982,576	789,082	0.95*	7955	1.01	9894	1.26*
chr11	109,987,751	838,642	0.97*	9844	1.25*	18,584	2.37*
chr12	85,119,472	680,602	1.02*	8318	1.06*	23,650	3.01*
chr13	84,213,851	632,048	0.96*	8163	1.04	9592	1.22*
chr14	81,216,349	651,816	1.02*	11,305	1.44*	10,472	1.33*
chr15	84,472,747	654,573	0.99*	16,919	2.16*	15,079	1.92*
chr16	77,710,258	607,581	1.00	10,006	1.27*	9106	1.16*
chr17	76,280,064	601,604	1.00	9669	1.23*	12,845	1.64*
chr18	65,811,054	550,950	1.07*	7713	0.98	8724	1.11*
chr19	64,845,320	534,764	1.05*	8046	1.03	12,726	1.62*
chr20	75,686,341	597,155	1.01	7396	0.94*	26,097	3.33*
chr21	69,078,422	556,979	1.03*	10,512	1.34*	9866	1.26*
chr22	61,598,339	443,835	0.92*	7706	0.98	11,706	1.49*
chr23	52,334,015	419,924	1.02*	8738	1.11*	15,969	2.03*

(continued on next page)

Download English Version:

<https://daneshyari.com/en/article/2820515>

Download Persian Version:

<https://daneshyari.com/article/2820515>

[Daneshyari.com](https://daneshyari.com)