The Proceedings of the International Conference on Creationism

Volume 7 Article 33

2013

New Research Evaluating Similarities Between Human and Chimpanzee DNA

Jeffrey Tomkins
Institute for Creation Research

Proceedings of the Seventh International Conference on Creationism.
Pittsburgh, PA: Creation Science Fellowship

NEW RESEARCH EVALUATING SIMILARITIES BETWEEN HUMAN AND CHIMPANZEE DNA

Jeffery Tomkins, Institute for Creation Research, 1806 Royal Lane, Dallas, TX 75229

KEYWORDS: Human, Chimpanzee, Genome, DNA Similarity, Primate Evolution

ABSTRACT

A preliminary study was performed by Tomkins comparing 40,000 chimpanzee genomic sequences against the human genome which indicated that reported levels of human-chimp DNA similarity were significantly lower than commonly reported. The present, follow-up study was then completed in which chimp chromosomes were sliced into new individual query files of varying string lengths and queried against their human chromosome homolog. This allowed for comparisons to be optimized irrespective of the linear order of genes and sequence features. The definition of similarity was the amount (percent) of optimally aligned chimp DNA. For the chimp autosomes, the amount of optimally aligned DNA sequence provided similarities between 66 and 76 percent, depending on the chromosome. Only 69 percent of the chimpanzee X chromosome was similar to human and only 43 percent of the Y chromosome. Genome-wide, only 70% of the chimpanzee DNA was similar to human under the most optimal alignment conditions. While, chimpanzees and humans share many localized protein-coding regions of high similarity, the overall extreme discontinuity between the two genomes defies evolutionary time-scales and dogmatic presuppositions about a common ancestor.

INTRODUCTION

Recent bioinformatics research has been completed in human-chimp DNA similarity in the following areas: 1) the chromosome 2 fusion model and 2) genome-wide DNA similarity. The chromosome two fusion model postulates that two small chimp 'like' chromosomes (2A and 2B) fused to form human chromosome 2 – explaining the discrepancy in chromosome numbers between human and chimp. In previous research, Tomkins and Bergman (2011a) showed how the so-called "fusion site" and "cryptic centromere site" on human chromosome 2 were highly questionable. Additional bioinformatics research on the prevalence of forward and reverse complement telomere repeats $(TTAGGG/CCCTAA)_N$ across the human genome has been completed. This study shows that interstitial (internal) telomere sequences (ITS) are suppressed as monomers, but enriched as repeats across the human genome (compared to random 6-base sequences), excluding chromosome endpoints (telomeres and sub-telomeres). In fact, ITS repeats much longer than those observed in the so-called fusion site on chromosome 2 are somewhat ubiquitous across the genome. It is postulated that ITS motifs may be a design feature associated with some undiscovered aspect of genome structure and function.

INTERSTITIAL TELOMERE SEQUENCE (ITS) RESEARCH

One of the most-cited DNA arguments for human evolution is the hypothetical head-to-head fusion of two small ape-like chromosomes (labeled 2A and 2B in chimp) that is thought to have resulted in human chromosome 2. A majority of the research that undergirds this has used indirect methods of analysis such as DNA probe hybridization, chromosomal banding (staining with microscopy), and limited DNA sequencing prior to the advent of high-throughput DNA sequencing technologies (Yunis & Prakash, 1982; Ijdo, *et al*., 1991).

The standard mammalian 6-base telomere sequence $(TTAGGG)_N$ is the essential DNA motif repeated in telomeres – the end-caps of nuclear chromosomes (Tomkins & Bergman 2011a). In 2002, a study was reported in which the complete re-sequencing of the fusion-related region of human chromosome two was undertaken (Fan *et al*., 2002). Researchers found that the fusion site was ambiguous in regards to its representation of an evolutionary recent end-to-end chromosomal fusion event based on estimated sequence degeneration over deep time. The authors concluded that "the head-to-head repeat arrays at the RP11-395L14 fusion site have significantly degenerated from the near perfect $(TTAGGG)_n$ arrays found in telomeres". They further stated, "Why are the arrays at the fusion site so degenerate if the fusion occurred within the telomeric repeat arrays less than ~6 Mya?".

Tomkins and Bergman recently (2011a) reanalyzed the fusion site DNA sequence and found that the main DNA signature comprises a region of only 798 bases. Outside this area, the presence of telomeric repeats completely breaks down and they are only found hundreds to thousands of bases apart. Human telomeres are 5,000 to 15,000 bases in length (Tomkins & Bergman 2011b) and an end-to-end fusion event should leave a signature thousands of bases in size.

Even within the small 798-base core area, the evidence for a fusion event is highly degenerate, as discussed above. An important attribute associated with this observation is the fact that the telomere motifs in the fusion site are largely monomeric and not found as continuous repeats. Of the 10 intact TTAGGG motifs on the left side of the 798-base fusion site, only two tandem occurrences can be found. Of the 43 intact CCCTAA motifs on the right side of the 798-base fusion site, twelve tandem motifs are found. Furthermore, the motifs are generally erratic and not in frame. For more details see Tomkins and Bergman (2011a).

An alternative explanation for the telomere-like features present at the putative fusion site is that they represent some type of internal genomic motif. When the 798-base fusion site was searched against the human genome using the BLASTN algorithm 85 matches occurred. Many of these matches occurred at internal sites (Tomkins & Bergman 2011a).

While the role that the telomere motif plays in chromosome end-cap maintenance and function is increasingly better understood, it's presence within internal locations of chromosomes is still somewhat of a mystery. These internally located motifs are called 'interstitial telomere sequence' (ITS) and are acknowledged to be common in animals, including humans (Lin & Yan, 2008). A majority of research characterizing ITS has utilized fluorescent in-situ hybridization (FISH) techniques, and has only revealed genomic information regarding the larger ITS regions of the human genome. However, a combined informatic and wet lab study in 2001 revealed 50 short

ITS sites with nearly pristine TTAGGG repeats across the human genome from 24 to 130 bases in length (Azzalin, *et al*., 2001). Therefore, we performed a second study on a more advanced version of the human genome to bioinformatically characterize ITS sequences genome-wide.

Using scanning software written by author Tomkins expressly for the purpose of identifying pristine telomere sequence along with motif repeats, every human chromosome was scanned. The telomere and sub-telomere regions of each chromosome were removed prior to scanning using a text editor so that only the internal regions were analyzed. In addition, a variety of random 6-base sequences were used as controls for comparison.

Resulting data indicated that, compared to random 6-base sequences, telomere motifs were slightly suppressed in their occurrence as monomers, but increased as repeats (ITS). Furthermore, the presence of ITS sites varied markedly depending on the chromosome. Genome-wide, ITS motifs occurring in pristine telomere repeats of 2 to 11 or more tandem motifs are found by the hundreds across the genome. In addition, a putative end-to-end fusion motif was found on chromosome 9 in addition to the well-documented site on chromosome 2. Clearly, ITS are considerably more common in the human genome that originally reported by Azalin, *et al*. in 2001.

Since the two most-cited claims for human-chimp ancestry are clearly false, we felt a detailed examination of the third major claim – a high percent similarity between human and chimpanzee genomes – was warranted.

GENOME-WIDE DNA SIMILARITY

A common claim is that the DNA of chimpanzees (*Pan troglodytes*) and humans (*Homo sapiens*) are nearly identical. However, this over-simplified dogma is becoming much less popular among primate evolutionists because current DNA research is showing much higher levels of discontinuity between human and chimp. This revision in the standard paradigm was characterized by primate evolutionist Todd Preuss when he made the following statement in a 2012 PNAS paper.

> *It is now clear that the genetic differences between humans and chimpanzees are far more extensive than previously thought; their genomes are not 98% or 99% identical.*

One of the major problems with past research in comparative DNA analysis between chimps and humans was recently reviewed in several reports (Bergman & Tomkins, 2012; Tomkins & Bergman, 2012). They found that there is a great deal of preferential and selective treatment of the data being analyzed. In general, only the most favorable data such as gene-rich sequences that exist in both species (homologs) is used from a larger data pool. Non-alignable regions and large gaps in DNA sequence alignments are typically omitted, thus increasing levels of similarity.

The key publication regarding the chimp genome was the 2005 *Nature* paper from the International Chimpanzee Genome Sequencing Consortium, which presented the comparative data with human in a highly selective and obfuscated format. The non-similar data from their

alignments was largely absent. Instead, the paper focused on hypothetical analyses for divergence rates and selective forces in highly similar homologous regions rather than reporting overall levels of discontinuity between chimp and human DNA.

Tomkins and Bergman (2012) included information from the human genome project along with data reported in the 2005 chimpanzee paper and derived an overall genome DNA similarity estimate of 80.6%, which they proposed as a very conservative figure. Geneticist Richard Buggs took a similar, but more exacting approach and came up with a figure of about 70% similarity genome wide (Buggs, 2008).

Since 2005, a comprehensive analysis between chimp and human DNA has yet to be done. However, an intriguing report has been privately published (e.g., Anonymous, 2012). This effort employed an algorithm that involved the random selection of 10,000 30-base sequences from each chimp chromosome and then determined their identity based on a query against their human chromosome counterpart. Excluding the Y chromosome, this study came up with an average 63% DNA identity (similarity) genome-wide. Although like the Buggs (2008) conclusions mentioned above, this information was not subject to peer-review, but it is included here, after validation, because the conclusions align with those of this paper despite the differences in technique.

In 2011, Tomkins queried 40,000 chimpanzee genomic DNA sequences against four versions of the human genome using a wide variety of BLASTN algorithm parameters. For just the aligned regions, an 86 to 89% DNA similarity was observed, depending on the algorithm parameter combinations. However, less than 20% of the total chimp DNA sequence aligned under the most optimum conditions. Given that the average length of the chimp sequences were 740 bases, these data indicate that regions of human-chimp DNA similarity must break down at stretches of less than 740 bases on average. The question then arises as to what query sequence lengths would be more optimal for comparing the chimp genome against human.

METHODS

To address that question, a new study was performed using a range of smaller sequence slices for queries to optimize the alignment of chimp DNA on human. This approach would also allow for matching of DNA sequence irrespective of gene/feature linear order. Using this approach, a comprehensive chromosome-by-chromosome genome comparison between chimpanzee and human was undertaken using optimized DNA sequence slices. The most recent version of the chimpanzee chromosome assembly (Pan_troglodytes-2.1.4), aligned and anchored to the genome human, was utilized. On an individual chromosome basis, new query sub-files were created with sequences in 50-base increments for all of the chimp chromosomes. The new query files also had all the ‘N’s removed from the chimp sequence that would have produced false alignments to the large spans of ‘N’s in the human genome assembly (GRCh37). For a complete description of the methods used, see Tomkins (2013).

RESULTS AND DISCUSSION

Figure 1 visually depicts each set of experiments per chromosome. The top percentages for

chimp DNA aligned to the human genome are listed in Table 1. For a complete discussion and comparison of these results to the known gene density and level of sequence completion for human chromosomes, see Tomkins (2013).

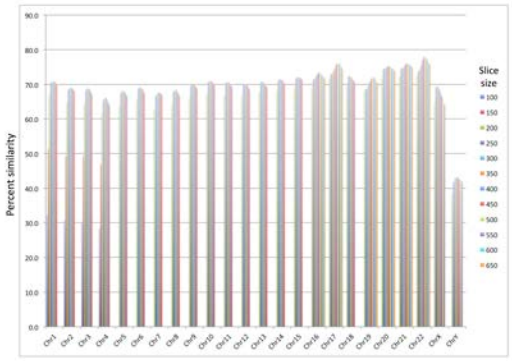

Figure 1. Percent of chimp sequence aligned to the respective human chromosomes using optimized sequence slices.

Table 1. Individual chromosome similarities for chimpanzee compared to human using optimized sequence slices and the BLASTN algorithm.

Chromosomes compared	Optimized slice size producing top similarity (number bases)	Percent optimally aligned chimp sequence
1	350	70.9
2A, 2B vs 2 (human)	300	69.0
3	300	68.9
4	300	66.1
5	300	68.2
6	300	69.2
7	350	67.3
8	300	68.4
9	350	70.1
10	300	71.0
11	300	70.8
12	300	70.1
13	300	70.8
14	300	71.6

15	350	72.0
16	450	73.3
17	500	76.1
18	250	72.5
19	500	72.0
20	400	75.2
21	500	76.2
22	450	77.9
X	300	69.4
Y	400	43.2

For the whole chimpanzee genome, only 70% of the DNA on average was similar to human under the most optimal sequence-slice conditions. This empirically based conclusion is well within the range of preliminary results suggested by other researchers (Buggs, 2008; Anonymous, 2012).

For the chimpanzee autosomes, the optimally aligned DNA sequence provided similarities between 66 and 76 percent, depending on the chromosome. In general, higher DNA similarities were obtained in the smaller and more gene-dense human chromosomes. However, there were several exceptions (chimp chromosomes 19 and 21) that defied this trend. These results indicate that not all gene-rich areas of the chimp and human genomes are highly similar.

Only 69 percent of the chimpanzee X chromosome was similar to human and 43 percent of the Y chromosome. The extreme Y-chromosome dissimilarity is supported by other work in which the MSY regions of the chimp and human Y-chromosomes were compared and found to be very dissimilar in not only DNA sequence, but also gene content (Hughes, *et al*., 2010).

In summary, the third major claim for human-chimpanzee common ancestry is false. The percent similarities among the two genomes is far less than that which has been typically reported. While it is true that certain small portions of the two genomes are highly similar, the bulk of the evidence shows the opposite.

REFERENCES

Anonymous (2012), An automatic comparison of the human and chimpanzee genomes. progettocosmo.altervista.org/index.php?option=content&task=view&id=130.

Azzalin, C.M., *et al*. (2001), Human intrachromosomal telomeric-like repeats: sequence organization and mechanisms of origin. *Chromosoma* 110:75-82.

Bergman, J. and Tomkins, J. (2012), Is the human genome nearly identical to chimpanzee? A reassessment of the literature. *Journal of Creation* 26:54-60.

Buggs, R. (2008), Chimpanzee? *Reformatorisch Dagblad;* www.refdag.nl/.chimpanzee_1_282611.

Fan, Y., *et al*. (2002), Genomic structure and evolution of the ancestral chromosome fusion site in 2q13-2q14.1 and paralogous regions on other human chromosomes. *Genome Res.* 12:1651–1662.

Hughes, J. F., *et al*. (2010), Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. *Nature* 463:536–539.

Ijdo, J.W., *et al*. (1991), Origin of human chromosome 2: an ancestral telomere-telomere fusion. *Proc. Natl. Acad. Sci.* 88:9051–9055.

Lin, W.L. and Yan, J. (2008), Endings in the middle: Current knowledge of interstitial telomeric sequences. *Mutation Research* 658:95-110.

Tomkins, J. (2011), Genome-Wide DNA Alignment Similarity (Identity) for 40,000 Chimpanzee DNA Sequences Queried against the Human Genome is 86 - 89%. *Answers Research Journal* 4:233-241.

Tomkins, J. and Bergman, J. (2011a). The chromosome 2 fusion model of human evolution - part 2: re-analysis of the genomic data. *Journal of Creation* 25:111-127.

Tomkins, J. and Bergman, J. (2011b), Telomeres: implications for aging and evidence for intelligent design. *Journal of Creation* 25(2):86–97.

Tomkins, J. and Bergman, J. (2012), Genomic monkey business—estimates of nearly identical human–chimp DNA similarity re-evaluated using omitted data. *Journal of Creation* 26:94-100.

Tomkins, J.P. 2013. Comprehensive Analysis of Chimpanzee and Human Chromosomes Reveals Average DNA Similarity of 70%. *Answers Research Journal* 6:63-69.

Yunis, J. J. and Prakash, O. (1982), The origin of man: a chromosomal pictorial legacy, *Science* **215**:1525–1530.