

Identification and Phylogenetic Analysis of Schizophrenia Associated Retrovirus (SZRV) Element in the Human GenBank Database

Jun-Seop Kim **Sang-Bae Kim** **InSong Koh**
bkjs1206@lycos.co.kr ksb003@korea.com insong@nih.go.kr

Section of Bioinformatics, Central Genome Center, National Institute of Health,
Seoul, 122-701, Korea

Keywords: SZRV, human genome, multiple alignment, phylogenetic analysis

1 Introduction

A significant proportion of the human genome consists of interspersed repetitive DNA sequences. One type of repetitive DNA elements is human endogenous retroviruses (HERVs), representing about 1% of the human genome. Their structures closely resemble those of retroviruses, and include internal sequences with homology to *gag*, *pol*, and *env* open reading frames (ORFs) flanked by long terminal repeats (LTRs). The internal sequences contain several transcriptional regulatory sequences [5]. Human genome mapping research has discovered “sequences” of HERVs (i.e. parts of their RNA), which are present in multiple copies and become part of human genome. HERV elements are generally do not cause any problem in their human host, but they may act as an agent of genome change and a potential source of genetic variation associated with certain diseases. A few HERV elements, such as schizophrenia-associated retrovirus (SZRV; schizophrenia), multiple-sclerosis retrovirus (MSRV; multiple sclerosis), endogenous retrovirus-9 (ERV-9) and HERV-W, have been isolated and associated with patients having specific diseases. SZRV element includes *pol* gene as its part and has 3 subtypes; SZRV-1, -3, and -4 [1, 2, 4].

Schizophrenia is a serious neuropsychiatric disease with uncertain etiology. Involvement of endogenous and/or exogenous retroviral-related sequences in the etiopathogenesis of some cases of schizophrenia has been postulated on the basis of clinical and epidemiological observations. The three SZRV subtypes were reported as candidate etiologic agents [3]. However, their distribution in the human genome has not been clearly identified. The aim of this study is to identify the distribution of the homologous sequences of SZRV subtypes in human genome, and to perform their phylogenetic analysis to find their association with SZRV element using the neighbor-joining method.

2 Methods

We obtained SZRV-1 (GenBank Accession No. AF135487), SZRV-3 (GenBank Accession No. AF188999), and SZRV-4 (GenBank Accession No. AF189000) from the GenBank database at National Center of Biotechnology Information (NCBI). Then, we used BLAST search tools in OMIGA program to identify homologous sequences of the SZRV element in the human genome. Homologous sequences were obtained and edited by performing GeneDoc program. The set of homologous sequences was processed to a treeguide file for phylogenetic analysis by running ClustalX program. Finally, we used TreeExplorer program to represent the phylogenetic tree for homologous sequences.

3 Results and Discussions

We investigated the distribution of homologous sequences of SZRV element in the human genome and identified 20 large sequences (> 2,000bp) and 27 small sequences (< 500bp) against SZRV-1 from the Human GenBank database (Fig. 1). However, no homologous sequences were found against SZRV-3

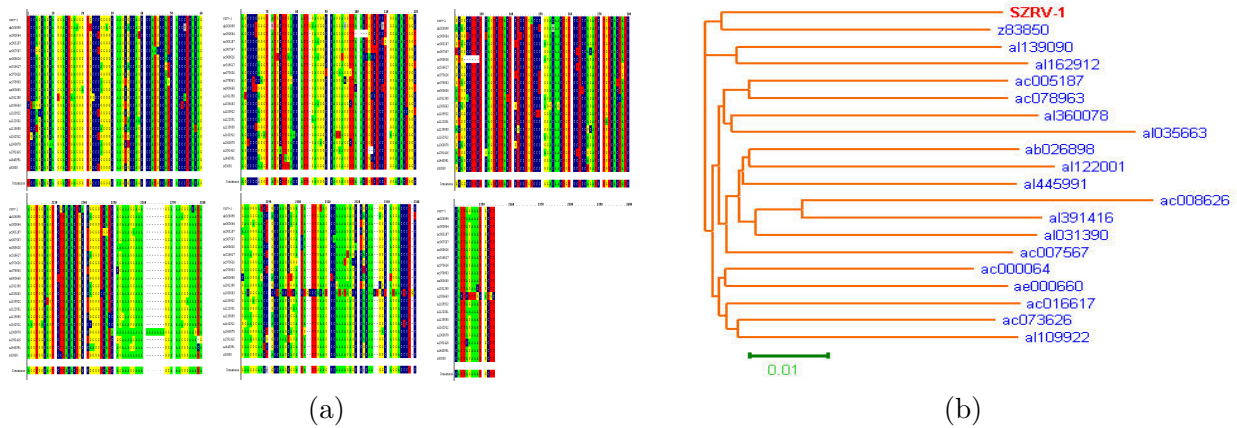


Figure 1: Multiple alignment and phylogenetic analysis of homologous sequences for SZRV-1 in the human genome. (a) Multiple alignment of SZRV-1 homologous sequences was performed by OMIGA program. (b) Phylogenetic tree was constructed for SZRV-1 homologous sequence with homologous sequences by neighbor-joining method in the Genbank database.

and SZRV-4 in the human genome. Twenty homologous sequences were identified at chromosome 1, 3, 4, 5, 6, 7, 12, 13, 19, 20, and X (Table 1). We identified the sequences at several locations (chromosome 4, 5, 7, 13, and 19), where were not previously reported. Although this result was not confirmed by biological experiments, it supports the idea that retroviral sequences may transpose to other sites on human genome. A high degree of homology (88- 92 %) between SZRV-1 and its 20 homologous sequences was found through investigation of multiple alignments in Human GenBank database. It suggests that these sequences may be associated with recovering transcriptional activation of pro-retrovirus as a putative *pol* gene in human.

Table 1: Chromosomal location and identities (%) of DNA sequences between SZRV-1 and its homologous sequences in Human GenBank database.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Location
1. SZRV-1	-	91	91	91	91	89	91	92	92	91	91	90	91	91	91	90	91	91	91	92	undefind
2. ab026898	-	91	91	91	89	90	92	91	91	91	89	90	92	90	90	91	91	92	91	92	3q21.3
3. ac000064	-	91	91	89	91	92	91	92	91	89	91	91	91	90	91	90	91	91	91	91	7q21
4. ac005187	-	91	89	91	92	93	92	91	90	91	91	90	90	91	91	92	92	92	92	92	4q25
5. ac007567	-	90	91	92	92	91	91	89	91	91	91	90	90	91	91	91	92	92	92	92	7q31
6. ac008626	-	88	90	89	89	90	87	88	89	88	88	88	88	88	88	88	88	88	88	88	19
7. ac016617	-	92	91	91	90	89	91	91	90	89	91	91	90	90	90	90	91	91	91	91	5
8. ac073626	-	92	92	92	90	92	92	91	91	91	91	91	91	91	91	91	91	91	91	92	7
9. ac078963	-	92	91	90	91	91	91	91	91	91	91	91	91	92	91	92	92	92	92	92	12q
10. ae000660	-	91	90	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	92	undefind
11. ab031390	-	89	90	92	90	90	90	92	92	91	91	92	92	91	91	91	91	91	91	91	6q
12. ab035663	-	89	89	89	89	89	89	89	89	89	89	89	89	89	89	89	89	89	89	89	20
12. ab035664	-	91	90	90	91	90	91	91	91	91	91	91	91	91	91	91	91	91	91	91	6q
14. al122001	-	91	90	90	91	92	91	91	91	91	91	91	91	91	91	91	91	91	91	91	1p
15. al139090	-	91	90	91	90	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	6
16. al162912	-	90	90	90	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	Xq
17. al360078	-	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	20
18. al391416	-	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	6
19. al445991	-	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	1
20. z83850	-	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	Xq22

Acknowledgements

This work was funded by the Bioinformatics Training Grant of Ministry of Health & Welfare, Korea.

References

- [1] Azoulay-Cayla, A., Is multiple sclerosis a disease of viral origin? *Pathol. Biol.*, 48(1):4–14, Review, 2000.
- [2] Costas, J. and Naveira, H., Evolutionary history of the human endogenous retrovirus family ERV9, *Mol. Biol. Evol.*, 17(2):320–30, 2000.
- [3] Deb-Rinker, P. *et al.*, Molecular characterization of a MSRV-like sequence identified by RDA from monozygotic twin pairs discordant for schizophrenia, *Genomics*, 61(2):133–144, 1999.
- [4] Perron, H. *et al.*, Nucleotide, protein molecular identification of a novel retrovirus repeatedly isolated from patients with multiple sclerosis. The Collaborative Research Group on Multiple Sclerosis, *Proc. Natl. Acad. Sci. USA*, 94(14):7583–7588, 1999.
- [5] Sverdlev, E.D., Perpetually mobile footprints of ancient infections in the human genome, *PBBS Letters*, 428:1–6, 1998.