A single-primer PCR-based retroviral-related DNA polymorphism shared by two distinct human populations

Paromita Deb, Timothy A. Klempan, Richard L. O'Reilly, and Shiva M. Singh

Abstract: Almost 10% of the human genome consists of DNA sequences that share homology with retroviruses. These sequences, which represent a stable component of the human genome (although some may retain the ability to transpose), remain poorly understood. We used degenerate primers specific to the two conserved regions (boxes 4 and 5) of the retroviral *pol* gene, common to all retroviruses, and PCR-amplified related sequences from individuals representing two distinct populations: Caucasians and Dogrib Indians. The large number of sequences that are reproducibly amplified represent numerous sites of retroviral integration in the human genome. In both populations studied, one of the two primers yielded a polymorphic band, present in ~30% of the samples, that has probably been present in the human genome since before the divergence of the two populations ~10 000 years ago. It was established that this polymorphism was due to priming-site differences and not to deletions. Further, this priming site is duplicated at two genomic sites (representing 341- and 343-bp fragments) with at least two alleles each. Such novel polymorphisms should provide useful markers and permit assessment of evolutionary mechanisms associated with retroviral-related genomic evolution.

Key words: Dogrib Indian, evolution, human genome, polymorphism, retrovirus.

Résumé : Presque 10 % du génome humain est composé de séquences d'ADN qui montrent de l'homologie avec des rétrovirus. Ces séquences forment une composante stable du génome humain (bien que quelques séquences aient conservé la capacité de transposition) mais elles demeurent mal connues. Des amorces dégénérées spécifiques de deux régions conservées (les boîtes 4 et 5) du gène *pol* rétroviral, lequel est présent chez tous les rétrovirus, ont été utilisées pour amplifier des séquences apparentées chez des individus représentant deux populations distinctes, les Caucasiens et les Indiens Dogrib. Le grand nombre de séquences qui peuvent être amplifiées de façon reproductible représentent de nombreux sites d'intégration rétrovirale dans le génome humain. L'emploi d'une seule des deux amorces a produit une bande polymorphe, présente chez ~30 % des individus chez les deux populations. Ce polymorphisme est vraisemblablement présent dans le génome humain depuis une date antérieure à la divergence des deux populations, soit environ 10 000 ans. Ce polymorphisme est dû à des différences au niveau des sites d'appariement et non pas à des délétions. De plus, ce site d'appariement est dupliqué en deux sites génomiques (produisant des fragments de 341 et 343 pb) avec deux allèles pour chacun. Ces nouveaux marqueurs devraient s'avérer utiles et permettre d'étudier les mécanismes évolutifs associés à l'évolution génomique liée aux séquences rétrovirales.

Mots clés : Indien Dogrib, évolution, génome humain, polymorphisme, rétrovirus.

[Traduit par la Rédaction]

Introduction

Molecular characterization of the human genome has revealed several families of DNA sequence elements that have homology to retroviruses (Tassabehji et al. 1994). These elements, known as human endogenous retroviral (HERV) sequences, are a stable and inherited part of the human genome and constitute a significant component of it (Lefebvre et al. 1995). The estimated proportion of such sequences in the human genome ranges from 0.1 to 10% in different reports. Additionally, different degrees of retroviral-related sequences have also been identified in eukaryotic genomes, in both animals and plants (Bennetzen 1996). It is also logical to argue that different families of HERV sequences may have become incorporated into the human genome over time, with many likely to be ~40 million years old, as implied by the sharing of such sequences among primate genomes. The mechanism of integration would involve reverse transcription of the viral RNA and recombination with the host DNA following infection. The process relies on the gene and enzyme, reverse transcriptase, which is common to all eukaryotic RNA viruses. It is argued that reverse transcription has had a profound effect on mammalian-genome evolution and that retroviruses play a crucial role in the on-

Corresponding Editor: B. Golding.

Received December 22, 1997. Accepted May 5, 1998.

P. Deb, T.A. Klempan, and S.M. Singh.[1] Molecular Genetics Unit, Department of Zoology, The University of Western Ontario, London, ON N6A 5B7, Canada.
R.L. O'Reilly. Department of Psychiatry, The University of Western Ontario, London, ON N6A 5B7, Canada.

[1]Author to whom all correspondence should be addressed (e-mail: ssingh@julian.uwo.ca).

© 1998 NRC Canada

going evolution, continued differentiation, and shaping of eukaryotic genomes (Leib-Mosch et al. 1992; Krieg et al. 1992). Once incorporated, the sequence could become a stable component of the host genome, with the potential for transposition from one genomic site to another. Retrotransposition of endogenous retroid elements of the human genome as a mutational mechanism remains poorly understood, but has been experimentally demonstrated in a few genetic diseases (O'Reilly and Singh 1996).

The integration process depends not only on the characteristics of retroviruses, but also on the sequence structure of host genomes. It has been suggested that the human genome may contain a set of sites (500–1000) that are used as targets for retroviral integration at high frequency (Shih et al. 1988). There are two features of the synthesis of retroviral DNA that might be especially error prone: removal of internal RNA primers and recombination (Coffin 1984). Both processes might introduce multiple mutations in a restricted region of the viral genome, obliterating them transcriptionally and (or) translationally. This may also account for a large number of deleted sequences among families of endogenous retroviruses and retrotransposons in the human genome.

The impact of retrotransposition on human-genome evolution is difficult to assess. A large number of basic questions with respect to such sequences remain speculative. These include: which (if any) of the human endogenous sequences can still undergo transposition and how often, how many copies of each of the retroviral sequences or related elements now exist, and what the distribution of such sequences is in the human genome. Are there preferred sites in the human genome for integration of retroviral sequences? Is the process of viral genome incorporation into the human genome still ongoing, and at what rate? Could it represent a continuum of evolutionary phenomena, and could these phenomena account for a mutational mechanism capable of causing genetic diseases? The answers to these and other related questions are expected to have implications in diverse areas, such as genome evolution, population differentiation, mutational mechanisms, and retroviral vector-based gene therapy, among others.

The presence of retroviral sequences at large numbers of sites (most still uncharacterized) that have presumably become integrated at different times during the evolution of the human genome offers the possibility of using them as markers. Such markers however, are expected to be useful only after their careful characterization, which forms the focus of this investigation. We used primers specific to the retroviral polymerase gene and identified a presence–absence polymorphism of a ~340-bp band in humans. Further, we assessed the sequence distribution and nature of this polymorphism in two distinct groups of individuals (Caucasian admixtures from southwestern Ontario and Dogrib Indians from the Northwest Territories, Canada). The two groups have been relatively isolated from each other since their divergence (Szathmary 1983), and the maximum European admixture in the Dogribs has been estimated to be 8.7% (Szathmary et al. 1983). The two groups not only differ by virtue of their ancestry (founder effect), but also with respect to population size, possible population subdivisions, and gene infusions. Molecular characterization of the band revealed the nature of this polymorphism, including the possible timing of this integration process.

Materials and methods

Subjects

This study included 82 apparently unrelated individuals of caucasian ancestry from southwestern Ontario, recruited by Dr. R. O'Reilly, St. Thomas, Ontario, and a random sample of 12 Dogrib Indian individuals from the Northwest Territories, originally recruited by Dr. E. Szathmary. The Dogribs represent the largest Indian group in the Northwest Territories and occupy the region east of the Mackenzie River between Great Slave and Great Bear lakes (Szathmary 1983). Over the years, most of the population has relied on hunting and fishing. This lifestyle is still reflected in the sizes of settlement populations, which fluctuate with the season of the year, particularly in the northern part of the Dogrib area. This life style is very different from that of the populations of southwestern Ontario, a region that probably represents the highest population density in Canada, with continued immigration, and a lifestyle based on agriculture and industry. Presumably there has been no significant interbreeding between the two populations since their divergence.

Isolation and amplification of genomic DNA

Frozen samples of genomic DNA of Dogrib individuals were obtained from Dr. R. Ferrell (Pittsburgh) and stored at 4°C in TE (10 mM Tris–HCl, 1 mM EDTA, pH 8.0) during the course of the experiments. The Dogrib blood samples were originally collected by Dr. Szathmary and were used in a number of studies, including Szathmary et al. (1983) and Torroni et al. (1992). White blood cells, separated by spinning down 15 mL of blood samples from each of the Caucasian individuals, were used to isolate genomic DNA at the University of Western Ontario, using a protocol by Jeanpierre (1987).

The assessment of retroviral sequences in genomic DNA relied on a pair of PCR primers specific to the reverse transcriptase (RT) polymerase gene common to all known retroviruses. Specifically, the two primers represented two relatively conserved regions of this gene, box 4 and box 5 (Xiong and Eickbush 1988). Also, the two primers were made degenerate at specific sites (Table 1), to accommodate the degeneracy of codons as well as any variations in DNA sequence among the different families of retroviruses. When used together in a PCR reaction using human DNA, the two primers generate a prominent band of ~133 bp, with a number of additional minor bands. Some of these could have resulted if a single primer annealed in two different orientations to generate individual-specific bands. Such bands would have the potential to represent a new source of individual variation not reported in humans.

In this study, we used the two primers individually in single-primer PCR reactions with the genomic DNA as template. The PCR conditions were optimized with different concentrations of magnesium and different PCR cycles. A common master mixture was made so that each 50-µL PCR reaction mixture would contain 22.5 µL of sterile double-distilled water, 5 µL of 10× PCR buffer (final concentration of 1×; Gibco BRL Life Technologies Inc.), 5 µL of a 200 µM dNTP mixture containing 100 mM of each nucleotide (final concentration of 20 µM; Pharmacia Biotech), 1.5 µL of 50 mM $MgCl_2$ (final concentration of 1.5 mM; Gibco BRL), 4 µL of a single 50 pmol/µL primer, 2 µL of [α-^{35}S]dATP (10 mCi/mL (1 Ci = 37 GBq); Amersham), 0.2 µL of Taq polymerase (5 U/µL; Gibco BRL), and 10 µL of 10 ng/µL genomic DNA template. PCR was performed in a Perkin-Elmer Cetus Thermocycler, using an initial denaturation at 95°C for 5 min; followed by 3 cycles of 94°C for 1 min, 37°C for 2 min, and 72°C for 5 min; 7 cycles of 94°C for 1 min, 37°C for 1 min, and 72°C for 5 min;

© 1998 NRC Canada

Table 1. Primers used in PCR reactions, with their nucleotide sequences and the regions they specify.

Primer region		Primer sequence
Retroviral primers		
I (Box 4)[a]		5′-CCAAGCTTGTNYTNCCNCARGG-3′
II (Box 5)		5′-CCAAGCTTRTCRTCCATRTA-3′
Internal primers		
341 bp sequence 1	Forward	5′-AAAATGTAGRTTGTGGCCCA-3′
	Reverse	5′-CAGCCATCTTTCTAAAATGCA-3′
343 bp sequence 2	Forward	5′-TGAAGCAYAGRCAGGCAGTTA-3′[b]
	Reverse	5′-AATGCATAGATTTCGTCCCC-3′

Note: Y, pyrimidine; R, purine; and N, any base in the degenerate primers. Underlined bases represent *Hin*dIII sites.

[a]In the single-primer PCR reactions, retroviral primer I generated the polymorphism mentioned in the text.

[b]Degenerate at two sites, according to the sequences obtained.

and 35 cycles of 94°C for 30 s, 55°C for 1 min, and 72°C for 1 min. A 5-min extension period at 72°C followed the cycles, and the samples were stored at 4°C until analyzed.

The products of the primary PCR reaction (6 µL) were incubated with 2 µL of gel loading dye at 80°C for 2 min. The radiolabeled PCR products were then run in 6% polyacrylamide (8 M urea) sequencing gels. Electrophoresis conditions were 50°C at a constant power of 80 W for 5 h. After electrophoresis, the gel was blotted on 3M paper (Bio-Rad) and dried at 80°C for 2 h. X-ray film (Kodak BioMax) was placed next to the dried gel in a cassette, exposed for 72 h at room temperature, and processed with GBX developer and fixer.

The band of interest was cut out from the filter paper with a clean razor blade for cloning and sequence characterization. The DNA from the band was eluted by boiling in 100 µL of sterile double-distilled water for 15 min. The supernatant was transferred to a new tube after a 2-min centrifugation at room temperature. The sample was precipitated, using 10 µL of 3 M sodium acetate, 5 µL of glycogen (10 mg/mL), and 450 µL of 100% ethanol at –80°C for 30 min, and centrifuged at 13 200 rpm for 10 min at 4°C. The DNA pellet was rinsed with 70% ethanol, dried, and resuspended in 20 µL of sterile double-distilled water. Resuspended DNA (5 µL) was used for reamplification with the same primer (in the absence of radio-isotope) and run on a 6% mini-acrylamide gel. The DNA from the amplified band was eluted in 20 µL of sterile double-distilled water overnight.

Cloning and sequencing of specific PCR products

Reamplified PCR product (~15 ng) was ligated into pCR 2.1 vector, using the TA cloning kit (Invitrogen), and introduced into One Shot INVαF′ bacterial cells provided in the kit. The plasmid DNA from selected bacterial colonies was isolated by alkaline lysis (Sambrook et al. 1989). The DNA pellet was purified, resuspended in 10 µL of sterile double-distilled water, and sequenced from both directions (T7 and M13 as forward and reverse primers, respectively), using the T7 Sequencing Kit (Pharmacia). Sequencing reactions were separated in 6% polyacrylamide (8 M urea) gels for 5–6 h. The gels were dried, exposed to X-ray film, and processed as before. Sequences were merged and analyzed by the University of Wisconsin Genetics Computer Group (GCG) software (Genetics Computer Group 1994).[2] Further characterization of the sequences included the use of internal primers in PCR and the assessment of the resulting PCR products, using appropriate diagnostic restriction enzyme digestions.

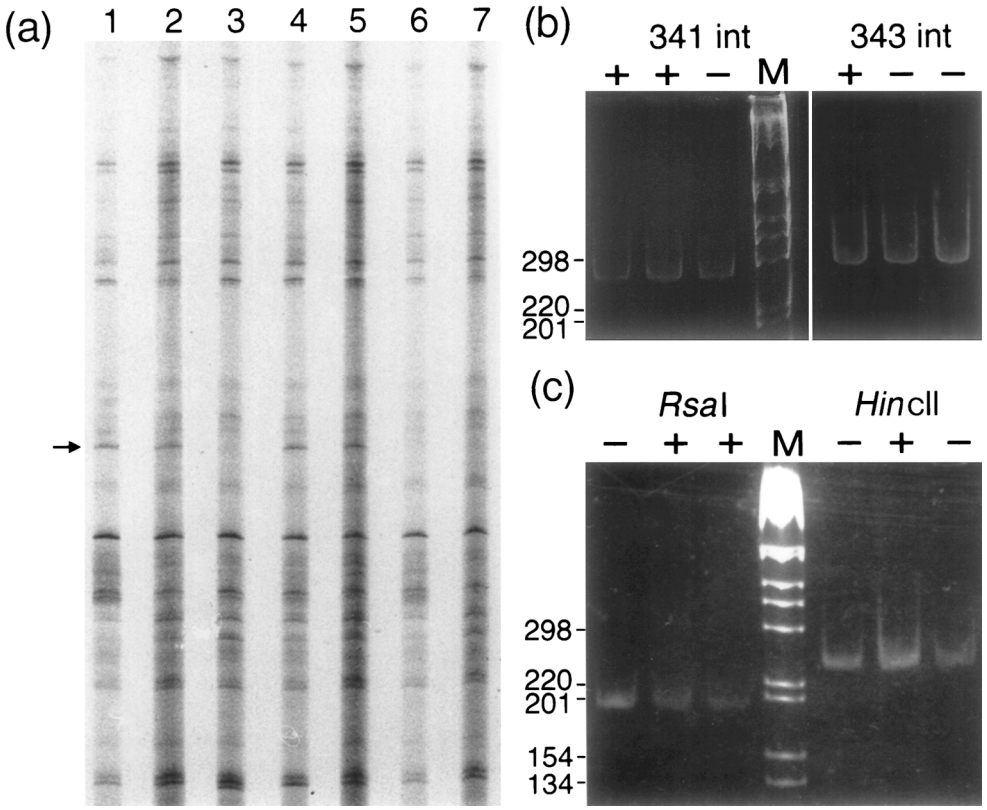
Results and discussion

Retroviral-related sequences were amplifiable from genomic DNA of all humans evaluated, using degenerate primers specific to the reverse transcriptase gene (Table 1). PCR products from the same DNA samples (and with different isolations of DNA samples from the same individual) run multiple times, as well as PCR products from DNA samples run after independent PCR reactions, all generated the same pattern of bands, reiterating the fact that these bands are highly reproducible. Primers specific to box 4 and box 5 of the retroviral reverse transcriptase gene (Xiong and Eickbush 1988) yielded one major band (~133 bp) and a number of other minor bands from every human genomic DNA (not shown). Although the major band represented the appropriate reverse transcriptase (RT) related fragment, some of the minor bands could be attributed to the annealing of a single primer at two sites on the genomic DNA in such a way that they could amplify a fragment of DNA in a suitable PCR reaction. In this study we attempted to assess the bands generated in single-primer PCR reactions. The single primer used in this study is specific to the box 4 region of the retroviral polymerase gene. Assessment of individual specificity of the multiple bands generated by this primer required the use of radioactivity during PCR and separation of the DNA bands on a sequencing gel (Fig. 1*a*). It is apparent from this Fig. 1 that the banding patterns obtained are very similar in the seven individuals shown. A similar pattern was observed for a total of 94 individuals examined in this study. Interestingly, one of the DNA bands with primer I (arrow in Fig. 1*a*) was found to be polymorphic, present in some individuals and not in others. This presence–absence polymorphism was found to be highly reproducible and individual-specific. It is present in ~30% of the individuals studied from both the populations (Caucasians: 25/82; Dogribs: 4/12). Further characterization of this band involved

[2]Genetics Computer Group. 1994. Wisconsin sequence analysis package, program manual, version 8. University of Wisconsin, Madison.

© 1998 NRC Canada

Fig. 1. (*a*) Representative region of a fluorogram generated by separation of ^{35}S-labeled PCR products from seven individuals using a degenerate retroviral primer. Note that each lane contains a large number of bands, one of which (arrow) is polymorphic in a sample of 94 individuals. This band is apparent in ~30% of the samples (both Caucasians and Dogrib Indians) analyzed. (*b*) A 280-bp product in random samples amplified with internal primers for sequence 1 (on the left, marked 341 int) and a 289-bp product amplified with internal primers for sequence 2 (on the right, marked 343 int); "+" and "–" indicate presence and absence of the polymorphic band (*a*), respectively, in the samples shown. (*c*) Internally amplified PCR products digested with *Rsa*I (left), showing larger fragment of ~190 bp, and *Hin*cII (right), showing larger fragment of ~230 bp; "+" and "–" indicate presence and absence of the polymorphic band (*a*) in the samples shown.



sequencing of DNA from this band, use of internal primers, restriction digestion by diagnostic enzymes, evaluation of sequence similarity, and comparison of the pattern between the two populations.

Fifteen clones each from one Caucasian and one Dogrib individual were sequenced to determine whether they shared common sequences with respect to this polymorphism. Four different sequences of two sizes, 341 bp (sequences 1 and 1′) and 343 bp (sequences 2 and 2′) were obtained. These sequences were aligned for homology matches using the GCG sequence analysis program. The 341 bp sequence 1 obtained from the Dogrib individual was 98.8% identical with the 341 bp sequence 1 obtained from the Caucasian. Their differences are confined to four sites only, including one in the primer region. Three of these sites could represent mutations that have occurred since the divergence of the two populations, as the possible PCR errors were ruled out by use of repeated experiments. Similarly, the 343 bp sequence 2 was also found to be very similar between the two individuals: 98.9% similarity between the Dogrib individual and the Caucasian. It is interesting to note that the 341 bp sequence 1 and 343 bp sequence 2 are very different from each other, with a percent identity of only 44.2% (Fig. 2). Additionally, the two sequences of 341 bp (1 and 1′) have an identity of 37.5% and the two sequences of 343 bp (2 and 2′) have an identity of 40% (Fig. 3). In fact, the only strong similarity between these two pairs of sequences includes the two primer ends.

Internal primers were used to further evaluate the nature of this polymorphism, since a presence–absence PCR polymorphism could arise owing to either the presence–absence of an unique sequence or a difference in the priming sites. Two pairs of internal primers were designed (Table 1) that were complementary to regions within the sequences obtained and interior to the retroviral primers (italics in Fig. 2). Only two types of sequences were examined using these internal primers: 341 bp sequence 1 and 343 bp sequence 2. Genomic DNA samples (representing presence as well as absence of the polymorphic band with the single retroviral primer) were used in PCR amplifications using the internal primers for these two sequences. Every genomic DNA assessed yielded PCR products of the expected sizes with their respective internal primers, i.e., of 280 and 289 bp for the two original sequences of 341 and 343 bp, respectively. A representative result of such a PCR reaction included in Fig. 1*b* shows that all genomic DNAs, representing presence (+) as well as absence (-) of the polymorphic band, yielded DNA bands of appropriate sizes after PCR with the internal

© 1998 NRC Canada

Fig. 2. The two different sequences (percent identity of 44.2%), each of which are conserved between the Caucasian and Dogrib Indian samples. (*a*) Sequence 1. (*b*) Sequence 2. Underlined bases represent variable sites. Boxes mark sites for diagnostic restriction enzymes: *Rsa*I in sequence 1 and *Hin*cII in sequence 2. Bases in italics represent internal primers.

(a)

1-CCAAGCTTGT TTTGCCTCAA GGGCTGTAAA *ATGTAGGTTG TGGCCCA*GGT

TTGACATGTG GTCCATAGTT CACTGACTCC AGTTCTATGC AGGGAGAGTA

GTAGCAAAAA GAGAGACATG TACTTACAAC TATATACAGT CTCATCTTGA

CAGAGTATGA AGTTTGAAGA AGGGTCATGA TAGATAAGCC CAGAGATGTA

GGTAGGGGAT ACGCAGGGTG TCACATTGAG GAAATTTAGA CCTACATCTA

TAAATATGAG TATCAGTAGA ATGAGAAAAT CATATT*TGCA TTTTAGAAAG*

*ATGGCTG*TGT TGAGAATTTC CCTGAGGAAA AACAAGCTTG G -341

(b)

1-CCAAGCTTGT TCTTCCTCAG GGAAC*TGAAG CACAGGCAGG CAGTTA*GAGT

ATACTGAGCA AAGAAGGGAC AGAAAAGAAA GAGGTTGGAA ACTGAGCCAG

GGACTGTATC TTACAAACCA CCATATGTCA TAATAGGAAG TTGAGATTGT

ATTCCATGCA CAGTGGCAAG TCCTTAAAGA ATTCCAAGCA GTAGAATAGC

ATGATGTCTA TTAAGTGGGC AGTAGTGATT GAAAGAAGGA AATCTATCAA

AGTTAACCAA AAGAAATAGA AGAGGCAAAA AATCTGTGGA TCAA*GGGGAC*

*GAAATCTATG CATT*ATCAAA ACCTTGAGGA AAAACAAGCT TGG -343

primers. This suggested that the polymorphism was not due to deletions, but was rather due to sequence differences in the priming sites. The identity of these amplified products was further verified with specific diagnostic restriction enzymes, each with only one site within the internally amplified sequences: *RSa*I for sequence 1 and *Hin*cII and *Nde*I (not shown) for sequence 2. As an example, the *Rsa*I digestion of the PCR product generated by the internal primers specific to the 341-bp band yielded two fragments with the expected sizes of ~90 and ~190 bp, and the *Hin*cII digestion of the PCR product generated by the internal primers specific to the 343-bp band yielded two fragments with the expected sizes of ~60 and ~230 bp (Fig. 1*c*).

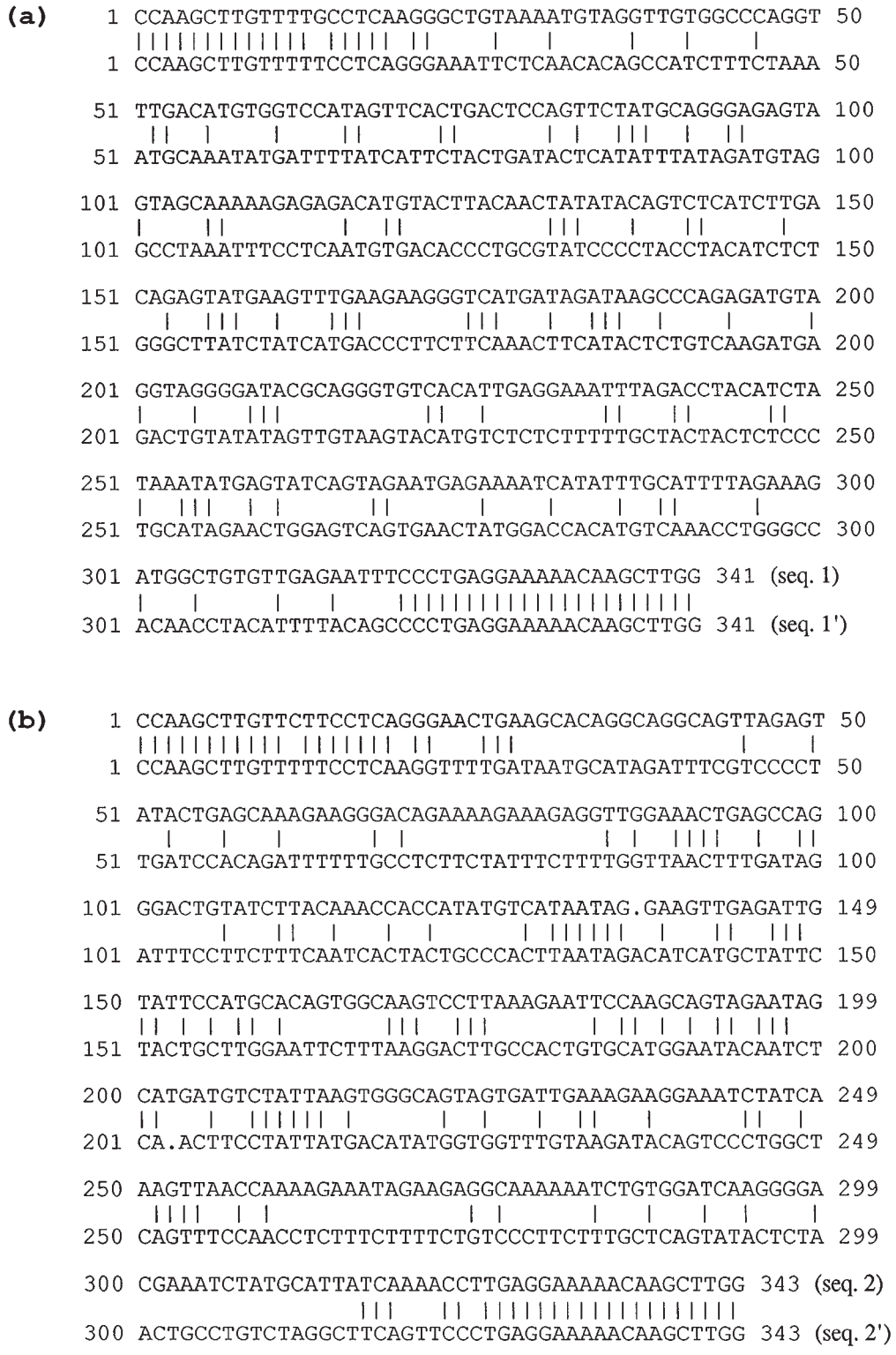
We focused on the two sequences that are nearly identical between the two populations (Fig. 2). BLAST searches at the DNA level did not generate any significant matches. These sequences were translated using all three open reading frames, and the longest resulting peptide sequences were assessed for similarities in a data-base search. The peptide sequence for sequence 1 is 113 amino acids long with no stop codons. The different sequences from the BLAST results included Equine infectious anemia virus, membrane protein of the Dengue virus, and the *env* polyprotein precursor, all of RNA origin. The peptide sequence for sequence 2 is 114 amino acids long, with 3 internal stop codons. Hence, the BLAST results were limited. However, this peptide showed homology to viral sequences as well, which included the surface and major surface antigen of the Hepatitis B virus.

Retroviral-related sequences must account for a significant proportion of the human DNA that has undergone extensive reorganizations (retrotransposition, insertion, deletion, and substitution, among others) during the evolution of the human genome. One would also expect this phenomenon of genome evolution and divergence to represent a continuum that may be assessed by evaluation of divergent populations that have remained isolated from each other for a relatively long period of time. In the following section we discuss our comparative results for the Caucasians from southwestern Ontario and the Dogribs of the Northwest Territories, Canada in the context of evolution and divergence of the two populations.

As indicated earlier, the Nadene American Indians, which include the Dogribs of northwest North America, differ from the Caucasians of southwestern Ontario in many ways. Southwestern Ontario has a relatively large population of mixed lineage, mainly European. The Athapaskan-speaking Dogrib population on the other hand, represents a smaller, more homogeneous group of individuals, whose ancestry is mostly Asian. The features that are unique between these two populations include the number and time of founding events, size of founder populations, and population size. Archeological evidence on the timing of ancestral American Indian migrations (1–3 in number) is ambiguous. However, traditional anthropological analysis has confirmed that American Indians came from Asia, probably crossing the Bering land bridge when it was exposed during an episode

© 1998 NRC Canada

Fig. 3. Comparison between the two types of sequences observed for the 341-bp (*a*) and 343-bp (*b*) products. Note that the only regions similar in both the products are at the two primer ends.



of glaciation (Wallace and Torroni 1992). Comparison of mtDNA diversity and analysis of linguistic diversity tends to place these migrations within the past 15 000 years before present (Wallace et al. 1985), and to place the time that the Nadene became genetically distinct at about 5259 – 10 500 years before present (Torroni et al. 1992).

Given that these two populations diverged so long ago and that they both have retroviral-related sequence polymorphisms that are nearly identical, suggests that the observed polymorphism precedes the divergence of these groups. It is likely that these retroviral sequences have been present in the human genome for a fairly long period of

© 1998 NRC Canada

time. Similar results have also been reported by Yeh et al. (1995), who observed similarities of the human endogenous retroviral multigenes in DNA prepared from various racial backgrounds, and suggest that most of these HERV-E genes were incorporated into the human genome early during evolution and that these sequences have not changed during human racial divergence. Why is it then that these sequences have been conserved for so long, with no major changes, in two distinct groups? There are two possibilities. It is likely that the time frame is not long enough to accumulate sequence differences. It is also likely that there is a lack of mutability within these sequences or that a unique selection process exists, whereby these sequences could be part of a larger functional domain.

The fact that all 94 individuals (representing Dogribs and Caucasians) studied showed more than one nucleotide sequence present as part of the polymorphic band, suggests that these sequences are not present at the same locus. Such results could be interpreted using a two-locus model, with the 341-bp sequence representing one locus and the 343-bp sequence representing the other. Both of these loci are polymorphic, representing sequences 1 and 1′ and 2 and 2′. However the Dogrib individual was found to be homozygous for the 343-bp sequence and heterozygous for the 341-bp sequence, while the Caucasian individual was found to be homozygous for the 341-bp sequence and heterozygous for the 343-bp sequence. These sequences could be present either in tandem in the same region or at different sites, including different chromosomes. Regardless of where these sequences are in the genome, or how many copies are present, they have been kept the same for a long period of time (~10 000 years) in the human genome. A follow-up study on primate sequences and chromosomal localization of the sites of the observed polymorphism should offer insight into the mechanisms associated with retroviral-based evolution of the human genome. As it stands, the observed polymorphism could be used as a reliable molecular marker in other studies.

Molecular characterization of the band establishes that this polymorphism is due to differences in the genome sequence confined to the priming sites. More importantly, distinct groups of sequences associated with this band would argue for sequential duplication, retrotransposition, or multiple-insertion events, all before the separation of the two populations. Such markers help in developing details of the history of human populations that are thought to have originated in Africa about 100 000 – 200 000 years ago with racial differentiation occurring within the past 100 000 years (Waddle 1994).

Acknowledgements

We thank Dr. E. Szathmary (currently of The University of Manitoba) and Dr. R. Ferrell (Pittsburgh) for DNA from the Dogrib samples and the Stanley Foundation for financial support.

References

Bennetzen, J.L. 1996. The contributions of retroelements to plant genome organization, function and evolution. Trends Microbiol. **4**: 347–353.

Coffin, J. 1984. Endogenous viruses. *In* RNA tumor viruses, vol. 1. *Edited by* R. Weiss, N. Teich, H. Varmus, and J. Coffin. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. pp. 1109–1203.

Jeanpierre, M. 1987. A rapid method for the purification of DNA from blood. Nucleic Acids Res. **15**: 9611.

Krieg, A.M., Gourley, M.F., and Perl, A. 1992. Endogenous retroviruses: potential etiologic agents in autoimmunity. FASEB J. **6**: 2537–2544.

Lefebvre, S., Hubert, B., Tekaia, F., Brahic, M., and Bureau, J.F. 1995. Isolation from human brain of six previously unreported cDNAs related to the reverse transcriptase of humanendogenous retroviruses. AIDS Res. Hum. Retroviruses, **11**: 231–237.

Leib-Mosch, C., Bachmann, M., Brack-Werner, R., Werner, T., Erfle, V., and Hehlmann, R. 1992. Expression and biological significance of human endogenous retroviral sequences. Leukemia (Basingstoke), **6**: 72S–75S.

O'Reilly, R.L., and Singh, S.M. 1996. Retroviruses and schizophrenia revisited. Am. J. Med. Genet. **67**: 19–24.

Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. Molecular cloning: a laboratory manual. 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Shih, C.C., Stoye, J.P., and Coffin, J.M. 1988. Highly preferred targets for retrovirus integration. Cell, **53**: 531–537.

Szathmary, E.J. 1983. Dogrib Indians of the Northwest Territories, Canada: genetic diversity and genetic relationship among subarctic Indians. Ann. Hum. Biol. **10**: 147–162.

Szathmary, E.J., Ferrell, R.E., and Gershowitz, H. 1983. Genetic differentiation in Dogrib Indians: serum protein and erythrocyte enzyme variation. Am. J. Phys. Anthropol. **62**: 249–254.

Tassabehji, M., Strachan, T., Anderson, M., Campbell, R.D., Collier, S., and Lako, M. 1994. Identification of a novel family of human endogenous retroviruses and characterization of one family member, HERV-K(C4), located in the complement C4 gene cluster. Nucleic Acids Res. **22**: 5211–5217.

Torroni, A., Schurr, T.G., Yang, C.C., Szathmary, E.J., Williams, R.C., Schanfield, M.S., Troup, G.A., Knowler, W.C., Lawrence, D.N., Weiss, K.M., and Wallace, D.C. 1992. Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. Genetics, **130**: 153–162.

Waddle, D.M. 1994. Matrix correlation tests support a single origin for modern humans. Nature (London), **368**: 452–454.

Wallace, D.C., and Torroni, A. 1992. American Indian prehistory as written in the mitochondrial DNA: a review. Hum. Biol. **64**: 403–416.

Wallace, D.C., Garrison, K., and Knowler, W.C. 1985. Dramatic founder effects inAmerindian mitochondrial DNAs. Am. J. Phys. Anthropol. **68**: 149–155.

Xiong, Y., and Eickbush, T.H. 1988. Similarity of reverse transcriptase-like sequences of viruses, transposable elements, and mitochondrial introns. Mol. Biol. Evol. **5**: 675–690.

Yeh, K.W., Yang, W.K., Huang, H.C., Feng, Y.N., Liu, J.C., Wu, F.Y.H., and Wu, C.W. 1995. Cloning and characterization of the endogenous retroviral-tRNA(Glu) multigene family from human genomes of different racial backgrounds. Gene (Amst.), **155**: 247–252.

© 1998 NRC Canada