analysis on genomic DNAs from the B-LCLs. The variable region of the Ig heavy chain is encoded by 3 gene clusters: approximately 40 functional variable (V) genes, 6 functional joining (J) genes and 25 functional diversity (D) genes. At the later stages of B lymphocyte differentiation, this region undergoes genomic rearrangements such that only one randomly selected gene from each gene cluster is connected to form the mature V(D)J variable gene, and the remaining "left over" genes are removed from the genome (summarized in OMIM #147070, http://www.ncbi.nlm.nih.gov/omim/). The PCR analysis identified V(D)J recombination in all B-LCLs (Figure 8). Both the results of the array CGH analysis and the PCR analysis were consistent with a previous report that many B-LCLs secrete Ig into the culture medium (Rosen et al., 1977). Since six of the lines showed a single PCR product, these lines were each presumed to consist of a single B cell clone. However, the other cell lines showed two or more PCR products indicating that they consisted of two or more B cell clones.

DISCUSSION

Long-term cell culture lines, such as immortalized cancer cell lines, generally possess abnormal karyotypes. However, some cell lines tend to retain a normal karyotype. It is well known that B-LCLs and pluripotent stem cell lines, such as embryonic stem (ES)

cells and induced pluripotent stem (iPS) cells, show this behavior. Due to this characteristic, B-LCLs have been used in many projects, such as the International Histocompatibility Working Group (http://www.ihwg.org/cellbank/cell_lines/blcl.html), International HapMap Project (http://hapmap.ncbi.nlm.nih.gov/) and Human Genome Diversity Project (http://www.cephb.fr/en/hgdp/table.php). High throughput and comprehensive genome analyses have recently been reported using B-LCL collections, e.g., genotyping of blood cell alloantigens (Kroll et al., 2001), detection of copy number variation (CNV) (Redon et al., 2006), and population genetics (Li et al., 2008).

It may be inevitable, however, that long-term cell cultures will accumulate chromosomal aberrations and/or genetic mutations. Additionally, it should be noted that there is still some controversy over whether B-LCLs are truly immortalized cells. Sugimoto et al. (2004) reported that B-LCLs could no longer proliferate following long-term culture, that is, they showed the so-called "crisis" observed in human primary cell cultures. By chance, some B-LCLs can overcome this crisis phase and continue to grow indefinitely in culture, i.e., show immortalization; however, such cells also showed malignant transformation with abnormal karyotypes (Sugimoto et al., 2004). These findings have prompted a vigorous debate on whether the genomes of B-LCLs are really stable.

To address this question, genome-wide evaluations of the genomic stability of B-LCLs have recently been initiated. Whole genome single nucleotide polymorphism (SNP) genotyping indicated that the discrepancies between B-LCLs and their parental PBMNCs were not statistically significant (Simon-Sanchez et al., 2006; Herbeck et al., 2009). In contrast, however, another study has reported different copy numbers of certain genes in B-LCLs and their parental PBMNCs (The Wellcome Trust Case Control Consortium, 2010). In relation to this issue, Bruder et al. analyzed CNVs in monozygotic twins and found somatic mosaicism for CNVs, i.e., CNVs were detected not only between individuals but also within an individual, with a rate of 70-80% of blood cells affected in the most severe cases and of 10-15% in the less affected cases (Bruder et al., 2008). CNVs form continuously in mitosis and meiosis by a diversity of mechanisms, and it is thought that CNVs are one of major driving forces in the rapid evolution of human beings (reviewed by Hastings et al., 2009).

In our array CGH analysis, several aberrations were detected in B-LCLs when they were compared with PBMNCs. However, most of these aberrations were also detected in B lineage cells when compared with non-B lineage cells. Therefore, it is highly likely that most of the aberrations detected in the B-LCLs were present in their parental B cells rather than having occurred de novo during establishment of the cell

cultures. However, with respect to some of the aberrations detected in some B-LCLs derived from the Sonoda-Tajima collection (Figure S1-S4), it was impossible to determine whether they occurred de novo during culture or were present in the parental B cells since insufficient numbers of parental B cells were available for analysis. Of note, a deletion in chromosome 22 (in the variable and joining regions of the Ig λ light chain) was detected in several B-LCLs but not in all lines, suggesting that EBV had infected the B cells at different stages of differentiation, i.e., before and after the rearrangement of the Ig λ light chain.

Interestingly, amplification at 5q35.3 was only detected in one of the two Japanese B-LCLs derived from the same individual (Figure 6). Of course, we cannot formally deny the possibility of de novo amplification during cell culture. However, this observation might indicate somatic mosaicism as described above (Bruder et al., 2008). Even if this is the case, it is not clear at present whether such somatic rearrangements are coupled with RAG transposition (Reddy et al., 2010) or some other mechanism (reviewed by Hastings et al., 2009). Moreover, with respect to aberrations of genes other than of the Ig chains in B-LCLs and primary B cells, the biological implications remain to be determined. When B-LCLs are utilized in genetic research, such aberrations need to be taken into account.

According to our array CGH results, the variation in copy numbers of some genes in B-LCLs compared to their parental PBMNCs might have been present in the parental B cells, and might have been detected if only B lineage cells had been collected and analyzed in a similar fashion to the present study. The possible consequences of any CNV between B cells and non-B cells have yet to be investigated. At present, we cannot formally exclude the possibility that such CNV occurred during the establishment and culture of the B-LCLs.

Overall, our analyses support the contention that B-LCLs provide a valuable source of genomic DNAs for a wide range of genetic studies. However, it is recommended that short-term cell cultures are utilized. Certainly, so-called immortalized B-LCLs with abnormal karyotypes (Okubo et al., 2001; Sugimoto et al., 2004) should not be used for genetic research. In addition, the occurrence of aberrations in B-LCLs when compared to whole mononuclear blood cells should be taken into account when the affected genomic loci are targeted for research.

The Sonoda-Tajima Cell Collection is a very valuable cell collection obtained from various ethnic populations across the world, particularly from South America. Since some of these populations in South America no longer exist, it will be impossible to prepare a similar collection in the future. Here, we describe establishment of B-LCLs

from the Sonoda-Tajima Cell Collection; currently, B-LCLs from more than 500 of the total of about 3,500 samples have been developed. All of these B-LCLs are available from the cell bank held at the Cell Engineering Division of RIKEN BioResource Center in Japan (http://www.brc.riken.jp/lab/cell/english/). At the moment, approximately 150 B-LCLs are immediately available, while the others are now under preparation for rapid distribution. In addition, we are willing to establish more B-LCLs on demand from scientists around the world. We believe that our development of this resource will contribute to various fields of science such as human genetics, human evolution, the history of human migrations across continents, the pharmacokinetics of ethnic minority groups, among others.

FUNDING

This work was supported by the Ministry of Education, Culture, Sports, Science, and Technology in Japan.

ACKNOWLEDGEMENTS

We thank Dr. I. Ishiwata at Ishiwata Hospital for providing the umbilical cord blood samples, Dr. Y. Kodama at Radiation Effect Research Foundation for suggestions and

comments on karyotype analysis, Dr. T. Yamamoto at Nagoya University for comments on CNVs, Dr. K. Moriwaki at RIKEN BioResource Center for general discussion and encouragement, H. Sone and J. Niikura for technical assistance, and other members of the Cell Engineering Division of RIKEN BioResource Center for help with various technical matters.

REFERENCES

Abe T. et al. 1994. Evidence for defects in V(D)J rearrangements in patients with severe combined immunodeficiency. J Immunol. 152: 5504-5513

Bird AG, McLachlan SM, Britton S. 1981. Cyclosporin A promotes spontaneous outgrowth *in vitro* of Epstein-Barr virus-induced B-cell lines. Nature. 289: 300-301

Bruder CEG. et al. 2008. Phenotypically concordant and discordant monozygotic twins display different DNA Copy-Number-Variation profiles. Am. J. Human Genet. 82: 763-771

Fujiyoshi T. et al. 1999. Characteristic distribution of HTLV type I and HTLV type II carriers among native ethnic groups in South America. AIDS Res Hum Retroviruses. 15: 1235-1239

Hastings PJ. et al. 2009. Mechanisms of change in gene copy number. Nat Rev Genet. 10:551-564

Herbeck JT. et al. 2009. Fidelity of SNP array genotyping using Epstein Barr Virus-transformed B-lymphocyte cell lines: Implications for genome-wide association studies. PLoS ONE. 4: e6915

Ijichi S. et al. 1993. Identification of Human T cell Leukemia virus type IIb infection in the Wayu, an aboriginal population of Colombia. Jpn J Cancer Res. 84: 1215-1218

Kiyoi H, Naoe T, Horibe K, Ohno R. 1992 Characterization of the immunoglobulin heavy chain complementarity determining region (CDR)-III sequences from human B cell precursor acute lymphoblastic leukemia cells. J Clin Invest. 89: 739-746

Komurian-Pradel F, Pelloquin F, Sonoda S, Osame M, de The G. 1992. Geographical subtypes demonstrated by RFLP following PCR in the LTR region of HTLV-1. AIDS Res Hum Retroviruses. 8: 429-434

Kroll H, Carl B, Santoso S, Bux J, Bein G. 2001. Workshop report on the genotyping of blood cell alloantigens. Transfus Med. 11: 211-219

Lalle P. et al. 1995. Genomic stability and wild-type p53 function of lymphoblastoid cells with germ-line *p53* mutation. Oncogene. 10: 2447-2454

Li H-C. et al. 1999. The presence of ancient human T-cell lymphotropic virus type I provirus DNA in an Andean mummy. Nat Med. 5: 1428-1432

Li JZ. et al. 2008. Worldwide human relationships inferred from genome-wide patterns of variation. Science. 319: 1100-1104

Masters JR. et al. 2001. Short tandem repeat profiling provides an international reference standard for human cell lines. Proc Natl Acad Sci USA. 98: 8012-8017

Miura T. et al. 1994. Phylogenetic subtypes of human T-lymphotropic virus type I and their relations to the anthropological background. Proc Natl Acad Sci USA. 91:

1124-1127

Miura T. et al. 1997. Molecular phylogeny of human T-cell leukemia virus type I and II of Amerindians in Colombia and Chile. J Mol Evol. 44: S76-S82

Nilsson K. 1979. The nature of lymphoid cell lines and their relationship to the virus. In: Epstein MA, Achong BG, editor. *The Epstein-Barr Virus*. Berlin: Springer-Verlag. p. 225-266,

Nilsson K. 1992. Human B-lymphoid cell lines. Hum Cell. 5: 25-41

Okubo M. et al. 2001. Clonal chromosomal aberrations accompanied by strong telomerase activity in immortalization of human B-lymphoblastoid cell lines transformed by Epstein-Barr virus. Cancer Genet Cytogenet. 129: 30-34

Osame M, Usuku K, Ijichi N, Amitani H, Igata A. 1986. HTLV-1 associated myelopathy, a new clinical entity. Lancet i:1031-1032

Reddy YVR, Perkins EJ, Ramsden DA. 2010. Genomic instability due to V(D)J recombination-associated transposition. Gene Dev. 20: 1575-1582

Redon R. et al. 2006. Global variation in copy number in the human genome. Nature. 444: 444-454

Rickinson AB. et al. 1984. T-cell-mediated regression of "spontaneous" and of Epstein-Barr virus-induced B-cell transformation *in vitro*: Studies with Cyclosporin

A. Cell Immunol. 87: 646-658

Rosen A, Gergely P, Jodal M, Klein G, Britton S. 1977. Polyclonal Ig production after Epstein-Barr virus infection of human lymphocytes *in vitro*. Nature. 267: 52-54

Simon-Sanchez J. et al. 2006. Genome-wide SNP assay reveals structural genomic variation, extended homozygosity and cell-line induced alterations in normal individuals. Human Mol Genet. 16:1-14

Sonoda S. Li HC. Tajima K. 2011. Ethnoepidemiology of HTLV-1 related diseases: Ethnic determinants of HTLV-1 susceptibility and its worldwide dispersal. Cancer Sci. 102: 295-301

Sugimoto M, Tahara H, Ide T, Furuichi Y. 2004. Steps involved in immortalization and tumorigenesis in human B-lymphoblastoid cell lines transformed by Epstein-Barr virus. Cancer Res. 64: 3361-3364

Switzer WM. et al. 1995. Phylogenetic relationship and geographic distribution of multiple Human T-cell Lymphotropic Virus type II subtypes. J Virol. 69: 621-632

The Wellcome Trust Case Control Consortium. 2010. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. Nature. 464: 713-720

Uchiyama T, Yodoi J, Sagawa K, Takatsuki K, Uchino H. 1977. Adult T-cell leukemia:

clinical and hematologic features of 16 cases. Blood. 50: 481-492

Yamashita M, Ido E, Miura T, Hayami M. 1996. Molecular epidemiology of HTLV-1 in the world. J Acquir Immune Defic Syndr Hum Retrovirol. 13: S124-131

Yoshino K. et al. 2006. Essential role for gene profiling analysis in the authentication of human cell lines. Hum Cell. 19: 43-38

Yunis JJ, Sawyer JR, Ball DW. 1978. The characterization of high-resolution G-banded chromosomes of man. Chromosoma. 67: 293-307

FIGURE LEGENDS

Fig. 1. Hypothetical pathways of prehistoric human migration.

Red arrows indicate hypothetical prehistoric migration routes of humans after leaving Africa approximately 70 thousand years ago. The numbers in boxes indicate the number of individuals who donated blood samples. The yellow circles indicate the approximate geographic location where the blood samples were collected.

Fig. 2. Geographic locations where blood samples were collected

Red circles indicate approximate locations where the blood samples were collected. The numbers in circles correspond to the location numbers in Table 1.

Fig. 3. Karyotype analysis

The karyotypes of the 3 cell lines YAN3268 (A), WY084 (B), and YAN3191 (C) were analyzed. A representative karyotype of the sample of 20 cells karyotyped and 50 mode-analyzed cells. The chromosome notation system follows the guidelines in ISCN 1991.

Fig. 4. Array CGH of chromosome 1 of the YAN3191 cell line

The acquired hybridization signal for chromosome 1 is shown. Dots indicate probes arrayed on the microarray. Red, green and black signals indicate probes that were called "amplified", "decreased" or "no change", respectively, at the corresponding genomic regions in the YAN3191 cell line compared to mononuclear cells derived from peripheral blood.

Fig. 5. An example of array CGH analysis showing the results for chromosome 14

(A) The acquired hybridization signal for the entire chromosome 14. (B, C) Expanded images around 14q11.2 (B) and 14q32.33 (C). The dots in A and the crosses in B and C indicate probes arrayed on the microarray. The red, green and black signals indicate probes that were called "amplified", "decreased" or "no change", respectively, in B-LCLs compared to mononuclear cells derived from peripheral blood in YAN3268, YAN3191 and YAN3143, and in $CD19^+$ cells compared to $CD19^-$ cells in Cord Blood-1, Cord Blood-2 and the Adult Japanese sample.

Fig. 6. Panel of aberration loci in each cell

Red and green columns indicate the loci called as amplifications and deletions, respectively. Yellow column indicates that the aberrations were detected under less

stringent statistical conditions. Grey columns indicate no statistical differences.

Fig. 7. FACS analysis to check the content of CD19(+) cells

A. FACS analysis of PBMNC stained with anti-CD3 and anti-CD19 antibodies. B and C, FACS analysis of CD19(+)-depleted cell population (B) and CD19(+) cell population (C) stained with anti-CD19 antibody after MACS beads separation, respectively. Lower column indicates statistical analysis for each Region shown in figures.

Fig. 8. V(D)J recombination

PCR products were separated on a gel and the amplified V(D)J recombination products are shown. Fibroblasts were used as a recombination-negative control. Mononuclear cells derived from peripheral blood cells (PBMNCs) and $CD19^{+}$ cells (B lineage cells) were used as multiclonal recombination-positive controls. If the cell line is monoclonal, then a single band is generated by PCR.

Table 1 Mongoloid minority populations included in this study

Country	Location	Tribe	Other name	Latitude (degrees N)	Longitude (degree W)	B-LCL Established M	F	Total
Venezuela	1	Sanema (Sanumá)	Sanumá, Chirichano, Guaika, Samatali, Samatari, Sanima, Tsanuma, Xamatari	3.5	64.5	13	9	22
	2	Ye'Kuana	Maquiritari, Cunuana, Defcuana, Maiongong, Maquiritai, Maquiritare, Pawana, Soto, Yefcuana	3.5	64.5	6	14	20
Colombia/Venezuela								
	3	Piaroa	Adole, Ature, Guagua, Kuakua, Quaqua	5	67	2	2	4
Colombia	4	Amorua;Guahibo	-	6	67.5	0	2	2
	5	Etnia Siquani;Guahibo	-	6	67.5	2	1	3
	6	Siquani;Guahibo	-	6	67.5	4	1	5
	7	Guahibo	Goahibo, Goahiva, Guaigua, Guajibo, Guayba, Sicuani, Sikuani, Wahibo	6	67.5	3	2	5
	8	Ticuna	Tikuna, Tucuna, Tukúna	4	69.5	6	14	20
	9	Wayu	Wayuu, Guajiro, Goajiro, Guajira, Uáira, Waiu, Wayúu, Wayuunaiki	11.5	72	36	62	98
	10	Wayuu Epiayu	-	11.5	72	0	2	2
	11	Wayuu Puchaina	-	11.5	72	1	1	2
	12	Wayuu Uriana	-	11.5	72	2	2	4
	13	Matapi	Yucuna, Yukuna	-1	69.5	1	0	1
	14	Miraña	Bora, Boro, Meamuyna	-1	69.5	0	1	1
	15	Cumbal	-	1	77	4	11	15
	16	Inga	Highland Inga	1	77	3	13	16
	17	Kamsa	Camsá, Camëntsëá, Coche, Kamemtxa, Kamse, Sibundoy	1	77	3	17	20
Ecuador	18	Cañar	Cañar Highland Quichua	-2.5	79	7	20	27
	19	Saraguro	Loja Quichua, Saraguro Quichua	-3.5	79.5	7	15	22
Bolivia	20	Aymara	-	(-16) - (-18.5)	68.5 - 69	15	13	28
	21	Chipaya	Puquina	-18.5	68.5	5	3	8
	22	Quechua	North La Paz Quechua	(-21.5) - (-21)	66 - 67.5	13	12	25
	23	Mestizo		(-20) - (-21)	65.5 - 66	2	1	3
Paraguay	24	Chaco (Lengua)	Enxet	-22.5	59.5	18	17	35
	25	Chaco (Nivaclé)	Ashlushlay, Axluslay, Chulupe, Chulupí, Chulupie, Churupí, Nivaklé	-22.5	59.5	9	2	11
	26	Chaco (Sanapaná)	Lanapsua, Quiativis, Quilyacmoc, Saapa, Sanam	-22.5	59.5	1	0	1
Chile	27	Atacama	Kunza, Atacameño, Likanantaí, Lipe, Ulipe	(-23) - (-23.5)	68	14	13	27
	28	Mapuche	Mapudungun, Araucano, Mapudungu	-39	71.5 - 72	26	33	59
	29	Huilliche	Huiliche, Veliche	-43	73	5	4	9
Argentine	30	Puna	-	-22.5	65.5	7	17	24
			Total			215	304	519

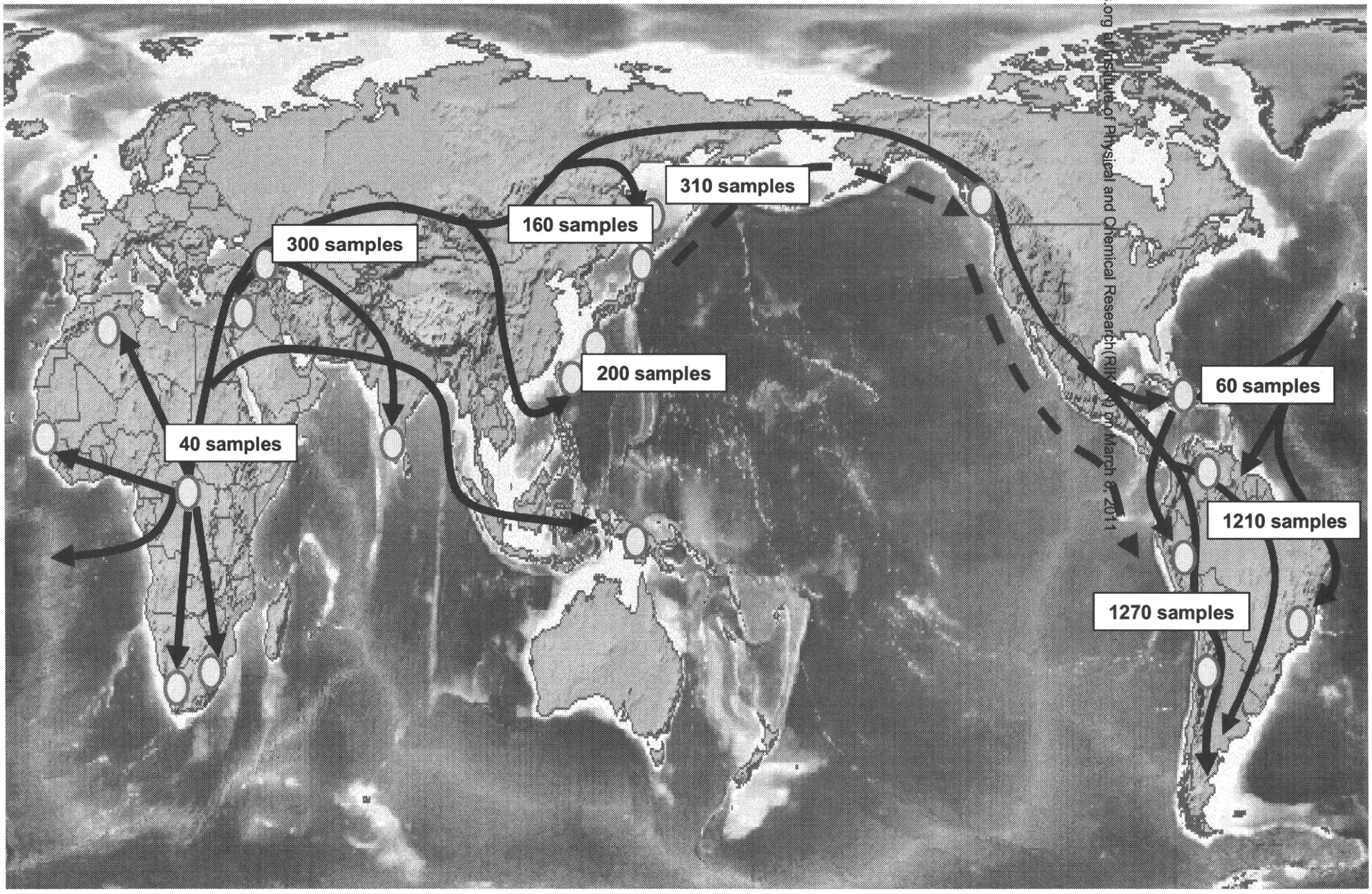
Other names were referred by Ethnologue country index (http://www.ethnologue.com/country_index.asp)
"-, not entried in Linguist group"

Table 2 Genomic loci where aberrations were detected

Chromosome	Cytoband	Start	Stop	pattern	Gene name
detected in all samples					
2	p11.2	88916534	89743016	deletion	immunoglobulin kappa J, V clusters
7	p14.1	38262501	38349233	amplification	T cell receptor gamma C, J, V clusters
14	q11.2	21485151	22050050	amplification	T cell receptor alpha V, J clusters
14	q32.33	105314054	106286079	deletion	immunoglobulin heavy chain V, D, J clusters
detected in over 5 samples					
7	q34	141988254	142202474	amplification	T cell receptor beta V, J clusters
16#$	q22.1	66945730	66967713	amplification/deletion	LOC100129324, Sphingomyelin Phosphodiesterase 3
22	q11.22	20717615	21576029	deletion	immunoglobulin lambda V, J clusters
detected in 2-4 samples					
2#	q35	218971508	218979039	amplification	MIRN26B, CTDSP1
3	p21.31	48577843	48600744	amplification	UCN2, COL7A1
detected only in Dann and PBMC					
5	q35.3	179147890	179177396	amplification	LTC4S, MGAT4B

#, aberration was not detected in B lymphocyte population
$, aberration was not detected in B-LCL established from adult Japanese

Figure 1



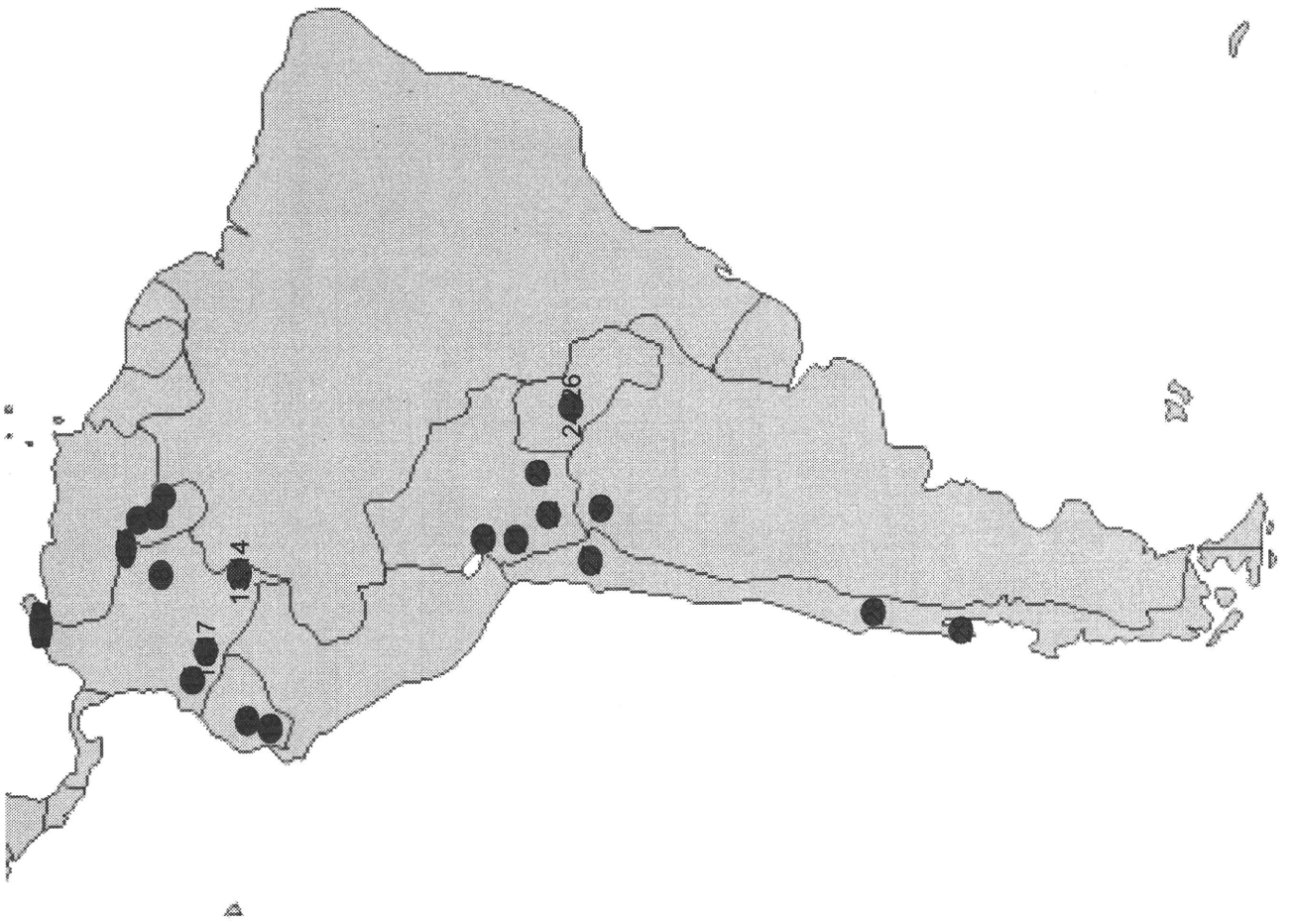
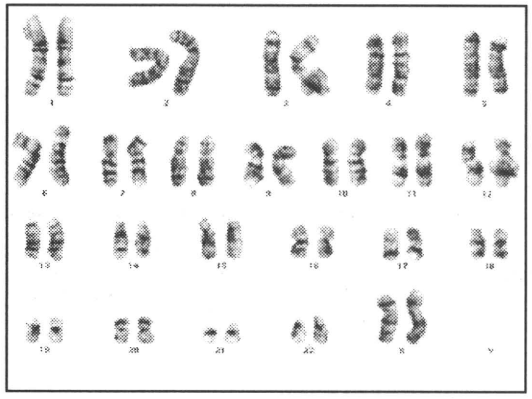


Figure 2

Figure 3

A



YAN3268

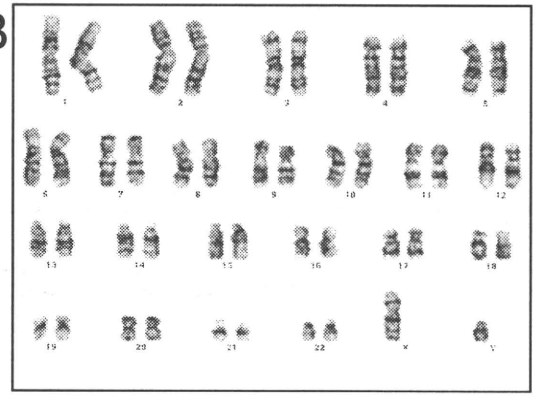
Mode of chromosome number

number of chromosomes	46			
number of analyzed cells	50			

Karyotype

46,XX [20]

B



WY084

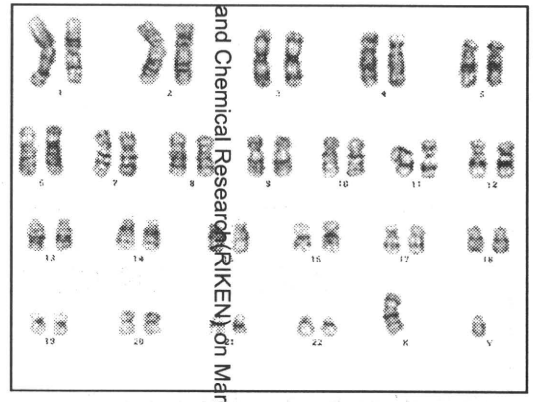
Mode of chromosome number

number of chromosomes	46	90	91	
number of analyzed cells	48	1	1	

Karyotype

46,XY [18]
90,XXYY,-6,-10 [1]
91,XXYY,-22 [1]

C



YAN3191

Mode of chromosome number

number of chromosomes	46	92		
number of analyzed cells	46	4		

Karyotype

46,XY,1qh+(presumable) [20]