

Sexagesimal scale for mapping human genome

Escala sexagesimal para mapear el genoma humano

RICARDO CRUZ-COKE

Servicio de Genética, Departamento de Medicina, Hospital Clínico, Universidad de Chile,
Avenida Santos Dumont 999, Santiago, Chile, e-mail: rcruzco@machi.med.uchile.cl

ABSTRACT

In a previous work I designed a diagram of the human genome based on a circular ideogram of the haploid set of chromosomes, using a low resolution scale of Megabase units. The purpose of this work is to draft a new scale to measure the physical map of the human genome at the highest resolution level. The entire length of the haploid genome of males is deployed in a circumference, marked with a sexagesimal scale with 360 degrees and 1296000 arc seconds. The radio of this circumference displays a semilogarithmic metric scale from 1 m up to the nanometer level. The base pair level of DNA sequences, 10^{-9} of this circumference, is measured in milliarsec unit (mas), equivalent to a thousand of arcsecond. The "mas" unit, correspond to 1.27 nanometers (nm) or 0.427 base pair (bp) and it is the framework for measure DNA sequences. Thus the three billion base pairs of the human genome may be identified by 1296000000 "mas" units in continuous correlation from number 1 to number 1296000000. This sexagesimal scale covers all the levels of the nuclear genetic material, from nucleotides to chromosomes. The locations of every codon and every gene may be numbered in the physical map of chromosome regions according to this new scale, instead of the partial kilobase and Megabase scales used today. The advantage of the new scale is the unification of the set of chromosomes under a continuous scale of measurement at the DNA level, facilitating the correlation with the phenotypes of man and other species.

Key words: human genome, DNA, chromosomes, genetic map.

RESUMEN

En un trabajo anterior yo diseñé un diagrama del genoma humano basado en un ideograma circular del conjunto haploide de cromosomas, usando una escala de baja resolución en megabases. El propósito de este trabajo es el de diseñar una nueva escala para medir el mapa físico del genoma humano al más alto nivel de resolución. La longitud completa del genoma haploide del varón es extendido en una circunferencia, marcada con una escala sexagesimal de 360 grados y 1296000 segundos de arco. El radio de esta circunferencia expone una escala métrica semiologarítmica desde 1 m hasta el nivel del nanómetro. El nivel de bases pares de las secuencias del ADN de esta circunferencia es medida en unidades milliarsec, "mas", equivalente a una milésima de segundo de arco. La unidad más equivalente a 1.27 nanómetro (nm) o 0.427 pares de bases (pb), y es el bastidor para medir las secuencias del ADN. Así, los 3 millones de pares de bases del genoma humano pueden ser identificados por 1296000000 unidades "mas" en continua correlación desde el número 1 al número 1296000000. Esta escala sexagesimal cubre todos los niveles del material genético nuclear desde los nucleótidos hasta los cromosomas. Las localizaciones de cada codon y de cada gen pueden ser numerados secuencialmente en las regiones cromosómicas del mapa físico del genoma, de acuerdo con esta nueva escala, en vez de usar las escalas parciales de kilobases y de megabases usadas actualmente. La ventaja de esta nueva escala es la unificación del conjunto de cromosomas bajo una continua escala correlativa de medidas a nivel del ADN, facilitando la correlación con los fenotipos del ser humano y otras especies.

Palabras clave: genoma, ADN, cromosomas, mapa genético.

INTRODUCTION

In a previous work I attempted to design a diagram of the whole structure of the human genome based in a circular ideogram of the haploid set of chromosomes (Cruz-Coke 1990). It was a low resolution chromosome map measured in Mb units. This map established the possibility to indentify the approximate position

of a gene to the nearest degree of the standard scale of the circle.

Recent progress in mapping the human chromosomes to a high level of resolution has reached the DNA sequence itself (Olson 1989). These researches used the partial and limited scales of kilobases and base pairs according to the respective levels of resolution. Nevertheless no attempt has been made to introduce an absolute universal

scale to cover all the levels of magnitude of the human genome, distributed in the chromosomes.

The purpose of this work is to draft a new high resolution scale, with universal and absolute characteristics, that can reach the base pairs levels of DNA sequences and may cover with a continuous measure the whole structure of human genome.

MATERIAL AND METHODS

The human genome containing some 3 billion (3×10^9) base pair of nucleotides, is composed by giant molecules extended along astronomical magnitudes equivalent to the distances of the solar system as shown in Table 1. Consequently, it seems reasonable to use astrophysical units of measurement in order to design an universal scale to cover the full extend of the genetic material.

Material

Within the solar system we determine absolute distances by using newtonian celestial mechanics. The angular measure is a basic element to determine astronomical distances, and it is originated when the circle's circumference is divided in 360 equal parts. One of these parts is an arc-degree ($^\circ$), subdivided in 60 arc minutes ($'$) and each arc minute into 60 arc seconds ($''$). Hence there are $360 \times 60 \times 60 = 1296000''$ in the full circle (Zeilek 1992).

The only direct method to estimate astronomical distances is the heliocentric parallax angle based in orbit of the earth around the sun. Parallax angle based in the change of an object apparent position, a star, when viewed from two differ-

ent location, the sun and the earth. The parallax angle is measured in arc seconds ($''$). The unit of distance is a "parsec" when parallax angle measure one arc second ($''$) and the distance from a star to sun is 206265 times the distance from earth to sun. The equation is given by: $ps = 1/p$, where ps is parcec and p parallax angle (Zeilek 1992).

"Alpha centauri" our nearest star, is 1.33 parsec with a parallax of 0.752". Planet Uranus, at the limit of naked eye resolution visibility from the earth, has a resolution angle of 306". The best modern telescope NNT at La Silla, Chile, may achieve resolutions at 0.3".

Method to construct a high resolution cytogenetic scale

In order to apply the astrophysical unit or measurement to human genetic material, the entire length of the haploid set of chromosomes, 22 autosomes, X & Y, is displayed in a circumference marked with a scale of 360 degrees and 1296000 arcseconds ($''$). Each chromosome was measured according to its relative length and ordered by size, and occupying a given corresponding arc and segment of the circle limited by the integer of a degree, as shown in Fig. 1.

The radius display the semilogarithmic metric scale from 1 meter up to a nanometer (10^{-9}) and the Angstrom unit (10^{-10}). The chromosome level is displayed at the micron level (10^{-6}) and is measure by degrees, the genes are measured by arcseconds at the nanometer level with kilobases. The higher resolution power at the Angstrom (10^{-10}) level is reached by the milliarsecond unit (mas), equivalent to a thousand of arcsec (Seilek 1992). Consequently, the

TABLE I

Universal magnitudes of solar and genome systems

Magnitudes universales del sistema solar y genoma

Unit	Scale	Distance	System	
Tera (T)	10^{12}	To Neptune	Solar	
Giga (G)	10^9	To Mars		
Mega (M)	10^6	To moon		
Kilo (K)	10^3	Human genome length	Genome	
meter	1			
milli (m)	10^{-3}	Chromosome		
micro (u)	10^{-6}			
nano (n)	10^{-9}			DNA thickness
pico (p)	10^{-12}			Atoms

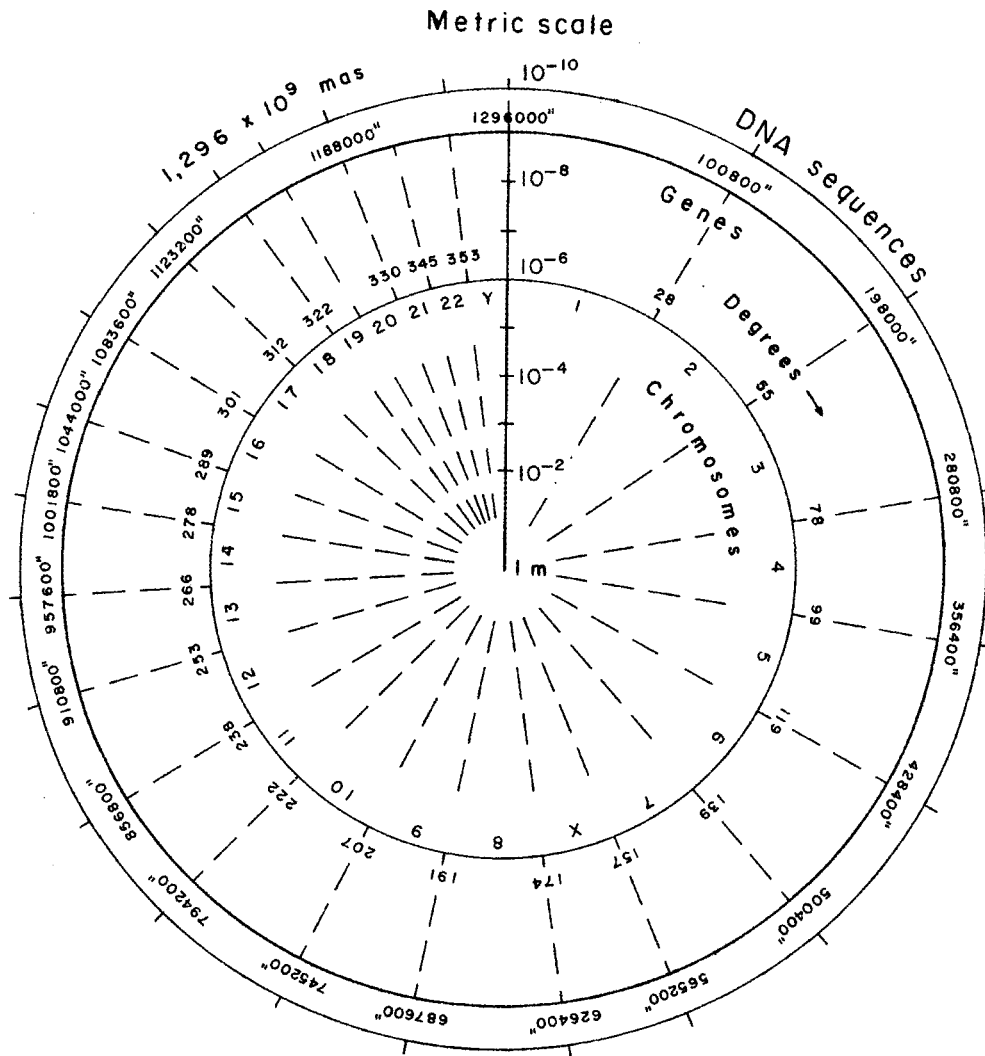


Fig. 1. A circular ideogram of the haploid set of chromosomes of the human male showing a sexagesimal scale of DNA sequences.

Ideograma circular de un conjunto haploide cromosómico de varón que muestra una escala sexagesimal de secuencias de ADN.

unit mas cover 1.27 nanometer and 0.432 base pair.

The milliarsec scale is the framework of the DNA sequences determines the fixed positions along a circular scale of correlative numbers from 1 to 1296000000 mas (milliarsec 0.001).

Levels of mapping human chromosomes

The chromosome maps are used to locate DNA fragments in regions of chromosome bands, at a low level resolution power between 1 to 10 Megabases (Mb). Clone probes determine chro-

sosome through the method of in situ hybridization. Higher resolution to a level of 1 to 100 kilobases (Kb) may be developed by macrorestriction maps using restriction enzymes and pulse field electrophoresis to separate Mb fragments and measure distance in kilobase pairs. The ultimate physical map to obtain the DNA sequence uses Sequence Tagged Sites (STS) reaching the disease gene sequence at the level of one base pair (Fig. 2). Thus, the modern methods to explore the microcosm descend the levels of magnitude through increasingly refined views until reaches the sequence itself (Olson 1989).

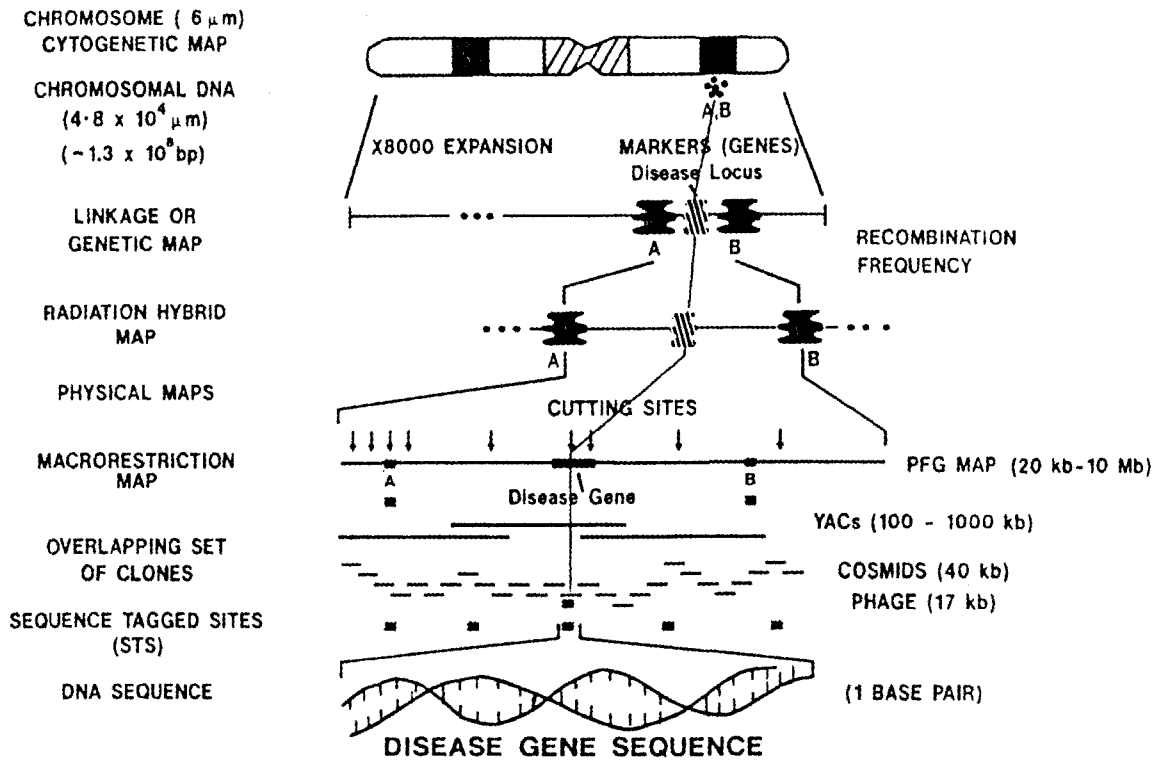


Fig. 2 Levels of mapping human chromosomes (after Olson 1989).

Niveles de mapeo de cromosomas humanos.

RESULTS

I have constructed in Fig. 3 a low resolution ideogram of chromosome 6 as shown in Fig. 1. Chromosome 6, located between the degrees 119 and 139 of the circumference, contains roughly 5.5 % of the chromosomes set of males. Contains 50 bands along some 55 million nm. The Major Histocompatibility Complex (MHC), with the HLA loci, is located according to the Denver notation (McKusick 1992) in region 6p of bands 21.31 to 21.33 (McKusick 1992). Its sexagesimal scale covers 20 degrees and 72000 arcseconds (") displayed along the 160 Megabases. Thus, this new scale adjusts well to the low resolution levels of chromosome regions (Ziegler & Orr 1993).

In Fig. 4 I have constructed a high resolution map of DNA sequences of HLA cluster at low resolutions level, the bands (regions) of short arm of chromosome 6 are marked with the Denver notation showing between 6p 21.31 - 21.33 the MHC cluster including the HLA system. The arcsecond scale ranged in this short arm from the numbers 428400" to 450000".

Descending to the high resolution level of 10-9 m, the HLA Cluster is depicted with all genes distributed along the 3.500 kilobases, which are correlated to the arcsec scale from an arbitrary starting point of number 440000" to 441512". Also is shown a partial nanometer scale.

DISCUSSION

I have chose the allele SQA1-101 of the locus DQ of class II, a predisposing gene to Diabetes (Pérez et al. 1995), to descend to the DNA sequences at the base pair level. A Sequence Specific Oligonucleotide (SSO) probe for allele DQA1, reaches the scale of milliarsec (mas) in the starting position 441200.001". A tentative estimated depicted the 18 bp of the probe with the 6 nm length and the mas scale.

This approach shows that it is possible to measure the DNA at the highest resolution level using three scales; base pair (bp) manometer (nm) and milliarsec (mas). DNA is a polymer composed by nucleotide base pairs separated by 0.34 nm

Chromosome 6

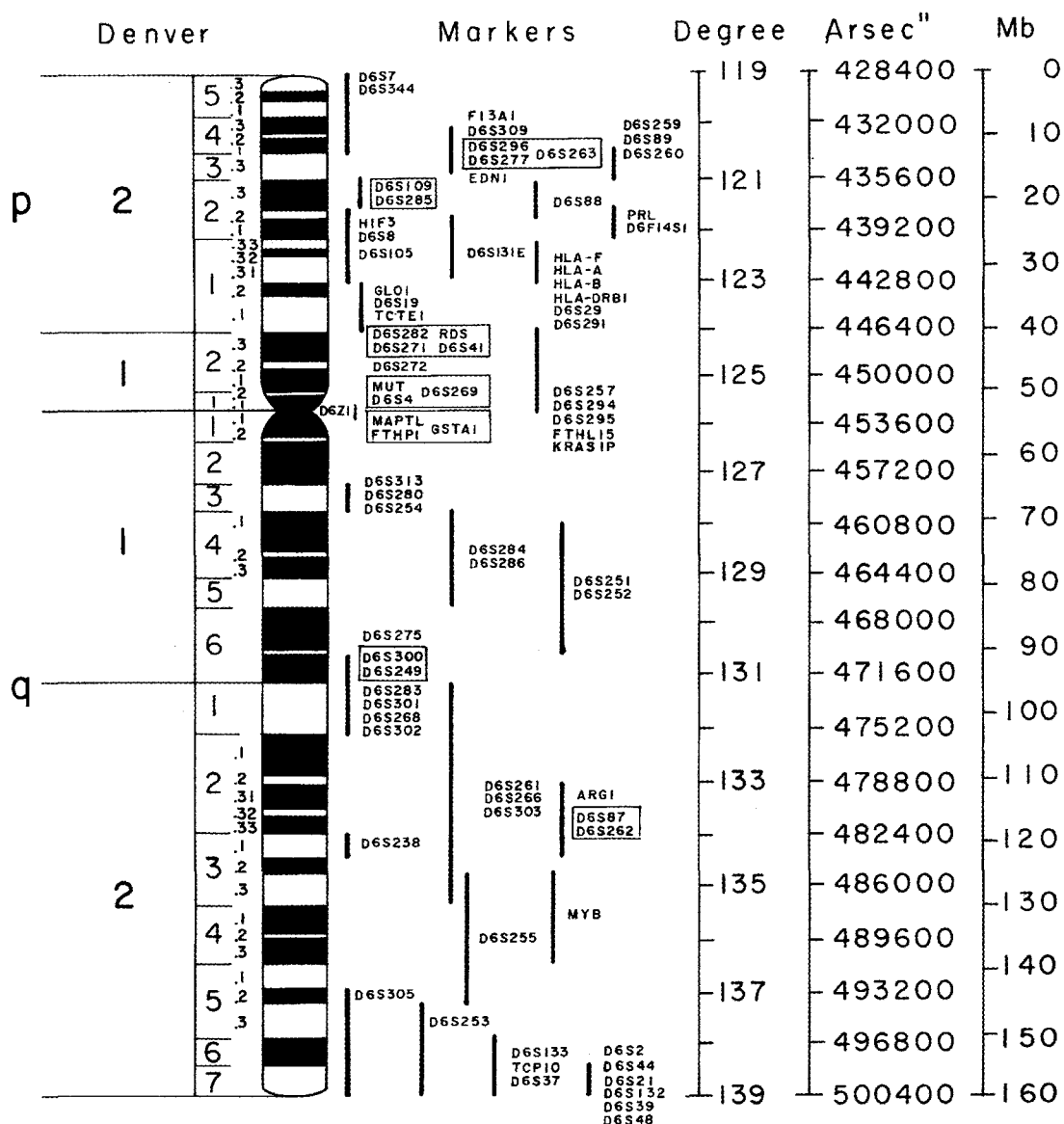


Fig. 3. A low resolution ideogram of Chromosome 6.

Ideograma de baja resolución del cromosoma 6.

with a thickness of 2 nm. The helix makes a complete turn every ten base pair, that is 3.4 nm. Thus, if the DNA molecule contains 3×10^9 bp, the total length may reach 1.02 m. One nanometer covers rough 3 bp, that is, a codon (Singer & Berg 1994).

As the arsec scale contains 1296000 positions and the entire length of DNA molecule contains

3×10^6 kb; $3.000.000/1.296.000 = 2.314$ arcseconds per kilobase. Consequently one arsec equates 0.432 kilobases. Thus, the milliarcsec unit (mas) covers 0.432 bp and 1.27 nm.

The conversion table for the high resolution levels units base pair (bp), nanometer (nm), and milliarcsec (mas) is shown in Table 2.

Sequencing the HLA cluster with the new sexagesimal scale

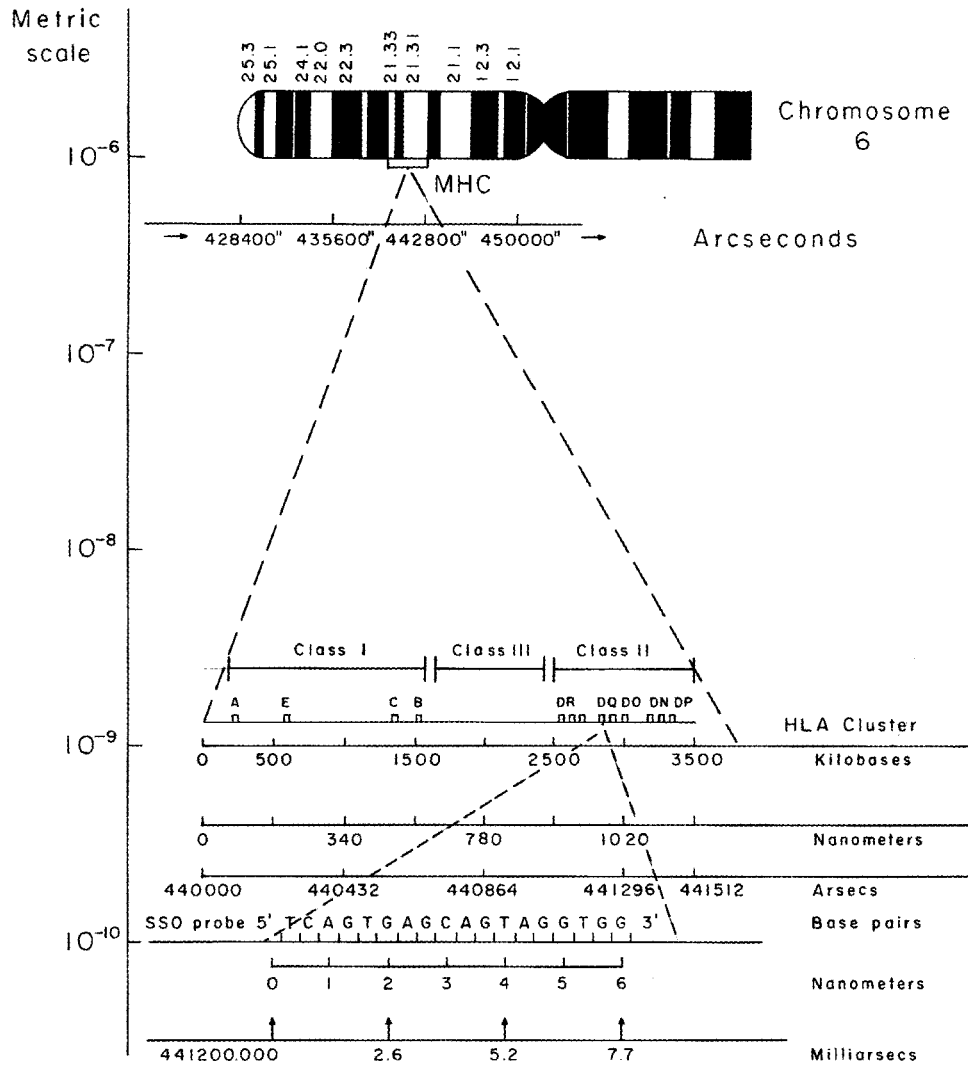


Fig. 4 A high resolution map of HLA cluster marked with a sexagesimal scale.

Un mapa de alta resolución del grupo HLA marcado con escala sexagesimal.

The advantages of the new sexagesimal scale are clear. The new scale uses sexagesimal units of measure applied in astronomical researches. The scale is circular, continuous and adjusted to the geometrical properties of the circle. The scale is absolute, permanent and universal, covering all the ranges of the genetic material.

The human genome is by nature, a changing structure and the present scale of Mb, Kb and bp are measuring only partial events and fractions of

the chromosome regions. On the contrary, the arcsec scale with a given fixed length of 1296000000 milliarsecs (mas) may work as a main frame genome, absolute and fixed. The Sequence Specific Oligonucleotide (SSO) probes may use this scale as a reference point to identify and locate introns and exons of alleles. The total exact length of the human genome is not yet known and the total number of base pairs, the same.

TABLE 2

Conversion table for bp, nm and mas units

Tabla de conversion para bp, nm y mas

bp	nm	mas
1	0.34	0.432
2	0.68	0.864
3	1.02	1.296
4	1.36	1.728
5	1.70	2.160
6	2.04	2.592
7	2.38	3.024
8	2.72	3.456
9	3.06	3.888
10	3.40	4.320

Using this scale, every gene or allele can be numbered correlatively according to a standard code of 10 digits, covering the magnitudes from 1 to 10^{-10} m. The standard digit codenunbers of the three scales compared, are shown in Table 3.

CONCLUSIONS

The introduction of this sexagesimal scale operating as a main frame at the DNA level, will probably help the researchers to positioning at the chromosome regions, all the genes and alleles at the codon level. This fact will facilitate the genetic correlation with phenotypes in man and other species. All the genes in the human species will be located in an unified genomic scale facilitating the interespecific analysis of DNA in the living world.

Invited Editors: R. Godoy-Herrera and G. Gajardo
Received March 5, 2000; accepted September 18, 2000

TABLE 3

Comparison of genomic scales

Comparación de escalas genomicas

Digit code numbers	1	2	3	4	5	6	7	8	9	10
Nanometers (nm)	1	0	2	0	0	0	0	0	0	0
Base pair (bp)	3	0	0	0	0	0	0	0	0	0
milliarsecs (mas)	1	2	9	6	0	0	0	0	0	0

M A I N F R A M E

RECONOCIMIENTO

En memoria del Profesor Danko Brncic quien me orientó e inspiró en la teoría y práctica de la genética para ser aplicada en medicina.

LITERATURE CITED

- CRUZ-COKE R (1990) A general diagram of human genome. *Journal of Medical Genetics* 27: 388-389.
- McKUSICK VA (1992) Mendelian inheritance in man. John Hopkins University Press, Baltimore, Maryland. 1168 pp.
- OLSON M, L HOOD, C CANTOR & D BOTSTEIN (1989) A common language for physical mapping of the human genome. *Science* 245: 1434-1435.
- PÉREZ F, E CARRASCO, M CAVILLAN, MD GUTIÉRREZ, G LARENAS, G LÓPEZ, M GARCÍA DE LOS RÍOS & M SERRANO (1995) Genetic predisposition to develop IDDM. *Revista Médica de Chile* 123: 1205-1213.
- SINGER M & P BERG (1994) Genes and genomes. Blackwell Scientific Publications, Oxford, United Kingdom. 928 pp.
- ZEILEK M, SA GREGORY & EP SMITH (1992) Introductory astronomy and astrophysics. Harcourt & Brace, Forth Worth, Texas. 504 pp.
- ZIEGLER A & HT ORR (1993) Second international chromosome 6 workshop. *Genome Digest* 1: 2-3.