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: rcruzcok@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 Mega-base 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 1,296,000 arc seconds. The radius 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 1,296,000,000 "mas" units in continuous correlation from number 1 to number 1,296,000,000. 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 sequentially in the physical map of chromosome regions according to this new scale, instead of the partial kilobase and Mega-base 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.

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 \(3 \times 10^9\) 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 different 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 parsec 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 semilogaritmic 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 milliarcsecond unit (mas), equivalent to a thousand of arcsec (Seilek 1992). Consequently, the

**TABLE 1**

Universal magnitudes of solar and genome systems

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

After Cruz-Coke (1990)
Fig. 1. A circular ideogram of the haploid set of chromosomes of the human male showing a sexagesimal scale of DNA sequences.

Figura 1. Un 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 chromosome 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).
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 arcseconds 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 441200.00". 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) nanometer (nm) and milliarsec (mas). DNA is a polymer composed by nucleotide base pairs separated by 0.34 nm
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 \(3 \times 10^9\) bp, the total length may reach 1.02 m. One nanometer covers roughly 3 bp, that is, a codon (Singer & Berg 1994).

As the arsec scale contains 1296000 positions and the entire length of DNA molecule contains \(3 \times 10^9\) kb; \(3.000.000/1.296.000 = 2.314\) arcseconds per kilobase. Consequently one arcsec equates 0.432 kilobases. Thus, the milliarsec unit (mas) covers 0.432 bp and 1.27 nm.

The conversion table for the high resolution levels units base pair (bp), nonometer (nm), and milliarsec (mas) is shown in Table 2.
Sequencing the HLA cluster with the new sexagesimal scale

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 arsec scale with a given fixed length of 1296000000 milliarcs (mas) may work as a main frame genome, absolute and fixed. The Sequence Specific Oligonucleotidic (SSO) probes may used 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.

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.
TABLE 2

Conversion table for bp, nm and mas units

<table>
<thead>
<tr>
<th>bp</th>
<th>nm</th>
<th>mas</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.34</td>
<td>0.432</td>
</tr>
<tr>
<td>2</td>
<td>0.68</td>
<td>0.864</td>
</tr>
<tr>
<td>3</td>
<td>1.02</td>
<td>1.296</td>
</tr>
<tr>
<td>4</td>
<td>1.36</td>
<td>1.728</td>
</tr>
<tr>
<td>5</td>
<td>1.70</td>
<td>2.160</td>
</tr>
<tr>
<td>6</td>
<td>2.04</td>
<td>2.592</td>
</tr>
<tr>
<td>7</td>
<td>2.38</td>
<td>3.024</td>
</tr>
<tr>
<td>8</td>
<td>2.72</td>
<td>3.456</td>
</tr>
<tr>
<td>9</td>
<td>3.06</td>
<td>3.888</td>
</tr>
<tr>
<td>10</td>
<td>3.40</td>
<td>4.320</td>
</tr>
</tbody>
</table>

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 codenames 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 interspecific analysis of DNA in the living world.

LITERATURE CITED