Research Journal of Medical Sciences 1 (1): 6-8, 2007

Genome Organisation

R. Blundell
Department of Physiology and Biochemistry,
Biomedical Science Building, Faculty of Medicine and Surgery,
University of Malta, Msida MSD06, Malta

The genome is a functional unit consisting of a hierarchy of subunits: one or more chromosomes, a number of genes, pseudogenes and "junk" DNA (repetitive/simple sequences). The sizes of the genomes vary from one organism to another. Some examples are shown in Table 1.

Table 1: Variation of genome sizes between different organisms[1]

Common name	Latin name	Genome sizes×10^9bp
Phage	*Bacteriophage lamda*	0.00005
Phage	*Bacteriophage T4*	0.00017
Virus	*SV 40 virus*	0.000005
Bacterium	*Escherichia coli*	0.004
Yeast	*Saccharomyces cerevisiae*	0.024
Fungus	*Neurospora crassa*	0.017
Alga	*Chlamydomonas reinhardi*	0.06
Maize	*Zea mays*	3.9
Onion	*Allium cepa*	16.8
Lily	*Lilium longiflorum*	36.1
Pea	*Pisum sativum*	4.5
Silkmoth	*Bombyx mori*	0.52
Fruitfly	*Drosophila melanogaster*	0.18
Crab	*Cancer borealis*	1.7
Rainbow trout	*Salmo gairdneri*	2.5
Tench	*Tinca tinca*	0.9
Lung fish	*Protopterus aethiopicus*	50.0
Xenopus	*Xenopus laevis*	3.15
Frog	*Rana pipiens*	7.6
Salamander	*Amphiuma means*	84.0
Alligator	*Allogator mississipiensis*	2.5
Chicken	*Gallus domesticus*	1.25
Mouse	*Mus musculus*	2.5
Human	*Homo sapiens*	3.0

Chromosomes and karyotyping: Chromosomes are condensed in the late prophase or metaphase stage of mitosis and thus are visible under the light microscope. The karyotype refers to the arrangment of condensed chromosomes.

In humans there are 24 types of chromosomes, 22 autosomes and 2 sex chromosomes. The X chromosomes are present in two copies in female somatic cells, but only one in male somatic cells. Y chromosomes are only present in male somatic cells and thus only present in males[2].

The size and position of the centromere is an important characteristic that distinguish one chromosome from another. The centromere can be midway between the ends of the chromosomes (a metacentric chromosome), somewhat closer to one end than to the other (an submetacentric chromosome), closer to one end (an acrocentric chromosome), or at the end (telocentric chromosomes). The last occurs in mammals but is absent in the human genome.

Several staining methods were developed that revealed substructures within the chromosomes. These substructures had the form of alternating dark and light zones of various widths. Q-bands are obtained when stained with quinacrine compounds. G bands are stained with Giemsa. Giemsa dark zones are (A+T) rich and are present in relatively few genes. Giemsa light zones are (G+C) rich and are present in about three-fourths of the genes. C bands correspond to constitutive heterochromatin, which is found in all centromeres and in the long arm of the Y chromosome.

Bernardi[3] has separated DNA fragments by equilibrium centrifugation and distinguished different families of DNA fragments characterised by their base composition. Warm-blooded vertebrates were found to contain DNA components that are high in GC content whereas cold-blooded vertebrates were found to have low GC levels. These components distinguished by their buoyant densities are called isochores. Most genes in warm-blooded vertebrates were found to be GC rich and therefore in the isochores with the highest GC content.

The number assigned to the chromosome is given in descending order depending on size. In the human karyotype, chromosome number 1 is the largest. There is an exception in that no. 21 is actually smaller than no. 22. Other nomenclatures are used to describe the long or short arm of each chromosome. Letter p (from the French for Petit) is used to define the short arm, while the letter q (because it follows p in the alphabet) is used to describe the long arm. Within each arm of a chromosome, numbers are assigned to large areas called regions and another set of numbers is used for bands within regions[4].

Bands are situated in various regions along the chromosome arms and specific landmarks delimit the regions. The landmarks are defined and consistent,

providing morphological features that are important in identifying chromosomes. Landmarks include the ends of the chromosome arms, the centromere and certain bands. The bands and regions are numbered from the centromere outward.

Genes: It is estimated that the human genome is made up of about 50.000 to 100.000 genes[5]. The human globin gene although relatively small is an example of a typical gene in higher eukaryotes. A globin gene typically has a 5' flanking region that includes a promotor, a coding sequence which is divided into three exons by two non-coding intervening sequences (IVS or introns) and a 3' sequence which is transcribed but not translated[6]. Exons are nucleotide sequences that are included in mature mRNA, whether they are translated or not. Introns are nucleotide sequences that are transcribed, but are spliced out of the pre-mRNA as part of the maturation process that occurs in the nucleus.

The number and size of both exons and introns vary among unrelated genes. Most exons are small, averaging less than 200 nucleotides. Exons that are larger than 200 nucleotides tend to occur at the 3' ends of genes and consist primarily of non-coding sequences[6].

The number of introns within a gene varies from zero (e.g., histone genes) to greater than 50 for collagen genes and at least 75 for the dystrophin gene. The ratio of exons DNA to intron DNA is a matter of some interest, because it defines the size of a gene. On average, there appears to be at least ten times as much intron DNA as exon DNA in mammals and the ratio may actually be larger[7].

Pseudogenes: The advent of gene cloning and rapid DNA sequencing has led to some surprising discoveries, one of which is the occurrence of pseudogenes. These are DNA sequences closely related to functional genes, but incapable of coding for the normal gene product because of one or more deletions, insertions, frameshifts, or stop codons[8].

There are two major classes of pseudogenes. The human globin pseudogenes exemplify the first class, which can be accounted for by gene duplication, followed by mutational inactivation of one of the gene copies in studies where a second normal copy is not selectively beneficial. This type of pseudogene retains the overall organisation of the parental gene in terms of exons and introns.

The second class contains the exon sequences of the functional gene, but not the introns. A poly $(A)^3$ sequence is usually present at the 3' end of the pseudogene. In other words, the pseudogene appears to have been derived from a messenger RNA[9]. A widely held hypothesis is that a messanger RNA is occasionally copied into a cDNA by reverse transciptase enzyme and the cDNA is subsequently inserted into the genome by integrase enzyme.

Interspersed genome-wide repeats: In all higher eurkaryotes a substantial fraction of the genome consist of repeated (or reiterated) sequences[10]. Families of repeats derived from transposable elements constitute a major portion of eukaryotic genomes, far exceeding exons in the proportion of the genomes devoted to them[11]. These elements can be classified in four categories:

- Short Interspersed Nuclear Elements (SINEs)
- Long Interspersed Nuclear Elements (LINEs)
- Long-Terminal Repeat (LTR) retro-virus like elements
- DNA transposons

In the human, there are about 1.100.000 Alu sequences (a SINE) and 590,000 Line I sequences (a LINE). LINE 1 sequences make up a greater proportion of the human genome than all the gene-coding sequences combined. LINEs account for 16.7% of the human genome, SINEs account for 11.7%, LTR for 4.6% while DNA transposons account for 1.6%[12].

The repeats may have at least three important functional and evolutionary roles. First some may evolve to become the regulatory regions of genes expressed in a tissue-specific manner[13]. Secondly, repeats play an important role in refashioning the genomic architecture by facilitating homologous recombination, translocations and perhaps gene conversion[12]. These repeats have also an important role in epigenetic phenomena, such as parental imprinting[14].

REFERENCES

1. Maclean, N., 1989. Genes and Gene Regulation. Edward Arnold.
2. Wagner, R.P., M.P. Maguire and R.L. Stallings, 1993. Chromosome. A Synthesis. Wiley-Liss.
3. Bernardi, G., 1989. The isochore organisation of the human genome. Annual. Rev. Gene., 23: 637-661.
4. McKusick, V.A., 1996. History of Med. Genetics. Chapter 1 in Principles and Practise of Medical Genetics, 3rd Edn., Eds. Rimoin, D.L., Connor, J.M., Pyertiz, R.E. Edinburgh: Chirchill Livingstone, pp: 1-30.
5. Rowen, L., G. Mahairas and L. Hood, 1997. Sequencing the Human Genome. Sci., 278: 605-607.
6. McKusick, V.A., 1986. The Gene Map of Homo Sapiens: Status and Prospects. The Molecular biology of Homo sapiens. Cold Spring Harbor Symp. Quant. Biol., 51: 15-27.

7. McConkey, E.H., 1993. Human genetics: The Molecular Revolution. James and Bartlett Publishers.
8. Honig, G.R. and J.C. Adams, 1986. Human Hemoglobin genetics. Springer-Verleg.
9. Lewin, R., 1983. Genetic probes become even sharper. Sci., 221: 4616.
10. Charlesworth, B., P. Sniegowski and W. Stephan, 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. Nature, 371: 215-220.
11. SanMiguel, P., A. Tikhonov, Y.K. Jin, N. Motchoulskaia and D. Zakharov *et al*., 1996. Nested retrotransposons in the intergenic regions of the maize genomes. Science, pp: 765-768.
12. Henikoff, S., E.A. Greene, S. Pietrokovski, P. Bork, T.K. Attwood and L. Hood, 1997. Gene Families: The Taxonomy of Protein Paralogs and Chimeras. Sci., 278: 609-614.
13. Brosius, J., 1991. Retroposons-seeds of Evolution. Sci., 251: 753.
14. Barlow, D.P., 1993. Methylation and imprinting: From Host Defence to Gene Regulation. Science, pp: 309-310.