ELSEVIER

Gene 295 (2002) 311–316

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

www.elsevier.com/locate/gene

An analysis of the genome of *Ciona intestinalis*

Giuliana de Luca di Roseto[1], Giuseppe Bucciarelli, Giorgio Bernardi*

Laboratorio di Evoluzione Molecolare, Stazione Zoologica Anton Dohrn, Villa Comunale, 80121 Naples, Italy

Received 31 March 2002; accepted 18 May 2002
Received by E. Olmo

Abstract

An analysis by CsCl density gradient centrifugation has shown that, at a fragment size of about 100 kb, the DNA of a urochordate, *Ciona intestinalis*, is remarkably homogeneous in base composition. Localization of 16 coding sequences from *C. intestinalis*, chosen so as to cover the distribution range of all available coding sequences for this organism, showed a nearly symmetrical distribution almost coinciding with the DNA distribution. Both distributions are remarkably different from those found in vertebrates, which are skewed towards high GC levels (to a greater extent in warm-blooded vertebrates). In order to account for this change in genome organization, we propose a working hypothesis that can be tested. Basically, we suggest that the genome duplication that occurred between urochordates and fishes was accompanied by a preferential integration of transposons in one compartment of the genome, which was made gene-poor (by lowering gene density) compared to the rest. Since the gene-poor compartment (the 'empty quarter') is characterized by a lower level of gene expression compared to the gene-rich compartment (the 'genome core') in the vertebrate genome, we further suggest, as a working hypothesis, that a compartmentalization according to gene expression already existed in urochordates.

Keywords: *Ciona intestinalis*; Density gradient centrifugation; Genome organization; Urochordate

1. Introduction

The genomes of warm-blooded vertebrates are characterized by a large compositional heterogeneity, which is discontinuous in that these genomes are made up of isochores, long, fairly homogenous regions, belonging to a small number of families. In other words, the genomes of mammals and birds are mosaics of isochores, whose GC levels range from 30% to over 60%. Gene density is strikingly non-uniform in these genomes, GC-poor isochores being characterized by low, GC-rich isochores by high gene concentrations. Gene concentration defines, therefore, two 'gene spaces'. In the human genome, the GC-richest isochore families H2 and H3 make up a 'genome core', which represents about 12% of the genome, whereas the GC-poorest isochore families L1, L2 and H1 form an 'empty quarter', which represents the remaining 88% of the genome. The two gene spaces are characterized by a number of distinct features. In the genome core, transcription is strong, recombination is high, replication is early, chromatin structure is open, introns are short, chromosomal location is mainly distal, whereas opposite features characterize the empty quarter.

All the properties just described for the genome of warm-blooded vertebrates seem to be shared by the genomes of cold-blooded vertebrates, a major difference being, however, that GC-rich isochores are not as GC-rich, nor as abundant. This difference indicates that the ancestral genome core of cold-blooded vertebrates underwent a compositional transition which led to the very GC-rich genome core of the warm-blooded vertebrates.

These findings, summarized in recent review articles (Bernardi, 2000a,b), raise two questions, concerning the time of appearance in the evolution of chordates of the two main features which characterize the genomes of vertebrates, namely (a) the formation of the two gene spaces, and (b) the origin of the compositional heterogeneity. In order to shed light on these questions, we have investigated the genome of a urochordate, *Ciona intestinalis*.

Abbreviations: PCR, polymerase chain reaction

* Corresponding author. Tel.: +39-81-583-3215; fax: +39-81-245-5807.
E-mail address: bernardi@alpha.szn.it (G. Bernardi).
[1] Present address: Università degli Studi di Napoli 'Federico II', Dipartimento di Biochimica e Biotecnologie Mediche (DBBM), Torre Biologica, II Policlinico, 80100 Naples, Italy.

0378-1119/02/$ - see front matter
PII: S0378-1119(02)00734-5

2. Materials and methods

2.1. Animals

Adult *C. intestinalis* were collected in the bay of Naples by the fishing service of our Institute. DNA was extracted from sperms using the Genomix kit (Talent, Trieste, Italy). Molecular weights of the DNA fragments were shown to be $\geq$ 100 kb by pulsed field electrophoresis on 1% agarose gel.

2.2. Equilibrium centrifugation in CsCl density gradient

The profile of the DNA distribution in a CsCl gradient was obtained by analytical ultra-centrifugation to sedimentation equilibrium, as previously described (Thiery et al., 1976; Sabeur et al., 1993). The relationship of Schildkraut et al. (1962), $\rho = (GC \times 0.098)/100 + 1.66$, was used to convert buoyant densities into GC levels. *Bacillus subtilis* phage 2C DNA ($\rho = 1.742$ g/cm^3) was used as a density marker. The program McCurveFit (Raner, 1992) was used to find the best fitting Gaussian curve for the analytical profile of *C. intestinalis* DNA.

The shallow gradient method (De Sario et al., 1995) was used to obtain a preparative CsCl profile of *C. intestinalis*. Twenty micrograms of DNA were loaded on each gradient. Centrifugation was carried out in a vertical VTi90 rotor at 20 °C and 35,000 rpm for 24 h, using a Beckman preparative ultracentrifuge with the brake off. About 60 fractions of 80 μl each were collected using a Hitachi DGF-U instrument. DNA fractions were denatured with 0.4 M NaOH and transferred to a positively charged nylon membrane (Appligene, Pleasantown, CA), using a dot-blot apparatus (BioRad, Richmond, CA). The membrane containing 100 μl of each fraction was hybridized with a coding sequence probe that was radiolabeled by the random oligo primer method. The Church and Gilbert (1984) solution was used in hybridization experiments. Filters were analysed with a PhosphorImager using the Image Quantification program. The percentages of hybridization signals were plotted against the corresponding fractions.

2.3. Distribution of genes in C. intestinalis

Hybridization experiments for each gene were done several times using different gradients. In order to compare the results, we fitted each shallow gradient profile with a Gaussian curve, so normalizing these profiles with DNA distribution as obtained by analytical ultracentrifugation. Using this approach, we could assess the GC level for all fractions of different gradients and localize all genes in the analytical profile.

2.4. Probes

Sequences used for hybridization experiments were single-copy coding sequences from *C. intestinalis*. The majority of genes were cloned in the restriction sites *Eco*RI–*Xho*I of pBluescript II SK$^+$ plasmid vector. This was the case of Hox3 (Locascio et al., 1999), Hox5 (Gionti et al., 1998), cam (Di Gregorio et al., 1998), msx-b (Aniello et al., 1999), rgc (Piscopo et al., 2000), rgg1 (Tanaka et al., 2000), trop (Di Gregorio and Levine, 1999), Su(H) (Corbo et al., 1998). On the other hand, titf1 CDS (Ristoratore et al., 1999) was cloned in *Bam*HI and *Xho*I sites; the z subunit of proteasome (Marino et al., 1999) was cloned in *Hin*dIII–*Not*I site and Tgasi CDS (Cariello et al., 1997) in *Not*I. In the case of β-catenin and cadherin (Imai et al., 2000), snail (Fujiwara et al., 1998) and distalless A (Caracciolo et al., 2000), we have used internal primers to amplify by PCR the fragments of interest ranging from 0.8 to 2kb.

We have also used as probes cDNA of *C. intestinalis* obtained at mobile larval stage and 18S rDNA fragment obtained from PCR on the basis of *C. intestinalis* 18S partial sequence in GenBank (Wada, 1998). The cDNA was obtained from mRNA poly(A$^+$) (extracted with Fast Track kit, Invitrogen) and copied into cDNA using Invitrogen Copy Kit.

2.5. Sequence data analyses

Complete sequences from GenBank (Release 126; 15 October, 2001) were processed using the ACNUC retrieval system (Gouy et al., 1985) and cleaned of redundancy; the program ANALSEQ (Gautier and Jacobzon, 1989) was used to determine the base composition of coding sequences (CDS). In particular, the histogram of GC_3 levels (percentages) was constructed by partitioning the values into bins of 5%, GC_3.

2.6. Orthogonal regression analyses

Regression lines were calculated by the orthogonal (major axis) regression formula $y\text{-}\langle y\rangle = s(x - \langle x\rangle)$, $s = (b + \sqrt{b^2 + 4c^2})/(2c)$, in which x and y are GC and GC_3, respectively, s is the slope, $\langle x\rangle$ and $\langle y\rangle$ are the coordinates of the centre of scatterplot mass, i.e. the average GC and GC_3 levels, $b = v_y - v_x$ is the difference of the sample variances, and c is the covariance. With no errors in the individual data points, the confidence intervals for the slopes would be around $s \pm 0.3$, as calculated using the formula of Jolicoeur (1990).

3. Results

3.1. Compositional distribution of large DNA fragments

The CsCl analytical profile of *C. intestinalis* DNA was characterized by a main peak with a modal buoyant density of 1.6945 g/cm^3 corresponding to 35.2% GC and by two minor peaks characterized by a lower and a higher buoyant densities (Fig. 1). The lighter peak probably is mitochondrial DNA, the heavier peak was demonstrated to correspond to ribosomal DNA by hybridization of an

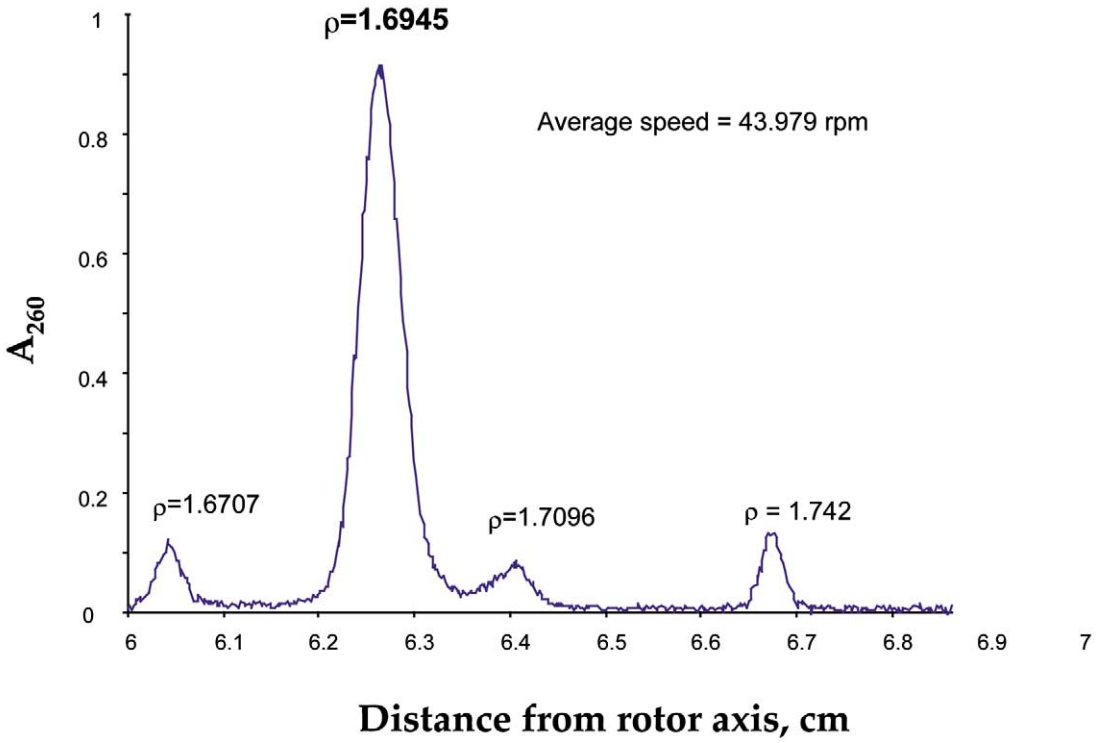


Fig. 1. Profile of *Ciona intestinalis* DNA as obtained by analytical ultracentrifugation to sedimentation equilibrium in a CsCl gradient. The lighter peak probably is mitochondrial DNA, the heavier peak is ribosomal DNA.

appropriate probe on a preparative CsCl gradient (not shown). The profile of the main DNA band of *C. intestinalis* covers a narrow range. The heterogeneity (measured as standard deviation) of this CsCl profile, was estimated as 2.4 by fitting a Gaussian curve (Fig. 2).

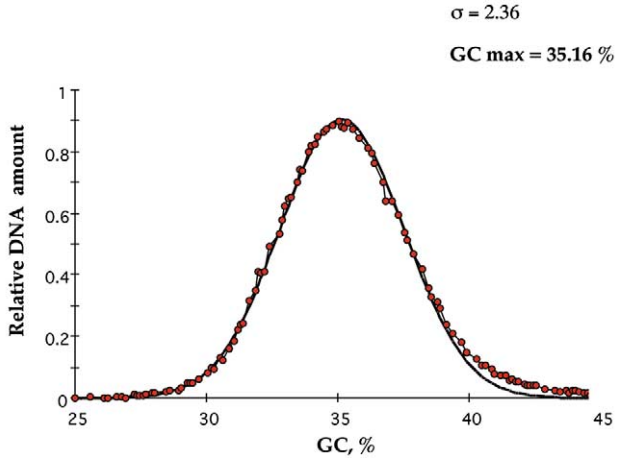


Fig. 2. The analytical profile of the main-band DNA from *Ciona intestinalis* (curve) is fitted with a Gaussian curve (circles) using the MacCurveFit program. The GC scale was obtained from the buoyant density values using the relation of Schildkraut et al. (1962). The standard deviation (the sum of standard deviations due to composition and to Brownian diffusion) value is 2.36.

3.2. Analysis of coding sequences for C. intestinalis

Fig. 3 shows the total GC and the GC_3 levels of the 105 coding sequences available at present in GenBank. The total GC range is 37.4–59.2%, while the GC_3 range is 28.1–58.1%. On this basis, we chose as probes 16 coding sequences for hybridization experiments, covering a GC range of 42.5–52.2%, and a GC_3 range of 28.1% to about 54%.

3.3. Distribution of genes in the C. intestinalis genome

DNA fractions from shallow gradient (Fig. 4) were transferred to hybridization filters, using the dot-blot method for the analysis of gene distribution.

Fig. 5 shows a plot of GC_3 levels of the 16 coding sequences analysed against the GC levels of the corresponding long sequences (GC fractions) in which the coding sequences were localized. The regression equation was used to position the GC_3 distribution relative to the CsCl profile (Fig. 6). The 'regional' gene concentration was then calculated by dividing each bar value of the GC_3 histogram (bins of 5% width) by the corresponding value on the CsCl profile. This approach (Zoubak et al., 1996) estimates gene concentrations across the compositional range of the genome and shows that gene density in *C. intestinalis* is only barely shifted (by 1% in GC) from the DNA peak obtained by analytical ultracentrifugation (Fig. 7).

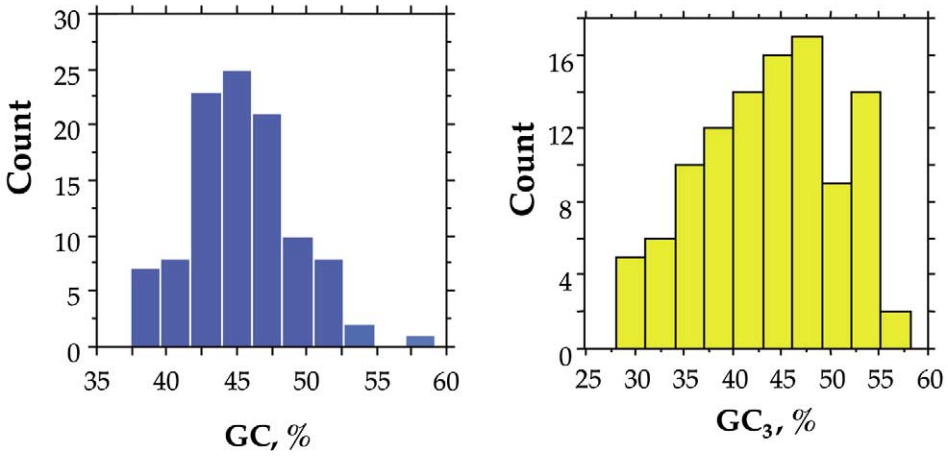


Fig. 3. Frequency histogram of all coding sequences available in GenBank as plotted against GC levels (left) or GC_3 levels (right).

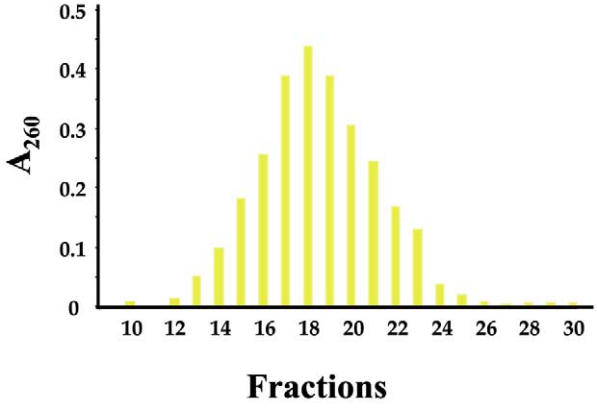
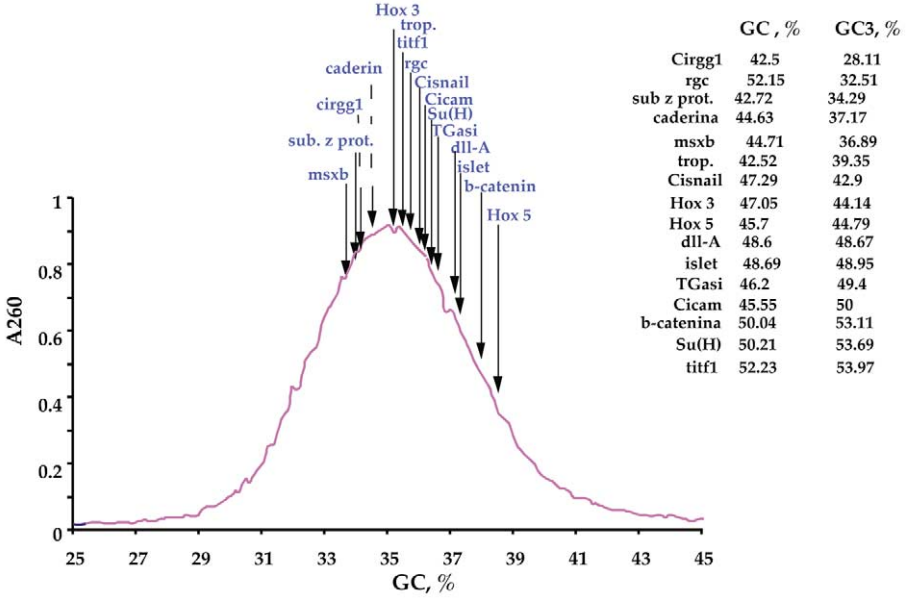


Fig. 4. (A) DNA profile of *Ciona intestinalis* using the shallow gradient method. (B) For the sake of comparison, the profile of *Brachydanio rerio* DNA is also displayed.

As an alternative approach to analyse the gene distribution, we also used the cDNA of *C. intestinalis* at mobile larval stage as a probe for hybridization (Fig. 8). This experiment also showed a slight shift of the hybridization peak towards GC-rich fractions.

4. Discussion

The results obtained in the present work demonstrate two important points: (1) at fragment size of about 100 kb, the genome of *C. intestinalis* is remarkably homogeneous in base composition; in fact, it is more homogenous than many bacterial DNAs; (2) the distribution of genes in the genome also is remarkably homogeneous. In both properties, the genome of *C. intestinalis* is very different from those of vertebrates, ranging from fishes to mammals and



	GC, %	GC3, %
Cirgg1	42.5	28.11
rgc	52.15	32.51
sub z prot.	42.72	34.29
caderina	44.63	37.17
msxb	44.71	36.89
trop.	42.52	39.35
Cisnail	47.29	42.9
Hox 3	47.05	44.14
Hox 5	45.7	44.79
dll-A	48.6	48.67
islet	48.69	48.95
TGasi	46.2	49.4
Cicam	45.55	50
b-catenina	50.04	53.11
Su(H)	50.21	53.69
titf1	52.23	53.97

Fig. 5. Localization of the 16 coding sequences used as probes for hybridization experiments on the DNA fragments in which these CDS were embedded. GC and GC_3 levels of the 16 coding sequences are also shown.

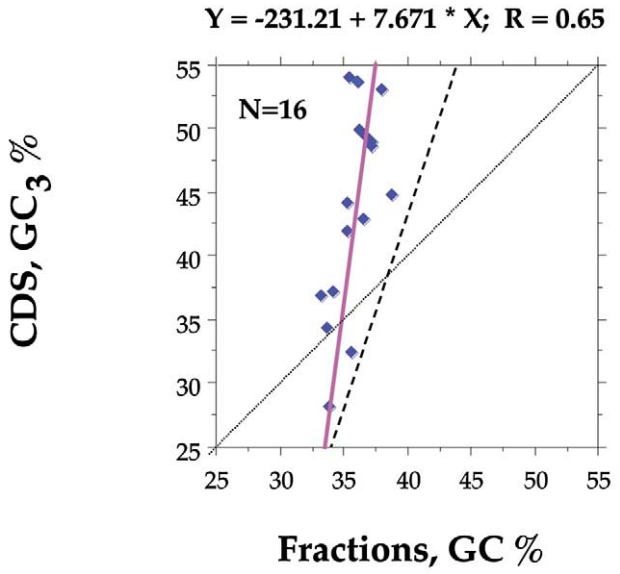


Fig. 6. Correlation between GC_3 levels of *Ciona intestinalis* coding sequences used for hybridization experiments ($n = 16$) and the GC levels of the DNA fractions in which genes were localized. The range of angles corresponding to a 5% confidence interval as calculated according to Jolicoeur (1990) is from 80.4 to 84.7°. The slope corresponds to an angle of 82.6°. The corresponding plot for human coding sequences and the diagonal are also shown (broken line).

birds, since the latter are characterized by a heterogeneity and an asymmetry in the distribution of both DNA molecules and genes, these features being much more pronounced in warm-blooded vertebrates. This raised the question as to which process led from the features of the urochordate genome to those of vertebrates. We present here a speculative working hypothesis which can be tested.

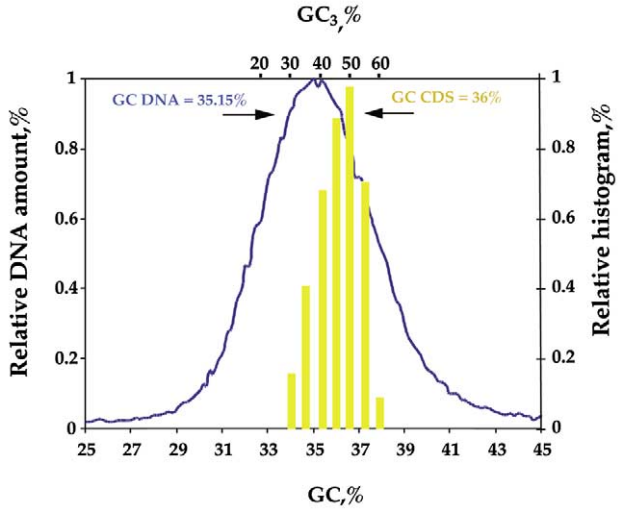


Fig. 7. Profile of gene concentration in *Ciona intestinalis* genome, as obtained by dividing the relative numbers of genes in each 5% GC_3 interval of the histogram of gene distribution (yellow bars) by the corresponding relative amount of DNA deduced from the CsCl profile (blue line).The positioning of the GC_3 histogram relative to the CsCl profile is based on the correlation of Fig. 6.

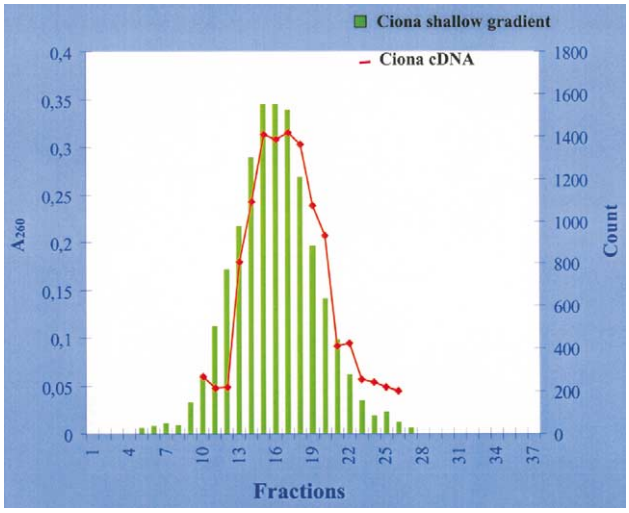


Fig. 8. Profile of gene distribution obtained using *Ciona intestinalis* cDNA at mobile larval stage as probe for hybridization experiment (red curve) The histogram shows the total DNA distribution obtained from shallow gradient method. In this case on the gradient 20 μg of DNA were loaded.

The very first consideration here is to take into account that genome size is 180 Mb in *C. intestinalis* and about 400 Mb in fishes with the smallest genomes, like *Arothron diadematus* (Pizon et al., 1984) and other fishes of the order *Tetraodontiformes* (*Fugu rubripes*, *Tetraodon viridiformis*). Since in the case of *A. diadematus*, the amounts of rapidly re-associating and intermediate re-associating sequences are very small (6 and 7%, respectively), in fact the smallest reported so far for a vertebrate genome, this is a strong indication of a genome duplication in the ancestral line leading to fishes, in agreement with the original proposal by Ohno (1970) and current ideas. This genome duplication was accompanied by an increase of intergenic sequences, as indicated by the fact that, neglecting cases of further polyploidizations, the 'average' genome size of fishes is around 1000 Mb (Bernardi and Bernardi, 1990a,b; Bucciarelli et al., 2002). This suggests that one possible and, at present, the most likely mechanism for the formation of two compartments in the vertebrate genome was that insertions of transposons in intergenic sequences (and introns) took preferentially place in one part of the genome, which was made gene-poor compared to the rest. Since in the vertebrate genome the gene-poor compartment, the 'empty quarter', is characterized by a lower level of gene expression compared to the gene-rich compartment, the 'genome core', as suggested by Bernardi and demonstrated by D'Onofrio (2002), we further suggest that two compartments characterized by different levels of gene expression already existed in urochordates. This working hypothesis is presented in the scheme of Fig. 9. The current investigations on genome sequence and expression in Ascidians should allow this point to be verified soon. Needless to say, our results do not pin down the precise evolutionary time of such changes in genome organization. Investigations along the same line on the genomes of *Cephalochordates* and *Agnathans* should provide this information.

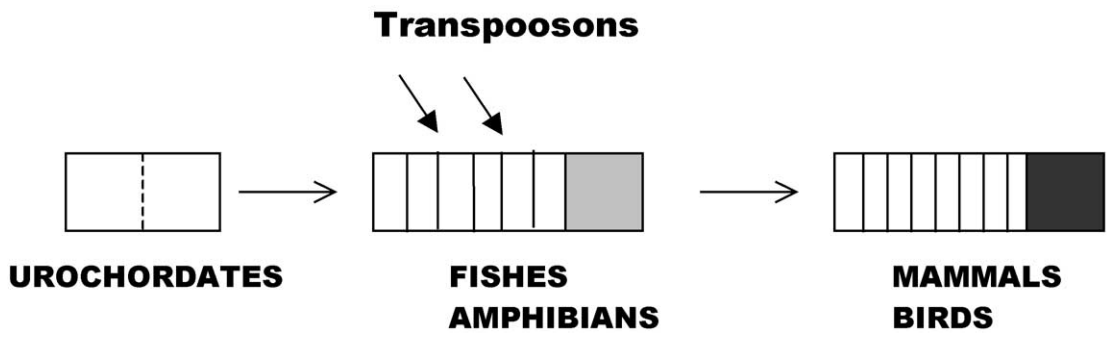


Fig. 9. A scheme of genome evolution from urochordates (where two components with different expression levels are supposed to have already existed) to fishes (where the ancestral 'empty quarter' increased in size, with a consequent decrease of gene concentration, and the 'genome core' slightly increased in GC level) to mammals and birds (where the 'genome core' underwent a further stronger increase in GC).

References

Aniello, F., Locascio, A., Villani, M.G., Di Gregorio, A., Fucci, L., Branno, M., 1999. Identification and developmental expression of Cimsx-b: a novel homologue of Drosophila msh gene in *Ciona intestinalis*. Mech. Dev. 88, 123–126.

Bernardi, G., 2000a. Isochores and the evolutionary genomics of vertebrates. Gene 241, 3–17.

Bernardi, G., 2000b. The compositional evolution of vertebrate genomes. Gene 259, 31–43.

Bernardi, G., Bernardi, G., 1990a. Compositional patterns in the nuclear genomes of cold-blooded vertebrates. J. Mol. Evol. 31, 265–281.

Bernardi, G., Bernardi, G., 1990b. Compositional transitions in the nuclear genomes of cold-blooded vertebrates. J. Mol. Evol. 31, 282–293.

Bucciarelli, G., Bernardi, G., Bernardi, G., 2002. An ultracentrifugation analysis of two hundred fish genomes. Gene 295, 153–162.

Caracciolo, A., Di Gregorio, A., Aniello, F., Di Lauro, R., Branno, M., 2000. Identification and developmental expression of three Distal-less homeobox containing genes in the ascidian *Ciona intestinalis*. Mech. Dev. 99 (1–2), 173–176.

Cariello, L., Ristoratore, F., Zanetti, L., 1997. A new transglutaminase-like from the ascidian *Ciona intestinalis*. FEBS Lett. 408, 171–176.

Church, G.M., Gilbert, W., 1984. Genomic sequencing, 1984. Proc. Natl. Acad. Sci, USA 81 (7), 1991–1995.

Corbo, J.C., Fujiwara, S., Levine, M., Di Gregorio, A., 1998. Suppressor of hairless activates brachyury expression in the Ciona embryo. Dev Biol. 203 (2), 358–368.

De Sario, A., Geigl, E.M., Bernardi, G., 1995. A rapid procedure for the compositional analysis of yeast artificial chromosomes. Nucleic Acids Res. 23, 4013–4014.

Di Gregorio, A., Levine, M., 1999. Regulation of Ci-tropomyosin-like, a Brachyury target gene in the ascidian, *Ciona intestinalis*. Development 126 (24), 5599–5609.

Di Gregorio, A., Villani, M.G., Locascio, A., Ristoratore, F., Aniello, F., Branno, M., 1998. Developmental regulation and tissue specific localization of calmodulin mRNA in the protochordate *Ciona intestinalis*. Dev. Growth Differ. 40, 387–394.

D'Onofrio, G., 2002. Expression patterns and gene distribution in the human genome. Gene in press.

Fujiwara, S., Corbo, J.C., Levine, M., 1998. The snail repressor establishes a muscle/notochord boundary in the Ciona embryo. Development 125 (13), 2511–2520.

Gionti, M., Ristoratore, F., Di Gregorio, A., Aniello, F., Branno, M., Di Lauro, R., 1998. Cihox5, a new *Ciona intestinalis* Hox related gene is involved in regionalization of the spinal cord. Dev. Genes Evol. 207, 515–523.

Gautier, C., Jacobzon, M., 1989. Publication interne, UMR CNRS 5558 Biometrie. Génétique et Biologie des Population, Université Claude Bernard Lyon-I,, Lyon, France.

Gouy, M., Gautier, C., Attimonelli, N., Lanave, C., Di Paola, G., 1985. ACNUC Portable retrieval system for nucleic acid sequence database: logical and physical design and usage. Comput. Appl. Biosci. 1, 167–172.

Imai, K., Takada, N., Satoh, N., Satou, Y., 2000. (beta)-Catenin mediates the specification of endoderm cells in ascidian embryos. Development 127 (14), 3009–3020.

Jolicoeur, P., 1990. Bivariate allometry: Interval estimation of the slopes of the ordinary and standardized normal major axes and structural relationship. J. Theor. Biol. 144, 273–285.

Locascio, A., Aniello, F., Amoroso, A., Manzanares, M., Krunhauf, R., Branno, M., 1999. Patterning the ascidian nervous system structure, expression and transgenic analysis of the CiHox-3 gene. Development 126, 4734–4748.

Marino, R., De Santis, R., Giuliano, P., Pinto, M.R., 1999. Follicle cell proteasome activity and acid extract from the cgf vitelline coat prompt the outset of self-sterility in *Ciona intestinalis* oocytes. Proc. Natl. Acad. Sci. USA 96, 9633–9636.

Ohno, S., 1970. Evolution by Gene Duplication, Springer, Berlin.

Piscopo, A., Branno, M., Aniello, F., Corrado, M., Piscopo, M., Fucci, L., 2000. Isolation and characterization of the cDNA for a *Ciona intestinalis* RNA binding protein: spatial and temporal expression during development. Differentiation. 66 (1), 23–30.

Pizon, V., Cuny, G., Bernardi, G., 1984. Nucleotide sequence organization in the very small genome of a tetraodontid fish, *Atothron diadematus*. Eur. J. Biochem. 140 (1), 25–30.

Raner, K., 1992. MacTutor Magazine 8 (3), 24.

Ristoratore, F., Spagnuolo, A., Aniello, F., Branno, M., Fabbrini, F., Di Lauro, R., 1999. Expression and functional analysis of Cititf1, an ascidian NK-2 class gene, suggest its role in endoderm development. Development. 126, 5149–5159.

Schildkraut, C.L., Marmur, J., Doty, P., 1962. Determination of the base composition of deoxyribonucleic acid from its buoyant density in CsC1. J. Mol. Biol. 4, 430–443.

Sabeur, G., Macaya, G., Kadi, F., Bernardi, G., 1993. The isochore patterns of mammalian genomes and their phylogenetic implications. J. Mol. Evol. 37, 93–108.

Tanaka, K.J., Kawamura, H., Nishikata, T., 2000. The transcript coding for an RNA-binding protein is localized in the anterior side of the ascidian 2-cell stage embryo. Dev. Genes Evol. 210 (8–9), 464–466.

Thiery, J.P., Macaya, G., Bernardi, G., 1976. An analysis of eukaryotic genomes by density gradient centrifugation. J. Mol. Biol. 108, 219–235.

Wada, H., 1998. Evolutionary history of free-swimming and sessile lifestyles in urochordates as deduced from 18S rDNA molecular phylogeny. Mol. Biol. Evol. 15 (9), 1189–1194.

Zoubak, S., Clay, O., Bernardi, G., 1996. The gene distribution of the human genome. Gene 174, 95–102.