J Mol Evol (1989) 28:469–486

Journal of
Molecular Evolution

Evolution of Chromosome Bands: Molecular Ecology of Noncoding DNA*

Gerald P. Holmquist

Beckman Research Institute of the City of Hope, Department of Biology, 1450 E. Duarte Rd., Duarte, California 91010, USA

Summary. Giemsa dark bands, G-bands, are a derived chromatin character that evolved along the chromosomes of early chordates. They are facultative heterochromatin reflecting acquisition of a late replication mechanism to repress tissue-specific genes. Subsequently, R-bands, the primitive chromatin state, became directionally GC rich as evidenced by Q-banding of mammalian and avian chromosomes. Contrary to predictions from the neutral mutation theory, noncoding DNA is positionally constrained along the banding pattern with short interspersed repeats in R-bands and long interspersed repeats in G-bands. Chromosomes seem dynamically stable: the banding pattern and gene arrangement along several human and murine autosomes has remained constant for 100 million years, whereas much of the noncoding DNA, especially retroposons, has changed. Several coding sequence attributes and probably mutation rates are determined more by where a gene lives than by what it does. R-band exons in homeotherms but not G-band exons have directionally acquired GC-rich wobble bases and the corresponding codon usage: CpG islands in mammals are specific to R-band exons, exons not facultatively heterochromatinized, and are independent of the tissue expression pattern of the gene. The dynamic organization of noncoding DNA suggests a feedback loop that could influence codon usage and stabilize the chromosome's chromatin pattern: DNA sequences determine affinities of → proteins that together form → a chromatin that modulates → rate constants for DNA modification that determine → DNA sequences. Theories of hierarchical selection and molecular ecology show how selection can act on Darwinian units of noncoding DNA at the genome level thus creating positionally constrained DNA and contributing minimal genetic load at the individual level.

Key words: Hierarchical selection — Chromosome bands — Base composition isochores — Natural selection — Replication time — Retroposons — Neutral mutation theory

Introduction

Each mammalian chromosome is a mosaic of compartments or bands that map as contiguous domains onto the chromosomal DNA molecule. Compartments are visualized as bands on metaphase chromosomes, as units of chromosome replication (Ganner and Evans 1971; Latt 1975; Lau and Arrighi 1981), as eight or more contiguous replicons that initiate replication in unison (Hand 1978; Dubey and Raman 1987), or as very long, >300 kb, DNA segments (isochores) of strikingly homogeneous base composition (Bernardi et al. 1985). Each chromosome compartment differs from its nearest neighbor compartments in GC content and replication time. Staining techniques distinguish three classes of compartments called C-bands, G-bands (Giemsa dark, quinacrine bright, AT-rich, late-replicating bands), and R-bands (Giemsa pale, quinacrine dull, GC-rich, early-replicating bands) (Fig. 1). C-bands, constitutive (centric) heterochromatin, contain tandemly repetitious satellite DNA and few or no genes. This review concerns differences (Table 1) between the 2000 different (Yunis 1981) euchromatic bands, R-bands, and G-bands that average 1300 kb of DNA per band.

Offprint requests to: G.P. Holmquist

* Presented in part at the FEBS Symposium on Genome Organization and Evolution, held in Crete, Greece, September 1–5, 1986

Fig. 1. Properties and stability of banding patterns. Patterns along human (-H) chromosome 1, HSA 1, revealed as follows: **C-H**, C-banding of centric heterochromatin; **L-H**, in situ hybridization using the long repeat sequence L1 (Manuelidis and Ward 1984); **A-H**, in situ hybridization using the short repeat *Alu* (Manuelidis and Ward 1984); **Q-H**, Q-banding where AT-rich DNA is quinacrine bright; **G-H**, trypsin-Giemsa banding where G-bands are dark and R-bands are light; **R-H**, replication banding where early-replicating R-band DNA stains darkly (von Kiel et al. 1985). The C-band that varies in size from individual to individual lacks L1 and *Alu* and is G^+, Q^-. There are several AT-rich isochore fractions, some AT richer than others, but all corresponding to G-band DNA; these fractions have an L1 or [1/*Alu*] content that increases with the AT-richness of the isochore (Soriano et al. 1983). Thus, those G-bands that are AT richer than other G-bands are also L1 richer. Consequently, the $[Alu]^{-1}$ and L1 concentration distributions more closely parallel Q-banding intensity than they parallel G-banding intensity. This explains why Q bright bands are almost always Giemsa dark but the degree of Q brightness does not parallel degree of G darkness. **G-A**, G-banding of cheetah chromosome AJU C1 (O'Brien et al. 1984); **R-F**, replication banding of domestic cat chromosome FCA C1 (von Kiel et al. 1985); **R-M**, **R-H**, replication banding of mouse chromosome MMU 4, and HSA 1 (von Kiel et al. 1985). The two cat C1p arms (short arms at top) band almost identically to human 1p and, with the exception of about 15% centric heterochromatin, the mouse MMU 4 does likewise; these chromosome arms from three different mammalian orders contain the same genes in the same linear array (O'Brien et al. 1984; von Kiel et al. 1985; Sawyer and Hozier 1986) and show that autosomal banding patterns can be conserved. We thank Laura Manuelidis (Manuelidis and Ward 1984), Horst Hameister (von Kiel et al. 1985), Charleen Moore, and Stephen J. O'Brien (O'Brien et al. 1984) for prints and permission to publish them.

Noncoding DNA constitutes up to 98% of G- and R-band DNA but its function or lack thereof remains unknown (Ohno 1972; Doolittle and Sapienza 1980; Orgel and Crick 1980). This review shows that noncoding DNA is highly organized around an evolutionarily stable banding pattern. Although it constitutes the bulk of this pattern, it still turns over relatively quickly. This dynamic organization of noncoding DNA suggests its function.

One function of G-bands involves facultative heterochromatinization of their tissue-specific genes. Late replication is associated with permanent transcriptional inactivity of constitutive heterochromatin or facultative inactivity of whole chromosomes such as the inactive X in females. Recent data indicate that individual replicons in G-bands are facultatively repressed like the inactive X-chromosome in mammals (Goldman et al. 1984; Holmquist 1987a; Goldman 1988). Unlike *Drosophila*, which neither has clustered replicons (Stineman 1981) nor seems to include late replication in its repertoire of common repression mechanisms (Holmquist 1987a), vertebrates have most of their tissue-specific genes in late-replicating G-bands. In G-bands, all 10 of the genes assessed were repressed and replicated late in most tissues; however, in cell types where these 10 genes were expressed, they replicated early and their surrounding conditionally early-replicating domain included proximal genes but was probably limited to the gene's residence replicon (Goldman et al. 1984; Holmquist 1987a; Goldman 1988; Hatton et al. 1988a,b). Housekeeping genes, constitutively active or activatable genes, are in R-bands where they almost always replicate early (Goldman et al. 1984; Holmquist 1987a; Goldman 1988; Hatton et al. 1988a,b). We initially called the class of genomic compartments that constitute the R-bands "the housekeeping subgenome" and G-bands "the ontogenic subgenome" (Goldman et al. 1984) because these compartment classes initially appeared functionally different (Goldman et al. 1984).

Genes in Bands

Differences in replication time and base composition allow the identification of R- and G-bands by replication banding (Ganner and Evans 1971; Latt 1975; Lau and Arrighi 1981) and quinacrine banding (Ganner and Evans 1971; Latt et al. 1974; Comings et al. 1975; Latt 1975; Lau and Arrighi 1981) (Fig. 1), and also allow physical separation of DNA from the two band classes. Replication time fractionation is based on bands being 1300 kb units of chromosomal replication (Holmquist 1987a, 1988a) and on the fact that 500+-kb-long segments of early-replicating DNA do not contain regions of late-replicating DNA (Kowalski and Cheevers 1976). During S-phase, R-bands replicate early and finish synthesis before G-bands initiate synthesis (Camargo and Cervenka 1982). Using synchronized cells labeled with the thymidine analog BrdUrd during early S-phase, the early-replicating BrdUrd-substituted DNA is denser and can be separated from the late-replicating unsubstituted half of the genome (Holmquist et al. 1982). Base composition fractionation is based on a different property of warm-blooded vertebrate's chromosome bands. Each band is composed of a long stretch of DNA of rather homogeneous base composition. When these long stretches or "isochores" are base composition-fractionated in ligand-assisted CsCl density gradients,

Table 1. Properties of bands and their DNA sequences

R-bands	G-bands
Staining sensitive to protein denaturation[a]	More resistant
DNase II and Hae III sensitive[b]	More resistant
Open to DNase I nick translation[c]	Closed
Quinacrine dull bands[d]	Quinacrine bright bands
dG + dC-rich DNA	dA + dT-rich DNA
Early replicating	Late replicating
Pachytene interchromomeres[e]	Pachytene chromomeres
Chiasmatic in meiosis I[f]	Achasmatic
Crossing over frequent[g]	Infrequent
Synaptic initiation[h]	
Homologous synapsis[i]	Homology independent
Abortions frequent in trisomic state[j]	More viable trisomics
High mitotic chiasmata frequency[k]	Low
High amplification rate[l]	Low amplification rate
Housekeeping genes (all)[m]	Tissue specific (most)
Upstream GGGCGG[m]	Rare Sp1 consensus sequence
CCGG-rich[n]	CCGG poor
CpG-rich islands in genes[o]	No or very few islands
Slow-sequence divergence rate[p]	Faster rate
dG + dC-rich third coding base[q]	dA + dT-rich third coding base
SINE-rich[r]	LINE-rich[r]
Short period interspersion pattern[s]	Mixed interspersion pattern
Alu[t]	L1 (human)[t]
B1 (mouse, hamster)[u]	L1 (rodent)[v]
Hepatitis B virus[w]	A36Fc repeat[x]
Bovine leukemia virus[y]	Mouse mammary tumor virus[z]
	Epstein-Barr virus[aa]
Shared SINEs[bb]	
Processed pseudogenes[cc]	
SV40[dd]	

The first four properties are for euchromatin in acid–alcohol-fixed chromosomes. G-, Q-, and replication banding patterns are congruent for most bands (Ganner and Evans 1971; Latt 1975; Comings 1978; Lau and Arrighi 1981; Schweizer 1981; Camargo and Cervenka 1982; Holmquist et al. 1982)

[a] Trypsin, chymotrypsin, strong detergents, or hot saline cause R-band DNA to become inaccessible to the cationic Giemsa dyes at a faster rate than for G-band DNA (Comings and Avelino 1975; Comings 1978; Holmquist 1987b)

[b] DNase II or micrococcyl nuclease digestion followed by Giemsa staining (Sahasrabbuddhe et al. 1978). Of the restriction endonucleases, only Hae III gives a reasonable G-banding pattern (Bianchi et al. 1985) with avian or mammalian but not anuran chromosomes (Schmid and de Almeida 1988) and is consistent with the frogs' lack of GC-rich R-bands

[c] DNase I digestion followed by nick translation with biotinylated precursors. Because the inactive X-chromosome in females undergoes trypsin G-banding like the active-X but is almost totally refractory to nick translation, nick translation banding is believed to reflect gene activity (Kerem et al. 1984)

[d] Base composition-sensitive fluorochromes produce the same or reciprocal patterns of bright fluorescence depending on their base specificity (Schweizer 1981)

[e] Meiotic chromomeres correspond to G-bands (Okada and Comings 1974; Fang and Jagiello 1988) and the interchromomeric DNA stains bright with chromomycin A_3–distamycin as do mitotic R-bands (Ambros and Sumner 1987). We do note that species without G-bands have meiotic chromomeres

[f] Fang and Jagiello (1988)

[g] Ashley (1988)

[h] Chandley (1986) in theory, with evidence (Ashley, in press)

[i] Ashley and Russell (1986), Ashley (1988)

[j] Hoehn (1975), Kuhn et al. (1985)

[k] Kuhn et al. (1985)

[l] Very early-replicating genes are most often amplified (Schimke et al. 1986)

[m] Table 2

[n] When digested with MspI, the number average molecular weight of early- and late-replicating hamster DNA is 2535 and 4330 bp, respectively (Holmquist and Caston 1986; Holmquist 1988a)

[o] Table 2

[p] Early-replicating genes (Table 2) and early-replicating interspersed repeats (Holmquist and Caston 1986) cross-hybridize rodent–human much more frequently than do their late-replicating counterparts. Genes that map to R-bands have a high GC content in their coding sequences (Ikemura and Aota 1988); genes with a high GC content in their coding sequences have a low silent substitution rate (Filipski 1988a,b)

[q] Bernardi et al. (1985), Aota and Ikemura (1986), Bernardi and Bernardi (1986a,b)

[r] Holmquist and Caston (1986)

[s] From fractionation of human DNA according to % renaturation (Strayer et al. 1983) and as expected from the LINE-depleted property of R-band DNA (Holmquist 1988a)

[t] Soriano et al (1983), Manuelidis and Ward (1984), Holmquist and Caston (1986), Korenberg and Rykowski (1988)

[u] Soriano et al. (1983), Holmquist and Caston (1986)

[v] Soriano et al. (1983), Vizard and Rosenberg (1984)

[w] Bernardi et al. (1985)

[x] A long interspersed repeat family of low copy number in humans (Holmquist and Caston 1986)

[y] Kettmann et al. (1979), Soriano et al. (1981)

[z] Salinas et al. (1987)

[aa] By in situ hybridization of ^{3}H-DNA to chromosomes (Shiraishi et al. 1985)

[bb] Some SINEs show no band preference (Holmquist and Caston 1986)

[cc] Goldman et al. (1984)

[dd] In three different cell lines containing one copy each of SV40 DNA, the replication time of SV40 DNA was early, rather early, and late (Marchionni and Roufa 1981)

they yield up to four different AT-rich isochore fractions, the AT-rich half of the genome, and up to four different GC-rich isochore fractions (Thiery et al. 1976; Cuny et al. 1981; Bernardi et al. 1985; Filipski et al. 1987; Salinas et al. 1987). Both band DNA fractionation schemes yield comparable results in that GC-rich or early-replicating fractions, henceforth both called R-band fractions, both contain about 85+% of the human *Alu* (ca. 500,000+ members and 5% of human DNA) or rodent B1 (ca. 100,000 members) repeat family (Table 1), whereas AT-rich or late-replicating DNA fractions, G-band fractions, contain most of the human or murine L1 repeat family (ca. 50,000 members) (Table 1). That *Alu* and L1 sequences are concentrated in R- and G-bands, respectively, was even visualized by in situ hybridization (Fig. 1). Table 2, an arranged list of

Table 2. Properties of genes

	1	2	3	4	5	6	7
				R-bandlike genes			
Hypoxanthine phosphoribosyl transferase	H	E		I	–	c	++
Adenine phosphoribosyl transferase	H	E		I	–	c	––
CAD	H	E					++
Dihydrofolate reductase	H	E		I	–	c	–+
Arginine succinate synthetase	H	E		I	t	c	–+
Glucose-6-phosphate dehydrogenase	H	E		I			–
Beta-tubulins	H	E		I	–	c	++
Phosphoglycerate kinase-1	H	E		I	–	c	++
Tyrosine amino transferase	H	E		I			++
Beta-actin	H	E		I	t	–	++
Gamma-actin	H	E					++
Metallothionein-I	H	E		I	t	c	
Superoxide dismutase	H			I	t	c	
Heat shock protein	H			I	t	c	
HMG CoA reductase	H			I	–	c	
Glyceraldehyde-3-phosphate dehydrogenase	H			I			
Triose phosphate isomerase	H			I	t	c	
Adenosine deaminase	H			I	–	c	
Histocompatibility class I	H			I	t		
c-myc	H	E	G	I	t	c	++
c-Ha-ras	H	E	G	I	–	c	++
c-abl	?	E	G				
c-src	H			I			
c-Ki-ras	H	E		I			
c-fos	H	E		I	t	c	
c-int	?			I			
c-raf	H	E					
All histones	H	E		I	t		+
Ribosomal proteins L-30, L-32	H			I	–	–	
Xenopus somatic 5s RNA	H	E					
Alpha-globin	T	E	G	I	t	c	–
c-sis	T	E	G	I	–	c	
c-myb, c-fes, c-rel	T	E					
c-mos	T	E	G				
c-fms	T	E					
Apolipoprotein A-I	T	E		3′	t	c	
Histocompatibility class II	T	E		3′	t		
Steroid 21 hydroxylase	T	E					
Pre-pro-opiomelanocortin	T		G	I	t		
Conalbumin (chicken)	T		G				
Collagen II	T			I	t	c	
Retinol binding protein	T			I	t	c	
Thy-1	T	E		I	t		
Enkephalin	T			I	t	–	
Vimentin (human)	T		G				+
				??-band			
Vimentin (rodent)	T		A	I	t	–	+
Placental lactogen	T	E		N	t		–
Skeletal muscle actin	T	L	G	3′	t	c	++
Immunoglobulin constant	T	E	A				
				G-bandlike genes			
Beta-globin	T	L	A	N	t	–	–
Immunoglobulin variable	T	L	A				
Gamma, delta-, epsilon-globins	T		A	N	t	–	
Ovalbumin (chicken)	T		A	N	t	–	
Cardiac actin	T		A				
Alpha-1-antichymotrypsin	T	L					–
Alpha-1-antitrypsin	T	L		N	t	–	–
Beta-casein	T	L					––
Phenylalanine hydroxylase	T	L					–+
Factor IX	T	L		N	t		
Fibronectin	T	L					

Table 2. Continued

	1	2	3	4	5	6	7
Myosin heavy chain cardiac alpha	T	L		N	t	–	
N-myc	T	L		N			
Alpha-amylase I	T	L		N			
Albumin	T	L					
Alpha-fetoprotein	T	L		N			
Major urinary proteins	T	L					
T-cell beta-chain constant and variable	T	L					
Nerve growth factor	T			N			
Insulin	T			N	t	c	
Interferons	T			N	t	–	
Fibrinogen alpha, gamma	T			N	t	c	
Fibrinogen beta	T			N	t	–	
Delta-crystallin	T			N			
Growth hormone	T			N	t		
Myoglobin	T			N	t	–	
Interleukin 2	T			N	t	–	
Xenopus oocyte 5s RNA	T	L					

Data are from human, or if unavailable, from mouse or hamster unless specified.

(1) Expression patterns: (H) housekeeping, or (T) tissue specific. (2) Replicaton times: early (E) or late (L) (Calza et al. 1984; Goldman et al. 1984; Iqbal et al. 1984, 1987; Holmquist 1987a; Goldman 1988; Hatton et al. 1988a,b). (3) Data on the base composition of a gene's isochore (Bernardi et al. 1985; Filipski et al. 1987) were summarized by pooling isochore data from the AT-richest (A) or the GC-richest (G) halves of the genome. (4) CpG islands are in or near the 5′ end (I), 3′ end only (3′), or are not (N) near the gene (Bird 1986; Stanton et al. 1986; Filipski et al. 1987; Gardiner-Garden and Frommer 1987). (5) Presence (t) or absence (–) of an upstream TATA box (Gardiner-Garden and Frommer 1987). (6) Presence (c) or absence (–) of CCGCCC box within 250 bp 5′ of the transcriptional start site (Gardiner-Garden and Frommer 1987). (7) A measure of divergence rate: the gene's cDNA sequence cross-hybridized rodent–human at high (++) or low (–+) stringency, 87% or 77% similarity criteria respectively, or failed to cross-hybridize (––) (Goldman et al. 1984).

To avoid bias, data are complete listings of all mammalian genes cited in the preceding references excepting Gardiner-Garden and Frommer (1987). Brain gene and embryo-specific gene data are unavailable. Most data on replication time or base composition are coincident determinations from two or more mammalian species. They were generally determined from large 30–100-kb DNA fragments and consequently reflect properties of the domain in which the gene's exon sequences reside. Some data on properties 4–6 (Gardiner-Garden and Frommer 1987) and the base composition around vimentin (Bernardi et al. 1985) showed species differences. When plotted in four-dimensional property space, the genes appear clustered into two common groupings, (H/T)EGI R-bandlike, and TLAN G-bandlike

A gene is defined as replicating early if it replicates when the cell has a DNA content between 2C, the diploid G1 value, and 3C, the DNA content when half of the genome has replicated. Predicting the replication time of active genes is accurate within this range. The replication times of the 55 tabulated genes reflect over 200 separate determinations in various cell types. In only two cases did an active gene replicate after the cell had accumulated a 3C DNA content; these are DHFR at 3.2C and H2a histone at 3.3C, both in L60T, a mouse L cell (Hatton et al. 1988a). Both these genes replicated earlier, before 3C, in eight and three other cell lines, respectively. The 3C transition also corresponds to when R-bands have finished synthesis and G-bands are initiating synthesis (Holmquist et al. 1982; Holmquist 1987a, 1988a). AT-rich DNA is defined as coming from the two AT-rich isochores L1 and L2 (CsCl buoyant densities less than 1.703 g/cm^3); GC-rich DNA is from the three denser GC-rich isochores H1, H2, and H3 (Bernardi et al. 1985). L1 + L2 account for slightly more [55–64% (Macaya et al. 1976; Bernardi et al. 1985)] of the genome than does the 50% of late-replicating DNA

Mouchiroud et al. (1987) suggested housekeeping genes are more abundant but not always located in R-bands because reasonable numbers of sequenced housekeeping genes have a low GC content in their exons. However, many of the GC-poor genes they reported could actually be in GC-rich isochores (GC-rich bands) or, like the immunoglobulin constant region, be in early-replicating but AT-rich DNA; these data are presently unavailable. Also, Holmquist (1988a) reviewed why the coincidence of G-bands, quinacrine bright bands, and late-replicating bands is a good generalization from low resolution, 300 bands/genome, banding studies but the Q-band and G-band identity may be only an 85% accurate generalization at higher, 2000 bands/genome, resolution. Hamster chromosomes have several late-replicating R-bands (Holmquist et al. 1982). Some AT-rich DNA, 55–64% of the genome, must be excluded from the late-replicating 50% of the genome by arithmetic. Also, CpG islands are not absolute indicators of position. In alpha-globin clusters, one mouse globin gene and the chicken alpha^D-globin gene uncharacteristically lack islands. Similarly, in β-globin clusters, the chicken ρ-globin gene uncharacteristically has an island. Without stressing exceptions, this table shows that gene function and sequence property groupings can be made and they do roughly correspond to a structure, the gross low resolution properties of R-bands and G-bands. The ??-band category, indicating four loci that do deviate from the (H/T) EGI and TLAN groupings, shows that these generalizations are not universal

several properties of 86 genes, also supports the general coincidence of the two band fractionation schemes, i.e., genes in early- or late-replicating DNA are usually in GC-rich or AT-rich isochores, respectively.

As staining reveals two band classes, Table 2 reveals two gene groupings corresponding to these band classes. When data are available, housekeeping genes are *always* accompanied by the three properties E, G, and I (Table 2); HEGI are R-bandlike genes. A

majority of tissue-specific genes have the opposite four properties, i.e., TLAN; these are G-bandlike genes. Many tissue-specific genes are uncharacteristically in R-bands (Holmquist 1987a, 1988a; Ikemura and Aota 1988). Those with properties TEGI are in the R-bandlike (early-replicating, GC-rich isochores) grouping (Table 2); these do not utilize late replication as a repression-associated mechanism (Holmquist 1987a)

Because tissue-specific genes can be in either the G-bandlike group or the R-bandlike group, one can determine if any one property of these genes is more highly correlated with an aspect of the gene's function, tissue specificity, or with its bandlike group. For example, many housekeeping genes lack TATA boxes, whereas almost all tissue-specific genes, including both G-bandlike and R-bandlike ones, have TATA boxes (an RNA polymerase II initiation site); TATA boxes suggest a function-associated property. Lack of an upstream CCGCCC box, a consensus sequence for Sp1 protein binding, seems from our present limited data more associated with just G-bandlike genes than with all tissue-specific genes (Table 2) and is thus a band-associated property.

From the data in Table 1, one cannot distinguish if only housekeeping genes or all R-bandlike genes diverge more slowly (cross-hybridize rodent–human more frequently) than do G-bandlike genes. However, irrespective of their housekeeping or tissue-specific function, rodent genes with a high GC content in their coding sequences have a much higher silent codon substitution rate than do genes with a high AT content (Filipski 1988a,b; Wolfe, Sharp, and Li, unpublished). Also, interspersed repeat families in early-replicating DNA usually cross-hybridized hamster–human, whereas families in late-replicating DNA seldom cross-hybridized (Holmquist and Caston 1986; Holmquist 1988a). The combined data suggest that the molecular clock runs relatively slowly[1] in R-band chromatin and this slow rate is not just confined to housekeeping exons.

CpG islands are several hundred-bp-long DNA tracts that extend into coding sequences and contain 10 times the average genomic CpG dinucleotide frequency (Bird 1986). CpG dinucleotides in islands are hypomethylated and account for most of the *Sma*I, *Nae*I, *Nar*I, and *Sac*II sites used in chromosome mapping (Brown and Bird 1986). Islands (Table 2) are included in R-bandlike genes, both housekeeping genes and R-bandlike tissue-specific genes, but not included in G-bandlike genes. The presence or absence of CpG islands is more highly correlated with a gene's bandlike group type than with a function, its tissue expression pattern. Such a correlation reflects what has been called a band-associated "compartmental strategy" (Bernardi and Bernardi 1986; Perrin and Bernardi 1987) of the vertebrate genome.

Codon usage is also a compartmental strategy (Bernardi and Bernardi 1986a,b; Mouchiroud et al. 1987). The base composition of (1) a gene, (2) its codons' second, first, and especially third position bases, i.e., the wobble bases that primarily determine codon usage, and (3) 200+ kb of flanking sequences (the base composition isochore fraction in which the gene is located) are all three so positively correlated (Bernardi and Bernardi 1985, 1986a,b; Bernardi et al. 1985; Filipski et al. 1987; Muto and Osawa 1987; Ikemura and Aota 1988) that the term biased AT/GC pressure was coined to describe this base compositional constraint of flanking DNA on genes (Muto and Osawa 1987). Prokaryotic genomes have an organism-specific synonymous codon usage frequency that, due to a lack of noncoding DNA, greatly influences their average base composition (Zuckerkandl and Pauling 1965; Grantham et al. 1980; Ikemura 1985; Bulmer 1987). Warm-blooded vertebrate genomes differ (Bernardi and Bernardi 1985, 1986a,b; Aota and Ikemura 1986; Ikemura and Aota 1988) in that (1) their average base composition is almost totally determined by noncoding DNA, (2) they have band-specific base composition, and (3) they have band-specific codon usage. The codon usage is a compartmental strategy (Bernardi and Bernardi 1985, 1986a,b) and is independent of each gene's maximum transcription rate or function.

Many sequenced human genes have been mapped to either a G-band or an R-band of low resolution karyotype.[5] Ikemura and Aota (1988) found 19/20 of such R-band loci had a global (gene and flanking sequences) GC content >50%, whereas 73/102 G-band loci had a global GC content <50%. Their data also confirmed that housekeeping genes are confined to R-band loci and that exons in R-band loci or G-band loci have the base composition and wobble base usage indicated by Bernardi et al. (1985).

When Did Q-bands Evolve?

What animals have what kinds of bands is a difficult question to answer (reviewed and referenced, Holmquist 1988a). To a cytogeneticist, the banded prepattern is like a photographic latent image that can be developed for visualization in many ways.

[1] The molecular clock's rate was predicted to follow a compartmental strategy. Cox (1972) speculated that genes on chromosomes will not be organized necessarily according to function but rather into regions of different intrinsic mutation rates. Proteins whose sequences are under strong selective pressures should have loci in areas of low mutation rate. Because 0.5 deleterious mutations per genome per generation is the maximum allowable in an equilibrium population and may limit the total number of genes an organism possesses, this strategy, by decreasing the net genetic load of all genes, would allow organisms to accumulate more genes

Often, for a particular nonmammalian species, one worker could not visualize the pattern, the prepattern was actually present, and a subsequent more persistent worker succeeded in displaying it. Although the mechanism of trypsin G-banding is not understood at the molecular level (Holmquist 1987b), the mechanisms of replication banding and quinacrine banding are understood, and an inability to cytologically demonstrate these patterns in a particular species can be confirmed at the molecular level. For example, cold-blooded vertebrate chromosomes are notable for their inability to reveal Q-bands or Hae III-induced G-bands. Careful studies using quinacrine or the base composition-sensitive energy transfer stain combination, mithramycin/chromomycin A3 (Schmid 1980), and the restriction endonuclease Hae III (Schmid and de Almeida 1988) revealed no base composition-dependent banding in frog chromosomes even though these chromosomes show excellent replication bands (Schempp and Schmid 1981) and G-bands (Stock and Mengden 1975). A lack of GC-rich DNA to serve as a Q-banding substrate was confirmed at the molecular level by Bernardi et al. (1985).

GC-rich DNA isochores were found in mammals and birds (Thiery et al. 1976; Olofsson and Bernardi 1983; Bernardi et al. 1985) but few or none were found in frogs and fish (Macaya et al. 1976; Bernardi et al. 1985). Bernardi et al. (1985, 1986b), Filipski et al. (1987), and Perrin and Bernardi (1987) compared the base compositions of genes in amphibians, birds, and mammals with the base compositions of the isochores in which they are located. They made these observations. (1) Amphibians and fish differ from mammals and birds by having no or little GC-rich isochore DNA. (2) Genes that ancestrally are in AT-rich isochores in amphibians but are in GC-rich isochores in mammals have an increased G+C frequency in their immediate flanking sequences, introns, and an even more pronounced increase in the codon third positions of their exons. For example, the human alpha-globin gene in the R-bandlike group (Table 2) has nine times the CpG frequency of the *Xenopus* gene. It has a GC content of 67% and a GC usage in its third coding base positions of 88% compared with 43.5% and 46.5%, respectively, in the *Xenopus* alpha-globin gene (Filipski et al. 1987). In the G-bandlike group, the base composition and third coding base GC usage of the human beta-globin gene is only 40% and 67%, respectively. The alpha-globin gene is tissue specific and identical in function to beta-globin, its relative by gene duplication. However, alpha-globin has adopted a codon strategy like that of a housekeeping gene; this is not because of what the alpha-globin protein does, but because of where the alpha-globin locus is located, this is a compartmental strategy. (3) In GC-rich isochores, the GC content of third codon bases is usually higher than that of the noncoding DNA. In their response to directional mutation pressure, the wobble bases of coding sequences seem less constrained than even bulk DNA. (4) The base composition of the retroposon families *Alu* and L1, and the retroviruses listed in Table 1, rather closely matched the base composition of their flanking sequences, the isochores in which these mobile elements are found (Bernardi et al. 1985). (5) Fish that have been isolated in warm, 37–40°C lakes, have more GC-rich isochore DNA than related species from colder waters (Bernardi and Bernardi 1986b). Bernardi et al. (1985) and Filipski et al. (1987) concluded that many separate mechanisms, including the fixation of both point mutations and mobile elements, have acted to increase the GC content of GC-rich isochores. This increase in GC-richness has occurred rather homogeneously along the entirety of each 200+-kb isochore sequence, and the increase may be an effect of warm temperatures. That CpG islands may be genomic fossils from early invertebrates[2] seems contrary to the recent evolution of GC richness in quinacrine dull bands.

Goldman (1988) emphasized the replicon or independently supercoilable loop as a unit of gene regulation. One important aspect of the isochore concept is it establishes the evolutionary unit of base composition homogeneity as being larger than a 100-kb loop. It implies that replicons in a cluster, loops in a rosette, communicate among themselves to allow gradual base composition changes to occur in concert among those loops in that individual cluster. This is the essence of bands as units of genome evolution.

When Did Replication Bands Evolve?

A replication band represents a cluster of replicons all of which initiate and terminate replication in synchrony. This can be visualized by fiber autoradiography of ^{3}H pulse-labeled DNA. Unlike mammalian DNA, *Drosophila* DNA shows no evidence for such clusters of replicons (Stineman 1981); *Drosophila* chromosomes also show no G-bands or replication bands. The lowest vertebrates that have revealed replication bands or G-bands are bony fish (Fig. 2).

I argued that replication bands reflect the acqui-

[2] CpG islands may "represent the last remnants of the long tracts of nonmethylated DNA that make up most of the invertebrate genome" because mammalian genes lacking islands (G-bandlike genes) have a low CpG density and the sparse CpG dinucleotides remaining are mostly methylated (Bird 1987). Further evidence coupling the advent of G-bands to the increased cytosine methylation frequency of vertebrate genomes is lacking

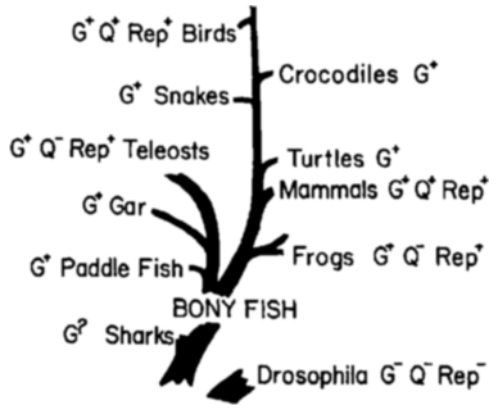

Fig. 2. Chromosome banding properties of extant taxa presented as they diverged from a line leading to extinct thecodonts. Modified from Holmquist (1988b) and including corrections: lack of G-bands in sharks was not rigorously tested, and B. Clark did not complete his Ph.D. thesis. Additions: replication banding in the urodele *Hynobius nigrescens* (Kuro-o et al. 1986) and improved Q-banding of chicken chromosomes (Ansari et al. 1986).

sition of a late replication-associated repression mechanism, facultative heterochromatinization of tissue-specific genes; this repression mechanism is lacking in *Drosophila* and would be advantageous for vertebrates that must repress many more tissue-specific genes, especially brain genes (Holmquist 1987a). Thus, facultative heterochromatinization of blocks of G-band genes would be a derived character with R-bands reflecting the primitive chromatin state. Redistributing preexisting tissue-specific genes away from housekeeping genes would have involved extensive genome reshuffling.

Meiotic chromomeres are beadlike compacted chromatin segments seen along meiotic chromosomes at the prophase stage of meiosis I. They are primitive, seen in almost all eukaryotes including *Neurospora* (Perkins and Barry 1979). Functionally, interchromomeric chromatin seems involved in synaptic initiation and homology-dependent synaptic pairing (Table 1). In mammals, mitotic G-bands correspond to AT-rich meiotic chromomeres and mitotic R-bands correspond to GC-rich meiotic interchromomeres (Table 1). The lack of mitotic replication bands in plants (Schweizer, personal communication), eukaryotes with meiotic chromomeres, and the general lack of chromomere-sized replicon clusters in lower eukaryotes (however, see Funderud et al. 1979) suggests that the temporal coordination of replicon clusters responsible for replication banding evolved around a preexisting pattern of meiotic chromomeres.

In summary, late-replicating bands are derived chromatin structures that were organized along a preexisting meiotic chromomere pattern and evolved in early deuterostomes or early chordates. They reflect the addition of late replication to the cells' preexisting repertoire of repression mechanisms. Chromatin differences facilitating visualization of bands by trypsin G-banding probably evolved later. More recently and restricted to homeotherms, Q-bandability evolved because of an increased GC content of early-replicating chromatin. This tidy summary has yet to be debated.

Dynamic Genome: Conserved Pattern

A dynamic concept of chromosome evolution emerges from recent data showing that major repeat families have originated and dispersed throughout the genome while the banding pattern remained unchanged.

Retroposition is the reverse flow of genetic information from RNA to DNA (Rogers 1985; Weiner et al. 1986). Nonviral retroposons make up a large fraction of the mammalian genome and seem innocuous in that they do not cis-regulate nearby genes. The process of retroposition involves transcription of a genomic sequence, reverse transcription of the RNA into an RNA–DNA hybrid, and subsequent insertion (genome breakage followed by ligation of the insert) of the DNA back into the genome (Rogers 1985). Genomic retroposon sequences or subsets thereof thus provide templates for further retroposition and the possible creation of many genomic copies. A family of interspersed repeat sequences created by retroposition corresponds to partial or complete copies of one cellular RNA species. The most successful families of short retroposons, such as *Alu,* have large family sizes and carry their promoters with them. They arose from polymerase III transcripts with sequences that include an internal polymerase III promoter (Rogers 1985). These retroposons have been quickly identified in novel vertebrate species, such as teleost fish, newts, and turtles, by in vitro transcription of genomic DNA with Pol III, as initiated from their internal promoters, followed by isolation via gel electrophoresis of discrete RNA molecular weight bands containing the transcribed retroposon sequences (Endoh and Okada 1986; Matsumoto et al. 1986).

Retroposon families can be classified by band preference. Of the 19 different hamster repeat families we used to probe replication time-fractionated DNA, two-thirds showed a band preference with SINEs, short interspersed repeats, in R-bands and LINEs, long interspersed repeats, usually in G-bands (Holmquist and Caston 1986). *Alu,* B1, and B2 families (in R-bands) and the L1 family (in G-bands) are partitioned (Soriano et al. 1983; Goldman et al. 1984; Holmquist and Caston 1986; Holmquist 1988a,b) so that 80–90% (Holmquist and Caston 1986) of the copies are in only one band type (Table 1 and Fig. 1).

New nonviral retroposon families are accumulating and growing in all mammalian orders (Rogers 1985; Deininger and Daniels 1986; Weiner et al. 1986). The *Alu* family, common to only primates, probably originated from 7SL RNA 65 million years (Myr) ago after primates diverged from rodents (Deininger and Daniels 1986) but before monkeys diverged from prosimians (Weiner et al. 1986). The monomer family of SINEs in the primate *Galago* originated from Met-tRNA and is absent from apes (Deininger and Daniels 1986). The B2a family of rodents originated from Ser-tRNA (Rogers 1985; Deininger and Daniels 1986), the rat identifier sequence, a SINE family, originated from Ala-tRNA and is specific to rats (Deininger and Daniels 1986), and the major rabbit SINE, the C family, seems specific to lagomorphs (Hardison and Printz 1985). This direct evidence for the independent origin and growth of *new* retroposon families since the radiation of the major mammalian orders, along with no direct evidence for deletion of old retroposons (Willard et al. 1987) supports the common view that retroposons only accumulate and are not excised (Deininger and Daniels 1986; Weiner et al. 1986).

A minority view, retroposons are excised, is more reasonable. Avian genomes contain few retroposons,[3] whereas retroposons are as common in salmon, tortoises, and frogs as they are in mammals (Endoh and Okada 1986; Matsumoto et al. 1986). Birds must have lost retroposons possibly by excision. Also, excision ultimately seems required because without some constraint on retroposon mass, our genome would accumulate retroposons and blow up like that of a lungfish. A genome size constraint was inferred for the enormous genomes of urodeles; species with larger genomes have correspondingly longer and selectively more disadvantageous development times (Cavalier-Smith 1985). Mammalian genomes vary in size within the limited range of 2.1–3.1 × 10^9 nuucleotide pairs/haploid genome (Lewin 1980) and much of this variation is due to satellite DNA. Even orthologous chromosome compartments in mice, humans, and cats seem constrained in size in that differential accordioning of

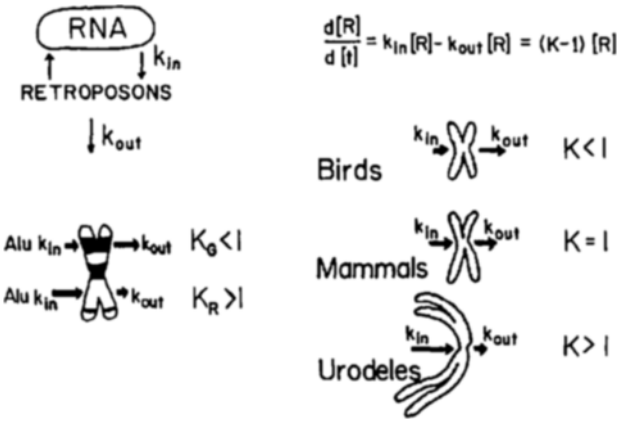

Fig. 3. An extreme view of chromosomes as stable patterns through which flow a karyotypically innocuous flux of mobile sequences, retroposons. Retroposon sequences of genomic concentration R generate, via RNA polymerase III, an RNA pool proportional to R. Retroposition into the genome via reverse transcriptase is then at a rate roughly proportional to R, i.e., $k_{in}[R]$. We presume a flux of retroposons $k_{out}[R]$ also leaves the genome at some rate proportional to R; the equilibrium rate constant is $K = k_{in}/k_{out}$. For *Alu*, $K_R > 1 > K_G$ describes the average history of K-values in R- and G-bands during the last 80 Myr in the lineage leading to humans. When extended to the C-value paradox (Cavalier-Smith 1985), with birds having smaller genomes [C = 0.8–1.3 pg of DNA/haploid genome (Lewin 1980)] and much less (Eden and Hendrick 1978; Epplen et al. 1978) interspersed repeat DNA than mammals [C = 2–3 pg of DNA (Lewin 1980)], the concept indicates that both the large and variable sizes of newt and salamander genomes [C = 18–69 pg in plethodontid salamanders (Mizuno and Macgregor 1974)] are explained by large overall K-values in urodeles. Although retroviral sequences can excise precisely by homologous recombination between flanking LTR sequences, there is no evidence that retroposons can do likewise. If retroposons are purged by chance inclusion in deletions, the equilibrium concept is not mathematically precise.

By sharing of a common RNA retroposition pool or by molecular drive, DNA sequences can migrate from populations of R-band sequences into populations of G-band sequences. Because a small migration rate between two large populations prevents allele frequency drift (Lewontin 1974), but the two band DNA populations *have* drifed apart in their base and mobile element frequencies, then the two band DNA populations behave as if isolated from each other (Holmquist et al 1982). This argument justified calling R-band DNA and G-band DNA "separate sub-genomes" (Goldman et al. 1984). Also, by the definitions of Grantham et al. (1980), the sub "genome and not the individual gene is the unit of selection for codon usage."

bands (Holmquist 1988b) is not readily apparent (Fig. 1). A constraint on retroposon mass suggests that in mammals but not urodeles (Fig. 3) retroposons should be excising from the genome about as fast as they are becoming fixed in the genome.

The fixation and possible deletion of retroposons during evolution may be rapid. The rate of turnover for L1 elements in mice seems to be only 3.3 Myr (Martin et al. 1985). Active and recently acquired endogenous copies of mouse mammary tumor virus (MMTV), copies specific to only certain mouse strains, are in both G- and R-bands, whereas the

[3] An almost total lack of SINEs in the avian genome (Epplen et al. 1978; Holmquist 1988a) but not in other reptiles or vertebrates indicates that mass excision of SINEs occurred. The avian genome's departure from the short period interspersion pattern of interspersed repeats, a pattern indicative of SINEs (Holmquist 1988b) and common to all other deuterostome genomes tested, is considered to be a relatively recent event restricted to this one Sauropsidian lineage (Bouchard 1982). CR1, the only reported SINE family in chickens, is present at only 1500–7000 copies and these are concentrated in GC-rich isochores (Olofsson and Bernardi 1983). Thus, numerous retroposons appear unessential to the continued existence of the banding pattern and have been deleted in birds. However, when retroposons are present, the banding pattern directs their dynamics

approximately 10 old MMTV copies, those highly methylated and common to all strains, are restricted to only one AT-rich isochore fraction of G-band DNA (Salinas et al. 1987). Because active MMTV provirions are detrimental to their host, selection against individuals with early-replicating (potentially active) MMTV sequences possibly could account for eventual partitioning of MMTV into late-replicating (usually inactive) G-bands. Unlike that of low copy number retroviruses such as MMTV, the substitutional load of genetic deaths required to similarly partition high copy number retroposons such as L1 is unreasonable (see Discussion). Consequently, retroposon partitioning is most likely due to innocuous differential retroposition/excision rates along the chromosome rather than to selective deaths of individuals with a retroposon in the wrong band.

The fixation of retroposons, although involving DNA breaks for their insertion, is karyotypically innocuous. HSA 1p has accumulated 30,000 copies of *Alu* and MMU 4 has accumulated about half that many copies of B1 and B2 since humans and mice diverged. Yet the banding patterns of HSA 1p and MMU 4 are almost identical (Fig. 1). Rodent and probably cat chromosomes lack *Alu* and contain completely different major SINE families (Rogers 1985; Deininger and Daniels 1986; Weiner et al. 1986) in the ecological niche of the human *Alu* family (Holmquist 1988b). Like MMU 4 in mice, several chromosome regions of the house cat, *Felis catus,* are almost identical to human chromosome regions by gene mapping, G-banding, and replication banding (O'Brien et al. 1984; von Kiel et al. 1985; Sawyer and Hozier 1986) (Fig. 1). Five hundred thousand (Schmid and Jelinek 1982) to 910,000 (Hwu et al. 1986) *Alu* copies have been inserted in the human lineage during primate evolution (Deininger and Daniels 1986), yet only 150 different chromosome rearrangements are evident between all different primate lineages (Dutrillaux 1979; Seuanez 1979). Humans may actually have fixed another 500,000 *Alu* copies and 50,000 L1 copies in the last 6 Myr since humans diverged from apes (Hwu et al. 1986). Yet aside from a few chromosome rearrangements, the ape and human G-banding patterns are identical. In plethodontid salamanders, wherein the C-value ranges from 18 pg to 69 pg DNA/haploid genome, the unbanded karyotypes are, with the exception of actual chromosome size, identical in chromosome number and arm length ratios in most of the species karyotyped (Mizuno and Macgregor 1974). A description of these data that defines the partitioning of mobile families of interspersed retroposons and the C-value itself in terms of rate constants for transposition (and possible excision) is shown in Fig. 3. It portrays the karyotype as a rather stable pattern through which flows a flux of karyotypically innocuous elements (Holmquist 1988b). Stable patterns in such dynamic systems require feedback loop(s) to stabilize them. Because retroposon sequences do not include a consensus sequence for their own integration (Rogers 1985; Weiner et al. 1986), another feedback loop must stabilize their patterns of insertion.

Chromatin Domains and Mutation Rates

Various DNA modification reactions have rate constants that vary along the chromosome. Near the glue gene cluster of *Drosophila* is a sharp 50-bp-wide boundary between rapidly and slowly evolving DNA. Sequenced DNA from five *Drosophila* species showed that the rate of fixation of base substitutions has differed by almost 10-fold across this boundary (Martin and Meyerowitz 1986). Because the frequency of small insertion or deletion events across this boundary was nearly constant, the different base substitution rates cannot be due to natural selection against individuals with altered DNA and must reflect the cell's ability to modulate the rate of DNA base substitution along the chromosome independently of deletion and insertion rates. Repair rates in mammals also vary along the chromosome. In monkey cells, aflatoxin B_1 adducts in alpha DNA of the centric heterochromatin are removed much more slowly than are adducts in the remainder of the genome (Zolan et al. 1982; Leadon et al. 1983), whereas no such difference was detected for thymidine dimer removal (Zolan et al. 1982). However, thymidine dimers are removed up to 10 times faster from active genes (c-abl, DHFR) than they are from inactive genes (c-mos) (Madhani et al. 1986; Mellon et al. 1986) and this fast repair domain extends at least 50 kb upstream from the human DHFR gene (Madhani et al. 1986; Mellon et al. 1986). Very early-replicating genes seem to amplify more easily and frequently than do other genes (Schimke et al. 1986). Thus, individual reactions that maintain or alter DNA sequences are often *independently* modulated along the chromosome and do not simply reflect overall patterns of chromatin accessibility. We now ask "does the molecular clock vary coincident with the banding pattern?" If a rate increase in only one DNA modification reaction, homologous recombination, can drive the collapse of a standard eukaryotic genome into the almost intronless *Saccharomyces* genome (Fink 1987), a genome void of most redundant sequences, then smaller rate changes varying along individual chromosomes could certainly mold an alternating pattern of bands.

Some chromatin proteins are band type specific. C-bands have very little nonhistone protein (Comings et al. 1977) and probably have some proteins

specific to them (Solomon et al. 1986). Fluorescent serum against T-antigen illuminates a G-banded pattern along chromosomes of SV40-transformed cells (D'Alisa et al. 1979). A fluorescent monoclonal antibody against a nuclear protein specifically illuminates R-bands (Schonberg et al. 1987). From this, along with evidence that rate constants for mutation, repair, and amplification may vary along the chromosome and knowledge that retroposition rate constants, directional base composition changes, inclusion of CpG islands, and even mutation rate are orchestrated by the banding pattern (Table 1), we infer that bands are chromatin macrodomains that compartmentally modulate the rates of DNA modification reactions.[4]

What force maintains the base composition uniformity in one isochore (band), whereas different isochores (bands) have different GC contents and different *Alu* concentrations? Why are gene sequence properties correlated with properties of the noncoding DNA in which the genes are imbedded? How can a stable banding pattern survive, whereas many of its elements are being constantly replaced? Answers require new concepts.

If chromatin modulates its DNA modification reactions, then a chromatin domain would ultimately influence its own DNA sequences. This would involve a self-stabilizing feedback loop where: DNA sequences determine the affinity of → proteins that bind to form → a chromatin that modulates → rate constants for DNA modification reactions that determine → DNA sequences. For example, if the unmethylated CpGs in CpG islands cooperatively bound a cytosine methylase-inhibiting protein, they would be spared the directional mutation of deamination, $C^{m}pG \rightarrow TpG$, thus preserving both themselves and local GC richness. Even a coding sequence could be influenced by local DNA via the chromatin domain that local DNA organizes. We now discuss how the force of natural selection, acting at the DNA level, could cause communities of mobile DNA sequences to cooperatively coevolve via this feedback loop to assist formation of stable chromatin patterns.

Discussion

Noncoding sequences along the chromosome are too ordered to result from fixation of selectively neutral events but would have produced an unreasonably high genetic load if natural selection had acted on individuals in which these events occurred.

[4] Sueoka (1988) suggested there might be several different directional mutation pressures (differential repair rates, etc.) in different locations of the vertebrate genome with the cause of the different pressures possibly residing in the structural elements of the different chromatin domains

Random Drift and Positional Constraints

The essence of Kimura's (1983) neutral mutation–random drift theory is that the selection intensity on 94–98% of all molecular mutants is so weak that mutation pressure and random drift prevail. DNA sequences unconstrained by natural selection should be polymorphic in the population, maintained in the species by mutational input and random extinction. Third codon bases did appear neutral and functionally unconstrained in 1983. But with additional data, Bernardi and Bernardi (1986) showed that those third codon bases that became fixed in mammalian populations are positionally constrained. Their base composition is positively correlated with the base composition of flanking DNA, i.e., fixed in tune with the compositional strategy of the genome. We can now estimate the percentage of nucleotides that are positionally constrained, i.e., have been fixed in R-bands, because their loci are in R-bands and not in G-bands. The 3% (Holmquist et al. 1982) difference in base composition implies 3% of the bases are G or C because their loci are in R-bands instead of G-bands. Eighty-five percent of the 5% of the genome that is *Alu* adds another 8.5%. Another approximately 8.5% in hamsters comprises a variety of low copy number SINE families all specific to R-bands (Holmquist and Caston 1986). Only our lack of data precludes extending this percentage past 15–20% and positional constraints on coding sequences have not been included. Positionally constrained DNA is so extensive that cytogeneticists use it to identify chromosomes (Fig. 1).

Kimura and Maruyama (1969) used Fick's diffusion equations as a function of time to argue that a new amino acid or base at almost any locus is not (functionally) constrained but is free to randomly diffuse through homologous loci of an interbreeding population over time without a substitutional cost contribution to genetic load. But diffusion equations also apply to spatial constraints. In the application of Fick's equation (Tanford 1961) to CsCl sedimentation equilibrium, the final concentration gradient is a balance between a sedimentaiton force concentrating CsCl at the periphery of the rotor and a force of diffusion tending to make the CsCl concentration homogeneous. At equilibrium, a concentration gradient implies a second force must act against the diffusive force to sustain the gradient. The bands represent concentration gradients of constrained DNA. The basic pattern has been at equilibrium in HSA 1p (Fig. 1) for over 80 Myr. What force maintains these gradients from mutational diffusion?

The High Cost of Natural Selection

There are only two possible contributions to the force that directs positional constraints. Mutational

input could be positionally ordered, and extinction of a sequence could involve ordered mutational output. Figure 3 shows these two mechanisms as competing for retroposons but it could similarly be applied to individual bases with the results appearing as a temperature-sensitive directional mutation pressure for increased G+C in homeotherms. This model ignores fixation and depicts rates of mutational input and mutational output as being positionally organized by some undescribed force. Genome size and longitudinal differentiation along the chromosomal DNA molecule become simple historical records of input/output ratios.

While applying it only to exon sequences, Bernardi and Bernardi (1986) suggested that natural selection acting on individuals is the force ordering mutational input–output. The DNA that responds to that force classically provides a function to the organism upon which selection acts. Haldane's (1957) substitutional cost theory was the basis for Ohno calling noncoding DNA nonfunctional junk (Ohno 1972) and stimulated the neutral mutation hypothesis (Kimura 1983); with a spontaneous mutation rate of 10^{-5}/locus·generation, "the genetic load for maintaining this DNA as functional and sequence constrained would be intolerable and therefore it must be functionless junk" (Ohno 1972). The substitutional load, the cost to a population of replacing one allele with a new more advantageous allele, is that about $30n$ individuals (n = the population size) carrying the old allele must die genetic deaths, while the new allele spreads through and becomes fixed in the population. Recent data allows one to calculate the cost of a selective advantage for positionally constrained DNA as 4×10^5 genetic deaths per generation to fix *Alu* sequences in R-bands since *Alu* first appeared in primates[5] and 3.8×10^5 genetic deaths per year to fix the increased GC content of our R-band compartments since humans diverged from frogs.[6] Unreasonable!!

Positionally constrained DNA differs from functionally constrained DNA in that it accumulates mutations rapidly and was hence called unconstrained DNA (Kimura 1983). Zuckerkandl (1986) realized the impossible substitutional cost of so many mutations becoming fixed in positionally constrained sequences and proposed that this DNA is functionally constrained as sequence motifs rather than as rigorously defined sequences. Many mutations could accumulate in a motif as nearly neutral mutations until sequence adulteration passed a certain threshold of selective disadvantage. At this time, a motif's accumulation of *many* mutations could be eliminated by natural selection acting to remove an entire motif, rather than each of its mutations individually, at a fraction of the usual substitutional cost. This is an extreme form of nonindependent selection (Smith 1968). In the preceding paragraph, we treated the *Alu* sequence as a unit motif and found that the genetic load for fixing whole motifs was unreasonably great so the load involved in fixing their individual base substitutions is academic.

The simplifying assumptions and exact equations of population genetics lead to absurd contradictions.[5,6] Whether positionally constrained DNA is advantageous, or neutral, or disadvantageous to the individual, it cannot exist. I propose that life involves such a complex hierarchy of systems interactions that selection and mutational load can be realized at lower, less costly levels than the individual. *Alu* can be positionally constrained without killing individuals.

Hierarchical Selection

Organisms of the same species represent heritably related, phenotypically similar Darwinian units.

[5] Retroposon–orthologous-site polymorphisms are rare in humans (Kominami, et al. 1983; Economou-Pachnis and Tsichlis 1985; Furano et al. 1986; Sawada and Schmid 1986; Trabuchet et al. 1987) implying that most retroposon sequences are already fixed in the population. For *Alu*, this is reasonable if the copies were all generated by three or four rather old bursts of retroposition from three or four distinct founder sequences (Sawada and Schmid 1986; Willard et al. 1987; Britten et al. 1988). Any neutral mutation becomes either eliminated from or effectively fixed in a population of size n in about $2n$ generations and any selective advantage or disadvantage speeds this process (Kimura 1983). For an *Alu* inserted in higher primates, this would be about $2 \times$ 100,000 individuals $\times$ 20 years/generation = 4 Myr. This estimate is consistent with finding no orthologous-site *Alu* polymorphisms for six sites, a comparison between humans and chimpanzees in the alpha-globin locus (Sawada and Schmid 1986). The high concentration of *Alu* in R-bands classically rules out fixation by selectively neutral events. Assume an *Alu* in an R-band is advantageous. Because *Alu* evolved 65 Myr ago (Deininger and Daniels 1986), the lineage leading to humans has fixed 5×10^5 *Alu* copies with an estimated substitutional cost of [(5 $\times$ 10^5 copies $\times$ 30 $\times$ 100,000 individuals in population)/(65×10^6 years)] = 2.3×10^4 genetic deaths/year. Alternatively, one could argue that *Alu* insertions occur homogeneously throughout the genome and are nearly neutral in R-bands, but that such insertions in G-bands represent deleterious mutations. The frequency with which mutations occur is approximately $2n$ times their rate of fixation (Kimura 1983): $2 \times$ 100,000 individuals $\times$ (5×10^5 *Alu*) $\times$ 20 years per generation/65×10^6 years = 3×10^4 deleterious mutations per generation for selection against *Alu* inserts in G-bands. The maximum allowable number of all deleterious mutations per genome per generation is 0.5 per gamete in an equilibrium population (Crow and Kimura 1970). So *Alu* cannot afford to be deleterious in G-bands

[6] R-band DNA is 3.24% GC richer than G-band DNA (Holmquist et al. 1982), an increase that occurred since frogs diverged from man 360 Myr ago. 1.4×10^9 bp of R-band DNA $\times$ 0.0324 = 4.5×10^7 times A or T → G or C. ($4.5 \times 10^7 \times 30 \times$ 100,000/3.60×10^8 years) = 3.8×10^5 genetic deaths per year for 360 Myr to directionally exchange A/T for G/C. If the increased GC has been fixed, the substitutional load is much greater than *Alu*'s contribution

They propagate usually very similar units and compete with both other species and members of their own species in a bounded ecosystem wherein their numbers are limited. Consider as a hypothetical example, the hundreds of mitochondria in a mammalian male's germ cell. These are also Darwinian units. They compete with each other while obeying all the rules of Darwinian units, and they are essential to the success of their bounded ecosystem, the cell. Dropping down to a lower hierarchical level, the mitochondrion is a smaller bounded ecosystem and its 10 DNA molecules behave like Darwinian units competing with each other and mutually contribution to the success of their bounded ecosystem, the mitochondrion. At this hierarchical level, consider a hypothetical petite-like mutation, a mitochondrial DNA molecule with a deletion of a few tRNA genes. The deletion decreases the replication time and consequently increases the selective advantage of that molecule in the mitochondrion. The petite-like molecule will increase its frequency at the expense of wild type molecules in that mitochondrion's progeny. But mitochondria with petite-like DNA compete with wild type mitochondria in the cell (at the next higher hierarchical level). If the selective advantage due to the deletion was positive, then these mitochondria would "win" in the cell lineage harboring them. But now these cell lineages are competing with other germ cell lineages in the organism's testicle. If the fitness of petite-like cells is low becaues they are anaerobic, then the petite-like cell lineages would die out. In both male and female mammalian gonads, 70–90% of the germ cells undergo atresia during mitotic proliferation or before meiosis I (Baker 1963); within wide limits, the loss of a germ cell lineage decreases only the variety but not the final number of gametes produced. Thus, intervening hierarchical levels of organization ensure protection of costly Darwinian units higher in the hierarchy from changes in smaller less expensive Darwinian units lower in the hierarchy. The hypothesized petite-like mutation would disappear with minimal mutational cost to the organism because the resulting genetic deaths were at the cellular level of hierarchical selection.

As an example of hierarchical selection, R-band trisomy is selected against at less cost than is G-band trisomy (Table 1). Selection against uncharacteristic activity of tissue-specific genes such as brain or liver-specific genes acts after midembryogenesis or after birth when these gene products are normally required. As housekeeping genes are active and usually required in all cell types, selection against their uncharacteristic activity should more often occur in the germ line or before implantation. Most human embryos trisomic for very large chromosome segments either fail to implant or abort before birth. Embryos trisomic for chromosomes containing small amounts of R-band chromatin, such as chromosomes 21, 18, or 13, often survive to birth, whereas embryos trisomic for smaller chromosomes but ones with more R-band chromatin, i.e., chromosomes 19, 20, and 22, either fail to implant or spontaneously abort in the first trimester (Hoehn 1975; Kuhn et al. 1985) (Table 1). In the currency of gestation and lactation time, the mutational cost of lowered reproductive fitness of R-band (housekeeping gene) trisomy is less than that of G-band trisomy.[7]

Returning to the mitochondrion example, assume all its 10 DNA molecules were fused into one large chromosome consisting of 10 tandem repeats. This would at first appear to remove the lowest level of the hierarchy so that an altered repeat could not increase in this mitochondrion's lineage at the expense of the nine wild type repeats. The large chromosome would now behave as the lowest unit of Darwinian selection. However, if the petite-like repeat could, by gene conversion, unequal sister chromatid exchange, transposition, slippage-replication, amplification, i.e., any of the mechanisms of molecular drive (Dover 1982), increase its frequency at the expense of the wild type repeats, then the repeats, although linked as one chromosome, would behave again like individual units. The petite-like repeat could again increase its frequency at the expense of wild type repeat units. If the mutation were a base substitution or insertion instead of a deletion, the base pair would act as the smallest Darwinian unit (Zuckerkandl 1976). Thus, non-Mendelian mechanisms of inheritance all share one fundamental systems property, they add a lower level to the hierarchy, allowing segments of chromosomes as small as base pairs the freedom to behave as Darwinian units or selfish DNA (Dawkins 1976; Orgel and Crick 1980), assuming the following criteria are met: (1) Limits on genome size limit the population size of these units. (2) The mutational load accompanying their competition is not effected at a costly organism level. Such a freeedom would actually ensure that a dynamic cooperative system would form. Ecosystems exemplify such systems.

Molecular Ecology of Ordered Mutational Input

Mobile DNA sequences as Darwinian units in bands are like individual plants as Darwinian units in floral

[7] If a dominant mutation in a housekeeping gene such as a ribosomal protein occurred in a secondary spermatogonia and conferred a sudden disadvantage to this cell lineage, the resulting reduction of genetic variability, a component of genetic load, in the population of individuals would be less, by the ratio of spermatogonia/individuals, than that from a dominant tissue-specific gene that killed only individuals

communities. (1) Sexual reproduction: parental strands exchange, daughter strands have sequence properties of both parental strands. Asexual reproduction: amplification or duplicative retroposition. (2) Retroposons occupy the equivalent of ecological niches in the genome. B1 and B2 or *Alu* are the dominant SINE family in the R-band niche of mice or men, respectively, whereas lagomorphs have C1 repeats instead of *Alu* or B1 (Sakamoto and Okada 1985). (3) Like individual plants in a plant community, individual repeats can come and go (birth and death), yet the spatial distribution of their family members can, like the boundaries of a plant community, remain sharp and constant (Fig. 3). (4) Interspersed repeats, such as the retroviral-like *ETn* elements in various murine species (Sonigo et al. 1987) or the nonviral retroposon L1 family in various mammalian orders (Martin et al. 1985), have coevolved (Holmquist and Dancis 1979) in a concerted fashion around a drifting family consensus sequence (Dover 1982, 1986), i.e., elements of the family have greater similarity to one another when from the same organismal species than when from different species. The mechanism of concerted evolution of interspersed repeats could involve selective amplification (duplicative transposition) of one or a few copies of each element with elimination of old elements and/or gene conversion between elements in different parts of the genome (Martin et al. 1985; Dover 1986; Sonigo et al. 1987). Like a plant population, the murine L1 family or *ETn* family appears as a popualtion of interbreeding Darwinian units sharing the same gene pool. (5) Bees and flowering plants have coevolved. They live together and help each other. Might two different sequence families, like bees and flowering plants, interspecifically coevolve and consequently cohabit the same compartments? This would drive local aggregations of synergistic sequence types into sequence communities. It would appear as a concentrating force in Fick's equilibrium equation, and be a prediction from molecular ecology. As 9 of 21 different SINE sequences and 11 of 12 different LINE sequences are concentrated in either hamster's R- or G-bands, respectively (Holmquist and Caston 1986), their cohabitation suggests synergistic coevolution and further examination of plant community properties.

Odum (1969) proposed that succession in an ecosystem has three general properties. (1) It is an orderly directional process of plant community development. (2) It culminates in a climax ecosystem with maximal information content, and, by the buffering effects of interspecies interactions, produces local homeostasis. He implied that survival of the fittest is often survival of the most cooperative with constellations of species coevolving synergistically. (3) Even though the external environment of temperature, rainfall, etc. dictates the subset of all plant communities that can exist, once a particular successional pattern or route to one particular climax community becomes established, the local community determines to a great extent the local biotic microenvironment to which individuals must adapt. This is positional constraint.

Succession in the southern Lake Michigan sand dunes (Olson 1958) reveals these principles. Succession from the monophyletic dune grass pioneering community to a climax forest of trees, bushes, and shrubs may branch into one of two major routes. These lead either to a mesophytic climax community of basswood–maple, beech, birch, and alkaline soil, or to a xerophytic community of pine, black oak, blueberry, huckleberry, and acidic soil (Olson 1958). Black oak leaf litter mulches by fungal degradation to an acidic humis through which cationic nutrient ions are easily leached by rain. The nutrient level in the acidic soil is just sufficient for the survival of black oak and blueberry but insufficient for the survival of maple. Conversely, litter from the mesophytic trees mulches by bacterial degradation to an alkaline humis. The higher pH keeps this soil's level of absorbed cationic nutrients high enough for the maple and beech to always have a competitive advantage over black oak. Viewed from above, the geographically patchy pattern of light and dark green areas representing the two climax communities is, like the chromosome's banding pattern, fairly stable over time (Olson 1958) even though the elements, individual plants, are continuously replaced. One community excludes the other's species. The areas can, like bands in Fig. 3, be described by their germination–growth-to-maturity equilibrium rate constants: mesophytic where $K_{\text{maple}} \geq 1$ and $K_{\text{black oak}} < 1$ or xerophytic where $K_{\text{maple}} < 1$ and $K_{\text{black oak}} \geq 1$. One loop in the self stabilizing feedback system is: plants modulate → soil pH that affects → nutrient level that determines → germination success of plants. Although the black oak tree functions as a canopy tree and grows best in high-nutrient soil, were it to be isolated and myopically studied in a greenhouse, its producing a nutrient-poor soil would enigmatically seem detrimental to its function as a tall canopy tree. However, in tune with the pH feedback loop, this property reflects the oak's community strategy of excluding maple and not the oak's canopy function. The enigma was created by studying the oak separate from its positional habitat.

The propensity of DNA sequences to follow a community strategy related to chromatin (soil) type has been called politeness (Zuckerkandl 1986). It involved a study of DNA sequences in their native habitat. Polite DNA (Zuckerkandl 1986) is sequences that, upon entering a chromatin domain, leave that domain's unique integrity undisturbed.

AT-rich retroposons are usually found politely imbedded in AT-rich domains because the compartmentalized genome has already maximized politeness. Table 2 actually demonstrates politeness. Each gene, like a computer file, contains coded sequences broadly indicating disc and sector locations where that gene "belongs." Alpha satellite (C-band) DNA behaves impolitely when inserted into euchromatin; there it causes instability, chromosome breaks, and sister chromatid exchanges (Heartlein et al. 1987). Although it accurately describes many observations, politeness is a passive concept and therefore insufficient to complete the active feedback loop required for community homeostasis and chromosome stability. A noncoding DNA sequence must actively complete the loop by favoring propagation of both itself and its fellow community members. For example, B1 sequences in murine R-bands should contribute to an R-band chromatin that increases the retroposition of both B1 and B2 sequences, whereas B2, in turn, should synergistically assist B1 into the same domain. Thus, as distinct from molecular drive that restricts natural selection to acting only on the external dynamics of populations of individuals (Dover 1986), molecular ecology demands that selection also act below the individual level on the internal dynamics of non-Mendelian DNA behavior. If such feedback systems involving DNA turnover became firm forces driving the genome into patterns of distinct chromatin domains, then each domain could influence even the habitat to which resident coding sequences must adapt.

C-bands are quite polymorphic (John and Miklos 1979). The often explosive population dynamics of their tandem repeats are like those of nonsynergistic pioneering species in early stages of succession (Odum 1969). Mammalian retroposon dynamics are usually much more like that of a climax ecosystem or even like symbionts in a population of hosts, stable, commensural, homeostatic, as if the chromatin domains they help form may also impart an advantage to the chordate host. An obvious advantage is the possible strategy of retroposons providing nearby genes with a more optimal chromatin environment. Genes inactive in most tissues and genes whose deleterious mutants would cause a mutational load at the organism level could, with the advent of G-bands, now be sequestered into facultative heterochromatin where their repression mechanism repertoire would include late replication. Their spectrum of chromatin proteins that bind to newly synthesized DNA could be modulated by protein availability during late S-phase (Gottesfeld and Bloomer 1982). Additionally, these genes could be in chromatin compartments with unique rates of mutation, amplification, repair, and eventual codon usage. An abundance of DNA, unconstrained by a protein-coding function, would be absolutely essential to the existence and homeostasis of such unique chromatin compartments and to the advantage such compartments would provide to the individual (Cox 1972).

In summary, the chromosome's backbone constituents are dynamic, whereas its stable banding pattern rigidly constrains both protein-coding sequences and noncoding DNA sequences. These data are inconsistent with the mainstream molecular theory that coding sequences are determined only by selection acting on protein function and with the neutral mutation hypothesis that predicts noncoding sequences be positionally unconstrained. Molecular ecology theory was presented in an extreme form to show that these data can be explained without evoking structural DNA and without emphasizing the covalent continuity of chromosomal DNA. Noncoding DNA in mammals is sufficiently organized and its organization is now sufficiently recognized to suggest its function.

Acknowledgments. I thank Giorgio Bernardi for everything including the EMBO meeting; and Emile Zuckerkandl, Art Riggs, Carl Schildkraut, Tomoko Ohta, Jan Filipski, and Susumo Ohno for discussion and unpublished data. I am totally indebted to the Beckman Research Institute for funding this research.

References

Ambros PF, Sumner AT (1987) Correlation of pachytene chromomers and metaphase bands of human chromosomes, and distinctive properties of telomeric regions. Cytogenet Cell Genet 44:223–228

Ansari HA, Takagi N, Sasaki M (1986) Interordinal conservation of chromosome banding pattern in *Gallus domesticus* (Galliformes) and *Melopsittacus undulatus* (Psittaciformes). Cytogenet Cell Genet 43:6–9

Aota S-I, Ikemura T (1986) Diversity in G+C content at the third position of codons in vertebrate genes and its cause. Nucleic Acids Res 14:6345–6355

Ashley T (1988) G-band position effects on meiotic synapsis and crossing over. Genetics 118:307–317

Ashley T, Russell LB (1986) A new type of nonhomologous synapsis in T(X;4) 1R1 translocation male mice. Cytogenet Cell Genet 43:194–200

Baker TG (1963) A quantitative and cytological study of germ cells in human ovaries. Proc Soc Lond Ser B 158:417–433

Bernardi G, Bernardi G (1985) Codon usage and genome composition. J Mol Evol 22:363–365

Bernardi G, Bernardi G (1986a) The human genome and its evolutionary context. Cold Spring Harbor Symp Quant Biol 41:479–487

Bernardi G, Bernardi G (1986b) Compositional constraints and genome evolution. J Mol Evol 24:1–11

Bernardi G, Olofsson B, Filipski J, Zerial J, Salinas J, Cuny G, Meunier-Rotival M, Rodier M (1985) The mosaic genome of warm-blooded vertebrates. Science 228:953–958

Bianchi MS, Bianchi NO, Pantelias GE, Wolff S (1985) The mechanism and pattern of banding induced by restriction endonucleases in human chromosomes. Chromosoma 91:131–136

Bird AP (1986) CpG rich islands and the function of DNA methylation. Nature 321:209–213
Bird AP (1987) CpG islands as gene markers in the vertebrate nucleus. Trends Genet 3:342–347
Bouchard RA (1982) Moderately repetitive DNA in evolution. Int Rev Cytol 76:113–193
Britten RJ, Baron WF, Stout DB, Davidson EH (1986) Sources and evolution of human *Alu* repeated sequences. Proc Natl Acad Sci USA 85:4770–4774
Brown RAW, Bird AP (1986) Long-range restriction site mapping of mammalian genomic DNA. Nature 322:477–481
Bulmer M (1987) Coevolution of codon usage and transfer RNA abundance. Nature 325:728–730
Calza RE, Eckhardt LA, DelGiudice T, Schildkraut CL (1984) Changes in gene position are accompanied by a change in time of replication. Cell 36:689–696
Camargo M, Cervenka J (1982) Patterns of DNA replication of human chromosomes. II. Replication map and replication model. Am J Hum Genet 34:737–780
Cavalier-Smith T (1985) The evolution of genome size. John Wiley and Sons, New York
Chandley AC (1986) A model for effective pairing and recombination at meiosis based on early replicating sites (R-bands) along chromosomes. Hum Genet 72:50–57
Comings DE (1978) Mechanisms of chromosome banding and implications for chromosome structure. Annu Rev Genet 12: 25–46
Comings DE, Avelino E (1975) Mechanisms of chromosome banding VII. Interaction of methylene blue with DNA and chromatin. Chromosoma 51:365–379
Comings D, Harris D, Okada TA, Holmquist GP (1977) Nuclear proteins IV. Deficiency of non-histone proteins in *Drosophila virilis* and mouse heterochromatin. Exp Cell Res 105: 349–365
Comings DE, Kovacs BW, Avelino E, Harris DC (1975) Mechanisms of chromosome banding. V. Quinacrine banding. Chromosoma 50:111–145
Cox EC (1972) On the organizations of higher chromosomes. Nature New Biol 239:133–134
Crow JF, Kimura M (1970) An introduction to popualtion genetics theory. Harper and Row, New York
Cuny G, Soriano P, Macay G, Bernardi G (1981) The major components of the mouse and human genomes. 1. Preparation, basic properties and compositional heterogeneity. Eur J Biochem 115:227–223
D'Alisa RM, Korf BR, Gershey EL (1979) T antigen banding on chromosomes of simian virus 40 infected muntjac cells. Cytogenet Cell Genet 24:27–36
Dawkins R (1976) The selfish gene. Oxford University Press
Deininger PL, Daniels GR (1986) The recent evolution of mammalian repetitive DNA elements. Trends Genet 2:76–80
Doolittle WF, Sapienza C (1980) Selfish genes, the phenotype paradigm and genome evolution. Nature 284:601–603
Dover G (1982) Molecular drive: a cohesive mode of species evolution. Nature 299:111–117
Dover GA (1986) Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. Trends Genet 2:159–165
Dubey DD, Raman R (1987) Factors influencing replicon organization in tissues having different S-phase durations in the mole rat, *Bandicota bengalensis*. Chromosoma 95:285–289
Dutrillaux B (1979) Chromosomal evolution in primates: tentative phylogeny form *Microcebus murinus* (Prosimian) to man. Hum Genet 48:251–314
Economou-Pachnis A, Tsichlis PN (1985) Insertion of an Alu SINE in the human homologue of the Mlvi-2 locus. Nucleic Acids Res 13:8379–8387
Eden FC, Hendrick JP (1978) Unusual organization of DNA sequences in the chicken. Biochemistry 17:5835
Endoh H, Okada N (1986) Total DNA transcription in vitro: a procedure to detect highly repetitive and transcribable sequences with tRNA-like structures. Proc Natl Acad Sci USA 83:251–255
Epplen JT, Leipoldt M, Engel W, Schmidtke J (1978) DNA sequence organisation in avian genomes. Chromosoma 69: 307–321
Fang JS, Jagiello GM (1988) An analysis of the chromomere map and chiasmata characteristics of human diplotene spermatocytes. Cytogenet Cell Genet 47:52–57
Filipski J (1988a) Why the rate of silent codon substitutions is variable within a vertebrate's genome. J Theor Biol 134:159–164
Filipski J (1988b) Sequence-based phylogeny in eukaryotic genomes. Nature 334:572
Filipski J, Salinas J, Rodier F (1987) Two distinct compositional classes of vertebrate gene-bearing DNA stretches, their structures and possible evolutionary origin. DNA 6:109–118
Fink GR (1987) Pseudogenes in yeast? Cell 49:5–6
Funderud S, Andreassen R, Haugli F (1979) DNA replication in *Physarum polycephalum*: electron microscopic and autoradiographic analysis of replicating DNA from defined stages of the S-period. Nucleic Acids Res 6:1417–1431
Furano AV, Somerville CC, Tsichlis PN, D'Ambrosio E (1986) Target sites for the transposition of rat long interspersed repeated DNA elements (LINEs) are not random. Nucleic Acids Res 14:3717–3726
Ganner E, Evans HJ (1971) The relationship between patterns of DNA replication and quinacrine fluorescence in the human complement. Chromosoma 35:326–341
Gardiner-Garden M, Frommer M (1987) CpG islands in vertebrate genomes. J Mol Biol 196:261–282
Goldman MA (1988) The chromatin domain as a unit of gene regulation. BioEssays 9:50–55
Goldman MA, Holmquist GP, Gray MC, Caston LA, Nag A (1984) Replication timing of mammalian genes and middle repetitive sequences. Science 224:686–692
Gottesfeld J, Bloomer LS (1982) Assembly of transcriptionally active 5s RNA gene chromatin in vitro. Cell 28:781–791
Grantham R, Gautier G, Gouy G, Mercier M, Pare A (1980) Codon catalog usage and the genome hypothesis. Nucleic Acids Res 8:r49–r62
Haldane JBS (1957) The cost of natural selection. J Genet 55: 511–524
Hand R (1978) Eucaryotic DNA: organization of the genome for replication. Cell 15:317–325
Hardison RC, Printz R (1985) Variability within the rabbit C repeats and sequences shared with other SINEs. Nucleic Acids Res 13:1073–1088
Hatton KS, Dhar V, Brown EH, Iqbal MA, Stuart S, Didamo VT, Schildkraut CL (1988a) The replication program of active and inactive multigene families in mammalian cells. Mol Cell Biol 8:2149–2158
Hatton KS, Dhar V, Gahn TA, Brown EH, Mager D, Schildkraut CL (1988b) The temporal order of replication of multigene families reflects chromosomal location and transcriptional activity. In: Cancer cells 6. Eukaryotic DNA replication. Cold Spring Harbor Laboratories, Cold Spring Harbor NY, pp 335–340
Heartlein MW, Stroh H, Ellin M, Latt SA (1987) Chromosome instability associated with human alphoid DNA transfected into the Chinese hamster genome. Am J Hum Genet 41:A122
Hoehn H (1975) Functional implications of differential chromosome banding. Am J Hum Genet 27:676–686
Holmquist GP (1987a) Role of replication time in the control of tissue-specific gene expression. Am J Hum Genet 40:151–173
Holmquist GP (1987b) The magic of cytogenetic technology.

In: Obe G, Basler A (eds) Cytogenetics. Springer-Verlag, Heidelberg, pp 31–47
Holmquist GP (1988a) DNA sequences in G-bands and R-bands. In: Adolph KW (ed) Chromosomes and chromatin. CRC Press, Boca Raton FL, pp 76–121
Holmquist GP (1988b) Mobile genetic elements in G-band and R-band DNA. In: Daniel A (ed) The cytogenetics of mammalian autosomal rearrangmeents. Alan R Liss, New York, pp 803–834.
Holmquist GP, Caston LA (1986) Replication time of interspersed repetitive seuqences. Biochim Biophys Acta 868:164–177
Holmquist GP, Dancis B (1979) Telomere replication, kinetochore organizers, and satellite DNA evolution. Proc Natl Acad Sci USA 76:4566–4570
Holmquist G, Gray M, Porter T, Jordan J (1982) Characterization of Giemsa dark- and light-band DNA. Cell 31:121–129
Hwu HR, Robers JW, Davidson EH, Britten RJ (1986) Insertion and/or deletion of many repeated DNA sequences in human and higher ape evolution. Proc Natl Acad Sci USA 83:3875–3879
Ikemura T (1985) Codon usage and tRNA content in unicellular and multicellular organisms. Mol Biol Evol 22:13–34
Ikemura T, Aota S (1988) Global variation in G+C content along vertebrate genome DNA: possible correlation with chromosome band structures. J Mol Biol 203:1–13
Iqbal MA, Plumb M, Stein J, Stein G, Schildkraut CL (1984) Coordinate replication of members of the multigene family of core and H1 human histone genes. Proc Natl Acad Sci USA 81:7723–7727
Iqbal MA, Chinsky J, Didamo V, Schildkraut CL (1987) Replication of proto-oncogenes early during the S phase in mammalian cell lines. Nucleic Acids Res 15:87–103
John B, Miklos GLG (1979) Functional aspects of satellite DNA and heterochromatin. Int Rev Cytol 58:1–114
Kerem BS, Goitein R, Diamond G, Cedar H, Marcus M (1984) Mapping of DNAase I sensitive regions on mitotic chromosomes. Cell 38:493–499
Kettmann R, Meunier-Rotival M, Cortadas J, Cuny G, Ghysdael J, Mammerickx M, Burney A, Bernardi G (1979) Integration of bovine leukemia virus DNA in the bovine genome. Proc Natl Acad Sci USA 76:4822–4826
Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge
Kimura M, Maruyama T (1969) The substitutional load in a finite population. Heredity 24:101–114
Kominami R, Muramatsu M, Moriwaki K (1983) A mouse type 2 *Alu* sequence (M2) is mobile in the genome. Nature 301:87–89
Korenberg JR, Rykowski MC (1988) Human genome organization: Alu, Lines, and the molecular structure of metaphase chromosome bands. Chromosoma 53:391–400
Kowalski J, Cheevers WP (1976) Synthesis of high molecular weight DNA strands during S phase. J Mol Biol 104:603–615
Kuhn EM, Therman E, Denniston C (1985) Mitotic chiasmata, gene density, and oncogenes. Hum Genet 70:1–5
Kuro-o M, Ikebe C, Kohno S (1986) Cytogenetic studies of Hynobiidae (Urodelia). Cytogenet Cell Genet 43:14–18
Latt, SA (1975) Flourescence analysis of late DNA replication in human metaphase chromosomes. Somatic Cell Genet 1: 293–321
Latt SA, Brodie S, Munroe SH (1974) Optical studies of complexes of quinacrine with DNA and chromatin: implications for the fluorescence of cytological chromosome preparations. Chromosoma 49:17–40
Lau YF, Arrighi FE (1981) Studies of mammalian chromosome replication. II. Evidence for the existence of defined chromosome replicating units. Chromosome 83:721–741
Leadon SA, Zolan ME, Hanawalt PC (1983) Restricted repair of aflatoxin B1 induced damage in alpha DNA of monkey cells. Nucleic Acids Res 11:5675–5689
Lewin B (1980) Eucaryotic chromosomes: gene expression. John Wiley and Sons, New York
Lewontin RC (1974) The genetic basis of evolutionary change. Columbia University Press, New York
Macaya G, Theiry JP, Bernardi G (1976) An approach to the organization of eukaryotic genomes at a molecular level. J Mol Biol 108:237–254
Madhani HD, Bohr VA, Hanawalt PC (1986) Differential DNA repair in transcriptionally active and inactive proto-oncogenes: c-abl and c-mos. Cell 45:417–423
Manuelidis L, Ward DC (1984) Chromosomal and nuclear distribution of the 1.9-kb human DNA repeat segment. Chromosoma 91:28–38
Marchionni MA, Roufa DJ (1981) Replication of viral DNA sequences integrated within the chromatin of SV-40-transformed Chinese hamster lung cells. Cell 26:245–258
Martin CH, Meyerowitz EM (1986) Characterization of the boundaries between adjacent rapidly and slowly evolving genomic regions in *Drosophila*. Proc Natl Acad Sci USA 83: 8654–8658
Martin SL, Voliva CF, Hardies SC, Edgell MH, Hutchison III CA (1985) Tempo and mode of concerted evolution in the L1 repeat family of mice. Mol Biol Evol 2:127–140
Matsumoto KI, Murakami K, Okada N (1986) Gene for lysine tRNA may be a progenitor of the highly repetitive and transcribable sequences present in the salmon genome. Proc Natl Acad Sci USA 83:3156–3160
Mellon I, Bohr VA, Smith CA, Hanawalt PC (1986) Preferential DNA repair of an active gene in human cells. Proc Natl Acad Sci USA 83:8878–8822
Mizuno S, Macgregor HC (1974) Chromosomes, DNA sequences, and evolution in salamanders of the genus *Plethodon*. Chromosoma 48:239–296
Mouchiroud D, Fichant G, Bernardi G (1987) Compositional compartmentalization and gene composition in the genome of vertebrates. J Mol Evol 26:198–204
Muto A, Osawa S (1987) The guanine and cytosine content of genomic DNA and bacterial evolution. Proc Natl Acad Sci USA 84:166–169
O'Brien SJ, Reeves RH, Simonson JM, Eichelberger MA, Nash WG (1984) Parallels of genomic organization and of endogenous retrovirus organization in cat and man. Dev Genet 4:341–354
Odum EP (1969) The strategy of ecosystem development. Science 164:262–270
Ohno S (1972) Evolutional reason for having so much junk DNA. In: Pfeiffer RA (ed) Modern aspects of cytogenetics: constitutive heterochromatin in man. F. K. Schattauer Verlag, Stuttgart, pp 169–190
Okada TA, Comings DE (1974) Mechanisms of chromosome banding III. Similarity between G-bands of mitotic chromosomes and chromomeres of meiotic chromosomes. Chromosoma 48:65–71
Olofsson B, Bernardi G (1983) Organization of nucleotide sequences in the chicken genome. Eur J Biochem 130:241–245
Olson JS (1958) Rates of succession and soil changes on southern Lake Michigan sand dunes. Bot Gaz 119:125–170
Orgel LE, Crick FHC (1980) Selfish DNA: the ultimate parasite. Nature 284:604–607
Perkins DD, Barry EG (1977) The cytogenetics of neurospora. Annu Rev Genet 19:133–224
Perrin P, Bernardi G (1987) Directional fixation of mutations in vertebrate evolution. J Mol Evol 26:301–310
Rogers JH (1985) The origin and evolution of retroposons. Int Rev Cytol 93:187–279
Sahasrabuddhe CG, Pathak S, Hsu TC (1978) Responses of

mammalian metaphase chromosomes to endonuclease digestion. Chromosoma 69:331–338

Sakamoto K, Okada N (1985) Rodent type 2 Alu family, rat identifier sequence, rabbit C-family and bovine or goat 73-bp repeat may have evolved from tRNA genes. J Mol Evol 22:134–140

Salinas J, Zerial M, Filipski J, Crepin M, Bernardi G (1987) Nonrandom distribution of MMTV proviral sequences in the mouse genome. Nucleic Acids Res 15:3009–3022

Sawada I, Schmid CW (1986) Primate evolution of the alpha-globin gene cluster and its Alu-like repeats. J Mol Biol 192: 693–709

Sawyer JR, Hozier JC (1986) High resolution of mouse chromosomes: banding conservation between man and mouse. Science 232:1632–1635

Schempp W, Schmid M (1981) Chromosome banding in amphibia VI. BrdU-replication patterns in Anura and demonstration of XX/XY sex chromosomes in *Rana esculenta*. Chromosoma 83:711–720

Schimke RT, Sherwood SW, Hill AB, Johnston RN (1986) Overreplication and recombination of DNA in higher eukaryotes: potential consequences and biological implications. Proc Natl Acad Sci USA 83:2157–2161

Schmid CW, Jelinek WR (1982) The Alu family of dispersed repetitive sequences. Science 216:1065–1070

Schmid M (1980) Replication banding in Amphibia. IV. Differentiation of GC- and AT-rich chromosome regions in Anura. Chromosoma 77:83–103

Schmid M, de Almeida CG (1988) Chromosome banding in Amphibia. Chromosoma 96:283–290

Schonberg SA, Fukuyama K, Hara N, Epstein AL (1987) Monoclonal antibodies to human nuclear proteins as probes for human chromosome structure. Am J Hum Genet 41:A138

Schweizer D (1981) Counterstain-enhanced chromosome banding. Hum Genet 57:1–14

Seuanez HN (1979) The phylogeny of human chromosomes. Springer-Verlag, Berlin

Shiraishi Y, Taguchi T, Ohta Y, Hirai K (1985) Chromosomal localization of the Epstein-Barr virus (EBV) genome in Bloom's syndrome B-lymphoblastoid cell lines transformed with EBV. Chromosoma 93:157–164

Smith JM (1968) "Haldane's dilemma" and the rate of evolution. Nature 219:114–116

Solomon MJ, Strauss F, Varshavsky A (1986) A mammalian high mobility group protein recognizes any stretch of six A-T base pairs in duplex DNA. Proc Natl Acad Sci USA 83:1276–1280

Sonigo P, Wain-Hobson S, Bougueleret L, Tiollais P, Jacob F (1987) Nucleotide sequence and evolution of ETn elements. Proc Natl Acad Sci USA 84:3768–3771

Soriano P, Macaya G, Bernardi G (1981) The major components of the mouse and human genomes. Eur J Biochem 115: 235–239

Soriano P, Meunier-Rotival M, Bernardi G (1983) The distribution of interspersed repeats is nonuniform and conserved in the mouse and human genomes. Proc Natl Acad Sci USA 80:1816–1820

Stanton LW, Schwab M, Bishop JM (1986) Nucleotide sequence of the human N-myc gene. Proc Natl Acad Sci USA 83:1772–1776

Stineman M (1981) Chromosomal replication in *Drosophila virilis* III. Organization of active origins in the highly polytene salivary gland cells. Chromosoma 82:289–307

Stock AD, Mengden GA (1975) Chromosome banding pattern conservatism in birds and nonhomology of chromosome banding patterns between birds, turtles, snakes and amphibians. Chromosoma 50:69–77

Strayer D, Heintz N, Roeder R, Gillespie D (1983) Three organizations of human DNA. Proc Natl Acad Sci USA 80: 4770–4774

Sueoka N (1988) Directional mutation pressure and neutral molecular evolution. Proc Natl Acad Sci USA 85:2653–2657

Tanford C (1961) Physical chemistry of macromolecules. John Wiley and Sons, New York

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

Trabuchet G, Chebloune Y, Savatier P, Lachuer J, Faure C, Verdier G, Nigon VM (1987) Recent insertion of an *Alu* sequence in the beta-globin gene cluster of the gorilla. J Mol Evol 25:288–291

Vizard DL, Rosenberg NL (1984) Temporal replication of an interspersed repeated sequence of mouse DNA. Biochim Biophys Acta 782:402–407

von Kiel K, Hameister H, Sommssich IE, Adolph S (1985) Early replication banding reveals a strongly conserved functional pattern in mammalian chromosomes. Chromosoma 93:69–76

Weiner AM, Deininger PL, Efstratiadis A (1986) Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Annu Rev Biochem 55:631–661

Willard C, Nguyen HT, Schmid CW (1987) Existence of at least three distinct *Alu* subfamilies. J Mol Evol 26:180–186

Yunis JJ (1981) Mid-prophase human chromosomes. The attainment of 2,000 bands. Hum Genet 56:293–298

Zolan ME, Cortopassi GA, Smith CA, Hanawalt PC (1982) Deficient repair of chemical adducts in alpha DNA of monkey cells. Cell 28:613–619

Zuckerkandl E (1976) Evolutionary processes and evolutionary noise at the molecular level II. A selectionist model for random fixations in proteins. J Mol Evol 7:269–311

Zuckerkandl E (1986) Polite DNA: functional density and functional compatibility in genomes. J Mol Evol 24:12–27

Zuckerkandl E, Pauling L (1965) Molecules as documents of evolutionary history. J Theor Biol 8:357–366

Received August 22, 1988/Revised and accepted December 5, 1988