

Review Article: Chromosome Bands, Their Chromatin Flavors, and Their Functional Features

G. P. Holmquist

Department of Biology, City of Hope Medical Center, Duarte, CA

Summary

To show that the input pattern of chromosomal mutations is highly organized relative to the band patterns along human chromosomes, a new term, "metaphase chromatin flavor," is introduced. Five different flavors of euchromatic metaphase bands are cytologically identified along a human ideogram. These are G-bands and, based upon combinations of extreme Alu richness and GC richness, four different R-band flavors. The two flavors with extremely GC-rich components, traditionally called "T-bands," represent only 15% of all bands. However, they contain 65% of mapped genes, 19 of 25 mapped oncogenes, most cancer-associated rearrangements, evolutionary rearrangements, meiotic chiasmata, and X-ray-induced breaks. Flavors with extremely Alu-rich flavors are also involved in melphalan-induced rearrangements, pachytene stretching, and mitotic chiasmata. Frequencies of CpG islands, CCGCC boxes, retroposon families, and genes are characteristic to each chromatin flavor and will facilitate alignment of genome sequences onto ideograms of chromatin flavor. The influence of chromatin flavor on the evolution of a gene's sequence is so strong that one can infer the flavor of the band in which a gene resides from the sequence of the gene itself. Correlation coefficients for many pairs of mapped genetic variables, while globally high, are quite low within bands of one flavor, implicating a concerted mode of evolution for bands of one chromatin flavor.

Introduction

The human genomic sequence is quite ordered. Fluorochrome bandings, using base composition-sensitive fluorochromes such as quinacrine, or in situ hybridization bandings, using retroposon probes, show base composition and retroposon distributions as glowing patterns of sequence order; this order is organized congruent to the familiar Giemsa band pattern (Summer 1990). Ordered mutational output (natural selection) and ordered mutational input are the only forces which, acting over evolutionary time, can possibly order our genome's sequence (Holmquist 1990). Historically, mutational input has been assumed to be random, leaving natural selection as the sole paradigm

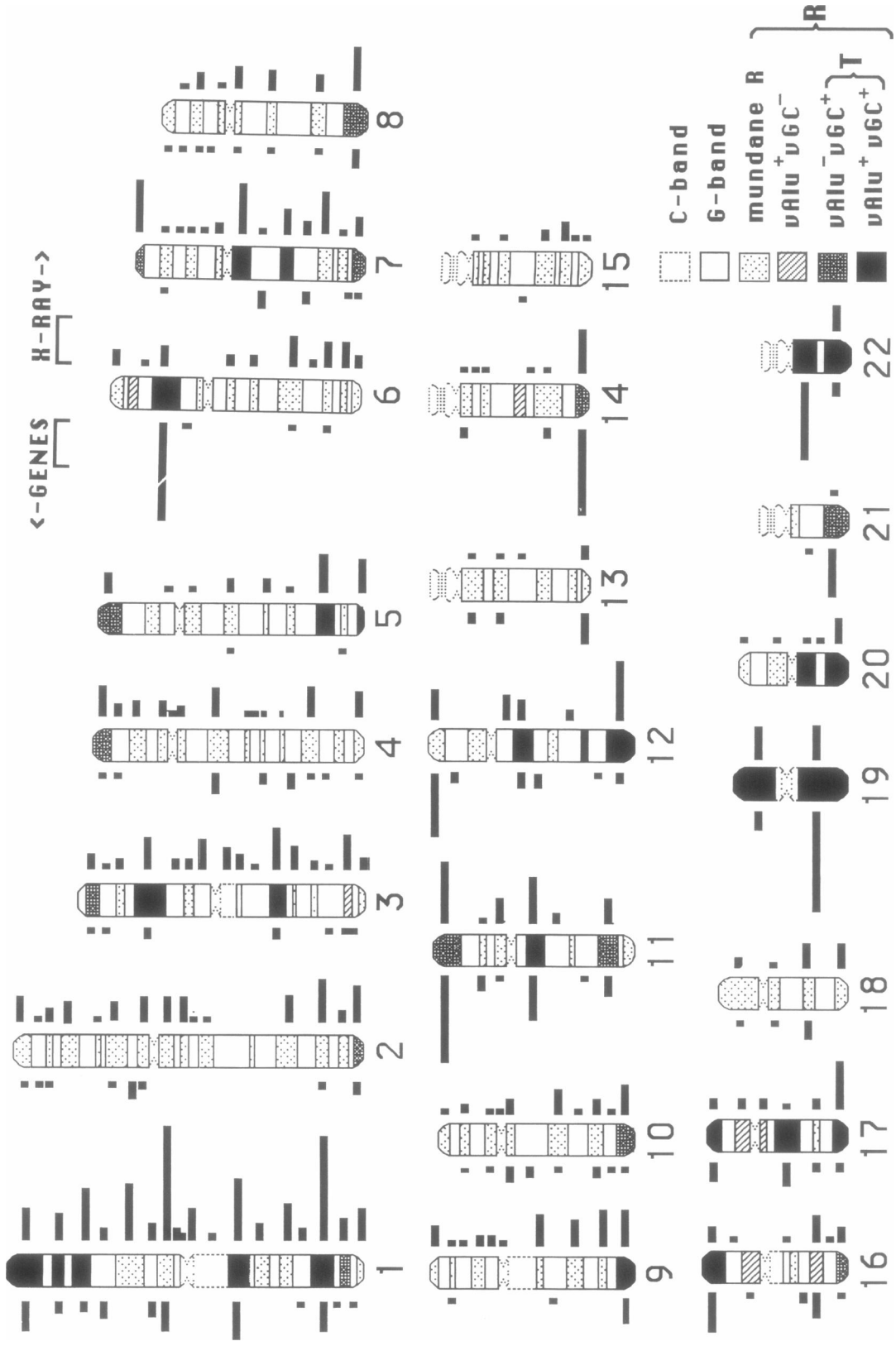
to explain accumulation of sequence order. Recent evidence shows that the input of molecular mutations varies from one region to another along the genome (Hanawalt 1990), but how this relates to chromosome bands has been uncertain. By showing that input of *chromosomal mutations* is highly organized relative to the band patterns along human chromosomes, we can infer that the input of *molecular mutations* is similarly organized. This explains the historical accumulation of localized base-composition differences and several other localized sequence differences; they are molecular fossils documenting a history of ordered mutations.

To recognize patterns of mutational input, I first identified five different chromatin flavors of metaphase bands: G-bands and four different R-band flavors. A metaphase ideogram of these chromatin flavors is presented, and the frequency distribution of almost every phenomena which has been mapped at single-metaphase-band resolution is superimposed upon the flavor ideogram. Metaphase bands of any one chromatin

Received December 10, 1991; revision received February 21, 1992.

Address for correspondence and reprints: G. P. Holmquist, Department of Biology, City of Hope Medical Center, Duarte, CA 91010.

© 1992 by The American Society of Human Genetics. All rights reserved. 0002-9297/92/5101-0004\$02.00



flavor show distinct frequencies of these mapped phenomena; the cytological distinction of chromatin flavor is indeed related to function. These newfound relationships are used to model how human genome sequence data and the base-composition isochores of Bernardi et al. (1985) will correspond to high-resolution ideograms. Finally, expectations from several models of chromosome evolution are compared with the data. Concerted mechanisms of evolution emerge as the forces driving formation of these band-congruent sequence patterns.

Chromatin Flavors

“A chromosome band designates a chromatid segment which presents distinct functional and structural characteristics resulting in specific staining properties after various banding techniques” (Drouin et al. 1991*b*, p. 65). “Chromatin flavor,” a new term, describes how a metaphase band stains by each of several techniques (Holmquist 1990). In human euchromatin, the major historically distinguished chromatin flavors were G-bands (GTG⁺, positive by G-banding using trypsin and Giemsa) and R-bands (GTG⁻). R-bands, originally defined as the reverse of quinacrine bright bands even before GTG-band methodology was developed (Dutrillaux and Lejeune 1971), are of several varieties. RHG-banding (Reverse banding by heat and Giemsa staining) produces roughly the reverse pattern of quinacrine banding or GTG-banding. RHG-banding is accomplished by heating the fixed chromosomes at 87.5°C for 5–30+ min in physiological saline (Dutrillaux and Covic 1974; Dutrillaux 1977). Here, the AT-richer DNA presumably denatures faster, reducing its affinity for Giemsa stain. With increasing heat-treatment times, sets of originally dark R-bands gradually fade, leaving T-bands (THG⁺-bands) as an extremely heat-resistant subset of R-bands

(Dutrillaux 1977). While the base composition of R-band DNA is more GC rich than G-band DNA, some R-bands are more GC rich than others. The very GC-richest 27% of R-bands were selectively displayed by simultaneously using the AT-specific fluorochrome DAPI and the GC-specific fluorochrome chromomycin A₃, followed by subtractive enhancement techniques (Ambros and Sumner 1987). At 300-bands/haploid-genome resolution, these very GC-rich bands were shown to correspond almost exactly to THG⁺-bands (Holmquist 1990), further supporting the proposed base-composition dependence of RHG-banding. Henceforth the terms “T-band” and “vGC⁺ band” will be used interchangeably for any metaphase band which is positive by either T-banding or vGC-banding.

In situ hybridization using a probe from the Alu family of short interspersed repeats illuminates an overall R-band pattern (Manuelidis and Ward 1984) with 28% of the R-bands—very Alu-rich bands (vAlu⁺ bands)—showing a much stronger signal than the average R-band (Korenberg and Rykowski 1988). After the vAlu⁺ and vGC⁺ bands₃₀₀ are indicated on an ideogram (fig. 1), we see that four classes of R-bands exist; these classes differ by their combination of extreme Alu richness and extreme GC richness.

The resolution at which a band is resolved will be indicated by a subscript, e.g., “vGC⁺₃₀₀” will describe a vGC⁺ band observed by using short metaphase chromosomes wherein only 300 bands/haploid genome could be resolved. We will discuss only five different chromatin flavors of euchromatic metaphase bands: one G-band₃₀₀ flavor and four different R-band₃₀₀ flavors. All G-bands are GTG⁺, with no further subclassification noted. R-bands are GTG⁻, with few exceptions (Drouin and Richer 1985; Drouin et al. 1991*b*). Mundane R-bands are GTG⁻ vAlu⁻ vGC⁻. The other three R-band flavors are special. GTG⁻

Figure 1 R-banding ideogram₃₀₀ (R-bands at 300-bands/genome resolution), along with gene frequency (*left histogram*) and X-ray break frequency (*right histogram*). Histograms are for 334 mapped genes (vertical bar = 10 genes) (Kidd et al. 1989) and 660 X-ray-induced exchanges and breaks (bar = 10 exchanges) (Barrios et al. 1989). C-bands and five metaphase chromatin flavors (G-band, mundane R, vAlu⁺ vGC⁻, vAlu⁻ vGC⁺, and vAlu⁺ vGC⁺) of euchromatic bands are distinguished by their response to the following banding techniques: C-banding; G-banding; T-banding, or a very GC-rich DNA-detection scheme; and in situ hybridization using an Alu probe. Flavors can be resolved at bands₄₅₀ (Holmquist 1990), but the lower bands₃₀₀ resolution of the Paris conference (Paris-conference-1971, 1972) was used here to accommodate mapping data for table 1. Criteria for flavor designations are from published photos, explained elsewhere (Holmquist 1990). Very Alu-rich R-bands, vAlu⁺, were determined from fig. 1 of Korenberg and Rykowski (1988). R-bands were designated as very GC-rich R-bands, vGC⁺, either if they were very GC rich by fluorochrome staining (from both fig. 4 of Ambros and Sumner [1987] and color-separated reprints of this figure as supplied by Peter Ambros) or if they were T-bands on the basis of classical THG- or THA-banding (from fig. 4 of Dutrillaux [1977], figs. 1 and 3 of Dutrillaux [1973], and fig. 1 of Bobrow and Madan [1973]). All flavor determinations (Holmquist 1990) were made before any data in table 1 were accumulated; they were not changed thereafter.

vAlu⁺ vGC⁻ bands differ from mundane R-bands by being very Alu rich. GTG⁻ vAlu⁻ vGC⁺ (only GC rich) or GTG⁻ vAlu⁺ vGC⁺ (both very Alu rich and very GC rich) are flavors which together are called "T-bands." With chromatin flavors now defined at 300-bands/genome resolution, I now ask a simple question: Are these purely cytological distinctions, as previously made by me (Holmquist 1990) from published figures and as presented in figure 1, significantly related to the frequency of functional features along the chromosome?

Mapping Events to Chromatin Flavors

Figure 1 is an R-band₃₀₀ ideogram with the R-bands indicated by various shades of darkness, to provide their representation with positive visual information. Special chromatin flavors are distinguished by the darker shadings. Frequency histograms of mapped genes or X-ray-induced exchanges are shown alongside the chromosomes (fig. 1). The darker-shaded flavors are where the action is. Table 1 summarizes the frequency of these and other phenomena for bands in each flavor grouping, and the remainder of this section details each of table 1's columns.

The 334 genes which have been mapped to an individual autosomal band₃₀₀ (Kidd et al. 1989) are rare in G-bands and are very frequent in T-bands (fig. 1 and table 1); brain genes are significantly absent from this sample. Gene density is independent of the vAlu⁺ flavor component and is six to seven times higher in T-bands than in mundane R-bands.

The densities of mapped oncogenes (Trent et al. 1989) are also high in T-bands, especially the vAlu⁻ vGC⁺ T-bands (table 1). Cancer exchange breakpoints (Trent et al. 1989), mapped chromosome exchanges associated with a specific neoplasia, also cluster in T-bands (table 1). The distribution of cancer exchange breakpoints more closely parallels oncogene density than it parallels gene density, in that the breakpoints are significantly more frequent in vAlu⁻ vGC⁺ bands than in vAlu⁺ vGC⁺ bands (table 1).

Blood cells from X-irradiated patients show chromosome aberrations in the subsequently induced metaphases. Breakpoints and exchange points were identified and mapped (Barrios et al. 1989); about half of these were translocation exchanges. The mapped breakpoints and exchange points are significantly clustered in T-bands and are scarce in G-bands (fig. 1 and table 1). Mapped hot spots from six separate studies of radiation-induced chromosome aberrations (mostly

chromosome exchanges) were also tabulated (Barrios et al. 1989). Tissue sources included *in vivo* and *in vitro*-irradiated lymphocytes, *in vivo*-irradiated fibroblasts, and lymphocytes from atomic-bomb survivors. T-bands are the more frequent sites of radiation-induced aberration hot spots, and G-bands are the least frequent (table 1). While hot spots and cold spots for such exchanges were known, their distribution congruent to a chromosome band pattern is new.

T-bands represent only 15% of all bands but contain 65% of mapped genes, 19 of 25 mapped oncogenes, 60% of cancer-associated breaks, and 42% of X-ray-induced breaks. Quite expectedly, many reports have shown that pairs of such mapped phenomena—e.g., oncogenes and cancer exchanges—have distributions which are both nonrandom and mutually coincident; these reports used these data, along with the false expectation that such coincident distributions should be rare, to argue for direct causal relations between the variables tested (Yunis and Soreng 1984; Heim and Mitelman 1987). Table 1 demonstrates that such coincident distributions are surprisingly very common. This commonness precludes support, by coincident distributions, for a cause-effect relationship between the two variables.

Forty-seven bands in the human ideogram were identified as having been involved in karyotype evolution; they were sites of chromosome exchanges or C-band insertions and were fixed in a species during primate evolution (Miro et al. 1987). These 47 bands are also frequently T-bands (table 1), even when telomeric bands are excluded from the calculation (Holmquist 1990). Selectionists would have predicted the opposite distribution; since exchanges are usually detrimental to genes at the exchange point (DeMarini et al. 1989), fixed rearrangements should preferentially involve gene-depleted regions: G-bands and mundane R-bands. One could explain the observed distribution by positing that T-band exchanges are advantageous. However, since primates' evolutionary exchanges also have a frequency distribution among the various chromatin flavors which is very similar to the distribution of X-ray-induced exchanges (table 1), the combined data are more easily explained as follows: Chromosomal mutations most frequently occur in T-bands, and neutral drift passively fixes chromosomal mutations in proportion to their occurrence.

Large genomic chunks have been conserved intact during mammalian radiation; cats, cows, humans, and even mice share many nearly identical chromosome segments containing the same genes and having

Table I

Properties Mapped to Chromatin Flavors

FLAVOR ^a	OBSERVED EXPECTED ^b													AVERAGE NO. OF SUBBANDS N = 1,114		
	ACTUAL FREQUENCY (%)			OBSERVED EXPECTED ^b												
	Frequency	All Genes	Oncogene	Cancer Exchange	X-ray Exchange	X-ray Hot Spot	Evolutionary Exchange	Mspl Breakpoint	Fragile X Sites	Chiasmata	Pachytene Elongated	Fanconi	Melphalan Break	Melphalan Exchange	Melphalan Hot Spot	
	N = 268	N = 334	N = 25	N = 181	N = 626	N = 92	N = 47	N = 999	N = 82	N = 2,070	N = 18	N = 287	N = 759	N = 891	N = 29	
G	1.00	.24	.48	.29	.17	.17	.51	.22	.68	.33	.00	.61	1.23	1.04	.14	2.8
G'	19.4	.39	.00	.23	.25	.22	.55	.23	1.07	.32	.00	1.20	1.66	1.63	2.13	3.3
R-mundane	37.7	.61	.21	.65	1.23	.95	.90	1.18	.97	1.18	.15	1.02	.36	.49	.55	4.3
vAlu ⁺ vGC ⁺	2.6	1.00	.80	1.48	.98	.42	1.65	1.23	.92	1.23	4.27	.81	.66	.99	1.32	4.0
vAlu ⁺ vGC ⁺	10.1	1.00	4.31	2.91	2.81	4.42	2.31	3.19	1.09	2.20	8.25	1.80	1.83	1.75	2.74	7.2
vAlu ⁻ vGC ⁺	5.2	1.00	3.44	5.92	2.60	1.87	2.86	1.95	2.34	3.01	.00	.54	.68	.69	.66	6.7

FREQUENCY OF (%)		
Fanconi	N = 67	2.3
Melphalan Break	N = 244	5.0
Melphalan Exchange	N = 281	4.1
Melphalan Hot Spot	N = 12	6.9
Fanconi	N = 67	14.8
Melphalan Break	N = 244	15.4
Melphalan Exchange	N = 281	19.8
Melphalan Hot Spot	N = 12	27.6
Fanconi	N = 67	6.3
Melphalan Break	N = 244	11.8
Melphalan Exchange	N = 281	7.3
Melphalan Hot Spot	N = 12	6.9

^a G' bands are simply G-bands which neighbor an R-band of special flavor. Fanconi and melphalan data were from RHG-banded chromosomes, so an optical illusion forced many T-band events to appear in neighboring G'-bands. In the last three rows, G'-bands are subclassified according to the flavor of their adjacent R-bands. The frequency of each type of G' band is followed by the percentage of total data in each G' subclass and is shown for R-banding data columns only. Mapped gene frequency in 6p21 was reduced from 35 (Kidd et al. 1989) to 21, to compensate for oversampling of this band.

^b The ideogram number of each autosomal band³⁰⁰, that band's chromatin flavor from fig. 1, and other mapped event frequencies as taken from the literature were entered into a data base. (Observed event frequency) in all bands of one flavor class)/(expected event frequency) if the event distribution were independent of chromatin flavor) is shown and should approximate 1.0 if the mapped events occurred randomly along the chromosome.

the same GTG-band and replication-band patterns (von Kiel et al. 1985; O'Brien et al. 1988; Threadgill et al. 1991). In addition, interstitial vGC⁺ bands in humans often correspond to terminal bands of acrocentrics in other primates (Ambros and Sumner 1987; Dutrillaux 1979). Noticing that karyotype evolution primarily involves centromeres, telomeres, and T-bands, I suggest that conserved chunks are not functionally conserved linkages but are simply genomic regions which, by chance, are flanked by exchange-prone elements.

The axial distribution of some chromatin flavors is nonrandom. vAlu⁻ vGC⁺ bands are significantly absent from the centromere proximal third of all chromosome arms, while vAlu⁺ vGC⁻ bands are absent from the distal third (table 2). T-bands were initially named for their terminal disposition (Dutrillaux 1973). Their very GC richness was suggested to arise at all chromosome ends, with evolutionary rearrangements subsequently internalizing interstitial T-bands as relics of a telomeric legacy (Ambros and Sumner 1987; Dutrillaux 1979). Without invoking a recent Alu accumulation in centric T-bands, this hypothesis is now inconsistent with the telomeric disposition of only vAlu⁻ vGC⁺ bands, and not vAlu⁺ vGC⁺ bands (table 2). Axial gradients of diminishing chromomere size (Lima-De-Faria 1975, 1976) or "equilocal" positioning of heterochromatic blocks (Schweizer and Loidl 1987) are also functions of centromere distance in a variety of species. These spatial constraints seem related to the three-dimensional organization of the somatic or meiotic interphase nucleus (Schweizer and Loidl 1987). Such interphase data for T-bands are lacking (Nakayasu and Berezney 1991).

Cells treated with the restriction endonuclease *MspI* reveal its clastogenicity by showing chromosome and chromatid breaks in their subsequent mitoses. Mapped breaks from metaphases observed 6 h posttreatment and 20 h posttreatment (Porfiro et al. 1989) are combined in table 1. The mapped breaks are infrequent in G-bands, consistent with *MspI* cutting late-replicating G-band DNA (after terminal digestion, $M_n = 4.3$ kb) less frequently than it cuts early-replicating DNA ($M_n = 2.5$ kb) (Holmquist 1988). Unlike X-rays, *MspI* breaks chromosomes significantly more frequently in vAlu⁺ vGC⁺ bands than in vAlu⁻ vGC⁺ bands.

Fragile sites, both rare ones and common ones (Sutherland and Ledbetter 1989), are combined in table 1. They occur with about equal frequency in all chromatin flavors (table 1). Although distributions of cancer-specific rearrangements (Yunis and Soreng 1984) and distributions of evolutionary breakpoints (Miro et al. 1987) have been reported to be coincident with fragile sites and T-bands, respectively, fragile-site distributions and T-bands are not themselves coincident (table 1).

Synapsis of homologues is a prerequisite for chiasma-associated crossing-over during meiosis and differs between R-bands and G-bands. Synapsis initiates in R-bands (Ashley 1990). In murine translocation heterozygotes, synaptic initiation in R-bands₃₀₀ is homology dependent, while it is homology independent in G-bands₃₀₀ (Ashley 1990).

At pachytene, when synapsis is complete, chromosomes are most apparent. These are local beadlike coilings or thickenings of the chromosome which correspond to the mitotic G-bands (Okada and Comings 1974; Ambros and Sumner 1987). This correspon-

Table 2

Flavor Frequency and Meiotic Chiasma Frequency

FLAVOR	NO. OF BANDS			NO. OF CHIASMATA			NO. OF CHIASMATA/BAND ± SEM		
	Distal	Medial	Proximal	Distal	Medial	Proximal	Distal	Medial	Proximal
G	10	26	30	31	97	49	3.10 ± 1.4	3.73 ± 1.1	1.63 ± .5
G'	16	23	9	50	55	15	3.12 ± 1.1	2.39 ± .9	1.67 ± .9
R-mundane	25	34	38	370	358	174	14.8 ± 1.6	10.5 ± 1.2	4.58 ± 1.0
vAlu ⁺ vGC ⁻ ...	0	3	4	0	35	31	...	11.7 ± 5.9	7.75 ± 4.6
vAlu ⁺ vGC ⁺ ...	10	6	8	251	77	51	25.1 ± 2.2	12.8 ± 2.3	6.38 ± 2.1
vAlu ⁻ vGC ⁺ ...	13	1	0	309	16	0	23.8 ± 2.8	16.0	...

NOTE.—Data are from table 1 and are here recalculated for the centromere distal, medial, and proximal thirds of each autosome arm. vAlu⁻ vGC⁺ bands are telic, vAlu⁺ vGC⁺ bands are evenly distributed, and vAlu⁺ vGC⁻ bands are centric. Chiasma density has two obvious components: distance from the centromere and chromatin flavor.

dence allows pachytene events to be mapped onto mitotic ideograms. Eighteen interchromomeric regions (R-bands) are exceptionally elongated at pachytene (Luciani et al. 1988), and 17 of these 18 correspond to bands with a $vAlu^+$ flavor component (table 1). Why very Alu-rich chromatin shows this stretching is unknown.

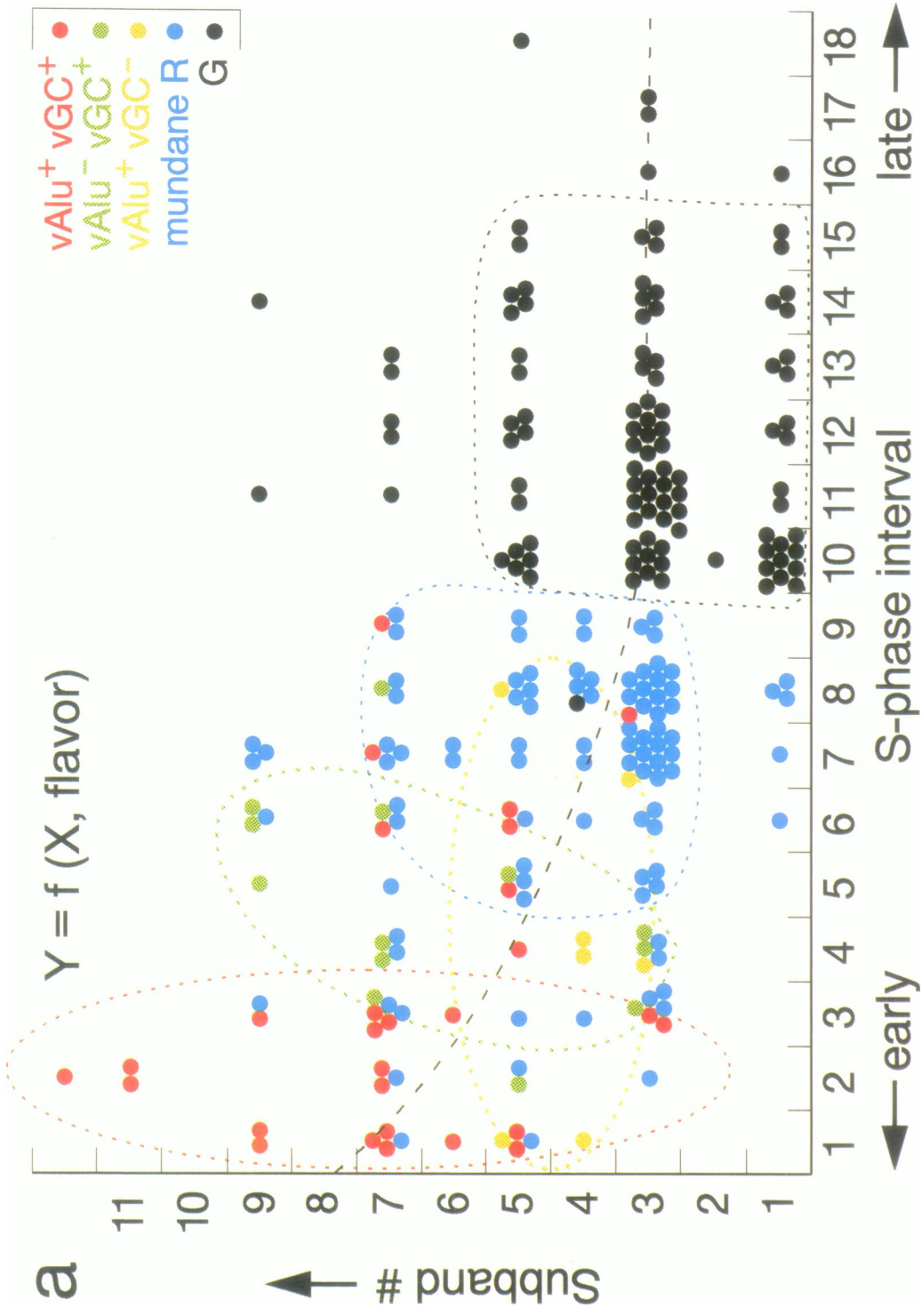
Meiotic chiasmata are the cytological manifestation of crossing-over. Their frequency along the genome was mapped by Fang and Jagiello (1988) and is a function of at least two variables: distance from the centromere (Laurie and Hulten 1985) and chromatin flavor (table 2). Distal bands of any one flavor generally have two or three times the chiasma frequency of centromere proximal bands of the same flavor; this reflects the axial component of chiasma density. However, within any one of the three equidistance sectors, chiasma density retains a strong dependence on chromatin flavor (table 2), with T-bands having six times the chiasma frequency of G-bands. Both chiasma frequency and gene frequency follow the chromatin flavor relationship, C-bands = $0 < G$ -bands $<$ mundane R-bands $<$ T-bands (fig. 2*b*). If the function of meiotic recombination is indeed to segregate genes, then the mechanism regulating the distribution of meiotic recombination frequency effectively serves to segregate genes while ignoring junk (Ohno 1972) DNA.

An optical illusion has historically clouded interpretations from cytogenetic mapping data. The human eye interprets events as occurring in unstained bands (Savage 1977), independent of the band of origin. For example, while radiation-induced exchanges seemed to be predominantly in R-bands after analysis of GTG-banded chromosomes (Holmberg and Jonasson 1973), when similar chromosomes were analyzed after RHG-banding, the breaks appeared predominantly in G-bands. When the same chromosomes were sequentially R- and G-banded during analysis, breaks seemed to appear mostly at G-band/R-band junctions (Buckton 1976; Dubos et al. 1978). The beauty of table 1 is that, within all R-band flavors, the illusion works similarly. Even though data so far discussed were gathered by GTG-banding alone, the T-band-versus-R-mundane conclusions were unaffected by scoring illusions. We can conclude, for example, that X-ray-induced breaks are more frequently associated with (in, at the junctions of, or very near) T-bands than with mundane R-bands.

The remaining break-distribution data in table 1 were obtained from RHG-banded chromosomes. The aberrations occurred in lymphocytes either spontane-

ously in Fanconi anemia patients (Dutrillaux et al. 1977) or after induction by the alkylating agent melphalan (Mamuris et al. 1989). In both cases, aberrations were more frequently scored in G-bands than in R-bands (table 1). While most of the melphalan-induced breaks and exchanges seemed to occur in G-bands by cytological localization of the exchange points, they were proved to be associated with R-bands, by showing that, when each chromosome was considered, exchange per chromosome/chromosome length and R-band content were positively correlated (Mamuris et al. 1989). The effect of the optical illusion, forcing R-band-associated breaks to be scored in G-bands, is evident from the special G' category of bands (table 1). G' -bands are not a unique chromatin flavor but are an analysis tool. They are simply those G-bands which border R-bands of special flavor. Any event which was associated with a special R-band instead of with a mundane R-band would, because of the optical illusion acting on R-banded chromosomes, be preferentially scored in a bordering G' -band instead of in a G-band. Since the Fanconi and melphalan-break frequencies in the G' category are significantly elevated relative to the G category (table 1), I presume that this elevation was caused by the optical illusion acting on events associated with special R-bands. At the bottom of table 1, G' -bands are further subdivided according to the special flavor of their bordering band, so that the optical illusion's diversion of scored events from each special R-band flavor can be discerned in G' -bands. Table 1 shows that melphalan-exchange distributions and the hot spots for melphalan-induced breaks and exchanges are different from that of radiation, with the $vAlu^+ vGC^-$ flavor showing even more associated events than does the $vAlu^- vGC^+$ flavor. Similar results showing a high frequency of mitotic chiasmata associated with $vAlu^+ vGC^-$ and $vAlu^+ vGC^+$ bands were previously obtained from an analysis (Holmquist 1990) of mapped somatic crossover events (Kuhn et al. 1985) from Bloom syndrome patients.

In conclusion, the five chromatin flavors presented in figure 1 as value judgments of the author are important distinctions, because the five flavors have significantly different frequencies of mapped chromosomal events. Distributions of chromosomal mutations are coincident with chromatin flavor patterns, and the relative mutation frequency associated with a particular chromatin flavor depends on the clastogen used or the metabolic defect involved.



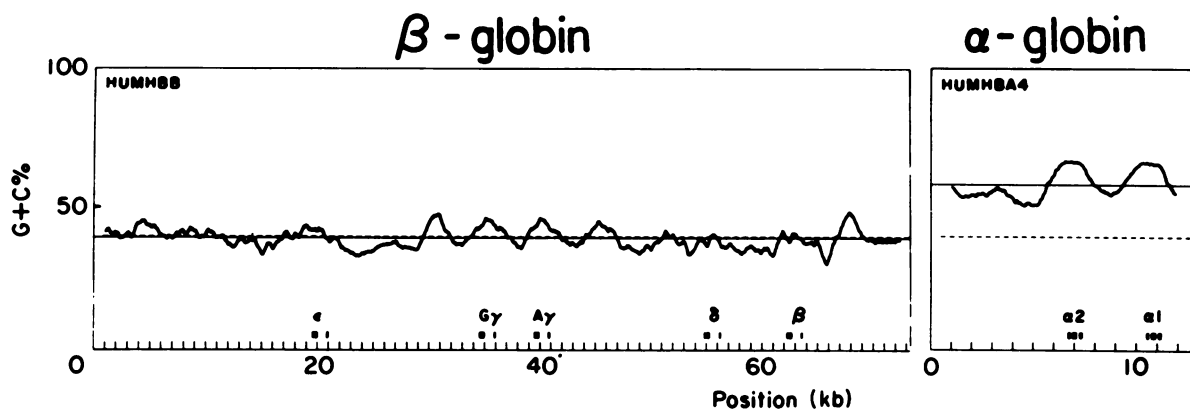


Figure 3 Base composition, averaged from a 2-kb interval moved in increments of 10 bp (Ikemura and Aota 1988). Local base-composition homogeneity, the essence of isochores, is evident throughout the alpha-globin locus (i.e., H3 isochore) and beta-globin locus (i.e., L2 isochore) (Bernardi et al. 1985) and indicates that continuous base-composition maps from genome sequencing will be superimposable upon ideogram maps, as in fig. 4. (Redrawn with permission from Ikemura and Aota [1988].)

Sequence Motifs and a Model of the Chromosome's Sequence

I have, until now, only discussed properties of low-resolution metaphase bands. To show how these data are relevant to genome sequencing, I must relate the flavor ideogram at 300 bands/genome (10-mb) resolution to the higher(2-mb)-resolution prophase bands and then relate prophase bands to DNA sequence distributions. I will contend that metaphase bands and prophase bands represent two different levels of chromosome organization and are related in name only.

The ideogram reflects genomic compartments (Bernardi et al. 1985), long stretches of locally homogeneous base composition (fig. 3) which, when contiguously pieced together, should appear as shown in figure 4. Human DNA, when fragmented and fractionated by ligand-assisted buoyant-density centrifugation, does not reveal a continuous distribution of base-composition differences. Instead, it resolves into five quite distinct modal density fractions (Macaya et al. 1976; Cuny et al. 1981): two light isochore fractions L1 and L2 and three heavy isochore fractions H1, H2, and H3 (fig. 4). The genome fraction present in each isochore fraction is independent of the length of the starting material, 16–200+ kb, indicating long stretches of rather homogeneous base composition. A pattern of base-composition differences along chromosomes has been resolved at a resolution of 300–450 bands/haploid genome (10-mb) (Korenberg and Engels 1978). It is visualized using AT-specific fluorochromes such as quinacrine, 33258 Hoechst, and DAPI or by GC-specific fluorochromes such as

chromomycin A₃ (Schweizer 1981). Ideograms from base-composition banding correspond almost exactly to standard GTG_{300–450} ideograms (Schweizer 1981). As modeled in figure 4, the isochores are presumed to

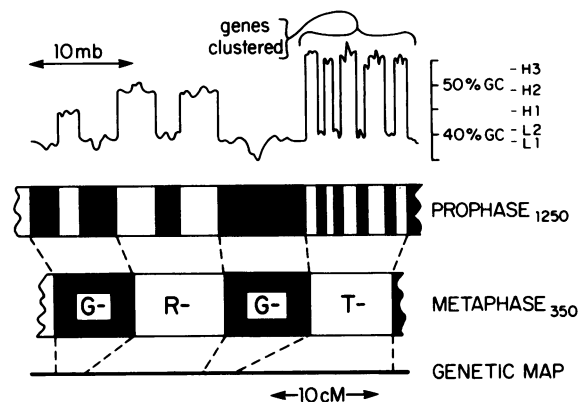


Figure 4 Model of base-composition variation along the genome, and its relation to cytogenetic and genetic maps (for similar models, see Gardiner et al. [1990a] and Ikemura and Wada [1991]). Each alternation of base composition corresponds to a base-composition isochore and to a 2.4-mb prophase replication band. Isochores L1 (30% of genome [Mouchiroud et al. 1988]) and L2 (35%) contain the late-replicating, tissue-specific genes and correspond to prophase G-bands. Isochores H1 (22%), H2 (10%), and H3 (3%) contain early-replicating genes and correspond to prophase R-bands. The prophase bands condense (fig. 5) into 8-mb metaphase bands, with R-bands condensing the most (Drouin et al. 1991a). Genes cluster in a special class of metaphase R-band called “T-bands.” Even late-replicating AT-rich genes may be concentrated in the thin prophase G-bands of these T-bands. Each metaphase band is modeled as a unit of concerted evolution, to contain only two isochores. Data relating isochores other than H3 to chromatin flavors are scanty (Gardiner et al. 1990a), and mundane R₁₂₅₀ bands may actually contain the L2 isochore.

correspond to prophase bands (Bernardi et al. 1985), and the base-composition transition between isochores is probably sharp (Ikemura et al. 1990). The most GC-rich isochore, H3, possesses the highest gene density (Bernardi et al. 1985; Perrin and Bernardi 1987) and is modeled as a T-band isochore fraction (fig. 4), as has been suggested elsewhere (Gardiner et al. 1990b; Holmquist 1990; Ikemura and Wada 1991).

A chromosome band is also a unit of chromosome replication, a cluster of temporally synchronous replicons which initiate and terminate replication during a small fraction of the S-phase (Holmquist 1988). Replication banding involves selective incorporation of the thymidine analogue BrdUrd into early- or late-replicon clusters before allowing the chromatin to condense into chromosomes. S-phase transit can be arrested in the middle of S-phase, with methotrexate or with high concentrations of BrdUrd or thymidine (Drouin et al. 1988, 1990). Arrest occurs after all R-bands have replicated but before G-bands have replicated. With such arrests producing BrdUrd-substituted R-bands or BrdUrd-substituted G-bands, high-resolution replication banding_{1,250} was obtained. These replication bands_{1,250} are almost exactly identical to or complementary to GTG_{-1,250} bands (Lemieux et al. 1990) (fig. 5), and the early-replicating bands constitute about 60% of the genome (Drouin et al. 1991b).

Early- and late-replicating DNA can be separated using colcemid-synchronized populations of cells, adding BrdUrd to the culture medium in the middle of S-phase, subsequently collecting metaphase cells, isolating DNA, and finally separating unsubstituted early-replicating DNA from substituted late-replicating DNA (Holmquist 1988). When results from replication-time fractionation and base-composition fractionation are compared, they are virtually identical. Genes or repeat families in the early 50%–60% of S-phase are in the GC-rich H1, H2, or H3 isochore fractions (35% of the genome), while genes or repeat families in the late 40%–50% fraction are in the L1 and L2 isochore fractions (65% of the genome) (Holmquist 1989b). Independent of cytogenetic mapping, these data allowed a molecular definition of R-band-like and G-band-like genes that is based upon their replication times and/or the base composition of their flanking DNA (Holmquist 1989b).

By means of these distinctive molecular attributes, 86 genes were listed whose replication time, or isochore, or CpG-island status is known (Holmquist 1989b). CpG islands are several hundred base-pair-

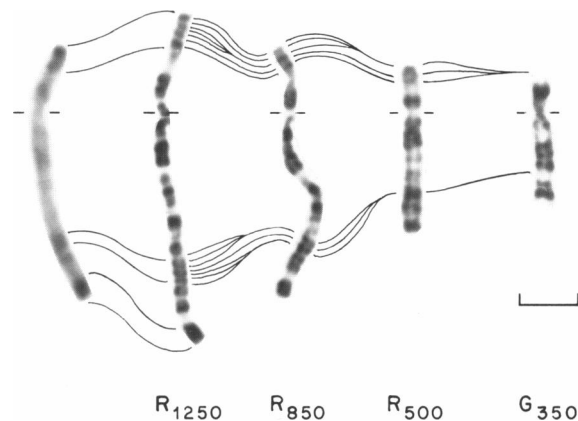


Figure 5 Replication banding of human chromosome 5, by using mid-S-phase arrest followed by release into BrdUrd, which produced R-banding at various stages of chromosome condensation. The opposite BrdUrd substitution pattern produced G-banding in a very condensed metaphase chromosome. The number of different resolved bands per haploid genome, as indicated by subscripts, decreases as the chromosomes condense (Drouin et al. 1991a). The decrease is due to fusions of many prophase subbands. The leftmost chromosome, with a length corresponding to a resolution of 850 bands/genome, was treated by THG-banding. It shows that, at prophase, the heat-resistant T₈₅₀-subbands are clustered. In shorter chromosomes, these clusters are fused and form a single metaphase T₃₀₀-band (not shown). The medial T₈₅₀-band cluster is more apparent as a heat-resistant T-band by acridine orange staining than by the THG technique (Dutrillaux 1973) shown here. (Scale bar = 4 μ m.) (Courtesy of Regen Drouin and C.-L. Richer.)

long sequences with unusually high and unmethylated concentrations of the dinucleotide CpG. With only four exceptions, the 86 genes fall into two distinct groups. The R-band-like group contains all house-keeping genes and about half of the tissue-specific genes: its genes are early replicating, are in GC-rich isochores, and have CpG islands. The G-band-like group contains only tissue-specific genes: these are late replicating, are in AT-rich isochores, and lack CpG islands. CpG islands represent band-indexing motifs. Also, in the R-band-like group, 21 of 25 sequenced genes have an upstream GGCGGG (consensus sequence for Sp1 binding), while, in the G-band-like group, only 2 of 11 sequenced genes have this sequence. Thus, CpG islands and Sp1-binding sequences both represent indexing sequence motifs. As predicted by Gasser and Laemmli (1987), these motifs indicate the kind of band (chromosome compartment) in which a gene resides. CpG island motifs, being sites for *NotI*, etc., are used for mapping by pulsed-field gel electrophoresis. Their scarcity in genes such as the

Duchenne muscular dystrophy gene indicates that such genes are G-band-like genes (Burmeister et al. 1988).

In the L1 and L2 (G-band-like) isochore fractions, all 16 sequenced genes have a G + C content of <70% for their third-coding-position bases (Bernardi et al. 1985). In the H1, H2, and H3 (R-band-like) isochore fractions, 11 of 14 genes have a third codon GC content of >70% (Bernardi et al. 1985). Recently, Ike-mura and Wada (1991) further confirmed these conclusions and, using genes which have been both sequenced and mapped to individual bands, showed that genes which map to the T-bands of figure 1 have a higher average G + C percentage at the third-codon position (71% G + C) than do genes which map to mundane R-bands (62% G + C) or to G-bands (55% G + C). The concentration of the retroposon family Alu increased by about a factor of six or greater along the isochore series L1, L2, H1, H2, H3, while that of the L1 family decreases by an approximately similar factor (Soriano et al. 1983; Zerial et al. 1986). Thus, retroposon concentration and codon usage are also like indexing sequences: they predict a residence band class.

When mitotic prophase initiates, its prophase chromosomes are long and reveal many fine subbands. The term "metaphase chromatin flavor" does not apply to these individual subbands but only to their metaphase fusion products. An early report discerned 2,000 GTG-bands (Yunis 1981). Such resolution was never repeated, and so 1,250 prophase subbands/genome (fig. 5), as modeled in figure 4, seems a reasonable current estimate for the maximum number of bands that genome sequencing will eventually unveil (Drouin et al. 1990). As the prophase chromosomes condense, subbands fuse (fig. 5) until only ~300–350 bands/genome can be discerned by late metaphase. Exactly how subbands seemingly disappear, even when visualized by replication banding, remains a mystery (Drouin et al. 1991a). A condensation metric, the number of subbands_{1,250} which fuse to form a band₃₀₀, was provided, by Regen Drouin, from high-resolution ideograms (Drouin and Richer 1989). This metric varies with metaphase chromatin flavor (table 1). The generic T₃₀₀ bands are the fusion product of about seven prophase subbands, a significantly large number of bands_{1,250} than observed for any other chromatin flavor, as shown in figures 2 and 5 and as modeled in figure 4.

High-resolution T-banding shows that THG⁺ subbands are not randomly distributed along the prophase

chromosome but are clustered (fig. 5). Each THG⁺_{1,250} cluster fuses during mitosis, to form a single T₃₀₀ band (fig. 5). This is the first direct evidence that R-subbands which belong to a single metaphase R-band are similar and generally distinct from other R-subbands as modeled in figure 4 for base composition. Present in situ resolution is insufficient to similarly show that vAlu⁺ bands_{1,250} are clustered, but the vAlu⁺₃₀₀ bands seem to be caused by more than one subband out of seven contributing to the vAlu⁺₃₀₀ signal. While prophase bands and metaphase bands share the name "band," they seem to represent distinctly different hierarchical levels of chromosome organization, with metaphase bands being clusters of "similar" prophase bands.

The GC content of mammalian R-bands in general and of their T-bands in particular has directionally increased since the divergence from reptiles and/or the appearance of warm-blooded animals (Bernardi et al. 1985, 1986; Perrin and Bernardi 1987). In primates, an unusually strong directional increase has occurred in clusters of R-subbands. Within a cluster, these subbands have undergone this strong GC-content increase in unison, in a fashion which coupled the base-composition increases of each R_{1,250}-band within a fusion cluster. This unit of seemingly concerted (Dover 1982; Ohta and Dover 1983) evolution approximately coincides with the metaphase band₃₀₀.

In conclusion, the black-and-white bands₃₀₀ of ideograms, rather than reflecting low-resolution preparation artifacts, represent a meaningful hierarchical level of chromosome organization; they are units of concerted chromosome evolution which display a property called "metaphase chromatin flavor." For defining this hierarchical level more exactly, a different resolution—e.g., bands₄₅₀, as indicated by isochore differences between 21q22.1 and 21q22.3 (Gardiner et al. 1990a)—may prove to be a more appropriate choice than bands₃₀₀. The band resolution which maximized value clustering in figure 2 would be the most appropriate choice.

The metaphase band₃₀₀ may also be a unit of gene concentration, with G-band-like genes possibly concentrated in the same metaphase bands₃₀₀ as are R-band-like genes. For example, the G-band-like (a molecular definition) gene beta-globin maps, by cytogenetic methods, to an R-band, 11p15.5. This seems an embarrassing contradiction (Bickmore and Sumner 1989). However, G-band genes in R-bands are, in my opinion, simply a problem of cytogenetic resolution and confusing nomenclature. I had divided 86 genes into

Table 3**G-Band-like Genes**

GENE ^a	BINARIZED PROPERTY ^b					BAND ^c	FLAVOR
	1	2	3	4	5		
α -Amylase 1	T	?	?	N	?	1p21	G-Band
α -1-Antichymotrypsin	T	L	?	?	?	14q32	vAlu ⁻ vGC ⁺
α -1-Antitrypsin	T	L	?	N	-	14q32	vAlu ⁻ vGC ⁺
β -Globin	T	L	A	N	-	11p15.5	vAlu ⁻ vGC ⁺
Fibrinogen α , γ	T	?	?	N	S	4q28	G-band
Fibrinogen β	T	?	?	N	-	4q28	G-band
γ , δ , ϵ -globins	T	L	A	N	-	11p15.5	vAlu ⁻ vGC ⁺
Immunoglobulin variant ...	T	L	A	?	?	14q32	vAlu ⁻ vGC ⁺
Insulin	T	?	?	N	?	11p15.5	vAlu ⁻ vGC ⁺
Interferon α	T	?	?	N	-	9p22	R-mundane
N-myc	T	L	?	N	?	2p24	G-band

^a All genes which were classified as G-band like on the basis of molecular attributes (Holmquist 1989b) and which were mapped to an individual band₃₀₀ (Kidd et al. 1989), along with their band of residence and its metaphase chromatin flavor.

^b For molecular classification. 1 = Tissue specific vs. Housekeeping; 2 = Late replicating vs. Early replicating; 3 = AT-rich isochore vs. GC-rich isochore; 4 = No CpG island vs. + CpG island; and 5 = absence (-) vs. presence S of upstream Sp1 site.

^c Most genes which are G-band like on the basis of molecular criteria map to R-bands, especially to the very GC-rich R-bands.

two groups based on their molecular properties. Eighteen of the R-band-like genes have been mapped cytogenetically to bands₃₀₀. In accord with results of similar studies by Ikemura and Wada (1991), 17 of 18 mapped to R₃₀₀-bands, usually T-bands of the vAlu⁺ vGC⁺ flavor (data not shown). G-band-like genes, mapped with similar resolution, also frequently mapped to T-bands (table 3). Ikemura and Wada's (1991) data neither strengthened nor weakened this supposition of many G-band-like genes in T-bands; they listed genes with a G + C percentage of <40% at the third-codon position and reported that 13 mapped to G-bands, that 5 mapped to mundane R-bands, and that 4 mapped to T-bands. One explanation which can be drawn from the limited data in table 3 is that many G-band-like genes are simultaneously located both in late-replicating G_{1,250} subbands and in T₃₀₀-bands, the G_{1,250}-subbands being the ones engulfed by the condensation of T₃₀₀-bands, as shown in figure 5 and as modeled in figure 4. The resolution subscripts introduced in this article provide a semantic to address problems such as, When does a G-band gene map to an R-band? Several kinds of GTG⁺_{1,250} DNA will be distinguishable to the genome-sequencing initiative. For example, GTG⁺_{1,250} THG⁺₃₀₀ DNA (DNA of prophase G-bands which are included in a metaphase

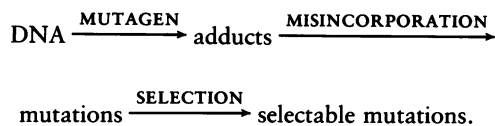
T-band) should be different than GTG⁺_{1,250} GTG⁺₃₀₀ DNA (DNA of prophase G-bands which are included in metaphase G-bands); the GTG⁺_{1,250} THG⁺₃₀₀ DNA could probably have a higher gene density. Here, our expectations of sequences in a G_{1,250}-band are influenced by the metaphase chromatin flavor of the larger metaphase band—i.e., prophase cluster—in which it resides.

Compartmentalized Sequence Evolution

Primary chromosomal mutations, gene frequency, oncogene frequency, and other cytological events show hot spots and cold spots. These have been alternatively ascribed to either the earliest-replicating, or quinacrine-dullest, or Alu-richest bands (Kuhn et al. 1985; Kuhn 1976; Korenberg and Rykowski 1988; Mamuris et al. 1989; Gardiner et al. 1990b) and are here established as congruent with a chromatin flavor pattern. The input of chromosomal mutations is ordered (table 1). Natural selection (DeMarini et al. 1989), another mechanism for ordering the genome, and neutral drift then act on this ordered input (fig. 6), to fix chromosomal mutations in the population. The fixed changes are recognized as karyotype evolution. If the input of molecular mutations is similarly

ordered, this would explain both why selectively neutral DNA shows base-composition patterns congruent with the banding pattern and why a gene's sequence, quite independent of the gene's function, is related to the kind of band in which the gene resides.

Mutageneticists study the following process:



Both the frequency of adduct production (Kootstra et al. 1989; Olivero et al. 1990; Pfeifer et al. 1991) and adduct-repair rates in mammals depend on the type of adduct, its precise location in the genome, and the functional state of the chromatin being repaired, as reviewed by Hanawalt (1990). For example, UV-induced cyclopyrimidine dimers in rodent cells are repaired much faster in active genes than in inactive genes and are repaired as much as 20 times faster when in the transcribed strand than when in the nontranscribed strand (Hanawalt 1990). Rapid adduct repair by precise repair mechanisms decreases the chance of mutation, the chance that a replication fork will come on an adduct, not recognize the adduct as the appropriate base, misincorporate an inappropriate base, and thus generate a heritable base change. Accordingly, UV photoproducts in the rapidly repaired, transcribed strand produce 10-fold fewer selectable HPRT mutants in cultured hamster cells than do photoproducts in the nontranscribed strand (van Zeeland et al. 1990). In cells deficient for repair of UV damage or in Cockayne syndrome cells which are deficient for selective repair of damage in transcribed strands (Venema et al. 1990), UV photoproducts in the transcribed strand produce opposite results: more selectable HPRT mutants. Thus, the rate of mutagen-induced molecular mutations is determined, at both the global level and the local chromatin level, by repair mechanisms. From this experimental proof, it can be inferred that local biases of mutational input unquestionably contribute to genome evolution. The questions which remain concern not the existence of mutational input patterns but the relative magnitude of their contribution to sequence evolution, their mechanisms of regulation, and their relation to band patterns.

The rate of the molecular clock (Zuckerandl 1986) varies with the banding pattern. Early-replicating genes and interspersed repeats cross-hybridize—e.g., rodent × human—much more frequently than do

their late-replicating counterparts (Goldman et al. 1984; Holmquist and Caston 1986). In rodents, the silent substitution rate is higher for G-band-like genes than for R-band-like genes (Filipski 1988; Wolfe et al. 1988). When a fluorescent mixed probe from a human chromosome library is used in the presence of excess unlabeled repetitive DNA to “paint” primate chromosomes by in situ hybridization, it paints both G- and R-bands of hominoids but paints only R-bands of lower primates (Wienberg et al. 1990). Thus, spontaneous molecular mutations in both genes and junk DNA are being fixed at different rates, faster in G-bands and slower in R-bands.

The band-congruent sequence order along the human genome is due to the quality (instead of the quantity) of fixed mutations, and it represents fossil evidence for a history of ordered mutational input and/or output along the genome. Four models to explain the evolution of this sequence organization will be presented; they are (1) natural selection *sensu strictu*, (2) direct cause-effect influences on mutational input, (3) selfish-DNA feedback loops for mutational input, and (4) concerted actions on mutation input into or output from chromatin flavors.

When mutational input was not recognized as ordered, only *random mutational input-natural selection models* could explain that genomes have become ordered (fig. 6). Bernardi et al. (1985; Bernardi 1986; Bernardi et al. 1988; Mouchiroud et al. 1988) argued that selection is responsible for the sequence patterns along the banded chromosomes. Natural selection *sensu strictu* immediately acts on an incremental phenotypic change which has altered the reproductive fitness of a mutant progeny; the phenotypic change is directly caused by a DNA sequence change, and so there is a direct cause-effect pathway between DNA sequence and fitness. For example, if AT→GC mutations incrementally increase fitness when they occur in or near mammalian genes, because they increase the thermal stability of genes and gene products (Bernardi 1986), then gene richness and percent GC will be causally related; the two variables—mapped genes/band and average GC content of a band—would be positively correlated. However, the genetic load associated with direct selection is too high to explain the rate at which GC and Alu have accumulated in preferred regions of the human genome, so alternate explanations have been explored (Holmquist 1989b).

Molecular ecology is the study of mutational input organized at the local level especially when the responsible chromatin organization is coincident with chro-

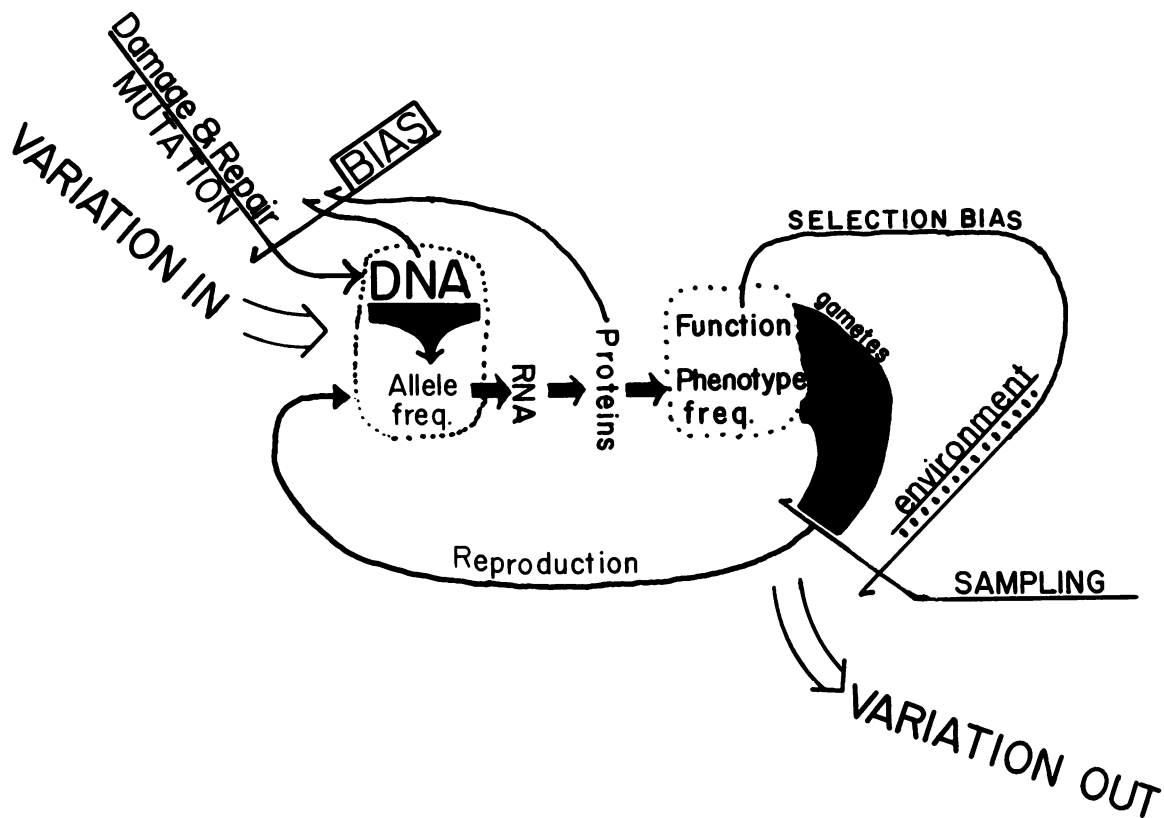


Figure 6 Flow of sequence variation through the generation cycle of an interbreeding population, demonstrating the requisite connectivity for generating self-organizing patterns. Variation out occurs mainly by random sampling and allows selectively neutral mutants to become fixed in the population by chance, a process called “neutral drift.” Any bias on how gametes are produced or sampled for inclusion in the next generation is called “natural selection.” Selection usually increases the rate of variation out but, more important, it can organize the variation which remains. Variation enters the population by mutational input, which is also biased. Global bias (*arrow from proteins to bias*) occurs when, e.g., a DNA polymerase misreads a particular base with equal infidelity throughout the genome, resulting in a directional mutation pressure toward an extreme base composition (Sueoka 1988). Local bias (*arrow from DNA to bias*) occurs when, e.g., a DNA sequence binds a protein which protects (*footprints*) local guanines from alkylating mutagens. Note that local bias can generate a self-sustaining feedback loop of selfish DNA. Historical accumulation of sequence order in a population’s genome can be due only to biased sampling and biased mutational input. The concepts of substitutional cost and mutational load are realizations that natural selection imposes arithmetic limits on the rate at which, without causing the dynamic system to crash, advantageous alleles can be pumped in or at which low-fitness alleles can be pumped out. I contend that a band represents a chromatin domain which directs its mutational input differently than do neighboring bands; this circumvents genetic loads and substitutional costs. Q-banding would represent a glowing example of a long-term input bias for G + C in R-bands; any gene which had existed in a Q⁺ band would, quite divorced from this gene’s function, bear the fossilized imprint of this history, in its codon usage and flanking sequences.

matin band patterns (Holmquist 1989a, 1989b). The molecular ecologist, like the plant ecologist, walks the landscape and notices a small number of distinct associations of sequence motifs, sequence communities. With limited data on species frequency, he chooses readily distinguishable variables: percent GC, Alu frequency, and replication time. He binarizes each variable, names sequence communities by their combination of binarized variables (dominant community

species), and asks how such communities came to be. As plant communities may be the direct effect of overlapping tolerance ranges (Grime 1979), Filipiski (1989) reviewed how *direct cause-effect models* explain several examples of sequence organization. As applied to chromosome compartments, such models predict that R-band DNA, originating from active (gene-rich) chromatin, would be more accessible to repair enzymes, would be repaired more quickly and

accurately, and, suffering fewer mutations, would have a slower molecular clock (Filipski 1988). More specifically, active domains could be more accessible to the base-mismatch repair system, described by Brown and Jiricny (1987, 1988), which preferentially changes mismatched A's and T's into G's and C's. Such mismatch repair, acting during recombination and recognized as gene conversion (Radman 1988), would drive recombinogenic domains to GC richness. Also as Filipski (1989) has suggested, retroposon insertion or base misincorporation might be related to either replication time (Wolfe et al. 1988; Bickmore and Sumner 1989) or speed of replication-fork progression (Schaaper and Radman 1989); low early-S-phase dATP concentrations (Wolfe et al. 1988) could cause selective misincorporation of G and C into early-replicating DNA.

An alternative to direct-cause-and-effect models is the *selfish-DNA feedback loop model* of mutational input (Holmquist 1989a, 1989b). It is similar to that which Odum (1969) used to explain the association of plant species into distinct communities. DNA sequences determine which proteins bind, which organizes a chromatin type, which determines differential "accessibility" to repair enzymes, which determines mutation quality, and which determines future DNA sequences. Local sequences could selfishly determine their own future neighbors, with sequence community members coevolving like bees and flowering plants but being independent of natural selection and genetic load (fig. 6). This is a self-organization model. It predicts the base-compositional homogeneity observed along long genomic stretches (fig. 3) (Ikemura and Aota 1988) but does not address the table 1 data concerning relations between different bands of one flavor.

Concerted phenomena are implicated in sequence evolution. When a dispersed family of DNA sequences maintains intrafamily sequence similarity while the family's consensus sequence diverges, this is recognized as concerted evolution, another example of self-organization. The molecular-drive (Dover 1986) mechanisms responsible for coupling a member's sequence to that of every other member in the family are gene conversion, intragroup exchange, duplication-deletion, selective amplification, etc. In model systems of multigene families, intragroup sequence variation decreases whenever the frequency of any one molecular-drive mechanism increases (Ohta and Dover 1983; Ohta 1990). Also, the genetic load incurred by selection, acting on individual family members to cause divergence of the family's consensus sequence, is less

when molecular drive acts than when it does not act (Ohta 1990). I will now generalize the scope of concerted sequence evolution, to include not only sequence homology but also all measurable chromosome properties; the generalized drive mechanisms will remain undefined. With this generalized definition, I recognize several sequence patterns which could have evolved by concerted mechanisms. These include the five base-composition-isochore fractions, approximately five band₃₀₀ chromatin flavors, and the individual clusters of several prophase RHG⁺_{1,250} subbands belonging to a single RHG⁺₃₀₀ band. The unique prediction of concerted evolution involves the variance rather than the mean of measured values; concerted mechanisms regulating mutational input or output would produce groups of chromatin types wherein intragroup variation is small compared with global genomic variation. Suppression of intragroup variation for each involved variable would concomitantly reduce this variable's correlation with any other variable, within one chromatin type.

Cause-Effect Relations during Sequence Evolution

To test predictions of the four models, six pairwise combinations of values for four variables were plotted in two dimensions. Two examples are shown in figure 2. I asked, in the fitted equation to the global data $Y = f(X, \text{chromatin flavor})$, Is the contribution of flavor minimal, $Y = f(X)$ (i.e., single direct cause), or consequential, $Y = f(X, \text{flavor})$ (multiple direct causes or other mechanisms)? I also asked whether the X, Y correlation occurred within flavors (cause→direct effect) or only between different flavors ([1] causes acting on the average value of a flavor rather than directly on values of the tested variables or [2] causes acting on values of tested variables, with concerted evolution minimizing variation within a flavor).

By simply ignoring the Y-axis in figure 2a, one sees that each flavor class—the red, green, yellow, and blue data points—has a different replication-time distribution. Thus, apparent replication time is $f(vGC, vAlu)$ and is a function of at least two variables, with the earliest replicating bands being both very GC rich and very Alu rich. Consequently, early replication alone is not the previously postulated single direct cause of either GC richness (Wolfe et al. 1988) or Alu richness (Bickmore and Sumner 1989). Also, Alu richness and GC richness are not multiple effects of a single direct cause.

Values of all six variables are significantly correlated; plots of all six variable pairs could be globally interpreted as $Y = f(X)$, where $f(X)$ is a single-valued continuous function of X , like the dashed lines in figure 2. These global correlations support a global cause-effect relationship between variables. For example, natural selection acting to cause increased recombination rates in gene-rich regions is consistent with the dotted line $Y = f(X)$ in figure 2*b*. This interpretation is misleading because, when each data point is additionally color coded to indicate the flavor of the contributing band, then Y is seen to be independent of X , within any one of four chromatin flavors (figs. 2*a* and *b*). The paired variables at first seemed correlated; however, only the mean values of the flavors are correlated.

Clustering of data points for paired variables from bands of similar flavor is apparent (fig. 2*a*), with many flavors often occupying their own almost individually unique two-dimensional range of values in variable space. Were the data plotted in four dimensions by using all four variables, then the individual clusters would become even more distinct than is apparent in these two-dimensional plots. Our ability to visualize a small number of distinct clusters—i.e., value ranges—has several implications: (1) The flavor distinctions are functionally consequential. (2) Bands of similar flavor could have been defined as a distinct group, on the basis of their distinct value ranges in variable space, without the cytological distinction of chromatin flavor. This implies that the essence of chromatin flavor can exist in rodent genomes even though these genomes lack the cytological trappings—Alu, H3 isochores (Bernardi et al. 1988; Mouchiroud et al. 1988), and probably T-banding—initially used to distinguish chromatin flavors. (3) If many more than the five already distinguished consequential flavors existed, then most presently distinguished flavors should show correlations between paired variables. They do not, and so a majority of the flavors—i.e., value clusters in figure 2—have probably been identified (4) Cause→direct-effect mechanisms, especially those involving single causes with multiple effects, predict more-regular distributions of the data points, instead of the discontinuous distributions we recognize as value clusters. Clustering is consistent with causes acting on flavors, instead of directly on local values of variables. Clustering is also consistent, with a very strong intraflavor coupling preventing direct effects on a variable's value from accumulating locally to affect only the resident band. (5) In an explanation of molecular

drive (Ohta and Dover 1983), the measured effect—sequence homology between members of a multigene family—was quite distinct from mechanisms such as recombination or gene conversion, which purportedly caused the effect of sequence homology. In figure 2 and table 1, which variables are causes and which variables are effects is unclear. If recombination rate and replication time are indeed causes, then the frequency distributions of these causal events are themselves concertedly distributed, imparting a complex beauty to genome evolution.

The independence of variable Y on variable X , within any one flavor (except for $vAlu^- vGC^+$) (fig. 2), has two interpretations. The first is that concerted mechanisms so strongly couple and suppress variation within a family of similarly flavored bands that correlations between paired variables within a flavor will always be weak. Such concerted coupling would also reduce genetic load (Ohta 1990). Within the $vAlu^- vGC^+$ flavor, the correlation between apparent replication time and subband number cannot be explained by weak coupling, because the correlation is in a direction opposite to that of the global correlation. The second interpretation is that causes act on chromatin flavors and only indirectly effect the measured variables. An example is the possibility that natural selection, not acting directly on gene density but acting on each flavor's average gene density, caused increased recombination in gene-rich flavors. Here, selection could have acted *indirectly* on values of the local variable, i.e., gene density. This possibility allows natural selection to drive sequence-pattern evolution with a lower associated mutational load (Zuckermandl 1986) than that associated with selection acting directly on the incremental fitness change of each Alu insertion or AT→GC change.

Extending the term “concerted evolution” from the confines of sequence homology to include sequence motifs, gene density, recombination rate, and replication time was purposeful and should be noted. Just as a particular recombination frequency is one value of the variable called “recombination rate,” the extended concept makes a particular DNA sequence into one value of the variable called “DNA sequences.” A chromatin flavor can now be mathematically defined, via the dynamic-flow diagram in figure 6, as a value range for each of many genomic variables. It is this range of values (dashed ellipses in fig. 2) which is heritable and concertedly propagated. When an organism is treated with SDS and proteinase K, the only measurable variable remaining is a DNA sequence.

Conclusions

The human genomic sequence is a linear array of ~ 2 -mb-sized-motif communities—prophase bands—wherein a gene's sequence reflects its residence. These communities are arranged into 10-mb clusters called “metaphase bands,” and most cytogenetic data are from these 10-mb clusters, rather than from their constituent 2-mb-motif communities. Probably only five community types and only five (or a few more) consequential clusters exist. In humans, the five different motif communities are distinguishable as base-composition isochores, and the different clusters are cytologically distinguishable as metaphase chromatin flavors; this will facilitate mapping relations between the sequenced genome and ideograms. Sequence evolution appears self-organizational, with concerted phenomena being instrumental in orchestrating the formation of sequence patterns. Sequence patterns which likely evolved by concerted phenomena include the five-base-composition-isochore fractions, approximately five band₃₀₀ chromatin flavors, the individual clusters of several prophase $\text{RGH}^+_{1,250}$ subbands belonging to a single RHG^+_{300} band, and possibly the homogeneous base composition within each prophase band itself. The forces governing mutational input are much more complicated than “single cause \rightarrow multiple effects” and possibly operate through a particular three-dimensional organization of the various chromatin types in mitotic (Manuelidis 1991; Nakayasu and Bezzney 1991) or meiotic (Schweizer and Loidl 1987; Ashley 1990) nuclei.

Acknowledgments

I thank Regen Drouin and Claud-Lise Richer for unpublished data, review, and total cooperation; Art Riggs, Jan Filipski, Laura Manuelidis, and Toshimichi Ikemura for review; and Amit Chopra for data collection. This work was funded by the City of Hope and the Upjohn Co.

References

- Ambros PF, Sumner AT (1987) Correlation of pachytene chromomeres and metaphase bands of human chromosomes, and distinctive properties of telomeric regions. *Cytogenet Cell Genet* 44:223–228
- Ashley T (1990) Prediction of mammalian meiotic synaptic and recombination behavior of inversion heterozygotes based on mitotic breakpoint data and the possible evolutionary consequences. *Genetics* 83:1–7
- Barrios L, Miro R, Caballin MR, Fuster C, Guedea F, Subias A, Egozcue J (1989) Cytogenetic effects of radiotherapy breakpoint distribution in induced chromosome aberrations. *Cancer Genet Cytogenet* 41:61–70
- Bernardi G (1986) Compositional constraints and genome evolution. *J Mol Evol* 24:1–11
- Bernardi G, Mouchiroud D, Gautier C (1988) Compositional patterns in vertebrate genomes: conservation and change in evolution. *J Mol Evol* 28:7–18
- Bernardi G, Olofsson B, Filipski J, Zerial J, Salinas J, Cuny G, Meunier-Rotival M, et al (1985) The mosaic genome of warm-blooded vertebrates. *Science* 228:953–958
- Bickmore WA, Sumner AT (1989) Mammalian chromosome banding—an expression of genome organization. *Trends Genet* 5:144–148
- Bobrow M, Madan K (1973) The effects of various banding procedures on human chromosomes studied with acridine orange. *Cytogenet Cell Genet* 12:145–156
- Brown TC, Jiricny J (1987) A specific mismatch repair event protects mammalian cells from loss of 5-methylcytosine. *Cell* 50:945–950
- (1988) Different base/base mispairs are corrected with different efficiencies and specificities in monkey kidney cells. *Cell* 54:705–711
- Buckton KE (1976) Identification with G and R banding of the position of breakage points induced in human chromosomes by *in vitro* X-irradiation. *Int J Radiat Biol* 29:475–488
- Burmeister M, Monaco AP, Gillard EF, van Ommen GJ, Affara NA, Ferguson-Smith MA, Kunkel LM, et al (1988) A 10-megabase physical map of human Xp21, including the Duchenne muscular dystrophy gene. *Genomics* 2:189–202
- Cuny G, Soriano P, Macaya G, Bernardi G (1981) The major components of the mouse and human genomes. I. Preparation, basic properties and compositional heterogeneity. *Eur J Biochem* 115:227–223
- DeMarini DM, Brockman HE, de Serres FJ, Evans HH, Stankowski LF Jr, Hsie AW (1989) Specific-locus mutations induced in eukaryotes (especially mammalian cells) by radiation and chemicals: a perspective. *Mutat Res* 220:11–29
- Dover G (1982) Molecular drive: a cohesive mode of species evolution. *Nature* 299:111–117
- (1986) Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. *Trends Genet* 2:159–165
- Drouin R, Lemieux N, Richer C-L (1988) High-resolution R-banding at the 1250-band level. I. Technical considerations on cell synchronization and R-banding (RHG and RBG). *Cytobios* 56:107–125
- (1990) Analysis of DNA replication during S-phase by means of dynamic chromosome bandings at high resolution. *Chromosoma* 99:273–280
- (1991a) Chromosome condensation from prophase

- to late metaphase: relationship to chromosome bands and their replication time. *Cytogenet Cell Genet* 57:91–99
- (1991b) High-resolution R-banding at the 1250-band level. III. Comparative analysis of morphologic and dynamic R-band patterns (RHG and RBG). *Hereditas* 114:65–77
- Drouin R, Richer C-L (1985) Analysis of high-resolution R-bands, obtained by heat-denaturation and Giemsa staining, on human prophase chromosomes. *Can J Genet Cytol* 27:83–91
- (1989) High-resolution R-banding at the 1250-band level. II. Schematic representation and nomenclature of human RBG-banded chromosomes. *Genome* 32:425–439
- Dubos C, Pequignot EV, Dutrillaux B (1978) Localization of gamma-ray induced chromatid breaks using a three consecutive staining technique. *Mutat Res* 49:127–131
- Dutrillaux B (1973) Nouveau système de marquage chromosomique: les bandes T. *Chromosoma* 41:395–402
- (1977) New chromosome techniques. In: Yunis JJ (ed) *Molecular structure of human chromosomes*. Academic Press, New York, pp 233–265
- (1979) Chromosomal evolution in primates: Tentative phylogeny from *Microcebus murinus* (Prosimian) to man. *Hum Genet* 48:251–314
- Dutrillaux B, Couturier J, Richer C-L, Viegas-Pequignot E (1976) Sequence of DNA replication in 277 R- and Q-bands of human chromosomes using BrdU treatment. *Chromosoma* 58:51–61
- Dutrillaux B, Couturier J, Viegas-Pequignot E, Schaison G (1977) Localization of chromatid breaks in Fanconi's anemia, using three consecutive stains. *Hum Genet* 37:65–71
- Dutrillaux B, Covic M (1974) Étude de facteurs influencant la dénaturation thermique ménagée des chromosomes. *Exp Cell Res* 85:143–153
- Dutrillaux B, Lejeune J (1971) Sur une nouvelle technique d'analyse du caryotype humain. *CR Acad Sci Paris [D]* 272:2638
- 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 (1988) Why the rate of silent codon substitutions is variable within a vertebrate's genome. *J Theor Biol* 134:159–164
- (1989) Evolution of DNA sequence contributions of mutational bias and selection to the origin of chromosomal compartments. In: Obe G (ed) *Advances in mutagenesis research*. Springer, Berlin, pp 1–54
- Gardiner K, Aissani B, Bernardi G (1990a) A compositional map of human chromosome 21. *EMBO J* 9:1853–1858
- Gardiner K, Horisberger M, Kraus J, Tantravahi U, Korenberg J, Rao V, Reddy S, et al (1990b) Analysis of human chromosome 21: correlation of physical and cytogenetic maps: gene and CpG Island distributions. *EMBO J* 9:25–34
- Gasser SM, Laemmli UK (1987) A glimpse at chromosomal order. *Trends Genet* 3:16–22
- Goldman MA, Holmquist GP, Gray MC, Caston LA, Nag A (1984) Replication timing of mammalian genes and middle repetitive sequences. *Science* 224:686–692
- Grime JP (1979) *Plant strategies and vegetation processes*. Wiley, New York
- Hanawalt PC (1990) Role of gene expression in the fine structure of DNA damage processing. *Biotechnol Hum Genet Predisposition Dis* 3:135–145
- Heim S, Mitelman F (1987) Nineteen of 26 cellular oncogenes precisely localized in the human genome map to one of the 83 bands involved in primary cancer-specific rearrangements. *Hum Genet* 75:70–72
- Holmberg M, Jonasson J (1973) Preferential location of X-ray induced chromosome breakage in the R-bands of human chromosomes. *Hereditas* 74:57–67
- Holmquist GP (1988) DNA sequences in G-bands and R-bands. In: Adolph KW (ed) *Chromosomes and chromatin*. CRC Press, Boca Raton, pp 76–121
- (1989a) DNA sequences in G- and R-bands: phylogenetic aspects. In: Fredga K (ed) *Chromosomes today*. George Allen & Unwin, London, pp 21–32
- (1989b) Evolution of chromosome bands: molecular ecology of noncoding DNA. *J Mol Evol* 28:469–486
- (1990) Mutational bias, molecular ecology, and chromosome evolution. In: Obe G (ed) *Advances in mutagenesis research*. Springer, Berlin, pp 95–126
- Holmquist GP, Caston LA (1986) Replication time of interspersed repetitive sequences. *Biochim Biophys Acta* 868:164–177
- 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
- Ikemura T, Wada K (1991) Evident diversity of codon usage patterns of human genes with respect to chromosome banding patterns and chromosome numbers: relation between nucleotide sequence data and cytogenetic data. *Nucleic Acids Res* 19:4333–4339
- Ikemura T, Wada KN, Aota SI (1990) Giant G + C% mosaic structures of the human genome found by arrangement of genbank human DNA sequences according to genetic positions. *Genomics* 8:207–216
- Kidd KK, Bowcock AM, Schmidtke J, Track RK, Ricciuti F, Hutchings G, Bale A, et al (1989) Report of the DNA committee and catalogs of cloned and mapped genes and DNA polymorphisms. *Cytogenet Cell Genet* 51:622–947
- Kootstra A, Lew LK, Nairn RS, MacLeod MC (1989) Preferential modification of GC boxes by benzo[a]pyrene-7,8-diol-9,10-epoxide. *Mol Carcinogenesis* 1:239–244
- Korenberg JR, Engels WR (1978) Base ratio, DNA content, and quinacrine-brightness of human chromosomes. *Proc*

- Natl Acad Sci USA 75:3382–3386
- Korenberg JR, Rykowski MC (1988) Human genome organization: Alu, Lines, and the molecular structure of metaphase chromosome bands. *Cell* 53:391–400
- Kuhn EM (1976) Localization by Q-banding of mitotic chiasmata in cases of Bloom's syndrome. *Chromosoma* 57: 1–11
- Kuhn EM, Therman E, Denniston C (1985) Mitotic chiasmata, gene density, and oncogenes. *Hum Genet* 70:1–5
- Laurie DA, Hulten MA (1985) Further studies on chiasma distribution and interference in the human male. *Ann Hum Genet* 49:203–214
- Lemieux N, Drouin R, Richer C-L (1990) High-resolution dynamic and morphologic G-bandings (GBG and GTG): a comparative study. *Hum Genet* 85:261–266
- Lima-De-Faria A (1975) The relation between chromomeres, replicons, operons, transcription units, genes, viruses and palindromes. *Hereditas* 81:249–284
- (1976) The chromosome field. IV. The distribution of non-disjunction, chiasmata and other properties. *Hereditas* 83:175–190
- Luciani M, Guichaoua MR, Cau P, Devictor B, Salagnon N (1988) Differential elongation of autosomal pachytene bivalents related to their DNA content in human spermatocytes. *Chromosoma* 97:19–25
- Macaya G, Thierry JP, Bernardi G (1976) An approach to the organization of eukaryotic genomes at a molecular level. *J Mol Biol* 108:237–254
- Mamuris Z, Prieur M, Dutrillaux B, Aurias A (1989) The chemotherapeutic drug melphalan induces breakage of chromosomes regions rearranged in secondary leukemia. *Cancer Genet Cytogenet* 37:65–77
- Manuelidis L (1991) A view of interphase chromosomes. *Science* 250:1533–1540
- Manuelidis L, Ward DC (1984) Chromosomal and nuclear distribution of the 1.9-kb human DNA repeat segment. *Chromosoma* 91:28–38
- Miro R, Clemente IC, Fuster C, Egozcue J (1987) Fragile sites, chromosome evolution, and human neoplasia. *Hum Genet* 75:345–349
- Mouchiroud D, Gautier C, Bernardi G (1988) The compositional distribution of coding sequences and DNA molecules in humans and murids. *J Mol Evol* 27:311–320
- Nakayasu H, Berezney R (1991) Mapping replicational sites in the eucaryotic cell nucleus. *J Cell Biol* 108:1–11
- O'Brien SJ, Seuanz HN, Womak JE (1988) Mammalian genome organization: an evolutionary view. *Annu Rev Genet* 22:323–351
- 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*. FK Schattauer, Stuttgart, pp 169–180
- Ohta T (1990) Some new aspects of population genetics arising from gene multiplicity. In: Takahata N, Crow JF (eds) *Population biology of genes and molecules*. Baifukan, Tokyo, pp 169–180
- Ohta T, Dover GA (1983) Population genetics of multigene families that are dispersed into two or more chromosomes. *Proc Natl Acad Sci USA* 80:4079–4083
- 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
- Olivero OA, Huitfeldt H, Poirier MC (1990) Chromosome site-specific immunohistochemical detection of DNA adducts in N-acetoxy-2-acetylaminofluorene-exposed Chinese hamster ovary cells. *Mol Carcinogenesis* 3:37–43
- Paris-conference-1971 (1972) Standardization in human cytogenetics. *Birth Defects* 8(7)
- Perrin P, Bernardi G (1987) Directional fixation of mutations in vertebrate evolution. *J Mol Evol* 26:301–310
- Pfeifer GP, Drouin R, Riggs AD, Holmquist GP (1991) In vivo mapping of a DNA adduct at nucleotide resolution: detection of pyrimidine (6-4) pyrimidone photoproducts by ligation-mediated polymerase chain reaction. *Proc Natl Acad Sci USA* 88:1374–1378
- Porfiro B, Tedeschi B, Vernole P, Caporossi D, Nicolletti B (1989) The distribution of MspI-induced breaks in human lymphocyte chromosomes and its relationship to common fragile sites. *Mutat Res* 213:117–124
- Radman M (1988) Mismatch repair and genetic recombination. In: Kucherlapati R, Smith GR (eds) *Genetic recombination*. American Society for Microbiology, Washington, DC, pp 169–192
- Savage JRK (1977) Assignment of aberration breakpoints in banded chromosomes. *Nature* 270:513–514
- Schaaper RM, Radman M (1989) The extreme mutator effect of *Escherichia coli* mutD5 results from saturation of mismatched repair by excessive DNA replication errors. *EMBO J* 8:3511–3516
- Schweizer D (1981) Counterstain-enhanced chromosome banding. *Hum Genet* 57:1–14
- Schweizer D, Loidl J (1987) A model for heterochromatin dispersion and the evolution of C-band patterns. *Chromosomes Today* 9:61–74
- 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
- Sueoka N (1988) Directional mutation pressure and neutral molecular evolution. *Proc Natl Acad Sci USA* 85:2653–2657
- Sumner AT (1990) *Chromosome banding*. Unwin Hyman, London
- Sutherland GR, Ledbetter DH (1989) Report of the Committee on Cytogenetic Markers. *Cytogenet Cell Genet* 51: 452–458
- Threadgill DS, Kraus JP, Kraweitz SA, Womak JE (1991)

- Evidence for the evolutionary origin of human chromosome 21 from comparative gene mapping in the cow and mouse. *Proc Natl Acad Sci USA* 88:154–158
- Trent JM, Kaneko Y, Mitelman F (1989) Report of the Committee on Structural Chromosome Changes in Neoplasia. *Cytogenet Cell Genet* 51:533–562
- van Zeeland AA, Vrieling H, van Rooijen ML, Venema J, Zdzienicka MZ, Simons JW, Mullenders LHF et al (1990) Influence of DNA excision repair of UV-induced mutation spectra in Chinese hamster cells. In: Mendelsohn ML, Albertini RJ (eds) *Mutation and the environment, part A: Basic mechanisms*. Wiley-Liss, New York, pp 249–254
- Venema J, Mullenders LHF, Natarajan AT, van Zeeland AA, Mayne LV (1990) The genetic defect in Cockayne syndrome is associated with a defect in repair of UV-induced DNA damage in transcriptionally active DNA. *Proc Natl Acad Sci USA* 87:4707–4711
- von Kiel K, Hameister H, Somssich IE, Adolph S (1985) Early replication banding reveals a strongly conserved functional pattern in mammalian chromosomes. *Chromosoma* 93:69–76
- Wienberg J, Jauch A, Stanyon R, Cremer T (1990) Molecular cytotaxonomy of primates by chromosomal *in situ* suppression hybridization. *Genomics* 8:347–350
- Wolfe KH, Sharp PM, Li WH (1988) Mutation rates vary among regions of the mammalian genome. *Nature* 337:283–285
- Yunis JJ (1981) Mid-prophase human chromosomes: the attainment of 2,000 bands. *Hum Genet* 56:293–298
- Yunis JJ, Soreng AL (1984) Constitutive fragile sites and cancer. *Science* 226:1199–1203
- Zerial M, Salinas J, Filipinski J, Bernardi G (1986) Gene distribution and nucleotide sequence organization in the human genome. *Eur J Biochem* 160:479–485
- Zuckerandl E (1986) Polite DNA: functional density and functional compatability in genomes. *J Mol Evol* 24:12–27