

TREBALLS
DE LA
SOCIETAT CATALANA
DE BIOLOGIA

FILIAL DE L'INSTITUT D'ESTUDIS CATALANS

VOLUM 45, 1994

BARCELONA
1994

EL LLENGUATGE DELS GENS: ÚS DE CODONS EN GENS DE VERTEBRATS

ANTONIO MARÍN,¹ JOSÉ LUIS OLIVER² I JUAN RAMÓN MEDINA¹

1. *Departamento de Genética y Biotecnia, Universidad de Sevilla*

2. *Unidad de Genética. Facultad de Ciencias, Universidad de Granada*

Versió catalana d'Esther Esteban

El DNA que forma el genoma nuclear dels vertebrats de sang calenta és un mosaic quant a la composició relativa de les quatre bases (adenina, A; citosina, C; guanina, G i timina, T, que és substituïda per uracil, U, en l'RNA) que són les peces claus i canviants dels graons bàsics de la informació genètica o nucleòtids. Així es compon de fragments molt llargs de DNA (de més de 300.000 bases) que es caracteritzen per nivells de G+C lleugerament diferents (ja siguin rics en G+C o pobres en G+C), i per una composició de bases força homogènia (Bernardi *et al.*, 1985a; Bernardi i Bernardi, 1986a, 1986b; Ikemura i Aota, 1988).

La variació en el contingut de G+C al llarg del genoma va acompanyada per canvis en l'ús dels codons (grup de tres bases que codifiquen un aminoàcid concret) en els gens que codifiquen proteïnes (Bernardi *et al.*, 1985; Bernardi i Bernardi, 1986a, 1986b; Ikemura i Aota, 1988; Bernardi i Bernardi, 1985b; Aota i Ikemura, 1986). Això és possible perquè el codi genètic (és a dir, l'equivalència entre grups de tres bases del DNA i els diferents aminoàcids de les proteïnes) és degenerat, fet que significa que diferents codons informen un mateix aminoàcid.

En les seqüències codificants de vertebrats, el contingut de G+C en la tercera posició del codó varia dins d'un ample rang i, conseqüentment, els models de tria de codó per un mateix aminoàcid en diferents gens d'un genoma concret sovint són molt diferents entre ells (Aota i Ikemura, 1986). Això és possible pels múltiples casos en què, sigui quina sigui l'última base, el triplet codifica un mateix aminoàcid.

PLANTEJAMENT DEL PROBLEMA

La compartimentació del genoma dels vertebrats homeotermes en zones segons la composició de bases o nucleòtids, ofereix una bona oportunitat per relacionar els canvis en l'ús dels codons amb les variacions de composició dins del mateix genoma. En un treball anterior (Marin *et al.*, 1991) es va mostrar l'existència de diferències, a nivell de grups de codons sinònims, en les relacions entre el contingut en G+C de les seqüències codificants i la tria de codons particulars.

A partir de les observacions experimentals per part de Kornberg i col·laboradors mesurant

freqüències d'aparició de dinucleòtids o dobles (és a dir, grups de dos nucleòtids o dues bases concretes) en molècules de DNA, s'han produït un nombre considerable d'estudis que examinen el fenomen de preferència de doblet en una varietat de genomes que inclouen els virals, mitocondrials, procariòtics i eucariòtics, analitzant en cada un d'ells les seqüències, ja siguin codificants o no codificants, de nucleòtids.

Aquests treballs han revelat que la distribució de la freqüència dels dinucleòtids en el DNA no és uniforme, sinó que alguns dinucleòtids resulten afavorits mentre que d'altres són desfavorits.

Es creu que l'origen de les preferències en els dinucleòtids es deu a biaixos mutacionals (és a dir, que la probabilitat de mutació o canvi d'una base a una altra no és a l'atzar ni igual per totes elles) en la maquinària de replicació del DNA, que poden estar influïts per efectes de les bases veïnes (Bulmer, 1986) o bé per canvis en el medi químic del nucli (com per exemple el contingut en precursors dels nucleòtids; Wolfe *et al.*, 1986). De fet, hi ha una considerable quantitat de proves a favor de l'existència d'un paper tant de selecció com de biaix mutacional.

En aquest treball es tracten aquests factors comparant les distribucions de dinucleòtids en les posicions II-III en el codó, en una àmplia mostra de gens de vertebrats. Els resultats obtinguts confirmen la influència de la base en posició central del codó en la tria de la tercera base del codó, tal i com ja havien apuntat altres autors (Nussinov, 1981; Lipman i Wilbur, 1983). Quan es va analitzar la distribució de les terceres bases del codó en els setze grups de codons definits per les seves primeres i segones bases, es va obtenir un model clar per als conjunts de codons amb una A o una U i una C o una G com a base central del codó; també va resultar aparent en aquestes anàlisis un efecte de la primera base en la distribució de les terceres bases del codó.

Els resultats presentats suggereixen l'existència d'una primera restricció estructural a nivell del DNA en la tria del codó, la qual estaria modulada per les restriccions degudes a la

composició global de bases per tal d'arribar a donar els models d'ús de codons que es troben en el conjunt del genoma (Bibb *et al.*, 1984) o del compartiment genòmic (Bernardi i Bernardi, 1986a, 1986b); s'ha trobat també un biaix molt fort en els codons de gens amb una elevada expressió (Gouy i Gautier, 1982; Bennetzen i Hall, 1982; Sharp *et al.*, 1986; Thomas *et al.*, 1988).

Fóra bo esmentar aquí que a inicis dels anys vuitanta Nussinov va suggerir que alguns aspectes de la tria entre els codons que codifiquen un mateix aminoàcid, poden explicar-se en termes de regularitats observades a nivell d'unitats senzilles de dinucleòtids (Nussinov, 1981), que serien una propietat inherent de les seqüències de nucleòtids (Nussinov, 1983). Més recentment, el fenomen generalment observat de preferència en el codó s'ha contemplat com un «simple reflex del principi de construcció de les seqüències codificants i que té molt poc a veure amb la selecció per se» (Ohno, 1988).

DADES I MÈTODES

El recull per part d'Ikemura i col·laboradors (Aota *et al.*, 1988) dels usos de codons en gens d'una gran varietat d'organismes proveeix el present treball amb una àmplia base de dades per investigar. En aquest estudi s'han utilitzat taules d'ús de codons per a gens del genoma nuclear d'*Homo sapiens* (HUM, amb un nombre de gens, $n = 407$), *Mus musculus* (MUS, $n = 184$), *Rattus norvegicus* (RAT, $n = 230$), *Bos taurus* (BOV, $n = 69$), i *Gallus gallus* (CHK, $n = 71$).

El primer objectiu és comparar el patró de tria de la tercera base del codó entre grups de codons sinònims. Per aquest propòsit s'han dividit provisionalment els aminoàcids codificats sis codons (arginina, leucina i serina) en grups de codons degenerats quatre vegades, és a dir, que quatre codons diferents codifiquen el mateix aminoàcid (Arg4, Leu4 i Ser4) i degenerats dues vegades (Arg2, Leu2 i Ser2);

TAULA I

Distàncies entre els grups de quartets de codons

	Arg4	Leu4	Ser4	Thr	Pro	Ala	Gly
HUM							
Leu4	4,87	—					
Ser4	4,42	6,38	—				
Thr	4,01	5,57	2,14	—			
Pro	4,27	6,33	2,47	2,62	—		
Ala	4,00	5,98	1,60	1,72	2,11	—	
Gly	3,23	4,22	2,87	2,36	3,00	2,34	—
Val	4,83	1,34	5,89	5,76	6,01	5,60	3,92
MUS							
Leu4	4,73	—					
Ser4	5,06	5,83	—				
Thr	4,70	5,33	2,61	—			
Pro	5,45	6,23	2,66	2,71	—		
Ala	4,39	5,39	2,10	1,94	2,59	—	
Gly	3,59	3,81	3,05	2,54	2,96	2,34	—
Val	4,34	1,30	5,56	5,10	5,88	4,82	3,75
RAT							
Leu4	4,33	—					
Ser4	5,07	6,57	—				
Thr	4,46	5,66	3,00	—			
Pro	4,43	5,36	3,34	2,48	—		
Ala	3,78	5,67	2,52	2,64	2,62	—	
Gly	2,70	3,76	3,53	2,46	2,66	2,66	—
Val	4,17	2,20	6,14	5,41	5,40	4,83	3,47
BOV							
Leu4	5,56	—					
Ser4	5,45	6,34	—				
Thr	4,23	5,56	2,34	—			
Pro	6,10	6,28	2,65	2,22	—		
Ala	5,43	6,20	1,84	1,94	2,10	—	
Gly	4,34	4,47	2,52	2,34	2,26	2,63	—
Val	5,21	1,58	5,25	4,46	5,04	5,37	3,35
CHK							
Leu4	7,81	—					
Ser4	6,35	6,83	—				
Thr	5,24	7,07	2,56	—			
Pro	5,96	8,46	3,39	2,83	—		
Ala	3,92	6,86	2,74	2,39	2,35	—	
Gly	3,67	5,26	3,94	2,74	3,50	2,92	—
Val	6,36	1,59	5,09	4,97	6,19	5,13	3,94

TAULA II

Distàncies entre els duets restringits a una pirimidina

	Ser2	Asn	His	Asp	Tyr	Cys
HUM						
Asn	0,69	—				
His	0,73	0,49	—			
Asp	0,50	0,41	0,45	—		
Tyr	0,48	0,60	0,58	0,55	—	
Cys	0,83	0,78	0,74	0,60	0,96	—
Phe	0,51	0,55	0,52	0,40	0,54	0,71
MUS						
Asn	0,64	—				
His	0,79	0,58	—			
Asp	0,65	0,32	0,51	—		
Tyr	0,77	0,68	0,77	0,71	—	
Cys	0,92	0,69	0,84	0,60	0,90	—
Phe	0,54	0,36	0,56	0,34	0,58	0,65
RAT						
Asn	0,48	—				
His	0,66	0,82	—			
Asp	0,52	0,73	0,63	—		
Tyr	0,56	0,44	0,85	0,86	—	
Cys	0,82	0,71	0,88	0,72	0,76	—
Phe	0,78	1,08	0,69	0,42	1,06	0,58
BOV						
Asn	0,42	—				
His	0,69	0,60	—			
Asp	0,68	0,51	0,61	—		
Tyr	0,62	0,30	0,88	0,59	—	
Cys	0,77	0,64	0,62	0,76	0,85	—
Phe	0,78	0,72	0,37	0,30	0,95	0,71
CHK						
Asn	0,30	—				
His	0,62	0,36	—			
Asp	0,94	0,62	0,64	—		
Tyr	0,49	0,31	0,41	0,67	—	
Cys	0,69	0,53	0,75	0,69	0,83	—
Phe	0,72	0,56	0,51	0,58	0,69	0,77

TAULA III

Distàncies entre els duets restringits a una purina

	Arg2	Leu2	Lys	Gln
HUM				
Leu2	1,40	—		
Lys	0,76	0,76	—	
Gln	0,92	0,66	0,39	—
GLu	0,63	0,75	0,38	0,45
MUS				
Leu2	1,51	—		
Lys	0,88	0,66	—	
Gln	1,23	0,49	0,49	—
GLu	0,78	0,78	0,38	0,59
RAT				
Leu2	1,67	—		
Lys	0,90	0,70	—	
Gln	1,05	0,64	0,39	—
GLu	0,63	0,98	0,27	0,41
BOV				
Leu2	1,01	—		
Lys	0,58	0,82	—	
Gln	1,27	1,18	0,59	—
GLu	0,80	0,89	0,38	0,71
CHK				
Leu2	0,90	—		
Lys	0,53	0,47	—	
Gln	0,90	0,53	0,34	—
GLu	0,70	0,61	0,25	0,48

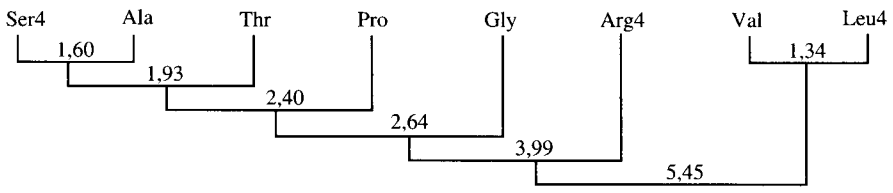
dins de cada grup, els diversos codons difereixen només en el tercer nucleòtid.

En un primer pas s'ha comparat de manera separada el model de tria de la tercera base del codó dins de tres tipus de grups de codons: quartets o grups de codons degenerats quatre vegades (en els quals la tercera posició del codó la pot ocupar qualsevol dels quatre nucleòtids), duets restringits a pirimidina (en els quals a la tercera posició del codó hi ha U o bé C), i duets restringits a purina (en els quals la darrera base

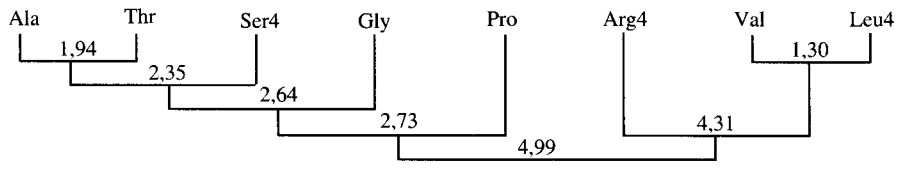
del codó és una A o una G). En aquesta anàlisi, el grup de codons de la isoleucina no s'hi inclou perquè és únic, en el sentit que la tria pot ser feta entre els acabaments de codons A-, U- i C-. La metionina i el triptòfan, que només tenen un codó cadascun, no són d'interès particular, així com els codons d'acabament, els quals només es presenten una vegada a cada gen.

Per comparar la tria de la tercera base del codó entre els grups de codons definits anteriorment, es transformaren els recomptes

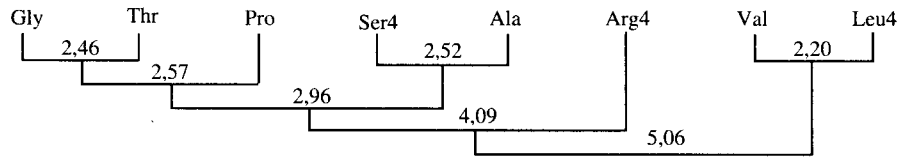
HUM



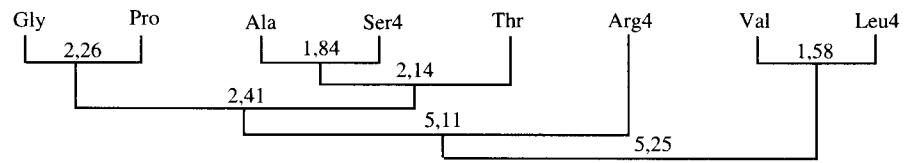
MUS



RAT



BOV



CHK

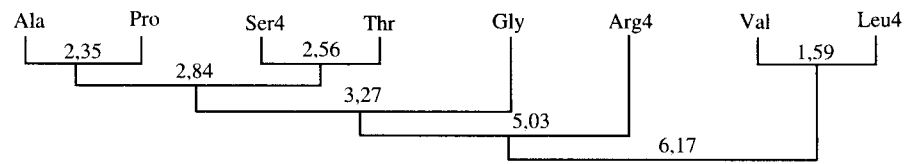
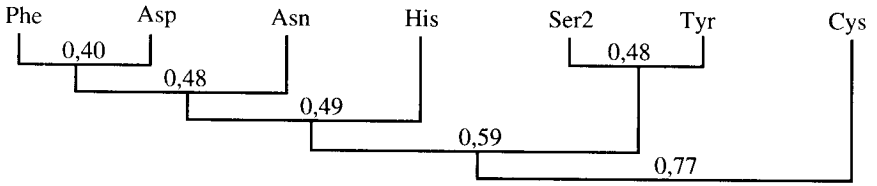
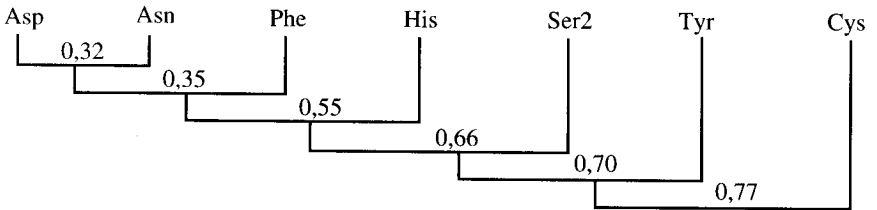


Figura 1. Dendrograms dels grups de quartets de codons.

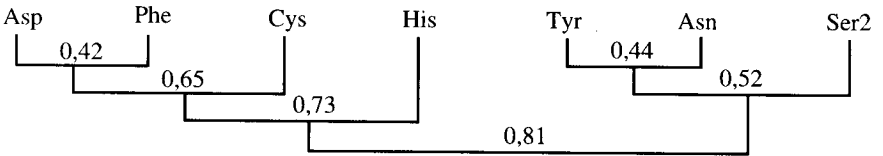
HUM



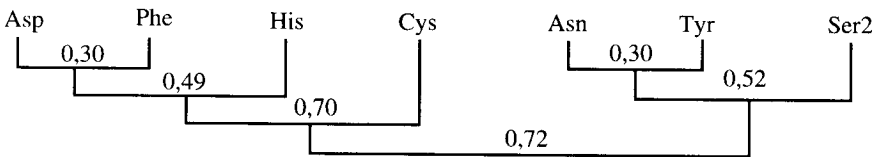
MUS



RAT



BOV



CHK

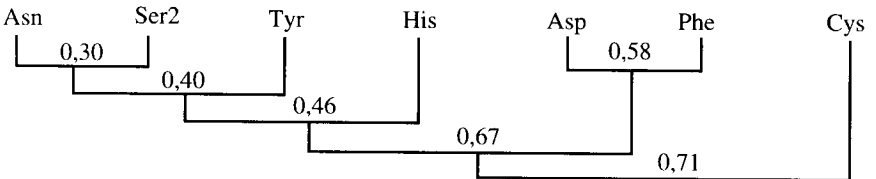
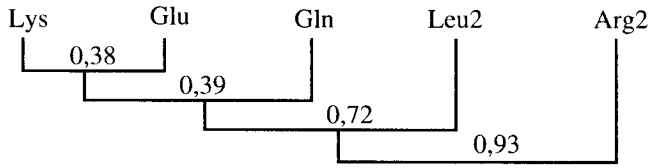
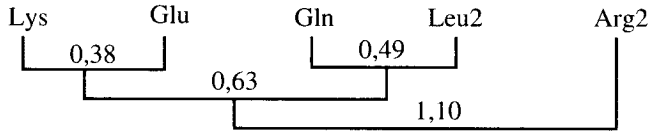


Figura 2. Dendrogrames dels duets restringits a la pirimidina.

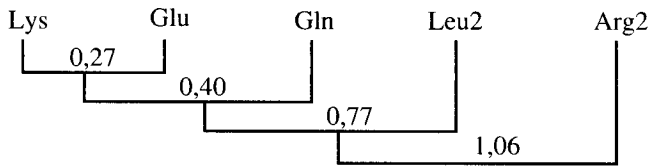
HUM



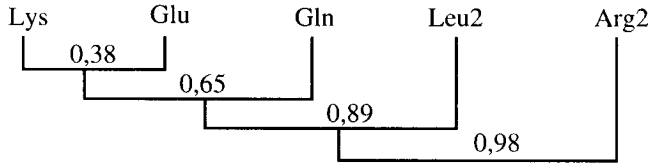
MUS



RAT



BOV



CHK

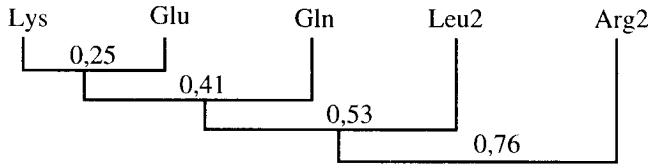


Figura 3. Dendrogrames dels duets restringits a la purina.

absoluts de cada codó per gen, de les taules d'Ikemura, en valors d'ús relatiu de codons sinònims (RSCU) de Sharp i col·laboradors (1986). El valor del RSCU en un gen és la freqüència observada d'aquell codó dividida per la freqüència esperada si l'ús de codons entre el grup apropiat de codons sinònims fos uniforme. Aquest fet fa ressaltar la tria particular de codons sense considerar la composició d'ami- noàcids del producte gènic, i permet comparar entre grups de dades de diferents dimensions.

Una mesura quantitativa de les diferències en la tria de la tercera base del codó entre dos grups de codons X i Y per a un genoma o grup de gens donat, és la suma dels quadrats de les diferències dels valors de RSCU de cada codó en cada gen, dividit pel nombre de gens analitzats:

$$D_{xy} = 1/N (S_i = 1, n (RSCU_{xi} - RSCU_{yi})^2),$$

on N és el nombre de gens considerats, n és el nombre de codons dins del grup de codons, i S representa la suma sobre i.

Calculant el D_{xy} entre cada parell de grups de

codons dins de cada classe (quartets, duets de pirimidina i duets de purina) es va obtenir una sèrie de matrius de diferències o de distàncies per a cada genoma. En aquestes matrius es va aplicar un algorisme de classificació per obtenir dendrogrames que mostressin de manera senzilla les relacions contingudes en les matrius de distància. L'algorisme utilitzat és l'UPGMA o *average linkage*, en el qual es van agrupant les dues classes amb menor distància, i la seva distància conjunta amb les classes restants és la mitjana de les que tenien abans d'ajuntar-se. S'ha calculat a través del programa BMDP1M (Dixon, 1981).

Les matrius de distàncies per als quartets, grups de codons restringits a les pirimidines (C o T) o restringits a la purina (A o G) s'indiquen en les taules i-iii, i els dendrogrames corresponents en les figures 1-3.

En un segon pas, es va aplicar l'esmentada estratègia d'anàlisi als setze blocs de codons del codi genètic, en el qual cada grup de codons es defineix per les seves bases primera i segona, mentre que la tercera posició pot ocupar-la qualsevol dels quatre nucleòtids. Aquest fet relaciona la tria de la tercera base amb les posicions

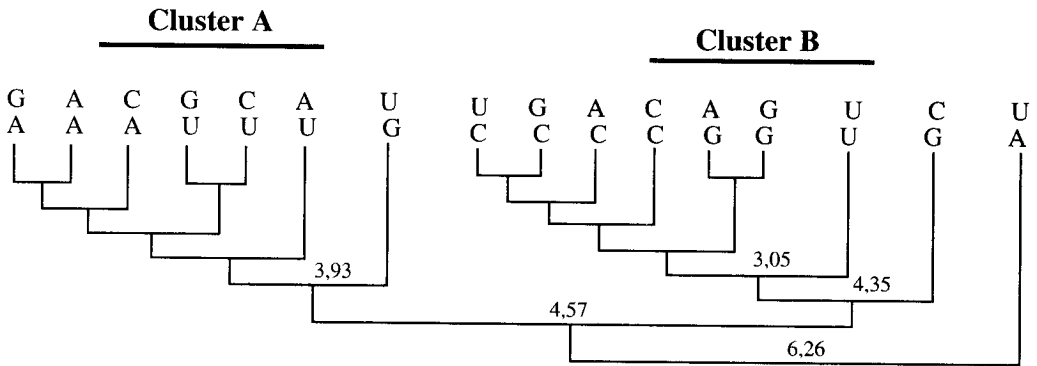


Figura 4. Dendrograma dels grups de codons obtingut per *average linkage* sobre la matriu de distàncies entre blocs. Només s'indiquen les distàncies mitjanes superiors a 3. Les dues primeres bases del codó es presenten en vertical.

TAULA IV

Distàncies entre grups de codons en gens de vertebrats. Les bases centrals del codó comparades s'indiquen entre parèntesis.

(U-C)					(A-G)				
	UU	AU	GU	CU		UA	AA	GA	CA
GC	2,54	3,90	5,22	5,88	GG	5,08	2,85	2,15	4,08
AC	3,02	4,04	5,40	5,66	AG	5,15	3,26	2,66	4,57
CC	3,49	4,49	5,78	6,24	CG	7,01	3,96	3,58	4,81
UC	3,03	4,36	5,78	6,35	UG	7,74	3,65	3,56	4,52
(U-A)					(U-U)				
	UU	AU	GU	CU		UU	AU	GU	CU
GA	3,63	2,29	2,26	2,06	GU	4,82	2,45	—	—
AA	4,22	2,49	2,43	2,00	AU	2,91	—	—	—
CA	5,26	2,79	2,63	1,85	CU	4,94	2,22	1,57	—
UA	3,80	6,06	8,34	9,18	UU	—	—	—	—
(U-G)					(C-C)				
	UU	AU	GU	CU		UC	AC	GC	CC
GG	2,95	3,06	2,26	4,13	GC	2,02	2,05	—	—
AG	3,24	3,40	4,31	4,49	AC	2,48	—	—	—
CG	4,73	4,01	4,72	4,91	CC	2,79	2,59	2,34	—
UG	4,34	3,59	4,34	3,89	UC	—	—	—	—
(C-A)					(A-A)				
	UC	AC	GC	CC		UA	AA	GA	CA
GA	3,98	3,33	3,42	3,44	GA	6,95	1,45	—	—
AA	5,08	4,29	4,56	4,39	AA	7,90	—	—	—
CA	6,66	5,86	6,12	6,12	CA	9,50	2,05	2,03	—
UA	4,52	4,39	3,46	4,76	UA	—	—	—	—
(C-G)					(G-G)				
	UC	AC	GC	CC		UG	AG	GG	CG
GG	3,12	2,45	2,48	2,90	GG	4,38	2,51	—	—
AG	3,77	2,86	3,02	3,28	AG	4,41	—	—	—
CG	4,92	4,36	4,12	4,80	CG	5,93	4,24	3,29	—
UG	5,88	5,54	5,46	5,43	UG	—	—	—	—

primera i segona del doblet, independentment de l'aminoàcid codificat. Els resultats es mostren a la taula iv i a la figura 4.

GRUPS DE QUARTETS DE CODONS

Els dendrogrames de la figura 1 revelen un patró de ramificació lògic en tots cinc genomes. Hi ha dos agrupaments principals: un d'ells format per la parella Valina i Leu4, i l'altre que conté la resta de quartets excepte Arg4, que es reuneix a més distància a un o altre dels dos agrupaments principals.

Els grups de codons de Valina i Leu4 tenen com a tret comú la base U en posició central en el codó, i ambdós són únics dins els quartets ja que mostren una preferència marcada per l'acabament de codó en G, de manera que augmenta el contingut en G+C en la seqüència codificant (Marin *et al.*, 1991).

El segon agrupament principal conté els grups de codons de la Ser4, treonina, prolina, alanina i glicina. Excepte aquesta darrera, la resta tenen la base C en posició central al codó; tots cinc dendrogrames mostren un clar biaix que afavoreix la tria de codons amb acabament en C sota pressió GC, és a dir quan el contingut de G+C és alt. El grup de codons de la glicina té la G com a base central, i s'observa en tots cinc genomes una clara tendència a un descens preferencial en l'ús d'A sobre el d'U en l'acabament dels codons quan hi ha pressió GC (Marin *et al.*, 1991).

Finalment, el grup de codons Arg4 es reuneix amb l'agrupament abans esmentat excepte en el genoma de *Mus*, on s'uneix a l'agrupament format per la Val-Leu4. No obstant això, en tots cinc genomes les distàncies mitjanes de l'Arg4 als dos agrupaments principals són molt grans. La base central del grup de codons de l'Arg4 és la G, igual que en el grup de codons de la glicina, i és amb aquest darrer grup de codons amb qui mostra la distància mínima; a més, en aquest grup de codons de l'Arg4 hi ha un certa tendència

a un descens preferencial en l'ús dels codons amb acabament en A sota pressió GC, tal i com passava en el grup de codons de la glicina (Marin *et al.*, 1991).

DUETS RESTRINGITS A UNA PIRIMIDINA

El dendrograma dels duets restringits a una pirimidina (figura 2) no presenta un patró de ramificació tan consistent com el descrit per als quartets. Tanmateix, es poden observar algunes tendències: per un costat, els grups de codons de la fenilalanina i l'àcid aspàrtic que es troben agrupats en quatre dels cinc genomes, i per l'altre, els grups de codons de l'asparagina, Ser2 i tirosina que s'agrupen en tres dels cinc genomes.

D'entre els duets de pirimidina es troben els doblets UA- que codifiquen la tirosina o els UAA i UAG que són codons de terminació; els doblets UG- que codifiquen la cisteïna o el codó de terminació UGA; i els doblets CA- que expressen histidina, i que hom suposa que augmenten la seva freqüència a causa de pressió mutacional en la 5-metil-citosina dels doblets CpG; per tant, al llarg de l'evolució, els doblets CG canviarien a TG en una de les cadenes del DNA i a AC en la cadena complementària (Bird, 1980; Nussinov, 1981; Bulmer, 1986).

DUETS RESTRINGITS A UNA PURINA

Tal i com mostra el dendrograma de la figura 3, els duets de lisina, àcid glutàmic i glutamina posseeixen un patró de tria de la tercera base del codó molt similar, i es troben agrupats en primer lloc —excepte en el genoma de *Mus*—; tots tres duets tenen en posició central la base A. A continuació, els grups de codons de la Leu2 i Arg2 que presenten la base central del codó diferent (U i G, respectivament) mostren una gran distància entre ells i el primer agrupament.

BLOCS DE CODONS

Finalment, hem aplicat l'anàlisi de distàncies als setze blocs de codons del codi genètic; cada grup de codons es troba definit per les seves bases primera i segona, mentre que la tercera base pot ser qualsevol dels quatre nucleòtids. En aquest pas hem reunit totes les seqüències dels cinc genomes, però els resultats són molt similars quan l'anàlisi es fa en genomes individuals. Les distàncies entre blocs de codons es presenten a la taula iv; estan ordenades en forma de comparacions per totes les possibles parelles de bases en la posició central del codó, per fer més entenedora la percepció dels efectes per separat de les bases segona i primera en la tria de la tercera. Veiem que en cada columna dels deu grups de comparacions les distàncies augmenten amb la primera base en l'ordre de G, A i les més grans amb C.

El dendrograma corresponent a les distàncies del grup de codons es troba representat a la figura 4; mostra dos agrupaments acusats, A i B, l'agrupament A conté els grups GA-, AA-, CA-, GU-, CU- i AU-; l'agrupament B inclou els grups UC-, GC-, AC-, CC-, AG- i GG-. La resta de grups queden fora dels agrupaments principals, i s'uneixen a ells a una distància considerable.

És important esmentar el fet que l'agrupament A conté només grups de codons que tenen l'A o l'U com a segona base, mentre que l'agrupament B es compon exclusivament de grups de codons amb G o C com a bases centrals. Un altre fet interessant és que dins de cada un dels dos agrupaments, els grups de codons que mostren complementarietat per a la primera i segona base es troben en dos subagrupaments separats, per exemple el grup de GA- amb el grup de CU- i el bloc CA- amb el GU- en l'agrupament A, i els grups UC- i CC- amb els seus complementaris respectius AG- i GG- en l'agrupament B.

Voldríem acabar aquest article fent notar que des dels treballs pioners de Grantham i col·laboradors (Grantham i Gautier, 1980;

Grantham *et al.*, 1980, 1981), el nostre coneixement sobre el fenomen de l'ús no aleatori de codons sinònims s'ha incrementat considerablement en els darrers anys (vegeu la revisió d'Ikemura, 1985), tanmateix, encara queden aspectes que no estan completament entesos. En aquesta contribució hem aplicat una mesura intuïtivament vàlida per guanyar un xic més de coneixement dels sistemes d'informació biològica; en particular, l'ordenació no aleatòria dels grups de codons mostrats en la figura 4 podria reflectir alguna característica dels codons independent del seu significat, i que suggeriria l'existència de possibles restriccions estructurals condicionades per l'aparell de traducció del missatge genètic.

AGRAÏMENTS

Voldríem donar les gràcies al Dr. T. Ikemura per facilitar-nos una còpia de les taules d'ús de codons (Aota *et al.*, 1988), al Dr. B. Cubero per ajudar-nos en la discussió i a la Sra. Puig-Troup per col·laborar en la preparació del manuscrit.

BIBLIOGRAFIA

- AOTA, S.; T. IKEMURA (1986). «Diversity in G+C Content at the Third Position of Codons in Vertebrate Genes and its Cause». *Nucleic Acids Research*, núm. 14, p. 6345-6355 i 8702.
- AOTA, S., T. GOJOBORI, F. ISHIBASHI, T. MARUYAMA; T. IKEMURA (1988). «Codon Usage Tabulated from the Genbank Genetic Sequence Data». *Nucleic Acids Research*, núm. 16, p. r315-r402.
- BENNETZEN, J. L.; B. D. HALL (1982). «Codon Selection in Yeast». *Journal of Biological Chemistry*, núm. 257, p. 3026-3031.
- BERNARDI, G., B. OLOFSSON, J. FILIPSKI, M. ZERIAL, J. SALINAS, G. CUNY, M. MEUNIER-ROTYVAL, F. RODIER (1985). «The Mosaic Genome of Warm-Blooded Vertebrates». *Science*, núm. 228, p. 953-958.
- BERNARDI, G.; G. BERNARDI (1985). «Codon Usage and Genome Composition». *Journal of Molecular Evolution*, núm. 22, p. 363-365.
- (1986a). «Compositional Constraints and Genome Evolution». *Journal of Molecular Evolution*, núm. 24, p. 1-11.

- (1986b). «The Human Genome and its Evolutionary Context». *Cold Spring Harbor Symposium on Quantitative Biology*, núm. 51, p. 479-487.
- BIBB, M. J., P. R. FINDLAY; M. W. JOHNSON (1984). «The Relationship between Base Composition and Its Use for the Simple and Reliable Identification of Protein-Coding Sequences». *Gene*, núm. 30, p. 157-166.
- BIRD, A. P. (1980). «DNA Methylation and the Frequency of CPG in Animal DNA». *Nucleic Acids Research*, núm. 8, p. 1499-1504.
- BULMER, M. (1986). «Neighboring Base Effects on Substitution Rates in Pseudogenes». *Molecular Biology and Evolution*, núm. 3, p. 322-329.
- DIXON, W. J. editor (1981). *BMDP Statistical Software*, Berkeley, University of California Press.
- GOUY, M.; C. GAUTIER (1982). «Codon Usage in Bacteria: Correlation with Gene Expressivity». *Nucleic Acids Research*, núm. 10, p. 7055-7074.
- GRANTHAM, R.; C. GAUTIER (1980). «Genetic Distances from mRNA Sequences». *Naturwissenschaften*, núm. 67, p. 93-94.
- GRANTHAM, R., C. GAUTIER, M. GOUY, M. JACOBZONE; R. MERCIER (1981). «Codon Catalog Usage and the Genome Hypothesis». *Nucleic Acids Research*, núm. 9, p. r43-r74.
- GRANTHAM, R., C. GAUTIER, M. GOUY, R. MERCIER; A. PAVE (1981). «Codon Catalog Usage Is a Genome Strategy Modulated for Gene Expressivity». *Nucleic Acids Research*, núm. 9, p. r43-r74.
- IKEMURA, T. (1985). «Codon Usage and tRNA Content in Unicellular and Multicellular Organisms». *Molecular Biology and Evolution*, núm. 2, p. 13-14.
- IKEMURA, T.; S. AOTA (1988). «Global Variation in G+C Content Along Vertebrate Genome DNA-Possible Correlation with Chromosome Band Structures». *Journal of Molecular Evolution*, núm. 203, p. 1-13.
- LIPMAN, D. J.; W. WILBUR (1983). «Contextual Constraints on Synonymous Codon Choices». *Journal of Molecular Biology*, núm. 163, p. 363-376.
- MARIN, A., J. BERTRANPETIT, J. L. OLIVER; J. R. MEDINA (1989). «Variation in G+C Content and Codon Choice: Differences Among Synonymous Codon Groups in Vertebrate Genes». *Nucleic Acids Research*, núm. 17, p. 6181-6189.
- NUSSINOV, R. (1981a). «The Universal Dinucleotide Asymmetry Rules and the Aminoacid Codon Choice». *Journal of Molecular Evolution*, núm. 17, p. 237-244.
- (1981b). «Eukariotic Dinucleotide Preference Rules and Their Implications for Degenerate Codon Usage». *Journal of Molecular Biology*, núm. 149, p. 125-131.
- (1984). «Doublet Frequencies in Evolutionary Distinct Groups». *Nucleic Acids Research*, núm. 12, p. 1749-1763.
- OHNO, S. (1988). «Codon Preferences is but an Illusion Created by the Construction Principle of Coding Sequences». *Proceedings of the National Academy of Sciences USA*, núm. 85, p. 4378-4382.
- SHARP, P. M., T. M. F. TUOHY; R. MOSURSKI (1986). «Codon Usage in Yeast - Cluster - Analysis Clearly Differentiates Highly and Lowly Expressed Genes». *Nucleic Acids Research*, núm. 14, p. 5125-5143.
- THOMAS, L. K., D. B. DIX; R. C. THOMPSON (1988). «Codon Choice and Gene Expression: Synonymous Codons Differ in Their Ability to Direct Aminoacylated-Transfer RNA Binding to Ribosomes in vitro». *Proceedings of the National Academy of Sciences USA*, núm. 85, p. 4242-4246.