
ATTI ACCADEMIA NAZIONALE DEI LINCEI
CLASSE SCIENZE FISICHE MATEMATICHE NATURALI

RENDICONTI

ANTONIO FAIELLA, DARIO ACAMPORA, MARIA PANNese,
MAURIZIO D'ESPOSITO, ANTONIO SIMEONE, EDOARDO
BONCINELLI

Organizzazione delle regioni genomiche umane codificanti omeodomini

*Atti della Accademia Nazionale dei Lincei. Classe di Scienze Fisiche,
Matematiche e Naturali. Rendiconti, Serie 8, Vol. 83 (1989), n.1, p. 293–298.*
Accademia Nazionale dei Lincei

<http://www.bdim.eu/item?id=RLINA_1989_8_83_1_293_0>

L'utilizzo e la stampa di questo documento digitale è consentito liberamente per motivi di ricerca e studio. Non è consentito l'utilizzo dello stesso per motivi commerciali. Tutte le copie di questo documento devono riportare questo avvertimento.

*Articolo digitalizzato nel quadro del programma
bdim (Biblioteca Digitale Italiana di Matematica)
SIMAI & UMI*

<http://www.bdim.eu/>

Embrilogia e morfogenesi. — *Organizzazione delle regioni genomiche umane codificanti omeodomini.* Nota (*) di ANTONIO FAIELLA, DARIO ACAMPORA, MARIA PANNESE, MAURIZIO D'ESPOSITO, ANTONIO SIMEONE e EDOARDO BONCINELLI, presentata dal Socio G. MONTALENTI.

ABSTRACT. — *Organization of human homeobox genes.* We have isolated et sequenced 17 human homeoboxes present in four genomic regions located on chromosomes 2, 7, 12 and 17. These regions are termed *HOX* complex loci. Here we show that these loci evolved from a single ancestral locus through large-scale duplication events. Comparison of human and *Drosophila* homeoboxes suggests that basic features of this ancestral locus might have predated the divergence of vertebrates and invertebrates.

KEY WORDS: Development; Gene evolution; Homeobox.

RIASSUNTO. — Abbiamo isolato e sequenziato 17 omeobox umane presenti in quattro regioni genomiche localizzate sui cromosomi 2, 7, 12 e 17. Queste regioni sono chiamate loci «HOX». Il nostro gruppo ha mostrato che questi loci evolvono da un singolo locus ancestrale attraverso eventi di duplicazione su larga scala. Il paragone tra le omeobox di *Drosophila* e di uomo suggerisce che le caratteristiche di base di questo locus ancestrale potrebbero aver predatato la divergenza tra vertebrati ed invertebrati.

L'omeobox è una sequenza di DNA di 183 paia di basi (bp) che è stata originariamente individuata in alcuni geni di *Drosophila* che controllano lo sviluppo di questo insetto [1]. Omeodominio è stato chiamato il dominio proteico di 61 aminoacidi codificato dalla sequenza omeobox. Questo si trova invariabilmente verso il carbossi-terminale della proteina codificata da questi geni e si è ipotizzato che abbia la proprietà di legarsi a sequenze specifiche di DNA con la probabile funzione di regolare l'espressione di altri geni [1, 2].

Numerose sequenze omeobox sono state individuate in organismi superiori quali il riccio di mare, l'anfibio *Xenopus laevis*, il topo e l'uomo. Noi ci siamo concentrati sullo studio di geni umani contenenti un omeobox. Recentemente abbiamo isolato quattro cloni di DNA complementare (cDNA), chiamati *c1*, *c8*, *c10* e *c13*, di cui abbiamo studiato la struttura e l'espressione in tessuti adulti e embrionali [3-5]. Nell'intento di comprendere l'organizzazione trascrizionale dei messaggeri corrispondenti ai quattro cloni abbiamo isolato le corrispondenti regioni genomiche e successivamente abbiamo esteso l'analisi alle regioni circostanti. Questa analisi ha condotto all'isolamento di un certo numero di omeobox umane e alla possibilità di collocarle nel loro contesto genomico e cromosomico. La fig. 1 mostra la mappa di quattro regioni cromosomiche, chiamate loci *HOX*, situate sui cromosomi 2, 7, 12 e 17. Queste contengono complessivamente 17 sequenze omeobox così distribuite: 4 in 40 chilobasi (kb) di *HOX-1*, 7 in 60 kb di *HOX-2*, 4 in 50 kb di *HOX-3* e 2 in 30 kb sul cromosoma 2. Questa organizzazione in gruppi appartenenti a loci genetici complessi non è affatto caratteristica delle omeobox umane. Un'organizzazione analoga è stata osservata nel topo e in misura per ora più limitata in *X. laevis*. Anche le omeobox di *Drosophila* appaiono raggruppate in regioni genomiche limitate. Ne sono state individuate 3

(*) Pervenuta all'Accademia il 3 agosto 1988.

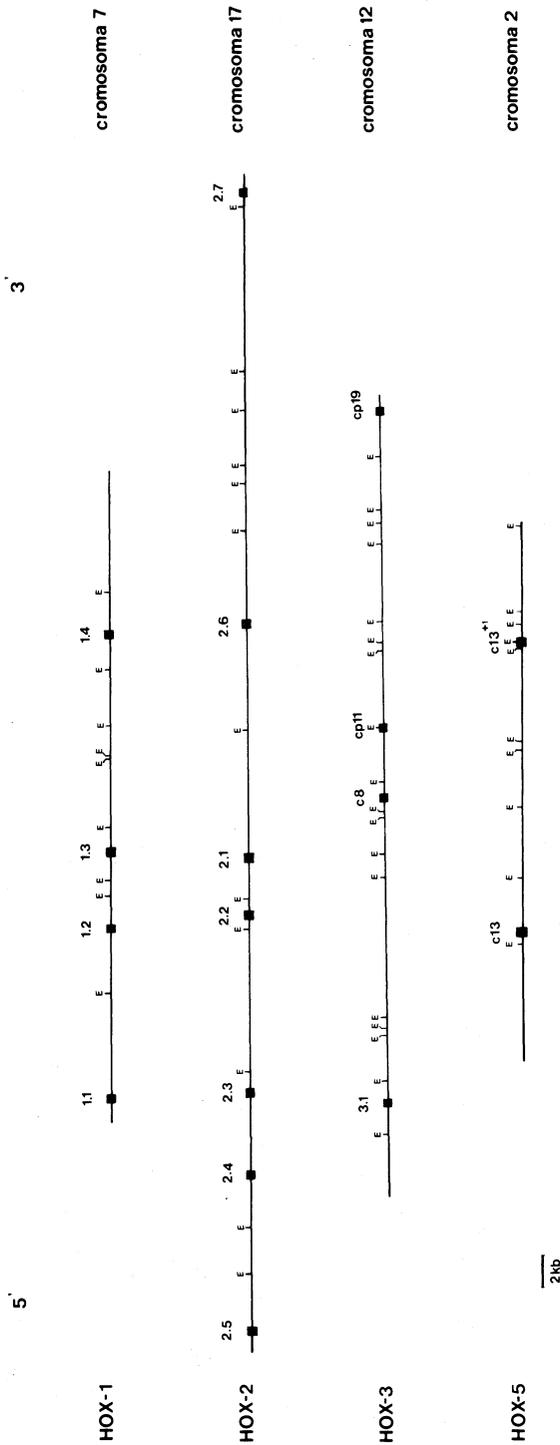


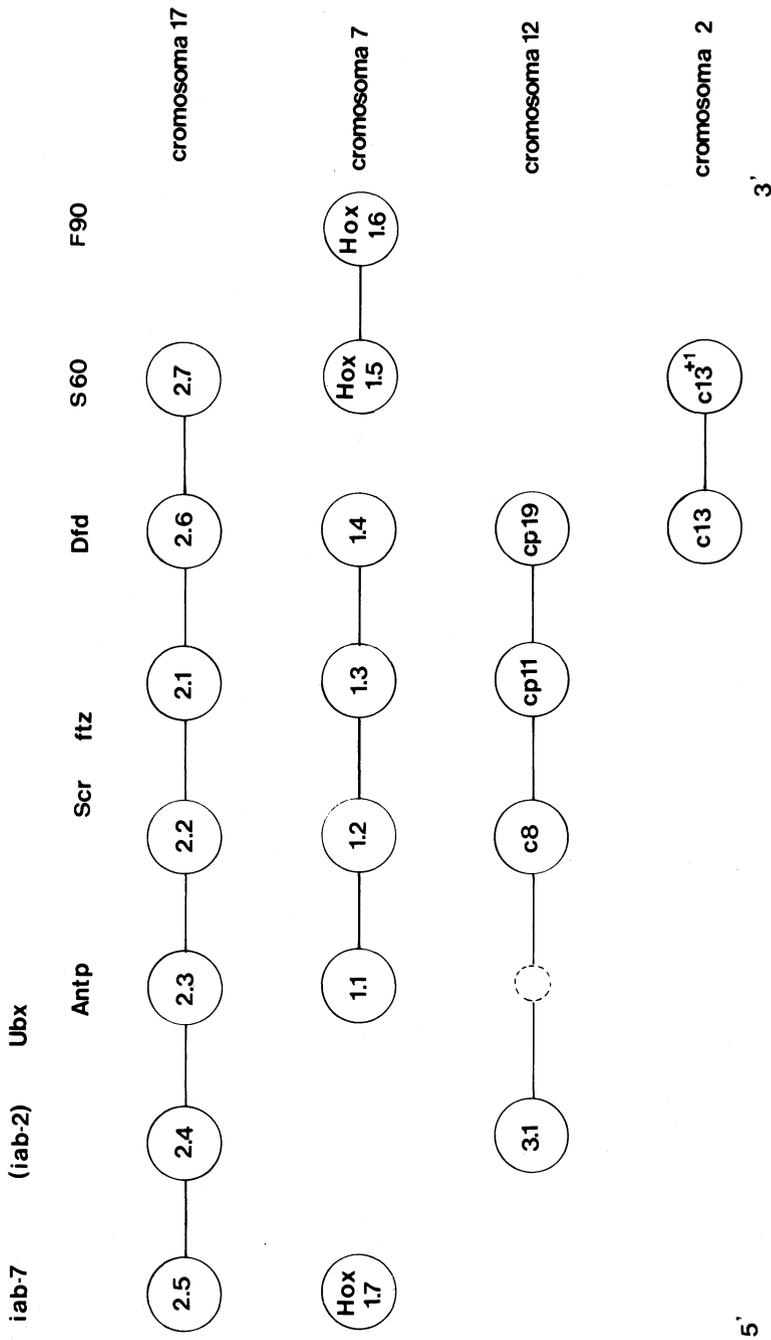
Fig. 1. - Mappa di restrizione di quattro regioni cromosomiche umane contenenti omeobox (quadrati neri). La direzione di trascrizione è da sinistra verso destra per tutte le omeobox. Le mappe sono ottenute dall'analisi di un certo numero di cloni cosmici sovrapposte isolati da una genoteca costruita nel vettore pcos2EMBL (9). Il locus *HOX-1* si trova sul cromosoma 7, *HOX-2* sul 17, *HOX-3* sul 12 e *HOX-5* sul 2. Le varie omeobox sono indicate con la loro designazione ufficiale (ad esempio *HOX-1.4* o *HOX-2.2*) laddove è possibile, mentre alcune conservano la loro designazione originaria perché non è stabilita ancora quella ufficiale: *c8*, *cp11*, *cp19*, *c13* e *c13 + 1*. E = *EcoRI*.

appartenenti ai geni *Ultrabithorax* (*Ubx*), *infraabdominal-2* (*iab-2*) e *iab-7* in una regione chiamata bithorax-complex (BX-C) e almeno 6 appartenenti ai geni *Antennapedia* (*Antp*), *Sex combs reduced* (*Scr*), *fushi tarazu* (*ftz*), *Deformed* (*Dfd*), *zerknüllt* (*zen*) e *labial* (*lab*) in una seconda regione chiamata Antennapedia-complex (ANT-C) [2].

Abbiamo determinato la sequenza nucleotidica delle 17 omeobox e da questa siamo risaliti alla sequenza aminoacidica primaria degli omeodomini codificati. In fig. 2 sono riportate le sequenze dei 17 omeodomini umani (quelli designati HOX più *c8*, *cp11*, *cp19*, *c13* e *c13 + 1*) insieme ad alcuni di topo (*Hox-1.5*, *1.6* e *1.7*), di *X. laevis* (MM3, Ac1, Xhox-1B, Xhox-1A) e di *Drosophila* (dove S60 rappresenta l'omeodominio di *zen* e F90-2 probabilmente quello di *lab*). È evidente l'alto grado di conservazione dei vari omeodomini. Nonostante questa relativa omogeneità, tra questi si possono individuare

	10	20	30	40	50	60	
Antp	RKRGRQTYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRMRMKWKKENK						TKGEP
HOX2.3	-----	-----Y-----					-A-PG
Hox1.1	-----	-----				-----H-	DESQA
MM3	-----	-----		-----V-----			ASSTL
c8	-R---I-S-----			-----N-----		-----SN	LTSTL
HOX2.2	TA-----	-----Y-----				-----S-	LLSAS
HOX1.2	GR-----			-----N-----			LINST
Ac1	-R---I-S-----			-----N-----		-----SN	LSSTL
Scr	T-Q-TS-----				-----M-----	-----L-H-	MASMN
cp11	G-S-TS-----			-----NN--N-----		-----DS-	M-SKE
HOX2.1	G-A-TA-----				-----S-----	-----D--	L-SMS
HOX1.3	G-A-TA-----				-----S-----	-----D--	L-SMS
Xhox-1B	G-A-TA-----				-----T-S-----	-----D--	
ftz	S-T-----		-----I-----	-----D-N-S-S-----		-----S-DRT	LDSS-
c13	P-S-TA-Q-V-----			-----T-S-----		-----DH-	LPNTK
cp19	P-S-AA-Q-V-----	-----Y-----		-----S-S-----		-----DHR	LPNTK
HOX2.6	P-S-TA-Q-V-----	-----Y-----		-----V-----	-----S-----	-----DH-	LPNTK
HOX1.4	P-S-TA-Q-V-----				-----T-S-V-----	-----DH-	LPNTK
Xhox-1A	A-S-TA-Q-V-----	-----Y-----		-----V-----	-----T-R-S-----	-----DH-	LPNTK
Dfd	P-Q-TA-H-I-----	-----Y-P-----		-----Q-T-V-S-----		-----D--	LPNTK
c13+1	S-V-TA-SA-LV-----		-----C-P-V-M-NL-N-----			-----Y-DQ-	A--IL
HOX2.7	S-A-TA-SA-LV-----		-----FV-P-V-M-NL-N-S-----			-----Y-DQ-	A--IL
Hox1.5	S--TA-P-LV-----		-----M-P-V-M-NL-N-----			-----Y-DQ-	G--IL
S60	L-S-TAF-SV-LV-----	N-KS-M-Y-T-----	-----QR-S-C-----			-----V-F-DIQ	
Hox1.6	PNAV-TNS-TK-LT-----		-----K-----AA-V-----	AS-Q-N-T-V-----		-----Q-RE-	EGLL-
F90-2	NNS-TNF-NK-LT-----		-----A-----	NT-Q-N-T-V-----		-----Q-RV-	EGLI-
HOX3.1	-RT-----S-----	-----L-P-K-----	VS-G-V-----			-----N	KDKL-
HOX2.4	-R-----S-----	-----L-P-K-----	VS-G-V-----			-----N	KDKF-
Ubx	-R-----	-----T-H-----	-----M-Y-----	-----E-----		-----L-IQ	AIK-L
HOX2.5	SRKK-CP-K-----	-----L-M-----	D-H-V-RL-N-S-V-----			-----M-M--	EQ-KE
Hox1.7	TRKK-CP-KH-----	-----L-M-----	D-Y-V-RL-N-V-----			-----M-I--	DRAKD
iab-7	VRKK-KP-SKF-----	-----L-A-VSKQK-W-L-RN-Q-----				-----V-N-NSQ	

Fig. 2. - Lista di sequenze primarie di omeodomini. È usato il codice a una lettera per gli aminoacidi. Tutte le sequenze sono riferite al dominio di *Antennapedia* che rappresenta una sorta di sequenza consenso: i trattini indicano identità. Dove disponibili sono riportati anche 5 aminoacidi che seguono immediatamente l'omeodominio. Sono raggruppati gli omeodomini che mostrano una maggiore omologia di sequenza.



Gruppi I II III IV V VI VII VIII
 Fig. 3. - Rappresentazione schematica di 20 omeodomini (cerchi) nella loro disposizione sui quattro cromosomi. Le distanze tra le omeobox non sono rispettate. La direzione di trascrizione è da sinistra a destra. Le regioni genomiche sono state allineate in modo che omeodomini dello stesso gruppo si trovino sulla stessa verticale. Sono stati inclusi tre omeodomini di topo (*Hox-1.5*, *1.6* e *1.7*) per completare il quadro dato che gli omologhi umani non sono ancora stati isolati e analizzati. Il cerchio piccolo tratteggiato indica un omeodominio previsto dallo schema ma ancora da individuare. Sopra lo schema sono riportati gli omeodomini di *Drosophila* appartenenti ai complessi BX-C e ANT-C. Ogni omeodominio di *Drosophila* è stato disposto vicino al gruppo (o ai gruppi) cui somiglia di più. S60 rappresenta l'omeodominio di *zer1*, mentre F90 dovrebbe appartenere a *lab*, anche se questo punto non è perfettamente chiaro. Il gene *iab-2* è indicato tra parentesi perché la sequenza del suo omeodominio non è disponibile. La sua affinità al gruppo II è quindi da considerarsi ipotetica.

delle sottofamiglie o gruppi. Membri dello stesso gruppo sono più conservati anche se appartenenti a specie diverse. Per restare agli omeodomini umani, ci siamo chiesti se membri dello stesso gruppo occupassero posizioni corrispondenti nei quattro loci *HOX*, cioè se il primo omeodominio al 5' di un locus fosse l'omologo del primo degli altri loci, il secondo l'omologo dei secondi e così via. Anche se il quadro non è ancora completo possiamo affermare che la risposta è affermativa. La fig. 3 mostra schematicamente gli omeodomini nella loro disposizione cromosomica. Sulla stessa verticale sono stati posti omeodomini appartenenti allo stesso gruppo. Per completare lo schema sono stati inclusi tre omeodomini di topo di cui non è stato ancora isolato l'omologo umano. Come si vede i gruppi di conservazione aminoacidica si succedono nello stesso ordine nei vari loci. Questi dati suggeriscono che i quattro loci umani si sono originati per duplicazione dell'intera regione cromosomica ancestrale. A questi eventi di duplicazione è seguita una distribuzione in vari cromosomi. Questa ipotesi è avvalorata dalla presenza di geni di varie catene del collagene [6] e protooncogeni della stessa famiglia in regioni cromosomiche adiacenti in tutti e quattro i casi.

L'ipotesi sull'origine evolutiva di questi loci può essere spinta un passo più avanti considerando gli omeodomini di *Drosophila*. La sequenza di questi è ovviamente più distante da quella dei geni umani di quanto lo sia quella di altri vertebrati. Ciò nonostante si può tentare di mettere gli omeodomini di *Drosophila* in relazione con il gruppo a cui più somigliano, o da cui meno differiscono. Il risultato di questa analisi è mostrato in fig. 3, dove gli omeodomini di *Drosophila* appartenenti ai complessi BX-C e ANT-C sono stati posti sopra alla verticale del gruppo (o dei gruppi) cui più somigliano. Il fatto sorprendente è che gli omeodomini di *Drosophila* allineati su questa base si vanno a collocare nello stesso ordine fisico dei corrispondenti geni sul cromosoma di *Drosophila*. A differenza degli omeodomini di vertebrato, questi sono ripartiti in due loci complessi, come abbiamo già ricordato. A seguito di queste osservazioni si può ipotizzare che una regione genomica contenente diverse omeobox, almeno quattro, preesistesse come tale in organismi vissuti prima della separazione fra Protostomi e Deuterostomi. Nella linea evolutiva che porta ai vertebrati questa regione si è mantenuta intatta e successivamente duplicata. Nella linea che porta ai Ditteri la regione si è spezzata all'altezza del gruppo III con la possibile duplicazione del corrispondente gene, dando luogo a *Ubx* e *Antp*. Nell'unico insetto diverso dalla *Drosophila* finora analizzato sotto questo profilo, *Tribolium castaneum*, la regione cromosomica appare ancora intatta, contenendo geni che in *Drosophila* sono distribuiti sui due complessi [7].

Il parallelismo strutturale tra le omeobox di *Drosophila* e quelle dei vertebrati potrebbe a sua volta sottendere un parallelismo più profondo. È noto da tempo che omeodomini appartenenti a geni omeotici di *Drosophila* raffigurati a destra della fig. 3 sono espressi in regioni anteriori del corpo dell'insetto e omeodomini via via più a sinistra sono espressi in regioni sempre più caudali. Questo isomorfismo fra ordine fisico dei geni sul cromosoma e regioni corporee che questi controllano non ha trovato finora spiegazione esauriente [2]. Questo isomorfismo sembra estendersi ai vertebrati. Alcuni omeodomini rappresentati sulla destra della fig. 3 sono espressi in regioni dell'embrione di *X. laevis*, topo eomo che sono anteriori rispetto a regioni dove sono

espressi omeodomini spostati più sulla sinistra della fig. 3 [8]. Attualmente siamo impegnati nell'appurare quanto generale possa essere questa osservazione e verificare le sue implicazioni per la comprensione del ruolo biologico svolto da questa famiglia genica.

BIBLIOGRAFIA

- [1] AKAM M., 1987. *The molecular basis for metameric pattern in the Drosophila embryo*. Development, 101: 1-22.
- [2] GEHRING W. J. and HIROMI Y., 1986. *Homeotic genes and the homeobox*. Ann. Rev. Genet., 20: 147-173.
- [3] SIMEONE A., MAVILIO F., BOTTERO L., GIAMPAOLO A., RUSSO G., FAIELLA A., BONCINELLI E. and PESCHLE C., 1986. *A human homeo box gene specifically expressed in spinal cord during embryonic development*. Nature, 320: 763-765.
- [4] MAVILIO F., SIMEONE A., GIAMPAOLO A., FAIELLA A., ZAPPAVIGNA V., ACAMPORA D., POIANA G., RUSSO G., PESCHLE C. and BONCINELLI E., 1986. *Differential and stage-related expression in embryonic tissues of a new human homeobox gene*. Nature, 324: 664-668.
- [5] SIMEONE A., MAVILIO F., ACAMPORA D., GIAMPAOLO A., FAIELLA A., ZAPPAVIGNA V., D'ESPOSITO M., PANNESE M., RUSSO G., BONCINELLI E. and PESCHLE C., 1987. *Two human homeobox genes, c1 and c8: structure analysis and expression in embryonic development*. Proc. Natl. Acad. Sci. USA, 84: 4914-4918.
- [6] CANNIZZARO L. A., CROCE C. M., GRIFFIN C. A., SIMEONE A., BONCINELLI E. and HUEBNER K., 1987. *Human homeobox-containing genes located at chromosome regions 2q31-2q37 and 12q12-12q13*. Am. J. Hum. Genet., 41: 1-15.
- [7] BEEMAN R. W., 1987. *A homoeotic gene cluster in the red flour beetle*. Nature, 327: 247-249.
- [8] GAUNT S. J., 1988. *Mouse homeobox gene transcripts occupy different but overlapping domains in embryonic germ layers and organs: a comparison of Hox-3.1 and Hox-1.5*. Development, 102: 159-174.
- [9] POUSTKA A., RACKWITZ H. R., FRISCHAUF A. M., KOHN B. and LEHRACH H., 1984. *Selective isolation of cosmic clones by homologous recombination in Escherichia coli*. Proc. Natl. Acad. Sci. USA, 81: 4129-4133.