Mammalian Genome 13, 456–462 (2002).
DOI: 10.1007/s00335-001-2157-0

Mammalian Genome
Incorporating Mouse Genome

Segmental paralogy in the human genome: a large-scale triplication on 1p, 6p, and 21q

Pierluigi Strippoli,* Pietro D'Addabbo,* Luca Lenzi, Sandra Giannone, Silvia Canaider, Raffaella Casadei, Lorenza Vitale, Paolo Carinci, Maria Zannotti

Research Center for Molecular Genetics "Fondazione CARISBO" at the Institute of Histology and General Embryology, University of Bologna, 40126, Bologna, Italy

Received: 7 December 2001 / Accepted: 17 April 2002

Abstract. Few cases of large-scale segmental paralogy have been reported in the human genome. We have identified a large (~500 kb) segment on human chromosome (HC) 21 (21q22) that is triplicated on HC 1 (1p35) and HC 6 (6p12-21). We also identified a new member of *CLIC* (Chloride Intracellular Channel) family on 21q, namely *CLIC6*. All three segments appear to include three functional members of three different gene families: *DSCR1*-like (Down Syndrome Candidate Region 1-like), *CLIC*, and *AML*/*Runt* (Acute Myeloid Leukemia/Runt). Molecular evolution analysis shows a common evolutionary origin for the triplicated regions. This finding of a further large-scale genomic triplication that went undetected at previously systematic automated searches provides a new model for gene divergence study and underlines the need for new tools to effectively detect inter-chromosomal similarity. An algorithm to overcome current limitations is proposed.

Introduction

Members of the same gene family are often dispersed in paralogous regions of the genome, suggesting a close evolutionary relationship (Strachan and Read 1999). In certain cases in which whole chromosomal segment is duplicated, some genes present in the segment can cluster, retaining their reciprocal associations in distant loci (synteny). Regions of segmental paralogy are difficult to identify owing to the divergences that occur during evolution. Identification of synteny often provides the main indication of cases of paralogy. The few reports of regions containing clusters of functional genes displaying interchromosomal segmental paralogy are summarized in Strachan and Read (1999) and IHGSC (2001). When Hattori et al. (2000) sequenced human chromosome (HC) 21, they found only a few small duplications, the largest (~100 kb) being similar to regions on HC 4, 7, 20, and 22. Furthermore, an automated system revealed the existence of paralogy regions between HC 21 and HC 6,11,13, and Y (see Fig. 13 in Venter et al. 2001).

We recently identified the *DSCR1*-like family (Strippoli et al. 2000a), which includes *DSCR1* on HC 21. The observation that the three *DSCR1*-like members lie in proximity of the three AML genes on 1p, 6p, and 21q has led us to characterize a further case of segmental paralogy: a ~500-kb region that is triplicated in humans on 1p35, 6p12-p21, and 21q22. Each region includes three functional members of three different human gene families, namely, *AML*, *DSCR1*-like, and *CLIC*. We have formally identified a new *CLIC* member (*CLIC6*) on HC 21. Phylogenetic analysis of the three families suggests that the segment on HC 21 is the most ancient. We also propose guidelines for a new algorithm to detect interchromosomal similarities that have previously gone undetected at automated analysis.

Materials and Methods

Sequence and map database searches. Between January 2001 and January 2002 we consulted the databases of the HC 21 Sequencing and Mapping Consortium (http://hgp.gsc.riken.go.jp/chr21), the NCBI (National Center for Biotechnology Information) (http://www.ncbi.nlm.nih.gov), and the Human Genome Project Working Draft (http://www.genome.ucsc.edu), inserting the nucleotide and amino acid sequences, of the DSCR1-like and AML genes as query. The identified EST sequences were manually assembled in consensus sequences, translated by using DNAId software for Macintosh, and assigned to the respective gene family. "Blast 2 sequences" software was used to identify the human ortholog of each non-human gene identified (http: //www.ncbi.nlm.nih.gov/blast).

Reverse transcription (RT) – polymerase chain reaction (PCR) amplification. RNA and cDNA sources were total RNA from human placenta (pool of two whole placentas), adult brain, and pancreas, and cDNA from human liver (all purchased from Clontech, Palo Alto, Calif.). One microgram of total RNA was reverse transcribed at 42°C for 60 min in a 25-μl final volume with cloned Moloney murine leukemia virus reverse transcriptase 200 U (Promega, Madison, Wis, used with companion buffer), 5 μM oligo dT-15, and 500 μM for each dNTP. The reaction was stopped by incubation at 95°C for 15 min. Amplification primer were designed with the Amplify software (ftp://ftp.ebi.ac.uk/pub/software/mac/). For PCR, the primer sequences were (5′-3′ direction): #1-TGGGGA CCCAACATCCCGAATC (forward) and #2-GTTTTCGTA TCCTTGCTCACTCAAC (reverse); the oligonucleotides were synthesized by Sigma-Genosys, Pampisford, Cambridgeshire, UK. PCR experiments were performed in 50 μl final volume, containing 5 μl reverse transcription mix or 5μl of cDNA, 1.25 U Taq Polymerase (TaKaRa, Shiga, Japan) with companion reagents (0.2 mM each dNTP, 1.5 mM $MgCl_2$, 1 × PCR buffer), and 0.3 μM of each primer. An initial 2-min denaturation step at 94°C was followed by amplification for 40 cycles (30 s 94°C, 30 s at 61°C, 30 s at 72°C) and final extension for 7 min at 72°C.

*The authors contributed equally to the work.

Correspondence to: P. Carinci; E-mail: carinci@alma.unibo.it

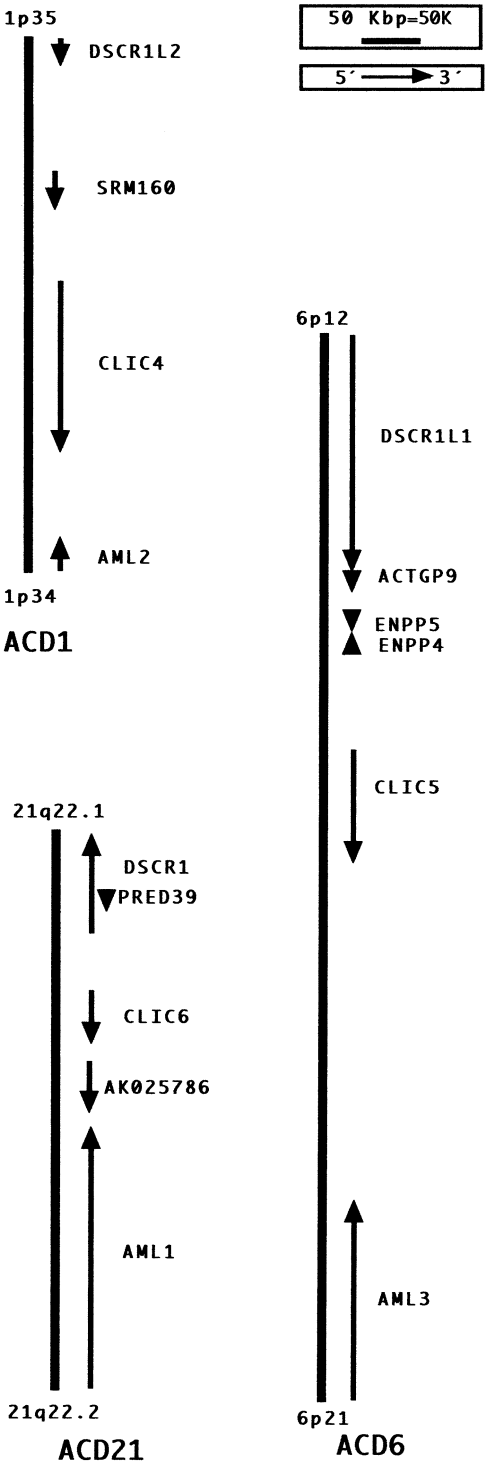

Fig. 1. Physical maps of ACD1, ACD6, and ACD21 clusters. Arrows indicate direction of gene transcription.

Sequencing and analysis. RT-PCR product was gel analyzed following standard methods and was purified with QIAquick kit (QIAGEN, Hilden, Germany). Automated sequence analysis of both DNA strands was performed with the same primers used in the PCR reaction. The BigDye chain-terminator method was used with an automated ABI 310 DNA sequencer (Perkin-Elmer, Foster City, Calif.). Each region was sequenced twice with two independent amplification reaction products as template.

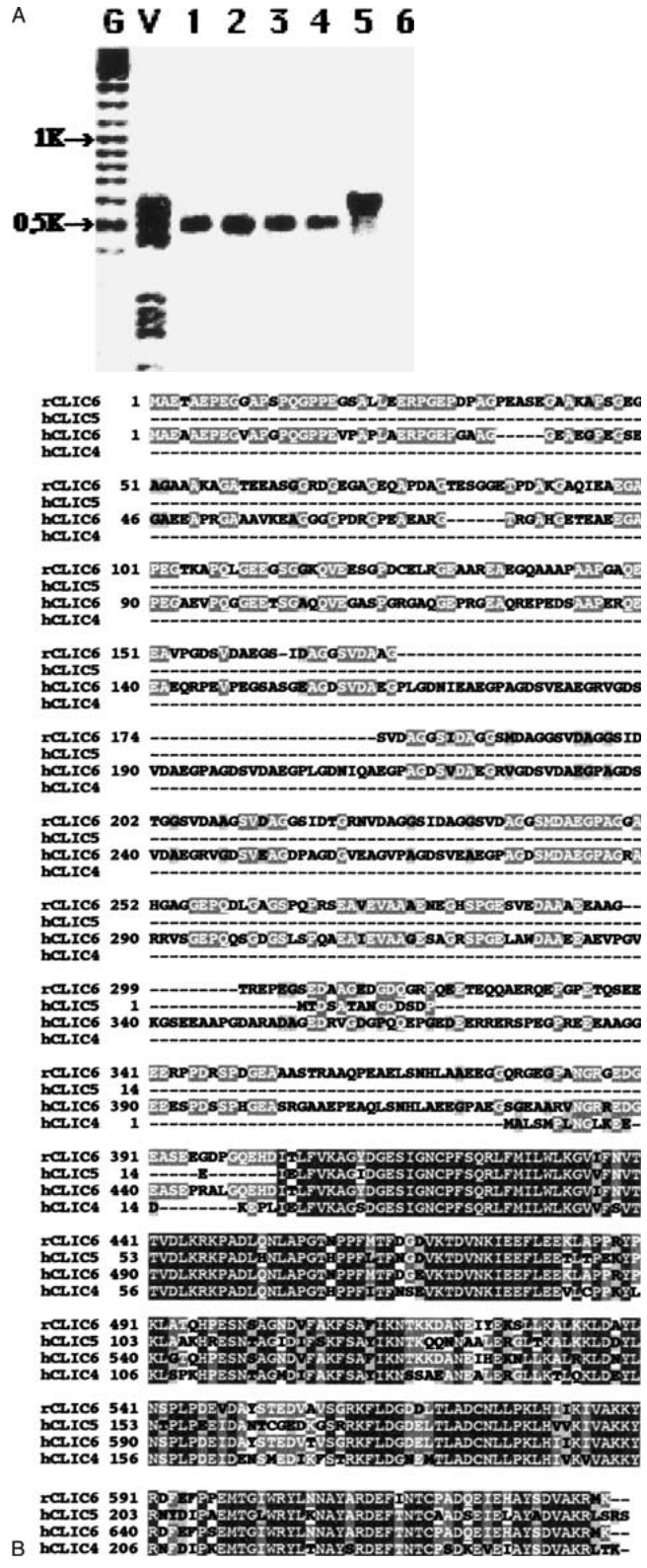

Fig. 2. (A) Gel electrophoresis of the *CLIC6* amplification products (ethidium bromide staining, photographic negative): G and V = respectively, "GeneRuler" and "Marker V" molecular weight markers; lanes 1–4, *CLIC6* from four human tissues: brain, pancreas, placenta, liver; lanes 5–6, positive (*B2M*) and negative PCR samples. **(B)** ClustalW alignment among three human CLIC proteins (hCLIC4, hCLIC5, and hCLIC6) and rabbit parchorin (rCLIC6), as obtained with ClustalW software and visualized with Macboxshade. Amino acids that are identical in all sequences are printed in white on black; those identical in two or three sequences are in white on gray; similar amino acids are in black on gray.

We defined the gene structure, using "Blast 2 sequences" software for comparing the gene and related genomic fragment sequences. Multiple alignments were performed with Clustal W software (version 1.7; Thompson et al. 1994). We searched for profiles, patterns and motifs using InterPRO (version 3.1, May 2001; Apweiler et al. 2001; http://www.ebi.ac.uk/interpro/scan.html). For secondary structure prediction, the target sequence was analyzed with PredictProtein software (http://dodo.cpmc.columbia.edu/). Signal peptides were searched with PSORT at the web site of the Tokyo University's Human Genome Center (http://psort.ims.u-tokyo.ac.jp/).

Molecular evolution analysis. The Block Maker software identified blocks with the highest local similarity within all pasted sequences. High-quality, neighbor-joining trees were made from the BLOCK alignments by using a routine in ClustalW. Kimura correction for multiple substitution was applied; 100 bootstrap values were calculated. This analysis was performed at the web site http://blocks.fhcrc.org/blocks/.

Results

The ACD gene cluster. The bioinformatic analysis of the *DSCR1*-like family revealed the existence of an undescribed cluster localized on 21q22.1 (in a region of ~534 kb) and triplicated in the human normal genome on 1p35 and 6p12-21 (Fig. 1). The three members of the *DSCR1*-like family (Strippoli et al. 2000a) share the same chromosomal localization as the three members of the *AML* family (Levanon et al. 1994, 2001). On HC 21, *DSCR1* is localized near to *AML1*; on HC 1, *DSCR1L2* and *AML2* delimit a region of ~450 kb; on HC 6, *AML3* and *DSCR1L1* define a region of ~1000 kb. The extension of each cluster was determined by searching for the existence of paralogy for flanking genes.

Following the observation of the presence of a *CLIC* family member within each of the three dusters, we have called the conserved gene cluster *ACD* (i.e., AML/CLIC/DSCR1-like), denominating the three individual clusters *ACD1*, *ACD6*, and *ACD21* according to their respective chromosomal localizations.

Figure 1 reports the known genes contained in each cluster. Note that a *CLIC* family member is always located between a member of *AML* and *DSCR1*-like genes. Differences among the clusters (data not shown) include: 1) lack of conservation of some small genes; 2) differences in exon/intron structure and alternative splicing isoform numbers among conserved genes; 3) dissimilarities in orientation of the three *DSCR1*-like genes.

Identification of human parchorin gene, CLIC6. Two members of the *CLIC* family (reviewed by Debska et al. 2001), namely, *CLIC4* and *CLIC5*, are localized within *ACD1* and *ACD6*, respectively. Sequence similarity analysis (using BLAST) allowed identification of a new gene in the *ACD21* cluster. This new *CLIC* family gene had been partially predicted by Hattori et al. (2000), and the corresponding assigned locus is *CLIC6* at the Human Nomenclature Committee. Our sequence comparison showed similarity between the partially predicted sequence of *CLIC6* and an *Oryctolagus cuniculus* (rabbit) gene coding for the 637-residues parchorin protein (the name derives from its high expression in parietal cells and the choroid plexus; Nishizawa et al. 2000). At its amino terminus (N-end), parchorin has a repeated peptide domain similar to human involucrin, which is not present in other *CLIC* family members. Comparison by using the TBLASTN algorithm of the rabbit parchorin peptide sequence and the HC 21 nucleotide sequence allowed reconstruction of the complete sequence of human *CLIC6*. Partial sequencing demonstrated that *CLIC6* is expressed and subject to splicing. We obtained the predicted 494-bp RT-PCR products from all analyzed human tissue types: brain, pancreas, placenta, and liver (Fig. 2a). The 447-bp *CLIC6* partial sequence obtained (excluding the primers within which it is exactly included) showed no ambiguity; it has been registered in GenBank with accession #AF426169. Sequence comparison between the obtained partial cDNA and the related genomic fragment sequence (#AP001720) allowed formal confirmation of the exon/intron limits for exons 4, 5, and 6, and recognition of the related splicing sites, exhibiting a standard GT/ AG rule. This demonstrates that *CLIC6* is a new functional human gene that is expressed as mRNA in various tissue types and is subject to splicing.

Sequence analysis of the CLIC6 gene. Table 1 reports limits, dimensions, and splicing sites for exons and introns in the human *CLIC6*. The mRNA (assembled starting from the identified exons) has an open reading frame of 2058 bp, encoding for a polypeptide of 686 amino acids. The molecular weight of the predicted protein is ~71 kDa, and the pI value is 4.23, indicating a highly acid nature. This is in agreement with *in vitro* data for rabbit CLIC6 (molecular weight, 64-kDa, pI 4.18—Nishizawa et al. 2000). The predicted human CLIC6 shows 63% sequence identity with rabbit parchorin (Fig. 2b). Its characteristic feature is the N-end domain, which shows similarity with the related rabbit region and with involucrin, a structural protein with an amino acid tandem repeat. The carboxyl terminus (C-end) of the protein presents a high level of similarity with the complete sequences of the other known *CLIC* products. The conserved domain runs from I453 to the C-end. Domains with high similarity also exist within the repeat, providing further confirmation that human CLIC6 is the ortholog of rabbit parchorin.

InterPRO analysis confirmed that CLIC6 is an actual member of the CLIC family, since it conserves the typical INTCLCHANNEL fingerprint composed by five peptide domains (Table 1b). The same search also revealed low similarity among CLIC6 and the C-end domains of both glutathione S-transferase (GST) and 5′-3′ exonuclease. Nevertheless, similarities with GST in small domains of CLIC family members do not seem to correspond with a functional motif. The repeated region of CLIC6 seems to be devoid of peptide domains resembling known functional domains. Secondary structure was predicted by using the PredictProtein algorithm. This showed that the repeated region forms a structure with many highly hydrophilic loops, whereas the conserved region forms four or five hydrophobic α-helixes. PSORT analysis did not reveal any known cell localization signal.

Molecular evolution analysis. We performed a dbEST analysis to find novel *DSCR1*-like and *AML* family genes. We predicted the entire or partial coding region for 25 unknown *DSCR1*-like and *AML* family members (Table 2). We observed that all *AML* family genes are confined to animals, whereas *DSCR1L* genes are also found in various plants, yeast and fungi. Using Block Maker, we constructed a neighbor-joining tree (Saitou and Nei 1987) for each gene family of the *ACD* clusters (Fig. 3). Supplementary data with details for each gene used in this analysis are available at our web site (http://apollo11.isto.unibo.it/suppl/ACD_cluster_data.htm). Only the *CLIC* family has more than three human members. The three *CLIC* genes in the cluster (ie., *CLIC4*, *CLIC5*, and *CLIC6*) show more similarity with one another than with other human members, suggesting that they may have a late common origin compatible with two duplication and divergence rounds. Moreover, the same chronological order of divergence was observed in all three families: the genes localized in ACD21 self-separated from the other two genes localized on HC 1 and

Table 1. a) Genome organization of human *CLIC6*. Positions are referred to HC 21 sequence #AP001720.

Intron	Positions	Size (bp)	Donor site	Acceptor site
1	262711–299226	36516	/gtaaagctcg	ccacttgtag/
2	299337–299890	554	/gtaagacaag	tcccacctag/
3	300017–300646	630	/gtaggcctca	acttttcaag/
4	300754–301303	550	/gtgagtacct	tctgttttag/
5	301486–308267	6782	/gttcatcttc	tcattttcag/
Exon	**Positions**	**Size (bp)**	**Exon start**	**Exon end**
1	<261337–262710	1374		CTTCGTCAAG/
2	299227–299336	110	/GCTGGTTATG	ACCTGAAAAG/
3	299891–300016	126	/GAAACCCGCA	CTCCCCCGAG/
4	300647–300753	107	/GTATCCCAAG	GCAAATGAGA/
5	301304–301485	182	/TTCATGAAAA	TATTATTAAG/
6	308268–308511	244	/ATTGTGGCCA	

Table 1. b) Peptide domain of human *CLIC6*. Data by InterPRO analysis.

Start aa	Stop aa	Size (aa)	INTCLCHANNEL
477	493	16	Motif 1
519	531	12	Motif 2
556	566	10	Motif 3
608	617	9	Motif 4
618	630	12	Motif 5

HC 6 in a more ancient era, whereas a common progenitor seems to have generated ACD1 and ACD6 clusters by divergence.

Discussion

Few reports exist on cases of large-scale segmental paralogy in the human genome. We present strong evidence of a further instance in the form of a ~500-kb segment of HC 21 that is triplicated on HC 1 and 6. The available gene maps suggested the presence of two members of the *CLIC* family on HC 1 and HC 6, both coding for proteins responsible for chloride transport within intracellular membranes. Hattori et al. (2000) noted the presence of a small *CLIC*-like sequence exactly between *AML1* and *DSCR1L1*. By means of spliced cDNA fragment cloning, we have formally identified *CLIC6* on HC 21.

The finding of *CLIC6* reveals a new gene association group triplicated in the human genome. The three segments are extended for ~500 kb, and each includes three functional members of three different gene families: *AML*, *CLIC*, and *DSCR1*-like. Hence, we propose the name ACD cluster for the gene association group. The orientation of the paralogous genes with respect to the direction of transcription is also similar, except for DSCR1. This is probably owing to an inversion event, which is consistent with the observation that local inversions are more common than large-scale duplications (Venter et al. 2001). The ACD cluster is a paralogous segment shared by three human chromosomes. It overlies about 1.4% of HC 21 and thus significantly enhances the fraction of interchromosomal duplications previously described for HC 21 (estimated of 1.6%; IHGSC 2001). Ac-

Table 2. ESTs of unknown members of *AML* and DSCR1-like families. ESTs providing information about the existence of the listed genes in various organisms.

Gene	Organism	ESTs Accession numbers	Prediction
AML1	*Rattus norvegicus*	BG016110	partial
AML2	*Rattus norvegicus*	BF521704	partial
DSCR1L	*Aspergillus nidulans*	AI213143, AA965762, AI213418, AA966905, AI211115, AA786787, AA786934, AA784828	partial
DSCR1L	*Blumeria graminis*	AW792760	partial
DSCRIL	*Dictyostelium discoideum*	C92200, AU052410	partial
DSCR1L	*Hordeum vulgare*	BE413499, AL503715	partial
DSCR1L	*Paralichthys olivaceus*	AU090836	partial
DSCR1L	*Pleurotus ostreatus*	AT003942	partial
DSCR1L	*Solanum tuberosum*	BF052789	partial
DSCR1L	*Strongyloides stercoralis*	BG227181	partial
DSCR1	*Bos taurus*	BF041330, BE485918, AV604722	partial
DSCR1	*Danio rerio*	AW233012, AI601661, BG307074	gapped (not overlapping ESTs)
DSCR1	*Oryzias lutipes*	AV670322, AU180786	gapped (not overlapping ESTs)
DSCR1	*Rattus norvegicus*	BE127094	partial
DSCR1	*Sus scrofa*	BF711037	partial
DSCR1	*Xenopus laevis*	BG515763, BG264239, BF613411	complete
DSCR1LI	*Bos taurus*	BF890409; BE723315, BF046352, BF042939,BF706654	partial
DSCR1L1	*Danio rerio*	BG307867, BF937618	partial
DSCR1L1	*Gallus gallus*	AL587640	partial
DSCR1L1	*Oryzias latipes*	AU169846	complete
DSCRILI	*Rattus norvegicus*	BF521938, AI501641, AI501566, AI138048, AI548610, AI113336, AI500795	partial
DSCR1L1	*Sus scrofa*	BF443461	partial
DSCR1L1	*Xenopus laevis*	AW643175, BG162439	partial
DSCR1L2	*Sus scrofa*	BF442222	partial
DSCR1L2	*Xenopus laevis*	AW644064	partial

cording to the HC 21 coordinate system (Hattori et al. 2000), the cluster lies between 21.4 Mb and 21.9 Mb, in the central part of HC 21, far from the centromeric and subtelomeric regions where most duplications have been observed. Although the available data on the mouse genome sequence still contains large gaps, preliminary analysis shows a colocalization of the *AML* and *DSCR1*-like family orthologous genes (Strippoli et al. 2000b) on murine Chromosomes 4, 17, and 16, in regions already identified as being related to the HC 1, HC 6, and HC 21 fragments, respectively. One of the families (*DSCR1*-like) is conserved from yeast to humans.

Clues regarding the evolutionary mechanisms that generated and maintained the linkage among the *ACD* cluster genes remain to be determined. *AML* genes codify for transcription factors playing a role in development and oncogenesis, *CLIC* genes code for ionic intracellular channels, and *DSCR1*-like proteins are a group of calcineurin inhibitors (Fuentes et al. 2000; Kingsbury and Cunningham 2000). Interestingly, most of the genes (including *DSCR1*, *AML1*, *AML2*, *AML3*, *CLIC4*, and *CLIC5*) are expressed during development. It cannot be excluded that conservation of the *ACD* cluster has played a role in an advanced evolutionary step, since it is absent in the genomes of lower species. However, the cluster might have been too late an event for disruption to occur. The maintenance of linkage may also have been due to coordinated regulation of expression. Functional studies are needed to elucidate this question.

We have analyzed the cladistic structure of all the known members of the three conserved gene families. The three cladistic trees overlap consistently with an early disjunction of a cluster of HC 21 genes, followed by the formation of new clusters on HC 1 and HC 6 (Fig. 3). Hitherto, our knowledge of rearrangements on HC 21 was restricted to the human-primate divergence era as a result of human-ape karyotype comparison (Richard and Dutrillaux 1998). Analysis of the EST in Table 2 and of the known genes suggests that an ancestral yeast gene of the *DSCR1*-like family was the first to appear. The known functions of AML1 in hemopoiesis and of AML3 in osteogenesis support the notion that the ancestral AML family appeared later, probably with the appearance of animals. At about this time, the two family progenitors presumably linked and arranged themselves around an ancestral *CLIC* family member. In the evolutionary steps from lower (*D. melanogaster*, *S. purpuratus*) to higher species (*X. laevis*, *O. latipes*), the *AML* and *DSCR1*-like families both came to have three members.

These observations reinforce the notion that the entire ACD cluster shares a common evolutionary origin. Following the recent challenges to the whole-genome double replication theory (Makalowski 2001), the most plausible explanation for the formation of the ACD cluster might be successive segmental duplications during two rounds of large-scale chromosomal rearrangements dated by Wang and Gu (2000) to 525 and 435 million years ago (evolutionarily just before the divergence between Tetrapods and Teleostei). The ACD cluster provides a major new model for gene evolution studies, in that divergence can be simultaneously observed in three different gene families whose members all remain functional while undergoing intra-familial functional differentiation.

A complete list of paralogous regions would provide essential knowledge. As well as allowing identification of other new genes, it would shed light on the history and mechanisms that shaped evolution. It should also help to explain phenotype diversification, both during evolution and in oncological patients with altered karyotypes. The present report underlines the need for more effective tools to detect interchromosomal similarity in the human genome. The two recognized auto-

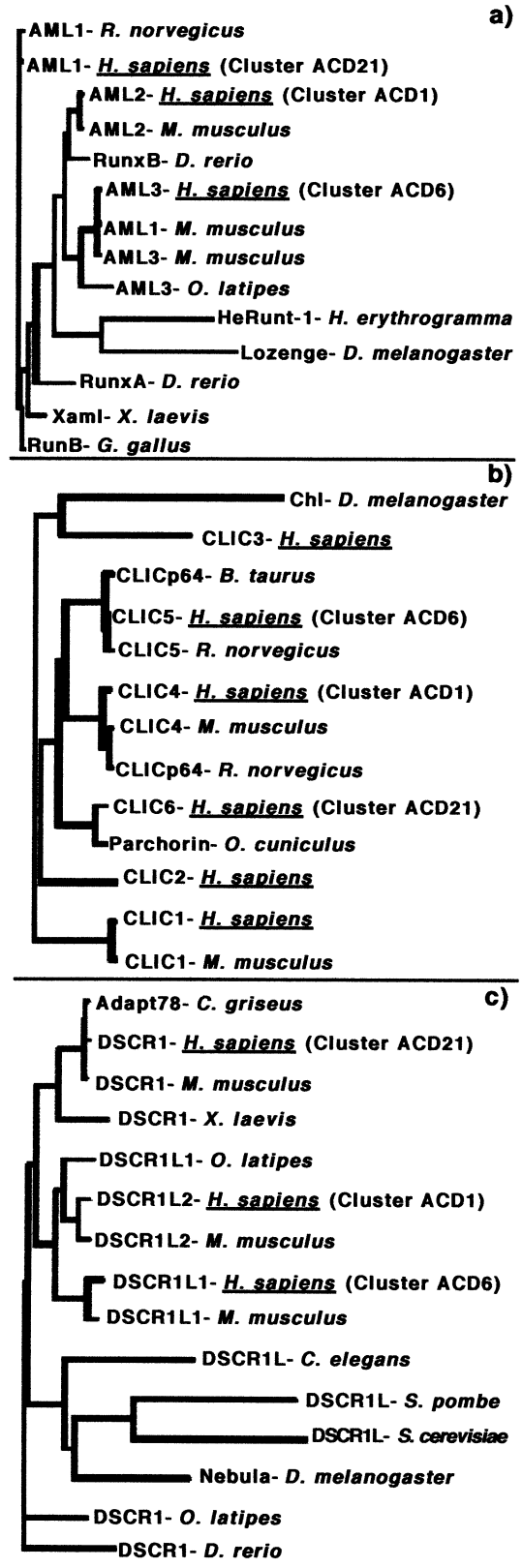

Fig. 3. Phylogenetic trees of AML **(A)**, CLIC **(B)**, and DSCR1-like **(C)** protein families, as obtained by Block Maker analysis. We used the following amino acid sequences as the input: human CLIC1, CLIC2, CLIC3, CLIC4, CLIC5, and CLIC6; murine CLIC1 and CLIC4; rabbit parchorin; CLIC5 and CLICp64 of *Rattus norvegicus*; bovine CLICp64; CG10997 (Chl) of *Drosophila melanogoster*.

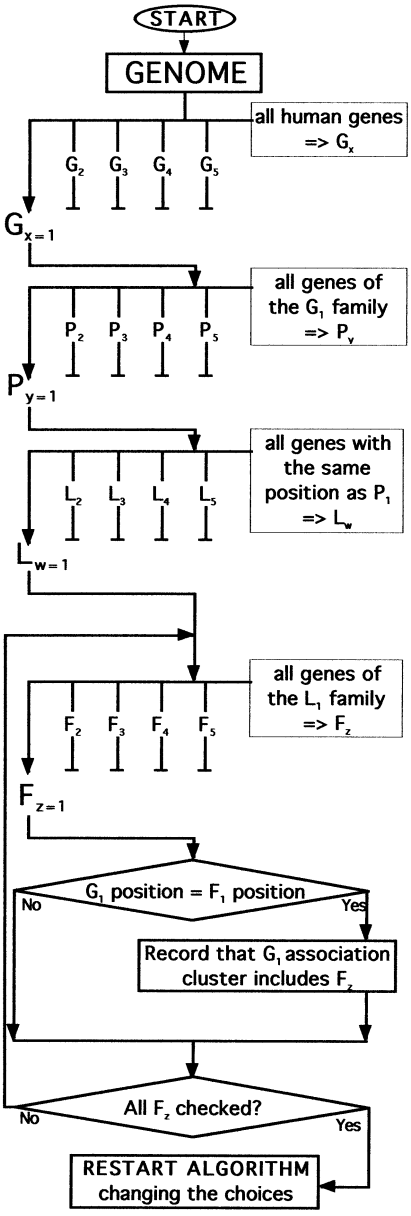

Fig. 4. Flowchart showing the six steps of the new algorithm for identification of paralogous regions: 1) start from a gene (G_x); 2) search all members of its gene family; 3) choose a gene (P_y) among these and search all genes that show similar location; 4) choose a gene (L_w) among these and search all members of its gene family; 5) select genes (F_z) showing the same location as G_x and record that they are included in the G_x association cluster (which is an indication of segmental paralogy); 6) restart algorithm, changing one or more choices. This algorithm can be made more specific by changing the "location range" (the location may be the arm, the band, or a range in centimorgans around the gene position).

mated methods—based on whole-genome sequence, identifying segments in the range of 90–99.5% identity (IHGSC 2001), and detection of all sets of three or more protein matches occurring in close proximity on two different chromosomes (Venter et al. 2001)—both failed to detect the ACD cluster. These two algorithms favor specificity over sensitivity and can only reveal duplications.

We propose an algorithm based on simple iterative recording of the presence of at least one member of two or more different gene families in at least two genome locations (see Fig. 4). This system has several advantages: the cut-off may be modified, simply indicating the size of the area that is considered a "location" (for example, a cytogenetic band location or a physical distance on the two sides of the reference gene); the system does not suffer from limitations due to low grade of sequence similarity or to the rigidity in the definition of the relative distance of the associated genes; the iterations may detect any number of replications (duplications, triplications, and so on). Preliminary manual simulation shows that, with the cytogenetic band as "location", this algorithm detects the triplication described here, as well as other known paralogy regions. Lack of an absolutely complete gene catalog, with integrated family classification and position coordinates, currently hinders automation. However, this problem should disappear once the human genome draft has been converted into an entirely finished and annotated sequence.

Acknowledgments. We are grateful to Gabriella Mattei for her excellent technical assistance with automatic sequencing, and to Robin M.T. Cooke for editing. This work was conducted in the Research Center for Molecular Genetics "Fondazione CARISBO" at the Institute of Histology and General Embryology, University of Bologna. The work has been funded by University of Bologna "Progetto giovani Ricercatori" grant to P. Strippoli and MURST (ex 60%) grants to P. Carinci.

References

Apweiler R, Attwood TK, Bairoch A, Bateman A, Birney E, et al. (2001) The InterPro database, an integrated documentation resource for protein families, domains and functional sites. Nucleic Acids Res 29, 37–40

Debska G, Kicinska A, Skalska J, Szewczyk A (2001) Intracellular potassium and chloride channels: an update. Acta Biochim Pol 48, 137–144

Fuentes JJ, Genesca L, Kingsbury TJ, Cunningham KW, Perez-Riba M, et al. (2000) DSCR1, overexpressed in Down syndrome, is an inhibitor of calcineurin-mediated signaling pathways. Hum Mol Genet 9, 1681–1690

Hattori M, Fujiyama A, Taylor TD, Watanabe H, Yada T, et al. (2000) The DNA sequence of human chromosome 21. Nature 405, 311–319

IHGSC (International Human Genome Sequencing Consortium) (2001) Initial sequencing and analysis of the human genome. Nature 409, 860–921

Kingsbury TJ, Cunningham KW (2000) A conserved family of calcineurin regulators. Genes Dev 14, 1595–1604

Levanon D, Negreanu V, Bernstein Y, Bar-Am I, Avivi L, Groner Y (1994) AML1, AML2, and AML3, the human members of the runt domain gene-family: cDNA structure, expression, and chromosomal localization. Genomics 23, 425–432

Levanon D, Glusman G, Bangsow T, Ben-Asher E, Male DL et al. (2001) Architecture and anatomy of the genomic locus encoding the human leukemia-associated transcription factor RUNX1/AML1. Gene 262, 23–33

Makalowski W (2001) Are we polyploids? A brief history of one hypothesis. Genome Res 11, 667–670

Nishizawa T, Nagao T, Iwatsubo T, Forte JG, Urushidani T (2000) Molecular cloning and characterization of a novel chloride intracellular channel-related protein, parchorin, expressed in water-secreting cells. J Biol Chem 275, 11164–11173

Richard F, Dutrillaux B (1998) Origin of human chromosome 21 and its consequences: a 50-million-year-old story. Chromosome Res 6, 263–268

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4, 406–425

Strachan T, Read AP (1999) Our place in the tree of life. In: Human Molecular Genetics, Oxford, UK: Bio Scientific Publishers, pp 315–350

Strippoli P, Lenzi L, Petrini M, Carinci P, Zannotti M (2000a) A new gene family including DSCR1 (Down Syndrome Candidate Region 1) and ZAKI-4: characterization from yeast to human and identification of DSCR1-like 2, a novel human member (DSCR1L2). Genomics 64, 252–263

Strippoli P, Petrini M, Lenzi L, Carinci P, Zannotti M (2000b) The murine DSCR1-like (Down syndrome candidate region 1) gene family: conserved synteny with the human orthologous genes. Gene 257, 223–232

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22, 4673–4680

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, et al. (2001) The sequence of the human genome. Science 291, 1304–1351

Wang Y, Gu X (2000) Evolutionary patterns of gene families generated in the early stage of vertebrates. J Mol Evol 51, 88–96