Late-onset and typical Huntington disease families from Crete have distinct genetic origins

ELEONORA KARTSAKI[1*], CLEANTHE SPANAKI[1*], MINAS TZAGOURNISSAKIS[1], APHRODITE PETSAKOU[1], NICHOLAS MOSCHONAS[2], MARCY MacDONALD[3] and ANDREAS PLAITAKIS[1]

Departments of [1]Neurology, and [2]Biology, University of Crete, Voutes Place, 71500 Heraklion, Crete, Greece; [3]Harvard Medical School, Molecular Neurogenetics Unit, Massachusetts General Hospital, Boston, MA, USA

Received November 2, 2005; Accepted November 22, 2005

Abstract. HD families in which late-onset occurs consistently in affected members are rare. The objectives of this work was to study such late-onset HD families encountered on Crete, and to trace their genetic origin. Nine late-onset HD kindreds (61 affected members) were studied along with two typical HD families (17 affected members). We genotyped 33 late-onset Cretan HD chromosomes, 9 Cretan typical HD chromosomes and 114 Cretan control chromosomes using 14 STR markers and 20 SNPs that map to 4p16.3. In contrast to the typical HD pedigrees, the late-onset HD families lacked anticipation and juvenile cases. The expanded CAG repeat (36-42 units) in these families remained either stable or it showed small increment instability, even when transmitted through the father. All late-onset HD chromosomes shared a conserved haplotype defined by the markers D4S95: 1090, D4S127: 157, rs362277: A, rs3025814: G, rs2530596: A that span a 0.277-Mb segment on 4p16.3. Coalescence analysis traced this haplotype to a founder who lived about 1000 years ago. In contrast, each of the two typical HD disease pedigrees derived from a different founder. Sequencing of a 5-kb DNA segment immediately upstream of the HD gene revealed a novel single nucleotide polymorphism at -1757 bp relative to the translation start site, which was more prevalent in Cretan than in North American chromosomes. All late-onset HD families on Crete arose from a common founder with the disease's mutation evolving over the centuries via small-increment instability. These findings suggest that cis-acting factors may be operational.

Introduction

Huntington's disease (HD) is an autosomal dominant neurodegenerative disorder with a worldwide prevalence of about 50-100 cases per million. The majority of HD patients experience their first symptoms between 30 and 50 years of age (mean = 38-40 years), but age at onset can vary among members of the same family with anticipation (onset of the disease at a younger age in children of affected parents) occurring predominantly on paternal transmissions (1,2). About 5-10% of HD cases are juvenile. These often show a distinct phenotype and a more severe clinical course than that of patients who experience disease onset in their mid-adult life. Nearly 25% of HD cases start after 50 years of age and only 3-5% after 60 years (3-7).

The disorder is caused by an expanded CAG repeat (>35 repeats) in the HD gene (within 4p16.3 region) that is unstable and tends to increase in size during meiotic transmission. The CAG repeat lies in the 5' end of the open reading frame (first exon) of the HD gene, which is known to have 67 exons and 66 introns (4-8). Observations on new (*de novo*) HD cases revealed that the disease's allele has arisen from abrupt expansion of parental mutable allele containing 27-35 CAG repeats (intermediate alleles) (8-10). However, it is presently unclear whether this constitutes the main process by which the HD mutation originated in most of existing HD families.

Previous observations on various populations have established that age at disease onset correlates inversely with the size of the CAG repeat (5-7), but this correlation is weak or absent for cases that start after 50 years of age (3). More recent studies have estimated that the size of the expanded and normal CAG repeats and their interaction can account for 68% of the variance of age at onset seen in HD (11). There are also indications that the unexplained variation in the age at onset is heritable and may relate to polymorphism of other genes (12).

We have previously reported (13) a form of HD occurring on the island of Crete that is characterized by disease onset 15-20 years later than that for typical HD. In these families, onset of disease in late life is explained by the expansion of the CAG repeat in the low pathological range (13). In addition, the rather consistent occurrence of late-onset cases in these families is explained by the observed intergenerational stability of the CAG repeat.

Here we studied further HD among natives of Crete and have identified additional late-onset HD kindreds as well as families affected by typical HD (mid-life onset and unstable expansion of the CAG repeat). The availability of a substantial number of families with late-onset and typical HD disease from a relatively small pool of people sharing a common

Correspondence to: Dr Andreas Plaitakis, Department of Neurology, University of Crete, School of Health Sciences, Section of Medicine, 71500 Heraklion, Crete, Greece
E-mail: plaitak@med.uoc.gr

*Contributed equally

Key words: Huntington, instability, late-onset, cis-acting factors

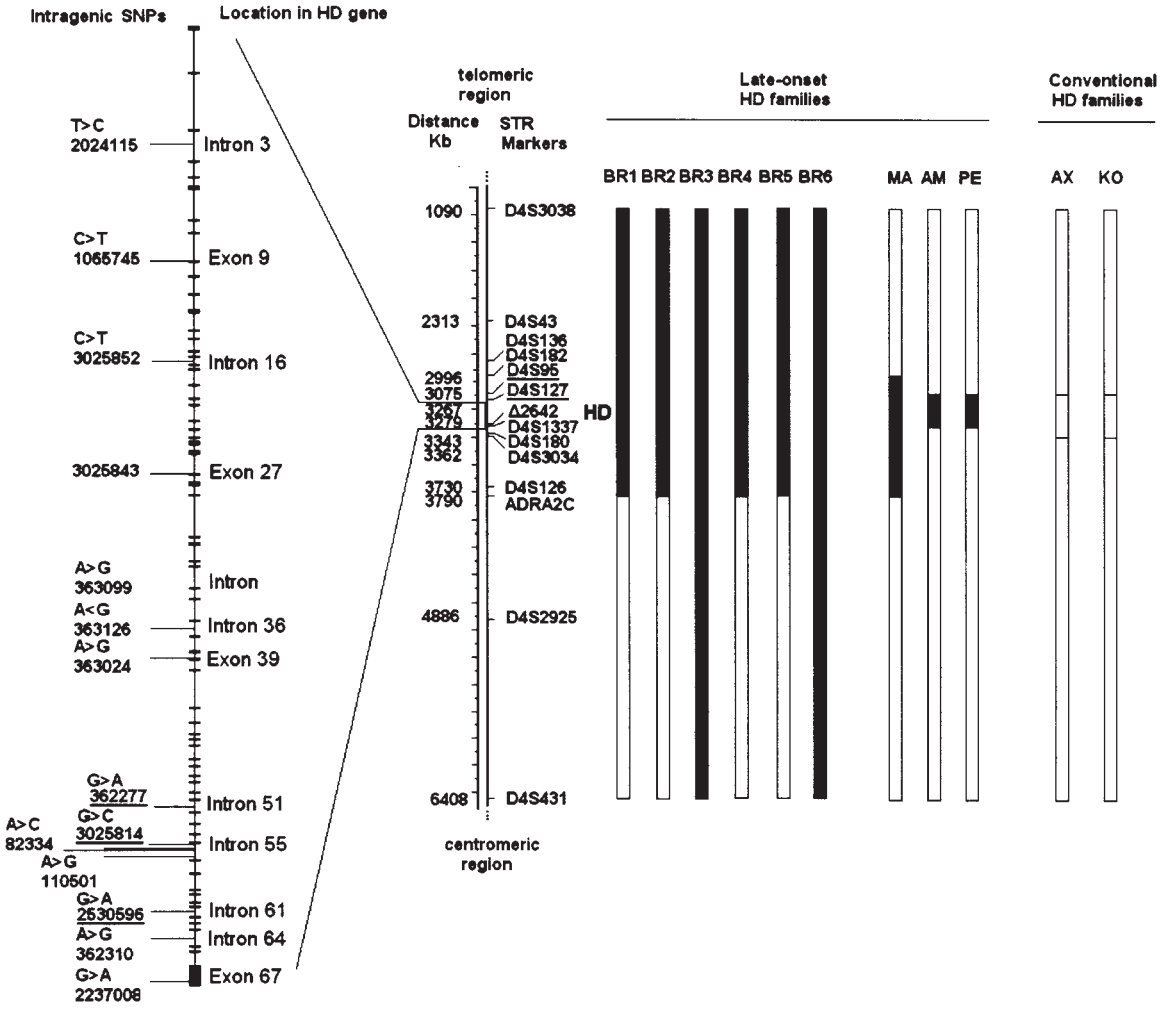

Figure 1. Physical map of the HD gene region. The figure shows the 4p16.3 region of the human chromosome 4 as a long vertical bar, not in scale. The 14 microsatellite markers used in our analysis are placed along the chromosome with kb positions. To the left of this 4p16.3 region, a vertical bar shows the location inside of the HD gene of the 15 SNPs studied. To the right of this 4p16.3 region, vertical bars indicate areas of the HD chromosomes shared by the late-onset HD families (enclosed in a black box); those shared by the conventional form of HD are enclosed in a white box. As shown here, all late-onset HD families that originated from a single village (BR1-BR6) display an extensive common haplotype flanked by polymorphic markers D4S3038 and ADRA2C, which are 2.7 Mb apart. Those from other villages (AM, MA and PE) share a narrower DNA segment that contains the STR markers D4S95, D4S127 and the intragenic SNPs rs362277, rs3025814, rs2530596 and that is flanked by the STR markers D4S95 and D4S3034, which are 0.336 Mb apart. D4S136 and D4S182 are located in 4p16.3 region, but their precise position remains uncertain (National Center for Biotechnology Information).

genetic background and the same environment, offered us the opportunity to trace the origin of the disease in this population. Results are described below.

Materials and methods

Clinical studies. Family histories were collected for individuals with dominantly inherited choreic disease and genealogical trees were constructed. Affected family members and unaffected relatives at risk were personally evaluated by us. In addition to the traditional neurologic examination, detailed evaluations of the eye movements and the mental status of the patients were done. The latter included the quantitative Mini-Mental State test. In several patients, brain CT scans or MRIs were obtained. The clinical diagnosis of HD was confirmed by two examiners (A.P. and M.T.). To establish age at disease onset with a reasonable certainty, multiple interviews were held with all available first-degree relatives (spouse, children and siblings), preferably those who lived with the patient at the time. Onset of the disease was generally considered to be the time at which the characteristic choreic movements first became noticeable. As with other HD studies, difficulties determining age at disease onset were encountered in a few cases in which the presenting symptom was personality changes.

Determination of the (CAG)n repeat size, genotyping and haplotyping. DNA was isolated from peripheral blood leukocytes by standard techniques. Chromosomes were screened for the HD mutation by PCR amplification as described by Riess *et al* (14). We genotyped 33 individuals from the late-onset families with a HD chromosome and 29 of their relatives at risk without the HD mutation using the polymorphic markers described below. In addition, we genotyped nine individuals with a HD chromosome, who were members of two conventional HD kindreds from Crete. To identify suitable markers for this analysis, we searched the available databases (National Center for Biotechnology Information,

Table I. Sequence of oligonucleotides used for STS analysis.

Marker	Primer		Anneal. temp. (˚C)	Product size (bp)
	Forward (5'-3')	Reverse (5'-3')		
D4S3038	GAAGACCAGCATTCGG	GGTTTAATACACAGTAATTGTTCA	55	209-227
D4S43	CTTCCTTTTCTCTCTGGATGC	ACATCACGCTTATCTTTGGGG	57	176-190
D4S136	CTGACTTGATCCAATCCAAAGGAAAG	TTGAACCTAGTAGGCGGAAGTTGCAC	60	225-249
D4S182	GGATCCAATCCAAAGGAAAGTTCC	TTTTTCTCCCCCCATGACACCATG	58	114-138
D4S95	GCATAAAATGGGGATAACAGTAC	GACATTGCTTTATAGCTGTGCCTCAGTTT	60	990-1600
D4S127	CCTCTGTTTGCAATCCATTT	GTCCCTTGCATGCCCTGGCT	55	145-161
Δ2642	GCTGGGGAACAGCATCACACCC	CCTGGAGTTGACTGGAGACGTG	58	109-112
D4S1337	ATGTGGCAGAAGTGCTCTTTGTG	AACAACCAGCAGGTGACTATCAG	62	156[a]
D4S180	CACATCTTCCTGTTTCTTTGAACATC	GAGAGACCCCAGAGTCCAGCAG	62	212[a]
D4S3034	CTGCCAATAAACTGGGT	TTGCTCACCAAAGAGGTT	55	182-196
D4S126	GGATCCTGTCACTGTACTCCAGCC	TGCTTAACCAGTTTGACCATGAGG	57	145-177
ADRA2C	CGCTGCCTCCCTTCCACCTGTTG	AGTGGGCAGGGCGGGGCAGGT	62	179-193
D4S2925	TCAGAAACCCCTACAGGAAA	TTTGATGAGTTATTCGGAGG	55	151-155[a]
D4S431	AGGCATACTAGGCCGTATT	TTCCCATCAGCGTCTTC	55	244-290

[a]STS markers not polymorphic in our control and patients samples.

Table II. Sequence of oligonucleotides used for SNP analysis.

Intragenic SNPs	Primer		Enzymes
	Forward (5'-3')	Reverse (5'-3')	
rs2024115	AAGCCCCTGGAGTAAATCTT	TGGAAGAGAAGGCTCCTATG	BsrI
rs1065745	TCATACACAGCACCAAGACC	AAGGAAGATGGAATGCAGAC	Bs1I
rs3025852	TGAAACCCATAGAGGAGAAAA	CCTCTCACCCAAAGAAAAAG	Tsp45I
rs3025843	TGAGTGACATGGGTTAGCTC	AATATCTTGAACCGCACCTG	BfaI
rs363099	CTAGATGAACTCAGCCCAGA	ATTTTTCCAGTGCTGACAAC	BsrBI
rs363126	AGCAGCACTTTCTTTGTTCA	CTGCAGAGGCAATCTACTGA	HphI
rs363124	CTGCAGGGAGAAGACTTTTA	TACCTGGTCTTGCTTCACTT	Tsp45I
rs362277	TATGCCCAGTGTAGAAAGCA	GACACCATTCTGTGGATCTT	AvrII
rs3025814	CTGCTCTTGTTGACATGTGG	CACCTGCTATGTCCCAGAG	BsrI
rs82334	CTGAAAGCTCTCGACACCA	GGAACCTGGTGTAATACCAAA	BsajI
rs82332	CACGTGTCTCTGGGACATAG	ACTAGTTGCCTGCTCATTCC	BSERI
rs110501	CCTCCCAAACCTACATTTTA	CGTCAGCCTTCATTGCTAGA	Tsp45I
rs2530596	CTCATTCAGTCCCTGGCTAC	CTAACTCAGCCTCCCCTGT	NspI
rs362310	GGAAAGTCAGAGCCTAAGCA	CACATAGGGATGTTTGTGGA	HhaI
rs2237008	ACTGAAGCGTGTTTCTTTCC	CTCATCTCAGCTCCAAGAGG	HhaI

and The Genome Database). Fourteen STR markers (D4S3038, D4S43, D4S136, D4S182, D4S95, D4S127, Δ2642, D4S1337, D4S180, D4S3034, D4S126, ADRA2C, D4S2925, D4S431) were identified mapping to a 5.5-Mb region on 4p16.3 that contains the HD mutation (Fig. 1). In addition 20 intragenic SNPs identified through the National Center for Biotechnology Information, the UCSC Genome Bioinformatics and the Ensemble Genome Browser, were also used in the genotypic analysis of our patient and control samples. Nine SNPs proved not to be polymorphic in our control and patients sample and were not pursued any further. The map order and distances between the markers are shown in Fig. 1. For genotyping we used PCR primers and conditions as described in Tables I-III. For markers D4S43, D4S136, D4S182, D4S95, D4S127,

Table III. Variation of intragenic SNPs in HD patients and their frequencies in Cretan controls.

SNP	Position in HD gene	Location in HD gene (kb)	Change	Late-onset HD	Conventional HD	Frequencies (%) in Cretan controls
rs2024115	Intron 3	3141	C/T	T	T	T: 75.9 C: 24.1
rs1065745	Exon 9	3159	C/T	C	C	C: 97.8 T: 2.2
rs363099	Intron n.p.		A/G	A	A	A: 83.6 G: 16.4
rs363124	Exon 39	3226	A/G	A	A	A: 87.3 G: 12.7
rs362277	Intron 51	3255	G/A	**A**	**G**	A: 39.2 G: 60.8
rs3025814	Intron 55	3261	C/G	**G**	**C**	G: 64.8 C: 35.2
rs82334	Intron 55	3262	A/C	A	A	A: 88.3 C: 11.7
rs82332	Intron 55	3262	A/G	A	A	A: 89.3 G: 10.7
rs110501	Intron 55	3262	A/G	A	A	A: 72.8 G: 27.2
rs2530596	Intron 61	3273	A/G	**A**	**G**	A: 46.5 G: 53.5
rs362310	Intron 64	3276	G/A	G	G	G: 13.5 A: 86.5
rs2237008	Exon 67	3281	G/A	G	G	G: 97 A: 3

n.p., not placed according to Sanger Institute - Ensembl Genome Browser database. Nine additional SNPs mapping to these regions were also studied. These did not prove to be polymorphic in our control and patient sample and were not pursued any further.

Δ2642, D4S180, D4S126 and ADRA2C, the PCR reaction was performed in 25 μl using 20 ng DNA, 10 pmoles of each primer (and one end-labeled primer with γ-[^{32}P]ATP used at 150000 cpm/reaction), 250 μM of each dNTPs, 1.5 mM $MgCl_2$, 1X AmpliTaq buffer and 1.25 units of Taq DNA polymerase (AmpliTaq-Cetus). For markers D4S3038, D4S3034, D4S1337, D4S2925 and D4S431, the PCR reaction was performed in 10 μl using 20 ng DNA, 1 pmole of each primer (and one 5' end-labelled primer with fluorescent dye IRD-800), 200 μM of each dNTPs, 1.5 mM $MgCl_2$, 1X AmpliTaq buffer and 1 unit of Taq DNA polymerase (AmpliTaq-Cetus). The radiolabeled PCR products were separated on 6% acrylamide sequencing gels, except for the D4S95 marker, which was analysed on 1.5% agarose gel. The IRD-800 labelled PCR products were separated on LI-COR 4200 system (LI-COR, Lincoln, NE). Proper Mendelian inheritance of each marker was manually confirmed. STR markers D4S180, D4S2925 and D4S1337 were not polymorphic in our control and patient DNA sample and were excluded from the construction of the final haplotype.

For SNP genotyping, the PCR reaction was performed in a volume of 20 μl using 60 ng DNA, 10 pmoles of each primer, 200 μM of each dNTPs, 2-2.5 mM $MgCl_2$, 1X AmpliTaq buffer and 0.5 unit of Taq DNA polymerase (AmpliTaq-Cetus). PCR conditions included an initial denaturation of 5 min at 94˚C, followed by 30-35 cycles of 1 min at 94˚C, 45 sec at 55˚C and 1.5 min at 72˚C and a final step of 10 min at 72˚C. PCR products of 200-400 bp in size, including the SNP under study, were digested with the appropriate enzyme supplied by New England Biolabs. Restriction digestions were performed in a 15-μl reaction volume with the use of 10 μl PCR product. Digested PCR products were electrophoresed on 1.5% agarose gel. The position of SNPs within the HD gene, the observed variance, primers and enzymes used in these analyses are shown in Tables II and III.

Allelic association and linkage disequilibrium analysis. Thirty-three late-onset HD chromosomes (n_D) and 114 normal chromosomes (n_N) were studied. Allele frequencies for the STR markers D4S3038, D4S43, D4S136, D4S182, D4S126, D4S3034 and D4S431, which flank the conserved haplotype of late-onset HD, were obtained directly by allele counting for both HD and control chromosomes. Variances of the frequencies were obtained with the formula: $p(1-p)/n$, where p is the observed frequencies and n is the number of chromosomes studied. Statistical comparison of allele frequencies between disease and control chromosomes was based on χ^2 and on two-tailed Fisher's exact test as appropriate. The alleles were classified into two groups, one for the associated allele and another for all other alleles combined. Since multiple alleles were tested at each marker locus, a statistical correlation for multiple tests was employed. A Bonferroni correction was applied by multiplication of the nominal significance level (p) by the number of alleles tested. Differences were considered significant when the aforementioned product was $p<0.05$.

The degree of LD, measured by δ, between each marker locus and the late-onset HD locus was assessed based on the formula: $\delta=(p_D-p_N)/(1-p_N)$, according to the method of Bengtsson and Thomson (15), where p_D is the frequency of the associated allele on disease chromosomes and p_N the frequency of the associated allele on normal chromosomes. We assume that any differences in δ represent effects of recombination, although mutation at the marker can also occur (16).

Recombination fraction (θ). The recombination fraction (θ) between each of the associated alleles at the marker loci studied and the disease locus was estimated using the Mapfun Program of the Fastlink Software Package that converts genetic distances to recombination fractions according to Haldane

Table IV. Sequence of oligonucleotides used for amplification and sequencing of the 5 kb DNA segment upstream of the HD gene.

Marker	Primer		Anneal. temp. (˚C)	Product	
	Forward (5'-3')	Reverse (5'-3')		Size (bp)	Position (+1 translation start)
HD0	CCGCCTCGCCGCGACACT	CTGGAAGGACTTGAGGGACTCGA	61	452	+54 to -398
HD1	TGCTTCTCGCTGCACTAATC	GTTCTGCCTCACACAGCAAG	56	968	-297 to -1264
HD2	GCCATCTTATGAGTCTGCCCACTG AGGTAACTAACAAC	GCCCCCCATCGG TCCCAGCG	71→65	1142	-1151 to -2292
HD3	GGAACAAATCACCCC AGCAC	AAAATTAAGTTCCAGCGAGGTG	58→56	944	-1883 to -2826
HD4	ATGGGTTG GAATAATTTGGTTTG	GCGAGAAGTGGAGCTGGTTAAG	56	995	-2683 to -3677
HD5	GAGAGGGTTTCATCTTGTTGGTC	ACTGTGTGCAGTTGTCATTCCAG	58	805	-3546 to -4350
HD6	CCCTAGAACCTAAGGAAACAGG	CACCATGTAGACCCTGATAACG	58	1000	-4168 to -5167
HD7	TAAAACACATGGCAATAATACCC	AAAAGAAGGAAACCCTGACG	51	975	-4874 to -5848

Map Function. The Genetic distances for the polymorphic markers analysed were estimated on the basis of their known physical distances (available from National Center for Biotechnology Information) using the conversion factor 1.6 cM/1 Mb. This factor was obtained on the basis of six genetic markers (D4S43, D4S95, D4S127, D4S180, D4S125 and D4S126 spanning the region of interest), for which both genetic distances compilied by Lonjou *et al* (17) and physical distances are known. For each marker, we calculated a cM:Mb ratio and the average value (1.6:1) was used to convert physical distances (in Mb) to genetic distances (in cM). The segment of the HD chromosome containing the conserved haplotype (D4S95: 2996 kb to SNP rs2530596: 3273 kb, approximately, from the telomeric end of 4p arm) of late-onset HD chromosome was considered a single locus bearing the disease mutation. The middle of this segment (3135 kb) served as the reference point for estimating the physical distance between each marker and the disease locus.

Age estimation. To estimate the number of generations (G) to the most recent common ancestor of the late-onset HD, we used the extent of linkage disequilibrium between flanking polymorphic marker loci and the disease locus in order to study the decay of LD over time. This method, originally described by Risch *et al* (18), employs the formula: $G=\ln\delta/\ln(1-\theta)$, where θ is the recombination fraction between disease locus and the marker analyzed, and δ is the degree of linkage disequilibrium (LD) at loci flanking the conserved core markers. As described above, δ was assessed based on the formula: $\delta=(p_D-p_N)/(1-p_N)$, where p_D is the frequency of the associated allele on disease chromosome and p_N the frequency of the same allele on normal chromosomes. A confidence interval for δ was calculated under the assumption of independence for the sample chromosomes. Since $\delta=1-(1-p_D)/(1-p_N)$ is the function of the ratio of two independent random variables, the variance of δ was approximated by the formula:

$$\mathrm{var}(\delta) = \frac{1}{(1-p_N)^4}\sigma^2_{pN}\left[\sigma^2_{pD} + (1-p_D)^2\right] + \frac{\sigma^2_{pD}}{(1-p_N)^2}$$

where $\sigma^2_{pD} = \dfrac{[p_D(1-p_D)]}{n_D}$ and $\sigma^2_{pN} = \dfrac{[p_N(1-p_N)]}{n_N}$

A 90% confidence interval for δ was obtained as $\delta\pm1.64\sigma_\delta$, where $\sigma_\delta{}^2=\mathrm{Var}(\delta)$. A 90% confidence interval for G was calculated by the use of the formula given above for the confidence interval for δ.

Analysis of a 5' end segment of the HD gene and upstream sequences. To see whether changes in the CAG repeat were present in the late-onset HD cases a 696-bp segment of the 5' end of the HD gene containing the CAG repeat was amplified using the PCR primers 5'-CCAAACTCACGGTCGGTGCA GCGGCTCCTCAG-3' (antisense) and 5'-ACACTTCACAC ACAGCTTCG-3' (sense). Amplified PCR products were subcloned in a TA plasmid vector (Invitrogen TA cloning kits) and used for sequencing using standard methods. In addition, to search for possible cis-acting elements that could influence CAG stability we sequenced a 5.5-kb DNA segment on 4p16.3 that extends immediately upstream from the translation start site (+1) of the HD gene. Sixteen primers were used to amplify eight consecutive overlapping DNA segments that extended from postition +54 to position -5515 relative to the translation start site +1 of the HD gene: primer sequences as well as the length and the positions of the eight DNA segments are presented in Table IV. Amplified segments were sequenced using the same sets of primers. The extended products were visualized on an ABI 3100 Genetic Analyzer and the results were compared to the available sequence in the database using blast software. Fifty to 100 control chromosomes were analysed for the three single nucleotide alterations identified. The presence of a G→A change at position -1757 was tested by digesting the 1083 bp PCR product for 50 DNA control samples with PvuII at 37˚C. DNA samples that were homozygous for the G→A change were digested to three bands of 533, 350 and 255 bp, whereas the heterozygous and the homozygous for the reported nucleotide were digested to 790, 533, 350, 255 bp and 790, 350 bp respectively. The insertion of cytosine (C base) between positions -662 and -663 was

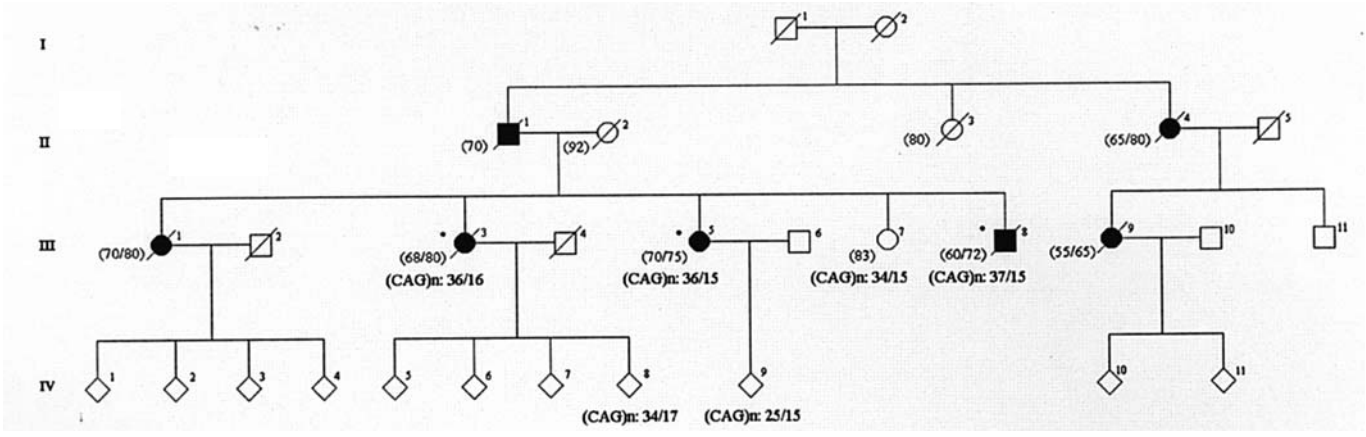

Figure 2. Pedigree of the newly identified family (MA) affected by late-onset HD, circles denote females; squares denote males and filled symbols indicate family members affected by Huntington's disease. A diagonal line through a symbol indicates that the individual is deceased. A dot at the right upper corner of the symbol indicates HD patients personally evaluated. Number typed in bolt below the symbols show the number of (CAG)n repeats. The first number in parentheses represents age at disease onset and the second number current age or age at death. If only one number is shown, this represents current age or age at death. For reasons of confidentiality, individuals of fourth generation are indicated by diamonds.

Table V. Age at disease onset for parent/offspring pairs in late-onset HD kindreds.

	Mean age at onset ± SEM	No. with age at onset ≥50 years (%)	No. with age at onset ≥60 years (%)
Parent (n=39)	55.8±7.03	29/31 (94)	14/31 (45)
Offspring (n=39)	54.8±8.46	31/39 (80)	14/39 (36)
Change: parent-child	0.97±6.67		
Mother (n=18)	56.7±5.06	13/14 (93)	5/14 (36)
Offspring (n=18)	55.9±6.78	15/18 (83)	6/18 (33)
Change: parent-child	0.83±7.94		
Father (n=21)	55.0±8.71	11/13 (85)	7/13 (54)
Offspring (n=21)	53.9±9.9	15/21 (71)	8/21 (38)
Change: parent-child	1.1±5.14		

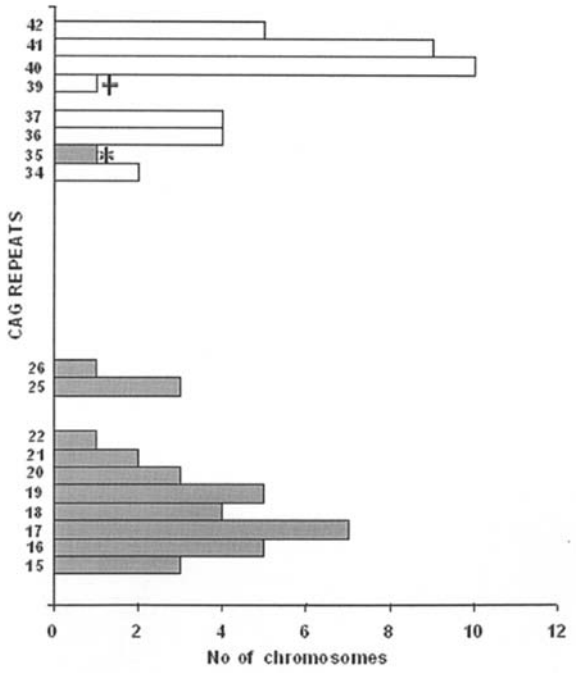

Figure 3. Distribution of the CAG repeats in HD chromosomes and normal chromosomes of late-onset HD patients. *Derived from contraction of a 36 repeat allele (see pedigree on Fig. 2). +'Normal' allele of a HD patient with a disease allele of 41 CAG repeats. Dark colour bars represent the number of normal chromosomes whereas light colour bars represent HD chromosomes.

tested by digesting the 927 bp PCR product with the BsrBI restriction enzyme. Homozygous control subjects for the C-insertion were digested to 382, 336 ad 219 bp. The presence of the A→C change at position -694 bp and the four polymorphic sites of the HD gene promoter region were tested by sequencing the two PCR products containing the alterations using the same set of primers and were visualized on a 3100 ABI Genetic Analyzer.

Results

Late-onset HD pedigrees. We studied nine kindreds with late-onset HD. Six of these families (kindreds BR1-6) originated from a single village located in the northeaster part of the island. Two families (kindreds AM and MA) were from two different villages, also located in the northeastern part of Crete and the last (kindred PE) was from a village in the south part of the Heraklion Prefecture. Families BR1-6 and AM were reported previously (13), whereas kindreds MA (Fig. 2) and PE are new.

The nine late-onset families encompassed 438 family members, 61 of which were affected by HD. The mean age of disease onset was 56.1±10.2 years. Over 80% and 40% of HD patients experienced their first symptoms after 50 and 60 years of age, respectively. Clinical data from 39 parent-child transmissions of HD studied revealed that age at disease onset in the parent and in the offspring were very similar. This was true for both paternal and maternal transmissions

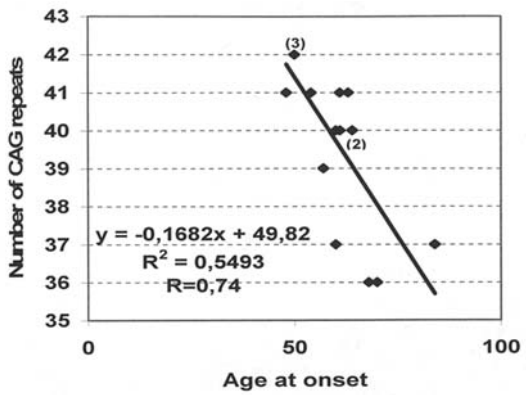

Figure 4. Comparison of CAG repeats size and age at onset in patients with the late-onset form of HD from Crete. Number in parentheses represents the number of HD patients with the same age at disease onset and the same number of CAG repeats.

Table VI. Meiotic transmission of the expanded CAG repeat.

A. Late-onset HD families	Paternal	Maternal	Total
Same size (CAG)n	6	2	8
Contraction by 1 repeat	1	2	3
Contraction by 2 repeats	-	1	1
Expansion by 1 repeat	1	-	1
Expansion by 2 repeats	1	-	1
Total	9	5	14

B. Conventional HD families	Paternal	Maternal	Total
Expansion by 3 repeats	1	-	1
Expansion by 13 repeats	1	-	1
Total	2		2

(Table V). Presently, 46 family members affected by HD are dead. Many of whom succumbed to their disease in the 12-14 years while on follow-up by us. Their average age at death was 68.0±10.5 years.

The size of the expanded CAG repeat (Fig. 3) in late-onset HD chromosomes (n=35) ranged from 36 to 42 repeats (with more than 2/3 of these alleles having ≥40 CAG repeats) and correlated inversely (r=0.74) with age at onset (Fig. 4). Two individuals at risk for HD in the late-onset pedigree MA inherited 34 repeat alleles that resulted from contraction of 36 alleles (Fig. 3). The length of the normal alleles in subjects with one HD chromosome ranged from 15 to 25 repeats (mean: 19 repeats). A HD patient with a CAG repeat of 41 trinucleotide units had a 'normal' allele in the intermediate range (35 CAG repeats).

Fourteen meiotic transmissions of the CAG repeat (nine paternal and five maternal) were studied. Twelve of these involved transmission of the expanded CAG repeat to presently asymptomatic children of HD patients and two to offspring affected by HD. Results revealed that the repeat was passed unaltered in 8 out of 14 transmissions (57%) and it changed slightly (by 1 or 2 repeats) in the remaining 6 (43%) (Table VI).

Conventional HD pedigrees. Two Cretan pedigrees (AX and KO) affected by conventional HD were studied. The average age at disease onset was 43.5 years in the AX kindred and 35.3 years in KO kindred. Anticipation was evident in both of these families with two juvenile HD cases occurring in the AX kindred and one in the KO kindred. DNA analysis showed expansion of the CAG repeat in the range of 37-54 repeats (mean 42). In two meiotic transmissions studied, the CAG expanded in the one by 3 repeats and in the second by 13 repeats (Table VI), with the latter expansion found in a juvenile HD case. Hence, these clinical and molecular genetic data for conventional HD on Crete are quite similar to those reported worldwide and clearly differentiate the late-onset form of the disorder from typical HD.

Microsatellite Markers and intragenic SNPs	Late-onset HD families									Conventional HD families	
	BR1	BR2	BR3	BR4	BR5	BR6	MA	AM	PE	AX	KO
D4S3038	221	221	221	221	221	221	221	221	209	221	219
D4S43	178	178	178	178	178	178	176	186	182	180	182
D4S136	235	235	235	235	235	235	233	233	233	233	233
D4S182	126	126	126	126	126	126	126	124	122	122	124
D4S95	1090	1090	1090	1090	1090	1090	1090	1090	1090	1090	1090
D4S127	157	157	157	157	157	157	157	157	157	159	159/154
rs362277	A	A	A	A	A	A	A	A	A	G	G/A
rs3025814	G	G	G	G	G	G	G	G	G	C/G	C
Δ2642	112	112	112	112	112	112	112	112	112	109	112
rs2530596	A	A	A	A	A	A	A	A	A	G/A	G
D4S3034	188	188	188	188	188	188	188	190	190	190	190
D4S126	169	169	169	169	169	169	169	161	165	169	161
ADRA2C	183	183	183	183	183	183	183	183	183	183	181
D4S431	262	252	256	250	250	256	254	256	254	250	254

Figure 5. Haplotypes of late-onset and conventional HD families from Crete. The subgroup of the BR1-BR6 late-onset HD families, which originated from a single village share an extensive haplotype enclosed by doted line boxes based on STS and SNP that span a 0.277 Mb segment on 4p16.3. A more narrow chromosomal area containing the STS markers D4S95, D4S127 and the intragenic SNPs rs362277, rs3025814, rs2530596, is shared by all nine late-onset HD families (solid line boxes). In contrast, the two families with conventional HD (AX and KO) share a distinct haplotype (enclosed in solid line boxes). STS markers and SNPs are listed from telomere (top) to centromere (bottom).

Table VII. Allele-frequencies for HD (late-onset and typical chromosomes) and for normal chromosomes.

Marker/alleles		Control chromosomes (n=114)	Normal chromosomes of late-onset HD patients and their relatives (n=80)	Late-onset HD chromosomes (n=33)	Δ (90% CI)	Typical HD chromosomes (n=9)	δ (90% CI)
		Frequency	Frequency	Frequency		Frequency	
D4S43	190	0.009	0	0		0	
	188	0.017	0	0		0	
	186	0.149	0.175	0.182		0	
	184	0.035	0.0875	0		0	
	182	0.395	0.4125	0.091		0.445	
	180	0.342	0.1625	0		0.555	
	178	0	0.0375	0.636[a]	**0.636** (0.498-0.774)	0	
	176	0.053	0.1125	0.091		0	
	174	0	0.0125	0		0	
D4S136	255	0	0.0123	0		0	
	249	0.017	0	0		0	
	247	0	0	0		0	
	243	0	0.0123	0		0	
	241	0.009	0	0		0	
	239	0.009	0	0		0	
	237	0.079	0.062	0		0	
	235	0.377	0.160	0.636[a]	**0.416**	0	
	233	0.394	0.629	0.364		1	
	229	0.009	0	0		0	
	227	0.053	0	0		0	
	225	0.053	0.074	0		0	
	223	0	0.0247	0		0	
D4S182	138	0.009	0	0		0	
	136	0.009	0.0125	0		0	
	132	0.009	0	0		0	
	130	0.018	0	0		0	
	128	0.018	0.0375	0		0	
	126	0.116	0.1	0.727[a]	**0.691**	0	
	124	0.446	0.45	0.182		0.445[a]	**0.43**
	122	0.241	0.325	0.091		0.555	
	120	0.027	0.025	0		0	
	118	0.062	0.025	0		0	
	116	0.027	0.025	0		0	
	114	0.018	0	0		0	
D4S126	177	0	0	0		0	
	175	0.009	0	0		0	
	173	0.018	0.0617	0		0	
	171	0.158	0.1728	0		0	
	169	0.131	0.185	0.727[a]	**0.686** (0.541-0.831)	0.555[a]	**0.488**
	167	0.140	0.098	0		0	
	165	0.219	0.247	0.091		0	
	163	0.193	0.185	0		0	
	161	0.096	0.05	0.182		0.445	
	159	0.009	0	0		0	
	155	0	0	0		0	
	153	0.018	0	0		0	
	145	0.009	0	0		0	
D4S431	268	0.009	0	0		0	
	266	0.019	0	0		0	
	264	0.047	0.074	0		0	
	262	0.009	0.0247	0.152[a]	**0.144** (0.04-0.248)	0	
	260	0.028	0	0		0.112	
	258	0.188	0.16	0		0	
	256	0.207	0.111	0.394		0	
	254	0.264	0.26	0.182		0.444	
	252	0.113	0.123	0.182		0	
	250	0.094	0.136	0.09		0.444[a]	**0.49**
	248	0.009	0.037	0		0	
	244	0.009	0.0246	0		0	

Allele-frequency estimates for mutation bearing and control chromosomes for the five markers that flank the conserved haplotype for both late-onset and conventional HD. [a]$p<0.001$ (χ^2). Numbers in parenthesis are the lower and upper bounds of the 90% confidence interval for δ.

Table VIII. Estimation of age of late-onset HD.

Marker/ alleles	Distance from late-onset HD locus: Megabases	Centimorgans	θ[a]	LD[b] δ	Estimated age[c]-G (90% CI)[d]	Mean age ± SD (90% CI)[d]
D4S43	0.800	1.28	0.0126	**0.636**	29	
178					(17-45)	**31±2.1**
D4S126	0.725	1.16	0.0115	**0.686**	33	**(29-35)**
169					(16-55)	
D4S431	3.540	5.664	0.0535	**0.144**	32	
262					(23-53)	

[a]Estimated on the basis of physical distance (in Mb) by use of the conversion factor 1.6 cM/Mb and Haldane's map function. Factor 1.6 was estimated on the basis of 7 genetic markers (D4S43, D4S182, D4S95, D4S127, D4S180, D4S125, and D4S126), spanning the region of interest, for which both physical and genetic distances are available. We did not use the STR markers D4S136 and D4S182 as they are not accurately placed in 4p16.3 according to the most current physical map (National Center for Biotechnology Information). [b]Calculated according to the method of Bengtsson and Thomson (15): $\delta=(p_D-p_N)/(1-p_N)$. [c]G is the age calculated by use of equation $G=\ln\delta/(\ln(1-\theta)$ [Risch *et al* (18)]. [d]An approximate CI for G was calculated with the above formula when the minimum CI for δ was used. Numbers in parenthesis indicate a 90% confidence interval for G.

Haplotype analysis. All 33 late-onset HD chromosomes shared a highly conserved haplotype defined by STR markers D4S95: 1090, D4S127: 157 and intragenic SNPs 362277: A, 3025814: G, 2530596: A, which span a 0.277-Mb on 4p16.3. This segment is flanked by the STR markers D4S95 and D4S3034 that are 0.336 Mb apart (Fig. 1). As shown in Fig. 1, the late-onset HD chromosomes of kindreds BR1-BR6 (derived from a single village) shared a wider chromosomal region, spanning 2.7 Mb between markers D4S3038 and ADRA2C.

In contrast to the late-onset kindreds, each of the two conventional HD families showed its own distinct haplotype (Figs. 1 and 5). The AX pedigree haplotype was defined by the markers D4S95: 1090, D4S127: 159, rs362277: G, rs3025814: C/G, Δ2642: 109, rs2530596: G/A and the KO pedigree by D4S95: 1090, D4S127: 159/155, rs362277: G/A, rs3025814: C, Δ2642: 112, rs2530596: G (Fig. 5).

Linkage disequilibrium analysis and age estimation. Table VII shows the allele frequencies for five STR markers flanking the conserved haplotype for the late-onset HD chromosomes (n=33), for the normal chromosomes of patients with late-onset HD and their relatives (n=80) and for control chromosomes (n=114), sampled at random from the Cretan population. A single allele for each marker locus tested was associated with the disease; it was found to be in striking excess among the late-onset HD chromosomes, as compared to the two groups of normal chromosomes (D4S43, allele 178, corrected p<0.001; D4S136, allele 235, p<0.008; D4S182, allele 126, p<0.001; D4S126, allele 169, p<0.001; D4S3034, allele 188, p=0.038 and D4S431, allele 262, p<0.009). The δ values obtained are shown in Table VII. The extent of LD for the polymorphic markers D4S43, D4S126 and D4S431, which are accurately placed on physical maps (National Center of Biotechnology Information-August 2003), were used to estimate the generations to the most recent common ancestor for late-onset HD. Table VIII shows the δ values and the recombination fractions used to calculate the number of generations (G) to the most

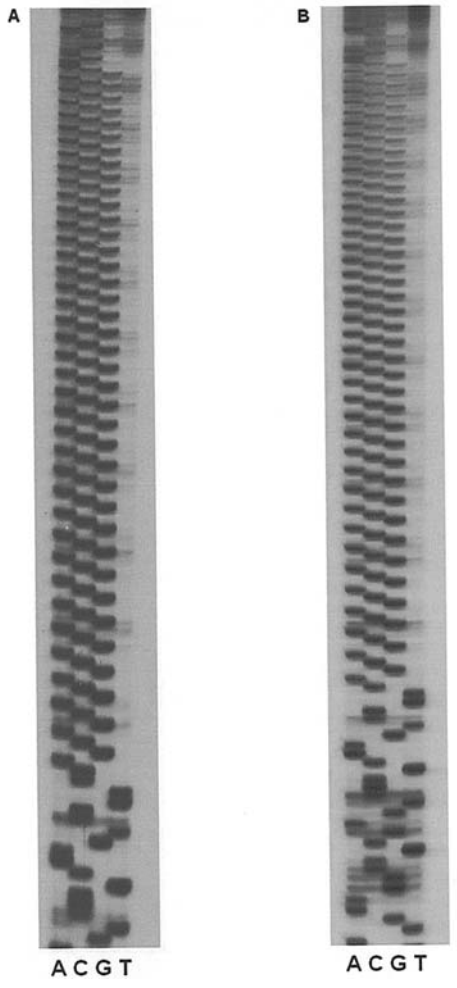

Figure 6. The CAG repeat of the late-onset HD chromosomes. TA vector clones containing part of the first exon of the HD chromosome were sequenced. A, Late-onset HD chromosome from BR1 family containing 38 CAG repeats. B, Late-onset HD chromosomes from the AM family containing 40 CAG repeats. As shown here, the (CAG)n repeat is not interrupted. Also the sequenced region of the 5' end of the HD gene (from -109 bp to +587 bp) that flanks the CAG shows no DNA alterations compared to published data (3).

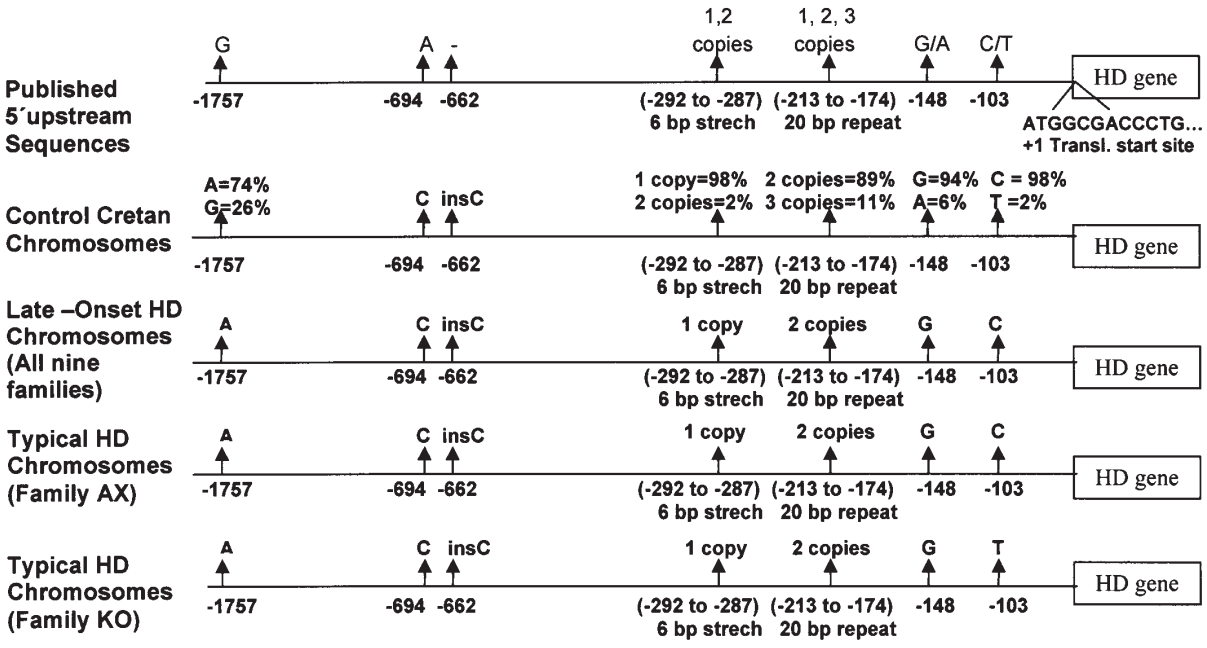

Figure 7. Haplotypes of 5' upstream HD gene sequences. In addition to previously described rare polymorphisms at positions -103, -148, -213 to -174 and -292 to -287, an SNP (G→A) was detected at position -1757. This was present in 74% of control chromosomes from Crete and in 52% of North American control chormosomes. Also, our sequence data differ from those previously described in that an A→C change was present at position -694 and an insertion of C between -662 and -663. These two latter changes were present in all 28 control chromosomes studied from North American subjects of European origin.

recent common ancestor for late-onset HD, as well as a minimum and a maximum confidence interval for *G*. The age estimates, obtained with the use of the three STR markers, were quite consistent, with the mean value being 31.0±2.1 generations (90% confidence interval: 29-35 generations).

The CAG repeat is not interrupted in late-onset HD. There is evidence that interruption of the CAG repeat in the SCA1 gene (19) and of the AGG repeat in the FMR1 gene (20) confers stability to the repeat. To see whether changes in the CAG repeat were present in our patients with late-onset HD, we sequenced a 5' end segment of the HD gene in these patients. Results showed that the elongated CAG repeat was uninterrupted (Fig. 6). Also none of these patients had the CAA→CAG mutation for the penultimate glutamine (19).

DNA polymorphisms of the 5' upstream sequences. Analysis of 5' upstream sequences (from position +54 to -5515 relative to the translation start site +1) of the HD gene, revealed the four known rare polymorphisms in the promoter region (Fig. 7) (21). These involved single base pair substitutions at positions -103 (C→T) and -148 (G→A), a 6-bp stretch repetition at position -292 to -287, and a 20-bp direct repeat element at position -213 to -174. In addition, we found a new single nucleotide polymorphism at position -1757 (G→A). This was present in all of the HD chromosomes we studied (late-onset and typical pedigrees). The G→A change at position -1757 was found in 74% of control chromosomes from Crete (n=206) and in 52% of chromosomes of North American subjects of European origin (n=118) ($p<0.001$) (Fig. 4). All HD chromosomes from all the nine late-onset HD families and from the typical HD kindred AX showed the same haplotype: -103: C, -148: G, 20 bp repeat (-213 to -174): 2 copies, 6 bp repeat (-292 to -287): 1 copy, -1757: A. In contrast, the typical HD family KO, had a T instead of C at position -103. This was present in only 2% of the control chromosomes from Crete (Fig. 7).

In addition, comparison of our sequence data with those previously reported in the databank revealed two single nucleotide changes consisting of an A→C change at position -694 and an insertion of C between -662 and -663 (always relatively to the translation starting site +1). These were present in all control and HD chromosomes from Crete and in all 28 chromosomes of North American subjects of European extraction studied. Hence, they should represent sequencing errors present in the data bank.

Discussion

On the island of Crete, we have observed that late-onset HD kindreds are more common than those affected by the typical (mid-life onset) form of the disorder. Our sample is quite representative of the disease on the island, as the University Hospital of Crete is a neurology reference center that draws patients from the entire island. Natives of Crete have a life expectancy similar (or even higher than) to that of other Western countries, maintain strong family ties and show low migration rates. This permitted us to obtain accurate genealogical information for each HD proband and to trace the disorder in their families.

Clinical data on 39 parent-child pairs (father: 21 mother: 18) from late-onset HD families revealed no significant differences with respect to the age at disease onset. None of these transmissions resulted in young-onset HD cases, regardless as to whether the disorder was inherited through the father or the mother. In addition, the study of 12 meiotic transmissions of the elongated CAG repeat to presently asymptomatic family members did not show expansions that could predict the

development of young-onset HD in these individuals. Most of these members with a HD chromosome were adults in their 30s to 50s. The present results (based on 51 HD transmission in total) contrast those derived from studies on conventional HD families showing the frequent instability of the CAG repeat particularly when inherited from the father (5,7). Snell *et al* (5) have reported a mean difference of 9.11 years for paternal transmissions of HD. Lack of anticipation in our late-onset HD kindreds can be explained by the propensity of the CAG repeat to retain its length or to change only slightly (by one or two CAG repeats) in its passage through successive generations.

Haplotypic analyses, using several polymorphic markers on 4p16.3, revealed that all late-onset HD families shared a highly conserved haplotype defined by the STR markers and SNPs (D4S95: 1090, D4S127: 157, rs362277: A, rs3025814: G, rs2530596: A) that span a 0.277-Mb segment on 4p16.3. In contrast, each of the two conventional HD families of Crete studied showed its own distinct haplotype suggesting two different founders. Interestingly, Δ2642 an intragenic marker present in 2% of controls from Crete was present in the AX (but not in the KO) typical HD family. The converse was true for the rare promoter region polymorphism at position -103 (C→T), which was present in the KO (but not in the AX) typical HD pedigree. This polymorphism was present in 2% of the controls from Crete.

Coalescence analysis for late-onset HD suggested a coalescence time of 31.0±2.1 generations ago (90% confidence interval = 29-35 generations). If we assume that a generation comprises 25 years, the coalescence time for the mutation of late-onset HD was 875 years ago or c. 1100 AD. This corresponds to the late Byzantine or early Venetian period on Crete. Given the dynamic nature of the HD mutation, our age estimation cannot determine when the HD mutation occurred, i.e. when the CAG repeat expanded from a normal to a pathological size. Instead, it estimates when the founder acquired in his/her HD chromosome DNA sequences that led to the late-onset HD. The linkage disequilibrium, found between several DNA markers located on 4p16.3 and the disease locus, is clearly not due to stratification that can occur in genetic isolates, as the normal chromosomes of the late-onset HD cases showed allele frequencies similar to those of the general population of Crete.

Although we found no significant differences between parent-child pairs in late-onset HD with respect to the age at onset, there was a slight trend toward earlier onset in the offspring than in the parent (mean change: 1.0 year, Table V). If indeed each successive generation developed the disease one year earlier, it is possible that it took over 30 generations (as suggested by the coalescence analysis) for the age of onset to be lowered to the present 56 years. As such, at the time of the origin of the disease-associated haplotype (about 30 generations ago), the founder would be expected to show his/her first symptoms at a very old age (8th-9th decade of life), should he/she have lived that long.

These observations suggest that the genesis of the HD mutation in late-onset HD proceeded via small-increment expansion of the CAG repeat, probably occurring over many generations. Should this mutation have arisen from elongation of normal mutable intermediate alleles containing 25-35 CAG repeats (alleles of this size were found in the 'normal' chromosomes of Cretan HD cases), this gradual process is likely to generate alleles in the low pathological range (just above the threshold for disease expression) thus resulting in elderly HD founders. However, if this is the prevailing mode of genesis of the HD mutation, why are late-onset HD pedigrees not prevalent? One possible explanation is that the late-onset-HD mutation in the population of Crete is a relatively recent event and that the occurrence of both typical and late-onset HD in this population may represent different stages of disease evolution.

In contrast to the above gradual model of HD mutation genesis, all new (*de novo*) HD cases reported in the literature (8-10,23-27) have been related to large-scale expansion of parental intermediate alleles (27-35 repeats). Specifically, in 37 new (sporadic) HD cases described (4,9,23-27), the mutated parental intermediate allele increased by 4-30 trinucleotide units, generating pathological alleles of 41-66 repeats. The mean age at disease onset for these *de novo* HD cases was 33±16 years (9,23,24) which is somewhat lower than that for familial HD cases. The propensity of intermediate alleles of parents of *de novo* HD to undergo large expansions during meiotic transmission is further evidenced by the increased recurrent risk for the offspring of the parent with the mutated allele (9,16,22,23-27). In addition, the newly expanded CAG repeat remains unstable during its transmission to the succeeding generation (7,11). On the other hand, intermediate alleles found in collateral branches of *de novo* HD families, are mostly stable during meiotic transmission or show small-increment instability (9,22-27).

At present, the molecular mechanisms that render the CAG repeat unstable in HD remain unclear. Goldberg *et al* (28), produced transgenic mice containing a full length human HD cDNA and found that the CAG repeat remained remarkably stable through 97 meioses, thus suggesting that genomic sequences outside the HD gene are responsible for the instability of the repeat in HD. More direct evidence implicating cis-acting elements in the CAG repeat instability that occurs in Kennedy syndrome have been obtained by La Spada *et al* (29) who reported that integration of a 70-kb segment of the human androgen receptor (AR) locus in mice led to an unstable transgenic line, whereas the integration of the AR gene without the flanking sequences produced stable transgenic lines. Also Libby *et al* (30) have suggested that a particular sequence motif in close proximity to the gene responsible for SCA-7 influences repeat instability at this locus. La Spada *et al* (31) have identified a 50-kb DNA region upstream of the fragile X gene (FRAXA) that appeared to be involved in FRAXA allele instability.

Our observations, showing that all late-onset HD families (unlike the typical HD kindreds) derived from the same ancestor, suggest that this form of HD is genetically determined via cis-acting elements. In search of cis-acting elements, we analyzed 5' upstream sequences in Cretan chromosomes and found, in addition to the four known rare polymorphisms in the promoter region of the HD gene (21), three new single nucleotide changes compared to sequences available in the databanks. Two of these changes (an A→C change at position -694 and an insertion of C between -662 and -663 relatively to the translation starting site +1) were sequence error as they

were present in all normal DNA samples from Crete as well as in all 28 North American chromosomes studied. Substitution of G for A at position -1757 was present in 74% of Cretan Controls and in 52% of North American subjects studied and in all HD chromosomes from Crete. As late-onset and typical (AX family) HD chromosomes showed the same haplotype (present in 71% of controls), these changes may not play a role in the CAG stability of the late-onset HD. However, we cannot exclude the possibility of cis-acting elements located at greater distances from the HD gene as shown for Kennedy syndrome and FRAXA (31).

In summary, in this small population sharing a common genetic background, late-onset HD kindreds are more prevalent than those with the typical, mid-life onset HD. Haplotypic analyses suggested that all late-onset HD cases derived from a single ancestor who lived about 1000 years ago. These findings suggest that late-onset HD is genetically determined and that the disease mutation evolved over the centuries via small increment instability.

Acknowledgments

This work was supported by an educational grant (EPEAEK Pythagoras) to C.S. from the Ministry of Education of Greece and by the Association for Research and Treatment of Neurologic Disorders of Crete (‘EY ZHN’). This study was also supported by NIH grants NS16367, the Huntington's Disease Center Without Walls, and NS 32765 to M.M. The authors thank patients and family members who participated in this study. They also thank Irini Skoula, Aggela Pasparaki and Athanasios Alegakis for help with this study, and Iris Plaitakis for editorial support.

References

1. Conneally PM: Huntington disease: genetics and epidemiology. Am J Hum Genet 36: 506-526, 1984.
2. Ridley RM, Frith CD, Crow TJ and Conneally PM: Anticipation in Huntington's disease is inherited through the male line but may originate in the female. J Med Genet 25: 589-595, 1988.
3. Kremer B, Squitieri F, Telenius H, *et al*: Molecular analysis of late onset Huntington's disease. J Med Genet 30: 991-995, 1993.
4. The Huntington's Disease Collaborative Research Group: A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 72: 971-983, 1993.
5. Snell RG, MacMillan JC, Cheadle JP, *et al*: Relationship between trinucleotide repeat expansion and phenotypic variation in Huntington's disease. Nat Genet 4: 393-397, 1993.
6. Stine OC, Pleasant N, Franz ML, *et al*: Correlation between the onset age of Huntington's disease and length of the trinucleotide repeat in IT-15. Hum Mol Genet 2: 1547-1549, 1993.
7. Duyao M, Ambrose C, Myers R, *et al*: Trinucleotide repeat length instability and age of onset in Huntington's disease. Nat Genet 4: 387-392, 1993.
8. Kremer B, Goldberg P, Andrew SE, *et al*: A worldwide study of the Huntington's disease mutation. The sensitivity and specificity of measuring CAG repeats. N Engl J Med 330: 1401-1406, 1994.
9. Goldberg YP, Andrew SE, Theilmann J, *et al*: Familial predisposition to recurrent mutations causing Huntington's disease: genetic risk to sibs of sporadic cases. J Med Genet 30: 987-990, 1993.
10. Rubinsztein DC, Leggo J, Coles R, *et al*: Phenotypic characterization of individuals with 30-40 CAG repeats in the Huntington disease (HD) gene reveals HD cases with 36 repeats and apparently normal elderly individuals with 36-39 repeats. Am J Hum Genet 59: 16-22, 1996.
11. Djousse L, Knowlton B, Hayden M, *et al*: Interaction of normal and expanded CAG repeat sizes influences age at onset of Huntington disease. Am J Med Genet 119: 279-282, 2003.
12. Li JL, Hayden MR, Almqvist EW, *et al*: A genome scan for modifiers of age at onset in Huntington disease: the HD MAPS study. Am J Hum Genet 73: 682-687, 2003.
13. Tsagournissakis M, Fesdjian OC, Shashidharan P and Plaitakis A: A stability of the Huntington's disease (CAG)n repeat in a late onset form occuring on the island of Crete. Hum Mol Genet 4: 2239-2243, 1996.
14. Riess O, Noerremolle A, Soerensen SA and Epplen JT: Improved PCR conditions for the strength of (CAG)n repeats causing HD. Hum Mol Genet 2: 1523, 1993.
15. Bengtsson BO and Thomson G: Measuring the strength of associations between HLA antigens and diseases. Tissue Antigens 18: 356-363, 1981.
16. Rannala B and Bertorelle G: Using linked markers to infer the age of a mutation (Review). Hum Mutat 18: 87-100, 2001.
17. Lonjou C, Collins A, Ajioka RS, *et al*: Allellic association under map error and recominational heterogeneity: a tale of two sites. Proc Natl Acad Sci USA 95: 11366-11370, 1998.
18. Risch N, De Leon D, Ozelius L, *et al*: Genetic analysis of idiopathic torsion dystonia in Askenazi Jews and their recent descent from a small founder population. Nat Genet 9: 152-159, 1995.
19. Chung MY, Ranum LP, Duvick LA, *et al*: Evidence for a mechanism predisposing to intergenerational CAG repeat instability in spinocerebellar ataxia type I. Nat Genet 5: 254-258, 1993.
20. Eichler EE, Holden JJ, Popovich BW, *et al*: Length of uninterrupted CGG repeats determines instability in the FMR1 gene. Nat Genet 8: 88-94, 1994.
21. Coles R, Leggo J and Rubinstzein DC: Analysis of the 5' upstream region of the Huntington's disease (HD) gene shows six new alleles which are unrelated to the age at onset of HD. J Med Genet 34: 371-374, 1997.
22. McNeil SM, Novelletto A, Srinidhi J, *et al*: Reduced penetrance of the Huntington's disease mutation. Hum Mol Genet 6: 775-779, 1997.
23. Goldberg YP, McMurray CT, Zeisler J, *et al*: Increased instability of intermediate alleles in families with sporadic Huntington disease compared to similar sized intermediate alleles in the general population. Hum Mol Genet 4: 1911-1918, 1995.
24. Goldberg YP, Kremer B, Andrew SE, *et al*: Molecular analysis of new mutations for Huntington's disease: intermediate alleles and sex origin effects. Nat Genet 5: 174-180, 1993.
25. Laccone F and Christian W: A recurrent expansion of a maternal allele with 36 CAG repeats causes Huntington disease in two sisters. Am J Hum Genet 66: 1145-1148, 2000.
26. Sanchez A, Mila M, Castellvi-Bel S, *et al*: Maternal transmission in sporadic Huntington's disease. J Neurol Neurosurg Psychiatry 62: 535-537, 1997.
27. Watanabe M, Satoh A, Kanemoto M, *et al*: *De novo* expansion of a CAG repeat in a Japanese patient with sporadic Huntington disease. J Neurol Sci 178: 159-162, 2000.
28. Goldberg YP, Kalchman MA, Metzler M, *et al*: Absence of disease phenotype and intergenerational stability of the CAG repeat in transgenic mice expressing the human Huntington disease transcript. Hum Mol Genet 5: 177-185, 1996.
29. La Spada AR, Peterson KR, Meadows SA, *et al*: Androgen receptor YAC transgenic mice carrying CAG 45 alleles show trinucleotide repeat instability. Hum Mol Genet 7: 959-967, 1998.
30. Libby RT, Monckton DG, Fu YH, *et al*: Genomic context drives SCA7 CAG repeat instability, while expressed SCA7 cDNAs are intergenerationally and somatically stable in transgenic mice. Hum Mol Genet 12: 41-50, 2003.
31. La Spada AR, Richards RI and Wieringa B: Dynamic mutations on the move in Banff. Nat Genet 36: 667-670, 2004.