

been shown that the expression of the siderophore receptor PheA is mediated by both the Fur (ferric uptake regulation) protein and a classic two-component regulatory system in *Pseudomonas putida* (15). The AirS protein has amino acid sequence similarities to the sensor-regulator proteins BvgS (54%) of *Bordetella pertussis* and LemA (56%) of *Pseudomonas syringae* (16). Structurally, AirS is also similar to BvgS and LemA as they all contain both an autophosphorylated histidine site (transmitter module) and a phospho-accepting aspartate site (receiver module) in their primary amino acid sequences (7). As part of a global regulatory network, BvgS and LemA have each been shown capable of responding to various environmental signals encountered, resulting in the coordinate regulation of transcription of their virulence factors during the bacterial infectious cycle (14, 16).

The ability to obtain adequate iron from the environment is closely linked to bacterial virulence. Clinical isolates of bacteria with iron-acquisition systems have enhanced virulence in animals. Additionally, aerobactin-producing strains of uropathogenic *E. coli* are more lethal in an animal model of ascending urinary tract infection than non-aerobactin-producing strains (17). In vivo, aerobactin has been detected in urine from patients with *E. coli* urinary tract infection (18). Furthermore, serum from patients with *E. coli* bacteriuria have antibodies against IRMPs (19). The rapid transcriptional activation of *airS* upon P-pili attachment and the requirement of *airS* for activation of the siderophore system indicate that *airS* may direct the coordinate regulation of uropathogenic *E. coli*'s iron-acquisition machinery in the urinary tract of the host.

REFERENCES AND NOTES

1. J. R. Johnson, *Clin. Microbiol. Rev.* **4**, 80 (1991).
2. J. A. Roberts *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 11889 (1994).
3. M. J. Kuehn *et al.*, *Nature* **356**, 252 (1992); E. Bullitt and L. Makowski, *ibid.* **373**, 164 (1995).
4. N. Stromberg *et al.*, *EMBO J.* **9**, 2001 (1990).
5. K. K. Wong and M. McClelland, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 639 (1994); P. Liang and A. B. Pardee, *Science* **257**, 967 (1992).
6. Attachment and RT-PCR procedures were as follows: DS17 or DS17-8 were grown at 37°C in LB broth overnight, and 1 ml of overnight culture was inoculated into 9 ml of fresh LB broth. Cultures were incubated for 3 hours more, and 0.5 ml of 10% glutaraldehyde-stabilized human group O RBCs or rabbit RBCs were added to the cultures. After 30 min to allow for attachment, cells were pelleted, and the pellets of bacteria and RBCs were lysed in 1 ml of TRIzol total RNA isolation reagent (Gibco BRL). Total RNA isolation was continued according to TRIzol instructions. Approximately 2 mg of total RNA harvested from each sample was reverse-transcribed and then used as templates for PCR with a non-specific set of primers as described in (5). The PCR parameters were 94°C for 1 min, 42°C for 1 min, and 72°C for 1 min for a total of 40 cycles, followed by 10 min of elongation at 72°C.
7. S. Nagasawa *et al.*, *Mol. Microbiol.* **6**, 799 (1992); K. Ishige *et al.*, *EMBO J.* **13**, 5195 (1994).
8. Specific attachment and RT-PCR procedures were as follows: Total RNA isolations were as described in (6). Approximately 2 mg of total RNA harvested from each sample was reverse-transcribed and then used as templates for PCR with the *airS* primer set. The primer sequences were as follows: *airS*-1 (5'-CAG-CACGCTGGATGAAGTCGT-3') and *airS*-2 (5'-AGCGCGATTGATCGCGTTTCAGGAAT-3'). For Fig. 1, B and C, petri dishes (100, 15 mm in diameter) were coated overnight at 4°C with 200 mg of GbO₄ or CDH dissolved in 3 ml of methanol. The DS17 or DS17-8 strains were grown in LB broth overnight at 37°C, and 1 ml of overnight culture of each was inoculated into 9 ml of fresh LB broth. Bacteria were grown for another 3 hours and then centrifuged. Nine milliliters of culture supernatant was removed, and the bacterial pellets were resuspended in the remaining 1 ml of culture media. The bacterial suspensions were added to the pre-coated petri dishes at 37°C for 30 min with gentle shaking (60 rpm). Bacteria were lysed on the plates with 5 ml of TRIzol solution, and total RNA was isolated as above. Specific *airS* primers are listed above. The pyruvate dehydrogenase primers were *pdh*-1 (5'-TTGCTCGAAACAC-GACACGCCCT-3') and *pdh*-2 (5'-ACGGCAATCT-GATACTGGAGTAC-3'). PCR products were resolved on 8% polyacrylamide gels.
9. A. H. Free and H. M. Free, in *Urinalysis in Clinical Laboratory Practice* (CRC Press, Cleveland, 1975), chap. 4, pp. 13-19.
10. J. B. Neilands, *Can. J. Microbiol.* **38**, 728 (1992).
11. M. L. Gueriot, *Annu. Rev. Microbiol.* **48**, 743 (1994).
12. J. A. Robledo *et al.*, *J. Urol.* **143**, 386 (1990).
13. C. D. Nau and J. Konisky, *J. Bacteriol.* **171**, 1041 (1989).
14. B. Arico *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6671 (1989); A. R. Melton and A. A. Weiss, *J. Bacteriol.* **170**, 6206 (1989); P. C. Giardina *et al.*, *ibid.* **177**, 6058 (1995).
15. C. R. Dean and K. Poole, *Mol. Microbiol.* **8**, 1095 (1993).
16. D. K. Willis *et al.*, *Mol. Plant-Microbe Interact.* **1**, 80 (1990).
17. J. Z. Montgonerie *et al.*, *Infect. Immun.* **46**, 835 (1984).
18. S. M. Opal *et al.*, *J. Infect. Disease* **161**, 794 (1990).
19. E. Griffiths *et al.*, *Infect. Immun.* **47**, 808 (1985).
20. C. M. Hamilton *et al.*, *J. Bacteriol.* **171**, 4617 (1989).
21. Y. Kohara *et al.*, *Cell* **50**, 495 (1987).
22. J. Brosius, *Gene* **27**, 151 (1984).
23. J. R. Neumann *et al.*, *Biotechniques* **5**, 444 (1987).
24. We thank M. Caparon for providing the kanamycin cassette, S. R. Kushner for the plasmid pMAK705, and J. W. St. Geme for critically reading the manuscript. Supported by NIH grant 5R01GM44655-05, the Swedish Medical Research Council, the Göran Gustafsson Foundation of Natural and Medical Science, and an unrestricted grant for infectious disease research from Bristol-Myers Squibb.

9 February 1996; accepted 24 June 1996

Pycnodysostosis, a Lysosomal Disease Caused by Cathepsin K Deficiency

Bruce D. Gelb,* Guo-Ping Shi, Harold A. Chapman, Robert J. Desnick

Pycnodysostosis, an autosomal recessive osteochondrodysplasia characterized by osteosclerosis and short stature, maps to chromosome 1q21. Cathepsin K, a cysteine protease gene that is highly expressed in osteoclasts, localized to the pycnodysostosis region. Nonsense, missense, and stop codon mutations in the gene encoding cathepsin K were identified in patients. Transient expression of complementary DNA containing the stop codon mutation resulted in messenger RNA but no immunologically detectable protein. Thus, pycnodysostosis results from gene defects in a lysosomal protease with highest expression in osteoclasts. These findings suggest that cathepsin K is a major protease in bone resorption, providing a possible rationale for the treatment of disorders such as osteoporosis and certain forms of arthritis.

Recent interest in developing treatments for osteoporosis and certain forms of arthritis has stimulated efforts to identify the gene defects underlying certain inherited sclerosing skeletal dysplasias. Among these, pycnodysostosis (Pycno) is a rare, autosomal recessive trait characterized by osteosclerosis, short stature, acro-osteolysis of the distal phalanges, bone fragility, clavicular dysplasia, and skull deformities with delayed suture closure (1). The disease has gained attention as the French painter, Henri de Toulouse-Lautrec, was retrospectively diagnosed as having Pycno, a diagnosis that has been the subject of recent debate (2).

Bone resorption, a process mediated by osteoclasts, is characterized by the solubilization of inorganic mineral and subsequent proteolytic degradation of organic matrix, primarily type I collagen. In Pycno, osteoclast numbers are normal as are their ruffled borders and clear zones, but the region of demineralized bone surrounding individual osteoclasts is increased (3). Ultrastructural examination of these osteoclasts reveals large, abnormal cytoplasmic vacuoles containing bone collagen fibrils. These findings

are consistent with the findings in Pycno, a diagnosis that has been the subject of recent debate (2). Bone resorption, a process mediated by osteoclasts, is characterized by the solubilization of inorganic mineral and subsequent proteolytic degradation of organic matrix, primarily type I collagen. In Pycno, osteoclast numbers are normal as are their ruffled borders and clear zones, but the region of demineralized bone surrounding individual osteoclasts is increased (3). Ultrastructural examination of these osteoclasts reveals large, abnormal cytoplasmic vacuoles containing bone collagen fibrils. These findings

B. D. Gelb and R. J. Desnick, Department of Human Genetics and Division of Pediatric Cardiology, Mount Sinai School of Medicine, New York, NY 10029, USA. G.-P. Shi and H. A. Chapman, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA

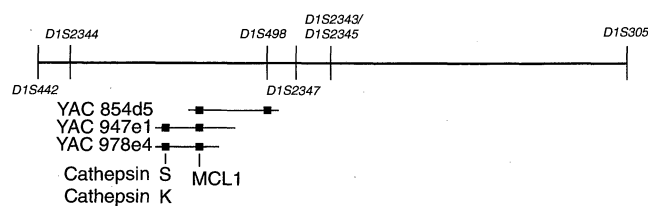
*To whom correspondence should be addressed at Department of Human Genetics, Mount Sinai School of Medicine, Box 1203, One Gustave L. Levy Place, New York, NY 10029, USA. E-mail: gelb@msvax.mssm.edu

suggest that Pycno osteoclasts function normally in demineralizing bone, but do not adequately degrade the organic matrix.

To identify the gene defect underlying Pycno, we performed a genome-wide search for the Pycno locus in a large, consanguineous Israeli Arab family with 16 affected relatives (4). A genomic region that was homozygous-by-descent for all affected individuals was localized to the pericentric region of chromosome 1 with a lod score (logarithm of the odds ratio for linkage) of 11.72 at *DIS498*. Ancestral recombinant events localized the Pycno region to 4 cM from *DIS442* to *DIS305* (Fig. 1). Independent linkage analysis of a large Mexican Pycno family confirmed linkage to the pericentric region of chromosome 1 and excluded the macrophage colony-stimulating factor *CSF1*, which is defective in the *op/op* mouse model of osteopetrosis (5). Recently, we refined the Pycno locus to a region of about 2 cM at chromosome 1q21 from *DIS2344* to *DIS2343/DIS2345* by analyzing the Israeli Arab family with new markers (6) (Fig. 1). This refinement excluded the interleukin-6 receptor gene that previously had been proposed as a candidate gene (4), but included *MCL1*, a *Bcl2* homolog relevant to monocyte/macrophage differentiation.

An additional putative candidate gene was lysosomal cathepsin S, a cysteine protease with expression in osteoclasts, which had been assigned to chromosome band 1q21 by fluorescence in situ hybridization analysis (7). To determine whether cathepsin S mapped to the Pycno critical region, we developed a cathepsin S sequence tagged site (STS) on the basis of the published 5'-flanking region sequence (7). STS content mapping was performed with a yeast artificial chromosome (YAC) contig spanning the Pycno region (8). The cathepsin S STS was amplified from two overlapping CEPH megaYACs, 978e4 and 947e1, and was mapped centromeric to *DIS498* in the Pycno critical region (Fig. 1). Sequencing of the cathepsin S coding region in affected members from the Israeli Arab family and from a consanguineous Moroccan Arab family by exon amplification from genomic DNA and by reverse transcriptase-polymerase chain reaction (RT-PCR) (9) revealed a C-to-T transversion of a CpG dinucleotide at position 343 in the cDNA in both families, predicting an Arg¹¹³ → Trp (R113W) substitution in the propeptide, near the putative cleavage site at Ile¹¹⁴ (10). Analysis of 18 unrelated normal Israeli Arab individuals living near the Israeli Arab Pycno family (11) demonstrated that 12 of 36 alleles were positive for R113W, consistent with its being a sequence polymorphism. In addition, cathepsin S activities in lymphoblasts from two Israeli Arab Pycno patients

Fig. 1. Pycnodysostosis critical region at chromosome 1q21. The microsatellite markers flanking and within the Pycno critical region are shown in the order established previously (6). Two overlapping CEPH megaYAC clones that were positive for the cathepsin S and K STSs are indicated and anchored by *DIS498*.



and three normal controls were not significantly different (210 mU/U α -galactosidase A $\pm$ 100 versus 340 ± 40 , respectively) (12). Thus, cathepsin S was ruled out as a candidate gene for Pycno.

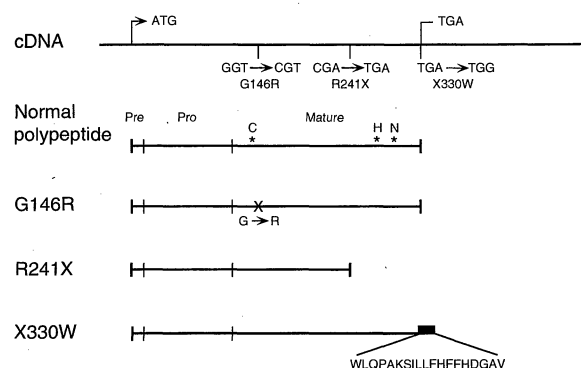
Among the other unmapped cathepsin genes, cathepsin K was a strong candidate for Pycno because this lysosomal cysteine protease is highly expressed in osteoclasts and was similar in sequence to cathepsin S, suggesting a possible tandem duplication of an ancestral cathepsin (13). A cathepsin K STS, developed on the basis of the published 3' untranslated region (13), was amplified from the same two overlapping YACs in the Pycno critical region that contained the cathepsin S STS. RT-PCR amplification and sequencing of the cathepsin K transcript from lymphoblast total RNA (14) derived from two Israeli Arab Pycno patients revealed an A-to-G transition at cDNA nucleotide 1095 (13), which predicted the substitution of the termination codon by a tryptophan residue (X330W) and the elongation of the COOH-terminus by 19 additional amino acids (Fig. 2).

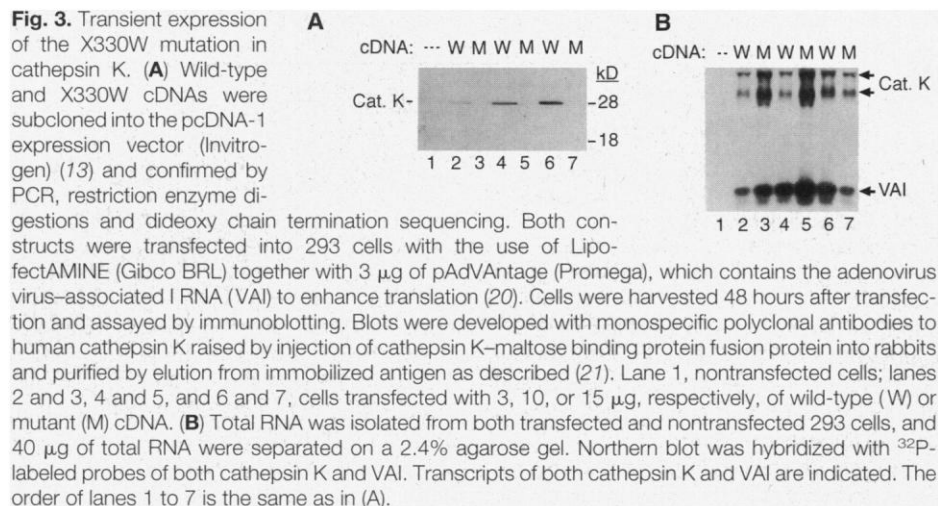
Evaluation of the X330W allele in the entire Israeli Arab Pycno family and 43 unrelated normal Arab control individuals revealed that it cosegregated with disease in the Pycno family and was not present in any of the 86 Arab control alleles (15).

Further evidence that cathepsin K mutations caused Pycno was the finding of two additional mutations in two unrelated Pycno families (Fig. 2). Two affected Moroccan Arab siblings had a missense mutation, a G-to-C transversion at nucleotide 541, predicting a Gly¹⁴⁶ → Arg (G146R) substitution. An American Hispanic patient with nonconsanguineous parents was found to be heterozygous for markers across the Pycno critical region. Sequence analysis demonstrated heteroallelism for the G146R mutation and a C-to-T transition of a CpG dinucleotide at nucleotide 826, predicting an R241X nonsense mutation. Restriction analysis of amplified segments from genomic DNA with Bam I for G146R and Ava I for R241X confirmed the RT-PCR results.

To assess the effects of the putative Israeli Arab cathepsin K mutation on enzyme activity, we introduced the X330W mutation into the full-length cDNA by site-directed mutagenesis. The normal and X330W cathepsin K alleles were transiently expressed in 293 cells, which do not have detectable cathepsin K protein (Fig. 3A). We observed cathepsin K protein by immunoblotting in the whole cell lysates of cells transfected with the normal allele. In contrast, no detectable protein was observed in lysates from cells transfected with the X330W construct. Northern analysis of total RNA from trans-

Fig. 2. Cathepsin K cDNA and polypeptide with Pycno mutations. The top line represents the cDNA with initiation and stop codons as indicated. The locations of the three mutations are indicated below the cDNA, showing both the affected codon and the predicted amino acid alteration. The normal cathepsin K polypeptide is shown with the pre-, pro-, and mature peptides. The three residues in the active site conserved among all cysteine protease cathepsins in the papain superfamily are indicated with an asterisk. The missense mutation G146R, found in both the Moroccan Arab and American Hispanic families, is shown with the position of the residue change near the active cysteine. The R241X mutation, found in the American Hispanic family, is shown with the predicted truncation of the mature peptide, eliminating two active site residues. The X330W mutation, found in the Israeli Arab family, is shown with the predicted elongation of the mature peptide by 19 residues. Abbreviations for the amino acid residues are as follows: A, Ala; D, Asp; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; P, Pro; Q, Gln; S, Ser; V, Val; and W, Trp.





fectected cells revealed comparable steady-state transcript levels between normal and mutant for both cathepsin K and the cotransfected adenovirus-associated I RNA gene (Fig. 3B).

The R241X mutation predicts polypeptide truncation at residue 241, and the loss of 89 amino acids including the completely conserved His²⁷⁶ and Asn²⁹⁶ residues of the active site (13). Such truncation mutations can also cause transcript instability and exon skipping (16). Thus, the R241X mutation presumably will be null for cathepsin K activity. The G146R mutation, found in the American Hispanic and Moroccan Arab families, occurred at a CpG dinucleotide and may prove to be a common mutation. Because this missense mutation would alter the charge of this residue, which resides near the active cysteine, and no member of the papain family has a basic residue in this position, it likely alters cathepsin K activity as well.

Recently, cysteine proteases have been implicated in bone resorption and remodeling (17). Among these, the recently identified cathepsin K was reported to be selectively expressed in osteoclasts and osteoclastomas (18). The finding that cathepsin K deficiency causes Pycno, taken together with the knowledge that cathepsin K is the only cysteine protease highly expressed in osteoclasts (13) and that it has the highest type I collagenolytic, elastinolytic, and gelatinolytic activities of cysteine proteases (19), suggests that cathepsin K is a major protease in bone matrix resorption.

Pycno, which is a pure skeletal dysplasia, can now be classified as a lysosomal disease. Most lysosomal enzymes are constitutively expressed in all cell types, and their clinical manifestations reflect the accumulation of undegraded substrate (or substrates) in various cells and tissues. In contrast, the phenotype of Pycno is restricted to bone, and there is minimal substrate accumulation in osteoclasts.

This finding is consistent with the principal site of normal cathepsin K function being the subosteoclastic space into which the enzyme is presumably secreted for bone matrix degradation. The cell-specific high expression of cathepsin K is unique among known lysosomal enzymes, defining Pycno as a lysosomal disorder caused by defective tissue-specific expression.

The treatment of Pycno has involved the symptomatic management of fractures and other skeletal problems. Future approaches to correct the abnormal bone metabolism in Pycno could include bone marrow transplantation to provide normal osteoclasts, or osteoclast-targeted enzyme or gene replacement strategies. The finding that cathepsin K deficiency causes osteosclerosis has implications for the treatment of osteoporosis and certain forms of arthritis in which cathepsin K release by osteoclasts or other cells may be injurious. In these common disorders, down-regulation of gene expression by antisense RNA strategies or the administration of specific enzyme inhibitors may decrease pathologic bone resorption.

REFERENCES AND NOTES

1. P. Maroteaux, and M. Lamy, *Presse Med.* **70**, 999 (1962); L. Andren, J. F. Dyming, K. E. Hogeman, B. Wendeberg, *Acta Chir. Scand.* **124**, 496 (1962).
2. P. Maroteaux and M. Lamy, *J. Am. Med. Assoc.* **191**, 715 (1965); P. Maroteaux, *Presse Med.* **22**, 1635 (1993); J. B. Frey, *Nature Genet.* **10**, 128 (1995).
3. V. Everts, D. C. Aronson, W. Beertsen, *Calcif. Tissue Int.* **37**, 25 (1985).
4. B. D. Gelb, J. G. Edelson, R. J. Desnick, *Nature Genet.* **10**, 235 (1995).
5. M. H. Polymeropoulos et al., *ibid.*, p. 238.
6. B. D. Gelb et al., *Hum. Genet.* **98**, 141 (1996).
7. G.-P. Shi et al., *J. Biol. Chem.* **269**, 11530 (1994).
8. B. D. Gelb and R. J. Desnick, unpublished results.
9. Cathepsin S exons 2 to 5 were amplified from genomic DNA and sequenced by cycle sequencing by means of an ABI 377 sequencer. For analysis of the remaining coding region, total lymphoblast RNA was extracted (RNazol, Tel Test) and first-strand cDNA was synthesized with the use of an oligo(dT) primer (Gibco BRL). Cathepsin S cDNA nucleotides 631 to

1231 were amplified with gene-specific primers. Sequence comparisons were carried out with the Auto Assembler software package (Applied Biosystems).

10. G.-P. Shi, J. S. Munger, J. P. Meara, D. H. Rich, H. A. Chapman, *J. Biol. Chem.* **267**, 7258 (1992).
11. The R113W substitution in the cathepsin S gene was assessed in genomic DNA by amplification of a 150-base pair product containing exon 4 and the adjacent intron 4 sequence with the use of antisense (5'-GCATTTAAAGAGCTCTACCTAGGGT-3') and sense (5'-CAGAGAAATATCAGATATAAGTCAAC-CCCAAT-3') primers; the latter introduced a T-to-C change (underlined) that created a Mun I restriction site in the allele with the sequence variation. The amplified product was isolated (Qiagen), digested, and analyzed on a 2% agarose gel.
12. Cathepsin S enzymatic activity was assayed in lymphoblast lysates by fluorimetry with the substrate Z-Val-Val-Arg-NHMe (Bachem Feinchemikalien AG, Bubendorf, Switzerland) after inactivation of cathepsins B, L, and H at 40°C [H. Kirschke and B. Wiederanders, *Methods Enzymol.* **244**, 500 (1994)]. Enzymatic activity of the control lysosomal enzyme, α -galactosidase A, was determined by fluorimetry [D. F. Bishop and R. J. Desnick, *J. Biol. Chem.* **256**, 1307 (1982)].
13. G.-P. Shi et al., *FEBS Lett.* **357**, 129 (1995); D. Bromme and K. Okamoto, *Biol. Chem. Hoppe-Seyler* **376**, 379 (1995); K. Tezuka et al., *J. Biol. Chem.* **269**, 1106 (1994); T. Inaoka, et al., *Biochem. Biophys. Res. Commun.* **206**, 89 (1995); Y.-P. Li et al., *J. Bone Miner. Res.* **10**, 1197 (1995). Cathepsin K (E.C. 3.4.22.38) was also designated cathepsin OC2, cathepsin X, cathepsin O, and cathepsin O2 by the above investigators who originally and independently identified this gene.
14. Cathepsin K cDNA was amplified by RT-PCR from Pycno lymphoblast total RNA. For PCR amplification, cathepsin K sense and antisense primers corresponding to the 5' and 3' untranslated regions (5'-GCCGCAATCCCGATGGA-3' and 5'-CCTTGAGG-ATATTGAAGGGAAGTCTAG-3', respectively) were synthesized. Nested sense and antisense PCR primers (5'-CCCCTGATGGTGTGCCCA-3' and 5'-CCCTTCCAAAGTGCATCGTTACACT-3', respectively) were used to reamplify the cathepsin K cDNA from the initial PCR product. Products were isolated and cycle-sequenced in both orientations.
15. We assayed the putative cathepsin K mutation in the affected Israeli Arab Pycno patients in family members and in unrelated normal Arab individuals by amplifying nucleotides 998 to 1170 from genomic DNA using flanking primers (5'-GGGAGAGAAACT-GGGGAAACA-3' and the antisense primer from the nested PCR) and then digesting with Hinf I (New England Biolabs).
16. H. C. Dietz et al., *Science* **259**, 680 (1993).
17. H. C. Blair, P. H. Schlesinger, F. P. Ross, S. L. Teitelbaum, *Clin. Orthop. Relat. Res.* **294**, 7 (1993); T. Goto, *Histochemistry* **101**, 33 (1994); K. Debari, T. Sasaki, N. Udagawa, B. R. Rifkin, *Calcif. Tissue Int.* **56**, 566 (1995); V. Everts, W. Beertsen, R. Schroder, *ibid.* **43**, 172 (1988); V. Everts et al., *J. Cell. Physiol.* **150**, 221 (1992); H. Kakegawa et al., *FEBS Lett.* **321**, 247 (1993).
18. F. H. Drake et al., *J. Biol. Chem.* **271**, 12511 (1996).
19. D. Bromme et al., *ibid.*, p. 2126.
20. R. J. Kaufman and P. Murtha, *Mol. Cell. Biol.* **7**, 1568 (1987).
21. J. S. Munger et al., *Biochem. J.* **311**, 299 (1995).
22. We thank J. Zhang for technical assistance; P. Maroteaux and D. Wilson for providing access to the Moroccan and American Pycno families, respectively; and S. J. Weiss for assistance in preparing cathepsin K antibodies. Supported in part by National Institutes of Health grants 5R37 DK34045 and R01 DK31775 (to R.J.D.) and R01 HL44816 (to H.A.C.), a grant from the National Center for Research Resources for the Mount Sinai General Clinical Research Center (5 M01 RR00071), a Basil O'Connor Starter Scholar Award from the March of Dimes (B.D.G.), an American Heart Association grant-in-aid (B.D.G.), a grant for the Mount Sinai Child Health Research Center (5 P30 HD 28822), and the Mount Sinai Children's Center Fund.

15 May 1996; accepted 24 July 1996