

**Supplementary material for  
The HIRAN domain and recruitment of chromatin  
remodeling and repair activities to damaged DNA**

**Lakshminarayan M. Iyer, M. Madan Babu and L. Aravind\***

National Center for Biotechnology Information, National Library of Medicine,  
National Institutes of Health, Bethesda, MD 20894, USA

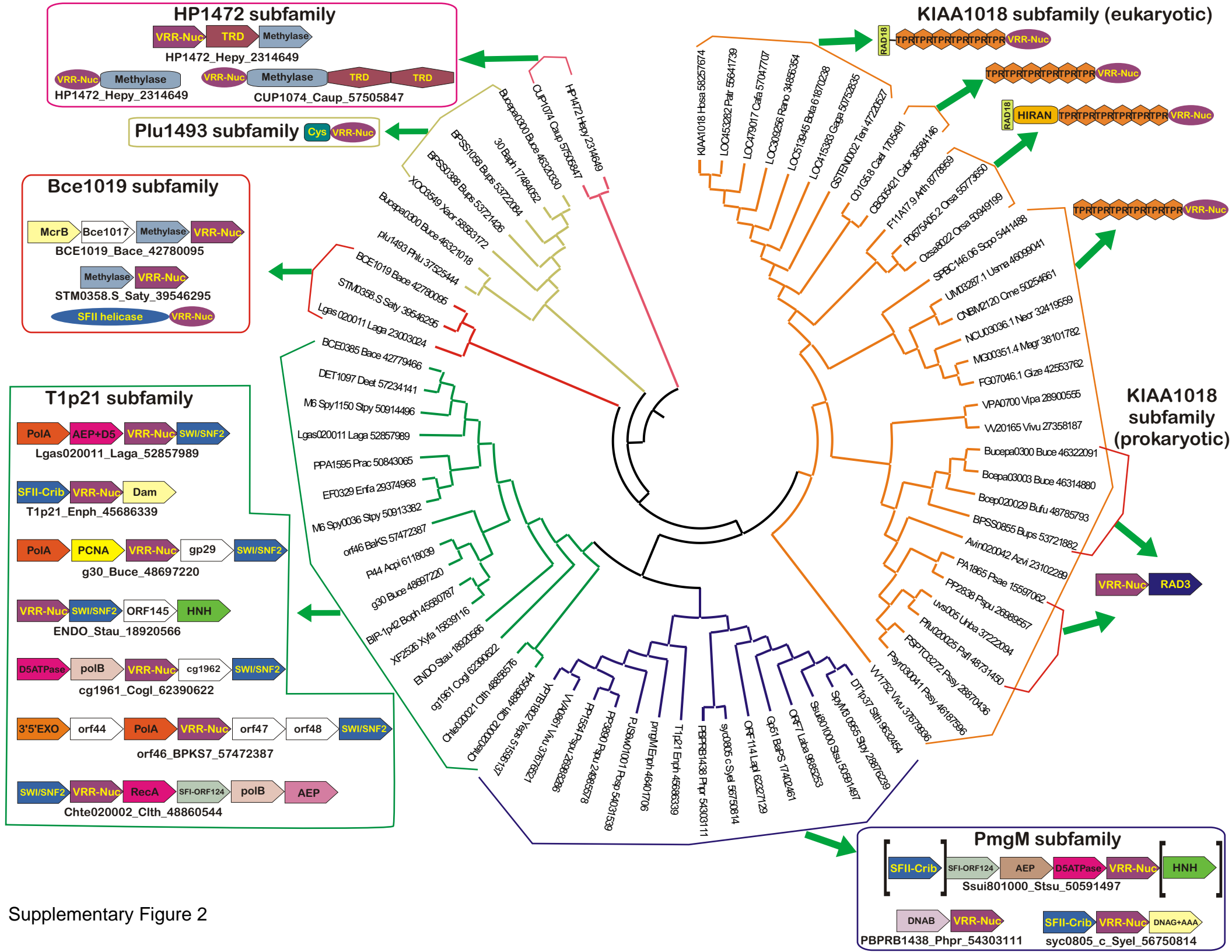
\*For correspondence:  
Email: [aravind@ncbi.nlm.nih.gov](mailto:aravind@ncbi.nlm.nih.gov)  
Telephone: (301) 594-2445  
Fax: (301) 480-9241

## Figure legends

**Supplementary Figure 1:** Multiple alignment of the members of the VRR-Nuc endonuclease family. The alignment shows the conservation of selected set of amino acids in the VRR-Nuc endonuclease from distantly related species. The different sub-families are: (a) T1p20, (b) PmgM, (c) KIAA1018, (d) BCE1019, (e) HP1472 and (f) plu1493 subfamily. The active site residues: Aspartic acid, Glutamic acid, Lysine and Glutamine are completely conserved in the different members and are highlighted accordingly. Proteins are denoted with their gene names, species abbreviation and gi numbers. The coloring reflects the amino acid conservation at 85% consensus. The consensus abbreviations and coloring scheme are as follows: h: hydrophobic residues (L,I,Y,F,M,W,A,C,V), l: aliphatic (L,I,A,V) and c: charged (K,E,R,D,H) residues, +: basic (K/R/H) residues, -: acidic (D,E) residues, and p: polar (S,T,E,C,D,R,K,H,N,Q) residues colored purple; s: small (S,A,C,G,D,N,P,V,T) and u: tiny (G,A,S) residues, colored green; b: big (L,I,F,M,W,Y,E,R,K,Q) residues shaded gray. Secondary structure assignments are as follows: H: Helix, E: Extended (Strand). Species abbreviations are as follows: Atha : Arabidopsis thaliana; Avin : Azotobacter vinelandii; BP12 : Staphylococcus aureus phage phi 12; BP315.2 : Streptococcus pyogenes phage 315.2; BPA PSE-1 : Acyrthosiphon pisum bacteriophage APSE-1; BPBIP-1 : Bordetella phage BIP-1; BPBcepC6B : Burkholderia cepacia complex phage BcepC6B; BPDT1 : Streptococcus thermophilus bacteriophage DT1; BPKS7 : Bacteriophage KS7; BPP1 : Enterobacteria phage P1; BPSA : Bacteriophage PSA; BPT1 : Enterobacteria phage T1; BPphi31 : Lactococcus bacteriophage phi31; BPphiE125 : Bacteriophage phiE125; BPphiJL-1 : Lactobacillus plantarum bacteriophage phiJL-1; Bcep : Burkholderia cepacia; Bcer : Bacillus cereus; Bfun : Burkholderia fungorum; Bpse : Burkholderia pseudomallei; Btau : Bos taurus; Cbri : Caenorhabditis briggsae; Cele : Caenorhabditis elegans; Cfam : Canis familiaris; Cglu : Corynebacterium glutamicum; Cneo : Cryptococcus neoformans; Cthe : Clostridium thermocellum; Cups : Campylobacter upsaliensis; Deth : Dehalococcoides ethenogenes; Efae : Enterococcus faecalis; Ggal : Gallus gallus; Gzea : Gibberella zeae; Hpyl : Helicobacter pylori; Hsap : Homo sapiens; Lgas : Lactobacillus gasseri; Mgri : Magnaporthe grisea; Ncra : Neurospora crassa; Osat : Oryza sativa; Pacn : Propionibacterium acnes; Paer : Pseudomonas aeruginosa; Pflu : Pseudomonas fluorescens; Plum : Photobacterium luminescens; Ppro : Photobacterium profundum; Pput : Pseudomonas putida; Psp. : Polaromonas sp.; Psyr : Pseudomonas syringae; Ptro : Pan troglodytes; Rnor : Rattus norvegicus; Spom : Schizosaccharomyces pombe; Spyo : Streptococcus pyogenes; Ssui : Streptococcus suis; Styp : Salmonella typhimurium; Telo : Thermosynechococcus elongatus; Tnig : Tetraodon nigroviridis; Ubac : uncultured bacterium; Umay : Ustilago maydis; Vpar : Vibrio parahaemolyticus; Vvul : Vibrio vulnificus; Xfas : Xylella fastidiosa; Xory : Xanthomonas oryzae; Ypse : Yersinia pseudotuberculosis.

**Supplementary Figure 2:** Neighbor-joining tree, domain organization and operon structure of the VRR-Nuc family of nucleases. Proteins are denoted by their gene names, species abbreviations and gi numbers as in Figure 1. Branches of the different VRR-Nuc subfamilies are represented in distinct colors in the neighbor-joining tree; cg1961: green, PmgM: blue KIAA1018: orange, BCE1019: red, HP1472: pink, and plu1493: Yellow. Operons are shown as boxed arrows with the arrow head pointing from the 5' end to the 3' end of the coding sequence. Operons and domain architectures without any names below them are shared by all the proteins in the referring bracket. Some non-standard domain abbreviations are explained below. SFIIcrib: superfamily II helicase with a conserved zinc ribbon at the C-terminus, polA: DNA polymerase of the polA superfamily, RAD18: RAD18-like zinc finger, DNAG+AAA: DNAG like primase fused to an AAA+ ATPase, cys: cysteine rich domain N-terminal to the plu1493 like proteins and TRD: Target recognition domain. With the exception of the polA protein in Bacteriophage KS7, these proteins have domain architectures similar to Poll. AEP+D5: Archaeo-eukaryotic primase fused to a D5 ATPase. Gene names enclosed in white boxed arrows represent poorly characterized genes.





Supplementary Figure 2