a0005

Hypervariable Region

F González-Candelas and FX López-Labrador, Universitat de València, Valencia, Spain

Glossary

d0005 **Antigens** Molecules or substances capable of triggering the production of antibodies by the immune system.

d0010 **Frequency-dependent selection** It occurs when selection acts differentially on genetic variants depending on their relative frequency in the population.

d0015 **Gene family** A set of genes present in several copies in a genome with related function and common origin.

d0020 **Homology** Relationship among genes, structures, or organisms resulting from shared ancestry.

d0025 **Positive selection** It occurs when selection acts favoring a beneficial variant over less-fit ones.

d0030 **Red-Queen situation** Image taken from Lewis Carrol's "Through the looking glass" in which the Red Queen keeps running just to stay in the same place. In biology this analogy is invoked when continuous changes are necessary to achieve survival and adaptation in the face of changing environmental conditions.

d0035 **Stem-loop structures** Paired and nonpaired portions of the secondary structure of RNAs. The pairing is due to complementarity of nucleotides in the same RNA strand.

d0040 **Synonymous/nonsynonymous substitutions** Changes at the nucleotide level in a protein-coding gene that do not or do alter the amino acid encoded in the corresponding codon.

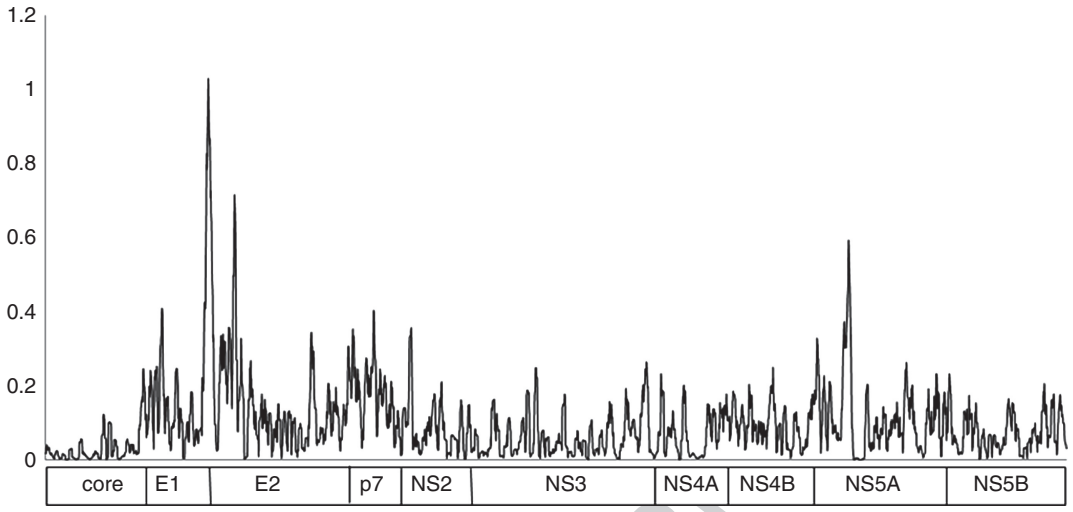
p0045 Hypervariable regions (HVRs) are contiguous portions in the genome or protein sequence of a species whose level of intraspecific variability is substantially higher than those of other regions. HVRs have been defined in all the domains of life, from viruses to eukaryotes. There is no absolute measure of variation to establish whether a given region is hypervariable and this designation is highly dependent on the level of knowledge and historical tradition in the study of the corresponding organisms. HVRs are important not only at the functional level, but also to discriminate among very closely related individuals yet preserving information on their relatedness. Thus, they can be very informative about the action of different evolutionary forces in organisms.

p0050 The first HVRs described and characterized were those in human immunoglobulins (Ig's), which are composed of light and heavy polypeptide chains, both including constant (C) and variable (V) regions. Variable regions include framework regions and short HVR stretches of around 120 amino acids that constitute the antibody binding sites for specific antigens at the ends of both the light and heavy chains (complementary determining regions or CDRs). Framework regions that have more stable, although also variable, amino acid composition separate the three CDRs and help in positioning them to contact the epitope of the antigen. Genetic mechanisms have evolved to finally produce up to about 1×10^{11} Ig molecules with different specificities from a somewhat modest number of genes. For their expression, these genes are reorganized from a library of gene segments located in separate chromosomes that, after additional recombination and mutations, generate such enormous diversity. Antigen receptors on the surface of B- and T-lymphocytes are based on (B-cells) or similar to (T-cells) immunoglobulin molecules and therefore multiple B- and T-lymphocyte cell clones with different antigen specificities can also be generated due to different amino acid composition in the CDRs. Variable domains of Ig's and T-cell receptors (TCRs) share sequence homology and striking structural similarity. A particular clone of T-cells also displays specificity to a particular epitope of a given antigen, and a high diversity of TCRs is also generated by means of combinatorial events between TCR genes and three HVRs (CDRs) positioned to contact antigens associated with major histocompatibility complex proteins.

p0055 Although HVRs are frequently associated with genes and proteins of the immune system, many other similar regions have been defined in eukaryotes. Hypervariability in gene families also appears frequently in systems evolved to recognize, or avoid recognition by, foreign molecules, be they venom-derived toxins or antigenic parasite surface proteins. Venom proteins in *Conus* snails (conopeptides) have evolved in highly variable gene families, thus facilitating their specificity against different preys. The variant surface glycoproteins (VSGs) of *Trypanosoma brucei*, the causative agent of African sleeping sickness, interact with the host's immune system alternating serologically different dominant cell-surface antigens by changing which VSGs are expressed. Other eukaryotic parasites, such as *Plasmodium* and *Babesia* spp., display surface antigens with HVRs that contribute to antigenic diversity resulting in immune evasion.

p0060 Portions of the ribosomal RNA (rRNA) genes are often referred to as HVRs and the same is true for the control region of animal mitochondrial DNA (mtDNA) as well as for many regions with tandemly repeated units (mini- and microsatellites). While rRNA genes are frequently used for species identification in genetic barcoding schemes, other HVRs are useful markers in within-species analyses of individual identity and population differentiation.

p0065 Bacterial and archaeal HVRs in the 16S rRNA genes are one of the main targets of next-generation sequencing approaches to characterize microbial biodiversity from uncultured samples. Additional HVRs in bacteria are found in genes and proteins involved in interspecific interactions. For instance, the multiprotein Dot/Icm type IV secretion system is the major virulence factor of intracellular pathogens such as *Legionella pneumophila* and *Coxiella burnetii*. It includes a large family of hypervariable genes (*fir*, functional homologues of *icmR*) which, despite not being homologous, share conserved



f0005 **Figure 1** Variabilityin the HCV polyprotein. The average number of nonsynonymous substitutions in each codon is shown for HCV-1 ($n = 91$) isolates taken from the HCV database at Los Alamos National Laboratory. HVRs in the E2 (envelope glycoprotein 2) and NS5A proteins are revealed by values of diversity larger than 0.5.

regulatory regions involved in the adaptation to specific environmental hosts. The secretion system is necessary to establish a specific type of vacuole that enables intracellular survival and replication of these bacteria. Many prokaryotes also use HVRs to defend themselves from their natural enemies. About 90% of the Archaea and 40% of the Bacteria harbor arrays of conserved direct DNA repeats (clustered, regularly interspaced, short palindromic repeats (CRISPRs)) interspersed by highly variable spacers. These spacers match sequences in bacteriophage genomes and transmissible plasmids, and they have been shown to provide protection against their invasion.

p0070 Viruses also frequently harbor HVRs in the coding regions of their genomes, frequently located in domains exposed to the environment. For instance, in the hepatitis C virus (HCV), there are three HVRs located in the envelope glycoprotein 2 and in the human immunodeficiency virus-1 (HIV-1), the most variable region (V3), is also found in the envelope, within the *env* gene (coding for gp120). These envelope regions contain antibody neutralization domains and epitopes for T-cell recognition and are involved in cell tropism. Another group of HVRs in viruses are those found in noncoding regions, usually at either the 5′ or 3′ ends of the genome. These noncoding HVRs are highly divergent as a result of substitutions, deletions, and insertions during replication cycles, normally within RNA stem-loop structures, which might play a role in the evolution, virulence, and pathogenicity of viruses such as avian infectious bronchitis virus (IBV), African swine fever virus (ASFV), tick-borne encephalitis virus (TBEV), or dengue virus (DENV).

p0075 Practically, in all organisms, high levels of variability appear essentially at two levels, functional and structural, which roughly correspond to variation generated within individuals and within populations. The mechanisms responsible for the generation and maintenance of variation in these two levels are necessarily different. Functional variability is generated by mechanisms such as alternative splicing, RNA editing, or module recombination, resulting in the expression of a more or less limited set of variants from a wide range of possibilities. Offspring usually inherit the capability of generating variation but not a limited set of the variants.

p0080 Structural variation can be generated by three main mechanisms, nucleotide substitutions, changes in the number of repeats, and recombination hotspots, and all of them are frequently found in association with HVRs. At the gene or genome level, mutation rates are not evenly distributed throughout the genome and there are some evidences that genetic variability is generated at higher frequency in HVRs than in other portions of the genome. In RNA viruses, increased mutation rates have been found associated with regions with low levels of secondary structure, while homopolymer runs and repetitive motifs increase the frequency of slippage-and-mispairing errors during replication. The same mechanism is the basis for the generation of variation in eukaryotic microsatellite regions. Similar structural features are found in association with recombination hotspots, but their dynamics can be very complex. For instance, human chromosomes are rich in recombination hotspots containing the motif CCTCCCT. A single transition T→C in the third position reduces the frequency of recombination and, in consequence, acts as a 'suppressor' mutation.

p0085 At the population level, genetic hypervariability results from two different processes, generation and maintenance. The latter is a more important contributor than the former to the total variation found at any instant in HVRs at the population level. Evolutionary genetics provides several clues for understanding how the generated variation can be maintained in a population. The number of variants and their frequencies usually departs from the expected values under the neutral theory of molecular evolution. Consequently, selection is thought to play a major role in the maintenance of genetic variability in HVRs. Heterosis, disruptive/diversifying selection, frequency-dependent selection, and selection in heterogeneous environments, both spatially and temporally, are the main selection regimes that can maintain genetic variation when acting in a finite population. All of them may act on HVRs, but their relative contributions to each specific region are largely unknown. The mutual feedback at the molecular level between hosts and parasites or predators leads to a Red Queen situation in which changes in the receptors of the

hosts are favored and, subsequently, changes in the recognition sequences of the parasite become advantageous. Negative frequency-dependent selection results from the advantage of the 'rare types' in hosts' populations; when a genotype increases its frequency in the population, it will be more easily targeted by parasites or predators capable of recognizing it and its advantage will diminish. These cycles contribute to the maintenance of variation in both host and parasite populations.

p0090 At an evolutionary timescale, HVRs in protein-coding regions are characterized by increased rates of nonsynonymous substitutions over synonymous changes. This results from the selection processes described previously which depend on changes at the amino acid sequence level. Tests based on the ratio between dN and dS, the relative rates of nonsynonymous and synonymous substitutions, respectively, will identify HVRs as regions evolving under strong positive selection.

p0095 Recent advances in sequencing technology and methods for analyzing the immense amount of data generated will certainly lead to the identification and further characterization of HVRs in practically all organisms. The few exceptions to this general pattern are represented by some bacterial pathogens (*Yersinia pestis*, *Bacillus anthracis*, and *Mycobacterium tuberculosis*, among others) and by species that have experienced intense and prolonged reductions of their population sizes, such as cheetahs or Speke's gazelles. The detailed analysis of HVRs in most other species will certainly provide important information to understand their relationships with other species, their physiology, and their evolution.

AU7 *See also:*

Further Reading

bib0005 (2010) Chapter 8: Immunology. In: Brooks GF, Carroll KC, Butel JS, and Morse SA (eds.) *Jawetz, Melnick, & Adelberg's Medical Microbiology*, 25th edn. AU8 McGraw-Hill.

bib0010 Achtman M (2008) Evolution, population structure and phylogeography of genetically monomorphic bacterial pathogens. *Annual Review of Microbiology* 62: 53–70.

bib0015 Blasco R, de la Vega I, Almazan F, Aguero M, and Viñuela E (1989) Genetic variation of african swine fever virus: Variable regions near the ends of the viral DNA. *Virology* 173: 251–257.

bib0020 Bowen DG and Walker CM (2005) Adaptive immune responses in acute and chronic hepatitis C virus infection. *Nature* 436: 946–952.

bib0025 Budowle B, Allard MW, Wilson MR, and Chakraborty R (2003) Forensics and mirochondrial DNA: Applications, debates, and foundations. *Annual Review of Genomics and Human Genetics* 4: 119–141.

bib0030 Chakravorty S, Helb D, Burday M, Connell N, and Alland D (2007) A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *Journal of Microbiological Methods* 69: 330–339.

bib0035 Conticello SG, Gilad Y, Avidan N, *et al.* (2001) Mechanisms for evolving hypervariability: The case of conopeptides. *Molecular Biology and Evolution* 18: 120–131.

bib0040 Feldman M and Segal G (2007) A pair of highly conserved two-component systems participates in the regulation of the hypervariable FIR proteins in different *Legionella* species. *The Journal of Bacteriology* 189: 3382–3391.

bib0045 Field MC and Boothroyd JC (1996) Sequence divergence in a family of variant surface glycoprotein genes from trypanosomes: Coding region hypervariability and downstream recombinogenic repeats. *Journal of Molecular Evolution* 42: 500–511.

bib0050 Jeffreys AJ, Neumann R, Panayi M, Myers S, and Donnelly P (2005) Human recombination hot spots hidden in regions of strong marker association. *Nature Genetics* 37: 601–606.

bib0055 McMichael AJ and Phillips RE (1997) Escape of Human Immunodeficiency Virus from immune control. *Annual Review of Immunology* 15: 271–296.

bib0060 Myers S, Bottolo L, Freeman C, McVean G, and Donnelly P (2005) A fine-scale map of recombination rates and hotspots across the human genome. *Science* 310: 321–324.

bib0065 Shurtleff AC, Beasley DWC, Chen JJY, *et al.* (2001) Genetic variation in the 3′ non-coding region of Dengue viruses. *Virology* 281: 75–87.

bib0070 Stoneking M (2000) Hypervariable sites in the mtDNA control region are mutational hotspots. *The American Journal of Human Genetics* 67: 1029–1032.

bib0075 Williams AK, Wang L, Sneed LW, and Collisson EW (1993) Analysis of a hypervariable region in the 3′ non-coding end of the infectious bronchitis virus genome. *Virus Research* 28: 19–27.

Relevant Websites

bib0080 http://www.hvrbase.org – HvrBase++. Database of HVRs in human and other primates mitochondrial genomes.

bib0085 http://www.hiv.lanl.gov – HIV databases. This website contains data on HIV genetic sequences, immunological epitopes, drug-resistance associated mutations, and vaccine trials.

bib0090 http://www.nlm.nih.gov – National Library of Medicine. Search for 'hypervariable' in the Medical Subject Headings to find relevant information of HVRs.

bib0095 http://tree.bio.ed.ac.uk – RNA Virus Database. It contains data and bioinformatics tools for almost 1000 viruses.

bib0100 http://vamps.mbl.edu – VAMPS (Visualization and Analysis of Microbial Population Structures). It includes specific databases of microbial SSU rRNA and the specific hypervariable regions therein.

Biographical Sketches

Fernando González-Candelas, obtained his PhD from the University of Valencia, Spain. He is Professor of Genetics at the University of Valencia (Instituto Cavanilles de Biodiversidad y Biología Evolutiva) and Senior Scientist at the Centro Superior de Investigación en Salud Pública (CSISP), Valencia, Spain. He is also a member of the CIBERESP (Network Center for Biomedical Research in Epidemiology and Public Health, ISCIII, Spain). His main research topics are evolutionary biology, molecular epidemiology, evolution of viruses and bacteria, bioinformatics, molecular evolution, conservation genetics, and population genetics.

F. Xavier López-Labrador obtained his PhD in microbiology from Hospital Clínic and Autonomous University of Barcelona. He is Senior Investigator of Spain's National Healthcare System (ISCIII-FIS) at Department of Genomics and Health, Centro Superior de Investigación en Salud PúblicaCentre AU9 Superior d'Investigació en Salut Pública (CSISP), Valencia, Spain. He is Associate Professor of Microbiology at the Medical School of the University of Valencia. He is also a member of the CIBERESP (Network Center for Biomedical Research in Epidemiology and Public Health, ISCIII, Spain). His main research topics are viral diagnostics, genotyping, genetic variation and antiviral resistance, hepatitis viruses, viral immunology, molecular epidemiology, and markers of disease progression.