Virus evolution

As new mass sequencing techniques reveal that the diversity of viruses is much greater than ever imagined, **Peter Simmonds** shows that the latest discoveries of new viruses are providing some clues to how viruses evolved.

In the spirit of Darwin 200, which marks the bicentenary of Charles Darwin's birth, I will describe a little of what we know about the evolution of viruses and their ultimate origins. One of the immediate problems facing such studies is the evident fact that viruses are hugely diverse in size, appearance, even the nature of their genetic material (DNA or RNA). From this, it is reasonably clear that they are a not a single evolutionary group, and cannot be easily added as a single unit to the tree of life with its three main divisions (*Bacteria*, *Archaea* and *Eukarya*). By the same token, it seems likely that different virus groups (e.g. animal RNA viruses, retroviruses, large DNA viruses, bacteriophages) may indeed have entirely separate evolutionary origins.

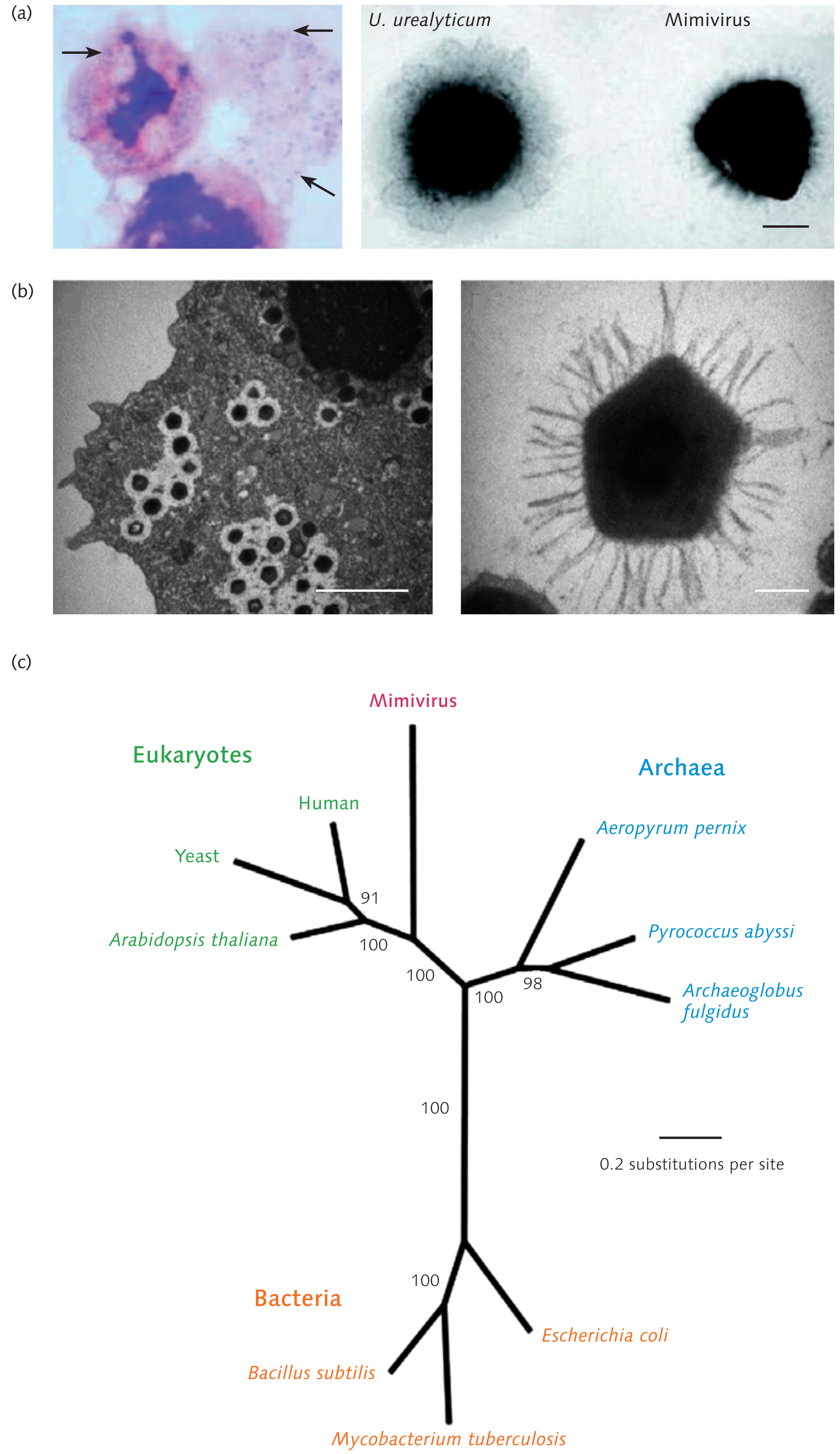
In this short article I will describe two areas where recent discoveries have produced tantalizing new insights into the origin and ubiquity of some of these groups. Through the application of new, mass-sequencing techniques and scope for large-scale environmental sampling for virus genomic sequences, we may finally be able to understand the extent and complexity of the 'virosphere' in which we live, and the extraordinary diversity of viruses that infect us.

The tale of a 'Gram-positive' virus

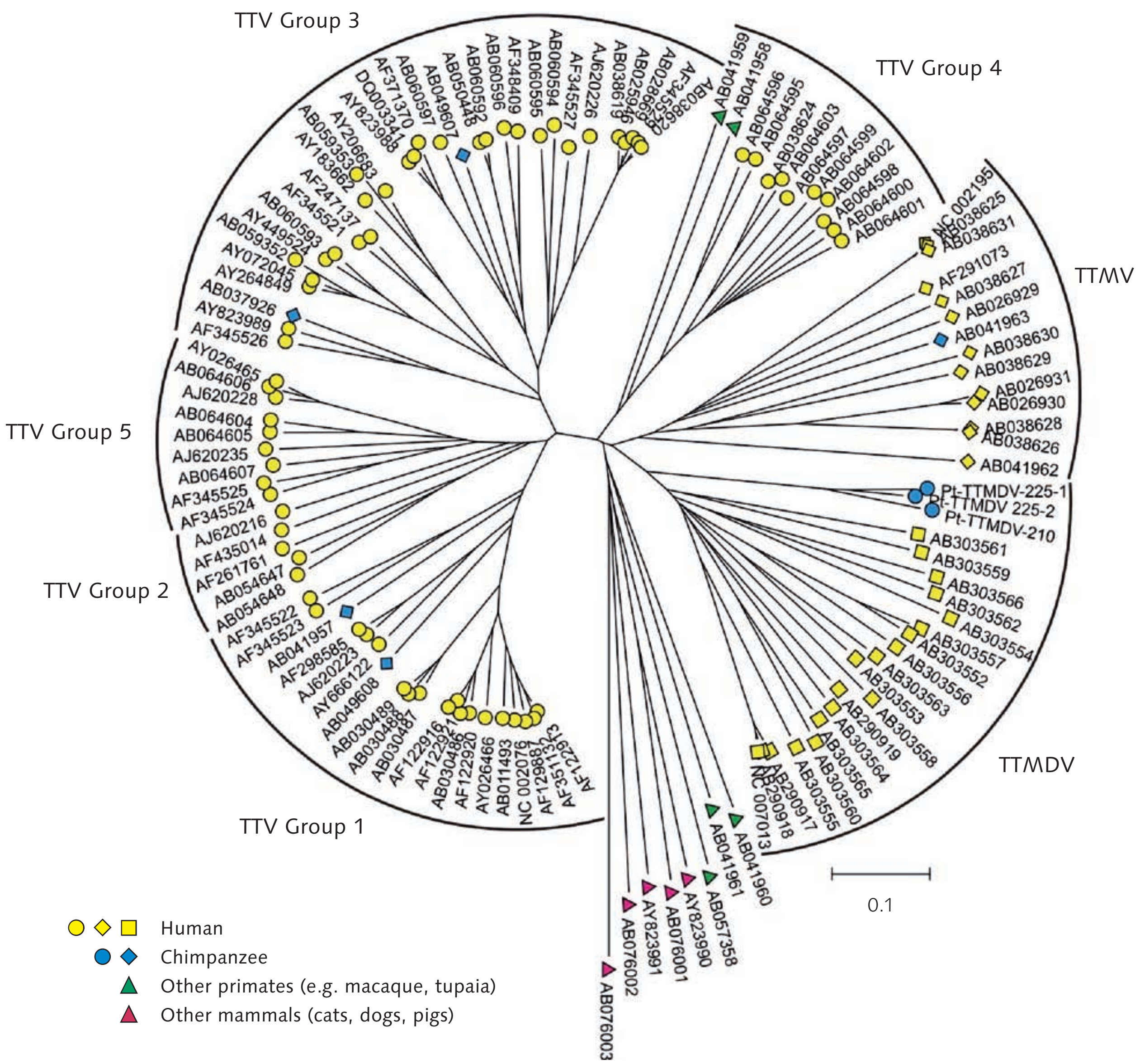
Amoebae growing in a water-cooling tower during the investigation of an outbreak of pneumonia in Bradford, UK, were found to harbour a number of intracellular *Legionella*-like organisms. There the issue might have been forgotten were it not for a attempt several years later by La Scola and colleagues in Marseilles to catalogue the bacteria in these and other environmental samples collected worldwide by ribosomal RNA (rRNA) sequencing. Puzzlingly, one of the Gram-positive amoeba-infecting bacteria with a diameter of around 500–700 nm (Fig. 1a) proved resistant to attempts to amplify the rRNA genes. Fortunately for future studies, the decision to sequence its genome revealed the presence of an entirely novel, large virus-like genome, 1.2 Mb in size and containing over 900 genes, far larger than many parasitic bacteria such as *Rickettsia*, *Ureaplasma* and some *Legionella* species with which it was first confused (Fig. 1a). Remarkably, the genome of what is now called mimivirus ('microbe-mimicking' virus) contained many genes involved in translation, such as amino acyl-tRNA synthetases, translation initiation, elongation and peptide release factors not found in other viruses, as well as a huge complement of genes (76 %) encoding proteins with no identifiable homology and unknown function.

What are mimiviruses? While they breach the size and complexity barrier that traditionally separates viruses and bacteria (Fig. 1a, b), they clearly aren't bacteria. Mimiviruses contain only seven genes homologous to a set of 61 core genes shared among bacteria, archaea and eukaryotes, and of these three, mimiviruses are actually most similar to those of eukaryotes (Fig. 1c). Remarkably though, they adopt a phylogenetic position ancestral to the split of the animal, plant and other component kingdoms of eukaryotes, implying an extraordinarily long evolutionary existence, possibly reaching back as far as the origin of life. Mimiviruses additionally lack ribosomal genes, an absence universal to viruses and which confine them to their obligate parasitic life cycle. Their genomes are linear with covalently closed ends, comparable to other large DNA viruses, such as asfarviruses (e.g. *African swine fever virus*) and poxviruses (e.g. smallpox virus), and distinct from the circular genomes of most bacteria and archaea. Finally, they replicate in the infected cell through construction of multiple virus assembly structures rather than by binary fission. This latter feature seems to preclude the existence of free-living mimiviruses in nature that might retain a greater proportion of ribosomal and other core genes required for autonomous replication.

A final remarkable feature of mimiviruses is their pervasive presence in the environment. Large-scale metagenomic sampling has revealed a plethora of viruses related to mimiviruses in seawater at frequencies that rival bacteriophages in detection frequency. What these viruses infect and whether there are indeed free-



▶ Fig. 1. (a) Left: mimiviruses (arrows) infecting the amoeba *Acanthamoeba polyphaga*, resembling Gram-positive particles. Right: comparison of the size of mimivirus with the intracellular bacteria *Ureaplasma urealyticum. Reproduced from Suzan-Monti* et al. *(2006) (see 'Further reading' for details) with permission from Elsevier.* (b) Left: appearance of mimiviruses in the cytoplasm of infected amoebae cells (bar, 2 μm. Right: electron micrograph of the non-enveloped icosahedral mimivirus virion surrounded by fibrils (bar, 100 nm). (c) Phylogenetic relationship of core genes of representative eukaryotes, bacteria and archaea with homologues in mimiviruses, showing their ancestral position relative to eukaryotes. *(b, c) Reproduced with permission from La Scola* et al. *(2003) (see 'Further reading' for details) with permission from AAAS*



▲ Fig. 2. Diversity of human and other mammalian TTV-like viruses (family *Anelloviridae*). The phylogenetic tree of ORF1 sequences reveals a large number of divergent types and groups (e.g. TTV, TTMDV, TTMV) as well as more divergent viruses infecting other primates and mammalian species. The greater number and diversity of TTV-like viruses found in humans (yellow circles) compared to other primates (blue) and other mammals (red) is undoubtedly the simple outcome of the greater attention being paid to human samples. Similar or probably greater diversity of TTV-like viruses certainly exists in these other species. *Adapted from Ninomiya* et al. *(2009) in* J Gen Virol *(see 'Further reading' for full details)*

living relatives built on the same structural plan are vital research questions for the future, as is the nature of the huge complement of unidentifiable gene sequences in the sea that probably belong to further uncharacterised virus groups and families. Because of their large size, filtering seawater for bacteria in metagenomic analyses doesn't produce only bacterial sequences.

The transition of the mimivirus from its humble, unsuspected origin in a water tank in the Midlands to one of the most abundant life-forms on the planet in the space of 4 years is a stark demonstration of how little we really know of viral diversity and evolution.

Unsuspected human viruses

The development and recent spectacular success of metagenomics and methodologically related molecular-based virus discovery methods has also greatly expanded the number and pace of discovery of human and veterinary viruses. The wealth of data emerging about the diversity of viruses and

the unsuspected existence of a range of newly discovered parvoviruses (human bocavirus, PARV4), picornaviruses, paramyxoviruses, coronaviruses (including SARS virus), three new polyomaviruses and several new anelloviruses again highlights our current ignorance of the true size of the 'virosphere' infecting humans and animals.

One of the remarkable characteristics of these newly discovered viruses is their frequent wide distribution in human and animal populations. In the picornavirus family, past or current infection with Saffold virus (SAFV), a newly discovered member of the genus *Cardiovirus*, was detected in >95% of children by the age of two, even though the virus itself has remained undetectable by conventional virus isolation methods applied over the last 50 years and whose actual existence was entirely unsuspected until 2 years ago. *Cosavirus*, a newly discovered genus of human picornaviruses, with dozens of serotypes and several species, which is likely to be as diverse genetically and clinically as human enteroviruses, has been detected in 50% or more of faecal samples from young children in locations all over the world (Pakistan, Nigeria, Egypt). Similarly, infections with TTV and related small DNA viruses, collectively termed anelloviruses, are ubiquitous in human, primate and probably all mammalian species, a finding more surprising for their persistence and ongoing lifelong replication in the lymphoreticular system (including bone marrow), and multiple infection in each person with a plethora of different types and species (Fig. 2). A wealth of further small, persistent and ubiquitous DNA viruses undoubtedly await to be discovered in the coming years.

The frequent ubiquity and non-pathogenicity of many of the recently discovered viruses may indeed be a particular attribute that has hindered their identification in the past. Most virus isolation and detection attempts have been necessarily clinically driven, aimed at identifying causes of various diseases with a suspected or demonstrated infectious aetiology, SARS and AIDS being two recent examples. Secondly, and particularly for DNA viruses, tight control of replication is a commonly used strategy to avoid immune recognition and elimination, but also one that hinders their detection and diagnosis. In the future, virus discovery programmes, such as those based on massively parallel sequencing of RNA and DNA in samples and tissues will reveal much more about the viruses that infect humans, their ecology and host adaptation, and, for some, their role in currently unexplained human and veterinary diseases.

These examples of the very large and the very small viruses give a flavour of their unexpected pervasiveness in the environment and their diversity. Viruses are numerically the most abundant and the most genetically diverse organisms on earth. With their short generation times, often huge population sizes and rapid sequence change over time, viruses additionally provide us with the opportunity to directly study evolution in action as they colonize diverse environments and adapt to new hosts.

Peter Simmonds
Centre for Infectious Diseases, University of Edinburgh, Summerhall, Edinburgh, EH9 1QH
(t 0131 650 7927; e peter.simmonds@ed.ac.uk)

Further reading

Claverie, J.M., Ogata, H., Audic, S., Abergel, C., Suhre, K. & Fournier, P.E. (2006). Mimivirus and the emerging concept of 'giant' virus. *Virus Res* 117, 133–144.

Ghedin, E. & Claverie, J.M. (2005). Mimivirus relatives in the Sargasso sea. *Virol J* 2, 62.

Jones, M.S., Lukashov, V.V., Ganac, R.D. & Schnurr, D.P. (2007). Discovery of a novel human picornavirus in a stool sample from a pediatric patient presenting with fever of unknown origin. *J Clin Microbiol* 45, 2144–2150.

Kapoor, A., Victoria, J., Simmonds, P., Slikas, E., Chieochansin, T., Naeem, A., Shaukat, S., Sharif, S., Alam, M.M., Angez, M., Wang, C., Zaidi, S., Shafer, R.W., Zaidi, S. & Delwart, E. (2008). A highly prevalent and genetically diversified *Picornaviridae* genus in South Asian children. *Proc Natl Acad Sci U S A* 105, 20482–20487.

La Scola, B., Audic, S., Robert, C., Jungang, L., De Lamballerie, X., Drancourt, M., Birtles, R., Claverie, J. M. & Raoult, D. (2003). A giant virus in amoebae. *Science* 299, 2033.

Ninomiya, M., Takahashi, M., Hoshino, Y., Ichiyama, K., Simmonds, P. & Okamoto, H. (2009). Analysis of the entire genomes of torque teno midi virus variants in chimpanzees: infrequent cross-species infection between humans and chimpanzees. *J Gen Virol* 90, 347–358.

Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Suzan, M. & Claverie, J.M. (2004). The 1.2-megabase genome sequence of mimivirus. *Science* 306, 1344–1350.

Simmonds, P. (2002). TT virus infection: a novel virus–host relationship. *J Med Microbiol* 51, 455–458.

Suzan-Monti, M., La Scola, B. & Raoult, D. (2006). Genomic and evolutionary aspects of *Mimivirus*. *Virus Res* 117, 145–155.

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., Wu, D., Paulsen, I., Nelson, K.E., Nelson, W., Fouts, D.E., Levy, S., Knap, A.H., Lomas, M.W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y.H. & Smith, H.O. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304, 66–74.

Zoll, J., Erkens Hulshof, S., Lanke, K., Verdyn Lunel, F., Melchers, W.J.G., Schoondermark-van de Ven, E., Roivainen, M., Galama, J.M. & Kuppeveld, F.J.M. (2009). Saffold virus, a human Theiler's-like cardiovirus, is ubiquitous and causes infection early in life. *PLoS Pathog* 5, doi:10.1371/journal.ppat.1000416.