Published in final edited form as:

Future Virol. 2013 February ; 8(2): 159–171. doi:10.2217/fvl.12.129.

The diversity of human RNA viruses

Mark E.J. Woolhouse and Kyle Adair

Centre for Immunity, Infection, Evolution, University of Edinburgh, Ashworth Laboratories, Kings Buildings, West Mains Rd, Edinburgh EH9 3JT, UK

Abstract

We still cannot answer the very basic question “how many kinds of RNA viruses are there?” even for those that infect humans. It is often suggested that there remains a large number of viruses in humans that we have not yet discovered or recognised, and that there is a much larger and rapidly evolving pool of potential new viruses in mammalian and avian reservoirs that humans are continually being exposed to. However, a careful examination of discovery rates of new human RNA virus species, genera and families challenges this view, raising the possibility that the diversity is much more limited. Moreover, there is some evidence that the cast of human viruses is dynamic, with existing viruses disappearing (at least from humans) and new viruses appearing (perhaps evolving) over timescales of decades. Most of these new viruses, however, remain rare; only a small (but highly significant) minority are capable of spreading extensively through human populations.

Keywords

species; evolution; biodiversity; surveillance; zoonosis; discovery curve; loss; emergence; nucleotide sequences

Virus discovery

The first viruses (then called ‘filterable agents’) were discovered at the end of the 19th century and the first in humans – yellow fever virus – was discovered in 1901 [1]. In the hundred years and more since then we have learnt an enormous amount about the nature of viruses, their biology, the ways that they interact with hosts and how they cause disease. But despite that enormous research effort (tens of thousands of published papers per year) we still do not know how many kinds of virus there are, even in the most-studied host species of all: us. This gap is especially surprising given the huge interest in biodiversity in general and the many methods that ecologists have developed to quantify or estimate it [2]. This paper is intended as a contribution to filling that gap, concentrating mainly on RNA viruses that infect humans.

The rationale for concentrating on RNA viruses is that at least some DNA virus taxa appear to have very different patterns of macroevolution. DNA virus taxa such as the anelloviruses and the papillomaviruses are highly diverse in humans. These viruses, together with some

other groups such as the herpes viruses, are believed to have been present in, evolved in and diversified within humans and human ancestors for many millions of years [3,4]. In contrast, most RNA viruses are likely to have been much more recent acquisitions [5], perhaps infecting humans for only thousands of years or less [6].

For clarity, we need to define “infection” (specifically, in this context, infection of humans). For viruses this is relatively straightforward since they are all obligate parasites, and we can take infection to mean entering and replicating in human cells. (We note that “infection” is not so easily defined for organisms which can replicate outside host cells). In practice, however, virus replication in host cells is not always demonstrated: two examples are virus detection through the use of serology (where replication is implied) and the use of high throughput sequencing (particularly, for example, when applied to gut contents) [7].

We should also note that viruses that can infect humans may not *only* infect humans. Indeed, the majority of “human” viruses occur naturally in non-human host ‘reservoirs’ (other mammals or birds) [8] and humans are no more than ‘incidental’ or ‘spillover’ hosts. Only a minority are capable of persisting in human populations (as endemic infections) or spreading through human populations (as epidemic infections) in the absence of a reservoir. We refer to this subset as ‘human-adapted’ viruses and, of course, it includes those that exert the greatest toll on human health, such as influenza A, measles and HIV-1.

The technologies available for discovering new viruses – in humans or any other host – have developed enormously since the early 20th century [9], progressing from filtration (developed in the 1890s) to complement fixation (1920s), tissue culture (1940s), monoclonal antibodies and electron microscopy (both in the 1970s), polymerase chain reaction (PCR) (1980s) and high throughput sequencing (2000s). Even so, the process of virus detection and identification is still mostly initiated by clinical observation; only quite recently have prospective studies of the diversity of human viruses been attempted and all of these have been relatively small-scale. However, it is widely held that the vast majority of human infections are not reported to medical practitioners and that no aetiological agent is formally identified for most of those that are. So it is unsurprising that we are still discovering new viruses that can infect humans (as is discussed in detail below). The key question, of course, is how many viruses remain to be discovered. We consider some approaches to answering this question below, but we begin by reviewing the taxonomic diversity of known viruses.

Number of species

Viruses are very different from other kinds of living organisms but we traditionally still apply the same approach to their taxonomic classification. The standard unit of taxonomy is the species. Rather than attempt a rigorous definition of a RNA virus “species” here, we simply follow the usage adopted by the International Committee on Taxonomy of Viruses (ICTV) [10]. The main ICTV criteria for defining RNA virus species are: relatedness based on sequence data; serological cross-reactivity; host range; and transmission route. None of these is an unequivocal marker of reproductive isolation, the usual meaning of “species”, and, worse still, they are used inconsistently across different virus taxa. This can lead to instances of ‘splitting’ (defining a high number of species relative to the amount of variation

present) or ‘lumping’ (the opposite) – which can distort the kinds of pattern we will consider later on.

Moreover, a focus on the “species” can mask some very important differences that occur within species. This is perhaps most apparent for the influenza A viruses: from a clinical or epidemiological perspective, seasonal influenza A such as H3N2 and H5N1 (‘bird flu’) or pandemic H1N1 (‘swine flu’) are very different from each other [11], but all are included within a single “species”. Even less variable virus species than influenza A may still contain functionally distinct subtypes with, for example, different antigenicities (i.e. different serotypes) or host preferences.

Despite these limitations, the “species” remains the most useful unit for studying virus diversity currently available and a simple count of species is then the obvious place to begin a review of human RNA virus diversity. A systematic search of the published literature – using a formal methodology and with pre-set inclusion and exclusion criteria (excluding, for example, viruses known to infect humans only as a result of deliberate exposure in the laboratory) – was first attempted in 2001 [12] and has since been updated using improved search tools and taking into account taxonomic revisions [8].

Using this approach we find that 158 RNA virus species as recognised by the ICTV [10] have been reported to infect humans (Table 1). These species are distributed among 47 genera and 17 families (with one genus not currently assigned to a family) (Table 1). For comparison, 91 DNA virus species occur in humans and these are distributed among 22 genera and 8 families.

Only a minority of the 158 human RNA virus species are human-adapted as defined in the previous section. Our best estimate is that 52 (33%) fall into this category, noting that 18 of these retain the ability to infect non-human hosts under natural conditions (and many more have proven capable of infecting non-human hosts in the laboratory). These 52 species are distributed among 27 genera and 11 families. For comparison, 79 (87%) human DNA virus species are human-adapted and these are distributed among 20 genera and 8 families.

Discovery curves

We are still regularly finding and recognising new species of human viruses [8]. Since 2007 at least 20 new “species” of human RNA virus have been proposed (Table 2), although all but two are still awaiting formal recognition by the ICTV. Historical trends in the discovery and recognition of diversity can be captured by plotting the discovery curve, a simple representation of the accumulation through time number of recognised species (or other taxa). Discovery curves are widely used in the ecological literature, but have only recently been examined for viruses [13]. The discovery curve for human RNA viruses is plotted in Figure 1. The curve begins in 1901 and the rate of discovery tends to increase up to the mid-1950s (there is a statistically significant breakpoint at 1954). After that the curve is surprisingly smooth for over half-a-century, despite numerous technological advances (listed above) and inevitable fluctuations in research effort. The mean discovery rate was 2.3 species per year between 1954 and 2006.

A key question is whether or not the rate of discovery is declining – this would be consistent with the idea that the species diversity of human RNA diversity has discernible limits. For species as recognised by the ICTV there is evidence of a fall in the discovery rate by >50% over the period 1954 to 2006 (for square root transformed counts $F_{1,51} = 5.87$, $p = 0.014$). This is an intriguing result as it implies that the majority of human RNA viruses may already have been discovered. This estimate is consistent with our previous studies [8,13] that fitted simple statistical models of the virus discovery process (for both RNA and DNA viruses) using maximum likelihood methods in order to estimate the discovery rate and, crucially, the size of the species pool, i.e. the total number of species that can infect humans. Based on data available in 2005, this approach generated an unbounded upper limit pool of human virus species, but a lower limit which is consistent with most human virus species (>70%) already having been discovered.

Both these projections, however, are sensitive to the inclusion of the recently reported but not formally recognised human viruses discussed above. But even if trends with respect to species should be regarded as provisional, trends with respect to higher taxa are much clearer: the rates of discovery of both new genera and new families show obvious declines (Figure 1). No new human RNA viruses have been found in a new genus since SARS coronavirus (the first human betacoronavirus) in 2003 [15]; and none have been found in a new family since human picobirnavirus (the first recognised member of the picobirnaviridae) in 1988 [15]. One interpretation of these patterns is that the boundaries to human RNA virus diversity (in terms of the number of families) are now largely known. This idea is developed more formally in the following section.

Diversity estimators

Various different estimators of biodiversity, as measured by the number of species, have been developed by ecologists over many years. Recently, Mora *et al.* [2] have suggested a new approach based on the numbers of higher taxa – genera, families, orders etc. – which they used to estimate the total number of species of Animalia. Their method involves fitting a statistical model to the hierarchical relationship between the numbers of higher taxa and assuming that this can be extrapolated to the species level.

Before applying this method to human viruses we first need to check whether the statistical patterns reported for the various taxa within the Animalia are also applicable to human viruses. In Figure 2 we show two representations of the taxonomic data: the numbers of species per genus and the numbers of genera per family. Both of these show a characteristic pattern that most genera (or families) are not very diverse but a small number are highly diverse; this pattern is well modelled by a simple power function [2]. One criticism of this approach is that the higher taxa are human constructs rather than truly natural entities, so any patterns emerging tells us more about the behaviour of taxonomists than about the workings of nature. Nonetheless, for well-studied taxa such as birds and mammals the model is empirically robust [2] and virus taxonomy shows no obvious anomalies compared with the taxonomies of other organisms.

Here we apply Mora *et al.*'s hyperexponential model to estimate the diversity of human RNA viruses. The model can be expressed as:

$$\log(\log S) = 2 \log(\log G) - \log(\log F)$$

where S is the estimated number of species, G is the observed number of genera, and F is the observed number of families. Yet higher taxa – orders, classes etc. – are not well defined for viruses so we do not use them here (cf. [2]). The combination of current values for G and F (47 and 17 respectively) is highly consistent with the pattern reported by Mora *et al.* for all Animalia. Using these values gives a best estimate of $S=187$ species (with jack-knife-estimated 95% confidence intervals of ± 13), implying that 79-91% of species are already known. However, both G and F are probably underestimated. To allow for this we make the plausible assumption that viruses which infect humans will eventually be found in all currently recognised genera and families containing RNA viruses which infect mammals [10]. This gives $G=60$ and $F=19$ and generates an estimate of $S=297$ species (± 29). Even this higher and probably more realistic estimate implies that 48-59% of species are already known.

We explore the sensitivity of our estimate of human RNA virus species richness to wider variation in G and F in Figure 3. Figure 3A shows that the estimate of species richness would only become substantially higher if we discovered human RNA viruses from many more genera, which is not happening at present (Figure 1). Figure 3B makes a related point: the estimate of species richness only becomes substantially higher if we are currently much too conservative (defined in terms of a scaling factor – see figure legend) in defining RNA virus genera within recognised families.

One way of interpreting these results is that they are consistent with the lower range of estimates obtained from extrapolation of the discovery curve (which takes no account of taxa above the species level) but are difficult to reconcile with the very high upper limits. In other words, the possibility of their being thousands of as yet undiscovered human RNA virus species is not compatible with current virus taxonomy (which generates the patterns shown in Figures 1 and 2). We return to this point below.

Loss of diversity

In addition to the continuing discovery of new species, some human RNA viruses seem to have disappeared from human populations. For a total of 52 species, we have been unable to find any record of human infection since 2006, even using an expanded search protocol that includes multi-language internet trawls [16]. These 'lost' viruses are highlighted in Table 1 and represent almost one-third of all ICTV-recognised human RNA virus species.

There are, of course, numerous reasons why a virus may not have recently been detected in humans, not least the possibility that no one has been looking (and we note that many of the 'lost' viruses occur in regions of the world with limited diagnostic capacity). Another possibility is that the original reports of human infection were mistaken (a possible example being cases of human foot-and-mouth disease reported in the 1960s [17] but, despite the

high-profile nature of this infection, never since). Others could simply be very rare: 15 species have only been reported in humans once, as single cases or outbreaks. However, as all 'lost' viruses are zoonotic, they could in principle re-emerge from their reservoir in the future. The only human-adapted virus to be lost (in this one instance as a result of human intervention) is SARS coronavirus, though this too is zoonotic and its potential re-emergence is a significant public health concern.

The implication of this is that the extant diversity of human RNA viruses is lower than the 158 species reported earlier. A crude estimate (using the upper limit of 52 lost species) of the average loss rate is slightly above 0.01 per species per year. Taken together with the current rate of discovery of just over 2 species per year, this implies that there is considerable turnover (sensu [18]) in the cast of extant human RNA viruses (best estimate: 15% per decade). However, as remarked above, almost all this turnover involves viruses for which humans are incidental or spillover hosts. Among this subset the turnover rate is appreciably higher (best estimate: 22% per decade).

Evolution and emergence of new human viruses

So far, implicit in our analyses has been the idea that there exists a fixed pool of human RNA virus species which we are still discovering. In reality, the composition of this pool is likely to change through time; just as some virus species appear to be being lost, others may be being added. There are two obvious ways in which new species may be added to the pool.

First, new human viruses may appear when humans are exposed to an existing virus species for the first time. The most important sources of new human viruses are other mammals and, to a lesser extent, birds [8]. We have limited knowledge of the diversity of RNA viruses in these reservoirs; although there are partial inventories for some domestic animals [19] (of which there are a few dozen species), we know much less about the viruses present in wild mammals (of which there are well over 5000 species). We also know that a substantial fraction of known mammal viruses is capable of infecting humans [12,19]. What is less clear is whether or not there are large numbers of such viruses which humans have yet to be exposed to.

Second, new human viruses may evolve from other viruses. In some instances this may simply reflect diversification within human populations: this is likely to apply to groups such as the human astroviruses [10], but it does not appear to be the dominant mode of speciation for many, perhaps most, human RNA viruses. More commonly, new RNA virus species arise as a result of jumps from one host species to another [5]. The most striking recent (in evolutionary terms) example of this among human viruses is the origin of the human immunodeficiency virus HIV-1 from its immediate ancestor in chimpanzees, SIV_{cpz} [20]. The origins of most of the other 51 'human-adapted' RNA virus species are much less well known, but many are suspected to have jumped from other host species. Moreover, the great majority of them fall into the category of "crowd diseases", which implies that they require relatively high host densities to persist [6]. In that case, an order-of-magnitude estimate of one such event per 100 years (rather than one per 1000 years or one per 10 years) is broadly consistent with human demographic history [6]. Of course, this does not preclude the

precursors of human-adapted viruses, such as SIVs, occurring as zoonotic infections prior to the evolution and emergence of the new species.

The overall balance between exposure and evolution as drivers of human RNA diversity is difficult to assess with any precision, particularly for those that are not human-adapted and exist primarily in non-human reservoirs. So we do not know what fraction of the 2 or so new species of human RNA virus discovered per year are of recent evolutionary origin. However, the idea that most new species appear as a consequent of evolution would be difficult to reconcile with falling discovery rates.

Future perspective

The immediate future seems certain to involve the discovery and/or recognition of new species of RNA viruses that infect humans. Indeed, given the recent increase in interest in novel viruses, combined with the considerable potential of high throughput sequencing as a virus discovery tool [7], it would be no surprise if the discovery rate increased in coming years. An important aid to interpreting any such trends will be the extent to which the accumulation of new species is accompanied by the addition of new genera and families containing species that can infect humans. If no or very few higher taxa are added then this would be consistent with the notion that the broad limits of human RNA diversity have already been identified and we are now just filling in the gaps. A technical caveat here is that discovery based on nucleotide sequencing is considerably easier for sequences closer to those that we already recognise than for sequences which are very different [7].

The recognition that human RNA viruses often have zoonotic origins has led to a parallel increase in interest in the viruses of animals other than humans. Our knowledge of the diversity of these viruses is less complete than it is for humans for domesticated species and far less complete for wildlife species. While we will surely see many more viruses added to the catalogue of diversity in coming years, it is difficult to imagine that our inventory of, say, viruses of all mammals and birds is likely to be any more than the tip of the iceberg for the foreseeable future. It may be that there is a case for concentrating our efforts on hosts most likely to be reservoirs of potentially epidemic human pathogens (as opposed to minor zoonotic infections). There is some suggestion that viruses acquired from other primates may be of special concern [8] but this is far from an absolute rule: SARS coronavirus, for example, is believed to have originated in bats [21] and influenza A is primarily a virus of wildfowl [22].

Another way that our understanding of virus diversity could change is through revisions to the way in which we characterise virus diversity. We may not be too far away from developing purely sequence-based taxonomies, although there is still the challenge of relating sequence differences to both reproductive isolation (i.e. “species” in the usual sense) and also to function (notably the abilities to infect, transmit from and cause disease in humans). The goal of being able to ‘read’ virus sequences and inferring their properties (such as an ability to infect humans or potential to cause disease) as a routine procedure still seems a very long way off. But recent studies of the substitutions required for avian H5N1 influenza A to become transmissible in ferrets hint that progress is possible [23].

As full genome sequences become available for a growing number of RNA viruses, it becomes ever more possible to undertake quantitative analyses of the patterns of sequence diversity. A seminal study of this kind [24] that analysed gene sequences from 42 mosquito-borne flaviviruses estimated that there were many hundreds and more likely thousands of as yet undiscovered species in that group. This approach, which is essentially a way of estimating under-sampling near the tips of a phylogenetic tree based on specified assumptions about the underlying macroevolutionary processes, has parallels with that of Mora *et al.*'s [2] analysis of taxonomic hierarchies. Further development and refinement of these kinds of approaches to make them as rigorous and robust as possible seems warranted.

Conclusion

Almost all previous discussions of the potential diversity of human RNA viruses, our own contributions included, emphasize the possibility that the 158 currently recognised species represent just a small fraction of a very large number of species capable of infecting humans of which we are currently unaware. This could indeed be the case, but here we acknowledge the alternative possibility that the number of species of RNA virus capable of infecting humans could, in fact, be relatively limited. Even so, in this context "limited" corresponds to many tens or even hundreds of additional species which, without extraordinary advances in technology or increases in effort, will keep virus hunters occupied for many decades at least. But it does imply that a complete inventory of the species diversity (as we currently define "species") of human RNA viruses is at least feasible.

The above argument is most easily defended if we think of RNA virus diversity as a fixed pool. But it is not: new species can evolve and have evolved, and some of these new species are able to infect humans. We might imagine that, given the relatively high rates of within-species evolution of RNA viruses, this happens very regularly. But this is not apparent in the existence of a huge diversity of known human-adapted viruses (i.e. those which are capable of persisting in human populations in the absence of a non-human reservoir). There are only 50 or so species of such viruses and even if, as is plausible, most of these have arisen in the past few millennia the appearance of a new one must still be a relatively rare event (though the possibility remains a clear and present threat to public health).

We can consider the same question in terms of the broader definition of a human virus that includes the majority that are zoonotic and cannot persist in human populations in the absence of a reservoir (which may be other mammals or birds). In this context it is more believable (given a very large pool of mammal and bird viruses on which to draw) that new species are appearing very frequently, even if most of these have little prospect of becoming truly human-adapted. But it seems that this is not a simple process of accumulation: there is evidence suggesting that zoonotic viruses that appear in human populations can also disappear again. If this is correct, it would inevitably result in (i) limits to the amount of standing diversity present in humans at any one time, and (ii) substantial turnover (perhaps on a time scale of just a few decades) in the cast of species to be found in humans.

These issues are fundamental to an informed evaluation of the threat to public health posed by emerging human RNA viruses. They will be resolved only by sustained (over decades)

and preferably greatly enhanced (or better targeted) surveillance of human RNA viruses and by large-scale, systematic studies of RNA virus diversity in other hosts, especially mammals and birds [25].

Acknowledgements

We are grateful to Margo Chase-Topping and Camilo Mora for statistical and modelling advice and to John Brownstein and colleagues for assistance with internet searches. This work is part of the VIZIONS project supported by the Wellcome Trust.

Bibliography

1. Reed W, Carroll J, Agramonte A. The etiology of yellow fever. An additional note. *J Am Med Assoc.* 1901; 36:431–440.
2. Mora C, Tittensor DP, Adi S, Simpson AGB, Worm B. How many species are there on earth and in the ocean? *PLoS Biology.* 2011; 9:e1001127. [PubMed: 21886479] [*Describes a novel approach to estimating taxonomic diversity]
3. Biagini P. Classification of TTV and related viruses (anelloviruses). *Curr Top Microbiol Immunol.* 2009; 331:21–33. [PubMed: 19230555]
4. Simmonds P. Reconstructing the origins of human hepatitis viruses. *Phil Trans R Soc Lond B.* 2001; 356:1013–1026. [PubMed: 11516379]
5. Kitchen A, Shackelton LA, Holmes EC. Family level phylogenies reveal modes of macroevolution in RNA viruses. *Proc Natl Acad Sci USA.* 2011; 108:238–243. [PubMed: 21173251] [***Interesting phylogenetic analysis providing convincing evidence for frequent host jumps by RNA viruses]
6. Wolfe ND, Dunavan CP, Diamond J. Origins of major human infectious diseases. *Nature.* 2007; 447:279–283. [PubMed: 17507975] [***Categorizes zoonotic pathogens based on their transmissibility within human populations]
7. Bexfield N, Kellam P. Metagenomics and the molecular identification of novel viruses. *Vet J.* 2011; 190:191–198. [PubMed: 21111643] [*Useful review of the pitfalls of using high throughput sequencing to identify novel viruses]
8. Woolhouse MEJ, Scott FA, Hudson Z, Howey R, Chase-Topping M. Human viruses: discovery and emergence. *Phil Trans R Soc Lond B.* 2012; 367:2864–2871. [PubMed: 22966141]
9. Fields, BN, Knipe, DM., Howley, PM., editors. *Storch GA Diagnostic virology. Fields Virology.* 5th ed. Lippincott, Williams & Wilkins; PA, USA: 2007. p. 565-604.
10. King, AMQ, Adams, MJ, Carstens, EB., Lefkowitz, EJ., editors. *Virus Taxonomy.* Elsevier; Amsterdam, The Netherlands: 2012.
11. van den Brand JMA, Stittelaar KJ, van Amerongen G, et al. Comparison of temporal and spatial dynamics of seasonal H3N2, pandemic H1N1 and highly pathogenic avian influenza H5N1 virus infections in ferrets. *PLoS ONE.* 2012; 7:e42343. [PubMed: 22905124]
12. Taylor LH, Latham SM, Woolhouse MEJ. Risk factors for human disease emergence. *Phil Trans R Soc Lond B.* 2001; 356:983–989. [PubMed: 11516376]
13. Woolhouse MEJ, Howey R, Gaunt E, Reilly L, Chase-Topping M, Savill N. Temporal trends in the discovery of human viruses. *Proc R Soc Lond B.* 2008; 275:2111–2115.
14. Peiris JSM, Lai ST, Poon LLM, et al. Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet.* 2003; 361:1319–1325. [PubMed: 12711465]
15. Pereira HG, Fialho AM, Flewett TH, Teixeira JMS, Andrade ZP. Novel viruses in human faeces. *Lancet.* 1988; 332:103–104.
16. Brownstein JS, Freifeld CC, Reis BY, Mandl KD. Surveillance sans frontieres: internet-based emerging infectious disease intelligence and the HealthMap project. *PLoS Medicine.* 2008; 5:1019–1024.
17. Armstrong R, Davie J, Hedger RS. Foot-and-mouth disease in man. *Brit Med J.* 1967; 4:529–530. [PubMed: 4294412]

18. Bakker C, Blair JM, Knapp AK. Does resource availability, resource heterogeneity or species turnover mediate changes in plant species richness in grazed grasslands? *Oecologia*. 2003; 137:385–391. [PubMed: 12955488]
19. Cleaveland S, Laurenson MK, Taylor LH. Diseases of humans and their domestic mammals: pathogen characteristics, host range and the risk of emergence. *Phil Trans R Soc Lond B*. 2001; 356:991–999. [PubMed: 11516377]
20. Keele BF, Van Heuverswyn F, Li Y, Baile E, et al. Chimpanzee reservoirs of pandemic and non-pandemic HIV-1. *Science*. 2006; 313:523–526. [PubMed: 16728595] [*Well researched case study of the origins of one of world's most important human viruses]
21. Sh Z, Hu Z. A review of studies on animal reservoirs of the SARS coronavirus. *Virus Res*. 2008; 133:74–87. [PubMed: 17451830]
22. Kruiken T, Holmes EC, McCauley J, Rimmelzwaan GF, Williams CS, Grenfell BT. Host species barriers to influenza virus infections. *Science*. 2006; 312:394–397. [PubMed: 16627737] [***Insightful review of the biology of influenza]
23. Herfst S, Schrauwen EJA, Linster M, et al. Airborne transmission of A/H5N1 influenza virus between ferrets. *Science*. 2012; 336:1534–1541. [PubMed: 22723413]
24. Pybus OG, Rambaut A, Holmes EC, Harvey P. New inferences from tree shape: numbers of missing taxa and population growth rates. *Syst Biol*. 2002; 51:881–888. [PubMed: 12554454] [***First attempt to develop a methodology for estimating virus diversity from phylogenetic data]
25. Morse SS, Mazet JAK, Woolhouse M, et al. Predicting and preventing the next pandemic zoonosis. *Lancet*. 380 (In press).
26. Chua KB, Crameri G, Hyatt A, et al. A previously unknown reovirus of bat origin is associated with an acute respiratory disease in humans. *Proc Natl Acad Sci USA*. 2007; 104:11424–11429. [PubMed: 17592121]
27. Yaiw KC, Crameri G, Wang L, et al. Serological evidence of possible human infection with *Tioman virus*, a newly described paramyxovirus of bat origin. *J Infect Dis*. 2007; 196:884–886. [PubMed: 17703419]
28. Finkbeiner SR, Kirkwood CD, Wang D. Complete genome sequence of a highly divergent astrovirus isolated from a child with acute diarrhea. *Virol J*. 2008; 5:117. [PubMed: 18854035]
29. Towner JS, Sealy TK, Khristova ML, et al. Newly discovered Ebola virus associated with hemorrhagic fever outbreak in Uganda. *PLoS Pathog*. 2008; 4:e1000212. [PubMed: 19023410]
30. Delgado S, Erickson BR, Agudo R, et al. Chapare virus, a newly discovered Arenavirus isolated from a fatal hemorrhagic fever case in Bolivia. *PLoS Pathog*. 2008; 4:e1000047. [PubMed: 18421377]
31. Kapoor A, Slikas E, Simmonds P, et al. A newly identified bocavirus species in human stool. *J Infect Dis*. 2008; 199:196–200.
32. Holtz LR, Finkbeiner SR, Kirkwood CD, Wang D. Identification of a novel picornavirus related to cosaviruses in a child with acute diarrhea. *Virol J*. 2008; 5:159–164. [PubMed: 19102772]
33. Finkbeiner SR, Li Y, Ruone S, et al. Identification of a novel astrovirus (Astrovirus VA1) associated with an outbreak of acute gastroenteritis. *J Virol*. 2009; 83:10836–10839. [PubMed: 19706703]
34. Kapoor A, Li L, Victoria J, et al. Multiple novel astroviruses in human stool. *J Gen Virol*. 2009; 90:2965–2972. [PubMed: 19692544]
35. Greninger AL, Runckel C, Chiu CY, et al. The complete genome of klassevirus – a novel picornavirus in pediatric stool. *Virol J*. 2009; 6:82. [PubMed: 19538752]
36. Briese T, Paweska JT, McMullan LK, et al. Genetic detection and characterization of Lujo virus, a new hemorrhagic fever-associated arenavirus from Southern Africa. *PLoS Pathog*. 2009; 5:e1000455. [PubMed: 19478873]
37. Klempa B, Koivogui L, Sylla O, et al. Serological evidence of human hantavirus infections in Guinea, West Africa. *J Infect Dis*. 2010; 201:1031–1034. [PubMed: 20187741]
38. Yu X-J, Liang M-F, Zhang S-Y, et al. Fever with thrombocytopenia associated with a novel Bunyavirus in China. *New Engl J Med*. 2011; 364:1523–1532. [PubMed: 21410387]

Executive summary

- *Virus discovery.* Viruses were first recognised over 100 years ago, but we still do not have a complete inventory even of those that infect humans.
- *Number of species.* There are currently 158 recognised species of RNA virus that can infect humans, representing 47 genera and 17 families. Less than a third of these, however, are human-adapted (in the sense that they can persist in humans in the absence of a non-human reservoir host).
- *Discovery curves.* We are still discovering/recognising new human RNA virus species but there is some evidence that the rate of discovery has fallen over the past 50 years, and it seems clear that this is the case for new genera and new families.
- *Diversity estimation.* The application of a method of estimating biodiversity from taxonomic structure raises the possibility that half or more of human RNA virus species have already been discovered.
- *Loss of diversity.* There is evidence that, as well new ones appearing, previously recognised species are disappearing from human populations, suggesting there may be significant turnover (on time scales of decades) in the cast of human virus species.
- *Evolution and emergence.* Exposure of humans to existing viruses in other hosts (mainly mammals and birds) is the most important source of new human RNA viruses, but the extent to which this process has been supplemented by the evolution of new species (in humans or other hosts) remains unclear – it is possible that this is a relatively rare event, which is consistent with the relatively small number (around 50) of human-adapted viruses.

Figure 1. RNA virus discovery curves: the cumulative number of RNA virus species (blue), genera (red) and families (green) found in humans between 1901 and 2006. Taxa are those used in the 9th ICTV report [9]. Adapted from [7].

Figure 2.

Taxonomic diversity of human RNA viruses. A) Numbers of species per genus for all 47 currently recognised genera. B) Numbers of genera per family for all 17 currently recognised families. Taxa are those used in the 9th ICTV report [9]. Adapted from [7].

Figure 3.

Sensitivity of species diversity estimator to changes in the numbers of recognised genera and families. A) Expected number of as yet undiscovered species (vertical axis) as a function of number of recognised genera (horizontal axis) and families (increasing from 15 to 25 as indicated by arrow). Expectations given the currently recognised numbers (47 and 17 respectively) and plausible upper limits (60 and 20) are shown (black circles). B) Expected number of as yet undiscovered species as a function of a scaling factor. The scaling factor has a central value of 2 (dotted vertical line) corresponding to the assumption that species

scale to genera as genera scale to family [2]. Values >2 indicate that families actually contain more genera than is currently recognised and values <2 that they contain fewer. The lower curve corresponds to the observed 47 genera and 17 families and the upper curve to the plausible estimate of 60 and 20 respectively.

Table 1

A list of currently recognised RNA virus species reported to infect humans. Those in grey have not, as far as we can determine, been reported in humans in the past 5 years.

Family	Genus	Species
Retro-	Deltaretro-	Primate T-lymphotropic virus 1
		Primate T-lymphotropic virus 2
		Primate T-lymphotropic virus 3
	Lenti-	Human immunodeficiency virus 1
		Human immunodeficiency virus 2
	Spuma-	Simian foamy virus
Picobirna-	Picobirna-	Human picobirnavirus
Reo-	Colti-	Colorado tick fever virus
	Orbi-	Changuinola virus
		Great Island virus
		Lebombo virus
		Orungo virus
	Rota-	Rotavirus A
		Rotavirus B
		Rotavirus C
	Seadorna-	Banna virus
Borna-	Borna-	Borna disease virus
Filo-	Marburg-	Lake Victoria Marburgvirus
	Ebola-	Reston ebolavirus
		Sudan ebolavirus
		Tai forest ebolavirus
		Zaire virus
Paramyxo-	Rubula-	Human parainfluenza virus 2
		Human parainfluenza virus 4
		Mumps virus
	Avula-	Newcastle disease virus
	Respiro-	Human parainfluenza virus 1
		Human parainfluenza virus 3
	Henipa-	Hendra virus
		Nipah virus
	Morbilli-	Measles virus
	Pneumo-	Human respiratory syncytial virus
	Metapneumo-	Human metapneumovirus
Rhabdo-	Vesiculo-	Chandipura virus
		Isfahan virus
		Piry virus

Family	Genus	Species
		Vesicular stomatitis Alagoas virus
		Vesicular stomatitis Indiana virus
		Vesicular stomatitis New Jersey virus
	Lyssa-	Australian bat lyssavirus
		Duvenhage virus
		European bat lyssavirus 1
		European bat lyssavirus 2
		Mokola virus
		Rabies virus
Arena-	Arena-	Guanarito virus
		Junin virus
		Lassa virus
		Lymphocytic choriomeningitis virus
		Machupo virus
		Pichinde virus
		Sabia virus
		Whitewater Arroyo virus
Bunya-	Orthobunya-	Bunyamwera virus
		Bwamba virus
		California encephalitis virus
		Caraparu virus
		Catu virus
		Guama virus
		Guaroa virus
		Kairi virus
		Marituba virus
		Oriboca virus
		Oropouche virus
		Shuni virus
		Tacaiuma virus
		Wyeomyia virus
	Hanta-	Andes virus
		Bayou virus
		Black creek canal virus
		Dobrava-Belgrade virus
		Hantaan virus
		Laguna Negra virus
		New York virus
		Puumala virus

Family	Genus	Species
		Seoul virus
		Sin Nombre virus
	Nairo-	Crimean-Congo haemorrhagic fever virus
		Dugbe virus
	Phelbo-	Candiru virus
		Punta Toro virus
		Rift Valley fever virus
		Sandfly fever Naples virus
Orthomyxo-	Influenzavirus A	Influenza A virus
	Influenzavirus B	Influenza B virus
	Influenzavirus C	Influenza C virus
	Thogoto-	Dhori virus
		Thogoto virus
Unassigned	Delta-	Hepatitis delta virus
Corona-	Alphacorona-	Human coronavirus 229E
		Human coronavirus NL63
	Betacorona-	Human coronavirus HKU1
		Human coronavirus OC43
		SARS coronavirus
	Toro-	Human torovirus
Picorna-	Entero-	Human enterovirus A
		Human enterovirus B
		Human enterovirus C
		Human enterovirus D
		Human rhinovirus A
		Human rhinovirus B
		Human rhinovirus C
	Cardio-	Encephalomyocarditis virus
		Theilovirus
	Aphtho-	Equine rhinitis A virus
		Foot and mouth disease virus
	Hepato-	Hepatitis A virus
	Parecho-	Human parechovirus
		Ljungan virus
	Kobu-	Aichi virus
Astro-	Mamastro-	Human astrovirus
		Human astrovirus 2
		Human astrovirus 3
		Human astrovirus 4

Family	Genus	Species
		Human astrovirus 5
		Human astrovirus 6
		Human astrovirus 7
		Human astrovirus 8
Calici-	Noro-	Norwalk virus
	Sapo-	Sapporo virus
Flavi-	Flavi-	Aroa virus
		Banji virus
		Dengue virus
		Ilheus virus
		Japanese encephalitis virus
		Kokobera virus
		Kyasanur forest disease virus
		Louping ill virus
		Murray Valley encephalitis virus
		Ntaya virus
		Omsk haemorrhagic fever virus
		Powassan virus
		Rio Bravo virus
		St Louis encephalitis virus
		Tick-borne encephalitis virus
		Usutu virus
		Wesselsbron virus
		West Nile virus
		Yellow fever virus
		Zika virus
	Hepaci-	Hepatitis C virus
Hepe-	Hepe-	Hepatitis E virus
Toga-	Alpha-	Barmah Forest virus
		Chikungunya virus
		Eastern equine encephalitis virus
		Everglades virus
		Getah virus
		Mayaro virus
		Mucambo virus
		O'nyong-nyong virus
		Pixuna virus
		Ross River virus
		Semliki Forest virus

Family	Genus	Species
		Sindbis virus
		Venezuelan equine encephalitis virus
		Western equine encephalitis virus
		Whataroa virus
	Rubi-	Rubella virus

Table 2

A list of putative RNA virus species first reported in humans between 2007 and 2011.

Species	Family	Reference
Melaka virus	Reoviridae	[26]
Tioman virus	Paramyxoviridae	[27]
Astrovirus MLB1	Astroviridae	[28]
Bundibugyo ebolavirus	Filoviridae	[29]
Chapare virus	Arenaviridae	[30]
Human cosavirus A	Picornaviridae	[31]
Human cosavirus B	Picornaviridae	[31]
Human cosavirus C	Picornaviridae	[31]
Human cosavirus D	Picornaviridae	[31]
Human cosavirus E1	Picornaviridae	[32]
Astrovirus VA1	Astroviridae	[33]
HMOAstrovirus A	Astroviridae	[34]
HMOAstrovirus B	Astroviridae	[34]
HMOAstrovirus C	Astroviridae	[34]
Klassevirus	Picornaviridae	[35]
Lujo virus	Arenaviridae	[36]
Sangassou virus	Bunyaviridae	[37]
SFTS virus	Bunyaviridae	[38]