Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Virology

Parallel molecular evolution and adaptation in viruses

Bernardo Gutierrez, Marina Escalera-Zamudio and Oliver G Pybus

Parallel molecular evolution is the independent evolution of the same genotype or phenotype from distinct ancestors. The simple genomes and rapid evolution of many viruses mean they are useful model systems for studying parallel evolution by natural selection. Parallel adaptation occurs in the context of several viral behaviours, including cross-species transmission, drug resistance, and host immune escape, and its existence suggests that at least some aspects of virus evolution and emergence are repeatable and predictable. We introduce examples of virus parallel evolution and summarise key concepts. We outline the difficulties in detecting parallel adaptation using virus genomes, with a particular focus on phylogenetic and structural approaches, and we discuss future approaches that may improve our understanding of the phenomenon.

Address
Department of Zoology, University of Oxford, Oxford, United Kingdom

Corresponding authors:
Gutierrez, Bernardo (bernardo.gutierrez@zoo.ox.ac.uk),
Pybus, Oliver G (oliver.pybus@zoo.ox.ac.uk)

Current Opinion in Virology 2019, **34**:90–96

This review comes from a themed issue on **Emerging viruses: inter-species transmission**

Edited by **Adolfo García-Sastre** and **Juergen A Richt**

For a complete overview see the Issue and the Editorial

Available online 28th January 2019

https://doi.org/10.1016/j.coviro.2018.12.006

1879-6257/© 2018 Elsevier B.V. All rights reserved.

Introduction

The processes that drive the cross-species transmission and emergence of viruses in natural systems are diverse and involve a range of ecological, evolutionary, and genetic factors. Amongst the evolutionary mechanisms, molecular adaptation by natural selection is a key phenomenon that entails the generation and spread of beneficial mutations that increase virus fitness in a specific environment. If rapidly evolving virus populations repeatedly experience comparable environmental changes and similar selective pressures, then viruses may exhibit what is known as parallel evolution.

The concept of parallel evolution is applied to organisms across the tree of life and can refer to all manners of phenotypes and traits at different levels of biological organisation [1]. Parallel evolution manifests itself in a variety of ways, from the repeated fixation of single point mutations or larger genetic changes (e.g. indels or genome rearrangements) [2], to the evolution of structural, functional or behavioural phenotypes. The terms 'parallel' and 'convergent' evolution are sometimes given distinct definitions: convergent evolution refers to the independent evolution of similar traits from different ancestral starting points [3], whereas parallel evolution describes the independent emergence of similar traits from the same state [4]. Here, we use 'parallel' to refer to both situations. Further, we use 'adaptation' to refer to evolutionary change through positive natural selection, and 'evolution' to encompass change via any process (e.g. by random genetic drift).

Parallel evolution in viruses may arise from adaptation to new host species [5] or different host demographic structures [6], evasion of host immune responses [7,8] or to circumvent anti-viral drugs [9]. These situations involve changes to the virus' environment that generate strong selective pressures, favouring the recurrent evolution of certain genotypes. Here, we summarise the theoretical concepts behind parallel molecular evolution in viruses, and explore some of the methods that can be used to identify such phenomena. We highlight the importance of phylogenetic methods to detect recurring evolutionary patterns generated by natural selection.

Parallel molecular evolution in viruses

Viruses are useful model systems for studying parallel evolution and adaptation. Many viruses, particularly RNA viruses, can adapt rapidly due to a combination of high mutation rates, large population sizes, short generation times, and large mutational selection coefficients [10••]. Further, the small genomes of RNA viruses (and some DNA viruses) may limit the range of genetic solutions available to viruses as they respond to environmental change. In contrast, the larger genomes of DNA viruses, such as Myxoma [11], may offer more potential genetic routes to the same phenotype.

Perhaps the best described example of virus parallel evolution is the development of anti-viral drug resistance by the human immunodeficiency virus (HIV). HIV exhibits all the features of rapid adaptation listed above [12•]; consequently, many HIV drug resistance mutations have been identified, which independently and repeatedly arise during chronic infection in different HIV patients [12•,13,14]. Interestingly, even within a genome as constrained as HIV's, different mutations can confer

Table 1

Selected examples of virus parallel evolution

Virus	Gene	Mutation	Details	Reference
			Laboratory conditions	
Vesicular stomatitis virus (VSV)	Various	Various	Independent populations sequentially passaged showed 19 loci that evolved shared alleles.	[64]
Middle East respiratory syndrome coronavirus (MERS-CoV)	S	S465F	Replicated sequential passaging in DDP4-expressing cells resulted in the mutation Ser → Phe in the spike protein	[65]
			Natural systems	
Human immunodeficiency virus (HIV)	Gag	M30R, Y30R	A change from Met or Lys → Arg observed in HIV groups M, N and O.	[18]
Rabies virus (RABV)	N L	L374S K200R	Substitutions in the nucleoprotein and polymerase observed in two independent zoonoses.	[5]
Influenza A virus (IAV)	HA	G228S	A Gly → Ser change observed in different HX subtypes that circulate in humans, compared to avian and equine sequences.	[66]
	PB2	E627K	A Glu → Lys change at position 627 of PB2 increases virulence in mammalian hosts, in both H5N1 and H3N2 subtypes.	[67]
West Nile Virus (WNV)	NS3	T249P	A single change (Thr→Pro) in the viral helicase increased virulence for American crows, and was observed in different bird outbreaks in Egypt, Romania, Russia, Israel and the United States.	[68]
Circulating vaccine-derived polioviruses (cVDPV)	VP1 VP2 VP3	F280Y F141Y T80K	Multiple parallel changes observed in vaccine derived polioviruses in 29 different epidemics in relation to a known, ancestral oral polio vaccine (OPV) strain sequence.	[69][a]

[a] An additional non-coding substitution was also reported by the authors and is not included in this table.

resistance to specific drugs [15]. HIV also displays rapid and repeatable evolution to host immune responses, as demonstrated by the parallel evolution in different patients of escape mutations to HLA-restricted T-cell responses [7].

Viruses that register frequent cross-species transmission events might also exhibit parallel evolution as a result of adaptation to new host environments [16]. Barriers to virus replication in a new host may include (i) lack suitable receptors for virus cell entry, (ii) innate and adaptive immune responses, (iii) reduced replication efficacy in specific cell types, and (iv) mechanisms of virion release from infected cells [17••]. For example, HIV-1 groups M, N and O are derived from separate cross-species transmissions to humans of simian immunodeficiency viruses (SIVs) in chimpanzees and gorillas. Following spillover, these groups independently acquired a Met → Arg amino acid change in the retroviral Gag protein [18]. Under experimental conditions, this mutation increases SIV replication in human lymphoid tissue [19]. Rabies virus (RABV) offers another example: the RABV lineage that infects dogs has changed host species on multiple occasions, causing outbreaks in other *Carnivora* species. Amongst these events, parallel amino acid changes were observed in two separate RABV zoonoses from dogs to ferret-badgers (a Leu → Ser mutation in the RABV nucleoprotein and a Lys → Arg change in the polymerase) [5].

Parallel evolution and adaptation may also target larger genome regions, and sites involved in post-translational modification. For example, the hemagglutinin of H7 subtype avian influenza virus (AIV) can acquire a N-glycosylation site when the virus jumps from wild to domestic birds [20], and highly pathogenic strains of AIV have emerged independently many times, through the evolution of different polybasic cleavage sites in hemagglutinin [21]. For HIV-1, it is now understood that, after zoonosis, groups M and O independently evolved the ability to antagonise human tetherin (an anti-viral restriction factor), but did so using entirely different retroviral accessory genes [22,23].

Parallel molecular evolution: selection and adaptation

Parallel molecular evolution has been described for many viruses both under laboratory conditions and in natural systems. Table 1 provides an illustrative selection of RNA virus examples. Whilst theoretical studies have focused on the ability of natural selection to increase the probability of parallel evolution [24], experimental studies have indicated the potential impact of other factors, including epistatic interactions amongst sites and heterogeneity of mutation rates across the genome [25••,26].

It is commonly hypothesised that natural selection is the key process involved in parallel molecular evolution. Negative selection eliminates many mutations from viral populations because strong functional constraints render most genetic changes disadvantageous. In contrast, positive selection favours the accumulation of mutations that are beneficial to virus replication and transmission. Determining whether a given site or genomic region has

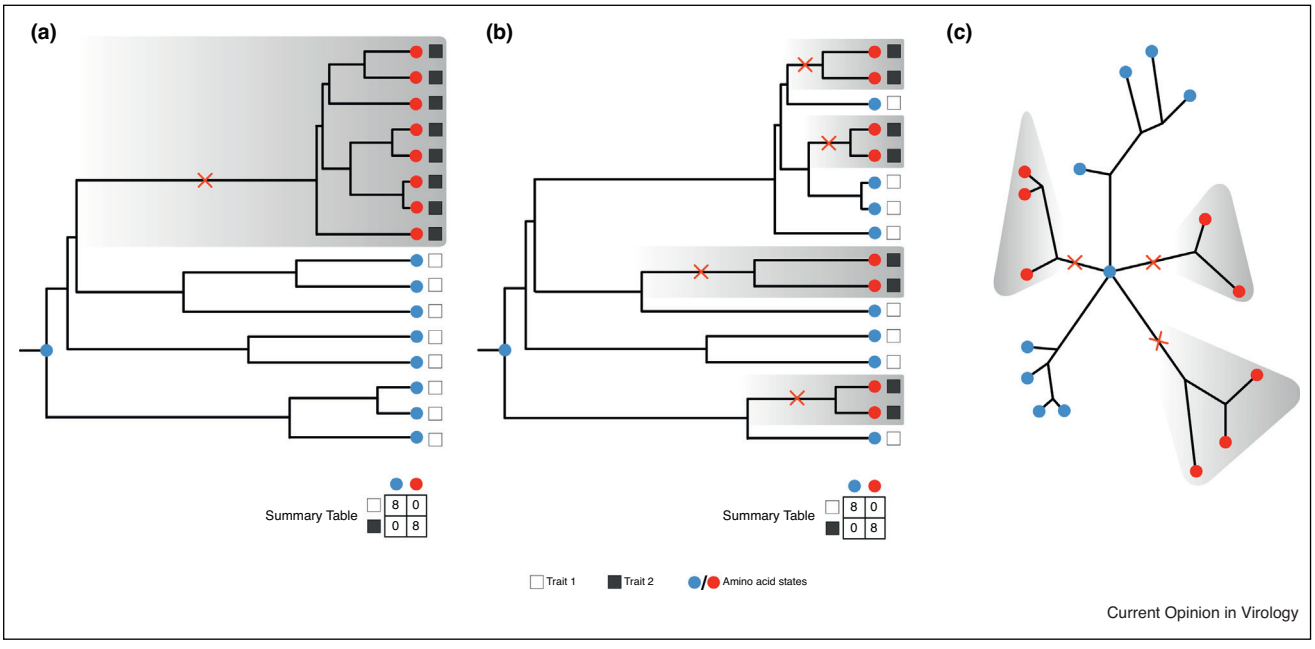
evolved under natural selection can be achieved by comparing the relative rates of synonymous (d_S) and non-synonymous (d_N) changes, quantified as a d_N/d_S ratio [27]. Because positive selection can act on specific sites at particular points in time, it is important to explore variation in d_N/d_S amongst sites and amongst virus lineages. Several methods are available to undertake such calculations [28–30]. If $d_N/d_S > 1$ for a specific codon, then multiple amino acid changes have likely occurred at that codon, and the codon is a candidate for the detection of parallel evolution (see below). However, d_N/d_S methods have limitations and may fail to detect instances of sites under selection [29]. In particular, some studies report that d_N/d_S tests have failed to identify sites that appear to have undergone parallel adaptation [5,31].

Detecting parallel molecular evolution

The robust identification of parallel evolution and adaptation can be broken down into four steps: (*i*) detection of parallel changes (usually amino acid changes), (*ii*) association of those changes with a phenotype or environmental change of interest, (*iii*) characterisation of the evolutionary forces behind the repeated occurrence of such changes, and (*iv*) evaluation of the functional effects of mutations in the context of host–pathogen biology. A powerful method for detecting repeated, independent evolution is comparative phylogenetics, which infers the evolutionary history of a given trait (including reconstruction of ancestral states) using parsimony or other phylogenetic models [32]. This approach is typically robust, so long as recombination or horizontal gene transfer (which can mimic parallel evolution) is absent.

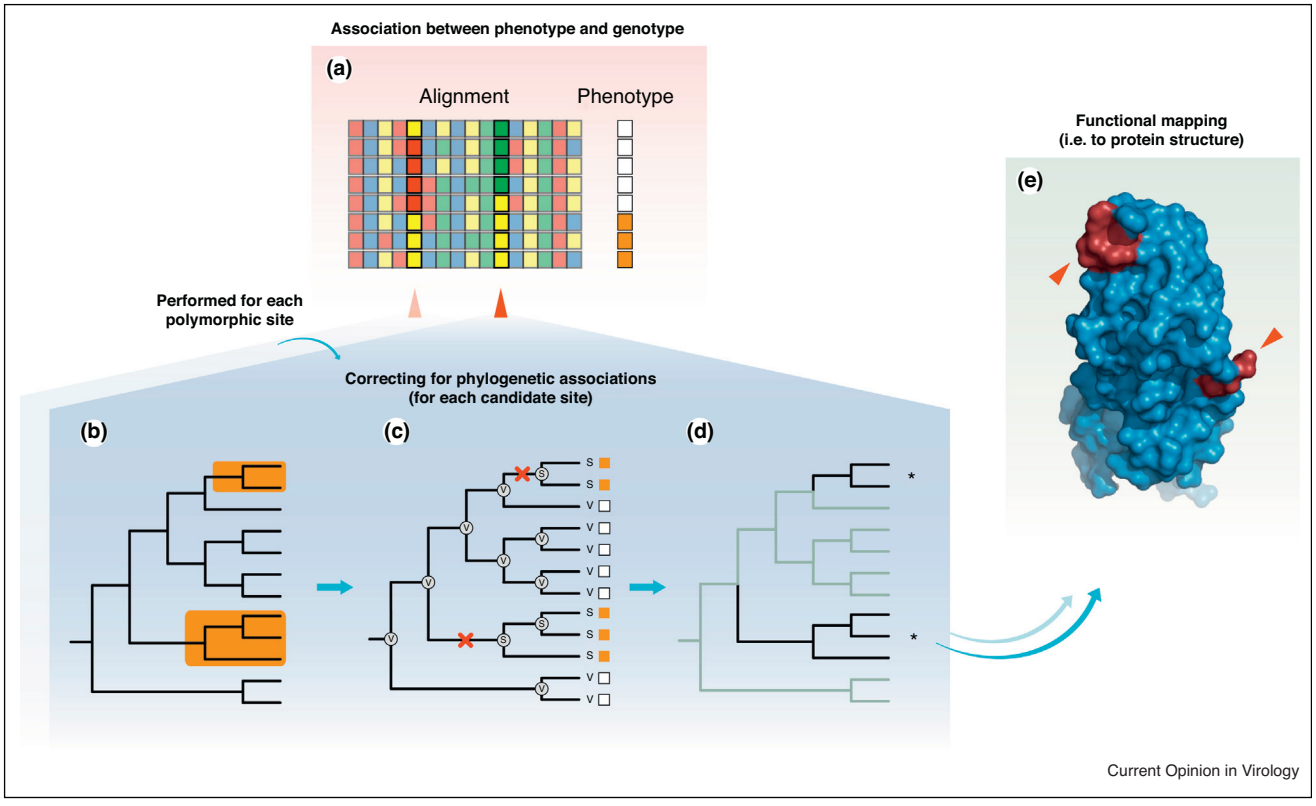
Steps (*ii*) and (*iii*) are less straightforward. It is tempting to detect associations between a mutation and a phenotype by calculating the mutation frequencies in viruses that do or do not exhibit the relevant phenotype. This generates a contingency table of observed counts (Figure 1) that is evaluated using a Chi-squared or Fisher's exact test. However, it has long been known in the evolutionary literature that this approach can lead to false positive conclusions, because the counts are not independent observations, but instead are correlated due to shared ancestry [33,34]. The problem can be resolved by considering the phylogenetic history of the trait in question (Figure 1). Specifically, recurrent acquisition of a mutation in the context of independent changes in phenotype (Figure 1b) indicates that the mutation is evolutionarily linked to the phenotype. However, if mutation and phenotypic change occur only once in the phylogeny (Figure 1a) then their association could be due to chance. Bhattacharya *et al.* [7] referred to the latter situation as the

Figure 1



Different evolutionary scenarios underlying the same apparent association between virus mutation and phenotype. **(a)** An apparently perfect association between the presence/absence of a mutation (blue/red dots) and a specific phenotype/environment (shading) can arise from a single evolutionary change (red cross). This association is typically represented and tested using a contingency table (bottom). However, each virus sequence state is *not* an independent observation due to this evolutionary history. **(b)** The same apparent association can arise through multiple independent mutations (red crosses) linked to change in phenotype/environment. This latter scenario is less likely to occur by chance and provides evidence for parallel adaptation. **(c)** Experimental evolution studies are often designed to observe changes (red crosses) in variants from a known ancestral strain (blue dot, centre) that is introduced to different environments (shading).

Figure 2



A framework for evaluating parallel evolution in virus sequences. **(a)** An alignment is estimated for a set of viral sequences. Polymorphic sites associated with a given phenotype or environment are identified. **(b)** A phylogeny is estimated, and the lineages in which the phenotype of interest is observed are identified (orange). **(c)** Ancestral states (here, S or V) are reconstructed for each of the polymorphic sites identified in step A. **(d)** Branches upon which a genetic change (red cross) is associated with the trait of interest are tested for positive selection. **(e)** Sites that exhibit evidence of molecular adaptation are mapped into a structure of the protein in question to determine possible functional effects.

'founder effect' and used a phylogenetically corrected statistical test to demonstrate that previously reported HLA-associated escape mutations in HIV [35] were not statistically robust. Another phenomenon that may cause mutations to appear artefactually linked to a specific phenotype or environment is genetic hitchhiking, whereby an observed mutation does not affect the associated phenotype, but is genetically linked to a mutation that does. Further, the phylogenetic association between a mutation and a phenotype may be weak if the phenotype is polygenic in nature, or depends on an interaction between the viral genotype and the environment.

It is useful to place virus experimental evolution studies in the same phylogenetic context (Figure 1c). Experimental evolution was, for example, used to identify adaptive mutations that increase viral fitness in vesicular stomatitis virus [26]. Such studies often initiate independent virus populations from the same source inoculum. The replicate populations are passaged in different environments and later scanned for shared evolutionary changes. This experimental design is, thus, equivalent to a star-shaped phylogeny with a known ancestor (Figure 1c). Since the replicate populations are not correlated by shared ancestry, differences amongst them are independent and can be tested using standard statistics.

Parallel evolution beyond the sequence level

As explained above, phylogenetic analyses of viral sequences may be able to detect parallel evolution, but are insufficient to establish whether a given trait has arisen repeatedly through adaptation. Experimental evidence, from molecular biology or animal challenge studies, is ultimately required to determine the functional consequences of parallel mutations. However, it may be possible to further refine candidate sites for experimental confirmation by undertaking *in silico* analyses that combine structural and evolutionary information. This is especially true if virus proteins evolve the same structural change (e.g. a change in charge) via a variety of mutational paths. Mapping of virus mutations onto resolved protein structures can provide information about the functional effects of genetic changes, including thermodynamic stability, or interactions with other proteins or biologically important molecules [36•].

For example, a recent study of human influenza B virus genomes observed parallel amino acid changes occurring three times in the HA gene of the Yamagata lineage [38]. Computational structural analyses showed that these were located in a major antigenic epitope [37]. These approaches include structural alignment methods [38] that can compare divergent virus proteins that exhibit little homology at the sequence level, but which share homology at the structural level. Other combined phylogenetic approaches can estimate the effect of specific mutations on protein stability, as demonstrated in a functional analysis of influenza A virus hemagglutinin gene variation [39]. We envisage that structural analyses could form part of a phylogenetically informed workflow for investigating the significance of parallel evolutionary changes (Figure 2), and for selecting appropriate candidates for experimental validation. The results obtained would also provide insights into the evolvability, mutability and robustness of virus proteins [40].

Other functional constraints, beyond those imposed by protein structure, could also play a role in parallel molecular evolution during emergence or zoonosis. Non-coding viral genetic elements, such as untranslated regions (UTRs) [41], internal ribosome entry sites (IRES) [42,43], or non-coding RNAs [44,45], can be functionally important for virus replication and infection. Flaviviruses provide many examples of the importance of RNA structures for viral replication [46] and host adaptation [41], some of which are proposed to have arisen through parallel evolution [47]; different functional roles have been inferred for RNA structures in other viruses [48,49]. Evaluating the effects of natural selection on RNA genome regions encoding for secondary structures represents an additional challenge because (i) our understanding of how such structures evolve is poor, and (ii) it is difficult to define neutrally evolving sites in such regions [50], which limits the use of d_N/d_S estimation methods, and our limited knowledge of the degree to which non-coding elements are structurally constrained. Various approaches have been developed to test for selection on non-coding regions [51] and to predict the secondary structure of RNA molecules [52,53], but to date their use has been limited in evolutionary analyses of viruses. Further development of these methods could improve our understanding of the role that RNA structures play in virus evolution [54].

Conclusions and future perspectives

Detecting parallel evolution in viruses, and inferring the mechanisms that underlie such changes, can benefit both basic and applied problems. If, in a rapidly evolving virus population, parallel mutations repeatedly arise in the context of the same change in phenotype or environment (i.e. Figure 1b), then at least some aspects of virus evolution are predictable. Such repeating patterns, at the sequence or structural level, may help to forecast virus emergence and cross-species transmission events in the future. Further, placing these changes within a structural context could be used to triage newly discovered viruses, enabling surveillance and research to focus on those with the greatest predicted propensity for cross-species transmission or emergence. We note, however, that instances of parallel evolution identified by computational means must be interpreted with caution; confirmation of the fitness costs of evolutionary change ultimately depends on experimental validation, as shown for some recent outbreak scenarios [55,56]. A number of experimental techniques are now available to link virus genotypes to phenotypes, including pseudotyping [57•], minigenome systems [58], viral reverse genetics [59], and site-directed mutagenesis [60]. These and related approaches are vital to determining the functional role of viral mutations and their effects on virus and host fitness.

Current challenges for inferring parallel molecular evolution are twofold. First, to better understand and predict parallel adaptation, and apply it to the specific scenario of virus emergence, we need a model that describes how the phenomenon depends on factors such as mutation rates, population sizes, selection coefficients and genome evolvability. Second, improvements in statistical methods and software are needed to make it easier to test parallel evolution hypotheses in the correct way. Advances have been made in testing specific phenomena associated with parallel evolution, such as coevolution [61] and episodic selection [28,30]; however, an integrated inference framework that maximises the accuracy of site identification, and is robust to confounding processes such as recombination and genetic hitchhiking is still required. New high throughput experimental approaches, such as deep mutational scanning, have been applied to virus pathogens and will provide more comprehensive information on the mutability of virus proteins [62•,63]. Results from similar empirical approaches, combined with the rapid growth of genome sequences for viruses from multiple hosts and ecosystems, will provide a broader evidence base for virus parallel evolution during virus emergence.

Acknowledgements

We thank Michael Golden for discussion and feedback. B.G. is supported by the 2017 Universities of Academic Excellence Scholarship Program of the Secretariat for Higher Education, Science, Technology and Innovation of the Republic of Ecuador (ARSEQ-BEC-003163-2017) and by Universidad San Francisco de Quito. M.E.Z. is supported by an EMBO Long Term Fellowship (ALTF376-2017). O.G.P. is supported by the European Union Seventh Framework Programme (FP7/2007–2013)/ European Research Council (614725-PATHPHYLODYN).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Stern DL: **The genetic causes of convergent evolution**. *Nat Rev Genet* 2013, **14**:751-764.

2. Wood TE, Burke JM, Rieseberg LH: **Parallel genotypic adaptation: when evolution repeats itself**. *Genetica* 2005, **123**:157-170.

3. Doolittle RF: **Convergent evolution: the need to be explicit**. *Trends Biochem Sci* 1994, **19**:15-18.

4. Zakon HH: **Convergent evolution on the molecular level**. *Brain Behav Evol* 2002, **59**:250-261.

5. Troupin C *et al.*: **Large-scale phylogenomic analysis reveals the complex evolutionary history of rabies virus in multiple carnivore hosts**. *PLoS Pathog* 2016, **12**:e1006041.

6. Kerr PJ: **Myxomatosis in Australia and Europe: a model for emerging infectious diseases**. *Antivir Res* 2012, **93**:387-415.

7. Bhattacharya T *et al.*: **Founder effects in the assessment of HIV polymorphisms and HLA allele associations**. *Science* 2007, **315**:1583-1586.

8. Leslie AJ *et al.*: **HIV evolution: CTL escape mutation and reversion after transmission**. *Nat Med* 2004, **10**:282-289.

9. Clutter DS *et al.*: **HIV-1 drug resistance and resistance testing**. *Infect Genet Evol* 2016, **46**:292-307.

10. •• Woolhouse ME *et al.*: **Assessing the epidemic potential of RNA and DNA viruses**. *Emerg Infect Dis* 2016, **22**:2037-2044.
This work provides an overview of DNA and RNA viruses with emergence potential, paired with simulated data to assess factors that determine the ability to establish outbreaks.

11. Kerr PJ *et al.*: **Evolutionary history and attenuation of myxoma virus on two continents**. *PLoS Pathog* 2012, **8**:e1002950.

12. • Wensing AM *et al.*: **2017 update of the drug resistance mutations in HIV-1**. *Top Antivir Med* 2017, **24**:132-133.
Summary of currently reported mutations associated with drug resistance in HIV-1, which includes many cases of parallel molecular evolution.

13. Jabara CB *et al.*: **Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID**. *Proc Natl Acad Sci U S A* 2011, **108**:20166-20171.

14. Palmer S *et al.*: **Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis**. *J Clin Microbiol* 2005, **43**:406-413.

15. Iyidogan P, Anderson KS: **Current perspectives on HIV-1 antiretroviral drug resistance**. *Viruses* 2014, **6**:4095-4139.

16. Longdon B *et al.*: **The evolution and genetics of virus host shifts**. *PLoS Pathog* 2014, **10**:e1004395.

17. •• Plowright RK *et al.*: **Pathways to zoonotic spillover**. *Nat Rev Microbiol* 2017, **15**:502-510.
This work explores the scenarios and factors that can prompt spillovers from animals to humans and the barriers that a pathogens sorts during zoonotic events.

18. Wain LV *et al.*: **Adaptation of HIV-1 to its human host**. *Mol Biol Evol* 2007, **24**:1853-1860.

19. Bibollet-Ruche F *et al.*: **Efficient SIVcpz replication in human lymphoid tissue requires viral matrix protein adaptation**. *J Clin Invest* 2012, **122**:1644-1652.

20. Lebarbenchon C, Stallknecht DE: **Host shifts and molecular evolution of H7 avian influenza virus hemagglutinin**. *Virol J* 2011, **8**:328.

21. Hulse DJ *et al.*: **Molecular determinants within the surface proteins involved in the pathogenicity of H5N1 influenza viruses in chickens**. *J Virol* 2004, **78**:9954-9964.

22. Sauter D *et al.*: **Tetherin-driven adaptation of Vpu and Nef function and the evolution of pandemic and nonpandemic HIV-1 strains**. *Cell Host Microbe* 2009, **6**:409-421.

23. Kluge SF *et al.*: **Nef proteins of epidemic HIV-1 group O strains antagonize human tetherin**. *Cell Host Microbe* 2014, **16**:639-650.

24. Orr HA: **The probability of parallel evolution**. *Evolution* 2005, **59**:216-220.

25. •• Bailey SF *et al.*: **What drives parallel evolution?: How population size and mutational variation contribute to repeated evolution**. *Bioessays* 2017, **39**:1-9.
Metanalysis of experimental evolution datasets to analyse the effect of population sizes on the probability of occurrence of parallel molecular evolution.

26. Cuevas JM, Elena SF, Moya A: **Molecular basis of adaptive convergence in experimental populations of RNA viruses**. *Genetics* 2002, **162**:533-542.

27. Salemi M, Vandamme A-M, Lemey P: edn 2. *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing*. Cambridge, UK; New York: Cambridge University Press; 2009 723 p.

28. Yang Z, dos Reis M: **Statistical properties of the branch-site test of positive selection**. *Mol Biol Evol* 2011, **28**:1217-1228.

29. Anisimova M, Bielawski JP, Yang Z: **Accuracy and power of bayes prediction of amino acid sites under positive selection**. *Mol Biol Evol* 2002, **19**:950-958.

30. Murrell B *et al.*: **Detecting individual sites subject to episodic diversifying selection**. *PLoS Genet* 2012, **8**:e1002764.

31. Crandall KA *et al.*: **Parallel evolution of drug resistance in HIV: failure of nonsynonymous/synonymous substitution rate ratio to detect selection**. *Mol Biol Evol* 1999, **16**:372-382.

32. Avise JC, Nicholson TH: *Evolutionary Pathways in Nature: A Phylogenetic Approach*. Cambridge: Cambridge University Press; 2006 286 p.

33. Harvey PH, Pagel MD: *The Comparative Method in Evolutionary Biology. Oxford Series in Ecology and Evolution*. Oxford; New York: Oxford University Press; 1991 239 p.

34. Pagel M: **Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters**. *Proc R Soc Lond B* 1994, **255**:37-45.

35. Moore CB *et al.*: **Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level**. *Science* 2002, **296**:1439-1443.

36. • Glusman G *et al.*: **Mapping genetic variations to three-dimensional protein structures to enhance variant interpretation: a proposed framework**. *Genome Med* 2017, **9**:113.
Proposed integrated framework for mapping the effects of structural variants of proteins in the context of the human genome (part of the GVto3D Workshop).

37. Langat P *et al.*: **Genome-wide evolutionary dynamics of influenza B viruses on a global scale**. *PLoS Pathog* 2017, **13**: e1006749.

38. Ma J, Wang S: **Algorithms, applications, and challenges of protein structure alignment**. *Adv Protein Chem Struct Biol* 2014, **94**:121-175.

39. Bloom JD, Glassman MJ: **Inferring stabilizing mutations from protein phylogenies: application to influenza hemagglutinin**. *PLoS Comput Biol* 2009, **5**:e1000349.

40. Tokuriki N, Tawfik DS: **Stability effects of mutations and protein evolvability**. *Curr Opin Struct Biol* 2009, **19**:596-604.

41. Villordo SM *et al.*: **RNA structure duplications and Flavivirus host adaptation**. *Trends Microbiol* 2016, **24**:270-283.

42. Kieft JS: **Viral IRES RNA structures and ribosome interactions**. *Trends Biochem Sci* 2008, **33**:274-283.

43. Pfingsten JS, Costantino DA, Kieft JS: **Structural basis for ribosome recruitment and manipulation by a viral IRES RNA**. *Science* 2006, **314**:1450-1454.

44. Akiyama BM *et al.*: **Zika virus produces noncoding RNAs using a multi-pseudoknot structure that confounds a cellular exonuclease**. *Science* 2016, **354**:1148-1152.

45. Tycowski KT *et al.*: **Viral noncoding RNAs: more surprises**. *Genes Dev* 2015, **29**:567-584.

46. Alvarez DE *et al.*: **Role of RNA structures present at the 3′UTR of dengue virus on translation, RNA synthesis, and viral replication**. *Virology* 2005, **339**:200-212.

47. Gritsun DJ *et al.*: **Molecular archaeology of Flaviviridae untranslated regions: duplicated RNA structures in the replication enhancer of flaviviruses and pestiviruses emerged via convergent evolution**. *PLoS One* 2014, **9**:e92056.

48. Gultyaev AP *et al.*: **RNA structural constraints in the evolution of the influenza A virus genome NP segment**. *RNA Biol* 2014, **11**:942-952.

49. Bailey D *et al.*: **Functional analysis of RNA structures present at the 3′ extremity of the murine norovirus genome: the variable polypyrimidine tract plays a role in viral virulence**. *J Virol* 2010, **84**:2859-2870.

50. Wong WS, Nielsen R: **Detecting selection in noncoding regions of nucleotide sequences**. *Genetics* 2004, **167**:949-958.

51. Zhen Y, Andolfatto P: **Methods to detect selection on noncoding DNA**. *Methods Mol Biol* 2012, **856**:141-159.

52. Bernhart SH *et al.*: **RNAalifold: improved consensus structure prediction for RNA alignments**. *BMC Bioinform* 2008, **9**:474.

53. Knudsen B, Hein J: **Pfold: RNA secondary structure prediction using stochastic context-free grammars**. *Nucleic Acids Res* 2003, **31**:3423-3428.

54. Simmonds P, Smith DB: **Structural constraints on RNA virus evolution**. *J Virol* 1999, **73**:5787-5794.

55. Diehl WE *et al.*: **Ebola virus glycoprotein with increased infectivity dominated the 2013-2016 epidemic**. *Cell* 2016, **167**:1088-1098 e6.

56. Urbanowicz RA *et al.*: **Human adaptation of Ebola virus during the West African outbreak**. *Cell* 2016, **167**:1079-1087 e5.

57. • Steffen I, Simmons G: **Pseudotyping viral vectors with emerging virus envelope proteins**. *Curr Gene Ther* 2016, **16**:47-55.
Review of the use of pseudotyping methods to investigate emerging viruses such as SARS-CoV, EBOV, CHIKV, MERS-CoV, amongst others.

58. Hoenen T *et al.*: **Minigenomes, transcription and replication competent virus-like particles and beyond: reverse genetics systems for filoviruses and other negative stranded hemorrhagic fever viruses**. *Antivir Res* 2011, **91**:195-208.

59. Stobart CC, Moore ML: **RNA virus reverse genetics and vaccine design**. *Viruses* 2014, **6**:2531-2550.

60. Carter P: **Site-directed mutagenesis**. *Biochem J* 1986, **237**:1-7.

61. Dutheil JY: **Detecting coevolving positions in a molecule: why and how to account for phylogeny**. *Brief Bioinform* 2012, **13**:228-243.

62. • Doud MB, Bloom JD: **Accurate measurement of the effects of all amino-acid mutations on influenza hemagglutinin**. *Viruses* 2016, **8**.
Example of the use of deep mutational scanning to determine the mutability of the hemagglutinin in Influenza A viruses, a powerful tool for predicting sites that could be subject to parallel evolution.

63. Haddox HK, Dingens AS, Bloom JD: **Experimental estimation of the effects of all amino-acid mutations to HIV's envelope protein on viral replication in cell culture**. *PLoS Pathog* 2016, **12**: e1006114.

64. Remold SK, Rambaut A, Turner PE: **Evolutionary genomics of host adaptation in vesicular stomatitis virus**. *Mol Biol Evol* 2008, **25**:1138-1147.

65. Letko M *et al.*: **Adaptive evolution of MERS-CoV to species variation in DPP4**. *Cell Rep* 2018, **24**:1730-1737.

66. Connor RJ *et al.*: **Receptor specificity in human, avian, and equine H2 and H3 influenza virus isolates**. *Virology* 1994, **205**:17-23.

67. Steel J *et al.*: **Transmission of influenza virus in a mammalian host is increased by PB2 amino acids 627K or 627E/701N**. *PLoS Pathog* 2009, **5**:e1000252.

68. Brault AC, Huang CYH, Langevin SA, Kinney RM, Bowen RA, Ramey WN, Panella NA, Holmes EC, Powers AM, Miller BR: **A single positively selected West Nile viral mutation confers increased virogenesis in American crows**. *Nat Genet* 2007, **39**:1162-1166.

69. Stern A, Yeh MT, Zinger T, Smith M, Wright C, Ling G, Nielsen R, Macadam A, Andino R: **The evolutionary pathway to virulence of an RNA virus**. *Cell* 2017, **169**:35-46.