The expected effect of deleterious mutations on within-host adaptation of pathogens

Judith M. Fonville[a,b,c#]

Centre for Pathogen Evolution, Department of Zoology, University of Cambridge, UK[a]; WHO Collaborating Center for Modeling, Evolution, and Control of Emerging Infectious Diseases, UK[b]; Department of Viroscience, Erasmus MC, The Netherlands[c]

Running Head: Expected effect of deleterious mutations on adaptation

#Address correspondence to Dr. Judith M Fonville, jmf77@cam.ac.uk.

Word count: Abstract: 237

Text introduction-acknowledgements: 6855

Keywords: deleterious, evolution, adaptation, epistasis, compensatory, fitness valley, fitness landscape, mutation order

ABSTRACT

Adaptation is a common theme in both pathogen emergence, for example in zoonotic cross-species transmission, and pathogen control, where adaptation might limit the effect of the immune response and antiviral treatment. When such evolution requires deleterious intermediate mutations, fitness ridges and valleys arise in the pathogen's fitness landscape. The effect of deleterious intermediate mutations on within-host pathogen adaptation is examined with deterministic calculations, appropriate for pathogens replicating in large populations with high error rates. The effect of deleterious intermediates on pathogen adaptation is smaller than their name might suggest: when two mutations are required, and each individual single mutation is fully deleterious, the pathogen can jump across the fitness valley by obtaining two mutations at once, leading to a proportion of adapted mutant that is 20-fold lower than for the situation where all mutants are neutral. The negative effects of deleterious intermediates are typically substantially smaller, and outweighed, by fitness advantages of the adapted mutant. Moreover, requiring a specific mutation order has a substantially smaller effect on pathogen adaptation than the effect of all intermediates being deleterious. These results can be rationalized when calculating the number of routes of mutation available to the pathogen, providing a simple approach to estimate the effect of deleterious mutations. The calculations discussed here are applicable when assessing the effect of deleterious mutations on the within-host adaptation of pathogens, for example in the context of zoonotic emergence, antigenic escape, and drug resistance.

IMPORTANCE

Adaptation is critical for pathogens after zoonotic transmission into a new host species, or to achieve antigenic immune escape and drug resistance. Using a deterministic approach, the effects of deleterious intermediate mutations on pathogen adaptation are calculated whilst avoiding commonly made simplifications that do not apply to large pathogen populations replicating with high mutations rates. Perhaps unexpectedly, pathogen adaptation does not halt when the intermediate mutations are fully deleterious. Negative effects of deleterious mutations are substantially outweighed by fitness gains of adaptation. To gain an understanding of the effect of deleterious mutations on pathogen adaptation, a simple approach is introduced that counts the number of routes available to the pathogen with and without deleterious intermediate mutations. This methodology enables a straightforward calculation of the proportion of the pathogen population that will cross a fitness valley or traverse a fitness ridge, without reverting to more complicated mathematical models.

INTRODUCTION

The fitness landscape of a pathogen is likely to have a rugged shape and consist of multiple optima. Reductions in fitness occur when underlying combinations of genetic mutations lead to an unfit or deleterious phenotype, creating depressions in the fitness landscape. One phenomenon causing sharp peaks and troughs in the fitness landscape is sign epistasis, where a beneficial adaptation involves a combination of individually deleterious mutations (1–5). In the case where intermediate mutations are less fit than the wild type and adapted virus, a fitness valley is created – a barrier of disadvantageous mutations hampering the access to other landscape regions (4, 6). If there is a specific order in which mutations can occur without compromising the fitness, for example where compensatory or obligatory co-mutations can remove the deleterious effect of another mutation, the landscape contains a fitness ridge. Such fitness valleys and ridges are commonplace in virology, as will be illustrated with examples drawn from the influenza field.

During zoonotic overspill infections of an avian influenza virus into humans, pressure exists for the pathogen to adapt to this possible new host (7–9). The virus was fit in its original host, and needs to be fit in the new host, but this adaptation process might require deleterious intermediate mutations. The need for adaptation of a zoonotic pathogen is illustrated by the requirement of a combination of mutations in avian A/H5N1 virus for airborne transmission

between mammals (10, 11). Interestingly, two of the mutations that were found necessary to confer airborne transmissibility, polymerase basic protein 2 (PB2) E627K and polymerase basic protein 1 (PB1) H99Y (11, 12), increased the fitness of the adapted virus if both mutations occurred together, as inferred from substantially larger plaque sizes than the wild type, yet each individual mutation decreased the fitness compared to the wild type virus (12). Similarly, Imai *et al.* showed that the receptor-binding mutations N224K and Q226L in the hemagglutinin gene (HA), required for an airborne transmissible phenotype of A/H5N1, reduced the stability of HA, but could be compensated for by mutation T318I in the HA stalk, which restored protein stability (10). Although the stabilizing mutation was not essential for virus survival, it did substantially increase viral fitness.

Another example of deleterious intermediate mutations is escape from pre-existing host immunity through fitness-decreasing mutations for antigenically variable pathogens (13, 14). For example, the altered receptor-binding avidity and lower replication resulting from the antigenic escape mutation HA K165E in A/Puerto Rico/8/1934(H1N1) could be compensated for by mutations in HA or the neuraminidase (NA) (15, 16), and stabilizing mutations were required to occur prior to the introduction of immune-escape mutations in influenza A/H3N2 virus (17). Similarly, there are numerous examples where antiviral-resistance conferring mutations come at a fitness cost for the virus, that can be compensated for by other mutations: several neuraminidase substitutions can

occur and have occurred as either permissive or compensatory mutations to counteract the adverse fitness effects of the oseltamivir-resistance mutation NA H275Y in influenza A/H1N1 virus (18–20); and similarly the I222V NA mutation in A/H3N2 partially restored the viral fitness-decreasing oseltamivir-resistance mutation NA E119V (21).

The name “deleterious mutation” may suggest that the existence of such mutations is unlikely, and thus the expectation that the crossing of a fitness valley comprised of individually deleterious mutations is difficult, if not impossible. Indeed, when evolution is described as an adaptive walk or directed evolution, adaptation consists of a series of incrementally neutral or beneficial mutations, and thus the crossing of the fitness valley would be technically impossible (22–26). Also the possibility of obtaining several mutations at once, to “jump across” a fitness valley, is not considered in some theoretical models (27–29). A methodological framework frequently used to study pathogen evolution making such assumptions is the “strong-selection-weak-mutation” (SSWM) approximation (30, 31). Models using the SSWM assumption describe the evolutionary trajectory of a population where selective sweeps cause the sequential fixation of advantageous mutations, whilst deleterious or neutral mutations are disregarded (32–34).

Here, we demonstrate how pathogens replicating in large population sizes and with high error rates, such as RNA viruses, can cross fitness valleys, based on a

simple, and appropriate, deterministic model of within-host pathogen evolution. Instead of following the evolution of a pathogen population toward fixation of certain mutations, as is for example done in SSWM models, we calculate the probability of a randomly drawn virion from the within-host virus population after initial infection with a single genotype to have obtained a set of mutations after a given number of replication rounds. This probability, when multiplied with the pathogen population size, gives the expected number of virions with this specific set of mutations. In other words, the probability is directly related to the proportion of viruses in the total within-host population with this set of mutations.

This probability of a virion to have a set of mutations is highly relevant, because increased proportions are likely to correspond to increased probability of spread of such mutants. When the bottleneck of transmission is narrow, for example in the case where a single virion is transmitted to the next host, the probability describes the likelihood that infection of the next host will begin with the adapted virus. Alternatively, if the bottleneck is wide, the expected proportion of adapted virus at the start of the infection of the next host can be calculated and used to estimate the chances of further adaptation.

In this manuscript, we calculate the effects of deleterious mutations, fitness valleys and fitness ridges on within-host pathogen evolution using a straightforward deterministic model (35, 36). Such a deterministic probability calculation is appropriate for studying the dynamics and evolution of large

155 populations with asexual reproduction at a high mutation rate, such as most RNA
156 viruses, because stochastic effects play a limited role. The proportion of the
157 adapted mutant is calculated for varying valley depths (i.e. the fitness of the
158 deleterious mutation) and breadths (i.e. number of deleterious mutations). Finally,
159 we also describe the probability of traversing a fitness ridge, for varying numbers
160 of mutations that need to be acquired in order.
161

METHODS

The within-host population dynamics of virus mutants were calculated as deterministic probabilities, based on the methodology described previously (35, 36). In this calculation, the errors made by the virus polymerase are represented by an error rate, and form the source of introduction of mutations, but the approach can equally be used for non-viral pathogens, where mutations are introduced through another mechanism. The probability of accumulating mutations and the within-host evolutionary dynamics of the virus population are explored as a function of the fitness of the wild type, intermediate and adapted mutants.

Calculating virus populations

A virus type *j* is a virus with a particular set of mutations. The probability of each virus type (N_j) after a replication round is given by the sum of contributions from each type in the previous replication rounds:

eq. 1 $$N_j(t) = \sum_i [N_i(t-1)\mu_{ij}]$$

Where μ_{ij} is the probability of type *i* mutating to type *j*, and each type contributes exactly its expected value. If the mutation rate μ_{ij} is low, the main contribution to the proportion of the population that is N_j at time *t* will be from the proportion of the population that was N_j at time *t-1*, and a smaller contribution from virus type

185 N_i at time *t-1* that mutated into type N_j. The probability of mutation μ is calculated
186 as follows:
187
188 **eq. 2** $$\mu_{ij} = \prod_{\{m=0\}}(1-r)\ \prod_{\{m=1\}} r$$
189
190 Where *r* is the polymerase error rate. Thus, μ_{ij} is the product of the probabilities
191 of non-mutation (1-*r*) for the set {*m*=0}, i.e. positions for which no mutation is
192 required, and of the probability for mutation (*r*) for the set of positions that need to
193 mutate {*m*=1}.
194
195 **Accounting for fitness values**
196 The deleterious and beneficial selection values were incorporated by adjusting
197 the "progeny" of each virion to express the fitness disadvantage or advantage in
198 each genome replication step, prior to the start of the next generation. The
199 starting population (generation zero) consists only of zero-mutant, the starting
200 virus. After the first replication round (*equation 1*), the population of each type N_i
201 is multiplied by its relative fitness f_i, and the population is normalized (such that $\sum_i$
202 N_{i_adj} = 1) through division by the sum of the fitness-weighted prevalence all
203 types:
204
205 **eq. 3** $$N_{i\_adj}(t) = \frac{N_i(t) * f_i}{\sum_i [N_i(t) * f_i]}$$
206

The N_{i_adj} represent the populations at the start of the next genome replication step, and are used as N_i in *equation 1* in the multiplication with the mutation matrix. When calculating the effect of deleterious mutations the fitness *f* is varied, and a "fully deleterious mutant" has a fitness of $f = 0$, which causes relative increases in the probability of the other virus types in the total virus population.

Unless otherwise noted, the fitness of the wild type (i.e. starting) virus, and the final type with the full set of mutations of interest, the "adapted virus", is neutral: $f = 1$.

Because this model normalizes the virus population via *equation 3*, and accounts for back-mutations in *equation* 2, the results are slightly different from the short-hand formula introduced in *equation 4* in the results section. For this reason, calculating the number of routes (see below) is a fast and informative approach to very closely approximate the probability of a certain set of mutations, but is not analytically identical to the modeling results.

Stochastic model runs

In addition to the deterministic modeling results above, a set of stochastic discrete-time multi-type branching process simulations were run, see also Russell *et al*. (35). The starting population of a single virion expanded exponentially with a branching factor of 32 (leading to 10^3 virions produced per infected cell, after the two genome replication steps), until the population size

exceeded 10^{10} virions, from which point onwards the branching factor was set to 1. For each genome replication step, the expected number of each mutant type was determined with a Poisson distributed random variable with the expectation value based on the mutation matrix shown in *equation 2* and the number of virions of each mutant type existing before the replication step, as was done for *equation 1*. This number was then multiplied by the relative fitness of each type, and rounded to the nearest integer, prior to starting the next genome replication step. We performed 10,000 stochastic runs for each of the 101 settings of fitness of the deleterious intermediate mutants (between 0 and 1 in steps of 0.01). The intensity of the shade of the pale red and blue colors is calculated based on the $\log_2$ of the number simulation runs that have the resulting proportion of double mutants for each fitness setting; the average proportion across the 10,000 runs per fitness setting is indicated with the line connecting the circles.

Determining the route

We introduce the terminology “through singles” to mean the process by which the two mutations are acquired through separate single mutations occurring in distinct replication rounds, and “through doubles” to mean to process where two mutations are achieved by mutating both sites in a single replication round. We investigated the probability of a double mutant to occur through doubles by setting the μ for single mutations to zero in the mutation matrix. The through singles probability is calculated as the difference between the probability when all routes are allowed, and the through doubles probability.

Similarly, to calculate how often the required set of mutations was achieved through a specific order, the fitness of any non-order mutant was set to zero. The difference between the probability calculated if any order is allowed and the probability when only a specific order is available determines the probability of non-order mutation routes.

The fraction of available routes is calculated as the number of available routes given the constraints divided by the number of original routes.

Parameter choice

The mutation rate is parameterized by the current best estimate for the influenza virus polymerase error rate ($r = 1\times10^{-5}$ mutations per site, per genome replication (37, 38)), and can trivially be adjusted for other mutation rates – indeed all results in the manuscript are *not* specific to influenza virus, or viruses in general, but to all large populations where mutations occur.

A “replication round” in this manuscript refers to any step in which RNA is synthesized, because in each round of replication polymerase errors can be introduced. For influenza viruses, where vRNA is replicated into cRNA and then cRNA is copied into vRNA, there are (at least) two replication rounds per cell cycle. Results are shown after 20 viral replication rounds, which corresponds to five days of influenza virus infection (where each replication round lasts around 6

hours, and virions exit the cell after 12 hours), but again, the number of replication rounds can be varied in the equations above.

RESULTS

Adaptation depends on the fitness of the deleterious intermediate mutations

When all mutations are neutral, a simple probabilistic calculation of mutation accumulation closely approximates the probability that any randomly drawn virion from the within-host virus population in an individual initially infected with a single genotype, would have mutated the *m* sites of interest over time (the number of replication rounds, *t*):

eq. 4 $p(m,t) = t^m r^m$

This equation multiplies the probability of getting *m* mutations (based on the polymerase error rate *r*), r^m, with the number of combinatorial options to acquire these *m* mutations over *t* generations (t^m), see also Russell *et al.* (35) and Gokhale *et al.* (39).

The probability that a given virion will have mutated $m = 2$ sites after $t = 20$ replication rounds with a polymerase error rate $r = 1\times10^{-5}$ is approximated by *equation 4* as 4×10^{-8}. Naturally, this probability of observing both sites mutated will be less if either of the individual mutations is deleterious. If both individual mutations are deleterious, the pathogen will have to get across a fitness valley. The fitness of each single deleterious mutant determines the likelihood of the

301 virion to cross this fitness valley. A fully deleterious mutation has a relative fitness
302 of 0, which means that no progeny is made from these virions at all, whilst for a
303 relative fitness of 0.5 half as much progeny descends from these virions
304 compared to virions with a relative fitness of 1.
305
306 *Figure 1A* explores how the probability of a pathogen to cross a fitness valley
307 depends on the deleterious effect of the intermediate mutations. In this scenario,
308 each individual mutation is equally deleterious, and the wild type (starting) and
309 the virus with the two required mutations (the “adapted virus”) have neutral
310 fitness. The blue line shows the deterministic probability of a virion to be a double
311 mutant as a function of the fitness of the intermediate mutants. In the neutral
312 scenario without any fitness valley, where the fitness of each intermediate mutant
313 is 1, the probability of the double mutant after 20 replication rounds is, as
314 approximated above, 4×10^{-8}. As the relative fitness of each intermediate mutant
315 decreases toward zero (fully deleterious), the probability that any random virion
316 in the virus mixture is a double mutant decreases to 2×10^{-9}. Note that, despite
317 the two intermediate mutants being fully deleterious, the probability of the double
318 mutant is only twenty-fold lower than without the fully deleterious fitness valley.
319 The pale region is composed of 10,000 stochastic model simulations for each of
320 the 101 different settings of fitness *f*. The average of these runs, indicated by the
321 connected circles, is somewhat lower than the deterministic calculations, most
322 visibly for intermediate values of deleterious fitness. These stochastic simulations

highlight that even though stochastic effects may play a role, double mutations do occur regularly, even when the intermediate mutants are fully deleterious.

Such double mutants can arise when both mutations were acquired simultaneously in a single replication round, the “through doubles” mechanism. The purple line in *Figure 1B* shows how much this mechanism of acquiring both mutations at once contributes toward the likelihood of a virion being a double mutant. Note that this probability is not affected by the relative fitness: because the deleterious intermediates were never formed when two mutations were obtained at once, the virions avoid having to incur the designated fitness cost. The fitness valley is not crossed, but the virus "jumps" over it. The cyan line describes the alternative “through singles” mechanism where the two single mutations were obtained in distinct replications rounds – the situation in which the virions did incur the deleterious cost of the intermediate, and actually crossed through the valley. In *Figure 1B*, it can be seen that this contribution depends strongly on the fitness of the intermediate single mutants. If the intermediate mutants are neutral, or have a high relative fitness, the through singles mechanism is the main contributor toward the probability of acquiring a double mutant (right hand side of *Figure 1B*). However, when the intermediate single mutants are highly deleterious the main contributor to the probability of a double mutant is the through doubles mechanism.

345 Returning to *Figure 1A*, it appears that the through doubles mechanism is less
346 sensitive to stochastic variations than the through singles mechanism, and the
347 deviation between the deterministic model and stochastic results is largest for
348 intermediate values of deleterious fitness. Here, the non-negligible deleterious
349 cost causes stochastic loss of single mutants before the second mutation occurs.
350 Such stochastic losses are less prominent for fit intermediates ($f = 1$) or identical
351 or similar to the losses calculated in the deterministic model for highly deleterious
352 intermediates.
353

354 **An intuitive understanding: counting the number of "routes"**

355 Although the "through doubles" and "through singles" mechanisms in *Figure 1*
356 both require two polymerase errors, the probability of which is r^2, the relative
357 contribution of through singles to obtain two mutations is larger than through
358 doubles at $f = 1$. This phenomenon can be understood by considering "the
359 number of routes". The through doubles route can happen once in each
360 replication round, and thus in t different ways (here 20). However, to get two
361 single mutations, there can be e.g. single mutations in two subsequent rounds
362 (for which there are 19*2 options – the factor of two accounts for which of the
363 mutations is first), or single mutations in two replication rounds separated by a
364 replication round without mutation (for which there are 18*2 options), and so on,
365 until there is one single mutation in round 1 and one in round 20 (for which there
366 are 2 options only). The sum of these possibilities is 380 routes, which when
367 combined with the 20 routes of through doubles, corresponds to 20^2 ways to

obtain two mutations in twenty replication rounds, i.e. the factor t^m in *equation 4*. Although the term t^m is nothing more than a combinatorial factor, it was found that explicitly analyzing the number of routes represented by this term is useful for reasoning about the expected effects of varying fitness valley shapes.

If the single mutants are fully deleterious, the effective number of routes to obtain a double mutant through singles is 0, because a double mutant can never arise from a single mutant if single mutants do not have progeny. In this situation, only the through doubles mechanism is possible, and thus 20 out of the original 400 routes remain, causing a reduction in the probability of a double mutant by a factor 0.05 (the probability decreased from 4×10^{-8} to 2×10^{-9}).

A general description to calculate the effective number of routes to obtain a double mutant through singles for any relative fitness *f* of the single mutations can be given as well: if single mutations happen in successive replication rounds (*delay* = 1), the fitness cost is incurred once, if they are separated by one replication round (*delay* = 2), the fitness cost is incurred twice, whilst if the single mutations are 19 generations apart (*delay* = 19), the fitness cost *f* is incurred 19 times. In total, the effective contribution to the number of routes through singles weighted by the incurred deleterious cost is given by:

eq. 5 Effective number of routes $= \sum_{delay=1}^{delay=(t-1)} [2 * (t - delay) * f^{delay}]$

Where t is the number of replication rounds, as before, and *delay* is the time lapse in replication rounds between the two single mutations (for t = 20 replication rounds, the maximum *delay* is 19). The factor of two reflects the fact that the single mutations can be acquired in two different orders; the term (t - *delay*) represents the number of options for any given delay (19 for a delay of 1, 18 for a delay of 2, etc.); while f^{delay} is the penalty term for the incurred fitness cost over *delay* rounds of replication.

Figure 2 illustrates how the number of effective routes is composed of the contribution of the through doubles mechanism (in grey), and the different single-single mutation routes. If the relative fitness of the intermediate mutants is 0, the through doubles mechanism is the only contributor to the number of routes, as was seen in *Figure 1*. Again, the effective number of routes for the through doubles mechanism is independent of the fitness of the deleterious single mutants, as any deleterious fitness cost is not incurred. *Figure 2* also demonstrates that the effective number of routes of two single mutations separated by many replication rounds (e.g. *delay* = 19, in pink) is substantially smaller than the effective number of routes for two subsequent single mutations (*delay* = 1, in red). The reason for this is twofold: first, if there are 20 replication rounds, there are 19*2 routes to generate two single mutations 1 generation apart, whilst there are only 1*2 routes to obtain two single mutations 19 generations apart. Second, any deleterious effect of the single mutants is incurred for more replication rounds if the delay between the two single mutations

414 is longer, and thus the contribution of these single-single routes with longer
415 delays decreases even more as the intermediate mutants become more
416 deleterious.
417
418 *Table 1* shows that the total number of routes increases (t^m) as more mutations
419 are required, listing the results for two to seven mutations required. It also shows
420 the fraction of routes remaining when all intermediate mutants are fully
421 deleterious. In the case where five mutations are required, for example, this
422 means that all individual and combined intermediates (and thus all single, double,
423 triple and quadruple mutants) are deleterious. Because all intermediate mutants
424 are fully deleterious, all mutations have to be acquired at once, for which there
425 are t options: so t out of t^m routes remain. Although there were initially many
426 routes to acquire 5 mutations (20^5), only 20 remain.
427
428 When comparing the fraction of available routes for the situation where 5
429 mutations are required, 6.3×10^{-6}, with the fraction when 2 mutations are required,
430 5.0×10^{-2}, it is clear that the fraction of the available number of routes decreases
431 greatly as the number of intermediate deleterious mutations increases. Note, in
432 addition to a larger number of deleterious intermediate mutants slowing down the
433 viral adaptation, there is also the increased difficulty of acquiring more mutations
434 in the first place (which is given by r^m).
435

When a set of mutations is required of which only some are deleterious, the ratio of the effective number of routes compared to the total number of routes when that subset of mutations was not deleterious is the same as the fraction of routes available for the number of deleterious mutations. As an example, consider the situation where 5 mutations are required, and two of the mutations need to be acquired as a double. When none of the 5 mutations are deleterious, there are 20^5 routes (t^m). When the two mutations are individually fully deleterious, the second mutation of the double pair needs to occur simultaneously with the first mutation of the pair. As a result, the timing of mutation, for which there are 20 options if there are 20 replication rounds, needs only to be established for 4 mutations, as the timing of the last mutation needs to be identical to the timing of the other mutation in the pair. Hence, when two of the five mutations need to be acquired as a double, there are 20^4 routes left, and the fraction $20^4/20^5$ is 0.05, see 2 mutations required in *Table 1*.

The effect of deleterious intermediates is outweighed by the fitness advantage of adaptation

In the previous calculations, we studied situations where the fully adapted mutant had neutral fitness, and the number of available routes could directly be used when calculating the probability of a virion being a fully adapted mutant. Next, we investigate whether the deleterious cost of an intermediate mutant can be outweighed by the fitness gain that would be obtained upon achieving the full set of mutations, for example as a result of obtaining a certain beneficial phenotype

such as antigenic escape or increased replication. In *Figure* 3, the probability of any virion being a double mutant after t = 20 replication rounds is indicated by color, as a function of the relative fitness advantage f of the double mutant, varied from 1 to 4, and the relative fitness of the deleterious single mutants, varied from 0 to 1.

Figure 3 demonstrates that the probability of a random virion having obtained two mutations after 20 replication rounds varied relatively little with the fitness of the deleterious intermediates (along the x-axis): for example, the maximum change in the neutral scenario (see Figure 1) was 20-fold, which corresponds to 1.3 units on a $\log_{10}$ axis. In contrast, the fitness gain of the double mutant causes changes across 9 orders of magnitude, and this fitness gain therefore appears to be the main determinant of the probability of a double mutant. The mechanism behind these observations is that, in contrast to the deleterious cost, which is often incurred only briefly, or avoided altogether by obtaining both mutations at once, the fitness benefit of the double mutant is incurred in every single replication round once it has arisen, hence exponentially increasing its presence in the pathogen population. For fitness f = 4 of the double mutant, the minimum probability of a virion to be a double mutant, across all fitnesses of the single mutants, was 0.99998.

Adaptation via fitness ridges: compensating mutations imposing order

Requiring mutations to occur in a specific order is a special case of deleterious mutations: imagine the scenario in which two mutations are required, whereby one single mutation α compensates for or removes the deleterious effect of the other single mutation β (both the single intermediate α and the double mutant $\alpha\beta$ have neutral fitness). To understand the effect of such imposed order on the probability of obtaining a certain mutant, again the number of effective routes calculation is helpful.

In the scenario where the double mutant and mutation α are neutral, while mutation of only site β is fully deleterious, the two ways toward the double mutant are to either get both mutations simultaneously, or to obtain the non-deleterious, compensating mutation α before mutating the site β. As explained above, there are 20 routes out of 400 to obtain both mutations at once, and half of the remaining 380 routes will have had the compensating mutation α prior to mutation of site β: in total 210 out of the 400 routes remain. Even though mutation β was fully deleterious, 52.5% of the routes are still available, thus incurring only a 2-fold reduction in the total proportion of double mutant.

Table 2 shows the number of allowed routes when all mutations need to be obtained in order, for situations where two to seven mutations are required. As more order is required, the reduction in fraction of allowed routes increases: for

503 stringent ordering of five mutations, only 1.3% of the routes remain, which is less
504 than the 52.5% for requiring order of two mutations.
505
506 Because imposing order does not necessarily require multiple mutations to occur
507 at once, the fraction of available routes is substantially larger in the situation
508 where order is required than for the situation where all intermediate mutants were
509 deleterious. For example, when requiring five mutations, 1.3% of the routes
510 remain if requiring specific ordering of these five mutations, whereas only
511 0.00063% of the routes remained when all intermediates were deleterious (see
512 *Table 1*). Thus, using available fitness ridges is always, and often considerably,
513 easier than jumping across or crossing a fitness valley.
514
515 The red line in *Figure 4A* also shows that the effect of requiring order on the
516 probability of a virion obtaining two mutations is relatively small, even if the non-
517 ordered single mutant is fully deleterious (compare $f = 0$ and $f = 1$), especially
518 when compared to the situation where both single mutants were deleterious (blue
519 line). The pale region is composed of 10,000 stochastic model simulations for the
520 fitness ridge, in red, and fitness valley, in blue. The average of the stochastic
521 runs, shown as circles, again indicate how traversing via a fitness ridge is
522 substantially more likely than jumping or crossing a fitness valley. Moreover,
523 stochastic effects play virtually no role for the outcome of a virion that can travel
524 via a fitness ridge, as the results are very similar to the deterministic model.
525

Figure 4B shows the contribution of the three different mechanisms that could lead to a double mutant. First, the virus could follow the imposed order and travel via two subsequent mutations along the fitness ridge. Second, the virus could simultaneously mutate both sites and jump across the surrounding fitness valley. Both of these mechanisms do not violate the imposed order, and their contributions in *Figure 4B* are independent of the fitness of the deleterious single mutant. Third, the virus could disobey the imposed order and obtain the non-ordered single mutation first, which will incur the deleterious cost for a certain time. The contribution of this latter mechanism depends on the fitness of the non-ordered deleterious single mutant, and becomes zero when $f = 0$. As f decreases, the ridge in the fitness valley becomes the main mechanism toward obtaining the set of mutations. In general, following the imposed order and travelling via the fitness ridge becomes more attractive as the fitness valley deepens and widens, as non-ordered single intermediates become even less viable, and obtaining simultaneous mutations even less likely.

DISCUSSION

Using probabilistic calculations on within-host genetic evolution, we found that the effect of a fitness valley of deleterious intermediate mutations on adaptation is much smaller than might be expected, and that the effect of requiring a specific order for mutations to occur is even smaller. In coinfected individuals, mechanisms such as reassortment (if the mutations of interest are on separate genes) and recombination are additionally affecting the ability of viruses to overcome fitness valleys, processes that are not currently included in the model. Instead, we calculated, based on within-host evolution of a single starting virus genotype infecting an individual, the probability of any virion getting a set of host-adaptation mutations. This probability is directly related to the expected proportion of adapted mutant in the total population. The equations and calculations presented in this work can be used in any situation with fitness valleys and ridges where the deterministic assumptions are fulfilled, and the population reproduces asexually. As a result, this method can aid the study on the effects of deleterious mutations in a wide range of pathogens, including, for example, tuberculosis and HIV (5, 40, 41).

The methodology of counting the number of routes is a straightforward approach to calculate the effect of deleterious intermediate mutants, and understand the ways by which pathogen populations traverse fitness valleys and ridges. For example, in the situation where a virus requires two mutations that are each

565 individually fully deleterious, the evolution is not halted, as this trap is avoided by
566 acquiring multiple mutations at once. In addition, if viruses need to follow a
567 specific order of mutation, the out-of-order intermediate mutants can be
568 described as fully deleterious. If only a handful of mutations need to be acquired
569 in order, the influence on the adaptation of the virus is minimal, because the
570 compensating mutations will occur beforehand without much difficulty: many
571 routes are still available. Indeed, a key finding of this research is that fitness
572 disadvantages of intermediate mutants sometimes have a great effect on the
573 proportion of adapted mutant, but only when a large number of intermediate
574 mutants are deleterious, and their fitness cost is large.
575
576 Although some studies recognize the importance of deleterious intermediates
577 and the crossing of fitness valleys to the overall adaptive evolution of pathogens
578 (27, 29, 42, 43), and the possibility of multiple simultaneous mutations to
579 overcome such fitness valleys (44), various other models assume a strong-
580 selection-weak-mutation paradigm (32–34, 45), ignoring any adaptive trajectories
581 that require the crossing of a fitness valley. Such assumptions might be
582 appropriate for small population sizes, or pathogens with low mutation rates (5,
583 30). However, for a pathogen with a large population size and high mutation rate,
584 these SSWM assumptions are substantially violated. For influenza virus for
585 example, the mutation rate is around 1 mutation per 10^5 nucleotides, per round of
586 genome replication ($r = 10^{-5}$) (37, 38), and the population size P easily exceeds

587 10^{10} virions in a single host, hence the SSWM conditions $4rP << 1$ or
588 $rP<<1/\ln(Ps)$, where s is the fitness increase, are not fulfilled (30, 46).
589
590 It should be noted though, that the assumptions of the deterministic
591 approximation are violated in parts of *Tables 1* and *2* (which showed the
592 numbers of routes for scenarios where up to 7 mutations are required). When the
593 inverse of the error rate to the power of the number of mutations required ($1/r^m$) is
594 larger than or comparable to the population size, stochastic variations may
595 become relevant. This was seen in *Figure 1A*, where for a population size around
596 10^{10} the two mutations were not acquired as readily as the deterministic model
597 would have suggested. The smaller the population size is in comparison to ($1/r^m$),
598 the higher the likelihood that stochastic effects decrease the expected proportion
599 observed in the pathogen population. Especially when large numbers of
600 mutations are required, the expected number of times the adapted mutant occurs
601 will be small, if not zero, when taking account of the population size. Moreover,
602 stochastic death of rare intermediate or fully adapted mutants will further affect
603 the observed proportions. In a single host, one can multiply the effective virus
604 population size, say 10^{10}, with the probability of interest, e.g. 6.65×10^{-7} for a
605 random virion being a double mutant when the intermediate single mutants were
606 fully deleterious, the starting mutant neutral and the double mutant has a fitness f
607 = 1.5, to get the expected number of virions with the mutations of interest, here
608 6650. In the context of transmission, where frequently small populations
609 consisting of less than a handful of virions are estimated to start new infections

610 for e.g. influenza virus, HIV and hepatitis C (47–50), the probability of any virion
611 being a mutant of interest is informative for epidemiological studies and risk
612 assessment. In the biologically implausible case (but just to clarify) that each
613 virion has the same chance of being transmitted and starting the next infection, if
614 only a single virion starts the next infection, the chance that the next host is
615 infected with only the adapted virus is identical to the proportion of this adapted
616 virus in the donor host.
617
618 Interestingly, evolutionary models have also been used to improve the
619 understanding of the developmental stages and processes in cancer, and to
620 increase the efficacy of treatment regimes (51). In the evolution of a cancerous
621 cell, there is often a fitness valley to be crossed before the cell is able to progress
622 to expansive, uncontrolled growth (51). As with many other evolutionary models,
623 models for cancer evolution are focused on population-level adaptation. The
624 cancer literature describes two main mechanisms for the population-level
625 crossing of the fitness valley: sequential fixation, whereby the full cancerous cell
626 population acquires one mutation, and only after fixation of the first mutation, the
627 second mutation becomes fixated; and stochastic tunneling, whereby the second
628 mutation establishes prior to fixation of the first mutation (52, 53). Stochastic
629 tunneling describes the probability of fixation on a population-level, and allows for
630 sequential but not necessarily simultaneous mutations, in contrast, the
631 deterministic calculations above describe the probability of obtaining multiple
632 mutations simultaneously by any single unit (cell or pathogen or virion), and can

633 be converted to an expected proportion in the population, but does not equate to
634 fixation.
635
636 An advantage of the deterministic calculations used here is that this approach
637 can easily be adjusted to encompass more complicated schemes of required
638 mutations and associated fitnesses. It is, for example, not limited to the
639 investigation of effects of deleterious intermediates, but can also be used when
640 individual mutations are neutral or beneficial, and the combined mutation is
641 synergistic, for example mutations at positions 138 and 229 in the non-structural
642 protein 1 (NS1) (54) and 147, 339 and 588 in PB2 (55) of influenza A/H5N1 virus
643 affecting virulence.
644
645 The implementation can be easily changed to model other situations, for example
646 i) where a mutation has a fitness effect when in the vRNA, but not when
647 occurring in the cRNA, as those molecules are not transcribed into mRNA and
648 translated into protein; or ii) to encompass delayed phenotypes of mutations (56,
649 57) whereby deleterious or advantageous fitness effects are not fully observed as
650 the respective proteins are only generated in a meaningful amount at a later time.
651 Such mechanisms might alter the likelihood of deleterious and adaptive
652 mutations occurring, for example by a deleterious mutation arising as non-
653 deleterious in the cRNA, and the compensatory mutation arising in the next
654 replication round, such that deleterious vRNA is never formed and both
655 mutations were effectively neutral.

In the stochastic model, the branching factor governs population growth, and as a result of the founder effect this leads to mutations arising earlier in time achieving higher proportions; an effect that can be seen as a banded simulation runs, for example in Figure 4A, where the top band shows that there are fewer instances (lower red intensity) but higher proportions of double mutants for mutants arising early. Again, the implementation of this model can be adjusted such that the branching factor varies in both steps of replication to match the specific parameters for the virus of interest.

The calculations enable both estimating the likelihood of crossing fitness valleys, as well as the probability of passing a narrow fitness ridge. The work presented here on assessing the effect of fitness ridges and required order is relevant, for example, in modeling antibiotic resistance (26), and pyrimethamine resistance of malaria (58). The equations can easily incorporate variable mutation rates (59), which may be useful to investigate different polymerase error rates of influenza virus (60), and to account for varying replication fidelity of HIV reverse transcriptase along different positions in the genome (61). This feature is also important in the evolutionary modeling of cancer, where disease progression often involves the acquisition of decreased genetic stability, and thus an altered mutation rate (62).

678 Counting the number of routes is also a method that could be applied to
679 determine the multiplicity of drug therapy, as the acquisition of drug-resistant
680 mutations might be avoided by a treatment regime shaping the fitness valley
681 deep and wide enough to prevent the pathogen from crossing, an approach that
682 has been described with combination therapy for example in influenza,
683 tuberculosis and HIV treatment (61, 63–65). In the context of drug therapy,
684 Ribeiro *et al.* already noted that for totally defective intermediate HIV mutants, all
685 higher order strains have to be produced directly from wild-type, i.e. only allowing
686 routes where all mutations are acquired at once; they also described that for
687 smaller selective disadvantages, a k+1 mutant is most likely produced from a k-
688 point mutant, i.e. a qualitative phrasing of our quantitation of the contribution to
689 the number of routes from obtaining two subsequent mutations for less-
690 deleterious intermediates (66).
691
692 The introduction of fitness valleys can also be exploited as a mitigation strategy
693 for infectious diseases. One could, for example, design vaccines that require a
694 pathogen to obtain destabilizing mutations to enable immune escape. Models of
695 pathogen evolution can help to establish whether such approaches will
696 completely stall adaptation of the pathogen, or with what likelihood the created
697 fitness valleys would be crossed. Moreover, such approaches could also be used
698 to explore alternative routes as a result of epistatic interactions that might allow
699 deleterious mutations to occur if acquired in the right order (15, 17, 20).
700

701 The successful and efficient invasion of zoonoses into the human population is
702 often thought to be constrained by the existence of deleterious mutations on the
703 path to adaptation. Therefore, calculations on the effects of fitness valleys are of
704 critical importance in pandemic risk assessment of emerging pathogens (8, 33,
705 35, 44, 67, 68), and in addition to inform the cost-benefit analyses of gain-of-
706 function experiments and dual-use research of concern.
707
708 In summary, the ability to calculate the effect of deleterious mutations and order,
709 and to understand the results with the description of the number of available
710 routes, helps to assess the expected impact of fitness valleys and ridges in
711 pathogen evolution, with applications in drug resistance, immune escape, and
712 zoonotic risks assessments.

ACKNOWLEDGEMENTS

The author declares no conflict of interest. I would like to acknowledge André Brown and Colin Russell for previous designs of the deterministic model, Sander Herfst and Gabriele Neumann for providing input on examples of deleterious mutations, and Ana Mosterín-Höpping, Leah Katzelnick, Ramona Mögling, David Pattinson and Derek Smith for careful reading of previous manuscript versions. This work was supported by the award of a Fellowship in Biomedical Informatics from the Medical Research Council UK (MR/K021885/1) and a Junior Research Fellowship from Homerton College Cambridge to JMF; and the award of HHSN272201400008C (NIAID Centres of Excellence for Influenza Research and Surveillance) to the Center for Pathogen Evolution.

REFERENCES

1. **Whitlock MC, Phillips PC, Moore FB-G, Tonsor SJ**. 1995. Multiple fitness peaks and epistasis. Annu Rev Ecol Syst **26**:601–629.

2. **Maisnier-Patin S, Andersson DI**. 2004. Adaptation to the deleterious effects of antimicrobial drug resistance mutations by compensatory evolution. Res Microbiol **155**:360–369.

3. **Wilke CO, Lenski RE, Adami C**. 2003. Compensatory mutations cause excess of antagonistic epistasis in RNA secondary structure folding. BMC Evol Biol **3**:3.

4. **Poelwijk FJ, Kiviet DJ, Weinreich DM, Tans SJ**. 2007. Empirical fitness landscapes reveal accessible evolutionary paths. Nature **445**:383–386.

5. **da Silva J, Coetzer M, Nedellec R, Pastore C, Mosier DE**. 2010. Fitness epistasis and constraints on adaptation in a human immunodeficiency virus type 1 protein region. Genetics **185**:293–303.

6. **de Visser JAGM, Krug J**. 2014. Empirical fitness landscapes and the predictability of evolution. Nat Rev Genet **15**:480–490.

7. **Paulson JC, de Vries RP**. 2013. H5N1 receptor specificity as a factor in pandemic risk. Virus Res **178**:99–113.

8. **Kuiken T, Holmes EC, McCauley J, Rimmelzwaan GF, Williams CS, Grenfell BT**. 2006. Host species barriers to influenza virus infections. Science **312**:394–397.

9. **Webby R, Hoffmann E, Webster R**. 2004. Molecular constraints to interspecies transmission of viral pathogens. Nat Med **10**:S77–S81.

10. **Imai M, Watanabe T, Hatta M, Das SC, Ozawa M, Shinya K, Zhong G, Hanson A, Katsura H, Watanabe S, Li C, Kawakami E, Yamada S, Kiso M, Suzuki Y, Maher EA, Neumann G, Kawaoka Y**. 2012. Experimental adaptation of an influenza H5 HA confers respiratory droplet transmission to a reassortant H5 HA/H1N1 virus in ferrets. Nature **486**:420–428.

11. **Herfst S, Schrauwen EJA, Linster M, Chutinimitkul S, de Wit E, Munster VJ, Sorrell EM, Bestebroer TM, Burke DF, Smith DJ, Rimmelzwaan GF, Osterhaus ADME, Fouchier RAM**. 2012. Airborne transmission of influenza A/H5N1 virus between errets. Science **336**:1534–1541.

12. **Linster M, van Boheemen S, de Graaf M, Schrauwen EJA, Lexmond P, Mänz B, Bestebroer TM, Baumann J, van Riel D, Rimmelzwaan GF, Osterhaus ADME, Matrosovich M, Fouchier RAM, Herfst S**. 2014. Identification, characterization, and natural selection of mutations driving airborne transmission of A/H5N1 virus. Cell **157**:329–339.

13. **Kryazhimskiy S, Dushoff J, Bazykin GA, Plotkin JB**. 2011. Prevalence of epistasis in the evolution of influenza A surface proteins. PLoS Genet **7**:e1001301.

14. **Das SR, Hensley SE, David A, Schmidt L, Gibbs JS, Puigbò P, Ince WL, Bennink JR, Yewdell JW**. 2011. Fitness costs limit influenza A virus hemagglutinin glycosylation as an immune evasion strategy. Proc Natl Acad Sci U S A **108**:E1417–E1422.

15. **Myers JL, Wetzel KS, Linderman SL, Li Y, Sullivan CB, Hensley SE**. 2013. Compensatory hemagglutinin mutations alter antigenic properties of influenza viruses. J Virol **87**:11168–11172.

16. **Hensley SE, Das SR, Gibbs JS, Bailey AL, Schmidt LM, Bennink JR, Yewdell JW**. 2011. Influenza A virus hemagglutinin antibody escape promotes neuraminidase antigenic variation and drug resistance. PLoS One **6**:e15190.

17. **Gong LI, Suchard MA, Bloom JD**. 2013. Stability-mediated epistasis constrains the evolution of an influenza protein. Elife **2**:e00631.

18. **Duan S, Govorkova EA, Bahl J, Zaraket H, Baranovich T, Seiler P, Prevost K, Webster RG, Webby RJ**. 2014. Epistatic interactions between neuraminidase mutations facilitated the emergence of the oseltamivir-resistant H1N1 influenza viruses. Nat Commun **5**:5029.

19. **Bloom JD, Gong LI, Baltimore D**. 2010. Permissive secondary mutations enable the evolution of influenza oseltamivir resistance. Science **328**:1272–1275.

20. **Butler J, Hooper KA, Petrie S, Lee R, Maurer-Stroh S, Reh L, Guarnaccia T, Baas C, Xue L, Vitesnik S, Leang S-K, McVernon J, Kelso A, Barr IG, McCaw JM, Bloom JD, Hurt AC**. 2014. Estimating the fitness advantage conferred by permissive neuraminidase mutations in recent oseltamivir-resistant A(H1N1)pdm09 influenza viruses. PLoS Pathog **10**:e1004065.

21. **Simon P, Holder BP, Bouhy X, Abed Y, Beauchemin CAA, Boivin G**. 2011. The I222V neuraminidase mutation has a compensatory role in replication of an oseltamivir-resistant influenza virus A/H3N2 E119V mutant. J Clin Microbiol **49**:715–717.

22. **Maynard Smith J**. 1970. Natural selection and the concept of a protein space. Nature **225**:563–564.

23. **Orr HA**. 2005. The genetic theory of adaptation: a brief history. Nat Rev Genet **6**:119–127.

24. **Romero PA, Arnold FH**. 2009. Exploring protein fitness landscapes by directed evolution. Nat Rev Mol Cell Biol **10**:866–876.

802 25. **Orr HA**. 1998. The population genetics of adaptation: the distribution of
803 factors fixed during adaptive evolution. Evolution (N Y) **52**:935–949.

804 26. **Weinreich DM, Delaney NF, DePristo MA, Hartl DL**. 2006. Darwinian
805 evolution can follow only very few mutational paths to fitter proteins.
806 Science **312**:111–114.

807 27. **Covert AW, Lenski RE, Wilke CO, Ofria C**. 2013. Experiments on the role
808 of deleterious mutations as stepping stones in adaptive evolution. Proc Natl
809 Acad Sci U S A **110**:E3171–3178.

810 28. **Gerrish PJ, Lenski RE**. 1998. The fate of competing beneficial mutations
811 in an asexual population. Genetica **102-103**:127–144.

812 29. **Cowperthwaite MC, Bull JJ, Meyers LA**. 2006. From bad to good: fitness
813 reversals and the ascent of deleterious mutations. PLoS Comput Biol
814 **2**:1292–1300.

815 30. **Gillespie JH**. 1983. Some properties of finite populations experiencing
816 strong selection and weak mutation. Am Nat **121**:691–708.

817 31. **Dean AM, Thornton JW**. 2007. Mechanistic approaches to the study of
818 evolution: the functional synthesis. Nat Rev Genet **8**:675–688.

819 32. **Franke J, Klözer A, de Visser JAGM, Krug J**. 2011. Evolutionary
820 accessibility of mutational pathways. PLoS Comput Biol **7**:e1002134.

821 33. **Park M, Loverdo C, Schreiber SJ, Lloyd-Smith JO**. 2013. Multiple scales
822 of selection influence the evolutionary emergence of novel pathogens.
823 Philos Trans R Soc Lond B Biol Sci **368**:20120333.

824 34. **Weinreich DM, Watson RA, Chao L**. 2005. Perspective: Sign epistasis
825 and genetic constraint on evolutionary trajectories. Evolution (N Y)
826 **59**:1165–1174.

827 35. **Russell CA, Fonville JM, Brown AEX, Burke DF, Smith DL, James SL,**
828 **Herfst S, van Boheemen S, Linster M, Schrauwen EJ, Katzelnick L,**
829 **Mosterín A, Kuiken T, Maher E, Neumann G, Osterhaus ADME,**
830 **Kawaoka Y, Fouchier RAM, Smith DJ**. 2012. The potential for respiratory
831 droplet-transmissible A/H5N1 influenza virus to evolve in a mammalian
832 host. Science **336**:1541–1547.

833 36. **Fonville JM, Burke DF, Lewis NS, Katzelnick LC, Russell CA**. 2013.
834 Quantifying the fitness advantage of polymerase substitutions in influenza
835 A/H7N9 viruses during adaptation to humans. PLoS One **8**:e76047.

836 37. **Sanjuán R, Nebot MR, Chirico N, Mansky LM, Belshaw R**. 2010. Viral
837 mutation rates. J Virol **84**:9733–9748.

838 38. **Drake JW**. 1993. Rates of spontaneous mutation among RNA viruses.
839 Proc Natl Acad Sci U S A **90**:4171–4175.

840 39. **Gokhale CS, Iwasa Y, Nowak MA, Traulsen A**. 2009. The pace of
841 evolution across fitness valleys. J Theor Biol **259**:613–620.

842 40. **Gagneux S, Long CD, Small PM, Van T, Schoolnik GK, Bohannan**
843 **BJM**. 2006. The competitive cost of antibiotic resistance in Mycobacterium
844 tuberculosis. Science **312**:1944–1946.

845 41. **Coffin JM**. 1995. HIV population dynamics in vivo: implications for genetic
846 variation, pathogenesis, and therapy. Science **267**:483–489.

847 42. **Weissman DB, Desai MM, Fisher DS, Feldman MW**. 2009. The rate at
848 which asexual populations cross fitness valleys. Theor Popul Biol **75**:286–
849 300.

850 43. **Loverdo C, Lloyd-Smith JO**. 2013. Evolutionary invasion and escape in
851 the presence of deleterious mutations. PLoS One **8**:e68179.

852 44. **Alexander HK, Day T**. 2010. Risk factors for the evolutionary emergence
853 of pathogens. J R Soc Interface **7**:1455–1474.

854 45. **Weinreich DM, Chao L**. 2005. Rapid evolutionary escape by large
855 populations from local fitness peaks is likely in nature. Evolution (N Y)
856 **59**:1175–1182.

857 46. **Desai MM, Fisher DS**. 2007. Beneficial mutation-selection balance and the
858 effect of linkage on positive selection. Genetics **176**:1759–1798.

47. **Varble A, Albrecht RA, Backes S, Crumiller M, Bouvier NM, Sachs D, Garcia-Sastre A, TenOever BR**. 2014. Influenza A virus transmission bottlenecks are defined by infection route and recipient host. Cell Host Microbe **16**:691–700.

48. **Wilker PR, Dinis JM, Starrett G, Imai M, Hatta M, Nelson CW, O'Connor DH, Hughes AL, Neumann G, Kawaoka Y, Friedrich TC**. 2013. Selection on haemagglutinin imposes a bottleneck during mammalian transmission of reassortant H5N1 influenza viruses. Nat Commun **4**:2636.

49. **Wang GP, Sherrill-Mix S a, Chang K-M, Quince C, Bushman FD**. 2010. Hepatitis C virus transmission bottlenecks analyzed by deep sequencing. J Virol **84**:6218–6228.

50. **Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, Sun C, Grayson T, Wang S, Li H, Wei X, Jiang C, Kirchherr JL, Gao F, Anderson JA, Ping L-H, Swanstrom R, Tomaras GD, Blattner WA, Goepfert PA, Kilby JM, Saag MS, Delwart EL, Busch MP, Cohen MS, Montefiori DC, Haynes BF, Gaschen B, Athreya GS, Lee HY, Wood N, Seoighe C, Perelson AS, Bhattacharya T, Korber BT, Hahn BH, Shaw GM**. 2008. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc Natl Acad Sci U S A **105**:7552–7557.

880 51. **Merlo LMF, Pepper JW, Reid BJ, Maley CC**. 2006. Cancer as an
881 evolutionary and ecological process. Nat Rev Cancer **6**:924–935.

882 52. **Iwasa Y, Michor F, Nowak MA**. 2004. Stochastic tunnels in evolutionary
883 dynamics. Genetics **166**:1571–1579.

884 53. **Komarova NL, Sengupta A, Nowak MA**. 2003. Mutation-selection
885 networks of cancer initiation: tumor suppressor genes and chromosomal
886 instability. J Theor Biol **223**:433–450.

887 54. **Fan S, Macken CA, Li C, Ozawa M, Goto H, Iswahyudi NFN, Nidom CA,**
888 **Chen H, Neumann G, Kawaoka Y**. 2013. Synergistic effect of the PDZ
889 and p85β-binding domains of the NS1 protein on virulence of an avian
890 H5N1 influenza A virus. J Virol **87**:4861–4871.

891 55. **Fan S, Hatta M, Kim JH, Halfmann P, Imai M, Macken CA, Le MQ,**
892 **Nguyen T, Neumann G, Kawaoka Y**. 2014. Novel residues in avian
893 influenza virus PB2 protein affect virulence in mammalian hosts. Nat
894 Commun **5**:5021.

895 56. **Valcárcel J, Ortín J**. 1989. Phenotypic hiding: the carryover of mutations
896 in RNA viruses as shown by detection of mar mutants in influenza virus. J
897 Virol **63**:4107–4109.

898 57. **Wilke CO, Novella IS**. 2003. Phenotypic mixing and hiding may contribute
899 to memory in viral quasispecies. BMC Microbiol **3**:11.

58. **Lozovsky ER, Chookajorn T, Brown KM, Imwong M, Shaw PJ, Kamchonwongpaisan S, Neafsey DE, Weinreich DM, Hartl DL**. 2009. Stepwise acquisition of pyrimethamine resistance in the malaria parasite. Proc Natl Acad Sci U S A **106**:12025–12030.

59. **Mansky LM, Cunningham KS**. 2000. Virus mutators and antimutators: roles in evolution, pathogenesis and emergence. Trends Genet **16**:512–517.

60. **Suárez P, Valcárcel J, Ortín J**. 1992. Heterogeneity of the mutation rates of influenza A viruses: isolation of mutator mutants. J Virol **66**:2491–2494.

61. **Ribeiro RM, Bonhoeffer S, Nowak MA**. 1998. The frequency of resistant mutant virus before antiviral therapy. AIDS **12**:461–465.

62. **Loeb LA**. 1991. Mutator phenotype may be required for multistage carcinogenesis. Cancer Res **51**:3075–3079.

63. **Schrag SJ, Perrot V, Levin BR**. 1997. Adaptation to the fitness costs of antibiotic resistance in Escherichia coli. Proc R Soc London Ser B **264**:1287–1291.

64. **Perelson AS, Rong L, Hayden FG**. 2012. Combination antiviral therapy for influenza: predictions from modeling of human infections. J Infect Dis **205**:1642–1645.

65. **Müller B, Borrell S, Rose G, Gagneux S**. 2013. The heterogeneous evolution of multidrug-resistant Mycobacterium tuberculosis. Trends Genet **29**:160–169.

66. **Ribeiro RM, Bonhoeffer S**. 1999. A stochastic model for primary HIV infection: optimal timing of therapy. AIDS **13**:351–357.

67. **Pepin KM, Lass S, Pulliam JRC, Read AF, Lloyd-Smith JO**. 2010. Identifying genetic markers of adaptation for surveillance of viral host jumps. Nat Rev Microbiol **8**:802–813.

68. **Holmes EC**. 2013. What can we predict about viral evolution and emergence? Curr Opin Virol **3**:180–184.

FIGURES:

Figure 1: A) The blue line shows the deterministic probability of any virion being a double mutant ($\log_{10}$) as a function of the relative fitness of the intermediate single mutants (a relative fitness of 0 means the single mutants are fully deleterious). The probability is shown after t = 20 replication rounds, in the situation where the starting (wild type) virus and the adapted (double) mutant are neutral (relative fitness = 1). The pale shades of blue indicate the results of 10,000 stochastic simulations for each of the 101 settings of the relative fitness f. The circles indicate the average of the stochastic runs for each fitness setting. **B)** The probability of any virion being a double mutant is split into a mechanism where two mutations were acquired in a single replication round ("through doubles"); and a mechanism where the two single mutations occurred in distinct replication rounds ("through singles").

Figure 2: The effective number of routes to a double mutant is shown as a function of the relative fitness of the intermediate single mutants (a relative fitness f = 0 means the single mutants are fully deleterious). The results are shown after t = 20 replication rounds, in the situation where the starting (wild type) virus and the adapted (double) mutant are neutral (f = 1). The colors illustrate the relative contributions of a double mutation at once (grey, delay = 0), and the single-single mutation routes, with an increasing delay of 1 until 19

951 replication rounds between the mutation events represented by the gradient
952 shown in the colorbar.
953
954 **Figure 3:** The probability of any virion being a double mutant ($\log_{10}$) is shown in
955 color as a function of the relative fitness of the deleterious intermediate single
956 mutants (x-axis) and the relative fitness of the double, i.e. host-adapted, mutant
957 (y-axis). The probability is shown after t = 20 replication rounds, and the starting
958 (wild type) virus is neutral (f = 1).
959
960 **Figure 4: A)** The deterministic probability of any virion being a double mutant
961 ($\log_{10}$) is shown as a function of the relative fitness of the deleterious
962 intermediate single mutant(s). The graph shows the situation when both single
963 mutants are deleterious ("fitness valley", see the blue line in *Figure 1*), or when
964 only the non-ordered single mutant is deleterious ("fitness ridge"). The probability
965 is shown after t = 20 replication rounds, the starting (wild type) virus, ordered
966 single mutant (for the fitness ridge) and double mutant are neutral (f = 1). The
967 pale shades of blue and red indicate the results of 10,000 stochastic simulations
968 for each of the 101 settings of the relative fitness f for the fitness valley, and
969 fitness ridge, respectively. The circles indicate the average of the stochastic runs
970 for each fitness setting. **B)** The probability of double mutant for the fitness ridge
971 (red line) is divided into the contribution toward this probability by the mechanism
972 where two mutations were acquired in a single replication round ("through
973 doubles"); as two single mutations in distinct replication rounds in order ("through

974 ordered singles"); and as two single mutations in distinct replication rounds
975 occurring in the incorrect order, where the deleterious single mutation is obtained
976 first and incurs the fitness cost ("through deleterious singles").

977 # TABLES

978

	# mutations required					
	2	3	4	5	6	7
Effective # of routes	20	20	20	20	20	20
Total routes	400	8×10^{3}	1.6×10^{5}	3.2×10^{6}	6.4×10^{7}	1.28×10^{9}
Fraction available	5.0×10^{-2}	2.5×10^{-3}	1.3×10^{-4}	6.3×10^{-6}	3.1×10^{-7}	1.6×10^{-8}

979

980 **Table 1:** The effective number of routes if all intermediate mutants are fully
981 deleterious, the number of available routes if all intermediate mutants were viable
982 (total routes), and the fraction of routes available (Effective # of routes/Total
983 routes). The results are shown for two to seven mutations required, for t = 20
984 replication rounds. The starting (wild type) virus and the adapted (double) mutant
985 are neutral (f = 1), and all possible intermediates are fully deleterious (f = 0).

	# mutations required					
	2	3	4	5	6	7
Allowed routes	210	1540	8855	42504	177100	657800
Total routes	400	8×10^{3}	1.6×10^{5}	3.2×10^{6}	6.4×10^{7}	1.28×10^{9}
Fraction available	0.525	0.193	0.055	0.013	0.003	5.14×10^{-4}

Table 2: The number of allowed routes if complete and strict order is required for all mutations, the number of available routes if all intermediate mutants were viable (total routes), and the fraction of routes available, for situations where two to seven mutations are required. As an example, for 4 mutations required, this would mean that mutation A has to occur prior to or simultaneously with mutation B; mutation B has to occur prior to or simultaneously with mutation C; and mutation C has to occur prior to or simultaneously with mutation D. The results are shown for t = 20 replication rounds, the starting (wild type) virus and each ordered mutant are neutral (f = 1), and all possible non-ordered intermediate mutants are fully deleterious (f = 0).