RESEARCH ARTICLE

# Imprecise recombinant viruses evolve via a fitness-driven, iterative process of polymerase template-switching events

Kirsten Bentley[1¤]*, Fadi Ghassan Alnaji[2◉], Luke Woodford[1◉], Siân Jones[1], Andrew Woodman[3], David J. Evans[1]*

1 BSRC and School of Biology, University of St Andrews, St Andrews, United Kingdom, 2 Department of Microbiology, University of Illinois, Champaign-Urbana, Illinois, United States of America, 3 Department of Biochemistry and Molecular Biology, Pennsylvania State University, Pennsylvania, United States of America

◉ These authors contributed equally to this work.
¤ Current address: School of Medicine, Cardiff University, Cardiff, United Kingdom
* bentleyk@cardiff.ac.uk (KB); d.j.evans@st-andrews.ac.uk (DJE)

## Abstract

Recombination is a common feature of many positive-strand RNA viruses, playing an important role in virus evolution. However, to date, there is limited understanding of the mechanisms behind the process. Utilising *in vitro* assays, we have previously shown that the template-switching event of recombination is a random and ubiquitous process that often leads to recombinant viruses with imprecise genomes containing sequence duplications. Subsequently, a process termed resolution, that has yet to be mechanistically studied, removes these duplicated sequences resulting in a virus population of wild type length genomes. Using defined imprecise recombinant viruses together with Oxford Nanopore and Illumina high throughput next generation sequencing technologies we have investigated the process of resolution. We show that genome resolution involves subsequent rounds of template-switching recombination with viral fitness resulting in the survival of a small subset of recombinant genomes. This alters our previously held understanding that recombination and resolution are independent steps of the process, and instead demonstrates that viruses undergo frequent and continuous recombination events over a prolonged period until the fittest viruses, predominantly those with wild type length genomes, dominate the population.

## Author summary

Viruses with positive-sense RNA genomes, such as poliovirus, have several mechanisms by which they evolve. One of these is the process of recombination involving the large-scale exchange of genetic information. Recombination occurs during replication when the viral polymerase, bound to the nascent RNA chain, switches from copying one genome to another. However, the polymerase does not always accurately switch between the two, resulting in sequence duplications or deletions, and genomes that are referred to as imprecise. Over multiple rounds of replication sequence duplications are lost and genomes are resolved to wild type length, but it is unclear how this occurs. Here we used synthetic

polioviruses containing defined sequence duplications to determine that the genome population undergoes repeated rounds of recombination until sequence duplications are lost and viruses with precise, wild type length genomes are selected for. This selection is based on the overall fitness of the virus population, with less fit imprecise viruses evolving more quickly. Our study suggests that recombination is a continual process where virus fitness drives the selection of a small subset of recombinant variants. These data are important for understanding how novel viruses evolve via recombination and how this process can be blocked to prevent novel and dangerous pathogens from arising.

## Introduction

RNA viruses usually exist as quasispecies due to error-prone RNA-dependent RNA polymerases (RdRps) which misincorporate nucleotides during genome replication. In many RNA viruses, the process of recombination generates additional diversity within the virus population. Recombination allows for a larger exchange of genetic material between closely related viruses and is proposed to be a mechanism by which genomes can be rescued from error catastrophe, which would otherwise have resulted from the accumulation of deleterious mutations introduced by the RdRp [1–3]. Recombination is widely accepted to occur via a copy-choice mechanism involving the RdRp switching from one template (the donor) to another (the acceptor) during negative-strand synthesis [4]. Sequence identity and RNA structure have been proposed to influence the template-switching event, although evidence is often contradictory and may reflect analysis of highly selected populations (reviewed in [5]).

Poliovirus, the prototype species C enterovirus of the *Picornaviridae* family, has proven to be a useful tool for the study of recombination. It has long been known that the three serotypes frequently recombine [6,7] and there are well-established reverse genetic techniques allowing markers for recombination to be engineered into the genome [8,9]. Poliovirus, like all enteroviruses, has a positive-strand RNA genome of approximately 7.5 Kb. The genome encodes a single polyprotein flanked by 5′ and 3′ untranslated regions (UTR) containing the signals for translation and replication. The polyprotein is co- and post-translationally cleaved to yield the structural proteins (P1: VP4, VP3, VP2, VP1) and the non-structural proteins (P2: 2A$^{pro}$, 2B, 2C and P3: 3A, 3B$^{VPg}$, 3C$^{pro}$ and 3D$^{pol}$).

We have developed an *in vitro* recombination assay–the poliovirus CRE-REP assay–that solely generates recombinant progeny viruses [10]. The CRE-REP assay requires co-transfection of two RNA molecules that independently cannot generate viable virus progeny; (1) a poliovirus serotype 1 (PV1)-based replicon in which the structural proteins have been replaced with a luciferase marker gene, and (2) a full-length poliovirus serotype 3 (PV3) genome in which the 2C *cis*-acting replication element (CRE; [11]) has been mutated so that the virus can no longer synthesise new positive-strand genomes [12]. When co-transfected into permissive cells the two non-infectious genomes produce negative strands, which can recombine to yield a population of viable recombinant viruses. Using this assay, we postulated that recombination is a biphasic process [10]. First, an RdRp strand-transfer event produces genomes that may contain sequence duplications of up to several hundred nucleotides. The imprecise and apparently random nature of this recombination event argues strongly against sequence identity or structure being a major influence [13]. Since *in vivo* recombination results in genome-length products we proposed that strand transfer is followed by a second event we termed resolution, in which sequence duplications are lost via an iterative process. This could be demonstrated experimentally *in vitro* [10]. Resolution results in a virus population containing a variety of

precise junctions between donor and acceptor templates [10]. Mechanistically the process of resolution is poorly defined, but two obvious scenarios can be envisaged: an *in-cis* internal deletion event involving individual genomes or an *in-trans* process requiring two or more genomes (Fig 1).

In recent studies involving the next generation sequencing of whole cell RNA preparations from co-infected cells, and parallel CRE-REP assays using highly modified templates, we provide further evidence that recombination *per se* (the strand transfer event) is not influenced by either the RNA donor or acceptor sequence or its structure [13]. At face value this contradicts numerous studies of recombinant virus genomes which have implicated local RNA structure, sequence identity or particular nucleotides at the recombination junction [14–20]. However, progeny recombinant virus genomes in these studies are likely to have undergone one or many rounds of resolution and, in contrast to the initial recombinant viral RNA population in the cell cytoplasm, are–by definition–functionally intact and packaged into infectious virus.

It therefore remained a possibility that recombination and resolution are mechanistically dissimilar events, in which case sequence identity, composition or structure might influence one and not the other. To address this we used a well-characterised imprecise recombinant genome, #105B a type 1/3 poliovirus recombinant containing a 249 nt genome duplication [10], to investigate resolution independently of the initial template-switching event of recombination. We demonstrate that resolution of the virus population is driven by viral fitness and is enhanced in the presence of ribavirin, a mutagen that decreases RdRp fidelity and has previously been demonstrated to enhance recombination [10]. In addition, we characterised the resolving/resolved virus population following co-infection of uniquely barcoded imprecise recombinant viruses. We determined that resolution is, at least in part, a process requiring further template-switching/recombination events. The biphasic recombination process therefore appears to involve iterative strand-transfer events in which the primary determinant of selection is the fitness of the resulting virus genome.

## Results

### Rate of resolution is influenced by virus fitness

To better understand the resolution process, we investigated whether resolution was influenced by the sequence background of the imprecise recombinants studied to date. We have previously identified and subsequently constructed a cDNA for a viable intertypic (PV3/PV1) imprecise recombinant designated, #105B (GenBank MZ165620). The #105B cDNA has a 249 nt sequence insertion at the VP1/2A cleavage boundary consisting of duplicated VP1 and 2A sequences, as well as a fragment of firefly luciferase sequence derived from the sub-genomic replicon used in the CRE-REP assay in which #105B was generated (Fig 2A, upper panel) [10].

We reasoned that any *in cis* resolution events might be influenced by sequence identity even if the original strand transfer event had not been. We therefore generated an intratypic (PV3/PV3) version of #105B for comparison, designated #PV3-105B, containing an identical sequence duplication to that of #105B but in a sequence background of PV3 only (Fig 2A, lower panel; GenBank MZ165621). One-step growth curve analysis of #105B and #PV3-105B showed a small but significantly lower ($P<0.05$, t-test) average maximum titre of $4.9\times10^7$ pfu/ml for #105B when compared to $2.2\times10^8$ and $6\times10^8$ pfu/ml for parental viruses PV3 and PV1 respectively (Fig 2B). In contrast, we observed no statistically significant loss in replication fitness of #PV3-105B, with a maximum virus titre of $3.3\times10^8$ pfu/ml. While maximum virus titres are only one determinant of virus fitness in tissue culture–and may not be truly representative
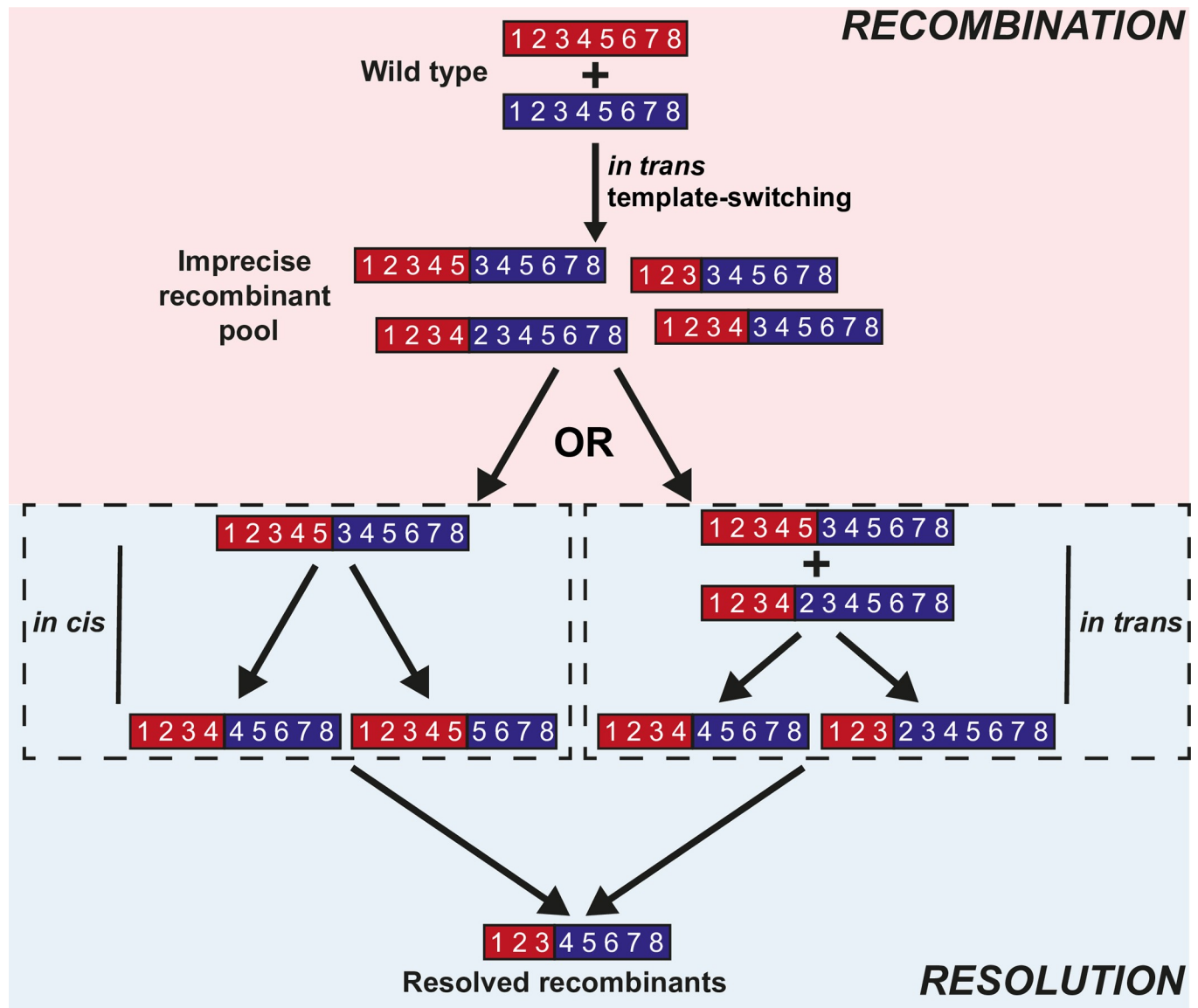
**Fig 1. Model of recombination and resolution.** Recombination initially occurs as an *in trans* template-switching event between two parental viruses (represented by numerical genome schematics). This event generates a mixed population of viruses containing a variety of imprecise recombinants with sequence duplications, as represented by the duplicated numbers of the schematic genomes. The pool of imprecise recombinants subsequently undergoes resolution where these duplications may be lost via an, as yet undetermined, *in cis* deletion process or via further rounds of *in trans* template-switching recombination events.

of virus replication in the gut–this suggests that sequence duplications alone may not have an observable effect on virus fitness. This is contrary to our original finding that imprecise recombinants with duplications have an associated fitness cost.

We next tested whether these differences in the magnitude of replication influenced the rate of resolution of the two imprecise recombinant viruses. Duplicate samples of #105B and #PV3-105B were serially passaged 15 times (p1-15) on HeLa cells and the region spanning the imprecise recombination site was amplified by RT-PCR from cell culture supernatant at each passage (Fig 2C). Resolution of imprecise recombinants is reflected in the increased gel mobility of the PCR fragment spanning the recombination site, as deletions accumulate in the virus
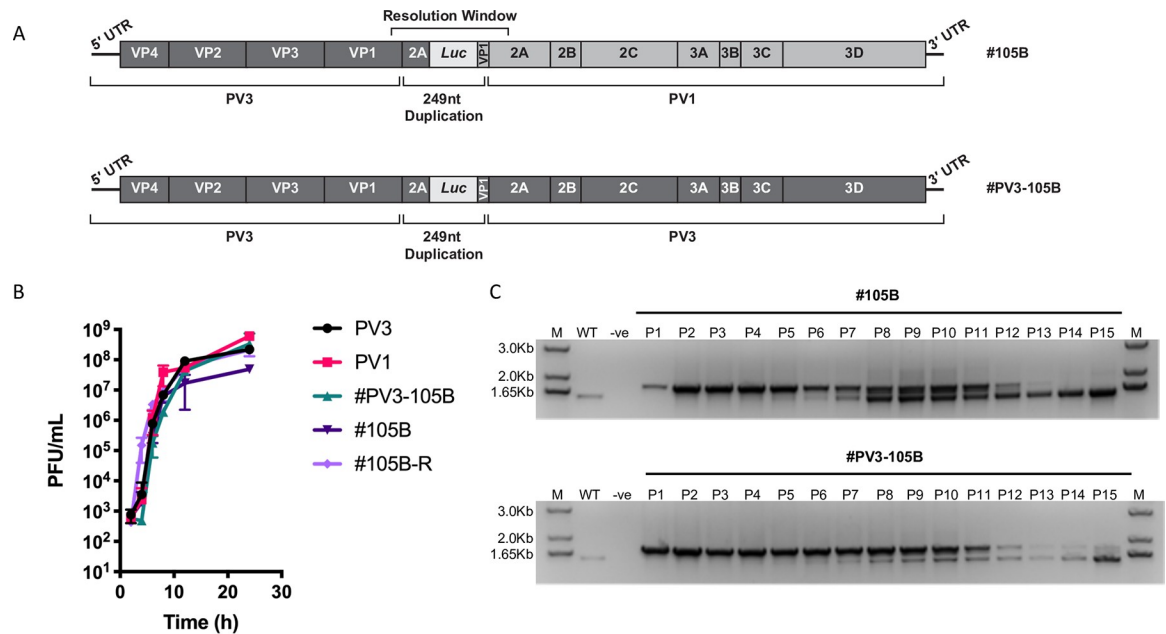
**Fig 2. Phenotypic characterisation of intra- and intertypic imprecise recombinants.** (A) Schematic representation of the genomes of imprecise recombinants #105B and #PV3-105B. Regions in dark grey are derived from PV3 and regions in light grey are derived from PV1. (B) Replication kinetics of imprecise clones. HeLa cells were infected with virus stock at multiplicity of infection of 5 and virus-containing supernatants harvested over 24 h. Virus titres were determined by plaque assay on HeLa cells. Error bars represent the standard deviation of 3 experiments. (C) Resolution of imprecise recombinant clones. Both #105B (upper panel) and #PV3-105B (lower panel) were serially passaged on HeLa cells for 15 passages. Viral RNA was extracted from supernatants and the potential region of recombination amplified by RT-PCR. The WT lane represents the 1470 nt product generated from a correct-genome length sample.

population. By p13, #105B had resolved to yield a single visible product representative of viruses with genomes of native 'wild type' length. In contrast, with #PV3-105B, which when resolved is indistinguishable from PV3, a minor product representing a greater-than-genome-length virus population was observable alongside resolving products up to, and including, p15. In both assays DNA products of intermediate size between unresolved and fully resolved products were visible. These were most prominent between p8 and p11 for #105B, and p9 and p11 for #PV3-105B (Fig 2C) and their gradual disappearance was suggestive of either an iterative resolution process or that they were 'dead-end' products subsequently out-competed in the virus population.

Analysis of the growth kinetics for the resolved #105B virus population (#105B-R) at p15 showed restoration of parental levels of virus replication, with a peak virus titre of $2.1 \times 10^8$ pfu/ml (Fig 2B). This demonstrated that at the population level, resolution is in part driven by virus replication fitness, with fitter imprecise viruses, i.e., #PV3-105B, under less selective pressure. We would therefore expect fitter imprecise recombinants to persist in the population for longer than those with a replication defect.

## Resolution of a clonal recombinant virus generates a diverse virus population

We have previously reported apparently intermediate products in the resolution process [10] similar to the transient PCR products that appeared between p8 and p11 during the resolution of #105B and #PV3-105B (Fig 2C). To characterise these intermediate products in greater detail and so gain a better understanding of the overall virus population during resolution, we

utilised Oxford Nanopore Technologies (ONT) high throughput sequence analysis of both #105B and #PV3-105B.

We selected individual or pooled passaged samples that would best represent the unresolved (p1) and intermediate resolved (#105B, p8-11; #PV3-105B, p10-13) populations. We also analysed #105B and #PV3-105B at p15. Furthermore, to investigate how a population that appears to be resolved may continue to evolve over time we additionally passaged #105B to p18 and included this population in the analysis.

An amplicon library was prepared according to Oxford Nanopore protocols for 1D analysis. The library consisted of uniquely barcoded PCR products, generated by amplifying a ~1.74 Kb region spanning the resolution window (see Materials and Methods), and representing the unresolved, intermediate, and resolved virus populations (S1 Table). The library was prepared such that those populations predicted to contain the most diversity in the range of recombinant sequences generated, had greater representation within the library.

Two independent libraries were generated, one for each set of passaged samples, and sequenced separately (reads available at SRA, BioProject PRJNA731157). For the analysis we opted to focus only on genomes that could yield viable infectious progeny virus, i.e., those in which any deletion still maintained the open reading frame. Any one deletion, however, could occur anywhere within the duplicated region of the resolution window spanning nts 3361–3720 (Fig 2A). Theoretically there were 14,209 possible deletions that fulfilled these criteria. These were created *in silico* and used as reference templates for alignment of the ONT reads (S1 Fig; and see Materials and Methods). After optimising the alignment of individual reads to the template population, individual consensus sequences were generated representing the resolving or resolved genomes within the virus population (S1 Fig).

As expected, the p1 populations for both #105B and #PV3-105B yielded single consensus sequences matching the input virus, supporting the validity of the analysis pipeline. In all other populations recombinant variants were designated as either fully resolved (i.e., the overall genome length was identical to that of the parental viruses due to the acquisition of a 249 nt deletion) or partially resolved (i.e., intermediate products in which some of the original 249 nt duplication was retained). Fully resolved, precise junctions for #105B were designated after the nucleotide position at which the recombinant genome sequence switched from matching the 5′ PV3 sequence to the 3′ PV1 sequence, with recombinants designated as the last nucleotide matching PV3 (e.g., PV3$^{3378}$). However, due to sequence identity (81%) between PV3 and PV1 in this region there was often ambiguity as to the exact crossover point, as previously described for such precise junctions [10]. Resolution to a precise junction required the deletion of 249 nt, which, to remain viable, must include the 137 nt of luciferase sequence derived from the PV1 donor. The remaining 112 nt, encompassing duplicated regions of VP1 and 2A, could be deleted from either the PV1 donor-derived or PV3 acceptor-derived sequence (resolution window; Fig 2A). Consequently, a precise junction could theoretically occur at any one of these 112 nt positions. However, due to the level of sequence identity between PV1 and PV3 in the VP1/2A duplicated region only 23 precise junctions could be unambiguously defined. As fully resolved #PV3-105B is indistinguishable from parental PV3, precise junctions could not be defined. In comparison, recombinants of both #105B and #PV3-105B maintaining even a small sequence duplication could inevitably be determined more accurately. These partially resolved variants were designated by the number of nucleotides deleted from the initial 249 nt duplication (e.g., #105BΔ162).

Despite the high numbers of theoretical recombinants, analysis of the #105B populations detected only seven partially resolved (Fig 3A) and eight fully resolved (Fig 3C) unique junctions (S2 Fig and S2 Table). Despite ~32% of the total ONT reads being generated to the partially resolved #PV3-105B samples, only 3 unique junctions (#PV3-105BΔ231, -Δ225 and
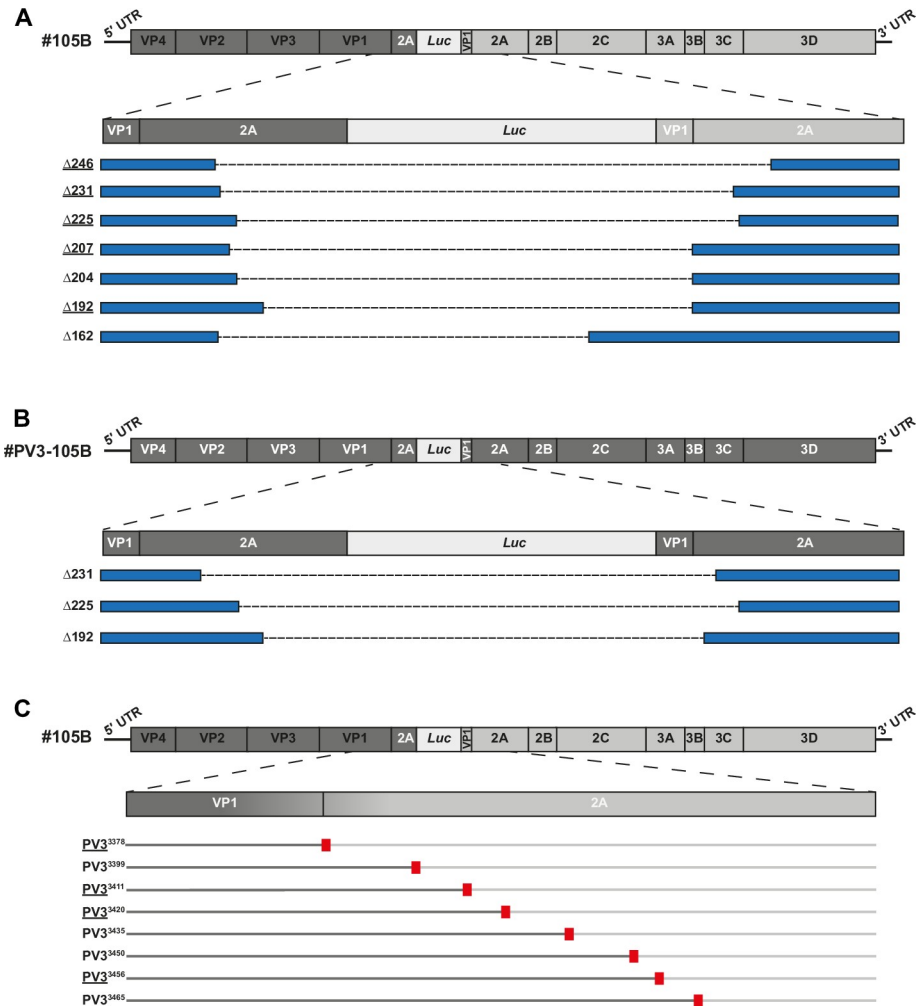
**Fig 3. Nanopore analysis of #105B and #PV3-105B recombinant populations.** The sequences of viable recombinant genomes were determined by consensus generation from Oxford Nanopore reads from all #105B and #PV3-105B passaging populations sequenced. Partially resolved genomes are shown for (A) #105B and (B) #PV3-105B. The original genome is shown at the top, where regions in dark grey are derived from PV3 and regions in light grey are derived from PV1, and an expanded view of the duplicated region shown below. Individual recombinants are denoted blue boxes with dashed lines representing the region of the duplication that has been lost. Numbers on the left-hand side represent the number of nt deleted from the 249 nt required to fully resolve the genome. Underlined recombinant names indicates recombinants detected in both replicate samples. (C) Resolved variants of #105B. Top schematic represents the genome of #105B with an expanded view of the VP1/2A region expected for a resolved recombinant. Resolved variants are represented by dual-coloured lines with sequence from PV3 in dark grey and sequence from PV1 in light grey. Red boxes denote the location of the resolved junction. Recombinant names, denoting the junction position, are given on the left. Underlined recombinant names indicates recombinants detected in both replicate samples.

-Δ192) were present with sufficient frequency to generate representative consensus sequences (Fig 3B). Notably, a few junctions were conserved in both #105B- and #PV3-105B-derived resolving samples; #105BΔ225 and #105BΔ192 have identical deletion sizes and junction positions to #PV3-105BΔ225 and #PV3-105BΔ192. In addition, #105BΔ231 and #PV3-105BΔ231 contain an identically sized deletion, although the junctions are in different genome locations (S2 Fig).

Considerable overlap was observed in both the resolved and partially resolved genomes detected between the independently generated and analysed samples derived from #105B
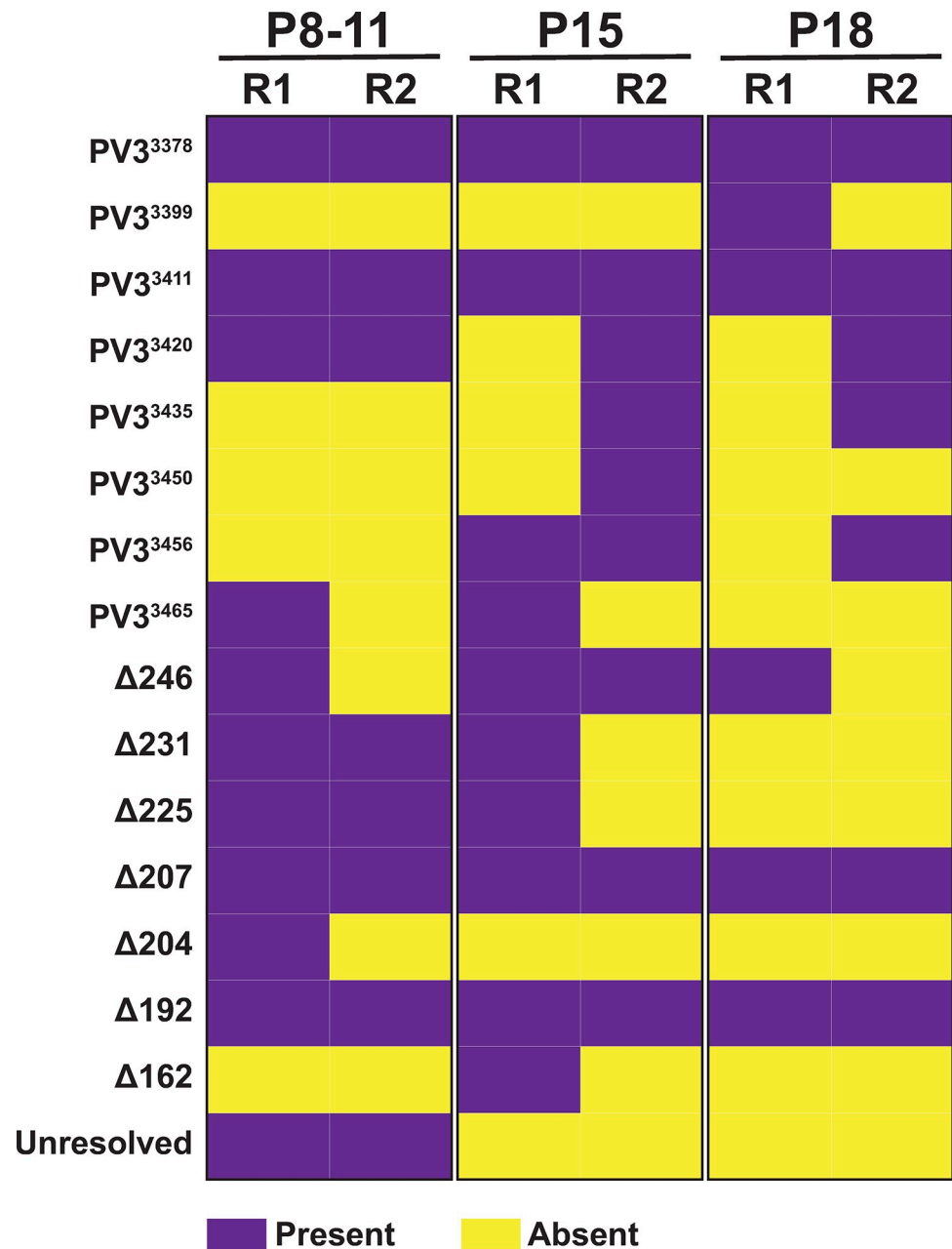
**Fig 4. Comparison of recombinants identified between #105B nanopore replicates.** Each recombinant variant identified from Oxford Nanopore analysis of the three #105B populations, p8-11, p15 and p18, are denoted as either present (purple) or absent (yellow) from passaging replicates (R) 1 and 2.

(Fig 4). Four of the fully resolved sequences (PV3³³⁷⁸, PV3³⁴¹¹, PV3³⁴²⁰ and PV3³⁴⁵⁶) and five of the partially resolved sequences (#105BΔ246, -Δ231, -Δ225, -Δ207 and -Δ192) were identified in both independent replicates. Despite this overlap, the frequencies at which particular variants were observed was not necessarily constant between replicates. We used consensus generation–which is based on ONT read number–as a proxy measure for variant frequency and plotted the relative percentages of each variant during the latter stages of passaging (S3 Fig). Based on the passaging methodology–where only 25% of supernatant was used to infect

subsequent passages–we may expect to observe a bottleneck in the population based not only on variant fitness, but also due to low-level variants not being transferred to the next passage. For example, resolved variant PV3$^{3456}$ represented 3% and 19% of the population at p15 for replicate 1 and 2 respectively. However, by p18 this variant was no longer detected in replicate 1, but had increased to 21% of the population for replicate 2. This suggests that in replicate 1 this virus was either not transferred upon passage, due to its initial low prevalence, or was potentially rapidly outcompeted by other variants acquired simultaneously.

Unexpectedly, we identified the presence of several partially resolved variants at p15. Products representative of these variants were not visible during gel analysis (Fig 2C, upper panel); however, this is most likely a consequence of the limited sensitivity of the gel analysis technique. A subset of these partially resolved variants also continued to be present at p18, indicating that even seemingly resolved populations may be continually evolving alongside recombinant variants with non-wild type length genomes. It is important to note that repeated passaging may also lead to the incorporation of point mutations through the natural error rates of the viral RdRp. It has previously been demonstrated that the normal mutation rate for poliovirus is 2.1 per 10,000 nt [21] and such mutations elsewhere in the genome may be beneficial for virus fitness, thus influencing the growth of particular variants and therefore the overall resolving population. The presence of such mutations may help to explain why some partially resolved variants remain at P18 and is similar to the generation of quasispecies observed for HCV (reviewed in [22]).

While strict selection criteria were applied to the data to minimize errors in recombinant junction or genome prediction caused by the high ONT error rate [23], these results suggest only a limited number of the theoretical viable variants were either generated or replicated to detectable levels during sequential virus passage. The similarities between partially resolved variants of #105B and #PV3-105B also suggested that specific recombinants were favoured regardless of the sequence of the backbone virus. As this was surprising, we used traditional Sanger sequencing to provide an independent analysis of the resolving population to determine that the ONT consensus sequences generated were accurate representations of the genomes extant in the virus population. Following PCR cloning, individual virus sequences were isolated from p9 and p18 of #105B, and p10 for #PV3-105B. Of the 20 clones analysed we were able to identify resolved variants PV3$^{3365}$ and PV3$^{3399}$, and partially resolved variants -Δ246, -Δ231, -Δ225 and -Δ192 from #105B, and partially resolved variant -Δ192 from #PV3-105B. Since these junctions were also all identified in the ONT analysis we concluded that consensus generation in the latter was representative of the underlying population of resolving recombinant molecules.

## Analysis of resolution by Illumina sequencing

The analysis of the resolving virus population by ONT showed it has significant promise in the study of recombination. For example, the long-read lengths enabled unresolved, partially resolved and fully resolved genomes to be detected in a single population. However, the high error rate inherent in ONT sequencing necessitated the use of strict criteria during pre-screening of primary data for consensus generation, including a threshold number of reads needed to form the consensus (see Materials & Methods). Inevitably this reduced sensitivity which we suspected resulted in missing less frequent resolving intermediates. In a recent study of recombination we had used paired-end Illumina sequencing to detect recombination junctions within total cell RNA. In this we were able to detect a much greater range of crossover junctions in the RNA population than were present in the resulting recombinant infectious virus population [13]. We therefore applied this method to investigate the process of resolution and to compare with results from the ONT analysis.
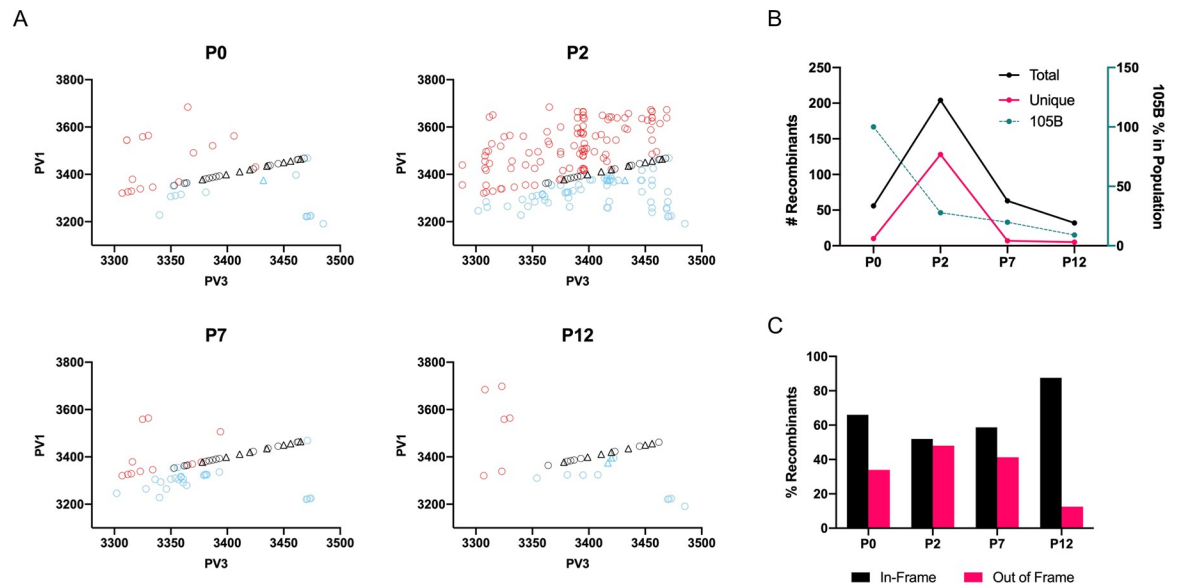
**Fig 5. Visualisation of #105B Illumina recombinant type.** (A) Recombinants detected via Illumina sequencing were plotted by type for p0, p2, p7, and p12; precise (open black circles), imprecise-insertion (open blue circles) or imprecise-deletion (open red circles). All recombinants detected with Oxford Nanopore as well as Illumina sequencing, irrespective of passage number, are indicated as precise (open black triangles) or imprecise-insertion (open blue triangles) recombinants. (B) Recombinant counts. The total number of independent junctions at each passage (left Y axis) was plotted against the passage number (x-axis) highlighting the level of diversity within the population at each passage. The proportion of #105B at each passage was plotted as a percentage of the total reads for #105B at p0 and after normalisation of all reads to p0 (right Y axis). (C) Comparison of the percentage of in-frame and out-of-frame recombinants present at each passage.

The imprecise recombinant #105B was passaged on HeLa cells until complete cytopathic effect (CPE) was observed for each cell monolayer, whereupon total RNA samples were harvested for subsequent RT-PCR and agarose gel analysis of the duplicated region as above. Under the conditions used in this assay, resolving genomes were apparent by p4 and resolution appeared near-complete by p10. After two further passages the virus populations from passages p2, p7 and p12 were selected for further analysis. In addition, a p0 sample–obtained following transfection of RNA–was analysed to determine the population diversity detectable prior to passaging. In contrast to the ONT analysis, which involved alignment to in-frame templates generated *in silico*, no constraints were placed on the resolving sequence junctions in the Illumina dataset which were identified using a bioinformatics pipeline based on ViReMa [13,24,25] (S4 Fig; reads available at SRA, BioProject PRJNA731157).

The unpassaged p0 population, harvested 24 h post-transfection was not clonal, but instead contained a range of resolved precise deletions, together with additional genomes bearing imprecise insertions or deletions (Fig 5A). By p2, the total amount of variation within the sequenced population had markedly increased, with subsequent decreases at p7, and again at p12 (Fig 5A and 5B). Changes in population diversity followed a similar pattern when measured both as the number of previously undetected unique junctions identified at each passage (Fig 5B), and as a measure of Shannon Entropy (S5 Fig). The level of #105B within the population was also reduced following each passage, as expected for the process of resolution (Fig 5B). A significant proportion of the junctions identified by VirReMa within the Illumina dataset were out-of-frame (~30–50% in p0, p2 and p7), with the proportion dropping to ~10% by p12 (Fig 5C).

It was clear from the p0 data that, if resolution is a stepwise process, it can occur extremely rapidly as all of the 23 fully resolved variant genomes that can be unambiguously identified (as

discussed above) could be detected within the 24 h period between RNA transfection and RNA extraction of the p0 sample. During subsequent passage the number of resolved variants detected fell from 23 at p0 to 15 at p12, presumably as the fittest variants came to dominate the population.

Overall, the Illumina sequence analysis of the RNA population during resolution revealed a comparable abundance, variety, and distribution of recombinant genomes as we identified during our study on recombination [13], indicating the same, or similar, mechanisms may underpin both processes.

## Ribavirin increases rate of resolution

RNA polymerase fidelity is a major determinant of recombination [10,26], with error-prone RdRp variants generating recombinants at an increased frequency [27,28]. One way to increase the error-rate of the native polymerase is to introduce sub-lethal concentrations of ribavirin, a well characterised guanosine analogue which induces base mismatches when incorporated into the viral genome and, at high levels, results in lethal mutagenesis by error catastrophe [21]. We have previously demonstrated that ribavirin increases the yield of recombinants in the CRE-REP assay by ~300% [10]. If the process of resolution also involves additional rounds of recombination or is influenced by the error-rate of the RdRp, we predicted that the presence of ribavirin would also increase the rate of resolution.

We therefore serially passaged #105B, recovered following transfection with *in vitro* produced RNA, in the presence of a range of sub-lethal concentrations of ribavirin and monitored the resolving population by RT-PCR. There was a clear dose-response with increasing levels of ribavirin resulting in the more rapid appearance of a fully resolved product with increased mobility when analysed by agarose gel electrophoresis (Fig 6). In addition, the presence of the large unresolved parental #105B product–still apparent at passage 10 in the absence of
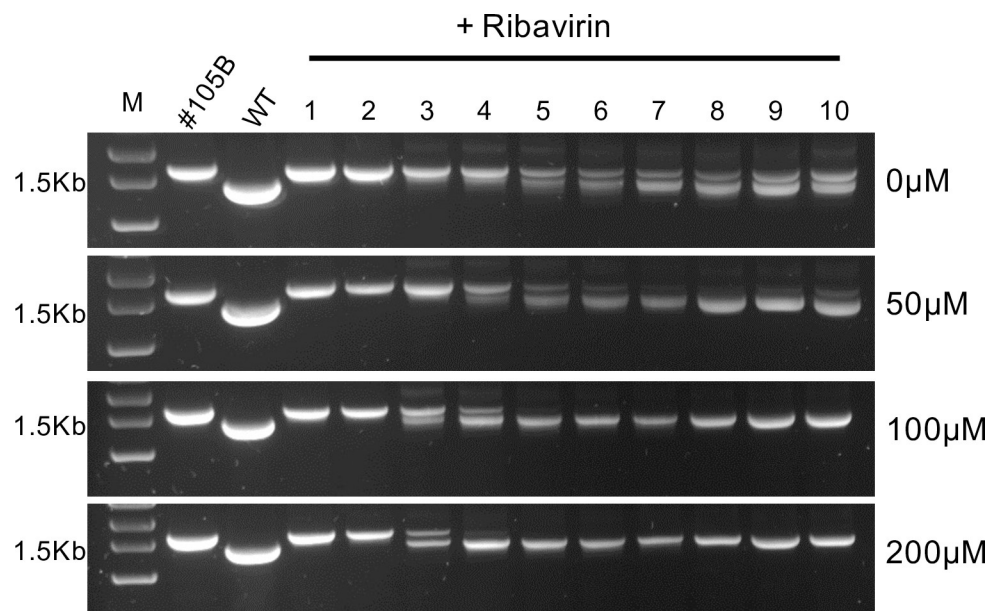


**Fig 6. The rate of resolution of #105B increases in the presence of Ribavirin.** #105B was serially passaged ten times on HeLa cells in the presence of 0, 50, 100 or 200 μM of ribavirin. Viral RNA was extracted, amplified by PCR and analysed by gel electrophoresis. WT = PV3, representing genome-length product. M = GeneRuler 1kb plus DNA markers (FisherScientific).

ribavirin–was also influenced by increasing concentrations of ribavirin, being lost by p5 or p4 with 100 μM or 200 μM ribavirin respectively. These results indicate that the process of resolution is influenced by the error rate of the viral polymerase and is at least suggestive that further rounds of *in trans* recombination might be involved in resolution. To investigate the latter further we explored the generation of recombinant genomes during the resolution of uniquely barcoded variants of #105B.

## Recombination is detected during resolution of barcoded viruses

The structure of the #105B genome means that all partially or fully resolved derivatives must effectively 'lose' sequences from within the resolution window. We therefore introduced modifications flanking this region that would allow the detection of resolving genomes that could only have arisen following a recombination event (Figs 2A and 7A). Taking advantage of the degeneracy of the triplet amino acid code we synthesised a small panel of synonymous variants encompassing five-codons on each side of the resolution window. Individual viruses were recovered from engineered cDNAs using standard techniques and their growth kinetics compared in parallel to #105B. Two uniquely barcoded cDNAs designated #105B$^{T1}$ and #105B$^{T2}$ (Fig 7A), encoding viruses with replication kinetics indistinguishable from the parental #105B (Fig 7B) were chosen for further study.

HeLa cells were co-infected with #105B and either #105B$^{T1}$ or #105B$^{T2}$ at a 1:1 ratio of each virus and a final MOI of 10. Virus was harvested and serially passaged, with each passage titred to maintain the MOI of 10 in order to maximise the likelihood that each cell would be infected with at least one copy of each #105B variant. Resolution was monitored by RT-PCR amplification of the resolution window (and flanking barcodes), and 48 individually cloned PCR
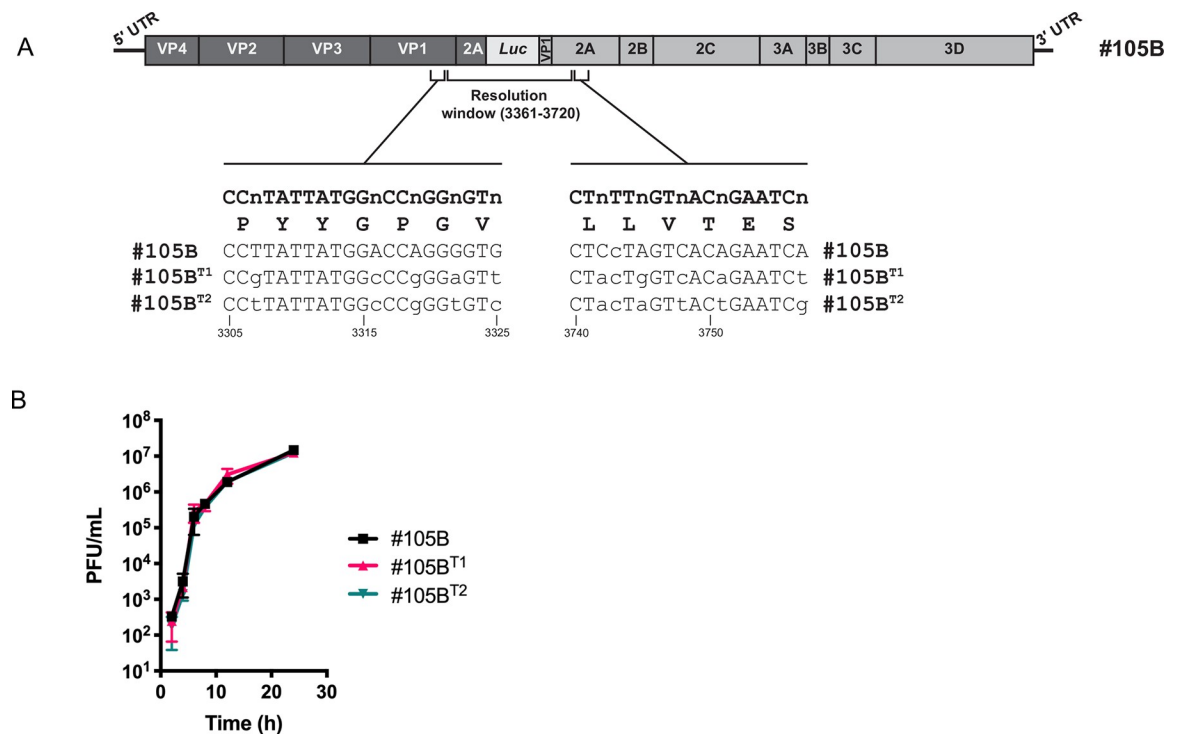


**Fig 7. Characterisation of barcoded #105B.** (A) Schematic representation of the sequences modified in #105B for the addition individual barcodes. (B) Replication kinetics of #105B and barcoded viruses.

products from resolving, or resolved, populations were sequenced for each of the two co-infections (n = 96).

Of the 96 sequences 89 consisted of a single clonal population, with the remaining 7 being mixed and excluded from further analysis. Of the 89 clonal sequences, 76 (85%) were unresolved (n = 5), partially resolved (n = 53) or completely resolved (n = 18) with flanking barcode regions matching only one or other of the parental viruses. These neither supported nor refuted an involvement for recombination in the resolution process as the resolved genomes could have been generated by recombination between identical parental viruses or via an *in cis* deletion process. However, the remaining 13 sequences (15%), all of which were partially (n = 11) or completely (n = 2) resolved, contained mismatched sequences in the barcoded regions indicative of a recombination event. Four of these 13 sequences were biological duplicates, leaving a total of 9 unique resolving genomes that must have arisen from a recombination event (Fig 8 and S3 Table). Interestingly, 2 of the junctions identified–partially resolved -Δ207 and fully resolved PV3$^{3420}$ –were identical to junctions previously observed when passaging #105B (compare Fig 3A and 3C to Fig 8), with -Δ207 also being detected from both barcoded co-infections. This observation was in agreement with the commonality of junctions observed between replicates of the #105B ONT analysis and again, implies that a limited
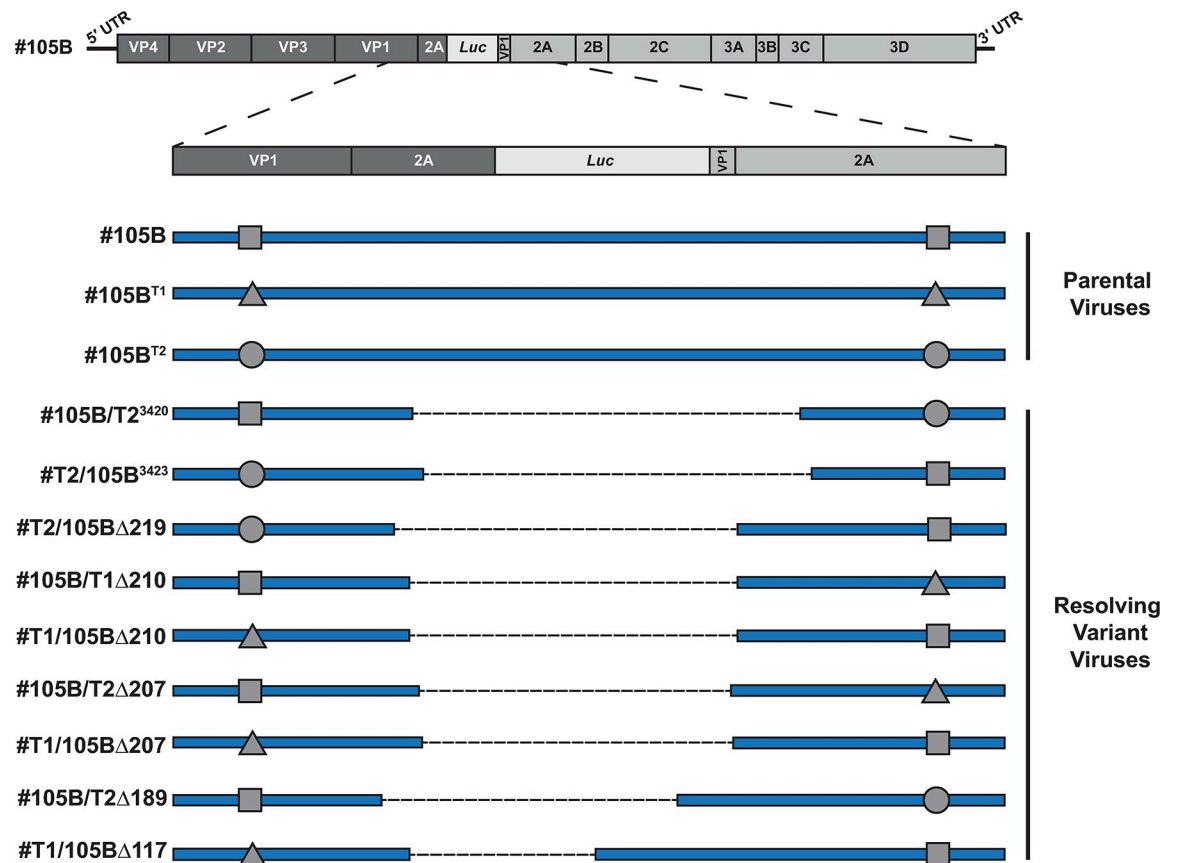


**Fig 8. Analysis of barcoded #105B recombinants.** Schematic representation of unique recombinant variants identified following co-infection of #105B with barcoded virus #105B$^{T1}$ or #105B$^{T2}$, and passage on HeLa cells until resolution was observed. Sequence names are given on the left with the top three, #105B, #105B$^{T1}$ and #105B$^{T2}$, representing parental constructs with modified sequences flanking the resolution window represented by a triangle (#105B$^{T1}$) or a circle (#105B$^{T2}$). Fully resolved or partially resolved sequences are shown with a dashed line representing the region that has been lost from the parental sequence, with appropriate shapes denoting the parental construct from which the 5′ and 3′ sequences were derived.

number of resolving variants are detected from infectious virus supernatant, despite the abundance of variants at the RNA level [13]. Taken together these data confirm that resolution occurs, at least in part, via further rounds of template-switching events.

## Discussion

The evolution of RNA viruses with single-stranded, positive-sense genomes is, in large part, driven by the process of recombination. This process can overcome the effects of deleterious mutations on individual genomes, thus continuing to provide the population diversity required to prevent error catastrophe [29], while also resulting in the generation of novel viruses with potentially novel phenotypes [30,31]. The mechanisms underpinning recombination have not been well characterised, but initial polymerase template-switching events frequently result in the insertion or deletion of genomic sequence, referred to as imprecise junctions [10]. We have previously demonstrated that viruses with imprecise insertions in the genome undergo a process–termed resolution–that leads to the loss of duplicated sequences and the creation of recombinant viruses with precise junctions and native or 'wild type' length genomes [10].

Recombination has previously been linked to the presence of RNA structures and/or regions of sequence identity [15–17,19,32–34], although there are a number of contradictory reports regarding the exact influence of such factors [18,20,35]. The majority of published studies on recombination have focused on analysing the end results of recombination i.e., viruses with wild type length genomes. If resolution is a distinct process from recombination, then it may be that the sequence and structure dependencies observed are a result of this second process and not the initial polymerase template-switching event. In the current study, we have therefore investigated this process of resolution in order to understand it better.

Throughout this study we have made use of a previously cloned imprecise intertypic type 3/1 poliovirus recombinant–#105B –containing a 249nt insertion spanning the P1/P2 junction boundary (Fig 2) [10]. To address whether sequence identity influenced the process of resolution we created an intratypic type 3/3 equivalent of #105B, designated #PV3-105B (Fig 2A), in which the duplicated viral sequences were all from poliovirus type 3. If sequence identity was a positive influence on resolution, we expected to see an increased rate of resolution for #PV3-105B when compared to #105B. Contrary to this expectation, we found that #PV3-105B resolved slower than #105B (Fig 2C). #PV3-105B also exhibited higher peak virus titres compared to #105B (Fig 2B). As we have previously demonstrated that resolution selects for increasingly fit viruses [10] we conclude that the slower resolution rate of #PV3-105B is a consequence of initial increased fitness, with concomitant reduced selection pressure for better replicating variants. Although this confirmed that resolution is primarily a fitness-driven process, the differing replication rates meant we could not conclusively rule out a role for sequence identity in the process.

The gradual appearance of lower molecular weight products over the course of #105B and #PV3-105B passaging, leading to wild type length genome populations supports a sequential resolution process, as previously suggested [10], with some intermediate products being non-viable. To investigate these, we employed ONT sequencing of amplicons generated from virus supernatant from a passaged resolving virus population. Although the long read-length of nanopore technology allowed analysis of the entire amplicon, the inherent high error rate necessitated the development of a novel analysis pipeline. We generated a database *in silico* that contained all 14,209 potential in-frame resolving and resolved genomes that could be present in supernatant, and used these as templates for alignment of ONT reads and the generation of variant consensus sequences (S1 Fig).

We identified a relatively small number of recombinant variants in total, 15 for #105B and 4 for #PV3-105B (Fig 3). We observed that the ONT analysis was sensitive to the threshold (number of reads to generate a consensus) used to define a junction (S1 Fig); reducing the threshold increased the range of junctions predicted, but introduced results potentially influenced by the template matched and the inherent high error rate of the technology. As the latter improves, the long ONT read lengths will likely offer the exciting potential for whole-genome recombination analysis.

One of the most interesting observations in the analysis of the #105B ONT data was the overlap observed in the variants isolated between different passaging replicates (Fig 4). The question arises as to whether these variants are (a) generated more readily or (b) have a fitness advantage, and therefore are amplified faster to detectable levels in the population. If resolution is a fitness driven process, then we would assume the latter. Resolved variant PV3$^{3378}$ was detected in both replicates at all passages analysed (Fig 4), with the greatest number of consensus sequences (S3 Fig), and therefore reads, generated, suggesting that this precise recombinant has a selective advantage. However, in direct competition assays with cloned recombinant PV3$^{3399}$ –a resolved variant detected less frequently and at lower levels–no fitness advantage was observed for cloned PV3$^{3378}$ (S6 Fig). This implies that more complex population dynamics are at play than can be determined by the sensitivity of such an assay.

Our ONT analysis of resolving viral genomes likely represented only a subset of the RNA variation present within the cell cytoplasm [13]. We therefore investigated the latter using Illumina next generation sequencing and observed significant levels of diversity within the RNA population (Figs 5A and S4). Based on the RNA population diversity observed at p0 via Illumina sequencing a large number of genome variants are generated immediately following transfection of a clonal RNA template (Fig 5A). Would the same infectious recombinants arise each time from different p0 RNA samples? We would expect this to be the case if those particular intermediates have advantages, either in their creation or their subsequent propagation, and would be an interesting topic for further study. We have shown that under the same passaging conditions parallel replicates generate similar resolving populations (Figs 4 and S3). However, altering those conditions–for example the volume of supernatant taken for subsequent passages–can influence the resolution process, as observed in the differences between the rate of resolution between the ONT and Illumina data sets. The viable variant population is therefore likely to be a result of both the input virus population and the stochasticity of the resolution process.

As recently demonstrated for recombination [13], viruses undergoing resolution generate an extensive population of diverse molecules (S4 Fig), the identity of which could arise from additional *in trans* polymerase-mediated template switching events. If that were the case, we would expect resolution to be influenced by factors known to influence recombination *per se*. Additionally we would also expect to be able to identify specific *in trans* crossover products with a suitable experimental system.

To test this hypothesis further we repeated the passaging of #105B in the presence of sublethal doses of ribavirin and analysed the resolving populations by gel electrophoresis (Fig 6). It is now well established that reduced polymerase fidelity results in increased error frequencies, which, in turn, is linked to an increase in recombination and correlates with recombination being a driving force behind virus evolution and avoidance of error-catastrophe-based population decline. By recapitulating this increased error rate through the use of ribavirin we observed an increase in resolution up to and including in the presence of 200 μM ribavirin, supporting our hypothesis that the mechanism underpinning resolution is further rounds of recombination.

Poliovirus polymerase mutations rates, both intrinsic and in the presence of mutagens such as ribavirin, have been well-studied [21,36–38]. Crotty *et al.* [21] demonstrated that the

mutation rates for poliovirus were increased from 2.1 per 10,000 nt to 2.5- and 9.3 per 10,000 nt in the presence of 100 μM and 400 μM ribavirin respectively–a range which encompasses the concentrations used in our study. Such an increase in the overall mutation rate may also increase the occurrence of point mutations elsewhere in the genome that are beneficial to the virus. Resolution may therefore be driven by selection of those variants that are lacking in duplications but also containing point mutations that improve virus fitness.

Notwithstanding this evidence, it remained possible that resolution occurs via a novel mechanism such as *in cis* deletion events. By its very nature recombination requires two replicating genomes but, only when there are differences between those two genomes can recombination be identified. To discriminate between *in cis* deletion and *in trans* recombination events during the resolution of #105B we engineered unique 'barcodes' into two copies of #105B genome (Fig 7) and passaged each virus in the presence of wild type #105B. Analysis of the resolving virus populations identified recombination events between barcoded and non-barcoded #105B genomes (Fig 8) in 15% of the progeny analysed. Of these, the two junctions identified in the three variants #105B/T2$^{3420}$, #105B/T2Δ207 and #T1/105BΔ207, had been observed previously (Fig 3A and 3C) supporting our theory that a select few recombinants with a replication advantage are likely to colonize a host.

Presuming recombination occurs during negative strand synthesis [4] the 3′ genome sequence can be regarded as being derived from the donor template, with the 5′ sequence derived from the acceptor. By its nature the CRE-REP assay could only generate viable genomes with a 5′ PV3 and 3′ PV1, and therefore it was not possible to establish if there was an inherent bias for templates to act as a donor or an acceptor. In the analysis of the barcoded recombinants we observed templates acting as both donor and acceptor based on the orientation of the 5′ and 3′ barcode sequences, with 5 examples of #105B as the donor and 4 examples of #105B as the acceptor, thus demonstrating, albeit with a limited number of samples, no obvious bias in the directionality of recombination as we have previously indicated [13].

It is clear, though perhaps unsurprising, that the resolution process can and does involve further rounds of recombination that necessitates an *in trans* strand transfer event. An iterative recombination process also explains the presence of out-of-frame recombinants continually detected throughout passaging of #105B (Fig 5C). Such junctions are possible when template-switching events occur at either position 1 or 2 of any individual codon and will be generated anew with each passage. However, as out-of-frame genomes will not go on to produce infectious progeny virions, and as shown in this study (Fig 5A), the relative percentage in the population will eventually be surpassed by the fitter, and replication competent, in-frame genomes.

While the evidence highlights a role for *in trans* template-switching events, it remains unclear as to whether *in cis* events are also driving the evolution of viruses with imprecise genomes. Ongoing studies investigating the behaviour of polymerases during RNA synthesis suggests that both *in trans* template-switching, and *in cis* copy-back type recombination events can occur [39,40], and such research may offer valuable insights into our further understanding of this fundamental process.

## Materials and methods

### Virus and cell culture

HeLa cells were maintained at 37˚C, 5% $CO_2$ in Dulbecco's Modified Eagle Medium (DMEM; Sigma) supplemented with 10% heat-inactivated FBS (FBS-DMEM; Life Technologies). Poliovirus type 1 (Mahoney), type 3 (Leon) and recombinant pT7-#105B were recovered following transfection of an *in vitro* transcribed RNA from full-length cDNA and virus stocks quantified by plaque assay on HeLa cells. Growth kinetics of viruses were determined following infection

of HeLa cells at a multiplicity of infection (MOI) of 5 in serum-free media (SF-DMEM). Unabsorbed virus was removed by washing with PBS and cells incubated in fresh FBS-DMEM at 37°C, 5% $CO_2$. Supernatants containing virus were harvested at various time points post-infection and quantified by plaque assay on fresh monolayers of HeLa cells.

### Serial passaging

Unless otherwise stated, for serial passaging HeLa cells were seeded at $1 \times 10^5$ cells/well in 12-well plates 18 h prior to infection. To generate passage 1, stock virus was infected at MOI 1, assuming a cell density upon infection of $5 \times 10^5$, for 1h at 37°C, 5% $CO_2$, and cells overlaid with 1mL/well FBS-DMEM. Supernatants were harvested at 8 h post-infection during early signs of CPE and 250 μL of each 1 mL supernatant was used to infect fresh monolayers of HeLa cells, as above, for subsequent passages.

### *in vitro* RNA transcription and transfection

PV1 and #105B variant plasmids were linearized with *Apa*I, and PV3 linearized with *Sal*I, and RNA transcribed using a HiScribe T7 High Yield RNA Synthesis kit (NEB) following the manufacturers' protocol. RNA transcripts were DNaseI (NEB) treated to remove residual template DNA and column purified using a GeneJET RNA Purification kit (ThermoFisher) prior to spectrophotometric quantification. For virus stocks, 1 μg of RNA was transfected into 80% confluent HeLa cell monolayers using Lipofectamine 2000 (Life Technologies) in a 3:1 Lipofectamine 2000:RNA ratio as per manufacturers' protocol. Virus supernatant harvested 24 h post-transfection and quantified by plaque assay on HeLa cells.

### Synthesis and infection of barcoded #105B

The region of #105B in which resolution could occur to yield a precise genome (resolution window) is located from nt 3361 to nt 3720 of #105B. Five amino acids, with maximum coding redundancy, were selected flanking either side of this region: 5′-PV3_VP1 Pro277, Gly280, Pro281, Gly282, Val283 and, 3′-PV1_2A Leu39, Leu40, Val41, Thr42, Ser44. A geneblock library was synthesised (IDT) with an n at the 3rd base position of each amino acid. Additionally, Leu40 was altered from TTA > CTn to ensure maximum redundancy without altering the amino acid. Individual DNA fragments were used to replace the equivalent WT sequence in pT7-#105B using HiFi Builder cloning technology (NEB). Colonies were screened and sequenced and two plasmids containing unique and divergent sequences were selected as pT7-#105B$^{T1}$ and pT7-#105B$^{T2}$.

HeLa cells were seeded at $5 \times 10^5$ per well in 6-well plates 18 h prior to assay. Cells were co-infected with virus inoculum at a final MOI 10 for 30 min at 37°C. Inoculum was removed and cells washed once with PBS. Fresh media was added and cells incubated for a further 8 h at 37°C, 5% $CO_2$. Virus-containing supernatant was harvested and quantified by plaque assay on fresh HeLa cells prior to subsequent passage at MOI 10. RNA was extracted from viral supernatants at passages 7 and 9 and the resolution window amplified by RT-PCR as described below. PCR products were cloned using the pGEM-T Easy System (Promega) and sequence determined by Sanger sequencing (Eurofins Genomics, Germany) with primer VP1(F): 5′-GGTGTATTGGCAGTTCGTGTTG.

### RNA extraction and RT-PCR

Viral RNA was isolated from clarified cell culture supernatants using an Omega Total RNA Isolation Kit 1 (Omega BioTek, VWR) and reverse transcribed at 42°C using an oligo dT

primer and SuperScript II reverse transcriptase (Life Technologies) as per manufacturers' protocol. The resolution window was amplified using standard Taq polymerase (NEB) with an initial denaturing at 95°C for 30 sec, followed by 30 cycles of 95°C for 30 sec, 50°C for 30 sec and 68°C for 90 sec, and a final extension at 68°C for 5 min with primers PV3(F): 5′-GCAAACAT CTTCCAACCCGTCC and PV1(R): 5′-TTGCTCTTGAACTGTATGTAGTTG, or PV3-IS(F): 5′-CCCAAACACGTACGTGTCTG and PV1-IS(R): 5′-TGTATCCGTCGAAGTGTGATGG for Illumina sequencing. PCR products were separated by 1% agarose gel electrophoresis and sequence determined by Sanger sequencing (Eurofins Genomics, Germany). Sequence analysis was completed using SnapGene, Version 4.0.8. Illumina sequencing was carried out and analysed as previously described [13].

## Ribavirin treatment

HeLa cells were seeded at 2 x $10^5$ per well in 12-well plates and infected with #105B at MOI 1 for 30 min. Cells were washed with PBS and fresh media added to all wells. Ribavirin was added to a final concentration of 0, 50, 100 or 200 μM and cells incubated for 24 h. Virus was subsequently blind passaged by infecting fresh monolayers of HeLa cells with a 1:4 dilution of supernatant from the previous passage, with Ribavirin added as above. At each passage, RNA was extracted from viral supernatants and the resolution window amplified by RT-PCR as described above.

## Next generation sequencing library preparation and sequencing

Viral RNA was extracted from serially passaged #105B and #PV3-105B virus supernatants and reverse transcribed as described above. For Illumina sequencing, samples were prepared and reads analysed as described in [13].

For Oxford Nanopore Technologies (ONT) analysis, primers for the initial PCR amplification of the resolution window, were adapted to include the Oxford Nanopore adaptor sequence and reactions cycled as described above. PCR products were purified using a Wizard SV Gel and PCR Clean-up System (Promega). Samples were barcoded using Oxford Nanopore PCR barcodes (EXP-PBC001) as per manufacturer's protocol. PCR products were purified as above and quantified on a NanoDrop ND1000. Amplicons were pooled in desired ratios (S1 Table) with a total of 1 μg of amplicons used to create a DNA library. Library construction was carried out as per Oxford Nanopore protocol SQK-LSK109. The library was run on a FLO-MIN106 flow cell primed as per protocol SQK-LSK109 with reads collected over an 8 h period.

## Oxford Nanopore Technologies sequence data processing

Raw fast5 files were basecalled and de-multiplexed using ONT Albacore Pipeline software (v 1.2.4). Passed reads were converted to FASTA format using the extract command in Nanopolish (v 0.6.1) [41].

A variants database was created of 14,209 potential in-frame, resolved/resolving sequences within the resolution window. ONT reads were screened against the variants database using Blastn. Reads with <90% sequence identity to a potential variant, and that were +/-10% of the min/max possible length of the PCR product, were removed. Remaining reads were extracted and ordered based on each potential variant and all the ONT reads which matched to it. Aligned reads were processed using Nanopolish to generate consensus sequences against best matched potential variants. A cut-off limit was applied, where only template matches which contained >100 ONT reads were processed to generate a consensus sequence. Using Nanopolish, all generated consensus sequences were verified as in-frame and confirmed as matching

the best potential variant template. A recombinant was defined as true if ≥10 identical consensus sequences were generated. Scripts available on request.

## Supporting information

**S1 Fig. Oxford Nanopore Technologies sequencing analysis pipeline.** Flowchart shows each step taken to generate recombinant sequences from Oxford Nanopore sequencing data. Boxes represent processes in the pipeline, diamonds represent decision points and ovals represent termination points.
(TIF)

**S2 Fig. Sequences of recombinant junctions identified from nanopore analysis.** Sequences denote the 30 nt spanning the junction of individual #105B (A) or #PV3-105B (B) resolving recombinants (designation in upper left corner). Uppercase sequences distinguish PV3 derived sequence from PV1 replicon derived sequence with numbers denoting the genome positions relative to full length PV3 and PV1. Numbers in the upper right corner indicate the number of additional nucleotides present in the recombinant genome, while underlined nucleotides indicate the reading-frame of the recombinant.
(TIF)

**S3 Fig. #105B variant population frequencies.** The frequency of each variant within the overall population was determined based on the number of consensus sequences generated for each variant. The number of consensus sequences for each variant was expressed as a percentage of the total number of consensus sequences generated at each passage (or pooled passages). Variants are shown for both Replicate 1 (black) and Replicate 2 (pink) passaging.
(TIF)

**S4 Fig. Parallel coordinates visualisation of #105B Illumina recombinants.** The location of each recombination junction was mapped respective to each parental genome. Each line represents a unique recombinant within the population of precise (black), imprecise-insertion (blue) and imprecise-deletion (red) recombinants. Each recombinant is shown as in-frame (1.0) or out-of-frame (0.0) with $\log_{10}$ read count below.
(TIF)

**S5 Fig. Shannon entropy diversity of Illumina sequenced resolving #105B populations.** The flattened data was used to calculate Shannon Entropy using the equation $H(X) = \sum_{i=0}^{N-1} Pi \ln pi$. The probability of each unique junction was calculated by dividing 1 over the total number of junctions, and plotted against passage number.
(TIF)

**S6 Fig. Sanger sequence analysis of competition assays between resolved recombinants.** Sequence trace analysis of the viral population following co-infection and 10 passages of 2 resolved variants, PV3$^{3378}$ and PV3$^{3399}$. Viruses were co-infected at ratios of 1:4 (PV3$^{3378}$: PV3$^{3399}$) (upper panel), 4:1 (PV3$^{3378}$: PV3$^{3399}$) (middle panel) and 1:1 (bottom panel) with traces showing the population sequence at p10. Grey boxes highlight the different nucleotides between PV3$^{3378}$ and PV3$^{3399}$ used to determine the dominant sequence in the population.
(TIF)

**S1 Table. Oxford Nanopore Technologies Library and Reads.**
(DOCX)

**S2 Table. Oxford Nanopore Variant Junctions.**
(DOCX)

**S3 Table. Barcoded #105B Variants.**
(DOCX)

## Author Contributions

**Conceptualization:** Kirsten Bentley, David J. Evans.

**Formal analysis:** Kirsten Bentley, Fadi Ghassan Alnaji, Luke Woodford.

**Investigation:** Kirsten Bentley, Fadi Ghassan Alnaji, Luke Woodford, Siân Jones, Andrew Woodman.

**Project administration:** David J. Evans.

**Supervision:** David J. Evans.

**Writing – original draft:** Kirsten Bentley.

**Writing – review & editing:** Kirsten Bentley, Fadi Ghassan Alnaji, Luke Woodford, David J. Evans.

## References

1. Xiao Y, Rouzine IM, Bianco S, Acevedo A, Goldstein EF, Farkov M, et al. RNA Recombination Enhances Adaptability and Is Required for Virus Spread and Virulence. Cell Host Microbe. 2016; 19 (4):493–503. https://doi.org/10.1016/j.chom.2016.03.009 PMID: 27078068; PubMed Central PMCID: PMC4840895.

2. Stern A, Yeh MT, Zinger T, Smith M, Wright C, Ling G, et al. The Evolutionary Pathway to Virulence of an RNA Virus. Cell. 2017; 169(1):35–46.e19. https://doi.org/10.1016/j.cell.2017.03.013 PMID: 28340348.

3. Kempf BJ, Watkins CL, Peersen OB, Barton DJ. Picornavirus RNA Recombination Counteracts Error Catastrophe. J Virol. 2019; 93(14). Epub 2019/06/28. https://doi.org/10.1128/JVI.00652-19 PMID: 31068422; PubMed Central PMCID: PMC6600191.

4. Kirkegaard K, Baltimore D. The mechanism of RNA recombination in poliovirus. Cell. 1986; 47(3):433–43. 0092-8674(86)90600-8 [pii]. https://doi.org/10.1016/0092-8674(86)90600-8 PMID: 3021340.

5. Bentley K, Evans DJ. Mechanisms and consequences of positive-strand RNA virus recombination. J Gen Virol. 2018; 99(10):1345–56. Epub 2018/08/29. https://doi.org/10.1099/jgv.0.001142 PMID: 30156526.

6. Cammack N, Phillips A, Dunn G, Patel V, Minor PD. Intertypic genomic rearrangements of poliovirus strains in vaccinees. Virology. 1988; 167(2):507–14. PMID: 2849237.

7. Cuervo NS, Guillot S, Romanenkova N, Combiescu M, Aubert-Combiescu A, Seghier M, et al. Genomic features of intertypic recombinant sabin poliovirus strains excreted by primary vaccinees. J Virol. 2001; 75(13):5740–51. https://doi.org/10.1128/JVI.75.13.5740-5751.2001 PMID: 11390576; PubMed Central PMCID: PMC114290.

8. Sergiescu D, Aubert-Combiescu A, Crainic R. Recombination between guanidine-resistant and dextran sulfate-resistant mutants of type 1 poliovirus. J Virol. 1969; 3(3):326–30. https://doi.org/10.1128/JVI.3.3.326-330.1969 PMID: 4305674; PubMed Central PMCID: PMC375771.

9. Ledinko N. Genetic recombination with poliovirus type 1. Studies of crosses between a normal horse serum-resistant mutant and several guanidine-resistant mutants of the same strain. Virology. 1963; 20:107–19. https://doi.org/10.1016/0042-6822(63)90145-4 PMID: 13929034.

10. Lowry K, Woodman A, Cook J, Evans DJ. Recombination in enteroviruses is a biphasic replicative process involving the generation of greater-than genome length 'imprecise' intermediates. PLoS Pathog. 2014; 10(6):e1004191. Epub 2014/06/12. https://doi.org/10.1371/journal.ppat.1004191 PMID: 24945141; PubMed Central PMCID: PMC4055744.

11. Goodfellow I, Chaudhry Y, Richardson A, Meredith J, Almond JW, Barclay W, et al. Identification of a cis-acting replication element within the poliovirus coding region. J Virol. 2000; 74(10):4590–600. https://doi.org/10.1128/jvi.74.10.4590-4600.2000 PMID: 10775595; PubMed Central PMCID: PMC111979.

12. Goodfellow I, Polacek C, Andino R, Evans D. The poliovirus 2C cis-acting replication element-mediated uridylylation of VPg is not required for synthesis of negative-sense genomes. J Gen Virol. 2003; 84(Pt 9):2359–63. https://doi.org/10.1099/vir.0.19132-0 PMID: 12917456.

13. Alnaji FG, Bentley K, Pearson A, Woodman A, Moore JD, Fox H, et al. Recombination in enteroviruses is a ubiquitous event independent of sequence homology and RNA structure. bioRxiv. 2020:2020.09.29.319285. https://doi.org/10.1101/2020.09.29.319285

14. Runckel C, Westesson O, Andino R, DeRisi JL. Identification and manipulation of the molecular determinants influencing poliovirus recombination. PLoS Pathog. 2013; 9(2):e1003164. https://doi.org/10.1371/journal.ppat.1003164 PMID: 23408891; PubMed Central PMCID: PMC3567174.

15. White KA, Morris TJ. RNA determinants of junction site selection in RNA virus recombinants and defective interfering RNAs. RNA. 1995; 1(10):1029–40. PMID: 8595558; PubMed Central PMCID: PMC1369329.

16. Nagy PD, Zhang C, Simon AE. Dissecting RNA recombination in vitro: role of RNA sequences and the viral replicase. EMBO J. 1998; 17(8):2392–403. https://doi.org/10.1093/emboj/17.8.2392 PMID: 9545250; PubMed Central PMCID: PMC1170582.

17. Carpenter CD, Oh JW, Zhang C, Simon AE. Involvement of a stem-loop structure in the location of junction sites in viral RNA recombination. J Mol Biol. 1995; 245(5):608–22. https://doi.org/10.1006/jmbi.1994.0050 PMID: 7844830.

18. Bujarski JJ, Kaesberg P. Genetic recombination between RNA components of a multipartite plant virus. Nature. 1986; 321(6069):528–31. https://doi.org/10.1038/321528a0 PMID: 2423879.

19. Figlerowicz M. Role of RNA structure in non-homologous recombination between genomic molecules of brome mosaic virus. Nucleic Acids Res. 2000; 28(8):1714–23. https://doi.org/10.1093/nar/28.8.1714 PMID: 10734190; PubMed Central PMCID: PMC102819.

20. Nagy PD, Bujarski JJ. Efficient system of homologous RNA recombination in brome mosaic virus: sequence and structure requirements and accuracy of crossovers. J Virol. 1995; 69(1):131–40. https://doi.org/10.1128/JVI.69.1.131-140.1995 PMID: 7983703; PubMed Central PMCID: PMC188556.

21. Crotty S, Cameron CE, Andino R. RNA virus error catastrophe: direct molecular test by using ribavirin. Proc Natl Acad Sci U S A. 2001; 98(12):6895–900. https://doi.org/10.1073/pnas.111085598 PMID: 11371613.

22. Domingo E, Sheldon J, Perales C. Viral quasispecies evolution. Microbiol Mol Biol Rev. 2012; 76 (2):159–216. https://doi.org/10.1128/MMBR.05023-11 PMID: 22688811; PubMed Central PMCID: PMC3372249.

23. Tyler AD, Mataseje L, Urfano CJ, Schmidt L, Antonation KS, Mulvey MR, et al. Evaluation of Oxford Nanopore's MinION Sequencing Device for Microbial Whole Genome Sequencing Applications. Sci Rep. 2018; 8(1):10931. Epub 2018/07/19. https://doi.org/10.1038/s41598-018-29334-5 PMID: 30026559; PubMed Central PMCID: PMC6053456.

24. Routh A, Johnson JE. Discovery of functional genomic motifs in viruses with ViReMa-a Virus Recombination Mapper-for analysis of next-generation sequencing data. Nucleic Acids Res. 2014; 42(2):e11. Epub 2013/10/19. https://doi.org/10.1093/nar/gkt916 PMID: 24137010; PubMed Central PMCID: PMC3902915.

25. Alnaji FG, Holmes JR, Rendon G, Vera JC, Fields CJ, Martin BE, et al. Sequencing Framework for the Sensitive Detection and Precise Mapping of Defective Interfering Particle-Associated Deletions across Influenza A and B Viruses. J Virol. 2019; 93(11). Epub 2019/05/15. https://doi.org/10.1128/JVI.00354-19 PMID: 30867305; PubMed Central PMCID: PMC6532088.

26. Kempf BJ, Peersen OB, Barton DJ. Poliovirus Polymerase Leu420 Facilitates RNA Recombination and Ribavirin Resistance. J Virol. 2016; 90(19):8410–21. Epub 2016/09/12. https://doi.org/10.1128/JVI.00078-16 PMID: 27412593; PubMed Central PMCID: PMC5021434.

27. Woodman A, Arnold JJ, Cameron CE, Evans DJ. Biochemical and genetic analysis of the role of the viral polymerase in enterovirus recombination. Nucleic Acids Res. 2016; 44(14):6883–95. Epub 2016/06/17. https://doi.org/10.1093/nar/gkw567 PMID: 27317698; PubMed Central PMCID: PMC5001610.

28. Poirier EZ, Mounce BC, Rozen-Gagnon K, Hooikaas PJ, Stapleford KA, Moratorio G, et al. Low-fidelity polymerases of alphaviruses recombine at higher rates to overproduce defective interfering particles. J Virol. 2015. Epub 2015/12/18. https://doi.org/10.1128/JVI.02921-15 PMID: 26676773.

29. Gnädig NF, Beaucourt S, Campagnola G, Bordería AV, Sanz-Ramos M, Gong P, et al. Coxsackievirus B3 mutator strains are attenuated in vivo. Proc Natl Acad Sci U S A. 2012; 109(34):E2294–303. Epub 2012/08/01. https://doi.org/10.1073/pnas.1204022109 PMID: 22853955; PubMed Central PMCID: PMC3427060.

30. Ge XY, Li JL, Yang XL, Chmura AA, Zhu G, Epstein JH, et al. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. Nature. 2013; 503(7477):535–8. Epub 2013/10/30. https://doi.org/10.1038/nature12711 PMID: 24172901; PubMed Central PMCID: PMC5389864.

31. Hahn CS, Lustig S, Strauss EG, Strauss JH. Western equine encephalitis virus is a recombinant virus. Proc Natl Acad Sci U S A. 1988; 85(16):5997–6001. https://doi.org/10.1073/pnas.85.16.5997 PMID: 3413072; PubMed Central PMCID: PMC281892.

32. Chuang CK, Chen WJ. Experimental evidence that RNA recombination occurs in the Japanese encephalitis virus. Virology. 2009; 394(2):286–97. Epub 2009/09/19. https://doi.org/10.1016/j.virol.2009.08.030 PMID: 19766282.

33. Rowe CL, Fleming JO, Nathan MJ, Sgro JY, Palmenberg AC, Baker SC. Generation of coronavirus spike deletion variants by high-frequency recombination at regions of predicted RNA secondary structure. J Virol. 1997; 71(8):6183–90. https://doi.org/10.1128/JVI.71.8.6183-6190.1997 PMID: 9223514; PubMed Central PMCID: PMC191880.

34. Cascone PJ, Haydar TF, Simon AE. Sequences and structures required for recombination between virus-associated RNAs. Science. 1993; 260(5109):801–5. https://doi.org/10.1126/science.8484119 PMID: 8484119.

35. Bruyere A, Wantroba M, Flasinski S, Dzianott A, Bujarski JJ. Frequent homologous recombination events between molecules of one RNA component in a multipartite RNA virus. J Virol. 2000; 74 (9):4214–9. https://doi.org/10.1128/jvi.74.9.4214-4219.2000 PMID: 10756034; PubMed Central PMCID: PMC111936.

36. Crotty S, Maag D, Arnold JJ, Zhong W, Lau JY, Hong Z, et al. The broad-spectrum antiviral ribonucleoside ribavirin is an RNA virus mutagen. Nat Med. 2000; 6(12):1375–9. https://doi.org/10.1038/82191 PMID: 11100123.

37. Wells VR, Plotch SJ, DeStefano JJ. Determination of the mutation rate of poliovirus RNA-dependent RNA polymerase. Virus Res. 2001; 74(1–2):119–32. https://doi.org/10.1016/s0168-1702(00)00256-2 PMID: 11226580.

38. Sanjuán R, Nebot MR, Chirico N, Mansky LM, Belshaw R. Viral mutation rates. J Virol. 2010; 84 (19):9733–48. Epub 2010/07/21. https://doi.org/10.1128/JVI.00694-10 PMID: 20660197; PubMed Central PMCID: PMC2937809.

39. Dulin D, Vilfan ID, Berghuis BA, Hage S, Bamford DH, Poranen MM, et al. Elongation-Competent Pauses Govern the Fidelity of a Viral RNA-Dependent RNA Polymerase. Cell Rep. 2015. Epub 2015/02/11. https://doi.org/10.1016/j.celrep.2015.01.031 PMID: 25683720.

40. Janissen R, Woodman A, Lee K-M, Moustafa I, Fitzgerald F, Huang P-N, et al. Induced copy-back RNA synthesis as a novel therapeutic mechanism against RNA viruses. bioRxiv. 2020:2020.02.12.946558. https://doi.org/10.1101/2020.02.12.946558

41. Loman NJ, Quick J, Simpson JT. A complete bacterial genome assembled de novo using only nanopore sequencing data. Nat Methods. 2015; 12(8):733–5. Epub 2015/06/15. https://doi.org/10.1038/nmeth.3444 PMID: 26076426.