Recombination in enteroviruses is a ubiquitous event independent of sequence homology and RNA structure

Fadi G. Alnaji[1]*, Kirsten Bentley[2], Ashley Pearson[2], Andrew Woodman[3], Jonathan D. Moore[4], Helen Fox[5], Andrew Macadam[5] and David Evans[2].

[1] Department of Microbiology, University of Illinois Champaign-Urbana, Champaign-Urbana, 601 E. John St, Illinois, 61801, USA.
[2] Biomedical Sciences Research Complex and School of Biology, University of St Andrew, St. Andrews KY16 9ST
[3] Department of Biochemistry & Molecular Biology, 201 Althouse Lab, University Park, PA 16802, USA.
[4] John Innes Centre, Cell and Developmental Biology, Norwich, NR4 7UH, UK.
[5] Division of Virology, National Institute for Biological Standards and Control, Potters Bar, Hertfordshire, EN6 3QG, UK.

* To whom correspondence should be addressed. Email: falnaji@illinois.edu and kb209@st-andrews.ac.uk

ABSTRACT

Recombination within RNA viruses is an important evolutionary process that can significantly influence virus fitness and has been repeatedly reported to compromise vaccine effectiveness. However, its precise mechanism is poorly understood. Here, we used an established poliovirus-based *in vitro* assay (CRE-REP) to investigate the molecular determinants of recombination and show that neither sequence identity, nor RNA structure, have any significant effect on recombination frequency. Since the CRE-REP assay is confined by the ability to detect infectious virus progeny, we utilized deep sequencing to study the recombinant genome population that arises early in infection and before any bottleneck of selection for viable progeny. We were able to detect and analyse hundreds of recombinants containing sequence insertions or deletions, or that were of wild type genome length. While we found higher diversity in recombination events than from CRE-REP assays, the analyses demonstrate no biases towards sequence or structure, in support of the CRE-REP assay findings. The results suggest that genome functionality and fitness are of greater importance in determining the identity of recombinants. These studies provide critical information that can improve our understanding of the recombination process, and consequently allow for the production of less recombinogenic and more stable vaccines.

INTRODUCTION

The rapid evolution of RNA viruses is attributable to both the error prone nature of viral RNA-dependent RNA polymerases (RdRp) and the extensive exchange of genetic information achieved through the processes of recombination and – in segmented viruses – reassortment. Together, these are important drivers of virus evolution, often linked to the emergence of novel pathogens and disease outbreaks [1-4]. Recombination predominates in the positive-strand (mRNA-sense) RNA viruses and is proposed to occur via distinct replicative or non-replicative mechanisms. In the latter the formation of a full-length genome likely involves processing and ligation by cellular enzymes, and the detailed mechanism and biological significance remain unclear [5, 6]. In contrast, replicative recombination between two compatible virus genomes co-infecting the same cell is increasingly well-studied. During negative-strand synthesis the RdRp undergoes a copy-choice strand-transfer event, which can eventually result in the formation of a viable, chimeric genome [7]. Recombination may influence virus virulence, transmission and cell or host tropism. Its universal presence in the positive-strand RNA viruses reinforces its importance as an evolutionary mechanism, which has been suggested may have originated to rescue genomes from deleterious mutations acquired during error-prone RdRp replication [8, 9].

The importance of recombination in nature is typified by the frequent isolation of recombinant forms of members of the *enterovirus* genus of the family *Picornaviridae*. The enteroviruses possess a ~7.5Kb genome encoding a single polyprotein flanked by 5′ and 3′ untranslated regions (UTR) containing signals necessary for viral translation and replication. Co- and post-translational processing of the polyprotein yields the P1 structural proteins (VP4, VP3, VP2 and VP1), and the P2 and P3 non-structural proteins (2Apro, 2B, 2C, 3A, 3BVpg, 3Cpro and 3Dpol). The P2 and P3 proteins carry out a variety of roles including the subversion of the cellular environment and the establishment of membrane-bound replication complexes within which the virus genome is replicated. While recombination events have been observed within phylogenetically distinct species of enteroviruses [10] these are primarily restricted to intraspecies recombinants. Poliovirus (PV), the prototype enterovirus of species C, exists as three serotypes (PV1-3) which have long been known to recombine, for example in recipients of the trivalent live-attenuated vaccine [11]. More recently, during monitoring of the poliomyelitis eradication campaign, the emergence of vaccine-derived recombinant polioviruses (VDRP) have been demonstrated arising from Sabin strains and co-circulating species C enteroviruses [12, 13]. This emphasises the need for an improved understanding of the recombination mechanism, with the goal of developing non-recombinogenic vaccines.

Poliovirus and other picornaviruses have been extensively used to study recombination [7, 14]. These include the initial demonstration of the copy-choice strand-transfer event and numerous studies interpreting the influence of RNA sequence and structure on the generation of viable recombinants. To facilitate the analysis of early events in recombination we developed an *in vitro* assay (the CRE-REP assay) that solely yielded recombinant viruses [15]. Using this we demonstrated that the process was biphasic. An initial imprecise recombination event generates hybrid genomes bearing duplications (which we termed imprecise junctions). These were subsequently lost during a secondary event, often iterative, which we termed resolution. The final resolved product contained a precise junction with no duplicated sequences. We additionally developed a defined biochemical assay which demonstrated that the RdRp was necessary and sufficient for recombination [16], and used both assays with a variety of RdRp variants to demonstrate that polymerase fidelity profoundly influences recombination frequency [8, 16, 17].

In previous studies RNA sequence identity and/or structure have been suggested as determining the location and frequency of the crossover junctions in recombinants [18-23]. However, since the CRE-REP assay indicates that viable recombinants involve both recombination *per se* and resolution events, this confounds inferring the influence of RNA sequence and structure on the strand-transfer event solely from the sequences at the precise junction in recombinant genomes. This may explain why these analyses can generate conflicting conclusions [18, 24], though this may also reflect variation in the processes involved between different viruses.

To directly address the influence of RNA structure and sequence identity on the strand-transfer event in recombination we have exploited the CRE-REP assay and modified the input genomes. By independently analysing genomes containing extensive regions of sequence identity, or templates modified to retain similar levels of identity but high or low levels of RNA structure, we were unable to demonstrate any significant influence on the rates or location of recombination. We therefore extended these studies to investigate the early RNA products arising from a natural co-infection of cells with two serotypes of poliovirus. In contrast to the generally clustered range of recombinants found using the CRE-REP assay [15] these were distributed throughout the genomic region analysed. These results support our contention that recombination is a promiscuous event that is not significantly influenced by either RNA sequence identity or structure of the parental genomes. We further propose that function and fitness are the primary determinants that influence the identity of recombinant genomes that are successfully replicated and predominate after genome recombination.

MATERIAL AND METHODS

Virus and cell culture

HeLa cells and L929 mouse fibroblast cells were maintained in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% heat-inactivated FBS (FBS-DMEM). Poliovirus type 1 (Mahoney) and type 3 (Leon) were recovered following transfection of *in vitro* transcribed RNAs from full-length cDNAs. Additional virus stocks for replication competent PV3-2A/2B junction modified viruses FLC/PV3^{L}, FLC/PV3^{H}, and FLC/PV3^{1} (see below) were similarly generated and all virus stocks quantified by plaque assay on HeLa cells. Growth kinetics of viruses were determined following synchronous 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. Recombinant viruses isolated from CRE-REP assays were biologically cloned by limit dilution in 96-well plates of confluent HeLa cells. Virus-containing supernatant was removed after 3 days incubation at 37°C, 5% CO_2 and stored at -80°C, and the remaining cell monolayer stained with a 0.1% crystal violet solution. Virus supernatants from the highest dilutions causing complete cytopathic effect (CPE) were utilised in further analysis.

For co-infection studies, poliovirus type 1 and type 3 were used to co-infect four T175 flasks at an MOI of 10. The cells were incubated for 30 minutes at 37ºC before the virus was removed and the cells were washed with PBS. Fresh DMEM media was added and the flasks were incubated at 37ºC for 5 h. After media removal, cells were washed with PBS, trypsinised and pooled in 2 ml DMEM media, giving a total of $7.6x10^{7}$ cells. The cells were lysed by three freeze-thaw cycles to extract virus and the debris was discarded. The supernatant contained the viruses was filtered through 0.2 µm filters and the resulting filtrate used for RNA extraction, followed by RT-PCR and NGS sequencing

Design of modified 2A/2B junction sequences

A 450 nt sequence, equivalent to nucleotides 3599 to 4045, of both PV1 and PV3 was designed using SSE [25] to generate sequences with the minimum and maximum amount of RNA structure possible by maintaining amino acid sequence and divergence between PV1 and PV3. The native PV1 and PV3 sequences were scrambled with CDLR in SSE. Each scramble was scrambled a further 100 times to generate a dataset of >100,000 sequences for each. Each dataset was processed using the reProcessSSEoutput.pl script, which renumbers each sequence and recalculates the divergence. Sequences were added to sqlite3 database (structureInfluence.DB) using populateDB.pl script, and processed with processDB.pl script. Within each table generated for PV1 and PV3 the average divergence was calculated, and the average divergence of the two tables was 0.14897947. Boundaries of +/- 1% were calculated and these included 6230 PV1-, and 6303 PV3-derived sequences respectively. The top 10 and bottom 10 energies from within this range were extracted from the database for each of PV1 and PV3, and entered into a matrix for comparison, and the identity of these 40 sequences calculated using SMS. A 'high' and 'low' structure sequence was then selected for synthesis for each of PV1 and PV3. Scripts available on request.

The same 450 nt sequence was altered for sequence identity by replacing the PV1 sequence with that of PV3 and *vice versa*. The sequence exchange was not 100% as there are 7 amino acid differences between PV1 and PV3 in this region – 2A residues 87, 116, 123 and 2B residues 3, 22, 26 and 30. At these amino acid positions the sequence was not altered between PV1 and PV3. Synthesis of all modified 450 nt DNA fragments was carried out by GeneArt (Life Technologies) and provided as sequence verified, plasmid DNA.

CRE-REP assay, plasmids, in vitro RNA transcription and transfection

The CRE-REP assay used to generate recombinant viruses has been described previously [15]. Wild type cDNAs of PV1 donor template, pRLucWT, and PV3 acceptor template, pT7/SL3, were used as a background for the generation of the 2A/2B modified cDNAs. Standard molecular biology techniques were used to replace the 450 nt sequence spanning the 2A/2B boundary with the synthesised DNAs described above to generate the CRE-REP cDNA variants referred to as $PV1^{L}$, $PV1^{H}$, $PV1^{3}$, $PV3^{L}$, $PV3^{H}$ and $PV3^{1}$. In addition, the relevant PV3 sequences were placed into a background of the replication competent PV3 cDNA, pT7/FLC, for growth kinetics analysis. All cDNAs were confirmed by sequence analysis prior to use.

Plasmids were linearized with *Apa*I (pRLucWT background) or *Sal*I (pT7/SL3 or pT7/FLC background) 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 all CRE-REP variant assays equimolar amounts of both template RNAs (based on 250 ng of acceptor) were prepared with Lipofectamine 2000 (Life Technologies) in a 3:1 Lipofectamine 2000:RNA ratio as per manufacturers' protocol, and transfected into 80% confluent L929 cell monolayers. Virus supernatant was recovered at 30 h post-transfection and virus quantified by plaque assay on HeLa cells. For virus stocks, 1 µg of RNA was transfected into 80% confluent HeLa cell monolayers using Lipofectamine 2000 as above, virus supernatant harvested at 12 h post-transfection and quantified by plaque assay on HeLa cells.

RNA extraction and RT-PCR

Viral RNA was isolated from CRE-REP clarified cell culture supernatant samples using a GeneJET RNA Purification Kit (ThermoFisher) and reverse transcribed at 42°C using an oligo dT primer and SuperScript II reverse transcriptase (Life Technologies) as per manufacturers' protocol. The region of recombination (VP1 to 2C) was amplified using primers PV3-F (5'-GCAAACATCTTCCAACCCGTCC-3') and PV1-R (5'-TTGCTCTTGAACTGTATGTAGTTG-3') and Taq polymerase (Life Technologies) with an initial denaturing at 94°C for 3 min, followed by 35 cycles of 94°C for 45 sec, 50°C for 30 sec and 72°C for 60 sec, and a final extension at 72°C for 10 min.

Following poliovirus co-infection, total RNA was extracted from cells by QIAamp Viral RNA Mini Kit (Qiagen) and reverse transcribed as above. Subsequently 5 µl from the cDNA synthesis reaction was used in the PCR amplification. The region of recombination was amplified using primers PV3-F2 (5′- CTCCAAAGTCCGCATTTACA-3′) and PV1-R2 (5′- ATCAGGTTGGTTGCTACA-3′) and Taq polymerase (Promega) with an initial denaturing at 95°C for 2 min, followed by 35 cycles of 95°C for 1 min, 58.1°C for 30 sec and 72°C for 1.5 min, and a final extension at 72°C for 5 min.

Sanger sequencing analysis

PCR products of recombinants from CRE-REP assays were analysed by Sanger sequencing (GATC Biotech, Germany), and the recombination junctions determined by aligning against parental PV1 and PV3 reference sequences using DNASTAR Lasergene software (v12.3.1).

NGS library preparation

Nextera XT DNA kit (Illumina, 2015a) was used to prepare the NGS samples for sequencing as per the manufacturer instructions. Briefly 1 ng from the amplicons was simultaneously fragmented and tagged with unique adapter sequences, followed by 10 cycles of PCR, before loading into the MiSeq instrument. The generated reads were uploaded to Illumina BaseSpace, a cloud-based genomics analysis and storage platform that directly integrates with all Illumina sequencers.

RESULTS

Influence of RNA structure and sequence on CRE-REP recombination

The CRE-REP assay [15] utilises a PV1 luciferase-encoding subgenomic replicon as the polymerase donor and a full-length PV3 genome (the acceptor) carrying an effectively lethal mutation in the *cis*-acting replication element (CRE) located within the 2C-coding region [26] (Fig 1A). The latter mutation means the acceptor template can only undergo negative-strand synthesis [27]. Co-transfection of *in vitro*-generated RNA from both templates yields viable recombinants if there is a strand-transfer event 5′ to the functional CRE element in the donor and 3′ to the P1 capsid-coding region in the acceptor template. These positions are separated by 1070 nt in the virus genome. We previously demonstrated that viable recombinants had junctions that predominantly clustered in one of two regions, spanning either the VP1/2A or 2A/2B polyprotein cleavage boundaries. We reasoned that by modifying the sequence identity or RNA structure within one of these regions in the donor and acceptor templates, we could investigate the local influences of sequence or RNA structure on recombination using the unmodified region as an internal control.

In previous studies we noted that recombinants at the VP1/2A border often incorporated heterologous sequences derived from the luciferase coding region of the donor [15]. Since this confounded the creation of identical donor and acceptor templates, we instead chose to modify a 450 nt window spanning the region encoding the 2A/2B junction (nucleotides 3599 – 4049). This encompassed all the imprecise recombinants characterised in our earlier study [15] and maintained an *Xba*I site, unique to the donor template, at nt 3572 that could be used for recombinant screening (Fig 1A). Sequences 5′ to the *Xba*I site are defined as Cluster 1 (C1) with those 3′ to nt 3572 defined as being within Cluster 2 (C2).

We exploited the redundancy of the amino acid triplet code to design 450 nt sequences, centred on the 2A/2B cleavage site, with high and low levels of RNA structure (see Materials and Methods). All designed sequences exhibited the same level of identity between themselves compared to their parental sequence (85.5%) and were selected to ensure the same level of sequence identity as is naturally present between the parental PV1 and PV3 CRE-REP partners (77.6% +/- 1%; Supplementary Table 1). In each case the encoded proteins were identical to the respective native PV1 donor or PV3 acceptor templates. Similarly, we engineered donor and acceptor templates with a 450 nt region of identity in the same region, with the exception of the retention of seven amino acid differences (see Materials and Methods) to avoid disruption of any *in cis* interactions that involved these residues. The resulting modified RNA templates for use in CRE-REP assays were: (1) the PV3-derived acceptor modified $PV3^{L}$, $PV3^{H}$ and $PV3^{1}$, and (2) the PV1-derived donor modified $PV1^{L}$, $PV1^{H}$ and $PV1^{3}$ (Fig 1B and S1).

The CRE-REP assay allows analysis of replicative recombination. Since it was possible that these modifications caused changes to the replication kinetics that could impact recombination rates, we analysed the influence on replication of the sequence changes engineered into the donor and acceptor templates. Luciferase-encoding donor RNA templates ($PV1^{L}$, $PV1^{H}$ or $PV1^{3}$) were generated *in vitro* and 250 ng independently transfected into HeLa cells. Resulting luciferase expression over an 8 h time course showed no significant difference (t-test) from the control unmodified WT template (PV1WT; Fig 2A). Since the CRE mutation harboured by the acceptor PV3 template in the CRE-REP assay is lethal we also engineered the acceptor template modifications ($PV3^{L}$, $PV3^{H}$ and $PV3^{1}$) into a replication-competent PV3 cDNA (pT7/FLC; PV3WT). Virus was recovered following transfection of HeLa cells, and the sequence and titre confirmed before conducting single-step growth assays in HeLa cells at MOI 5. Over a 24 h time course no significant (t-test) difference in replication was observed when comparing viruses with the modified acceptor sequences and PV3WT (Fig 2B). These studies confirmed that the sequences introduced to the CRE-REP donor and acceptor templates had no significant influence on virus replication.

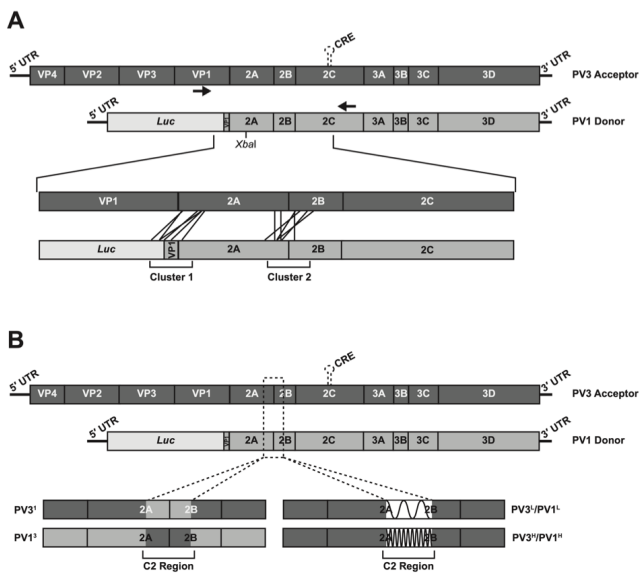

Fig 1. Schematic representation of CRE-REP assays. (A) Intertypic CRE-REP assay showing the genomes of the PV3 acceptor and PV1 donor RNAs. Black arrows represent position of primers used to amplify recombinant genomes. The lower expanded image illustrates the recombination window and highlights examples of recombination events in Cluster 1 and Cluster 2. The position of a unique *Xba*I site used for screening is shown in the donor template. (B) Modified CRE-REP assays highlighting the region of modification within Cluster 2 and the resulting acceptor and donor templates with altered sequence identity or RNA secondary structure.

To generate recombinant virus populations, we conducted CRE-REP assays in parallel. For each assay one unmodified template (either donor or acceptor) was paired with one modified donor or acceptor template. Assays are referred to by the name of the modified template of the RNA pair. Donor and acceptor RNAs were co-transfected into murine L929 cells (permissive but not susceptible to poliovirus) in equimolar amounts and virus-containing supernatant harvested 30 h post-transfection. A total of three independent co-transfections were carried out for each assay and the yield of recombinant virus determined by plaque assay (Fig 2C, D). The unmodified CRE-REP control assay (WT) generated an average of 2.1 x 10^3 pfu/ml (range 0.7 x 10^3 to 4.7 x 10^3 pfu/ml). Using unpaired t-test analysis we show that the total yield of recombinants from CRE-REP assays with template modifications did not significantly differ from the WT control assay, demonstrating no obvious advantage or disadvantage in recombinant generation resulting from the modifications

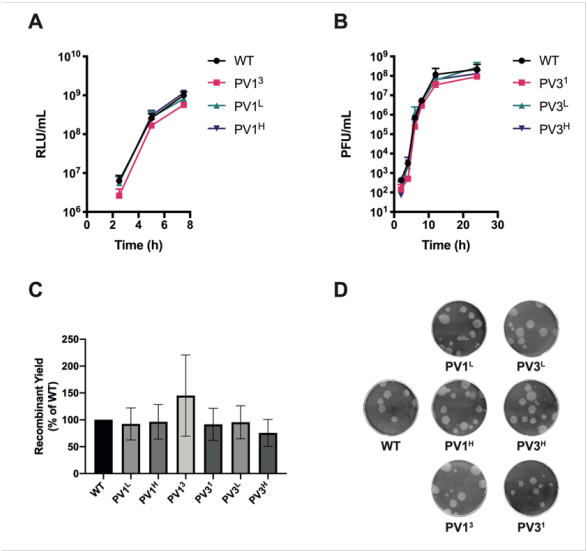

Fig 2. Phenotypic characterisation of donor and acceptor RNAs in modified CRE-REP assays. (A) Replication kinetics for donor RNAs. HeLa cells were transfected with 200 ng RNA/well in 24-well plates and luciferase activity measured over an 8 h time course. Error bars represent the standard deviation of three experiments with samples assayed in triplicate. (B) Replication kinetics for acceptor RNAs. HeLa cells were infected with virus stock at MOI 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 three experiments. (C) Recombinant yield from modified CRE-REP assays. RNAs were co-transfected into L929 cells in equimolar amounts and virus-containing supernatants harvest 30 h post-transfection. Recombinant virus yields were determined by plaque assay on HeLa cells and expressed as a percentage of the WT assay. Error bars represent standard error of the mean of three co-transfection experiments. (D) Plaque morphology comparison between recombinant viruses obtained from each of the CRE-REP assays.

Analysis of recombination junctions in the modified CRE-REP assay

The overall yield of recombinants (Fig 2C) includes all those that map between the donor CRE and acceptor P1 regions. To determine whether this total obscured significant increase or decrease in the proportion of recombination events within the C1 and C2 regions we sequenced a representative sample of recovered viruses. A total of 420 recombinants (60 from each CRE-REP assay) were generated by limit dilution of cell-free supernatant pooled from the three independent co-transfections of respective RNA partners. Following RNA extraction and reverse transcription, the entire recombination region was amplified by PCR and the products analysed by gel electrophoresis. The majority of products were larger than the distance separating the primers used (1470 nt.) indicating the presence of imprecise recombination junctions (data not shown) containing duplicated sequences of varying lengths. This was not unexpected and was in agreement with previous studies using the CRE-REP assay [15]. Subsequent sequence analysis of all the amplified products was carried out to determine the location, and type, of junction for each recombinant virus.

Following sequencing, 119 were discarded due to ambiguities in the sequence that indicated the virus population was not clonal. Approximately one third of the remaining recombinants were present more than once in the dataset. These identical sequences could either arise independently due to unique recombination events or may reflect early recombination products which have undergone additional rounds of replication, so increasing their proportion in the mixed virus progeny. Since the experimental design could not discriminate between these, we considered both the data in its entirety (unflattened) and after discarding identical sequences (flattened).

Of the 301 sequences determined in the unflattened data set 53% mapped to the C1 region and 47% to the C2 region. In comparison, in the flattened data, containing only unique recombinants (n = 205), 58% of junctions mapped to the C1 region and 42% to the C2 region (Fig 3B). Analysis of each individual CRE-REP assay showed that within the flattened data, recombination junctions in the parental (WT) assay were divided 62% and 38% to the C1 and C2 regions respectively, a figure close to that previously reported [15] (Fig 3C). The average distribution in each of the CRE-REP assays using modified templates was 58% and 42% (+/- 7%) in the C1 and C2 regions. We found no significant difference between the overall unflattened and flattened data sets (Fig 3B; Fishers Exact Test) and analysis of the flattened data demonstrated that there were also no statistically significant (Fishers Exact Test) favoured or disfavoured modified assay pairings that altered the distribution of junctions between the C1 and C2 regions from that seen with the unmodified WT templates (Fig 3C).

We therefore concluded that neither sequence identity, nor the gross level of RNA structure, play a major role in influencing the location of recombination junctions within viable recombinants that were selected and packaged and subsequently propagated.

Isolation of recombinants following virus co-infection

If, as the studies described above suggest, RNA sequence and structure do not influence the process of recombination we would expect these junctions to have been functionally selected *i.e.* on viability or relative fitness, from a much more diverse spectrum of crossover events. Such a diverse population potentially includes both out-of-frame junctions and deletions, neither of which would be represented in the viable recombinant progeny. To investigate this further and to determine whether the CRE-REP assay reflects the recombination junctions that arise during natural co-infection we developed an assay based upon the next generation sequencing (NGS) analysis of a mixed virus population arising after cell infection with serotypes 1 and 3 poliovirus.

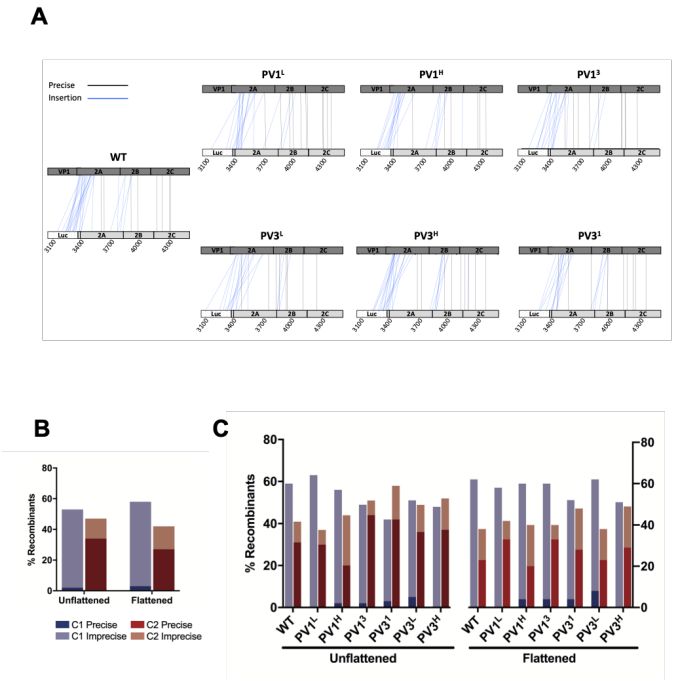

Fig 3. Analysis of recombination junctions from CRE-REP assay. (A) Parallel coordinates visualization of recombinants. The location of each recombination junction was mapped respective to each parental genome. Each line represents a unique recombinant within the population of precise (black) or imprecise-insertion (blue) recombinants. (B) Recombinant junctions by location and type. Junctions were defined as precise or imprecise and graphed according to location in Cluster 1 (C1) or Cluster 2 (C2). Unflattened data was compared to flattened data, with duplicate sequences removed, using Fishers Exact Test and showed no significant difference in the distribution of recombinants. (C) Recombinant junctions by CRE-REP assay. Junctions were defined as for (B) and graphed by individual CRE-REP assay. Each modified assay was compared to wild type using Fishers Exact Test and showed no significant difference in the distribution of recombinants, either between assays or unflattened and flattened data sets.

Recombination is an infrequent event and it was expected that the vast majority of progeny from a mixed infection would be non-recombinant parental genomes. To favour the isolation of recombinants we co-infected HeLa cells with PV1 and PV3 at a high MOI of 10 to ensure co-infection. Total cell RNA was harvested from 7.6 x 10^7 cells 5 hours post-infection (h.p.i), based upon studies by Egger and Bienz showing that recombinants can be detected at 2.5 h.p.i and are released from 6 h.p.i [28]. RNA was reverse transcribed and amplified using an optimised pair of recombinant-specific primers to amplify a region of the genome spanning the sequence 3235-4548 (see Materials and Methods) of the same recombination orientation as that isolated in the CRE-REP assay i.e. 5′-PV3/PV1-3′. Because total cell RNA was harvested without passaging virus on fresh cells there was no requirement for a recombinant genome to be replication competent for detection with the recombinant-specific PCR assay. In subsequent analysis by gel electrophoresis we observed a distinct product of ~1.3 kb, representing the expected distance between primers for precise recombination products, as well as both larger (up to ~3 kb) and smaller (~0.45 kb) products. In addition, there was diffuse background staining between these products suggestive of a smaller amount of variable sized PCR products (Fig 4A). To control for RT or PCR artefacts (negative control) an equimolar mixture of *in vitro* transcribed full-length PV1 and PV3 RNA was mixed and analysed following parallel RT-PCR reactions and failed to produce any visible products following gel electrophoresis.

The PCR products, including the smears, above and below the expected product of 1.3 kb were excised from the gel, cloned and representative samples were sequenced. All were derived from imprecise recombinant type 3/1 poliovirus genomes, with the size difference reflecting the presence of insertions or deletions respectively (data not shown). In previous studies we had observed in-frame insertions of up to 249 nt [15]. The identification of recombination products ranging from ~1.7 kb larger to ~0.85 kb smaller than expected for a precise junction, suggested a much greater variety in recombination events at the RNA level than previously observed when analysing packaged and released virus isolated in the CRE-REP assay.

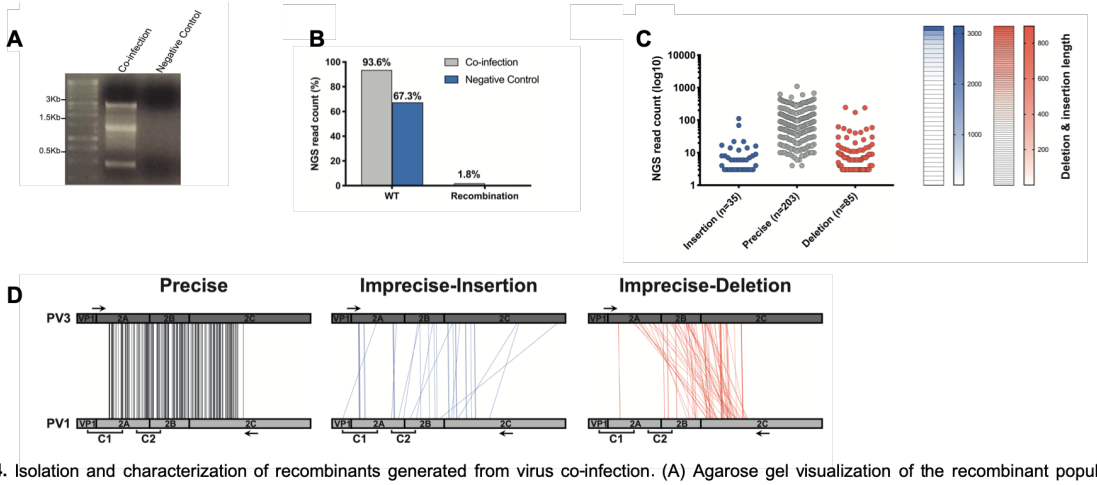

Fig 4. Isolation and characterization of recombinants generated from virus co-infection. (A) Agarose gel visualization of the recombinant population. PCR amplification of the region between nucleotides 3235-4548 followed by agarose gel screening. The negative control is an equimolar mixture of *in vitro* transcribed full-length PV1 and PV3 RNA (B) NGS read quantification comparison. The number of NGS reads that either mapped to WT genomes or were detected by ViReMa expressed as a percentage of the total reads in the co-infection and the negative control sample (C) Recombination type and length analysis. The number of precise and imprecise (both insertion and deletion) recombinants was plotted against the number of NGS reads (left panel); each dot represents a unique recombinant. On the right panel, heatmaps show the distribution of sequence lengths of insertion (blue) and deletion (red) within detected imprecise recombinants. Each cell represents a unique recombinant. (D) Parallel coordinates visualization of recombinants from co-infection studies. 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.

The complexity of recombination events revealed by NGS sequence analysis

To analyse the recombinant RNA population we developed a pipeline based upon the aligner Bowtie2 [29] and the Viral Recombination Mapper (ViReMa) [30]. We used a panel of simulated and experimental control datasets to optimize each step of the pipeline as described previously [31]. Fragmented PCR products amplified from poliovirus co-infected cells, or mixed control RNAs generated *in vitro*, were sequenced on Illumina MiSeq (2 x 250 nt paired-end reads). Subsequently, sequencing data were quality filtered, and paired reads were pooled into one fastq file prior to input to the analysis pipeline.

We initially aligned the entire dataset (~1.5 million reads) against the WT PV3 and PV1 reference sequences using Bowtie2 to eliminate those reads that perfectly matched either of the parental virus sequences. These reads constituted 93.6% and 67.3% of the total NGS reads of the co-infection and control samples, respectively (Fig 4B, Table 1). These uninformative reads were designated “WT-like” reads and were excluded from the data set. The unmatched reads were saved and analysed by ViReMa.

	Negative Control	Co-infection	Technical Replicate
Total reads	797386	1488344	1023536
Aligned to PV1	11754	279958	46028
Aligned to PV3	525214	1111733	909263
Not Mapped*	260418	69292	61749
Recombination reads	0	27361	6496

* Reads derived from cellular genome or tagged ‘unknown’ for a variety of reasons, including high levels of sequence variation and premature read truncation

Table 1. Summary of the NGS sequencing read numbers detected for all samples used in this study.

ViReMa works by iteratively matching a read to one input template and then, once the match cannot be extended, matching the remainder of the read to the second input template, returning the positions matched. Because a ~25 nt seed sequence is needed to initiate the mapping process, junctions that fall within the first and last 25 nt of the PCR amplicons (and their corresponding NGS reads after fragmentation) will be missed in the data analysis. Of the reads not designated "WT-like", ~30% contained identifiable PV3/1 recombination junctions (~1.8% of the entire dataset) with the remainder being unmapped by ViReMa (Fig 4B, Table 1). For completeness we analysed the latter using basic bioinformatic strategies (e.g. BLAST). Despite the use of virus-specific PCR amplification 11.6% matched mammalian host cells sequences, reflecting the difficulty in avoiding cellular read contaminants. A further 0.3% were located in the terminal 25 nt regions of NGS reads overlooked during ViReMa scanning with the remainder tagged as 'unknown' for a variety of reasons, including high levels of sequence variation and premature read truncation.

Comparable filtering and analysis of the control dataset from the RNAs synthesised *in vitro* failed to detect any recombination junctions. This confirmed that, at least with the primer pair used for this analysis, RT and PCR artefacts were not present in detectable amounts. We additionally failed to detect junction sequences in the reverse orientation (i.e. 5′-PV1/PV3-3′) in either the recombination or control datasets using ViReMa, confirming that the amplified viral sequences had resulted from a virus-specific PCR reaction.

Within the ~27 x 10^3 junction sequences identified (Table 1) there were 323 unique junctions between PV3 and PV1. These were initially grouped into those with no sequence duplication (so called precise junctions [15], n=203), those with sequence insertions (imprecise insertions, n=35), or those with sequence deletions (imprecise deletions, n=85). The two 'imprecise' groups exhibited a range of different size insertions (1 – 3150 nt) or deletions (1 – 896 nt; Fig 4C). All 323 recombinants were mapped using parallel coordinate diagrams, which clearly showed that recombination junctions, whether precise or imprecise, were distributed throughout the analysis region (Figure 4D), with no evidence for the clustering around the encoded polyprotein cleavage sequences seen in previous analysis [15].

To validate these observations, we performed a technical replicate by independently extracting total cell RNA from the original co-infection sample (Fig S2, Table 1). After amplification and fractionation, the population of recombinant molecules were sequenced and analysed using the same pipeline. Despite fewer NGS reads overall, and a concomitant reduction in ViReMa-detected crossover events, we were still able to identify hundreds of recombinants between PV3 and PV1. Correlation analysis of the NGS reads for each junction in the two replicates, normalised to the total reads per junction, demonstrated the method produced statistically reproducible results (Spearman R=0.88). All subsequent analysis used the first sequenced population containing 323 unique junctions.

These results indicate the presence of a wide variety of recombination junctions in dually infected cells in culture. The combination of recombinant-specific RT-PCR amplification, Illumina sequencing and a pipeline exploiting ViReMa to identify recombination junctions provided a robust and reproducible strategy to investigate the results of single stranded positive sense RNA virus recombination.

Analysis of recombination junctions following NGS

The NGS analysis of junction sequences provides an independent approach to determine the randomness, or otherwise, of the recombination process. To formally test whether the observed crossover junctions between PV3 and PV1 were randomly distributed across the analysed genomic region, we compared the data with a modelled random dataset. Precise recombinants – those in which there are no insertions or deletions with regard to the parental genomes – were compared with the random modelled data using a sliding cumulative score across the window of recombination (Fig 5A). No significant difference was found between the two populations (P=0.06; Mann–Whitney U test),

strongly implying that precise recombination junctions are randomly located in the RNA population within the targeted region. In support of this we did not find any bias in the nucleotide composition of the donor or acceptor sequences in the region immediately flanking both types of recombination junctions (data not shown).

ViReMa defines recombination junctions by scanning donor and template sequences from 5′ to 3′ for mismatches [30]. As a consequence, junctions are assigned to the first mismatched nucleotide 3′ to a region of sequence identity. Within the recombination window there are 219 short regions of sequence identity between 2 and 32 nt in extent. It was therefore necessary to investigate whether these regions of localized sequence identity influenced the distribution of precise recombination junctions. First, we made the assumption that recombination was equally likely to occur between any nucleotide within a region of sequence identity and its adjacent 3′ nucleotide. We then normalized the data based on our assumption; for example, if a precise junction with 100 NGS reads occurs next to a 4 nt stretch of sequence that is identical between PV3 and PV1, then this can be considered as 5 precise junctions each with 20 NGS reads; 4 junctions correspond to the identical sequence length and one junction to the adjacent nucleotide used by ViReMa to report the junction site (see Fig S3 for the normalization method).

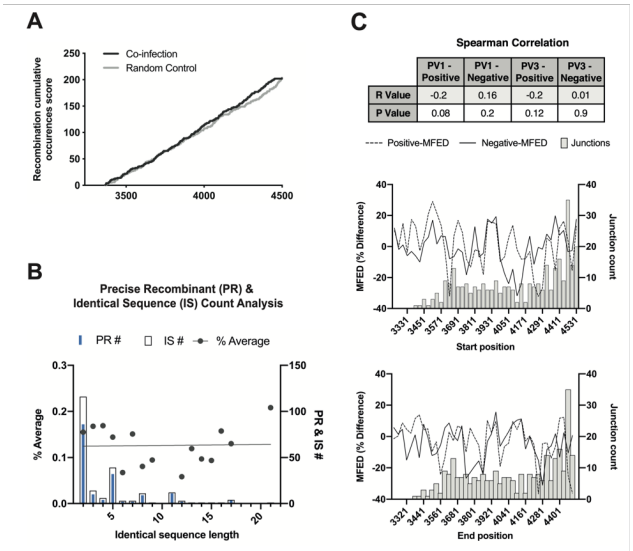

Fig 5. Analysis of recombination junction sequences generated from virus co-infection. (A) Determining the randomness of precise recombinant junctions. The occurrence of recombination on the genome was given a score of 1 while no occurrence is 0. The number of recombination occurrences was calculated in a cumulative way i.e. the occurrence of recombination at each position is added to the occurrence at the previous position. This was applied to two populations; the precise recombinants from the virus co-infection sample and a random recombination model generated by Excel (the same amount of recombination junctions were randomized). The cumulative recombinant count (y-axis) was plotted against the locations on the PV1 (donor) amplicon (x-axis). (B) Sequence identity and recombination. Individual counts of identical sequences (IS) of different lengths were summed within the amplified region and plotted against the count of precise junctions that mapped to each different length of sequence identity. For each of the latter the average number of junctions per nucleotide was calculated assuming recombination was equally likely to occur at each position within identical donor and template sequences (also see S3). The linear regression was calculated to compare the slopes between the fitted line and the perfect line. (C) The influence of RNA structure on recombination. The MFED figure, a measure of RNA structure, was calculated for a 100 nt sliding window (at 30 nt increments) for both PV1 and PV3 positive and negative strands (dashed and solid lines respectively) and plotted against the number of total recombinants within the same sliding window. A Spearman correlation coefficient was calculated between the recombinant count and the MFED values on either strand.

From our dataset, we calculated the average number of reads mapping to precise junctions within the 982 nt region in which recombinants were identified. The total number of precise junctions in this window is 203, while the cumulative total of reads that map to these junctions is 26,571; giving an average of ~131 reads per precise junction (i.e. 26571/203). Of these reads, ViReMa mapped the majority (24,987) to the first mismatched 3′ nt following one of the 219 short regions of sequence identity. We next calculated the average number of precise junctions mapping per unit length to the regions of sequence identity as described above. To determine if there was a correlation between the length of sequence identity and the distribution of junctions defined by ViReMa, we used linear regression to compare the slope between the perfect line and the line generated by analysis of the sequencing data. None of the precise recombinants mapped to the most extensive region of sequence identity (32 nt). Of the remainder, we found no significant difference in the slope (P=0.93), suggesting that recombination does not preferentially occur within regions of sequence identity between templates (Fig 5B).

In further independent analysis of the imprecise recombinants, we looked for biases in their location within the recombination window and for any obvious patterns in the extent of the insertion or

deletion with respect to the parental genomes. Of the 120 imprecise recombinants identified 21% maintained the polyprotein open reading frame (9 insertions and 16 deletions). We could detect no correlations suggesting that recombination is anything other than a random process, uninfluenced by reading frame or sequence identity, or sequence length (Fig S4).

Finally, to complement the reverse genetic analysis of the influence of RNA structure on recombination within the modified Cluster 1 and 2 regions (above), we looked for any potential influence of the underlying RNA structure in the recombinants identified by NGS. We calculated the Mean Folding Energy Difference (MFED) – an indicator of the amount of RNA structure in the virus genome [32], by comparison with the mean of 999 randomised sequences of the same nucleotide composition – using a sliding 100 nt window of the donor and acceptor sequences, for both the sense and antisense strands, and overlaid a plot of this with the recombinant count of all types of recombination within the same sliding window. A Spearman correlation was calculated between the MFED of either positive or negative strand and the number of junction counts (flattened data) within the indicated sliding window. As with the CRE-REP experimental approach (above) we could detect no correlation between RNA structure and recombination (Fig 5C).

Discussion

Recombination is an important evolutionary process in the single-stranded positive-sense RNA viruses [33, 34]. As an evolutionary process, recombination is probably ubiquitous, and may be a necessary rescue mechanism to accommodate – and escape – the build-up of deleterious mutations that arise from the error-prone polymerase [35, 36]. By creating hybrid genomes, recombination also contributes to maintaining or increasing the diversity of the viral quasispecies, a key contributor to virus fitness [37, 38]. Finally, as a consequence of the formation of hybrid genomes, recombination has the potential to create novel virus variants, with unique genotypes and phenotypes [39-41]

Despite this, the molecular mechanisms underlying recombination remain relatively poorly understood. Several previous studies have implicated RNA secondary structure or sequence similarity in influencing the location of recombination junctions (reviewed in [42, 43]). It has been suggested that RNA structures, such as hairpins or stem-loops present in either the donor or acceptor templates, may act as favourable sites for polymerase dissociation and re-initiation respectively [19, 44, 45], while homology between acceptor and donor templates is proposed to promote regions of dimerization that support precise switching of the polymerase [24, 46, 47]. In addition, recombination hotspots have been found in which specific nucleotides are more frequently associated at the junction site [18]. Although the terms homologous and non-homologous are often applied to explain the occurrence of recombination junctions in regions of the parental genomes that are similar or dissimilar, evidence for the direct contribution of sequence identity to the process of template switching remains elusive. The activity of the polymerase itself has also been studied using a well-defined biochemical assay [48], which has demonstrated roles for both polymerase speed and fidelity in the frequency of recombination [16, 38, 49, 50].

In previous studies we developed a system (the CRE-REP assay) that enables only the recovery of recombinants from two partially compromised input genomes. We used this system to demonstrate the generation of recombinants between type 1 and type 3 poliovirus that contained short genome duplications at the crossover junction. Subsequent passage resulted in the gradual loss of the duplicated sequences, a process we designated the resolution of imprecise recombinants [15]. We reported that, within the region of the genome analysed, recombinants formed clusters spanning the encoded polypeptide proteolytic cleavage sites. We proposed that this reflected a selection mechanism that operated at a functional level, to ensure production of replication competent virus genomes that

could subsequently be encapsidated to generate viable virus particles.

In the current study we have extended the use of the CRE-REP assay by exploiting the observed clustering of recombination junctions to investigate the contribution RNA sequence structure and template identity have on the process. In complementary experiments we used a recombinant-specific PCR assay, coupled with next generation sequencing of the population of amplified molecules, to investigate the recombinant products arising from mixed infection with poliovirus types 1 and 3. By initiating this analysis just five hours post-infection, a time previously associated with the presence of hybrid genomes and mixed replication complex [28] , we were able to survey the range of genomes present before significant selection due to replication, encapsidation or other events in the virus life cycle requiring a functional genome. Our results broadly support our hypothesis that the production of viable recombinant progeny viruses starts with an essentially random strand crossover event that generates a mixed population of hybrid genomes. These are then subjected to rigorous functional selection for those capable of replicating, being encapsidated and forming a virion able to initiate a productive infection in a new cell.

Numerous studies have identified the region at the junction of the P1 and P2 domains of the poliovirus polyprotein as a recombination hotspot [10, 51, 52]. Logically this might be expected as recombination would result in the juxtaposition in the new genome of two functional 'modules'; the structural proteins that form the virus particle, and the non-structural proteins that direct genome replication. Using the CRE-REP assay we demonstrated that within this area, viable functional recombinants initially clustered at the region encoding the VP1/P2A junction or the P2A/P2B junction [15]. We used the former as an *in cis* control, while engineering the sequences of the parental genomes to investigate their influence on recombination in the P2A/P2B region (Fig 1, 2). A 450 nt window in the donor and recipient parental sequence was substituted to generate genomes – with indistinguishable replication kinetics – that were 98.5% identical (cf. 77.6% for the parental type 1 and type 3 genomes in the same region). Likewise, taking advantage of codon degeneracy, we engineered equally divergent synonymous sequences at the nucleotide level into this region, but that were either highly structured or highly unstructured.

Using these input genomes in the CRE-REP assay we observed no significant variation in the yield of recombinants overall, or in the distribution of recombinants between the cluster 1 and cluster 2 regions of the area of recombination. Had either RNA sequence identity or structure been a significant influence on the mechanism of recombination we would have expected the distribution of recombinants to be skewed towards the cluster spanning the modified sequences that favoured polymerase template switching. Since this was not observed we conclude that recombination is not directly influenced by either RNA structure or sequence identity. Sequence analysis of 205 unique recombination junctions derived from these studies revealed that 29.8% were precise, and 70.2% were imprecise insertions. As expected, due to a presumed loss of functionality, we found no recombinants with genomes containing deletions. Between the flattened and unflattened data sets 96 junctions were isolated more than once, as represented by duplicate sequences. Of these the majority (66%) were isolated twice, 22% were isolated three times, 10% isolated 4 times and 2% isolated 5 times. Although we cannot determine whether these junctions arose as independent recombination events, or represent early events or fitter viruses whose growth therefore outpaced the rest of the population, the fact that removing them from the data set did not significantly alter the ratio of cluster 1 to cluster 2 junctions would suggest that these viruses had either a fitness advantage or were a result of increased infection events during the limit dilution cloning process.

By design, the CRE-REP assay generates viable progeny viruses for analysis, though these may have to undergo additional selection – the resolution process – to produce viruses sufficiently fit to compete with wild-type parental genomes [15]. As such, the CRE-REP assay does not allow the initial population of hybrid genomes, from which viable recombinants are selected, to be characterised. To address this, we developed a recombinant-specific assay targeting the same region of the genome as

analysed in the CRE-REP assay, to analyse cells co-infected with unmodified type 1 and type 3 poliovirus. Five hours after co-infection of HeLa cells, total cell RNA was harvested, reverse transcribed and recombinant viral RNAs amplified by PCR. This population of molecules was characterized using next generation sequencing.

Of the 1.5 million reads analysed, ~1.8% contained recombination junctions identifiable using ViReMa [30]. Approximately one third of these were imprecise, exhibiting a much wider range of insertions than previously identified [15] of up to +3150 nt. Of the imprecise junctions identified, 70% were deletions of up to -896 nt with respect to the parental genomes, a class of recombinant not identified previously using the CRE-REP assay, which is dependent upon the formation of infectious virus progeny. As expected, and in contrast to the CRE-REP assay, a significant proportion (79%) of all imprecise genomes lacked an intact open reading frame (ORF) due to the presence of either insertions or deletions that no longer rendered the multiple of 3 nucleotides. These results indicate that, within hours of cell co-infection, a widely divergent pool of recombinant genomes exist within the infected cell. A proportion of these – those with deletions in the polyprotein, or those with a truncated ORF – are evolutionary *cul-de-sacs* which cannot contribute to the viable progeny virus population.

It is not possible to meaningfully quantify the proportion of initial recombinant genomes that have precise or imprecise junctions, or that are in-frame or out-of-frame. Analysis of the numbers of NGS reads containing unique junctions is confounded by the rate at which an individual junction arises through polymerase template switching, the ability of the resulting genome to be replicated and the inherent biases in NGS library preparation. Nevertheless, even cursory analysis of the distribution of the junctions observed in the NGS dataset demonstrates a significantly greater range of junctions than we had previously observed in the CRE-REP assay. This result is not unexpected considering the rigorous selection imposed by the requirement, in the CRE-REP assay, to generate infectious virus progeny for analysis. Precise junctions were present throughout the region analysed, and imprecise junctions contained a much wider range of lengths of insertion or deletions (Fig 4) than previously observed. More detailed analysis of those junctions identified in the context of the multiple discrete short regions of sequence identity between the donor and recipient sequences, or to the RNA structure of either the positive or negative strands, provided no evidence supporting a role for sequence identity or RNA structure in influencing the location of recombination junctions, at least when analysed 5 hours post-infection.

The CRE-REP assay demonstrated that the production of a replication-competent recombinant genome was not a single-step process. Instead, it likely comprised an initial strand-transfer event followed by one or more stages of selection that enabled genomes with increased fitness – relative to the initial product – to be amplified. Similar conclusions were reached using a broadly analogous assay with divergent type (enterovirus parental genomes) [53]. RNA structure or sequence identity could contribute to any or all stages in this cascade of events, all of which are driven by functional constraints on the initial recombinant genome. Indeed, several previous studies have concluded that RNA structure [18, 19, 23] or sequence identity [19, 54, 55] were implicated in recombination.

A biphasic recombination process confounds the inference of the role for RNA structure or sequence identity in the events leading to the production of viable recombinants as it may be involved in genome function rather than recombination *per se*. For example, studies that map crossover junctions to structured regions of recombinant genome may conclude that they are causative, when the association may actually be correlative, but reflective of the genome modularity and relatively high fitness of a crossover in that specific region. Similarly, with genomes that exhibit variably extensive regions of sequence identity within an area in which recombinants are mapped, the functional selection may result in a clustering of crossovers in a region of sequence identity, not because of the identity *per se*, but because it creates a fit progeny genome, capable of competing well with others in the population.

We addressed these confounding issues using two complementary strategies. By modifying the parental genomes in the CRE-REP assay, whilst retaining an *in cis* control region, we could directly

analyse a role for RNA structure or sequence identity in recombination. We see no evidence for either. Secondly, by analysing the very early products of recombination – without the prerequisite for formation of replication competent genomes – using next generation sequencing, we obtained a snapshot of the diversity of genomes present within a co-infected cell. Analysis of these showed that crossover junctions were significantly more variable that the subset represented in viable progeny, but that again there was no correlation between the RNA structure or sequence identity of the parental genomes and the location of the recombination crossover.

A number of interesting questions remain about this important evolutionary mechanism which will require further study. If recombination is sequence and structure independent does the same apply to the iterative resolution process that follows? As initially proposed, resolution is an *in cis* event, but could equally well occur *in trans* as a recombination event between identical recombinants. If recombinants are generated at random as we propose, what are the predominant functional criteria that determine their selection? At a more mechanistic level, how does the template switching event occur?

DATA AVAILABILITY

The raw sequencing data is available in the NCBI sequence read archive (SRA) with accession number: PRJNA648771

REFERENCES

1. Hahn, C.S., et al., *Western equine encephalitis virus is a recombinant virus.* Proc Natl Acad Sci U S A, 1988. **85**(16): p. 5997-6001.
2. Ge, X.Y., et al., *Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor.* Nature, 2013. **503**(7477): p. 535-8.
3. Li, K.S., et al., *Genesis of a highly pathogenic and potentially pandemic H5N1 influenza virus in eastern Asia.* Nature, 2004. **430**(6996): p. 209-13.
4. Smith, G.J., et al., *Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic.* Nature, 2009. **459**(7250): p. 1122-5.
5. Gmyl, A.P., et al., *Nonreplicative RNA recombination in poliovirus.* J Virol, 1999. **73**(11): p. 8958-65.
6. Gmyl, A.P., et al., *Nonreplicative homologous RNA recombination: promiscuous joining of RNA pieces?* RNA, 2003. **9**(10): p. 1221-31.
7. Kirkegaard, K. and D. Baltimore, *The mechanism of RNA recombination in poliovirus.* Cell, 1986. **47**(3): p. 433-43.
8. Xiao, Y., et al., *RNA Recombination Enhances Adaptability and Is Required for Virus Spread and Virulence.* Cell Host Microbe, 2016. **19**(4): p. 493-503.
9. Stern, A., et al., *The Evolutionary Pathway to Virulence of an RNA Virus.* Cell, 2017. **169**(1): p. 35-46.e19.
10. Simmonds, P. and J. Welch, *Frequency and dynamics of recombination within different species of human enteroviruses.* J Virol, 2006. **80**(1): p. 483-93.
11. Cammack, N., et al., *Intertypic genomic rearrangements of poliovirus strains in vaccinees.* Virology, 1988. **167**(2): p. 507-14.
12. Arita, M., et al., *A Sabin 3-derived poliovirus recombinant contained a sequence homologous with indigenous human enterovirus species C in the viral polymerase coding region.* J Virol, 2005. **79**(20): p. 12650-7.
13. Kew, O., et al., *Outbreak of poliomyelitis in Hispaniola associated with circulating type 1 vaccine-derived poliovirus.* Science, 2002. **296**(5566): p. 356-9.

14. Sergiescu, D., A. Aubert-Combiescu, and R. Crainic, *Recombination between guanidine-resistant and dextran sulfate-resistant mutants of type 1 poliovirus.* J Virol, 1969. **3**(3): p. 326-30.
15. Lowry, K., et al., *Recombination in enteroviruses is a biphasic replicative process involving the generation of greater-than genome length 'imprecise' intermediates.* PLoS Pathog, 2014. **10**(6): p. e1004191.
16. Woodman, A., et al., *Biochemical and genetic analysis of the role of the viral polymerase in enterovirus recombination.* Nucleic Acids Res, 2016. **44**(14): p. 6883-95.
17. Kempf, B.J., O.B. Peersen, and D.J. Barton, *Poliovirus Polymerase Leu420 Facilitates RNA Recombination and Ribavirin Resistance.* J Virol, 2016. **90**(19): p. 8410-21.
18. Runckel, C., et al., *Identification and manipulation of the molecular determinants influencing poliovirus recombination.* PLoS Pathog, 2013. **9**(2): p. e1003164.
19. White, K.A. and T.J. Morris, *RNA determinants of junction site selection in RNA virus recombinants and defective interfering RNAs.* RNA, 1995. **1**(10): p. 1029-40.
20. Figlerowicz, M., *Role of RNA structure in non-homologous recombination between genomic molecules of brome mosaic virus.* Nucleic Acids Res, 2000. **28**(8): p. 1714-23.
21. Dedepsidis, E., et al., *Correlation between recombination junctions and RNA secondary structure elements in poliovirus Sabin strains.* Virus Genes, 2010. **41**(2): p. 181-91.
22. Shapka, N. and P.D. Nagy, *The AU-rich RNA recombination hot spot sequence of Brome mosaic virus is functional in tombusviruses: implications for the mechanism of RNA recombination.* J Virol, 2004. **78**(5): p. 2288-300.
23. Simon-Loriere, E., et al., *RNA structures facilitate recombination-mediated gene swapping in HIV-1.* J Virol, 2010. **84**(24): p. 12675-82.
24. Nagy, P.D. and J.J. Bujarski, *Efficient system of homologous RNA recombination in brome mosaic virus: sequence and structure requirements and accuracy of crossovers.* J Virol, 1995. **69**(1): p. 131-40.
25. Simmonds, P., *SSE: a nucleotide and amino acid sequence analysis platform.* BMC Res Notes, 2012. **5**: p. 50.
26. Goodfellow, I., et al., *Identification of a cis-acting replication element within the poliovirus coding region.* J Virol, 2000. **74**(10): p. 4590-600.
27. Goodfellow, I., et al., *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): p. 2359-63.
28. Egger, D. and K. Bienz, *Intracellular location and translocation of silent and active poliovirus replication complexes.* J Gen Virol, 2005. **86**(Pt 3): p. 707-18.
29. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nature Methods, 2012. **9**: p. 357-359.
30. Routh, A. and J.E. Johnson, *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): p. e11.
31. Alnaji, F.G., 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).
32. Simmonds, P., A. Tuplin, and D.J. Evans, *Detection of genome-scale ordered RNA structure (GORS) in genomes of positive-stranded RNA viruses: Implications for virus evolution and host persistence.* RNA, 2004. **10**(9): p. 1337-51.
33. Worobey, M. and E.C. Holmes, *Evolutionary aspects of recombination in RNA viruses.* J Gen Virol, 1999. **80 (Pt 10)**: p. 2535-43.
34. Simon-Loriere, E. and E.C. Holmes, *Gene duplication is infrequent in the recent evolutionary history of RNA viruses.* Mol Biol Evol, 2013. **30**(6): p. 1263-9.

35. Kempf, B.J., et al., *Picornavirus RNA Recombination Counteracts Error Catastrophe.* J Virol, 2019. **93**(14).
36. Crotty, S., C.E. Cameron, and R. Andino, *RNA virus error catastrophe: direct molecular test by using ribavirin.* Proc Natl Acad Sci U S A, 2001. **98**(12): p. 6895-900.
37. Vignuzzi, M., et al., *Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population.* Nature, 2006. **439**(7074): p. 344-8.
38. Lauring, A.S. and R. Andino, *Quasispecies theory and the behavior of RNA viruses.* PLoS Pathog, 2010. **6**(7): p. e1001005.
39. McWilliam Leitch, E.C., et al., *The association of recombination events in the founding and emergence of subgenogroup evolutionary lineages of human enterovirus 71.* J Virol, 2012. **86**(5): p. 2676-85.
40. Baker, J.C., *Bovine viral diarrhea virus: a review.* J Am Vet Med Assoc, 1987. **190**(11): p. 1449-58.
41. Meyers, G., et al., *Viral cytopathogenicity correlated with integration of ubiquitin-coding sequences.* Virology, 1991. **180**(2): p. 602-16.
42. Bentley, K. and D.J. Evans, *Mechanisms and consequences of positive-strand RNA virus recombination.* J Gen Virol, 2018. **99**(10): p. 1345-1356.
43. Muslin, C., et al., *Recombination in Enteroviruses, a Multi-Step Modular Evolutionary Process.* Viruses, 2019. **11**(9).
44. Nagy, P.D., C. Zhang, and A.E. Simon, *Dissecting RNA recombination in vitro: role of RNA sequences and the viral replicase.* EMBO J, 1998. **17**(8): p. 2392-403.
45. Carpenter, C.D., et al., *Involvement of a stem-loop structure in the location of junction sites in viral RNA recombination.* J Mol Biol, 1995. **245**(5): p. 608-22.
46. Cascone, P.J., T.F. Haydar, and A.E. Simon, *Sequences and structures required for recombination between virus-associated RNAs.* Science, 1993. **260**(5109): p. 801-5.
47. Draghici, H.K. and M. Varrelmann, *Evidence for similarity-assisted recombination and predicted stem-loop structure determinant in potato virus X RNA recombination.* J Gen Virol, 2010. **91**(Pt 2): p. 552-62.
48. Arnold, J.J. and C.E. Cameron, *Poliovirus RNA-dependent RNA polymerase (3D(pol)). Assembly of stable, elongation-competent complexes by using a symmetrical primer-template substrate (sym/sub).* J Biol Chem, 2000. **275**(8): p. 5329-36.
49. Kim, H., et al., *RNA-Dependent RNA Polymerase Speed and Fidelity are not the Only Determinants of the Mechanism or Efficiency of Recombination.* Genes (Basel), 2019. **10**(12).
50. Fitzsimmons, W.J., et al., *A speed-fidelity trade-off determines the mutation rate and virulence of an RNA virus.* PLoS Biol, 2018. **16**(6): p. e2006459.
51. Santti, J., et al., *Evidence of recombination among enteroviruses.* J Virol, 1999. **73**(10): p. 8741-9.
52. Joffret, M.L., et al., *Common and diverse features of cocirculating type 2 and 3 recombinant vaccine-derived polioviruses isolated from patients with poliomyelitis and healthy children.* J Infect Dis, 2012. **205**(9): p. 1363-73.
53. Holmblat, B., et al., *Nonhomologous recombination between defective poliovirus and coxsackievirus genomes suggests a new model of genetic plasticity for picornaviruses.* MBio, 2014. **5**(4): p. e01119-14.
54. Nagy, P.D. and A.E. Simon, *New insights into the mechanisms of RNA recombination.* Virology, 1997. **235**(1): p. 1-9.
55. Lai, M.M., *RNA recombination in animal and plant viruses.* Microbiol Rev, 1992. **56**(1): p. 61-79.