The Kinetic and Thermodynamic Aftermath of Horizontal Gene Transfer Governs Evolutionary Recovery

Sarah M. Doore¹ and Bentley A. Fane*¹
¹School of Plant Sciences and the BIO5 Institute, University of Arizona
*Corresponding author: E-mail: bfane@email.arizona.edu.
Associate editor: Eduardo Rocha

Abstract

Shared host cells can serve as melting pots for viral genomes, giving many phylogenies a web-like appearance due to horizontal gene transfer. However, not all virus families exhibit web-like phylogenies. Microviruses form three distinct clades, represented by ϕX174, G4, and α3. Here, we investigate protein-based barriers to horizontal gene transfer between clades. We transferred gene G, which encodes a structural protein, between ϕX174 and G4, and monitored the evolutionary recovery of the resulting chimeras. In both cases, particle assembly was the major barrier after gene transfer. The G4/ϕXG chimera displayed a temperature-sensitive assembly defect that could easily be corrected through single mutations that promote productive assembly. Gene transfer in the other direction was more problematic. The initial ϕXG4G chimera required an exogenous supply of both the ϕX174 major spike G and DNA pilot H proteins. Elevated DNA pilot protein levels may be required to compensate for off-pathway reactions that may have become thermodynamically and/or kinetically favored when the foreign spike protein was present. After three targeted genetic selections, the foreign spike protein was productively integrated into the ϕX174 background. The first adaption involved a global decrease in gene expression. This was followed by modifications affecting key protein–protein interactions that govern assembly. Finally, gene expression was re-elevated. Although the first selection suppresses nonproductive reactions, subsequent selections promote productive assembly and ultimately viability. However, viable chimeric strains exhibited reduced fitness compared with wild-type. This chimera’s path to recovery may partially explain how unusual recombinant viruses could persist long enough to naturally emerge.

Key words: virus assembly, horizontal gene transfer, phiX174, experimental evolution.

Introduction

Horizontal gene transfer (HGT) can lead to drastic phenotypic changes, such as metabolic reprogramming, antibiotic resistance, and virulence (Frost et al. 2005). In viruses, the nature of these transfer events can be dramatic: RNA viruses have recombined with DNA viruses (Diemer and Stedman 2012), plant viruses have recombined with animal viruses (Gibbs and Weiller 1999), and nearly complete satellite virus genomes have been found in other virus genomes (Swanson et al. 2012). Due to the parasitic nature of viruses, shared hosts can act as melting pots for viral genes, resulting in genomes produced by multiple gene transfer events. The mosaicism of double-stranded (ds) DNA bacteriophage genomes has been particularly well documented and can result in web-like phylogenies (Hendrix et al. 1999; Juhala et al. 2000; Lefeuvre et al. 2009; Casjens and Thuman-Commike 2011). The mechanism of recombination seems to act independently of selection, which will strongly favor mosaic genomes that can produce functional proteins and complexes (Hendrix et al. 2000).

Microviruses are small, single-stranded (ss) DNA, icosahedral bacteriophages that infect and lyse Enterobacteriaceae. Unlike the dsDNA bacteriophages, the approximately 50 known microviruses do not exhibit a web-like phylogeny. Instead, the individual phages primarily fall into three distinct clades, represented by ϕX174, G4, and α3 (Rokyta et al. 2006). Although this implies a low incidence of HGT, at least two possible transfer events have been previously documented (Rokyta et al. 2006). The most notable involves the gene encoding the external scaffolding protein, which appears to have originated in the ϕX174 clade before spreading to the G4- and α3-like viruses. The second transfer event may involve gene G, which encodes the major spike protein. Bacteriophage ID2, a member of the G4 clade, may have acquired a ϕX174-like major spike G gene (fig. 1A). Alternatively, the ID2 G gene may have originated outside the three major clades. Thus, HGT has shaped the evolution of the microviruses but not extensively enough to conceal relationships between clades. Recombination between microviruses should be possible: Sequences are similar enough to facilitate homologous recombination but dissimilar enough that a gene transfer event would be detectable. Thus, the dearth of HGT may indicate that proteins fail to efficiently function after transfer. In this scenario, natural recombinants would not survive to be sampled. This is the most widely accepted hypothesis explaining HGT barriers in bacteriophages (Hendrix et al. 1999; Juhala et al. 2000).

In order to investigate protein-based barriers to HGT, transfer events have been engineered. The resulting viruses were subsequently characterized and experimentally evolved (Rokyta and Wichman 2009; Springman et al. 2012). For example, Rokyta and colleagues exchanged coat F genes between two G4-like phages. Although the chimeras were less...
fit than the parental strains, they could still form plaques. Higher-fitness variants were experimentally evolved. The initial mutations primarily affected at least one of the two scaffolding proteins. This suggests that reduced fitness was due to inefficient particle assembly: The newly introduced protein may have raised a thermodynamic and/or kinetic barrier in the productive assembly pathway. While inefficient, these reactions are still more favorable than competing nonproductive reactions, which exist in every system. If the newly introduced protein raises morphogenetic barriers too high, off-pathway reactions are now energetically favored. If this occurs, assembly intermediates are siphoned into off-pathway kinetics traps. To overcome this type of barrier, nonproductive interactions must be suppressed before productive interactions can be optimized. To our knowledge, this type of barrier to HGT has not been explored.

All microviruses share similar structures and assembly pathways. Morphogenesis requires two types of scaffolding proteins, which divide the pathway into two phases (fig. 1B). The early phase is mediated by the internal scaffolding protein B. Five copies of B protein, along with one copy of DNA pilot protein H, bind to the underside of the 95 coat protein pentamer, producing the 95* particle. B protein induces a conformational change that facilitates interactions between the coat and the 6S G protein pentamer (Gordon et al. 2012). Joining the 6S pentamer atop the 95* intermediate produces the 125* particle (Cherwa et al. 2008). During the second phase of assembly, 240 copies of external scaffolding protein D organize twelve 125* particles into a procapsid (Cherwa, Organtini, et al. 2011). Afterwards, genome packaging occurs with the concurrent loss of the internal scaffolding protein. With the release of the external scaffolding protein, the capsomers undergo a radial collapse to produce the mature virion (McKenna et al. 1994; Dokland et al. 1997; Hafenstein and Fane 2002).

Major spike G and DNA pilot H proteins mediate host cell attachment and penetration (Inagaki et al. 2000, 2003; Sun et al. 2014; Young et al. 2014). Protein G sequences have diverged significantly between clades. The diverged residues primarily affect the 5-fold related G–G protein contacts in the 6S pentamer. Despite this divergence, the protein structures are nearly superimposable, and coat-spike protein interactions are highly conserved (McKenna et al. 1992, 1996; Bernal et al. 2003). Due to this conservation, diverse G proteins can likely interact with other phage proteins across species lines. Whether these interactions facilitate productive assembly or nonproductive reactions cannot be predicted. In this study, the major spike G genes were exchanged between ϕX174 and G4. These two proteins exhibit approximately 40% identity on the amino acid level. Our results indicate that the nature of the thermodynamic and/or kinetic aftermath was determined by the direction of transfer and corresponded to clade diversity. However, assembly was always the critical determinant of viability and fitness. The ϕX174 G protein may have elevated a thermodynamic and/or kinetic barrier when introduced into the G4 system. As the elevation was not severe, only inefficient interactions had to be optimized. This was easily accomplished in one mutational step. In contrast, when the G4 G protein was introduced into the ϕX174 morphogenetic pathway, nonproductive reactions were most favored. In this case, viability was ultimately achieved through a series of targeted genetic selections. Inhibitory reactions had to be suppressed before productive reactions could be promoted. The results of this study may partially explain how a complex series of adaptations can produce a new viral species.

**Results**

The G4-ϕXG Chimera Displayed a Temperature-Sensitive Assembly Defect

The G4-ϕXG chimera contained the ϕX174 major spike G gene within the G4 background. This chimera displayed a
severe temperature-sensitive (ts) phenotype, only forming plaques below 33°C. Even at lower temperatures (28°C), its fitness (doublings/hour) was greatly reduced compared with wild-type G4 (table 1). Expression of the cloned G4 G gene was both necessary and sufficient to restore plaque formation at higher temperatures (table 1). To determine whether the ts phenotype reflected a functional, postassembly defect, attachment and eclipse assays were performed. Particles assembled at the permissive temperature were assayed at the restrictive temperature. For attachment assays, cells and virions were mixed. As a function of time, the cells were removed by centrifugation. The titer of the supernatant reflects the unattached particles. For eclipse assays, particles were preattached to cells under conditions that prevent DNA delivery (16°C in nonnutrient buffer). As a function of time after temperature shift and media exchange, uneclipsed particles, which retain infectivity, were chemically dissociated from the cells. No significant differences in either attachment or eclipse kinetics were observed between the chimeric and wild-type virions (fig. 2A). This indicated that protein function within the context of the mature virion was not responsible for the ts phenotype. Thus, a possible defect in particle assembly was investigated.

To characterize morphogenesis, lysis-resistant cells were infected with either wild-type or chimeric G4 at a restrictive temperature (35°C). Infected cell extracts were analyzed by rate-zonal sedimentation in 5–30% sucrose gradients. After fractionation, assembled particles and morphogenetic intermediates were detected by UV spectroscopy and/or sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). The sedimentation profile of wild-type G4-infected cell extracts exhibited a large peak at 114S, indicating the presence of virions (114S). Virions were also detected, albeit minimally, in chimera-infected cell extracts (fig. 2B). Thus, the foreign spike protein does not completely abrogate assembly at the restrictive temperature. Moreover, the specific infectivity of these particles did not significantly differ from wild-type (4.0 × 10^{12} plaque forming units [pfu]/OD_{280} vs. 1.0 × 10^{12} pfu/OD_{280}). These data indicate that, once assembled, chimeric particles exhibit wild-type properties. G4-ψXG-infected cell extracts
were also examined for early assembly intermediates (fig. 1B). All three early intermediates, 6S, 9S* and 12S* particles, were detected (fig. 2C). However, no particle accumulated to gross excess. Typically, a complete block in the 12S*→procapsid transition results in a significant excess of 12S* particles over the 9S* and 6S particles (Cherwa et al. 2008; Cherwa and Fane 2009; Cherwa, Organtini, et al. 2011). This suggests that the foreign spike protein raised at least two thermodynamic and/or kinetic barriers in the assembly pathway: The 6S + 9S*→12S* and the 12S*→procapsid transitions.

### Altered Coat-Scaffolding Protein Interactions Rescued Chimeric Virion Assembly

To further investigate the nature of the assembly defect, a second-site genetic analysis was performed. Eight independently grown stocks were used, from which three suppressors were independently isolated multiple times (table 1). Thus, the selection likely identified the second-site suppressors with stronger suppressing phenotypes (Luria and Delbruck 1943). All suppressors were located in the coat protein (fig. 2D) and restored plaque formation at the restrictive temperature. However, they did not confer a fitness benefit at the permissive temperature (table 1). The atomic structure of the viral procapsid (McKenna et al. 1996; Dokland et al. 1997, 1999) and the results from previous studies (Burch et al. 1999; Holder and Bull 2001; Rokyta et al. 2009; Rokyta and Wichman 2009; Caudle et al. 2014; Miller et al. 2014) suggest that these mutations affect early morphogenesis (see Discussion). One of these (S141F) is a known suppressor of defective internal scaffolding protein function (Burch et al. 1999). Located at the coat-internal scaffolding protein interface (McKenna et al. 1996; Burch et al. 1999), it likely acts on the level of the 6S + 9S*→12S* transition. The other substitutions involved proline residues, which commonly suppress early assembly defects (Gordon et al. 2012), and are often recovered in selections for increased fitness at higher temperatures (Holder and Bull 2001; Rokyta et al. 2009; Rokyta and Wichman 2009; Caudle et al. 2014; Miller et al. 2014). Located in the insertion loops that cover the upper surface of the coat protein, they could alter coat-external scaffolding protein interactions and facilitate the 12S*→procapsid transition.

After altering these coat-scaffolding protein interactions, a second class of coat protein mutations appeared twice independently in the P35S5S background. These mutations, RS5→P and RS6→P, were always coisolated. In the ϕX174 X-ray structure, the side chain eta-nitrogens of these conserved arginine residues (source) interact with C-terminal amino acids in the spike protein (target). The target residues have diverged between ϕX174 and G4 (McKenna et al. 1994). These R→P substitutions would eliminate multiple coat-spike protein interactions. Thus, after overcoming the morphogenetic barrier, these mutations may restructure the coat-spike interface. Alterations within the spike protein C-terminus were also crucial to the evolution of the ϕX174-G4G chimera (see below).

### The ϕX-G4G Chimera Initially Required the Exogenous Expression of Two ϕX174 Genes

Unlike the G4-ϕXG chimera, the isolation of the ϕX-G4G chimera was much more problematic. This chimera contained the G4 major spike gene in the ϕX174 background. All attempts to isolate it in cells expressing the ϕX174 G gene resulted in large G gene deletions, suggesting that the foreign major spike protein was inhibitory. The major spike protein interacts with both the coat and the DNA pilot proteins (McKenna et al. 1992). If the newly introduced G4 spike protein interacts nonproductively with either of these ϕX174 proteins, it could siphon them off the productive assembly pathway. In this case, it would be possible to recover the ϕX-G4G chimera if the depleted protein’s intracellular level were elevated. In wild-type-infected cells, coat protein levels are very high and likely in excess in relation to protein H (Uchiyama et al. 2009; Cherwa, Young, et al. 2011). Thus, the limiting agent would likely be protein H and/or an H protein-containing intermediate. To test this hypothesis, we attempted to isolate the ϕX-G4G chimera in cells expressing both the ϕX174 spike G and DNA pilot H genes, where it was easily recovered. The strain could only form plaques on cells expressing both genes. To ensure that this phenotype was not an artifact of the construction scheme, a second chimera was produced using an alternate restriction site at the 3’-end of gene G. This strain also required the coexpression of both ϕX174 genes for plaque formation.

The assembly pathway was analyzed in chimeric phage-infected cells in the absence of exogenous gene expression. Unlike the extracts generated from wild-type-infected cells, no assembled large particles were detected (data not shown). Extracts were also examined for the presence of soluble early assembly intermediates. Early assembly intermediates were readily detected in the extracts of wild-type-infected cells (fig. 3A) but were below detection levels in the mutant extracts (fig. 3B). However, viral proteins were readily found in the insoluble fraction that could not be analyzed by rate-zonal sedimentation (fig. 3C). These observations suggested that the assembly intermediates were aggregating. Thus, with the introduction of the new protein, a competing off-pathway reaction has become energetically more favored than productive assembly. A similar molecular phenotype has been observed with a previously characterized H protein mutant that is not efficiently assembled into early assembly intermediates (Cherwa, Young, et al. 2011).

### Isolation of ϕX174-G4G Mutants Rescued by the Exogenous Expression of Only One ϕX174 Gene

To select for mutants that can suppress the off-pathway reaction involving protein H, we selected for strains that no longer required exogenous H gene expression. Approximately 10⁹ pfu was plated in 10⁸ aliquots on cells expressing only the ϕX174 G gene. Approximately 3,000 plaques were examined for nonwild-type plaque morphologies. Of these, 55 were selected for additional screening. Plaques were picked and stabbed into indicator lawns to characterize growth with or without complementation. From this screen,
only three phages maintained a complementation-dependent phenotype: Thus, these variants were present at a frequency between 10^{-9}/C0 to 10^{-10}/C0 within the stocks used in the selection (table 2). Genomes were sequenced to confirm the presence of the foreign major spike G gene. One mutant deleted the foreign gene, which was previously observed (see above). Two phages retained the foreign major spike gene but had undergone identical recombination events near the 3’-end. These mutants represent two independent isolations, as they were derived from independently produced stocks. Ninety-four percent of the resulting recombinant protein is derived from G4 (fig. 4). Ten out of 15 C-terminal residues are conserved between G4 and ‘X174: Thus, only five amino acids differ between the recombinant and the wild-type G proteins. The recombinant gene and protein are denoted as ‘X31'. A chimeric gene in an already-chimeric genome. Both phages displayed a ts phenotype.

In addition to the recombinant G gene, both independently isolated χ^2 strains contained a single base change at the end of gene C. This mutation results in a premature stop codon for glutamine codon 79 (CAA→TAA: och(C)Q79). The full-length C protein is 86 amino acids in length. In a wild-type infection, protein C mediates the switch from dsDNA synthesis to ss genomic DNA synthesis. As the mutation was not present in the parental strains, this suggests that the recombinant gene and premature stop codon were coselected. The ‘X31 G and ΔG strains were present at very low frequencies (10^{-9}–10^{-10}) within the stocks used in the selection. The premature stop codon was isolated three times independently: Twice in ‘X31 G strains and once in a ΔG strain. The frequency of coisolation suggests that the och(C)Q79 mutation may facilitate the passage through a narrow bottleneck. The mechanism by which it may accomplish this is described below.

A similar selection was performed using cells expressing only the ‘X174 H gene, which is the other gene product required by the parental strain. However, only wild-type recombinants were recovered. Thus, we were unable to isolate a modified chimeric strain that could form plaques using the foreign G protein, or a selected variant, when supplied with elevated levels of the H protein. If a strain with this phenotype

| Table 2. Plating Efficiencies for ‘X174 Wild-Type, Chimeras, and the Deletion Mutant. |
|--------------------------------------|-----------------|-----------------|-----------------|-----------------|
| Strain                              | Expressed Plasmid |                 |                 |                 |
|                                    | None             | ϕXG             | ϕXH             | ϕXGH            |
| ‘X174 wild-type                     | 1.0              | 1.0             | 1.0             | 1.0             |
| ϕX-G4G                              | <10^{-10}        | ~10^{-9}–10^{-10}| ~10^{-11}       | 1.0             |
| ϕX-GΔG                              | ND               | 1.0             | ND              | 1.0             |
| ϕX-G4G/χ^2                          | <10^{-11}        | 1.0             | <10^{-11}       | 1.0             |

Note.—Plating efficiency is defined as assay titer/most permissive titer.

Fig. 3. Assembly of the ϕX-G4G chimera. SDS-PAGE of small particle fractions from extracts of wild-type-infected cells (A) and chimera-infected cells (B). The location of the 12S*, 9S*, and 6S particles are given in the figure. M represents a marker of viral structural proteins: Coat (F), DNA pilot (H), major spike (G), internal scaffolding (B), and external scaffolding (D) proteins, as indicated. (C) Whole-cell lysate of chimera-infected cells. M represents a marker of viral structural proteins as indicated. ‘Uninf’ is a whole cell lysate of uninfected cells; WCL represents the lysate of chimera-infected cells.

Fig. 4. Alignment of the C-terminal regions of the G4, ‘X174, and recombinant (χ^2) G proteins. Recombination to produce χ^2 occurred at the indicated site.
could arise, it preexisted below a frequency of $10^{-11}$ in the stocks used in these selections (table 2).

The Premature Stop Codon in Protein C Leads to a Decrease in Overall Gene Expression

To determine the adaptive role of the premature stop codon, the och(C)Q79 mutation was placed in a wild-type background. Previous studies demonstrate that smaller naturally occurring C proteins are stronger inhibitors of ds replicative form (RF) DNA synthesis (Doore et al. 2014). A decrease in RF DNA, the template for transcription, could lead to decreased viral mRNA levels and, ultimately, in decreased production of the inhibitory protein. To investigate this hypothesis, levels of RF DNA and viral transcripts were examined in wild-type and och(C)Q79-infected cells. Cells harbored an empty plasmid, which was used as an internal control. ssDNA genomes, which are concurrently packaged and synthesized, are always associated with virions, which were removed during sample preparation. In wild-type infections, RF levels range from 15 to 20 copies per infected cell (Puga and Tessman 1973; Fujisawa and Hayashi 1976, 1977). RF DNA was characterized by two methods. In the first method, lysis-resistant cells were infected and incubated for 90 min, giving sufficient time for RF DNA to reach readily detectable levels. Plasmid and RF DNA were extracted and analyzed by gel electrophoresis. Unlike the wild-type control, RF DNA was undetectable in och(C)79-infected cells (fig. 5A). To quantify this phenomenon on a more physiologically relevant time scale, a second method was employed. Cells were infected and harvested 4 and 8 min postinfection. DNA was extracted and analyzed by quantitative polymerase chain reaction (qPCR), with primer pairs annealing to either the viral genome or the host cell plasmid, which served as an internal control. Viral genomes were quantified relative to plasmid DNA. At 4 min postinfection, wild-type- and och(C)Q79-infected cells displayed similar viral:plasmid DNA ratios, indicating a similar input of phage genomes. This ratio remained relatively constant in wild-type-infected cells, demonstrating that viral and plasmid DNA were synthesized at comparable rates. However, in och(C)Q79 infections, the viral:plasmid DNA ratio dropped 7-fold (fig. 5B). This suggests a much lower rate of viral RF DNA synthesis.

To quantify the effect of reduced viral DNA synthesis on the level of viral transcripts, total RNA was extracted and analyzed by quantitative reverse transcription (qRT)-PCR. In these experiments, viral mRNA for genes A, B, D, and G was quantified relative to a plasmid-encoded transcript. The och(C)Q79-infected cells contained much lower quantities of viral mRNA compared with the wild-type control (fig. 5C). This reduction was present at both time points even though RF DNA levels at min 4 were comparable to those observed in the wild-type infection. Moreover, the 7-fold reduction in RF DNA synthesis produced a dramatic decrease in viral transcripts: $10^{-2} – 10^{-3}$ the level found in wild-type-infected cells. This reduction was present at both time points even though RF DNA levels at min 4 were comparable to those observed in the wild-type infection. Moreover, the 7-fold reduction in RF DNA synthesis produced a dramatic decrease in viral transcripts: $10^{-2} – 10^{-3}$ the level found in wild-type-infected cells. This disparity may partially reflect the relative transcription-template efficiencies of different RF molecules (Hayashi Y and Hayashi M 1971; Puga and Tessman 1973; Fujisawa and Hayashi 1977). In both in vitro and in vivo studies (Hayashi Y and Hayashi M 1971; Puga and Tessman 1973), the most efficient template for transcription is the covalently closed, super-coiled RF DNA. This species of RF would be greatly reduced in the presence of a more active C protein, which inhibits both dsDNA replication and the requisite strand closure to produce a covalently closed circle (Fujisawa and Hayashi 1977). Next, to determine whether this mutation ultimately affected protein levels, coat protein concentrations were examined 90 min

![Figure 5](image-url)

**Fig. 5.** Characterization of the och(C)Q79 mutation. (A) Agarose gel of RF DNA extracted from wild-type and och(C)Q79-infected cells at 90 min postinfection. Representative of $n = 2$ experiments. (B) Quantification of WT and och(C)Q79 RF DNA at 4 and 8 min postinfection. Bars represent standard error; $n = 3$. (C) Quantification of viral transcripts at 4 and 8 min postinfection. Bars represent standard error; $n = 5$. (D) SDS-PAGE illustrating viral coat protein levels at 90 min postinfection. For densitometry, coat protein bands were normalized to the indicated host cell band.

Doore and Fane. doi:10.1093/molbev/msv130 MBE
postinfection. The coat protein could be detected in och(C)Q79-infected cells, albeit at reduced levels (fig. 5D). These results indicate that the premature stop codon leads to a global decrease in gene expression through the inhibition of RF DNA synthesis, which ultimately reduces the concentration of viral proteins in the infected cell.

Exogenous Expression of the $\chi^2$ G Gene Does Not Rescue och(C)Q79/$\chi^2$ G on the Level of Plaque Formation

Although the och(C)Q79/$\chi^2$ G no longer required the exogenous expression of gene H, it retained a complementation-dependent phenotype in relation to gene G. Thus, the $\chi^2$ major spike protein may not be efficiently incorporated into particles and/or its intracellular concentration may be too low as a consequence of the och(C)Q79 mutation. To determine whether elevated intracellular protein concentrations rescued plaque formation, the $\chi^2$ G gene was cloned and assayed for the ability to rescue och(C)Q79/$\chi^2$ G. As can be seen in table 3, no rescue was observed. These data indicate that further adaptations through changes in structural and/or scaffolding proteins are required. Moreover, the $\chi^2$ G protein may be inhibitory, as a majority of the protein is still encoded by G4. If true, a reduction in viral gene expression could facilitate complementation by the cloned $\varphi$X174 G gene. The inhibitory protein, being synthesized at a lower level, would be less likely to compete. To assay for inhibitory effects, wild-type $\varphi$X174 burst sizes were determined in cells expressing the cloned G4 or $\chi^2$ G genes and compared with two control infections: Cells with no plasmid and cells expressing the cloned $\varphi$X174 G gene. Bursts sizes were consistently 0.1–0.4 the level of the control infections ($n = 4$, data not shown).

In similar experiments, the expression of a cloned $\varphi$X174 G gene had no effect on wild-type G4 burst sizes. The results of these inhibition assays are consistent with the phenotypes of both the $\varphi$X-G4G and G4-$\varphi$XG chimeric strains.

Isolation of $\varphi$X174-G4G Mutants that Form Plaques in Cells Expressing the G4 Major Spike Gene

Due to the complexity of the assembly pathway, a foreign major spike protein could affect multiple morphogenetic stages. A series of genetic selections were taken to address these possibilities. For these experiments, the two original och(C)Q79/$\chi^2$ G isolates served as the parental strains and were propagated independently. The och(C)Q79/$\chi^2$ G mutant, which encodes a truncated C protein and the $\chi^2$ G protein depicted in figure 4, only forms plaques on cells expressing a cloned $\varphi$X174 G gene. Cloned $\varphi$X174 G gene expression is both necessary and sufficient for plaque formation. One isolate ultimately produced lineages B and D, whereas the other isolate produced lineage T (see below). In propagating these strains to high enough titers to conduct selections, rapid adaptation was evident. Plaque size became very heterogeneous. For example, in three independently grown populations of lineage D, large plaque variants represented at least 50% of the population. Sequencing results indicated that this variant was caused by an R→C substitution at amino acid 161 in the coat protein. Thus, the phage populations used in the following selections were already mixed, making it difficult to precisely ascertain which mutation(s) arose first.

Six different selections were performed as described above. They included: 1) Selections on wild-type cells (C122) expressing no exogenous viral genes; 2 and 3) selections on cells expressing either the internal scaffolding B or external scaffolding D gene, respectively, as mutations affecting scaffolding-structural protein interactions have predominated in other studies; 4) selections on cells expressing gene H, as aforementioned results indicate that protein H was limiting; 5 and 6) selections on cells expressing a cloned $\chi^2$ G or the G4 G gene, respectively. Approximately $1.3 \times 10^4$ plaques obtained from selections 1–5 were examined for altered plaque morphologies and 700 were subjected to further phenotypic screens as described in Materials and Methods. All 700 phages lost the gene G complementation-dependent phenotype. Moreover, they did not acquire cold sensitive, temperature sensitive, or novel complementation-dependent phenotypes in screens 2–5 (screen 6 is described below). Thus, they exhibited a primarily wild-type phenotype. This suggested that the foreign $\chi^2$ G gene was replaced by the $\varphi$X174 G gene. Before the selections, phage had to be propagated in cells expressing a cloned $\varphi$X174 G gene, which is where the recombination event would have occurred. To verify this conclusion, 18 phage genomes were sequenced and found to contain the wild-type $\varphi$X174 G gene. If these selections could yield a virus that retained the $\chi^2$ G gene, we estimate that the mutant existed below a frequency of $10^{-11}$ in the stocks used for the selections.

The sixth selection, which of course was the last one tried, readily yielded mutants (frequency $\sim 10^{-8}$ within the stocks used in these selections) that were rescued by the exogenous expression of the G4 G gene (G4Gututilizer). These mutants were not rescued by cloned $\chi^2$ G gene expression (table 3). However, the cloned $\varphi$X174 G gene could still complement them. Therefore, the substrate specificity for productive assembly was expanded. Based on whole-genome sequences, these strains had acquired multiple mutations (table 4). Only two strains—one each from lineages D and T—were chosen for more detailed analyses and subsequent selections.

Most mutations affected two regions of the genome. The first was the coat gene, which acquired mutations at sites involved in coat-external scaffolding protein interactions (Fane et al. 1993; Dokland et al. 1999; Burch and Fane 2003; Cherwa et al. 2008). The second was the interacistronic region between genes J and F. This region encodes a transcription

| Table 3. Plating Efficiencies for $\varphi$X174 Wild-Type and Chimera Derivatives. |
|-----------------|-------|-------|-------|
| Strain          | Expressed Plasmid |     |
|                 | None | $\varphi$XG | G4G | $\chi^2$G |
| $\varphi$X174 wild-type | 1.0 | 1.0 | 1.0 | 1.0 |
| och(C)Q79/$\chi^2$G | <10$^{-10}$ | 1.0 | 10$^{-8}$ | <10$^{-11}$ |
| och(C)Q79/$\chi^2$G/G4Gut | 10$^{-3}$ | 1.0 | 1.0 | 10$^{-3}$ |
terminator (TJ). Both base substitutions and a large deletion were recovered (table 4). These mutations would theoretically disrupt the secondary structure of the resulting mRNA, which can affect intracellular mRNA populations (Hayashi et al. 1983, 1989). TJ deletions have frequently arisen in other experimentally evolved strains (Bull et al. 1997; Wichman et al. 1999, 2005). Thus, alterations in this region appear to be a common step in many adaptations. A deletion of the terminator could lead to more messages for downstream genes. To investigate this possibility, qRT-PCR was performed with a TJ deletion in an otherwise wild-type background. Ratios were determined for a downstream gene (G) relative to a gene upstream from the terminator (B). No statistically significant difference in G:B ratios was detected, suggesting that the effect on transcript levels may be subtle (data not shown).

The deletion could also create a translational coupling arrangement: The stop codon of gene J now overlaps with the gene F ribosome binding site. To investigate an effect on translation, coat protein levels were monitored 40–80 min postinfection in wild-type and ΔJ-F-infected cells. In wild-type-infected cells, coat protein levels approximately doubled every 20 min compared with the host cell protein used for normalization. In contrast, coat protein levels in the ΔJ-F-infected cells nearly tripled between time points (fig. 6). Thus, the deletion may mitigate the effect of the och(C)Q79 mutation by increasing the translation of specific structural genes. G protein levels were too low and obscured by host bands to accurately follow the kinetics of its accumulation in whole cell lysates. As can be seen in table 4, several lineages eventually lost the ochre mutation, suggesting that it was primarily beneficial immediately after the G4 G gene replaced the ϕX174 gene.

Mutants that No Longer Required Exogenous Expression of Any Major Spike Gene Could Be Isolated and Were of Two Types

From the previous selection, two mutants (one each from lineages D and T) were screened for the ability to assemble without the exogenous expression of any spike gene. The parental mutants displayed very similar phenotypes, requiring the exogenous expression of a cloned ϕX174 or G4 G gene to form plaques. Thus, they obtained the ability to be complemented by G4 G protein. However, the genotypes differed. In addition to the och(C)Q79 and χV genes found in both strains, each parental strain (T4-10 and D-2) contained two additional coat protein mutations and a mutation in the J-F intercistronic region. However, these mutations differed between the two strains. A more detailed description is given in table 4. In both lineages, viable (complementation-independent) chimeras appeared at a frequency of approximately $10^{-3}$ in the stocks used in the selection. Lineage D produced two types of viable chimeras, whereas lineage T produced only one. The variant common to both lineages underwent an additional recombination event, in which the entirety of the major spike gene was now of G4 origin. These are denoted Lineage D and Lineage T $\phi X\cdot G4 GV$ (V: viable or Victory). The entire genomes were sequenced and no other mutations were detected. Lineage D also yielded a strain that retained the chimeric major spike gene ($\chi V$). This variant is denoted $\phi X\cdot G4 G/\chi V$. This strain acquired four additional mutations: One at the very end of gene F and three in the F-G intercistronic region, which contains the TJ transcription terminator and the gene G ribosome binding site.

The ultimate goal of this study was to characterize the requisite adaptations to productively integrate a structurally and functionally related but foreign protein into the $\phi X174$ system. As single plaques were selected throughout the course of these studies, the characterized mutants may neither represent the fittest possibilities nor equally fit alternatives. As can be seen in table 5, all viable chimeras exhibited reduced fitness compared with wild-type. It is likely that additional mutations are required before fitness losses are completely recovered.

Discussion

The Thermodynamic and/or Kinetic Aftermath of HGT

Viruses assemble through well-honed pathways, with proteins coevolving within the system to maintain efficient interactions (Fane and Prevelige 2003; Zlotnick and Fane 2010; Prevelige and Fane 2012). To accomplish this, productive reactions must be favored over the non-productive reactions that siphon intermediates off-pathway. In microviruses, scaffolding proteins suppress off-pathway reactions by preventing the formation of protein aggregates and kinetic traps (Gordon et al. 2012; Gordon and Fane 2013). When interacting components share evolutionary histories, the introduction of a structurally and functionally related but foreign protein will likely alter the thermodynamics and/or kinetics of assembly. In a simple case, the newly introduced protein may interact inefficiently. This could lead to a reaction having a higher activation energy: A kinetic barrier. Alternatively, the relative free energies of the products and reactants may change: A thermodynamic barrier. As long as the inefficient interactions remain more favorable than the off-pathway reactions, single mutations lower these barriers and restore efficient assembly. Several examples of this have been described in microviruses and other systems (Rokyta et al. 2009; Rokyta and Wichman 2009; Springman et al. 2012). Conversely, if the introduced protein elevates the activation energy of a productive reaction over those governing off-pathways reactions, and/or produces a very stable off-pathway product, adaptations will likely be more complex. To restore morphogenesis, and hence fitness, off-pathway reactions may have to be suppressed before or while productive reactions are made more favorable, which was observed in this study.

The number of proteins with which the newly introduced component must interact is related to the number of steps in an ordered pathway. This may accurately predict the number of nodes that can be perturbed by HGT and has insightfully been used to predict the success of transfer events (Jain et al. 1999; Cohen et al. 2011). However, it cannot completely explain posttransfer phenotypes. Levels of gene expression have also successfully predicted HGT outcomes but not in all cases...
## Table 4. List of Mutations Recovered in \( \psi X\text{-}G4G/\chi^2 \) Mutant Lineages.

<table>
<thead>
<tr>
<th>Selections and Strains</th>
<th>nt Change</th>
<th>aa Change *</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Complemented by ( \psi X\text{-}G )</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parental strain</td>
<td>C367T</td>
<td>(C)Q79och</td>
<td>Decreases expression of all viral genes [b]</td>
</tr>
<tr>
<td>Lineage T</td>
<td>C367T</td>
<td>(C)Q79och</td>
<td>No changes from parental strain</td>
</tr>
<tr>
<td>Lineage B</td>
<td>C367T</td>
<td>(C)Q79och</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T790C</td>
<td>(D)T134A</td>
<td>( D_1\text{-}G ) contact site [c]</td>
</tr>
<tr>
<td></td>
<td>A2276G</td>
<td>(F)T425A</td>
<td>Suppresses defective ( D ) protein interactions [d,e]</td>
</tr>
<tr>
<td>Lineage D</td>
<td>C367T</td>
<td>(C)Q79och</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C1484T</td>
<td>(F)R161C</td>
<td>( F_4\text{-}D ) contact site [e]</td>
</tr>
<tr>
<td><strong>Complemented by ( G4\text{-}G )</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lineage D</td>
<td></td>
<td></td>
<td>All mutations from the previous selection, plus . . .</td>
</tr>
<tr>
<td>Isolate D1</td>
<td>C1602T</td>
<td>(F)A200V</td>
<td>( F_4\text{-}D ) contact site [e]</td>
</tr>
<tr>
<td></td>
<td>G2084T</td>
<td>(F)T361F</td>
<td>Loop structure between ( \alpha 4 ) and ( \alpha 5 ) [f] in ( G4\text{-}G )</td>
</tr>
<tr>
<td></td>
<td>C4053A</td>
<td>(G)T11T</td>
<td></td>
</tr>
<tr>
<td>Isolate D2</td>
<td>G972A</td>
<td>—</td>
<td>Potential disruptor of ( T_j ) [b]</td>
</tr>
<tr>
<td></td>
<td>A1664G</td>
<td>(F)D154G</td>
<td>( F_1\text{-}D ) contact site [e]</td>
</tr>
<tr>
<td></td>
<td>C2280T</td>
<td>(F)S426L</td>
<td>Suppresses defective ( D ) protein interactions [d,e]</td>
</tr>
<tr>
<td>Isolate D3</td>
<td>C2280T</td>
<td>(F)S426L</td>
<td>Suppresses defective ( D ) protein interactions [d,e]</td>
</tr>
<tr>
<td></td>
<td>C4053A</td>
<td>(G)T11T</td>
<td>in ( G4\text{-}G )</td>
</tr>
<tr>
<td>Lineage T</td>
<td></td>
<td></td>
<td>All mutations from previous selection, plus . . .</td>
</tr>
<tr>
<td>Isolate T2-2</td>
<td>T367A</td>
<td>(C)och79K</td>
<td>Likely restores gene expression</td>
</tr>
<tr>
<td></td>
<td>G1473A</td>
<td>(F)R157H</td>
<td>( F_1\text{-}D ) contact site [e]</td>
</tr>
<tr>
<td></td>
<td>A1637G</td>
<td>(F)M212V</td>
<td>Suppresses defective ( D ) protein interactions [d,e]</td>
</tr>
<tr>
<td>Isolate T2-3</td>
<td>T367A</td>
<td>(C)och79K</td>
<td>Likely restores gene expression</td>
</tr>
<tr>
<td></td>
<td>C1602T</td>
<td>(F)A200V</td>
<td>( F_4\text{-}D ) contact site [e]</td>
</tr>
<tr>
<td></td>
<td>C2162T</td>
<td>(F)H387Y</td>
<td>( F_4\text{-}D ) contact site [e]</td>
</tr>
<tr>
<td>Isolate T3-2</td>
<td>C367T</td>
<td>(C)Q79och</td>
<td>Decreases expression of all viral genes</td>
</tr>
<tr>
<td></td>
<td>A1637G</td>
<td>(F)M212V</td>
<td>Suppresses defective ( D ) protein interactions [d,e]</td>
</tr>
<tr>
<td>Isolate T4-1</td>
<td>T367A</td>
<td>(C)och79K</td>
<td>Likely restores gene expression</td>
</tr>
<tr>
<td></td>
<td>G1266T</td>
<td>(F)G88V</td>
<td>( F_4\text{-}D ) contact site [e]</td>
</tr>
<tr>
<td>Isolate T4-8</td>
<td>C367T</td>
<td>(C)Q79och</td>
<td>Decreases expression of all viral genes</td>
</tr>
<tr>
<td></td>
<td>G1080A</td>
<td>(F)R26H</td>
<td>Start of ( \beta ) C in coat protein [g]</td>
</tr>
<tr>
<td></td>
<td>C1602T</td>
<td>(F)A200V</td>
<td>( F_4\text{-}D ) contact site [d]</td>
</tr>
<tr>
<td>Isolate T4-10</td>
<td>C367T</td>
<td>(C)Q79och</td>
<td>Decreases expression of all viral genes</td>
</tr>
<tr>
<td></td>
<td>G1266A</td>
<td>(F)G88D</td>
<td>( F_4\text{-}D ) contact site [e]</td>
</tr>
<tr>
<td></td>
<td>G2910A</td>
<td>(G)Q171Q</td>
<td>in ( \chi^2G )</td>
</tr>
<tr>
<td></td>
<td>C2162T</td>
<td>(F)H387Y</td>
<td>( F_4\text{-}D ) contact site [e]</td>
</tr>
<tr>
<td><strong>No complementation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-( \chi^2G )</td>
<td></td>
<td></td>
<td>All mutations from Isolate D2, plus . . .</td>
</tr>
<tr>
<td></td>
<td>G2371C</td>
<td>—</td>
<td>( F-G ) intercistronic region</td>
</tr>
<tr>
<td></td>
<td>G2379A</td>
<td>—</td>
<td>( F-G ) intercistronic region</td>
</tr>
<tr>
<td></td>
<td>G2387T</td>
<td>—</td>
<td>Ribosome binding site for ( \chi^2G )</td>
</tr>
<tr>
<td>D-G4GV</td>
<td></td>
<td></td>
<td>All mutations from Isolate D2, plus the recombination of ( G4\text{-}G ) into the genome</td>
</tr>
<tr>
<td>T-G4GV</td>
<td></td>
<td></td>
<td>All mutations from Isolate T4-10, plus the recombination of ( G4\text{-}G ) into the genome</td>
</tr>
</tbody>
</table>

*a The letter in parenthesis represents the protein affected by the given nucleotide change. Functions of each protein are: C, switch from dsDNA to ssDNA synthesis; D, internal scaffolding protein; F, coat protein; G, major spike protein. — represents an intercistronic region.

[b] This study.

cDokland et al. (1997, 1999).

dUchiyama et al. (2007).

eCherwa et al. (2008); Cherwa and Fane (2009).

The nucleotide sequence for this codon differs in the \( \psi X\text{-}174 \) strain used in these studies (Fane and Hayashi 1991) from the sequence published by Sanger et al. (1978) and has been described in Fane and Hayashi (1991).

gMcKenna et al. (1992, 1994).

[h] Hayashi et al. (1989).
Densitometry, coat protein bands were normalized to the indicated host wild-type or some productive assembly. This presumably replenished H protein levels and permitted thus plaque formation initially required the expression of protein assembly intermediates into insoluble aggregates. Thus, second-site suppressors that re-

Nonproductive, off-pathway reactions predominated stored efficient assembly were easily isolated. In contrast, off-pathway reactions. Thus, second-site suppressors that re-

However, early assembly intermediates remained soluble, sug-

Early assembly defect. The internal scaffolding, external scaf-

Different selections, designed to individually elevate proteins on gene H translation, through the upstream G gene, are particularly pronounced (Tessman et al. 1967; Burgess and Denhardt 1969). As the H protein is already the limiting assembly component, decreased expression of gene G, and by proxy gene H, would dramatically reduce particle morphogenesis.

After the Suppression of Off-Pathway Reactions, Assembly Can Be Optimized and Gene Expression Can Be Elevated

To move forward, och(C)Q79/χ2 G phage was used in five different selections, designed to individually elevate proteins involved in assembly: The internal scaffolding, external scaffolding, DNA pilot, G4 major spike or χ2 major spike proteins. Due to the lack of a cloned complementing coat gene, this condition could not be examined. Only the selection that elevated the level of the G4 G protein was effective, and mutants were obtained at a frequency of 10−8 from the stocks used in these selections. These variants could be complemented by either the foreign or the native spike gene, demonstrating an expanded substrate specificity for productive assembly. Multiple mutations had accumulated, most of which would promote efficient assembly (table 4): This now appeared to be the primary adaptive mechanism, rather than directly suppressing off-pathway reactions. Although it was not possible to determine the temporal order at which mutations arose, all strains contained mutations that likely coinfected cells, it could persist as a defective interfering particle. Defective interfering particles can rise to appreciable levels in high multiplicity of infection (MOI) ϕX174 cultures (Wichman et al. 2005).

Although the second mechanism was more complex, it also involved lowering the intracellular levels of the inhibitory protein. This was accomplished through a premature stop codon in gene C, och(C)Q79. The C protein inhibits RF DNA synthesis (Fujisawa and Hayashi 1977; Doore et al. 2014). The truncated C protein displayed increased activity, which resulted in dramatic reductions of intracellular RF DNA molecules and transcripts. In addition to reduced expression, the foreign gene recombined with the cloned ϕX174 G gene, exchanging the 3′ region of the foreign gene for ϕX174 sequence (ϕX-G4G/χ2). This would directly affect the coat spike protein interface (McKenna et al. 1992, 1994, 1996). Mutations in this region also occurred in the G4-ϕXG chimera, albeit after the initial suppression of the ts phenotype.

Genes expressed at low levels are more likely to be successfully incorporated into foreign backgrounds (Park and Zhang 2012; Yang et al. 2012). Preserving the balance of interacting components is a well-documented adaptation (Floor 1970; Sternberg 1976; Cherwa, Young, et al. 2011). Thus, it may have been more advantageous to decrease the expression of the entire genome, not just the transfected gene. Moreover, weak transcription terminators, translational coupling, and spatially competing ribosome binding sites make it difficult to alter the expression of one gene without altering the expression of others (Hayashi et al. 1988; Burch and Fane 2000; Brown et al. 2010, 2013; Doore et al. 2014). Polar effects on gene H translation, through the upstream G gene, are particularly pronounced (Tessman et al. 1967; Burgess and Denhardt 1969). As the H protein is already the limiting assembly component, decreased expression of gene G, and by proxy gene H, would dramatically reduce particle morphogenesis.

Off-Pathways Reactions Introduced by HGT May Determine the Trajectory of Evolutionary Recovery

After movement of the ϕX174 major spike gene into G4, the resulting chimera displayed a ts early assembly defect. However, early assembly intermediates remained soluble, suggesting that the newly introduced G protein did not promote off-pathway reactions. Thus, second-site suppressors that restored efficient assembly were easily isolated. In contrast, nonproductive, off-pathway reactions predominated ϕX174-G4G assembly immediately after HGT, siphoning H protein assembly intermediates into insoluble aggregates. Thus, plaque formation initially required the expression of the native ϕX174 G gene as well as the cloned H gene, which presumably replenished H protein levels and permitted some productive assembly.

The Suppression of Off-Pathway Reactions as the First Step in the Recovery of ϕX174-G4G

Before productive assembly could be optimized, off-pathways reactions were suppressed by one of two observed mechanisms. The first mechanism involved a deletion of the foreign gene (ϕXΔG): A dramatic yet highly effective way to eliminate the protein promoting the off-pathway reactions. However, unless a supply of ϕX174 major spike G protein is guaranteed, the mutant is an evolutionary dead-end. In

Table 5. Fitness Values of Wild-Type and Viable Chimeras at 33°C.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Fitness</th>
</tr>
</thead>
<tbody>
<tr>
<td>ϕX174 wild-type</td>
<td>14.5 ± 0.8</td>
</tr>
<tr>
<td>G4 wild-type</td>
<td>15.3 ± 0.5</td>
</tr>
<tr>
<td>D-G2V</td>
<td>6.7 ± 0.6</td>
</tr>
<tr>
<td>D-G4GV</td>
<td>6.9 ± 0.5</td>
</tr>
<tr>
<td>T-G4GV</td>
<td>6.2 ± 0.9</td>
</tr>
</tbody>
</table>

Note.—Values are reported as doublings/hour with standard error, n = 3.
restore gene expression. These included the loss of the premature stop codon in gene C and mutations in the J-F intercistronic region, which was characterized in greater detail (table 4 and fig. 6).

Once ϕX-G4G/χ2 was able to utilize the foreign major spike G protein, two events led to a complementation-independent (viable) phenotype. In the first, a recombination event occurred between the viral genome and the plasmid containing the G4 major spike gene. This reintroduced the entirety of the foreign gene into the genome and removed the χ2 gene from the system. Identical recombinants were observed in two independent lineages. Plaque formation did not require any additional mutations even though the total level of intracellular G protein would be lower than in the preceding evolutionary stage. In the preceding stage, G protein was produced from the genome and the plasmid. However, two species of G protein were present: G4 and χ2, which could lead to the production of suboptimal spike protein pentamers. In the second event, the χ2 gene was retained, but with additional mutations. Throughout these studies, the χ2 protein appeared to be a less-fit hybrid. Presumably, the additional mutations, which are located in the directly upstream region, increase the concentration of the protein to drive assembly reactions forward.

The Emergence of New Viruses

The ability for these chimeras to persist in nature is difficult to predict, as their success would depend on environmental conditions. For example, the G4-ϕXG could have a greater chance of emerging in a low MOI environment, where it could recover fitness through single mutations without as much competition. Conversely, the ϕX-G4G chimera may require a unique set of conditions, such as sustained coinfections with wild-type phages. These could supply the wild-type protein and/or assembled procapsids, into which the chimeric genome could be packaged. However, specific conditions that would allow this chimera to persist in nature long enough to emerge are difficult to ascertain.

New viruses frequently arise through recombination events (Koonin et al. 2006; Holmes and Drummond 2007; Martin et al. 2011; Diemer and Stedman 2012; Krupovic 2012; Roux et al. 2013), but only those that survive to be sampled can be appreciated. Occasionally a very unusual recombiant emerges. For example, a plant virus and a vertebrate virus recombined to produce a circovirus (Gibbs and Weil 1999), and recombination events between DNA and RNA viruses have been documented in extemophiles (Diemer and Stedman 2012). Between the original recombination event and the ultimate emergence of a new virus, multiple evolutionary mechanisms may be required. The results of this study may represent some of these evolutionary mechanisms.

Materials and Methods

Phage Plating, Media, Buffers, and Stock Preparation

Media, buffers, plating, and stock preparation have been previously described (Fane and Hayashi 1991).

Bacterial Strains and Plasmids

The Escherichia coli C strains BTCC 122 (Su—), BAF5 (SupE), and BAF8 (SupF) have been previously described (Fane and Hayashi 1991). BAF30 is a recA derivative of C122 (Fane et al. 1992). The YR7211 cell line contains a mutation in the mraY gene rendering it resistant to E protein-mediated lysis (Bernhardt et al. 2000, 2001). Clones of the viral genes ϕX174 B (Novak and Fane 2004), D (Cherwa and Fane 2009), and H (Cherwa, Young, et al. 2011) have been previously described.

The ϕX174 and G4 G genes were cloned by PCR amplification with primers that introduced an Ncol site before the gene’s ribosome binding site and a downstream HindIII site (ϕX174) or a BglII site (G4) after the stop codon. The amplified fragments were digested with Ncol and HindIII or BglII, then ligated into pSE420 (Invitrogen) DNA digested with the same enzymes. Gene expression is under lac promoter control. The cocloning of the ϕX174 G and H genes was performed in two steps. After gene G was cloned into pSE420, gene H was cloned into the same plasmid. The gene was amplified with primers that spanned a naturally occurring SacII site near the 3’-end of gene G and a downstream primer that introduced a HindIII site after the gene H stop codon. The amplified fragment was digested with SacII and HindIII, then ligated into the ϕX174 G plasmid digested with the same enzymes.

Generation of ssDNA, RF DNA, DNA Isolation, Rate Zonal Sedimentation, and Protein Electrophoresis

The protocols for ssDNA (Fane and Hayashi 1991) and dsDNA RF DNA (Burch et al. 1999) isolation and purification, rate zonal sedimentation, and protein electrophoresis protocols (Uchiyama and Fane 2005) have been previously described.

Construction of G4-ϕXG and ϕX-G4G Strains

Oligonucleotide-mediated mutagenesis was conducted as previously described (Fane et al. 1993), but in multiple rounds. In the first round of mutagenesis, a PciI restriction site was designed to span the gene G start codon. An amber mutation at codon 3 was also inserted for screening purposes. Mutagenized DNA was transfected into the amber-suppressing host BAF5 (SupE). Progeny were screened for amber phenotypes by stabbing phage into Su+ and BAF5 (SupE) lawns. The amber mutation was subsequently eliminated by selecting for am+ revertants. All genotypes were verified by direct DNA sequence analyses. After purification, ssDNA was purified for the second round of mutagenesis. These mutagenic primers were designed to place an Nhel restriction site in G4 (G4-ϕXG) or an Ncol restriction site in ϕX174 (ϕX-G4G) at the 5’-end of gene H along with an amber mutation in either codon 11 or codon 3, respectively. Protocols, selections, and reversions were performed as described above, but with transfection and screening on BAF8 (SupF) or BAF30 ϕXH. Corresponding restriction sites were introduced to the foreign gene G through oligonucleotides used in the PCR reactions. RF DNA and PCR products were cut with PciI and either Nhel
or Ncol. Following ligation, DNA was transfected into BAF30 pG4G for G4-ϕXG or BAF30 ϕXG for ϕXG-G4G (see Results). An alternate cloning scheme was also used for generating the ϕX-G4G strain. An AleI site naturally occurs at the 3′-end of gene G in ϕX174. RF DNA from ϕX174(G)Pci was digested with PciI and AleI. Corresponding sites were introduced into the G4 G gene through PCR primers. After ligation, DNA was transfected into BAF30 pϕXG.

Isolation of Second-Site Suppressors and Viable Chimeric Strains

To isolate second-site suppressors of the ts phenotype, approximately 10⁶ pfu of G4-ϕXG was plated at a restricted temperature, 35 °C. Second-site suppressors were identified by comparing mutant and parental genome sequences. To isolate ϕX-G4G mutants that were viable on restrictive cell lines, approximately 10⁻¹⁰⁸ pfu of the various ϕX-G4G strains was plated on ϕXG, pG4G, or C122 at 28, 33, and 37 °C. To isolate χG/G(och/C)Q79 mutants that could grow on restrictive cell lines, approximately 10⁻⁷–10⁻⁸ pfu was plated on C122 or cell lines containing either a cloned ϕX174 B, ϕX174 D, ϕX174 H, G4 G, or χG G gene. Putative mutants, based on reduced plaque morphology, were further screened for weak growth, ts or complementation-dependent phenotypes. When no genes were exogenously expressed in the selection, plaques were tooth-picked into indicator lawns incubated at 28, 37, and 42 °C. When a gene was exogenously expressed, plaques were assayed for a complementation-dependent phenotype.

Attachment, Eclipse, and Fitness Assays

Fitness assays (Bull et al. 1997), attachment assays (Hafenstein et al. 2004), and eclipse assays (Cherwa et al. 2009) were performed as previously described. Eclipse and attachment assays were performed in lysis resistant hosts (Bernhardt et al. 2000, 2001).

In Vivo Characterization of Assembly Intermediates and Whole Cell Lysates

Protocols for generating infected cell extracts, ultracentrifugation parameters, particle detection, and in vivo kinetics have been previously described (Uchiyama and Fane 2005; Uchiyama et al. 2007).

Isolation of RNA, cDNA Synthesis, and qPCR

For each infection, 100 ml of BAF30 cells harboring an empty pSE420 vector were grown to 1 x 10⁸ cells/ml. Cells were centrifuged and washed twice with chilled HFB buffer (Fane and Hayashi 1991), then resuspended in 1:100 volume of chilled HFB (0.06M NH₄Cl, 0.09M NaCl, 0.1M KCl, 0.1M Tris-HCl (pH 7.4), 1.0 mM MgSO₄, 1.0 mM CaCl₂) buffer with 10 mM MgCl₂ and 5.0 mM CaCl₂. Phage was added at an MOI of 0.1 and allowed to attach for 30 min at 16 °C. Samples were centrifuged at 13,000 rpm for 10 min to remove unattached phage, and pellets were resuspended in ice-cold TK (1.0% tryptone, 0.5% KCl) media (Fane and Hayashi 1991) with 10 mM MgCl₂ and 5.0 mM CaCl₂. This resuspension was then added to 9.0 ml prewarmed (33 °C) TK media with 10 mM MgCl₂ and 5.0 mM CaCl₂. At 4 and 8 min postinfection, 0.5 ml of sample was removed and added to 1.0 ml RNAProtect Bacterial Reagent (Qiagen) per manufacturer’s instructions. Total RNA was extracted using NucleoSpin RNA kit (Macherey-Nagel) per manufacturer’s instructions, but with the DNase incubation extended to 1 h. RNA was quantified, and quality was determined by 260/280 and 260/230 ratios. cDNA synthesis was performed using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems). Negative controls included reactions without RNA template or without reverse transcriptase.

qPCR was performed on an ABI 7300 Real-Time PCR Sequence Detection System, using Brilliant II SYBR Green with High Rox (Agilent) as a detector dye. Thermal cycling conditions started with a 50 °C hold for 2 min and a 95 °C hold for 3 min to activate the DNA polymerase. Thermal cycling conditions followed with a 50 °C hold for 2 min and a 95 °C hold for 3 min to activate the DNA polymerase, followed by 40 cycles of 95 °C for 15 s and 60 °C for 1 min. Dissociation curves were also performed for each sample to ensure no mispriming occurred. Relative quantification was performed using the ΔΔC_T method, with the ampR gene from pSE420 used as the reference gene.

Acknowledgments

This research was supported by National Science Foundation grants MCB 0948399 and 1408217 to B.A.F. The authors thank Dr Celeste J. Brown for assistance in designing qPCR experiments, the University of Arizona Statistics GIDP for assistance in statistical analysis, and Kai Doore for proofreading.

References


Krupovic M. 2012. Recombination between RNA viruses and plasmids might have played a central role in the origin and evolution of small DNA viruses. Bioessays 34:867–870.


