LIFE IN OUR PHAGE WORLD

A CENTENNIAL FIELD GUIDE TO THE EARTH'S MOST DIVERSE INHABITANTS

FOREST ROHWER
MERRY YOULE
HEATHER MAUGHAN
NAO HISAKAWA

ILLUSTRATIONS BY
LEAH L PANTÉA
BEN DARBY

Life in Our Phage World

*A centennial field guide to the Earth's
most diverse inhabitants*

by

Forest Rohwer

Merry Youle

Heather Maughan

Nao Hisakawa

Illustrations by Leah L Pantéa and Benjamin Darby

Table of Contents

Introduction

Two in the morning in a roadside hotel in the middle of California's Eastern Sierra Mountains. Over the last 48 hours, Mya Breitbart, Tom Schoenfeld, and I had driven over a thousand miles, then carried several hundred pounds of filters, pumps, car batteries, and water up and down steep slopes in the 95° F plus temperatures, all so that we could sit next to much hotter springs for several hours watching the pumps run. Now Tom has fallen asleep with Cheaters playing on the TV and Mya is in the bathroom, finishing the filtering for the day. When she is almost done, I jokingly say, "Just one more thing." She throws a pipetter at me and collapses on the floor. Tom doesn't stir. We'll grab some sleep and then get up at 5 am, drink a lot of coffee, and head back out to hunt the most voracious predators on the planet.

Much of biology is about feeding the phages[1]. By killing nonillions of Bacteria, they have major effects on global energy and nutrient cycles. Phages are the friend of the underdog. When a bacterial strain prospers and threatens to take over the local community, their phages feast and decimate that strain, thereby successfully maintaining microbial diversity in the face of a winner-take-all threat. This behavior can be a nuisance. When we populate a million dollar lysine fermenter with our bacterial workers of choice, one phage invader can multiply and crash the worker population in a couple of hours. But the phages must be forgiven for such pranks as so many of the major breakthroughs in biology over the past century emerged from the study of phage. Trace most any aspect of molecular biology back to its roots, and there you'll find a phage. Phages were there early on to provide experimental proof that nucleic acid, not protein, was the genetic material and to assist in the recognition of the triplet genetic code. Later they were used to uncover mechanisms of gene regulation, protein binding to DNA, protein folding, assembly of macromolecular structures, and genetic recombination. They have demonstrated evolution by flagrant horizontal gene transfer and provided proof that mutations arise independent of—not as a response to—the pressure of natural selection.
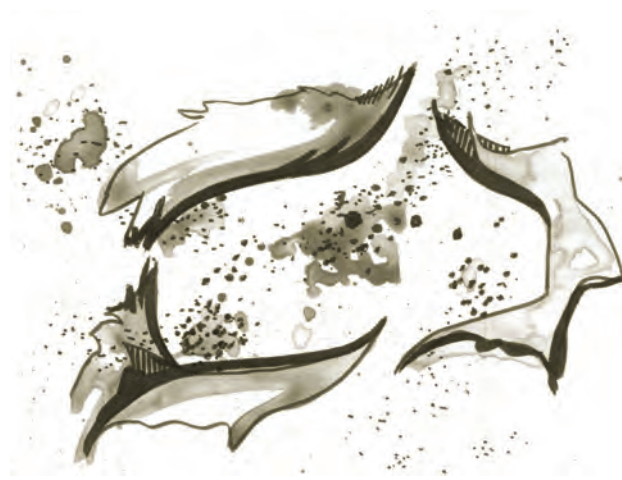
Enzymes from phage launched the molecular biology revolution and remain essential tools for genetic engineering. Phage genomes were the test subjects used for the first genomic and shotgun genomic/metagenomic sequencing, the first fully synthetic life forms, etc. Phage

---

[1] We use the term 'phage' sensu lato to encompass all microbial viruses, i.e., the bacteriophages (Bacteria-eaters), viruses of the Archaea, and viruses of single-celled eukaryotes.

biologists were at the forefront of advances in cancer biology. Most of the stuff of life itself—the global pool of genetic diversity—is encoded by phage. Closer to our individual homes, in the last ten years we finally came to realize that of all the varied genes we carry in our own bodies, the majority reside within our phages. Phages are essential bionts within the human—and every other—holobiont.

Despite their paramount importance to human health, to science, and to all life on the planet, the phage field remains a niche area of study. One reason that phages (as well as most viruses that don't make us or our domesticates sick) remain overlooked is that you can't just go out or look inside and observe them. When outside a host cell, they travel as virions so small that seeing them requires an electron microscope or other sophisticated and costly equipment. Most can't be cultured and interrogated in the lab because their hosts are not known or not yet culturable. Community metagenomics, likewise, is still relatively difficult and costly. This inability to 'see' phages leads to a disconnect between them and all other life forms. Most scientists and others just don't think of them as alive. So this major component of life is reduced to its inert intercellular transport form that is then subjected to biochemical analysis and described in lifeless terms, leaving us blind to their nature as active agents. This is somewhat of a travesty, as these bits of biochemistry are the most successful predators on the planet. They are promiscuous and engage in kinky sex games (e.g., homologous and illegitimate recombination with related and completely alien genomes, orgies of hundreds of genomes). Humans observing the virions perceive them to be inert. But these 'inert' particles, given contact with a potential host, reveal their true nature as complicated nanomachines primed for action. Their performance is precise; milliseconds or nanometers mean the difference between life and death.
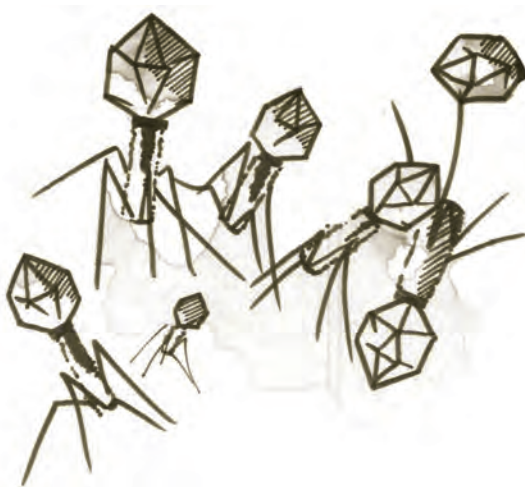
It is not possible to understand the biological world without 'seeing' the phages. This book provides a glimpse of the rich and diverse phage life that has been sampled over the past one hundred years. The overall organization of the book parallels the phage life cycle. It arbitrarily starts with their virions on the prowl, observes them as their genome enters and takes over

a host cell, describes their replication, then applauds as the progeny virions assemble and make their escape into the world. For each stage we chose a few diverse phages to feature.

Field guide pages provide basic information for each of these phages, the kinds of information a naturalist would have at hand for any life form they wanted to study. For each we also relate a lively, thoroughly researched story revealing some of this phage's secrets for success. Terms in boldface within the stories are defined in our glossary. Each story plays out visually in an illustration by San Diego fine artist Leah L. Pantéa. These illustrations are rich in detailed information intended to complement your reading of the text. The 30 featured phages were selected to illustrate the great diversity that exists in even the small fraction of the phage world that has been characterized. Although you will find well-studied phages such as λ, T4, and T7 in these pages, we made no attempt to include the wealth of information available for them; there are many good books that already do this. Each chapter ends with one or two longer, personalized perspectives. Each informs about a particular aspect of The Big Picture and relates part of the recent history of phage research. What makes them so delightful to read is that each is infused with the excitement and humor that has characterized phage research and phage researchers.
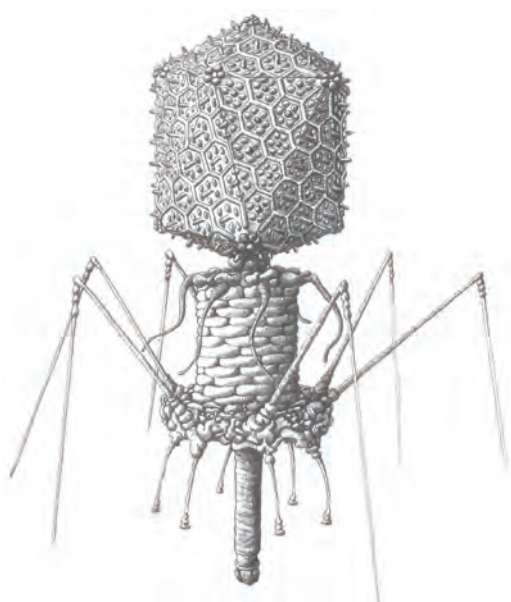
Since we envisioned an Audubon-like field guide to the phages of the world, the portraits of the 30 phages were rendered in pen and ink by Benjamin Darby, an imaginative San Diego artist. As typical of a field guide, he emphasized important or identifying characters of each specimen and added a touch of elegance. When no photo or virion structure was available for that particular phage, we turned to its close relatives for a stand-in. Such a field guide would also group the objects of study into related groups. This is not so easy to do for the phages. Observable virion morphology is not an adequate basis for such classification as great phage diversity lurks within each virion type.

The recent accumulation of genome data provides another handle on phage taxonomy, but application of this approach remains challenging. The now familiar Tree of Life portrays the evolutionary relationships among all members of the three domains based on the rRNA genes that they all carry. A similar tree could in theory be constructed for the phages if any single gene were present in all phage genomes, but there is no such gene, thus there can be no such tree. At most, such 'signature genes' can serve to elucidate relationships within closely-related groups. Moreover, the evolution of viruses has not followed the same strict

pattern of vertical descent from a common ancestor as predominates in many cellular organisms. Phages may not all share a common ancestor, and moreover they have exchanged genes horizontally. This argues for a different approach to their classification. To that end, we compared the genomes of 1220 phages and built a taxonomic tree based on their similarities (*see page 8-8*). For each featured phage in the field guide, we show its relationship to all the other 1219 phages on that tree and also zoom in on its local tree neighborhood.
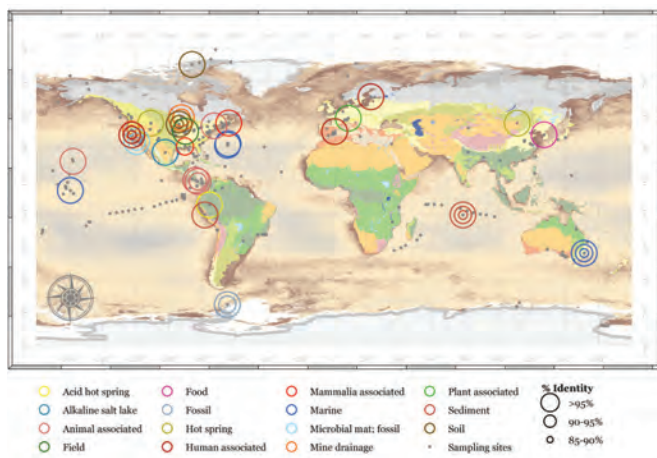
In the tradition of other field guides, we have included a global map showing the known geographic range of each featured phage as well as the habitats where it has been found so far. These 'sightings' (*see page 8-20*) are based on BLAST hits between that phage genome and publicly-available metagenomes from around the world. For a guide to interpreting these maps, see Appendix A4 (*page 8-20*). While the data displayed here is interesting, more important is what is missing. Most of the globe and many ecosystems have not been sampled nor have their phage communities been characterized. Microbes have been found everywhere people have looked on Earth—on the land, in the sea, in the air, inside rocks and inside host cells—even under extreme conditions previously thought to be unable to support life. Wherever there are microbes, there are phages. For phage explorers, most of the Earth remains a terra incognita. It is time to get to work and put phages on the map.

To portray phage genomes as the lively, evolving molecules that they are, we present two versions of each phage's genome. First, an artist-created overview shows the variety of genome structures used by our featured
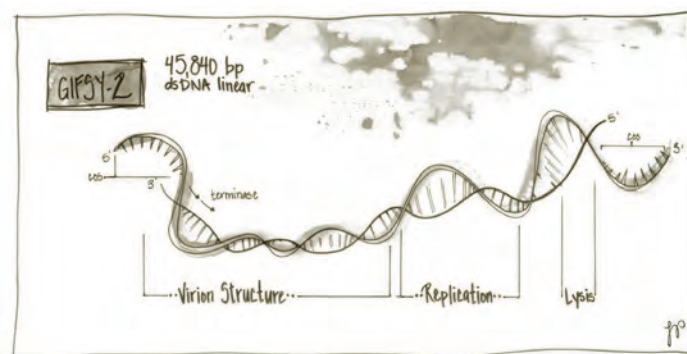
phages when traveling by virion between hosts (e.g., linear or circular, single-stranded or double-stranded, sticky ends, direct or inverted terminal repeats). Here we have also delineated functional modules and highlighted landmarks that are featured in the stories or are well appreciated among phageophiles. Each overview is followed by a detailed genome map that allows for admiration of each gene including information (if available) about its function, its homology with other phage genes, and/or (if applicable) the localization of its protein product in the virocell. Genes are counted as open reading frames (ORFs) if they encode a protein and as RNAs if their transcripts are not translated (e.g., tRNAs).

To emphasize the dynamic nature of phage in the writings, we have developed a lexicon based on ethology (*see page 8-27*) and used its terms in our writing. The goal of this writing style is to bring each phage to life, without seriously compromising scientific accuracy. It is also to remind us that there are many phage behaviors that we expect to observe, but haven't studied yet. In some cases, we can link a particular behavior to one or more genes, but the genetic basis for many remains to be discovered. No doubt clues are hiding in the ~80% of phage genes that are completely novel.

The first 100 years of phage research have fundamentally changed our lives and our understanding of the natural world. In the near future we expect to see a new synthesis in biology that puts phage at the center of the field, no longer to languish in a dimly lit corner as a biological novelty, an after-thought. But that will occur only when many people, such as yourself, include the phage in your research, in your study, in your teaching, and in your understanding of life on Earth. The second century of phage study is beginning. Be there.

Fast-forward to a decade after the Sierra Phage Hunting expedition and I'm walking around in a Wisconsin winter in shorts; –20° F is not a great place to make a San Diego fashion statement. My latest phage hunting had taken me to the Arctic and I am still waiting for my winter clothing to be shipped back from Russia. Mya has gone on to become a leader in the field of phage ecology, despite a history of throwing things at her PhD advisor. I am crunching through the snow with Tom and his ever-excitable business partner David Mead. Together they had built Lucigen into a leading company in the realm of enzymes and cloning, now expanding into diagnostics. Many of their products are based on enzymes originally found in phages isolated from hot springs, enzymes such as DNA polymerases that are also primases and reverse transcriptases, incredibly efficient ligases, and many others. But neither one of them is talking about their business successes. They are both happily planning yet another sampling trip to find yet more weird and wonderful phages. They know the phages are out there, waiting for someone to notice.

Chapter 2: On the Prowl

Enterobacteria Phage RB49

a Myophage that senses environmental cues to decide when to extend its long tail fibers

Genome
dsDNA; linear
164,018 bp
279 predicted ORFs; 0 RNAs

Encapsidation method
Packaging; T = 13 capsid

Common host
Escherichia coli

Habitat
Mammalian intestines

Lifestyle
Lytic

Ben Darby

ICTV Family
Myoviridae

% Identity

⭕ Animal associated   ⭕ Feces   ⭕ Marine

• Sampling sites

◯ >95%

◯ 90-95%

◦ 85-90%

For map details,
see Appendix A4.

RB49   164,018 bp
ds DNA linear

dtr   3'
5'

wac

dtr
5'
3'   ⌐..Replication..⌐   ⌐..Virion Structure..⌐   ⌐..Replication..⌐

Enterobacteria Phage RB49

2,000   4,000   6,000   8,000   10,000   12,000   14,000

rIIA, membrane (POG0387)   p5   p7 (POG0390)   dda, helicase (POG0377)   61, primase (POG0084)   41, helicase (POG0025)
39, topoisomerase (POG0382)   p9   srd, anti-sigma (POG2446)   p17   uvsX, recombination (POG0097)
p2   dexA, exonuclease (POG0982)   56, dCTP pyrophosphatase (POG0397)   p18   dc1   dc2
p3 (POG3796)   p6.1   modB, ADP ribo. (POGs 0393,2445)   p19 (POG0398)   p23, head vertex (POG0402)
p4 (POG3797)   p10   p12 (POG2190)

16,000   18,000   20,000   22,000   24,000   26,000   28,000   30,000

43, DNA polymerase (POG0359)   45, sliding clamp (POG0409)   46, endonuclease (POG0083)   p36   a-gt.4 (POG0710)   p44 (POGs 1685,1686)
uvsX, recombination (POG0097)   rpbA, RNAP-binding (POG0410)   47, endonuclease (POG0082)   p46 (POG3798)
p25   44, clamp loader SSU (POG0996)   p31   40, endonuclease (POG0184)   p45 (POG1687)
regA, tl. repressor (POG1681)   p033 (POG1683)   p35   p38   55, sigma factor (POG1006)
62, clamp loader SSU (POG0408)   p37   a-gt.5 (POG1684)

32,000   34,000   36,000   38,000   40,000   42,000   44,000

p48   p49   p51   nrdD, NTP reductase (POG0420)   p57   frd.2   p59   p65   p68   p70   p72   p76
nrdG, NTP reductase (POG0419)   p54   p55   dam, methylase (POG0240)   p67   p69   p71   p74   p78
p46 (POG3798)   p50   49, recomb. endonuclease (POG1000)   p64   p66   p73   p77
nrdH, glutaredoxin (POG0161)   nrdC, thioredoxin (POG0161)   p66   p75
p53 (POG1376)   p60

46,000   48,000   50,000   52,000   54,000   56,000   58,000   60,000

p79, LTF assembly (POG2450)   tk.4 (POG1382)   nrdC.11 (POGs 0431,1177)   p104   p109   p113   p116   p120
p78   tk, thymidine kinase (POG1030)   vs.1 (POG1454)   p95   vs.8 (POG2931)   p103   p107   p110   p115   p118   p122
p80   rI.-1 (POG0375)   regB, RNase (POG0436)   vs.6 (POG0440)   p106   dmd, mRNA degradation (POG2441)
p82   vs, tRNA synthetase mod. (POG2422)   p102, endolysin (POG0055)   p105   p108   p114 (POG1983)   p117   p119
p84 (POG3800)   p89 (POG0435)   p96   p112   p121
p83 (POG3799)   hypothetical protein   p97   p111
rI, lysis inhibition reg. (POG0433)   p100 (POG1480)
p81

p141
p142
53, baseplate wedge (POG0452)   p136 (POG0238)   8, baseplate wedge (POG0455)
5.1 (POG0374)   6, baseplate wedge (POG0229)   p143, methyltransferase (POG3801)
p130   5, BP hub & tail lysozyme (POG0453)   7, baseplate wedge (POG0454)   p140

62,000   64,000   66,000   68,000   70,000   72,000   74,000

p125   2, DNA end guard (POGs 1802,1010)
1, dNMP kinase (POG0020)
3, tail & sheath (POG1038)
57B (POG0450)   4, head completion (POG0451)
p127, TF chaperone (POG1799)
dmd, mRNA degradation (POG2441)
p124

9, BP wedge/TF connector (POG1804)   16, terminase SSU (POG1039)
11, BP wedge & tail pin (POG0457)   14, neck (POG0461)   20, portal vertex (POG1001)
p143, methyltransferase (POG3801)   p147, short tail fibers (POG0458)   13, neck (POG0460)   17, terminase LSU (POG0252)   19, tail tube (POG1005)
10, BP wedge & tail pin (POG0456)   wac, fibritin (POG0459)   15, tail (POG0462)   18, tail sheath (POG0356)

76,000   78,000   80,000   82,000   84,000   86,000   88,000   90,000

On the Prowl   Entry   Takeover   Replication   Replication   Assembly   Escape   Transfer   Non-coding   Multiple   Unknown
Lytic   Lysogenic   RNA   RNA

Testing the Waters

Merry Youle

Mise-en-Scène: *All life forms sense key environmental cues and then respond appropriately. Phages are no exception. They keep tabs on the external environment, some then choosing to promote or delay adsorption depending on conditions. For these phages, to extend their tail fibers or not to extend: that is the question. Temperate phages are even more sophisticated, weighing factors such as host physiology and abundance as they make their lysis/lysogeny decision.*

The iconic image of a phage seen on T-shirts and coffee mugs is that of a T4 **virion** tumbling through the milieu, its six 'claws' outstretched, poised for a deadly encounter with a hapless *E. coli*. However, such images can be misleading. Consider a more restrained possibility: a phage holding most of its tail fibers close to its tail or head, gingerly extending just one at a time to test the waters. This demure strategy offers some advantages. When extended, the tail fibers are more susceptible to damage (Kellenberger et al., 1965) and they slow virion diffusion. More importantly, there is no need for all six to be deployed to search for prey, as one extended fiber surveys almost as large a volume as does six. So which strategy do the phages choose: travel with all tail fibers extended, a few, or none?

While on the prowl, Podophage T7 extends individual tail fibers sequentially, just one at a time, to scout for prey (Hu et al., 2013). When it contacts a potential host, it walks along the cell surface like a six-legged dancer lightly balancing on only one leg at a time (http://www.youtube.com/watch?v=Gy42CoyqKjE). Each fiber in turn binds reversibly, and only weakly, but even weak interactions can provide enough 'gravity' to keep the phage exploring the surface rather than drifting away. This approach decreases the search space from three dimensions to two. When by chance a tail fiber encounters T7's speci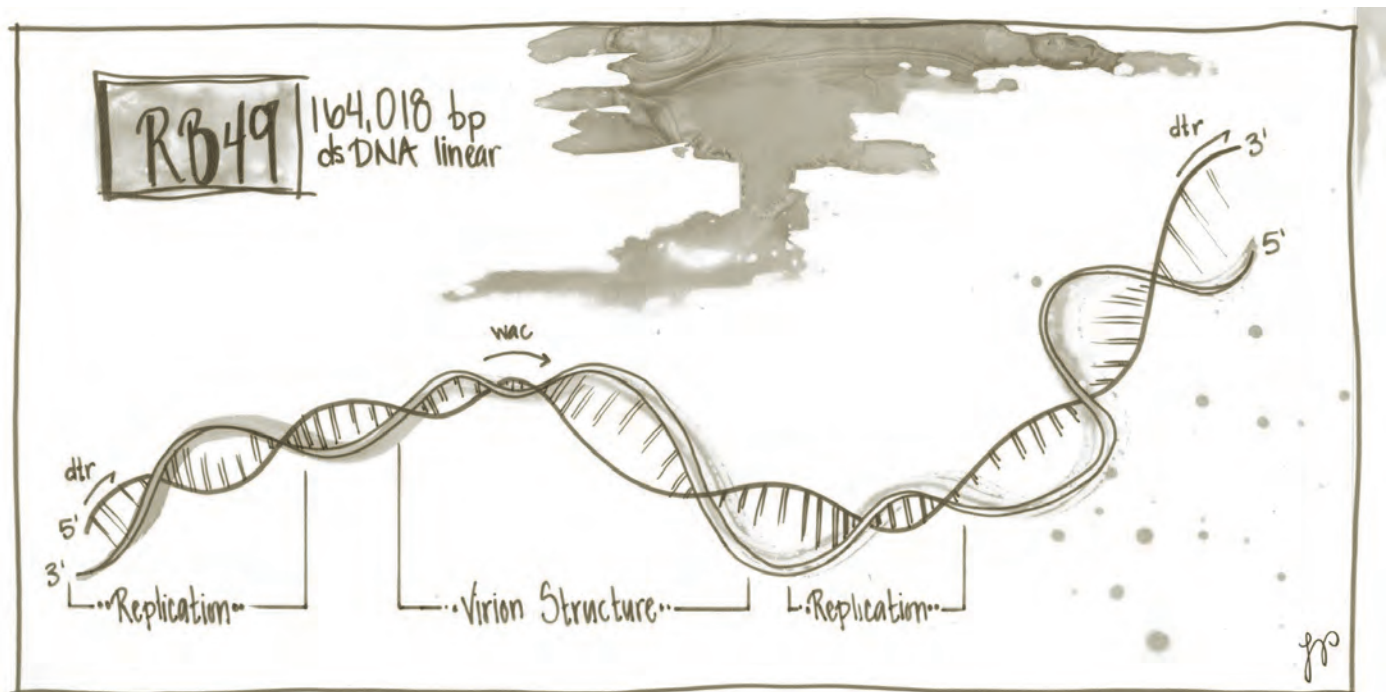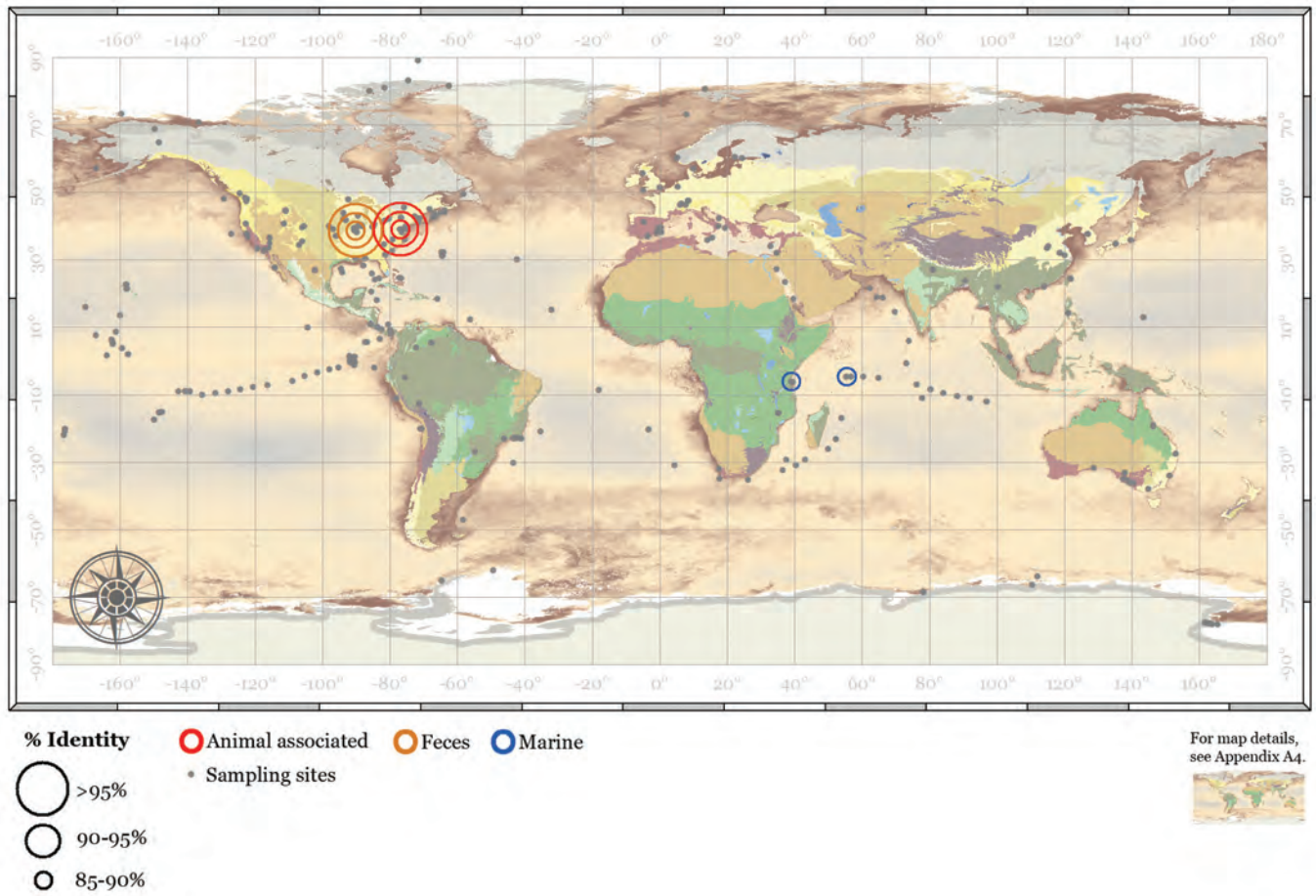fic **receptor**, walking comes to a halt. Now all six tail fibers bind and soon an infection is underway. Is T7 the exception or the norm?

Whiskers

Consider the T4-like phages, phages such as RB49. Being Myophages, their situation is a bit different. Their tail is a complex macromolecular machine, typically about 144 nm long and composed of at least 430 polypeptide chains. Each tail bears three sets of fibrous structures: six long tail fibers (LTFs) essential for host recognition and the initial reversible **adsorption**; six short tail fibers (STFs) required for irreversible adsorption; and six whiskers. As their name suggests, the whiskers are located at the phage 'neck' and are short, only 53 nm long. They don't interact directly with the host surface, but nevertheless they play a key role when on the prowl for a host.

These whiskers are stiff bristles, each one built from three parallel molecules of the Wac (<u>w</u>hisker <u>a</u>ntigen <u>c</u>ontrol) protein (Efimov et al., 1994). Although simpler than the LTFs, they nevertheless comprise three distinct regions. The middle 80% of the protein chain is a coiled coil α-helical structure that constitutes most of the length of the bristle (Letarov et al., 2005). The C-terminal domain at the distal end of each chain serves as a **foldon** that ensures correct folding and trimerization. The N-terminal domains of all the whiskers form a wheel-like collar around the neck, with the domains of adjacent whiskers linked by one copy of an unidentified protein (Kostyuchenko et al., 2005). This arrangement spaces the whiskers evenly and anchors them to the capsid.

The whiskers are put to work right away to assist with the last assembly step: the attachment of the LTFs. This maneuver is a bit of a trick. Try to picture docking one end of a ~144 nm long LTF to the baseplate of a preassembled virion within the crowded cytoplasm of the host cell. These phages align each LTF for attachment by using both ends of a whisker to 'grasp' it at specific locations (Kostyuchenko et al., 2005).
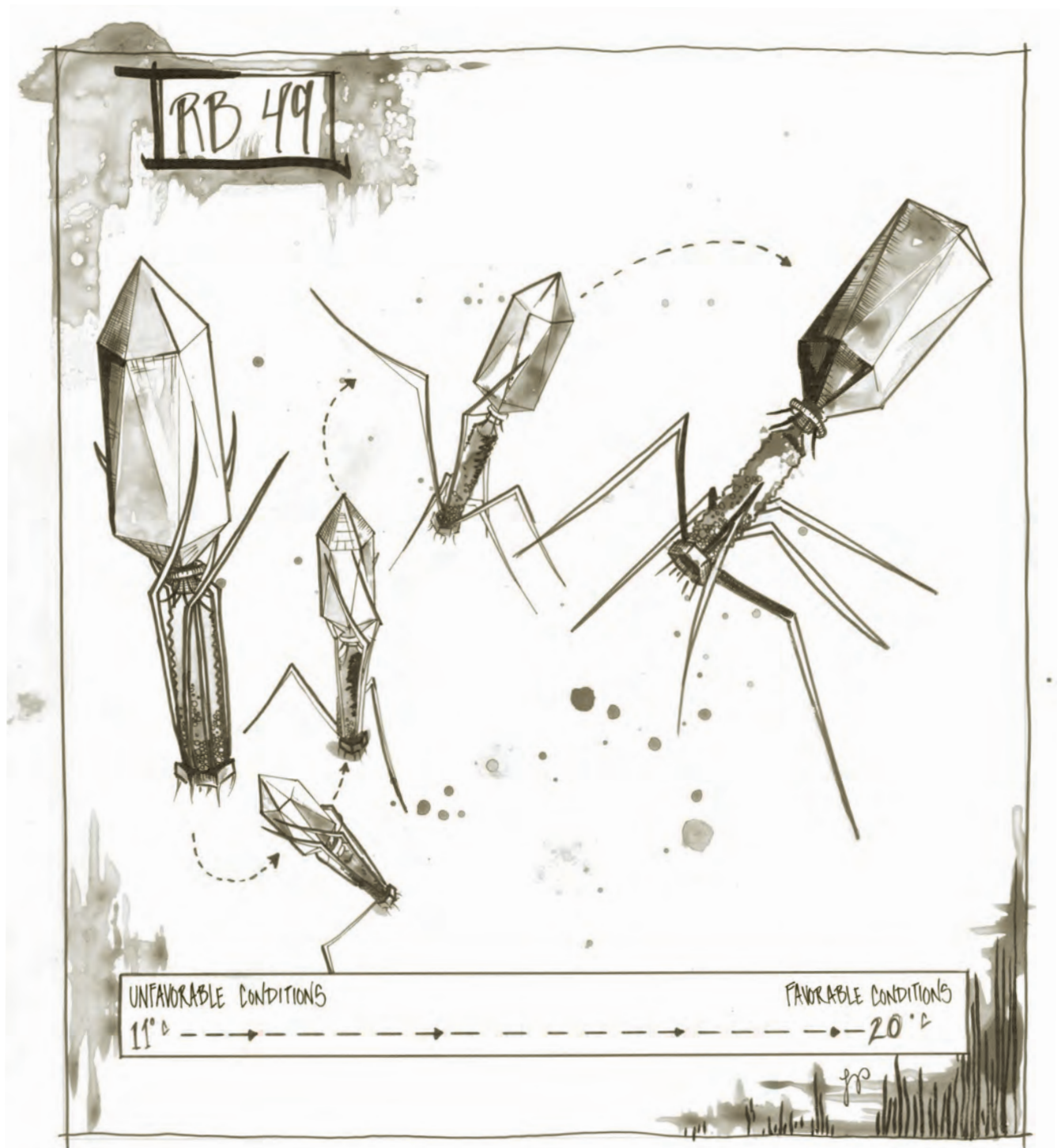
Environmental sensing

After host lysis, freshly minted RB49 virions set off into the world to repeat the cycle of carnage. Should we imagine them adrift in search of new hosts with all six tail fibers displayed? Or, T7-like, with most LTFs held close? Because it has whiskers, RB49 can choose. It adaptively retracts or extends its LTFs depending upon the environmental conditions it encounters. If the phage judges the environment to be adverse, its whiskers hold the LTFs in the retracted position where they form a 'jacket' around the tail sheath, slightly overlapping the head. Such introverted virions are not infective (Kellenberger et al., 1965). This also shelters the LTFs from damage.

What environmental conditions do these phages monitor? For one, they perform a litmus test, retracting their LTFs when the pH drops to 5 or below (Kellenberger et al., 1965). Likewise they consider 0.10 M salt hospitable, but retract their LTFs if the salt concentration decreases to 0.01 M (Conley and Wood, 1975). If the temperature drops from 20° C to 11° C, they respond by retracting. These are reversible responses, not permanent inactivation. When favorable conditions return, they unfurl their LTFs and infectivity is restored.

Some T4 strains have a more refined mechanism that tests the environment for a specific compound required for infectivity. These phages keep their LTFs retracted by binding them to their tail sheath until they 'sense' the presence of the cofactor (Brenner et al., 1962). For one such phage (T4B), a single molecule of tryptophan per LTF is sufficient to disrupt this binding and allow LTF extension (Kellenberger et al., 1965). It is likely that other phages use different cofactors when hunting in the intestinal milieu.

Shades of gray

A dynamic picture emerges for RB49 and the many other T4-like phages. When on the prowl, if conditions are unfavorable, their whiskers hold the LTFs close, thereby preventing adsorption. However, this need not be an all-or-none response. Perhaps when in the gut, influenced by multiple environmental signals, RB49 might take a cue from the tryptophan-requiring T4 strains and modulate its response. Depending on the tryptophan concentration, those strains extend one, two, three, or more LTFs. Even when denied tryptophan, only 80-85% retract all their LTFs, which still leaves 15-20% one-legged virions able to contact a host (Kellenberger et al., 1965).

Whiskers are typically described in the literature as "rudimentary" sensory devices, implying they are primitive or undeveloped. In actuality, they are a sophisticated and economical mechanism enabling 'inert' virions to respond adaptively to diverse external clues. They raise the question: Are T4 and its relatives sentient beings?

Cited references

Brenner, S., Champe, S., Streisinger, G., and Barnett, L. (1962). On the interaction of adsorption cofactors with bacteriophages T2 and T4. Virology 17, 30-39.

Conley, M.P., and Wood, W.B. (1975). Bacteriophage T4 whiskers: A rudimentary environment-sensing device. Proc Natl Acad Sci USA 72, 3701-3705.

Efimov, V.P., Nepluev, I.V., Sobolev, B.N., Zurabishvili, T.G., Schulthess, T., Lustig, A., Engel, J., Haener, M., Aebi, U., and Venyaminov, S.Y. (1994). Fibritin encoded by bacteriophage T4 Gene *wac* has a parallel triple-stranded α-helical coiled-coil structure. J Mol Biol 242, 470-486.

Hu, B., Margolin, W., Molineux, I.J., and Liu, J. (2013). The bacteriophage T7 virion undergoes extensive structural remodeling during infection. Science 339, 576-579.

Kellenberger, E., Bolle, A., Boy De La Tour, E., Epstein, R., Franklin, N., Jerne, N., Reale-Scafati, A., Sechaud, J., Bendet, I., and Goldstein, D. (1965). Functions and properties related to the tail fibers of bacteriophage T4. Virology 26, 419-440.

Kostyuchenko, V.A., Chipman, P.R., Leiman, P.G., Arisaka, F., Mesyanzhinov, V.V., and Rossmann, M.G. (2005). The tail structure of bacteriophage T4 and its mechanism of contraction. Nat Struct Mol Biol 12, 810-813.

Letarov, A., Manival, X., Desplats, C., and Krisch, H. (2005). gpwac of the T4-type bacteriophages: Structure, function, and evolution of a segmented coiled-coil protein that controls viral infectivity. J Bacteriol 187, 1055-1066.

Recommended review

Leiman, P, F Arisaka, M van Raaij, V Kostyuchenko, A Aksyuk, S Kanamaru, M Rossmann. 2010. Morphogenesis of the T4 tail and tail fibers. Virol J 7:355.

10^{24}

productive viral infections
per second on Earth

Hendrix, RW. 2010. Recoding in bacteriophages.
in JF Atkins, RF Gesteland, editors. *Recoding:
Expansion of Decoding Rules Enriches Gene Expression*.
Springer. p. 249-258.

The Phage Metagenomic Revolution

Matthew B. Sullivan[†]

Abstract: *Phages in nature are abundant and important as modulators of microbial population structure and metabolic outputs, yet quantifying their impacts in complex and interacting communities remains a major challenge. Fortunately, phage ecological methods have now advanced from counting 'dots' to, for instance, linking phages to their hosts in a population- and genome-based framework. Metagenome-enabled methodologies have led to the realization that phages directly manipulate microbial metabolisms through encoding their own 'auxiliary metabolic genes.' Other applications have organized the vast unknown phage sequence space into countable protein clusters and demonstrated that cyanophage genome sequence space is sufficiently structured to allow populations to be counted. These latter advances in particular allow the field to leverage and test decades of ecological and evolutionary theory to accelerate progress not only for phage research but for the fields of ecology and evolution, as well. It is time for studies of Earth's micro- and nanoscale ecosystem inhabitants to begin leading the life sciences!*

[†] Department of Ecology and Evolutionary Biology, The University of Arizona, Tucson AZ
Email: mbsulli@email.arizona.edu
Website: http://eebweb.arizona.edu/Faculty/mbsulli/

"This changes everything!" I remember saying this to phylogeneticist Ken Halanych while sitting in a Woods Hole classroom back in 2000. Most people were still elated from having survived the big 'Y2K' scare—the fear pandemic that computers everywhere might implode and all human existence would melt down to chaos! But I was talking about a groundbreaking paper that had just come out in *Limnology and Oceanography* (Rohwer et al. 2000). It was by a relatively little-known postdoc, Forest Rohwer, in which he presented the first marine phage genome and its ecological context. They reported that marine phages share features (genes) with non-marine phages, that phosphate scavenging genes appear critical for their survival in P-limited marine waters, and that virion structural proteins might be unrecognizable in the genomes of environmental phages. Looking back, their findings foretold much that we would slowly tease apart in the decade following.

Genome envy

To me, this paper meant so much because I was in my second year of PhD training at MIT and the Woods Hole Oceanographic Institution, and I had just isolated my first marine cyanobacterial phages (cyanophages) of marine *Prochlorococcus.* While I was productively chugging through basic phage characterization of a select few isolates, I yearned for more. I realized, having just read Rohwer's paper, that what I wanted was a genomic map for some of my cyanophages like the one he had for Roseophage SIO1. The power of genomics was alluring; it would be so informative, particularly for my environmental phages that lacked the foundation of decades of knowledge accumulation and genetic tool development. With a genome sequence in hand, you immediately could start thinking about what that phage was doing and what it might look like (getting electron micrographs of environmental phages can be challenging). You could even develop hypotheses about why that particular phage was successful in the environment. Not all the answers are there in the genome alone, of course, as most phage genes are 'unknown,' but still pieces and parts of the story are typically apparent in the average genome. Moreover, the genomic novelty added by each new environmental phage isolate that was sequenced allowed the first predictions of the size of the global pool of *phage* genomes (the global virome). In 2003 a bold estimate—this also from Rohwer—reckoned that it might comprise two billion proteins (Rohwer 2003). This number was an extrapolation from a scant 14 mycobacteriophage genomes, but the possibility of phage sequence space being that large was intriguing. It would make phages the largest source of genetic diversity on Earth.

But I was serious. I did want genomes for my four new *Prochlorococcus* phages. I approached my PhD adviser, Penny Chisholm, about this and she immediately recognized the challenges. At that stage, I was hardly capable of amassing the required quantities of these phages; non-optimal (in hindsight) culturing conditions meant that we were lucky if we could get 1 nanogram of DNA—a thousand-fold less than was needed for genome sequencing at that time! So Penny set up an opportunity for me to actually visit and collaborate with Forest—I was so nervous !!!—to learn how to sequence, annotate, and make sense of phage genome sequences. Even more daunting, for that mission I was allotted two weeks.
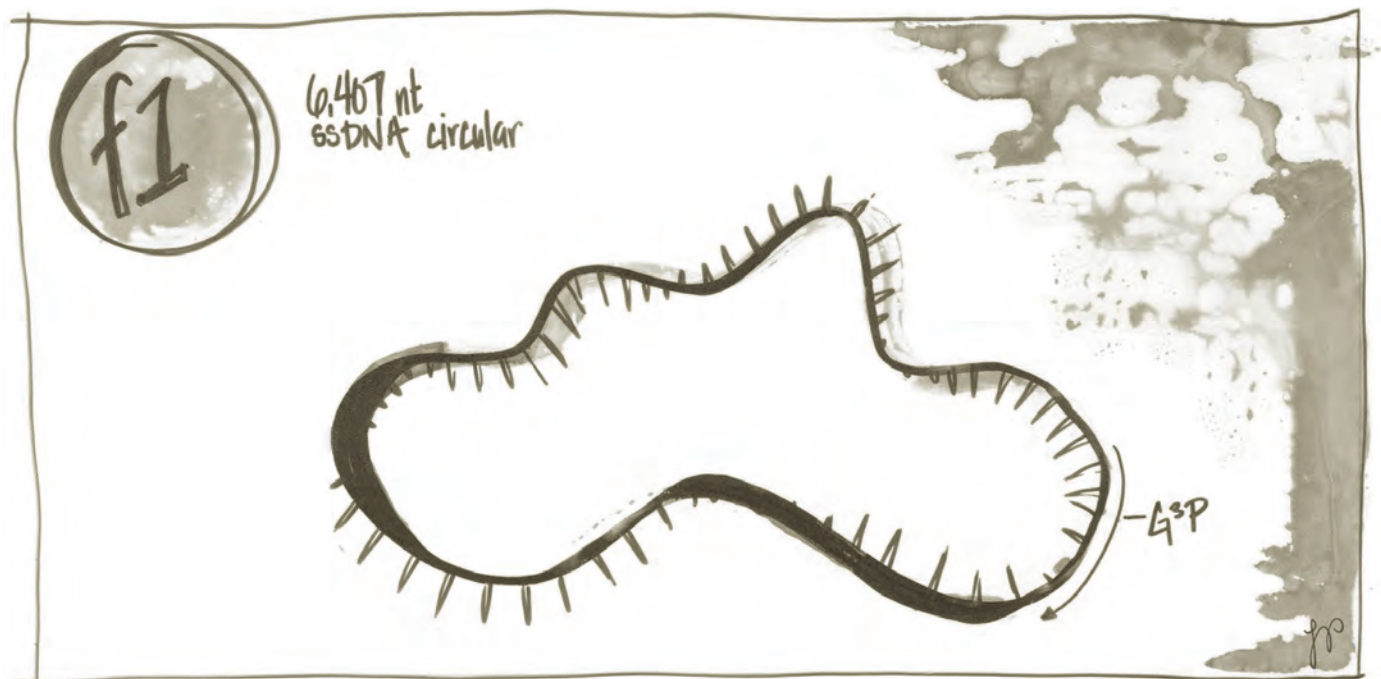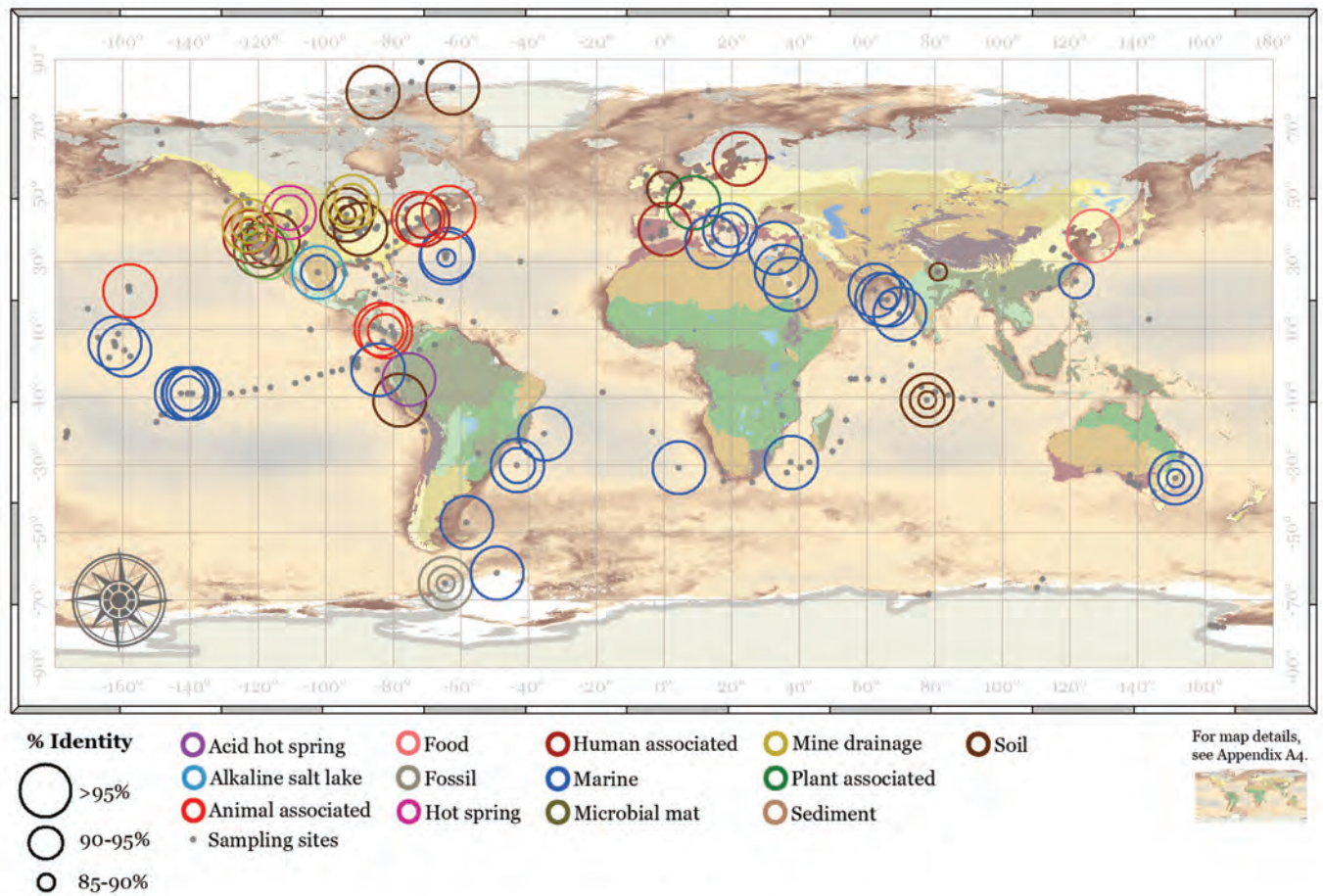
Upon arrival, Forest and I had coffee and chatted, and I was having a great time. He thought about all things phage, and in a totally offbeat way compared to the more microbe-centric world I normally lived in. It was great to finally get to put phage first. Anyone who knows Forest will not be surprised that, rather than work directly on my cyanophage genomes (he hates photosynthesis anyway), he convinced me that we should do two other things instead. First, we should make a web page that walked folks through the steps of getting a phage genome sequenced. (Now, over a decade later, the *Guide to Phage Genomics* is still a top Google hit (http://www.sci.sdsu.edu/PHAGE/guide.html). I loved that he would set me to doing this, as in the process I actually did really think through the isolate-to-genome process. Moreover, that we were providing a community resource exemplified a founding principle of the Luria and Delbruck school of phage biology: the creation of resources for the common scientific good. They had reasoned that rather than scientists competing with each other, the best way to study phage biology—since it is so hard—would be for the field to share anecdotal information and the subtleties of lab protocols so that we could all make more progress towards our particular research goals. I loved it. The second thing Forest suggested I do was help write a marine phage genomics review, my first paper as a PhD student. This was such an early stage in the field and there was so little to review that I knew every publication on the topic inside and out.

Eventually, Forest did send me down the path of learning how to do environmental phage genomics. The first step was on my way back East when I made a detour through Madison, Wisconsin, to spend a week learning how to make clone libraries from nothing—eventually branded as nanocloning. This training was with David Mead, president and founder of what was then a much smaller Lucigen Corporation, and I learned so much. David was incredibly patient with me, and I was soon on my way towards getting genomes of four marine cyanophages. Next step on the path was to figure out how to sequence, an operation that back then was neither easy or cheap. Fortunately, Penny smoothed the way for me as she had gotten a Community Sequencing Program grant funded. This meant that we sent our nanoclone libraries for each phage to the Department of Energy's Joint Genome Institute for Sanger sequencing and assembly. What came back was a finished genome of high quality. It was wonderful to step back in at that point as manually closing the genome could have been a *lot* of work, particularly with so little DNA. I was lucky.

Is this annotation for real?

So, by 2002 I had genomes for all four of my first cyanophages. Now it was time to visit Forest again. On this visit, though, I had my own genomes to look at, and one of Forest's grad students (Mya Breitbart) was there to hold my hand through the process. Back then annotating a phage genome was a manual, brute force process. Mya was amazing, helping me through all the tough spots. I then flew back to MIT with annotated genomes in hand and with new ideas swirling around in my head about who these phages are and what they might be doing. I was so struck by the fact that these cyanophages, isolated as they were from the middle of the low-nutrient, open ocean using marine cyanobacteria as a host, strongly resembled T4-like and T7-like phages of heterotrophic *E coli* isolated from sewage. This was cool—a universality, in a sense—from sewage to open ocean—and some commonalities among the phages infecting bacteria.

Ahh, but there was this problem, or at least I thought it was a problem. The issue was that these

Enterobacteria Phage f1

an Inophage whose progeny extrude from the host & use parental coat proteins parked in the membrane

Genome

ssDNA; circular

6,407 nt

10 predicted ORFs; 0 RNAs

Encapsidation method

Co-condensation

Common host

Escherichia coli

Habitat

Mammalian intestines & sewage

Lifestyle

Non-lytic

Ben Darby

Enterobacteria phage M13

Enterobacteria phage f1

Enterobacteria phage fd

Enterobacteria phage Ike

Enterobacteria phage I22

Pseudomonas phage Pf3

Ralstonia phage p12J

Xanthamonas phage Cf1c

Spiroplasma kunkelii virus SkV1_CR2-3x

Spiroplasma phage 1C74

Spiroplasma phage 1R8A2B

Spiroplasma phage SVTS2

Spiroplasma phage SVGII3

Spiroplasma phage SVGII3

Pseudomonas phage φ2954

Pseudomonas phage φ12

Pseudomonas phage φ13

Pseudomonas phage φ6

Pseudomonas phage φ8

Rhodococcus phage RRH1

Pseudomonas phage Pf1

Cellulophaga phage φ12a:1

Cellulophaga phage φ12:2

Cellulophaga phage φ48:1

Cellulophaga phage φ18:1

Stenotrophomonas phage φSMA9

Stenotrophomonas phage φSHP2

Ralstonia phage PE226

Ralstonia phage RSS1

Ralstonia phage RSS2

Vibrio phage CTX

Vibrio phage KSF-1φ

Vibrio phage VCY-φ

Vibrio phage VfO3K6

Vibrio phage VfO4K68

Vibrio phage Vf33

Vibrio phage Vfi2

Vibrio phage VSKK

Vibrio phage VSK

Vibrio phage VGJφ

Vibrio phage fs1

Vibrio phage NDi-fs1

Vibrio phage VEJφ

ICTV Family

Cystoviridae
Inoviridae
Microviridae
Siphoviridae
Unassigned

% Identity

○ >95%

○ 90-95%

○ 85-90%

○ Acid hot spring
○ Alkaline salt lake
○ Animal associated
· Sampling sites

○ Food
○ Fossil
○ Hot spring

○ Human associated
○ Marine
○ Microbial mat

○ Mine drainage
○ Plant associated
○ Sediment

○ Soil

For map details,
see Appendix A4.

6,407 nt
ssDNA circular

f1

–G³P

Enterobacteria Phage f1

○ Circular genome

g9p, minor capsid
g7p, minor capsid
g8p, major capsid (POG0002)
g10p   g5p, ssDNA-binding (POG0001)
g2p, replication initiation   g3p, attachment
g1p, maturase (POGs 0003,0004)
g6p, minor capsid   g4p, assembly (POG0005)

1,000   2,000   3,000   4,000   5,000   6,000

| On the Prowl | Entry | Takeover | Replication Lytic | Replication Lysogenic | Assembly | Escape | Transfer RNA | Non-coding RNA | Multiple | Unknown |

f1: A Pilus Phage

Heather Maughan & Merry Youle

Mise-en-Scène: *The Ff phages are a closely related group of filamentous phages (e.g., f1, fd, and M13) that coexist amiably with their host. They replicate perpetually as* **episomes***, and their progeny depart without cell lysis.*

A respectable male *E. coli* moves through the gut, and begins to feel frisky. As its pick-up line, it extends a long conjugative F pilus from its surface to hook up with a female. Thus preoccupied, it is unaware of an f1 phage scouting for **pili** nearby. Once f1 catches a pilus tip, it attaches and holds on for a free ride to its target. As *E. coli* retracts the pilus, it unwittingly escorts f1 to its **secondary receptor** in the **periplasm**. This *E. coli* has been fooled, bringing home an infection instead of a mate.

The single-stranded DNA genome of f1 phages travels inside a skinny cylindrical virion 760-900 nm long and only 4.3-6.3 nm in diameter (Marvin, Hohn 1969). Its simple protein shell is built from 2700 copies of the α-helical major coat protein (g8p) arranged in a closely packed helical array similar to overlapping scales on a fish (Glucksman, Bhattacharjee, Makowski 1992). When entering a host, this frugal phage deposits these coat proteins in the cell membrane (CM) for reuse later to coat progeny as they exit. The two virion ends are adorned with different sets of minor coat proteins, those necessary for entering a host at the 'distal' end and those for exiting at the 'proximal' end. Phage f1 exits quietly without killing its host, using a process that mirrors its entry.

Trailing a pilus to the door

The distal end of the virion sports 3-5 copies of g3p, a multi-tasking protein that contacts the receptors during infection (Gray, Brown, Marvin 1981; Rakonjac et al. 2011) and forms a pore in the CM for genome delivery (Glaser-Wuttke, Keppner, Rasched 1989). Each g3p contains three domains (the N-terminal D1, middle D2, and C-terminal D3) that act in succession during infection as the phage worms its way inside (Marvin 1998).

While f1 is on the prowl in the gut, the g3p N-terminus is exposed to the environment with all three of its domains safely tucked in and held close together, the short flexible linker regions between them forming relaxed loops. D2 acts first by attaching to the tip of a passing pilus (Lubkowski et al. 1999; Deng, Perham 2002). As *E. coli* retracts the pilus, the hitchhiking phage passes through the outer membrane (OM). When D2 grabs the pilus this frees the receptor-binding domain (D1) to dangle with its binding site exposed (Eckert et al. 2005). As the first end of the virion enters the periplasm, f1 peeks under the OM and fishes with D1 for its secondary receptor: the C-terminal domain of TolA, a periplasm-spanning bacterial protein (Holliger, Riechmann 1997; Riechmann, Holliger 1997). D1 binds TolA, which in turn frees D3 and allows it to contact the CM for the next step—DNA entry.

Dissolution on entry

Now f1 is poised to thread its DNA into the cell through a CM pore formed cooperatively by the D3 domains of the multiple g3p proteins at hand (Jakes, Davis, Zinder 1988; Glaser-Wuttke, Keppner, Rasched 1989). Unlike the case for most phages, as f1's genome enters the cell it does not leave its capsid at the door, nor does it bring the capsid along with it into the cell. Instead, as the DNA enters, the capsid disassembles with the assistance of host proteins—the rate-limiting step for phage infection (Click, Webster 1998). Phage f1 stashes monomers of the major coat protein (g8p) and some of the minor capsid proteins in the CM for retrieval and reuse by its progeny as they emerge.

Replication

With its small genome, f1 relies on host proteins for many essential functions including replication. Diverted host enzymes convert the phage's ssDNA into a double-stranded template that serves for
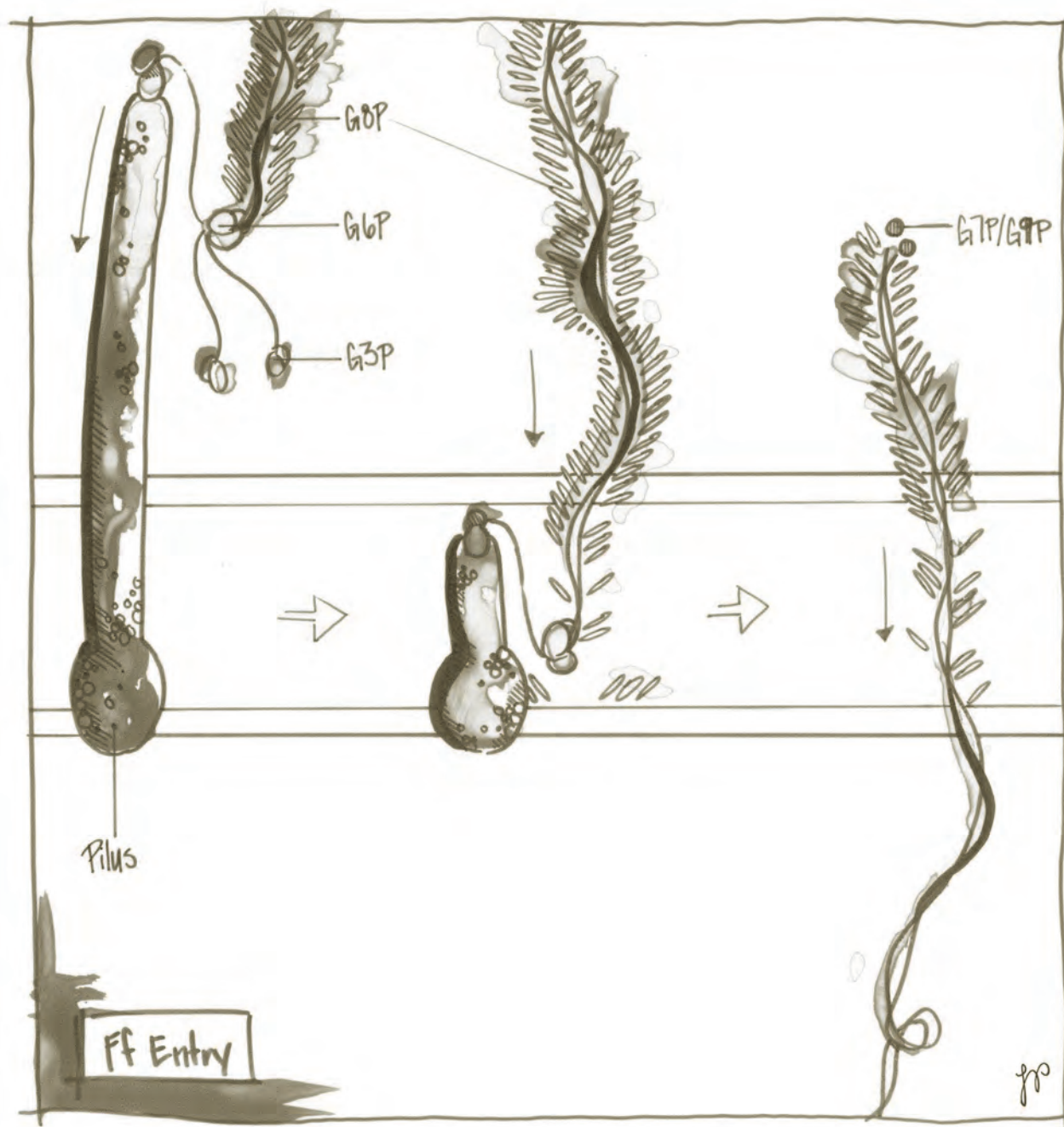
both transcription and the synthesis of new single-stranded genomes (Marvin, Hohn 1969; Russel, Linderoth, Sali 1997). Initially f1 produces new genomes at an exponential rate by converting each ssDNA copy into a double-stranded replicative form. At the same time, f1 actively synthesizes abundant copies of its ssDNA binding protein (g5p), enough copies to soon coat the newly-minted genomes with g5p dimers. Only a short dsDNA hairpin at the proximal end of the genome lacks this interim protein coat (Russel 1991). This hairpin structure serves as the packaging signal that leads the g5p-coated genomes to the CM for final assembly and export. Phage f1 keeps the replication machinery on task indefinitely to support ongoing continuous phage production, generation after generation. About a thousand progeny phage extrude from each cell each generation, altering membrane properties without bringing significant harm to the accommodating *E. coli* host (Marvin, Hohn 1969).
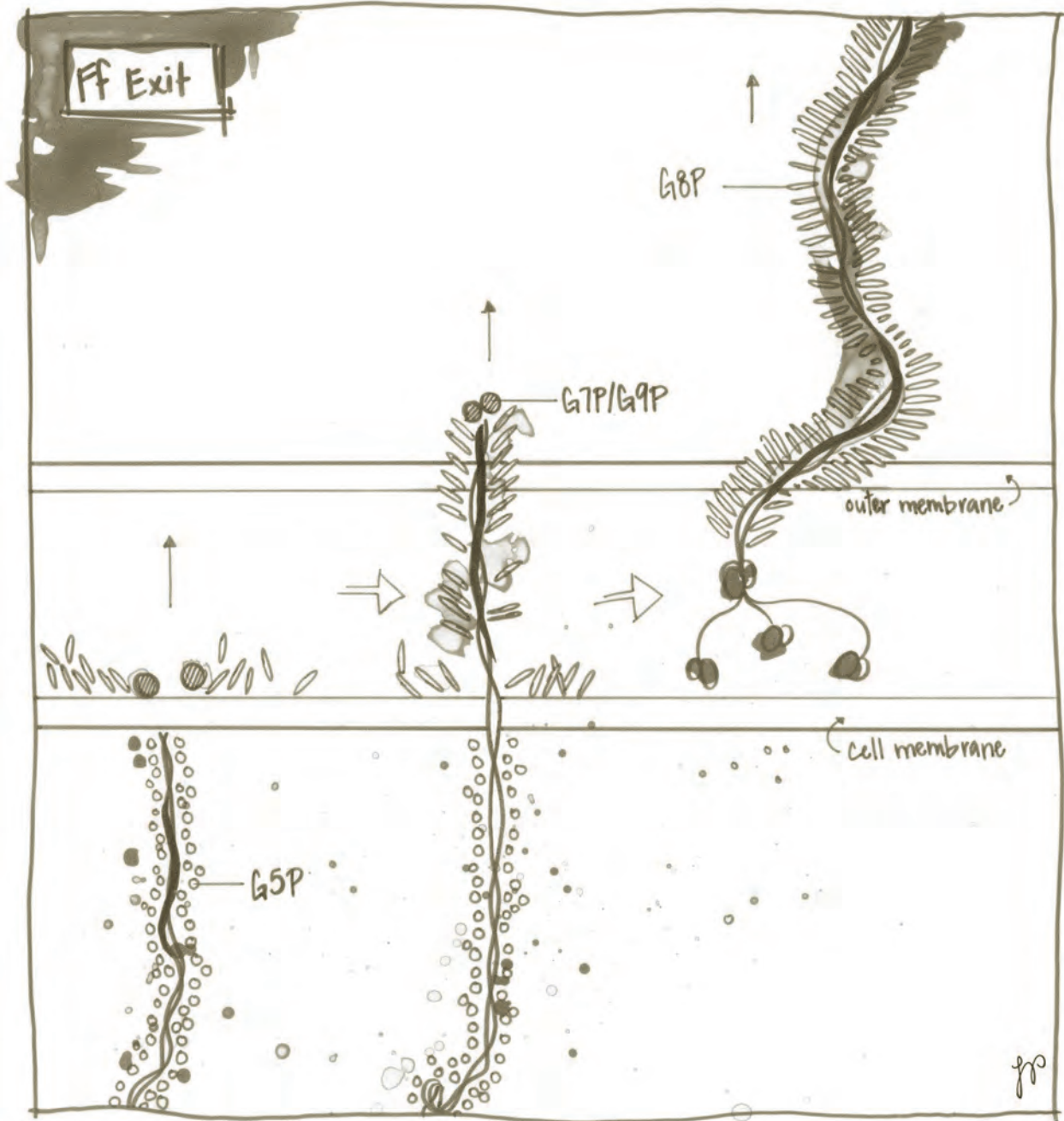
Construction on exit

As f1's cocooned genome approaches the CM with its packaging signal in the lead, it finds the needed virion components waiting as membrane-associated or integral membrane proteins. At the CM, the proximal end acquires its two minor coat proteins (g7p and g9p). Then the DNA passes through the CM, in the process shedding one skin as it acquires another. The g5p dimers are left behind in the cytoplasm, each one replaced by a copy of the major

coat protein g8p. Even though a few of these coat proteins were deposited in the membrane during infection by the parent phage, most were freshly made and anchored in the CM in anticipation. As the extruded proximal end navigates through the periplasm, it identifies its escape hatch in the OM. Although most filamentous phages exit through a borrowed host secretion channel, the f1 phages encode an efficient one of their own, their g4p secretin (Marciano, Russel, Simon 1999). With an aver-

age diameter of 14 nm, these secretins are an open road to freedom (Nickerson et al. 2012).

Since the extruding phage is more than 700 nm long, the leading proximal end clears the ~24 nm host envelope long before the trailing end has reached the CM. When the trailing end of the genome finally arrives there, the virion is 'pinched off' by the addition of the two minor coat proteins unique to this end (g6p and g3p). Protein g3p is crucial here (Rakonjac, Model 1998). Without its participation in terminating and releasing each virion, multiple virions fuse end to end to yield a 'polyphage' that looks suspiciously like a pilus.

Kin?

Some intriguing parallels between pili and filamentous phages hint at an evolutionary link between them (Bradley 1967; Rakonjac, Model 1998). The architecture of both includes a hollow cylinder composed of hundreds (if a phage) or thousands (if a pilus) of copies of a small protein arranged in a helical array. These composite structures disassemble to monomers that are inserted into the CM, where they sit tight until called upon to re-emerge and construct a new pilus or phage. To extrude a filamentous phage, membrane-embedded coat protein monomers are recruited and added one by one to the helical shell surrounding the ssDNA genome as it extrudes from the cell through a secretin OM pore. When the end of the DNA is reached, the structure is cut free from the cell as a completed virion that sets out to seek its fortune in the world. A growing pilus, likewise, extends through a secretin pore by the addition of protein monomers at its base. Pilus extension, however, is followed by retraction, the reverse process in which the pilus disassembles at its base and the proteins return to the membrane. If indeed these mechanisms share a common evolutionary history, which came first—the pilus or the phage?

Cited references

Bradley, DE. 1967. Ultrastructure of bacteriophage and bacteriocins. Bacteriol Rev 31:230-314.

Click, EM, RE Webster. 1998. The TolQRA proteins are required for membrane insertion of the major capsid protein of the filamentous phage f1 during infection. J Bacteriol 180:1723-1728.

Deng, LW, RN Perham. 2002. Delineating the site of interaction on the pIII protein of filamentous bacteriophage fd with the F-pilus of *Escherichia coli*. J Mol Biol 319:603-614.

Eckert, B, A Martin, J Balbach, FX Schmid. 2005. Prolyl isomerization as a molecular timer in phage infection. Nat Struct Mol Biol 12:619-623.

Glaser-Wuttke, G, J Keppner, I Rasched. 1989. Pore-forming properties of the adsorption protein of filamentous phage fd. Biochim Biophys Acta 985:239-247.

Glucksman, MJ, S Bhattacharjee, L Makowski. 1992. Three-dimensional structure of a cloning vector. X-ray diffraction studies of filamentous bacteriophage M13 at 7 Å resolution. J Mol Biol 226:455-470.

Gray, CW, RS Brown, DA Marvin. 1981. Adsorption complex of filamentous fd virus. J Mol Biol 146:621-627.

Holliger, P, L Riechmann. 1997. A conserved infection pathway for filamentous bacteriophages is suggested by the structure of the membrane penetration domain of the minor coat protein g3p from phage fd. Structure 5:265-275.

Jakes, K, N Davis, N Zinder. 1988. A hybrid toxin from bacteriophage f1 attachment protein and colicin E3 has altered cell receptor specificity. J Bacteriol 170:4231-4238.

Lubkowski, J, F Hennecke, A Pluckthun, A Wlodawer. 1999. Filamentous phage infection: Crystal structure of g3p in complex with its coreceptor, the C-terminal domain of TolA. Structure 7:711-722.

Marciano, DK, M Russel, SM Simon. 1999. An aqueous channel for filamentous phage export. Science 284:1516-1519.

Marvin, D, B Hohn. 1969. Filamentous bacterial viruses. Bacteriol Rev 33:172-209.

Marvin, DA. 1998. Filamentous phage structure, infection and assembly. Curr Opin Struct Biol 8:150-158.

Nickerson, NN, SS Abby, EP Rocha, M Chami, AP Pugsley. 2012. A single amino acid substitution changes the self-assembly status of a type IV piliation secretin. J Bacteriol 194:4951-4958.

Rakonjac, J, NJ Bennett, J Spagnuolo, D Gagic, M Russel. 2011. Filamentous bacteriophage: Biology, phage display and nanotechnology applications. Curr Issues Mol Biol 13:51-76.

Rakonjac, J, P Model. 1998. Roles of pIII in filamentous phage assembly. J Mol Biol 282:25-41.

Riechmann, L, P Holliger. 1997. The C-terminal domain of TolA is the coreceptor for filamentous phage infection of *E. coli*. Cell 90:351-360.

Russel, M. 1991. Filamentous phage assembly. Mol Microbiol 5:1607-1613.

Russel, M, NA Linderoth, A Sali. 1997. Filamentous phage assembly: Variation on a protein export theme. Gene 192:23-32.

Recommended reviews

Marvin, D, B Hohn. 1969. Filamentous bacterial viruses. Bacteriol Rev 33:172-209.

Marvin, D, M Symmons, S Straus. 2014. Structure and assembly of filamentous bacteriophages. Prog Biophys Mol Biol 114:80-122.

Rakonjac, J, NJ Bennett, J Spagnuolo, D Gagic, M Russel. 2011. Filamentous bacteriophage: Biology, phage display and nanotechnology applications. Curr Issues MolBiol 13:51-76.

Into the Devil's Kitchen:
A Personal History of Archaeal Viruses

by Kenneth Stedman[†]

Abstract: *Archaeal viruses are eccentric in both their virion structures and their genomes (and in their selection of researchers allowed to study them). Even the arguably best-researched archaeal virus, the lemon-shaped Fusellovirus SSV1, is replete with unsolved mysteries. My virus hunting career began at the side of Wolfram Zillig, the pioneer in the field, and over the last 20 years, together with other researchers, we have discovered many viruses of extremophilic Archaea. While more undoubtedly remain to be found, the field is poised to move from the discovery of new viruses to the exploration of the unique replication and host interactions of these fascinating nanobes.*

[†]Department of Biology and Center for Life in Extreme Environments, Portland State University, Portland, OR
Email: kstedman@pdx.edu
Website: http://web.pdx.edu/~kstedman/

Prologue

It was mid-September of 2003. I was hunting for viruses in a solfataric field in Lassen Volcanic National Park with a new Ph.D. student, Adam Clore, and an undergraduate student, Random Diessner. We had tortuously made our way around the moonscape-like environment to get near to a promising bubbling murky spring (Fig. 1). Now it was time to go in to "Devil's Kitchen," one of the main hydrothermal areas of the park. I carefully led my students across the fragile ground towards a promising spring, only to have my boot break through the thin crust of soil into the boiling acidic mud beneath.

Act 1. Archaeal viruses: Extremely different

Why was I endangering myself and my students to collect a small amount of hot, acidic, muddy water? The danger was real. The namesake of a nearby thermal area, Mr. Bumpass, lost both of his legs after falling into some of this stuff. We were hunting new viruses that infect *Sulfolobus* and its relatives. Members of the *Sulfolobales* are among some of the first-discovered and best-studied Archaea. The crenarchaeon *Sulfolobus* thrives in boiling acidic springs at 80° C and at pH 3 or even lower—quite remarkable in itself. The viruses that infect these thermoacidophiles are even more extraordinary with their unique shapes and genomes. They are so divergent, in fact, that an unprecedented ten new virus families were proposed

Figure 1. Overview of Devil's Kitchen, Lassen Volcanic National Park. September 2003. The sampled spring of interest indicated with an orange arrow. Photo credit: K. Stedman.

to accommodate them (Prangishvili 2013). Their virions offer an incredibly diverse assortment of shapes (Fig. 2). The relatively rare types with the familiar icosahedral capsid architecture include the *Sulfolobus* turreted icosahedral virus (STIV) that I discovered in Yellowstone National Park (Rice et al. 2004). Some, such as the aptly-named *Sulfolobus islandicus* rod-shaped virus (SIRV), are indeed rod-shaped (Prangishvili et al. 1999). These are outdone by those with filamentous vi-

rions whose length is twice the diameter of the cells that they infect; some of these, the *Sulfolobus islandicus* filamentous virus (SIFV), for instance, have nano-sized claw-like structures at their termini (Arnold et al. 2000). There are also amazing bottle-shaped virions such as ABV (Haring et al. 2005a). The majority of archaeal viruses have spindle or lemon-shaped virions of varying sizes, with or without long slender tails. One of these, the *Acidianus* two-tailed virus (ATV), 'grows' tails after exiting its host provided it is at the usual hot spring temperature (*see page 6-31*; Haring et al. 2005b). David Prangishvili has recently written an excellent review of this bizarre world of archaeal viruses (Prangishvili 2013). Here I will feature the *Sulfolobus* spindle-shaped viruses (SSVs), aka the Fuselloviruses, that are the main focus of my research group (Stedman, Prangishvili, Zillig 2006).

Act 2. SSV1: A lemon full of puzzles

The best-studied Fusellovirus is SSV1. Its genome is unique (Fig. 3); only one of the 35 open reading frames (ORFs), or putative genes, is clearly homologous to sequences found in any other viral or cellular genome (Palm et al. 1991). This one, the SSV1 integrase gene, is homologous to the well-studied integrase of phage λ, but possesses a few quirks of its own. First, the attachment site that is cleaved when the viral genome integrates into the host genome lies within the integrase gene itself. Thus, during integration, the integrase gene is disrupted and presumably inactivated (Reiter, Palm 1990). Another intriguing aspect concerns the structure and activity of the tetrameric functional form of the integrase. That the four monomers act in trans (Letzelter, Duguet, Serre 2004; Eilers, Young, Lawrence 2012) makes SSV1's integrase more similar to the eukaryotic flp-like recombinases than to λ integrase. Moreover, the SSV1 integrase gene is not essential for viral reproduction; if it is deleted, viral infection appears to proceed normally. However, the integrase gene must play some as yet unknown obscure role as viruses lacking this gene are at a competitive disadvantage relative to the wild-type (Clore, Stedman 2007). In contrast to phage λ, SSV1 does not have a lytic replication phase, but releases its virions, apparently without

Figure 2. Novel virion shapes in archaeal viruses as drawn based on photographs or diagrams in the cited references. Virion #1: *Sulfolobus* turreted icosahedral virus, STIV (Rice et al. 2004); virion #2: *Sulfolobus islandicus* rod-shaped virus, SIRV (Prangishvili et al. 1999); virion #3: *Sulfolobus islandicus* filamentous virus, SIFV (Arnold et al. 2000); virion #4: *Acidianus* bottle-shaped virus, ABV (Häring et al. 2005a); virion #5: *Acidianus* two-tailed virus, ATV (Häring et al. 2005b); virion #6: *Sulfolobus* spindle-shaped virus, SSV (Martin et al. 1984; Schleper, Kubo, Zillig 1992).

host lysis, by budding at the cellular membrane. Nevertheless, as is the case for phage λ, UV irradiation can induce increased SSV1 virion production up to 100-fold (Martin et al. 1984; Schleper, Kubo, Zillig 1992). The molecular mechanism of this induction is not clear, that of the assembly of the SSV1 virion even less so.

Other SSV1 genes whose function is known include those that encode the three virion structural proteins (VP1, VP2, and VP3). The VP1 protein is the major capsid protein and VP3 the minor capsid protein; together they make up the majority of the proteins in the distinctive, lemon-shaped capsid (Reiter et al. 1987a). The location of VP1 and VP3 in the capsid is not clear. There are many more copies of VP1 than VP3, indicating that the latter may be concentrated at the termini of the particle or in locations of pentagonal symmetry. Preliminary single-particle cryo-EM reconstruction data sug-