VIRUS DYNAMICS, ARCHAEAL POPULATIONS, AND WATER CHEMISTRY OF THREE ACIDIC HOT SPRINGS IN YELLOWSTONE NATIONAL PARK

by

Jamie Christine Snyder

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctorate of Philosophy

in

Microbiology

MONTANA STATE UNIVERSITY
Bozeman, MT

May 2005

© COPYRIGHT

by

Jamie Christine Snyder

2005

All Rights Reserved

APPROVAL

of a dissertation submitted by

Jamie Christine Snyder

This dissertation has been read by each member of the dissertation committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the College of Graduate Studies.

Dr. Mark J. Young, PhD

Approved for the Department of Microbiology

Dr. Tim Ford, PhD

Approved for the College of Graduate Studies

Dr. Bruce McLeod, PhD

STATEMENT OF PERMISSION TO USE

In presenting this dissertation in partial fulfillment of the requirements for a doctoral degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library. I further agree that copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for extensive copying or reproduction of this dissertation should be referred to ProQuest Information and Learning, 300 North Zeeb Road, Ann Arbor, Michigan 48106, to whom I have granted "the exclusive right to reproduce and distribute my dissertation in and from microform along with the non-exclusive right to reproduce and distribute my abstract in any format in whole or in part."

Jamie Christine Snyder

ACKNOWLEDGEMENTS

I would like to thank all of the members of my committee: Dr. Mark Young, Dr. Michele Hardy, Dr. Robert Sharrock, Dr. David Ward, and Dr. Martin Lawrence. Thanks for the helpful feedback and your support over the years. I would specifically like to thank my advisor, Mark Young, for all of the guidance and support throughout this research.

I would like to thank all the current and past members of the lab: Debbie Willits, Sue Brumfield, Dr. Eric Gillitzer, Josh Spuhler, Blake Wiedenheft, George Rice, Ben Widener, Hulda Thorisdottir, Jennifer "Furry" Fulton, Darla Cadman, Heather Spuhler, Lars Liepold, Jen Revis, Mir Anwar, Dr. Ken Stedman, Ellie (Redfield) Kurth, Amber Roche, Tom Arnold, Chris Crowley, Carli Rognlie, Janet Mendonca, Na Li, and Loren Barber.

A special thanks to: Dr. Eric Gillitzer, for taking the time to critically read this dissertation; Josh Spuhler, for always being willing to go to Yellowstone and collect samples, I enjoyed every trip; Debbie Willits, for critically reading this dissertation, for being helpful when-ever necessary, and for being a great friend; and Dr. Ken Stedman for sending me SSV-like sequences and water chemistry data from Lassen National Park, and for answering numerous questions along the way. I would also like to thank all the people who worked on the Yellowstone project, it has been both exciting and rewarding working with you.

I would like to thank Dr. Frank Roberto at the Idaho National Lab in Idaho Falls, Idaho for sequencing my 96-well plates, for being a great mentor, and for being a friend. I would like to thank Dr. Matt Lavin in the Plant Sciences and Plant Pathology department at MSU for helping me get started with the phylogenetic programs, the statistical programs, and helping me decipher the results. I would like to thank Christie Hendrix at the Research Permit Office in Yellowstone National Park for always being more than helpful in allowing us to conduct our research in the thermal areas of Yellowstone.

Finally, I would like to thank the members of my family whom over all the years have supported me, and the choices that I have made. Specifically, I would like to thank my husband, Dan McCarthy, for his patience and willingness to put his plans on hold while I finished my degree.

TABLE OF CONTENTS

1. VIRUSES OF ARCHAEA: BACKGROUND AND GOALS OF THIS RESEARCH……1

INTRODUCTION AND OVERVIEW……1
MAJOR QUESTIONS AND HYPOTHESIS……5
BACKGROUND: ARCHAEA……5
Korarchaeota and Nanoarchaeota……6
Euryarchaeota……8
Crenarchaeota……9
Archaea in Yellowstone National Park……10
BACKGROUND: ARCHAEAL VIRUSES……12
Euryarchaeotal viruses……14
Halophilic viruses……16
Methanogenic viruses……17
Thermophilic euryarchaeotal viruses……18
Crenarchaeotal viruses……18
Thermoproteus viruses……19
Acidianus viruses……20
Sulfolobus viruses……21
Uncharacterized viruses……30
Archaeal viruses in Yellowstone……31
MONITOR SITE SELECTION, RATIONALE, AND CHARACTERISTICS……32
SUMMARY……35

2. POPULATION DYNAMICS OF VIRUSES IN CULTURE……37

INTRODUCTION……37
MATERIALS AND METHODS……39
Sample collection and culturing……39
DNA extraction and viral PCR……40
Cloning and sequencing……41
Sequence analysis……42
RESULTS……42
Rabbit Creek and Ragged Hills virus culturing bias……42
Viral culturing bias using two medias……45
Rabbit Creek culturing bias……45
Ragged Hills culturing bias……48
SSV1 changes in *Sulfolobus solfataricus* P2 culture……52
CONCLUSIONS……57

3. CRATER HILLS MONITOR SITE: VIRUS DYNAMICS, HOST POPULATIONS, AND WATER CHEMISTRY……62

INTRODUCTION……62
MATERIALS AND METHODS……63
Site location……63
pH and temperature monitoring……63
Sample collection and water chemistry analysis……63
Total nucleic acid extraction from environmental samples……64

TABLE OF CONTENTS – CONTINUED

Primer design for amplification of viruses……64
PCR sensitivity of the primer sets……67
Archaeal 16S rDNA primers and PCR……68
Cloning PCR products……69
DNA sequencing……70
Sequence and phylogenetic analysis……71
Viral sequences……71
Archaeal host sequences……73
RESULTS……73
Temperature, pH, and water chemistry of the Crater Hills monitor site……73
Archaeal population in the Crater Hills monitor site……79
Archaeal population determined by the 20F/UA1406R primer set……79
Archaeal population determined by the 348Fa/UA1406R primer set……82
Population dynamics of viruses in the Crater Hills monitor site……88
SSV-like population dynamics in the Crater Hills monitor site……88
Detailed analysis of SSV-like population dynamics……104
SIRV-like population dynamics in the Crater Hills monitor site……108
Detailed analysis of SIRV-like population dynamics……119
Comparing SSV-like and SIRV-like population dynamics……124
CONCLUSIONS……125

4. RABBIT CREEK MONITOR SITE: VIRUS DYNAMICS, HOST POPULATIONS, AND WATER CHEMISTRY……133

INTRODUCTION……133
MATERIALS AND METHODS……134
Site location……134
pH and temperature monitoring……134
Sample collection and water chemistry……134
Total nucleic acid extraction from environmental samples……134
Primer design and PCR for viruses……135
Archaeal 16S rDNA primers and PCR……135
Cloning PCR products……136
DNA sequencing……136
Sequence and phylogenetic analysis……136
Viral sequences……136
Archaeal host sequences……138
RESULTS……138
Temperature, pH, and water chemistry of the Rabbit Creek monitor site……138
Archaeal population in the Rabbit Creek monitor site……144
Archaeal population determined by the 20F/UA1406R primer set……145
Archaeal population determined by the 348Fa/UA1406R primer set……148
Population dynamics of viruses in the Rabbit Creek monitor site……151
SSV-like population dynamics in the Rabbit Creek monitor site……151
Detailed analysis of SSV-like population dynamics……170
SIRV-like population dynamics in the Rabbit Creek monitor site……175
Detailed analysis of SIRV-like population dynamics……184
Comparing SSV-like and SIRV-like population dynamics……189
CONCLUSIONS……190

TABLE OF CONTENTS - CONTINUED

5. RAGGED HILLS MONITOR SITE: VIRUS DYNAMICS, HOST POPULATIONS, AND WATER CHEMSITRY……..197

INTRODUCTION……197
MATERIALS AND METHODS……198
Site location…..……..198
pH and temperature monitoring…….198
Sample collection and water chemistry……..198
Total nucleic acid extraction from environmental samples……..…199
Primer design and PCR for viruses……199
Archaeal 16S rDNA primers and PCR……...200
Cloning PCR products……200
DNA sequencing……200
Sequence and phylogenetic analysis……...201
Viral sequences……201
Archaeal host sequences……..202
RESULTS……203
Temperature, pH, and water chemistry in the Ragged Hills monitor site……..…203
Archaeal population in the Ragged Hills monitor site……...…209
Population dynamics of viruses in the Ragged Hills monitor site……214
SSV-like population dynamics in the Ragged Hills monitor site……214
Detailed analysis of SSV-like population dynamics……233
SIRV-like population dynamics in the Ragged Hills monitor site……...239
Detailed analysis of SIRV-like population dynamics……..251
Comparing SSV-like and SIRV-like population dynamics……255
CONCLUSIONS……257

6. COMPARISON OF THREE ACIDIC HOT SPRINGS IN YELLOWSTONE NATIONAL PARK……265

INTRODUCTION……..265
COMPARISON OF WATER CHEMISTRY IN THREE ACIDIC HOT SPRINGS……266
Water chemistry comparison with YNP monitor sites……...274
Water chemistry in Lassen National Park……274
Water chemistry of Lake Victoria, Africa……276
COMPARISON OF ARCHAEAL POPULATIONS IN THREE ACIDIC HOT SPRINGS……...277
COMPARISON OF SSV-LIKE POPULATIONS IN THREE ACIDIC HOT SPRINGS……283
Comparison of SSV-like populations using 3F/4R primer set……...284
Comparison of SSV-like populations using 7F/8R primer set……...…297
COMPARISON OF SIRV-LIKE POPULATIONS IN THREE ACIDIC HOT SPRINGS……...…301
DISCUSSION……....310
Implications of this research……..315
SUMMARY……...317

7. CONCLUSIONS TO QUESTIONS AND HYPOTHESIS……318

QUESTION 1……318
QUESTION 2……....319
QUESTION 3……....319
QUESTION 4……....320
HYPOTHESIS……...320

TABLE OF CONTENTS – CONTINUED

REFERENCES CITED……322

APPENDICES……..337

APPENDIX A: SUPPLEMENTARY PHYLOGENETIC ANALYSES FROM THE CRATER HILLS MONITOR SITE……..…..338

APPENDIX B: SUPPLEMENTARY PHYLOGENETIC ANALYSES FROM THE RABBIT CREEK MONITOR SITE……...…373

APPENDIX C: SUPPLEMENTARY PHYLOGENETIC ANALYSES FROM THE RAGGED HILLS MONITOR SITE……...406

LIST OF TABLES

Table Page

1.1. Viruses isolated from the Euryarchaeotes……..15

1.2. Viruses isolated from the Crenarchaeotes……..…19

1.3 Viruses and virus-like particles detected in environmental samples and enrichment cultures from Yellowstone…….31

3.1. Number of base pairs amplified with universal SSV primers in four sequenced SSV isolates…….66

3.2. Water chemistry values measured in the Crater Hills monitor site over time……75

3.3. The percent identities of environmental 16S rDNA clones from Crater Hills to published 16S rDNA sequences……....85

3.4. The presence of a primer bias in two sample collected from the Crater Hills monitor site……...87

3.5. The number of SSV-like clones sequenced from the Crater Hills monitor site over time……91

3.6. The comparison of functional and non-functional SSV-like ORFs from the Crater Hills monitor site……92

3.7. The SSV-like 3F/4R sequence abundance from the Crater Hills monitor site over time……...105

3.8. The numbers of transitions and transversions of SSV-like sequences in each sample from Crater Hills……....107

3.9. The number of transitions and transversions of SSV-like sequences over time from Crater Hills…….108

3.10. The number of SIRV-like clones sequenced from the Crater Hills monitor site over time……..112

3.11. The SIRV-like sequence abundance from the Crater Hills monitor site over time……..121

LIST OF TABLES – CONTINUED

Table Page

3.12. The numbers of transitions and transversions of SIRV-like sequences in each sample collected from Crater Hills……123

3.13. The number of transitions and transversions of SIRV-like sequences collected over time from Crater Hills……123

3.14. The matches between universal 16S rDNA primers and archaeal sequences detected in Yellowstone……128

4.1. Water chemistry values measured in the Rabbit Creek monitor site over time……..141

4.2. The percent identities of 16S rDNA sequences from Rabbit Creek to published 16S rDNA sequences……150

4.3. Number of SSV-like clones detected over time in the Rabbit Creek monitor site……151

4.4. The number of functional and non-functional SSV-like ORFs from Rabbit Creek……..152

4.5. Sequence abundance of SSV-like 3F/4R sequences in the Rabbit Creek monitor site over time……..…171

4.6. Transitions and transversions of SSV-like 3F/4R sequences in each sample from Rabbit Creek……...173

4.7. Transitions and transversions of SSV-like 3F/4R sequences collected from the Rabbit Creek monitor site over time……..173

4.8. Transitions and transversions of SSV-like 7F/8R sequences in each sample from Rabbit Creek……...174

4.9. Transitions and transversions of SSV-like 7F/8R sequences collected from the Rabbit Creek monitor site over time……..174

4.10. Number of SIRV-like clones sequenced from the Rabbit Creek monitor site over time……....177

LIST OF TABLES – CONTINUED

Table Page

4.11. Sequence abundance of SIRV-like sequences in the Rabbit Creek monitor site over time……186

4.12. Transitions and transversions of SIRV-like sequences in each sample from Rabbit Creek……...…188

4.13. Transitions and transversions of SIRV-like sequences collected from the Rabbit Creek monitor site over time……..189

5.1. Water chemistry values measured in the Ragged Hills monitor site over time……..204

5.2. The percent identities of 16S rDNA sequences from Ragged Hills compared to published 16S rDNA sequences…….211

5.3. Number of SSV-like clones sequenced from the Ragged Hills monitor site……..216

5.4. The number of functional and non-functional SSV-like ORFs from Ragged Hills……217

5.5. Sequence abundance of SSV-like 3F/4R sequences in Ragged Hills over time……..…235

5.6. Transitions and transversions of SSV-like 3F/4R sequences in each sample from Ragged Hills……...237

5.7. Transitions and transversions of SSV-like 3F/4R sequences collected from the Ragged Hills monitor site over time……..237

5.8. Transitions and transversions of SSV-like 7F/8R sequences in each sample from Ragged Hills……...238

5.9. Transitions and transversions of SSV-like 7F/8R sequences collected from the Ragged Hills monitor site over time……..239

5.10. Number of SIRV-like clones sequenced from the Ragged Hills monitor site……...240

LIST OF TABLES – CONTINUED

Table Page

5.11. Sequence abundance of SIRV-like sequences in Ragged Hills over time……...252

5.12. Transitions and transversions of SIRV-like sequences in each sample from Ragged Hills……...255

5.13. Transitions and transversions of SIRV-like sequences collected from the Ragged Hills monitor site over time……..255

6.1. The dates of the chemical analysis in each of the monitor sites over time……...…267

6.2. The averages of water chemistry values measured in each monitor site over time……...269

6.3. A comparison of water chemistry measurements from Yellowstone and Lassen National Parks……...275

6.4. A comparison of water chemistry in the three monitor sites and Lake Victoria, Africa…….276

6.5. The number of 16S rDNA sequences collected from each monitor site and the percent identity to published sequences……279

6.6. The number of 16S rDNA sequences detected from all three sites and the sites they were detected……...281

6.7. The number of SSV-like and SIRV-like clones sequenced from each of the monitor sites……..286

6.8. The numbers of transitions and transversions in the viral sequences from all three sites……...292

6.9. The groups of identical SSV-like sequences from the three monitor sites over time…….293

6.10. The nucleotide substitution models for each data set from each monitor site……..294

LIST OF TABLES – CONTINUED

Table Page

6.11. The groups of identical SIRV-like sequences from the three monitor sites over time……303

LIST OF FIGURES

Figure Page

1.1. Tree of life illustrating the three domains of life: Bacteria, Eukarya, and Archaea……...7

1.2. Two major sub-domains of Archaea: Euryarchaeotes and Crenarchaeotes…….7

1.3. Genome map and electron micrograph of SSV isolated from Japan (SSV1)……..…22

1.4. Genome map and electron micrograph of SSV isolated from Iceland…….……...25

1.5. Genome map and electron micrograph of SSV isolated from Yellowstone National Park (SSV-RH)……...26

1.6. Genome map and electron micrograph of SSV isolated from Kamchatka, Russia (SSV-K1)……...…….26

1.7. Electron micrographs of SIRVs from Iceland and SIRV-like viruses from Yellowstone……...27

1.8. Electron micrographs of SIFVs from Iceland and SIFV-like viruses from Yellowstone……...29

1.9. Electron micrograph and genome map of a newly discovered icosahedral virus from Yellowstone (STIV) …….……30

1.10. A map of Yellowstone illustrating the locations of the sites chosen for this research……..33

2.1. Environmental and cultured SSV-like 3F/4R DNA sequences from Rabbit Creek……..43

2.2. Environmental SSV-like 7F/8R DNA sequences from Ragged Hills compared to the four sequenced SSV isolates……...…44

2.3. SSV-like 3F/4R DNA sequences from Rabbit Creek cultured in two different medias……...……46

LIST OF FIGURES – CONTINUED

Figure Page

2.4. Environmental and cultured (in two medias) SSV-like 3F/4R DNA sequences from Rabbit Creek……..…….47

2.5. Environmental and cultured SSV-like 7F/8R DNA sequences from Rabbit Creek……..…48

2.6. Cultured SSV-like 3F/4R and 7F/8R DNA sequences from Ragged Hills……...……49

2.7. Environmental and cultured SSV-like 7F/8R DNA sequences from Ragged Hills……..50

2.8. Cultured SSV-like 3F/4R and 7F/8R DNA sequences from Rabbit Creek and Ragged Hills…….…….51

2.9. SSV1 3F/4R DNA sequences obtained from an infected *S. solfataricus* P2 culture……....53

2.10. SSV1 3F/4R protein sequences obtained from an infected *S. solfataricus* P2 culture……....54

2.11. SSV1 7F/8R DNA sequences obtained from an infected *S. solfataricus* P2 culture……....55

2.12. SSV1 7F/8R protein sequences obtained from an infected *S. solfataricus* P2 culture……...….56

3.1. The alignment of four sequenced SSV isolates and the 3F/4R primer set, which amplifies a portion of the largest ORF in SSVs……..…65

3.2. The alignment of four sequenced SSV isolates and the 7F/8R primer set, which amplifies the viral coat proteins……....65

3.3. A genomic map of SSV-RH showing the locations of the universal SSV primer sets 3F/4R and 7F/8R…….66

3.4. Alignment of primers used to amplify the coat protein of SIRV-like viruses in YNP……....67

LIST OF FIGURES – CONTINUED

Figure Page

3.5. Temperature and pH over time at the Crater Hills monitor site and a picture of the site……..74

3.6. Water chemistry values obtained from the Crater Hills monitor site over the course of this research……...76

3.7. Comparison of sulfate and aluminum concentrations in the three monitor sites over time……..…78

3.8. The average water chemistry values measured over time in the Crater Hills monitor site…….78

3.9. An overview of the archaeal population over time in the Crater Hills monitor site measured with the 20F/UA1406R primer set……....79

3.10. Phylogenetic analysis of environmental 16S rDNA clones amplified with 20F/UA1406R from Crater Hills over time……...81

3.11. Crater Hills host population based on nearly full-length or full-length 16S rDNA clones over time…….82

3.12. An overview of the archaeal population over time in the Crater Hills monitor site measured with the 348Fa/UA1406R primer set……...83

3.13. Phylogenetic analysis of environmental 16S rDNA clones amplified with 348Fa/UA1406R from Crater Hills……84

3.14. An overview of the SSV-like DNA sequences detected over time in the Crater Hills monitor site……...…89

3.15. An overview of the SSV-like protein sequences detected over time in the Crater Hills monitor site……...89

3.16. SSV-like DNA and protein sequences detected from Crater Hills on August 6, 2002……...90

3.17. SSV-like DNA sequences detected in the Crater Hills monitor site over time…….96

LIST OF FIGURES – CONTINUED

Figure Page

3.18. SSV-like protein sequences detected in the Crater Hills monitor site over time…….97

3.19. A phylogenetic comparison of the DNA sequences of the SSV-like functional and non-functional ORFs…….101

3.20. A phylogenetic comparison of the protein sequences of the SSV-like functional and non-functional ORFs…….102

3.21. SSV-like sequence abundance from the Crater Hills monitor site over time……106

3.22. An overview of the SIRV-like DNA sequences present in the Crater Hills monitor site over time……...109

3.23. An overview of the SIRV-like protein sequences present in the Crater Hills monitor site over time…….110

3.24. SIRV-like sequences detected from Crater Hills on August 6, 2002……..111

3.25. SIRV-like DNA sequences detected in the Crater Hills monitor site over time……...114

3.26. SIRV-like protein sequences detected in the Crater Hills monitor site over time……...115

3.27. SIRV-like sequence abundance from the Crater Hills monitor site over time……122

3.28. SSV-like sequences detected over time from the Crater Hills monitor site; sequences from October 2002 are red…….339

3.29. SSV-like sequences detected over time from the Crater Hills monitor site; sequences from February 2003 are red……340

3.30. SSV-like sequences detected over time from the Crater Hills monitor site; sequences from May 2003 are red……...341

LIST OF FIGURES – CONTINUED

Figure Page

3.31. SSV-like sequences detected over time from the Crater Hills monitor site; sequences from July 2003 are red……...342

3.32. SSV-like DNA sequences detected over time from the Crater Hills monitor site; sequences from August 2003 are red……..343

3.33. SSV-like protein sequences detected over time from the Crater Hills monitor site; sequences from August 2003 are red……..344

3.34. SSV-like DNA sequences detected over time from the Crater Hills monitor site; sequences from September 2003 are red……..345

3.35. SSV-like protein sequences detected over time from the Crater Hills monitor site; sequences from September 2003 are red……346

3.36. SSV-like DNA sequences detected over time from the Crater Hills monitor site; sequences from October 2003 are red……..347

3.37. SSV-like protein sequences detected over time from the Crater Hills monitor site; sequences from October 2003 are red……348

3.38. SSV-like DNA sequences detected in the Crater Hills monitor site from August 2002 through October 2003……..349

3.39. SIRV-like sequences detected over time from the Crater Hills monitor site; sequences from October 2002 are red……..350

3.40. SIRV-like DNA sequences detected over time from the Crater Hills monitor site; sequences from January 2003 are red……..351

3.41. SIRV-like protein sequences detected over time from the Crater Hills monitor site; sequences from January 2003 are red……..352

3.42. SIRV-like DNA sequences detected over time from the Crater Hills monitor site; sequences from February 2003 are red……...353

3.43. SIRV-like protein sequences detected over time from the Crater Hills monitor site; sequences from February 2003 are red……..354

LIST OF FIGURES – CONTINUED

Figure Page

3.44. SIRV-like DNA sequences detected over time from the Crater Hills monitor site; sequences from May 2003 are red……..…355

3.45. SIRV-like protein sequences detected over time from the Crater Hills monitor site; sequences from May 2003 are red……..356

3.46. SIRV-like DNA sequences detected over time from the Crater Hills monitor site; sequences from July 2003 are red……...357

3.47. SIRV-like protein sequences detected over time from the Crater Hills monitor site; sequences from July 2003 are red……..358

3.48. SIRV-like DNA sequences detected over time from the Crater Hills monitor site; sequences from August 2003 are red……..359

3.49. SIRV-like protein sequences detected over time from the Crater Hills monitor site; sequences from August 2003 are red……..360

3.50. SIRV-like DNA sequences detected over time from the Crater Hills monitor site; sequences from September 2003 are red……361

3.51. SIRV-like protein sequences detected over time from the Crater Hills monitor site; sequences from September 2003 are red……362

3.52. SIRV-like DNA sequences detected over time from the Crater Hills monitor site; sequences from October 2003 are red……363

3.53. SIRV-like protein sequences detected over time from the Crater Hills monitor site; sequences from October 2003 are red……364

3.54. SIRV-like DNA sequences detected over time from the Crater Hills monitor site; sequences from November 2003 are red……365

3.55. SIRV-like protein sequences detected over time from the Crater Hills monitor site; sequences from November 2003 are red……366

3.56. SIRV-like DNA sequences detected over time from the Crater Hills monitor site; sequences from February 2004 are red……...367

LIST OF FIGURES – CONTINUED

Figure Page

3.57. SIRV-like protein sequences detected over time from the Crater Hills monitor site: sequences from February 2004 are red……..368

3.58. SIRV-like DNA sequences detected over time from the Crater Hills monitor site; sequences from April 2004 are red…….369

3.59. SIRV-like protein sequences detected over time from the Crater Hills monitor site; sequences from April 2004 are red…….370

3.60. SIRV-like DNA sequences detected over time from the Crater Hills monitor site; sequences from June 2004 are red……..371

3.61. SIRV-like protein sequences detected over time from the Crater Hills monitor site; sequences from June 2004 are red……..372

4.1. Pictures of the Rabbit Creek monitor site taken in July 2001 and June 2004……...139

4.2. Temperature and pH over time at the Rabbit Creek monitor site and a picture of the site…….140

4.3. Water chemistry values obtained from the Rabbit Creek monitor site over the course of this research…….142

4.4. Iron concentrations over time in the three monitor sites……..143

4.5. Manganese concentrations over time in the three monitor sites……..143

4.6. Average water chemistry values measured over time in the Rabbit Creek monitor site……144

4.7. An overview of the archaeal population detected over time in the Rabbit Creek monitor site……..145

4.8. Rabbit Creek host population based on nearly full-length or full-length 16S rDNA clones……146

4.9. Phylogenetic analysis of environmental 16S rDNA sequences amplified with the 20F/UA1406R primer set from Rabbit Creek……147

LIST OF FIGURES – CONTINUED

Figure Page

4.10. Phylogenetic analysis of environmental 16S rDNA sequences amplified with 348Fa/UA1406R from Rabbit Creek……...149

4.11. An overview of the SSV-like DNA sequences detected in the Rabbit Creek monitor site over time……153

4.12. An overview of the SSV-like protein sequences detected in the Rabbit Creek monitor site over time……..153

4.13. SSV-like 3F/4R DNA and protein sequences detected in Rabbit Creek in July 2002……154

4.14. SSV-like 7F/8R DNA and protein sequences detected from Rabbit Creek in July 2002……155

4.15. SSV-like 3F/4R DNA sequences detected over time in the Rabbit Creek monitor site……..156

4.16. SSV-like 3F/4R protein sequences detected over time in the Rabbit Creek monitor site……..157

4.17. SSV-like 7F/8R DNA sequences collected from the Rabbit Creek monitor site over time……162

4.18. SSV-like 7F/8R protein sequences collected from the Rabbit Creek monitor site over time……163

4.19. A phylogenetic comparison of SSV-like 3F/4R DNA sequences representing functional and non-functional ORFs……...167

4.20. The overall population structure of SSV-like 3F/4R DNA sequences detected over time in Rabbit Creek……..168

4.21. A phylogenetic comparison of SSV-like 3F/4R protein sequences representing functional and non-functional ORFs……...169

4.22. SSV-like 3F/4R sequence abundance from the Rabbit Creek monitor site over time……...172

LIST OF FIGURES – CONTINUED

Figure Page

4.23. An overview of the SIRV-like DNA sequences detected over time in the Rabbit Creek monitor site……..175

4.24. An overview of the SIRV-like protein sequences detected over time in the Rabbit Creek monitor site……..176

4.25. SIRV-like DNA and protein sequences obtained from Rabbit Creek in July 2002……178

4.26. SIRV-like DNA sequences collected from the Rabbit Creek monitor site July 2002 through June 2004……...180

4.27. SIRV-like protein sequences collected from the Rabbit Creek monitor site July 2002 through June 2004……..181

4.28. Sequence abundance plot for SIRV-like sequences from Rabbit Creek over time……187

4.29. SSV-like 3F/4R sequences detected over time from the Rabbit Creek monitor site; sequences from October 2002 are red……..374

4.30. SSV-like 3F/4R sequences detected over time from the Rabbit Creek monitor site; sequences from January 2003 are red……...375

4.31. SSV-like 3F/4R DNA sequences detected over time from the Rabbit Creek monitor site; sequences from August 2003 are red……..376

4.32. SSV-like 3F/4R protein sequences detected over time from the Rabbit Creek monitor site; sequences from August 2003 are red……..377

4.33. SSV-like 3F/4R DNA sequences detected over time from the Rabbit Creek monitor site; sequences from October 2003 are red……378

4.34. SSV-like 3F/4R protein sequences detected over time from the Rabbit Creek monitor site; sequences from October 2003 are red……379

LIST OF FIGURES – CONTINUED

Figure Page

4.35. SSV-like 3F/4R DNA sequences detected over time from the Rabbit Creek monitor site; sequences from November 2003 are red……380

4.36. SSV-like 3F/4R protein sequences detected over time from the Rabbit Creek monitor site; sequences from November 2003 are red……381

4.37. SSV-like 3F/4R DNA sequences detected over time from the Rabbit Creek monitor site; sequences from January 2004 are red……382

4.38. SSV-like 3F/4R protein sequences detected over time from the Rabbit Creek monitor site; sequences from January 2004 are red……383

4.39. SSV-like 3F/4R DNA sequences detected from the Rabbit Creek monitor site from July 2002 through June 2004……384

4.40. SIRV-like sequences detected over time from the Rabbit Creek monitor site; sequences from October 2002 are red……..385

4.41. SIRV-like DNA sequences detected over time from the Rabbit Creek monitor site; sequences from January 2003 are red……...386

4.42. SIRV-like protein sequences detected over time from the Rabbit Creek monitor site; sequences from January 2003 are red……...387

4.43. SIRV-like DNA sequences detected over time from the Rabbit Creek monitor site; sequences from May 2003 are red……388

4.44. SIRV-like protein sequences detected over time from the Rabbit Creek monitor site; sequences from May 2003 are red……389

4.45. SIRV-like DNA sequences detected over time from the Rabbit Creek monitor site; sequences from June 2003 are red……390

4.46. SIRV-like protein sequences detected over time from the Rabbit Creek monitor site; sequences from June 2003 are red……391

LIST OF FIGURES – CONTINUED

Figure Page

4.47. SIRV-like DNA sequences detected over time from the Rabbit Creek monitor site; sequences from July 2003 are red……..392

4.48. SIRV-like DNA sequences detected over time from the Rabbit Creek monitor site; sequences from July 2003 are red……..393

4.49. SIRV-like DNA sequences detected over time from the Rabbit Creek monitor site; sequences from August 2003 are red……394

4.50. SIRV-like protein sequences detected over time from the Rabbit Creek monitor site; sequences from August 2003 are red……395

4.51. SIRV-like DNA sequences detected over time from the Rabbit Creek monitor site; sequences from October 2003 are red……...396

4.52. SIRV-like protein sequences detected over time from the Rabbit Creek monitor site; sequences from October 2003 are red……...397

4.53. SIRV-like DNA sequences detected over time from the Rabbit Creek monitor site; sequences from November 2003 are red……..398

4.54. SIRV-like protein sequences detected over time in the Rabbit Creek monitor site; sequences from November 2003 are red……...399

4.55. SIRV-like DNA sequences detected over time from the Rabbit Creek monitor site; sequences from January 2004 are red……...400

4.56. SIRV-like protein sequences detected over time from the Rabbit Creek monitor site; sequences from January 2004 are red……...401

4.57. SIRV-like DNA sequences detected over time from the Rabbit Creek monitor site; sequences from March 2004……402

4.58. SIRV-like protein sequences detected over time from the Rabbit Creek monitor site; sequences from March 2004……403

4.59. SIRV-like DNA sequences detected over time from the Rabbit Creek monitor site; sequences from June 2004 are red……404

LIST OF FIGURES – CONTINUED

Figure Page

4.60. SIRV-like protein sequences detected over time from the Rabbit Creek monitor site; sequences from June 2004 are red……405

5.1. Temperature and pH of the Ragged Hills monitor site over time and a picture of the site……203

5.2. Water chemistry values obtained from the Ragged Hills monitor site over the course of this research……205

5.3. Chloride and arsenic concentrations over time in the three monitor sites……207

5.4. Average water chemistry values measured over time in the Ragged Hills monitor site……208

5.5. An overview of the archaeal population detected over time in the Ragged Hills monitor site……209

5.6. Ragged Hills host population based on nearly full-length or full-length 16S rDNA clones……211

5.7. Phylogenetic analysis of environmental 16S rDNA genes from the Ragged Hills monitor site……212

5.8. An overview of the SSV-like DNA sequences detected in the Ragged Hills monitor site over time……215

5.9. An overview of the SSV-like protein sequences detected in the Ragged Hills monitor site over time……215

5.10. SSV-like 3F/4R DNA and protein sequences detected in Ragged Hills in October 2002……218

5.11. SSV-like 7F/8R DNA and protein sequences detected from Ragged Hills in October 2002……220

5.12. SSV-like 3F/4R DNA sequences detected over time in the Ragged Hills monitor site……223

LIST OF FIGURES – CONTINUED

Figure Page

5.13. SSV-like 3F/4R protein sequences detected over time in the Ragged Hills monitor site……...224

5.14. SSV-like 7F/8R DNA sequences collected from the Ragged Hills monitor site over time……..226

5.15. SSV-like 7F/8R protein sequences collected from the Ragged Hills monitor site over time……..227

5.16. A phylogenetic comparison of SSV-like 3F/4R DNA sequences representing functional and non-functional ORFs ……..230

5.17. A phylogenetic comparison of SSV-like 3F/4R protein sequences representing functional and non-functional ORFs……...232

5.18. SSV-like 3F/4R sequence abundance from the Ragged Hills monitor site over time……...236

5.19. An overview of the SIRV-like DNA sequences detected over time in the Ragged Hills monitor site……..241

5.20. An overview of the SIRV-like protein sequences detected over time in the Ragged Hills monitor site……..242

5.21. SIRV-like DNA and protein sequences obtained from Ragged Hills in October 2002……..243

5.22. SIRV-like DNA sequences collected from the Ragged Hills monitor site October 2002 through June 2004…….248

5.23. SIRV-like protein sequences collected from the Ragged Hills monitor site October 2002 through June 2004……249

5.24. Sequence abundance plot for SIRV-like sequences from Ragged Hills over time……..253

5.25. SSV-like 3F/4R sequences detected over time from the Ragged Hills monitor site; sequences from November 2002 are red……407

LIST OF FIGURES – CONTINUED

Figure Page

5.26. SSV-like 7F/8R sequences detected over time from the Ragged Hills monitor site; sequences from November 2002 are red……408

5.27. SSV-like 3F/4R DNA sequences detected over time from the Ragged Hills monitor site; sequences from February 2003 are red……...409

5.28. SSV-like 3F/4R protein sequences detected over time from the Ragged Hills monitor site; sequences from February 2003 are red……...410

5.29. SSV-like 3F/4R DNA sequences detected over time from the Ragged Hills monitor site; sequences from April 2003 are red……411

5.30. SSV-like 3F/4R protein sequences detected over time from the Ragged Hills monitor site; sequences from April 2003 are red……412

5.31. SSV-like 3F/4R DNA sequences detected over time from the Ragged Hills monitor site; sequences from May 6, 2003 are red……..413

5.32. SSV-like 3F/4R protein sequences detected over time from the Ragged Hills monitor site; sequences from May 6, 2003 are red……..414

5.33. SSV-like 3F/4R DNA sequences detected over time from the Ragged Hills monitor site; sequences from November 2003 are red……415

5.34. SSV-like 3F/4R protein sequences detected over time from the Ragged Hills monitor site; sequences from November 2003 are red……416

5.35. SSV-like 3F/4R DNA sequences detected in the Ragged Hills monitor site from October 2002 through June 2004……417

5.36. SIRV-like sequences detected over time from the Ragged Hills monitor site; sequences from November 2002 are red……418

LIST OF FIGURES – CONTINUED

Figure Page

5.37. SIRV-like DNA sequences detected over time from the Ragged Hills monitor site; sequences from February 2003 are red……...419

5.38. SIRV-like protein sequences detected over time from the Ragged Hills monitor site; sequences from February 2003 are red……...420

5.39. SIRV-like DNA sequences detected over time from the Ragged Hills monitor site; sequences from April 2003 are red……421

5.40. SIRV-like protein sequences detected over time from the Ragged Hills monitor site; sequences from April 2003 are red……422

5.41. SIRV-like DNA sequences detected over time from the Ragged Hills monitor site; sequences from May 6, 2003 are red……..423

5.42. SIRV-like protein sequences detected over time from the Ragged Hills monitor site; sequences from May 6, 2003 are red……..424

5.43. SIRV-like DNA sequences detected over time from the Ragged Hills monitor site; sequences from May 13, 2003 are red……425

5.44. SIRV-like protein sequences detected over time from the Ragged Hills monitor site; sequences from May 13, 2003 are red……426

5.45. SIRV-like DNA sequences detected over time from the Ragged Hills monitor site; sequences from June 2003 are red……..427

5.46. SIRV-like protein sequences detected over time from the Ragged Hills monitor site; sequences from June 2003 are red……..428

5.47. SIRV-like DNA sequences detected over time from the Ragged Hills monitor site; sequences from July 2003 are red……..429

5.48. SIRV-like protein sequences detected over time from the Ragged Hills monitor site; sequences from July 2003 are red……...430

5.49. SIRV-like DNA sequences detected over time from the Ragged Hills monitor site; sequences from August 4, 2003 are red……431

LIST OF FIGURES – CONTINUED

Figure Page

5.50. SIRV-like protein sequences detected over time from the Ragged Hills monitor site; sequences from August 4, 2003 are red…….432

5.51. SIRV-like DNA sequences detected over time from the Ragged Hills monitor site; sequences from August 26, 2003 are red……...433

5.52. SIRV-like protein sequences detected over time from the Ragged Hills monitor site; sequences from August 26, 2003 are red……...434

5.53. SIRV-like DNA sequences detected over time from the Ragged Hills monitor site; sequences from October 2003 are red……435

5.54. SIRV-like protein sequences detected over time from the Ragged Hills monitor site; sequences from October 2003 are red……436

5.55. SIRV-like DNA sequences detected over time from the Ragged Hills monitor site; sequences from November 2003 are red……437

5.56. SIRV-like protein sequences detected over time from the Ragged Hills monitor site; sequences from November 2003 are red……438

5.57. SIRV-like DNA sequences detected over time from the Ragged Hills monitor site; sequences from April 2004 are red…….439

5.58. SIRV-like protein sequences detected over time from the Ragged Hills monitor site; sequences from April 2004 are red…….440

5.59. SIRV-like DNA sequences detected over time from the Ragged Hills monitor site; sequences from June 2004 are red……..441

5.60. SIRV-like protein sequences detected over time from the Ragged Hills monitor site; sequences from June 2004 are red……..442

6.1. A comparison of the average water chemistry values measured in the three monitor sites over time……..268

6.2. The water chemistry values measured at the Crater Hills, Rabbit Creek, and Ragged Hills monitor sites…….271

LIST OF FIGURES – CONTINUED

Figure Page

6.3. A comparison of the archaeal populations detected within each monitor site……...278

6.4. SSV-like 3F/4R DNA sequences obtained from the three monitor sites over time……...285

6.5. The comparison of the overall SSV-like 3F/4R population structures within each monitor site……...287

6.6. A species richness curve for the SSV-like viruses detected over time in all three monitor sites……...289

6.7. SSV-like distribution at all three monitor sites over time……290

6.8. SSV-like 3F/4R protein sequences obtained from the three monitor sites over time……...296

6.9. SSV-like 7F/8R protein sequences detected in the Rabbit Creek and Ragged Hills monitor sites……...300

6.10. A species richness curve for the SIRV-like viruses detected over time in the three monitor sites……..304

6.11. SIRV-like distribution at all three monitor sites over time……..305

6.12. SIRV-like protein sequences detected in all three monitor sites over time……...308

6.13. A total community dissimilarity index for both viruses sampled over time within the three monitor sites……...309

ABSTRACT

Few viruses from Archaea have been described and only a limited number have been characterized in molecular detail. Most hyperthermophilic viruses isolated from the archaeal sub-domain Crenarchaeota belong to the *Sulfolobales* family. We have focused on the isolation and analysis of viruses replicating in *Sulfolobus*. *Sulfolobus* is an acidophilic hyperthermophile that lives optimally at 80°C and pH 3. The SSVs (*Sulfolobus* spindle-shaped viruses) are the viruses most commonly isolated from thermal acidic areas around the world. SSVs are dsDNA spindle-shaped viruses with genomes of approximately 15kb. They are also the predominant virus isolated from samples collected in Yellowstone National Park (YNP). The SIRVs *(Sulfolobus islandicus* rod-shaped viruses) are also commonly isolated from YNP. These viruses have linear dsDNA genomes of approximately 33kb.

I have used these two viruses to investigate the interaction of time and viral population dynamics. For a comparison between geographically isolated acidic hot springs, we chose to study three sites in YNP. These three sites were monitored continuously for changes in virus populations, archaeal host populations, and water chemistry. Each monitor site was chemically unique, but the water chemistry did not significantly change over time. The archaeal population in each of the three hot springs was unique and was found to slowly change over time. We have discovered that although there can be a substantial change in viral population structures in between the collection of samples, it does not seem to be correlated with a changing environmental parameter. There also appears to be a strong culturing bias present in the SSVs when we compare the virus population monitored directly from the environment to the viruses isolated from an enrichment culture established from a particular hot spring. A long term goal is to contribute to the understanding of the relationships between time and resident viral populations within hot spring environments. We hope this work will contribute a further understanding of life in extreme environments.

CHAPTER 1

VIRUSES OF ARCHAEA: BACKGROUND AND OBJECTIVES OF THIS RESEARCH

Introduction and overview

The overall objective of this research is to examine in detail the interplay of time and the environment on viral populations. Specifically, this research is directed towards understanding viral dynamics in high temperature acidic environments present in Yellowstone National Park (YNP). This involves the study of viruses from environments that are poorly understood and provides a unique opportunity to contribute to our understanding of these unusual environments. Currently, there is little understanding as to how organisms function in these extreme environments. Therefore, the analysis of viruses from these environments will provide an important component in understanding how these communities function. In addition, it is likely that viral populations will be sensitive sensors of change in the environment. It is envisioned that this research will contribute to our understanding of viral diversity in high temperature acidic environments and contribute to our knowledge of life in extreme environments.

This research is the first attempt at examining viral dynamics in high temperature acidic environments. In fact, to our knowledge, this is the first example of examining viral dynamics in any natural environment. The initial sampling for viruses and host sequences in thermal areas in YNP is the first time this has been conducted at this level

over time. This may result in a new understanding concerning how environmental changes affect microbial communities.

Research concerning diversity of viruses has been conducted on viruses from the marine environment, however none of this research has been carried out continuously over time. Breitbart *et. al.* (2002) studied the diversity of marine viruses off the coasts of La Jolla and San Diego, California, however only one sample was collected at each location. Their results revealed significant diversity in the two locations. Most (>65%) of the sequences detected were not significantly similar to previously reported sequences (Breitbart *et. al.* 2002). This study revealed significant diversity in the two sampling points, but did not show how viral diversity may change over time.

Studies on influenza virus have been performed over time to investigate how the virus changes (Ferguson *et. al.* 2003). This research followed changes that occur in the surface haemagglutinin glycoprotein (the main antibody target of the human immune system). This protein is under strong selection by the immune system and different strains of influenza were found to evolve at different rates. The diversity of the viral population was found to be low due to density-dependent constraints on the population size. The researchers believe that if the population size were not constrained, the amount of viral diversity would rapidly increase (Ferguson *et. al.* 2003). However, this study investigated changes in the influenza virus over many years and from varying sampling sources. In addition, these studies involved analysis of cultured isolates, which may have strongly biased the results. Our sampling scheme involves collecting samples from the same sites monthly to detect changes in the virus populations. A comparison of the

strong selection pressures in a pathogenic mammalian virus system to the viral-microbial community analysis examined in this study may prove insightful.

Recently, the diversity and annual dynamics of a group of coastal bacterioplankton was measured (Thompson *et. al.* 2005). This group of organisms was found to consist of at least a thousand distinct genotypes. Individual genotypes rarely recurred in samples, and allelic distribution did not show spatial or temporal separation. It will be interesting to determine whether such extensive genomic variation is a general characteristic of microbial populations. We plan to assess the diversity of archaeal viruses over time and determine whether radical changes are apparent in viruses living in extreme environments.

A study published on the diversity of halo-viruses mentions the use of laboratory strains to isolate viruses (Dyall-Smith *et. al.* 2003). The diversity of the halo-viruses may be greatly underrepresented because of the culturing capability of hosts and whether the isolated viruses can infect the laboratory-cultured organisms. The diversity of cyano-viruses has also been studied using laboratory-cultured organisms (Waterbury and Valois 1993; Wilson *et. al.* 1993). These previous reports study the diversity of viruses from an environment, but attempt to do so using laboratory-cultured isolates. The true diversity of the viruses may not be completely understood because of limitations of the viruses infecting the laboratory strains. In this research, we will determine whether a culturing bias exists for one hyperthermophilic virus isolated from thermal acidic areas.

Our research differs significantly from these previous reports. First, the environments in YNP are considerably less understood than are the marine or human

environments. In contrast to marine environments and humans, we suspect the YNP environments will be dominated by archaeal organisms (detailed below). Second, the organisms in the thermal environments in YNP are poorly understood. Third, our measurements of diversity will take place over time to determine what changes occur in the viral community over changing environmental parameters (i.e., temperature and pH). Our examination of viral dynamics will be monitored at a much higher level of resolution both in terms of time and virus number. We will also investigate the effect of culturing the viruses isolated from YNP thermal environments as compared to a culture-independent analysis.

To our knowledge, research of this type has never been conducted in the environments of YNP or anywhere else. Three thermal acidic sites in YNP were chosen and monitored monthly for changes in viral populations within a site and between sites, and to predict correlation with changes in the environment. Over the course of two years, the diversity and identity of potential host populations were studied, and the diversity of the viral populations were compared to previous time points in the same site and other sites to determine how and if change varies in virus population depending on time and location. The water chemistry of the three hot springs was also monitored on a monthly basis so that environmental change could be determined. The overall goal of this project was to examine the interplay of the dynamics of viral populations, host populations, and water chemistry in multiple YNP thermal sites.

Major questions and hypothesis

There are four major questions that this research proposes to address:

1. How do virus populations change over time within a specific YNP thermal feature?
2. Do different viruses in the same hot spring undergo the same dynamics or do they respond differently?
3. Do viruses in different thermal features have the same dynamics?
4. Can changes in viral populations be correlated to changes in host populations or water chemistry?

We hypothesize that changes in viral population structures will be a sensitive indicator of change in the environment.

Background: Archaea

Prior to the mid-1970s, life on Earth was classified based on complex morphologies and the fossil record. At this time, life was characterized into five kingdoms: animals, plants, bacteria, protists, and fungi. Microbial morphologies, unlike their multicellular counterparts, were considered too simple or un-interpretable to serve as a valid characteristic upon which to base phylogeny. Nucleic acid sequencing has changed all of that. The plethora of information provided by nucleic acid sequences has led to a different system of classifying organisms; therefore, classical phenotypical criteria are being supplemented with molecular criteria. Nucleic acid sequences can reveal evolutionary relationships that classical phenotypical criteria cannot. Phylogenetic

analysis divides life on this planet into three domains of life, the Bacteria, the Eukarya, and the Archaea, each containing two or more kingdoms (figure 1.1) (Woese *et. al.* 1990). As opposed to the Eukarya and Bacteria, little information is known about the Archaea, their role in the environment, and their associated viruses. There are currently 16 published complete genomes of Archaea (Fukui *et. al.* 2005; Futterer *et. al.* 2004; Galagan *et. al.* 2002; Deppenmeier *et. al.* 2002; Slesarev *et. al.* 2002; Fitz-Gibbon *et. al.* 2002; Robb *et. al.* 2001; Kawarabayasi *et. al.* 2001, 1999, and 1998; She *et. al.* 2001; Kawashima *et. al.* 2000; Ng *et. al.* 2000; Smith *et. al.* 1997; Klenk *et. al.* 1997; Bult *et. al.* 1996), and three genomes not yet published (*Nanoarchaeum equitans* Kin4-M, *Pyrococcus abyssi* GE5, and *Thermoplasma acidophilum* DSM 1728). These archaeal genomes are raising many questions about archaeal gene structure, expression, and function. Clearly, much of archaeal biochemistry remains to be elucidated.

The Archaeal domain is divided into two major kingdoms, the Euryarchaeota and the Crenarchaeota, and two proposed kingdoms, the Korarchaeota and the Nanoarchaeota (figure 1.2). These groups are defined primarily based on 16S rDNA sequences (Garrity 2001; Woese *et. al.* 1990).

Korarchaeota and Nanoarchaeota

The Korarchaeota branch near the root of the archaeal tree; therefore, they may display novel biological properties relevant to our understanding of ancient organisms (Barns 1996). This group of organisms is represented solely by 16S rRNA sequences, as cultures have not been obtained for any of these organisms. Since the original discovery of Korarchaeota sequences in Yellowstone (Barns 1996), they have been detected from

numerous environments around the world (Nercessian 2003; Hjorleifsdottir 2001; Marteinsson 2001).

Figure 1.1. A tree of life showing the three domains of life.

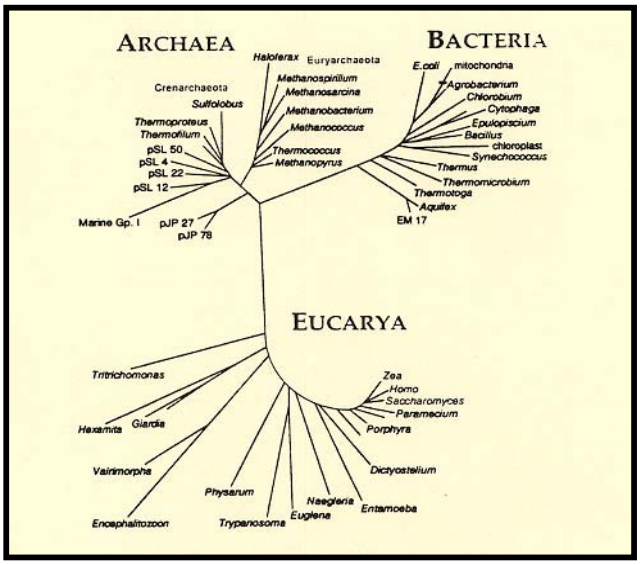
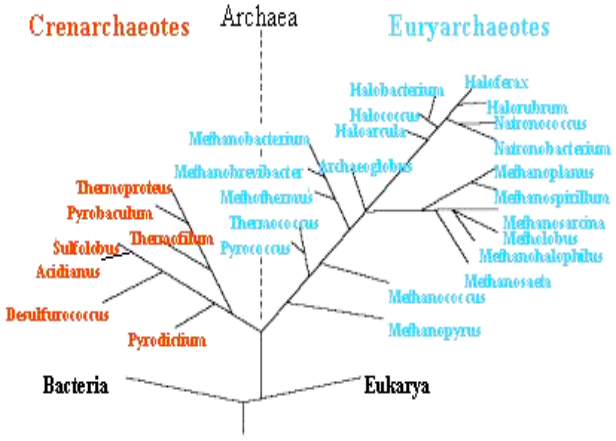


Figure 1.2. Two kingdoms of Archaea: Euryarchaeota and Crenarchaeota.



The Nanoarchaeota are likely small symbiotic microorganisms that are larger than some known viruses but smaller than any other described microorganism. Currently, this novel kingdom is represented by one species, "*Nanoarchaeum equitans*". This organism was cultivated from a submarine hot vent (Huber *et. al.* 2002). "*N. equitans*" is a nano-sized symbiont that grows attached to the surface of a new *Igniococcus* species. In a recent survey of high temperature environments, Nanoarchaea were found to be widely distributed (Hohn *et. al.* 2002). Small cell size and reduced genome size are common features observed in many symbiotic bacteria, but had not been described for archaea or hyperthermophiles. The high growth temperature and anaerobic growth conditions of Nanoarchaeota correlate with early environmental conditions. Therefore, this organism may represent a primitive form of microbial life.

Euryarchaeota

The Euryarchaeota comprise many phylogenetically diverse groups, but mainly consist of the halophiles, organisms that tolerate high salt environments, and the methanogens, methane-producing microorganisms (Garrity 2001). The extremely halophilic archaea are a diverse group of organisms that inhabit high saline environments, such as solar salt evaporation ponds and natural salt lakes, or artificial saline habitats, such as the surfaces of heavily salted foods. All extremely halophilic archaea are chemo-organotrophs and most species are obligate aerobes. Methanogens produce methane as part of their energy metabolism. The diversity of habitats from which methanogens can be isolated from is quite large, for example marsh, swamp, and lake sediments, rumens of cattle, sheep, and deer, large intestine of humans, geothermal sources, and endosymbionts

of various anaerobic protozoa (Whitman 1985). Methanogens are strict anaerobes and most grow at mesophilic temperatures, although species have been isolated from temperatures up to 110°C (Burggraf *et. al.* 1991).

Most Euryarchaeota isolated to date do not live at high temperatures. However, there are some hyperthermophilic Euryarchaeota. These include members of the orders *Thermococcales* and *Archaeoglobales*, and the organisms *Methanopyrus* and *Methanothermus*. Interestingly, these hyperthermophilic euryarchaeotes branch near the root on the archaeal tree (figure 1.1).

Crenarchaeota

Previous research indicates that crenarchaeotes are detected in acidic hot springs worldwide (Wiedenheft *et. al.* 2004; Whitaker *et. al.* 2003; Bettstetter *et. al.* 2003; Snyder *et. al.* 2003; Rachel *et. al.* 2002; Rice *et. al.* 2001; Zillig *et. al.* 1998, 1996, and 1994). The Crenarchaeota contain organisms at both ends of the temperature spectrum (Snyder *et. al.* 2003). Most known crenarchaeotes are hyperthermophiles that grow optimally above 80ºC. Hyperthermophilic crenarchaeotes have been isolated from both terrestrial and marine environments. Most known hyperthermophilic crenarchaeotes are obligate anaerobes and their metabolism is either chemo-organotrophic or chemo-lithotrophic. The most studied orders from terrestrial volcanic habitats are the aerobic, sulfur-oxidizers *Sulfolobales*, which includes *Sulfolobus* and *Acidianus*, and the anaerobic, sulfur-reducers *Thermoproteales*. The most thermophilic of all known archaea has been isolated from submarine volcanic habitats, and belongs to the order

Desulfurococcales. Some of these organisms have optimal growth temperatures above 100°C.

Marine and cold-dwelling crenarchaeotes have been observed from many non-thermal environments (Schleper *et. al.* 1997; Preston *et. al.* 1996; DeLong *et. al.* 1994 and 1992; Fuhrman *et. al.* 1993). These organisms are planktonic and occur in large numbers.

Archaea in Yellowstone National Park

Archaea have long been studied in YNP and are well represented in the thermal environments found within YNP (Spear *et. al.* 2005; Wiedenheft *et. al.* 2004; Whitaker *et. al.* 2003; Snyder *et. al.* 2003; Bettstetter *et. al.* 2003; Blank *et. al.* 2002; Rice *et. al.* 2001; Reysenbach *et. al.* 2000; Huber *et. al.* 1998; Barns *et. al.* 1996 and 1994). The abundance of thermal features (>10,000) in YNP makes it an ideal environment in which to study archaea, and more specifically, hyperthermophilic archaeons. The diversity of archaea in YNP has been studied largely in part due to Norman Pace and colleagues (Spear *et. al.* 2005; Blank *et. al.* 2002; Reysenbach *et. al.* 2000; Barns *et. al.* 1996 and 1994). A recent study investigated the community composition of hot springs (>70°C) in YNP (Spear 2005). In order to characterize the microbial population, a variety of universal 16S PCR primers were utilized. Based on 16S rRNA gene analysis, it was discovered that archaeal organisms comprised a small fraction of the total community (~2%).

Using a cultivation-independent protocol, many previously unknown organisms have been discovered. For example, Barns *et. al.* (1996) discovered 19 novel archaeal

16S rRNA sequences from Obsidian Pool, a high temperature (74-93°C), near neutral pH hot spring in YNP. Among these sequences some, termed Korarchaeota, were found that branch prior to the diversification of all known archaea in phylogenetic analyses. Another study carried out in Obsidian Pool revealed that archaea may not be the dominant organisms in hydrothermal pools (Hugenholtz *et. al.* 1998). Hyperthermophilic microbial diversity was also measured in numerous thermal silica-depositing springs around YNP; this research also indicated that archaea were rare in these communities (Blank *et. al.* 2002).

Most of the information concerning Archaea in high temperature environments (≥75°C) comes from studies of high temperature, neutral pH hot springs (Reysenbach *et. al.* 2000; Barns *et. al.* 1996). Reysenbach *et. al.* (2000) found that archaea form only a small part of the total biomass of the community in a high temperature (83°C) and neutral pH hot spring, Calcite Springs. However, it was determined the archaeal diversity exceeded that of the bacterial diversity. There is little information concerning archaeal organisms in acidic (pH 1-4), high temperature (≥75°C) hot springs within YNP. Most organisms that have been identified in acidic hot springs within YNP belong to the family *Sulfolobales* (Snyder *et. al.* 2003; Rachel *et. al.* 2002; Rice *et. al.* 2001); however, most of that information is based only on 16S rDNA sequences. No research has been conducted on the dynamics of archaeal organisms within the acidic hot springs in YNP.

We are primarily interested in the thermophilic crenarchaeon, *Sulfolobus*. Previous research has shown this organism is readily cultured from numerous acidic hot springs in YNP (Rice *et. al.* 2001). *Sulfolobus* was first isolated and characterized from a

hot spring in YNP (Brock *et. al.* 1972; R. Bizzoco, personal communication). Since then, *Sulfolobus* has been isolated from acidic hot springs around the world. *Sulfolobus* grows optimally at 80°C and pH 3 (Garrity 2001), although it can be found in environments with a temperature range from 60-90°C and a pH range from 1-5 (Rice *et. al.* 2001). Many choose to study *Sulfolobus* because of its ease to grow in the laboratory. *Sulfolobus* is an aerobic heterotroph that oxidizes sulfur to produce sulfuric acid.

The diversity of *Sulfolobus* within YNP and around the world has recently been investigated by Whitaker *et. al.* (2003) who compared numerous *Sulfolobus* isolates from volcanic areas around the world (Lassen National Park, YNP, Iceland, Italy, and Kamchatka, Russia). The population of *Sulfolobus* within YNP was diverse, however, it was determined that gene flow among *Sulfolobus* populations around the world is quite limited (Whitaker 2003). This study suggests that a long period of geographical isolation has resulted in the divergence of the studied *Sulfolobus* strains.

Background: Archaeal viruses

To date, we have a rudimentary knowledge of viruses that replicate in archaeal hosts. An understanding of archaeal viruses may prove useful for unraveling archaeal biochemistry. The recent discovery of many novel archaeal viruses, especially among members of the extremely thermophilic Crenarchaeota, is likely to lead to a more complete understanding of not only archaea, but also the biochemical adaptations required for life in extreme environments, and to new insights into both host and virus evolution.

The detailed analysis of viruses often leads to a better understanding of a particular cellular environment because viruses are molecular parasites of the cells in which they replicate. Since all viruses are dependent on the host cell's biochemical machinery for their replication, they are highly adapted to the unique characteristics of their host's cellular environment. The relative ease of both genetic and biochemical analyses of viruses as compared to their more complex hosts makes them very attractive systems for unraveling unique aspects of archaea.

When compared to more than 4,000 different viruses isolated from Eukarya and Bacteria, relatively few viruses have been isolated from Archaea. Approximately 39 archaeal viruses or virus-like particles (VLPs) have been identified, but few have been characterized in detail (Snyder *et. al*. 2003). The lack of archaeal viruses described to date is most likely due to the limited effort in isolating archaeal viruses and not a reflection of any biological restriction limiting viral diversity and number in the archaea.

The isolation of archaeal viruses from extreme environments is often problematic. Unlike detection of abundant viruses from marine environments (Brietbart *et. al*. 2002; Wommack and Colwell 2000), direct isolation of crenarchaeal viruses by filtration has been difficult (Rice *et. al*. 2001). A reason for this failure may be because these viruses have evolved to minimize the time they exist free of their hosts in hostile environments. Another potential limitation to isolation is that few potential archaeal hosts have been cultivated or those that have been cultured require conditions that are not conducive to the isolation of viruses. True plaque assays for the isolation of viruses from high temperature crenarachaeota have been difficult to develop because of the difficulty of

growing host lawns on solid media at temperatures >75°C, and the fact that the majority of creanarchaeal viruses described are not lytic. Successful plaque-like growth-inhibition assays for the detection of *Sulfolobus* viruses have been developed (Zillig *et. al.* 1994). However, not all growth inhibition zones are the result of virus replication. For example, anti-archaeal compounds (sulfolobicins) produced by some *Sulfolobus* strains limit the growth of related *Sulfolobus* strains (Prangishvili *et. al.* 2000). This inhibition of growth sometimes appears indistinguishable from virus-induced growth inhibition zones. In addition, the fact that many of the archaeal viruses are unrelated, with respect to morphology, to previously described viruses makes it more difficult to depend on established protocols. Similarly, unlike ribosomal RNA gene sequences, there are no known universal nucleic acid signatures for viruses, making identification based on nucleic acid hybridization impossible. Given these limitations, it is no surprise that most archaeal viruses have been discovered by large-scale random screening, screening potential hosts that can be cultured, or by direct visualization with an electron microscope (Snyder *et. al.* 2003).

Euryarcheaotal viruses

Approximately 27 viruses and VLPs have been isolated from euryarchaeal hosts (table 1.1). All euryarchaeal viruses described to date are DNA viruses (excluding ϕCh1 isolated from *Natrialba* species which may contain an RNA component (Klein *et. al.* 2002; Witte *et. al.* 1997)). Known euryarchaeal viruses possess circular or linear double-stranded (ds) genomes that range in size from 4.4-230 kilobase (kb) pairs. The complete or nearly complete genome sequence has been determined for five of these viruses

(Dyall-Smith *et. al.* 2003; Klein *et. al.* 2002; Tang *et. al.* 2002; Luo *et. al.* 2001; Pfister *et. al.* 1998). The predicted gene products encoded by these genomes have few similarities to other known sequences in the public databases.

Table 1.1. Viruses of the Euryarchaeotes.

Virus	*Host*	*Head/tail (nm)*	*Genome Size*	*References*
ϕH	*Halobacterium salinarium*	64/170	linear 59kb DNA	Stolt 1995 Zillig 1988
ϕN	*H. salinarium*	55/80	linear 56kb DNA	Wenke 1988
Hs1	*H. salinarium*	50/120	N.D.[a]	Torsvik 1974;1988
Ja1	*H. salinarium*	90/150	linear 230kb DNA	Wais 1975
Hh1	*H. salinarium*	60/100	37kb dsDNA	Pauling 1982
Hh3	*H. salinarium*	75/50	29kb dsDNA	Pauling 1982
S45	*H. salinarium*	40/70	N.D.	Daniels 1984
HF1	*Halobacterium* sp.	58/94	linear 80kb DNA	Dyall-Smith 2003 Nuttall 1993
HF2	*Halorubrum coriense*	58/94	linear 77kb DNA	Nuttall 1993 Tang 2002
His1	*Haloarcula hispanica*	74x44/7[b]	linear 14kb DNA	Bath 1998
His2	*Haloarcula hispanica*	62x69/0	16.2kb DNA	Dyall-Smith 2003
SH1	*Haloarcula hispanica*	55[c]	27kb DNA	Dyall-Smith 2003
ϕCh1	*Natrialba magadii*	70/130	linear 55kb DNA Small RNAs	Klein 2002 Witte 1997
ψM1	*Methanothermobacter marbugensis*	55/210	linear 30kb DNA	Meile 1989
ψM2	*M. marbugensis*	55/210	linear 26kb DNA	Jordan 1989
ϕF1	*M. thermoautotrophicus*	70/160	linear 85kb DNA	Nolling 1993
ϕF3	*M. thermoautotrophicus*	55/230	circular 36kb DNA	Nolling 1993
PG	*Methanobrevibacter smithii*	N.D.	~50kb dsDNA	Bertani 1986
PSM1	*Methanobrevibacter smithii*	N.D.	35kb dsDNA	Knox 1986
VTA	*Methanococcus voltae*	40/61	4.4kb DNA	Eiserling 1999
PAV1	*Pyrococcus abyssi*	120/80[b]	circular 18kb dsDNA	Geslin 2003a

[a]N.D.=Not determined; [b]dimensions of assymetrical head; [c]spherical virus particle; table modified from Snyder (2003).

Most characterized viruses from the euryarchaeota have typical "head and tail" morphology, similar to that of most bacterial viruses, and belong to the previously recognized viral families *Myoviridae* and *Siphoviridae* (table 1.1). The best studied of the euryarchaeotal viruses are the ϕH from *Halobacterium salinarium* (Zillig *et. al.* 1988), HF2 from *Haloferax volcanii* (Tang *et. al.* 2002), and ϕCh1 from the halo-

alkaliphile *Natrialba magadii* (Klein *et. al*. 2002). The *Methanothermobacter marburgensis* virus ψM1 and its variant ψM2 have also been extensively studied (Pfister *et. al*. 1998).

Halophilic viruses. Halo-virus ϕH is the most extensively studied of the viruses from extreme halophiles (Zillig *et. al*. 1988). Halophage ϕH belongs to the family *Myoviridae* and has a characteristic icosahedral head and a contractile tail with tail fibers. It has a highly variable genome due to recombination with its host (Pfeifer 1987). Halo-virus ϕH is temperate and its genome is stably maintained as a circular plasmid and does not integrate into the host chromosome (Schnabel *et. al*. 1984).

The haloviruses HF1 and HF2 were originally discovered as part of an Australian salterns survey for lytic viruses (Nuttall and Dyall-Smith 1993). These viruses are currently unclassified, but are likely to be placed in the *Siphoviridae* family. Like most other halophilic viruses, they have a typical head and tail morphology, and a linear dsDNA genome. The two viruses are morphologically identical, and have similar genome sizes and protein composition. They differ in host ranges, plaque morphologies, and sensitivity to chloroform (Nuttall and Dyall-Smith 1993).

The virus ϕCh1 was originally isolated from *Natrialba magadii* from the soda lake Lake Magadii in Kenya (Klein *et. al*. 2000). Unlike other known viruses of archaea, it possesses both an RNA and DNA component of its genome (Witte *et. al*. 1997). The DNA genome is similar to other halophages; it is double-stranded and linear, and was found to contain sequences similar to the ϕH genome (Klein *et. al*. 2002). Total nucleic acid extracted from ϕCh1 showed two bands on an agarose gel, one DNA band and one

RNA band (Witte *et. al*. 1997). Hybridization of the RNA fragments to total DNA from *N. magadii* and ϕCh1 showed that the virion-associated RNA is host encoded.

Methanogenic viruses. There are considerably fewer viruses described from methanogens than the halophiles or even the crenarchaeotes. It is unclear as to why this is the case, but it could be due to the difficulty involved with culturing methanogens. The best studied of the methanogenic viruses are ψM1 and its variant ψM2 that infect *Methanothermobacter marburgensis* (Stedman *et. al*. 2005). ψM1 is a lytic virus that was isolated from an anaerobic sludge at a temperature of 55-60°C (Meile *et. al*. 1989). The virus has an icosahedral head and a flexible tail. Its genome is approximately 30.4kb of double-stranded linear DNA. The virus ψM2, a variant of ψM1, has a completely sequenced genome of 26kb (Pfister *et. al*. 1998).

Other viruses from methanogens are less characterized. A small virus-like particle was observed in cultures of *Methanococcus voltae*, but it was very low titer (Eiserling *et. al*. 1999). A virus-like particle was found in another culture of *Methanococcus voltae* strain A3 (Wood *et. al*. 1989). It has morphology very similar to the SSV1 virus from *Sulfolobus* hosts (described below). The double-stranded DNA genome is 23kb, and was found to integrate into the host genome (Wood *et. al*. 1989). Two lytic viruses of *Methanothermobacter thermoautotropicus*, ϕF1 and ϕF3 were discovered from a sludge reactor at 55ºC (Nolling *et. al*. 1993). They both have typical head and tail morphologies, and double-stranded DNA genomes. ϕF1 has a much wider host range than does ϕF3.

Thermophilic euryarchaeotal viruses. The *Thermococcales*, including *Pyrococcus*, comprise some of the most thermophilic organisms known to date. Virus-like particles have been observed in *Pyrococcus* enrichment cultures from samples collected from deep-sea hydrothermal vents (Geslin *et. al.* 2003b). The complete sequence has been obtained for virus-like morphology isolated from *Pyrococcus abyssi* (Geslin *et. al.* 2003a). This virus-like particle, named PAV1, is lemon-shaped (120nm x 80nm) and resembles the virus SSV1, isolated from *Sulfolobus shibatae* (see below). It contains circular dsDNA of 18kb, which is present as a non-integrated form. An integrated form of the virus could not be detected in the host chromosome, and virus production was not induced by UV irradiation or mitomycin C.

Crenarchaeotal viruses

Like the viruses infecting euryarchaeotes, known crenarchaeotal viruses possess DNA genomes. Few of these viruses have been fully characterized. The first viruses isolated from the crenarchaeota infected extremely thermophilic members of the genera *Thermoproteus* (Janekovic *et. al.* 1983), *Sulfolobus* (Zillig *et. al.* 1994), and *Acidianus* (Zillig *et. al.* 1994). Unlike the viruses isolated from euryarchaeota, none of these viruses have been shown to possess head and tail morphologies (table 1.2). Instead, these viruses have a range of morphologies from helical stiff rods to irregular spindle-shapes. The complete or nearly complete genome sequences have been determined for several of these viruses (Rice *et. al.* 2004; Wiedenheft *et. al.* 2004; Haring *et. al.* 2004; Stedman *et. al.* 2003; Bettstetter *et. al.* 2003; Peng *et. al.* 2001; Arnold *et. al.* 2000b; Prangishvili *et. al.* 1999; Palm *et. al.* 1991; Neumann and Zillig 1990). Analysis of their viral genomes

indicates that they are unrelated to previously known viruses. The unique morphology, unusual genome structures, and genome sequences required the formation of three new viral families, *Fuselloviridae*, *Rudiviridae*, and *Lipothrixviridae*. It is likely that further analysis of new crenarchaeotal viruses will lead to the recognition of new viral families.

Thermoproteus viruses. Originally found in a high temperature mud hole in Iceland, *Thermoproteus tenax* is a chemolithotrophic, strictly anaerobic crenarchaeote that grows optimally at 88ºC (Zillig *et. al.* 1981). *T. tenax* has been found to harbor four viruses, TTV1, TTV2, TTV3, and TTV4 (table 1.2) (Janekovic *et. al.* 1983). TTV1, TTV2, and TTV3 have linear double-stranded DNA genomes that are unrelated according to hybridization analysis (Stedman *et. al.* 2005). Each of these viruses has a different protein composition (Arnold *et. al.* 1999).

Table 1.2. Viruses of the Crenarchaeota.

Virus	*Host*	*Dimensions (nm)*	*Genome Size*	*References*
SSV1	*Sulfolobus shibatae*	spindle 60/90	ccc 15.5kb dsDNA	Martin 1984
SSV2	*S. islandicus*	spindle 55/80	ccc 14.8kb dsDNA	Arnold 1999 Stedman 2003
SSV3	*S. islandicus*	spindle 55/80	ccc 15kb dsDNA	Zillig 1998
SSV-RH	*S. solfataricus*	spindle 60/90	ccc 16kb dsDNA	Wiedenheft 2004
SSV-K1	*Sulfolobus* sp.	spindle 60/90	ccc 17.8 kb dsDNA	Wiedenheft 2004
SIRV1	*S. islandicus*	stiff rod 780/23	linear 32kb dsDNA	Prangishvili 1999
SIRV2	*S. islandicus*	stiff rod 900/23	linear 35kb dsDNA	Prangishvili 1999
SIRV-Y1	*S. solfataricus*	stiff rod 900/25	linear 33kb dsDNA	Rice 2001
SIFV	*S. islandicus*	filamentous 1950/24	linear 41kb dsDNA	Arnold 2000b
SNDV	*S. neozealandicus*	droplet 110-185/95-75	ccc 20kb dsDNA	Arnold 2000a
STIV	*Sulfolobus* sp.	icosahedral 70	ccc 19kb dsDNA	Rice 2004
TTV1	*Thermoproteus tenax*	flexible rod 40/400	linear 15.9 dsDNA	Janekovic 1983
TTV2	*T. tenax*	filamentous 20/1250	linear 16kb dsDNA	Janekovic 1983
TTV3	*T. tenax*	filamentous 30/2500	linear 27kb dsDNA	Janekovic 1983
TTV4	*T. tenax*	stiff rod 30/500	linear 17kb dsDNA	Zillig 1988
PSV	*Pyrobaculum* sp. *Thermoproteus* sp.	spherical 100nm	linear 28kb dsDNA	Haring 2004
DAFV	*Acidianus ambivalens*	filamentous 2200/27	linear 56kb dsDNA	Zillig 1994
AFV1	*Acidianus* sp.	filamentous 900/24	linear 21kb dsDNA	Bettstetter 2003

**Table modified from Snyder (2003).

The best studied of these viruses, TTV1, belongs to the *Lipothrixviridae* family. The virions are flexible rods with a central core enclosed by an envelope derived from host lipids (Zillig *et. al.* 1988). TTV1 persists in a carrier state in its host and lysis is induced when sulfur in the culture is consumed (Janekovic *et. al.* 1983). TTV1 has a 16kb linear dsDNA genome that is bound by two highly basic proteins (Arnold *et. al.* 1999).

TTV2 and TTV3 also belong to the *Lipothrixviridae* viral family. TTV2 and TTV3 are morphologically similar to each other. Both are flexible filaments composed of linear DNA surrounded by protein and an outer lipid envelope. TTV4 virions are stiff rods, and are very virulent causing rapid lysis of host cells (Janekovic *et. al.* 1983). It is stable and infectious after one hour of autoclaving at 120ºC (Arnold *et. al.* 1999). In general, this group of viruses has been difficult to study mostly because of the difficulty in maintaining the viruses in culture.

A novel virus, named *Pyrobaculum* spherical virus (PSV), has been found to infect anaerobic hyperthermophilic archaea of the genera *Pyrobaculum* and *Thermoproteus* (Haring *et. al.* 2004). The spherical, enveloped virus is about 100nm in diameter. The viral nucleic acid was sequenced and contains 28kb of linear DNA. Because of the unique properties of this virus, a new viral family is proposed, the *Globuloviridae*.

Acidianus viruses. Members of the archaeal genus *Acidianus* are thermophilic acidophiles that can grow both heterotrophically and autotrophically either aerobically by

oxidizing sulfur to sulfuric acid or anaerobically by reducing sulfur to H_2S (Zillig *et. al.* 1986; Segerer *et. al.* 1985).

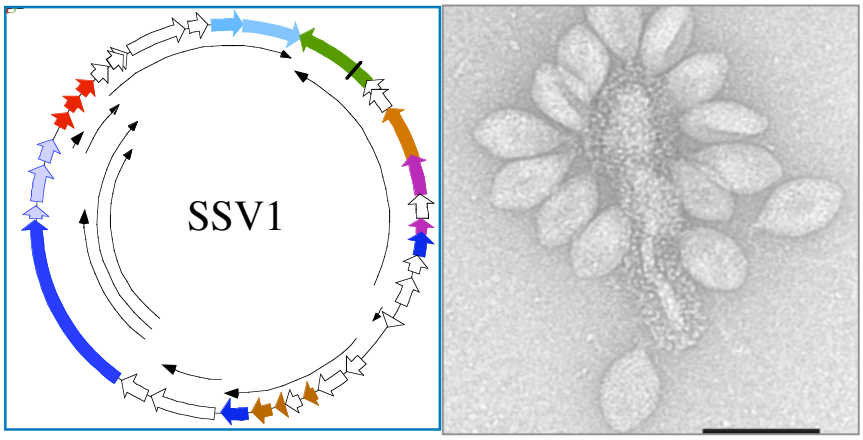
A large, filamentous virus about 2.2μm long and 27nm wide was found in an *Acidianus* isolate from Iceland (Zillig *et. al.* 1994). This virus, DAFV (*Desulfurolobus ambivalens* filamentous virus), has very similar morphology to TTV2 and SIFV (see below). It resembled SIFV in both protein profile and lipid content; however, there is no sequence data available for DAFV. AFV1 (*Acidianus* filamentous virus 1) was isolated from an *Acidianus* strain collected from a hot spring in YNP (Bettstetter *et. al.* 2003). The genome is 20.1kb of linear double-stranded DNA that has been completely sequenced. The filamentous virions are 900nm long and 24nm wide covered with a lipid envelope and contain at least five different proteins. The virions are believed to attach to host cells with the help of unusual claw-like termini. The host range is confined to *Acidianus* isolates from YNP.

Sulfolobus viruses. Most viruses isolated to date from crenarchaeal hosts have been from *Sulfolobus* (table 1.2) (Wiedenheft *et. al.* 2004; Rice *et. al.* 2004; Snyder *et. al.* 2003; Rice *et. al.* 2001; Zillig *et. al.* 1998, 1996, and 1994). This is most likely due to the relative ease of isolating and culturing the aerobic *Sulfolobus* as compared to anaerobic members of the crenarchaeotes. *Sulfolobus* is an acidophilic hyperthermophile that optimally reproduces in a pH range from 1-5 and a temperature range from 60-90ºC (Garrity 2001). *Sulfolobus* has been isolated from numerous thermal environments around the world with these parameters.

Wolfram Zillig and colleagues have been very successful at isolating viruses that infect *Sulfolobus* from Japan, Iceland, and New Zealand (Zillig *et. al.* 1998, 1996, and 1994). Since the original discovery of *Sulfolobus* in YNP, this organism had not been extensively studied within YNP until recently.

The first described and most well studied archaeal virus is *Sulfolobus* spindle-shaped virus 1 (SSV1). SSV1 was originally isolated from a hot spring in Beppu, Japan (Yeats *et. al.* 1982). SSV1 virus particles are approximately 60nm x 90nm in size (figure 1.3). The SSV1 genome was the first archaeal virus to be completely sequenced (Palm *et. al.* 1991). The genome is 15.5kb of circular dsDNA comprising 34 open reading frames (ORFs) (figure 1.3).

Figure 1.3. Genome map (left) of SSV1 virus illustrating 34 ORFs and EM micrograph (right) of SSV1 particle (Palm *et. al.* 1991).



The SSV1 genome is present in infected *Sulfolobus* cells as an episomal plasmid (both positively and negatively supercoiled) and as an integrated provirus. The biological function of each of the forms is unknown, however this was the first virus described that

contains positively supercoiled DNA (Nadal *et. al.* 1986). The viral genome integrates specifically into an arginyl-tRNA gene of the host (Reiter and Palm 1990). It is possible the integrated viral DNA is a template for transcription, the highly positively supercoiled form is used for encapsidation, and the negatively supercoiled form is used during replication of the virus (Snyder *et. al.* 2003).

SSV1 contains two viral coat proteins, VP1 and VP3, and a DNA binding protein, VP2 (shown in red in figure 1.3). These three genes along with the viral integrase (shown in green in figure 1.3) are the only four ORFs in SSV1 that have been assigned a function. None of the remaining ORFs match genes in the public databases. Because of the lack of sequence similarity, the function of the ORFs is being investigated by structural analysis utilizing X-ray crystallography (Kraft *et. al.* 2004).

There is limited information concerning SSV1 gene expression. Nine major transcripts have been detected and mapped onto the genome (black arrows in figure 1.3) (Reiter *et. al.* 1987). The nine transcripts start from six promoter sites. The major transcripts are T1 and T2, which encode for the structural proteins; T3 encodes ORF a291; T4 encodes for the largest ORF c792; and T5 transcribes the integrase gene. The elucidation of the SSV1 replication cycle remains to be determined. SSV1 particle production in culture can be induced by UV irradiation (Martin *et. al.* 1984) or treatment with mitomycin C (Grogan *et. al.* 1990; Yeats *et. al.* 1982). However, the induction of the virus does not cause host cell lysis (Grogan *et. al.* 1990; Yeats *et. al.* 1982).

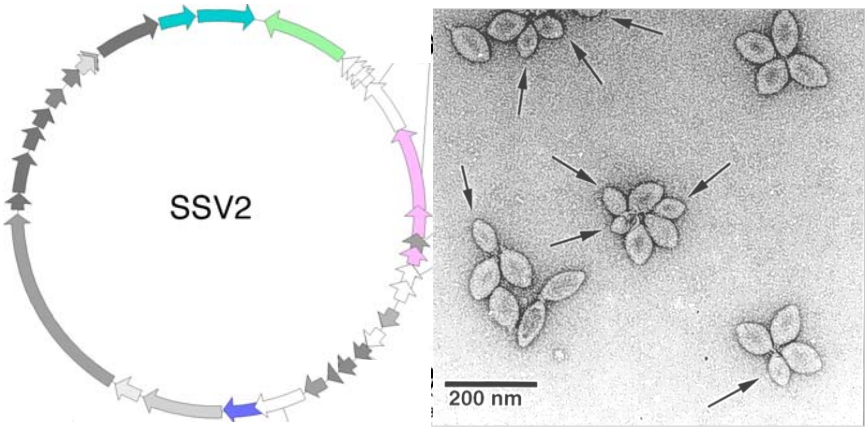
Sulfolobus cells can be successfully transfected with SSV1 DNA by electroporation (Schleper *et. al.* 1992). The SSV1 virus has been used to create a number

of shuttle vectors (Stedman *et. al.* 1999). By a serial selection technique, pBluescript S/K$^+$ (Stratagene, La Jolla, CA) was inserted into a nonessential ORF in the SSV1 genome. The construct will replicate in both *Sulfolobus* and *E. coli*. The chimeric DNA is packaged within virus particles and spreads within a culture leading to infection of nearly all cells. Because SSV1 can be induced with UV irradiation and can be present in high copy number, it is very appealing for complementation and expression studies.

A virus named SSV2 with similar morphology and genome size as SSV1 was found in samples collected from Iceland (figure 1.4) (Stedman *et. al.* 2003). The SSV2 genome is 55% identical to the SSV1 genome at the nucleotide level. The SSV2 genome is circular dsDNA, however, it is only 14.8kb in size (figure 1.4). The ORFs are arranged in a similar order on the genome indicating these two viruses belong to the same family (*Fuselloviridae*). The integrase gene and the viral coat proteins VP1 and VP3 are conserved between SSV2 and SSV1. However, SSV2 does not contain VP2. The SSV2 genome integrates into a glycyl tRNA gene instead of an arginyl tRNA gene (as SSV1 does) (Wiedenheft *et. al.* 2004; Stedman *et. al.* 2003). SSV2 can be induced with UV irradiation; however, not to the extent that SSV1 can be induced (Stedman *et. al.* 2003).

SSV-like viruses are commonly detected in enrichment cultures established from YNP (Rice *et. al.* 2001). One SSV named SSV-RH, isolated from a spring in the Ragged Hills area within Norris Geyser Basin, has been completely sequenced and its genome compared to other SSVs (Wiedneheft *et. al.* 2004).

Figure 1.4. Genome map (left) of SSV2 and electron micrograph (right) both showing similarity to SSV1 (Stedman *et. al*. 2003).



Other SSVs have been isolated and their genomes sequenced from other parts of the world, including YNP (SSV-RH) and Kamchatka, Russia (SSV-K1) (Wiedenheft *et. al*. 2004). Both of these SSVs are morphologically similar and contain similar genome sizes and arrangements (figures 1.5 and 1.6, respectively). In fact, all four sequenced SSVs (SSV1, SSV2, SSV-RH, and SSV-K1) have similar morphologies and co-linear genomes. The SSV-RH and SSV-K1 genomes integrate into the *Sulfolobus* host genome; however, SSV-RH integrates into a leucyl tRNA gene and SSV-K1 integrates into an aspartic acid tRNA and a glutamic acid tRNA (Wiedenheft *et. al*. 2004).

When the complete genomes of four SSVs (SSV1 from Japan; SSV2 from Iceland; SSV-RH from Yellowstone; and SSV-K1 from Russia) were compared, approximately one-half of the genome was found to be similar (Wiedenheft *et. al*. 2004). It is tempting to speculate that the conserved halves of the genome are necessary for life at high temperatures and low pH. The other half of the genome was found to have no similarities to each other or to other sequences in the public databases. Perhaps these

genes are what are necessary for each individual virus to survive in their respective environments. These ORFs may have diverged long ago allowing the adaptation of the SSVs to a unique environment, and now they show no similarity to each other. This research revealed that viruses in the same family could be drastically different. There is no correlation between genetic and geographic isolation. This is in contrast to what Whitaker *et. al.* (2003) determined for the host organism, *Sulfolobus*.

Figure 1.5. Genome map (left) of SSV isolated from Ragged Hills, YNP and micrograph (right) showing similarity to SSV1 (Wiedenheft *et. al.* 2004).

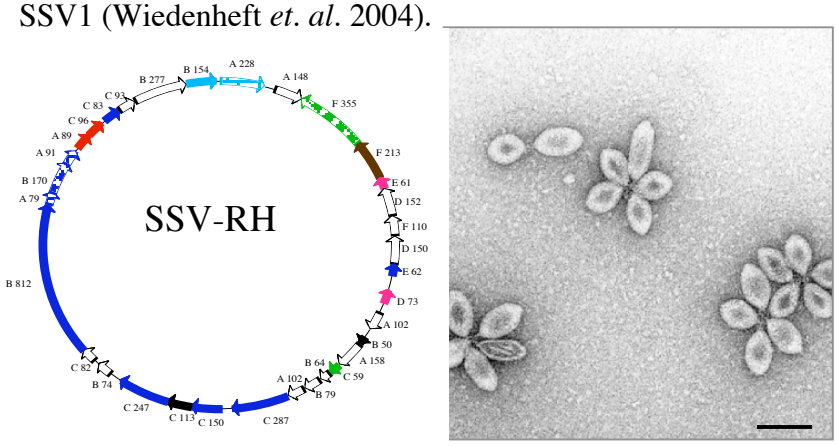
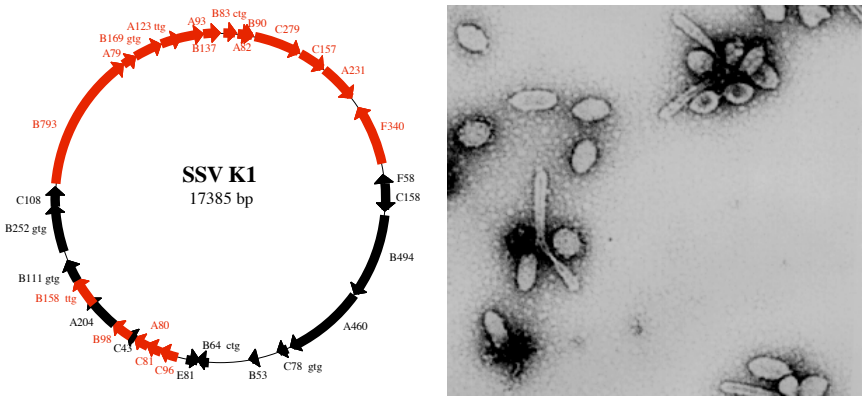
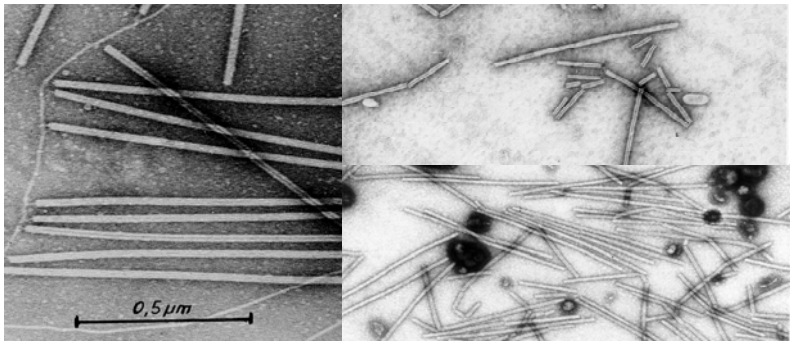


Figure 1.6. SSV isolated from Kamchatka Russia. The genome map (left) and the micrograph (right) show its similarity to others SSVs (Wiedenheft *et. al.* 2004).



The non-enveloped, stiff rod-shaped, linear double stranded DNA viruses SIRV1 and SIRV2 were first described by Zillig *et. al.* (1994). Although morphologically similar to plant viruses (for example, TMV) (figure 1.7), SIRV (*Sulfolobus islandicus* rod-shaped virus) viruses form a novel virus family, the *Rudiviridae* (www.ncbi.nlm.nih.gov/ICTV/). SIRV1, with a genome of 32kb is about 100nm shorter than SIRV2, with a genome of 35kb. The virions consist of a tube-like super-helix formed by the DNA complexed with a single basic protein. The linear DNA molecules have covalently closed hairpin ends. The SIRV viruses are not lytic and are present as carrier states in their original hosts. The genomes of both SIRV1 and SIRV2 remain unchanged in their original host, however, their genomes mutate drastically when the viruses are switched to different *Sulfolobus* hosts (Prangishvili *et. al.* 1999).

Figure 1.7. Electron micrographs of SIRV viruses. The left figure represents virus isolated from Iceland (Prangishvili *et. al.* 1999), and the two right figures illustrate the similarity in morphology of SIRVs found in YNP (Rice *et. al.* 2001).



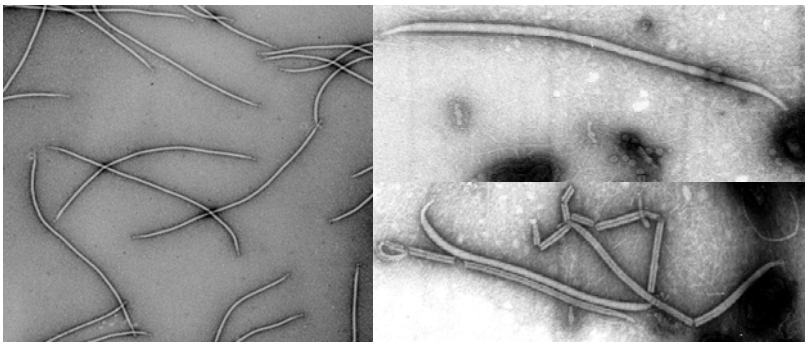
The replication cycle of rudiviruses is better understood than the replication of the SSVs. Head-to-head and tail-to-tail concatamers were found indicating replication similar to the genomes of some eukaryal viruses (Peng *et. al.* 2001). The replicative

intermediates are resolved by a virus-encoded Holliday junction resolvase (Birkenbihl *et. al.* 2001). These similarities may indicate evolutionary relationships between rudiviruses and some eukaryal viruses.

SIRV-like viruses are commonly detected in enrichment cultures established from YNP (Rice *et. al.* 2001). One SIRV-like virus, named SIRV-Y1, isolated from Norris Geyser Basin is currently being sequenced and its genome will be compared to those of SIRV1 and SIRV2 (M. Young lab, unpublished).

Sulfolobus islandicus filamentous virus (SIFV) was originally isolated from a hot spring in Iceland (Zillig *et. al.* 1994). The SIFV particles are ~30nm x 2,000nm. The virus particles are almost twice as long as the diameter of its host. SIFV possesses an envelope comprised of host lipids and mop-like tail fibers located at its termini (figure 1.8). The SIFV genome is 41kb and encodes for 74 ORFs (Arnold *et. al.* 2000b). The SIFV genome does not integrate into its host genome, but is maintained in a carrier state. The SIFV1 genome has been 96% sequenced as the ends of the DNA seem to be modified in an uncharacterized way. As with other archaeal viruses, the predicted SIFV encoded genes have little or no similarity to other genes in the public databases. However, there was similarity between ORFs in SIFV and some ORFs in SIRV (Peng *et. al.* 2001). SIFV-like viruses have been also detected from YNP enrichment cultures, although not as prominently as SSVs and SIRVs. SIFV-like viruses have also been detected in numerous anaerobic cultures from YNP (J. Spuhler, personal communication).

Figure 1.8. Electron micrographs of SIFV1 (left) isolated from Iceland (Arnold *et. al*. 2000b) and SIFV-like particles detected in YNP enrichment cultures (right) (Rice *et. al*. 2001).

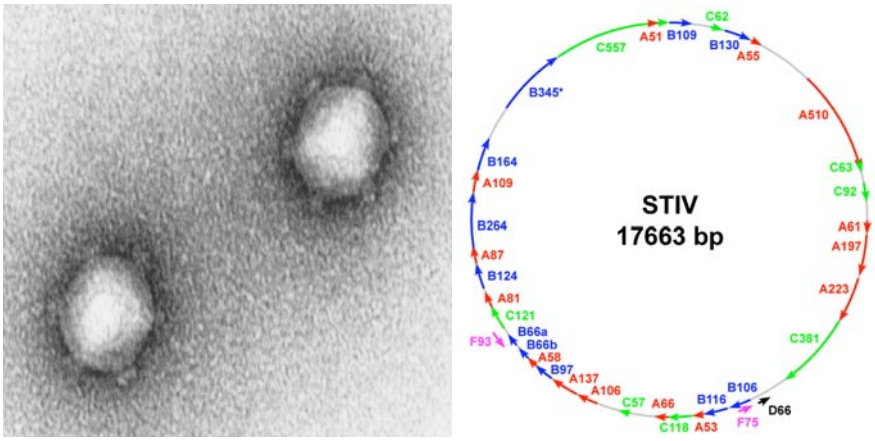


A novel virus, SNDV (*Sulfolobus neozealandicus* droplet-shaped virus) was isolated from a hot spring in New Zealand (Arnold *et. al*. 2000a). The host range of SNDV is confined to several *Sulfolobus* isolates from New Zealand. It contains a 20kb DNA genome that is believed to be heavily modified. SNDV has an unusual morphology, a "bearded" droplet of 110-185nm long by 75-90nm wide with many thin fibers on the pointed end of the virion. This virus has not been detected in hot springs outside of New Zealand.

A newly discovered virus, STIV (*Sulfolobus* turreted iscosahedral virus) has recently been completely sequenced and the structure has been solved (Rice *et. al*. 2004). This virus was isolated from a hot spring in the Rabbit Creek thermal area in YNP. The genome is circular dsDNA of 17.3kb, encoding for an estimated 36 ORFs (figure 1.9). A viral coat protein has been identified. The structure of this virus is unlike other previously described archaeal viruses with turrets extending from the five-fold axis of the

virion (figure 1.9). The function of the turrets is unknown at this time, but is postulated to play a role in the infection of *Sulfolobus* cells.

Figure 1.9. Electron micrograph (left) of newly discovered STIV from YNP and the genome map of the virus (right) (Rice *et. al.* 2004).



Uncharacterized viruses. Numerous virus-like particles detected in *Sulfolobus* enrichment cultures from YNP have been detected that have morphologies unlike any other crenarchaeotal viruses (Snyder *et. al.* 2003). However, there have also been new viral morphologies have been detected that are not associated with *Sulfolobus* hosts. Numerous virus morphologies found in *Acidianus* enrichment cultures isolated from YNP have been identified (Rachel *et. al.* 2002). These include viruses with head-and-tail morphologies, which would be the first example of a head-and-tail viral morphology in the crenarchaeotes. Novel virus-like particles have also been observed in suspected hyperthermophilic *Thermoproteales* cultured from YNP samples. The morphologies of these virus-like particles are similar to the lemon-shaped SSV morphologies, and morphologies similar to that of SIRV, a stiff rod (W. Zillig, personal communication).

Archaeal viruses in Yellowstone

The virus and VLPs detected from YNP to date are listed in table 1.3. Many of these virus morphologies have been observed from other acidic thermal features worldwide. These include SSV-, SIRV-, SIFV-, and AFV-like virus morphologies. Several other virus particle morphologies have only been detected thus far in YNP (table 1.3). We, along with other researchers, detected viral morphologies that had previously been unseen (Rice *et. al.* 2004; Snyder *et. al.* 2003; Bettstetter *et. al.* 2003; Rachel *et. al.* 2002). The viral diversity detected to date in YNP high temperature acidic environments likely represents a fraction of the total viral diversity present in these environments. No doubt more viruses will be discovered as more effort is directed at their isolation and characterization.

Table 1.3. Viruses and virus-like particles (VLPs) detected in environmental samples and enrichment cultures established from thermal acidic environments in YNP.

Virus or VLPs	*Host(s)*	*Morphology Type*	*Abundance*	*Other sites*	*References*
SSV-RH	*S. solfataricus*	spindle	Frequently detected	Japan, Iceland, Russia	Wiedenheft 2004
SSV-like VLP	*Thermoproteus* spp. *Sulfolobus* spp. *Acidianus* spp.	spindle	Frequently detected	Lassen	Zillig, unpublished; Stedman, unpublished; Rachel 2002
SIRV-Y1	*S. solfataricus*	stiff rod	Common	Iceland	Rice 2001
SIRV-like VLPs	*Sulfolobus* spp. *Thermoproteus* spp. *Acidianus* spp.	stiff rod	Common	Iceland	Rice 2001; Zillig, unpublished; Rachel 2002
SIFV-like VLP	*Sulfolobus* spp. *Acidianus* spp.	filamentous	Rare	Iceland	Zillig, unpublished Rice 2001
STIV	*Sulfolobus* spp.	icosahedral	Rare	None	Rice 2001; Rice 2004
AFV1	*Acidianus* spp.	filamentous	Common	None	Bettstetter 2003
Small icosahedral VLP	*Sulfolobus* spp.	icosahedral	Rare	None	Rice 2001
Asymetrically-shaped VLP	*Sulfolobus* spp.	asymmetric sprindle	Rare	None	Rice 2001

In previous research, a survey of viruses inhabiting many extremely thermal acidic sites in YNP was completed (Rice *et. al.* 2001). In this survey, SSV-like and SIRV-like morphologies were commonly detected in cultures established from YNP. SSVs and SIRVs appear to have a worldwide distribution. The genomes of four SSVs (Wiedenheft *et. al.* 2004; Stedman *et. al.* 2003; Palm *et. al.* 1991) and two SIRVs (Peng *et. al.* 2001; Pransgishvili *et. al.* 1999) have been sequenced. The ability to detect SSV-like and SIRV-like viruses in enrichment cultures from YNP and the availability of sequence data from each virus type resulted in the decision to monitor SSV-like and SIRV-like viruses in YNP.

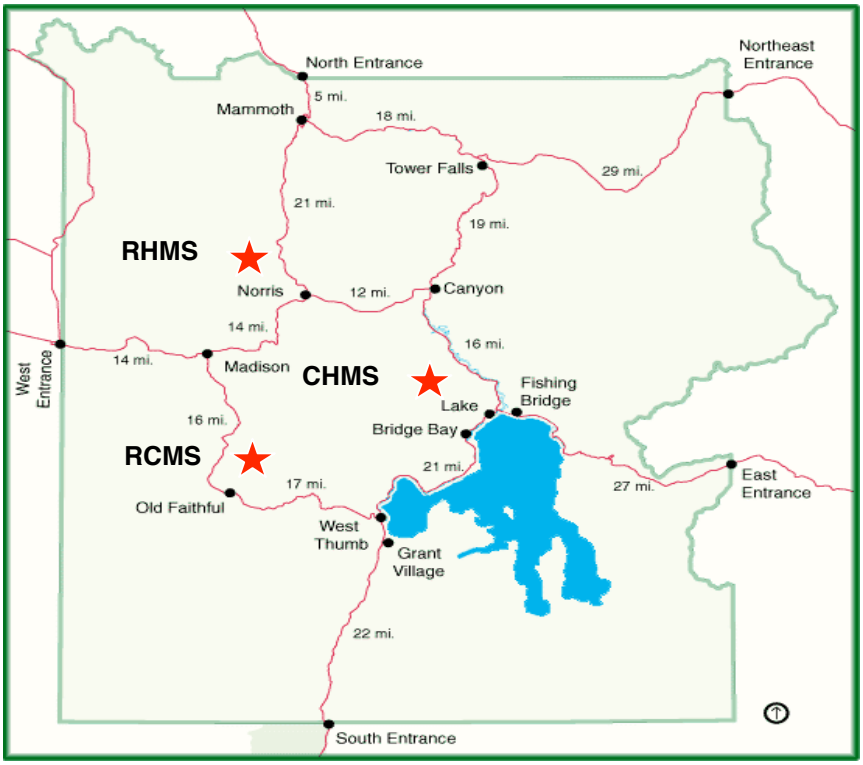
Site selection, rationale, and characteristics

Three geographically distinct thermal areas in YNP were chosen as monitoring sites for this research (represented by the red stars in figure 1.10). The rationale for choosing these sites includes the success in establishing *Sulfolobus* enrichment cultures, and the ability to detect SSV-like and SIRV-like viruses in these locations (Rice *et. al.* 2001).

Crater Hills is located within the Hayden Valley near the Mud Volcano geothermal area (44° 39.317′N and 110° 28.951′W). The Crater Hills area is characterized by a temperature range of 75 to 85°C and a pH range of 2 to 6 (M. Young lab, unpublished). The Crater Hills area has a very barren landscape of approximately 1km^2. Some of the hot springs are quite large (~300m^2), while others are small (<1m^2) and seem to come and go depending on the season. Crater Hills is a very active thermal

area, with the thumping of water being felt through the ground near some of the larger pools. The area surrounding the monitor site is quite active, and we have noticed a general expansion of thermal features in a south-westerly movement over the past few years. The colors of the individual hot springs in the Crater Hills area range from bright yellow to gray to aquamarine to white to dark gray. The Crater Hills area is surrounded by lush meadows; however, the thermal area is quite barren and there is a sharp distinction between the meadows and the thermal area.

Figure 1.10. A map illustrating locations that were monitored over time during this research. The three sites were monitored for water chemistry, archaeal populations, and virus populations. RHMS=Ragged Hills monitor site; RCMS=Rabbit Creek monitor site; CHMS=Crater Hills monitor site.



The Rabbit Creek monitor site is located within the Midway Geyser Basin (44° 31.287′N and 110° 48.647′W) near Old Faithful. This thermal area is characterized by high temperatures (75-90°C) and a pH range of 2 to 5 (M. Young lab, unpublished). This environment is a mature geothermal area and appears to be highly altered. There are no dead trees remaining from the emergence of the thermal area, indicating it is a mature thermal site. The presence of this acidic sulfate-rich thermal site was recorded as early as 1935 (Allen and Day 1935). There is an astounding amount of diversity within hot springs in the Rabbit Creek thermal area; some are gray bubbling 'mud pots', springs with red products, aquamarine colored springs, and steam dominated fumaroles. These features are quite diverse in appearance, having a wide range of sizes, colors, turbidity, clarity, and consistency. The hot springs ranged in clarity and color from completely transparent to very murky browns, grays, and reds. The mud pots ranged in color from red to white to brown. The individual springs and mud pots ranged in size from very small pools, less than ten centimeters in diameter, to medium sized pools as much as five meters in diameter.

Ragged Hills is located in Norris Geyser Basin (44° 43.653′N and 110° 42.862′W). This thermal area is characterized by temperatures ranging from 65-80°C and the pH ranges 2 to 5 (M. Young lab, unpublished). The Ragged Hills area is composed of silica-cemented gravels and until recently there was no evidence of any past thermal activity. The Ragged Hills thermal area is one of the most geothermally active areas within YNP. The site is a new geothermal feature that developed in June 1999. Over time has grown to its current size of 300m^2. There are abundant dead trees killed

off from the recent emergence of the thermal area. The site is located across the northwest ridge of a small saddle connecting the Ragged Hills area of Norris Geyser Basin to another small hill to the north. The flowing springs in the area have a diverse range of colors including clear, hazy blue, murky grey, murky brown, and murky white. There is also a wide array of colors found in the outflow channels, often changing downstream as well as through time. The outflow channels are often brown, yellow, red, orange and white.

These three sites in YNP were chosen because they are all thermal acidic hot springs and we were able to culture *Sulfolobus* and its viruses. They were also chosen for their uniqueness as well, such as their age and their chemical composition. By comparing these sites we could determine whether there is a difference in virus populations across different hot spring systems in YNP. The virus population structure, potential host populations, and the water chemistry of each of the three monitor sites was extensively studied.

Summary

1. High temperature acidic environments support crenarchaeal organisms and are poorly understood.
2. Few crenarchaeal viruses have been isolated, and little is known about these highly unusual viruses.
3. The role of viruses in any environment, and particularly YNP high temperature, acidic environments, is unknown.

4. The objective of this study is to examine for the first time, the dynamics of crenarchaeal viruses in high temperature, acidic environments within YNP.

The chapters in this dissertation are separated based on study site. Chapter 2 provides results of a hyperthermophilic virus culturing bias. Chapters 3, 4, and 5 provide data and results from the Crater Hills, Rabbit Creek, and Ragged Hills monitor sites, respectively. The materials and methods for each study site were identical. These methods are described in detail within Chapter 3. Chapters 3, 4, and 5 are structured beginning with water chemistry data, archaeal population data, followed by virus population data, and ending with a discussion of the results and concluding remarks. Chapter 6 provides a comparison of the three study sites and future implications of this research. Finally, Chapter 7 provides answers to the four major questions addressed in this research and whether the hypothesis was accepted or rejected.

CHAPTER 2

POPULATION DYNAMICS OF VIRUSES IN CULTURE

Introduction

A detailed understanding of the role of SSVs (*Sulfolobus* spindle-shaped viruses) in high temperature acidic environments requires an understanding of the diversity of viruses present in that environment. As a first step in understanding that diversity, we were interested in determining whether a culturing bias exists for these hyperthermophilic viruses in the acidic hot springs of Yellowstone National Park (YNP). We used two genes in the SSV genome to monitor changes in virus populations as a consequence of culturing environmental samples.

It is important to understand if there is a culturing bias in these organisms. Unlike the marine environment that can contain up to 10^6 to 10^8 virus particles/mL sea water (Breitbart *et. al.* 2002; Wommack and Colwell 2000), the thermal acidic areas contain very low amounts of free virus (Rice *et. al.* 2001). Therefore, in order to study viruses from thermal acidic environments one must culture the host organism. The presence of a culturing bias in these viruses could dramatically change what one would expect to see in the environment.

To date, no studies have determined if there is a strong selection bias when viruses of hyperthermophiles are replicated in culture as compared to their natural environments. Most investigators use viruses replicating in culture to examine fundamental aspects of viral replication and virus-host interactions. However, viruses

replicating in culture may not accurately represent the virus behavior and ecology found in the natural environment. Interest in the role of viruses in microbial ecology has been heightened by the completion of multiple microbial genomes, which has implicated the importance of viruses in horizontal gene transfer (Hughes and Friedman 2003; Nelson *et. al*. 1999; Peng *et. al*. 2001).

SSVs provide an ideal model system for investigating the effect of culturing on viral diversity. The broad collection of SSV genomes available has allowed a detailed comparison of the similarities and differences between geographically distributed SSVs (Wiedenheft *et. al*. 2004). SSV open reading frames (ORFs) show little or no sequence similarity to genes in the public database. However, there is a common set of eighteen gene shared by all four sequenced SSV isolates. In addition to this common set of genes, each SSV genome contains ORFs that are unique to that particular isolate. The distinct genetic signature of these viruses and the novel spindle-shaped morphology has resulted in the creation of a new viral family, the *Fusselloviridae* (www.ncbi.nlm.nih.gov/ICTV/). To date, all of the sequenced SSVs have been determined from cultured isolates. However, we do not know if these cultured isolates are representative of SSVs found in the environment.

A detailed understanding of the ecological role of SSVs in high temperature acidic environments requires an understanding of the diversity of viruses present in that environment. As a first step in understanding that diversity, we were interested in determining whether a culture bias exists for these hyperthermophilic viruses in the acidic hot springs of YNP.

Materials and Methods

Sample collection and culturing

Environmental samples were collected from the Rabbit Creek and Ragged Hills monitor sites. The Rabbit Creek site was 75°C and pH 3.3 on January 20, 2003, and the Ragged Hills site was 70°C and pH 2.7 on October 17, 2002. Five hundred milliliter water samples were collected from each hot spring using methods previously described (Rice *et. al.* 2001). Upon return to the laboratory, approximately 1mL of the environmental sample was inoculated into 60mL of enrichment medium. The media used was Zillig media, which contains per liter: 3g $(NH_4)_2SO_4$, 0.5g KH_2PO_4 x $3H_2O$, 0.5g $MgSO_4$ x $7H_2O$, 0.1g KCl, 0.01g $Ca(NO_3)_2$ x $4H_2O$, supplemented with rare elements, 1.8mg $MnCl_2$ x $4H_2O$, 0.05mg $CuCl_2$ x $2H_2O$, 0.03mg $NaMoO_4$ x $4H_2O$, 0.22mg $ZnSO_4$ x $7H_2O$, 0.03mg $VOSO_4$ x $5H_2O$, 4.5mg $Na_2B_4O_7$ x $10H_2O$, and 0.01mg $CoSO_4$ x $7H_2O$ and adjusted to pH 3.2 with sulfuric acid (Zillig *et. al.* 1994). The cultures were incubated at 80°C for 7-10 days until growth was visually evident. After the 7 days of incubation, 1mL of the established culture was passaged into 60mL of fresh media. The remainder of the primary culture was used for DNA extraction. The first passage culture was grown for 7-10 days, after which DNA was extracted.

We also wanted to determine whether the culturing media made a difference in possible culturing bias. Environmental samples were collected from the Rabbit Creek and Ragged Hills monitor sites. The Rabbit Creek site (sample named RCMS7) was 64°C and pH 3.8 on November 13, 2003. The Ragged Hills site (sample named RHMS9) was 65°C and pH 3.5 on November 13, 2003. We tested two media: Zillig medium

(described above) supplemented with 10% yeast extract, 10% sucrose, and 10% tryptone and DSMZ *Sulfolobus* medium 88 (http://www.dsmz.de/media/med088.htm). Medium 88 is very similar to Zillig medium, per liter it contains: 1.3g $(NH_4)_2SO_4$, 0.28g KH_2PO_4, 0.25g $MgSO_4$ x $7H_2O$, 0.017g $CaCl_2$ x $2H_2O$, 1.8mg $MnCl_2$ x $4H_2O$, 0.05mg $CuCl_2$ x $2H_2O$, 0.03mg $NaMoO_4$ x $4H_2O$, 0.22mg $ZnSO_4$ x $7H_2O$, 4.5mg $Na_2B_4O_7$ x $10H_2O$, 0.02g $FeCl_3$ x $6H_2O$, and 0.01mg $COSO_4$, 1g yeast extract, and adjusted to pH 2 with 10N sulfuric acid.

A *Sulfolobus solfataricus* P2 culture infected with SSV1 was used to determine how much the virus would change in culture. A glycerol stock of a previously infected culture of *S. solfataricus* was inoculated into Zillig medium supplemented with 10% tryptone (previous research indicated media supplemented with tryptone resulted in the best virus production, data not shown). The P2 culture was incubated at 80°C for 7-10 days until growth was visually evident. After growth was detected, 1mL of the established culture was passaged into 60mL of fresh media.

DNA extraction and viral PCR

DNA was extracted from the environmental samples using the UltraClean™ Water DNA Extraction Kit (Mo-Bio, Solana Beach, CA). The DNA was extracted from the cultures using a rapid alkaline lysis method described previously (Stedman *et. al.* 1999). The yield of DNA extracted from the environment and the cultures was estimated by ethidium bromide staining of agarose gels.

Two primer sets designed from the alignment of the four sequenced SSV isolates were used to amplify SSV sequences from culture and the environment. Primer set

UnvSSV#7F (5'-attcagattctg(a/t)at(a/t)cagaac-3') and UnvSSV#8R (5'-gatgagtgcgttagg(c/g)ga-3') flank the viral coat proteins and amplify 650 to 1,069 nucleotides, depending on the presence or absence of the VP2 gene in the isolate (Wiedenheft 2004). Primer set UnvSSV#3F (5'-caatcgccataggctacgg-3') and UnvSSV#4R (5'-cgttta(c/t)tactataacggtac-3') amplifies approximately 256 nucleotides of the largest ORF in the SSV genome (b812 in SSV-RH). PCR was performed with a Perkin-Elmer DNA Thermal Cycler 9700, using 40 cycles of 94°C for 1 minute, 45°C for 1 minute, and 72°C for 1.5 minutes. Reaction mixtures contained 10X *Taq* buffer, 10mM dNTPs, 50pmol/μL of both the forward and reverse primer, approximately 10ng of total extracted DNA, and 1 unit of *Taq* DNA polymerase. PCR products were visualized using agarose gels to determine concentrations and the correct size of the amplicon.

Cloning and sequencing

The PCR products were cloned using the TOPO TA Cloning Kit (Invitrogen, Carlsbad, CA). A 2mL deep 96-well plate was inoculated with *E. coli* transformed with the cloned PCR products. Plasmids were purified from *E. coli* using the Millipore 96-well plasmid prep kit (Millipore, Billerica, MA). The plasmid DNA was sequenced with primers available on the vector (T7 and M13R) and BigDye Terminator Mix (Applied Biosystems, Foster City, CA) on an ABI 9700 automated sequencer (Applied Biosystems).

Sequence analysis

Sequences were edited and analyzed using Sequencher 4.2.2 (Gene Codes Corporation, Ann Arbor, MI). Sequence alignments were performed using ClustalX (Higgins *et. al.* 1996). The alignments were subsequently analyzed by PAUP 4.0b10 (Swofford 2002) and a maximum parsimony analysis was conducted. Bootstrap values were obtained by resampling the data 10,000 times. A maximum likelihood analysis was conducted using MrBayes (Huelsenbeck and Ronquist 2001) with Bayesian clade credibility values.

Results

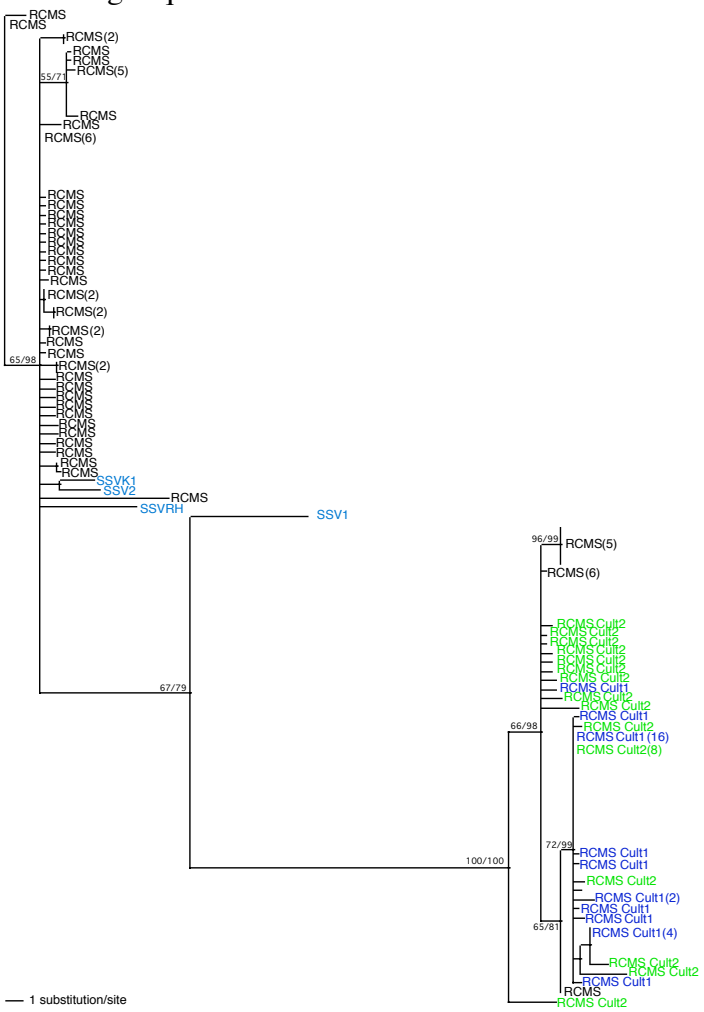
Rabbit Creek and Ragged Hills culturing bias

We detected a strong culturing bias in SSVs between the environment and the primary enrichment culture. Seventy-four PCR generated SSV-like 3F/4R clones from the Rabbit Creek site were phylogenetically analyzed. SSV-like diversity in the site is dominated by one major group and two minor groups (figure 2.1).

Phylogenetic analysis of 27 clones from the primary enrichment culture reveals an obvious culture bias (blue in figure 2.1). SSV-like diversity observed in the hot spring was reduced to a single group when grown in an enrichment culture. The primary enrichment culture was associated with only one of the environmental groups. This trend was maintained when 22 additional clones were analyzed from the first passage of the primary enrichment culture (green in figure 2.1). The enrichment culture clones form a separate group distinct from those observed in the environment, and this separation is

well supported by high bootstrap values (figure 2.1). Clearly, there is a strong culturing bias when one compares the diversity of SSVs in the environment to SSVs in culture.

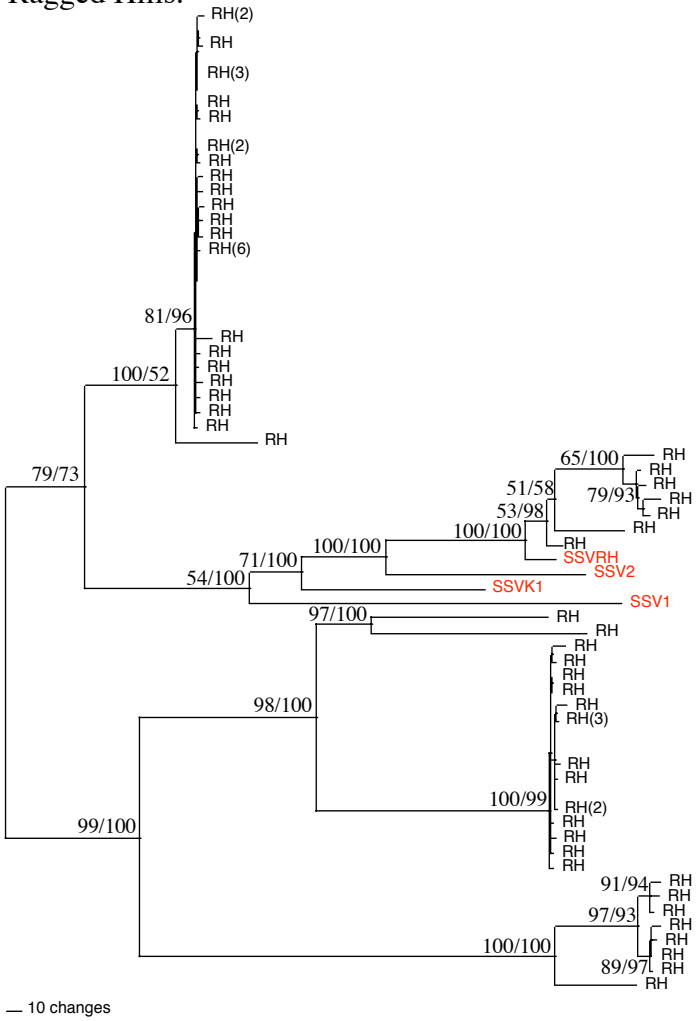
Figure 2.1. SSV 3F/4R sequences collected from Rabbit Creek monitor site. Blue=sequences from the primary enrichment; green=sequences from the secondary enrichment culture. The remaining sequences were obtained from the environment.



A culturing bias is also apparent when analyzing sequence from the SSV viral coat protein region. Direct analysis of 63 environmental clones from the Ragged Hills site identifies four dominant groups (figure 2.2). These environmental clones were

compared to coat protein genes from the sequenced SSV isolates (Wiedenheft *et. al.* 2004). The only discernible common thread among these four sequenced isolates is that they were maintained in culture. Interestingly, phylogenetic analysis of this region places all four viral isolates in the same group, despite their geographic separation (figure 2.2; sequenced isolates in red). These results suggest that culturing imposes a strong selective pressure in these viruses regardless of geographical origin.

Figure 2.2. SSV coat proteins from the Ragged Hills monitor site. Red=coat proteins from the four sequenced SSV isolates; remaining sequences were obtained from Ragged Hills.



There was a dramatic reduction in diversity in the primary enrichment culture (figure 2.1). This suggests that culturing SSVs from the environment alters the population structure. It is likely that culturing introduced a selection pressure that resulted in a bottleneck of SSV-like sequences that are not representative of the diversity present in the natural environment. Taken together, these results indicate that describing the diversity and population structure based on viruses isolated from cultures would severely underestimate the viral diversity actually present in that particular environment.

Viral culturing bias using two medias

Rabbit Creek culturing bias. Two enrichment cultures were obtained with the sample collected from Rabbit Creek in November 2003. Both SSV-like 3F/4R and 7F/8R sequences were investigated from these cultures. There does not seem to be a correlation with media or number of passages on the viral sequences obtained from culture (figure 2.3). The viruses cultured in Zillig media and media 88 group together with some outliers. There were some instances where the SSV-like 3F/4R sequences were identical whether they were cultured in Zillig media or media 88 or whether they were the primary enrichment culture or the first passage of the primary enrichment culture (represented by RCMS7_mix in figure 2.3).

A culturing bias is evident when comparing the viral sequences obtained from the cultures to the viral sequences obtained from environmental samples (figure 2.4). The environmental SSV-like sequences from the Rabbit Creek monitor site comprise two groups that are distinct from the cultured sequences (figure 2.4; purple represents the

environmental sequences). It is interesting to point out that two of the sequenced SSV isolates (SSV2 and SSV-K1; shown in blue in figure 2.4) group with environmental sequences from Rabbit Creek. SSV1 is related to a sequence detected in culture (figure 2.4). This may be a result of this virus being maintained in culture for several years before the genome was sequenced. SSV-RH is also related to a large group of cultured SSV-like sequences from Rabbit Creek (figure 2.4).

Figure 2.3. SSV 3F/4R sequences obtained from enrichment cultures from Rabbit Creek. Purple= SSVs from Zillig-YTS media; red=SSVs from Media 88; black=SSVs in both medias. M88=media 88; ZYTS=Zillig media + YTS; 1st=primary enrichment culture; 2nd=secondary enrichment culture; mix=identical sequences obtained from both medias.

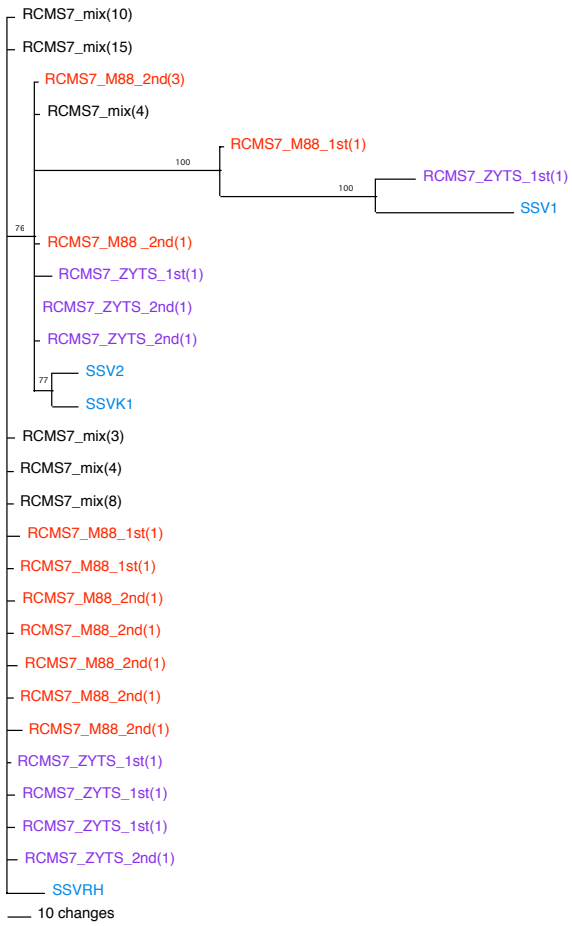
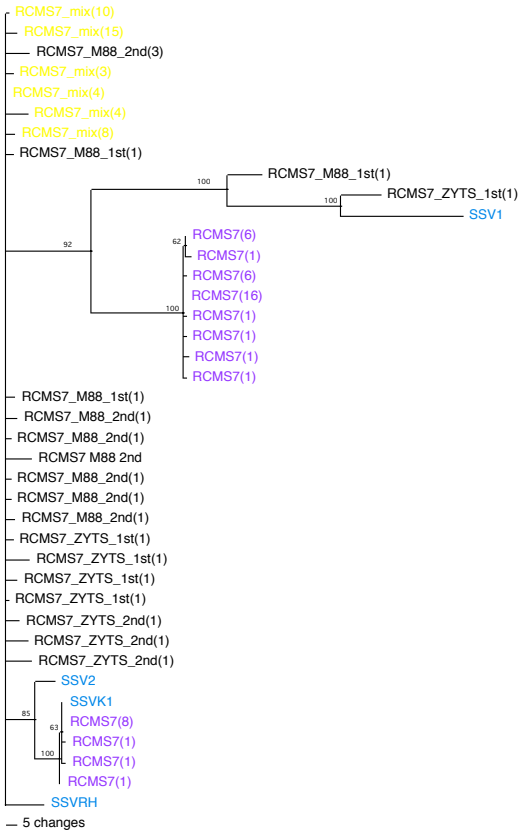


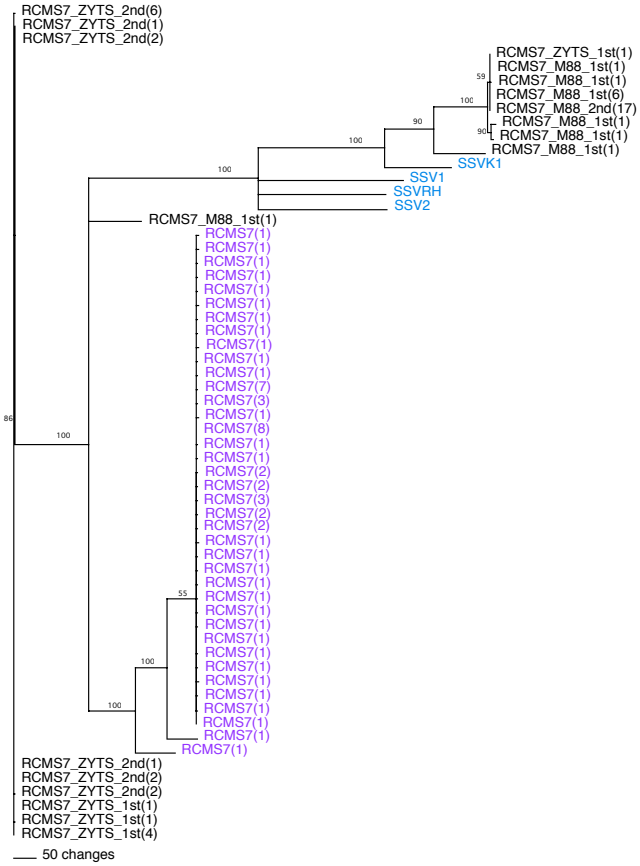
Figure 2.4. SSV 3F/4R sequences from cultures and the environment from Rabbit Creek. Purple=environmental sequences; yellow and black=cultured sequences. See legend for figure 2.3 for description of names.



When comparing the viral coat protein sequences (primer set 7F/8R) a culturing bias is also present (figure 2.5). The environmental sequences (represented by purple in figure 2.5) are all grouped together without any cultured sequences. The environmental viral coat protein sequences make up one group, which is different than the sequences from ORF c792 (primer set 3F/4R; figure 2.4) that compose two groups. This could be due to the fact that the viral coat proteins are more conserved in nature than is c792. In both instances an obvious culturing bias is present which does not seem to be dependent

on the media or the number of passages. In this case, the four sequenced SSV isolates group closely with Rabbit Creek sequences that have been cultured. Again, this could be a result of the sequenced isolates remaining in culture for many years before they were sequenced.

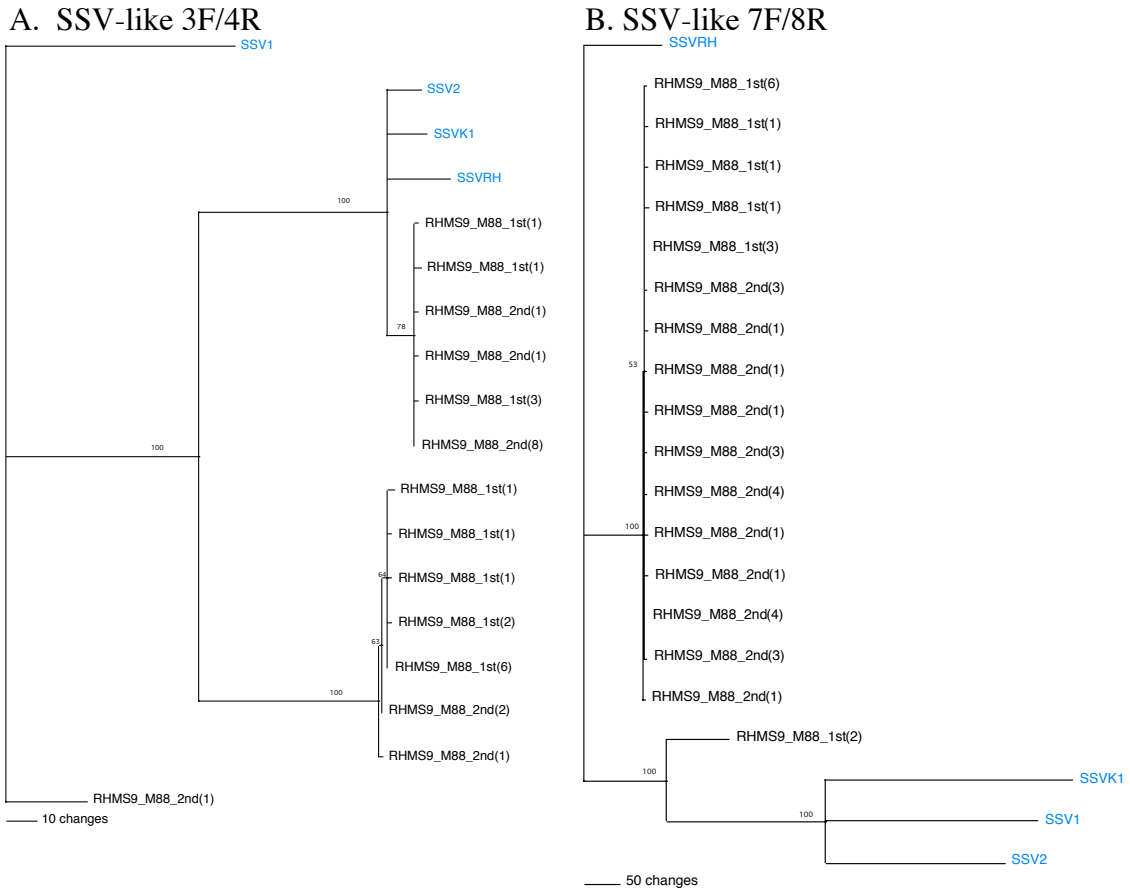
Figure 2.5. SSV 7F/8R sequences from Rabbit Creek. Purple=environmental SSV 7F/8R sequences; black= cultured SSV 7F/8R sequences. See legend for figure 2.3 for description of names.



Ragged Hills culturing bias. Similar to the Rabbit Creek samples, there does not seem to be a correlation in number of passages to viral sequences (figure 2.6). It is important to point out that a culture could not be established in Zillig media from the

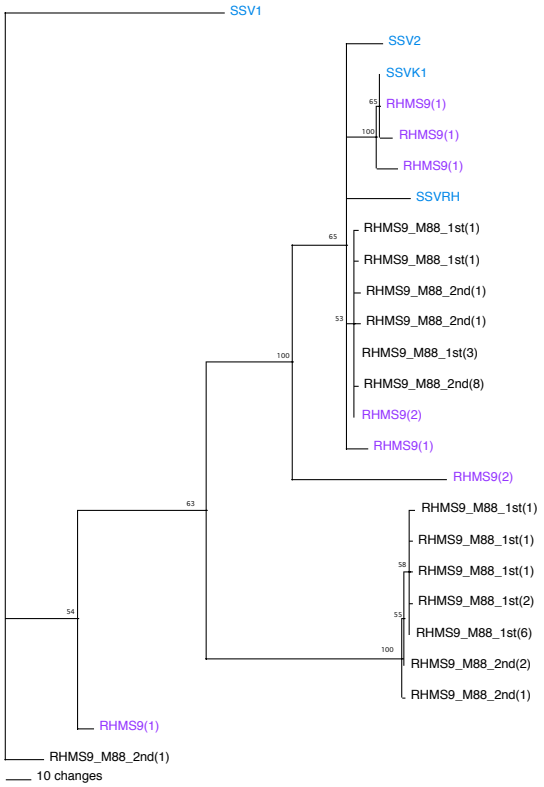
Ragged Hills monitor site. A possible reason for this is the predominant organism detected by 16S rDNA in the Ragged Hills monitor site is not *Sulfolobus* (see Chapter 5), which is the organism Zillig media was specifically designed to grow. However, there does not seem to be a large difference in the first or second passage in media 88. There are two dominant groups with one outlying sequence (figure 2.6a). This trend is slightly different when comparing the viral coat protein sequences (figure 2.6b). The viral sequences are all grouped into one major group with one outlying sequence. This again may be the result of the coat proteins being more conserved than ORF c792. This reveals that the viral sequences do not necessarily change with an inoculation into a fresh media.

Figure 2.6. SSV sequences from a Ragged Hills culture.
M88=media 88; 1st=primary enrichment culture; 2nd=secondary enrichment culture.



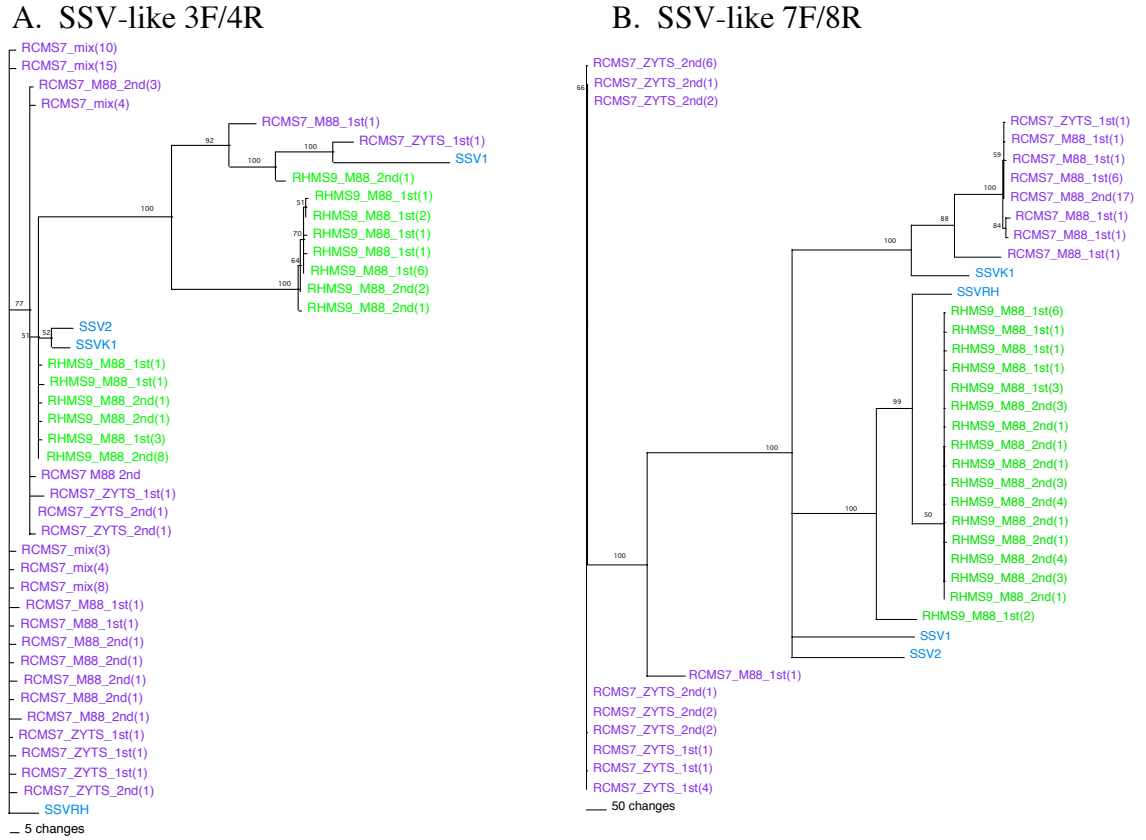
A viral culturing bias is not as evident in the samples from Ragged Hills as it was in the samples from Rabbit Creek (figure 2.7). Although there are very few environmental sequences from Ragged Hills, the cultured sequences and the environmental sequences group together. The largest group is composed of three of the four sequenced isolates (represented by blue in figure 2.7), environmental viral sequences, and cultured viral sequences. One group of viral sequences is composed of only cultured sequences, which still lends to the presence of a culturing bias in these viruses. Unfortunately, the viral coat proteins were never amplified from the RHMS9 culture or the environmental sample.

Figure 2.7. Environmental and cultured SSV 3F/4R sequences from Ragged Hills. Purple= environmental SSVs; black=cultured SSVs. See figure 2.6 for a description of names.



When comparing the cultured viral sequences from Rabbit Creek and Ragged Hills there are differences between the two sites. Even though the viruses are cultured in the same medias, there are definite differences between the cultured viral sequences from the two monitor sites (figure 2.8). There is a clear distinction between the viral sequences from the two monitor sites (figure 2.8; cultured Rabbit Creek (RCMS7) sequences are shown in purple; cultured Ragged Hills (RHMS9) sequences are shown in green, and the four sequenced isolates are shown in blue). As with individual cultures, there are more viral groups when the 3F/4R primer set is used (figure 2.8a), revealing perhaps that the c792 ORF is not as conserved as the viral coat proteins (figure 2.8b).

Figure 2.8. Cultured SSV sequences from Rabbit Creek (purple) and Ragged Hills (green). A=SSV-like 3F/4R; B=SSV-like 7F/8R. M88=media 88; ZYTS=Zillig media + YTS; mix=sequence found in both medias; 1st=primary enrichment; 2nd=secondary enrichment.



SSV1 changes in *Sulfolobus solfataricus* P2 culture

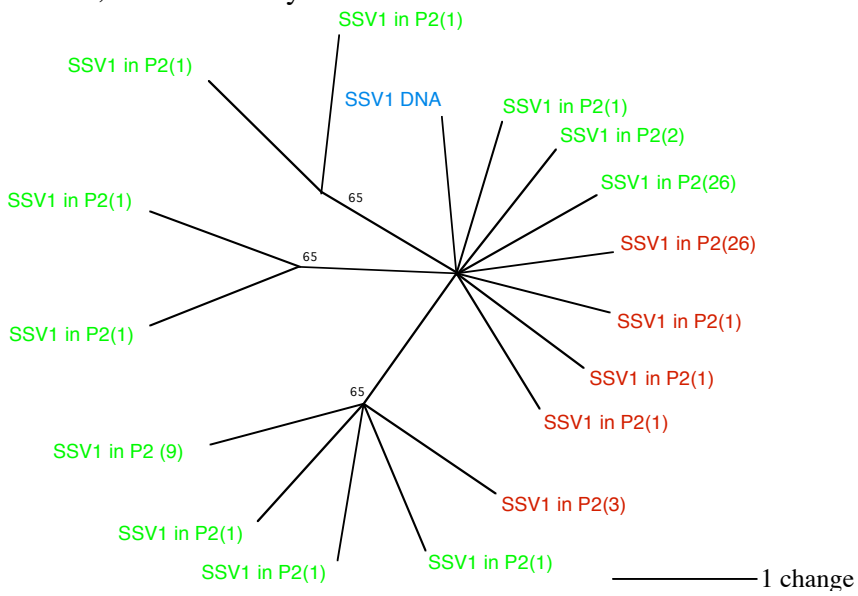
The changes occurring in SSV1 were monitored in a *Sulfolobus solfataricus* P2 culture over time. A *Sulfolobus solfataricus* P2 culture previously infected with SSV1 DNA was inoculated into Zillig medium supplemented with tryptone. Purified SSV1 DNA was PCR amplified, cloned, and sequenced with the sequences obtained from culture. The first enrichment culture yielded 45 c792 (3F/4R) clones and 42 coat protein (7F/8R) clones. The second enrichment culture yielded 32 3F/4R clones, but no coat protein clones were obtained. It is curious as to why the 7F/8R primers did not work on the secondary enrichment culture, and it will be further discussed later.

When 42 SSV1 3F/4R clones were sequenced and compared to the sequence from SSV1 DNA, there are four groups of sequences (figure 2.9; primary enrichment culture is represented by green). Twenty-nine of the SSV1 sequences closely resembled wild type SSV1. The sequences in this group each contained one nucleotide change from the SSV1 DNA. Two groups each contained only two sequences each, and each contained three changes from the wild type SSV1 sequence (figure 2.9). A larger group contained 11 sequences, and also contained three changes in comparison to SSV1 (figure 2.9).

When 32 clones from the secondary enrichment culture were added to the analysis, only two groups of sequences were represented (figure 2.9; secondary enrichment culture is represented by red). Most of the sequences (29) showed only one change from the SSV1 DNA (figure 2.9), and three other sequences showed three changes from SSV1 (figure 2.9). So, over time in culture, the viral sequences had

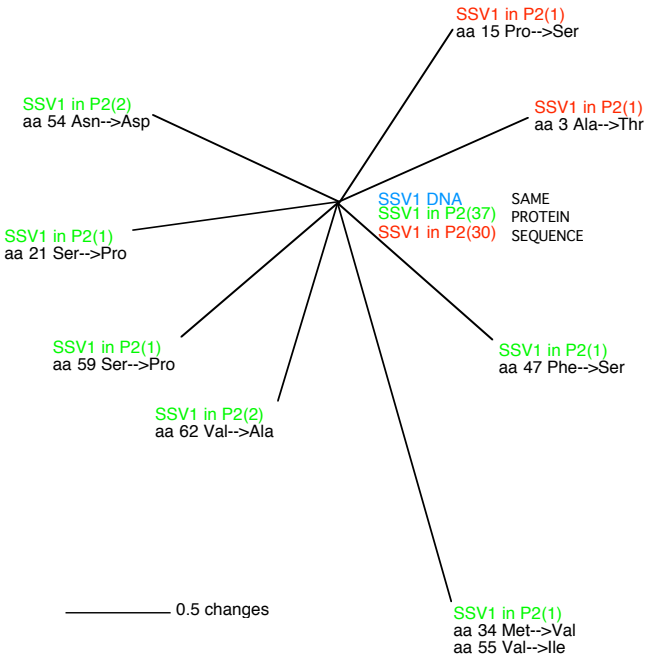
decreased from four groups to only two groups. This result is similar to what was observed in culturing SSV-like viruses from the natural environment.

Figure 2.9. SSV1 3F/4R DNA sequences obtained from a *Sulfolobus solfataricus* P2 infected culture. Green=primary culture; red=secondary culture.



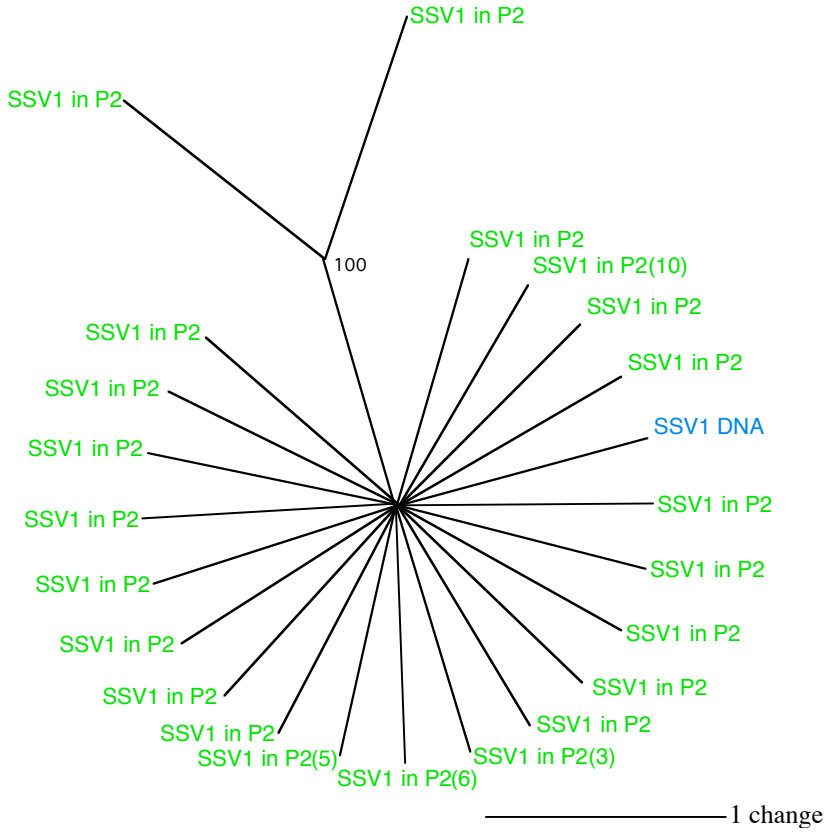
The SSV1 3F/4R protein sequences (85 amino acids) obtained from the cultures show even fewer changes. When comparing all sequences obtained from the primary and secondary enrichment cultures, only one sequence contained more than one amino acid change (figure 2.10). Most of the sequences (37 from the primary culture and 30 from the secondary culture) were identical to the wild type SSV1 3F/4R protein sequence (figure 2.10). Five sequences from the primary culture revealed only one amino acid change (represented by green in figure 2.10), and one sequence contained two amino acid changes (represented by red in figure 2.10). The amino acid changes do not appear to be centralized in one area, but instead throughout the entire sequence.

Figure 2.10. SSV1 3F/4R protein sequences obtained from an infected culture of *S. solfataricus* P2. Green=primary enrichment culture; red=secondary enrichment. The change(s) in each sequence are listed below the sequence name.



The SSV1 coat proteins (primer set 7F/8R) were only amplified from the primary enrichment culture. When 42 coat protein clones from culture are compared to the wild-type coat protein sequence of SSV1, there are even fewer changes than in the 3F/4R sequences (figure 2.11). Forty of the clones show only one nucleotide change in the coat protein sequence, and two sequences show three changes from the wild type SSV1 DNA sequence (figure 2.11). The forty clones that revealed only one change in the coat protein sequence, each showed a different nucleotide change, so it seems the virus may be radiating out (figure 2.11).

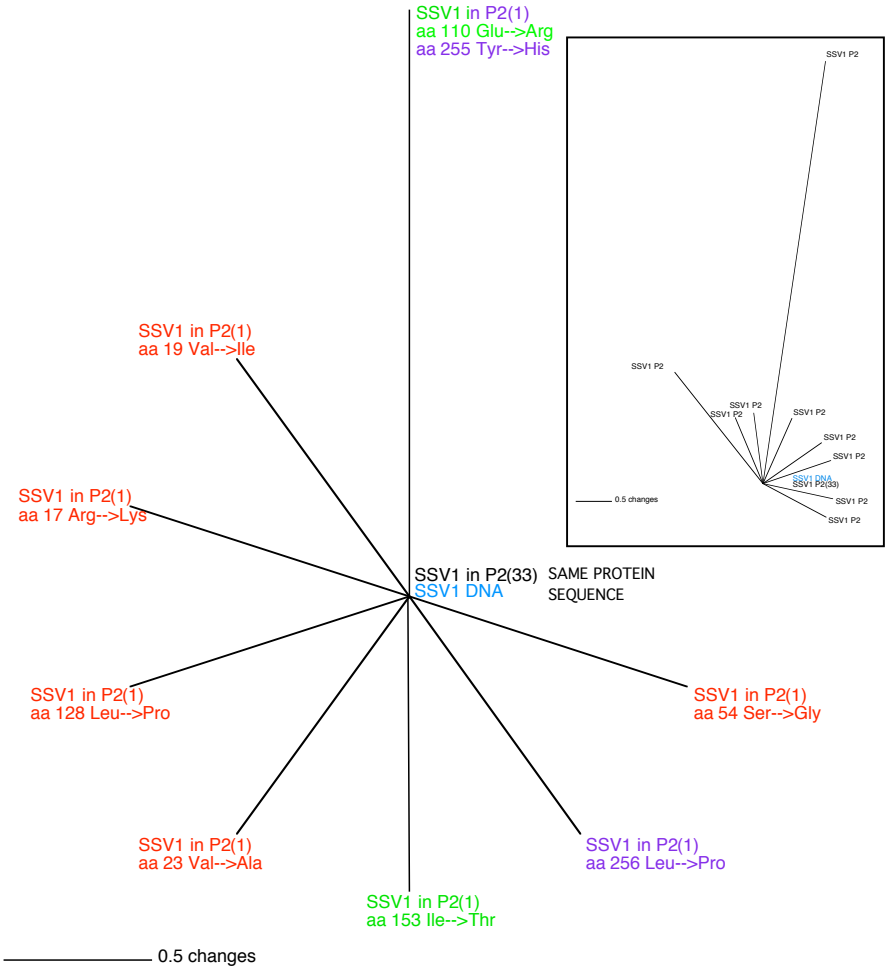
Figure 2.11. SSV1 7F/8R DNA sequences from a *Sulfolobus solfataricus* P2 infected culture.



When the coat protein sequences are translated to protein and compared to the protein sequence of SSV1, only a few changes are observed (figure 2.12). When 42 protein sequences are compared, one was very different from the other sequences (figure 2.12 inset). Upon closer investigation it was found this sequence was missing VP3, and therefore, it was removed from further analysis (figure 2.12). Only one cultured coat protein sequence contained two changes from the wild type SSV1 coat protein sequence (figure 2.12). The remaining sequences only showed one amino acid difference from wild type sequence (figure 2.12). Five sequences obtained from culture showed one

amino acid change in VP1, one sequence showed one amino acid change in VP3, one sequence had a change in VP2, and one sequence contained one change in VP3 and one change in VP2 (figure 2.12).

Figure 2.12. SSV1 7F/8R protein sequences from a *S. solfataricus* infected culture. All of the sequences are from the primary enrichment. Inset: phylogenetic tree of all sequences retrieved from culture; however, the sequence with the very long branch was missing VP3, and therefore, was left out of the analysis. Red=sequences with a change in VP1; green=sequences with a change in VP3; and purple=sequences with a change in VP2.



In reviewing the changes in SSV1 coat protein sequences, there is not a particular region of the protein sequence that appears more susceptible to change. The changes occurred throughout the sequence, and appeared to be random. The coat proteins appear to be more conserved than the 3F/4R sequences in SSVs.

Conclusions

There are several possibilities that may explain the observed culturing bias in the SSV-like populations. One possibility is that the host range of SSVs extends beyond the *Sulfolobus* genus. Sequence analysis of the 16S rDNA gene indicates that *Acidianus* and *Stygiolobus* are dominant members of the archaeal community in the two high temperature acidic environments sampled (see Chapters 4 and 5). However, these organisms were not detected in the enrichment cultures. All enrichment cultures were exclusively *Sulfolobus solfataricus* as determined by 16S rDNA sequence analysis. In contrast, Rachel *et. al*. (2002) established cultures from environmental samples from a hot springs in the Crater Hills region of YNP (44° 39′ 13.3″N and 110° 28′ 39.8″W) that were exclusively *Acidianus* by 16S rDNA analysis. They observed viruses with similar morphology as SSVs associated with these *Acidianus* hosts. Using similar culturing methods, designed to culture *Sulfolobus* species, they were only able to culture *Acidianus* species. A 60 x 90nm SSV-like spindle-shaped virus-like morphology was observed in these enrichment cultures. However, it has not yet been determined if these viruses are members of the *Fusselloviride*.

Another explanation for our results is the possibility that the viral sequences were present in low numbers in the environment, and we enriched for those viruses in the enrichment cultures. This appears to be congruent with the low abundance of *Sulfolobus* in the acidic hot springs in YNP.

It has long been known that there is a culturing bias in microorganisms, however, this trend has not been shown in viruses until now. Clearly, microbial diversity in enrichment cultures seldom reflects the diversity of microbial life present in the environment. As shown by this research, the same appears to be true for these hyperthermophilic viruses. There was a dramatic reduction in diversity in the primary enrichment culture (figure 2.1). This suggests that culturing SSVs from the environment alters the population structure. Culturing introduced a selection pressure that resulted in a bottleneck of SSV sequences that are not representative of the diversity present in the natural environment.

By amplifying these genes from extracted DNA, the presence of virus in the enrichment cultures is not guaranteed. PCR does not tell us that we are detecting viral sequences; we could be detecting other sequences that are present in the environment. However, the sequences appear to be viral because of the size of the amplicon, and the sequences significantly match SSV-like sequences in the public databases. Perhaps, we are underestimating the diversity of the viruses and not over-estimating because we are always using the same primers to amplify potentially changing viral sequences.

The culturing bias in SSVs from the environment was present regardless of the media used to enrich the viruses. However, when the cultured viral sequences from

Rabbit Creek and Ragged Hills are compared, there are distinct differences between the viral sequences from each location. Perhaps this distinction is representative of the different chemical environments of each acidic hot spring.

It is interesting the viral coat proteins were not detected more often from the cultured viruses enriched from Ragged Hills. Perhaps we are amplifying related viruses from different archaeal organisms. The sensitivity of the SSV 7F/8R primer set was found to be 10,000X lower than the SSV 3F/4R primer set (see Chapter 3). This is likely the explanation for detecting more SSV-like 3F/4R sequences in the environment and enrichment cultures.

By culturing viruses from two acidic hot springs in YNP, it was determined a culturing bias does exist for this hyperthermophilic virus. Over time, the diversity of the SSVs in culture was reduced to a single group of viruses. To determine how rapidly that event could have occurred, a sampling of the enrichment culture everyday over a time course would have been useful. Sampling the enrichment culture daily and monitoring how the viral sequences changed could have established the rate of the adaptation of the viruses to the new environment.

By studying SSV1 in culture, the sequences (both 3F/4R and 7F/8R) do not change significantly over time. In both instances, only one SSV sequence from the enrichment cultures showed more than one amino acid change from the wild type SSV1 sequence. There were not any sequences that contained three amino acid changes. The coat proteins of SSVs were found to be more conserved (in culture) than the sequences

obtained from the largest ORF in the viruses (c792 in SSV1). The changes occurring in each sequence appears to be random and not focused within one area of the gene.

The results from the SSV1 sequencing reveals that the viral sequences do change in culture, but not drastically. One might expect that in such a controlled environment, such as a laboratory culture, the viruses would not change at all. But, I believe that changes over time are inevitable. The glycerol stock of *Sulfolobus* strain P2 infected with SSV1 has been frozen away at -80°C for numerous years, so it is certainly reasonable that the virus may have changed some in that time or perhaps being inoculated into fresh culturing media will result in changes as well. It is also possible the replication of the virus resulted in changes. It is assumed these viruses utilize the polymerase of their host organism, which does possess a proofreading capability. However, that does not exclude the chance that mistakes are made and not corrected. In fact, many viruses encode a polymerase without a proofreading capability. This increases the chance that a variant of the virus may be better able to adapt to its surroundings.

The SSV1 sequences in this study do not appear to drastically change in culture over time. This result is not necessarily surprising; a laboratory culture is a very controlled environment in that the culture is kept at a constant temperature and pH without many fluctuations. In the natural environment, fluctuations in temperature and pH are inevitable. Therefore, it will be interesting to determine how these viruses may change in a natural system when they must respond to changing environmental parameters.

It has long been known that most microorganisms cannot be cultured from the environment, and that culturing dramatically reduces and changes the microbial diversity. But, until now it was unknown whether viruses were subject to the same occurrence. We have shown that by culturing viruses from the environment, the viral sequences may not be representative of the viruses in the natural environment. So, conclusions regarding the viruses based on culturing experiments must be made cautiously.

CHAPTER 3

CRATER HILLS MONITOR SITE: VIRUS DYNAMICS, HOST POPULATIONS, AND WATER CHEMISTRY

Introduction

A thermal acidic feature within the Crater Hills area was selected as a monitor site. The reasons for selecting this particular site included the monitor site is both thermal and acidic, the site is easily accessible year round, and we were successful in culturing *Sulfolobus* and its viruses from this hot spring (Rice *et. al.* 2001). Over the course of two years, we monitored an acidic hot spring within the Crater Hills thermal area in Yellowstone National Park (YNP) for changes in the viral and host populations, and water chemistry. The overall goal of this research is to determine if there are significant changes in viral population structures over time. The hypothesis being tested is whether there is an association between virus changes and observed changes in the host populations and water chemistry of an individual thermal feature.

The viruses monitored in Crater Hills over time were SSV-like and SIRV-like viruses. These viruses were chosen because of the ease of culturing *Sulfolobus* and its associated viruses from YNP, there was genomic data available for both virus types from geographically isolated thermal areas worldwide (Wiedenheft *et. al.* 2004; Peng *et. al.* 2001; Prangishvili *et. al.* 1999; M. Young lab, unpublished), and these two viruses were commonly detected in primary enrichment cultures established from YNP (Rice *et. al.* 2001).

Materials and Methods

Site location

The Crater Hills monitor site is located within Hayden Valley near the Mud Volcano geothermal area (44° 39.317′N, 110° 28.951′W).

pH and temperature monitoring

The temperature of the Crater Hills monitor site was continuously monitored with a data logger. The data logger recorded the measured temperature approximately every 20 minutes. The pH of the monitor site was measured each time a sample was collected.

Sample collection and water chemistry analysis

Hot spring water samples were collected by mixing the water in the pool and collecting a surface sample. Approximately 500mL of hot spring water was collected from each time point for both total nucleic acid extraction and water chemistry analyses.

The chemistry samples were analyzed by Energy Laboratories (Billings, MT). Each sample was analyzed for inorganics (chloride and sulfate), organics (dissolved organic carbon [DOC]), nutrients (ammonia, total nitrogen [expressed as nitrate + nitrite], and phosphorus), and metals (aluminum, antimony, arsenic, barium, boron, cadmium, calcium, copper, iron, lead, magnesium, manganese, mercury, molybdenum, nickel, potassium, selenium, silicon, sodium, and zinc). Energy laboratories supplied us with protocols and sampling kits containing four bottles: two plastic 500mL bottles, one plastic 250mL bottle, and one 500mL glass bottle. The water in the glass bottle was filtered and preserved with H_2SO_4, the water in the plastic 250mL bottle was not filtered

and was preserved with H_3PO_4, the water in one of the plastic 500mL plastic bottles was filtered and preserved with HNO_3, and the water in the other plastic 500mL bottle was not filtered or preserved (to test for dissolved metals). The samples were kept at 4°C until being sent to Energy Laboratories.

Total nucleic acid extraction
from environmental samples

Total nucleic acid from the Crater Hills monitor site was extracted using the UltraClean™ Water DNA Extraction Kit (Mo-Bio, Solana Beach, CA) as manufacturer recommends. Briefly outlined, ~500mL water is filtered through a 0.2μm filter, and the filter is bead-beated with SDS to break open cells. The subsequent steps include precipitating the nucleic acid with sodium acetate, washing with ethanol, and resuspending in 1mL 10mM Tris-Cl (pH 8.2). The integrity and yield of the extracted DNA was analyzed by agarose gel electrophoresis.

Primer design for amplification of viruses

SSV-like sequences were amplified from the total nucleic acid extracted from the environment. Based on an alignment of four SSV genomes from geographically isolated areas around the world (Wiedenheft *et. al.* 2004), two "universal" primer sets were designed (figures 3.1 and 3.2) to amplify two different regions in SSVs (figure 3.3). The primer set named UnvSSV#3F (5'-caatcgccataggctacgg-3') and UnvSSV#4R (5'-cgttta(c/t)tactataacggtac-3'), amplifies ~250-260 base pairs (bps) in the largest open reading frame (ORF) of each sequenced isolate (figure 3.1 and table 3.1). The second primer set, UnvSSV#7F (5'-attcagattctg(a/t)at(a/t)cagaac-3') and UnvSSV#8R (5'-

gatgagtgcgttagg(c/g)ga-3'), amplifies the coat protein genes (figure 3.2). Depending on the presence of VP2, these amplicons range in size from 682 bps to 1,069 bps (table 3.1).

Figure 3.1. The alignment of four sequenced SSV isolates and the 3F/4R primer set, which amplifies a portion of the largest ORF in SSVs.



Figure 3.2. The alignment of four sequenced SSV isolates and the 7F/8R primer set, which amplifies the viral coat proteins.



Figure 3.3. A genomic map of SSV-RH showing the locations of the universal SSV primer set 3F/4R and 7F/8R.

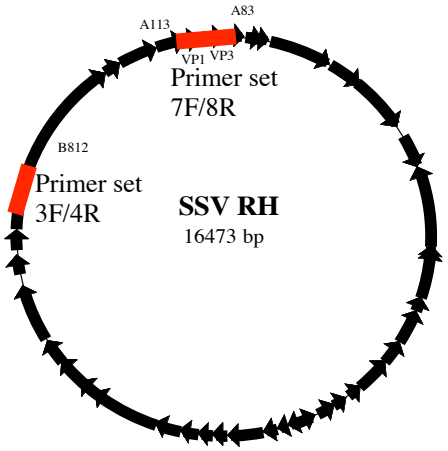
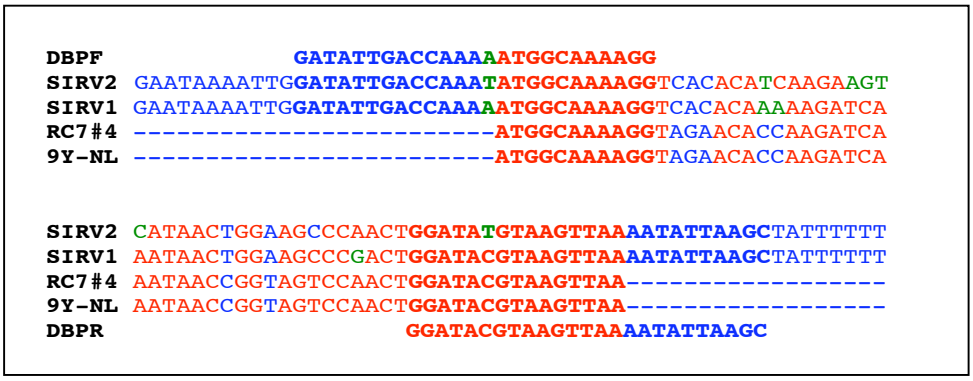


Table 3.1. Number of bases amplified by universal SSV primers in four sequenced isolates.

	Amplicon Size in SSVs	
	UnvSSV#3F/4R	*UnvSSV#7F/8R*
SSV1 (Japan)	253 bps (ORF c792)	1,069 bps (vp1; vp3; vp2)
SSV2 (Iceland)	259 bps (ORF c809)	682 bps (vp1; vp3)
SSVRH (YNP)	259 bps (ORF b812)	692 bps (vp1; vp3)
SSVK1 (Russia)	259 bps (ORF b793)	825 bps (vp1; vp3)

SIRV-like viral sequences were amplified from environmental extracts of total nucleic acid. One primer set that amplifies the coat protein was designed based on the alignment of SIRV1, SIRV2, and two partially sequenced samples from YNP (figure 3.4). The primers were designed based on the assumption that the start and the stop of the coat protein gene would be conserved. The DBP-F (5'-gatattgaccaaaaatggcaaaagg-3') and DBP-R (5'-gcttaaatattttaacttacgtatcc-3') primer set amplifies the entire 405 base pairs (bps) of the coat protein (the start and stop of the coat protein are underlined in each respective primer).

Figure 3.4. Alignment of primers used to amplify the coat protein of SIRV-like viruses in YNP.



*RC7#4 and 9Y-NL are two partially sequenced isolates from YNP.

PCR sensitivity of the primer sets

Once the primer sets were chosen, the sensitivity was measured before attempting to use them on environmental samples. To test the sensitivity of the aforementioned primer sets, purified DNA from SSV1, SSV2, SSV-K1, and SSV-RH was used to test the limits of the 3F/4R and 7F/8R primers. The sensitivity of the DBPF/R primers was tested using purified SIRV1 DNA. The purified DNA from each of the four SSVs and SIRV1 was serially diluted to 10^{-12} from a starting concentration of 1µg. The viral template concentration ranged from 1µg to 10^{-18} µg (0.001 femptogram). The PCR reactions were conducted in both a standard PCR machine (Perkin-Elmer DNA Thermal Cycler 9700, Wellesley, MA) and a real time PCR machine (RotorGene, Corbett Research, Mortlake, Australia) to determine if there was a significant difference. The standard PCR protocol was as follows: 5µL 10X PCR buffer with $MgCl_2$, 10mM dNTPs, 50pmol/µL each primer, 1µL diluted viral template, 1 unit *Taq*, and ddH_2O to bring the final reaction volume to 50µL. The reaction conditions started with an initial 2 minute denaturation at 95°C followed by 35 cycles of 1 minute at 95°C, 1 minute at 40°C (SSV primers) or

50°C (SIRV primers), and 2 minutes at 72°C, followed by a final extension step of 7 minutes at 72°C.

The real time PCR conditions were similar except the addition of SYBR Green (Molecular Probes, Carlsbad, CA). The real time PCR protocol was as follows: 2.5μL 10X PCR buffer with $MgCl_2$, 5mM dNTPs, 7.5pmol/μL each primer, 1μL diluted viral template, 0.25 units of *Taq*, SYBR Green, and water to bring the reaction volume to 25μL. The reaction parameters for the real time PCR started with an initial denaturation at 95°C for 2 minutes, followed by 40 cycles of 95°C for 20 seconds, 40°C (for SSV primers) or 50°C (for SIRV primers) for 30 seconds, and 72°C for 45 seconds (to acquire SYBR), a hold at 72°C for 3 minutes, and finally a melt from 55-99°C. The melt cycle allows the user to distinguish between real products and primer dimers based on the temperature of the DNA when it melts. The samples were either analyzed by agarose gel electrophoresis or visualized by fluorescence on the real time PCR machine.

The same PCR conditions used to test the primers were used to amplify SSV-like and SIRV-like DNA from the environment. The concentration of environmental DNA was estimated between 1pg and 10ng. Only the standard PCR machine was used for environmental PCR because of difficulties in cloning DNA intercalated with SYBR Green.

Archaeal 16S rDNA primers and PCR

16S rDNA gene sequence analysis was used to determine the identity of the potential archaeal host populations in the Crater Hills monitor site. Two primer sets were chosen to amplify 16S rDNA from the total DNA extracted from the hot spring samples,

20F (5'-ttccggttgatcctgccgg-3') and UA1406R (5'-gacgggcggtgtgtrca-3'), and 348Fa (5'-tccaggccctacggg-3') and UA1406R (Baker *et. al*. 2003). The forward primers (20F and 348Fa) are archaeal specific and the reverse primer (UA1406R) is universal. The reaction parameters were identical to the PCR conditions for the viral genes, with the exception of the annealing temperature and the extension time. The annealing temperature for the primer set 20F/UA1406R was 60°C and 55°C annealing temperature was used for primer set 348Fa/UA1406R. The extension times for primers was increased to 1.5 minutes. The primer set 20F/UA1406R should yield a product of ~1,400 bps with 348Fa/UA1406R yielding a 1,000 bps product. The amplicons were analyzed by agarose gel electrophoresis.

Cloning PCR products

The PCR products (both viral and 16S rDNA) were cloned into the pCR2.1 vector via TOPO-TA cloning (Invitrogen, Carlsbad, CA). The vector is provided linearized, supplied with overhanging thymidines (T) on the 3' ends, and Topoisomerase I covalently bound to the vector. This results in an easy ligation with PCR products originally amplified with *Taq* because *Taq* has a nontemplate-dependent terminal transferase activity that adds a single deoxyadenosine (A) residue to the 3' end of each amplicon. The ligation is efficient, occurring at room temperature in five minutes. The vector encodes for both ampicillin and kanamyacin resistance, and encodes for the lacZα gene, which allows for blue-white screening of the transformants. Every time a positive PCR product was generated, the product was ligated into the TOPO-TA vector,

transformed into *E. coli* (One Shot® TOP10, Invitrogen), and the white colonies were screened for insert.

Following transformation into *E. coli* and plating on selective media, colonies were inoculated into 2mL deep 96-well plates (Millipore, Billerica, MA), and incubated in 1mL LB^{KAN} for 24 hours. The cells were then pelleted by centrifuging and the media was removed. The protocol to purify plasmid follows the manufacturer's directions (Millipore). Briefly, the cells are resuspended and lysed. The RNA is removed with RNase, and the solution neutralized. The following steps involve filtering and washing the plasmids. Once completed, there is a 96-well plate with a different plasmid in each well.

DNA sequencing

Once the 96-well plasmid preps are complete, a second screen is done to estimate insert ratio on the plate. A random sampling of plasmid from the plate is digested with EcoR1 (Promega, Madison, WI) to test insert size and DNA quality. The plasmid DNA is sequenced using either M13R (5'-caggaaacagctatgac-3') or T7 Promoter (5'-ccctatagtgagtcgtatta-3') primers. The DNA sequencing protocol is as follows: 5μL ddH_2O, 5pmol/μL primer (M13R or T7 Promoter), 2μL plasmid DNA, 1μL 5X sequencing buffer, and 1μL BigDye Terminator Mix (Applied Biosystems, Foster City, CA). The sequencing reaction is composed of three different temperatures repeated 40 times: 96°C for 10 seconds, 50°C for 5 seconds, and 60°C for 4 minutes. Once the sequencing reaction is complete the samples are filtered and cleaned up with 0.3mM EDTA using Multiscreen® SEQ_{384} sequencing reaction cleanup plates (Millipore) (as

recommended by the manufacturer). The reactions are resuspended in 20μL 0.3mM EDTA and sequenced on an ABI 9700 automated sequencer (Applied Biosystems) in the lab of Dr. Francisco Roberto at the DOE-INL in Idaho Falls, Idaho.

Sequence and phylogenetic analysis

Viral sequences. The environmental SSV-like and SIRV-like sequences were edited using Sequencher version 4.2.2 (Gene Codes Corporation, Ann Arbor, MI). Once vector sequence was removed and sequences edited, the sequences were aligned (in Sequencher) to determine how many (if any) were 100% identical matches. The sequences were subsequently exported to Sequence Alignment Editor (Rambaut 1997) to manually verify the alignments produced by Sequencher. The sequences were then exported to ClustalX (Thompson *et. al.* 1997) to produce an alignment. The alignments were subsequently analyzed by PAUP 4.0b10 (Swofford 2002) and a maximum parsimony analysis was conducted. Bootstrap values were obtained by resampling the data 10,000 times, and a bootstrap 50% majority-rule consensus tree was produced.

Using Modeltest version 3.6 (Posada *et. al.* 1998) on a Macintosh platform, the appropriate nucleotide substitution model for maximum likelihood was selected. Modeltest is used to test the model of evolution the data best fits. This program tests 56 different models of evolution by performing hierarchical hypothesis testing. At each level the null hypothesis is accepted or rejected until the model that best fits the data is determined. Based on the nucleotide substitution model, a maximum likelihood analysis was conducted using MrBayes version 3.0 (Huelsenbeck and Ronquist 2001) on a

Macintosh platform. Clade credibility values were derived using MrBayes where two million permutations were sampled every ten thousand generations.

The viral DNA sequences were translated with Sequence Alignment Editor (Rambaut 1997) and maximum parsimony analyses were carried out as described above. The environmental SSV-like 3F/4R protein sequences were 85 to 87 amino acids from the largest ORF in the SSVs (812 amino acids predicted in the SSV-RH ORF). When the SSV-like DNA sequences were translated to protein, some sequences resulted in a truncated protein sequence. The translated DNA sequences that resulted in an interrupted ORF (those with stop codons throughout) are termed non-functional ORFs, and the DNA sequences that resulted in un-interrupted ORFs (no stop codons) are termed functional ORFs. The distinction between functional and non-functional ORFs was used to determine which viral sequences might have come from integrated copies of the virus. The interrupted SSV-like ORFs were not used in the analyses, unless otherwise noted. The environmental SIRV-like coat protein sequences were 135 amino acids in length. This represents the full-length of the coat protein. The molecular evolutionary analyses (rates of transitions and transversions) for both viruses were conducted using MEGA version 3 (Kumar *et. al.* 2004) on a PC based platform.

Phylogenetic trees were constructed sequentially in order to compare the new viral sequences to the viral sequences detected in earlier samples. In all of the following phylogenetic trees, the sequences from the sample being discussed will be presented in red, sequences from previous samples will be presented in black, and the sequences from the completely sequenced isolates (SSV1, SSV2, SSV-RH, and SSV-K1 for the 3F/4R

and 7F/8R primer sets; SIRV1 and SIRV2 for the DBPF/R primer set) will be blue. The black sequences are only presented as a reference to more easily determine how or if the viral population changed since the last sample was collected. I will define a viral group/clade as a monophyletic cluster of viral sequences. The viral group is supported by high bootstrap values and Bayesian clade credibilities.

In the DNA analysis, numbers next to principle branches show bootstrap support >50% using maximum parsimony and Bayesian clade credibilities using maximum likelihood. In the protein analysis, the number at the principle branches show bootstrap support >50% using maximum parsimony. The frequency of a particular sequence is noted in parentheses following the name of the sequence.

Archaeal host sequences. 16S rDNA sequences were edited using Sequencher version 4.2.2 (Gene Codes Corporation). The environmental 16S rDNA sequences were compared to published 16S rDNA sequences (http://www.ncbi.nlm.nih.gov/BLAST/). Environmental sequences were aligned to published 16S rDNA genes with ClustalX. Alignments were subsequently analyzed by PAUP 4.0b10 (Swofford 2002) and a maximum parsimony analysis was conducted as described above.

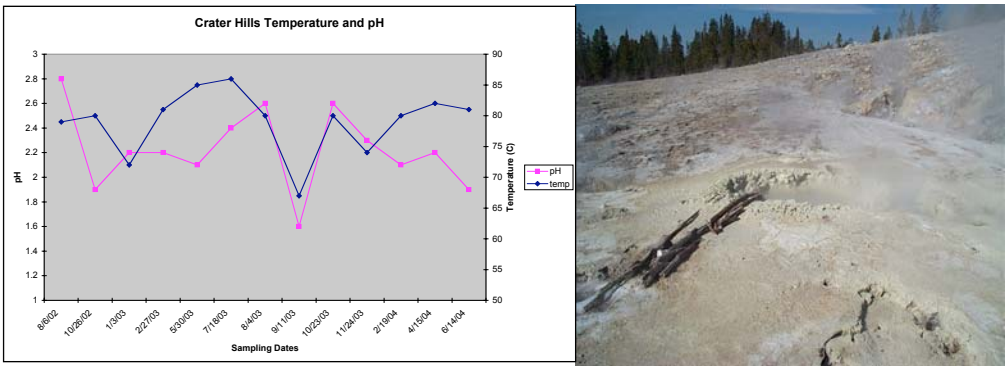
Results

Temperature, pH, and water chemistry of the Crater Hills monitor site

This monitor site was sampled thirteen times from August 2002 through June 2004. The site ranged in temperature from 67 to 86°C and the pH ranged from 1.6 to 2.6

(figure 3.5). The chemistry analysis revealed this site was distinct from the other two monitor sites in this study. Surprisingly, the chemistry did not significantly change during this research.

Figure 3.5. Temperature and pH over time at the Crater Hills monitor site (left) and a picture of the site (right).



The water chemistry analysis started January 3, 2003 and continued through June 14, 2004 for a total of 11 samples (table 3.2). Of the metals tested, several were not detected, or detected sporadically at very low levels (antimony, barium, calcium, copper, magnesium, mercury, molybdenum, nickel, and selenium) (table 3.2).

This site is a sulfate-dominated system with very high levels of sulfate in all samples collected from this hot spring (figure 3.6a). The average sulfate concentration was 1022mg/L with a range of 427-1610mg/L. The chloride levels remained low during this research (0-7mg/L) with one spike in the August 2003 sample (see figure 3.6a), and an average chloride concentration of 1.8mg/L. The dissolved organic carbon (DOC) concentration varied from 3-9.6mg/L with an average concentration of 5.7mg/L (table 3.2).

Table 3.2. Water chemistry values measured at the Crater Hills monitor site from January 2003 through June 2004.

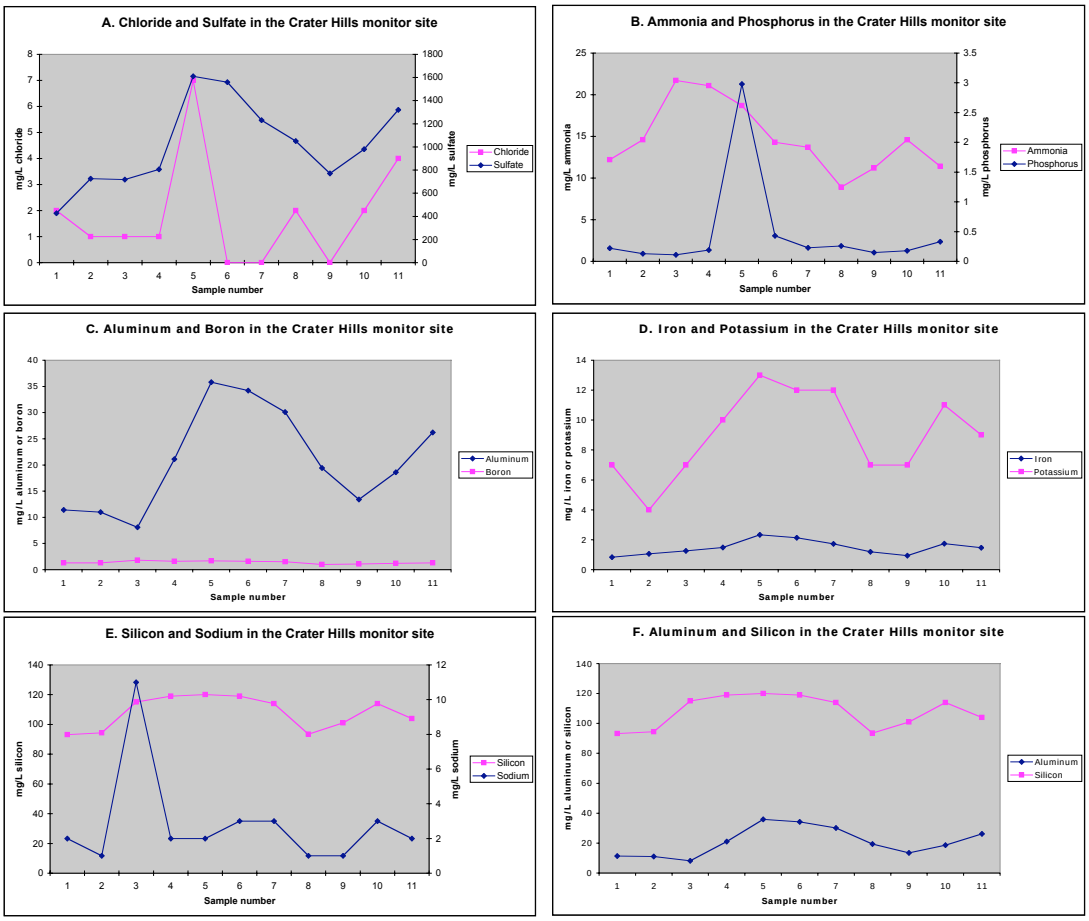
Sample numbers	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*
Sample dates	*01/03*	*02/03*	*05/03*	*07/03*	*08/03*	*09/03*	*10/03*	*11/03*	*02/04*	*04/04*	*06/04*
pH	2.2	2.2	2.1	2.4	2.6	1.6	2.6	2.3	2.1	2.2	1.9
Temp (C)	72	81	85	86	80	67	80	74	80	82	81
Chloride	2	1	1	1	7	0	0	2	0	2	4
Sulfate	427	726	719	807	1610	1560	1230	1050	770	980	1320
DOC	3	6.4	5.2	6.6	6.8	9.6	6.5	3.1	5.89	3.9	6.2
Ammonia	12.2	14.6	21.7	21.1	18.7	14.3	13.7	8.9	11.2	14.6	11.4
Nitrogen	0	0	0	0	0.07	0	0	0.1	0	0	0.07
Phosphorus	0.22	0.13	0.11	0.19	2.98	0.43	0.23	0.26	0.15	0.18	0.33
Aluminum	11.4	11	8.1	21.1	35.8	34.2	30.1	19.4	13.4	18.6	26.2
Antimony	0	0	0	0	0	0	0	0	0	0	0
Arsenic	0	0	0.015	0.021	0.009	0.009	0.017	0.007	0.007	0	0.034
Barium	0	0	0	0	0	0.1	0	0	0	0	0.1
Boron	1.3	1.3	1.8	1.6	1.7	1.6	1.5	1	1.1	1.2	1.3
Cadmium	0.002	0.001	0.004	0	0	0	0	0	0	0	0
Calcium	0	0	0	0	0	2	0	0	0	0	0
Copper	0	0	0	0	0.01	0	0.01	0	0	0	0
Iron	0.83	1.06	1.25	1.48	2.33	2.14	1.72	1.2	0.94	1.74	1.47
Lead	0	0	0.01	0	0.03	0.02	0.02	0.03	0	0	0.03
Magnesium	0	0	0	0	0	0	0	0	0	0	0
Manganese	0.01	0.02	0.02	0.02	0.04	0.03	0.04	0.02	0.02	0.03	0.03
Mercury	0	0	0	0	0	0	0	0	0	0	0
Molybdenum	0	0	0	0	0	0	0	0	0	0	0
Nickel	0	0	0	0	0	0	0	0	0.04	0	0
Potassium	7	4	7	10	13	12	12	7	7	11	9
Selenium	0	0	0	0	0	0	0	0	0	0	0
Silicon	93.2	94.4	115	119	120	119	114	93.5	101	114	104
Sodium	2	1	11	2	2	3	3	1	1	3	2
Zinc	0.09	0.03	0.12	0.06	0.04	0.24	0.06	0.02	0.06	0.06	0.02

*Sample numbers refer to numbers on corresponding graphs.
*Chemical species measured in mg/L.

The nutrients measured in the hot spring were ammonia, nitrogen, and phosphorus (table 3.2). The ammonia levels peaked in the May 2003 sample and have since been decreasing (figure 3.6b) with an average ammonia concentration of 14.8mg/L and a range of 8.9-21.7mg/L. Nitrogen, which was expressed as nitrate and nitrite, was only detected three times at very low levels (0.07mg/L-0.1mg/L) (table 3.2). The phosphorus concentrations were consistently low except for the sample collected in August 2003

(figure 3.6b). The average phosphorus concentration was 0.47mg/L; however, that may be high because of the spike in phosphorus in August 2003. If that sample is omitted, the overall average of phosphorus concentration was 0.22mg/L.

Figure 3.6. Water chemistry values obtained from the Crater Hills monitor site over the course of this research. The sample number corresponds to the sampling date in table 3.2. All values are represented in mg/L.



The metals detected consistently over time were aluminum, arsenic, boron, iron, lead, manganese, potassium, silicon, sodium, and zinc (table 3.2). The aluminum concentrations were the highest detected in the three monitor sites with an average

aluminum concentration of 20.8mg/L (figure 3.6c). Boron was detected continuously ranging from 1-1.8mg/L with an average of 1.4mg/L (figure 3.6c).

The highest iron concentration was 2.33mg/L in August 2003, and then it began to decrease until April 2004 (see figure 3.6d). The iron concentrations ranged from 0.83-2.33mg/L with an overall average of 1.47mg/L. Potassium was consistently detected (figure 3.6d), ranging in concentration from 4 to 13mg/L with an average of 9mg/L.

Silicon remained very constant in this monitor site (figure 3.6e). The silicon concentrations ranged from 93.2 to 120mg/L with an average of 108mg/L. Sodium was also consistently detected; however, there was one spike in concentration in the May 2003 sample, but another spike was not detected (figure 3.6e). The sodium was very low, ranging from 1-3mg/L with the spike at 11mg/L. Arsenic, lead, and manganese were all detected; however, at low levels (table 3.2). The last metal tested was zinc, and it was consistently detected at levels of 0.02-0.12mg/L (table 3.2).

Overall, the water chemistry of this monitor site was distinct from the chemistry of the other two monitor sites (see Chapters 4 and 5). The Crater Hills monitor site is a vapor-dominated system with no visible inflow or outflow channels. During the monitoring period, this monitor site was dominated by high sulfate and aluminum levels (figure 3.7). This hot spring was also characterized by the lowest observed pH values of the three monitor sites throughout the sampling time (1.6-2.6), and the temperatures ranged between 67-86°C (figure 3.5). Surprisingly, the chemical composition did not significantly change over time. A summary of the average water chemistry values measured over time is presented in figure 3.8.

Figure 3.7. Comparison of sulfate (A) and aluminum (B) levels in the three monitor sites during the course of this research. The sample numbers correspond to the sampling dates in table 3.2.

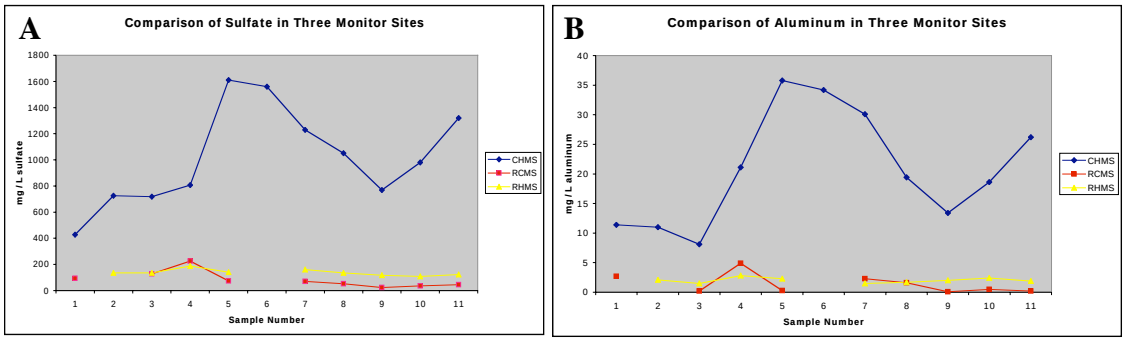
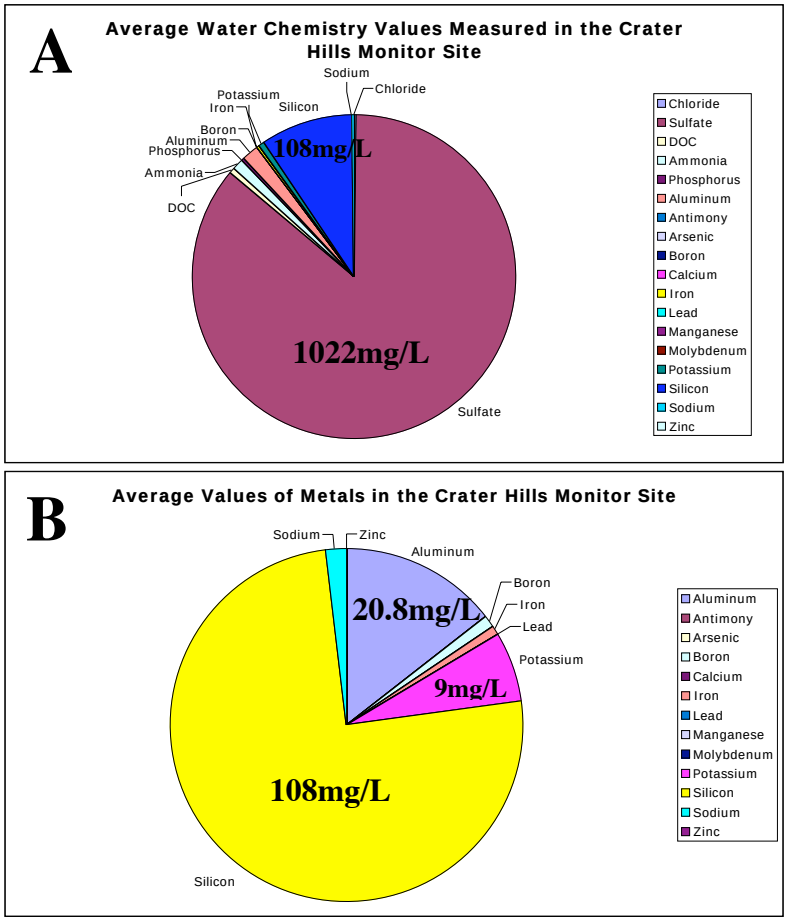


Figure 3.8. The average water chemistry values measured over time in the Crater Hills monitor site. A. The average of all the chemical species measured. B. The average values of the metals detected in the monitor site. All values are measured in mg/L.

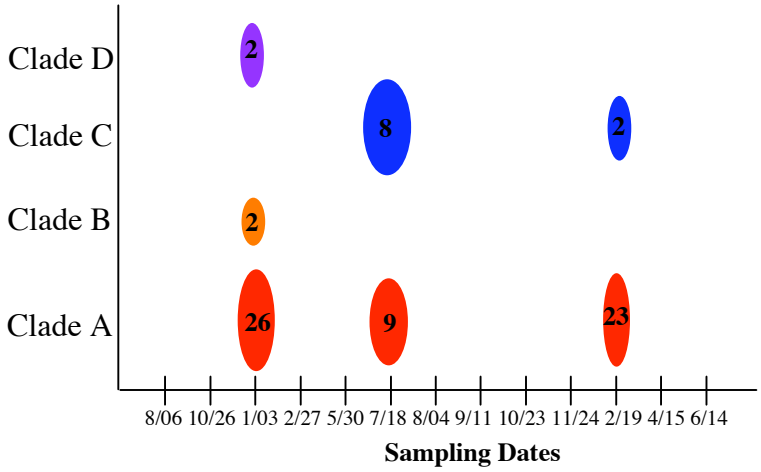


Archaeal population in the Crater Hills monitor site

To determine the archaeal host population, 16S rDNA sequence analysis was conducted on specific sampling dates. Six of thirteen samples were used to determine the dominant archaeal population over time in Crater Hills.

Archaeal population determined by the 20F/UA1406R primer set. The 16S rDNA genes from three samples were amplified with the 20F/UA1406R primer set. An overview of the results with this primer set is illustrated in figure 3.9.

Figure 3.9. The archaeal population in Crater Hills over time measured with the 20F/1406R primer set. The numbers indicate how many 16S rDNA clones most closely matched the published16S rDNA sequence. The absence of sequences does not reflect an absence of 16S rDNA sequences from the environment; not all samples were used for 16S rDNA gene sequence analysis.



The first sample was collected on January 3, 2003, and 30 full-length 16S rDNA clones were sequenced and compared against 16S rDNA genes in the public database (figure 3.11). Most of the environmental sequences (26) belong in clade A, two belong in clade B, and two sequences belong to clade D (figure 3.10). Surprisingly, only one

Sulfolobus-like 16S rDNA sequence was detected in this sample (clade D in figure 3.10). Rather, 28 out of 30 sequences (93%) most closely matched *Caldococcus noboribetus* (accession number D85038) with 89-95% identity (clade A in figure 3.9 and table 3.3). The remaining two clones matched published 16S rDNA sequences for *Sulfolobus solfataricus* (accession number X03235), and *Acidianus convivator* (accession number AJ634763) at low percentages (clade D in figure 3.9). The sequence that aligned best to *A. convivator* was only a 93% match, and the environmental sequence that most closely matched *S. solfataricus* was only a 92% match (table 3.3). These low percentage matches are indicative of archaeal species that are similar to these named species, but clearly distantly related.

Seventeen clones were obtained from July 18, 2003, and once again, *Caldococcus*-like sequences dominated the environmental library representing 88% of the clones (15/17) and matching with 89-97% identity to the published *Caldococcus* 16S rDNA gene (clades A and C in figure 3.10 and table 3.3). The other sequences detected in this sample matched *Vulcanisaeta distributa* (accession number AB063640) at 94-95% identity (clade C in figure 3.10 and table 3.3).

Twenty-five 16S rDNA clones were obtained from February 19, 2004 (table 3.3). The population was similar to that measured in July 2003. The environmental sequences belong in clades A and C (figure 3.10). Again, the dominant sequences most closely matched *Caldococcus* (23/25) with identities ranging from 85-96% (clades A and C in figure 3.10 and table 3.3). The other two sequences most closely matched *Stygiolobus azoricus* (accession number X90480) (84% identity) and *Vulcanisaeta* (96% identity)

(clade C in figure 3.10 and table 3.3). The archaeal population in the environmental clone libraries appears to be predominantly *Caldococcus*-like organisms (figure 3.11).

Figure 3.10. Environmental 16S rDNA sequences amplified with the 20F/ UA1406R primer set from the Crater Hills monitor site.

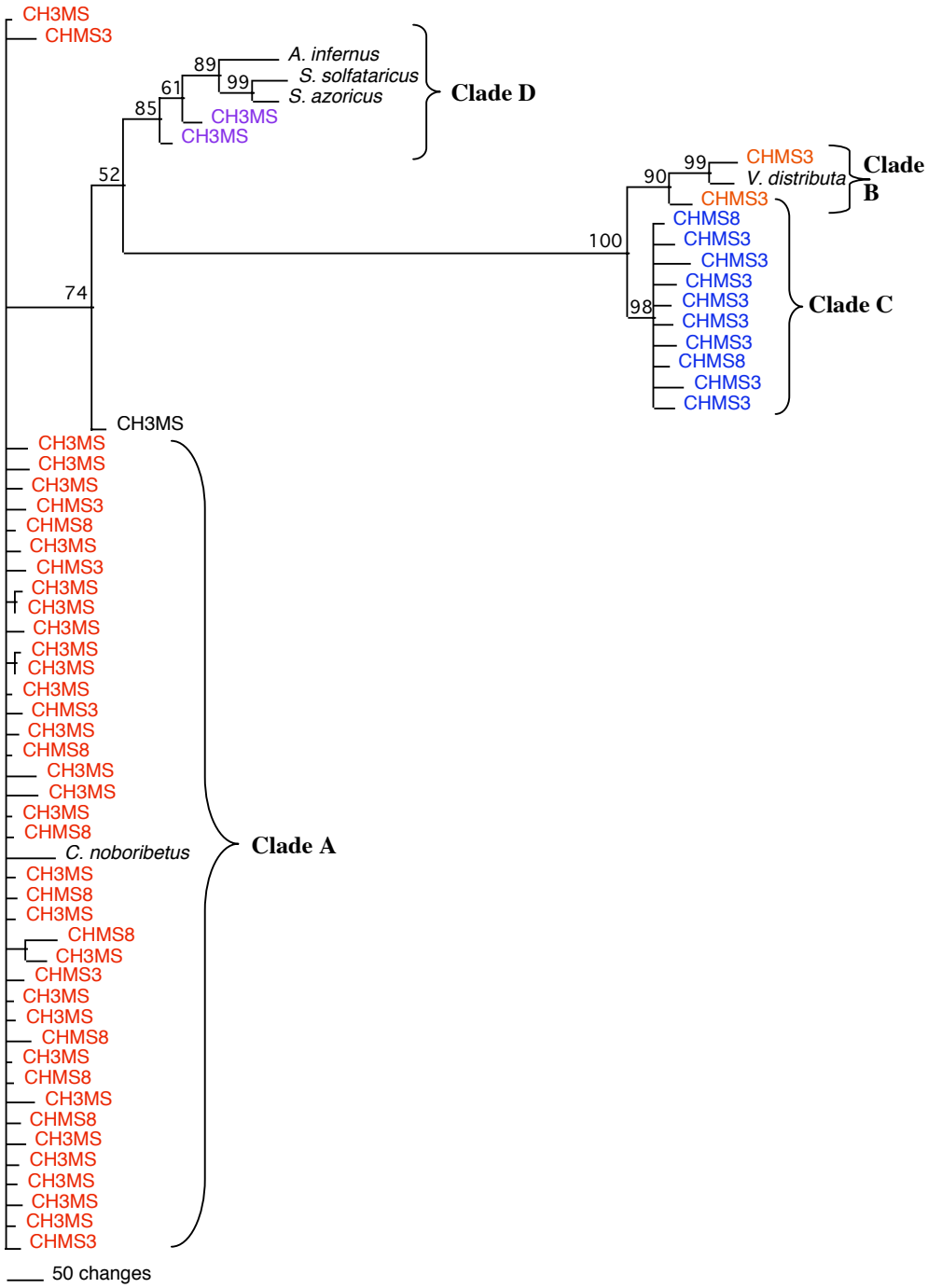
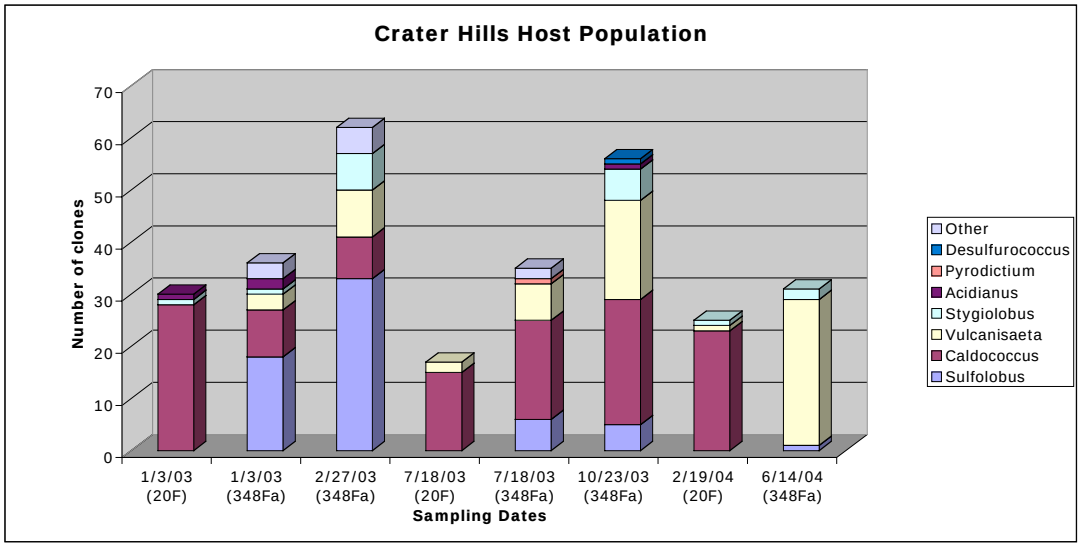


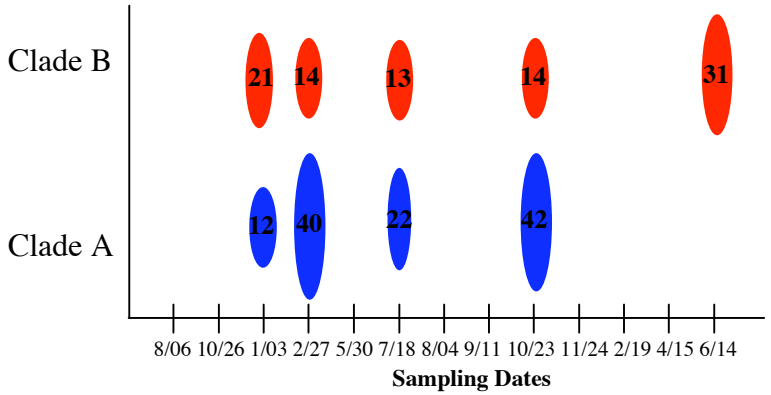
Figure 3.11. Crater Hills host population based on closest match in the public database (nearly full-length or full-length 16S rDNA clones over time).



Archaeal population determined by the 348Fa/UA1406R primer set. The 16S rDNA genes from five samples were amplified with the 20F/UA1406R primer set. An overview of the results with this primer set is illustrated in figure 3.12.

Thirty-six 16S rDNA clones from January 2003 were analyzed (table 3.3). The environmental sequences belong to clades A and B (figure 3.13). Eighteen *Sulfolobus*-like environmental sequences were 87-93% identical to the published *S. solfataricus* sequence (clades A and B in figure 3.13 and table 3.3). Nine environmental sequences most closely matched *C. noboribetus* with 90-95% identity (clade B in figure 3.13 and table 3.3). Three *V. distributa* sequences were detected; two of these clones showed 98% identity to the published *V. distributa* 16S rDNA sequence (clade B in figure 3.13 and table 3.3). In contrast to the sequences amplified with 20F/UA1406R, *S. solfataricus*-like sequences make up the majority of our environmental library (figure 3.11).

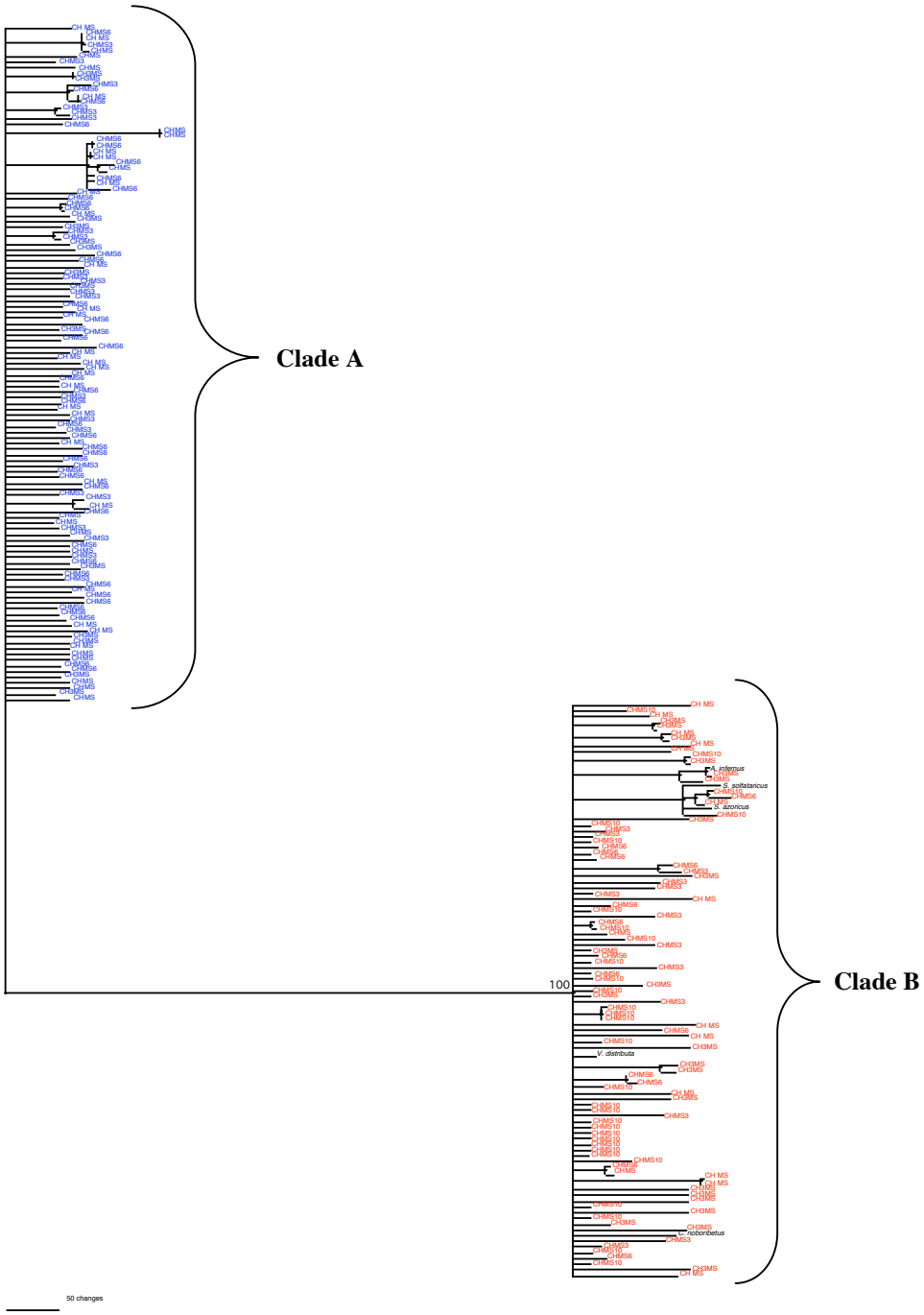
Figure 3.12. The archaeal population in Crater Hills over time measured with the 348Fa/1406R primer set. The numbers indicate how many 16S rDNA clones most closely matched the published16S rDNA sequence. The absence of sequences does not reflect an absence of 16S rDNA sequences from the environment; not all samples were used for 16S rDNA gene sequence analysis.



Another sample used for host analysis was collected on February 27, 2003. There were sixty-two clones obtained from this sample (table 3.3). The environmental sequences belong to clades A and B (figure 3.13). Fifty-two percent of the sequenced environmental clones (32/62) most closely matched *S. solfataricus* at 88-94% identity (clades A and B in figure 3.13 and table 3.3). Nine environmental clones most closely matched *V. distributa* at 93-98% identity (clade B in figure 3.13 and table 3.3). Of the nine clones, there were only two that matched the published *V. distributa* 16S rDNA gene at 98% (table 3.3). Eight clones most closely matched *C. noboribetus* at 91-96% identity, and seven clones most closely matched *S. azoricus* at 87-95% identity (clade B in figure 3.13 and table 3.3). Five clones were most similar to uncultured archaeal sequences with only 93% identity to the uncultured clones, and one clone was most similar to *Sulfolobus*

tokodaii (accession number AB022438) with 92% identity (clade A in figure 3.13 and table 3.3).

Figure 3.13. Environmental 16S rDNA sequences amplified with the 348Fa/UA1406R primer set from the Crater Hills monitor site.



The sample collected in July 2003 was also used for 16S rDNA gene sequence analysis (table 3.3). The dominant population in July 2003 appears to be *Caldococcus*-like organisms regardless of the primer set, but the 348Fa/UA1406R primer set amplified a more diverse population of species including *S. solfataricus*-like sequences. Nineteen *Caldococcus*-like sequences were amplified from this sample (clade B in figure 3.13 and table 3.3). The environmental sequences matched the published *Caldocccus* sequence at 88-95% identity (table 3.3). Seven environmental clones most closely matched *Vulcanisaeta* with 92-98% identity (clade B in figure 3.13 and table 3.3). Only two of these clones matched the published *Vulcanisaeta* sequence at 98% identity. Six environmental clones were 91-93% identical to *S. solfataricus* (clade A in figure 3.13 and table 3.3). The remaining clones matched *Pyrodictium occultum* (1/36) at 91% identity, and environmental clones isolated from Monteserrat (2/36) at 87% identity (clade A in figure 3.13 and table 3.3).

Table 3.3. The percent identities of environmental 16S rDNA clones from the Crater Hills monitor site to published 16S rDNA sequences in the public database.

Sampling Date	**Primer**	**Clones**	***Caldococcus***	***Vulcanisaeta***	***Sulfolobus***	***Acidianus***	***Other*****
1/3/03	20F	30	28 (89-95%)	ND	1 (92%)	1 (93%)	ND
1/3/03	348Fa	36	9 (90-95%)	3 (93-98%)	18 (87-93%)	2 (95-99%)	4 (88-93%)
2/27/03	348Fa	62	8 (91-96%)	9 (92-98%)	32 (88-94%)	ND	12 (87-95%)
7/18/03	20F	17	15 (89-97%)	2 (94-95%)	ND	ND	ND
7/18/03	348Fa	36	19 (88-95%)	7 (92-98%)	6 (91-93%)	ND	3 (87-91%)
10/23/03	348Fa	56	24 (92-94%)	19 (87-98%)	5 (91-93%)	1 (99%)	7 (83-90%)
2/19/04	20F	25	23 (86-96%)	1 (96%)	ND	ND	1 (84%)
6/14/04	348Fa	31	ND*	28 (92-98%)	1 (92%)	ND	2 (90-94%)

*ND=Not Detected

**Other=Clones matched *Desulfurococcus*, *Pyrodictium*, and environmental clones from Monteserrat.

A similar pattern was observed in the sample from October 23, 2003 (figure 3.11). Overall, *Caldococcus*-like sequences dominated the environmental library (figure 3.11). Twenty-four environmental clones matched *C. noboribetus* with 92-94% identity (clade B in figure 3.13 and table 3.3). *Vulcanisaeta*-like sequences were also detected in this sample (figure 3.11). Ten of the 19 clones were 98% identical to the published *V. distributa* 16S rDNA sequence (clade B in figure 3.13 and table 3.3). There were five *Sulfolobus*-like sequences detected in this sample (figure 3.11). All of these clones showed low identity to the published *Sulfolobus* sequences (clade A in figure 3.13 and table 3.3). The remaining clones showed low identity to *Desulfurococcus* and *Stygiolobus* 16S rDNA sequences (clade A in figure 3.13 and table 3.3).

In the last sample, the archaeal population in Crater Hills has shifted dramatically (figure 3.11). Thirty-one 16S rDNA clones were sequenced from the sample collected on June 14, 2004. All of the environmental 16S rDNA sequences belong to clade B (figure 3.13). Twenty-eight of the environmental clones most closely matched *Vulcanisaeta distributa* with 94-98% identity (figure 3.11 and table 3.3). Twenty-one of the clones matched the published *V. distributa* 16S rDNA sequence with 98% identity. Two sequences most closely resembled *Stygiolobus* (94% identity) and one other sequence matched *Sulfolobus* with only a 92% identity (figure 3.11 and table 3.3).

Overall, the archaeal population within this monitor site was diverse (figure 3.11). *Sulfolobus*-like, *Caldococcus*-like, and *Vulcanisaeta*-like sequences were the commonly detected 16S rDNA sequences in environmental libraries from this site. However, as discovered in the other monitor sites (see Chapters 4 and 5), the percent identities to

published 16S rDNA sequences were typically low (table 3.3). All of the detected sequences are likely representative of crenarchaeotes with low pH and high temperature growth optimums. Information concerning electron donors and acceptors for most of these organisms is unknown. However, based on the water chemistry, there are numerous potential donors and acceptors. Based on the pH, temperature, and water chemistry, it is reasonable the 16S rDNA sequences represent *Sulfolobus*-like, *Caldococcus*-like, and *Vulcanisaeta*-like organisms.

The presence of a primer bias was detected in the sample from January 2003; however, a bias was not as obvious in the sample collected in July 2003 (table 3.4). The 20F/UA1406R primer set resulted in the detection of four clades (figure 3.10) while the 348Fa/UA1406R primer set only detected two clades over time (figure 3.13). It is unknown why both primer sets were not successful in all samples. Perhaps, the archaeal population changed enough that one primer set would not amplify the 16S rDNA sequences. The optimal way to determine the archaeal population in this monitor site would investigate more samples and use numerous 16S rDNA primer sets.

Table 3.4. The presence of a primer bias in the January 2003 sample from Crater Hills; however, the same bias is not present in the July 2003 sample.

	CH3MS (January 2003)		***CHMS3 (July 2003)***	
	20F/1406R	**348Fa/1406R**	**20F/1406R**	**348Fa/1406R**
Sulfolobus	ND*	47%	ND	17%
Vulcanisaeta	ND	14%	12%	20%
Caldococcus	93%	14%	88%	52%
Stygiolobus	3%	4%	ND	ND
Acidianus	3%	9%	ND	ND
Uncultured**	ND	9%	ND	5%

*ND=Not Detected

**Uncultured=sequences most closely matched uncultured 16S rDNA sequences.

Population dynamics of viruses in the Crater Hills monitor site

SSV-like population dynamics in the Crater Hills monitor site. Total DNA was extracted from water collected from the Crater Hills monitor site and used as a source of template for PCR amplification of SSV-like sequences. There was not a discernable difference between the real time PCR and the standard PCR protocols. The primers were shown to be very sensitive on the purified viral DNA templates. The SSV primer set 3F/4R successfully amplified 0.1pg purified viral DNA while the 7F/8R primer set had a detection limit of 1ng purified viral DNA. The 3F/4R primer set was 10,000X more sensitive than the 7F/8R primer set. Therefore, the universal SSV primer set 3F/4R was used to measure the total diversity of SSVs in this monitor site. It is important to point out that the universal SSV primer set 7F/8R did not result in positive PCR products from this hot spring. There are numerous possibilities as to why this may be the case and they will be addressed later. An overview of SSV-like 3F/4R phylogenetic results is provided in figures 3.14 and 3.15.

The first sample was collected on August 6, 2002. Forty clones of SSV-like 3F/4R sequences were obtained that had ten different sequences (table 3.5). A phylogenetic tree was constructed with the clones from this sample illustrating one major group of SSV-like sequences (figure 3.16a). The 3F/4R sequences from the four sequenced SSV isolates are distinct from the SSV-like 3F/4R sequences from Crater Hills (figure 3.16a). When these sequences are translated to amino acids, there are a total of four inferred functional protein sequences and no non-functional protein sequences (table 3.6). The protein analysis resembles the DNA-based analysis (figure 3.16b).

Figure 3.14. The different groups of SSV-like viruses detected in the Crater Hills monitor site over time based on DNA sequence. Only functional SSV-like ORFs were included in the analysis. The colors of the groups correspond to figure 3.17. The empty sampling dates refer to the inability to amplify SSV-like 3F/4R sequences from the environment.

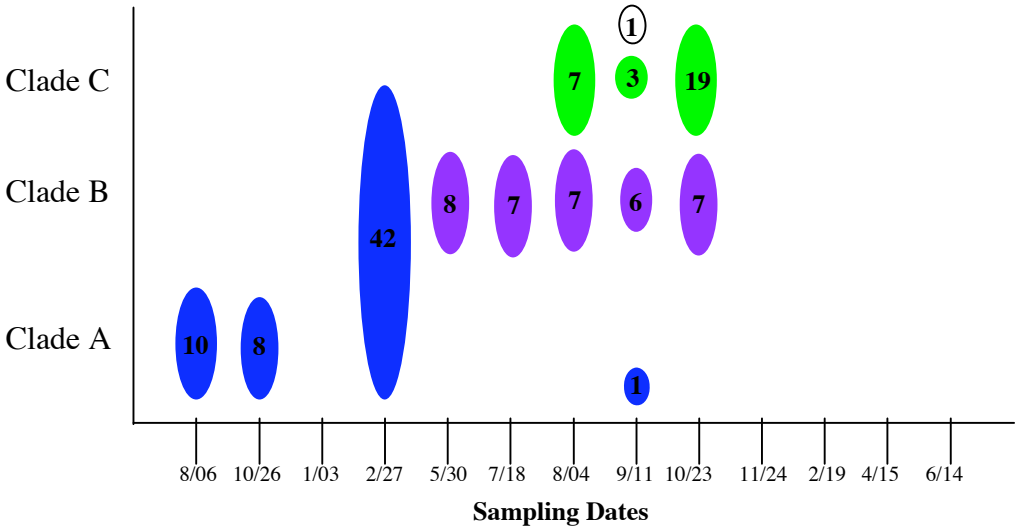


Figure 3.15. The different groups of SSV-like viruses detected in the Crater Hills monitor site over time based on protein sequence. The colors of the groups correspond to figure 3.18. The empty sampling dates refer to the inability to amplify SSV-like 3F/4R sequences from the environment.

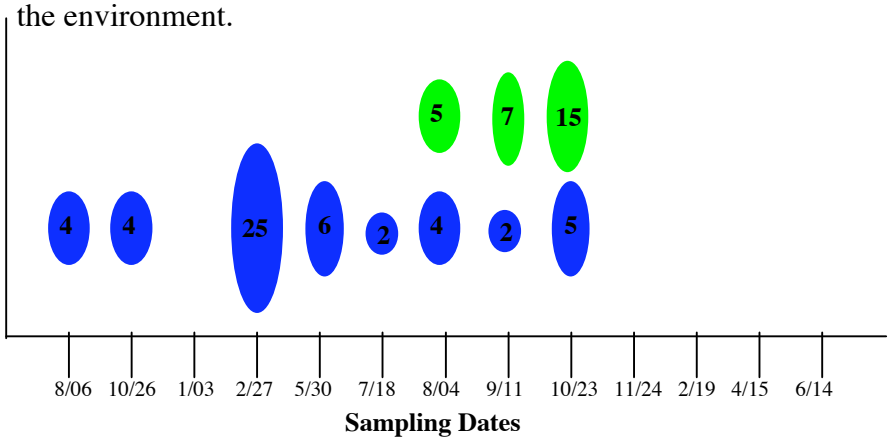
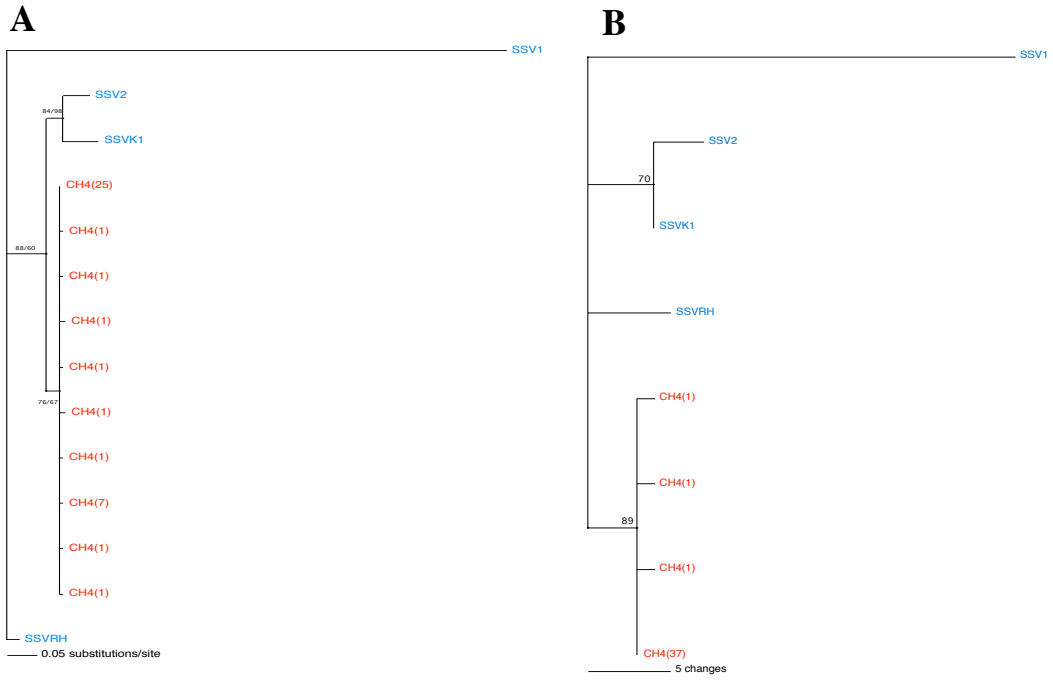


Figure 3.16. The SSV-like population detected in Crater Hills on August 6, 2002. The parentheses indicates the number of times a particular sequence was detected. (A) DNA sequences, and (B) inferred protein sequences.



Few changes in the SSV-like population structure were observed in the next sample collected from this monitor site. The second sample was collected on October 26, 2002, and 16 SSV-like 3F/4R clones were sequenced (table 3.5). When these sequences are added to the clones from the previous time point, they belong in the same group (clade A in figure 3.17). Even though two months had passed and the pH decreased by one unit (table 3.5), it does not appear that many changes occurred in the SSV-like population. When the sequences from this sample are translated, there are four inferred functional protein sequences and one non-functional protein sequences (table 3.6). The October 2002 protein-based analysis is similar to the DNA-based phylogenetic analysis in that the sequences belong to the largest group of protein sequences detected in this

monitor site (represented in blue in figure 3.17). There was one protein sequence from October 2002 that was identical to a sequence detected in August 2002 (figure 3.18; the * is positioned by the two identical sequences). The non-functional ORF contained two stop codons resulting from an insertion of an adenine (A).

Table 3.5. Number of SSV-like clones sequenced from the Crater Hills monitor site and how many unique DNA sequences in each library.

Sample Name	Collection Date	pH	Temp (°C)	Number of Clones	Unique Sequences
CH4	Aug 6, 2002	2.8	79	40	10
CHMS	Oct 26, 2002	1.9	80	16	8
CH3MS	Jan 3, 2003	2.2	72	0	0
CH_MS	Feb 27, 2003	2.2	81	71	52
CHMS2	May 30, 2003	2.1	85	10	8
CHMS3	July 18, 2003	2.4	86	26	7
CHMS4	Aug 4, 2003	2.6	80	56	22
CHMS5	Sept 11, 2003	1.6	67	17	12
CHMS6	Oct 23, 2003	2.6	80	56	31
CHMS7	Nov 24, 2003	2.3	74	0	0
CHMS8	Feb 19, 2004	2.1	80	0	0
CHMS9	Apr 15, 2004	2.2	82	0	0
CHMS10	June 14, 2004	1.9	81	0	0
Total Number of Clones Sequenced				**292**	**128**

The third sample was collected on January 3, 2003, the pH was 2.2 and the temperature was 72°C (table 3.5). However, no SSV-like sequences were amplified from this time point (figure 3.13). We do not believe this can be attributed to an experimental or technical problem. SSV-like 3F/4R sequences were not amplified even after repeated attempts, and both 16S rDNA and SIRV-like coat protein sequences were amplified from the same total environmental DNA. It is more likely the SSV-like population was at an abundance below detection level (<0.1pg in PCR reaction).

The SSV-like population began to expand within clade A in the fourth sample collected on February 27, 2003. When 71 SSV-like 3F/4R clones from this sample are added to the clones from the previous two sampling points, most of the SSV-like population is very similar to the two previous time points (clade A in figure 3.17). However, the SSV-like sequences within clade A are beginning to show differences. There are more outliers from this time point than the previous two; however, it might be assumed similar SSV-like sequences were present in the previous sampling time points, but there were not enough clones to detect them (40 clones from August 2002 and only 16 clones from October 2002, compared to 71 clones from February 2003) (table 3.5). Another possible reason for the outlying sequences could be that they are amplified from integrated copies of the SSVs.

Table 3.6. The comparison of SSV-like functional and non-functional protein sequences detected from the Crater Hills monitor site. For a description of terms, see text.

Sample Date	Sample Name	Unique DNA Sequences	Functional Protein Sequences	Non-Functional Protein Sequences
August 6, 2002	CH4	10	4	0
October 26, 2002	CHMS	8	4	1
February 27, 2003	CH_MS	52	25	10
May 30, 2003	CHMS2	8	6	0
July 18, 2003	CHMS3	7	2	1
August 4, 2003	CHMS4	23	9	9
September 11, 2003	CHMS5	12	9	0
October 23, 2003	CHMS6	31	20	5
TOTAL		**128**	**74**	**26**

This sample contained 25 inferred functional protein sequences and ten non-functional protein sequences translated from 52 unique DNA sequences (table 3.6). The protein-based analysis on this sample does not reflect the changes obvious in the DNA-

based analysis. The inferred protein sequences from this sample are placed in the group of protein sequences from earlier samples (blue in figure 3.18). The environmental SSV-like sequences with long branch lengths represent the interrupted ORFs (see Appendix A, figure 3.29). Upon closer investigation of these sequences, they contain numerous stop codons.

The emergence of a new SSV-like clade, and the apparent disappearance of clade was observed when the fifth sample was collected. This sample was collected on May 30, 2003, and only 10 clones were sequenced from SSV-like viruses (table 3.5). The SSV-like sequences from this sample (May 2003) comprise a distinct group from the rest of the SSV-like sequences that were collected at previous time points (clade B in figure 3.17). It is interesting to point out that the sequences collected on May 30, 2003 are more similar to SSV-K1 (isolated from the Kamchatka peninsula in Russia) than they are to sequences detected three months previously from the same hot spring. Perhaps with more clones we would see SSV-like 3F/4R sequences related to the sequences in clade A from previous months. Regardless, the formation of a new viral clade is still evident. The formation of the new SSV-like group is only apparent in the DNA-based analysis. When the sequences from this sample are translated, there are six inferred functional protein sequences that fall into the same group as all of the previous protein sequences thus far (represented in blue in figure 3.18).

The SSV-like population shows change in the following sampling time. When 26 SSV-like 3F/4R clones from July 18, 2003 (table 3.5) are added to the previous sequences, they are related to the sequences from May 2003 from clade B (figure 3.17).

When the DNA sequences from this sample are translated, there are two inferred functional protein sequences and one non-functional protein sequence (table 3.6). The functional protein sequences group with all of the protein sequences detected previously in this monitor site (represented in blue in figure 3.18). The non-functional protein sequence was similar to interrupted ORFs detected in February 2003 (see Appendix A, figure 3.31). Upon closer investigation, the ORF contains a frameshift that resulted in numerous stop codons throughout the sequence.

The sample collected in August 2003 reveals the viruses are undergoing change in this monitor site. Previous to this sampling point, there had been little change within the SSV-like population. The sample collected on August 4, 2003 resulted in 56 SSV-like 3F/4R clones (table 3.5). When these SSV-like sequences are added to the DNA-based phylogenetic analysis, more change in the viral population is apparent. Sequences detected from this time point fall into the group closely related to SSV-K1 (clade B in figure 3.17). Another new group of viral sequences is evident that is more closely related to SSV1, and is very distinct from all of the previous sequences detected thus far (clade C in figure 3.17). Again, it is interesting to point out that some SSV-like sequences from this sample are more closely related to SSV1 or SSV-K1 than they are to other viral sequences from this monitor site (figure 3.17).

Based on the functional protein sequences, there has been a significant change in the SSV-like population (figure 3.18). When the sequences from this time point are translated, there are nine functional inferred protein sequences and 10 non-functional protein sequences (table 3.6). The inferred protein sequences detected from this sample

represent the two distinct groups of viruses detected in this monitor site (figure 3.16). Like the DNA-based analysis, the protein-based analysis reveals some SSV-like protein sequences are closely related to SSV1 (represented in green in figure 3.18). Three of the 10 non-functional ORFs contain numerous stop codons, and seven of the interrupted ORFs contain a first position transversion in amino acid seven changing a glycine (GGA) to a stop codon (TGA) (see Appendix A, figure 3.32).

Surprisingly, there is little change in the viral population when 17 clones from September 11, 2003 are added to the analysis (table 3.5). The pH drastically dropped from 2.6 in August 2003 to 1.6 at this sampling point and the temperature also dropped from 80°C to 67°C (table 3.5). More viral sequences detected in this sample are related to the group composed of SSV-like sequences closely related to SSV-K1 (clade B in figure 3.17). Environmental viral sequences are detected that belong in the new group detected in August 2003 that is closely related to SSV1 (clade C in figure 3.17), and there are a couple of sequences that are most related to sequences first detected during the February 2003 sampling point (clade A in figure 3.17).

When the viral SSV-like 3F/4R DNA sequences from this sample are translated, there are nine inferred functional protein sequences (table 3.6). These nine sequences fall into the two previously established groups of protein sequences (figure 3.18). Two protein sequences from this sample were identical to protein sequences detected in May 2003, and one protein sequence was identical to a sequence detected in August 2003 (figure 3.18; the * represents the groups of identical sequences).

Figure 3.17. The functional ORFs from SSV-like DNA sequences detected in Crater Hills over time. The number in parentheses indicates how many times a particular sequence was detected. Inset: an unrooted phylogram to show the three major SSV-like virus groups detected in Crater Hills.

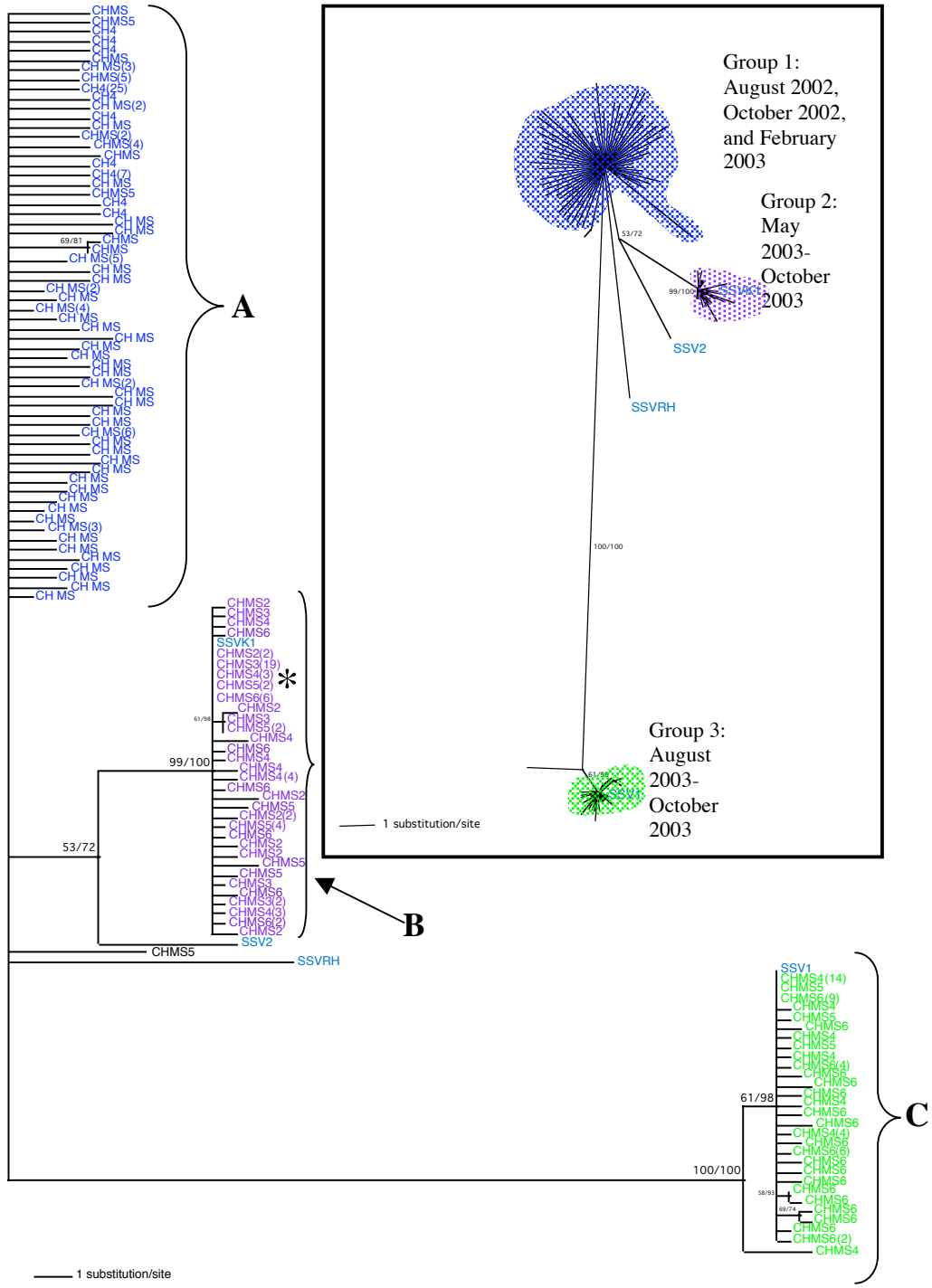
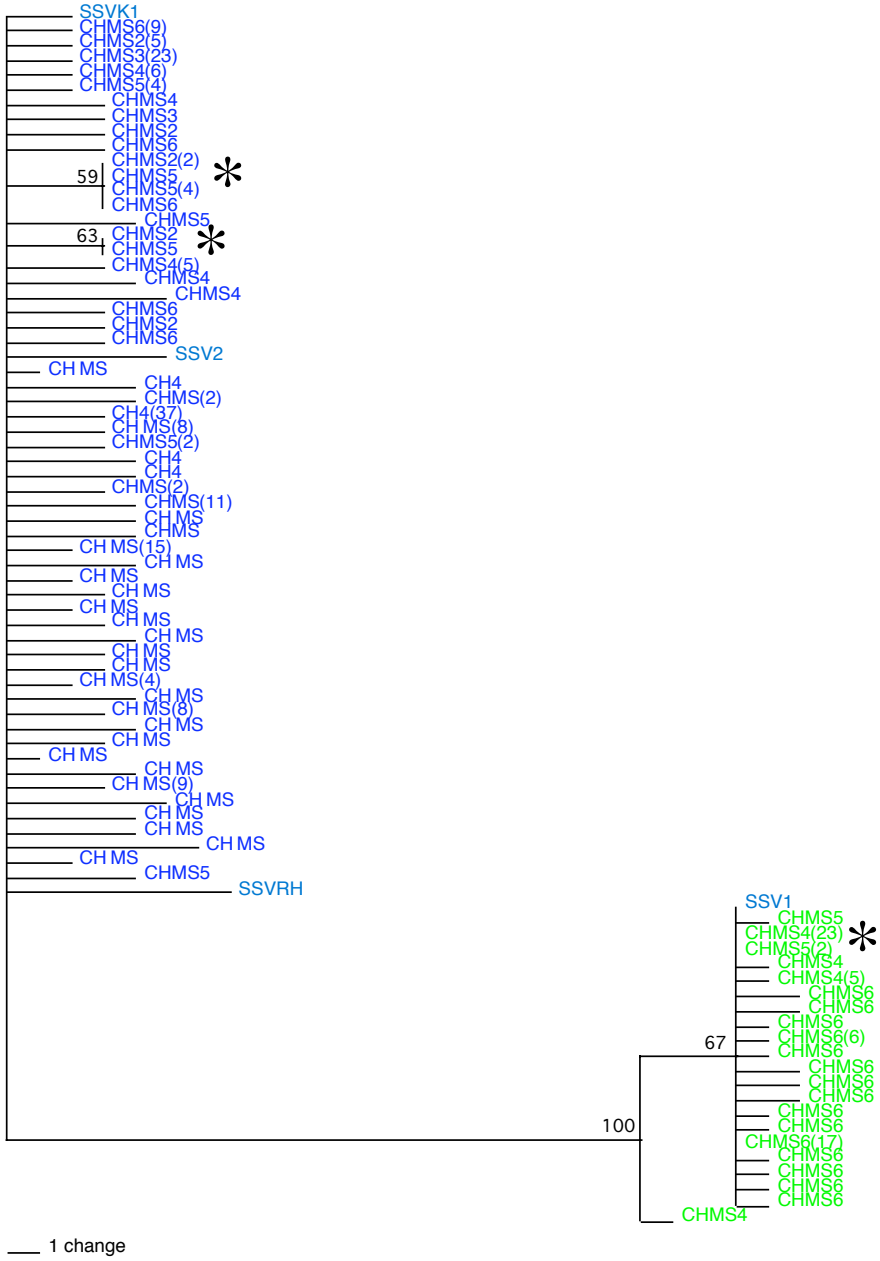


Figure 3.18. The SSV-like functional protein sequences detected in Crater Hills from August 2002 through October 2003. Each color represents a group of different viral sequences. The * represent identical protein sequences detected over time. The number in parentheses indicates how many times a particular protein sequence was detected.



The SSV-like virus population is showing little change in the October 2003 sample. This sample was collected on October 23, 2003, and 56 SSV-like 3F/4R clones were sequenced (table 3.5). The SSV-like 3F/4R sequences collected from this time point fall into previously established groups (clades B and C in figure 3.17). There are two sequences detected in this sample that are distinctly different from SSVs sequenced previously (see Appendix A, figure 3.36), but they might represent integrated copies of virus.

When the SSV-like 3F/4R sequences from this sample are translated, there are 20 inferred functional protein sequences and five non-functional sequences (table 3.6). As with the DNA-based analysis, the protein-based phylogenetic tree shows little change from the previous months (figure 3.18). There was only one protein sequence detected in this sample with an identical sequence to those found in previous samples (figure 3.18; the * represents the identical group of protein sequences). Three of the interrupted ORFs contain numerous stop codons, and two of the non-functional sequences reveal a transversion in the first position of a glycine (GGA) that results in a stop codon (TGA) (see Appendix A, figure 3.36).

Four more samples were collected during the time period from November 2003—June 2004, but there were not any SSV-like 3F/4R sequences detected from these last four samples (table 3.5). The environmental DNA extracted from these samples resulted in 16S rDNA and SIRV-like coat protein sequences. Therefore, I do not believe it was an experimental issue, but rather the abundance of the SSV-like viruses dropped below our detection limits (<0.1pg/PCR reaction).

In summary, from August 2002 to June 2004 a total of 292 SSV-like 3F/4R clones were sequenced (table 3.5). There appeared to be three dominant groups, based on 3F/4R DNA sequence, of SSV-like sequences over time with a few outlying sequences (figure 3.17). In May 2003, a switch from clade A to clade B was observed (figure 3.14). In August 2003, a switch from a single clade (clade B) to a mixture of two clades (clades B and C) was observed (figure 3.14). When comparing protein sequences, there are two distinct groups, comprised of 74 unique protein sequences, of viruses over time within this monitor site (figure 3.18). Unlike the DNA-based phylogenetic analysis, one clade of viral protein sequences was consistently detected with a second clade detected in August 2003 (figure 3.15). The complete phylogenetic analysis of SSV-like sequences included analysis of sequences after collection of every sample. For a step-by-step view of the changes in the SSV-like population over time, see Appendix A.

The SSV-like 3F/4R sequences detected in Crater Hills vary with time. The first group (clade A; represented by all blue sequences in figure 3.17) is represented by sequences from August 2002, October 2002, and February 2003; these were the dates of the first three samples collected from Crater Hills (table 3.5). There are also two sequences from the September 2003 sampling point (figure 3.17). The second group of sequences (clade B) is related to SSV-K1 and contains sequences from May 2003 through September 2003 with one large group of identical sequences (represented by the * in figure 3.17). The third group of SSV-like sequences from Crater Hills is similar to SSV1 from Japan (clade C in figure 3.17). This group of viruses is only composed of samples collected later in the study (August 2003, September 2003, and October 2003).

It is interesting the environmental SSV-like sequences were not more similar to SSV-RH. It was suspected the YNP environmental SSV-like sequences would be more related to SSV-RH than to the other three sequenced SSV isolates (SSV1, SSV2, and SSV-K1) because it was isolated from YNP.

It is interesting to compare the phylogenetic positioning of functional and non-functional SSV-like 3F/4R DNA sequences (figure 3.19). The non-functional ORFs were included in a DNA-based phylogenetic analysis. Many of the non-functional ORFs are similar to functional ORFs from this monitor site (figure 3.19). When the non-functional ORFs are not included in the SSV-like DNA analysis, the topology of the phylogenetic tree is similar, but the presence of three distinct groups of SSV-like viruses is more obvious (figure 3.17). The long branch lengths of the interrupted ORFs made it difficult to separate some of the viral groups close to the backbone of the tree because of the close branches (figure 3.19). However, without the long branches in the DNA-based analysis, it is much easier to visually determine the presence of three SSV-like viral groups present from August 2002 to October 2003 (figure 3.17).

Phylogenetic analysis of the inferred protein sequences revealed two major groups of SSV-like viruses (figures 3.18). The virus group closely related to SSV1 (represented by green in figure 3.18) appeared in this monitor site in August 2003, but the other major group of protein sequences was consistently detected (figures 3.15 and 3.18). By translating the SSV-like 3F/4R sequences retrieved from the environment, we could determine what sequences may be representing old integrated forms of SSVs. The interrupted ORFs were not included in most of the protein-based analyses. However,

they were included in an analysis of all the SSV-like inferred protein sequences isolated from this monitor site (figure 3.20).

Figure 3.19. All SSV-like 3F/4R DNA sequences detected in the Crater Hills monitor site over time. This analysis included the SSV-like DNA sequences that resulted in non-functional ORFs. The interrupted ORFs are colored green. The clades from figure 3.14 are indicated on this tree.

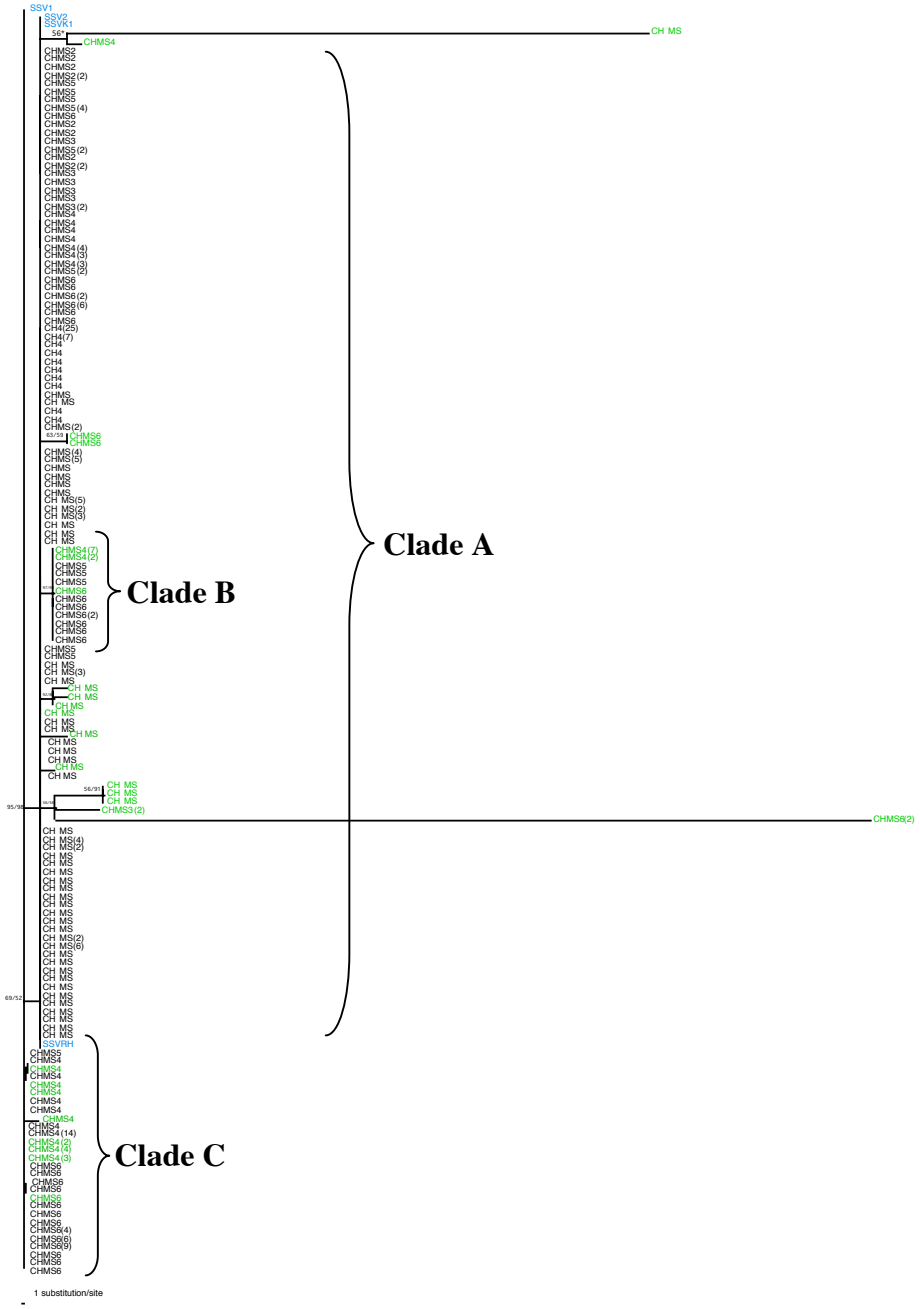
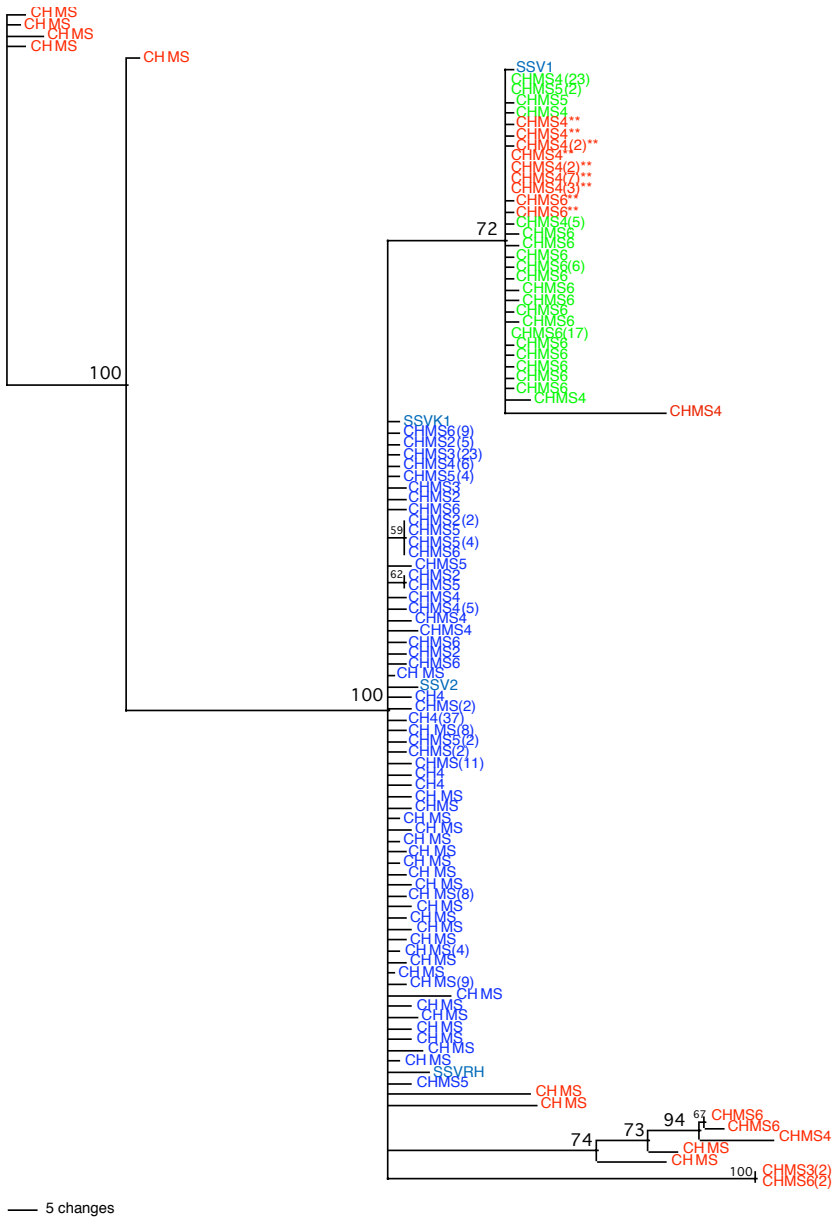


Figure 3.20. All of the SSV-like protein sequences detected in the Crater Hills monitor site. The colors of the two protein groups are the same as figure 3.18, and the red sequences represent the interrupted ORFs detected over time.



There are a couple of interesting results when the 26 non-functional protein sequences from all of the samples collected (table 3.6) are added to analysis (figure 3.20). Fourteen of the interrupted SSV-like ORFs are very distinct from the inferred functional protein sequences (figure 3.20; the interrupted ORFs are represented in red). When these inferred protein sequences are investigated more closely, there are many insertions, deletions, transversions, and transitions resulting in numerous stop codons throughout the sequence. The other non-functional protein sequences are closely related to the group of functional protein sequences related to SSV1 (figure 3.20; represented in red with **). It was unexpected that these sequences were closely related to functional protein sequences when some of the non-functional sequences were so distinct (figure 3.20). However, when these protein sequences are more closely examined, there is a first position transversion in amino acid seven changing it from a glycine (GGA) to a stop codon (TGA). Besides that change, these protein sequences are very similar to the group of functional protein sequences related to SSV1 (represented by green in figure 3.20). I do not believe the errors in the sequences can be contributed to PCR or sequencing errors. During the course of this research, there were not any interrupted ORFs detected in the SIRV-like (discussed below) or the 16S rDNA sequences. If the mistakes were the result of sequencing or PCR errors, they should also be detected in these other sequences.

Of the 74 protein sequences isolated from Crater Hills, there were only three sequences that were detected in more than one sample (figure 3.18; the * represents groups of identical protein sequences). The largest group of SSV-like viruses based on 3F/4R inferred protein sequences was detectable in the samples that SSV-like sequences

were detected (represented in blue in figure 3.18). However, the smaller group related to SSV1 was not detected until August 2003 (represented in green in figure 3.18). It is also interesting to note the changes in DNA sequences did not necessitate changes in the protein sequences (figures 3.14 and 3.15).

Detailed analysis of SSV-like population dynamics. From August 2002 to June 2004, there was a total of 292 SSV-like 3F/4R clones sequenced from this site (table 3.5). Using Modeltest version 3.6 (Posada *et. al.* 1998), the nucleotide substitution model for maximum likelihood estimations was determined to be a general time reversible (GTR) model (Rodriguez *et. al.* 1990) with heterogeneity between sites (gamma=2.1351), invariant sites (I=0.0091), base frequencies A=0.2773, C=0.1954, G=0.2382, and T=0.2891, and a substitution model A-C=1.4833, A-G=1.6068, A-T=0.9438, C-G=0.4277, C-T=2.3008, and G-T=1.0000.

Of the 292 SSV-like 3F/4R clones sequenced during the course of this research, 128 groups of identical sequences were detected. The most commonly detected SSV-like sequence was found 33 times, which composes 11.3% of the SSV-like population (table 3.7). This group of SSV-like sequences is only composed of sequences from August and October 2002 (table 3.7). The next most frequently detected sequence was found 32 times, which composes 10.9% of the SSV-like population (table 3.7). This group of sequences was found in most of the samples that were collected from this monitor site, so it is likely this is a group of SSV-like viruses that were consistently present during our sampling period (table 3.7). A group of 24 sequences was also detected, which composed 8.2% of the SSV-like virus population (table 3.7). This group was composed

of sequences from later samples (August, September, and October 2003); therefore, it seems to be a group that was identified later, but persisted until the end of the sampling (table 3.7).

Table 3.7. Crater Hills SSV-like 3F/4R sequence abundance illustrating the groups of identical sequences, how many times a group this size was detected, the percent of the SSV-like population, and what samples the group contained.

CRATER HILLS SSV-like SEQUENCE ABUNDANCE			
Identical Sequences	Times Detected	Percent of SSV-like Population	Samples[a]
33	1	11.3%	CH4;CHMS
32	1	10.9	CHMS2;CHMS3;CHMS4;CHMS5;CHMS6
24	1	8.2	CHMS4;CHMS5;CHMS6
10	1	3.4	CHMS4;CHMS5
10	1	3.4	CH4;CHMS;CH_MS
8	1	2.7	CHMS3;CHMS4;CHMS6
7	1	2.4	CHMS4
5	3 unique[b]	1.7% each[c]	
4	3 unique	1.4% each	
3	7 unique	1.1% each	
2	11 unique	0.68% each	
1	97 unique	0.34% each	

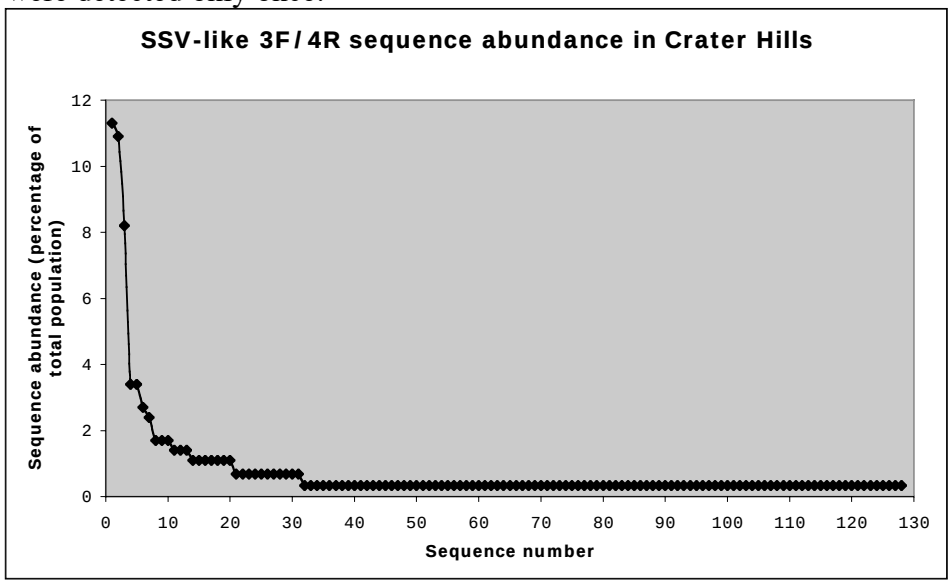
[a]sequences from these samples were present in the indicated groups
[b]each of the smaller groups were detected this many times, however, they are unique

Two unique groups composed of 10 sequences each were also detected (table 3.7). The first group of ten was composed of sequences from August and September 2003, and the second group of ten identical sequences included earlier samples (August 2002—February 2003) (table 3.7). The groups of ten each made up 3.4% of the SSV-like population. One group of eight sequences from July—August 2003 and October 2003 made up 2.7% of the SSV-like population, and one group of seven sequences from

August 2004 composed 2.4% of the SSV-like population (table 3.7). There were other groups of identical sequences detected ranging in size from two to five identical sequences (table 3.7).

Unique SSV-like sequences made up the majority of the SSV-like population in the Crater Hills monitor site (figure 3.21). There were 99 unique SSV-like 3F/4R sequences, which translates to 33% of the SSV-like population over the course of sampling (figure 3.21).

Figure 3.21. A sequence abundance plot for SSV-like 3F/4R sequences from Crater Hills monitor site. The left end of the x-axis represents most commonly detected sequences, and the right end of the x-axis represents the most rare sequences. Most of the SSV-like sequences in Crater Hills were detected only once.



In comparing the SSV-like DNA sequences to the SSV-like protein sequences, one can predict the changes in the DNA sequence will most likely be in the third position in the codon. The rate of transitions (Ti) and transversions (Tv) is almost equal to each

other (Ti=3.9076 and Tv=3.8548; Ts/Tv=0.8). When comparing the base changes in each position of the codon, it is as predicted, the third position base is changing more frequently than the first or second base positions (table 3.8).

Table 3.8. Numbers of transitions and transversions in SSV-like sequences collected from the Crater Hills monitor site.

SAMPLE DATE	IDENTICAL BASES	TRANSITIONS			TRANSVERSIONS		
		1st	2nd	3rd	1st	2nd	3rd
08/06/2002	258/259	1	0	0	0	0	0
10/26/2002	257/259	0	0	0	1	0	1
02/27/2003	251/259	3	0	3	1	0	1
05/30/2003	255/259	0	0	2	1	0	1
07/18/2003	258/259	0	0	0	0	1	0
08/04/2003	205/255	7	4	16	8	5	10
09/11/2003	215/257	6	2	14	6	4	9
10/23/2003	221/253	5	2	11	3	2	7

In comparing the 128 unique SSV-like DNA sequences (3F/4R) obtained over time, 194 bps (out of an average of 256 bps) did not change in all of the samples (table 3.9). When comparing all of the SSV-like sequences, 76% of the bases in the 3F/4R sequence remained unchanged (table 3.9). Transversions occurred slightly more (35) compared to transitions (27). Nineteen changes occurred in the first position of the codon; there were eight transitions and 11 transversions (table 3.9). Only 14 changes occurred in the second position with five transitions and nine transversions (table 3.9). A total of 28 changes occurred in the third position; there were 14 transitions and 14 transversions (table 3.9). The first five samples collected from Crater Hills contained sequences that were very similar to each other, but the last three samples revealed as increased diversity within the sequences (table 3.8). As expected, the third position

changes occurred more frequently, therefore explaining the fewer number of groups based on protein sequences (figure 3.18).

Table 3.9. The numbers of transitions and transversions for all SSV-like sequences collected from Crater Hills throughout this research.

Sample Range	Position	Identical Base Pairs	Ts Pairs	Tv Pairs
Aug 2002-Oct 2003 (256 bps)	*Overall*	*194/256*	*27*	*35*
	1st	66/85	8	11
	2nd	72/86	5	9
	3rd	56/85	14	14

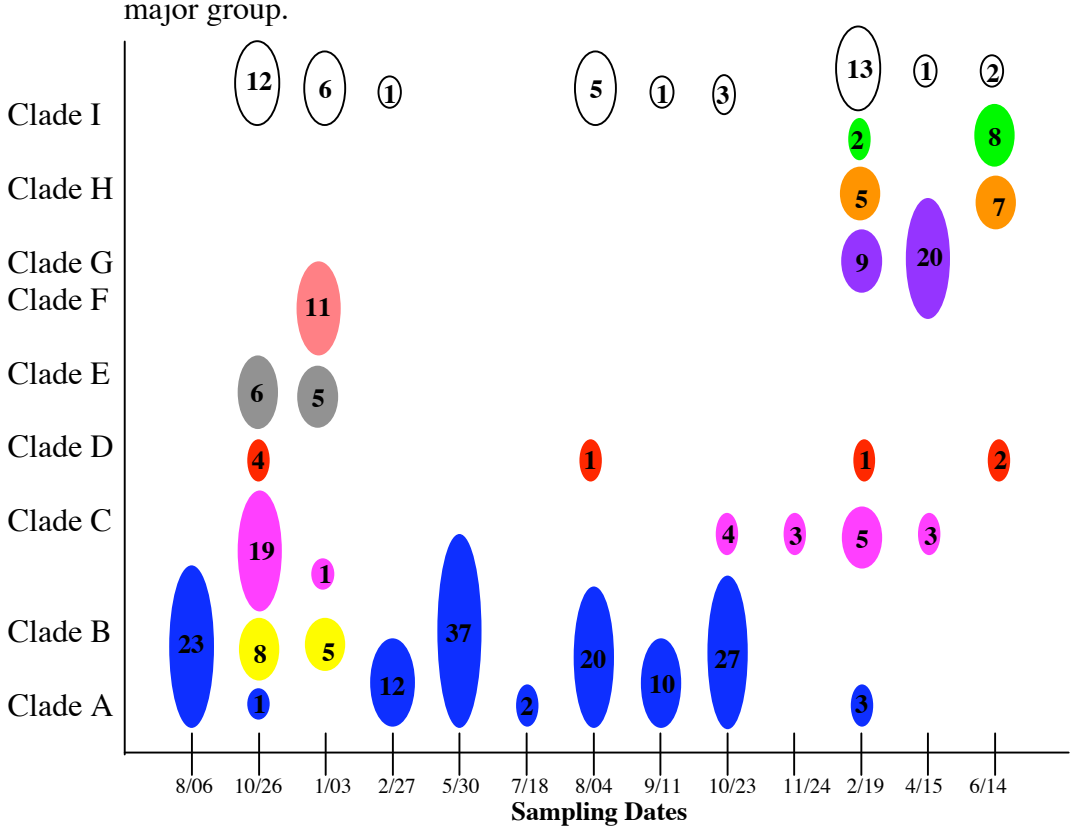
Ts=transitions; Tv=transversions

SIRV-like population dynamics in the Crater Hills monitor site. The same total environmental DNA extracted from the monitor site and used for the SSV-like 3F/4R PCR was used to amplify the coat protein gene from SIRV-like viruses. The primer set designed to amplify the coat protein gene of SIRV-like viruses (DBPF/R) was found to amplify 0.001pg (1 femptogram) of purified viral template. This primer set was 100X more sensitive than the SSV 3F/4R primer set and 1,000,000X more sensitive than the SSV 7F/8R primer set. An overview of the phylogenetic results from SIRV-like sequences detected in Crater Hills is illustrated in figures 3.22 and 3.23.

The first sample was collected on August 16, 2002 (table 3.10). Forty-three clones were analyzed and compared to the SIRV1 and SIRV2 coat proteins sequenced from viruses isolated in Iceland. As seen in figure 3.24a, the SIRV-like sequences show little diversity within the monitor site population; however, they are distinctly different from the viruses isolated from Iceland. The inferred protein sequences of the SIRV-like coat proteins from August 2002 are comparable to the DNA sequences (figure 3.24b).

Both the DNA-based and protein-based analysis revealed similar phylogenetic structures. The DNA sequences may reveal a slightly more diverse SIRV-like population.

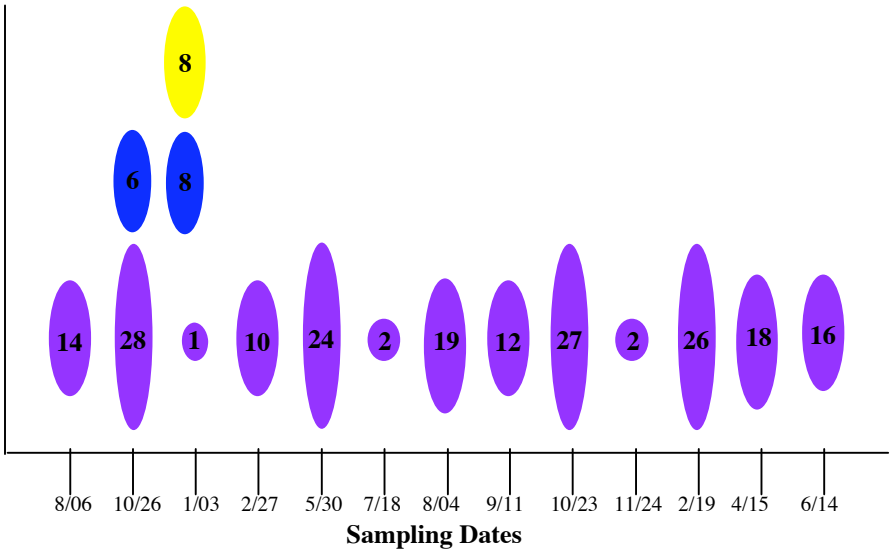
Figure 3.22. The different groups of SIRV-like viruses detected in the Crater Hills monitor site over time. The colors of the groups correspond to figure 3.25. The clear ovals refer to outlying sequences not belonging to a major group.



The next sample reveals diversification of the SIRV-like coat protein sequences. Eighty-four clones were obtained from October 26, 2002 (table 3.10) and analyzed with the previous sampling point. Within two months, the SIRV-like population has dramatically changed (figure 3.22). There is only one SIRV-like sequence from October 2002 that clusters with the sequences from August 2002 (clade A in figure 3.25). The SIRV-like population is much more diverse than it was in August 2002, forming at least

four new groups (clades B, C, D, and E) with numerous outlying sequences (figure 3.25). However, the inferred protein sequences do not reveal as much diversity indicating that most base changes are in the third position of the codon. There is a new group present in the protein tree (protein group represented by blue in figure 3.26). There was one protein sequence detected in this sample that was identical to the dominant sequence detected in the August 2002 sample (figure 3.26; the identical sequence is marked with *).

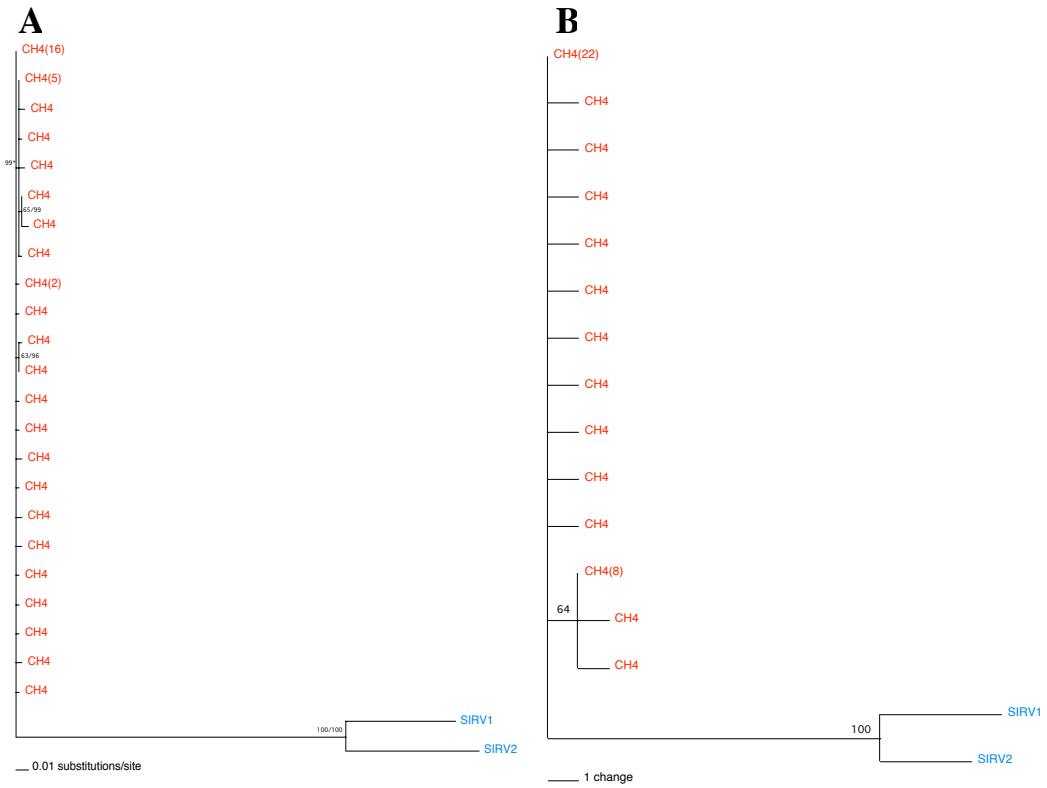
Figure 3.23. The different groups of SIRV-like viruses detected in the Crater Hills monitor site over time based on protein sequences. The colors of the groups correspond to figure 3.26.



Continued change is apparent in the SIRV-like population in the next sample. This sample was collected in January 2003, and 46 clones of SIRV-like coat protein sequences were obtained (table 3.10). When analyzed with sequences from August 2002 and October 2002, the SIRV-like sequences from January 2003 represented numerous groups (figure 3.22). Twenty-three SIRV-like coat protein sequences formed a new group (clade F in figure 3.25), seventeen sequences were associated with the October

2002 sequences in a dominant group from that time point (clade B in figure 3.25), and six sequences were placed into another group with October 2002 sequences (clade E in figure 3.25). Similar to the DNA-based phylogenetic analysis, the inferred protein sequences reveal the formation of a new group and new sequences added to existing groups (figure 3.26). The new protein group (colored orange in figure 3.26) is distinct from the other two groups detected in this sample (represented by purple and blue in figure 3.26). The dominant inferred protein sequence detected in the previous two samples was not detected in this sample.

Figure 3.24. The SIRV-like sequences obtained from Crater Hills in August 2002. The number in parentheses indicates how many times a sequence was detected. (A) DNA sequences, and (B) inferred protein sequences.



Although the diversity of SIRV-like population has been steadily increasing, the diversity decreases in the fourth sample collected from this monitor site. This sample was collected in February 2003. Eighty-three SIRV-like coat protein clones were obtained (table 3.10). This time point revealed a change in the viral population. All of the SIRV-like sequences from this sample were more closely related to sequences that had been detected in August 2002 (clade A in figure 3.25). All of these sequences were very similar to each other, in fact of the 83 clones, 50 clones had identical sequences. The inferred amino acid sequence also reveals the decrease in diversity from numerous groups to one group (represented in purple in figure 3.26). However, the dominant inferred protein sequence detected in the first two samples was detected in this sample after being undetectable in January 2003 (figure 3.26; the identical sequences are represented by *).

Table 3.10. Number of SIRV-like clones obtained from the Crater Hills monitor site, number of unique sequences in each library, and number of unique protein sequences.

Sample Name	Collection Date	pH	Temp (°C)	Number of Clones	Unique Sequences	Protein Sequences
CH4	Aug 6, 2002	2.8	79	43	23	14
CHMS	Oct 26, 2002	1.9	80	84	50	34
CH3MS	Jan 3, 2003	2.2	72	46	28	17
CH_MS	Feb 27, 2003	2.2	81	83	13	10
CHMS2	May 30, 2003	2.1	85	116	37	24
CHMS3	July 18, 2003	2.4	86	78	2	2
CHMS4	Aug 4, 2003	2.6	80	60	27	19
CHMS5	Sept 11, 2003	1.6	67	41	15	12
CHMS6	Oct 23, 2003	2.6	80	62	34	27
CHMS7	Nov 24, 2003	2.3	74	42	3	2
CHMS8	Feb 19, 2004	2.1	80	67	42	26
CHMS9	Apr 15, 2004	2.2	82	43	23	18
CHMS10	June 14, 2004	1.9	81	47	19	16
Total Number of Clones Sequenced				**812**	**280**	**211**

The diversity of the SIRV-like population continues to decrease in the next sample. The fifth sample was collected on May 30, 2003 (table 3.10). One hundred and sixteen SIRV-like coat protein clones were sequenced from this sampling point. As seen with the February 2003 time point, these clones closely resembled the August 2002 viral sequences (clade A in figure 3.25). This sample contained numerous identical sequences. For example, one sequence was represented 52 times in the environmental library. There were 37 unique DNA sequences that translated to 24 unique inferred protein sequences (table 3.10). All of the sequences from this sample belong in the largest group of SIRV-like protein sequences (represented in purple in figure 3.26). As with previous samples, only one sequence was detected from this sample that was identical to protein sequences detected in previous samples (figure 3.26).

The sixth sample reveals the lowest amount of SIRV-like diversity detected in one sample thus far. This sample was collected on July 18, 2003, and 78 clones were obtained; however, only two different sequences were detected from this sample (table 3.10). One sequence was detected 69 times and the other sequence was detected only nine times. The two sequences grouped with the SIRV-like sequences from May 2003 (clade A in figure 3.25). The inferred protein sequences are also similar to sequences from previous samples with all belonging to the largest group of protein sequences (represented by purple in figure 3.26). One of the two DNA sequences from this sample is identical at the protein level to the most abundant and dominant group of SIRV-like sequences (figure 3.26).

Figure 3.25. All of the SIRV-like DNA sequences collected from the Crater Hills monitor site. The different colors represent different groups of viruses and correlate with figure 3.22.

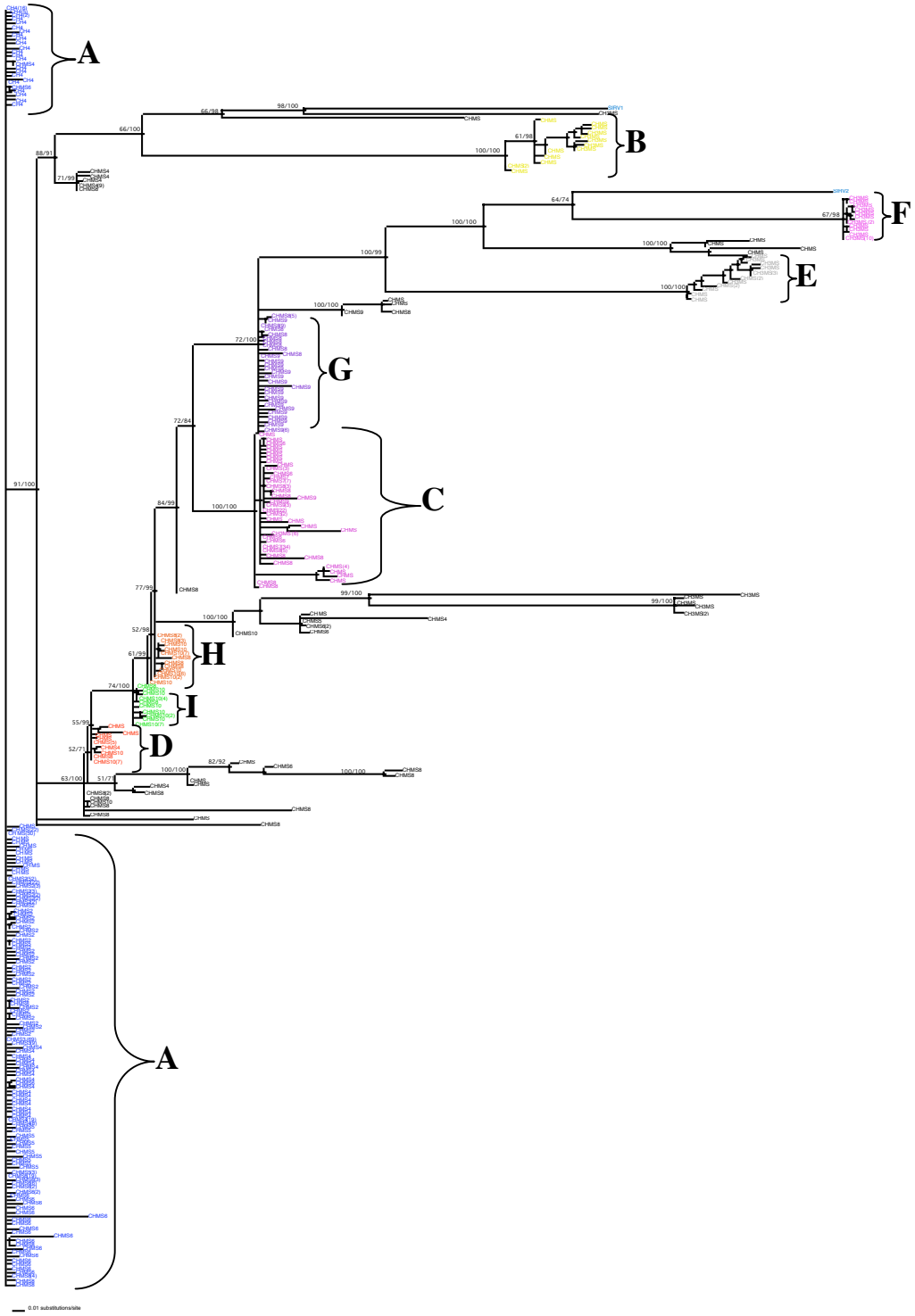
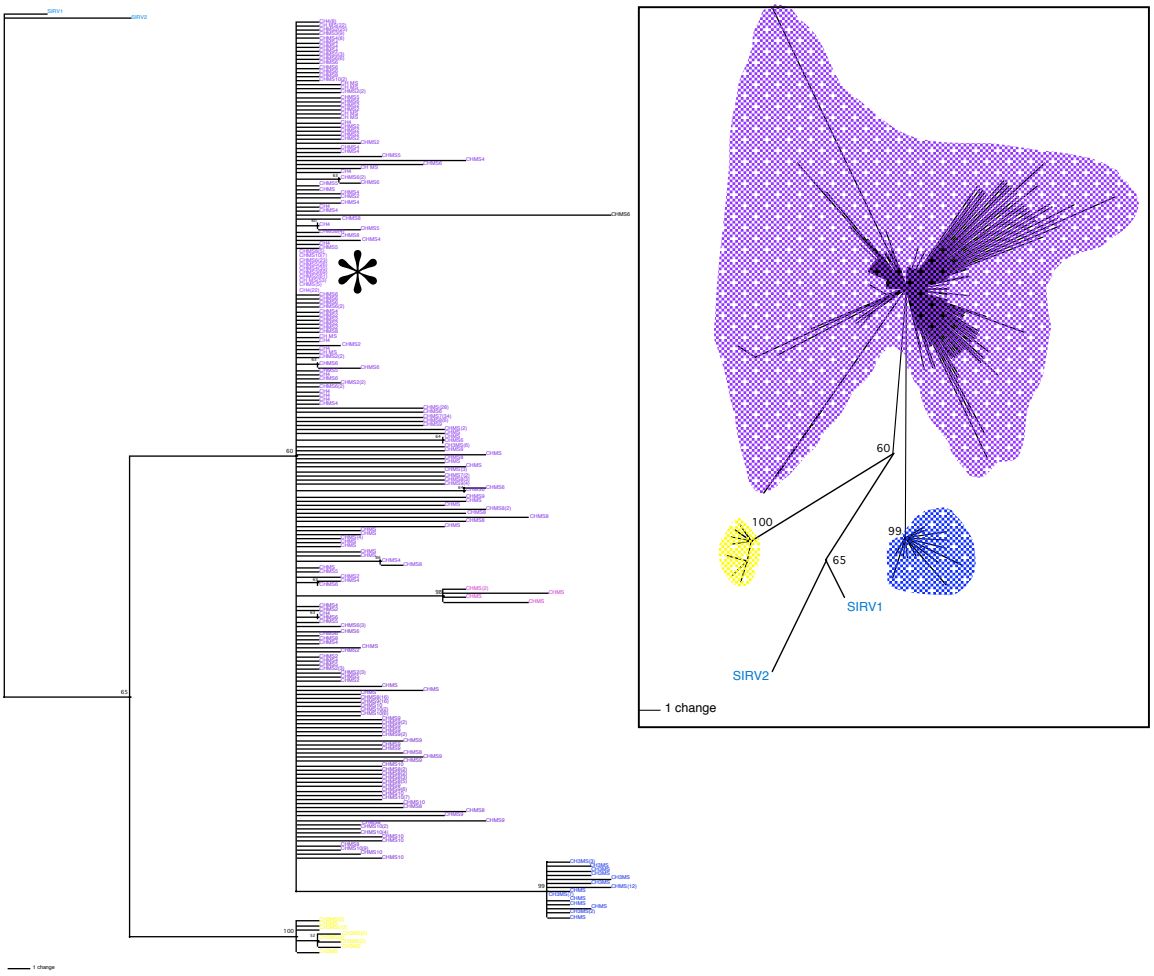


Figure 3.26. All of the SIRV-like protein sequences collected from the Crater Hills monitor site. The * represents identical sequences. Inset: unrooted phylogram illustrating the three SIRV-like groups based on protein sequence.



After months of little change in the SIRV-like population, changes in the virus population begin to appear. The next sample was collected on August 4, 2003 (table 3.10). Sixty SIRV-like coat protein clones were sequenced from this sample comprising 27 different identical DNA sequences (table 3.10). The majority of the sequences from this sample (August 2003) fall into a previously established group (clade A in figure 3.25). However, there are sequences that belong in a group that is quite distinct from the

other August 2003 sequences (clade D in figure 3.25). The inferred protein sequences of the SIRV-like coat protein are much more conserved than the DNA sequences; all the protein sequences from this sample belong to one group (represented in purple in figure 3.26).

The diversity of SIRV-like sequences decreases in the eighth sample. This sample was collected on September 11, 2003 and yielded 41 SIRV-like coat protein clones (table 3.10). All of the sequences detected from this sample belong in previously established SIRV-like viral groups (clade A in figure 3.25). The inferred protein-based phylogenetic tree also shows the September 2003 sequences in the largest group of SIRV-like viruses (represented in purple in figure 3.26).

The ninth sample shows diversification of SIRV-like sequences compared to the previous samples. The next sample was collected in October 2003, and 62 SIRV-like coat protein clones were sequenced (table 3.10). DNA sequence of the SIRV-like coat protein reveals two groups of SIRV-like viruses in this sample (clades A and C in figure 3.25). Four coat protein sequences are in a group of SIRV-like coat protein sequences that was detected in October 2002 and January 2003 (clade C), and the other sequences from this sample are in the predominant group of SIRV-like viruses (clade A) (figure 3.25). Inferred protein sequence reveals that all of the SIRV-like sequences from this sample belong in the largest group of SIRV-like viruses (represented in purple in figure 3.26). There were three inferred protein sequences in this sample that were identical to protein sequences that had previously been detected in other samples (figure 3.26).

The diversity of SIRV-like sequences decreases in the next sample. The tenth sample was collected in November 2003, and 42 SIRV-like coat protein clones were sequenced (table 3.10). Even though 42 clones were sequenced, there were only three unique sequences detected from this time point. When comparing this sample to previous samples, the diversity of the SIRV-like viral population has been decreased to a single group (clade C) (figure 3.25). Even though there is only one viral group detected in this sample, a change in the SIRV-like population occurred. Until this sample we had been consistently detected SIRV-like sequences belonging to clade A (figure 3.22). However, we do not detect any SIRV-like sequences belonging to clade A in this sample (figure 3.22). There are only two inferred protein sequences from this sample (table 3.10). Both of the protein sequences belong in the largest group of protein sequences (represented in purple in figure 3.26).

Significant change in the SIRV-like population is apparent in the next sample. The next sample was collected in February 2004 and 67 SIRV-like coat protein clones were sequenced (table 3.10). When compared to SIRV-like sequences from previous samples, the SIRV-like population has greatly diversified since the last sample was collected in November 2003 (figure 3.22). As opposed to the November 2003 sample in which only one group of SIRV-like viruses was detected (clade C), the February 2004 sample reveals viruses from at least four major groups and numerous minor groups (clades A, C, D, G, H, and I) (figure 3.25). This sample still contained SIRV-like sequences from groups detected in previous samples (clades A, C, and D) (figure 3.25). However, there were three new groups of sequences (clades G, H, and I) (figure 3.25).

When translated, the DNA sequences collected in February 2004 result in 26 unique inferred protein sequences (table 3.10). Most of the protein sequences are detected for the first time. However, all the protein sequences belong to the largest group of protein sequences detected thus far (figure 3.26).

There is still diversity in the SIRV-like population, however, not to the extent that was observed in the February 2004 sample. The twelfth sample was collected in April 2004 and 43 SIRV-like coat protein clones were sequenced (table 3.10). The SIRV-like viruses are represented by sequences belonging to clades C and G with one outlying sequence (figure 3.25). The inferred SIRV-like protein sequences show much less diversity in the SIRV-like population. There were 18 unique inferred protein sequences obtained from this sample (table 3.10). There were no protein sequences from this sample that were identical to protein sequences that had been detected in previous samples. The sequences from this sample belong to the same large group as most of the other SIRV-like sequences (represented in purple in figure 3.26).

Little change in the SIRV-like sequences is detected in the last sample collected from this monitor site. This sample was collected in June 2004 (table 3.10). Forty-seven clones were sequenced from this sample and compared to sequences obtained from previous samples. Most of the SIRV-like sequences from this sample fall into two separate groups that are related to sequences from February 2004 (clades H and I in figure 3.25). Some of the other sequences from this sample are related to samples from October 2002 and August 2003 (clade D in figure 3.25). There were 16 unique inferred protein sequences detected in this sample (table 3.10), and all of them were related to the

largest group of protein sequences (represented in purple in figure 3.26). One sequence (detected seven times) was identical to sequences that had previously been detected in this monitor site (figure 3.26; the * represents the group of identical sequences).

In summary, the SIRV-like population was very dynamic over time. It is interesting to note that changes in the DNA sequences did not always result in changes in the inferred protein sequences (figures 3.22 and 3.23). It appears as though the SIRV-like population in Crater Hills was undergoing a population re-structuring in the last three samples collected from the monitor site (figure 3.22). It would be interesting to determine whether this is a temporary shift or a permanent change. Overall, there were nine groups of SIRV-like viruses (figure 3.25). The complete phylogenetic analysis of SIRV-like coat protein sequences included analysis of sequences after every sample. For a step-by-step view of changes in the SIRV-like population over time, see Appendix A.

Detailed analysis of SIRV-like population dynamics. After almost two years of regularly sampling and sequencing SIRV-like coat proteins from this site, we detected many different groups of SIRV-like viruses (figure 3.25). One large group of SIRV-like DNA sequences was most commonly detected (clade A) (represented by blue in figure 3.25), eight smaller groups were detected (clades B-I), and finally, many groups containing four or fewer sequences were detected throughout the two years (figure 3.25). There were three major groups of SIRV-like viruses based on inferred amino acid sequence (figure 3.26). One protein sequence (represented by * in figure 3.26) was continuously detected from this monitor site.

From August 2002 to June 2004, there was a total of 812 SIRV-like coat protein clones sequenced from this monitor site (table 3.10). Using Modeltest version 3.6 (Posada *et. al.* 1998), the nucleotide substitution model for maximum likelihood estimations was determined to be K81uf (Kimura 1981) with heterogeneity between sites (gamma=0.8990), base frequencies A=0.3276, C=0.1814, G=0.1833, and T=0.3077, and a substitution model A-C=G-T=1.0000, A-G=C-T=2.8331, and A-T=C-G=1.2451.

For approximately two years (August 2002—June 2004), this monitor site was sampled and the SIRV-like viruses were detected in every sample (table 3.10). The largest group of identical DNA sequences was composed of 250 SIRV-like coat proteins, which translates to 31.4% of the total SIRV-like population that was sampled during the two years of this research (table 3.11). This sequence was detected during the first year and a half, but was no longer detected after October 2003 (table 3.11).

Another group of identical sequences detected composed of 75 identical sequences, which makes up 9.4% of the SIRV-like population in our samples (table 3.11). The third largest group of sequences was composed of 62 identical sequences, which correlates to 7.8% of the SIRV-like population (table 3.11). This group was detected in earlier samples, and then in late samples, but was not detected in the middle of the sample collection.

Smaller groups of identical sequences were also detected, ranging in size from two to 20 identical sequences. A large portion of the SIRV-like population was composed of single sequences (figure 3.27). A total of 236 unique SIRV-like coat

protein DNA sequences were detected over the course of this research, which translates to 29% of the SIRV-like population.

Table 3.11. Crater Hills SIRV-like sequences illustrating the groups of identical sequences, how many times a group this size was detected, the percent of the SIRV-like population, and what samples the group contained.

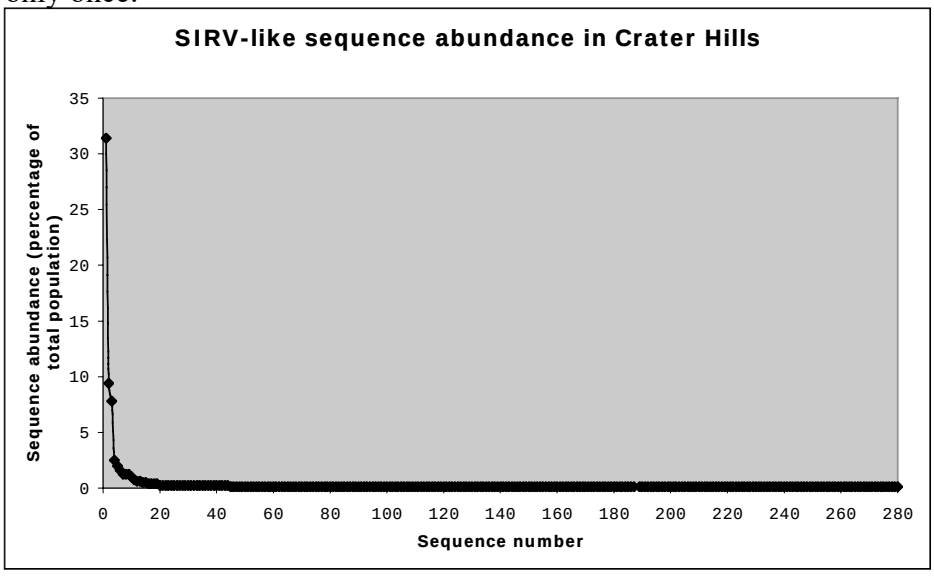
CRATER HILLS SIRV-like SEQUENCE ABUNDANCE			
Identical Sequences	Times Detected	Percent of SIRV-like Population	Samples[a]
250	1	31.4%	CH4;CH_MS;CHMS2—CHMS6
75	1	9.4	CH4;CH_MS;CHMS2—CHMS6
62	1	7.8	CHMS;CHMS6—CHMS8
20	1	2.5	CHMS8;CHMS9
16	1	2	CHMS;CHMS7—CHMS9
13	1	1.6	CHMS;CHMS8;CHMS10
10	3 unique[b]	1.25% each[c]	
8	1	1	
6	1	0.75	
5	2 unique	0.63% each	
4	2 unique	0.5% each	
3	4 unique	0.38% each	
2	25 unique	0.25% each	
1	236 unique	0.125% each	

[a]sequences from these samples were present in the indicated groups
[b]each of the smaller groups were detected this many times, however, they are unique
[c]each of the unique groups makes up this percentage of the population

There were 211 unique protein sequences detected in this monitor site (table 3.10). There were nine viral groups based on DNA sequences, but only three distinct protein groups detected in this monitor site (figures 3.22 and 3.23). Upon closer inspection, most of the viral clades (clades A, C, D, G, H, and I) translated to protein sequences belonging in the largest protein group (purple in figures 3.23 and 3.26).

Clades B and E were (detected in October 2002 and January 2003) translated to protein sequences belonging to a minor group (blue in figures 3.23 and 3.26). Finally, clade F was only detected in January 2003 and the sequences from this clade comprised a unique protein group in this monitor site (yellow in figures 3.23 and 3.26). Even though the DNA sequences of the SIRV-like coat protein were constantly changing, the inferred protein sequences appear to be more stable. A large group of identical sequences was detected in most of the samples (figure 3.26; the * represents the large group of identical protein sequences).

Figure 3.27. A sequence abundance plot for SIRV-like sequences from Crater Hills. The left end of the x-axis represents most commonly detected sequences, and the right end of the x-axis represents the most rare sequences. Most of the SIRV-like sequences in Crater Hills are detected only once.



Using the program MEGA version 3 (Kumar 2004), the changes in the SIRV-like DNA sequences in each sample collected from Crater Hills were calculated (table 3.12). When comparing all of the sequences, 88% of the bases in the SIRV-like coat protein

sequences remained unchanged (table 3.13). Transversions occurred somewhat more (25) compared to transitions (22). Eight changes occurred in the first position of the codon; there were four transitions and four transversions (table 3.13). Only seven changes occurred in the second position with two transitions and five transversions (table 3.13). A total of 32 changes occurred in the third position; there were 16 transitions and 16 transversions (table 3.13). The third position changes occurred more frequently, explaining the fewer number of groups based on protein sequences (figure 3.26).

Table 3.12. Numbers of transitions and transversions in SIRV-like sequences in each sample collected from the Crater Hills monitor site.

SAMPLE DATE	IDENTICAL BASES	TRANSITIONS			TRANSVERSIONS		
		1st	2nd	3rd	1st	2nd	3rd
08/06/2002	403/405	1	1	0	0	0	0
10/26/2002	350/405	5	2	18	4	6	19
01/03/2003	343/405	6	2	21	6	7	21
02/27/2003	402/405	1	1	0	0	1	0
05/30/2003	402/405	1	1	0	0	1	0
07/18/2003	404/405	0	0	0	0	1	0
08/04/2003	391/405	2	1	5	1	1	5
09/11/2003	402/405	1	1	0	0	1	0
10/23/2003	386/405	2	1	6	2	3	6
11/24/2003	403/405	0	0	1	0	1	0
02/19/2004	372/405	2	2	12	2	4	11
04/15/2004	394/405	1	1	3	1	2	3
06/14/2004	395/405	1	1	4	1	2	2

Table 3.13. The numbers of transitions and transversions for all SIRV -like sequences collected from Crater Hills.

Sample Range	Position	Identical Base Pairs	Ts Pairs	Tv Pairs
Aug 2002-June 2004	*Overall*	*358/405*	*22*	*25*
	1st	127/135	4	4
	2nd	128/135	2	5
	3rd	103/135	16	16

Ts=transitions; Tv=transversions

Comparing SSV-like and SIRV-like population dynamics

Unlike the SSV-like population, we were able to detect SIRV-like sequences in every sample collected over the two years of this research. We sequenced 292 SSV-like 3F/4R clones (table 3.5) and 812 SIRV-like clones (table 3.10) from this monitor site over time. It appears that SIRV-like viruses are permanent residents in this hot spring. Or, perhaps the SSV-like population occasionally dropped below our detection limits, and therefore, appeared to be absent from the environment. The detection limit of the SSV 3F/4R primers was 100X lower than the SIRV DBPF/R primers.

The SSV-like population was characterized by three distinct viral groups over time (figure 3.17). We detected changes on the DNA level, but the inferred protein sequences revealed modest change (figure 3.18). The SIRV-like population was much more diverse than the SSV-like population (figures 3.25 and 3.26). The SIRV-like population was characterized by approximately nine viral groups over time with numerous outlying sequences (figure 3.25).

There was one consistency between the two viral populations. Each virus population appeared to fluctuate over time. Occasionally both virus populations were relatively stable from sample-to-sample. The SSV-like population was mostly composed of sequences detected only one time (table 3.7). However, the SIRV-like population was dominated by a large group of viral sequences detected throughout this research (table 3.11).

The virus populations do not appear to be changing based on a changing environmental parameter. For example, on September 11, 2003 the pH of the monitor

site decreased from 2.6 (in August 2003) to 1.6 and the temperature decreased from 80°C (in August 2003) to 67°C. Even this seemingly drastic change did not alter the viral populations detected in this monitor site.

Conclusions

This monitor site proved to be a very dynamic thermal area with many unexpected results. This was the first detailed study that we know of investigating a thermal feature within the Crater Hills thermal area. This site was sampled continuously for approximately two years (August 2002—June 2004). During that time, the water chemistry, the dominant archaeal population, and changes in the population structures of two viruses were continually monitored.

The water chemistry was characterized by high sulfate and high aluminum concentrations (figure 3.7). The sulfate levels in this monitor site were at higher concentrations than the other two monitor sites (figure 3.7). The monitor site is a vapor-dominated system with no visible inflow or outflow channels. This site is turbulent due to out-gassing and therefore likely to be well mixed. From July 2003 through August 2003, there was an increase in the aluminum concentration followed by a decrease over the next four samples (figure 3.6c). This cycle appears to be repeating itself in the April 2004 and June 2004 samples (figure 3.6c); perhaps aluminum is present in high concentrations seasonally. Surprisingly, the water chemistry did not significantly change over time, and most of the observed changes could be contributed to seasonal fluctuations. The pH and temperature of this monitor site were relatively consistent. The

pH was measured between 1.6 and 2.8, and the temperature was measured at 67-86°C. With the exclusion of the conditions on September 11, 2003 (pH 1.6 and temperature 67°C), the site was very stable over the course of our sampling. By analyzing the water chemistry at this hot spring, we learned it is a unique system dominated by high sulfate concentrations.

This monitor site contained a fairly diverse archaeal population that changed over time. The full-length (~1,400bps) and almost full-length (~1,000bps) 16S rDNA sequences showed low identity to published 16S rDNA sequences (table 3.3). The 16S rDNA clones showed low identities to published 16S rDNA sequences with the exclusion of some high identity matches to *Vulcanisaeta* (table 3.3). This result gives an indication of the large amount of sequences that are not present in the existing public databases due to the limited work done in high temperature acidic environments.

The 16S rDNA sequences most commonly detected in environmental libraries constructed from this hot spring over time were *Caldococcus*-like and *Vulcanisaeta*-like sequences. *Caldococcus* was originally described in 1996 from a Noboribetsu hot spring (Aoshima *et. al*. 1996). The growth of *Caldococcus* was observed between pH 1.5 to 4.0 with optimal pH being pH 3.0, and the growth temperatures ranged from 70-96°C with an optimal temperature at 92°C. Based on these growth characteristics, it is very reasonable to expect *Caldococcus* or *Caldococcus*-like organisms in this monitor site.

Vulcanisaeta was originally isolated from hot springs in Japan and grows optimally at 85-90°C and pH 4.0-4.5 (Itoh *et. al*. 2002). These conditions do not seem to match the conditions of Crater Hills when it was sampled in June 2004 (pH 1.9 and

81°C); however, perhaps this organism has a larger growth range in nature. Current research is aiming towards culturing *Caldococcus* and *Vulcanisaeta* from the Crater Hills monitor site.

We detected an apparent primer bias in the beginning of this research (figure 3.11). By using both 16S rDNA primer sets on the January 2003 environmental DNA, different dominant host populations were indicated based on the primer set used in PCR assay (figure 3.11 and table 3.4). However, when both primer sets were used on the July 2003 environmental DNA, a consensus appeared concerning the dominant host population (figure 3.11 and table 3.4). The 20F/UA1406R primer set did not detect *Sulfolobus* sequences (table 3.4). The reason for this remains unexplained; each primer set matches published *Sulfolobus* 16S rDNA sequences 100% (table 3.14). It is curious that we were able to detect 16S rDNA sequences similar to published organisms with primers that did not match the published 16S rDNA sequence (table 3.14). For example, *S. azoricus* was detected in the January 2003 sample using the 20F primer. However, the 20F sequence does not align to the published *Stygiolobus* sequence (table 3.14). However, most of the 16S rDNA clones from this monitor site did not match published sequences very highly (table 3.3). The 16S rDNA sequences from Crater Hills are related to the published sequences, but not the same species.

The host population in Crater Hills was not as we had expected. By sequencing 16S rDNA genes from the environment, we learned this hot spring was dominated by different hosts throughout this study. *Sulfolobus*-like sequences were detected in the

beginning; we saw a shift to *Caldococcus*-like organisms, and finally *Vulcanisaeta*-like sequences dominated in the last sample.

Table 3.14. The matches between the 16S rDNA primers and the published 16S rDNA sequences. The numbers indicate where on the gene the primer aligns.

	20F	**348Fa**	**UA1406R**
S. solfataricus	+++++ (2-21)	++++ (339-353)	+++++ (1354-1370)
A. infernus	++ (110-128)	+++ (323-337)	+++++ (1341-1355)
S. azoricus	No match	++++ (311-325)	+++++ (1329-1343)
C. noboribetus	+++++ (2-21)	++++ (338-352)	++++ (1353-1369)
V. distributa	++ (248-266)	++++ (319-333)	+++++ (1339-1355)

+++++: perfect match; ++++: one mismatch; +++: two mismatches; ++: three or more mismatches

The archaeal population appeared to drastically change within a month from *Caldococcus*-like sequences to *Sulfolobus*-like sequences (figure 3.11). But, two different primer sets were utilized in these samples. The January 2003 sample, which was predominantly *Caldococcus*-like sequences, was amplified with 20F/UA1406R, and the February 2003 sample, which was predominantly *Sulfolobus*-like sequences, was amplified with 348Fa/UA1406R. It must be mentioned that the environmental sequences within both samples are not matching *Caldococcus* or *Sulfolobus* at very high percentages. We are likely detecting a primer bias between the two samples and not representing what is occurring in the natural environment. Initially, the amplification of archaeal 16S rDNA genes was attempted with both 16S rDNA primer sets. However, in each instance one primer set was successful and the other was not.

Based on 16S rDNA gene sequence analysis from this monitor site, *Sulfolobus* is not a dominant member in this particular hot spring over time. In fact, our limited sampling did not always detect *Sulfolobus*. Surprisingly, only one environmental 16S rDNA clone from this monitor site was similar to *Sulfolobus* sequences that had been previously detected in YNP (Whitaker *et. al.* 2003). We had been assuming the dominant host organism was *Sulfolobus* due to the ease of culturing this organism directly from environmental samples. These results lead to serious questions concerning the virus populations. We are easily detecting *Sulfolobus* viruses, however, we are unable to detect large numbers of *Sulfolobus* based on 16S rDNA sequence. Even the sequences that are most similar to published *Sulfolobus* 16S rDNA sequences are found at very low percentages (table 3.3). There are many approaches to addressing this anomaly. First, the virus sequences may be easily detected because they are present in a higher copy number than the *Sulfolobus* 16S rDNA gene. It has previously been shown that SSV1 is present at high copy numbers in its host organism in culture (Yeats *et. al.* 1982). Or, perhaps SSVs and SIRVs have a broader host range than just *Sulfolobus*. This would explain why we retrieve SSV-like and SIRV-like sequences, but not *Sulfolobus* sequences. Perhaps the viruses that we are amplifying infect other archaeal species. For example, Rachel *et. al.* (2002) discovered viruses infecting *Acidianus* organisms from Crater Hills that were morphologically similar to SSVs and SIRVs. However, I believe that we are amplifying *Sulfolobus* viruses. Throughout this research, *Sulfolobus*-like sequences were the only 16S rDNA sequences detected in each of the monitor sites (see

also Chapters 4 and 5). Even thought the host populations in each monitor site differed, we were able to amplify viral sequences from each of the monitor sites.

SSVs and SIRVs continuously monitored in Crater Hills were constantly undergoing change. SSV-like sequences were detected in eight samples. The fact that primer set 3F/4R yielded positive results, but the 7F/8R primer set did not, is interesting. The most likely reason for this result is the different sensitivity levels of the 'universal' SSV primer sets. The 3F/4R primer set was 10,000X more sensitive on purified viral template than was the 7F/8R primer set. It is likely the SSV-like coat protein sequences would be amplified with more sensitive primers.

It is curious as to why SSV-like sequences were no longer detected after the October 2003 sample (table 3.5). Perhaps the SSV-like viral population dropped below the detection limits of our primers (<0.1pg). Another possibility is the SSV-like viral population changed with the dominant host population. *Sulfolobus*-like 16S rDNA sequences were detected in this hot spring through October 2003 (figure 3.11). Although these sequences were not detected in large numbers and did not show a high identity to published *Sulfolobus* 16S rDNA sequences, they may be representing organisms closely related to *Sulfolobus* that are subject to infection by the same viruses. The absence of *Sulfolobus*-like sequences correlates with the ability to detect SSV-like sequences.

We do not believe the absence of SSV-like 3F/4R sequences reflects an experimental issue. The fact that SIRV-like and 16S rDNA sequences were amplified from the same samples indicates the source DNA is suitable for PCR and the technique is

not flawed. The most likely explanation is the SSV-like viral population dropped below what we could detect with our primers (<0.1pg).

SIRV-like sequences were detected in every sample collected from this monitor site (table 3.10). The primers designed to amplify the entire SIRV-like coat protein (405bps; 135aa) reliably detected this gene. This primer set was found to be very sensitive on purified viral template (0.001pg). SIRVs have not been shown to integrate into their host cell's chromosome; instead they are maintained within the cell as a stable carrier state (Prangishvili *et. al.* 1999). Over the course of this study, there were not any interrupted ORFs detected among the SIRV-like coat protein sequences obtained from this hot spring. This supports our claim that the interrupted SSV-like ORFs may be from integrated copies of the virus, instead of sequencing or PCR errors.

The SIRV-like population within Crater Hills is dynamic. The changes in the SIRV-like population are detected immediately and continue through numerous samples (figure 3.22). The fourth sampling point reveals the diversity of SIRVs has decreased from numerous groups to only one (figure 3.22). This dominant group of viruses is detected within the next samples (figure 3.22). The seventh sample the SIRV-like population diversifies again, revealing numerous groups of SIRV-like viruses (figure 3.22). However, this increased diversity is only detected in one sample because in the next sample (September 2003) the viral diversity drops again to only one group of viruses (figure 3.22). The remaining samples reveal the SIRV-like population is constantly changing and fairly diverse (figure 3.22).

There were numerous groups of SIRV-like viruses detected over time within this monitor site (figure 3.25). Unlike the SSV-like viruses, which had three distinct groups of viruses, the SIRV-like viruses are not as clearly distinguished. It appears that perhaps the SIRV-like population is dominated by a large group of viruses with many smaller groups of viruses, and finally a large number of singlets in the population. Perhaps the singlets are present and ever-changing in case an environmental parameter changes, the virus can take advantage and become dominant in the population.

Both SSVs and SIRVs were shown to change over time. Our data revealed that the population structures of SSVs and SIRVs are different in the same thermal area even though they are both archaeal viruses living at low pH and high temperature. The SIRVs were more diverse than the SSVs in this particular system. It will be interesting to see how this is maintained over a longer period of time, and how this compares to the other monitor sites.

In conclusion, we discovered both viral populations within this hot spring are dynamic and diverse. The SSV-like population was only detected through October 2003; however, the population changed and showed diversity within that time. The SIRV-like population was detected in each sample, and was shown to be very diverse. Even though there may be questions concerning the hosts of these viruses in nature, we were able to show the viral populations do change over time.

CHAPTER 4

RABBIT CREEK MONITOR SITE: VIRUS DYNAMICS, HOST POPULATIONS, AND WATER CHEMISTRY

Introduction

A thermal acidic feature within the Rabbit Creek area was selected as a monitor site. The reasons for selecting this site included the following: the thermal area is at least 70 years old, the site is characterized by high temperature and low pH, the area is easy to access year round, and we have previously cultured *Sulfolobus* and its viruses from this hot spring. The goal of this research is to determine if there are significant changes in viral population structures over time. The hypothesis we are testing is whether there is a relationship between virus population dynamics and observed changes in the host populations and water chemistry of an individual thermal feature. From July 2002 through June 2004, we monitored an acidic hot spring within the Rabbit Creek thermal area in Yellowstone National Park (YNP) for changes in virus populations, host populations, and water chemistry.

The Rabbit Creek area has been known as a geothermal site for more than 70 years (Allen and Day 1935). Because of the age of this thermal site, examination of this site allows an opportunity to monitor viral population dynamics in an established thermal feature as compared to younger thermal sites (see Chapters 3 and 5).

The population dynamics of SSV-like and SIRV-like viruses were monitored over time in the Rabbit Creek monitor site. These viruses were chosen because in an initial

survey of acidic thermal areas in YNP, both SSV-like and SIRV-like viruses were commonly detected in enrichment cultures, and sequence data for both viruses is available. Four SSV genomes from isolated thermal areas have been sequenced (Weidenheft *et. al.* 2004), and two SIRVs from Iceland have been sequenced (Peng *et. al.* 2001).

Materials and Methods

Site location

The Rabbit Creek monitor site is located within the Midway Geyser Basin geothermal area (44° 31.287′N, 110° 48.647′W).

pH and temperature monitoring

The temperature of the Rabbit Creek monitor site was continuously monitored via a data logger. The data logger recorded the measured temperature every 20 minutes. The pH of the monitor site was measured each time a sample was collected.

Sample collection and water chemistry

The hot spring water samples were collected and water chemistry analyses were performed as previously described (see Chapter 3).

Total nucleic acid extraction from environmental samples

Total nucleic acid from the Rabbit Creek monitor site was extracted using the UltraClean™ Water DNA Extraction Kit (Mo-Bio, Solana Beach, CA) following

manufacturer's instructions (for a brief description, see Chapter 3). The integrity and yield of the extracted DNA was determined by display on agarose gels.

Primer design and PCR for viruses

SSV-like sequences were PCR amplified from the extracted environmental nucleic acids. Two "universal "primer sets were chosen to amplify environmental SSV-like sequences (figures 3.1 and 3.2). The UnvSSV#3F and UnvSSV#4R primer set amplifies ~250-260 base pairs (bps) in the largest ORF of each sequenced isolate (figure 3.1 and table 3.1). The UnvSSV#7F and UnvSSV#8R primer set amplifies the coat proteins of the SSV viruses (figure 3.2). Depending on the presence of VP2, the amplicon ranges in size from 682 bps to 1,069 bps (table 3.1). A standard PCR protocol was employed containing 1pg-10ng total environmental nucleic acid (described in Chapter 3). The products were analyzed by agarose gel electrophoresis.

SIRV-like sequences were also amplified from the total nucleic acid extracted from the environment. Using the primer set DBPF/R (figure 3.4), the PCR reaction and cycling conditions are previously described (see Chapter 3). The DBPF/R primer set amplifies the entire coat protein gene of SIRVs (405bps). The products were analyzed by agarose gel electrophoresis.

Archaeal 16S rDNA primers and PCR

16S rDNA gene sequence analysis was used to determine the identity of the resident archaeal populations in the Rabbit Creek monitor site. Two primer sets (20F/UA1406R and 348Fa/UA1406R) (Baker *et. al.* 2003) were chosen to amplify the

16S rDNA gene from the extracted total nucleic acid. The PCR parameters were identical to previously described protocols (see Chapter 3). The primer set 20F/UA1406R yields a product of 1,400 bps with 348Fa/UA1406R yielding a 1,000 bps product. The amplicons were analyzed by agarose gel electrophoresis.

Cloning PCR products

Viral and 16S rDNA PCR products were cloned via TOPO-TA cloning into the pCR2.1 vector using the standard protocol (Invitrogen, Carlsbad, CA). The transformation into *E. coli* and plasmid purification was previously described (see Chapter 3).

DNA sequencing

The screening of the plasmids for insertion of PCR products was performed as previously described (see Chapter 3). DNA sequencing conditions were as described previously (see Chapter 3).

Sequence and phylogenetic analysis

Viral sequences. SSV-like and SIRV-like sequences were edited, aligned, and a maximum parsimony analysis was conducted as previously described (see Chapter 3). A maximum likelihood analysis was conducted using MrBayes version 3.0 (Huelsenbeck and Ronquist 2001) as described in Chapter 3.

Viral DNA sequences were translated, aligned, and a maximum parsimony analysis was conducted as described previously (see Chapter 3). The SSV-like 3F/4R

protein sequences were 85 to 87 amino acids from the largest ORF in the SSVs (812 amino acids predicted in the SSV-RH ORF). When the SSV-like DNA sequences were translated to protein, some sequences resulted in a truncated protein sequence. The translated DNA sequences that resulted in an interrupted ORF (those with stop codons throughout) are termed non-functional ORFs, and the DNA sequences that resulted in un-interrupted ORFs (no stop codons) are termed functional ORFs. The distinction between functional and non-functional ORFs was used to determine which viral sequences might have come from a previously integrated copy of the virus. These interrupted ORFs were not used in the protein analysis, unless otherwise noted. The SIRV coat proteins were also translated into amino acid sequence (135 amino acids). The molecular evolutionary analyses (rates of transitions and transversions) were conducted using MEGA version 3 (Kumar *et. al*. 2004).

Phylogenetic trees were constructed sequentially in order to compare the new viral sequences to the viral sequences detected at previous sampling times. In the phylogenetic trees, the sequences from the sample being discussed will be presented in red, sequences from previous samples will be presented in black, and the sequences from the completely sequenced isolates (SSV1, SSV2, SSV-RH, and SSV-K1 for 3F/4R and 7F/8R primer sets; SIRV1 and SIRV2 for the DBPF/R primer set) will be represented by blue. The black sequences are only presented as a reference to make it visually easier to determine how or if the viral population has changed since the previous sampling times. A viral group/clade will be defined as a monophyletic cluster of environmental viral

sequences. This group is supported by high bootstrap values. The viral sequences within a specific group/clade contain 5-10% sequence difference.

In the DNA analysis, numbers next to principle branches present bootstrap support >50% using maximum parsimony and Bayesian clade credibilities using maximum likelihood. In the protein analysis, the number at the principle branches show bootstrap support >50% using maximum parsimony. The frequency of a particular sequence is noted in parentheses following the name of the sequence. I will define a viral group as a monophyletic cluster of environmental virus sequences.

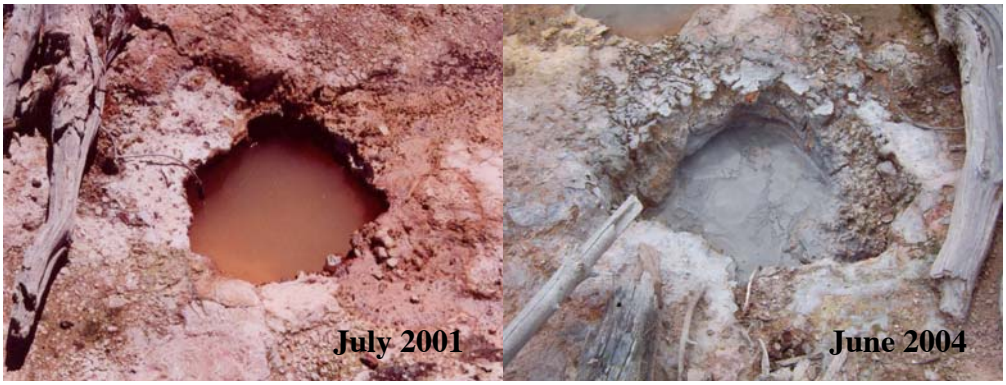
Archaeal host sequences. 16S rDNA sequences were edited and compared to published 16S rDNA sequences (http://www.ncbi.nlm.nih.gov/BLAST/). The environmental 16S rDNA sequences obtained from this site were analyzed as previously described (see Chapter 3).

Results

Temperature, pH and water chemistry of the Rabbit Creek monitor site

The Rabbit Creek monitor site was sampled twelve times from July 2002 through June 2004. Although this site is old, it has changed significantly in three years during this research (figure 4.1). The monitor site ranged in temperature from 64 to 90°C and pH values ranged from 3.2 to 5.1 (figure 4.2).

Figure 4.1. Pictures of the Rabbit Creek monitor site taken three years apart. The left picture was taken in July 2001, and the right picture was taken in June 2004.

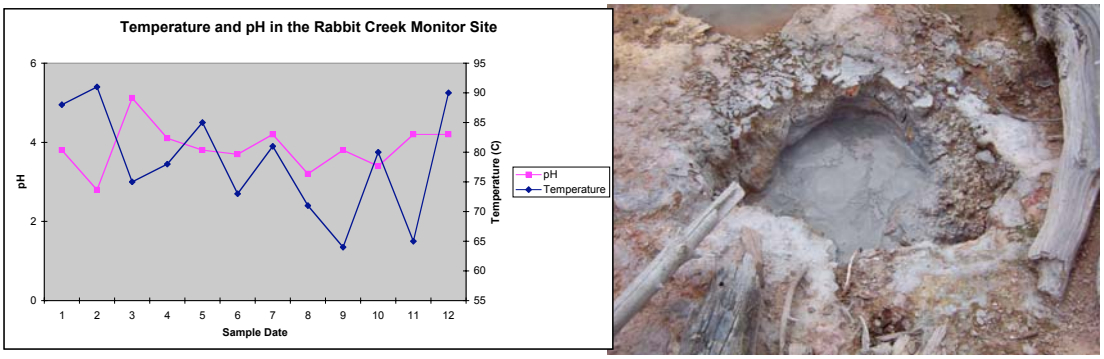


The water chemistry indicated a well-weathered geothermal site. The chemistry analysis started January 20, 2003 and continued through June 3, 2004 for a total of nine samples. Of the metals tested, several were not detected (or detected sporadically at very low levels) (table 4.1). Metals not detected include antimony, barium, boron, cadmium, copper, lead, magnesium, mercury, molybdenum, nickel, and selenium. The chloride levels remained low (0-6mg/L) during this research (figure 4.3a) with the average chloride concentration of 3mg/L. The sulfate levels in this monitor site were also at low levels (23-225mg/L) during this research (figure 4.3a) with the average sulfate concentration at 87mg/L. The dissolved organic carbon (DOC) concentration varied from 1.22-9.2mg/L with an average concentration of 5.3mg/L (table 4.1).

Essential nutrients were also monitored in the hot spring water. The nutrients measured were ammonia, nitrogen, and phosphorus (table 4.1). Ammonia levels remained constant with the exception of an increase in ammonia concentration in the July 2003 sample (figure 4.3b). The ammonia concentration ranged from 1.9-16.3mg/L with

an average concentration of 5.3mg/L. Nitrogen, expressed as nitrate and nitrite, was not detected (table 4.1). Phosphorus concentrations were higher from this monitor site in comparison to the other two sites (figure 4.3b) with the average phosphorus concentration at 0.7mg/L.

Figure 4.2. Temperature and pH over time at the Rabbit Creek monitor site (left) and a picture of the monitor site (right). The sampling dates correspond to the sampling dates in table 4.1.



The metals regularly detected were aluminum, arsenic, calcium, iron, manganese, potassium, silicon, sodium, and zinc (table 4.1). Most of these metals were lower concentrations in Rabbit Creek as compared to the other two monitor sites. The average aluminum concentration was 1.4mg/L with a range of 0.1-4.9mg/L (figure 4.3c). The concentrations of arsenic ranged from 0.02-0.06mg/L (figure 4.3c) with an average of 0.04mg/L. The concentration of calcium was low with an average calcium concentration of 1.5mg/L (figure 4.3c).

Iron concentrations were the highest measured in the three monitor sites. The iron concentrations ranged from 1.96-45.5mg/L with an overall average of 12.2mg/L (figure 4.3d). Potassium was constantly detected at lower levels than the other two monitor sites

(figure 4.3d), ranging in concentration from 3 to 14mg/L, with an overall average of 5.8mg/L.

Table 4.1. Water chemistry values measured at the Rabbit Creek monitor site from January 2003 through June 2004.

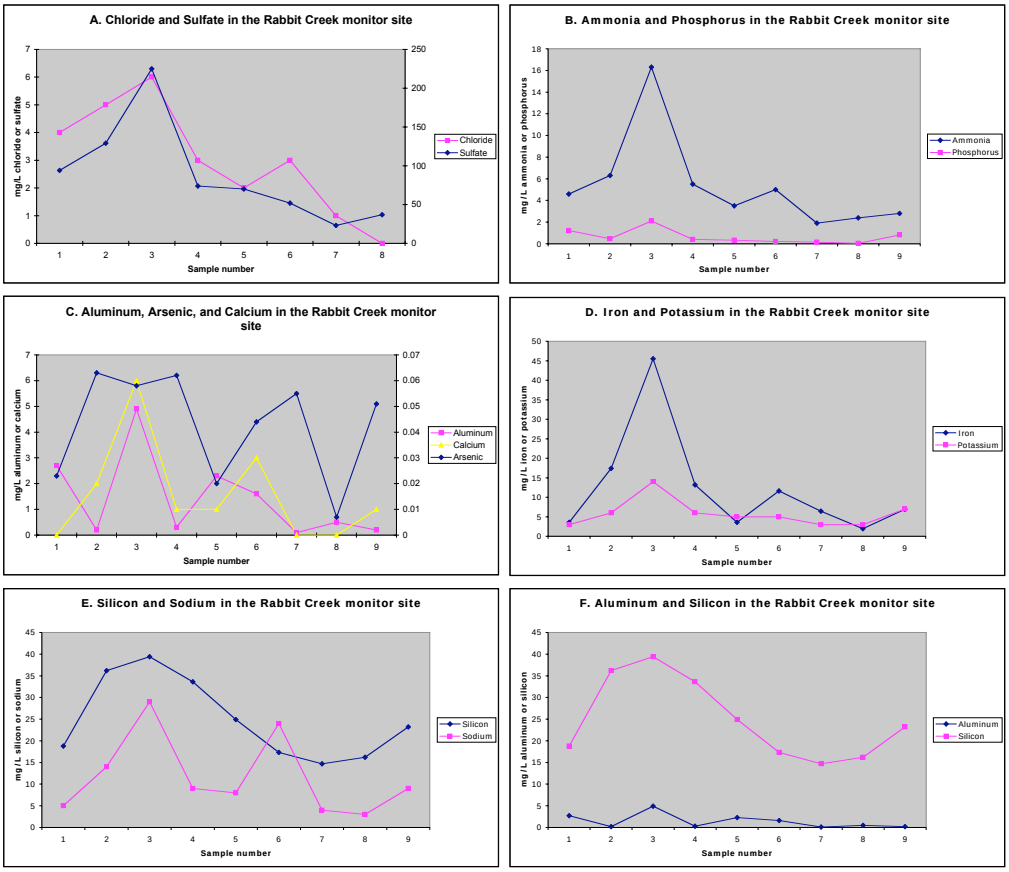
Sample Numbers	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***
Sample Dates	*01/03*	*06/03*	*07/03*	*08/03*	*10/03*	*11/03*	*01/04*	*03/04*	*06/04*
Chloride	4	5	6	3	2	3	1	0	3
Sulfate	94	129	225	74	70	52	23	37	45
DOC	3.2	5.7	9.2	5.3	5.1	5.2	5	1.22	8.1
Ammonia	4.6	6.3	16.3	5.5	3.5	5	1.9	2.4	2.8
Nitrogen	0	0	0	0	0	0	0	0	0
Phosphorus	1.23	0.46	2.1	0.41	0.034	0.21	0.16	0.03	0.81
Aluminum	2.7	0.2	4.9	0.3	2.3	1.6	0.1	0.5	0.2
Antimony	0	0	0	0	0	0	0.008	0	0.006
Arsenic	0.023	0.063	0.058	0.062	0.02	0.044	0.55	0.007	0.051
Barium	0	0	0.2	0	0	0	0	0	0
Boron	0	0	0	0	0	0.01	0	0	0
Cadmium	0	0	0.001	0	0	0	0	0	0
Calcium	0	2	6	1	1	3	0	0	1
Copper	0.07	0.01	0	0	0	0	0	0	0.02
Iron	3.55	17.4	45.5	13.2	3.56	11.6	6.43	1.96	6.93
Lead	0	0.01	0	0	0	0	0	0	0
Magnesium	0	0	2	0	0	0	0	0	0
Manganese	0.01	0.09	0.33	0.07	0.07	0.11	0.03	0.03	0.05
Mercury	0	0	0	0	0	0	0	0	0
Molybdenum	0	0	0	0	0	0	0	0	0.009
Nickel	0	0	0	0	0	0	0	0	0
Potassium	3	6	14	6	5	5	3	3	7
Selenium	0	0	0	0	0	0	0	0	0
Silicon	18.8	36.2	39.4	33.6	24.9	17.3	14.7	16.2	23.2
Sodium	5	14	29	9	8	24	4	3	9
Zinc	0.03	0.09	0.32	0.08	0.09	0.12	0.07	0.1	0.28

*Sample numbers refer to numbers on corresponding graphs.
*Chemical species measured in mg/L.

Silicon appeared to be cyclic (figure 4.3e), and it was the lowest of the three monitor sites. The silicon concentrations ranged from 14.7 to 39.4mg/L with an average

of 24.9mg/L. Sodium was also cyclically detected (figure 4.3e), ranging in concentration from 3-29mg/L. Manganese and zinc were both consistently detected at low levels (table 4.1).

Figure 4.3. Selected water chemistry values measured over time in Rabbit Creek. The sample numbers correspond to the sampling dates listed in table 4.1. All values are measured in mg/L.



In summary, the water chemistry supports the supposition that this is a geothermally older site than either Crater Hills (see Chapter 3) or Ragged Hills (see Chapter 5). The water chemistry suggests a de-mineralized environment brought on by high temperature leaching of underlying rocks. This monitor site is a vapor-dominated

system with no visible out-flowing channel. This site is dominated by high iron levels (figure 4.4) and high manganese levels (figure 4.5). This hot spring was also characterized by the highest observed temperature values of the three monitor sites throughout the sampling time (64-91°C), and the highest measured pH values of 2.8-5.1 (figure 4.2). The chemistry of this monitor site did not significantly change during the course of this research. It was determined the chemical composition was unique from the other monitor sites. For an overview of the average water chemistry values obtained, see figure 4.6.

Figure 4.4. Iron concentrations over time in the three monitor sites. The sample numbers correspond to the sampling dates in table 4.1.

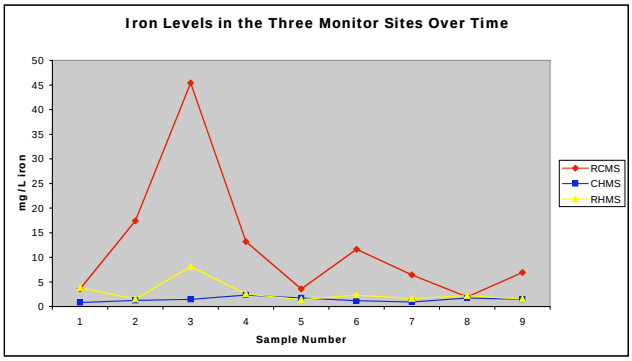


Figure 4.5. Manganese concentrations in the three monitor sites over time. The sample numbers correspond to the sampling dates in table 4.1.

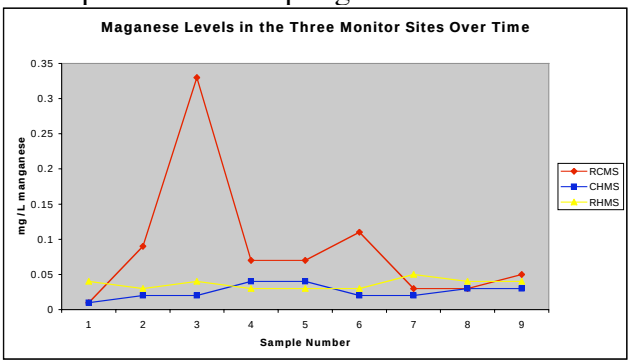
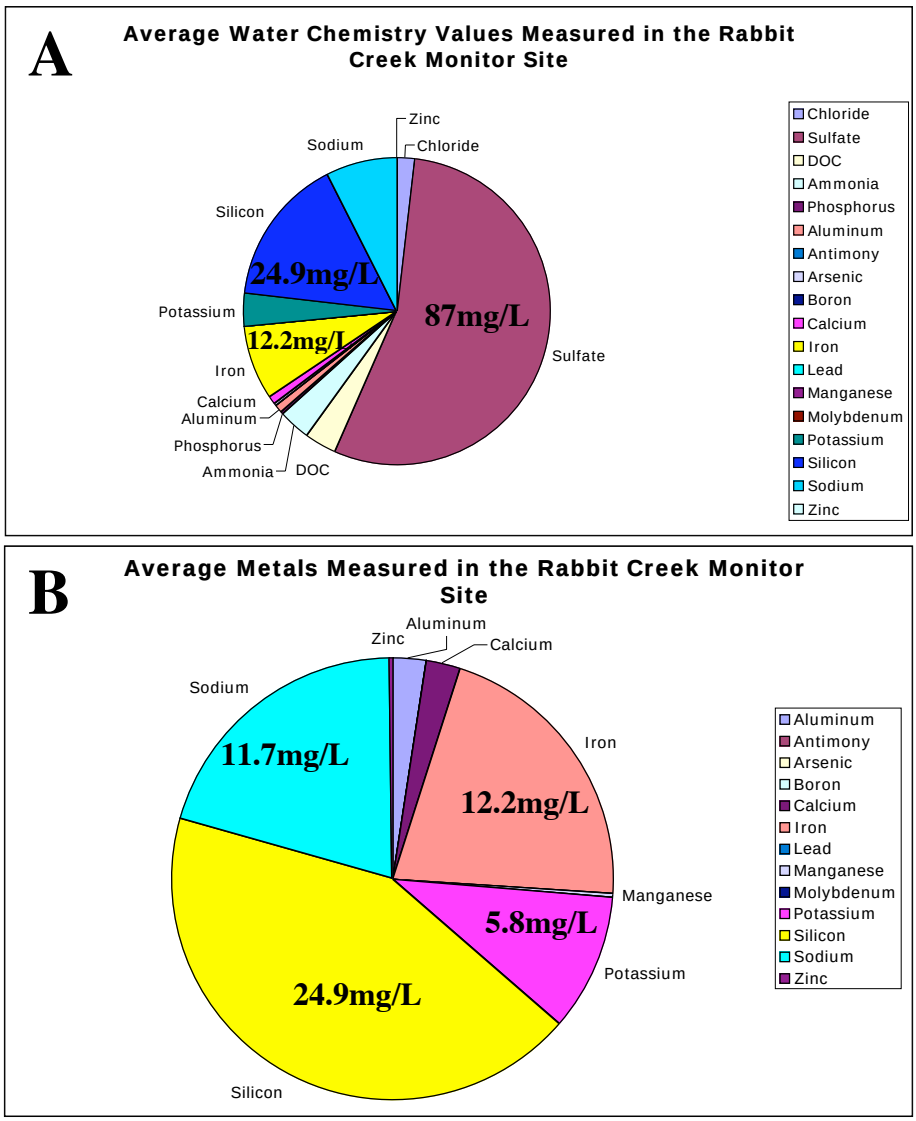


Figure 4.6. Average water chemistry values measured in the Rabbit Creek monitor site over time. A. Average values of all chemical species measured during this research. B. Average values of the metals measured in this research. All values are mg/L.

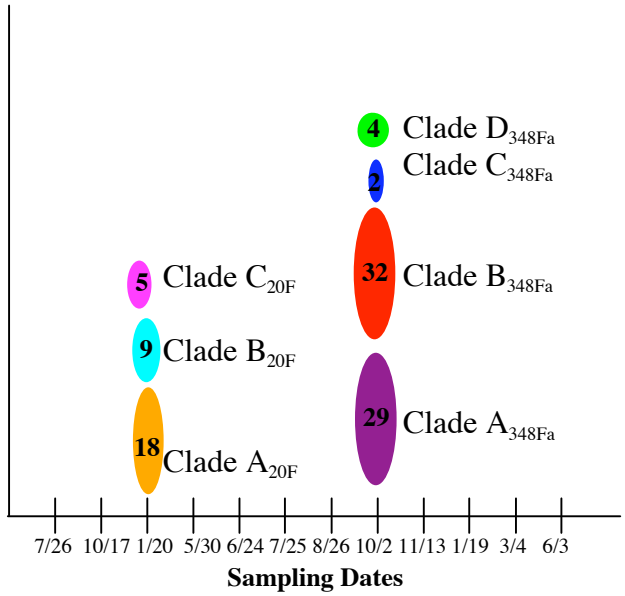


Archaeal population in the Rabbit Creek monitor site

To determine the host population, 16S rDNA sequence analysis was conducted on specific sampling dates. Only two of twelve samples were used to determine what the dominant archaeal population was over time. All of the samples collected from Rabbit

Creek were used in 16S rDNA PCR. However, only two samples resulted in 16S rDNA products. This result was unexpected and the reasons are unknown. The environmental DNA was suitable for PCR because viral sequences were detected in all of the Rabbit Creek samples. Numerous 16S rDNA primers were used on the samples. An overview of the host diversity is presented in figure 4.7.

Figure 4.7. The archaeal population in Rabbit Creek over time. The numbers indicate how many 16S rDNA clones most closely matched the published16S rDNA sequence. The absence of sequences in most samples reflects the inability to amplify 16S rDNA sequences from that particular time point and not a reflection of an absence of sequences in the environment.



Archaeal population determined by the 20F/UA1406R primer set. The first sample used for host identification was collected on January 20, 2003. Thirty-two full-length 16S rDNA clones were obtained from this sample (figure 4.8). When compared to 16S rDNA sequences available in the public database, all of the collected sequences

closely matched *Acidianus* species (96-99% identity), which divides into three clades (clades A, B, and C in figure 4.9 and table 4.2). The species of *Acidianus* detected were *A. infernus* (accession number X89852), *A. ambivalens* (accession number D85506), *A. tengchongensis* (accession number AF226987), and *A. brierleyi* (accession number D26489). *A. infernus* and *A. ambivalens* essentially have published 16S rDNA sequences with 99% identity; however, other published *Acidianus* species range from 90-91% identity to other *Acidianus* species. Fifteen of thirty-two clones most closely matched *A. brierleyi* with an identity of 97-99% (clade A in figure 4.9). Twelve of these clones matched the published *A. brierleyi* 16S rDNA sequence at 98-99% identity. Five environmental clones matched *A. infernus* with 96-99% identity (clades B and C in figure 4.9). Only three of these clones were 98-99% identical to the published *A. infernus* 16S rDNA gene. Nine clones most closely matched *A. ambivalens* with 98-99% identity (clades B and C in figure 4.9), and the remaining three clones matched *A. tengchongenesis* with 99% identity (clade A in figure 4.9).

Figure 4.8. Rabbit Creek host population based on nearly full-length or full-length 16S rDNA clones over time.

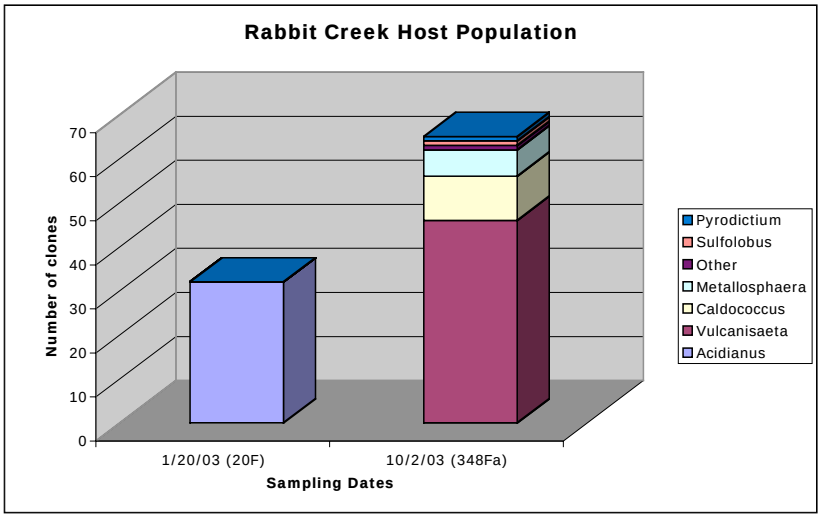
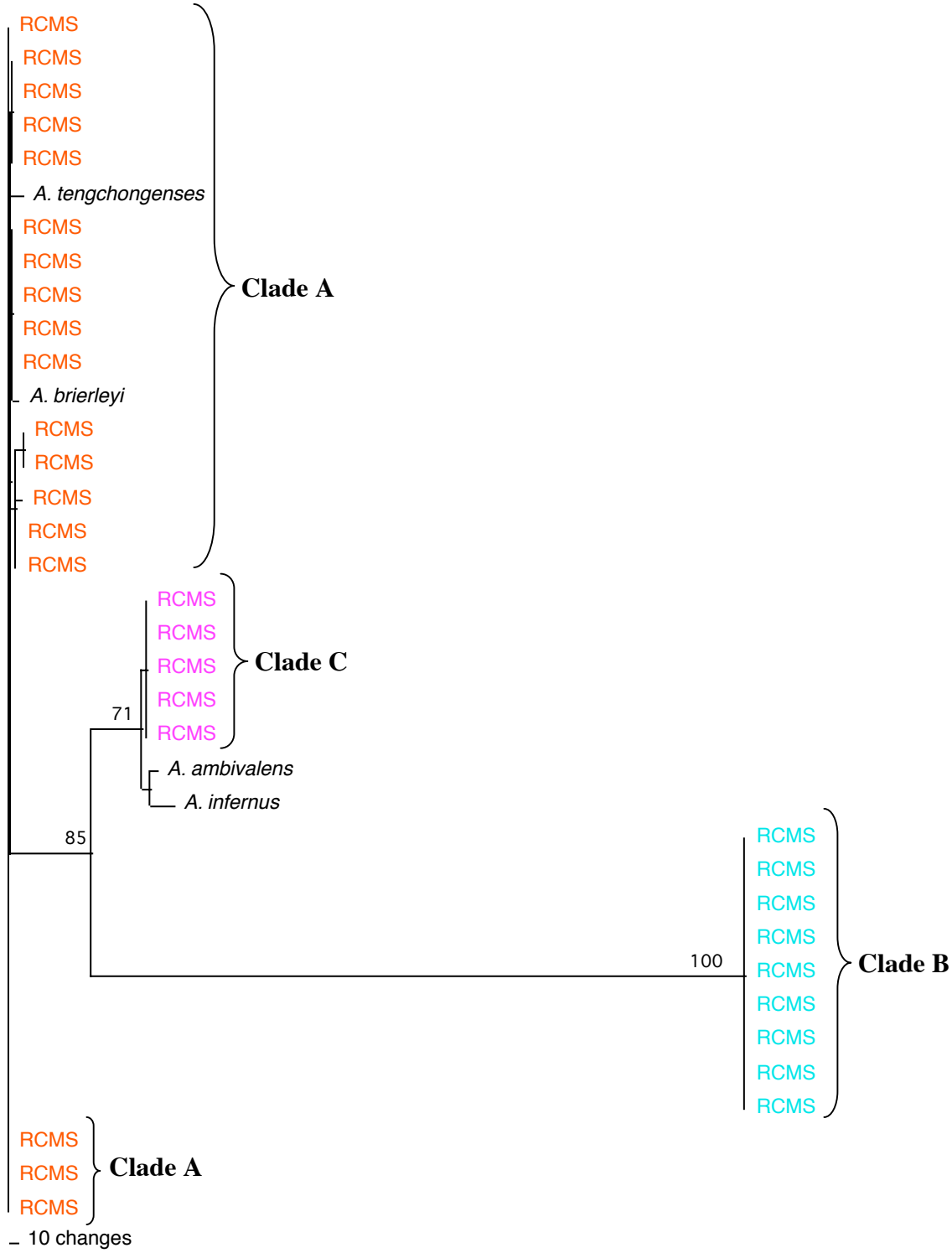
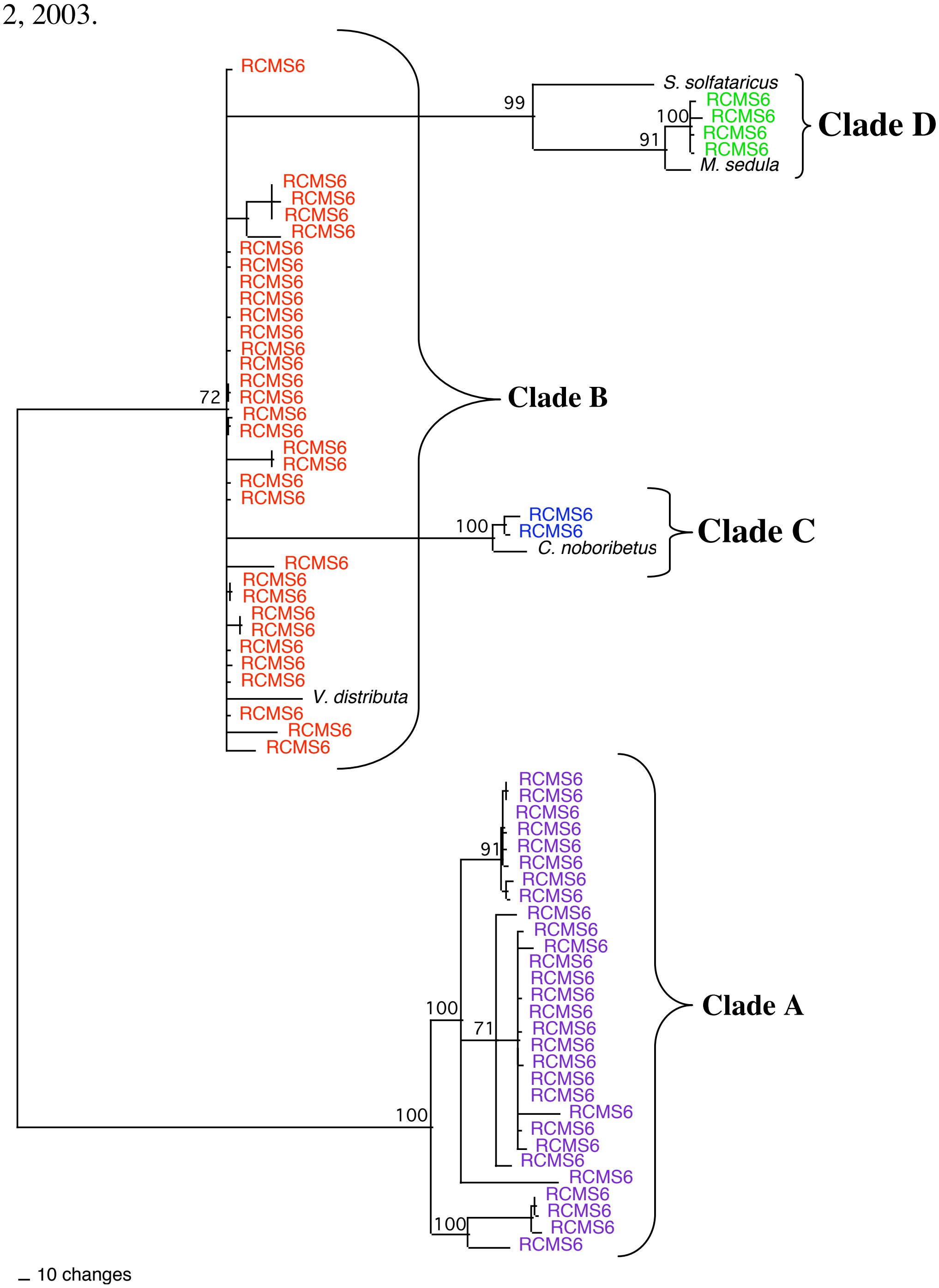


Figure 4.9. Environmental 16S rDNA sequences amplified with the 20F/UA1406R primer set from the Rabbit Creek monitor site on January 20, 2003.



Archaeal population determined by the 348Fa/UA1406R primer set. The other sample used for 16S rDNA sequence analysis was collected on October 2, 2003. There were 65 16S rDNA clones sequenced from this sample. There were numerous attempts to amplify 16S rDNA sequences with 20F/UA1406R without success. When the sequences from this sample were compared against the public database, the result was a more diverse population (figure 4.8). The majority of the population most closely matched *Vulcanisaeta distributa* (accession number AB063640). Forty-six clones (70.7% of the clone library) from this sample showed 94-98% identity to *V. distributa* (clades A and B in figure 4.10 and table 4.2). Thirty-three of these clones matched the published *V. distributa* 16S rDNA sequence at 98% identity (clade B in figure 4.10). Ten clones (15.3%) matched *Caldococcus noboribetus* (accession number D85038) with 90-97% identity (clades A and C in figure 4.10). Six clones (9.2%) showed 97% identity to *Metallosphaera prunae* (accession number X90482) (clades A and D in figure 4.10). The three remaining sequences each matched a different published 16S rDNA sequence at low identity (clade A in figure 4.10); one clone matched an unidentified Archaeon at only 86% identity, another clone showed 95% identity to *Sulfolobus islandicus*, and the remaining clone showed 91% identity to *Pyrodictium occultum*. The environmental sequences from this sample belong to four clades (figure 4.10). Clades B, C, and D are closely related to a published 16S rDNA sequence, and clade A is not closely related to any published 16S rDNA sequences. Perhaps Clade A represents an undiscovered archaeal population within this monitor site.

Figure 4.10. Environmental 16S rDNA sequences amplified with the 348Fa/UA1406R primer set from the Rabbit Creek monitor site on October 2, 2003.



There was a primer bias in the PCR amplification of the archaeal host population from the Rabbit Creek monitor site (figure 4.8 and table 4.2). The two samples contained drastically different host populations, but each was amplified with a different primer set. The 348Fa/UA1406R primer set resulted in the detection of four clades (figure 4.10) and the 20F/UA1406R primer set resulted in the detection of three clades (figure 4.9). Even though a similar number of clades were detected with each primer set, the 348Fa/UA1406R primer set resulted in a more diverse population (figures 4.9 and 4.10). It is unknown why only two samples were positive for 16S rDNA sequences from this monitor site. PCR amplification of the 16S rDNA sequence was attempted numerous times with both primer sets. SIRV-like DNA sequences were amplified from each sample collected from this monitor site suggesting that the environmental DNA was sufficient quality for PCR amplification. The 16S rDNA sequences detected within the monitor site likely represent crenarchaeal organisms that live in high temperature acidic conditions. To further understand and identify the archaeal population, more samples would need to be screened and both primer sets would need to be used.

Table 4.2. The percent identities of environmental 16S rDNA clones from the Rabbit Creek monitor site to published 16S rDNA sequences in the public database.

	RCMS (20F/UA1406R) January 20, 2003 Total clones=32	**RCMS6 (348Fa/UA1406R) October 2, 2003 Total clones=65**
Acidianus species	32 (96-99%)	ND
Vulcanisaeta distributa	ND	46 (94-98%)
Caldococcus noboribetus	ND	10 (90-97%)
Metallosphaera prunae	ND	6 (97%)
Sulfolobus islandicus	ND	1 (95%)
Pyrodictium occultum	ND	1 (91%)
Unidentified Archaeon	ND	1 (86%)

*ND=Not Detected

Population dynamics of viruses in the Rabbit Creek monitor site

SSV-like population dynamics in the Rabbit Creek monitor site. Total DNA was extracted from water collected from the Rabbit Creek monitor site as described above. Universal SSV primer sets 3F/4R and 7F/8R were used to measure the total diversity of SSVs in this monitor site (table 4.3). The universal SSV 3F/4R primer set resulted in the detection of viral sequences in seven out of twelve total sampling points. The coat protein primer set (7F/8R) was successful on only two samples collected from Rabbit Creek. An overview of SSV-like 3F/4R phylogenetic analysis is presented in figures 4.11 and 4.12.

Table 4.3. Number of SSV-like clones sequenced from the Rabbit Creek monitor site and how many unique DNA sequences in each library.

Sample Name	Collection Date	pH	Temp (°C)	Number of Clones 3F/4R	Unique Sequences	Number of Clones 7F/8R	Unique Sequences
RC7	July 26, 2002	3.8	88	34	15	50	28
RCM	Oct 17, 2002	2.8	91	27	15	0	0
RCMS	Jan 20, 2003	5.1	75	67	45	0	0
RCMS2	May 30, 2003	4.1	78	0	0	0	0
RCMS3	June 24, 2003	3.8	85	0	0	0	0
RCMS4	July 25, 2003	3.7	73	0	0	0	0
RCMS5	Aug 26, 2003	4.2	81	5	4	0	0
RCMS6	Oct 2, 2003	3.2	71	21	10	0	0
RCMS7	Nov 13, 2003	3.8	64	44	12	61	38
RCMS8	Jan 19, 2003	3.4	80	10	8	0	0
RCMS9	Mar 4, 2003	4.2	65	0	0	0	0
RCMS10	June 3, 2004	4.2	90	0	0	0	0
Total Number of Clones Sequenced				**208**	**94**	**112**	**65**

The first sample was collected on July 26, 2002, and 34 SSV-like 3F/4R clones were sequenced (table 4.3) and compared to the four sequenced SSV isolates (figure 4.13a). Some of the environmental SSV-like sequences are more related to SSV-RH and

SSV1 than they are to SSV-like sequences collected from the same hot spring (figure 4.13a). When the 3F/4R DNA sequences from this sample are translated, there are nine inferred functional protein sequences and one sequence that resulted in non-functional ORF (table 4.4). The diversity of inferred functional protein sequences is very similar to the pattern of the DNA sequences from this sample (figure 4.13b). The non-functional ORF from this sample is similar to the un-interrupted ORFs (figure 4.13a; the • represents the non-functional sequence). Upon closer investigation, the interrupted ORF has three stop codons due to an insertion causing a frameshift in the sequence.

Table 4.4. The comparison of SSV-like 3F/4R functional and non-functional protein sequences detected from the Rabbit Creek monitor site. For a description of terms, see text.

Sample Date	Sample Name	Unique DNA Sequences	Functional Protein Sequences	Non-Functional Protein Sequences
July 26, 2002	RC7	15	9	1
October 17, 2002	RCM	15	8	4
January 20, 2003	RCMS	45	25	14
August 26, 2003	RCMS5	4	1	3
October 2, 2003	RCMS6	9	7	1
November 13, 2003	RCMS7	12	4	2
January 19, 2004	RCMS8	8	3	3
TOTAL		**112**	**49**	**28**

SSV-like viral coat proteins (7F/8R) were also detected during the first sampling point (table 4.3). Fifty SSV-like coat protein clones were obtained and compared to the coat proteins from the completely sequenced isolates (figure 4.14a). The four sequenced SSV isolates show a different relatedness to the environmental sequences based on coat protein analysis compared to the 3F/4R analysis. SSV2 and SSV-K1 are not closely related to the Rabbit Creek environmental sequences based on 3F/4R (figure 4.13a). This

result was expected, as it has been determined the ORFs in the four sequenced SSV genomes did not show geographic correlation (Wiedenheft 2004). All of the four sequenced isolates are related to the environmental 7F/8R sequences (figure 4.14a).

Figure 4.11. The different groups of SSV-like viruses detected in the Rabbit Creek monitor site over time based on DNA sequences. The colors of the groups correspond to figure 4.15. The sampling dates without sequences represent the inability to amplify SSV-like sequences from the monitor site.

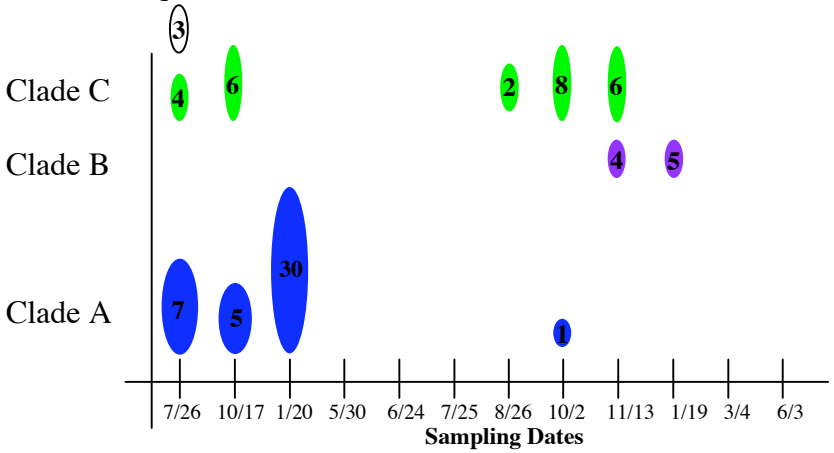
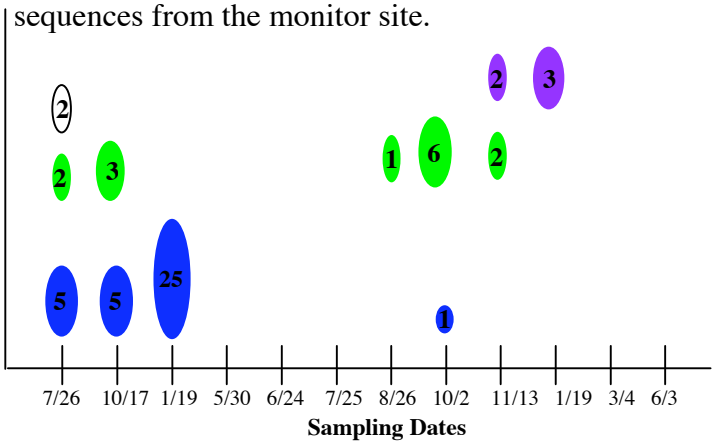


Figure 4.12. The different groups of SSV-like viruses detected in the Rabbit Creek monitor site over time based on protein sequences. The colors of the groups correspond to figure 4.16. The sampling dates without sequences represent the inability to amplify SSV-like sequences from the monitor site.



When the SSV-like coat protein sequences are translated, a very different tree is obtained (figure 4.14b). The coat proteins are distinct from SSV1, SSV2, and SSV-RH, and more related to SSV-K1, which was not observed in the DNA analysis (figure 4.14a). It is interesting to mention that when the 50 coat protein DNA sequences are translated to protein sequence, only 16 sequences result in functional proteins (figure 4.14b). These results indicate we may be amplifying coat protein sequences from viruses integrated into their host's chromosome. There were numerous stop codons in the non-functional 7F/8R protein sequences. This may be indicative of a past integration event.

Figure 4.13. SSV-like 3F/4R sequences collected in July 2002 from the Rabbit Creek monitor site. The numbers in parentheses indicate how many times a particular sequence was detected. A. DNA sequences; the • represents the non-functional ORF. B. Inferred protein sequences.

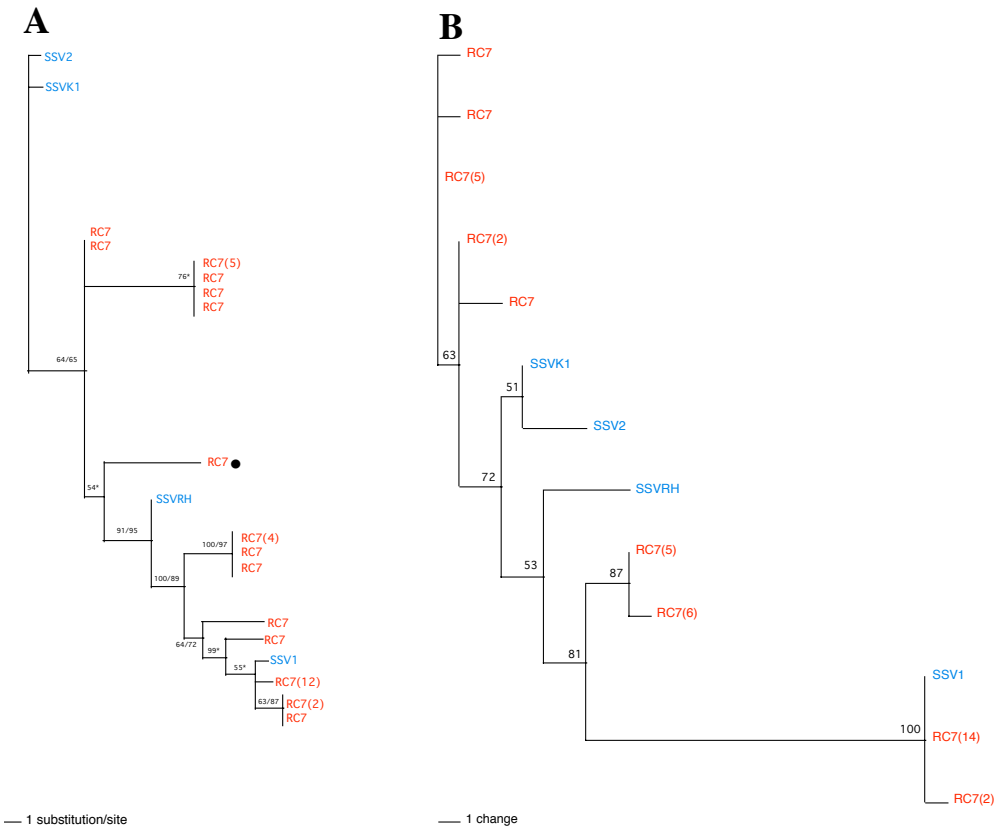
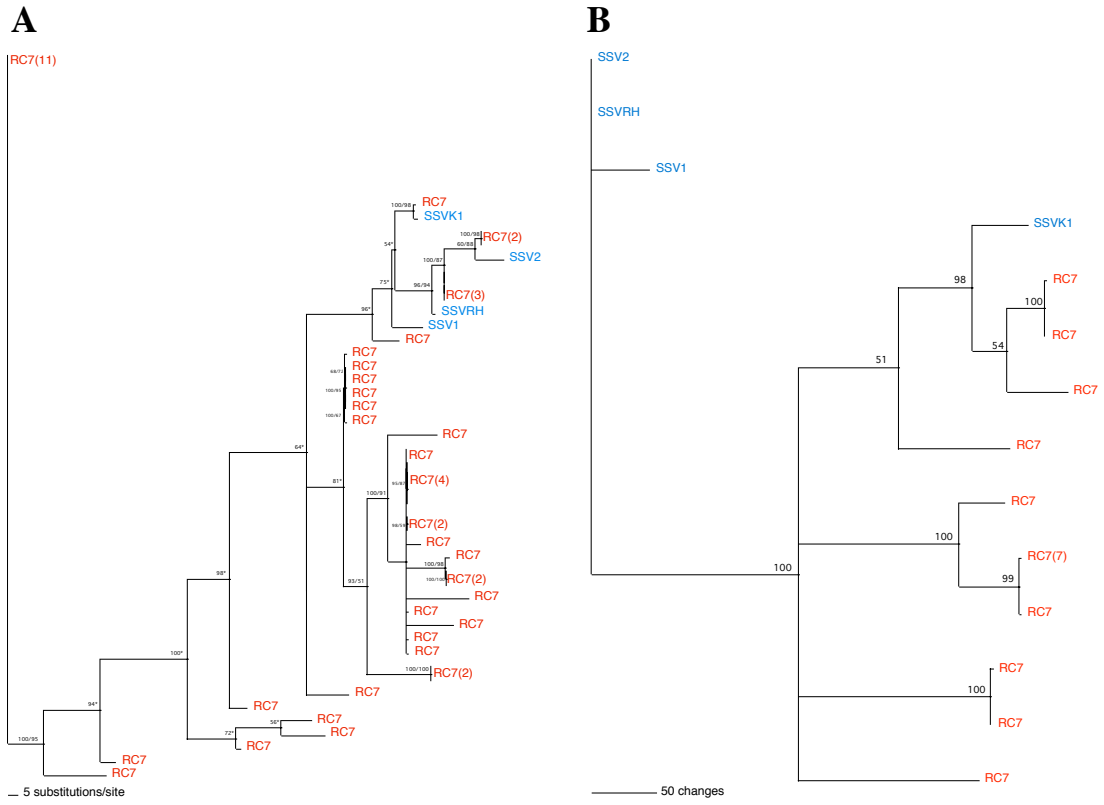


Figure 4.14. SSV-like 7F/8R sequences collected in July 2002 from Rabbit Creek. The numbers in parentheses indicate how many times a particular sequence was detected. A. DNA sequences. B. Inferred protein sequences.



The second sample was collected on October 17, 2002, and 27 SSV-like 3F/4R clones were sequenced (table 4.3). When the 27 clones from this sample were compared to the sequences from July 2002, the viral population structure did not drastically change. The dominant viral groups present in July 2002 are still present in October 2002 (clades A and C in figure 4.15). The inferred protein analysis resulted in a similar tree topology; most of the sequences from July 2002 were similar to the October 2002 sequences (represented by the blue and green in figure 4.16). There are three protein sequences that

are detected in this sample that are identical to protein sequences detected in the previous sample (figure 4.16; identical sequences represented by *).

Figure 4.15. The functional ORFs from SSV-like DNA sequences detected in Rabbit Creek over time. The number in parentheses indicates how many times a particular sequence was detected. The colors correspond to figure 4.11.

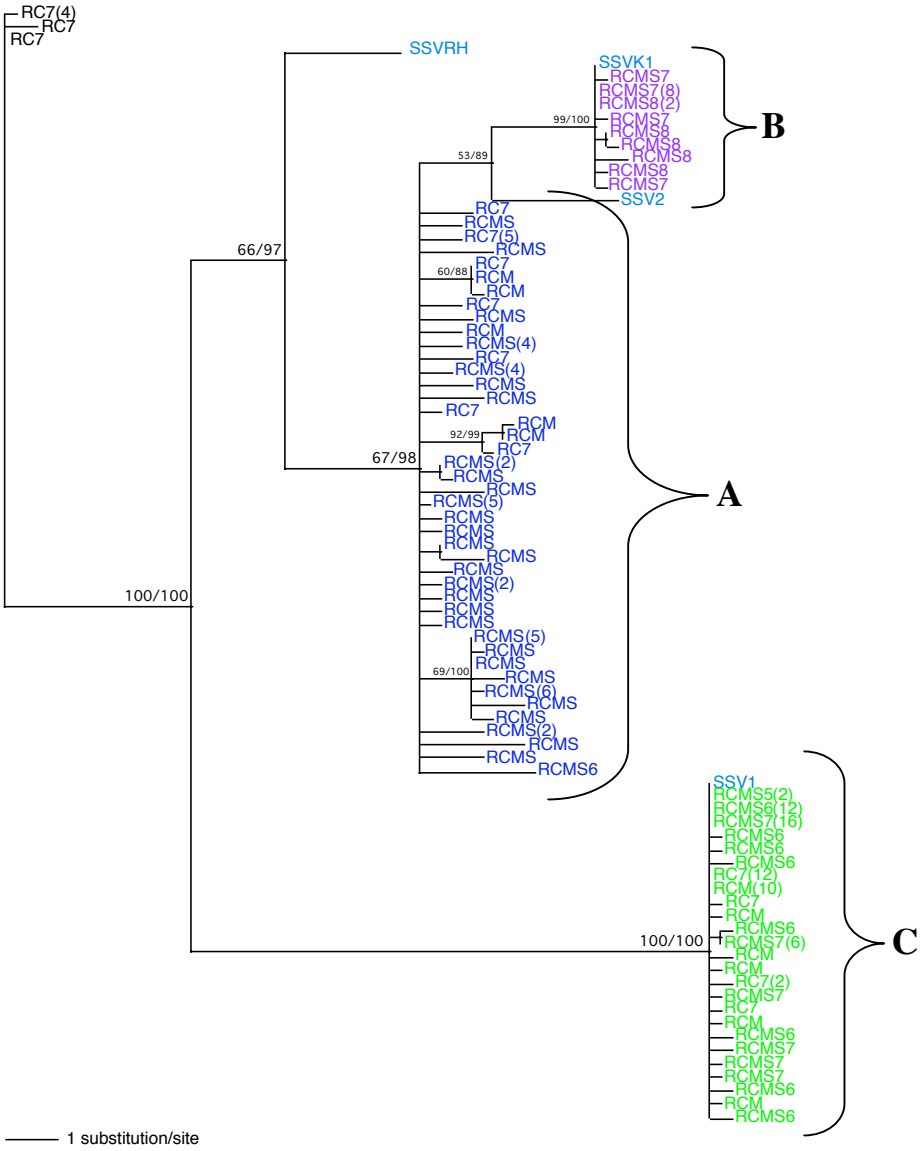
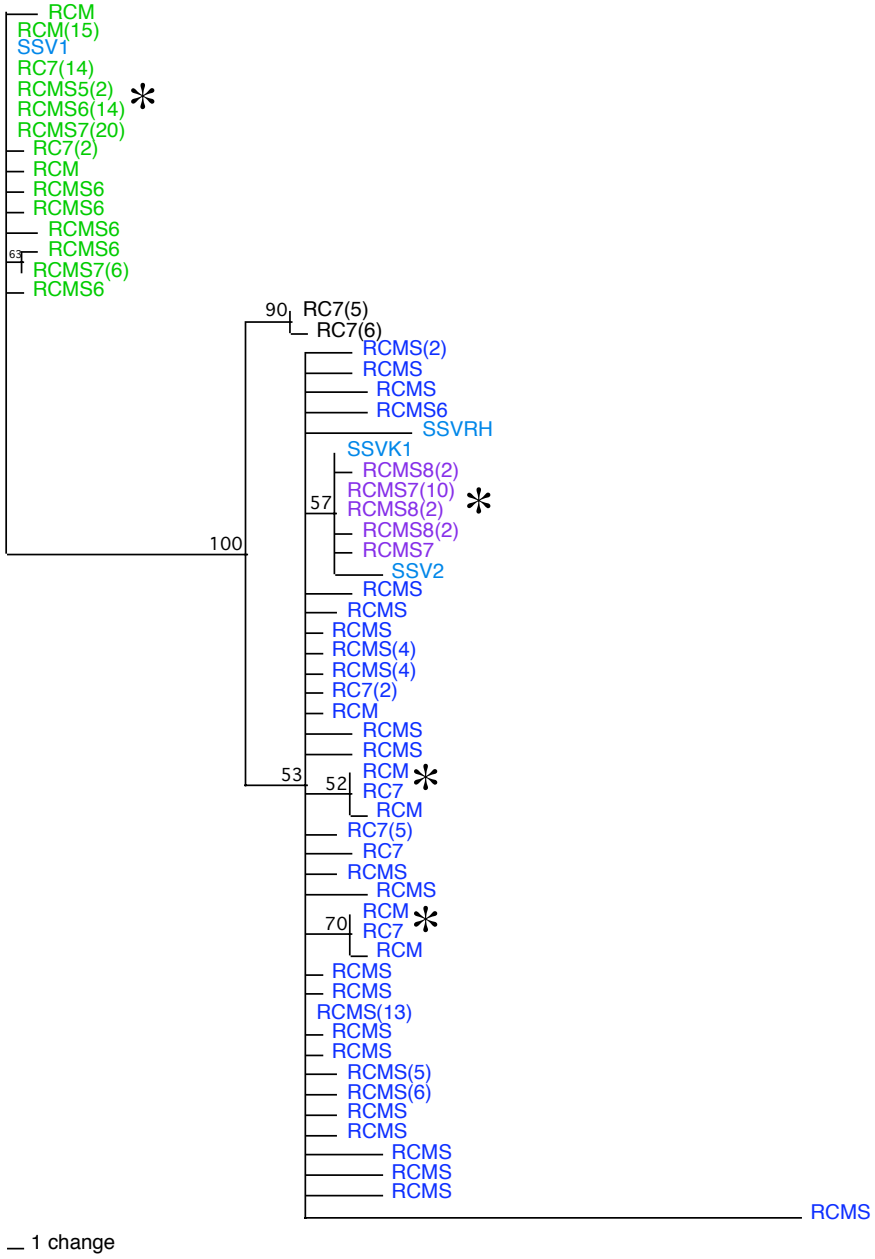


Figure 4.16. The groups of SSV-like 3F/4R protein sequences detected from July 2002 to January 2004 in Rabbit Creek. The * represent the identical protein sequences detected. The numbers in parentheses indicate how many times a protein sequence was detected. The colors correspond to figure 4.12.



The translated sequences from this sample resulted in eight functional ORFs and four non-functional ORFs (table 4.4). Two of the non-functional ORFs are obvious in the phylogenetic analysis because they have long branch lengths, and they are related to the non-functional sequence from the previous sample (see Appendix B, figure 4.29). These two sequences contain numerous stop codons, and the two remaining non-functional sequences are similar to the functional ORFs, except at codon 68. There is a second position transversion and a third position transition that results in a leucine (TTA) being changed to a stop (TAG).

The third sample was collected on January 20, 2003, and 67 SSV-like 3F/4R clones were sequenced (table 4.3). This month most of the SSV-like sequences are similar to the sequences detected in previous months (clade A in figure 4.15). The translation of the DNA sequences resulted in 25 functional ORFs and 14 non-functional ORFs (table 4.4). All of the sequences from this sample are included in the largest group of protein sequences (represented in blue in figure 4.16). There were not any sequences in this sample identical to protein sequences detected in earlier samples (July 2002 and October 2002). Most of the non-functional ORFs from this sample are obvious in the phylogenetic analysis (see Appendix B, figure 4.30). One has a very long branch length and is related to the interrupted sequence from the July 2002 sample. The remaining non-functional ORFs compose two groups in the phylogenetic tree. These groups are strictly composed of the non-functional sequences from this sample excluding the two non-functional sequences from October 2002. Upon closer investigation, all of these sequences were found to contain numerous stop codons.

Three more samples were collected over the next six months (February 2003—July 2003), but no SSV-like sequences were detected in these samples (table 4.3). This is most likely due to a low abundance of SSV-like viruses in the natural environment (<0.1pg). It is not expected to be the result of a technical issue as SIRV-like sequences (discussed below) were detected in these samples.

The next sample that contained SSV-like sequences was collected on August 26, 2003, but only five SSV-like 3F/4R clones were obtained and sequenced (table 4.3). When these SSV-like sequences were compared to sequences detected in previous samples, they are related to sequences from July 2002, October 2002, and SSV1 (clade C in figure 4.15). Over the past few months, the SSV-like population in Rabbit Creek shifted from representing clade A to sequences representing clade C (figure 4.15). However, conclusions regarding the SSV-like population must be made carefully because of the lack of clones.

Of the five clones sequenced from this sample, only one represented a functional protein sequence (table 4.4). The inferred protein-based analysis indicates this sequence is related to protein sequences from previous samples and SSV1 (represented in green in figure 4.16). In fact, the sequence is identical to SSV-like 3F/4R protein sequences detected in previous samples (figure 4.16; the * represents the identical sequences). When looking at the DNA-based phylogenetic analysis, it is not obvious what sequences are non-functional (see Appendix B, figure 4.31). The non-functional ORFs contain a first position transversion at nucleotide 21, which results in a glycine (GGA) being

changed to a stop codon (TGA). However, besides this one change, the remaining sequence is almost identical.

When SSV-like sequences collected from October 2003 (table 4.3) were compared to the sequences isolated from earlier samples, 20 clones are related to SSV1 and Rabbit Creek sequences from July 2002, October 2002, and August 2003 (clade C in figure 4.13). One sequence from this sample belonged to clade A (figure 4.15). This sample contained seven functional protein sequences and one sequence that resulted in a non-functional ORF (table 4.4). The protein sequences from this sample belong in two groups of viral sequences (represented by blue and green in figure 4.16). Most of the protein sequences are identical to a sequence that has been detected in previous samples (figure 4.16; the identical sequences are represented by *). The other protein sequence is located within the largest group of sequences (represented in blue in figure 4.16). The non-functional sequence contains a first position transversion that results in a glycine (GGA) being changed to a stop codon (TGA) at nucleotide position 19 (see Appendix B, figure 4.33).

The next sample was collected in November 2003 resulted in 44 SSV-like 3F/4R clones (table 4.3). When compared to sequences from previous sampling points, most of the new sequences belong in groups previously established (clade C in figure 4.15). The majority of the November 2003 sequences resembled sequences from July 2002, October 2002, August 2003, and October 2003 (clade C in figure 4.15). However, 11 of the sequences were most similar to SSV-K1 (clade B in figure 4.16). There were four inferred functional protein sequences and two non-functional protein sequences in this

sample (table 4.4). The inferred protein sequences detected from this sample belonged to two groups of protein sequences represented in this monitor site. Two inferred protein sequences belonged to a group of protein sequences detected previously in this monitor site (represented by green in figure 4.16). The remaining protein sequences belonged to a new group of protein sequences related to SSV-K1 (represented in purple in figure 4.16). The non-functional sequences contain a first position transversion that causes a glycine (GGA) to become a stop codon (TGA) at nucleotide position 19.

The SSV-like viral coat proteins (7F/8R) were detected in November 2003 and 61 clones were sequenced (table 4.3). When compared to the SSV-like coat protein sequences from July 2002, the sequences from this sample are very different from the sequences detected earlier (figure 4.17). As opposed to the July 2002 sample, there were not any sequences from this sample that resulted in non-functional ORFs. The inferred protein analysis also reveals the differences between the two sampling times (figure 4.18). Most of the protein sequences from this sample do not show any relatedness to sequences from the July 2002 sample (figure 4.18). Perhaps it is the change in viral sequences that occurred over one year. The SSV-like viral coat proteins were not detected again from this monitor site (table 4.3).

The last sample that SSV-like sequences were detected was collected on January 19, 2004 (table 4.3). Only ten clones were sequenced from this sample, and they represent one group in the SSV-like population (clade B in figure 4.15). The SSV-like sequences show relatedness to SSV-K1 (figure 4.15). Translating the sequences from this sample resulted in three functional ORFs and three non-functional ORFs (table 4.4).

The three inferred functional protein sequences belonged to the group of sequences in this monitor site related to SSV-K1 (represented in purple in figure 4.16). The three non-functional ORFs from the January 2004 sample are grouped with the non-functional sequences from July 2002 (see Appendix B, figure 4.37). These three sequences each contain five stop codons that are a result of an insertion of a cytosine (C) at nucleotide position 49, which resulted in a frameshift.

Figure 4.17. SSV-like 7F/8R DNA sequences collected from Rabbit Creek. The red sequences represent those collected in November 2003. The numbers in parentheses indicate how many times a sequence was detected.

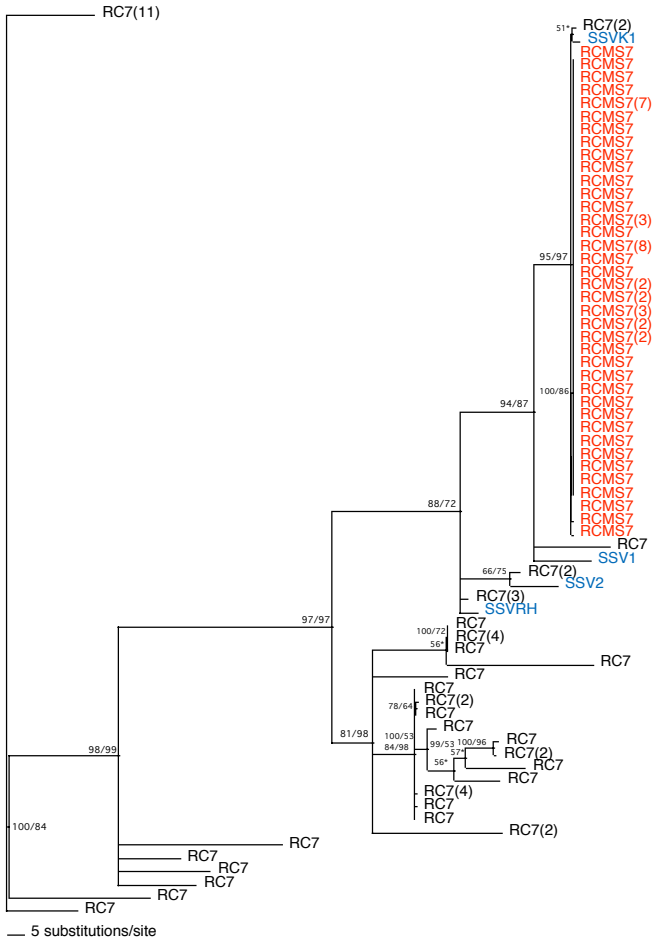
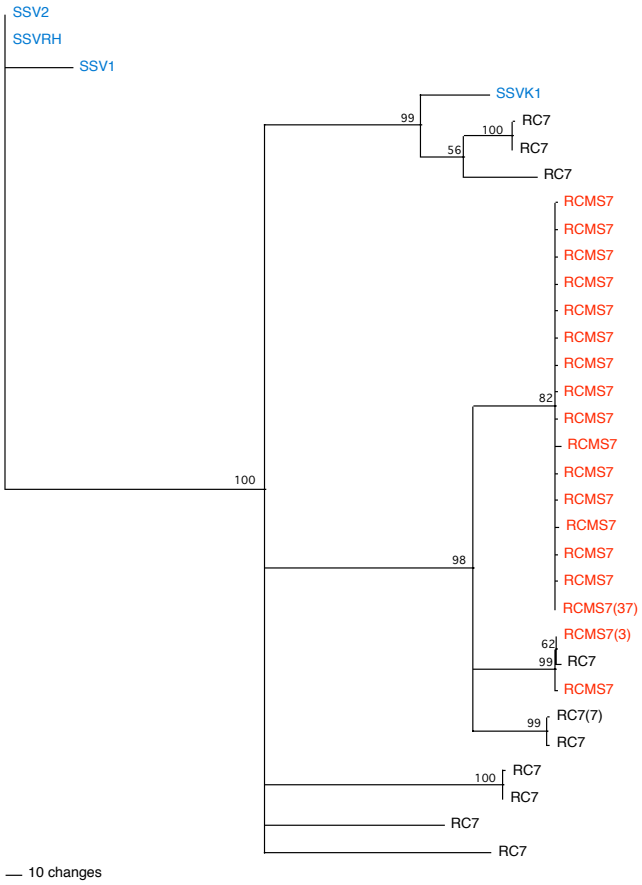


Figure 4.18. SSV-like 7F/8R protein sequences collected from Rabbit Creek. The red sequences represent those collected in November 2003. The numbers in parentheses indicate how many times a protein sequence was detected.



Two more samples were collected on March 4, 2004 and June 3, 2004; however, there were no SSV-like sequences (3F/4R or 7F/8R) detected from either of these samples (table 4.3). I believe this result reflects the low abundance of the viruses in the natural environment and not a technical error.

In summary, twelve samples were collected over the course of the two years of this research (July 2002—June 2004). Based on the SSV-like 3F/4R primer set, there

were three groups of SSV-like DNA sequences during the course of sampling this monitor site (figure 4.15). Inferred protein analysis of the 3F/4R SSV-like sequences reveals the presence of two major groups of viruses over time with a smaller group related to SSV-K1 (figure 4.16). The complete phylogenetic analysis of SSV-like sequences included analysis of sequences after every sample. For a step-by-step view of the changes in the SSV-like population over time, see Appendix B.

It is interesting to compare the phylogenetic positioning of functional and non-functional DNA sequences (figure 4.19). This DNA-based analysis was performed including the non-functional SSV-like sequences (figure 4.19; green represents the non-functional sequences). When the non-functional ORFs are removed from the SSV-like 3F/4R DNA analysis, the topology of the phylogenetic tree is similar, but the presence of distinct groups of SSV-like viruses is more obvious (figure 4.15). Without the long branches in the DNA-based analysis, it is much easier to visually determine the presence of different SSV-like viral groups present from July 2002 to January 2004 (figure 4.15). Instead of four viral groups, as could be estimated from figure 4.17, there are only three major SSV-like viral groups (figure 4.15). The SSV-like 3F/4R sequences from this monitor site are more closely related to SSV1, SSV2, and SSV-K1 than they are to the SSV from YNP (SSV-RH) (figure 4.15).

The SSV-like sequences appear to vary through time (figure 4.20). The first group related to SSV1 (clade C; represented by green in figure 4.20) is represented by sequences from July—October 2002 and August—November 2003. The largest group of viral sequences is composed of sequences from July—October 2002, January 2003, and

October 2003 (clade A; colored blue in figure 4.20), and the minor group related to SSV-K1 (clade B; colored purple in figure 4.20) contains sequences from November 2003 and January 2004. The other, smaller group of SSV-like sequences contains only three sequences from July 2002 (colored black in figure 4.15). By looking at the phylogenetic analysis without the non-functional ORFs, it is more obvious that there were three distinct groups of SSV-like viruses that vary over time (figure 4.20).

It is more difficult to conclude how many virus groups were present based on the SSV-like coat protein DNA analysis because the coat proteins were only detected twice (one year apart) (table 4.3). However, we can estimate there were approximately two groups of SSV-like viruses based on coat protein DNA sequences (figure 4.17). The two groups vary with time; the sample from July 2002 is more diverse than the November 2003 (figure 4.17).

Of the 49 unique SSV-like 3F/4R protein sequences, there were only four protein sequences that were detected in more than one sample (figure 4.16; the * represents groups of identical protein sequences). By translating the 3F/4R sequences retrieved from the environment, we could determine what sequences may be representing old integrated forms of SSVs. These interrupted ORFs were not included in most of the inferred protein-based analyses. However, these sequences were included in an analysis of all the SSV-like protein sequences (figure 4.21).

When the 28 non-functional protein sequences (table 4.4) are added to analysis, there are several interesting results (figure 4.21). Nineteen of the interrupted SSV-like ORFs are very distinct from the functional protein sequences (figure 4.21; the interrupted

ORFs are represented in red). When these protein sequences are investigated more closely, there are many insertions, deletions, transversions, and transitions resulting in numerous stop codons throughout the sequence. Five of the non-functional protein sequences are closely related to the group of functional protein sequences related to SSV1 (figure 4.21; represented in red with **). It was unexpected that these SSV-like sequences were closely related to functional protein sequences when most of the non-functional sequences were so distinct. However, when these protein sequences are more closely examined, there is a first position transversion at nucleotide 19 changing the codon from a glycine (GGA) to a stop codon (TGA). Besides that change, these protein sequences are very similar to the group of functional protein sequences (represented by green in figure 4.21) related to SSV1. The remaining two non-functional ORFs are located in the largest group of protein sequences from Rabbit Creek (represented in blue in figure 4.21; represented in red with *). When these sequences are more closely investigated, it was found an adenine (A) was deleted at nucleotide position 153 which caused a frameshift. Then, at nucleotide position 182 a cytosine (C) was inserted, which resulted in the re-establishment of original frame of the ORF. However, because of these insertions and deletions, two stop codons resulted.

It is not likely the differences in functional and non-functional protein sequences are the result of PCR or sequencing errors. As described for Crater Hills (see Chapter 3), the SIRV-like sequences did not result in non-functional protein sequences. SIRVs have not been shown to integrate into their host's chromosome (Prangishvili 1999). If the

differences were the result of PCR or sequencing errors, we would also expect to see errors in the SIRV-like sequences.

Figure 4.19. All SSV-like 3F/4R DNA sequences detected in the Crater Hills monitor site over time. This analysis included the SSV-like DNA sequences that resulted in non-functional ORFs. The interrupted ORFs are colored green. The viral clades from figure 4.15 are indicated.

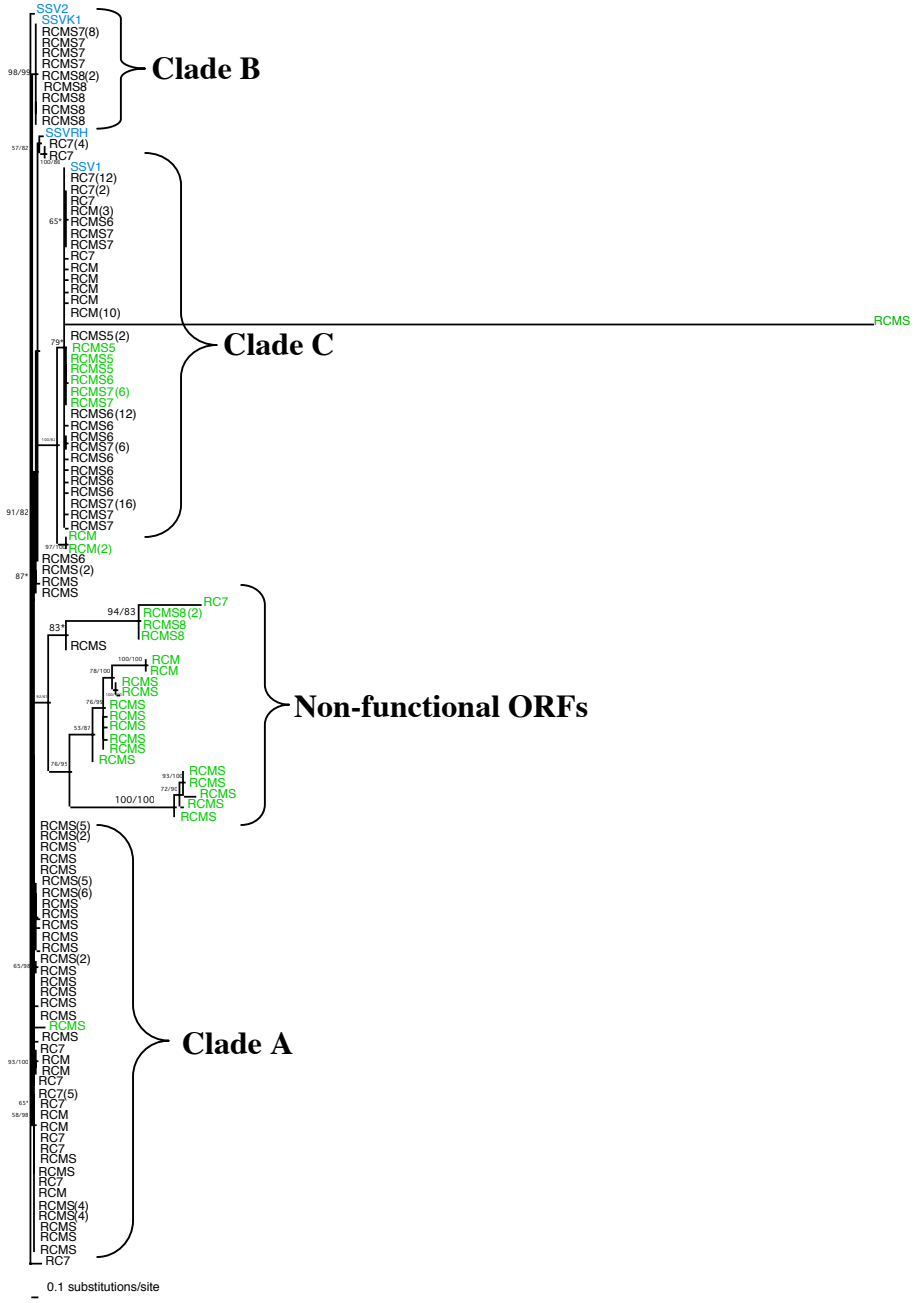


Figure 4.20. SSV-like 3F/4R DNA sequences collected over time in Rabbit Creek. The sequences that resulted in non-functional ORFs are not included in this analysis. The unrooted phylogram makes it easy to see the presence of three major SSV-like groups in the Rabbit Creek monitor site over time.

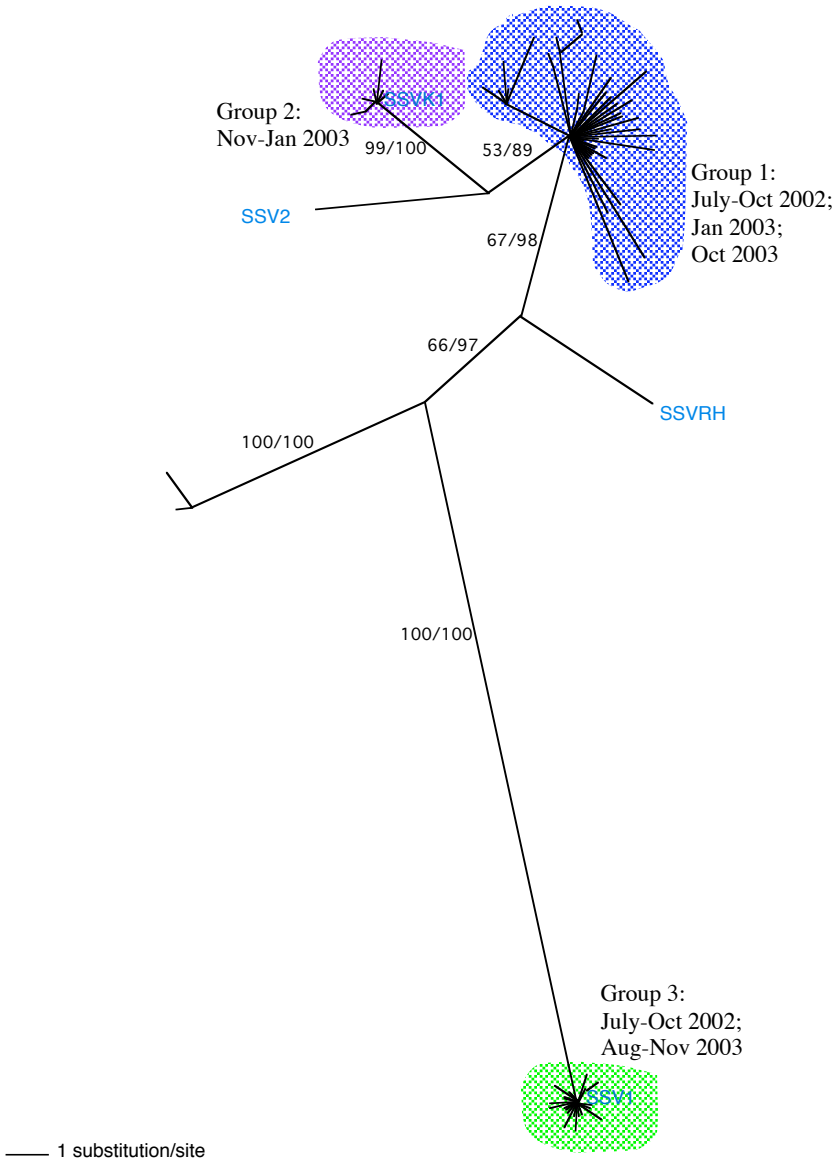
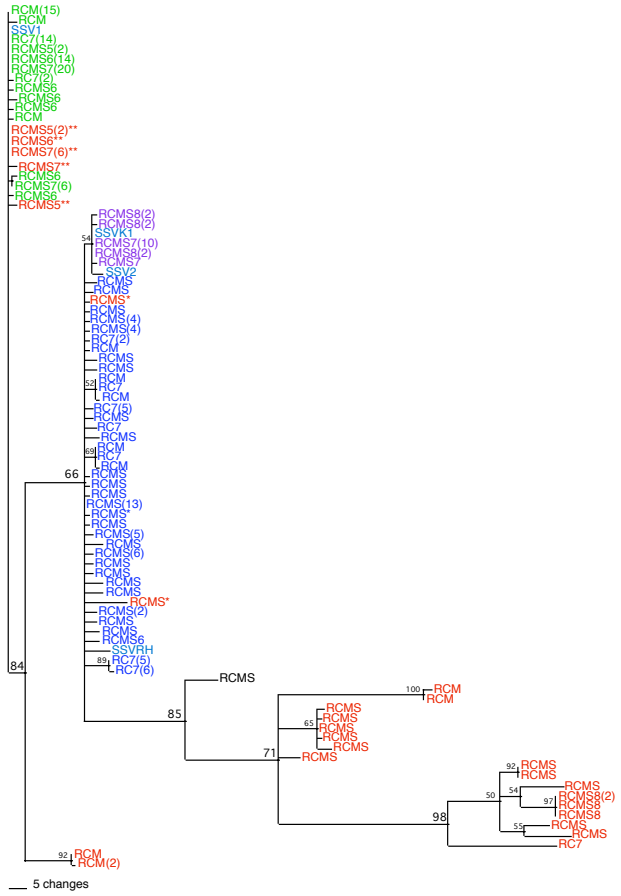


Figure 4.21. SSV-like 3F/4R protein sequences detected in Rabbit Creek from July 2002 to January 2004. The red sequences represent non-functional ORFs. The numbers in parentheses indicate how many times a protein sequence was detected.



Phylogenetic analysis of the inferred amino acid sequences of the SSV-like coat proteins reveal the presence of one major group with a couple of outlying sequences (figure 4.18). More SSV-like coat protein sequences would be necessary to better determine the number of SSV-like groups in this monitor site. There are numerous reasons as to why the SSV-like 3F/4R sequences were more commonly detected. First, the viral sequences amplified by 3F/4R and 7F/8R are in different parts of the viral

genome (figure 3.3), therefore, they may be under different selection pressures. Second, the length of the PCR amplicon for 3F/4R is smaller than the 7F/8R amplicon (~250bps for 3F/4R and ~800bps for 7F/8R; table 3.1). Finally, the 3F/4R primer set was more sensitive in amplifying sequences from purified viral template (0.1pg) compared to the 7F/8R primer set (1ng purified template).

Detailed analysis of SSV-like population dynamics. From July 2002 to June 2004, there was a total of 208 SSV-like 3F/4R clones sequenced from this monitor site (table 4.3). Using Modeltest 3.6 (Posada *et. al.* 1998), the nucleotide substitution model for maximum likelihood estimations was determined to be HKY+G (Hasegawa *et. al.* 1985) with heterogeneity between sites (gamma=6.3662), base frequencies A=0.2608, C=0.2123, G=0.2163, and T=0.3105, and the ratio of transitions to transversions is 0.8384.

The SSV-like 3F/4R sequences represented 94 groups of identical sequences. The largest group of sequences was composed of 52 identical sequences, which comprised 25% of the SSV-like population over time in this monitor site (table 4.5). This group of sequences contained sequences from most samples (excluding the January 2003 sample) (table 4.5). The second largest group of sequences was comprised of only 10 identical sequences, which makes up 4.8% of the SSV-like population (table 4.5). However, unlike the larger group, this group was only composed of sequences that were sampled later in this research. Therefore, this group may represent a new group of viruses. There were six different groups of six sequences, and each group comprised 2.9% of the total

SSV-like population (table 4.5). These groups contained samples from early and later collection dates (table 4.5).

Table 4.5. Rabbit Creek SSV-like 3F/4R sequence abundance illustrating the groups of identical sequences, how many times a group this size was detected, the percent of the SSV-like population, and what samples the group contained.

RABBIT CREEK SSV-like SEQUENCE ABUNDANCE			
Identical Sequences	Times Detected	Percent of SSV-like Population	Samples[a]
52	1	25%	RC7;RCM;RCMS5;RCMS6;RCMS7
10	1	4.8	RCMS7;RCMS8
6	1	2.9	RC7;RCMS
6	1	2.9	RC7;RCM;RCMS
6	1	2.9	RCMS
6	1	2.9	RCMS
6	1	2.9	RCMS7
6	1	2.9	RCMS7
5	2 unique[b]	2.4% each[c]	
4	3 unique	1.9% each	
2	7 unique	0.96% each	
1	73 unique	0.48% each	

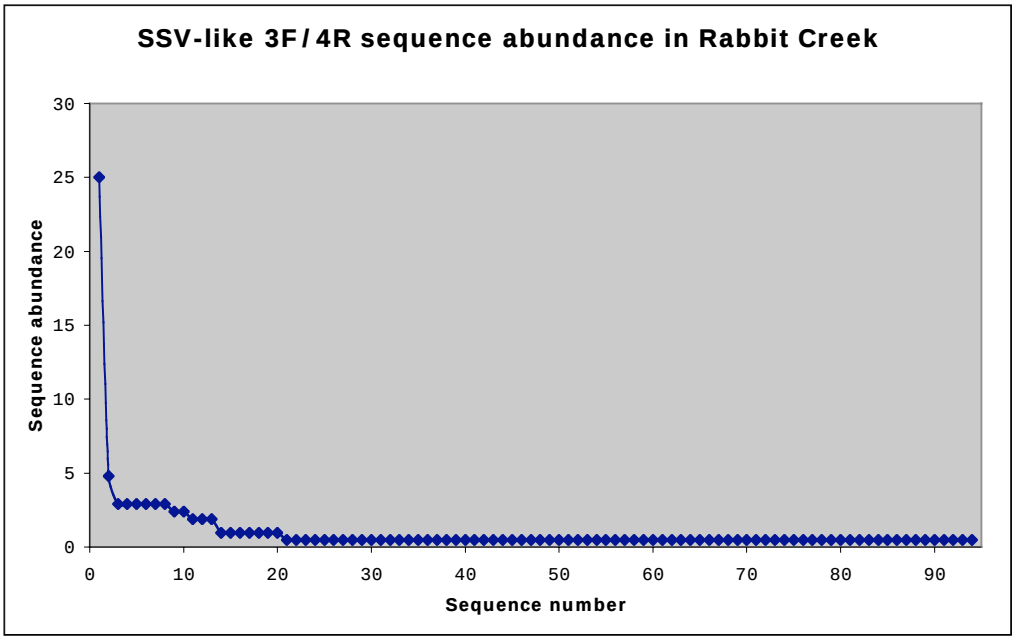
[a]sequences from these samples were present in the indicated groups
[b]each of the smaller groups were detected this many times, however, each is unique
[c]each of the unique groups makes up this percentage of the population

There were other smaller groups of identical sequences (table 4.5). They ranged from groups of two to five identical sequences; however, similar to the SSV-like population in Crater Hills (see Chapter 3), the unique SSV-like 3F/4R sequences made up the majority of the SSV-like population (figure 4.22). There were 73 unique SSV-like 3F/4R sequences, which translates to 35% of the sampled SSV-like population.

In comparing the 112 unique SSV-like DNA sequences (3F/4R) obtained over time, 175 bps (out of an average of 255 bps) did not vary (tables 4.6 and 4.7). When

comparing the base changes in each position of the codon, the third position base changes more frequently than the first or second base positions (tables 4.6 and 4.7). The rates of transitions and transversions varied with individual samples; some samples contained more changes than other samples (table 4.6).

Figure 4.22. A sequence abundance plot for SSV-like 3F/4R sequences from Rabbit Creek monitor site. The left end of the x-axis represents most commonly detected sequences, and the right end of the x-axis represents the most rare sequences. Most of the SSV-like sequences in Rabbit Creek are detected once.



When comparing all of the sequences, only 67% of the bases in the 3F/4R sequence remained unchanged (table 4.7). Transversions occurred more (50) compared to transitions (31). Twenty-five changes occurred in the first position of the codon; there were 10 transitions and 15 transversions (table 4.7). Only 22 changes occurred in the second position with seven transitions and 15 transversions (table 4.7). A total of 34 changes occurred in the third position; there were 15 transitions and 19 transversions

(table 4.7). As expected, the third position changes occurred more frequently, therefore explaining the fewer number of groups based on protein sequences (figure 4.16).

Table 4.6. Number of transitions and transversions in SSV-like 3F/4R sequences collected from the Rabbit Creek monitor site.

SAMPLE DATE	IDENTICAL BASES	TRANSITIONS			TRANSVERSIONS		
		1st	2nd	3rd	1st	2nd	3rd
07/26/2002	191/256	10	4	14	11	9	16
10/17/2002	208/255	8	3	12	7	5	13
01/20/2003	230/258	4	3	5	6	5	6
08/26/2003	252/253	0	0	1	0	0	0
10/02/2003	229/253	4	2	6	3	2	6
11/13/2003	196/254	8	4	19	9	5	13
01/19/2004	181/256	8	6	4	20	18	21

Table 4.7. The numbers of transitions and transversions for all SSV-like 3F/4R sequences collected from Rabbit Creek throughout this research.

Sample Range	Position	Identical Base Pairs	Ts Pairs	Tv Pairs
July 2002-Jan 2003 (255 bps)	*Overall*	*175/255*	*31*	*50*
	1st	60/85	10	15
	2nd	64/86	7	15
	3rd	51/85	15	19

Ts=transitions; Tv=transversions

Only two of the twelve samples were positive for the SSV-like coat protein (primer set 7F/8R) (table 4.3). Using Modeltest version 3.6 (Posada *et. al.* 1998), the nucleotide substitution model for maximum likelihood estimations was determined to be TVM+I+G (Posada *et. al.* 1998) with heterogeneity between sites (gamma=0.2560), invariant sites (I=0.0029), base frequencies A=0.2868, C=0.2079, G=0.2310, and T=0.2743, and a substitution model A-C=G-T=1.0000, A-G=1.4098, A-T=C-G=0.9561, and C-T=1.7129.

In comparing the 65 unique SSV-like coat protein DNA sequences (7F/8R) obtained over time, 512 bps (out of an average of 667 bps) did not change in all of the samples. Unexpectedly, the third position base does not change more frequently than the first or second base positions. In the first sample collected in July 2002, 205 bps were identical over an average of 561 total base pairs (36% of the bases were identical) (table 4.8). The second sample collected in November 2003 revealed less change when compared to the first sample (table 4.8); 662 bps out of 706 bps were identical (93%).

Table 4.8. Number of transitions and transversions in SSV-like 7F/8R sequences collected from the Rabbit Creek monitor site.

SAMPLE DATE	IDENTICAL BASES	TRANSITIONS			TRANSVERSIONS		
		1st	2nd	3rd	1st	2nd	3rd
07/26/2002	205/561	50	36	41	72	79	79
11/13/2003	662/706	6	6	4	9	8	11

When comparing all of the sequences, 77% of the bases in the coat protein (7F/8R) sequence remained unchanged (table 4.9). Transversions occurred more frequently (99) compared to transitions (56). Fifty-five changes occurred in the first position of the codon; there were 22 transitions and 33 transversions. Only 48 changes occurred in the second position with 17 transitions and 31 transversions. A total of 53 changes occurred in the third position; there were 18 transitions and 35 transversions.

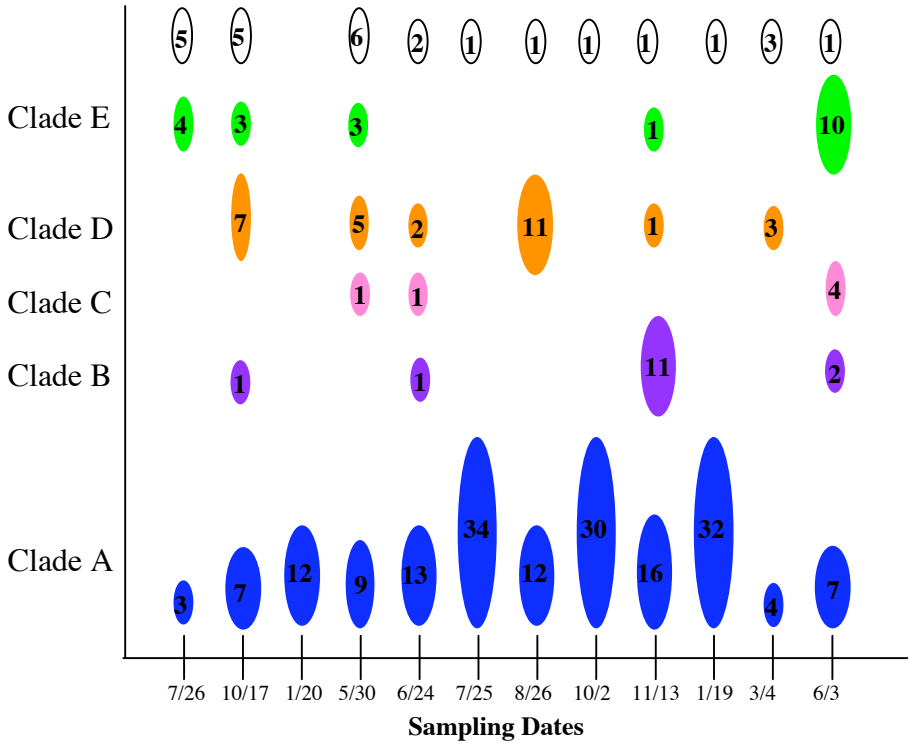
Table 4.9. Transitions and transversions for all SSV-like 7F/8R sequences collected from Rabbit Creek throughout this research.

Sample Range	Position	Identical Base Pairs	Ts Pairs	Tv Pairs
July 2002-November 2003 (667 bps)	*Overall*	*512/667*	*57*	*99*
	1st	168/222	22	33
	2nd	174/222	17	31
	3rd	170/222	18	35

Ts=transitions; Tv=transversions

SIRV-like population dynamics in the Rabbit Creek monitor site. SIRV-like coat proteins were detected in all the samples (table 4.10). As observed in Crater Hills, all of the SIRV-like coat proteins translated to un-interrupted ORFs. An overview of the DNA-based and protein-based phylogenetic analyses is represented in figures 4.23 and 4.24.

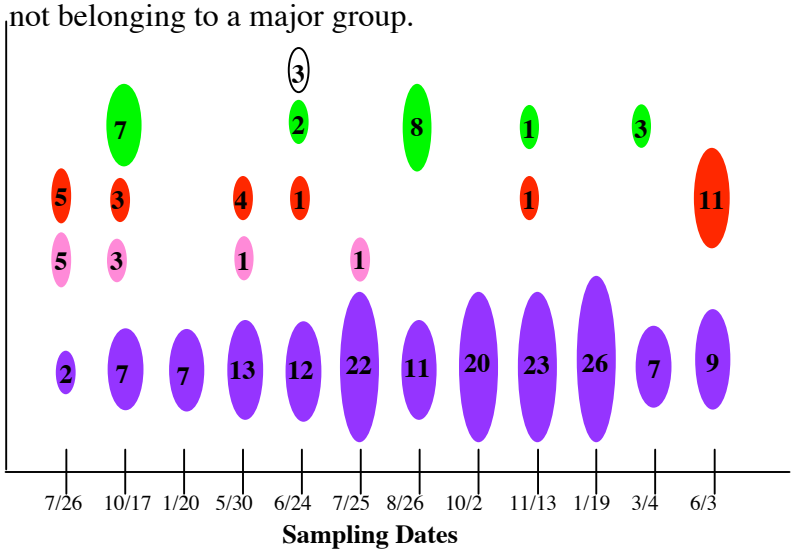
Figure 4.23. The different groups of SIRV-like viruses detected in the Rabbit Creek monitor site over time. The colors of the groups correspond to figure 4.26. The clear ovals refer to outlying sequences not belonging to a major group.



Forty-four SIRV-like coat protein sequences were collected from the first sampling point in July 2002 (table 4.10). The viruses form three major groups in Rabbit Creek with one group showing diversity (figure 4.25a). The YNP sequences are distinct

from the two Icelandic sequences (SIRV1 and SIRV2) (figure 4.25a). The 12 unique DNA sequences translated to 12 unique inferred protein sequences, which illustrated similar topology as the DNA-based phylogenetic analysis (figure 4.23b).

Figure 4.24. The different groups of SIRV-like viruses detected in the Rabbit Creek monitor site over time based on protein sequences. The colors of the groups correspond to figure 4.27. The clear oval refers to outlying sequences not belonging to a major group.



Change in the SIRV-like population is detected in the second sample. In October 2002, twenty-nine SIRV-like clones were sequenced (table 4.10). The groups from July 2002 are still represented in October 2002 (clades A and E) (figure 4.26). A new group of SIRV-like coat protein sequences is also detected (clade D in figure 4.26). The 22 SIRV-like coat protein DNA sequences translated to 20 unique inferred protein sequences (table 4.10), and four groups of viral protein sequences are evident (represented by the purple, pink, red, and green groups in figure 4.27).

Table 4.10. Number of SIRV-like clones obtained from the Rabbit Creek monitor site, number of unique DNA sequences in each library, and the number of unique protein sequences.

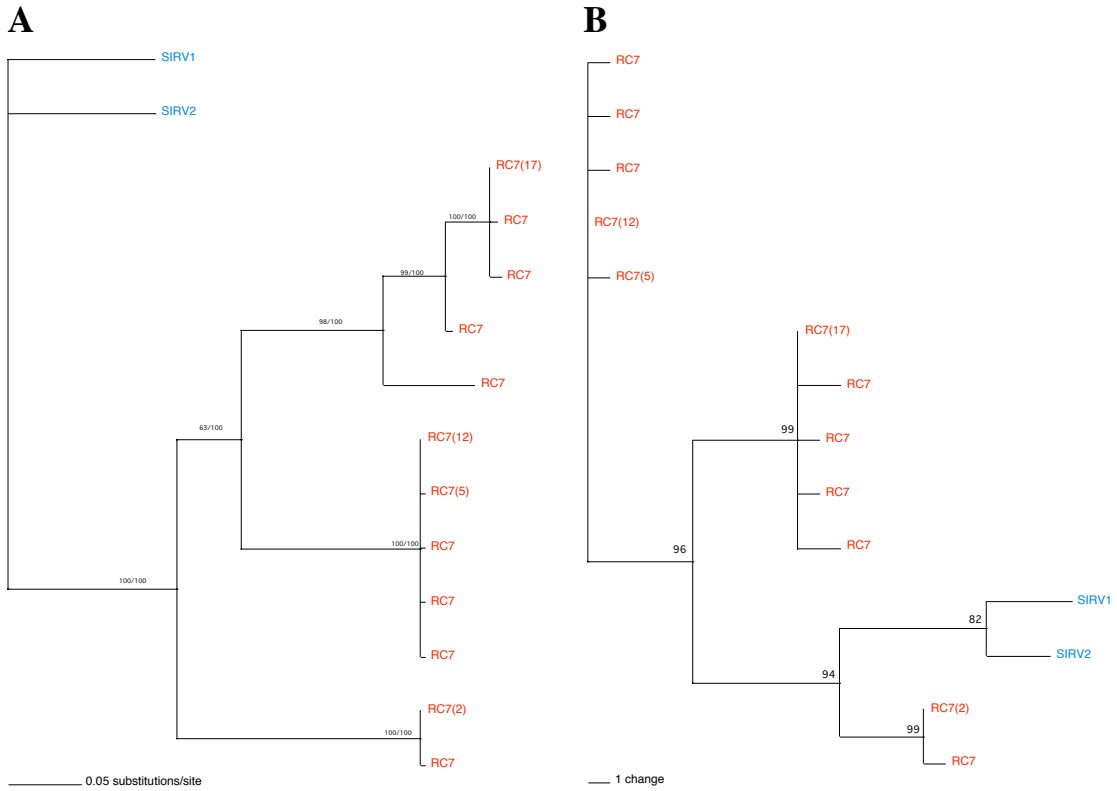
Sample Name	Collection Date	pH	Temp (°C)	Number of Clones	Unique Sequences	Protein Sequences
RC7	July 26, 2002	3.8	88	44	12	12
RCM	Oct 17, 2002	2.8	91	29	22	20
RCMS	Jan 20, 2003	5.1	75	66	12	7
RCMS2	May 30, 2003	4.1	78	39	24	18
RCMS3	June 24, 2003	3.8	85	21	19	18
RCMS4	July 25, 2003	3.7	73	145	35	23
RCMS5	Aug 26, 2003	4.2	81	80	24	19
RCMS6	Oct 2, 2003	3.2	71	98	31	20
RCMS7	Nov 13, 2003	3.8	64	57	30	25
RCMS8	Jan 19, 2004	3.4	80	51	33	26
RCMS9	Mar 4, 2004	4.2	65	12	10	10
RCMS10	June 3, 2004	4.2	90	46	46	20
Total Number of Clones				**688**	**240**	**198**

The diversity of SIRV-like viruses decreases in the next sample. The third sample was collected in January 2003, and sixty-six SIRV-like clones were sequenced (table 4.10). The population of SIRV-like viruses has been decreased to one group at this time point (clade A in figure 4.26). The inferred protein-based analysis resembles the DNA-based analysis. The seven protein sequences collected from January 2003 are all closely related to protein sequences from July and October 2002 (represented by the purple group in figure 4.27). There was one protein sequence that was identical to sequences detected from previous samples (figure 4.27; the * represents the group of identical protein sequences).

In the next sample collected in May 2003, the diversity of the SIRV-like viruses in Rabbit Creek increases (figure 4.23). The fourth sample was collected in May 2003 yielded 39 SIRV-like coat protein sequences (table 4.10). When these sequences were

compared to sequences sampled from previous times, the diversification of SIRV-like viruses becomes apparent. SIRV-like sequences were detected that represented clades A, C, D, and E (figure 4.26). The 39 unique DNA sequences translated to 24 unique inferred protein sequences (table 4.10). Like the DNA sequences, the protein sequences diversified in this sample (figure 4.24). There were protein sequences detected in this sample that were identical to sequences detected in previous samples that represented each group of viral protein sequences (figure 4.27; the * represents identical sequences).

Figure 4.25. SIRV-like sequences collected in July 2002 from Rabbit Creek. The numbers in parentheses indicate how many times a particular sequence was detected. A. DNA sequences. B. Inferred protein sequences.



The SIRV-like diversity measured in the previous sample is maintained in the next sample (figure 4.23). The fifth sample resulted in only 21 SIRV-like coat protein clones (table 4.10). The SIRV-like sequences from this sample belong in viral groups that had been previously established (clades A, B, C, and D in figure 4.26). Eighteen unique inferred protein sequences were translated from the 19 unique DNA sequences obtained from this sample (table 4.10). Most of the inferred SIRV-like protein sequences belong in the largest group of protein sequences (represented in purple in figure 4.27). One sequence in this sample was identical to the group of identical sequences detected from each sample collected thus far (figure 4.27; the * represents the identical sequences).

The sixth sample reveals a decrease in the diversity in the SIRV-like population (figure 4.23). The sixth sample was collected in July 2003 (table 4.10). One hundred and forty-five SIRV-like coat protein clones were sequenced from this sample, and 90 clones were identical. When all of the clones from July 2003 were compared to sequences from previous samples, the majority of the July 2003 SIRV-like sequences belong in one major group (clade A in figure 4.26). The 35 unique DNA sequences translated into 23 unique inferred protein sequences (table 4.10). The SIRV-like protein sequences have decreased to represent one major group (represented in purple in figure 4.27). However, one protein sequence belonged to a group detected in July and October 2002 (represented by pink in figure 4.27). Two protein sequences from this sample were identical to protein sequences detected previously (figure 4.27; the * represents the identical sequences).

Figure 4.26. SIRV-like DNA sequences collected from Rabbit Creek from July 2002 to June 2004. The numbers in parentheses indicate how many times a sequence was detected. The colors correspond to figure 4.23.

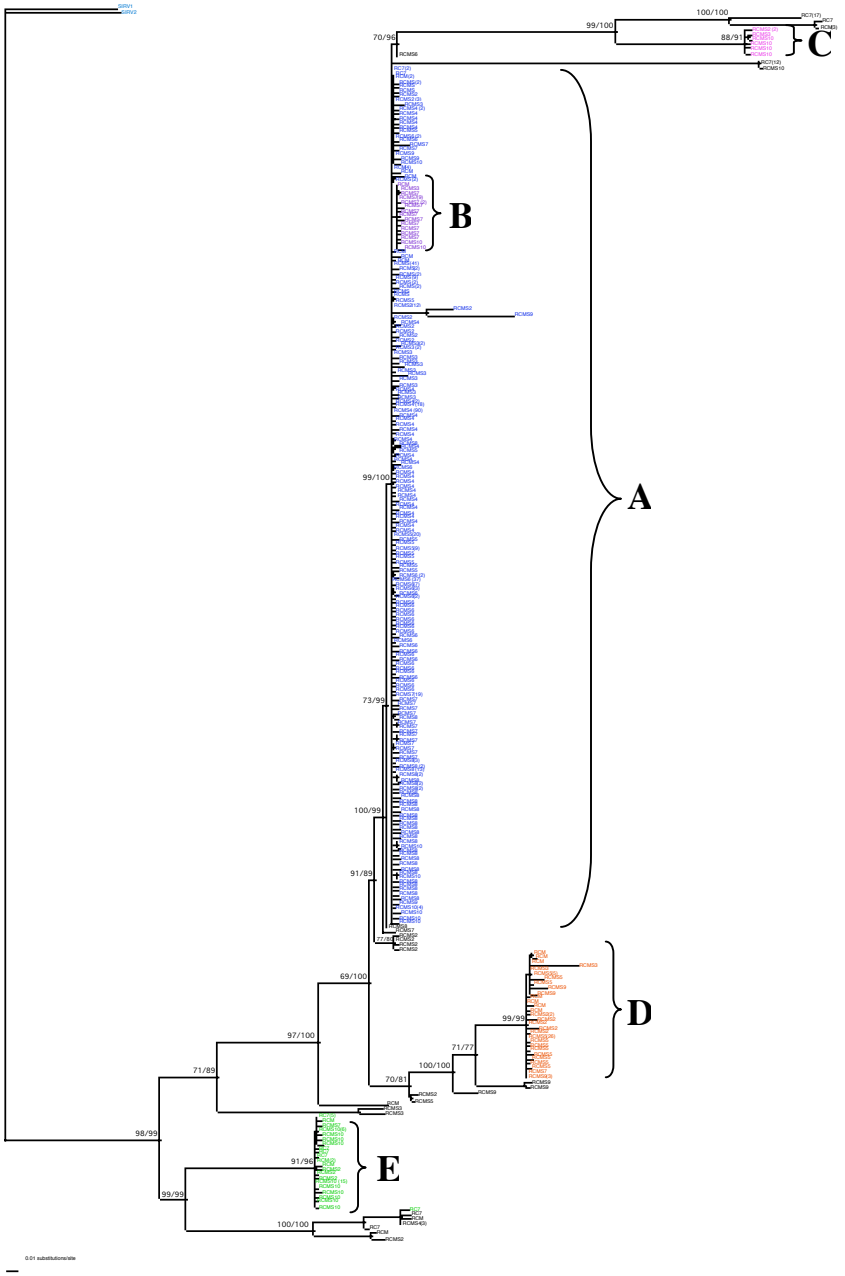
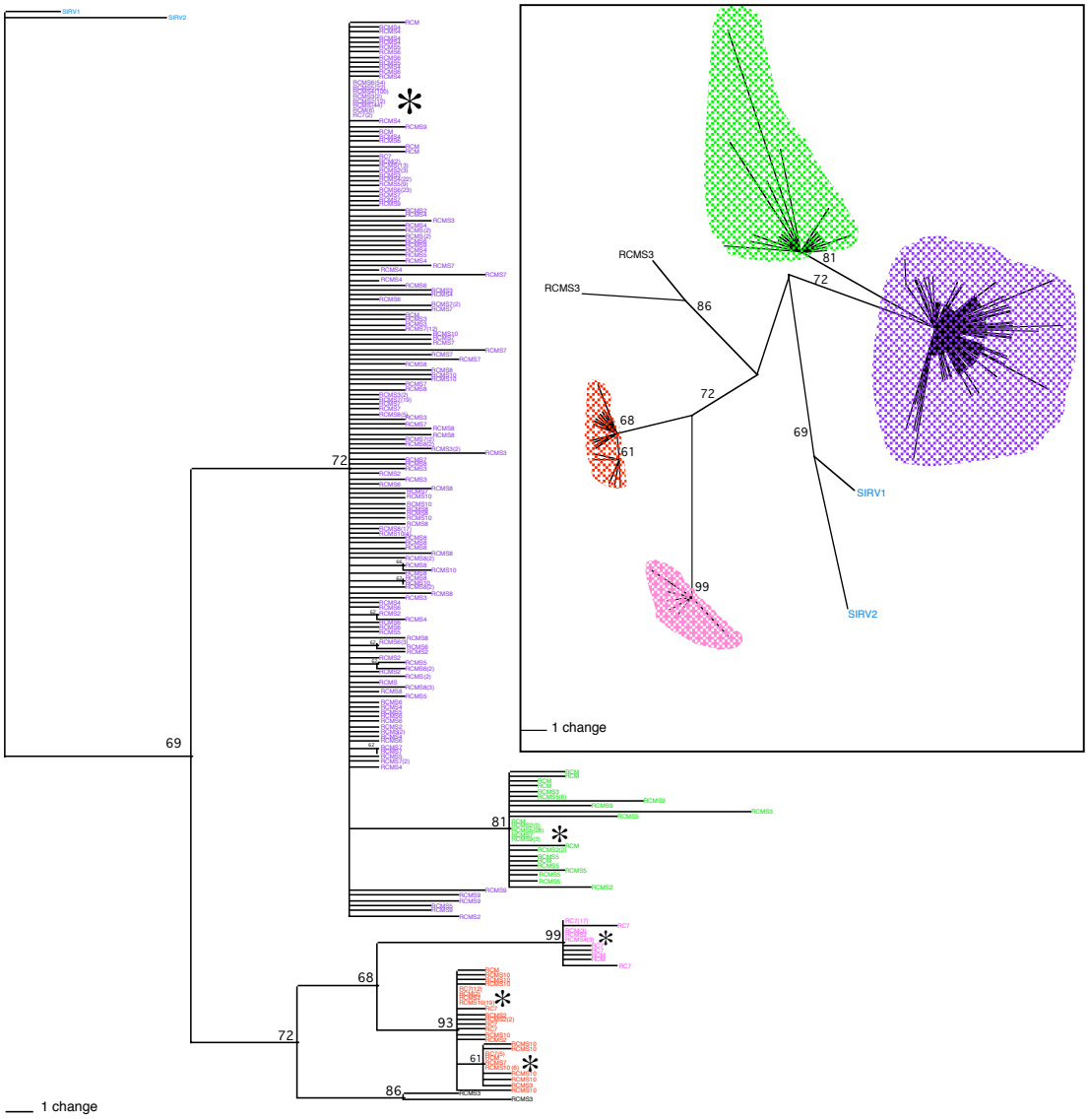


Figure 4.27. SIRV-like protein sequences collected from Rabbit Creek from July 2002 to January 2004. The * represent identical protein sequences. Inset: unrooted phylogram of SIRV-like protein sequences.



In August 2003, the SIRV-like population represents two groups of SIRV-like sequences (figure 4.23). The seventh sample yielded 80 SIRV-like coat protein clones (table 4.10). When these sequences were compared to sequences obtained from earlier samples, the SIRV-like population now represents two groups of viruses (clades A and D

in figure 4.26). The groups present in previous samples are still evident. Half of the SIRV-like sequences (40) from August 2003 are associated with the largest group of SIRV-like sequences (clade A in figure 4.26). The remaining coat protein sequences are associated with a smaller group of SIRV-like sequences (clade D in figure 4.26). Nineteen unique inferred protein sequences were detected from this sample (table 4.10). Most of them belong in the largest group of SIRV-like coat protein sequences (represented by purple in figure 4.27). However, some protein sequences belong in a group detected in earlier samples (represented by green in figure 4.27). Once again, the largest group of identical sequences that has been continually detected in this hot spring was detected in this sample (figure 4.27; the identical sequences are represented by *).

The SIRV-like viral diversity has decreased in October 2003 (figure 4.23). The sample collected on October 2, 2003 was the eighth sample (table 4.10). The diversity of SIRV-like sequences has been decreased to a single group (clade A in figure 4.26). There were 20 unique inferred protein sequences detected in this sample (table 4.10). The decrease in SIRV-like diversity is also evident in the protein sequences (figure 4.24). There were 54 cloned protein sequences detected in this sample that were identical to the largest group of identical SIRV-like sequences (figure 4.27; the * represents the identical sequences).

There is more diversity in the November 13, 2003 sample when compared to the previous sample collected in October 2003 (figure 4.23). The ninth sample contained 57 SIRV-like coat protein clones (table 4.10). This sample shows diversification of the SIRV-like viruses compared to the sample collected in October 2003 (figure 4.23). The

majority of the sequences from this sample (35) belong in the dominant group of SIRV-like viruses (clade A in figure 4.26). The sequences appear to be further sub-divided within this sub-group, revealing more diversity (clade B in figure 4.26). Twenty-five unique inferred protein sequences were obtained from this sample (table 4.10). They represent three groups in the SIRV-like sequences (figure 4.24). Most of the SIRV-like inferred protein sequences belong to the largest group of viruses (represented by purple in figure 4.27). Two of the inferred protein sequences each belong to a different smaller group of viruses; each of these sequences was shown to be identical to sequences previously detected from this monitor site (figure 4.27; the * represents the identical sequences).

The diversity of the SIRV-like population significantly decreases in the next sample collected in January 2004 (figure 4.23). Fifty-one SIRV-like coat protein clones were sequenced from this sample (table 4.10) and compared to sequences from previous samples. The population of SIRV-like viruses has decreased to represent one major group (clade A in figure 4.26). There were 26 unique inferred protein sequences detected in this sample (table 4.10). The inferred protein sequences represent one group of viruses (represented by purple in figure 4.27). There were not any sequences detected in this sample that were identical to sequences detected in previous samples.

The diversity of SIRV-like viruses increases in the next sample (figure 4.23). This sample was collected on March 4, 2004, and only 12 clones were sequenced (table 4.10). The diversity of SIRV-like viruses in this monitor site has increased to represent

two groups of viruses (clades A and D in figure 4.26). The inferred protein sequences also reveal the increase in diversity of SIRV-like viruses (figure 4.24).

The last sample continues to reveal change in the SIRV-like population (figure 4.23). The sample was collected on June 3, 2004 and 46 clones were sequenced (table 4.10). Some SIRV-like sequences belonged to the largest group of viruses (clade A); however, there were also sequences that represented smaller groups of viral sequences (clades B, C, and E) (figure 4.26). The inferred protein sequences also reveal an increasing diversity (figure 4.24). Approximately half of the protein sequences from this sample represent the largest group of protein sequences (the purple group in figure 4.27). The other half of protein sequences from this sample represent a group detected in earlier samples from this monitor site (the red group in figure 4.27).

In summary, the SIRV-like population was detectable in every sample collected from this hot spring (table 4.10). Like the SIRV-like viruses in Crater Hills (see Chapter 3), the diversity of viruses appears to fluctuate between several (3-4) groups to few (1-2) groups of viruses. It is interesting to note that changes in the SIRV-like DNA sequences did not always result in changes in the SIRV-like proteins sequences (figures 4.21 and 4.24). The complete phylogenetic analysis of SIRV-like sequences included analysis of sequences after every sample. For a step-by-step view of the changes in the SIRV-like population over time, see Appendix B.

Detailed analysis of SIRV-like population dynamics. After two years of regularly sampling (July 2002—June 2004), and sequencing SIRV-like coat protein clones from this site, we detected five different groups of SIRV-like viruses (figure 4.26). However,

there were only four groups of SIRV-like viruses based on inferred amino acid sequences (figure 4.27). The viral DNA sequences belonging in clade B translated to protein sequences belonging in the largest group of protein sequences (blue in figure 4.27). The largest group of identical protein sequences (represented by * in figure 4.27) was only detected in early samples from this monitor site (July 2002—October 2003). There were four other groups of identical sequences detected in this monitor site over time; all of these sequences belonged to the smaller groups of viral protein sequences (figure 4.27; the identical sequences are represented by **).

Over the two years of sampling, one large group of SIRV-like viruses was most commonly detected (clade A; represented by blue in figure 4.26), four smaller groups were detected (clades B, C, D, and E), and finally, some groups containing four or fewer sequences were detected (figure 4.26). The environmental population of SIRV-like viruses is similar to that of Crater Hills (see Chapter 3) and Ragged Hills (see Chapter 5), which are comprised largely of one group of sequences and numerous smaller groups of sequences. However, the SIRV-like population is not as diverse as the SIRV-like population in either Crater Hills (see Chapter 3) or Ragged Hills (see Chapter 5).

Using Modeltest version 3.6 (Posada *et. al.* 1998), the nucleotide substitution model for maximum likelihood estimations was determined to be K81uf+G (Kimura 1981) with heterogeneity between sites (gamma=6.3662), base frequencies A=0.3188, C=0.2040, G=0.1945, and T=0.2826, and a substitution model A-C=G-T=1.0000, A-G=C-T=3.7182, and A-T=C-G=1.5007.

There were 688 SIRV-like coat protein clones sequenced during the course of two years (July 2002—June 2004). These sequences composed 239 groups of identical sequences. The largest group of SIRV-like coat protein sequences contained 207 identical sequences, which correlates to 30% of the total SIRV-like population sampled during this research (table 4.11). This group consists of sequences collected in July 2002 to October 2003, but in recent samples, it was no longer detected.

Table 4.11. Rabbit Creek SIRV-like sequences illustrating the groups of identical sequences, how many times a group this size was detected, the percent of the SIRV-like population, and what samples the group contained.

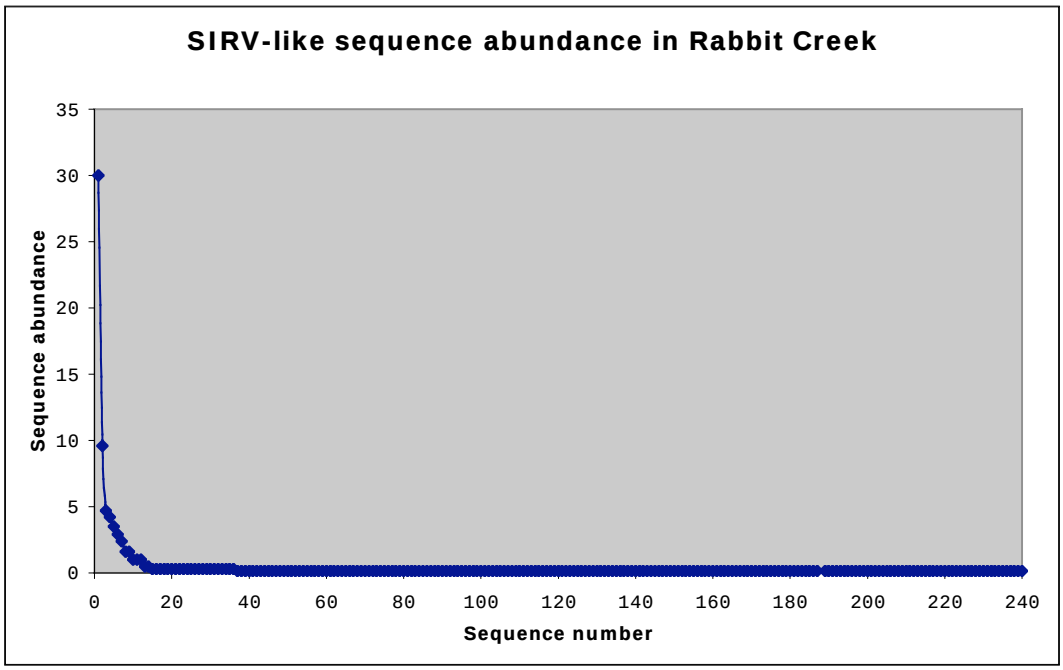
RABBIT CREEK SIRV-like SEQUENCE ABUNDANCE			
Identical Sequences	Times Detected	Percent of SIRV-like Population	Samples[a]
207	1	30%	RC7;RCM;RCMS—RCMS6
66	1	9.6	RC7;RCMS—RCMS6;RCMS9
32	1	4.7	RCM;RCMS2;RCMS5;RCMS7;RCMS9
29	1	4.2	RC7;RCM;RCMS10
24	1	3.5	RCMS3;RCMS7—RCMS8
20	1	2.9	RC7;RCMS4
17	1	2.4	RCMS8;RCMS10
11	2 unique[b]	1.6% each[c]	
7	3 unique	1% each	
3	2 unique	0.47% each	
2	22 unique	0.29% each	
1	203 unique	0.15% each	

[a]sequences from these samples were present in the indicated groups
[b]each of the smaller groups were detected this many times, however, they are unique
[c]each of the unique groups makes up this percentage of the population

Another group of identical viral sequences that was present during the majority of the sampling time was composed of 66 sequences, which consisted of 9.6% of the total

population (table 4.11). This group consisted of samples from the middle of the research; however, some samples from later in the research were also in this group (table 4.11). There was a group containing 32 identical sequences of SIRV-like coat proteins (table 4.11), and it comprised 4.7% of the total population of SIRV-like viruses. Another group of 24 identical sequences was detected in some samples representing 3.5% of our sampled SIRV-like population (table 4.11). Other groups of identical sequences ranged from two to 20 identical sequences comprising 0.29% to 2.9% of the total population (table 4.11). The unique sequences comprised 29.5% of the total population of SIRV-like coat protein sequences (figure 4.28).

Figure 4.28. A sequence abundance plot for SIRV-like sequences from Rabbit Creek. The left end of the x-axis represents most commonly detected sequences, and the right end of the x-axis represents the most rare sequences. Most of the SIRV-like sequences in Rabbit Creek are detected only once.



Over the course of sampling, there were 198 unique SIRV-like protein sequences (table 4.10). Seven protein sequences were detected more than once. However, the largest group of identical sequences was only detected in the July 2002—October 2003 samples (figure 4.27; the * represents the largest group of identical protein sequences).

Using MEGA version 3 (Kumar *et. al.* 2004), changes in the SIRV-like sequences were calculated (tables 4.12 and 4.13). When comparing all of the sequences, 89% of the bases in the SIRV coat protein sequence remained unchanged (table 4.13). Transversions occurred more (24) compared to transitions (19). Nine changes occurred in the first position of the codon; there were four transitions and five transversions (table 4.13). Only seven changes occurred in the second position with two transitions and five transversions (table 4.13). A total of 26 changes occurred in the third position; there were 13 transitions and 13 transversions (table 4.13).

Table 4.12. Numbers of transitions and transversions in SIRV-like sequences in each sample collected from the Rabbit Creek monitor site.

SAMPLE DATE	**IDENTICAL BASES**	**TRANSITIONS**			**TRANSVERSIONS**		
		1st	2nd	3rd	1st	2nd	3rd
07/26/2002	345/405	6	1	20	6	5	23
10/17/2002	342/405	6	2	22	6	6	21
01/20/2003	403/405	1	0	0	0	1	0
05/30/2003	349/405	6	2	20	5	5	18
06/24/2003	365/405	4	2	13	4	4	13
07/25/2003	395/405	1	1	3	1	1	3
08/26/2003	374/405	2	2	10	2	4	11
10/02/2003	403/405	0	1	0	0	1	0
11/13/2003	390/405	2	1	5	2	2	4
01/19/2004	402/405	1	1	0	1	0	0
03/04/2004	369/405	3	2	13	2	3	12
06/03/2004	357/405	5	2	17	5	4	16

Table 4.13. The numbers of transitions and transversions for all SIRV-like sequences detected over time in Rabbit Creek.

Sample Range	Position	Identical Base Pairs	Ts Pairs	Tv Pairs
July 2002-June 2004	*Overall*	*362/405*	*19*	*24*
	1st	126/135	4	5
	2nd	127/135	2	5
	3rd	109/135	13	13

Ts=transitions; Tv=transversions

Comparing SSV-like and SIRV-like population dynamics

Unlike the SSV-like population, we were able to detect SIRV-like sequences at every sample collected over the two years of this research. We sequenced 688 SIRV-like coat protein clones (table 4.10) compared to 320 SSV-like clones (208 3F/4R clones and 112 7F/8R clones; table 4.3). It appears that SIRV-like viruses are more permanent residents in this hot spring. The SSV-like population consisted of three groups of viruses based on 3F/4R sequences (figure 4.11). The SIRV-like population was much more diverse than the SSV-like population (figure 4.23). There was one consistency between the two viral populations; the changes in the diversity of each virus population appeared to fluctuate. The increased diversity of SIRV-like viruses is not likely a result of the increased number of sequenced SIRV-like clones. The overall population structures of the SSV-like viruses in each monitor site were very similar. It is likely if a larger number of SSV-like clones were sequenced, the same population structure would be evident.

There is no obvious correlation between viral populations and environmental changes. The changes occurring in the SSV-like populations do not seem to correlate with an environmental change. Some changes in the SIRV-like population may be attributed to a change in the environment. In October 2002, two new groups of SIRV-

like DNA sequences were detected (figure 4.23). The pH had decreased by one unit between the samples collected in July 2002 and the sample from October 2002 (table 4.10). Perhaps the decrease in pH can be correlated to the formation of a new group of SIRV-like viral sequences. The SIRV-like population was diverse until a decrease in diversity occurred in the January 2003 sample (figure 4.23). The pH in October 2002 was measured at 2.8, and the pH in January 2003 was 5.1 (table 4.10). Perhaps this increase of several units of pH caused the decrease of SIRV-like viral diversity. These examples illustrate that perhaps some changes are environmentally linked. However, there are other examples that show changes in the virus population do not relate with environmental change. Therefore, it is not known whether environmental conditions are significant enough to change a virus population.

Conclusions

The Rabbit Creek monitor site proved to be a very dynamic thermal area. This thermal feature was chosen as part of this research for two reasons. The first reason is the ability to culture *Sulfolobus* and detect viruses from this particular hot spring, and the second reason being the age of this site. The other monitor sites chosen in this research are thought to be younger in comparison. This site was sampled continuously for approximately two years. During that time, the water chemistry was monitored, the dominant archaeal population was monitored, and changes in the population structures of two viruses were continually investigated.

The water chemistry was characterized by high iron and manganese concentrations as compared to the other two monitor sites (figures 4.4 and 4.5). Surprisingly, the water chemistry did not significantly change over the monitoring period. Most of the observed changes in water chemistry could be contributed to seasonal fluctuations in water levels affecting the concentration and possibly the precipitation of metals. The only exception was the sample collected in July 2003 where most of the metals and nutrients showed a dramatic increase in concentration (table 4.1 and figure 4.3). This appeared to be an isolated incident, as a similar spike did not occur in other samples. This is an example of how unstable and unpredictable the thermal areas are in YNP. This monitor site is a vapor-dominated system without visible inflow or outflow channels. The pH and temperature of this monitor site were not as consistent as the other two monitor sites (see Chapters 3 and 5). The pH range was measured between 2.8 and 5.1, and the temperature was measured at 64-91°C.

By analyzing the water chemistry at this hot spring, we learned it is a unique system dominated by high iron and low metals compared to the other two monitor sites. These results suggest that this is an older geothermal feature compared to Crater Hills (see Chapter 3) and Ragged Hills (see Chapter 5).

The environmental libraries from this site contained a diverse archaeal population dominated by *Acidianus*-like and *Vulcanisaeta*-like sequences. Most of the full-length (~1,400bps) and almost full-length (~1,000bps) 16S rDNA sequences showed high identity to published 16S rDNA sequences (table 4.2). All of the 16S rDNA sequences matched published 16S rDNA sequences at 86-99% identity (table 4.2). The archaeal

population detected in environmental libraries from this monitor site was not as diverse as the Crater Hills monitor site (see Chapter 3), but it was more diverse than the Ragged Hills monitor site (see Chapter 5).

The 16S rDNA sequences most commonly detected in this hot spring over time were *Acidianus*-like and *Vulcanisaeta*-like sequences. *Acidianus* species are chemolithotrophic and S^0 dependent by oxidation or reduction of S^0. Cells grow at temperatures ranging from 65 to 96°C, with optimal growth occurring around 90°C. The pH suitable for growth ranges from 1.0 to 5.5, with optimal growth at pH 2 (Segerer *et. al*. 1986). *Acidianus* species are abundant in thermal areas around the world (Segerer *et. al*. 1986; Rachel *et. al*. 2002). As mentioned in Chapter 3, *Vulcanisaeta* grows optimally at 85-90°C and pH 4.0-4.5 (Itoh *et. al*. 2002). Based on the conditions in Rabbit Creek, it is reasonable to predict these organisms or similar organisms could survive in this monitor site.

The detected archaeal population was dependant on primer usage and the sampling date (table 4.2). The host population changed drastically in the 10 months between each sample date; however, the use of two different primers may contribute to the difference in populations. The growth and pH ranges for archaeal organisms of interest are consistent with the temperatures and pHs measured on January 20, 2003 and October 2, 2003 (table 4.2). However, more 16S rDNA sequences would need to be analyzed to further characterize the archaeal population. Current research is aiming towards culturing *Vulcanisaeta* and *Acidianus*.

Based on 16S rDNA sequences from this monitor site, *Sulfolobus* is not a dominant organism in this hot spring. In fact, only one of the samples contained clones resembling *Sulfolobus* (table 4.2). This clone had 95% identity to the published *S. islandicus* 16S rDNA sequence (table 4.2). This was the only sequence detected from this monitor site that was similar to a *Sulfolobus* sequence previously detected within YNP (Whitaker *et. al*. 2003). Based on results from previous studies, we assumed that *Sulfolobus* and *Sulfolobus*-like organisms would dominate these systems. But like the Crater Hills monitor site (see Chapter 3), *Sulfolobus* species do not appear to be the dominant archaeal organism in the Rabbit Creek monitor site.

The results concerning the host population lead to serious implications concerning the virus population. While easily detecting *Sulfolobus* viruses in the environment, we are unable to detect large numbers of *Sulfolobus* species based on 16S rDNA sequence. There are many potential reasons for this apparent anomaly. First, the virus sequences may be easily detected because they are present in a higher copy number than the *Sulfolobus* 16S rDNA gene. Or, perhaps SSVs and SIRVs have a broader host range than just *Sulfolobus*. However, we believe the viral sequences we amplified infect *Sulfolobus*. Each of the monitor sites consisted of its own unique archaeal population. Regardless, we were able to amplify both SSVs and SIRVs consistently from each monitor site. Furthermore, *Sulfolobus*-like 16S rDNA sequences were the only 16S rDNA sequences detected in all three monitor sites.

Both virus populations (SSVs and SIRVs) that were continuously monitored were constantly undergoing change. SSV-like sequences were only detected in seven samples

using the 3F/4R primer set, and only two samples using the 7F/8R primer set. The fact that primer set 3F/4R yielded more dependable results is consistent with the results in the other two monitor sites (see Chapters 3 and 5). A plausible explanation is the detection limit of the 3F/4R primers was 10,000X greater on purified viral template (0.1pg) than the detection limit of the 7F/8R primer set (1ng). We can say, with confidence, the SSV-like sequences are not from host organisms. The sequences have been compared to numerous organisms that may be found within these environments and there are no matches detected.

The entire SSV-like population was undetectable from May 2003 to August 2003 (table 4.3), and from March and June 2004 (table 4.3). There is not an obvious explanation as to why the SSV-like population was undetectable in these samples especially because SIRV-like sequences were detected in each sample (table 4.10). The Rabbit Creek monitor site was the most dynamic of the three monitor sites as far as highs and lows in pH and temperature. Perhaps because of this, the SSV-like viruses could not establish a permanent, stable population. We do not attribute the absences of SSV-like sequences to experimental errors. We were able to amplify SIRV-like sequences from every sample collected from this monitor site. This indicates the source DNA is suitable for PCR and it was not experimental technique. Perhaps the SSV-like viruses are present in the environment at a lower abundance that SIRV-like viruses.

As opposed to the SSV-like sequences, the SIRV-like sequences were detected in every sample collected from this monitor site (table 4.10). The primers designed to amplify the entire coat protein (405bps and 135aa) of SIRV reliably detected this gene.

SIRVs have not been shown to integrate into their host cell's chromosome; instead they are maintained within the cell as a stable carrier state (Prangishvili *et. al.* 1999). Over the course of this study, there were not any interrupted ORFs detected among the SIRV-like coat protein sequences obtained from this hot spring. This helps to support our claim that the interrupted SSV-like ORFs may be from integrated copies of the virus (as discussed above), instead of sequencing or PCR errors. If the interrupted SSV-like ORFs were a result of PCR or sequencing errors, we would expect to see the same errors occurring in the SIRV-like sequences.

The SIRV-like population was consistently detected from this monitor site (table 4.10). The diversity in the SIRV-like population is detected immediately in the first two samples (figure 4.23). The third sampling point reveals a decrease in the diversity of SIRVs from numerous groups to only one (figure 4.23). The next sample reveals an explosion of diversity in the SIRV-like population (figure 4.23). This diversity is maintained through the next two samples (figures 4.23); the diversity is reduced to a single group in the sixth sample from this monitor site (figure 4.23). The diversity decreases in the seventh sample (figure 4.23). The next samples fluctuate between numerous groups of SIRV-like viruses and one group of viruses (figure 4.23).

There were numerous groups of SIRV-like viruses detected over time within this monitor site. The SIRV-like viruses were composed of four distinct groups based on protein sequences (figure 4.24). There is one large group of SIRV-like viruses (colored purple in figure 4.27) and numerous smaller groups (colored pink, red, and green in figure 4.27).

In conclusion, we discovered the viral populations within this hot spring are dynamic and diverse. The SSV-like population was only detected sporadically through January 2004. The SSV-like population changed and showed diversity within the samples that contained SSV-like sequences. The SIRV-like population was detected in all of the samples and was shown to be very diverse. Even though there may be questions concerning the hosts of these viruses in nature, we were able to show the viral populations do change over time. It will be interesting to determine how the virus populations change in the youngest of our study sites, Ragged Hills.

CHAPTER 5

RAGGED HILLS MONITOR SITE: VIRUS DYNAMICS, HOST POPULATIONS, AND WATER CHEMISTRY

Introduction

A thermal acidic feature within the Ragged Hills area was selected as a monitor site. The reasons for selecting this particular site included the following: the monitor site is a new thermal feature, it is both thermal and acidic, the site is large and easy to access year round, and we are successful in culturing *Sulfolobus* and its viruses from this hot spring. Over the course of two years, we monitored an acidic hot spring within the Ragged Hills thermal area in Yellowstone National Park (YNP) for changes in the viral and host populations, and water chemistry. The goal of this research is to determine if there are significant changes in viral population structures over time. The hypothesis we are testing is whether there is a correlation between virus changes and observed changes in the host populations and water chemistry of an individual thermal feature.

The Ragged Hills area is one of the most geothermally active areas within YNP. The Ragged Hills area is a new geothermal feature that developed in June 1999. Because its relative young age, examination of this site affords a unique opportunity to examine viral population dynamics in a new thermal feature as compared to older thermal sites (see Chapters 3 and 4).

SSV-like and SIRV-like viruses were chosen to monitor during this study. These viruses were selected because (i) in an initial screening, both SSV-like and SIRV-like

virus particles were commonly observed in primary enrichment cultures established from this region, (ii) *Sulfolobus* species could be routinely cultured, and (iii) genomic sequence data was available for each virus type from a diversity of geographically distributed thermal features worldwide (Wiedenheft *et. al*. 2004; Peng *et. al*. 2001; Prangishvili *et. al*. 1999; M. Young lab, unpublished).

Materials and Methods

Site location

The monitor site is located across the northwest ridge of a small saddle connecting the Ragged Hills area of Norris Geyser Basin to another small hill to the north (44° 43.653′N and 110° 42.862′W).

pH and temperature monitoring

The temperature of the Ragged Hills monitor site was continuously monitored with a data logger. The data logger recorded the measured temperature approximately every 20 minutes. The pH of the monitor site was measured each time a sample was collected.

Sample collection and water chemistry

Hot spring water samples were collected extracted as previously described (see Chapter 3). The water chemistry samples were collected and analyzed as previously described (see Chapter 3).

Total nucleic acid extraction
from environmental samples

Total nucleic acid from the Ragged Hills monitor site was extracted using the UltraClean™ Water DNA Extraction Kit (Mo-Bio, Solana Beach, CA) per the manufacturer's protocol (for a brief outline, see Chapter 3). The integrity and yield of the extracted DNA was estimated by display on agarose gels.

Primer design and PCR for viruses

SSV-like sequences were PCR amplified from total environmental nucleic acid extracts. Two "universal" primer sets were chosen to amplify different regions of SSVs (figures 3.1 and 3.2). The primer set termed UnvSSV#3F and UnvSSV#4R amplifies ~250-260 base pairs (bps) in the largest open reading frame (ORF) of each sequenced isolate (figure 3.1 and table 3.1). The primer set termed UnvSSV#7F and UnvSSV#8R amplifies the coat proteins of SSVs (figure 3.2). The amplicon ranges in size from 682 bps to 1,069 bps (table 3.1) depending on the presence of the VP2 gene. The PCR protocol utilized a standard reaction mixture of 50μL containing 1pg-10ng total environmental nucleic acid. The reaction conditions were described previously (see Chapter 3). PCR products were analyzed by agarose gel electrophoresis.

SIRV-like sequences were also amplified from the total nucleic acids extracted from the environment. Using the DBPF/R primer set (figure 3.4), the PCR reaction and cycling conditions were previously described (see Chapter 3). PCR products were analyzed by agarose gel electrophoresis.

Archaeal 16S rDNA primers and PCR

16S rDNA gene sequence analysis was used to determine the identity of the potential archaeal host populations in the Ragged Hills monitor site. Two primer sets (20F/UA1406R and 348Fa/UA1406R) (Baker *et. al.* 2003) were chosen to amplify the 16S rDNA gene from the total nucleic acid extracted. The forward primers (20F and 348Fa) are archaeal specific and the reverse primer (UA1406R) is universal. The reaction parameters were identical to the PCR conditions previously mentioned (see Chapter 3). The primer set 20F/UA1406R should yield a product of 1,400 bps, with 348Fa/UA1406R primer set yielding a 1,000 bps product. The amplicons were analyzed by agarose gel electrophoresis.

Cloning PCR products

Viral and 16S rDNA PCR products were cloned via TOPO-TA cloning into the pCR2.1 vector using the standard protocol (Invitrogen, Carlsbad, CA) as previously described (see Chapter 3). The transformation of the plasmids and plasmid purification was previously described (see Chapter 3).

DNA sequencing

An additional screening of the plasmid preps was done as previously described (see Chapter 3). The sequencing reaction and conditions were identical to previously described methods (see Chapter 3).

Sequence and phylogenetic analysis

Viral sequences. Environmental SSV-like and SIRV-like sequences were edited, and aligned as previously described (see Chapter 3). The sequences were exported to ClustalX (Thompson *et. al.* 1997) for an alignment, and a maximum parsimony analysis was conducted using PAUP 4.0b10 (Swofford 2002) as previously described (see Chapter 3). A maximum likelihood analysis was conducted using MrBayes version 3.0 (Huelsenbeck and Ronquist 2001) as previously described (see Chapter 3).

Viral DNA sequences were translated, aligned, and a maximum parsimony analysis conducted as previously described (see Chapter 3). The environmental SSV-like 3F/4R protein sequences were 85 to 87 amino acids from the largest ORF in the SSVs (812 total amino acids predicted in the SSV-RH ORF). When the SSV-like DNA sequences were translated to protein, some sequences resulted in a truncated protein sequence. The translated DNA sequences that resulted in an interrupted ORF (those with stop codons throughout) are termed non-functional ORFs. DNA sequences that resulted in un-interrupted ORFs (no stop codons) are termed functional ORFs. The distinction between functional and non-functional ORFs was used to determine which viral sequences might have come from integrated copies of the virus. Interrupted SSV-like ORFs were not used in the protein analysis, unless otherwise noted. The environmental SIRV-like coat protein sequences were 135 amino acids. The molecular evolutionary analyses (rates of transitions and transversions) for both viruses were conducted using MEGA version 3 (Kumar *et. al.* 2004).

Phylogenetic trees were constructed sequentially in order to compare the new viral sequences to previously detected viral sequences. In all of the phylogenetic trees, sequences from the sample being discussed will be presented in red, sequences from previous samples will be presented in black, and the sequences from the completely sequenced isolates (SSV1, SSV2, SSV-RH, and SSV-K1 for 3F/4R and 7F/8R primer sets; SIRV1 and SIRV2 for the DBPF/R primer set) will be represented by blue. The black sequences are only presented as a reference to make it visually easier to determine how or if the viral population has changed since the last sample was collected. A virus group/clade will be defined as a group of monophyletic viral sequences supported by high bootstrap and Bayesian clade credibility values.

In the DNA analysis, numbers next to principle branches show bootstrap support >50% using maximum parsimony and Bayesian clade credibilities using maximum likelihood. In the protein analysis, the number at the principle branches show bootstrap support >50% using maximum parsimony.

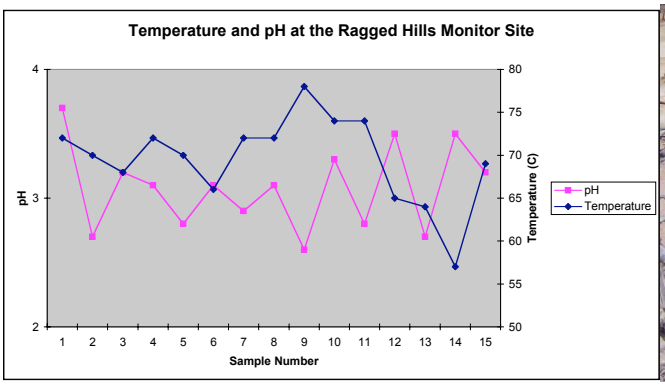
Archaeal host sequences. 16S rDNA sequences were similarly edited to remove vector sequence. The environmental sequences were compared to published 16S rDNA sequences (http://www.ncbi.nlm.nih.gov/BLAST/). Environmental sequences were analyzed as previously described (see Chapter 3).

Results

Temperature, pH, and water chemistry in the Ragged Hills monitor site

We sampled the Ragged Hills monitor site fifteen times from October 2002 through June 2004. During this time, the monitor site ranged in temperature from 57 to 78°C and the pH ranged from 2.6 to 3.7 (figure 5.1). In general, this monitor site has been cooling off during this study. The water chemistry of this monitor was unique in comparison to the other monitor sites, but stable over time. Most of the chemical changes occurring in this monitor site can be attributed to seasonal fluctuations. This monitor site is the youngest of the three study sites, and that is apparent in the water chemistry.

Figure 5.1. Temperature and pH over time from Ragged Hills monitor site (left) and a picture of the monitor site (right). The sample numbers correspond to the sampling dates in table 5.1.





Water chemistry analysis started February 17, 2003 and continued through June 3, 2004 for a total of 12 samples (table 5.1). Of the metals tested, several were not detected (or detected sporadically at very low levels) (barium, calcium, copper, lead, magnesium,

mercury, nickel, and selenium) (table 5.1). This monitor site is a chloride-sulfate dominated system and we observed the chloride levels remained high (569-668mg/L) during this research (figure 5.2a) with the average chloride concentration at 611mg/L. The average sulfate concentration was 135mg/L with a range of 109-187mg/L (figure 5.2a). The dissolved organic carbon (DOC) concentration varied from 0.3-4.9mg/L with an average concentration of 2.5mg/L (table 5.1).

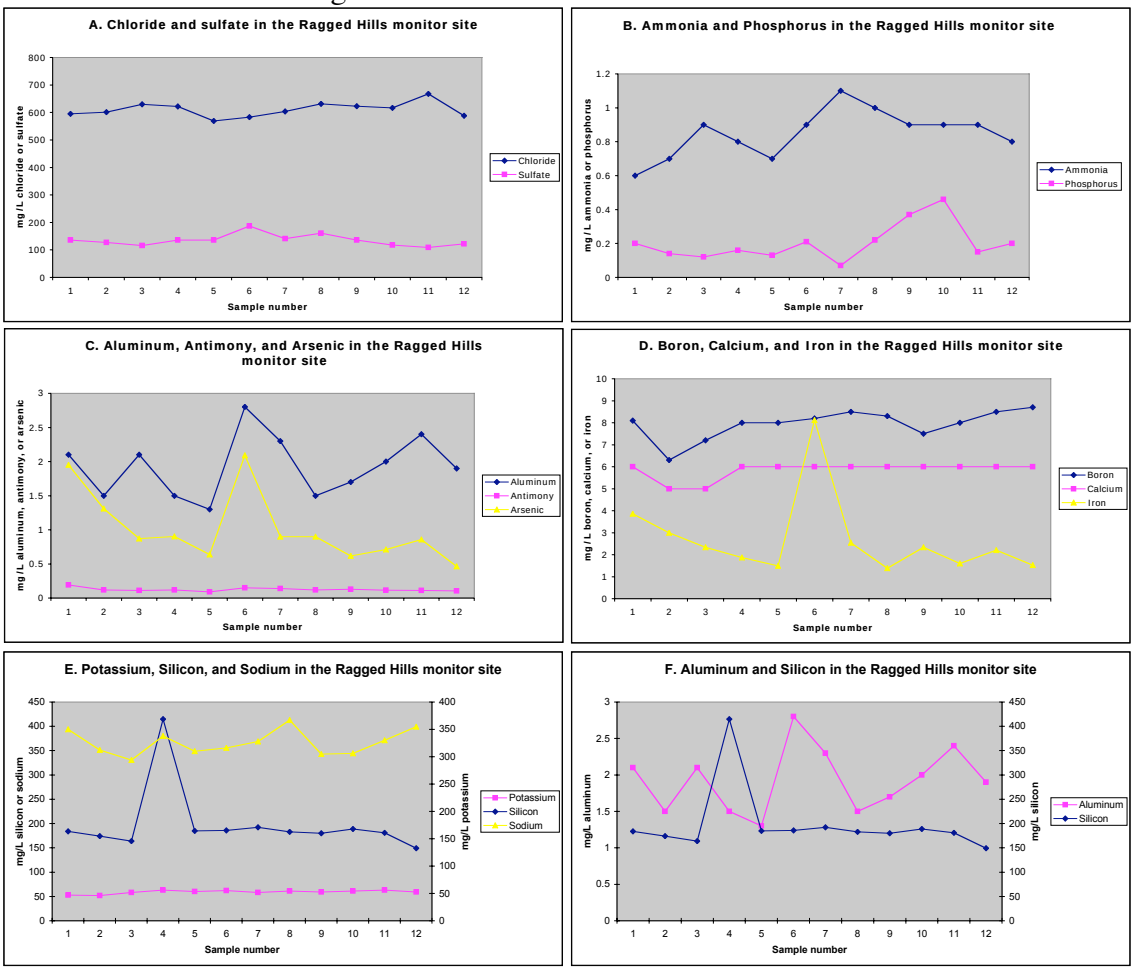
Table 5.1. Water chemistry values measured at the Ragged Hills monitor site from February 2003 through June 2004.

Sample numbers	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Sample dates	*02/03*	*04/03*	*05/03*	*05/03*	*06/03*	*07/03*	*08/03*	*10/03*	*11/03*	*01/04*	*04/04*	*06/04*
Chloride	595	601	630	622	569	583	604	631	623	617	668	588
Sulfate	136	127	116	136	136	187	141	161	136	118	109	122
DOC	0.7	1.1	1.3	2.7	4.7	4.2	3.9	2.5	4.9	3.3	0.3	0.5
Ammonia	0.6	0.7	0.9	0.8	0.7	0.9	1.1	1	0.9	0.9	0.9	0.8
Nitrogen	0	0	0	0	0	0	0.08	0	0	0	0	0
Phosphorus	0.2	0.14	0.12	0.16	0.13	0.21	0.07	0.22	0.37	0.46	0.15	0.2
Aluminum	2.1	1.5	2.1	1.5	1.3	2.8	2.3	1.5	1.7	2	2.4	1.9
Antimony	0.19	0.12	0.11	0.12	0.09	0.15	0.14	0.12	0.13	0.116	0.112	0.106
Arsenic	1.95	1.31	0.871	0.904	0.638	2.09	0.899	0.899	0.615	0.712	0.857	0.464
Barium	0	0	0	0	0	0	0	0	0	0	0	0
Boron	8.1	6.3	7.2	8	8	8.2	8.5	8.3	7.5	8	8.5	8.7
Cadmium	0	0	0	0	0	0	0.001	0	0	0	0	0
Calcium	6	5	5	6	6	6	6	6	6	6	6	6
Copper	0	0	0	0	0	0	0	0	0	0.01	0	0
Iron	3.86	2.99	2.34	1.87	1.49	8.12	2.54	1.38	2.34	1.6	2.21	1.52
Lead	0	0	0	0	0.01	0	0	0	0	0	0	0
Magnesium	0	0	0	0	0	0	0	0	0	0	0	0
Manganese	0.04	0.03	0.03	0.05	0.03	0.04	0.03	0.03	0.03	0.05	0.04	0.04
Mercury	0	0	0	0	0	0	0	0	0	0	0	0
Molybdenum	0.332	0.184	0.141	0.14	0.144	0.297	0.241	0.175	0.18	0.131	0.145	0.112
Nickel	0	0	0	0	0	0	0	0	0	0	0	0
Potassium	53	52	58	63	60	62	58	61	59	61	63	59
Selenium	0	0	0	0	0	0	0	0	0	0	0	0
Silicon	184	174	164	415	185	186	192	183	180	189	181	149
Sodium	350	312	294	338	310	316	328	367	305	306	330	355
Zinc	0.03	0.01	0.03	0.04	0.1	0.08	0.03	0	0.03	0.12	0.03	0.01

*Sample numbers correlate with numbers on corresponding graphs.
*Chemical species measured in mg/L.

The nutrients assayed were ammonia, nitrogen, and phosphorus (table 5.1). Ammonia levels remained steady throughout the study with an average ammonia concentration of 0.85mg/L over a range 0.6-1.1mg/L (figure 5.2b). Nitrogen, expressed as nitrate and nitrite, was detected once at a low level of 0.08mg/L (table 5.1). Phosphorus concentrations at this monitor site were consistently lower in comparison to the other two sites (figure 5.2b) with the average phosphorus concentration of 0.20mg/L.

Figure 5.2. Select water chemistry values detected in Ragged Hills monitor site over time. The sample numbers correspond to the sampling dates listed in table 5.1. All values are measured in mg/L.



The metals detected consistently over time were aluminum, antimony, arsenic, boron, iron, manganese, molybdenum, potassium, silicon, sodium, and zinc (table 5.1). Many of these metals were in the highest concentration as compared to the other monitor sites. The highest concentrations of arsenic were detected, ranging from 0.46-2.1mg/L (figure 5.2c) with an average of 1.0mg/L. Boron and calcium were detected at the highest levels (table 5.1). Boron was constant in this hot spring with an average concentration of 7.9mg/L and a range between 6.3-8.7mg/L (figure 5.2d). The calcium concentrations were very constant (6mg/L) with the exception of the April 2003 and May 6, 2003 samples where the calcium dropped by 1mg/L (figure 5.2d).

Silicon remained constant (figure 5.2e) with the exclusion of the May 13, 2003 sample. The silicon concentrations ranged from 149 to 415mg/L, with an average of 198.5mg/L. Sodium was also consistently detected (figure 5.2e) ranging from 294-367mg/L.

The average aluminum concentration was 1.9mg/L with a range of 1.3-2.8mg/L (figure 5.2c). This was the only monitor site in which antimony and molybdenum were detected (table 5.1). The antimony levels ranged from 0.10-0.19mg/L with an average of 0.12mg/L (figure 5.2c), and molybdenum ranged from 0.13-0.33mg/L with an average of 0.18mg/L (table 5.1). Iron in Ragged Hills was maintained at fairly low levels except for the July 2003 sample (figure 5.2d). In July 2003, the iron concentrations approach the concentrations detected in the Rabbit Creek monitor site (see Chapter 4); however, the increased iron is only detected in July 2003 (figure 5.2d).

Similar to manganese detected in Crater Hills and Rabbit Creek, manganese was consistently detected in the samples albeit very low concentrations (table 5.1). Potassium was constantly detected at much higher levels than the other two monitor sites (figure 5.2e), ranging in concentration from 52 to 63mg/L with an average potassium concentration of 59mg/L. The last metal assayed was zinc, and it was consistently detected at levels of 0-0.1mg/L (table 5.1).

In general, the chemistry of this monitor site was distinct from the chemistry of the Crater Hills (see Chapter 3) and Rabbit Creek (see Chapter 4) monitor sites. Surprisingly, the water chemistry did not significantly change over time. Similar to the other monitor sites, this site is a vapor-dominated system without any visible outflow channels. This site is dominated by high chloride concentrations (figure 5.3 and high arsenic concentrations (figure 5.3). This site was characterized by the lowest observed temperature values of the three monitor sites throughout the sampling time (57-78°C). Measured pH values were 2.8-3.7 (figure 5.1). The results from the water chemistry analysis reveals, that indeed, this is a young geothermal site. An overview of the average water chemistry values over time is illustrated in figure 5.4.

Figure 5.3. The chloride and arsenic levels in the three monitor sites over time. The sample number corresponds to the sampling date in table 5.1.

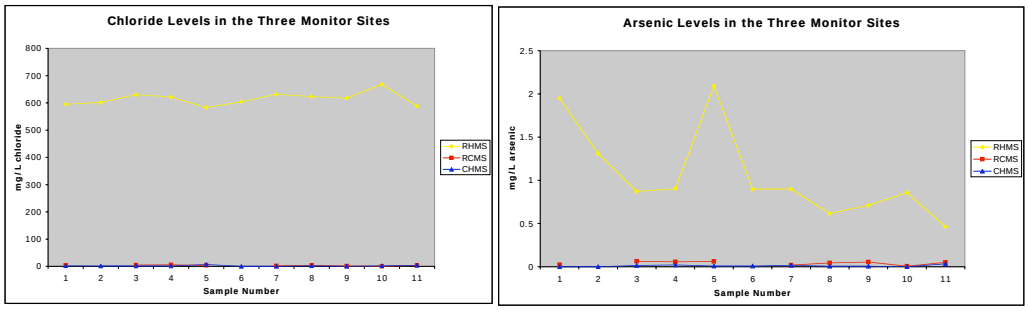
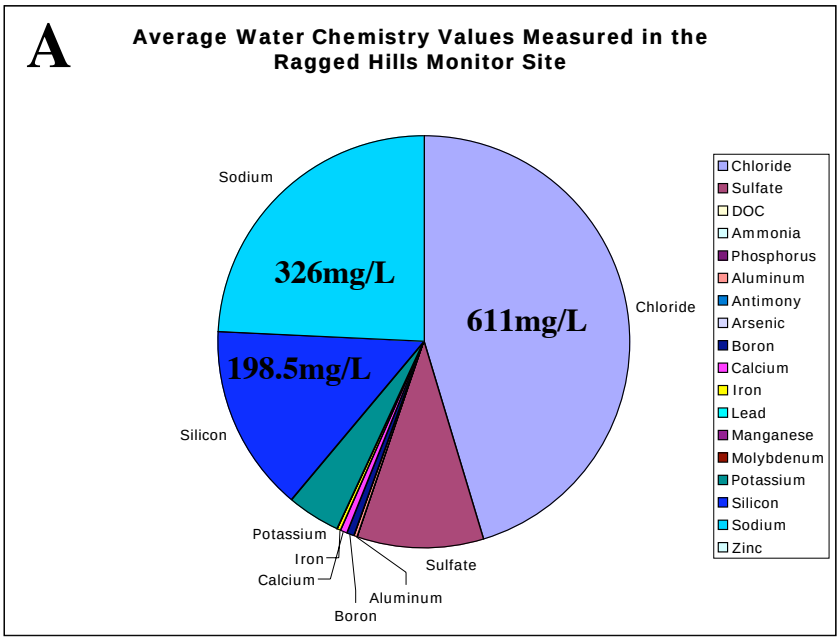
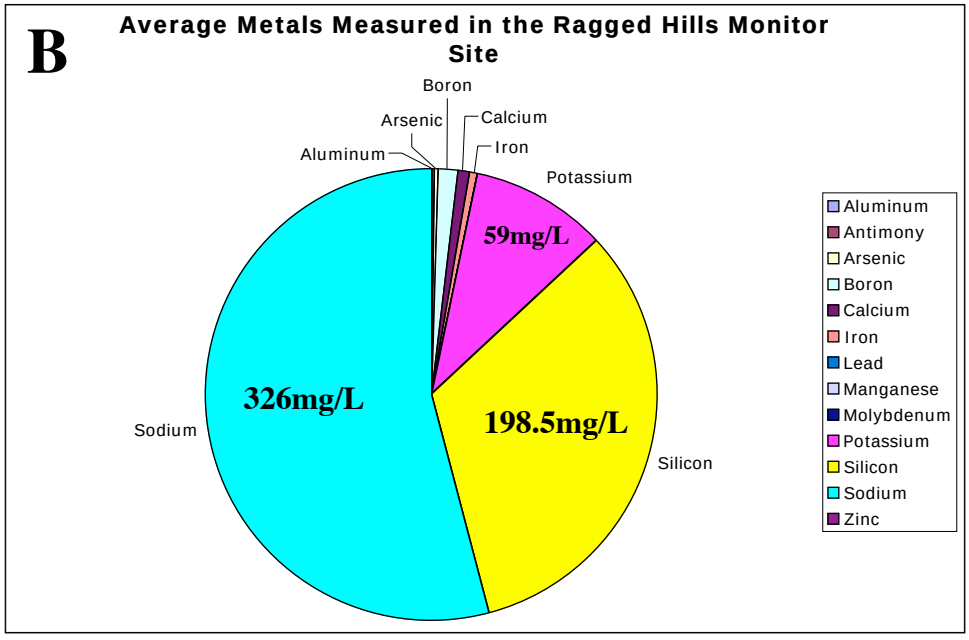


Figure 5.4. The average water chemistry values measured over time in the Ragged Hills monitor site. A. Average of all the chemical species measured. B. Average values of the metals detected in the Ragged Hills monitor site. All values are mg/L.

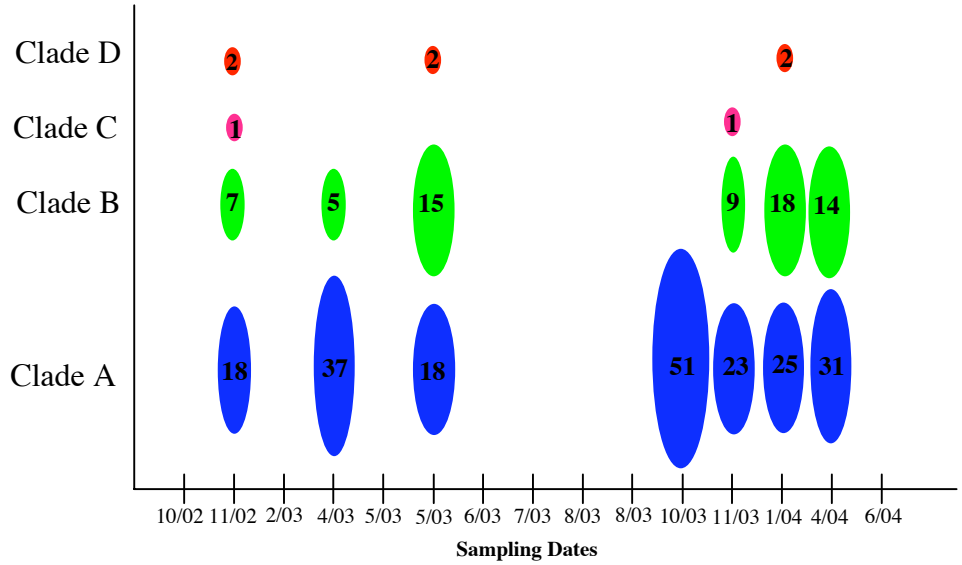




Archaeal population in the Ragged Hills monitor site

The archaeal host population was estimated by 16S rDNA sequence analysis. Seven of fifteen samples were used to determine the dominant archaeal host population over time. Only the 20F/UA1406R primer set was successful on samples collected from this monitor site. The 348Fa/UA1406R primer set was attempted numerous times without success. An overview of the dominant archaeal populations detected in this research over time is presented in figure 5.5.

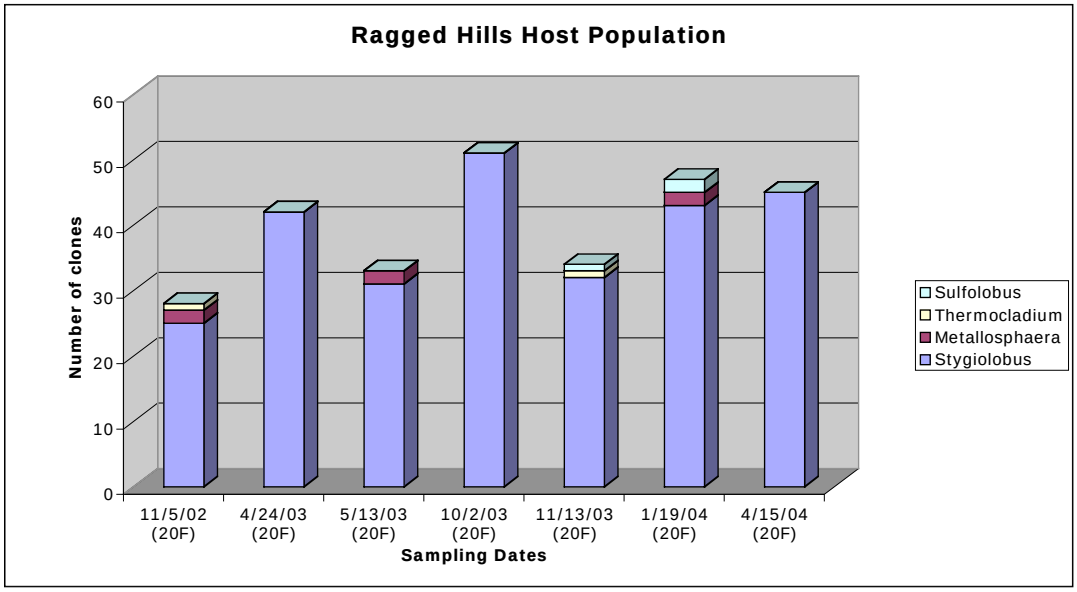
Figure 5.5. The archaeal population in Ragged Hills over time. The numbers indicate how many 16S rDNA clones belong to each clade. Only selected samples were used in the archaeal host analysis, therefore, the absence of sequences on some sampling dates does not reflect the absence of the sequences in the environment.



The first sample was collected on November 5, 2002. Twenty-eight full-length 16S rDNA clones were obtained and compared to 16S rDNA genes in the public database. Surprisingly, no *Sulfolobus*-like 16S rDNA sequences were detected (figure

5.6). Rather, 25 out of 28 sequences (89%) most closely matched *Stygiolobus azoricus* (accession number X90480) at 95-99% identity (figure 5.6 and table 5.2). These sequences belonged to clades A and B (figure 5.7). Thirteen of the environmental clones were ≥98% identical to the published *S. azoricus* 16S rDNA sequence (clade A in figure 5.7). Two of the remaining clones matched *Metallosphaera sedula* (accession number X90481) with 94% and 95% identity (clade D in figure 5.7) (table 5.2). The last clone (clade C in figure 5.7) most closely matched *Thermocladium modestius* (accession number AB005296) at 96% identity (table 5.2). These lower percentage matches are indicative of archaeal species that are similar to these named species, but clearly distantly related.

Figure 5.6. The archaeal population present in Ragged Hills at specific times based on 16S rDNA sequences.



The next sample used for host analysis was collected on April 24, 2003, and 42 16S rDNA clones were sequenced. The host population at this time was similar to that observed in November 2002 (figure 5.6). One hundred percent of the sequenced environmental clones (42/42) most closely matched *Stygiolobus azoricus* (figure 5.6). Thirty-seven of the environmental clones (clade A in figure 5.7) were ≥98% identical to the published *S. azoricus* 16S rDNA sequence (table 5.2).

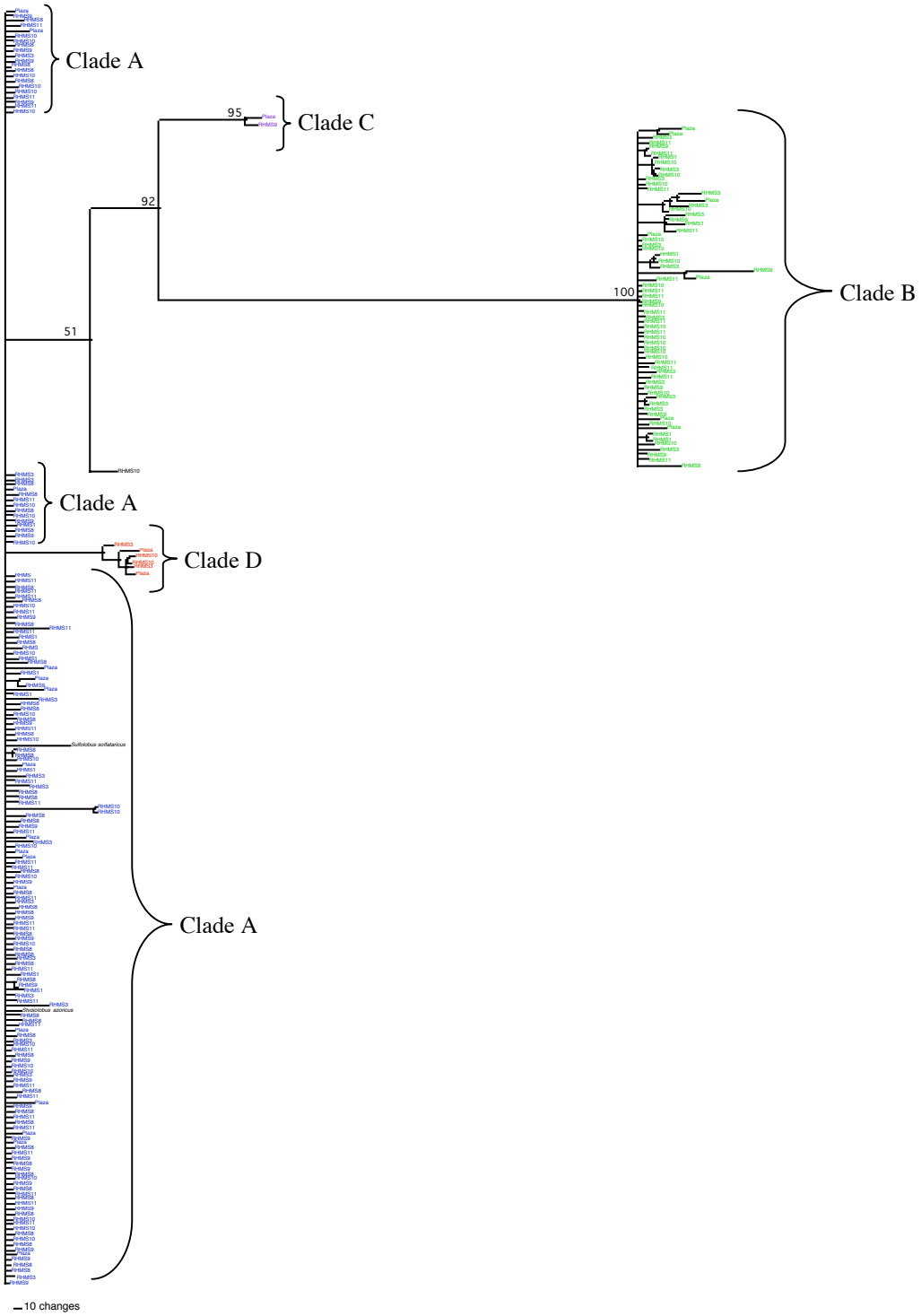
Table 5.2. The percent identities of environmental 16S rDNA clones from the Ragged Hills monitor site to published 16S rDNA sequences in the public database.

Sampling Date	**Clones**	***Stygiolobus***	***Thermocladium***	***Metallosphaera***	***Sulfolobus***
11/5/02	28	25 (95-99%)	1 (96%)	2 (94-95%)	ND
4/24/03	42	42 (95-99%)	ND	ND	ND
5/13/03	33	31 96-99%)	ND	2 (94-97%)	ND
10/2/03	51	51 (97-99%)	ND	ND	ND
11/13/03	34	32 (98-99%)	1 (97%)	ND	1 (98%)
1/19/04	47	43 (95-99%)	ND	2 (96%)	2 (98%)
4/15/04	45	45 (95-99%)	ND	ND	ND

*ND=Not Detected

For the sample collected on May 13, 2003, 33 clones were sequenced from this sample, and once again, *Stygiolobus*-like sequences dominated the population (clades A and B in figure 5.7). Ninety-three percent of the clones in this environmental library (31/33) were most similar to *Stygiolobus* at high percentages (96-99% identity) (figure 5.6 and table 5.2). Twenty-five clones from this sample were 98-99% identical to the published *S. azoricus* 16S rDNA sequence (table 5.2). The other organism (clade D in figure 5.7) detected in this sample was *Metallosphaera sedula* (2/33) at 94% and 97% identity (table 5.2).

Figure 5.7. The 16S rDNA sequences obtained from the Ragged Hills monitor site over time. The clades and colors correspond to figure 5.6.



On October 2, 2003, all of the environmental clones (clade A in figure 5.7) most closely resembled *Stygiolobus* at 97-99% identity (figure 5.6 and table 5.2). All, but two clones, showed 98-99% identity to the published *S. azoricus* 16S rDNA sequence. On November 13, 2003, 34 16S rDNA clones were obtained. Most of the sequences most closely matched *Stygiolobus* (32/34) with identities ranging from 98-99% (clades A and B in figure 5.7 and table 5.2). The other two sequences most closely matched *Thermocladium modestius* (97% identity) and *Sulfolobus solfataricus* (accession number X03235) (98% identity) (clades B and C in figure 5.7 and table 5.2), which is the first time that a *Sulfolobus*-like sequence has been detected from this monitor site during this research.

Stygiolobus-like organisms dominated the 47 full-length 16S rDNA clones collected on January 19, 2004 (figure 5.6). Ninety-one percent (43/47) of the clones matched *Stygiolobus azoricus* with 95-99% identity (clades A and B in figure 5.7 and table 5.2). Forty-one clones from this sample were 98-99% identical to the published *S. azoricus* 16S rDNA sequence. Of the remaining clones, two most closely matched *Metallosphaeara sedula* with 96% identity (clade D in figure 5.7), and the other two clones matched *Sulfolobus solfataricus* (98% identity) (clade A in figure 5.7 and table 5.2). For only the second time, *S. solfataricus*-like sequences were detected.

The last sample from Ragged Hills used for 16S rDNA analysis was collected on April 15, 2004. Forty-five 16S rDNA sequences were cloned from this sample and all of these clones most closely matched *Stygiolobus azoricus* (figure 5.6). Forty-two clones

from this sample were more than 98% identical to the published *S. azoricus* 16S rDNA sequence (clades A and B in figure 5.7 and table 5.2).

We detected an obvious primer bias from this monitor site. Surprisingly, the 348Fa/UA1406R primer set never resulted in the amplification of 16S rDNA sequences from Ragged Hills. This is especially perplexing because based on the published *S. azoricus* 16S rDNA sequence, the 348Fa primer matches better than the 20F primer (table 3.14). In spite of this, all of the *S. azoricus*-like 16S rDNA sequences from this site were amplified using the 20F/UA1406R primer set.

Overall, the archaeal population within the Ragged Hills monitor site was the least diverse of the three study sites (see Chapters 3 and 4). The clone libraries from this site were dominated by *Stygiolobus* and *Stygiolobus*-like 16S rDNA sequences (figure 5.6). In fact, a *Sulfolobus* 16S rDNA sequence was not detected from this monitor site until October 2003. Based on the 16S rDNA gene analysis, *Sulfolobus*-like sequences are not present in high numbers within this monitor site.

Population dynamics of viruses in the Ragged Hills monitor site

SSV-like population dynamics in the Ragged Hills monitor site. Total DNA was extracted from water as described previously. Universal SSV primer sets 3F/4R and 7F/8R were used to measure the total diversity of SSVs in this monitor site. The SSV 3F/4R primer set was successful in six samples out of 15 total collected samples (table 5.3), and the 7F/8R primer set only amplified sequences from two samples (table 5.3). An overview of SSV-like 3F/4R phylogenetic results is provided in figures 5.8 and 5.9.

Figure 5.8. The different groups of SSV-like viruses detected in the Ragged Hills monitor site over time based on DNA sequences. Only functional SSV-like ORFs were included in the analysis. The colors of the groups correspond to figure 5.12. The empty sampling dates represent the inability to amplify SSV-like 3F/4R sequences from the environment.

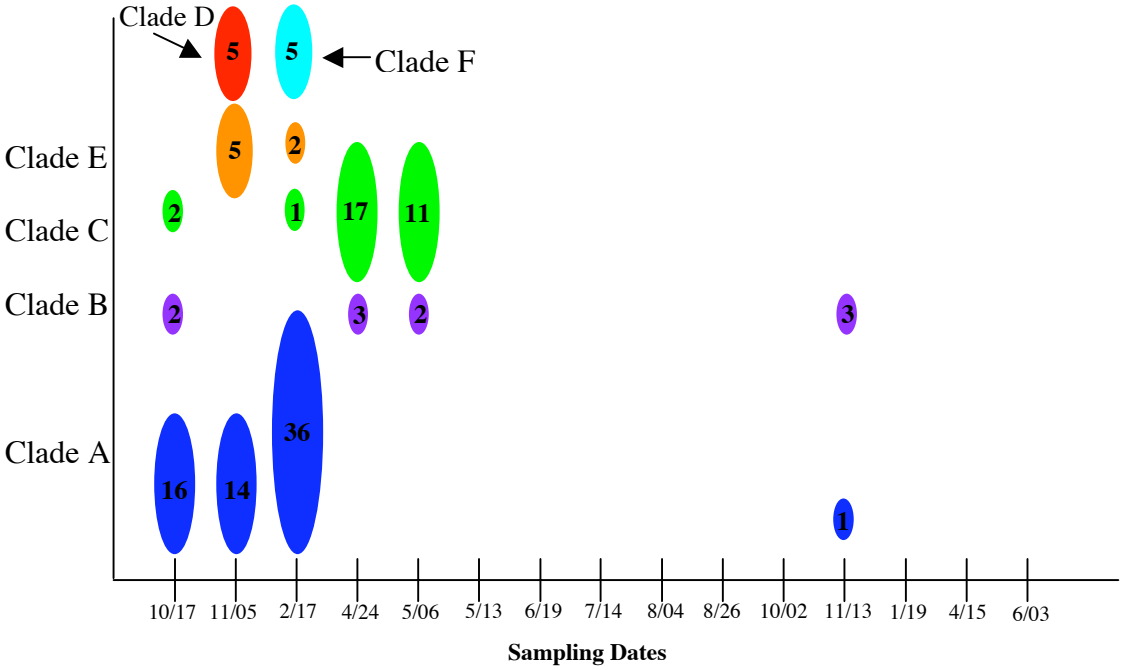
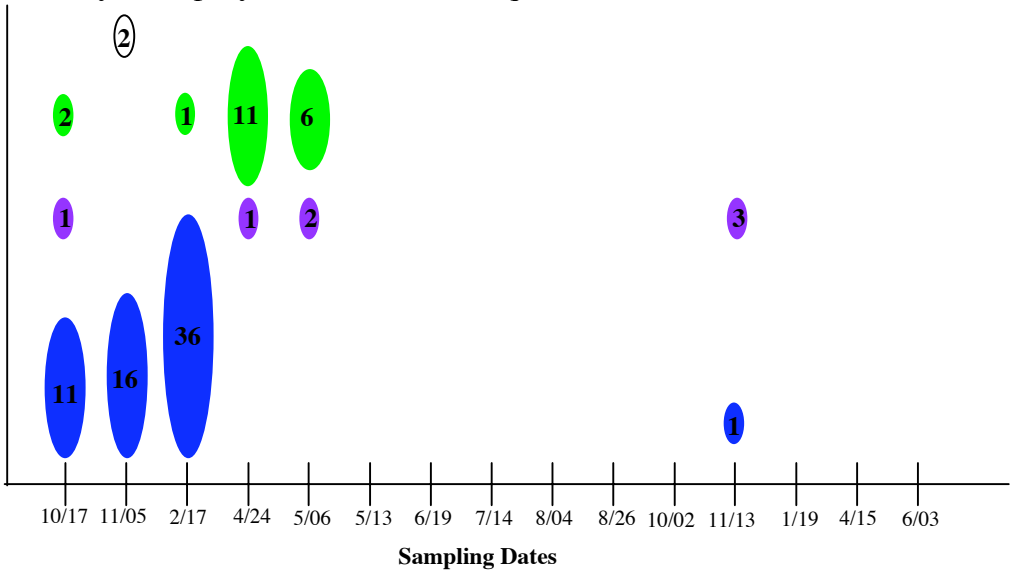


Figure 5.9. The different groups of SSV-like viruses detected in the Ragged Hills monitor site over time based on protein sequences. The colors of the groups correspond to figure 5.13. The empty sampling dates represent the inability to amplify SSV-like 3F/4R sequences from the environment.



The first sample was collected on October 17, 2002 and 76 cloned SSV-like sequences were amplified with the 3F/4R primer set (table 5.3). As shown in figure 5.10a, there appears to be one dominant population of SSV-like sequences with a couple of outliers. It is interesting to note the similarity between some of the environmental sequences to the four sequenced isolates (figure 5.10a; sequenced isolates in blue). SSV-K1 is closely related to two environmental sequences, SSV1 is also closely related to two environmental sequences, and SSV2 and SSV-RH are not closely related to any environmental sequences from October 2002 (figure 5.10a). It is interesting that SSV-RH is not more closely related to the SSV-like environmental sequences because it was isolated from the same geothermal area within YNP (Wiedenheft 2004). The SSV-like 3F/4R sequences from this sample belong in clades A, B, and C (figure 5.12).

Table 5.3. Number of SSV-like clones sequenced from the Ragged Hills monitor site and how many unique DNA sequences in each library.

Sample Name	Collection Date	pH	Temp (°C)	Number of Clones 3F/4R	Unique Sequences	Number of Clones 7F/8R	Unique Sequences
RH	Oct 17, 2002	3.7	72	76	24	56	12
Plaza	Nov 5, 2002	2.7	70	54	27	57	33
RGMS	Feb 17, 2003	3.2	68	64	44	0	0
RHMS1	Apr 24, 2003	3.1	72	43	21	0	0
RHMS2	May 6, 2003	2.8	70	39	18	0	0
RHMS3	May 13, 2003	3.1	66	0	0	0	0
RHMS4	June 19, 2003	2.9	72	0	0	0	0
RHMS5	July 14, 2003	3.1	72	0	0	0	0
RHMS6	Aug 4, 2003	2.6	78	0	0	0	0
RHMS7	Aug 26, 2003	3.3	74	0	0	0	0
RHMS8	Oct 2, 2003	2.8	74	0	0	0	0
RHMS9	Nov 13, 2003	3.5	65	9	7	0	0
RHMS10	Jan 19, 2004	2.7	64	0	0	0	0
RHMS11	Apr 15, 2004	3.5	57	0	0	0	0
RHMS12	June 3, 2004	3.2	69	0	0	0	0
Total Number of Clones Sequenced				**285**	**109**	**113**	**48**

From October 2002, there are 14 inferred functional protein sequences and five sequences that resulted in non-functional ORFs (table 5.4). The functional protein sequences form one group with the exception of three outliers, which are more similar to one of the completely sequenced isolates (figure 5.10b). The non-functional ORFs in this sample are not obvious in the phylogenetic tree and are similar to the un-interrupted ORFs (figure 5.10a; the • represents the non-functional ORFs). Upon closer investigation, four of the five interrupted ORFs have one stop codon. The first sequence contains an insertion of guanidine (G) at position 214, which results in a premature stop codon. Three sequences contain a transition or transversion that results in a stop codon, but the frame of the sequence is maintained. The fifth interrupted ORF contains two stop codons due to the deletion of a guanidine (G) at position 174, which results in a frameshift.

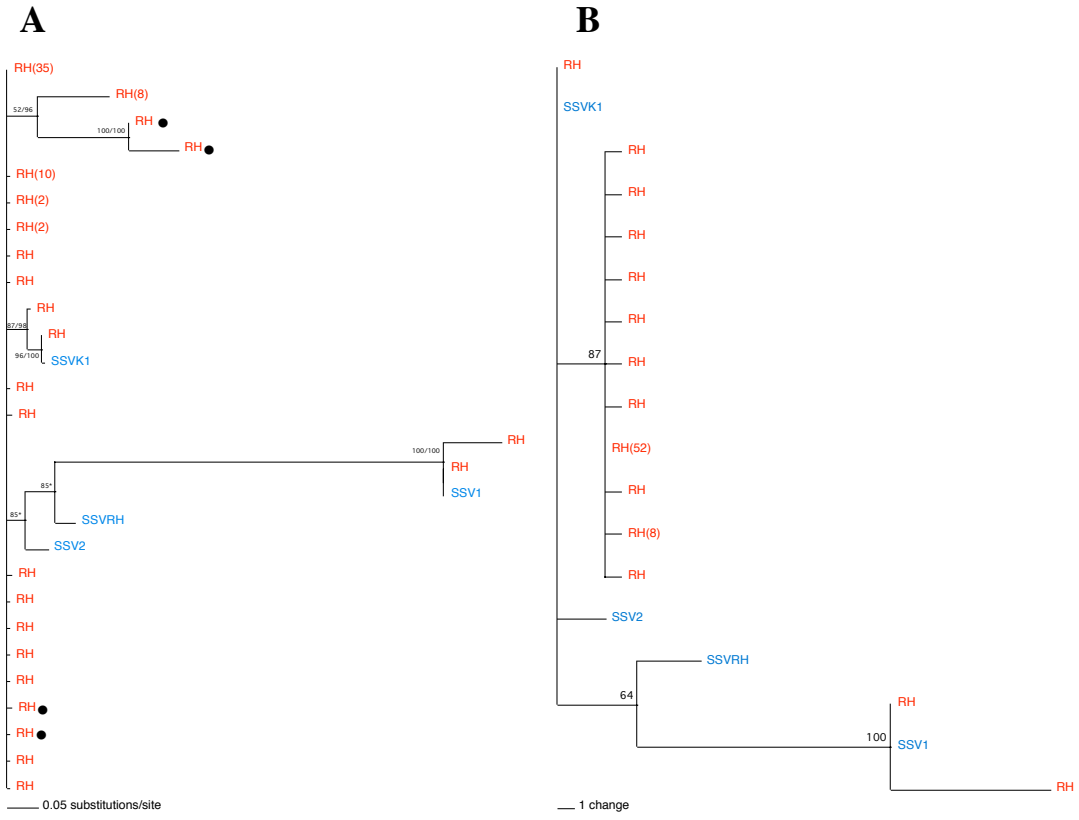
Table 5.4. The comparison of SSV-like 3F/4R functional and non-functional protein sequences detected from the Ragged Hills monitor site.

Sample Date	Sample Name	Unique DNA Sequences	Functional Protein Sequences	Non-Functional Protein Sequences
October 17, 2002	RH	24	14	5
November 5, 2002	Plaza	27	18	3
February 17, 2003	RGMS	44	37	0
April 24, 2003	RHMS1	21	12	1
May 6, 2003	RHMS2	18	8	5
November 13, 2003	RHMS9	7	4	2
TOTAL		**109**	**87**	**16**

Fifty-six SSV-like coat protein (primer set 7F/8R) environmental clones were also obtained from this sample (table 5.3). The coat protein sequences of the SSV-like viruses show different results compared to the 3F/4R primer set (figure 5.11a). In the coat

protein (7F/8R) analysis, SSV1 and SSV-K1 are not closely related to environmental sequences (figure 5.11). Instead, the sequenced SSV isolate from Ragged Hills (SSV-RH) is most closely related to a smaller group of environmental sequences, and SSV2 shows similarity to one sequence from Ragged Hills (figure 5.11). This result is consistent with previous results from comparing the four sequenced isolates (Wiedenheft *et. al.* 2004).

Figure 5.10. SSV-like 3F/4R sequences collected from Ragged Hills in October 2002. The numbers in parentheses represents the number of times the sequence was detected. The • represents DNA sequences that are translated to non-functional ORFs. (A) DNA sequences, and (B) inferred protein sequences.

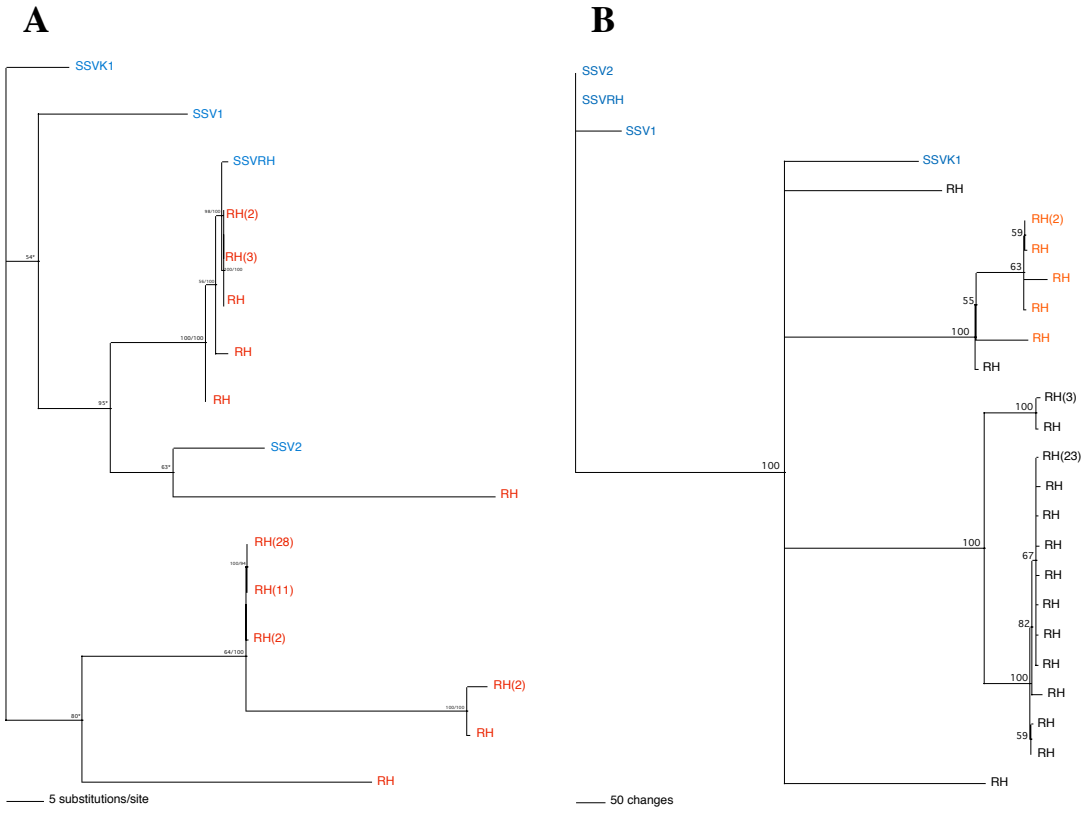


The SSV-like 7F/8R inferred protein sequences from October 2002 reveal that three of the sequenced isolates (SSV2, SSV-RH, and SSV1) are distinct from the YNP sequences (figure 5.11b). The environmental SSV-like coat protein sequences form two major groups based on amino acid sequence (figure 5.11b). Interestingly, the smaller group is comprised of SSV-like sequences that contain three coat proteins (indicated by orange in figure 5.11b). It was unexpected that SSV1 did not group with these sequences, but on closer inspection, the YNP sequences are quite distinct from the Japanese isolate. These seven SSV-like coat protein sequences were the only sequences detected throughout this research (from all three monitor sites) that contained three coat proteins.

The second sample was collected on November 5, 2002, and 59 SSV-like environmental 3F/4R sequences (table 5.3) were compared to the SSV-like sequences from October 2002. The sequences from this sample belong to SSV-like 3F/4R clades A, D, and E (figure 5.12). There are two smaller groups of environmental SSV-like sequences detected in November 2002 that were not detected in October 2002 (clades D and E in figure 5.12). The inferred protein analysis resulted in a similar tree topology when comparing the October and November 2002 sequences. Most of the sequences from November 2002 are related to the October 2002 sequences (blue group in figure 5.13). However, there were two protein sequences that were outliers of the large group (figure 5.13). There was one protein sequence detected in this sample that was identical to a protein sequence from October 2002 (figure 5.13; identical sequence represented by *). The translated sequences from this sample resulted in 18 functional ORFs and three

non-functional ORFs (table 5.4). Two of the non-functional ORFs are somewhat obvious in the phylogenetic analysis; however, they resemble a group of functional sequences (see Appendix C, figure 5.25). These two sequences contain numerous stop codons. The remaining non-functional sequence groups with many functional sequences. Upon closer inspection of the sequence, three stop codons occurred in the beginning of the sequence, but the frame was maintained.

Figure 5.11. SSV-like coat protein (7F/8R) sequences collected from Ragged Hills in October 2002. The orange represents sequences composed of three coat proteins. The number in parentheses represents the number of times the sequence was detected. (A) DNA sequences, and (B) Inferred protein sequences.



Fifty-seven SSV-like coat protein clones (7F/8R) were also sequenced from the November 2002 sample. Most of the coat protein sequences (7F/8R) from November 2002 are similar to October 2002 sequences (figure 5.14); however, there is the formation of two new groups. The inferred protein analysis reveals the November 2002 protein sequences are similar to the October 2002 sequences (figure 5.15). None of the coat proteins from this sample consisted of three coat proteins.

The third sample was collected on February 17, 2003, and 73 SSV-like 3F/4R clones were sequenced (table 5.3). The SSV-like sequences from this sample belong to clades A, E, and F (figure 5.12). We no longer detect the minor group of SSVs that was detected in November 2002 (clade D in figure 5.12). However, we detect a new group of SSV-like 3F/4R sequences (clade F in figure 5.12). The DNA sequences resulted in 37 functional ORFs and no non-functional ORFs (table 5.4). Most of the inferred protein sequences from this sample belong in the largest group of viruses (represented by blue figure 5.13). However, this sample contained one protein sequence that was most closely related to SSV1 (represented by green in figure 5.13). More protein sequences that are identical to each other were detected in this sample (represented by * in figure 5.13). The SSV-like viral coat proteins were not detected in this sample or any of the future samples collected from this monitor site (table 5.3).

There is continued change in the SSV-like population with the next sample. The fourth sample collected in April 2003 yielded 43 SSV-like 3F/4R clones (table 5.3). When the sequences from this sample were compared to SSV-like sequences previously detected, two groups that are distinctly different are present (clades B and C in figure

5.12). The largest group of viruses detected from this monitor site (clade A) was not detected in this sample. The protein-based analysis indicates these sequences are highly related to protein sequences from previous samples (groups represented by purple and green in figure 5.13). There are two groups of protein sequences; the larger group is closely related to SSV-K1 (purple in figure 5.13), and the other group is more closely related to SSV1 (green in figure 5.13).

This sample contained inferred protein sequences that were identical to sequences previously detected from this monitor site (figure 5.13; the * represents identical protein sequences). When the DNA sequences from this sample were translated, 12 functional protein sequences and one non-functional protein sequence resulted (table 5.4). When looking at the DNA-based phylogenetic analysis, it is not obvious what sequence represents the non-functional ORF (see Appendix C, figure 5.29). The non-functional ORF contains a second position transversion at nucleotide 167, which results in a serine (TCA) being changed to a stop codon (TAA).

The diversity of SSV-like 3F/4R sequences decreases with the 39 clones collected on May 6, 2003 (table 5.3). The sequences from May 2003 belong in previously established groups (clades B and C in figure 5.12). This sample contained eight inferred functional protein sequences and five sequences that resulted in non-functional ORFs (table 5.4). The functional protein sequences from this sample belong in the previously established groups in this monitor site (represented by purple and green in figure 5.13). Most of the protein sequences from this sample are identical to a sequence that has been detected in previous samples (figure 5.13; the identical sequences are represented by *).

Figure 5.12. SSV-like 3F/4R DNA sequences from Ragged Hills with new group colors. Colors correspond to figure 5.8. Inset: An unrooted phylogram of SSV-like sequences from Ragged Hills.

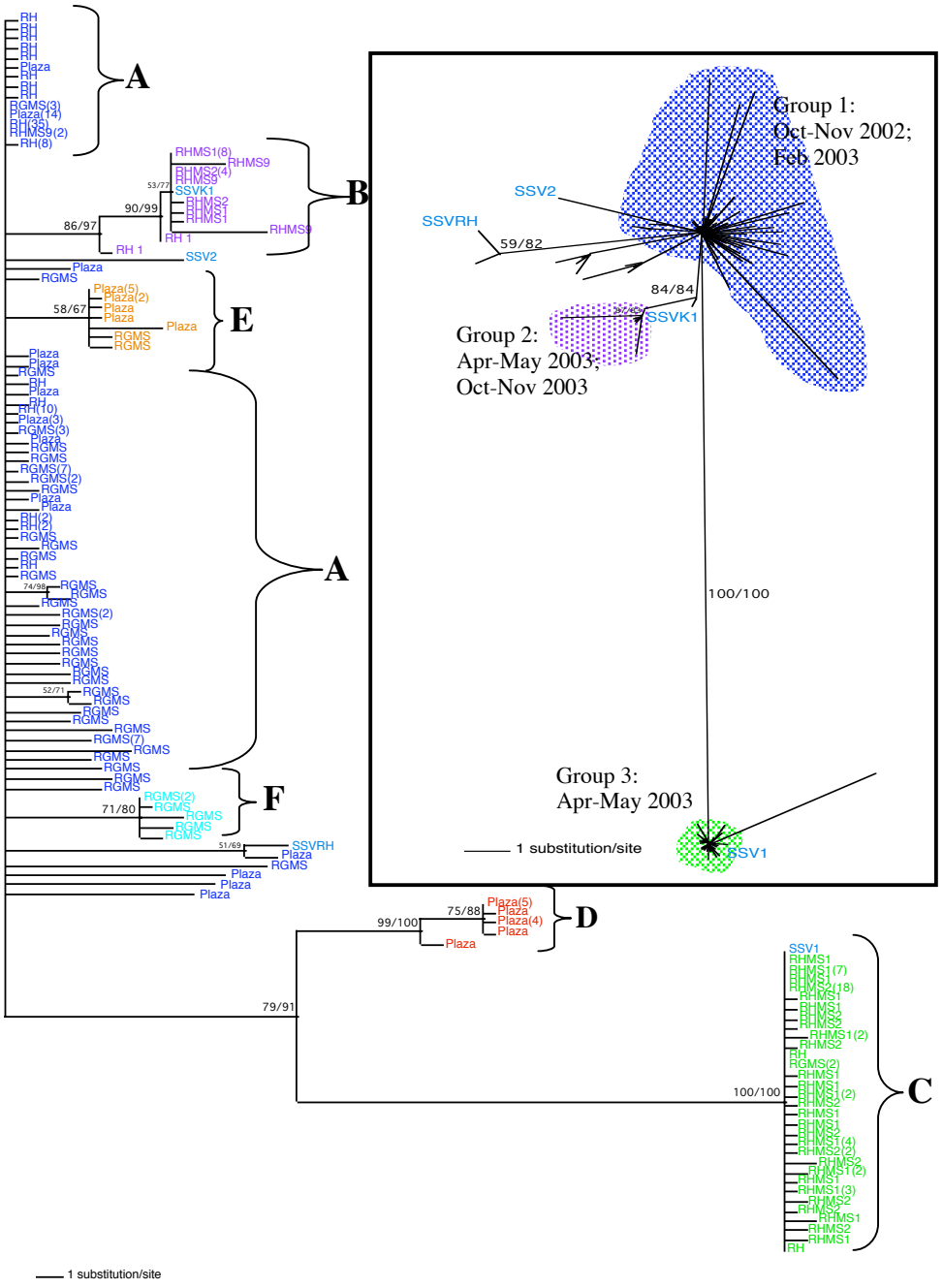
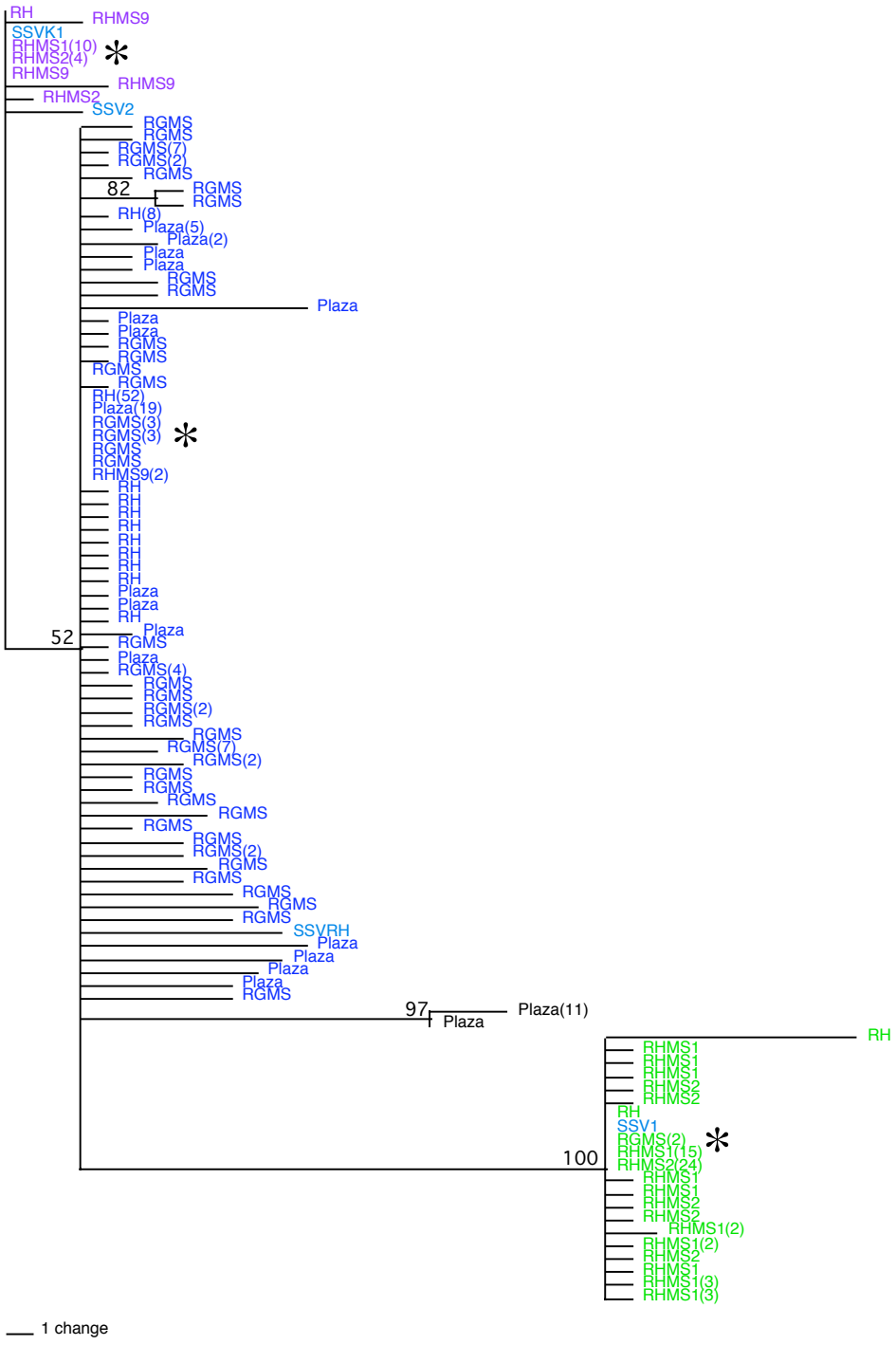


Figure 5.13. SSV-like 3F/4R protein sequences from Ragged Hills. The different groups represent the different groups of viruses detected over time. The * represents identical sequences.



The DNA sequences resulting in non-functional ORFs are similar to functioning ORFs from this sample (see Appendix C, figure 5.31). Two sequences contain a change from a tyrosine (TAT) to a stop codon (TAA) at nucleotide position 252. The three remaining sequences all contain one base change that results in a stop codon.

Six more samples were collected during the time period from May 2003—October 2003, but no SSV-like 3F/4R sequences were detected from these samples (table 5.3). These samples were extensively analyzed. We believe the lack of detected SSV-like sequences is not likely a technical issue because (i) there were numerous attempts to amplify SSV-like sequences (typically ≥20 PCR reactions), (ii) the controls used in PCR amplifications indicated our primers were very sensitive (see Chapter 3), and (iii) we were able to amplify 16S rDNA and SIRV-like sequences from the same extracted environmental nucleic acid. We believe the lack of SSV-like sequences is due to the fact these viruses were present in very low abundance in the environment during this time (<0.1pg/PCR reaction).

The next sample that SSV-like sequences were collected was on November 13, 2003, but only nine SSV-like 3F/4R clones were sequenced (table 5.3). When these nine SSV-like sequences were compared to sequences from previous samples, they were similar to sequences that had been previously identified (clades A and B in figure 5.12). However, in comparison with the sample collected on May 6, 2003, the large SSV-like group (clade A) is detected again (figure 5.12). There were only four functional protein sequences and two non-functional protein sequences in this sample (table 5.4). The functional protein sequences detected from this sample belonged to two previously

established groups (represented by blue and purple in figure 5.13). One of the non-functional protein sequences contains a single base change in a glycine (GGT) at amino acid seven resulting in a stop codon (TGA), and the other non-functional sequence contained four stop codons (see Appendix C, figure 5.33).

Figure 5.14. SSV-like 7F/8R DNA sequences from Ragged Hills. The different colors represent the three major groups of virus detected over time. The number in parentheses indicates how many times that particular sequence was detected.

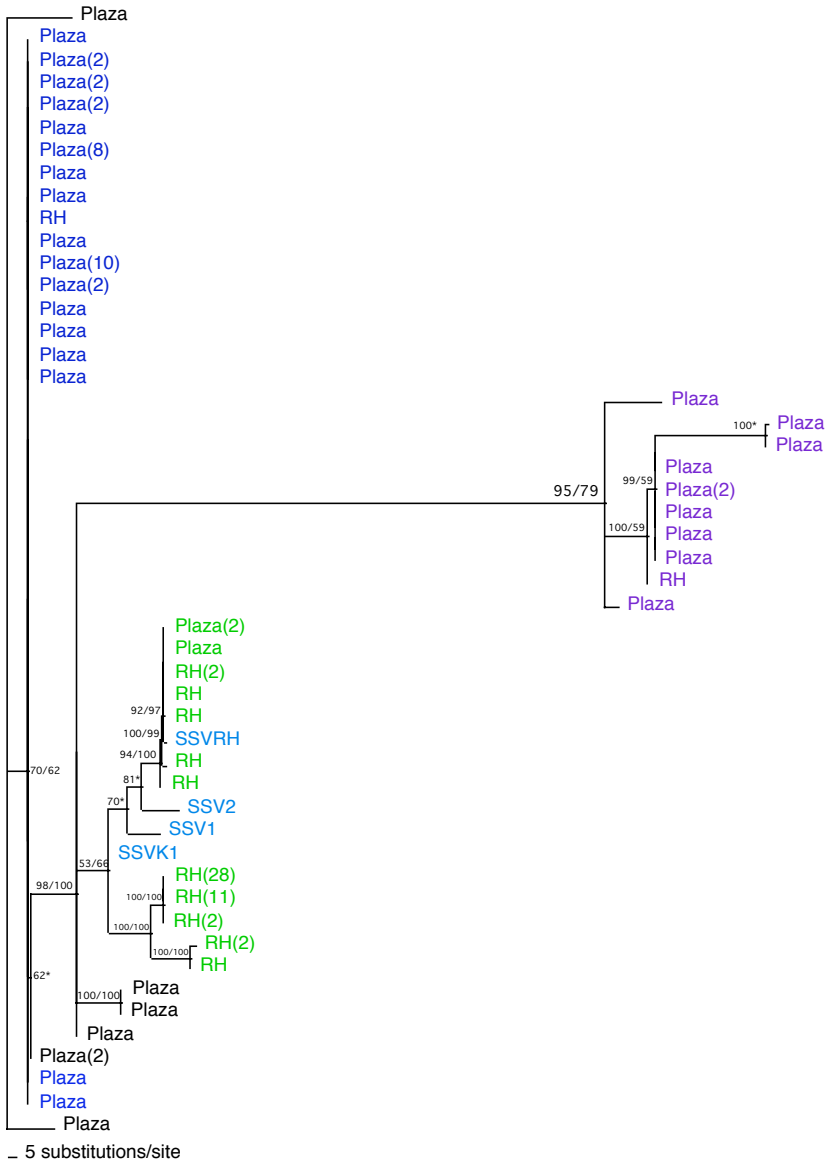
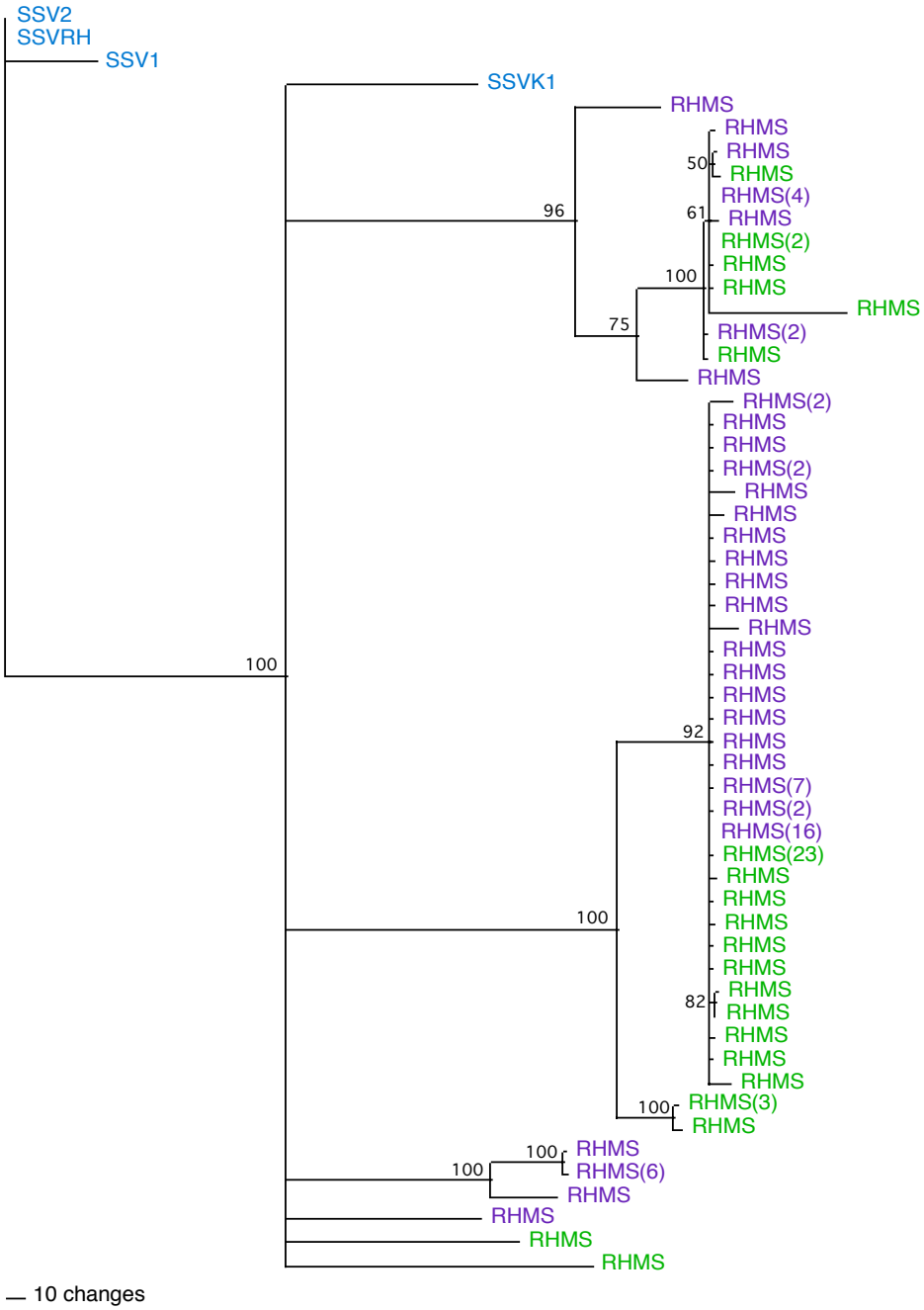


Figure 5.15. SSV-like 7F/8R protein sequences from Ragged Hills. The green represents the sequences collected in October 2002, and the purple represents the sequences collected in November 2002.



Three more samples were collected from the Ragged Hills monitor site (January 2004-June 2004), however, there were not any SSV-like sequences detected from these samples (table 5.3). However, as mentioned above, we do not contribute this absence to an experimental or technical problem. It is most probable the SSV-like viruses were in low abundance during this time period (<0.1pg/PCR reaction).

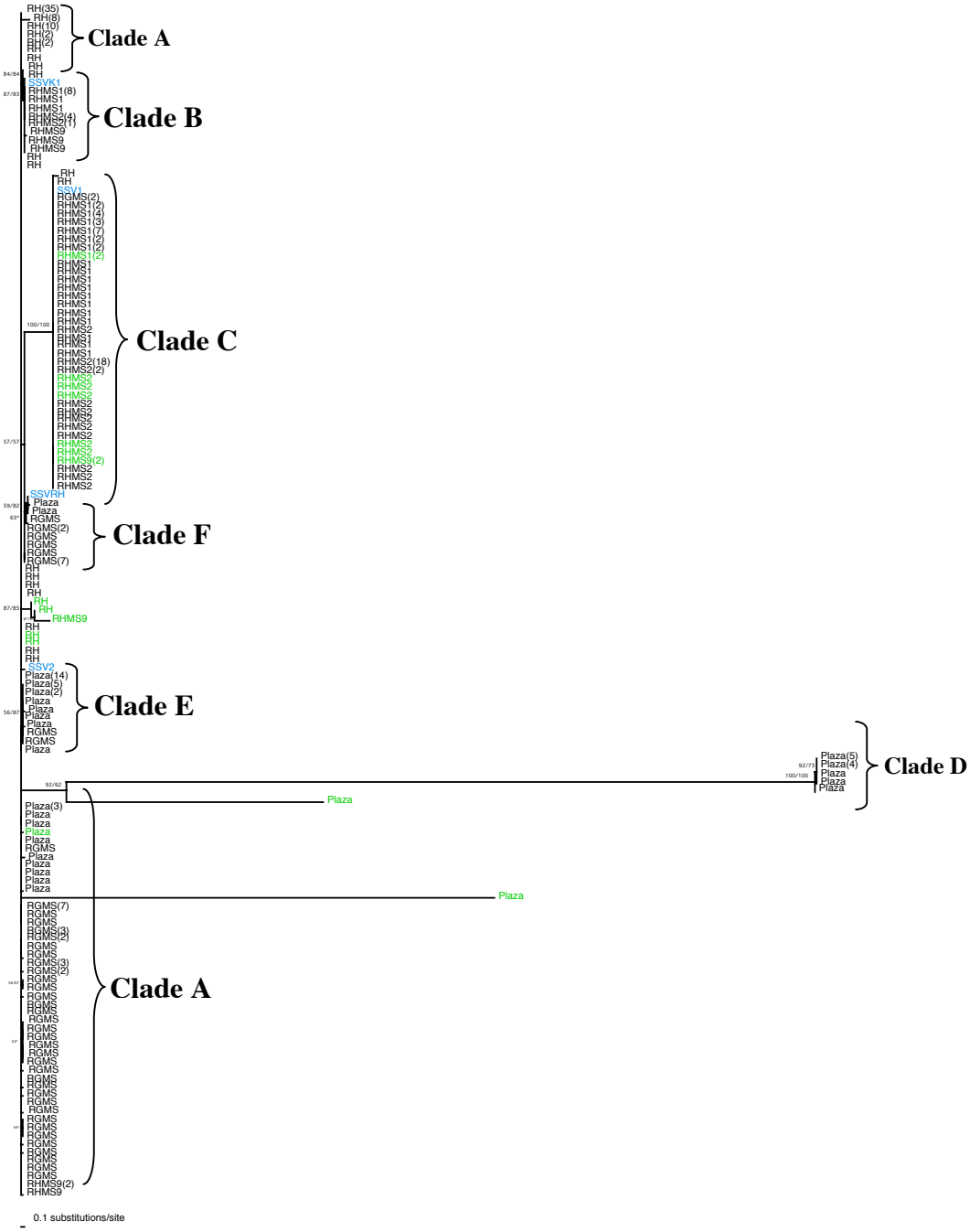
In summary, fifteen samples were collected from this monitor site over the course of two years of this research (October 2002—June 2004). Based on the SSV-like 3F/4R sequences, there were three major and two minor groups of SSV-like 3F/4R DNA sequences during the course of sampling this monitor site (figure 5.12). Surprisingly, each of the dominant groups is more closely related to a completely sequenced SSV isolate than they are to each other (figure 5.12).

It is more difficult to conclude how many viral groups were present over time using the SSV-like coat protein analysis (7F/8R). The SSV-like coat proteins were only detected twice in the early samples collected from this monitor site (table 5.3). However, we can estimate there were approximately three major groups of SSV-like viruses based on coat protein DNA sequences (figure 5.14). The three groups contain SSV-like coat protein sequences from both samples (October and November 2002) (figure 5.14). The four sequenced isolates all belong to one group of SSV-like coat proteins from this monitor site (figure 5.14). The complete phylogenetic analysis of SSV-like sequences included analysis of sequences after collection of every sample. For a step-by-step view of the changes in the SSV-like population over time, see Appendix C.

It is interesting to compare the phylogenetic positioning of functional and non-functional SSV-like 3F/4R DNA sequences (figure 5.16). This DNA-based analysis was performed including the non-functional SSV-like sequences (figure 5.16). Some of the non-functional ORFs were obvious in the phylogenetic analysis lying outside of functional virus groups. However, some non-functional ORFs were very similar to functioning ORFs (figure 5.16). When the non-functional ORFs are not included in the SSV-like 3F/4R DNA analysis, the topology of the phylogenetic tree is similar, but the presence of distinct groups of SSV-like viruses is more apparent (figure 5.12). Without the long branches in the DNA-based analysis, it is much easier to visualize the presence of different SSV-like viral groups present from October 2002 to November 2003 (figure 5.12).

SSV-like 3F/4R DNA sequences vary with time. The first group related to SSV1 (clade C in figure 5.12) is represented mostly by sequences from April 2003 and May 2003, and contains two sequences from October 2002. The other groups of sequences were harder to visualize in the phylogenetic tree containing the interrupted ORFs (figure 5.16), but when the interrupted ORFs are omitted from the analysis, the groups are more distinguishable (figure 5.12). The largest group from this monitor site (clade A in figure 5.12) represents the first three samples (October and November 2002 and February 2003). Another group related to SSV-K1 contains sequences from April, May, October, and November 2003 (figure 5.12). Three smaller groups (clades D, E, and F) were detected in November 2002 and February 2003, and not detected again (figure 5.8).

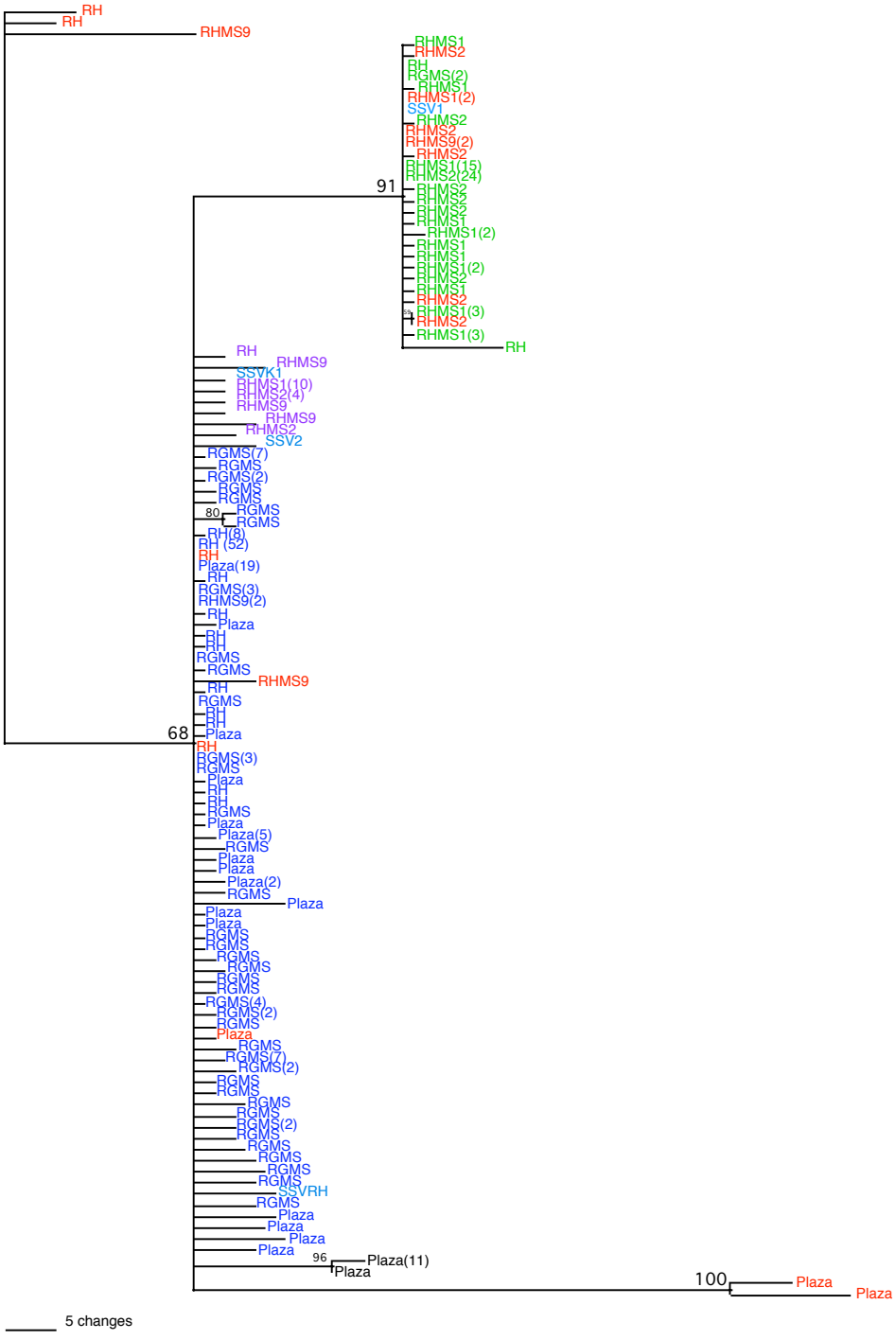
Figure 5.16. SSV-like 3F/4R DNA sequences from Ragged Hills. This phylogenetic analysis contained both functional and non-functional ORFs. The sequences colored green represent the DNA sequences that when translated result in an interrupted ORF. The viral clades from 5.12 are indicated on this tree.



Phylogenetic analysis of the SSV-like 3F/4R inferred protein sequences revealed three major groups over the duration of sampling (figure 5.13). These groups were consistently detected over time (figure 5.9). The inferred amino acid sequences of the SSV-like coat proteins (7F/8R) reveal the presence of two major groups with a couple of outlying sequences (figure 5.15). Three of the four sequenced SSV isolates are distinct from the SSV-like 7F/8R environmental sequences (figure 5.15). This separation is supported by a 100% bootstrap value (figure 5.15). More coat protein sequences would be necessary to better determine the number of SSV-like groups in this monitor site. It is interesting to note that changes in the DNA sequence of SSV-like viruses did not always result in a change in the inferred protein sequence.

Of the 87 SSV-like 3F/4R protein sequences, there were only three protein sequences that were detected in more than one sample (figure 5.13; the * represents groups of identical protein sequences detected more than once). The three SSV-like 3F/4R protein groups detected over time were consistently present (figure 5.9). The interrupted ORFs were not included in most of the protein-based analyses; however, they were included in an analysis of all the SSV-like 3F/4R protein sequences (figure 5.17). When the 16 non-functional protein sequences (table 5.4) are added to analysis, there are a couple of interesting results (figure 5.17). Only five of the interrupted SSV-like ORFs were very distinct from the functional protein sequences (figure 5.17; the interrupted ORFs are represented in red). When these protein sequences are investigated more closely, there are many insertions, deletions, transversions, and transitions resulting in numerous stop codons throughout the sequence.

Figure 5.17. SSV-like 3F/4R protein sequences from Ragged Hills. The red sequences represent the non-functional ORFs.



Seven of the non-functional protein sequences are closely related to the group of functional protein sequences related to SSV1 (figure 5.17). It was surprising that these sequences were closely related to functional protein sequences when some of the non-functional sequences were so distinct. However, when these protein sequences are more closely examined, there is a first position transversion at nucleotide 19 changing it from a glycine (GGA) to a stop codon (TGA). But, besides that change, these protein sequences are very similar to the group of functional protein sequences (represented by green in figure 5.17) related to SSV1. The remaining non-functional ORFs are located in the largest group of protein sequences (figure 5.17). When these sequences are more closely investigated, it was found a third position transversion at nucleotide 252 occurred resulting in a tyrosine (TAT) being changed to a stop codon (TAA).

It is possible that the non-functional SSV-like sequences are the result of PCR or sequencing errors. However, I believe this is unlikely. As discussed in previous chapters, when the SIRV-like sequences are translated there are not any non-functional sequences (see Chapter 3, 4, and below). If there were significant errors due to PCR or sequencing, we would expect to see them in the SIRV-like sequences and the 16S rDNA sequences. It is more likely the non-functional ORFs are due to the detection of integrated viral sequences. These sequences would no longer be under selection, and therefore, would accumulate errors.

Detailed analysis of SSV-like population dynamics. From October 2002 to June 2004, there was a total of 285 SSV-like 3F/4R clones sequenced (table 5.3). Using Modeltest version 3.6 (Posada *et. al.* 1998), the nucleotide substitution model for

maximum likelihood estimations was determined to be a general time reversible (GTR) model (Rodriguez *et. al.* 1990) with heterogeneity between sites (gamma=1.9455), no invariant sites, base frequencies A=0.2781, C=0.1927, G=0.2068, and T=0.3224, and a substitution model A-C=1.5710, A-G=1.9212, A-T=1.2963, C-G=0.8857, C-T=2.4901, and G-T=1.0000.

Of the 285 SSV-like 3F/4R clones sequenced, 109 identical sequences were detected. The most commonly detected SSV-like sequence was found 54 times, which composes 19% of the SSV-like population (table 5.5). This group is composed of sequences from October 2002 through February 2003 and then not detected again until November 2003 (table 5.5). The next most frequently detected sequence was found 29 times, which composes 10.1% of the SSV-like population (table 5.5). This group of sequences was found in most of the samples that were collected early from this monitor site (October 2002-May 2003) (table 5.5). A group of 16 sequences was also detected, which composed 5.6% of the SSV-like virus population (table 5.5). This group was composed of sequences from early samples (October 2002-February 2003 samples). This group was present early, but was undetectable in later samples (table 5.5). One group was composed of 13 identical sequences detected in April and May 2003, and detected again in November 2003 (table 5.5). One group of eight identical sequences was only detected in the October 2002 sample (table 5.5).

Table 5.5. Ragged Hills SSV-like 3F/4R sequence abundance illustrating the groups of identical sequences, how many times a group this size was detected, the percent of the SSV-like population, and what samples the group contained.

RAGGED HILLS SSV-like SEQUENCE ABUNDANCE			
Identical Sequences	Times Detected	Percent of SSV-like Population	Samples[a]
54	1	19%	RH;Plaza;RGMS;RHMS9
29	1	10.1	RH;RGMS;RHMS1;RHMS2
16	1	5.6	RH;Plaza;RGMS
13	1	4.5	RHMS1;RHMS2;RHMS9
8	1	2.8	RH
7	1	2.4	RGMS
7	1	2.4	RGMS
6	1	2.1	RHMS1;RHMS2
6	1	2.1	RGMS
5	2 unique[b]	1.7% each[c]	
4	1 unique	1.4% each	
3	1 unique	1% each	
2	11 unique	0.7% each	
1	85 unique	0.35% each	

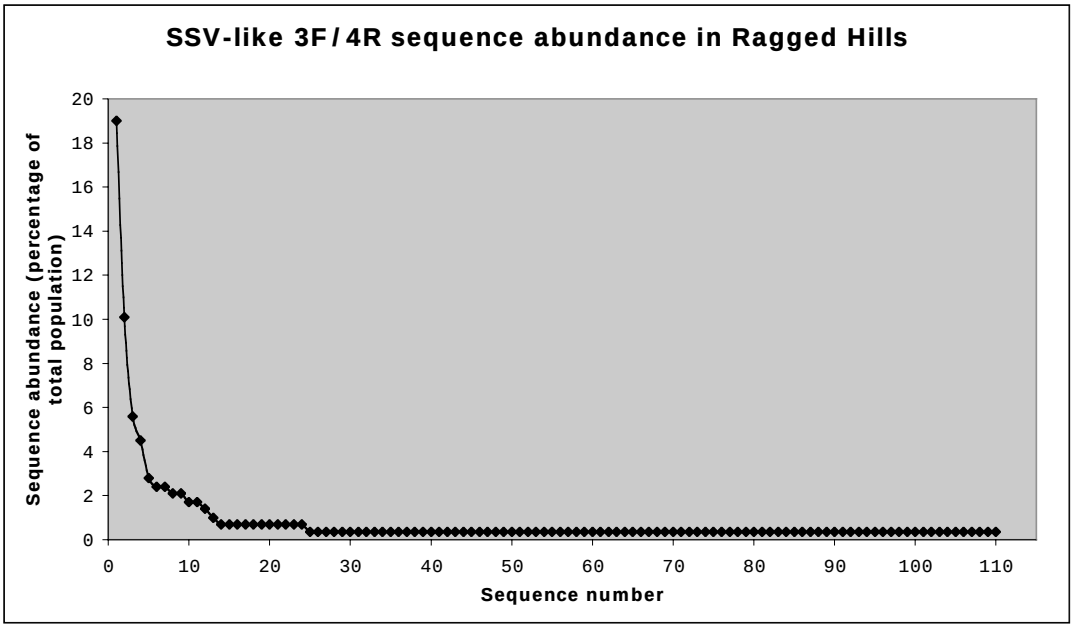
[a]sequences from these samples were present in the indicated groups
[b]each of the smaller groups were detected this many times, however, they are unique

Two unique groups composed of seven sequences each were also detected (table 5.5). Both groups were each composed of sequences from February 2003 (table 5.5). Two groups of six sequences were also detected in April-May 2003, and February 2003 (table 5.5). There were other groups of identical sequences detected ranging in size from two to five identical sequences (table 5.5).

Like the SSV-like sequences from the Crater Hills (see Chapter 3) and Rabbit Creek (see Chapter 4) monitor sites, unique sequences made up the majority of the SSV-

like population (figure 5.18). There were 95 unique SSV-like 3F/4R sequences, which translates to 33% of the SSV-like population over the course of sampling.

Figure 5.18. A sequence abundance plot for SSV-like 3F/4R sequences from Ragged Hills monitor site. The left end of the x-axis represents most commonly detected sequences, and the right end of the x-axis represents the most rare sequences. Most of the SSV-like sequences in Ragged Hills are detected once.



In comparing the 109 unique SSV-like 3F/4R DNA sequences obtained over time, 201 bps (out of an average of 256 bps) did not change in all of the samples (table 5.6). The rate of transitions (Ts) and transversions (Tv) are almost equal to each other (Ts/Tv=0.8). When comparing the base changes in each position of the codon, the third position base changes more frequently than the first or second base positions (table 5.6). The rates of transitions and transversions varied with individual samples (table 5.6).

Table 5.6. Numbers of transitions and transversions in SSV-like 3F/4R DNA sequences collected from the Ragged Hills monitor site.

SAMPLE DATE	IDENTICAL BASES	TRANSITIONS			TRANSVERSIONS		
		1st	2nd	3rd	1st	2nd	3rd
10/17/2002	231/258	3	2	6	5	4	6
11/05/2002	213/258	6	4	8	8	9	10
02/17/2003	245/258	3	1	5	1	1	2
04/24/2003	236/253	2	2	5	2	2	3
05/06/2003	214/253	6	2	13	6	3	9
11/13/2003	243/258	2	0	6	2	1	2

When comparing all of the sequences, 78% of the bases in the 3F/4R sequence remained unchanged (table 5.7). Transversions occurred slightly more (31) compared to transitions (24). Sixteen changes occurred in the first position of the codon; there were seven transitions and nine transversions (table 5.7). Only 12 changes occurred in the second position with four transitions and eight transversions (table 5.7). A total of 26 changes occurred in the third position; there were 13 transitions and 13 transversions (table 5.7). As expected, the third position changes occurred more frequently, therefore explaining the fewer number of groups based on protein sequences (figure 5.13).

Table 5.7. The numbers of transitions and transversions for all SSV-like 3F/4R sequences collected during this research at Ragged Hills.

Sample Range	Position	Identical Base Pairs	Ts Pairs	Tv Pairs
Oct 2002-Nov 2003 (256 bps)	*Overall*	*201/256*	*24*	*31*
	1st	68/85	7	9
	2nd	74/86	4	8
	3rd	59/85	13	13

Ts=transitions; Tv=transversions

From October 2002 to June 2004, there was a total of 285 SSV-like 7F/8R clones sequenced (table 5.3). Using Modeltest version 3.6 (Posada *et. al.*1998), the nucleotide

substitution model for maximum likelihood estimations was determined to be HKY+G (Hasegawa *et. al*. 1985) with heterogeneity between sites (gamma=0.6804), and base frequencies A=0.2872, C=0.2218, G=0.2132, and T=0.2779.

In comparing the 48 unique SSV-like DNA coat protein sequences (7F/8R) obtained over time, 441 bps (out of an average of 717 bps) did not change in all of the samples. The rate of transitions (Ts) and transversions (Tv) was (Ts/Tv=0.5). In the first sample collected in October 2002, 354 bps were identical over an average of 639 total base pairs (55% of the bases were identical). The first position showed more changes in comparison to the second and third positions (table 5.8). The second sample collected in November 2002 revealed slightly less change when compared to the first sample. Within this sample 363 bps out of 588 bps were identical (61%). There were 81 transitions and 144 transversions, with the majority of the changes occurring in the third position of the codon (table 5.8).

Table 5.8. The numbers of transitions and transversions in SSV-like 7F/8R DNA sequences collected from the Ragged Hills monitor site.

SAMPLE DATE	**IDENTICAL BASES**	**TRANSITIONS**			**TRANSVERSIONS**		
		1st	2nd	3rd	1st	2nd	3rd
10/17/2002	354/639	36	32	33	60	62	63
11/05/2002	363/588	30	25	26	45	49	51

When comparing all of the SSV-like coat protein sequences, only 61% of the bases in the sequence remained unchanged. Transversions occurred twice as often (184) compared to transitions (92) (table 5.9). Ninety-one changes occurred in the first position of the codon; there were 31 transitions and 60 transversions. Ninety-two changes

occurred in the second position with 29 transitions and 63 transversions. A total of 92 changes also occurred in the third position; there were 32 transitions and 60 transversions (table 5.9).

Table 5.9. The numbers of transitions and transversions for all SSV-like 7F/8R sequences collected from Ragged Hills during this research.

Sample Range	Position	Identical Base Pairs	Ts Pairs	Tv Pairs
October 2002-November 2002 (717 bps)	*Overall*	*441/717*	*92*	*183*
	1st	14/239	31	60
	2nd	146/239	29	63
	3rd	147/239	32	60

Ts=transitions; Tv=transversions

During the monitoring period, there were 87 unique SSV-like 3F/4R inferred protein sequences (table 5.4). There were six viral clades based on 3F/4R DNA sequences, but only three clades based on protein sequences (figures 5.8 and 5.9). Upon closer inspection, clades B and C represent minor protein groups (purple and green, respectively in figures 5.9 and 5.13). The largest protein group detected in this monitor site over time included clades A, D and E (blue in figures 5.9 and 5.13). Like the other two monitor sites, changes in SSV-like DNA sequences do not always result in changes in the protein sequences (figures 5.8 and 5.9).

SIRV-like population dynamics in the Ragged Hills monitor site. SIRV-like coat proteins were detected in all but one sample collected from the Ragged Hills monitor site (table 5.10). A total of 663 SIRV-like coat protein clones were sequenced from the Ragged Hills monitor site (table 5.10). As observed in Crater Hills (see Chapter 3) and Rabbit Creek (see Chapter 4), all of the SIRV-like coat proteins translated to un-

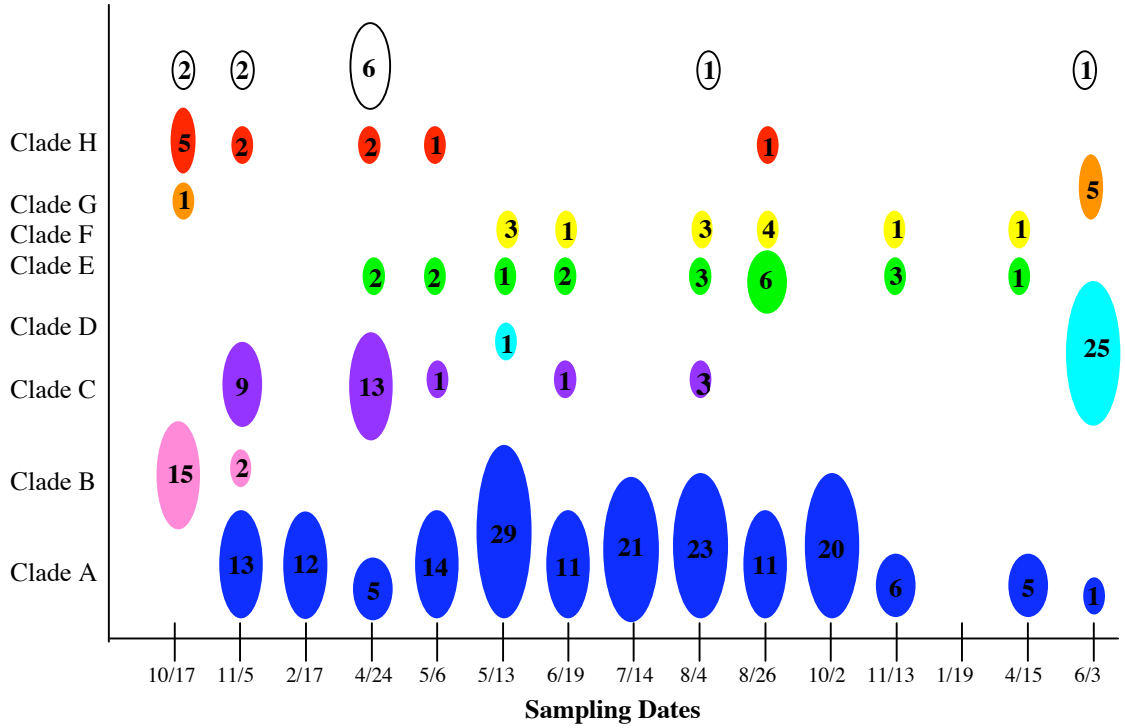
interrupted ORFs. For an overview of the SIRV-like phylogenetic analysis in the Ragged Hills monitor site, see figures 5.19 and 5.20.

Table 5.10. Number of SIRV-like clones obtained from the Ragged Hills monitor site, number of unique DNA sequences in each library, and the number of unique protein sequences in each library.

Sample Name	Collection Date	pH	Temp (°C)	Number of Clones	Unique Sequences	Protein Sequences
RH	Oct 17, 2002	3.7	72	56	22	16
Plaza	Nov 5, 2002	2.7	70	43	28	21
RGMS	Feb 17, 2003	3.2	68	60	12	9
RHMS1	Apr 24, 2003	3.1	72	41	28	22
RHMS2	May 6, 2003	2.8	70	37	18	12
RHMS3	May 13, 2003	3.1	66	76	33	28
RHMS4	June 19, 2003	2.9	72	25	15	13
RHMS5	July 14, 2003	3.1	72	75	21	17
RHMS6	Aug 4, 2003	2.6	78	60	33	25
RHMS7	Aug 26, 2003	3.3	74	50	22	17
RHMS8	Oct 2, 2003	2.8	74	54	20	15
RHMS9	Nov 13, 2003	3.5	65	24	10	9
RHMS10	Jan 19, 2004	2.7	64	0	0	0
RHMS11	Apr 15, 2004	3.5	57	14	7	6
RHMS12	June 3, 2004	3.2	69	48	32	19
Total Number of Clones Sequenced				**663**	**240**	**208**

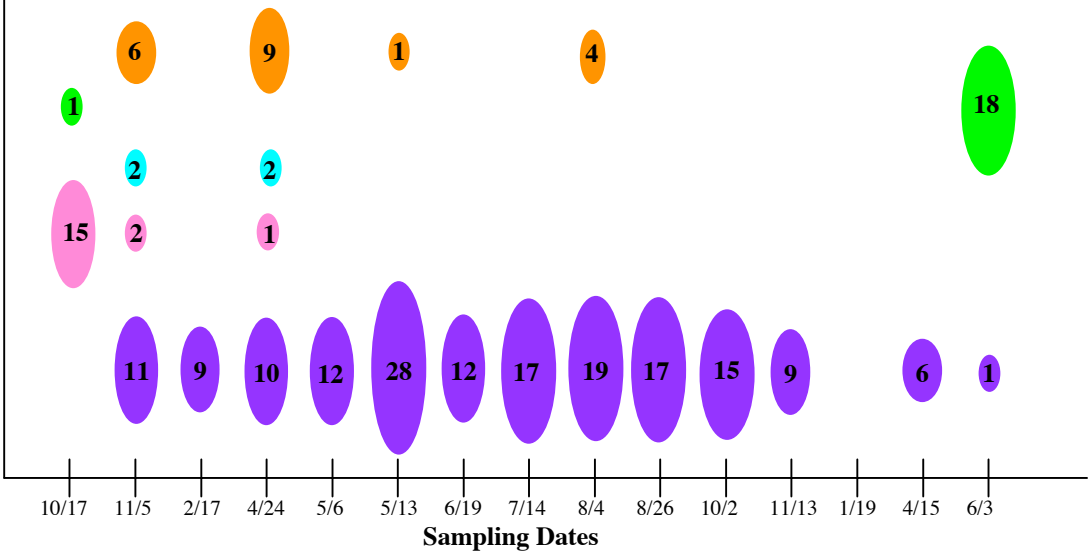
The first sample collected in October 2002 yielded 70 environmental clones of the SIRV-like coat proteins (table 5.10). This revealed one large population of SIRV-like viruses with a couple of outlying sequences from the major group (figure 5.21a). The 21 unique DNA sequences translated to 16 unique inferred protein sequences, which illustrated the same large group, excluding one outlying sequence that branches away from the other SIRV-like sequences (figure 5.21b).

Figure 5.19. The different groups of SIRV-like viruses detected in the Ragged Hills monitor site over time. The colors of the groups correspond to figure 5.22. The clear ovals refer to outlying sequences not belonging to a major group.



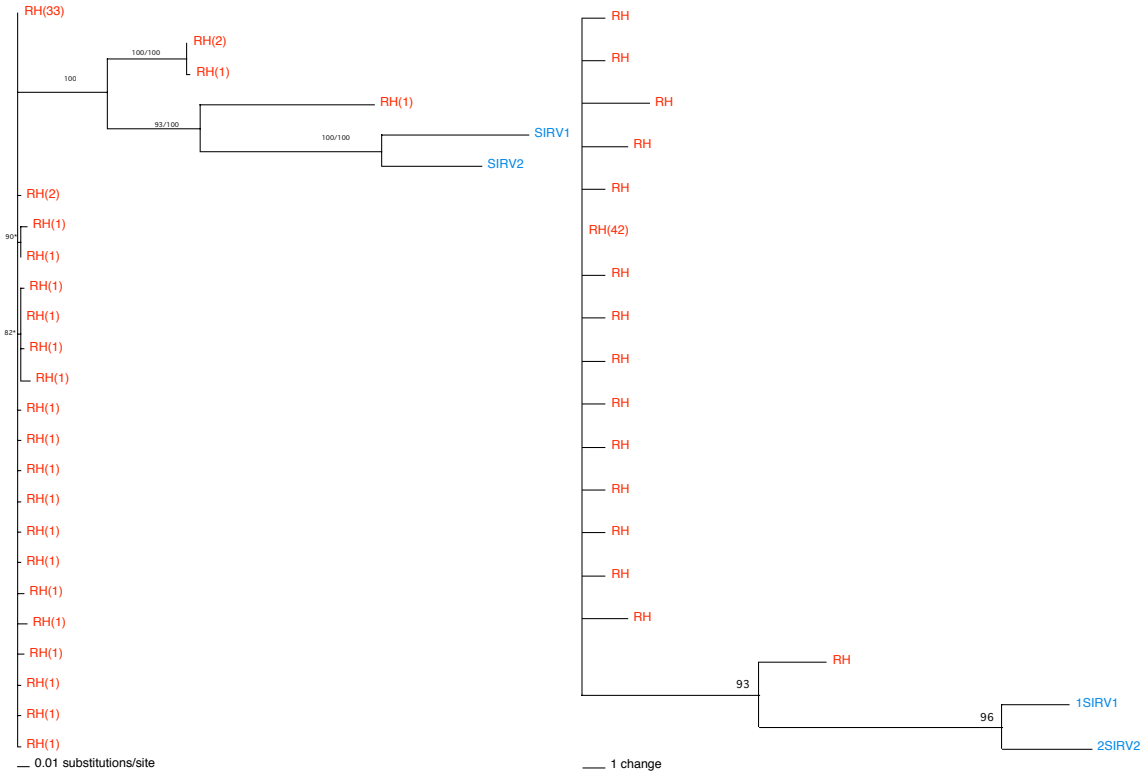
Change in the SIRV-like population is detected in the second sample. This sample was collected in November 2002 and reveals new SIRV-like virus groups in this hot spring (clades A, C, and H in figure 5.22). When 48 SIRV-like coat protein clones are compared to the sequences from October 2002, three new groups are evident (figure 5.22). Some of the November 2002 sequences are similar to the October 2002 sequences (clade B in figure 5.22), but most of them form new groups (clades A, C, and H in figure 5.22). The 28 SIRV-like coat protein DNA sequences translated to 21 unique inferred protein sequences (table 5.10), and three new groups are also evident based on these sequences (represented in purple, turquoise, and orange in figure 5.23).

Figure 5.20. The different groups of SIRV-like viruses detected in the Ragged Hills monitor site over time. The colors of the groups correspond to figure 5.23. The absence of SIRV-like sequences in January 2004 reflects the inability to amplify SIRV-like sequences from this sample.



The diversity of SIRV-like viruses is decreased in the next sample. The third sample was collected in February 2003 (table 5.10). Sixty SIRV-like coat protein sequences were obtained from this sample representing 12 unique DNA sequences (table 5.10). When the sequences from this sample are compared to SIRV-like sequences collected from previous dates, the diversity of the SIRV-like virus population is reduced to a single group (clade A in figure 5.22). The February 2003 sequences are most related to one of the groups of sequences collected in November 2002 (clade A in figure 5.22). The inferred protein-based analysis resembles the DNA-based analysis in that only one group was represented in this sample (represented in purple in figure 5.23). The nine inferred protein sequences collected from February 2003 are all closely related to protein sequences from November 2002 (figure 5.23).

Figure 5.21. SIRV-like sequences from Ragged Hills collected in October 2002.
A. DNA sequences B. Protein sequences



In the next sample collected in April 2003, the diversity of the SIRV-like viruses increases (clades A, C, E, H, and a few outliers in figure 5.22). Forty-one SIRV-like coat protein sequences collected in this sample represent each of the three major groups (clade A, C, and E in figure 5.22). Like the DNA-based analysis, the protein-based analysis represents increasing diversity in the SIRV-like population in this sample. The sequences detected in this sample belong to four different groups of protein sequences (represented by purple, pink, turquoise, and orange in figure 5.23). The 28 unique DNA sequences translated to 22 unique inferred protein sequences (table 5.10). There were three protein

sequences from this sample that were identical to sequences from previous samples (figure 5.23b; the * represents the identical sequences).

The SIRV-like diversity measured in previous samples is maintained in the fifth sample. This sample was collected in May 2003, and 37 SIRV-like coat protein clones were sequenced from this sample (table 5.10). When these SIRV-like clones are compared to sequences from previous samples, most belong in previously established groups (clades A, C, E, and H in figure 5.22). The majority of the clones belong in the larger group (clade A) represented by sequences from November 2002, February 2003, and April 2003 (figure 5.22). Twelve unique inferred protein sequences were translated from the 18 unique DNA sequences obtained from this sample (table 5.10). The SIRV-like protein sequences belong in the largest group of protein sequences from this monitor site (purple in figure 5.23). One sequence in this sample was identical to the group of identical sequences from February and April 2003 (figure 5.28b).

The next sample does not reveal significant change in the SIRV-like population. The sixth sample yielded 76 SIRV-like coat protein sequences (table 5.10). When these sequences were compared to sequences from previous samples, major change in the SIRV-like virus population is not detected. Most of the sequences belong in the largest group of SIRV-like viruses (clade A in figure 5.22). Most of the remaining sequences belong in minor groups (clades E and F in figure 5.22). The single remaining sequence is quite distinct from the other SIRV-like sequences obtained from this sample (clade D in figure 5.22). The thirty-three unique DNA sequences translated into 28 unique inferred protein sequences (table 5.10). Most of the protein sequences belong in the largest group

of viral protein sequences detected from this site (represented in purple in figure 5.23). One remaining SIRV-like protein sequence was unique from the others detected in this sample (represented in orange in figure 5.23). Two protein sequences from this sample were identical to protein sequences detected previously (figure 5.23; the * represents the identical sequences).

In June 2003, the SIRV-like population remains similar to previous samples. This seventh time point resulted in only 25 SIRV-like coat protein clones (table 5.10). When compared to previous samples, the population shows little change from May 13, 2003. The majority of the detected sequences from this sample belong in the largest group of SIRV-like sequences (clade A in figure 5.22). Other SIRV-like sequences from this sample belong to clades C, E, and F (figure 5.22).

Thirteen unique SIRV-like protein sequences were detected from June 2003 (see table 5.10). Most of them belong in the largest group of SIRV-like coat protein sequences (represented in purple in figure 5.23). SIRV-like sequences detected from this sample were identical to sequences detected previously in this monitor site (figure 5.23; the * represents the groups of identical sequences).

The previous sample collected in June 2003 contained sequences representing four groups in the SIRV-like population (clades A, C, E, and F in figure 5.22). However, the SIRV-like diversity has decreased dramatically in July 2003. The eighth sample contained 75 SIRV-like coat protein clones (table 5.10). All of the SIRV-like sequences belong to one group, which is the major group of SIRV-like sequences (clade A in figure 5.22). There were 17 unique inferred protein sequences detected in this sample (table

5.10). The decrease in SIRV-like diversity is also evident in the protein sequences. All of the SIRV-like protein sequences from this sample belong in the largest group of viral protein sequences (represented by purple in figure 5.23).

There is more diversity in the August 4, 2003 sample when compared to the previous sample collected in June 2003. Sixty clones were sequenced from this sample resulting in 33 unique DNA sequences (table 5.10). Most of the sequences from this sample represent the largest group of SIRV-like viruses (clade A in figure 5.22). Other SIRV-like sequences belong to three groups (clades C, E, and F in figure 5.22).

The SIRV-like diversity has also increased in the protein-based analysis in August 2003. Twenty-five unique inferred protein sequences were obtained from this sample (table 5.10). The sequences represent three groups in the SIRV-like sequences. Most of the protein sequences belong to the largest group of identical SIRV-like sequences (represented in purple in figure 5.23; the * represents identical sequences).

The diversity of the SIRV-like population does not significantly change in the next sample. The tenth sample was collected on August 26, 2003, and there were 50 SIRV-like coat protein clones sequenced from this sample (table 5.10). When these SIRV-like sequences were compared to sequences obtained from previous time points, the same population structure is evident as in the earlier sample collected in July 2003. Most of the SIRV-like DNA sequences belong to the largest group of sequences detected in this monitor site (clade A in figure 5.22). Other SIRV-like sequences detected in this sample represented clades E, F, and H (figure 5.22). There were seventeen unique inferred protein sequences detected in this sample (table 5.10). The inferred protein

sequences represent one group of viruses (represented in purple in figure 5.23). One sequence from this sample was identical to the largest group of SIRV-like viruses (figure 5.23; the * represents the identical sequences).

The diversity of SIRV-like viruses decreases in the next sample. This sample was collected on October 2, 2003, and 54 clones were sequenced (table 5.10). The diversity of SIRV-like viruses in this monitor site has decreased to one group, which is the largest group of SIRV-like viruses (clade A in figure 5.22). The inferred protein sequences also reveal the decrease in diversity (represented in purple in figure 5.23).

The diversity of SIRV-like viruses increases in the twelfth sample collected on November 13, 2003. Twenty-four clones were sequenced resulting in only 10 unique DNA sequences (table 5.10). When these SIRV-like sequences are compared to sequences from previous samples, they represent three groups of SIRV-like sequences (clades A, E, and F in figure 5.22). The inferred protein analysis reveals that although the DNA sequences were separated into two groups, the amino acid sequences belong to the largest group of SIRV-like protein sequences (represented in purple in figure 5.23). More sequences were detected from the largest group of identical SIRV-like protein sequences (figure 5.23; the * represents the groups of identical sequences.

The next sample collected on January 19, 2004 failed to yield positive PCR products for the SIRV-like coat protein. However, this same source of DNA yielded positive results for the 16S rDNA PCR. This suggests that the source DNA is competent for PCR.

Figure 5.22. SIRV-like DNA sequences collected from October 2002 through June 2004 from Ragged Hills. The group colors correspond to figure 5.19.

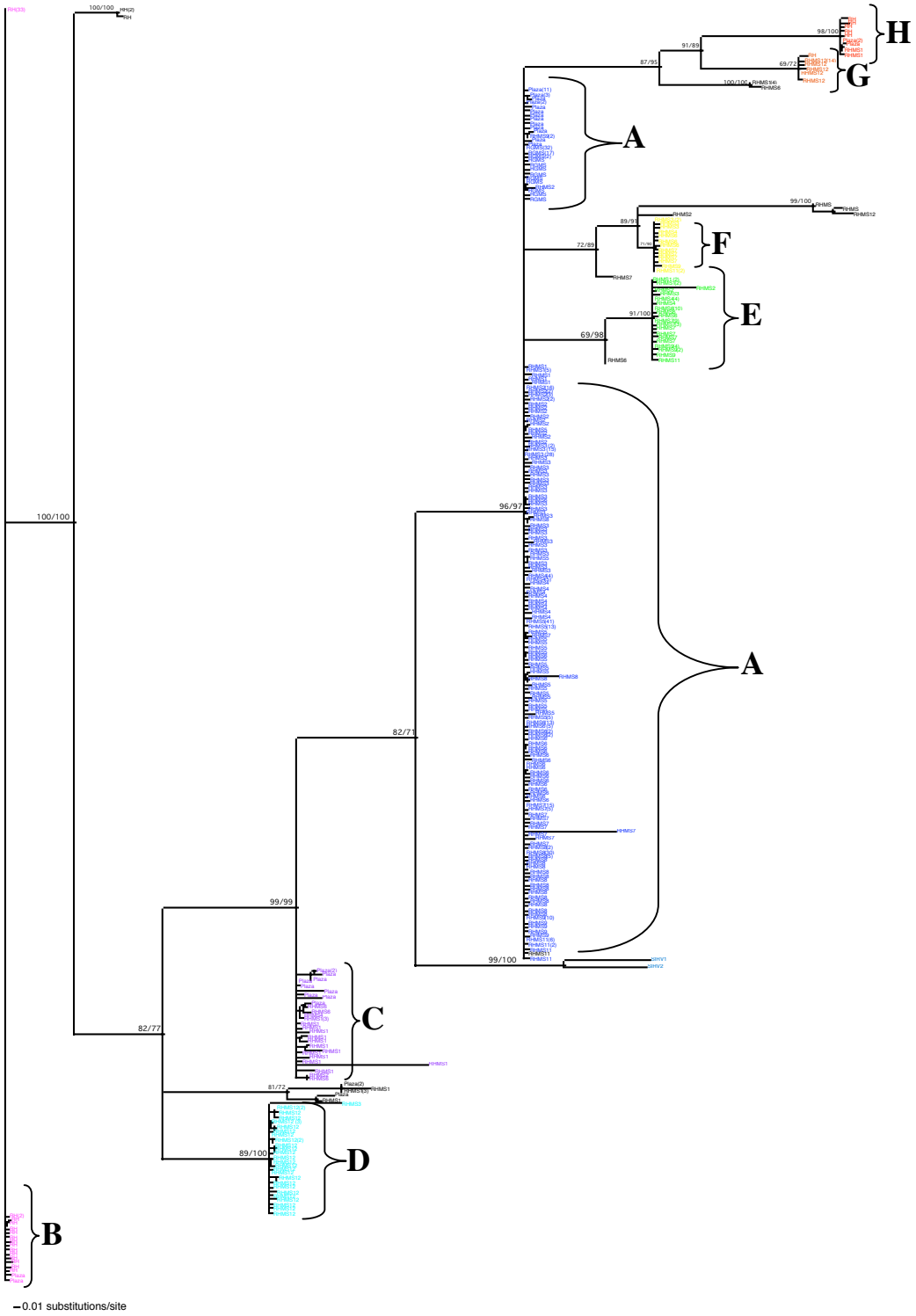
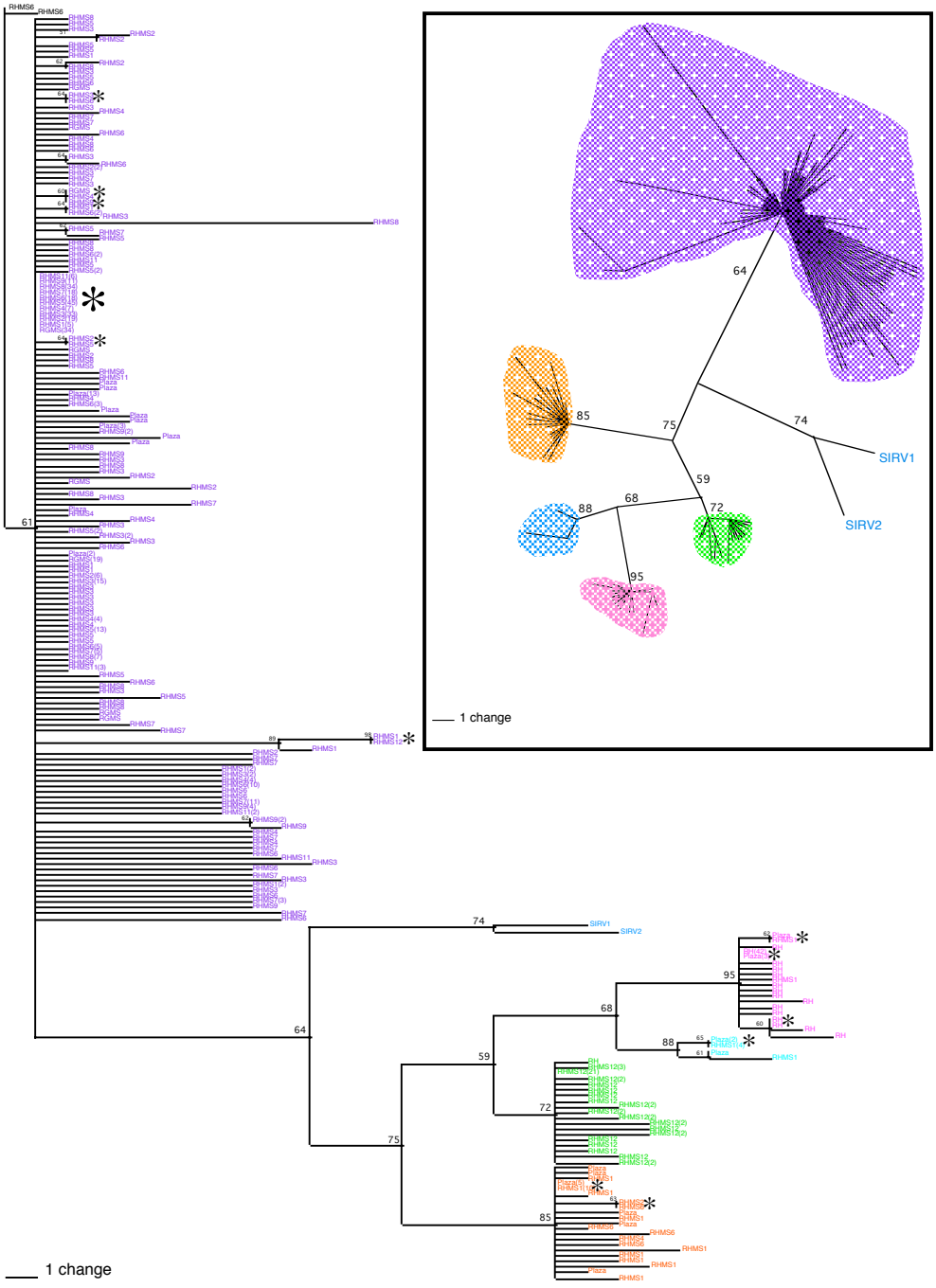


Figure 5.23. SIRV-like protein sequences collected from October 2002 through June 2004 from Ragged Hills. The * represents identical sequences. Inset: An unrooted phylogram of SIRV-like groups detected over time in the Ragged Hills monitor site.



The sample collected on April 15, 2004 contained only seven unique DNA sequences out of 14 clones (table 5.10). There has not been noteworthy change in the SIRV-like virus population for some months. When the SIRV-like sequences from April 2004 were compared to sequences from previous samples, they belong in three groups (clades A, E, and F in figure 5.22). Most of the SIRV-like DNA sequences belong to the largest group of SIRV-like sequences (clade A), and a few of the sequences belong to smaller groups (clades E and F) that were also detected in the November 2003 sample (figure 5.22). The inferred protein sequences belong in the largest group of SIRV-like protein sequences (represented in purple in figure 5.23).

The last sample begins to show change in the SIRV-like population. The sample was collected on June 3, 2004 and 48 clones were sequenced (table 5.10). These SIRV-like sequences composed two new groups that had not been previously detected (clades D and G in figure 5.22). Most of the sequences from this time point composed a fairly large group (clade D) of SIRV-like sequences, while the other group is much smaller (clade G) (figure 5.22). Surprisingly, there were no sequences detected that belonged to previously established SIRV-like groups. The inferred protein sequences revealed change, as did the DNA sequences (figure 5.23). A new group of SIRV-like protein sequences (represented in green in figure 5.23) was detected in this sample along with one protein sequence that belonged to the largest group of SIRV-like protein sequences (represented in purple in figure 5.23).

In summary, during the course of two years eight clades of SIRV-like viruses were detected (figure 5.22). One of these clades (clade A) was maintained throughout

most of the study (excluding the October 2002 sample). Clades B, C, D, E, F, G, and H were only present transiently (figure 5.8 and 5.22). The complete phylogenetic analysis of SIRV-like coat protein sequences included analysis of sequences after every sample. For a step-by-step view of the changes in the SIRV-like population over time, see Appendix C.

Detailed analysis of SIRV-like population dynamics. After almost two years of regularly sampling and sequencing 663 SIRV-like coat protein clones from this site, we detected eight different groups of SIRV-like viruses (figure 5.22). There are eight groups of SIRV-like viruses based on DNA sequence; however, there are only five groups of SIRV-like viruses based on inferred amino acid sequence (figure 5.23). The SIRV-like population is not as diverse as the SIRV-like population in Crater Hills (see Chapter 3); however, it is more diverse than the SIRV-like viruses in Rabbit Creek (see Chapter 5).

Using Modeltest version 3.6 (Posada *et. al.* 1998), the nucleotide substitution model for maximum likelihood estimations was determined to be a transversional model with heterogeneity between sites (gamma=0.9389), base frequencies A=0.3368, C=0.2082, G=0.1758, and T=0.2792, and a substitution model A-C=0.3905, A-G=C-T=1.8530, A-T=0.7612, C-G=0.9023, and G-T=1.000.

The SIRV-like population contained one large group of 203 identical sequences, which comprised 30.6% of the total population (table 5.11). This large group (clade A in figure 5.22) was maintained in most of the samples collected from this monitor site, with the exception of the samples collected in October and November 2002. Since this population was present in the remaining samples, it is surprising it was undetectable in

the January 19, 2003 sample in which SIRV-like coat proteins were not detected (table 5.6). Another group (clade A in figure 5.22) of identical viral sequences that was present during the majority of the sampling time was composed of 77 sequences, which consisted of 11.5% of the total population (table 5.11).

Table 5.11. Ragged Hills SIRV-like sequences illustrating the groups of identical sequences, how many times a group this size was detected, the percent of the SIRV-like population, and what samples the group contained.

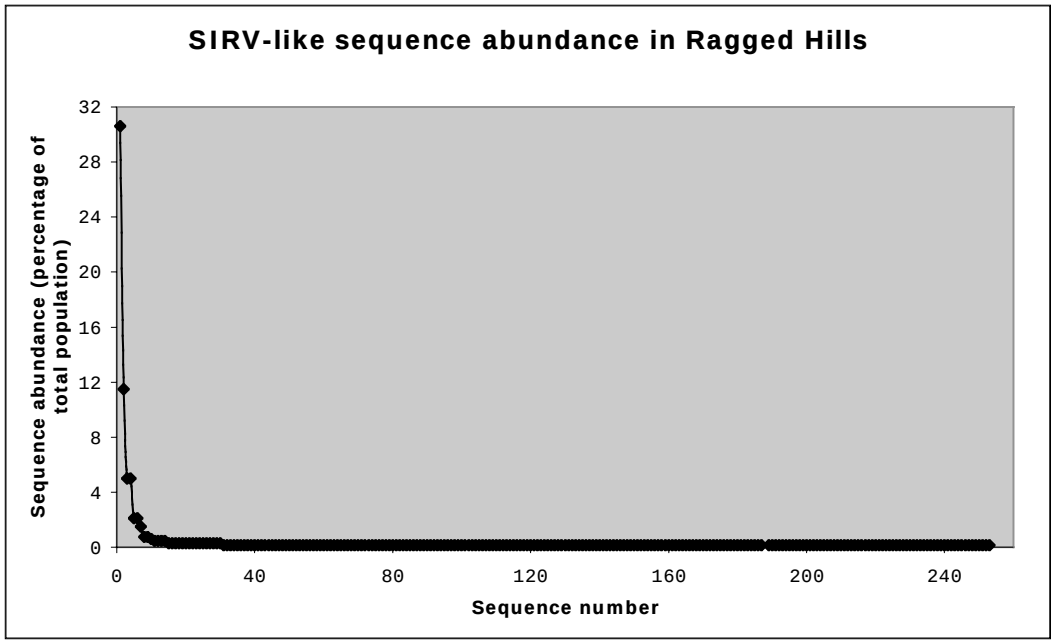
RAGGED HILLS SIRV-like SEQUENCE ABUNDANCE			
Identical Sequences	Times Detected	Percent of SIRV-like Population	Samples[a]
203	1	30.6%	RGMS;RHMS1—RHMS9;RHMS11—12
77	1	11.5	Plaza;RGMS;RHMS1—RHMS9; RHMS11—RHMS12
33	1	5	RH;Plaza
33	1	5	RHMS1;RHMS3;RHMS4;RHMS6; RHMS7;RHMS9;RHMS11
14	1	2.1	RHMS12
14	1	2.1	Plaza;RHMS4;RHMS6
10	1	1.5	RHMS1;RHMS7;RHMS9
5	2 unique[b]	0.75% each[c]	
4	1	0.6	
3	4 unique	0.45% each	
2	16 unique	0.3% each	
1	222 unique	0.15% each	

[a]sequences from these samples were present in the indicated groups
[b]each of the smaller groups were detected this many times, however, they are unique
[c]each of the unique groups makes up this percentage of the population

There were two distinct groups each containing 33 identical sequences of SIRV-like coat proteins, and each of them comprised 5% of the total population of SIRV-like

viruses (table 5.11). One of these groups of 33 identical sequences was in clade B and the other group belongs to clade E (figure 5.22). Other groups of identical sequences ranged from two to 14 identical sequences comprising 0.3% to 2.1% of the total population (table 5.11). Similar to Crater Hills and Rabbit Creek, the unique sequences made up a large portion of the total population. The unique sequences comprised 33% of the total population of SIRV-like coat protein sequences. The population structure of the SIRV-like viruses is similar to the other monitor sites (figure 5.24).

Figure 5.24. A sequence abundance plot for SIRV-like sequences from Ragged Hills. The left end of the x-axis represents most commonly detected sequences, and the right end of the x-axis represents the most rare sequences. Most of the SIRV-like sequences in Ragged Hills are detected only once.



Over the course of sampling the Ragged Hills monitor site, there were 208 unique protein sequences (table 5.10). Twelve protein sequences were detected more than once; but only the largest group of identical sequences was detected throughout most of the

samples (figure 5.23; the large * represents the largest group of identical protein sequences detected in Ragged Hills).

There were eight viral clades of SIRV-like viruses based on DNA sequences, but only five groups were present based on protein sequences (figures 5.19 and 5.20). The largest group of protein sequences detected in this monitor site (purple in figures 5.20 and 5.23) included sequences from clades A, E, F, and H. A protein group (pink in figures 5.20 and 5.23) detected in early samples (October 2002, November 2002, and April 2003) represented sequences from clade B and a few outlying sequences. Another minor protein group (turquoise in figures 5.20 and 5.23) represented outlying DNA sequences collected in November 2002 and April 2003.

The remaining protein groups were not consistently detected in this monitor site. A viral protein group detected sporadically (orange in figures 5.20 and 5.23) contained sequences from clade C and some outlying sequences. Finally, the protein group that appears to be dominant at the conclusion of this research (green in figures 5.20 and 5.23) contains sequences from clades D and G.

Using the program MEGA version 3 (Kimura *et. al.* 2004), the changes in the SIRV-like sequences were calculated (table 5.12). When comparing all of the sequences, 86% of the bases in the SIRV-like coat protein sequence remained unchanged (table 5.13). Transversions occurred somewhat more (30) compared to transitions (27). Eleven changes occurred in the first position of the codon; there were five transitions and six transversions (table 5.13). Only seven changes occurred in the second position with two

transitions and five transversions (table 5.13). A total of 38 changes occurred in the third position; there were 19 transitions and 19 transversions (table 5.13).

Table 5.12. The numbers of transitions and transversions in SIRV -like sequences in each sample collected from the Ragged Hills monitor site.

SAMPLE DATE	IDENTICAL BASES	TRANSITIONS			TRANSVERSIONS		
		1st	2nd	3rd	1st	2nd	3rd
10/17/2002	392/405	2	1	4	1	1	4
11/05/2002	346/405	6	2	20	7	4	20
02/17/2003	403/405	1	0	0	0	1	0
04/24/2003	334/405	7	3	24	7	5	24
05/06/2003	373/405	4	2	11	2	3	11
05/13/2003	388/405	2	1	6	1	2	5
06/19/2003	371/405	3	2	11	3	3	11
07/14/2003	402/405	1	0	1	0	1	0
08/04/2003	363/405	4	3	14	3	4	14
08/26/2003	373/405	2	2	11	2	4	11
10/02/2003	401/405	0	1	0	0	2	1
11/13/2003	373/405	2	2	10	3	4	11
04/15/2004	374/405	2	2	10	2	4	11
06/03/2004	391/405	2	1	4	1	2	4

Table 5.13. The numbers of transitions and transversions for SIRV-like coat protein sequences collected from Ragged Hills.

Sample Range	Position	Identical Base Pairs	Ts Pairs	Tv Pairs
Oct 2002-June 2004	*Overall*	*348/405*	*27*	*30*
	1st	124/135	5	6
	2nd	127/135	2	5
	3rd	97/135	19	19

Ts=transitions; Tv=transversions

Comparing SSV-like and SIRV-like population dynamics

Unlike the SSV-like population, we were able to detect SIRV-like sequences at every sample, excluding one, collected over the two years of this research (tables 5.3 and 5.10). It appears that SIRV-like viruses are more permanent residents in this hot spring. There were 663 SIRV-like coat protein clones compared to 398 total SSV-like clones

(285 3F/4R clones and 113 7F/8R clones) (tables 5.3 and 5.10). However, the SIRV DBPF/R primer set was 100X more sensitive that the SSV 3F/4R primer set and 1,000,000X more sensitive than the SSV 7F/8R primer set.

The SSV-like population revealed diversity even though it was not detected in all samples (figure 5.8). The SIRV-like population was much more diverse than the SSV-like population (eight SIRV-like clades compared to five SSV-like 3F/4R clades) (figures 5.8 and 5.19). However, there was one consistency between the two viral populations, each virus population appeared to be dynamic, changing over the course of the monitoring period. Occasionally both virus populations were relatively stable from sample-to-sample.

The SIRV-like population appeared, at times, to respond to environmental change. The sample collected in November 2002 reveals new virus groups not previously detected in this hot spring. Perhaps the new groups of viral sequences could be representative of the drop in pH from 3.7 in October 2002 to 2.7 in November 2002 (table 5.10). Also, in July 2003 the pH increased from 2.9 in June 2003 to 3.1 in July 2003 and the diversity of SIRV-like viruses decreased from two groups to only one group. However, a change in an environmental parameter was not necessary to detect change in the SIRV-like population. Although there are examples of environmental change followed by the detection of a new viral group, I do not think there is enough evidence to support temperature and pH changes are significant enough to change a virus population in nature. The viral populations within the hot springs in YNP appear to be changing regardless of a change in an environmental parameter.

Conclusions

This monitor site proved to be a very dynamic thermal area. This site was sampled continuously for approximately two years. During that time, the water chemistry was monitored, the dominant archaeal population was determined, and changes in the population structures of two viruses were continually investigated.

The water chemistry was characterized by high chloride and high arsenic concentrations (figure 5.3). Surprisingly, the water chemistry did not significantly change over time, and most of the observed changes could be contributed to seasonal fluctuations. This monitor site is a vapor-dominated system without visible inflow or outflow. This site is also very turbulent due to out-gassing and likely to be well mixed. In July 2003, there was a spike in sulfate, aluminum, iron, and molybdenum concentrations (table 5.1); it will be interesting to determine whether these are cyclic events or an isolated occurrence. This was the only monitor site that antimony and molybdenum were detected (table 5.1). The presence of these metals is indicative of undergoing rapid demineralization. The pH and temperature of this monitor site were relatively consistent. The pH was measured between 2.6 and 3.7, and the temperature was measured at 57-78°C. During this research, we noticed a general cooling of this monitor site. Even though this monitor site is acidic and thermal, it is chemically unique with respect to the monitor sites in Crater Hills (see Chapter 3) and Rabbit Creek (see Chapter 4).

By measuring the water chemistry at this hot spring, we learned it is a unique system dominated by high chloride and sodium compared to the other two monitor sites.

This monitor site also contained the highest concentrations of most metals that were analyzed in this study. This result is indicative of the young age of this monitor site. Surprisingly, the water chemistry of this hot spring did not significantly change over the two years of this study.

This site maintained a fairly constant, low complexity archaeal population dominated by *Stygiolobus*-like sequences. The full-length (~1,400bps) 16S rDNA sequences showed high identity to published 16S rDNA sequences (table 5.2). All of the sequences that matched *Stygiolobus azoricus* were 95-99% identical to the published *S. azoricus* 16S rDNA gene (table 5.2). Of 269 *Stygiolobus*-like environmental 16S rDNA clones, 237 matched the published *S. azoricus* 16S rDNA sequence with 98-99% identity. The archaeal population detected in clone libraries established from this monitor site was not as diverse as the Crater Hills (see Chapter 3) or Rabbit Creek (see Chapter 4) monitor sites. In fact, there were only four sequences (*Stygiolobus*, *Sulfolobus*, *Metallosphaera*, and *Thermocladium*) detected from this hot spring during the course of this research (table 5.2).

Surprisingly, the other 16S rDNA primer set (348Fa/UA1406R) never resulted in positive PCR products from this monitor site. It is especially perplexing because the dominant member of this archaeal community appears to be *Stygiolobus*-like organisms, and the published *Stygiolobus azoricus* 16S rDNA gene does not match the primer 20F, but matches 348Fa (table 3.14). The dominant organisms in this monitor site are likely closely related to *Stygiolobus* because the percent identity match of most the full-length

16S rDNA genes are very high (98-99%). These percentages are considered quite high when comparing 16S rDNA genes, and therefore most likely represent *Stygiolobus*.

The 16S rDNA sequences most commonly detected in this hot spring over time were *Stygiolobus*-like sequences. *Stygiolobus azoricus* was originally isolated from solfataric fields on São Miguel Island, Azores (Segerer *et. al*. 1991). This organism grows obligately chemolithotrophically by reduction of S^0 with H_2, and it was the first strictly anaerobic member of the order *Sulfolobales*. *Stygiolobus* grows at temperatures ranging from 57-89°C, with optimal growth occurring around 80°C. The pH suitable for growth ranges from 1.0 to 5.5, with optimal growth at pH 2.5 to 3. At the time of isolation of this organism, its growth range seemed to be restricted to the solfataras on São Miguel Island because the researchers were unable to isolate it from YNP, Iceland, Italy, and numerous other thermal areas worldwide (Segerer *et. al*. 1991). However, based on this data, it is quite possible that *Stygiolobus* is not restricted to the Azores. The growth and pH ranges for this organism are concurrent with the temperatures and pHs measured in Ragged Hills (figure 5.1). Current research is aiming towards culturing *Stygiolobus* from the Ragged Hills monitor site.

Similar to the other monitor sites (see Chapters 3 and 4), *Sulfolobus*-like sequences are not commonly detected in this site. The 20F/UA1406R primer resulted in the detection of three *Sulfolobus*-like sequences from Ragged Hills (table 5.2). Only two of seven samples contained clones resembling *Sulfolobus* (table 5.2). However, the environmental sequences matching *Sulfolobus* were 98% identical to the *S. solfataricus* 16S rDNA gene. The results concerning the host population lead to serious implications

concerning the virus population. We are easily detecting *Sulfolobus* viruses, however, we are unable to detect large numbers of *Sulfolobus* species based on 16S rDNA sequence. There are many potential reasons for this apparent anomaly. First, the virus sequences may be easily detected because they are present in a higher copy number than the *Sulfolobus* 16S rDNA gene. It has previously been shown that SSV1 is present at high copy numbers in its host organism in culture (Yeats 1982). Or, perhaps SSVs have a broader host range than just *Sulfolobus*. This would explain why we retrieve SSV-like and SIRV-like sequences, but few *Sulfolobus* sequences. It is also a possibility that these viruses may infect hyperthermophilic bacterial organisms. Interestingly, previous research has indicated archaeal organisms are not large constituents of the microbial population in hot springs within YNP (Spear *et. al.* 2005; Blank *et. al.* 2002; Hugenholtz *et. al.* 1998). However, these studies were not conducted in high temperature acidic environments. We consider alternative hosts unlikely since *Sulfolobus*-like sequences were the only archaeal 16S rDNA sequences common to all three monitor sites.

SSV-like sequences were only detected in six samples using the 3F/4R primer set, and only two samples using the 7F/8R primer set. The fact that primer set 3F/4R yielded more dependable results is consistent with the results in the other two monitor sites (see Chapters 3 and 4). We hypothesized the 7F/8R primer set, which should amplify the SSV-like coat proteins, would amplify SSV-like sequences in most samples. One would hypothesize the structural proteins of a virus to be very conserved, but since we know so little about these viruses ORF b812 (in SSV-RH) may be more conserved than the coat proteins. The most likely explanation for this result is the SSV 3F/4R primer set was

10,000X more sensitive on purified viral template than the SSV 7F/8R primer set. Potentially, if the 7F/8R primer set was as sensitive, the SSV-like coat protein sequences would have been detected more frequently.

The entire SSV-like population was undetectable after May 2003 even with the universal SSV 3F/4R primer set (table 5.3). There was one instance where nine SSV-like 3F/4R clones were obtained from the November 2003 sample. Perhaps the temperature drop from 70°C on May 6, 2003 to 66°C on May 13, 2003 was enough to eliminate the SSV-like population or the viral population decreased below our detection limit (<0.1pg/PCR reaction). However, the temperature of the monitor site slowly recovered to the mid-70's within a couple of months, but the SSV-like population was still undetectable. Perhaps the viral population (and the host population) was unable to recover following the decrease in temperature. We do not believe the absence of SSV-like sequences is the result of technical issues. There were numerous attempts to PCR amplify the SSV-like sequences, and the same source DNA was used for the amplification of SIRV-like and 16S rDNA sequences. It is possible the SSV-like population decreased below the detection limit of our primers. In order to determine whether this was the case, primers with greater sensitivity would need to be designed.

By extracting total environmental nucleic acids, it is difficult to determine whether the amplified SSV-like sequences are from an episomal form of the viral genome or from integrated form of the virus. However, by translating the sequences, we could limit the number of integrated viral sequences in the phylogenetic methods. This distinction is important because the ability of SSVs to integrate into their host cell's

chromosome (Muskhelishvili *et. al.* 1993; Wiedenheft *et. al.* 2004). In our clone libraries, some of the translated SSV-like ORFs contained numerous stop codons resulting in an interrupted ORF, and therefore, probably a non-functional viral sequence. There were a total of 17 sequences that were considered interrupted ORFs detected from our samples collected from Ragged Hills (table 5.3).

Some of the SSV-like non-functional ORFs are closely related to the functional ORFs (figure 5.16). Each of the SSV-like viral groups detected from this monitor site contains both functional and non-functional SSV-like ORFs (figure 5.16). Most of the non-functional sequences are very similar to the functional ORFs except that an error is made early in the sequence resulting in a premature stop codon. However, the remaining sequence is almost identical to the functioning ORFs. It is unlikely these errors are the result of PCR or sequencing errors. SIRV-like sequences from this monitor site never resulted in interrupted ORFs. If the errors in the SSV-like sequences were the result of PCR or sequencing errors, one would expect to see the same types of errors in the SIRV-like sequences. Perhaps these types of sequences represent a more recent event of integration into the host chromosome, and not as many errors have accumulated because of the recent integration.

As was the case in Rabbit Creek (see Chapter 4), it is more difficult to predict the number of SSV-like groups based on coat protein analysis. The 7F/8R primers were designed to amplify the viral coat proteins, but they were found to be unreliable in this research. Only two samples yielded a positive product for the coat proteins (table 5.3). The two samples were collected at the very beginning of this research (October and

November 2002) and then not detected again. Unlike the Rabbit Creek SSV-like coat protein sequences that were very distinct from each other, the sequences from Ragged Hills were closely related. The molecular analysis revealed that 61% of the bases in the coat protein sequence remained identical through the two samples. Perhaps the coat protein sequences from this monitor site were more similar to each other because they were sampled within one month of each other. In contrast, the coat protein sequences from Rabbit Creek were sampled over one year apart. However, it is difficult to make any conclusions concerning the SSV-like groups present in Ragged Hills because of the lack of coat protein sequences.

As opposed to the SSV-like sequences that were only detected in six samples, the SIRV-like sequences were detected in every sample, excluding one (table 5.10). The primers designed to amplify the entire coat protein (405bps; 135aa) of SIRV-like viruses reliably detected this gene. SIRVs have not been shown to integrate into their host cell's chromosome; instead they are maintained within the cell as a stable carrier state (Prangishvili *et. al.* 1999). Over the course of this study, there were not any interrupted ORFs detected among the SIRV-like coat protein sequences obtained from this hot spring. This supports our claim that the interrupted SSV-like ORFs may be from integrated copies of the virus (as discussed above), instead of sequencing or PCR errors.

Changes in the SIRV-like population are detected (figure 5.19). The third sampling point (February 2003) reveals a decrease in the diversity of SIRVs from numerous groups to only one (figure 5.19). The next sample (April 2003) reveals an explosion of diversity in the SIRV-like population (figure 5.19). This diversity is

maintained through several samples (figures 5.19). The diversity is reduced to a single group in the eighth sample (July 2003) from this monitor site (figure 5.19). In the next samples the SIRV-like population diversifies again, revealing numerous groups of SIRV-like viruses (figure 5.19). However, this diversity is short-lived because in the very next sample (October 2003) the viral diversity drops again to only one group of viruses (figure 5.19). The next sample (November 2003) reveals two groups of SIRV-like viruses (figure 5.19), and the detection of a new group occurs in the final sample (figure 5.19).

There were numerous groups of SIRV-like viruses detected over time within this monitor site. Unlike the SSV-like viruses, which had three distinct groups of viruses, the SIRV-like viruses are not as simple. It appears there are eight major groups of SIRV-like viruses based on the sequence of the coat protein (figure 5.22).

In conclusion, we discovered the viral populations within this hot spring are dynamic and diverse. The SSV-like population was only detected through November 2003; however, the population changed and showed diversity within that time. The SIRV-like population was detected in all, but one sample, and was shown to be very diverse.

CHAPTER 6

COMPARISON OF THREE ACIDIC HOT SPRINGS IN YELLOWSTONE NATIONAL PARK

Introduction

This research involved studying two virus populations (SSVs and SIRVs), archaeal host populations, and water chemistry over time in three distinctive, geographically isolated, acidic hot springs in Yellowstone National Park (YNP). This is the first type of this study conducted in any natural environment over time. The overall goal of this research was to investigate the interaction of time and environment on virus populations. However, this research has many other potential impacts as well. We know very little about the life in the acidic hot springs within YNP. This research will help expand our knowledge about the virus populations, archaeal host populations, and the water chemistry of the three monitor sites. Understanding viral diversity and how the viral populations change over time will help in further understanding the microbial community as a whole in these environments. This research may also contribute to understanding how these viruses have evolved over time. Even though this research encompassed only a two year period, there were many changes obvious in such a short time.

It was hypothesized that simultaneous monitoring of viral populations, host populations, and water chemistry would lead to new insights in how these environments operate. The 16S rDNA sequences obtained from this research helps further define the

archaeal communities in these hot springs. In many cases, assumptions concerning the dominant microbial population were challenged by this analysis. The water chemistry data collected for this research was aimed at achieving a better understanding of how these acidic hot springs fluctuate over time with respect to their chemistry. This was the first time that any of the three sites had been monitored continually over time for virus populations, archaeal populations, and water chemistry.

Comparison of water chemistry in three acidic hot springs

The water chemistry from each of the three monitor sites was continuously monitored for two years. Each monitor site was chemically unique. Originally, these hot springs were chosen as monitor sites because they all had the temperature (70-95°C) and pH (1-5) ranges that are appropriate for the growth of *Sulfolobus* species. Due to sampling logistics, the samples for chemical analysis at all three sites could not always be collected on the same day; therefore, comparisons are based on the closest sampling times (table 6.1). An overview of the water chemistry in each site is illustrated in figure 6.1.

Each monitor site displayed a wide variety of water chemistry. Of the three sites, chloride was highest in the Ragged Hills monitor site (figure 6.2a). The average chloride concentration in Ragged Hills was 611mg/L compared to only 1.8mg/L in Crater Hills and 3mg/L in Rabbit Creek (table 6.2). High sulfate levels were characteristic of the Crater Hills monitor site over time (figure 6.2b). The average sulfate concentration in Crater Hills was 1,022mg/L (table 6.2). This compares to 87mg/L sulfate in Rabbit

Creek and 135mg/L sulfate in Ragged Hills (table 6.2). Dissolved organic carbon (DOC) fluctuated in each monitor site over time (figure 6.2c). However, based on the average DOC concentration over time, the Crater Hills (5.74mg/L) and Rabbit Creek (5.3mg/L) monitor sites contained twice as much DOC as the Ragged Hills monitor site (2.5mg/L) (table 6.2).

Table 6.1. The dates of chemical analysis for each monitor site. The sample numbers correspond to figure 6.2a-r. RHMS=Ragged Hills monitor site; RCMS= Rabbit Creek monitor site; CHMS=Crater Hills monitor site.

Sample Number	Water Chemistry Samples Collected		
	RHMS	**RCMS**	**CHMS**
1	02/17/03	1/20/03	1/3/03
2	4/24/03		2/27/03
3	5/6/03		
4	5/13/03		5/30/03
5	6/19/03	6/24/03	7/18/03
6	7/14/03	7/25/03	8/4/03
7	8/4/03	8/26/03	9/11/03
8	10/2/03	10/2/03	10/23/03
9	11/13/03	11/13/03	11/24/03
10	1/19/04	1/19/04	2/19/04
11	4/15/04	3/4/04	4/15/04
12	6/3/04	6/3/04	6/14/04

The nutrient concentrations varied in the monitor sites over time. Nitrogen (nitrate and nitrite) was not detected in any of the monitor sites (table 6.2). Average ammonia concentrations were three times higher in the Crater Hills monitor site (14.8mg/L in Crater Hills compared to 5.3mg/L and 0.85mg/L in Rabbit Creek and Ragged Hills, respectively) (table 6.2). The ammonia in Crater Hills and Rabbit Creek appeared to fluctuate over time, as opposed to the low, constant levels in the Ragged

Hills monitor site (figure 6.2d). Phosphorus concentrations in the three monitor sites were constantly low (table 6.2). Interestingly, the phosphorus concentrations in Crater Hills and Rabbit Creek spiked in the sixth sample from each site (figure 6.2e). This corresponds to the July 25, 2003 sample collected from Rabbit Creek and the August 4, 2003 sample from Crater Hills (table 6.1). The spike in phosphorus did not occur in all the monitor sites even though a sample from Ragged Hills was collected the same day as the Crater Hills sample (table 6.1).

Figure 6.1. A comparison of the average water chemistry values measured in the three monitor sites over time. For more details, see table 6.2. A. Crater Hills, B. Rabbit Creek, and C. Ragged Hills. All values are mg/L.

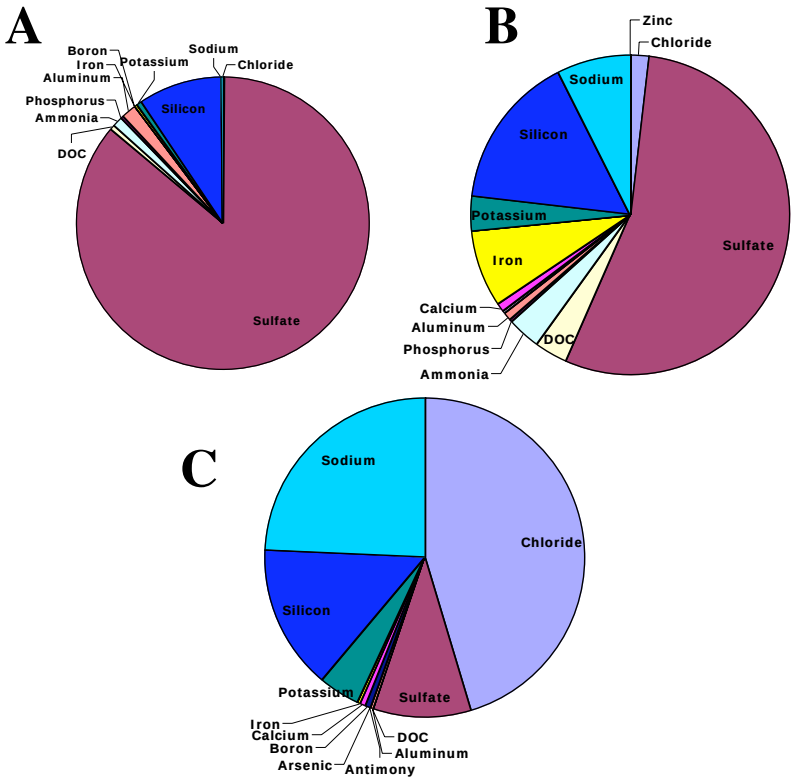


Table 6.2. The overall averages of water chemistry analysis from each of the three monitor sites. All values are mg/L. CHMS= Crater Hills monitor site, RCMS=Rabbit Creek monitor site, and RHMS=Ragged Hills monitor site.

	CHMS	RCMS	RHMS
Chloride	1.8	3	611
Sulfate	1022	87	135
DOC	5.74	5.3	2.5
Ammonia	14.8	5.3	0.85
Nitrogen	0	0	0
Phosphorus	0.47	0.7	0.203
Aluminum	20.8	1.4	1.9
Antimony	0	0	0.125
Arsenic	0.01	0.04	1.01
Barium	0	0	0
Boron	1.4	0	7.9
Cadmium	0	0	0
Calcium	0	1.5	5.8
Copper	0	0	0
Iron	1.47	12.2	2.7
Lead	0.012	0	0
Magnesium	0	0	0
Manganese	0.02	0.08	0.03
Mercury	0	0	0
Molybdenum	0	0	0.18
Nickel	0	0	0
Potassium	9	5.8	59
Selenium	0	0	0
Silicon	108	24.9	198.5
Sodium	2.8	11.7	326
Zinc	0.07	0.13	0.04

The chemical uniqueness of each monitor site is further illustrated in the dissolved metals detected in each site. Seven metals (barium, cadmium, copper, magnesium, mercury, nickel, and selenium) analyzed for in the chemical analysis were not detected (table 6.2). Aluminum in Crater Hills appears to increase and decrease cyclically (figure 6.2f). The aluminum concentrations increase during the summer months (May—August 2003), decrease during the fall and winter months (September 2003—February 2004),

and begin to increase again in the spring (April—June 2004). This correlates with lower water levels in the summer and higher water levels in the winter months at the Crater Hills monitor site. Aluminum in the Ragged Hills monitor site (1.9mg/L) was similar to the aluminum measured in Rabbit Creek (1.4mg/L), but a fluctuation over time was not apparent in either site (figure 6.2f). One of the two metals only detected in the Ragged Hills monitor site was antimony (figure 6.1g and table 6.2). The Norris Geyser Basin is known for its high concentrations of arsenic, and indeed, the Ragged Hills monitor site contained the highest levels of arsenic (1.01mg/L arsenic in Ragged Hills compared to 0.01mg/L and 0.04mg/L in Crater Hills and Rabbit Creek, respectively) (figure 6.2h). Both of these results correlate with the fact the Ragged Hills monitor site is a younger thermal site, and undergoing rapid de-mineralization.

The differences between the three monitor sites are further illustrated with boron, calcium, and iron concentrations. Boron was not detected in the Rabbit Creek monitor site, but was detected in both the Crater Hills and Ragged Hills sites (table 6.2). Boron was present in higher concentrations in Ragged Hills (7.9mg/L) than it was in Crater Hills (1.4mg/L), but it remained relatively stable over time in both monitor sites (figure 6.2i). Calcium was detected consistently and at the highest levels in the Ragged Hills monitor site (5.8mg/L), and it appeared to fluctuate at low levels in the Rabbit Creek monitor site (figure 6.2j). The fluctuation of calcium in Rabbit Creek was not correlated with pH change, temperature change, or seasonal change. Iron was detected in all of the three monitor sites in each sample (figure 6.2k). Previous research had indicated that the Rabbit Creek area contained high concentrations of iron. However, except for the sample

collected on July 25, 2003 from Rabbit Creek (45.5mg/L), the iron levels in all three sites are comparable (figure 6.2k).

Figure 6.2. The water chemistry values measured at the Crater Hills (CHMS), the Rabbit Creek (RCMS), and the Ragged Hills (RHMS) monitor sites. The sample numbers correlate with table 6.1. All values are measured in mg/L.

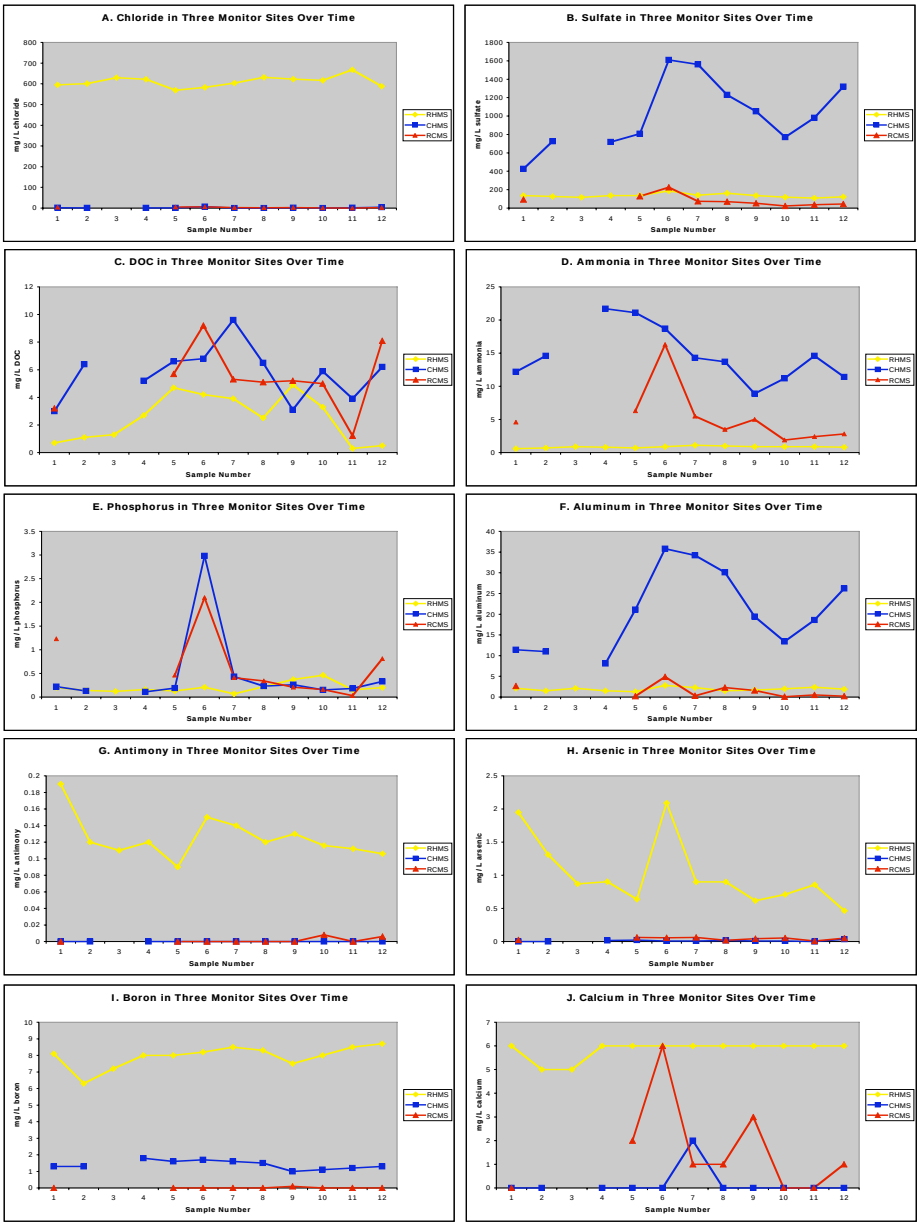
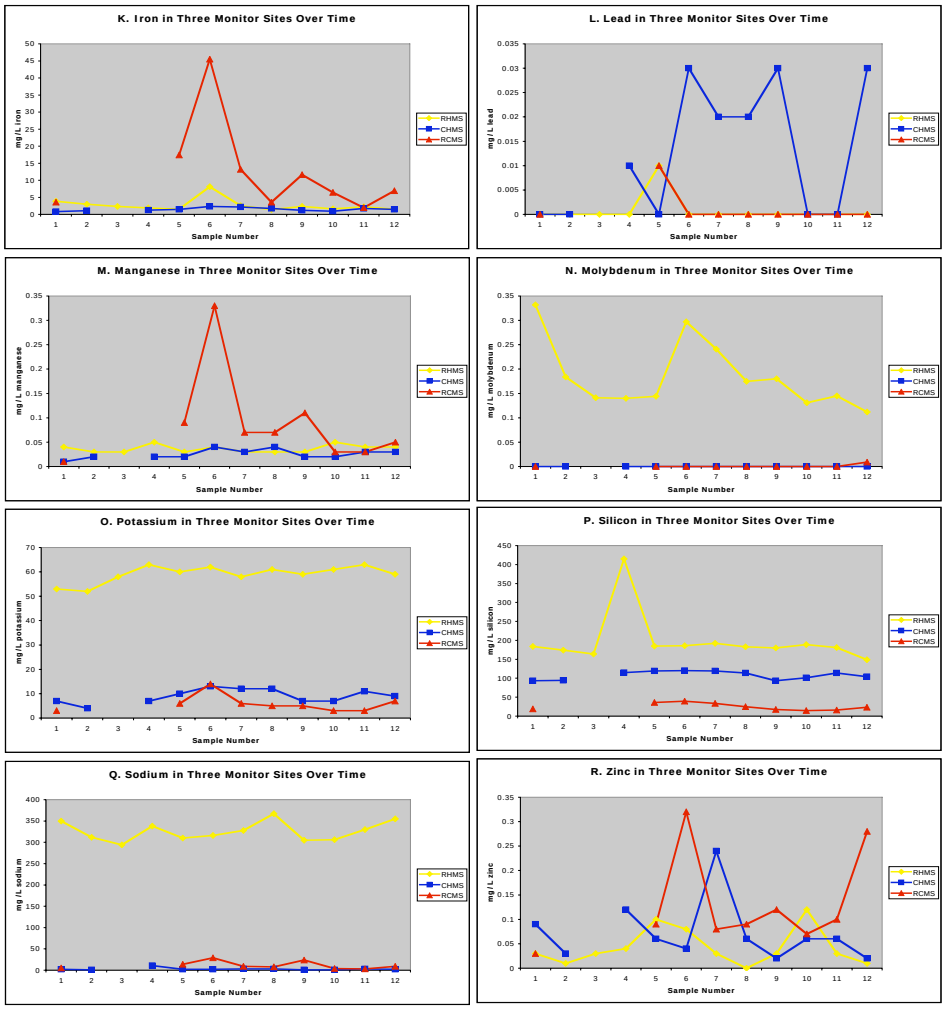


Figure 6.2 (cont).



Lead, manganese, and molybdenum were detected in the monitor sites at very low levels (table 6.2). Lead was constantly detected in Crater Hills over time albeit very low levels (0.012mg/L), and it was only detected once in Ragged Hills and Rabbit Creek (figure 6.2l). It is interesting to note the samples containing the lead in Ragged Hills (June 19, 2003) and Rabbit Creek (June 24, 2003) were collected within one week of each other (tables 6.1 and 6.2). Manganese was detected in all the monitor sites at low

levels (figure 6.2m). The manganese levels in Rabbit Creek (0.08mg/L) were slightly higher than the other two sites (0.02mg/L and 0.03mg/L in Crater Hills and Ragged Hills, respectively). One sample collected from Rabbit Creek on July 25, 2003 contained three times more manganese than other samples (figure 6.2m). However, this does not correlate with pH or temperature changes and the spike was not detected again. Molybdenum (along with antimony) was only detected in Ragged Hills (table 6.2). The molybdenum concentrations appear to be fairly unstable in the Ragged Hills monitor site with many fluctuations (figure 6.2n). Again, this could be contributed to the young age of this thermal site.

The high levels of potassium, silicon, and sodium in the Ragged Hills monitor site indicate the young age of this thermal feature. Potassium was detected in all of the monitor sites over time (figure 6.2o). The highest levels were detected in Ragged Hills (59mg/L), but the levels in all sites remained constant over time (figure 6.2o). The Ragged Hills monitor site also contained the highest levels of silicon (198.5mg/L) and sodium (326mg/L) (figures 6.2p and 6.2q). The silicon concentrations remained stable in each of the monitor sites (figure 6.2p). Sodium appeared to spike every five months in Ragged Hills (figure 6.2q). Zinc was also detected in each of the three monitor sites at low levels (figure 6.2r). However, over time, zinc concentrations were the highest in the Rabbit Creek monitor site (0.13mg/L) (table 6.2).

In summary, the chemical composition of each of the three monitor sites is unique in comparison to each other (figure 6.1). The Crater Hills monitor site contained the highest concentrations of sulfate, ammonia, aluminum, and lead (table 6.2). The Rabbit

Creek monitor site contained the highest levels of phosphorus, iron, manganese, and zinc (table 6.2). Finally, the Ragged Hills monitor site contained the highest concentrations of chloride, antimony, arsenic, boron, calcium, molybdenum, potassium, silicon, and sodium (table 6.2). In short, each of the monitor sites is characterized by a high temperature and low pH; however, chemically they are very distinct from each other.

Although a lot was learned about these acidic hot springs, this water chemistry analysis was not exhaustive. There are numerous other chemical components that could be analyzed to give a better understanding of each monitor site. For example, analyzing chemical speciation and gasses could be useful in learning more about these environments.

Water chemistry comparison with YNP monitor sites

Water chemistry in Lassen National Park. It is illustrative to compare the water chemistry of these three monitor sites with another thermal environment. Water chemistry was measured in Lassen National Park (LNP), California. The chemical analysis was conducted in the same way as the samples collected from YNP. The sample from Devil's Kitchen in LNP was collected on June 19, 2004, and the sample from Boiling Springs in LNP was collected on June 21, 2004 (table 6.3). The results from these two springs were compared to the water chemistry of the three monitor sites in YNP (table 6.3). Only the last sample from each YNP monitor site was used as a basis for comparison.

Table 6.3. The water chemistry measurements in Lassen National Park (LNP) and Yellowstone National Park (YNP). The sampling dates compared were the closest of the samples collected from each location. The chemical species are measured in mg/L.

	LNP Devil's Kitchen	LNP Boiling Springs	YNP Rabbit Creek	YNP Crater Hills	YNP Ragged Hills
Sample Date	**June 19, 2004**	**June 21, 2004**	**June 3, 2004**	**June 14, 2004**	**June 3, 2004**
Chloride	1	0	3	4	588
Sulfate	543	636	45	1320	122
DOC	2.5	1.2	8.1	6.2	0.5
Ammonia	12.2	0.8	2.8	11.4	0.8
Nitrogen	0.09	0.06	0	0.07	0
Phosphorus	0.11	0.06	0.81	0.33	0.2
Aluminum	0.9	21.1	0.2	26.2	1.9
Antimony	0	0	0.006	0	0.106
Arsenic	0	0	0.051	0.034	0.464
Barium	0	0	0	0.1	0
Boron	5.4	0.2	0	1.3	8.7
Cadmium	0	0	0	0	0
Calcium	60	6	1	0	6
Copper	0.04	0.04	0.02	0	0
Iron	21.2	29.8	6.93	1.47	1.52
Lead	0	0	0	0.03	0
Magnesium	24	4	0	0	0
Manganese	1.03	0.16	0.05	0.03	0.04
Mercury	0	0	0	0	0
Molybdenum	0	0	0.009	0	0.112
Nickel	0	0.04	0	0	0
Potassium	15	0	7	9	59
Selenium	0.007	0	0	0	0
Silicon	142	49	23.2	104	149
Sodium	38	0	9	2	355
Zinc	1.46	0.28	0.28	0.02	0.01

The acidic hot springs in LNP and YNP have very similar water chemistry (table 6.3). This result was not unexpected. The acidic hot springs in LNP and YNP share many characteristics. For example, both areas contain many acidic thermal sites that are ideal conditions for the growth of *Sulfolobus* and other archaeal organisms. Also, the geology and predicted age of each thermal area is similar, therefore the chemistry of the water should be similar.

Water chemistry of Lake Victoria, Africa. It is also interesting to compare the water chemistry of the three monitor sites to a non-thermal environment. Limited water chemistry is known about Lake Victoria in Africa (Talling *et. al*. 1965). Surprisingly, the differences in water chemistry of the three sites in YNP and Lake Victoria are similar to the differences observed between the three monitor sites (table 6.4). There are slight differences in the water chemistry between the two environments. For example, magnesium was not detected in YNP, but it was present in Lake Victoria (table 6.4). Potassium and sulfate concentrations were higher in the YNP monitor sites compared to Lake Victoria (table 6.4). Unfortunately, the chemical analysis of Lake Victoria was not as extensive as the water chemistry analysis in the three monitor sites. It is suspected the metal concentrations would be quite different because of the underlying rock present in YNP.

Table 6.4. A comparison of the average water chemistry measured in the three monitor sites within YNP, and a non-thermal environment, Lake Victoria, Africa. CHMS=Crater Hills, RCMS=Rabbit Creek, and RHMS=Ragged Hills. Chemistry values are measured in mg/L.

	CHMS	**RCMS**	**RHMS**	**Lake Victoria, Africa**
pH*	1.9	4.2	3.2	8-8.8
Calcium	0	1.5	5.8	5-15
Magnesium	0	0	0	1.59-8
Sodium	2.8	11.7	326	10.4-13.5
Potassium	9	5.8	59	3.7-4.2
Sulfate	1022	87	135	2.3
Chloride	1.8	3	611	3.9-7
Nitrogen**	0	0	0	0.011-0.029
Phosphorus	0.47	0.7	0.203	0.047-0.067

*pH of each monitor site in June 2004.
**Expressed as nitrate + nitrite.

In summary, the monitor sites in YNP are distinct from each other. The acidic hot springs within YNP are chemically similar to acidic hot springs in LNP (table 6.3), but distinct from non-thermal environments (table 6.4). Although each of the YNP monitor sites is thermal and acidic, they are chemically distinct. The differences could be attributed to local variations in underlying rocks, water levels, and time. The variation in water chemistry could be supporting the variation in microbial communities in the three monitor sites. Finally, the chemistry of the three monitor sites is not highly dynamic over time.

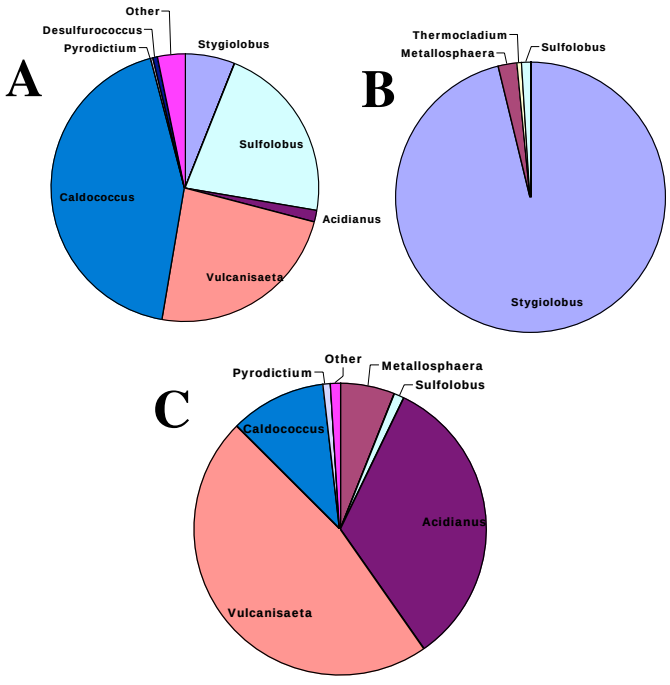
Comparison of archaeal populations in three acidic hot springs

The dominant archaeal population of each monitor site was unique. For an overview of the archaeal host populations within each monitor site, see figure 6.3. Perhaps the unique water chemistry of each monitor site is driving the unique archaeal population. 16S rDNA gene sequence analysis was used to identify the dominant archaeal population in each monitor site. Two primer sets were chosen to amplify full-length (20F/UA1406R; 1,400 bps) or almost full-length (348Fa/UA1406R; 1,000 bps) sequence of the 16S rDNA gene. The environmental sequences were compared to 16S rDNA sequences in the public database.

The Crater Hills monitor site contained the most diverse archaeal population over time (table 6.5 and figure 6.3). The majority of the environmental sequences most closely matched *Caldococcus noboribetus* (126 clones at 86-97% identity) and *Vulcanisaeta distributa* (69 clones at 87-98% identity) (table 6.5). *Sulfolobus*-like

sequences were detected 63 times in Crater Hills during the course of this research (table 6.5). Because of the low sequence identity to published *Sulfolobus* 16S rDNA sequences, it is likely this population within Crater Hills is closely related to *Sulfolobus*, perhaps another member of the *Sulfolobales*. The potential host population in Crater Hills is composed of archaeal organisms that favor low pH and high temperature environments. All of the sequences isolated from this monitor site represent organisms that grow anaerobically (excluding *Sulfolobus*). Most organisms in this hot spring are currently undescribed, but based on 16S rDNA sequences, they are related to described archaeal organisms.

Figure 6.3. A comparison of the archaeal populations detected within each monitor site. A. Crater Hills monitor site; B. Ragged Hills monitor site; C. Rabbit Creek monitor site.



The archaeal population within the Ragged Hills monitor site was the least diverse of the three monitor sites (table 6.5 and figure 6.3). Over the course of this study, only four distinct 16S rDNA sequences were collected from Ragged Hills. The most commonly detected sequence represented *Stygiolobus azoricus* (269 clones at 95-99% identity) (table 6.5 and figure 6.3). *Sulfolobus* is indeed present in Ragged Hills, just at low numbers (table 6.5 and figure 6.3). Surprisingly, the youngest of the three monitor sites is currently dominated by a single species. Perhaps the low nutrient level of the Ragged Hills monitor site and high concentration of metals (table 6.2) does not support a diverse microbial community. Like the Crater Hills monitor site, 16S rDNA sequences from the Ragged Hills site belonged primarily to anaerobic archaeal organisms.

Table 6.5. The organisms detected in each of the three monitor sites and the percentage identities to published 16S rDNA sequences. ND=not detected.

Organism Name	**Number of clones and percent match range**		
	Crater Hills	**Ragged Hills**	**Rabbit Creek**
Caldococcus	126 (86-97%)	ND	10 (90-97%)
Vulcanisaeta	69 (87-98%)	ND	46 (94-98%)
Sulfolobus	63 (87-94%)	3 (98%)	1 (95%)
Acidianus	4 (93-99%)	ND	32 (96-99%)
Metallosphaera	ND	6 (94-97%)	6 (97%)
Pyrodictium	1 (90%)	ND	1 (91%)
Stygiolobus	15 (88-95%)	269 (95-99%)	ND
Thermocladium	ND	2 (96-97%)	ND
Desulfurococcus	1 (83%)	ND	ND

Only two samples from the Rabbit Creek monitor site were used for 16S rDNA gene analysis, and the archaeal population detected in Rabbit Creek was dependant on the primer set. The first sample was exclusively composed of *Acidianus* sequences that were

highly related to published *Acidianus* 16S rDNA sequences (table 6.5 and figure 6.3). The archaeal population of Rabbit Creek shifts drastically in the next sample used for 16S rDNA analysis. This sample yielded a more diverse population, similar to that found in Crater Hills. The majority of the environmental clones most closely resembled *V. distributa* (table 6.5 and figure 6.3). Although only two samples were examined, the Rabbit Creek monitor site contained a diverse collection of archaeal organisms. Like the other two monitor sites, anaerobic organisms dominated the archaeal community based on 16S rDNA analysis. There appeared to be a dramatic shift in the archaeal population within Rabbit Creek; however, different primers were used in each sample. For unknown reasons, 16S rDNA PCR was only successful on these two samples even after numerous attempts with varying 16S rDNA primers. In order to more fully understand the archaeal population in Rabbit Creek, both primer sets would need to be used on more samples.

Overall, the environmental sequences collected in this research most closely resembled nine published 16S rDNA sequences and some matched uncultured archaeal clones (table 6.6). Most of the organisms were detected in two or more monitor sites, excluding *Thermocladium*-like sequences that were only detected in Ragged Hills and a *Desulfurococcus*-like sequence only detected in Crater Hills (table 6.6). *Sulfolobus* was the only 16S rDNA sequence detected in all three monitor sites (table 6.6).

It was predicted that *Sulfolobus* species would be commonly detected in these monitor sites. Indeed, one reason these three particular hot springs were chosen for this study was based on the fact that *Sulfolobus* could be easily cultured from each of the sites. However, based on environmental 16S rDNA gene analysis, the dominant

organism in these hot springs is not *Sulfolobus* or even *Sulfolobus*-like species (table 6.6). This research has revealed a broader range of the archaeal diversity present in these extreme thermal environments within YNP. Previous to this research, some species had not been detected outside the location of original isolation. For example, *Stygiolobus azoricus* was originally discovered in the Azores and had not been detected in other locations worldwide (Segerer *et. al.* 1991). However, it was found to be quite abundant in the YNP hot springs, especially in the Ragged Hills monitor site.

Table 6.6. The organisms detected in this research, how many times they were detected, and the location(s) they were detected.

Closest match*	***Number of Sequences***	***Location(s)***
Caldococcus noboribetus	136	Crater Hills; Rabbit Creek
Vulcanisaeta distributa	115	Crater Hills; Rabbit Creek
Sulfolobus solfataricus	67	Crater Hills; Ragged Hills; Rabbit Creek
Acidianus species	36	Crater Hills; Rabbit Creek
Metallosphaera sedula	12	Ragged Hills; Rabbit Creek
Pyrodictium occultum	2	Crater Hills; Rabbit Creek
Stygiolobus azoricus	284	Crater Hills; Ragged Hills
Thermocladium modestius	2	Ragged Hills
Desulfurococcus mobilis	1	Crater Hills
Uncultured archaeal clones	11	Crater Hills; Rabbit Creek

*closest match from NCBI BLAST based on 1,000-1,400bps of environmental 16SrDNA sequence.

The three monitor sites each contained unique microbial communities and unique water chemistries. The 16S rDNA genes sequenced in this study represent archaeal organisms that are anaerobic thermophilic acidiophiles (excluding the aerobic *Sulfolobus*). *Stygiolobus*-like, *Acidianus*-like, *Vulcanisaeta*-like, *Caldococcus*-like, and *Sulfolobus*-like sequences were the most commonly detected 16S rDNA genes (table 6.6). The information considering electron donors and acceptors is limited in these organisms. However, *Vulcanisaeta* has been shown to utilize $FeCl_3$ and sulfate as possible electron acceptors (Itoh *et. al.* 2002). It is reasonable the water chemistry of Rabbit Creek and Crater Hills would support *Vulcansaeta*-like populations.

Many of the organisms that were detected utilize elemental sulfur. *Caldococcus* requires sulfur in order to maintain a high growth rate (Aoshima *et. al.* 1996). *Sulfolobus* and *Acidianus* also utilize elemental sulfur to produce sulfuric acid (Garrity 2001). *Stygiolobus* grows anaerobically with elemental sulfur to produce hydrogen sulfide (Garrity 2001). Even though the chemistry analysis did not examine sulfur, the Crater Hills monitor site contains visible amounts of elemental sulfur in the hot spring and crusted around the border of the site.

Although there is little information concerning the growth of these organisms in nature, the water chemistry of each monitor site is likely to support the growth of the detected archaeal organisms. As mentioned previously, analyzing the chemical speciation and gasses present in the monitor sites will aid in the understanding of these microbial communities. Further understanding the chemistry in the acidic hot springs within YNP may help achieve a better understanding of archaeal organisms in nature.

It is likely that archaea are dominant in the high temperature acidic environments within YNP. Previous research has revealed that hyperthermophilic bacteria are dominant in many hot springs within YNP (Spear *et. al.* 2005; Hugenholtz *et. al.* 1998; Reysenbach *et. al.* 2000). However, these studies were conducted within high temperature (~70°C) and neutral pH environments. Current research with 16S rDNA PCR with universal primers (Spear *et. al.* 2005) has revealed the monitor sites contain exclusively archaeal organisms (M. Young lab, unpublished). Therefore, unlike neutral hot springs, the bacterial community in the acidic hot springs is likely not significant.

The archaeal host populations in Crater Hills and Rabbit Creek changed during this study. The dynamics of the potential host populations cannot be correlated to water chemistry. It is currently unknown what may be driving the changes in the host population structures. Perhaps viruses are driving the changes in the host populations.

These three acidic hot springs within YNP each contain a unique collection of archaeal organisms. We were underestimating the diversity of these communities when we assumed *Sulfolobus* species would be dominant because that was the organism easily cultured. Instead, I found archaeal communities that, for the most part, are currently undescribed. We hope to learn more about each of the common archaeal organisms detected in each of the monitor sites by culturing each organism and looking for viruses.

Comparison of SSV-like populations in three acidic hot springs

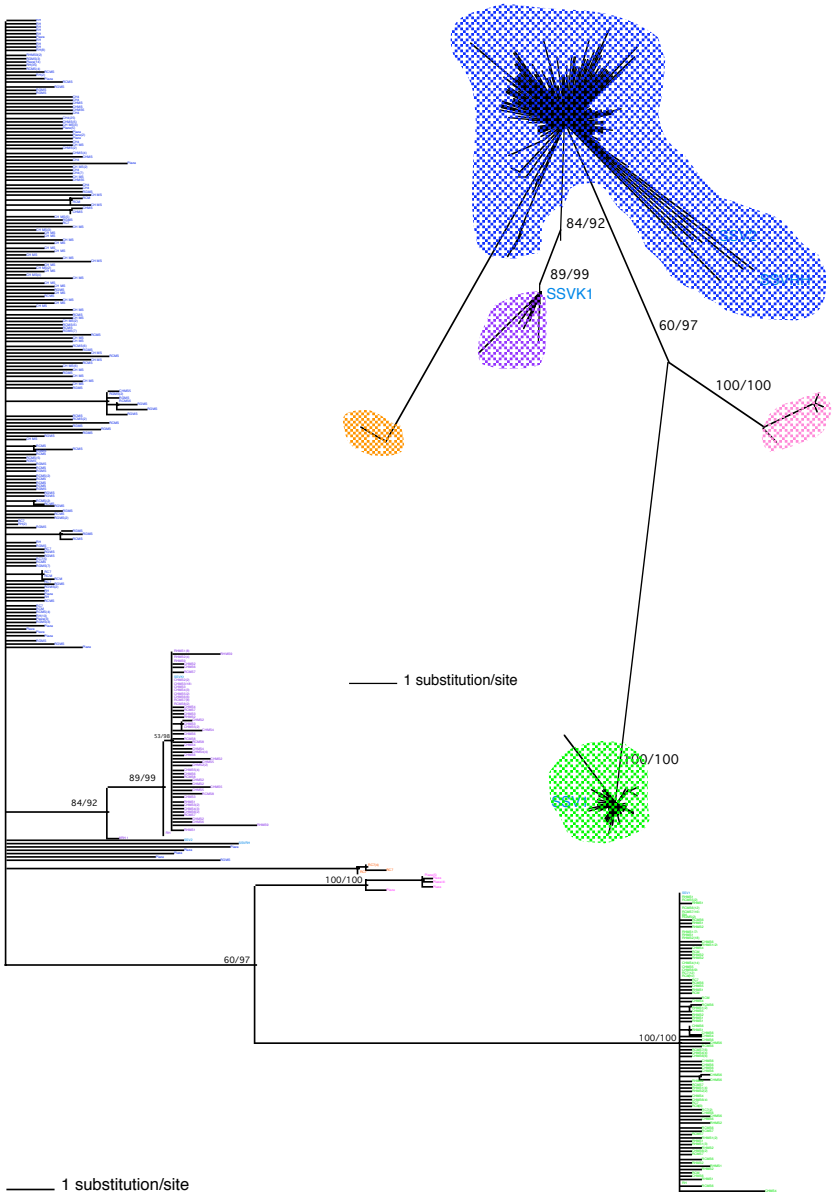
SSVs were monitored continuously over time in three acidic hot springs in YNP. Previous research revealed the most commonly isolated viruses from samples collected in

YNP were SSV-like viruses (Rice *et. al.* 2001). Four SSVs have been sequenced from various locations around the world and their genomes compared (Wiedenheft *et. al.* 2004). This allowed for the design of two sets of "universal" SSV primers to amplify SSV-like sequences from the environment. The 3F/4R primer set amplifies approximately 255bps of the largest ORF in the viruses (b812 in SSV-RH). This primer set was the most successful at amplifying SSV-like sequences from the three monitor sites. The 7F/8R primer set amplifies the coat proteins of the viruses. The amplicon ranged from approximately 600-1,000bps, depending on the presence of VP2 (Wiedenheft *et. al.* 2004). This primer set was not as successful at amplifying environmental viral sequences. In fact, a 7F/8R product was never obtained from the Crater Hills monitor site, and only four samples from the Ragged Hills and Rabbit Creek monitor sites resulted in a product. This result is likely a result of difference in sensitivity of the SSV 3F/4R and 7F/8R primer sets on purified viral templates. SSV 3F/4R had a detection limit of 0.1pg, and 7F/8R was 10,000X less sensitive (1ng).

Comparison of SSV-like populations using 3F/4R primer set

The 'universal' SSV 3F/4R primer set was successful at amplifying SSV-like sequences from the three monitor sites. However, not every sample was positive for SSV-like sequences with these primers. Overall, there were three major clades of SSV-like viruses detected in the three monitor sites over time (figure 6.4). The three major SSV-like clades were detected in all three monitor sites, but at different times. The Ragged Hills and Rabbit Creek monitor sites each contained one SSV-like clade not detected in the other sites.

Figure 6.4. The SSV-like DNA 3F/4R sequences obtained from the three monitor sites over time. A maximum likelihood analysis was conducted using MrBayes version 3 (Huelsenbeck 2001) with Bayesian clade credibility values, and a maximum parsimony analysis conducted with PAUP 4.0b10 (Swofford 2002) with 10,000 bootstrap replicates. The numbers at main nodes represent the Bayesian clade credibility and the maximum parsimony bootstrap. Inset: an unrooted phylogram to better illustrate the three major groups of SSV-like 3F/4R sequences.



The Crater Hills monitor site contained the best results for the SSV-like 3F/4R data. Eight of fourteen samples were positive for SSV-like sequences for a total of 292 SSV-like 3F/4R clones (table 6.7). Most of the detected SSV-like sequences were obtained in the first five samples collected from the Ragged Hills monitor site. The Ragged Hills monitor site resulted in approximately the same number of clones as the Crater Hills monitor site (285 SSV-like 3F/4R sequences; table 6.7); however, only six of fifteen total samples yielded positive results. The Rabbit Creek monitor site resulted in 208 SSV-like 3F/4R sequences (table 6.7) in seven out of twelve samples collected from this site.

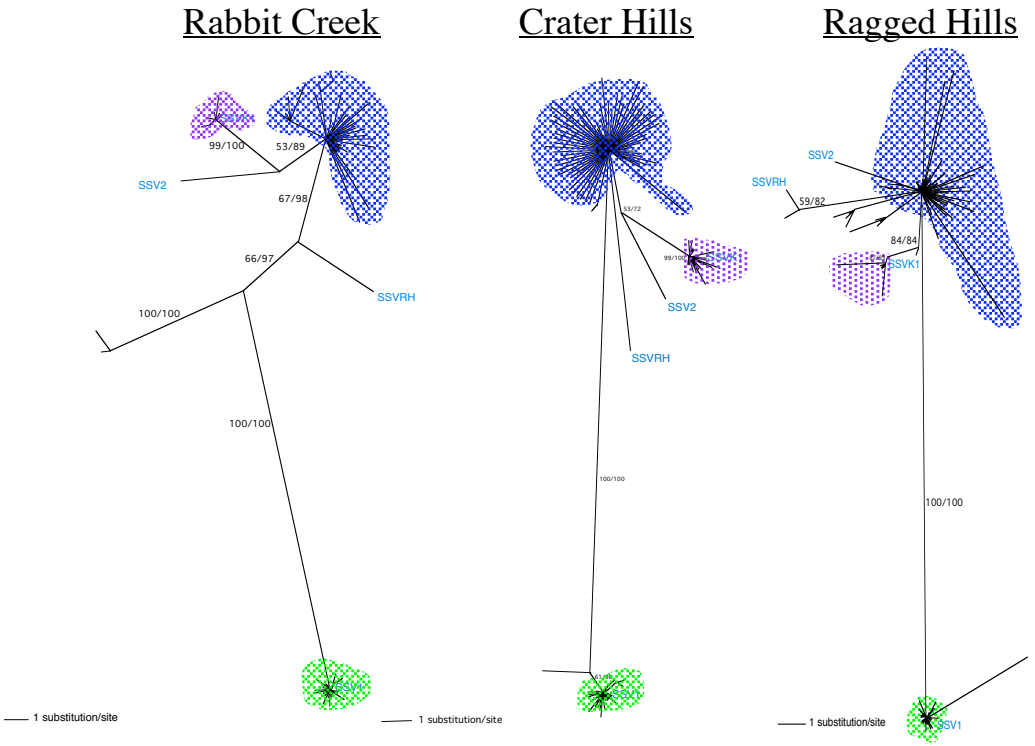
Table 6.7. The viral clones sequenced from each monitor site during this research.

	Crater Hills	**Rabbit Creek**	**Ragged Hills**	**Total**
SSV 3F/4R	292	208	285	**785**
SSV 7F/8R	0	112	113	**225**
SIRV DBPF/R	812	688	663	**2163**

The three monitor sites contain SSV-like virus populations that are dynamic and changing in a short period. The SSV-like populations in the three monitor sites that were studied for approximately two years were found to have similar overall population structures (figure 6.5). However, the presence or absence of a particular virus clade at a particular time varied widely. It is, therefore, impossible to predict which virus clades would be present at a particular site at a particular sampling time. The sequence abundance graphs of the SSV-like sequences from each of the three monitor sites reveal that sequences detected only one time (singletons) were the most common sequences

detected in each of the three monitor sites (figures 3.21, 4.22, and 5.18) usually composing more than a third of our sampled population. This unusually high number of singletons commonly indicates a high migration rate or a high mutation rate in these organisms.

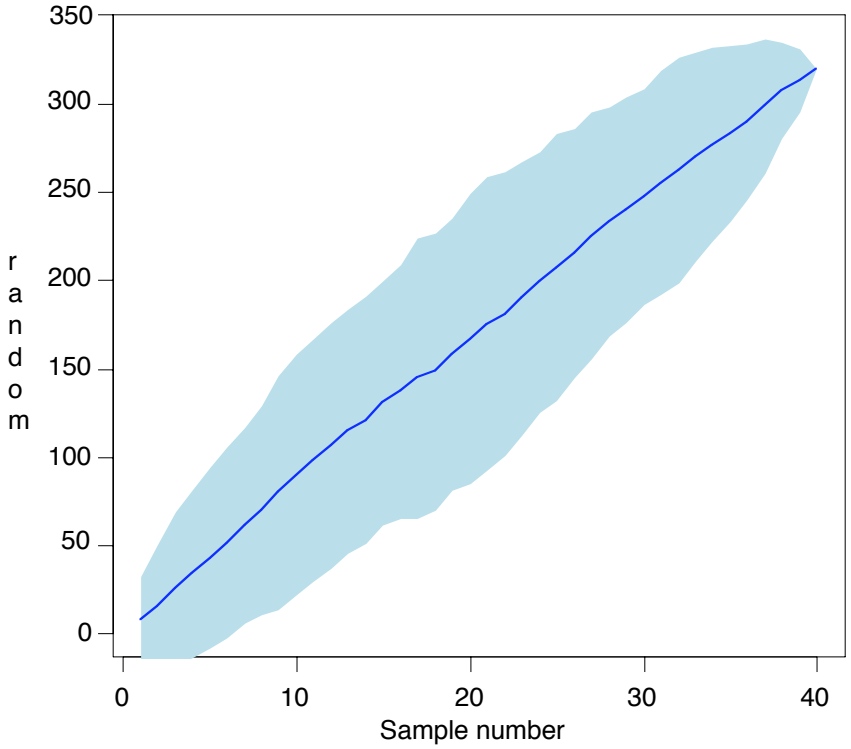
Figure 6.5. Unrooted phylograms illustrating the different groups of SSV-like 3F/4R sequences detected from the three monitor sites over time. A maximum likelihood analysis was conducted using MrBayes version 3 (Huelsenbeck and Ronquist 2001) with Bayesian clade credibility values, and a maximum parsimony analysis conducted with PAUP 4.0b10 (Swofford 2002) with 10,000 bootstrap replicates. The numbers at main nodes represent the Bayesian clade credibility and the maximum parsimony bootstrap. The colors correspond to figure 6.4.



I have shown that the SSV-like population can be drastically different from month to month. I sequenced 785 SSV-like 3F/4R clones (table 6.7) during this research. However, we did not capture the total diversity of SSV-like sequences in these hot springs (figure 6.6). A species accumulation curve (generated using the Vegan program version 1.7-50; http//:cc.oulu.fi/~jarioka/softhelp/vegan/html/00Index.html) for the SSV-like viruses from all study sites illustrates the large diversity of sequences (figure 6.6). It also reveals that a saturation of the viral community was not achieved (figure 6.6). It is unknown how many more clones would need to be sequenced to capture the total viral diversity within these hot springs. However, it could be potentially a very large number. When the SSV-like 3F/4R sequences from the monitor sites are combined, a Zipf distribution results (figure 6.7). The Zipf distribution indicates that there is one dominant sequence and the other sequences are present in much lower numbers (Cancho *et. al.* 2003). This indicates the presence of a dominant group of SSV-like viruses with other SSV-like viral populations present in lower numbers.

This is in contrast to SSV1 sequences once placed in culture, which do not change significantly over time. Nevertheless, it must be remembered that organisms behave very differently in culture than they would in the natural environment. In the environment, organisms are consistently dealing with a changing environment, and potentially competition with other organisms. In culture, these pressures are not as severe, and therefore the organisms might not have to adapt as much as they would in a natural setting.

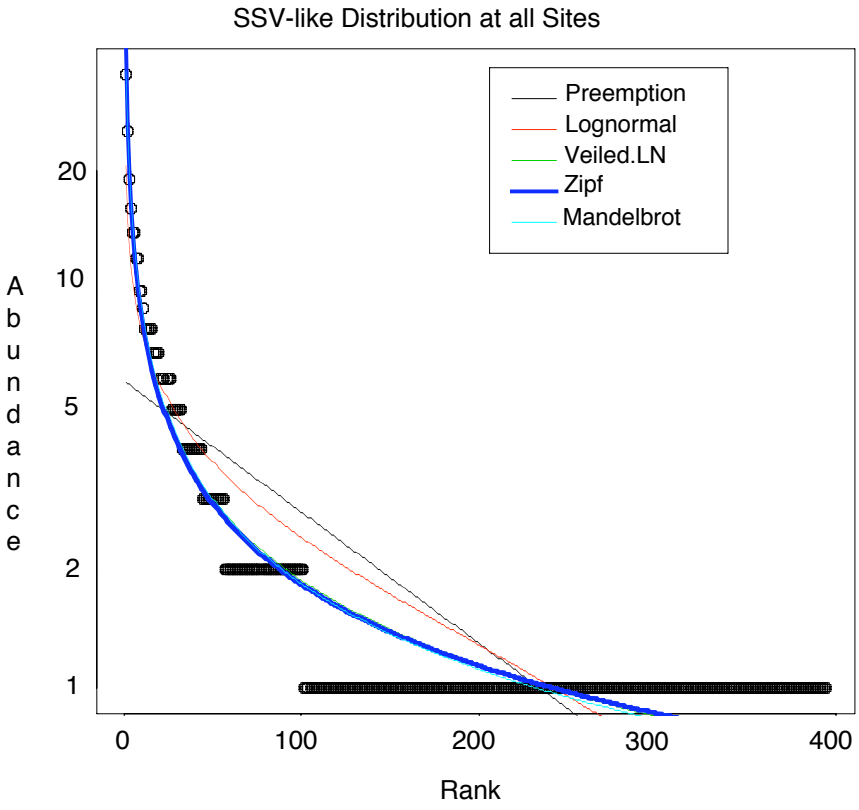
Figure 6.6. A sequence accumulation curve for the SSV-like viruses detected over time in all three monitor sites. This indicates the total diversity of SSV-like sequences in YNP was not captured.



When all of the SSV-like 3F/4R sequences obtained from the three different monitor sites are compared, there are three major clades of SSV-like viruses within YNP with two outlying, smaller groups (figure 6.4). The differences between the clades of SSV-like viruses represent differences of 5-10% in the 3F/4R sequence. One group of viruses represented by sequences from all the monitor sites is more closely related to SSV-K1 than to other YNP sequences (colored purple in figure 6.4). Another group that is also represented by sequences from all three acidic hot springs is more closely related to SSV1 than to other YNP sequences (colored green in figure 6.4). Finally, the largest

major group of viruses includes SSV-RH and SSV2 (colored blue in figure 6.4). There were two smaller, outlying groups of SSV-like sequences from YNP that represented sequences from Ragged Hills (colored pink in figure 6.4) and Rabbit Creek (colored orange in figure 6.4).

Figure 6.7. SSV-like distribution at all three monitor sites over time. All models were attempted, but the Zipf model best fit the data.



The data from this research indicates the differences between the four sequenced SSV isolates are not due to geographical isolation. One reason the four sequenced SSV isolates may be distinct from each other is the fact that prior to sequencing the viral genome, each of the viruses was maintained in culture (Palm *et. al.* 1991; Stedman *et. al.*

2003; Wiedenheft *et. al*. 2004). We detected a difference between cultured SSV-like sequences enriched from two monitor sites within YNP (Rabbit Creek and Ragged Hills; see Chapter 2). However, environmental SSV-like sequences do not reveal differences based on monitor site (figure 6.4).

The four sequenced SSV isolates do not represent all major viral types present in the individual environments (YNP, Japan, Iceland, and Kamchatka). The results are biased when a single cultured isolate is considered a geographical representative. A single SSV genome cannot be used to define the viral population structure and diversity within a thermal area. It is likely if a similar study were conducted in Japan, Iceland, or Kamchatka, the results would be similar to the results from YNP. Likewise, if one SSV was cultured and the genome was sequenced from each of the three monitor sites, the differences between the three genomes may be similar to the differences between the sequenced SSV genomes from Japan, Iceland, Kamchatka, and YNP.

Based on the 3F/4R analysis, the SSV-like populations in the three acidic hot springs within YNP are dynamic. Each site underwent a shift in SSV-like clades present over time. It is curious as to why the SSV-like viruses were not always detected in each of the monitor sites at each of the sampling times. However, the SSV-like 3F/4R population was missing from the monitor sites at similar times. Therefore, it is likely the viruses were not present in the environment during those months or they were present below our detection level. SIRV-like and 16S rDNA sequences were detected from the samples not containing SSV-like 3F/4R sequences indicating the source DNA was suitable for PCR. The amplification of the 3F/4R gene was attempted numerous times

before the absence of SSV-like sequences was declared, and a positive control of spiking environmental samples with known templates was successful. Therefore, it is unlikely the absence of SSV-like sequences is due to experimental error.

When comparing all of the SSV-like sequences collected from the YNP monitor sites, 77% of the bases in the 3F/4R sequence were identical over time (table 6.8). There were 320 unique SSV-like DNA sequences detected from 785 SSV-like 3F/4R clones. The largest group of identical sequences was composed of 88 sequences comprising 11.2% of the sequenced clones. This large group consisted of sequences from all three monitor sites (table 6.9). There were smaller groups of identical sequences ranging from two to 58 identical SSV-like sequences, but the majority of the SSV-like sequences (256) were detected only once. SSV-like singletons comprised 32.6% of the clones that were sequenced during the two years of this research. The base frequencies from SSV-like 3F/4R sequences in each monitor site were similar (table 6.10). Surprisingly, the SSV-like populations in each monitor site had different nucleotide substitution models (table 6.10); perhaps indicating differences in individual environments will cause evolution to proceed differently.

Table 6.8. Numbers of identical bases, transitions, and transversions over time in each of the monitored virus populations.

	Identical	**Transitions**			**Transversions**		
	Bases	1st	2nd	3rd	1st	2nd	3rd
SSV 3F/4R	199/258	8	4	20	8	5	14
SSV 7F/8R	329/509	22	22	22	38	34	42
SIRV	343/405	4	0	20	7	7	24

Table 6.9. The groups of identical SSV-like sequences from the three monitor sites over time. There were also groups of two to five identical SSV-like sequences and 256 singletons. CHMS=Crater Hills monitor site, RCMS=Rabbit Creek monitor site, and RHMS=Ragged Hills monitor site.

# of Identical Sequences	SSV-like sequences from CHMS	SSV-like sequences from RCMS	SSV-like sequences from RHMS
88	24	52	12
58	0	4	54
52	32	10	10
38	30	0	5
22	0	16	6
16	10	6	0
15	5	5	5
15	2	6	7
14	7	6	1
13	0	6	7
11	0	5	6
10	10	0	0
10	8	1	1
8	0	0	8
7	1	6	0
6	1	1	4

There are similarities and differences between the SSV-like populations in the three monitor sites over time based on DNA sequences. We detected the largest group of SSV-like sequences from each monitor site early in the research (July 2002—February 2003), but it was undetectable for seven to nine months (figures 3.14, 4.11, and 5.8; the largest group of SSV-like viruses is indicated in blue). When the group was detected again, it was done so at very low numbers; only one SSV-like sequence from this group was detected from each monitor site in September 2003 (Crater Hills), October 2003 (Rabbit Creek), and November 2003 (Ragged Hills) (figures 3.14, 4.11, and 5.8).

Table 6.10. The base frequencies and nucleotide substitution models for each data set from each monitor site. The models were selected using the AIC criterion in Modeltest version 3.06 (Posada *et. al.* 1998). CHMS=Crater Hills monitor site, RCMS= Rabbit Creek monitor site, and RHMS=Ragged Hills monitor site.

	SSV 3F/4R			**SSV 7F/8R**		**SIRV DBPF/R**		
	CHMS	**RCMS**	**RHMS**	**RCMS**	**RHMS**	**CHMS**	**RCMS**	**RHMS**
A	0.2773	0.2608	0.2781	0.2868	0.2872	0.3276	0.3188	0.3368
C	0.1954	0.2123	0.1927	0.2079	0.2218	0.1814	0.2040	0.2082
G	0.2382	0.2163	0.2068	0.2310	0.2132	0.1833	0.1945	0.1758
T	0.2891	0.3105	0.3224	0.2743	0.2779	0.3077	0.2826	0.2792
Model	*GTR+I+G*	*HKY+G*	*GTR+G*	*TVM+I+G*	*HKY+G*	*K81uf+G*	*K81uf+G*	*TVM+G*

In Crater Hills, the SSV-like sequences related to SSV-K1 (represented by purple in figure 6.4) and the sequences related to SSV1 (represented by green in figure 6.4) appeared later in the study (May 2003) and were detectable through October 2003 (figure 3.14). In the Rabbit Creek monitor site, SSV-like sequences related to SSV-K1 were not detected until November 2003 (figure 4.11), but the SSV-like sequences related to SSV1 were detected in most samples that SSV-like sequences were detected (figure 4.11). However, in the Ragged Hills monitor site, SSV-like sequences related to SSV-K1 or SSV1 were detected in most samples that SSV-like sequences were detected (figure 5.8). These results illustrate that although the same groups of SSV-like sequences were detected in each of the monitor sites over time, they were not present in all the sites at the same time.

The Ragged Hills monitor site contained SSV-like sequences that were not detected in Crater Hills or Rabbit Creek (figure 5.8; orange, red, and turquoise ovals), but these groups were only detected in November 2002 and January 2003. It is interesting to note the similarity in the absence of SSV-like sequences in the Rabbit Creek and Ragged

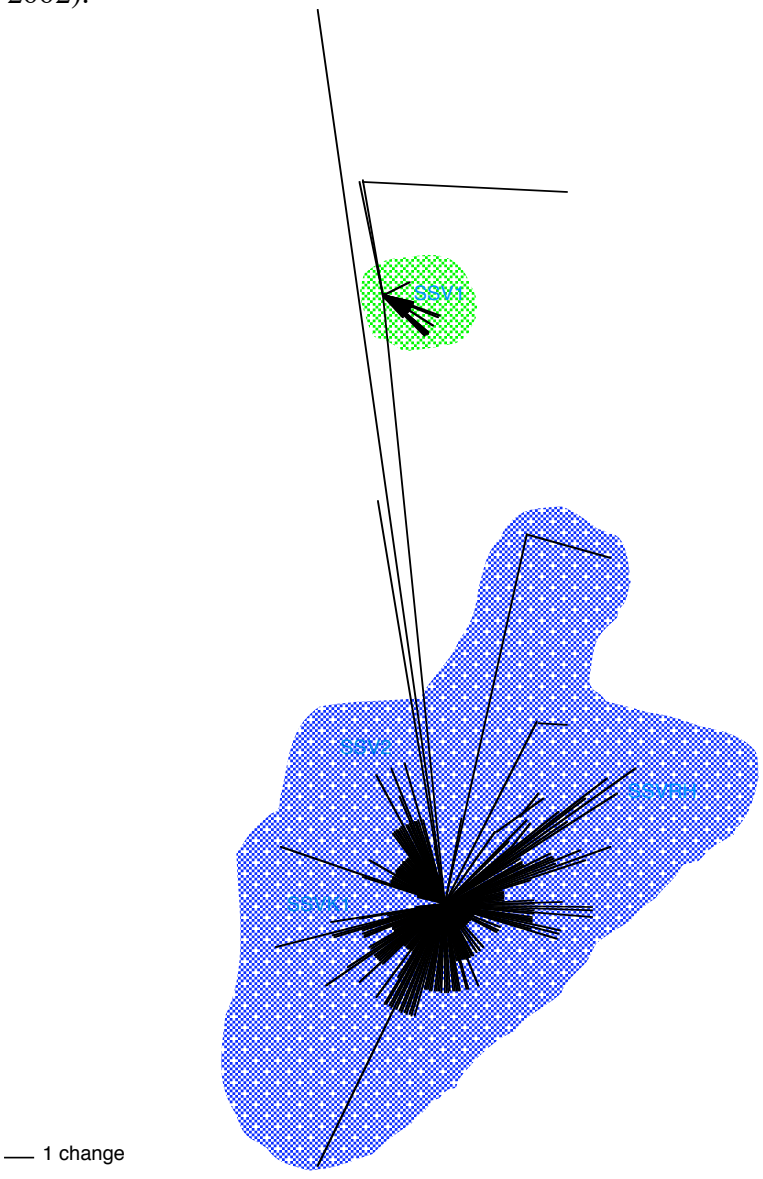
Hills monitor sites (figures 4.11 and 5.8, respectively). SSV-like sequences were undetectable in these sites during the summer and fall of 2003, but they were detected again in the winter of 2003 (figures 4.11 and 5.8). The SSV-like sequences in Crater Hills had the opposite trend; sequences were detected through fall 2003, and then undetectable (figure 3.14).

It is interesting to note that differences in the SSV-like DNA sequences detected in the three monitor sites did not always result in changes in the inferred protein sequences (figures 3.14 and 3.15, 4.11 and 4.12, and 5.8 and 5.9). When the SSV-like 3F/4R inferred protein sequences from all three monitor sites are compared, there are two major clades of SSV-like protein sequences (figure 6.8). The group of SSV-like DNA sequences related to SSV-K1 from the three monitor sites belongs to the largest group of protein sequences detected in this research (blue in figure 6.8).

A concern is whether we were amplifying viral sequences from virions or whether they were integrated into their host organisms. All four SSV viruses that have been completely sequenced have been shown to integrate into tRNA genes of *Sulfolobus* (Wiedenheft *et. al.* 2004). By extracting total DNA and performing PCR on the entire environmental sample, it is possible, if not likely, that proviruses were amplified and sequenced as well as episomal forms of viral DNA. The interrupted ORFs detected in this research may be representing old SSV integration events that may have acquired mutations over time (She *et. al.* 2001). However, recent events of integration may look like functional (uninterrupted) ORFs. Using the methods employed in this research, it

would be difficult to determine if the SSV-like sequences are from viruses or from recent integration events.

Figure 6.8. SSV-like 3F/4R inferred protein sequences from the three monitor sites collected during this research. A maximum parsimony analysis was conducted with 10,000 bootstrap replicates (Swofford 2002).



In summary, the SSV-like 3F/4R sequences detected from the three monitor sites changed over time. The population structures of the SSV-like sequences detected in all three monitor sites was comparable (figure 6.5). However, the clades of viruses were not always present in the three sites at the same time, making it impossible to predict the presence or absence of viral clades at a particular time. There is an enormous amount of diversity in SSV-like sequences within the three monitor sites over time.

Comparison of SSV-like populations using 7F/8R primer set

The primers that were designed to amplify the coat proteins of SSVs were not as successful as the 3F/4R primer set in amplifying SSV-like sequences from the environment (table 6.7). The SSV-like coat proteins were never detected from the Crater Hills monitor site. They were only detected four times during the course of this research, twice from the Ragged Hills monitor site, and twice from the Rabbit Creek monitor site. This result brings up the possibility that the 3F/4R primers may amplify sequences in many thermophilic organisms, and may not be specific to the SSV-like viruses. However, when compared to genomes and sequences of described thermophilic organisms, there is not a region that would be amplified by these primers. Furthermore, the environmental 3F/4R sequences were the correct size (~255bps) and they closely matched known SSV sequences (SSV1, SSV2, SSV-RH, and SSV-K1). Even though archaeal populations changed and varied between monitor sites, 3F/4R sequences were detected. These results indicate that the 3F/4R sequences were viral.

Perhaps the inability to consistently detect SSV-like coat proteins in the environment is a result of the sensitivity of the 7F/8R primer set (1ng/PCR reaction). The

primers designed to amplify the viral coat proteins were not as sensitive on purified viral templates as the 3F/4R primers (0.1pg/PCR reaction) (see Chapter 3). Presumably if the coat protein primers (7F/8R) were as sensitive as the 3F/4R primer set, more coat protein sequences would have been detected.

It is difficult to compare the two samples of SSV-like coat protein sequences from Rabbit Creek because of the separation in sampling times. The first sample that contained SSV-like coat protein clones was collected on July 26, 2002, and the coat proteins were not detected again until November 13, 2003 (table 4.3). There was more than a year in between the two samples containing SSV-like coat protein sequences. Regardless, over time the SSV-like coat protein sequences in Rabbit Creek changed from a diverse population to a population represented by one group (figure 4.17).

There were also only two samples collected from Ragged Hills that contained SSV-like coat protein sequences (table 5.3). However, as opposed to Rabbit Creek, the samples containing SSV-like coat protein sequences from Ragged Hills were both collected at the beginning of this research. The samples collected from Ragged Hills on October 17, 2002 and November 5, 2002 were the only samples from this monitor site that were positive for the SSV-like coat protein sequence. For the most part, the sequences collected in November 2002 were similar to the sequences collected from October 2002 (figure 5.14).

Perhaps a drastic change was not obvious in Ragged Hills as it was in Rabbit Creek because of the proximity in the collection of samples. In Rabbit Creek, the two samples containing SSV-like coat protein sequences were collected over one year apart;

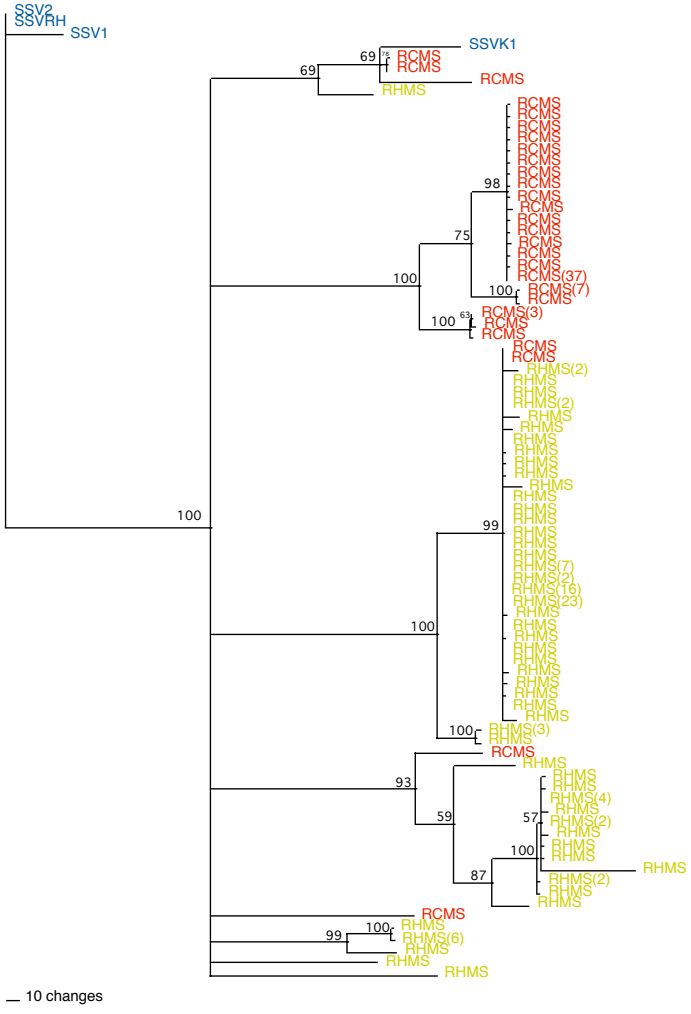
however, in Ragged Hills, the samples were collected within three weeks of each other. In any case, to better determine changes occurring in the SSV-like populations based on changes in the coat proteins, more clones would need to be acquired over time and sequenced.

In comparing the limited number of SSV-like coat protein sequences from Rabbit Creek and Ragged Hills, they appear to be geographically separated. When the inferred amino acid sequences from SSV-like coat proteins from Rabbit Creek and Ragged Hills are compared, they appear to be separated primarily based on location (figure 6.9). The major group from Rabbit Creek is only composed of Rabbit Creek sequences (red in figure 6.9), the major group from Ragged Hills (yellow in figure 6.9) contains only two Rabbit Creek sequences, and some of the minor groups contain sequences from both monitor sites (figure 6.9). Three of the sequenced isolates (SSV1, SSV2, and SSV-RH) are distinct from the YNP environmental sequences, but SSV-K1 is related to a minor group of Rabbit Creek coat protein sequences (figure 6.9). When comparing all of the SSV-like 7F/8R sequences collected from the monitor sites, only 64% of the bases in the coat protein DNA sequences were identical over time (table 6.7).

These results indicate the number of SSV-like 7F/8R clones examined in this study was severely limited. Presumably if more 7F/8R clones were sequenced and analyzed, the results would be comparable to the 3F/4R results indicating that similar groups of SSV-like viruses are present in the monitor sites over time. The SSV-like coat protein results are analogous to comparing one sequenced SSV isolate from each monitor site. They would presumably be quite different, and the differences may be contributed

to geographic isolation. However, the differences in viral sequences between thermal areas are not likely a result of geographic isolation, but rather a large diversity of viruses present in these environments. Overall, to better understand the diversity and dynamics of SSV-like viruses based on coat protein sequences, more clones need to be collected, sequenced, and analyzed.

Figure 6.9. The amino acid sequences of SSV-like coat proteins from Ragged Hills and Rabbit Creek. A maximum parsimony analysis conducted with PAUP 4.0b10 (Swofford 2002) with 10,000 bootstrap replicates. RHMS=Ragged Hills Monitor Site; RCMS=Rabbit Creek Monitor Site.



This research has indicated that SSV-like viral populations within hot springs in YNP are very diverse and dynamic. In fact, the virus populations are changing at such a high rate that it would be difficult to characterize the viral community at any one point in time. The overall SSV-like population structure, based on 3F/4R sequence, is similar in three hot springs that are geographically isolated from each other within YNP. However, the viral clades detected within the monitor sites were present in the sites at varying times. This result indicates non-selective migration of these viruses between the thermal sites within YNP. In order to fully understand the dynamics based on SSV-like 7F/8R sequences, more sequences will need to be analyzed and the design of more sensitive primers might be useful.

Comparison of SIRV-like populations in three acidic hot springs

The second virus monitored over time in this research was the SIRV-like population in YNP. Previous research revealed SIRV-like viruses were the second most commonly detected virus in enrichment cultures from YNP (Rice *et. al.* 2001). The complete genomes of two SIRVs have been sequenced from Iceland (Peng *et. al.* 2001), and one is currently being sequenced from the Norris Geyser Basin in YNP (M. Young lab, unpublished). One primer set was chosen to amplify SIRV-like sequences from the environment. The primer set, named DBPF/R, amplifies the entire coat protein (405 nucleotides and 135 amino acids). This primer set was very successful at amplifying SIRV-like sequences from the three monitor sites. This primer set was the most sensitive of the three viral primer sets used in this research (SIRV DBPF/R 0.001pg; SSV 3F/4R

0.1pg, and SSV 7F/8R 1ng). As opposed to the SSV-like viruses, the SIRV-like viruses were detected in all samples collected from the monitor sites during this study, excluding one sample from Ragged Hills. Overall, there were 2,163 SIRV-like coat protein clones from the monitor sites that were sequenced during this research (table 6.7).

When comparing all of the SIRV-like sequences collected from the monitor sites, 84% of the bases in the coat protein sequence were identical over time (table 6.8). From 2,163 SIRV-like coat protein clones there was 722 unique DNA sequences. The largest group of identical sequences was composed of 660 clones (table 6.11). This group contained sequences from each of the three monitor sites, and comprised 30.5% of the total SIRV-like clones sequenced during this research. Other groups of identical sequences ranging from two to 218 sequences were also detected in the sampled SIRV-like population (table 6.11). The unique DNA sequences (613) composed 28.34% of the SIRV-like clones sequenced during this research. In contrast to the SSV-like 3F/4R sequences, SIRV-like singletons were not the most commonly detected group over time. Instead, the largest group of identical sequences (comprised of 660 sequences from all three monitor sites) was the most commonly detected sequence over time (table 6.11).

Similar to what was detected with SSV-like viruses, the SIRV-like viral diversity was very large. The SIRV-like populations in the three monitor sites are also very dynamic over time. Even though over 2,100 SIRV-like coat protein clones were sequenced, we did not capture the total amount of diversity present in the hot springs (figure 6.10). However, the sequences we did collect reveal an astounding amount of diversity in these viruses over time (figure 6.10). When all of the SIRV-like sequences

detected in all three monitor sites are combined, a Mandelbrot distribution results (figure 6.11). The Mandelbrot distribution is similar to the Zipf distribution (as seen in the SSV-like viruses) in that one sequence occurs often while many others occur rarely (Mandelbrot 1966). This indicates the presence of a dominant, resident group of SIRV-like viruses, and the presence of numerous smaller groups of viruses that may be present transiently.

Table 6.11. The groups of identical SIRV-like sequences from the three monitor sites over time. There were also groups of two to five identical sequences and 613 singletons. CHMS=Crater Hills monitor site, RCMS= Rabbit Creek monitor site, and RHMS=Ragged Hills monitor site.

# of Identical Sequences	SIRV-like sequences from CHMS	SIRV-like sequences from RCMS	SIRV-like sequences from RHMS
660	250	203	207
218	75	72	77
127	62	32	33
53	10	29	14
42	4	24	14
38	0	3	35
33	16	7	10
22	0	20	2
20	0	17	3
20	20	0	0
17	1	11	5
16	2	11	3
13	13	0	0
12	10	1	1
10	10	0	0
8	8	0	0
7	7	0	0
7	7	0	0
7	0	7	0
7	0	7	0
6	6	0	0

The SIRV-like viruses were more diverse than the SSV-like viruses in the three monitor sites over time. The SIRV-like populations sampled in Crater Hills and Ragged Hills were found to be the most diverse with approximately nine and eight viral groups detected over time, respectively (figures 3.22 and 5.19). The Rabbit Creek SIRV-like population was the least diverse with only five groups of viruses detected over time (figure 4.23). The largest group of SIRV-like sequences from all three monitor sites was detected consistently throughout this research. Smaller groups of SIRV-like viruses appeared and disappeared over time, but the largest group seemed to be a permanent resident of the viral community in each monitor site.

Figure 6.10. A species accumulation curve for the SIRV-like viruses detected over time in all three monitor sites. This indicates the total diversity of SSV-like sequences in YNP was not captured.

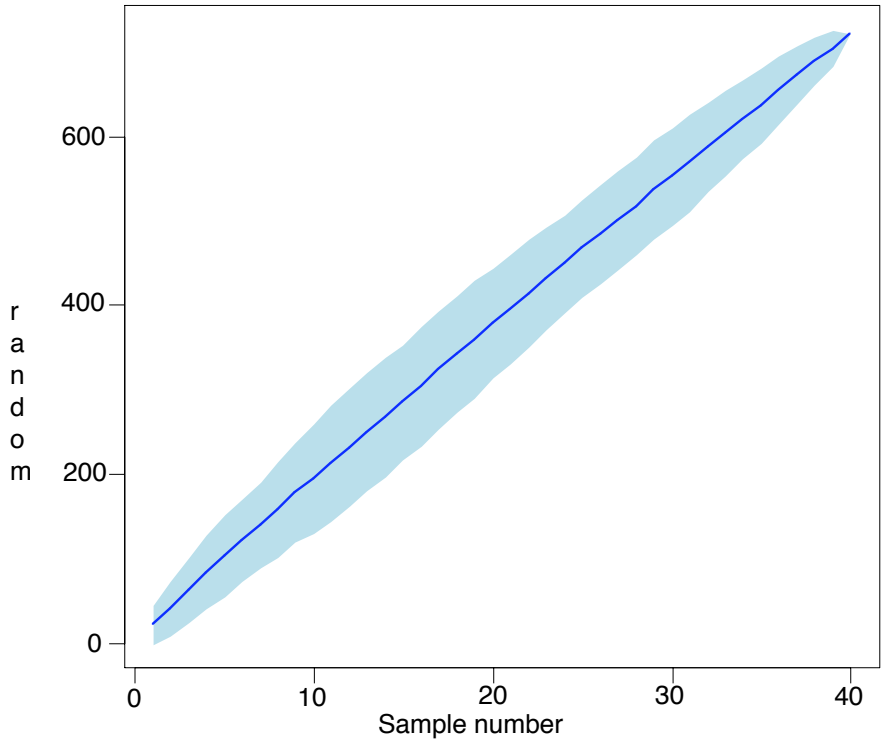
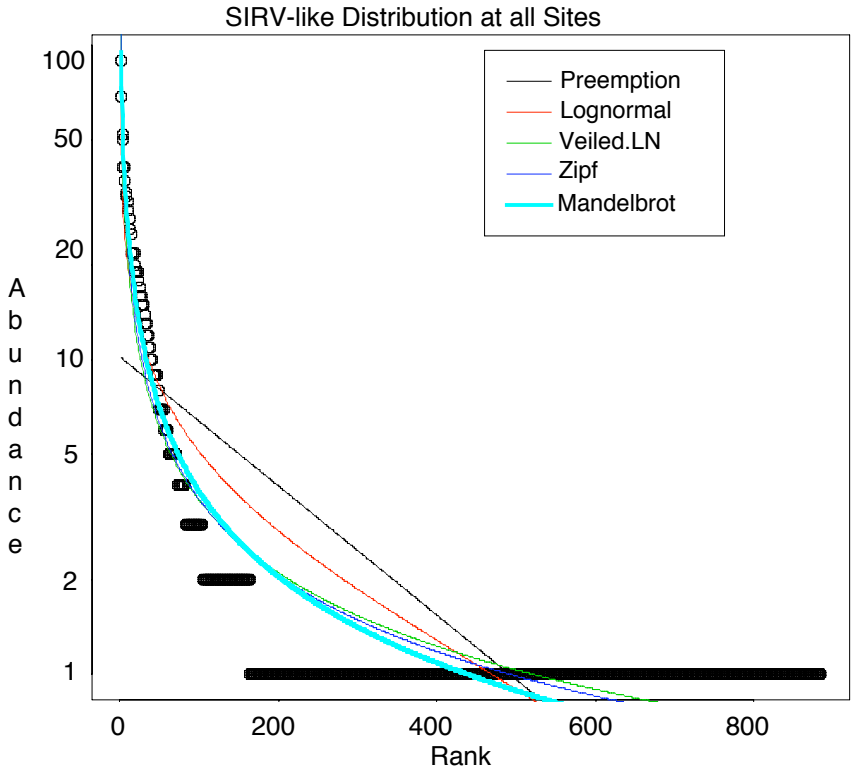


Figure 6.11. The SIRV-like distribution in all monitor sites over time. Of all the models tested, the Mandelbrot distribution best fit the data.



The Rabbit Creek SIRV-like population was the most stable of the three monitor sites (figure 4.23). The largest group of viruses was consistently detected from this monitor site over time (figure 4.23; represented by dark blue ovals). The minor groups of viruses were also consistently detected, however, their presence appeared to fluctuate over time (figure 4.23). It will be interesting to determine whether the dominant group (represented by blue in figure 4.23) is still consistently detected over time or whether it disappears and another group of SIRV-like viruses becomes dominant.

The Ragged Hills and Crater Hills SIRV-like populations were both diverse over time. The eight SIRV-like viral groups in Ragged Hills were detected in most of the

samples collected from this monitor site (figure 5.19). As found in Rabbit Creek, one group was reliably detected throughout this study (blue ovals in figure 5.19). However, many of the smaller groups were also commonly detected in Ragged Hills (figure 5.19). In the last sample collected from Ragged Hills (June 3, 2004), a new group of SIRV-like viruses (represented by turquoise in figure 5.19) is obvious and appears to be dominant. It will be interesting to determine whether this indicates a change occurring in the dominant SIRV-like population in Ragged Hills.

The Crater Hills SIRV-like population was the most diverse of the three monitor sites over time. The population in Crater Hills was also dominated by one group of SIRV-like sequences throughout this research (figure 3.22; represented by blue ovals). However, this dominant group of viruses was not as consistently detected as it was in the other two monitor sites (figure 4.23 and 5.19). The diversity in Crater Hills was continually cycling from many groups of sequences to only one or two groups of SIRV-like sequences (figure 3.22). In the last three samples collected from Crater Hills (February 2004, April 2004, and June 2004), a shift in the SIRV-like population is apparent. The group of viruses that has dominated Crater Hills throughout this research (represented by blue in figure 3.22) is no longer detected, and three new groups appear (represented by purple, orange, and dark blue in figure 3.22).

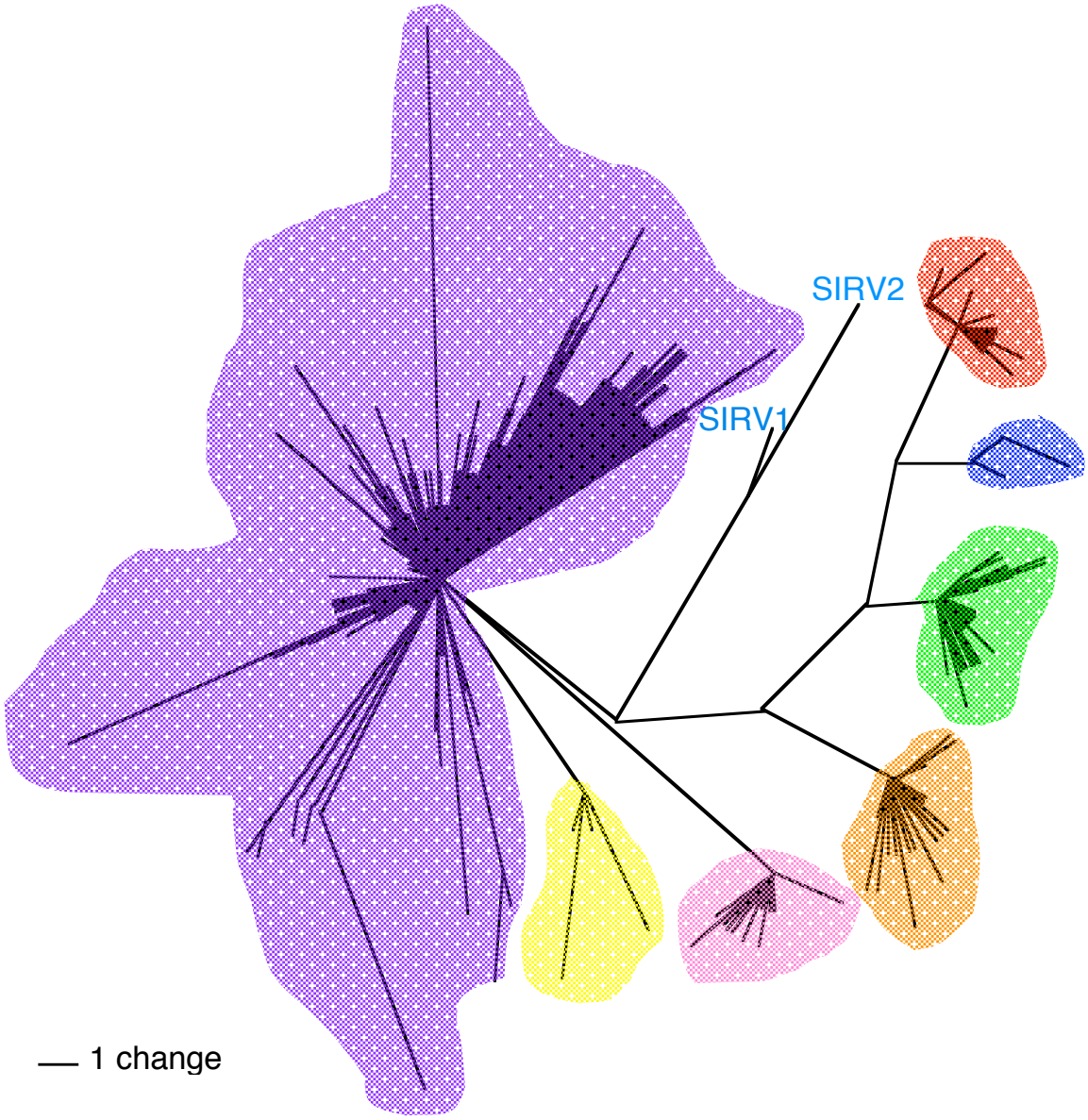
The SIRV-like inferred protein sequences detected in each site did not show the drastic changes visible in the DNA sequences over time (figure 3.23, 4.24, and 5.20). Overall, the SIRV-like protein sequences comprised seven groups over time (figure 6.12). The largest group (purple in figure 6.12) was consistently detected from each of the

monitor sites throughout this study, excluding the October 17, 2002 sample from Ragged Hills (figures 3.23, 4.24, and 5.20). This was the only group of protein sequences that was detected in all three monitor sites. The remaining six groups of SIRV-like protein sequences were detected in one or two monitor sites, but at varying times. Two groups of SIRV-like protein sequences were only detected in the Rabbit Creek and Ragged Hills monitor sites (represented by pink and green in figure 6.12). These two groups were not present in Rabbit Creek and Ragged Hills at the same times (figures 4.24 and 5.20). Two SIRV-like protein groups were exclusively found in Crater Hills (represented by yellow and blue in figure 6.12). These two groups were detected in October 2002 and January 2003 in Crater Hills, but were not detected again (figure 3.23). One group of SIRV-like protein sequences was exclusively detected in Ragged Hills (represented by orange in figure 6.12). This group was detected sporadically in the samples collected from Ragged Hills (figure 5.20). Finally, one group was exclusively detected in the Rabbit Creek monitor site (represented by red in figure 6.12). The presence of this group fluctuated over time in Rabbit Creek (figure 4.24). However, the majority of protein sequences collected in June 2004 from Rabbit Creek belong to this group (figure 4.24).

The SIRV-like populations appear to be undergoing a change in the last samples collected Crater Hills and Ragged Hills (figures 3.22 and 5.19). It is interesting the SIRV-like population shifts in February 2004 in Crater Hills, but a change in Ragged Hills is not apparent until June 2004. It will prove to be interesting to determine whether the SIRV-like population in each site underwent a permanent change in the dominant

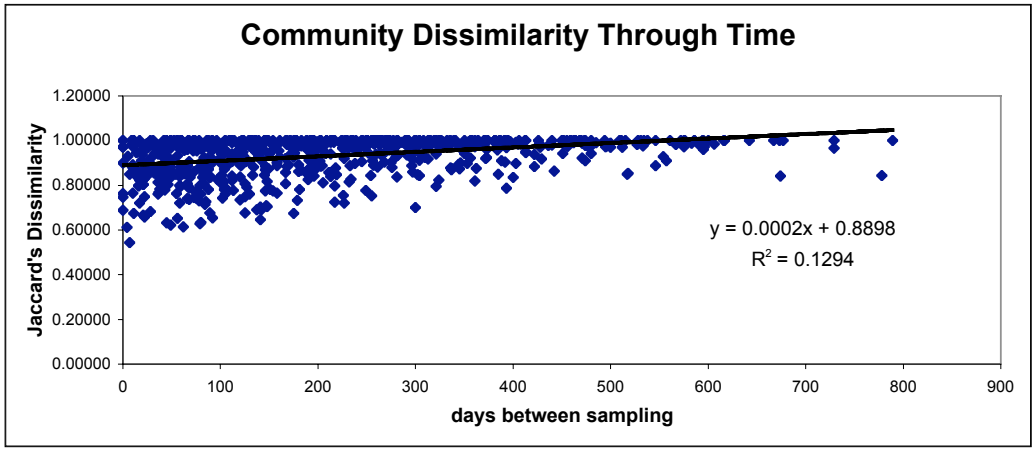
virus group or whether the change was temporary, and whether a change in the SIRV-like population occurs within the Rabbit Creek monitor site.

Figure 6.12. An unrooted phylogram of SIRV-like protein sequences from all three monitor sites collected during this research. A maximum parsimony analysis conducted with PAUP 4.0b10 (Swofford 2002) with 10,000 bootstrap replicates.



The SIRV-like populations in the three monitor sites were very diverse and dynamic over time. Identical SIRV-like sequences were detected from each monitor site, but they were not detected during the same time. Like the SSV-like sequences in YNP, it would be impossible to predict the SIRV-like population at a particular time and place. The SIRV-like protein sequences were more conserved than the DNA sequences. However, each monitor site contained SIRV-like protein groups that were exclusive to that particular monitor site. Presumably these protein groups would be detected in all three monitor sites with enough sampling and analysis of sequences.

Figure 6.13. A dissimilarity index of the SSV-like and SIRV-like sequences collected over time from all three monitor sites over time.



When the two virus populations are combined from all three monitor sites, the time between collection of samples and differences in viral sequences are not correlated (figure 6.13). The Jaccard's Dissimilarity Index measures dissimilarity in sequences over time (Faith *et. al.* 1987). However, this index does not factor shared absences of sequences, therefore a bias is present towards shared similarity. Regardless, even

sequences collected on the same day from different monitor sites (0 on the x-axis in figure 6.13) can be 100% dissimilar. This result indicates that we are detecting high change over in these viral sequences over time.

Discussion

This was the first attempt to continuously monitor two populations of hyperthermophilic viruses in three acidic hot springs in YNP. Based on results of previous research, we chose to monitor SSV-like and SIRV-like viruses from Crater Hills, Rabbit Creek, and Ragged Hills over the course of two years. We also monitored the dominant archaeal populations and the water chemistry in each of the sites.

Through this research, we were able to demonstrate the ability to successfully amplify both viral and archaeal DNA from high temperature acidic environments. Based on sequenced viral genomes, primers were designed to amplify conserved genes in SSVs and SIRVs. These primers were used to determine the viral population structure and how these virus populations might change over time in three acidic hot springs in YNP. We were able to demonstrate that the virus populations in these thermal areas are very dynamic.

The water chemistry was analyzed each time a sample was collected from each of the monitor sites. There were significant differences in the water chemistry of each site, but each site was relatively stable over time. Each monitor site had its own unique archaeal community, which may be influenced by the unique chemical environment.

However, despite the chemical uniqueness of each monitor site, identical viral sequences were detected in all monitor sites.

The archaeal populations in each monitor site were determined via 16S rDNA gene sequence analysis. Based on previous research, it was expected the dominant members of each site would be *Sulfolobus* species. However, *Sulfolobus* was present in each site at low abundance based on its 16S rDNA frequency in clone libraries. A total of nine different organisms were detected from the three sites, however, most of the environmental sequences showed low identity to published 16S rDNA sequences. The archaeal populations in these hot springs are clearly related to known organisms, but most of the organisms inhabiting the acidic thermal springs in YNP are likely unidentified species. Perhaps the difference in the dominant archaeal populations reflects the unique chemical environment of each hot spring. Most of the organisms we detected were found to be present in more than one monitor site; however, *Sulfolobus*-like sequences were the only ones that were detected in all three sites (table 6.6). Although each monitor site was dominated by different archaeal 16S rDNA sequences, both SSVs and SIRVs were detected. These results suggest that *Sulfolobus* species are not dominant organisms of the microbial communities within the monitor sites, but the viruses infecting *Sulfolobus* are present in high copy numbers.

The SSV-like population in the monitor sites was detected by sequencing two different conserved genes in the SSV genome. The primer set 3F/4R amplifies a portion of the largest predicted ORF in each virus, and the 7F/8R primer set amplifies the viral coat proteins. Surprisingly, the 3F/4R primer set yielded the best results; the coat protein

primers were successful in only four samples collected during this research. The SSV-like sequences in the three monitor sites were not always detectable. Overall, the SSV-like sequences from each site were similar and showed the same population structures (figure 6.5). Each monitor site contained a large group of sequences not closely related to any of the four sequenced SSV isolates, a group of SSV-like sequences closely related to SSV-K1, and a distinct group related to SSV1 (figure 6.5). This result indicates there is migration of these viruses throughout thermal environments around the world.

Changes in the SSV-like 3F/4R DNA sequences did not always result in changes in the inferred protein sequences. Over time, there were two major groups of protein sequences detected in the three monitor sites (figure 6.8). The first group was very large and was closely related to SSV2, SSV-RH, and SSV-K1 (figure 6.8). The second group of protein sequences was distinct from the larger group and closely related to SSV1 (figure 6.8). These results indicate SSV-like 3F/4R inferred protein sequences were more conserved over time or the changes in the SSV-like DNA sequences are occurring in the third position of the codons. Indeed, molecular analysis of the SSV-like DNA sequences reveals 34 changes in the third position compared to 25 changes in the first and second positions of the codon combined (table 6.8).

It had been previously hypothesized the genetic differences between the four sequenced SSV isolates were due to the fact the four viruses had been isolated from each other and had evolved independently because of functioning in different chemical environments (Wiedenheft *et. al.* 2004). My results suggest the differences in the four

sequenced SSV isolates is the result of a culturing bias and the large amount of viral diversity present in thermal environments.

The SIRV-like population was monitored by sequencing the viral coat protein in each of the three monitor sites. As opposed to the SSV-like sequences, SIRV-like coat protein sequences were detectable in all samples (excluding one from Ragged Hills) collected during this research. Each monitor site contained a dominant group of SIRV-like viruses that was consistently detected for most of this research, and a few smaller groups of viruses that appeared and disappeared over time (figures 3.22, 4.23, and 5.19). Near the conclusion of this research, a change in the dominant SIRV-like group in Crater Hills and Ragged Hills was detected (figures 3.22, and 5.19). These changes do not correlate with chemical changes or changing host populations. It will be interesting to determine whether this was a temporary or permanent change in the SIRV-like viruses, and if over time the SIRV-like population in Rabbit Creek will undergo a similar change.

There were numerous SIRV-like groups based on DNA sequences in each monitor site (nine in Crater Hills, eight in Ragged Hills, and five in Rabbit Creek); however, there were only seven groups of viruses based on inferred protein sequences (figure 6.12). The majority of SIRV-like protein sequences belonged in a large group that was detectable in all samples collected from each monitor site during this research (excluding the October 17, 2002 sample collected from Ragged Hills) (see figures 3.23, 4.24, and 5.20). This was the only group of SIRV-like protein sequences that was detected in all three monitor sites over time. The remaining groups of SIRV-like protein sequences represent viruses detected in one or two sites (figure 6.12).

The SIRV-like populations in the three monitor sites were more diverse than the SSV-like populations detected over time. Perhaps SIRV-like viruses are more diverse than SSV-like viruses as part of a replication strategy. If numerous variants are produced, the likelihood of a variant surviving in the environment will increase.

The SSV-like and SIRV-like viruses in the three monitor sites were very diverse. Even though we continually sampled for two years from three hot springs, we did not capture the diversity of virus sequences present in the environment. The amount of viral diversity that was detected was astounding (figures 6.6 and 6.10). In order to fully understand the viral populations in these hot springs, many more sequences need to be collected and analyzed. It is difficult to predict how many more viral sequences would need to be analyzed; it could potentially be a very large number. Over time the viral population structures were similar in each of the monitor sites. However, because the viruses are constantly changing, it would be very difficult to attain a complete understanding of the viral community within one sample, or even numerous samples. Therefore, attempting to describe a viral community over time is difficult because of the frequent change.

In this study, we were able to detect significant changes within one month of sampling. By studying viruses in these environments, we may begin to further understand how they function in the natural environment. However, the two viruses studied in this research were so dynamic that it would be difficult to achieve a complete understanding of the total virus populations in nature.

Implications of this research

To our knowledge, this is the first study to monitor a virus population in any natural environment over a two year period. Until now, it had been assumed that a single virus clade of a particular virus type would exist in an environment at any one time. However, this research has revealed that multiple virus clades can be simultaneously present in an environment at any one time. Moreover, this research revealed that virus population structure for any particular virus type may change drastically over a relatively short period of time (for example, less than 30 days).

There are two possible explanations for the considerable amount of diversity of viruses observed within the acidic hot springs within YNP: migration and immigration of viruses into and out of each spring or mutation within the resident virus population. The diversity of the SSV-like and SIRV-like viruses in YNP could be attributed to migration and immigration of these viruses between different acidic thermal features. The large number of viral sequences detected only one time in this research suggests that the viruses are not subject to high selection pressures in the acidic thermal environments of YNP. In other words, there may not be an advantage of one viral clade over another viral clade in these environments. The other plausible explanation for the diversity of viruses within YNP is a high rate of mutation within the resident virus population. The diversity of these viruses may be a result of the frequent change occurring in viruses (during replication, for example). However, such rampant change has been suggested for RNA viruses, but not for DNA viruses. It seems unlikely that mutations within the resident virus population could be the reason for change in the dominant virus group. It is more

probable that each monitor site is subject to migration and immigration of viruses. Both of these models propose that there is little or no selection of viruses within these environments.

The results of this study suggest biogeographical isolation of viruses may be dependent on the particular virus type. The SSV-like viruses from acidic thermal areas around the world do not appear to be biogeographically isolated. SSV-like sequences were detected from YNP that were more similar to the SSVs from Japan and Kamchatka than they were to other SSV-like sequences from YNP. Given a large enough sampling size, there is no evidence for biogeographical isolation of SSVs. I would hypothesize that repeating this research in Japan or Kamchatka would reveal similar results in the overall population structure of SSV-like viruses. In contrast, the SIRV-like viruses may be more subject to biogeographical isolation than are the SSV-like viruses. Even though more SIRV-like sequences were analyzed, there were not any sequences from YNP that were closely related to the SIRVs isolated from Iceland (figure 6.12). The different conclusions from each virus type may be a consequence of differences in replication cycles or differences in selection pressures on each virus.

The results of this research impact our current understanding of biodiversity. It is likely the diversity of organisms inhabiting thermal acidic features worldwide is much higher than previously thought. Significant diversity was detected in viral sequences over a two year period. In fact, we were unable to sample the total diversity of either virus as indicated by sequence abundance curves (figures 6.6 and 6.10). The large amount of diversity in viruses may be more easily detected because of the short

generation time of viruses compared to other organisms. Viral diversity detected in the ocean has also been significant. Viruses in the marine environment have been detected in numbers as high as 10^8 virus particles/mL seawater. Viruses may represent the dominant sequences on the planet, both in numbers and diversity.

Summary

Throughout this research, the water chemistry of the three monitor sites was relatively stable over time. However, the water chemistry in each site was unique. The archaeal population slowly changed and was unique in each site, perhaps reflecting the unique water chemistry in each monitor site. Even though the water chemistry and dominant archaeal populations were unique in each site, the detected viral population structures were similar. The viruses rapidly changed over time, so it would be difficult to predict the population structure of these viruses at any particular time and place.

CHAPTER 7

CONCLUSIONS TO QUESTIONS AND HYPOTHESIS

Question 1

Our goal was to determine whether virus populations in high temperature acidic environments within YNP change over time. To our knowledge, this is the first time anyone has attempted to monitor any population directly in the environment. We were able to detect significant changes within two virus populations (SSV-like and SIRV-like viruses) in three monitor sites over a two year period. The overall population structures of the SSV-like viruses from the three monitor sites were similar. However, each site demonstrated independent viral population dynamics. The SSV-like viruses were not detected in all the monitor sites at the same times. The SIRV-like viruses were more diverse and dynamic than the SSV-like viruses. Similar to the SSV-like viruses, identical SIRV-like sequences were detected in all the monitor sites over time. However, these identical sequences were not detected at the same time in each monitor site. The virus populations changed frequently, even more so than the potential host populations. In fact, we have determined it would be difficult to fully characterize an individual viral population over time because of the frequent and rapid change occurring in the virus populations.

Question 2

A second objective of this research was to determine whether different viruses undergo the same dynamics or whether they change independently of each other. We discovered that change in the two virus populations is independent. The SSV-like viruses appeared to be more structured and the same groups were found in all monitor sites over time. The SIRV-like viruses were more diverse than the SSV-like viruses. Larger groups of SIRV-like viruses were detected in each monitor site, but we also detected numerous groups present in only one of the sites. Moreover, changes in one viral population did not necessitate changes in the other viral population.

Question 3

We also investigated if there was a difference in viral populations inhabiting different thermal acidic sites in YNP. The overall population structure of each virus in different hot springs was not significantly different. However, even though the population structure of both viruses over time was similar in each hot spring, the viral population structures at individual time points were unique. Overall, there were three major clades of SSV-like viruses detected during this research in the monitor sites. However, all of the viral clades were not present in the monitor sites at the same times. Like the SSV-like viruses, the SIRV-like viruses comprised numerous groups that were common to all three monitor sites. However, the presence of the SIRV-like clades varied widely. The overall population structures of each virus should not be used to characterize

a viral community over time because changes in the viral populations occurred frequently during smaller time scale.

Question 4

Finally, we investigated whether the changes in virus populations could be correlated to changing host populations or changing water chemistry. We determined the virus populations in the thermal acidic environments in YNP are very dynamic. However, those changes appear to be independent of water chemistry or host populations. Attempts to correlate viral change with water chemistry did not produce significant results. The water chemistry of the three monitor sites was relatively constant during this research. The water chemistry in each monitor was unique. Surprisingly, we were able to detect both viruses (SSV-like and SIRV-like viruses) in the three sites. Both virus populations could change drastically even when the water chemistry did not significantly change. Likewise, the changes in host populations could not be correlated to changes in virus populations. Each of the monitor sites contained a unique dominant archaeal population, which could vary over time. Regardless, we were able to amplify both viruses from each monitor site. Even with changing dominant hosts, virus populations were still detectable.

Hypothesis

We hypothesized that changes in viral population structures would be a sensitive indicator of change in the environment. Through this research we determined the viral

populations are changing independently of the factors we measured (host populations and water chemistry). In fact, the viruses are changing at such high rates that it would be difficult to fully describe a virus population in the high temperature acidic sites within YNP. Therefore, because there are drastic changes in virus populations within short periods, they would not be sensitive indicators of change within a high temperature acidic environment.

REFERENCES CITED

Allen, E.T. and A.L. Day (1935). The Hot Springs of the Yellowstone National Park. Washington, D.C., Carnegie Institution of Washington.

Aoshima, M., Y. Nishibe, M. Hasegawa, A. Yamagishi and T. Oshima (1996). "Cloning and sequencing of a gene encoding 16S ribosomal RNA from a novel hyperthermophilic archaebacterium NC12." *Gene* **180**: 183-187.

Arnold, H.P., K.M. Stedman and W. Zillig (1999). Archaeal Phages. Encyclopedia of Virology, 2nd edition. R.G. Webster and A.Granoff. London, Academic Press, Ltd: 76-89.

Arnold, H.P., U. Ziese and W. Zillig (2000a). "SNDV, a novel virus of the extremely thermophilic and acidophilic archaeon *Sulfolobus*." *Virology* **272**: 409-416.

Arnold, H.P., W. Zillig, U. Ziese, I. Holz, M. Crosby, T. Utterback, J.F. Weidmann, J.K. Kristjanson, H.P. Klenk, K.E. Nelson and C.M. Fraser (2000b). "A novel lipothrixvirus, SIFV, of the extremely thermophilic crenarchaeon *Sulfolobus*." *Virology* **267**: 252-266.

Baker, G.C., J.J. Smith and D.A. Cowan (2003). "Review and re-analysis of domain-specific 16S primers." *Journal of Microbiological Methods* **55**: 541-555.

Barns, S.M., C.F. Delwiche, J.D. Palmer and N.R. Pace (1996). "Perspectives on archaeal diversity, thermophily, and monophyly from environmental rRNA sequences." *Proceedings of the National Academy of Science USA* **93**: 9188-9193.

Barns, S.M., R.E. Fundyga, M.W. Jefferies and N.R. Pace (1994). "Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment." *Proceedings of the National Academy of Science USA* **91**: 1609-1613.

Bath, C. and M.L. Dyall-Smith (1998). "His1: an archaeal virus of the Fusselloviridae family that infects *Haloarcula hispanica*." *Journal of Virology* **72**: 9392-9395.

Bertani, G. and L. Baresi (1986). "Looking for gene transfer mechanisms in methanogenic bacteria." Archaeabacteria '85. O. Kandler and W. Zillig. Stuttgart, Gustav Fisher Verlag: 398.

Bettstetter, M., X. Peng, R.A. Garrett and D. Prangishvili (2003). "AFV1, a novel virus infecting hyperthermophilic archaea of the genus *Acidianus*." *Virology* **315**: 68-79.

Birkenbihl, R., K. Neet, D. Prangishvili and B. Kemper (2001). "Holliday Junction resolving enzymes of archaeal viruses SIRV1 and SIRV2." *Journal of Molecular Biology* **309**: 1067-1076.

Blank, C.E., S.L. Cady and N.R. Pace (2002). "Microbial composition of near-boiling silica-depositing thermal springs throughout Yellowstone National Park." *Applied and Environmental Microbiology* **68**: 5123-5135.

Breitbart, M., P. Salamon, B. Andresen, J.M. Mahaffy, A.M. Segall, D. Mead, F. Azam and F. Rohwer (2002). "Genomic analysis of uncultured marine viral communities." *Proceedings of the National Academy of Sciences USA* **99**: 14250-14255.

Brock, T.D., K.M. Brock, R.T. Belly and R.L. Weiss (1972). "*Sulfolobus*: A new genus of sulfur-oxidizing bacteria living at low pH and high temperature." *Archives of Microbiology* **84**: 54-68.

Bult, C.J., O. White, G.J. Olsen, L. Zhou, R.D. Fleischmann, G.G. Sutton, J.A. Blake, L. M. FitzGerald, R.A. Clayton, J.D. Gocayne, A.R. Kerlavage, B.A. Dougherty, J.-F. Tomb, M.D. Adams, C.I. Reich, R. Overbeek, E.F. Kirkness, K.G. Weinstock, J.M. Merrick, A. Glodek, J.L. Scott, N.S.M. Geoghagen, J.F. Weidman, J.L. Fuhrmann, D. Nguyen, T.R. Utterback, J.M. Kelley, J.D. Peterson, P.W. Sadow, M.C. Hanna, M.D. Cotton, K.M. Roberts, M.A. Hurst, B.P. Kaine, M. Borodovsky, H.-P. Klenk, C.M. Fraser, H.O. Smith, C.R. Woese and J.C. Venter (1996). "Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*." *Science* **273**: 1058-1073.

Burgraff, S., K.O. Stetter, P. Rouviere and C.R. Woese (1991). "*Methanopyrus kandleri*: an archaeal methanogen unrelated to all other known methanogens." *Systematic and Applied Microbiology* **14**: 346-351.

Cancho, R.F.i. and R.V. Sole (2003). "Least effort and the origins of scaling in human language." *Proceedings of the National Academy of Science USA* **100**: 788-791.

Daniels, L.L. and A.C. Wais (1984). "Restriction and modification of halophage S45 in *Halobacterium*." *Current Microbiology* **10**: 133-136.

DeLong, E.F. (1992). "Archaea in coastal marine environments." *Proceedings of the National Academy of Science USA* **89**: 5685-5689.

DeLong, E.F., K.Y. Wu, B.B. Prezelin and R.V.M. Jovine (1994). "High abundance of Archaea in Antarctic marine picoplankton." *Nature* **371**: 695-697.

Deppenmeier, U., A. Johann, T. Hartsch, R. Merkl, R.A. Schmitz, R. Martinez-Arias, A. Henne, A. Wiezer, S. Baumer, C. Jacobi, H. Bruggemann, T. Lienard, A. Christmann, M. Bomeke, S. Steckel, A. Bhattacharyya, A. Lykidis, R. Overbeek, H.P. Klenk, R.P. Gunsalus, H.J. Fritz and G. Gottschalk (2002). "The genome of *Methanosarcina mazei*: evidence for lateral gene transfer between bacteria and archaea." *Journal of Molecular Microbiology and Biotechnology* **4**: 453-461.

Dyall-Smith, M., S.-L. Tang and C. Bath (2003). "Haloarchaeal viruses: how diverse are they?" *Research in Microbiology* **154**: 309-313.

Eiserling, F., A. Pushkin, M. Gingery and G. Bertani (1999). "Bacteriophage-like particles associated with the gene transfer agent of *Methanococcus voltae* PS." *Journal of General Virology* **80**: 3305-3308.

Faith D.P., P.R. Minchin, and L. Belbin (1987). "Compositional dissimilarity as a robust measure of ecological distance." *Vegetatio* **69**: 57-68.

Ferguson, N.M., A.P. Galvani and R.M. Bush (2003). "Ecological and immunological determinants of influenza evolution." *Nature* **422**: 428-433.

Fitz-Gibbon, S.T., H. Ladner, U.-J. Kim, K.O. Stetter, M.I. Simon and J.H. Miller (2002). "Genome sequence of the hyperthermophilic crenarchaeon *Pyrobaculum aerophilum*." *Proceedings of the National Academy of Science USA* **99**: 984-989.

Fuhrman, J.A., K. McCallum and A.A. Davis (1993). "Novel major archaebacterial group from marine plankton." *Nature* **356**: 148-149.

Fukui, T., H. Atomi, T. Kanai, R. Matsumi, S. Fujiwara and T. Imanaka (2005). "Complete genome sequence of hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1 and comparison with *Pyrococcus* genomes." *Genome Research* **15**: 352-363.

Futterer, O., A. Angelov, H. Liesegang, G. Gottschalk, C. Schleper, B. Schepers, C. Dock, G. Antranikian and W. Liebl (2004). "Genome sequence of *Picrophilus torridus* and its implications for life around pH 0." *Proceedings of the National Academy of Science USA* **101**: 9091-9096.

Galagan, J.E., C. Nusbaum, A. Roy, M.G. Endrizzi, P. Macdonald, W. FitzHugh, S. Calvo, R. Engels, S. Smirnov, D. Atnoor, A. Brown, N. Allen, J. Naylor, N. Stange-Thomann, K. DeArellano, R. Johnson, L. Linton, P. McEwan, K. McKernan, J. Talamas, A. Tirrell, W. Ye, A. Zimmer, R.D. Barber, I. Cann, D.E. Graham, D.A. Grahame, A.M. Guss, R. Hedderich, C. Ingram-Smith, H.C. Kuettner, J.A. Krzycki, J.A. Leigh, W. Li, J. Liu, B. Mukhopadhyay, J.N. Reeve, K. Smith, T.A. Springer, L.A. Umayam, O. White, R.H. White, E.C.d. Macario, J. G. Ferry, K.F. Jarrell, H. Jing, A.J.L. Macario, I. Paulsen, M. Pritchett, K.R. Sowers, R.V. Swanson, S.H. Zinder, E. Lander, W.W. Metcalf and B. Birren (2002). "The genome of *M. acetivorans* reveals extensive metabolic and physiological diversity." *Genome Research* **12**: 532-542.

Garrity, G.M. (2001). The Archaea and the Deeply Branching and Phototrophic Bacteria. New York, Springer.

Geslin, C., M.L. Romancer, G. Erauso, M. Gaillard, G. Perrot and D. Prieur (2003a). "PAV1, the first virus-like particle isolated from a hyperthermophilic euryarchaeote, "*Pyrococcus abyssi*"." *Journal of Bacteriology* **185**: 3888-3894.

Geslin, C., M.L. Romancer, M. Gaillard, G. Erauso and D. Prieur (2003b). "Observation of virus-like particles in high temperature enrichment cultures from deep-sea hydrothermal vents." *Research in Microbiology* **154**: 303-307.

Grogan, D., P. Palm and W. Zillig (1990). "Isolate B12, which harbours a virus-like element, represents a new species of the archaebacterial genus *Sulfolobus*, *Sulfolobus shibatae*, sp. nov." *Archives in Microbiology* **154**: 594-599.

Haring, M., X. Peng, K. Brugger, R. Rachel, K.O. Stetter, R.A. Garrett and D. Prangishvili (2004). "Morphology and genome organization of the virus PSV of the hyperthermophilic archaeal genera *Pyrobaculum* and *Thermoproteus*: a novel virus family, the *Globuloviridae*." *Virology* **323**: 233-242.

Hasegawa, M., K. Kishino and T. Yano (1985). "Dating the human ape splitting by a molecular clock of mictochondrial DNA." *Journal of Molecular Evolution* **22**: 160-174.

Higgins, D.G., J.D. Thompson and T.J. Gibson (1996). "Using CLUSTAL for multiple sequence alignments." *Methods in Enzymology* **266**: 383-402.

Hjorleifsdottir, S., S. Skirnisdottir, G.O. Hreggridsson, O. Holst and J.K. Kristjansson (2001). "Species comparison of cultivated and non-cultivated bacteria from short filaments in an Icelandic hot spring at 88 degrees C." *Microbial Ecology* **42**: 117-125.

Hohn, M.J., B.P. Hedlund and H. Huber (2002). "Detection of 16S DNA sequences representing the novel phylum "Nanoarchaeota": indication for a wide distribution in high temperature biotopes." *Systematic and Applied Microbiology* **25**: 551-554.

Huber, H., M.J. Hohn, R. Rachel, T. Fuchs, V.C. Wimmer and K.O. Stetter (2002). "A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont." *Nature* **417**: 63-67.

Huber, H. and K.O. Stetter (1998). "Hyperthermophiles and their possible potential in biotechnology." *Journal of Biotechnology* **64**: 39-52.

Huelsenbeck, J.P. and F. Ronquist (2001). "MRBAYES: Bayesian inference of phylogenetic trees." *Bioinformatics* **17**: 754-755.

Hugenholtz, P., C. Pitulle, K.L. Hershberger and N.R. Pace (1998). "Novel division level bacterial diversity in a Yellowstone hot spring." *Journal of Bacteriology* **180**: 366-376.

Hughes, A.L. and R. Friedman (2003). "Genome-wide survey for genes horizontally transferred from cellular organisms to baculoviruses." *Molecular and Biological Evolution* **20**: 979-987.

Itoh, T., K.-i. Suzuki and T. Nakase (2002). "*Vulcanisaeta distributa* gen. nov., sp. nov., and *Vulcanisaeta souniana* sp. nov., novel hyperthermophilic, rod-shaped crenarchaeotes isolated from hot springs in Japan." *International Journal of Systematic and Evolutionary Microbiology* **52**: 1097-1104.

Janekovic, D., S. Wunderl, I. Holz, W. Zillig, A. Gierl and H. Neumann (1983). "TTV1, TTV2, and TTV3, a family of viruses of the extremely thermophilic, anaerobic, sulfur reducing archaebacterium *Thermoproteus tenax*." *Molecular and General Genetics* **192**: 39-45.

Jordan, M., L. Meile, and T. Leisinger (1989). "Organization of *Methanobacterium thermoautotrophicum* bacteriophage ΨM1 DNA." *Molecular and General Genetics* **220**: 161-164.

Kawarabayasi, Y., Y. Hino, H. Horikawa, K. Jin-no, M. Takahashi, M. Sekine, S.-i. Baba, A. Ankai, H. Kosugi, A. Hosoyama, S. Fukui, Y. Nagai, K. Nishijima, R. Otsuka, H. Nakazawa, M. Takamiya, Y. Kato, T. Yoshizawa, T. Tanaka, Y. Kudoh, J. Yamazaki, N. Kushida, A. Oguchi, K.-i. Aoki, S. Masuda, M. Yanagii, M. Nishimura, A. Yamagishi, T. Oshima and H. Kikuchi (2001). "Complete genome sequence of an aerobic thermoacidophilic crenarchaeon *Sulfolobus tokodaii* strain 7." *DNA Research* **8**: 123-140.

Kawarabayasi, Y., Y. Hino, H. Horikawa, J. Yamazaki, Y. Haikawa, K. Jin-no, M. Takahashi, M. Sekine, S.-i. Baba, A. Ankai, H. Kosugi, A. Hosoyama, S. Fukui, Y. Nagai, K. Nishijima, H. Nakazawa, M. Takamiya, S. Masuda, T. Funahashi, T. Tanaka, Y. Kudoh, J. Yamazaki, N. Kushida, A. Oguchi, K.-i. Aoki, K. Kubota, Y. Nakamura, N. Nomura, Y. Sako, and H. Kikuchi (1999). "Complete genome sequence of an aerobic hyper-thermophilic crenarchaeon, *Aeropyrum pernix* K1." *DNA Research* **6**: 83-101, 145-152.

Kawarabayasi, Y., M. Sawada, H. Horikawa, Y. Haikawa, Y. Hino, S. Yamamoto, M. Sekine, S.-i. Baba, H. Kosugi, A. Hosoyama, Y. Nagai, M. Sakai, K. Ogura, R. Otsuka, H. Nakazawa, M. Takamiya, Y. Ohfuku, T. Funahashi, T. Tanaka, Y. Kudoh, J. Yamazaki, N. Kushida, A. Oguchi, K.-i. Aoki, T. Yoshizawa, Y. Nakamura, F. T. Robb, K. Horikoshi, Y. Masuchi, H. Shizuya and H. Kikuchi (1998). "Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, *Pyrococcus horikoshii* OT3." *DNA Research* **5**: 55-76.

Kawashima, T., N. Amano, H. Koike, S.-i. Makino, S. Higuchi, Y. Kawashima-Ohya, K. Watanabe, M. Yamazaki, K. Kanehori, T. Kawamoto, T. Nunoshiba, Y. Yamamoto, H. Aramaki, K. Makino and M. Suzuki (2000). "Archaeal adaptation to higher temperatures revealed by genomic sequence of *Thermoplasma volcanium*." *Proceedings of the National Academy of Science USA* **97**: 14257-14262.

Kimura, M. (1981). "Estimation of evolutionary distances between homologous nucleotide sequences." *Proceedings of the National Academy of Science USA* **78**: 454-458.

Klein, R., U. Baranbyi, N. Rossler, B. Greineder, H. Scholz and A. Witte (2002). "*Natrialba magadii* virus ϕCh1: first complete nucleotide sequence and functional organization of a virus infecting a haloalkaliphilic archaeon." *Molecular Microbiology* **45**: 851-863.

Klein, R., B. Greineder, U. Baranyi and A. Witte (2000). "The structural protein E of the archaeal virus ϕCh1: evidence for processing in *Natrialba magadii* during virus maturation." *Virology* **276**: 376-387.

Klenk, H.P., R.A. Clayton, J.-F. Tomb, O. White, K.E. Nelson, K.A. Ketchum, R.J. Dodson, M. Gwinn, E.K. Hickey, J.D. Peterson, D.L. Richardson, A.B. Kerlavage, D.E. Graham, N.C. Kyrpides, R.D. Fleischmann, J. Quackenbush, N. H. Lee, G.G. Sutton, S. Gill, E.F. Kirkness, B.A. Dougherty, K. McKenney, M. D. Adams, B. Loftus, S. Peterson, C.I. Relch, L.K. McNeil, J.H. Badger, A. Glodek, L. Zhou, R. Overbeek, J.D. Gocayne, J.F. Weidman, L. McDonald, T. Utterback, M.D. Cotton, T. Spriggs, P. Artiach, B.P. Kaine, S.M. Sykes, P.W. Sadow, K.P. D'Andrea, C. Bowman, C. Fujii, S.A. Garland, T.M. Mason, G.J. Olsen, C.M. Fraser, H.O. Smith, C.R. Woese and J.C. Venter (1997). "The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*." *Nature* **390**: 364-370.

Knox, M.R. and J.E. Harris (1986). Isolation and characterization of a bacteriophage of *Methanobrevibacter smithii*. XIV International Congress on Microbiology, Manchester.

Kraft, P., D. Kummel, A. Oeckinghaus, G.H. Gauss, B. Wiedenheft, M. Young and C.M. Lawrence (2004). "Structure of D63 from *Sulfolobus* spindle-shaped virus 1; surface properties of the dimeric four-helix bundle suggest an adapter protein function." *Journal of Virology* **78**: 7438-7442.

Kumar, S., K. Tamura and M. Nei (2004). "MEGA3: Integrated Software for Molecular Evolutionary Genetics Analysis and Sequence Alignment." *Briefings in Bioinformatics* **5**: In Press.

Luo, Y., P. Pfister, T. Leisinger and A. Wasserfallen (2001). "The genome of archaeal prophage ΨM100 encodes the lytic enzyme responsible for autolysis of *Methanothermobacter wolfeii*." *Journal of Bacteriology* **183**: 5788-5792.

Mandelbrot, B. (1966). *in* Readings in Mathematical Social Sciences. P. F. Lazarsfield and N. W. Henry. Cambridge, MA, MIT Press: 151-168.

Marteinsson, V.T., J.K. Kristjansson, H. Kristmannsdottir, M. Dahlquist, K. Saemundsson, M. Hannington, S.K. Petursdottir, A. Geptner and P. Stoffers (2001). "Discovery and description of giant submarine smectite cones on the seafloor in Eyjafjordur, Northern Iceland, a novel thermal microbial habitat." *Applied and Environmental Microbiology* **67**: 827-833.

Meile, L., U. Jenal, D. Studer, M. Jordan and T. Leisinger (1989). "Characterization of ψM1, a virulent phage of *Methanobacterium thermoautotrophicum* Marburg." *Archives of Microbiology* **152**: 105-110.

Muskhelishvili, G., P. Palm and W. Zillig (1993). "SSV1-encoded site-specific recombination system in *Sulfolobus shibatae*." *Molecular and General Genetics* **237**: 334-342.

Nadal, M., G. Mirambeau, P. Forterre, W.-D. Reiter and M. Duguet (1986). "Positively supercoiled DNA in a virus-like particle of an archaebacterium." *Nature* **321**: 256-258.

Nelson, K.E., R.A. Clayton, S.R. Gill, M.L. Gwinn, R.J. Dodson, D.H. Haft, E.K. Hickey, J.D. Peterson, W.C. Nelson, K.A. Ketchum, L. McDonald, T.R. Utterback, J.A. Malek, K.D. Linher, M.M. Garrett, A.M. Stewart, M.D. Cotton, M.S. Pratt, C.A. Phillips, D. Richardson, J. Heidelberg, G.G. Sutton, R.D. Fleischmann, J.A. Eisen and C.M. Fraser (1999). "Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*." *Nature* **399**: 323-329.

Nercessian, O., A.L. Reysenbach, D. Prieur and C. Jeanthon (2003). "Archaeal diversity associated with in situ samples deployed on hydrothermal vents on the East Pacific Rise." *Environmental Microbiology* **5**: 492-502.

Neumann, H. and W. Zillig (1990). "Structural variability in the genome of the *Thermoproteus tenax* virus TTV1." *Molecular and General Genetics* **222**: 435-437.

Ng, W. V., S.P. Kennedy, G.G. Mahairas, B. Berquist, M. Pan, H.D. Shukla, S.R. Lasky, N.S. Baliga, V. Thorsson, J. Sbrogna, S. Swartzell, D. Weir, J. Hall, T.A. Dahl, R. Welti, Y.A. Goo, B. Leithauser, K. Keller, R. Cruz, M.J. Danson, D.W. Hough, D.G. Maddocks, P.E. Jablonski, M.P. Krebs, C.M. Angevine, H. Dale, T.A. Isenbarger, R.F. Peck, M. Pohlschroder, J.L. Spudich, K.-H. Jung, M. Alam, T. Freitas, S. Hou, C.J. Daniels, P.P. Dennis, A.D. Omer, H. Ebhardt, T.M. Lowe, P. Liang, M. Riley, L. Hood and S. DasSarma (2000). "Genome sequence of *Halobacterium* species NRC-1." *Proceedings of the National Academy of Science USA* **97**: 12176-12181.

Nolling, J., A. Groffen and W.M.d. Vos (1993). "ϕF1 and ϕF3, two novel virulent, archaeal phages infecting different thermophilic strains of the genus *Methanobacterium*." *Journal of General Microbiology* **139**: 2511-2516.

Nuttall, S.D. and M.L. Dyall-Smith (1993). "HF1 and HF2: novel bacteriophages of halophilic archaea." *Virology* **197**: 678-684.

Palm, P., C. Schleper, B. Grampp, S. Yeats, P. McWilliam, W.-D. Reiter and W. Zillig (1991). "Complete nucleotide sequence of the virus SSV1 of the archaebacterium *Sulfolobus shibatae*." *Virology* **185**: 242-250.

Pauling, C. (1982). "Bacteriophages of *Halobacterium halobium*: isolated from fermented fish sauce and primary characterization." *Canadian Journal of Microbiology* **28**: 916-921.

Peng, X., H. Blum, Q. She, S. Mallok, K. Brugger, R.A. Garrett, W. Zillig and D. Prangishvili (2001). "Sequences and replication of genomes of the archaeal rudiviruses SIRV1 and SIRV2: relationships to the archaeal lipothrixvirus SIFV and some eukaryal viruses." *Virology* **291**: 226-234.

Pfeifer, F. (1987). Genetics of Halobacteria. Halophilic Bacteria. F. Rodriguez-Valera. Boca Raton, CRC Press: 105-133.

Pfister, P., A. Wasserfallen, R. Stettler and T. Leisinger (1998). "Molecular analysis of *Methanobacterium* phage ψM2." *Molecular Microbiology* **30**: 233-244.

Posada, D. and K.A. Crandall (1998). "MODELTEST: testing the model of DNA substitution." *Bioinformatics* **14**: 817-818.

Prangishvili, D., H.P. Arnold, D. Gotz, U. Ziese, I. Holz, J.K. Kristjansson and W. Zillig (1999). "A novel virus family, the Rudiviridae: structure, virus-host interactions and genome variability of the *Sulfolobus* viruses SIRV1 and SIRV2." *Genetics* **152**: 1387-1396.

Prangishvili, D., I. Holz, E. Stieger, S. Nickell, J.K. Kristjansson and W. Zillig (2000). "Sulfolobicins, specific proteinaceous toxins produced by strains of the extremely thermophilic archaeal genus *Sulfolobus*." *Journal of Bacteriology* **182**: 2985-2988.

Preston, C.M., K.Y. Wu, T.F. Molinski and E.F. DeLong (1996). "A psychrophilic crenarchaeon inhabits a marine sponge: *Cenarchaeum symbiosum* gen. nov., sp. nov." *Proceedings of the National Academy of Science USA* **93**: 6241-6246.

Rachel, R., M. Bettstetter, B.P. Hedlund, M. Haring, A. Kessler, K.O. Stetter and D. Prangishvili (2002). "Remarkable morphological diversity of viruses and virus-like particles in hot terrestrial environments." *Archives of Virology* **147**: 2419-2429.

Rambaut, A. (1997-2002). Sequence Alignment Editor. Oxford, University of Oxford.

Reiter, W.-D. and P. Palm (1990). "Identification and characterization of a defective SSV1 genome integrated into a tRNA gene in the archaebacterium *Sulfolobus* sp. B12." *Molecular and General Genetics* **221**: 65-71.

Reiter, W.-D., P. Palm, S. Yeats and W. Zillig (1987). "Gene expression in archaebacteria: physical mapping of constitutive and UV-inducible transcripts from the *Sulfolobus* virus-like particle SSV1." *Molecular and General Genetics* **209**: 270-275.

Reysenbach, A.-L., M. Ehringer and K. Hershberger (2000). "Microbial diversity at 83C in Calcite Springs, Yellowstone National Park: another environment where the *Aquificales* and "Korarchaeota" coexist." *Extremophiles* **4**: 61-67.

Rice, G., K. Stedman, J. Snyder, B. Wiedenheft, D. Willits, S. Brumfield, T. McDermott and M. J. Young (2001). "Viruses from extreme thermal environments." *Proceedings of the National Academy of Sciences USA* **98**: 13341-13345.

Rice, G., L. Tang, K. Stedman, F. Roberto, J. Spuhler, E. Gillitzer, J.E. Johnson, T. Douglas and M.J. Young (2004). "The structure of a thermophilic archaeal virus shows a double-stranded DNA viral capsid type that spans all domains of life." *Proceedings of the National Academy of Sciences* **101**: 7716-7720.

Robb, F.T., D.L. Maeder, J.R. Brown, J. DiRuggiero, M.D. Stump, R.K. Yeh, R.B. Weiss and D.M. Dunn (2001). "Genomic sequence of hyperthermophile, *Pyrococcus furiosus*: implications for physiology and enzymology." *Methods in Enzymology* **330**: 134-157.

Rodriguez, F., J. L. Oliver, A. Marin and J. R. Medina (1990). "The general stochastic model of nucleotide substitution." *Journal of Theoretical Biology* **142**: 485-501.

Schleper, C., K. Kubo and W. Zillig (1992). "The particle SSV1 from the extremely thermophilic archaeon *Sulfolobus* is a virus: demonstration of infectivity and of transfection with viral DNA." *Proceedings of the National Academy of Sciences USA* **89**: 7645-7649.

Schleper, C., R.V. Swanson, E.J. Mathur and E.F. DeLong (1997). "Characterization of a DNA polymerase from the uncultivated psychrophilic archaeon *Cenarchaeum symbiosum*." *Journal of Bacteriology* **179**: 7803-7811.

Schnabel, H. and W. Zillig (1984). "Circular structure of the genome of phage ϕH in a lysogenic *Halobacterium halobium*." *Molecular and General Genetics* **193**: 422-426.

Segerer, A., A. Neuner, J.K. Kristjansson and K.O. Stetter (1986). "*Acidianus infernus* ge. nov., sp. nov., and *Acidianus brierleyi* comb. nov.: Facultatively aerobic, extremely acidophilic thermophilic sulfur-metabolizing Archaebacteria." *International Journal of Systematic Bacteriology* **36**: 559-564.

Segerer, A., K.O. Stetter and F. Klink (1985). "Two contrary modes of chemolithotrophy in the same archaebacterium." *Nature* **313**: 787-789.

Segerer, A.H., A. Trincone, M. Gahrtz and K.O. Stetter (1991). "*Stygiolobus azoricus* gen. nov., sp. nov. represents a novel genus of anaerobic, extremely thermoacidophilic archaebacteria of the order *Sulfolobales*." *International Journal of Systematic Bacteriology* **41**: 495-501.

She, Q., R.K. Singh, F. Confalonieri, Y. Zivanovic, G. Allard, M.J. Awayez, C.C.-Y. Chan-Weiher, I.G. Clausen, B.A. Curtis, A.D. Moors, G. Erauso, C. Fletcher, P. M.K. Gordon, I.H.-d. Jong, A.C. Jeffries, C.J. Kozera, N. Medina, X. Peng, H.P. Thi-Ngoc, P. Redder, M.E. Schenk, C. Theriault, N. Tolstrup, R.L. Charlebois, W.F. Doolittle, M. Duguet, T. Gaasterland, R.A. Garrett, M.A. Ragan, C.W. Sensen and J. Van der Oost (2001). "The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2." *Proceedings of the National Academy of Science USA* **98**: 7835-7840.

Slesarev, A.I., K.V. Mezhevaya, K.S. Makarova, N.N. Polushin, O.V. Shcherbinina, V. V. Shakhova, G.I. Belova, L. Aravind, D.A. Natale, I.B. Rogozin, R.L. Tatusov, Y.I. Wolf, K.O. Stetter, A.G. Malykh, E.V. Koonin and S.A. Kozyavkin (2002). "The complete genome of hyperthermophile *Methanopyrus kandleri* AV19 and monophyly of archaeal methanogens." *Proceedings of the National Academy of Science USA* **99**: 4644-4649.

Smith, D.R., L.A. Doucette-Stamm, C. Deloughery, H. Lee, J. Dubois, T. Aldredge, R. Bashirzadeh, D. Blakely, R. Cook, K. Gilbert, D. Harrison, L. Hoang, P. Keagle, W. Lumm, B. Pothier, D. Qiu, R. Spadafora, R. Vicaire, Y. Wang, J. Wierzbowski, R. Gibson, N. Jiwani, A. Caruso, D. Bush, H. Safer, D. Patwell, S. Prabhakar, S. McDougall, G. Shimer, A. Goyal, S. Pietrokovski, G. M. Church, C.J. Daniels, J.-i. Mao, P. Rice, J. Nolling and J.N. Reeve (1997). "Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: functional analysis and comparative genomics." *Journal of Bacteriology* **179**: 7135-7155.

Snyder, J.C., K. Stedman, G. Rice, B. Wiedenheft, J. Spuhler and M.J. Young (2003). "Viruses of hyperthermophilic Archaea." *Research in Microbiology* **154**: 474-482.

Spear, J.R., J.J. Walker, T.M. McCollom and N.R. Pace (2005). "Hydrogen and bioenergetics in the Yellowstone geothermal ecosystem." *Proceedings of the National Academy of Science USA* **102**: 2555-2560.

Stedman, K., D. Prangishvili and W. Zillig (2005). Viruses of Archaea. The Bacteriophages, 2nd edition, volume I. R. Calendar. New York, Oxford Press: In press.

Stedman, K.M., C. Schleper, E. Rumpf and W. Zillig (1999). "Genetic requirements for the function of the archaeal virus SSV1 in *Sulfolobus solfataricus*: construction and testing of viral shuttle vectors." *Genetics* **152**: 1397-1405.

Stedman, K.M., Q. She, H. Phan, H.P. Arnold, I. Holz, R.A. Garrett and W. Zillig (2003). "Relationships between fuselloviruses infecting the extremely thermophilic archaeon *Sulfolobus*: SSV1 and SSV2." Research in Microbiology **154**: 295-302.

Stolt, P. and W. Zillig (1995). "Archaeabacterial Bacteriophages" Encyclopedia of Virology Plus. R.G. Webster and A. Granoff. London, Academic Press: 50-58.

Swofford, D. L. (2002). PAUP* Phylogenetic Analysis Using Parsimony (*and Other Methods). Sunderland, MA, Sinauer Associates.

Talling, J.F. and I.B. Talling (1965). "The chemical composition of African lake waters." *Internationale Revue der Gesamte Hydrobiologie* **50**: 421-463.

Tang, S.L., S.D. Nuttall, K. Ngui, C. Fisher, P. Lopez and M.L. Dyall-Smith (2002). "HF2: a double-stranded DNA tailed haloarchaeal virus with a mosaic genome." *Molecular Microbiology* **44**: 283-296.

Thompson, J.D., T.J. Gibson, F. Plewniak, F. Jeanmougin and D.G. Higgins (1997). "The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools." *Nucleic Acids Research* **24**: 4876-4882.

Thompson, J.R., S. Pacocha, C. Pharino, V. Klepac-Ceraj, D.E. Hunt, J. Benoit, R. Sarma-Rupavtarm, D.L. Distel and M.F. Polz (2005). "Genotypic diversity within a natural coastal bacterioplankton population." *Science* **307**: 1311-1313.

Torsvik, T. and I.D. Dundas (1974). "Bacteriophage of *Halobacterium salinarium*." *Nature* **248**: 680-681.

Torsvik, T. and I.D. Dundas (1980). "Persisting phage infection in *Halobacterium salinarium* str. 1." *Journal of General Virology* **47**: 29-36.

Wais, A.C., M. Kon, R.E. MacDonald, and B.D. Stollar (1975). "Salt-dependant bacteriophage infecting *Halobacterium cutirubrum* and *H. halobium*." *Nature* **256**: 314-315.

Waterbury, J.B. and F.W. Valois (1993). "Resistance to co-occurring phages enables marine *Synechococcus* communities to coexist with cyanophages abundant in seawater." *Applied and Environmental Microbiology* **59**: 3393-3399.

Wenke, H. and D. Oesterhelt (1988). "Isolation of a halobacterial phage with a fully cytosine-methylated genome." *Molecular and General Genetics* **211**: 407-414.

Whitaker, R.J., D.W. Grogan and J.W. Taylor (2003). "Geographic barriers isolate endemic populations of hyperthermophilic Archaea." *Science* **301**: 925-926.

Whitman, W. B. (1985). Methanogenic Bacteria. The Bacteria, Vol VIII, Academic Press, Inc.: 3-23.

Wiedenheft, B., K. Stedman, F. Roberto, D. Wilits, A.-K. Gleske, L. Zoeller, J. Snyder, T. Douglas and M. Young (2004). "Comparative genomic analysis of hyperthermophilic archaeal Fuselloviridae viruses." *Journal of Virology* **78**: 1954-1961.

Wilson, W.H., I.R. Joint, N.G. Carr and N.H. Mann (1993). "Isolation and molecular characterization of five marine cyanophages propagated on *Synechococcus* sp. strain WH7803." *Applied and Environmental Microbiology* **59**: 3736-3743.

Witte, A., U. Baranyi, R. Klein, M. Sulzner, C. Luo, G. Wanner, D.H. Kruger and W. Lubitz (1997). "Characterization of *Natronobacterium magadii* phage ϕCh1, a unique archaeal phage containing DNA and RNA." *Molecular Microbiology* **23**: 603-616.

Woese, C.R., O. Kandler and M.L. Wheeler (1990). "Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya." *Proceedings of the National Academy of Sciences USA* **87**: 4576-4579.

Wommack, K.E. and R.R. Colwell (2000). "Virioplankton: viruses in aquatic ecosystems." *Microbiology and Molecular Biology Reviews* **64**: 69-114.

Wood, A.G., W.B. Whitman and J. Konisky (1989). "Isolation and characterization of an archaebacterial virus-like particle from *Methanococcus voltae* A3." *Journal of Bacteriology* **171**: 93-98.

Yeats, S., P. McWilliam and W. Zillig (1982). "A plasmid in the archaebacterium *Sulfolobus acidocaldarius*." *The EMBO Journal* **1**: 1035-1038.

Zillig, W., H.P. Arnold, I. Holz, D. Prangishvili, A. Schweier, K. Stedman, Q. She, H. Phan, R. Garrett and J.K. Kristjansson (1998). "Genetic elements in the extremely thermophilic archaeon *Sulfolobus*." *Extremophiles* **2**: 131-140.

Zillig, W., A. Kletzin, C. Schleper, I. Holz, D. Janekovic, J. Hain, M. Lanzendorfer and J. K. Kristjansson (1994). "Screening for Sulfolobales, their plasmids, and their viruses in Icelandic solfataras." *Systematic and Applied Microbiology* **16**: 609-628.

Zillig, W., D. Prangishvili, C. Schleper, M. Elferink, I. Holz, S. Albers, D. Janekovic and D. Gotz (1996). "Viruses, plasmids, and other genetic elements of thermophilic and hyperthermophilic Archaea." *FEMS Microbiology Reviews* **18**: 225-236.

Zillig, W., W.-D. Reiter, P. Palm, F. Gropp, H. Neumann and M. Rettenberger (1988). Viruses of Archaebacteria. The Bacteriophages. R. Calendar. New York, Plenum Press: 517-558.

Zillig, W., K.O. Stetter, S. Wunderl, W. Schulz, H. Preiss and H. Scholz (1980). "The *Sulfolobus*-"Caldariella" Group: Taxonomy on the basis of the structure of DNA-dependent RNA polymerases." *Archives in Microbiology* **125**: 259-269.

Zillig, W., J. Tu and I. Holz (1981). "Thermoproteales--a third order of thermoacidophilic archaebacteria." *Nature* **293**(5827): 85-86.

Zillig, W., S. Yeats, I. Holz, A. Boeck, M. Rettenberger, F. Gropp and G. Simon (1986). "*Desulfurolobus ambivalens* new genus new species and autotrophic archaebacterium facultatively oxidizing or reducing sulfur." *Applied Microbiology* **8**: 197-203.

APPENDICES

APPENDIX A

SUPPLEMENTARY PHYLOGENETIC ANALYSES FROM THE CRATER HILLS MONITOR SITE

Figure 3.28. The SSV-like population detected from Crater Hills over time. The red sequences were collected in October 2002. A. DNA sequences; the • represents a non-functional ORF. B. Inferred protein sequences; the * represents identical protein sequences.

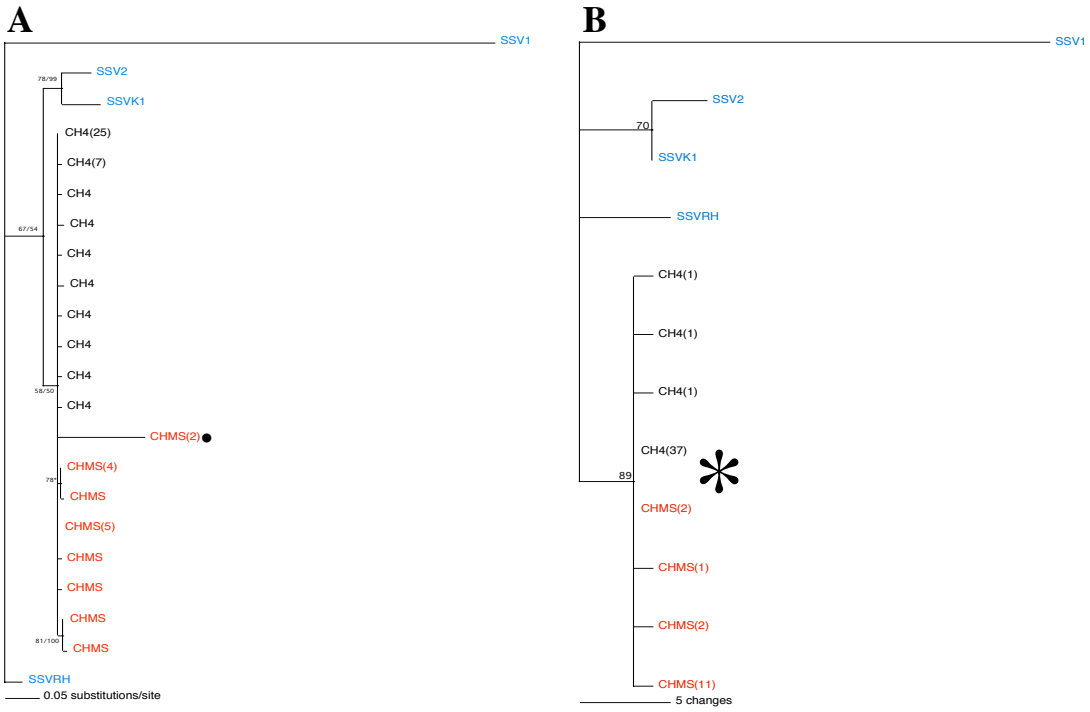


Figure 3.29. The SSV-like population detected in Crater Hills over time. The red sequences are those collected in February 2003. A. DNA sequences; the • represents non-functional ORFs. B. Inferred protein sequences.

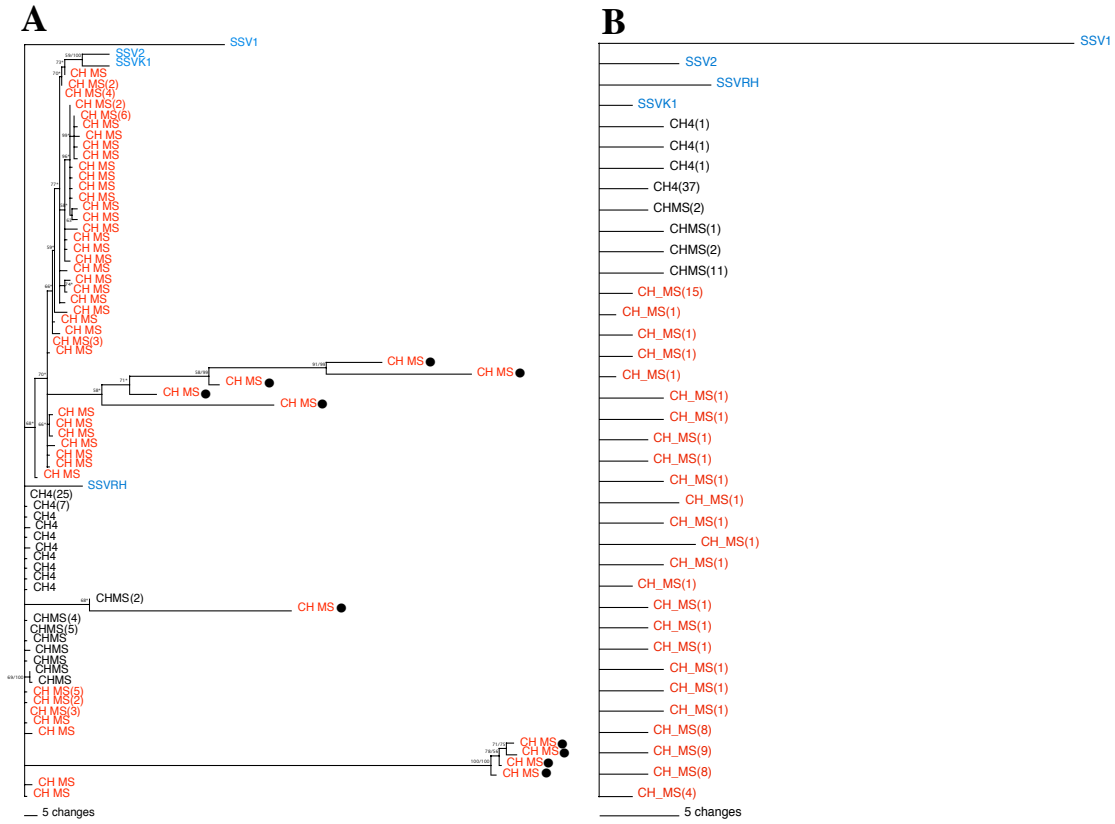


Figure 3.30. The SSV-like sequences detected in Crater Hills over time. The sequences from May 2003 are colored red. A. DNA sequences. B. Inferred protein sequences. The number in parentheses indicates how many times a sequence was detected.

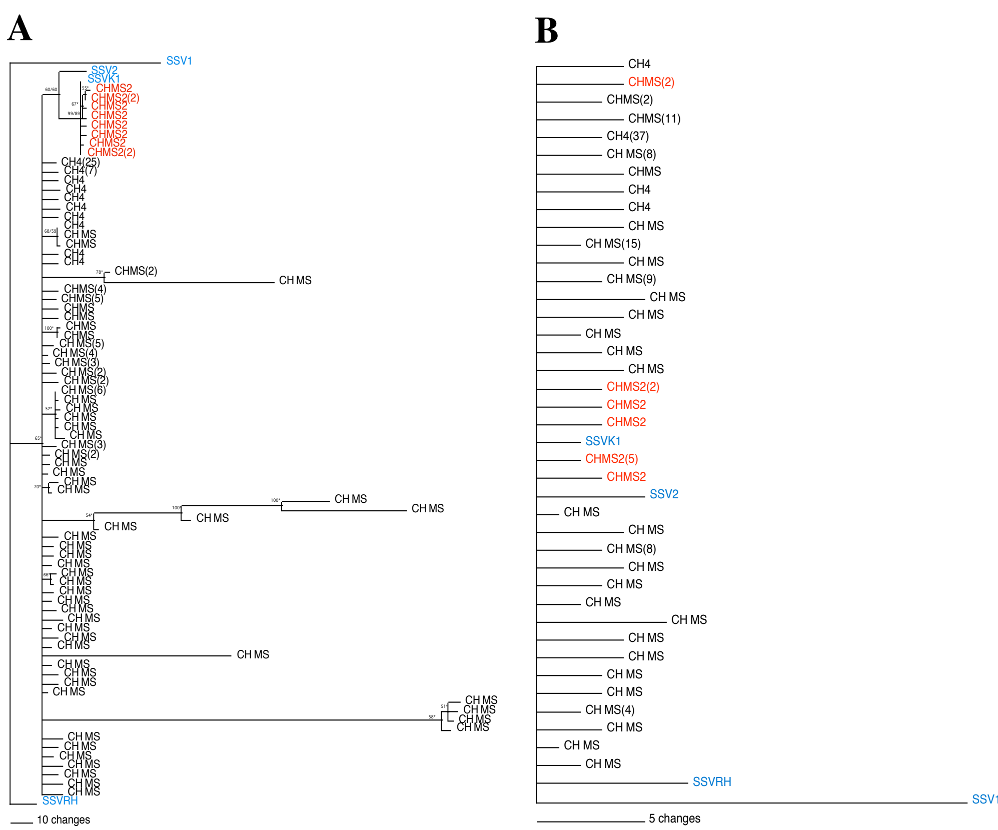


Figure 3.31. The SSV-like sequences detected in Crater Hills over time. The red sequences represent those collected in July 2003. A. DNA sequences; the • represents non-functional ORFs. B. Inferred protein sequences.

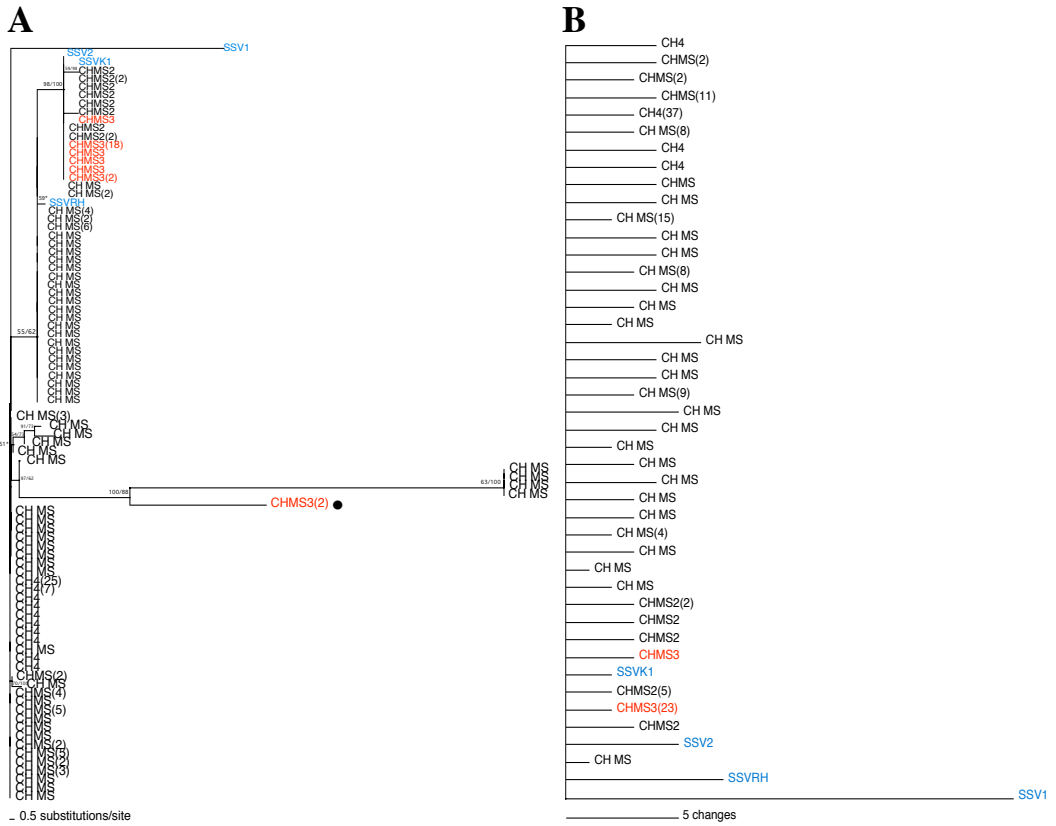


Figure 3.32. The SSV-like DNA sequences detected in Crater Hills over time. The red sequences were detected in August 2003. The number in parentheses represents the number of times a particular sequence was detected. The • represents non-functional ORFs.

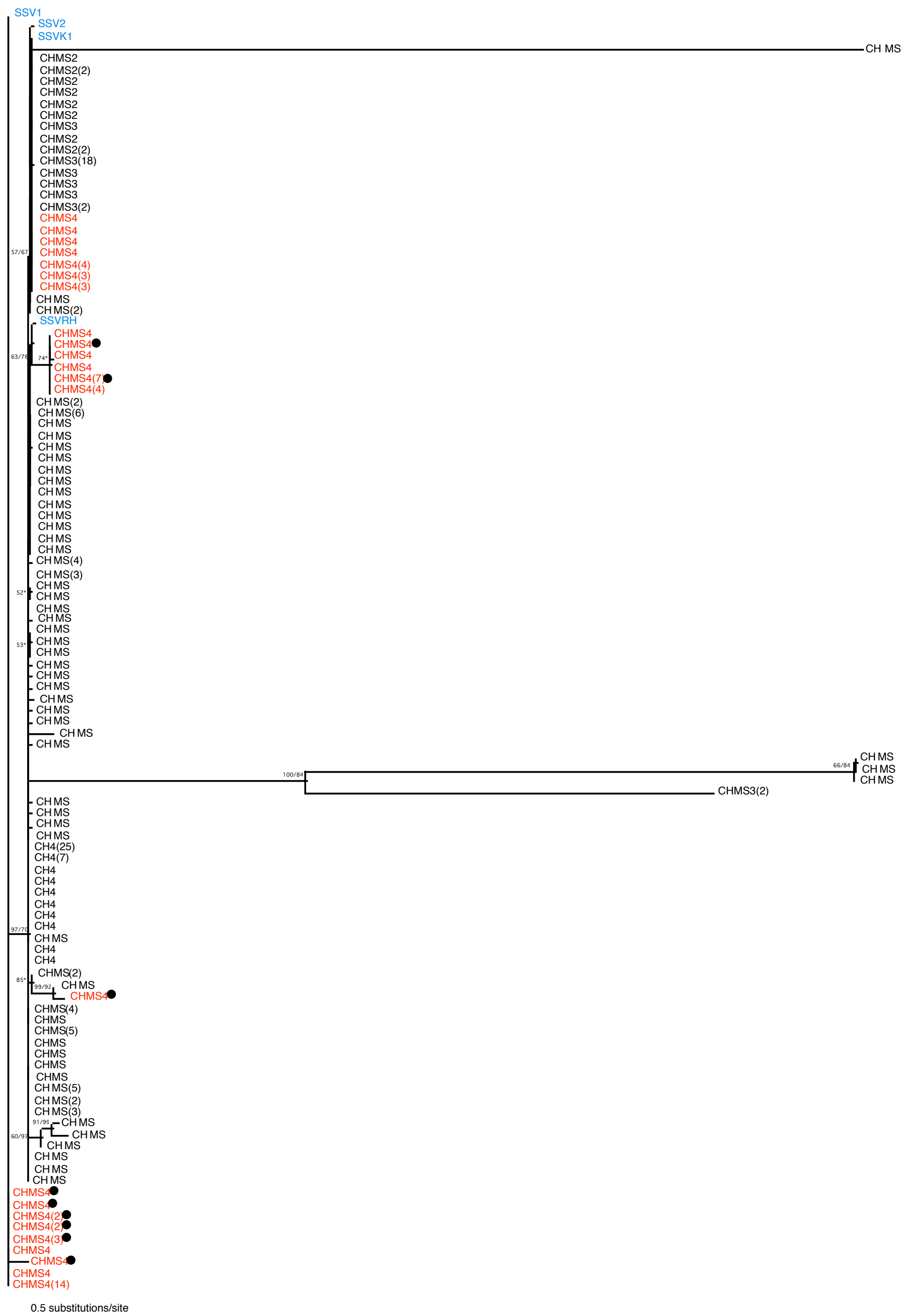


Figure 3.33. The SSV-like protein sequences detected in Crater Hills over time. The red sequences were collected in August 2003. The number in parentheses represents the number of times a particular sequence was detected.

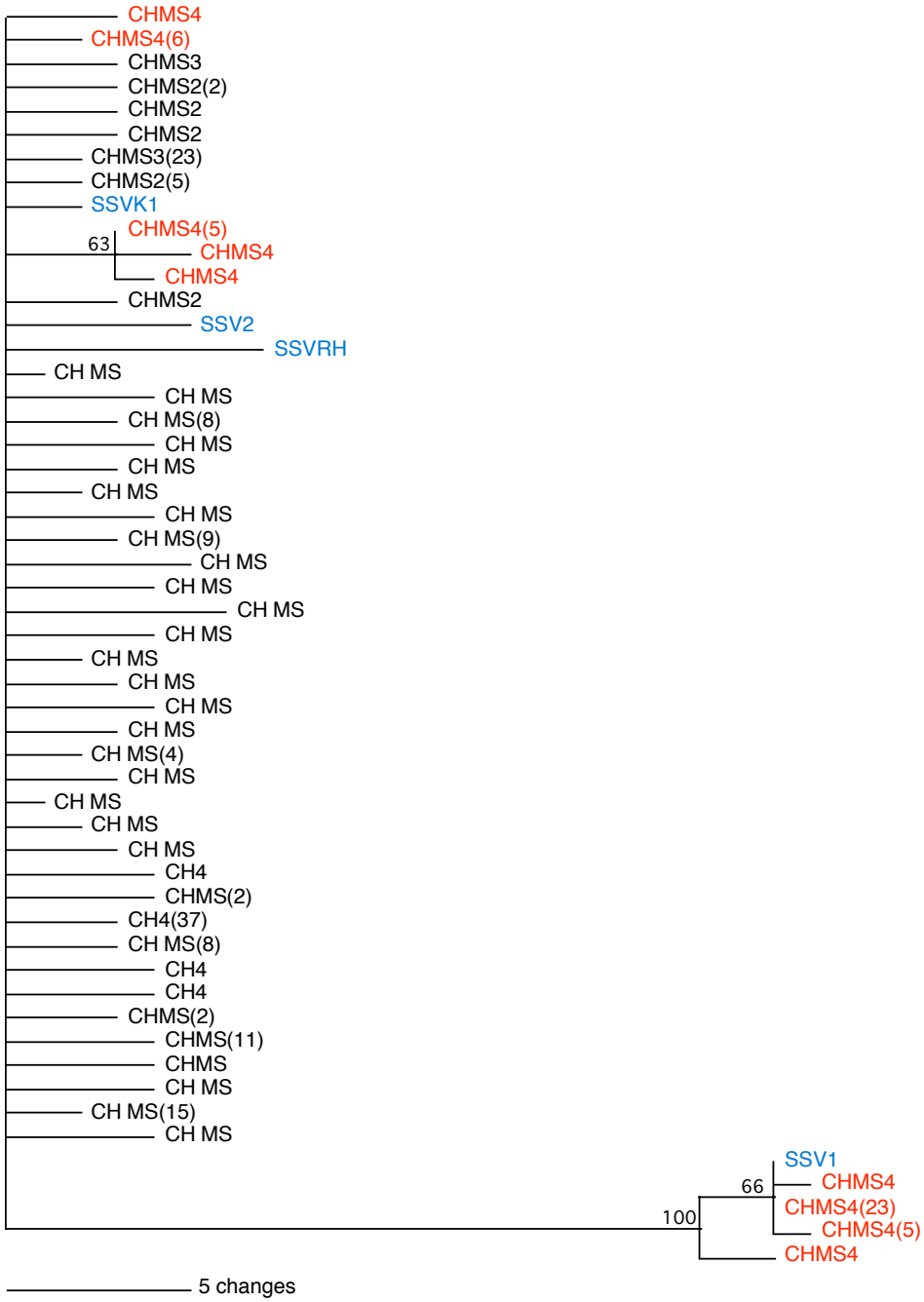


Figure 3.34. The SSV-like DNA sequences detected in Crater Hills over time. The sequences in red were collected in September 2003. The numbers in parentheses represents the number of times a sequence was detected.

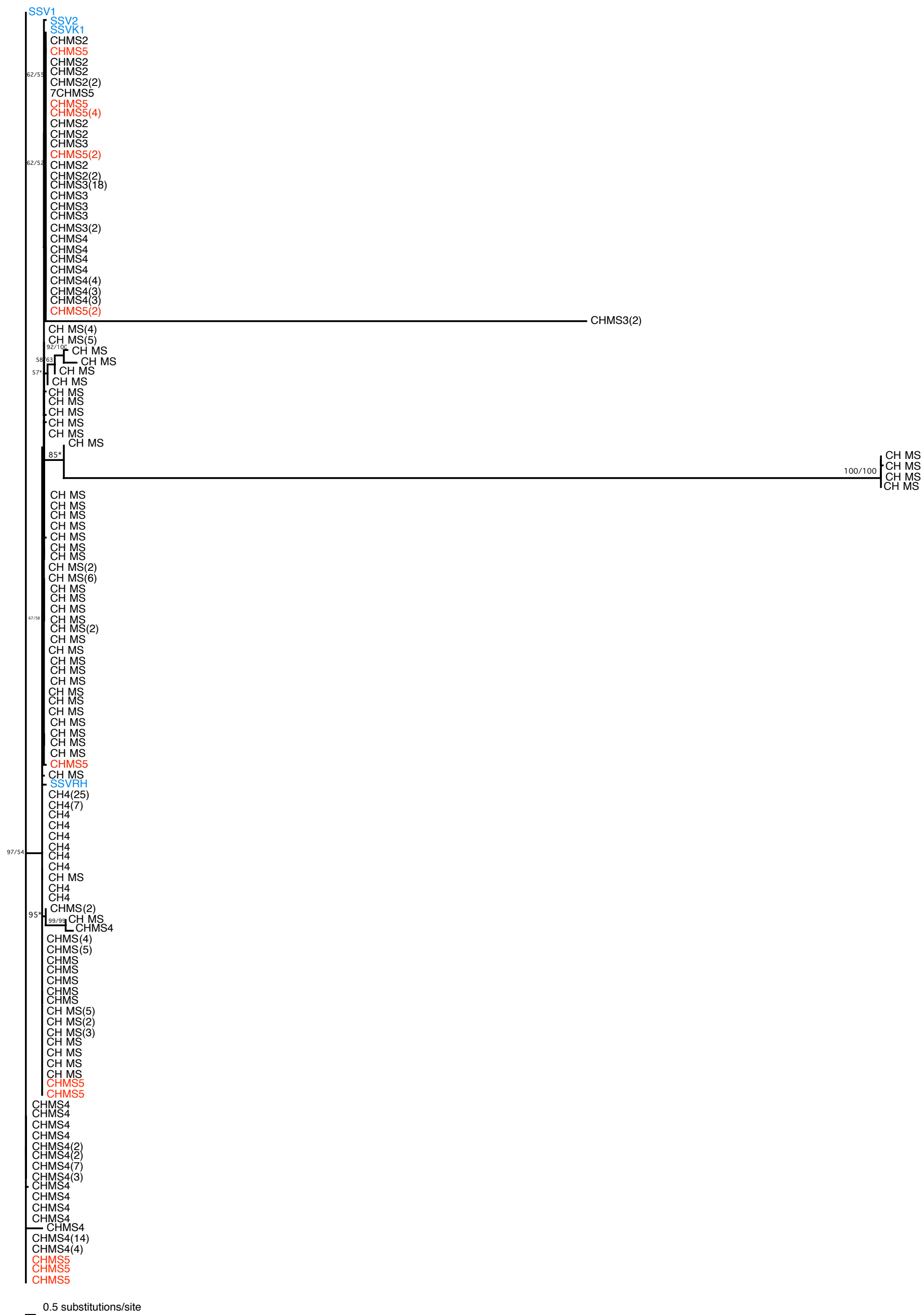


Figure 3.35. The SSV-like protein sequences detected in Crater Hills over time. The sequences from September 2003 are colored red. The * represents identical sequences. The numbers in parentheses represents the number of times a sequence was detected.

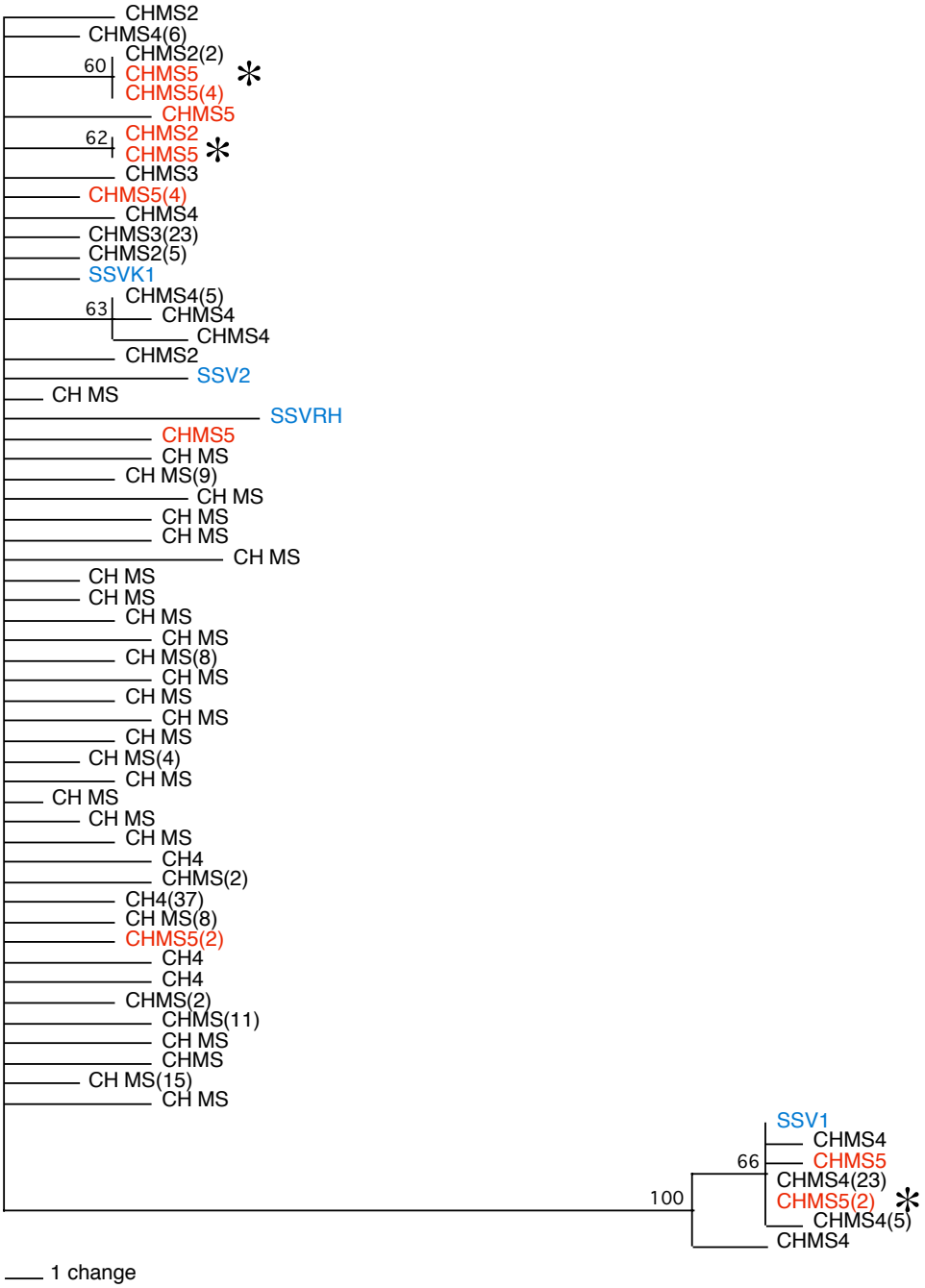


Figure 3.36. The SSV-like DNA sequences detected in Crater Hills over time. The sequences obtained from October 2003 are represented by red. The • represents the non-functional ORFs. The numbers in parentheses represents the number of times a sequence was detected.

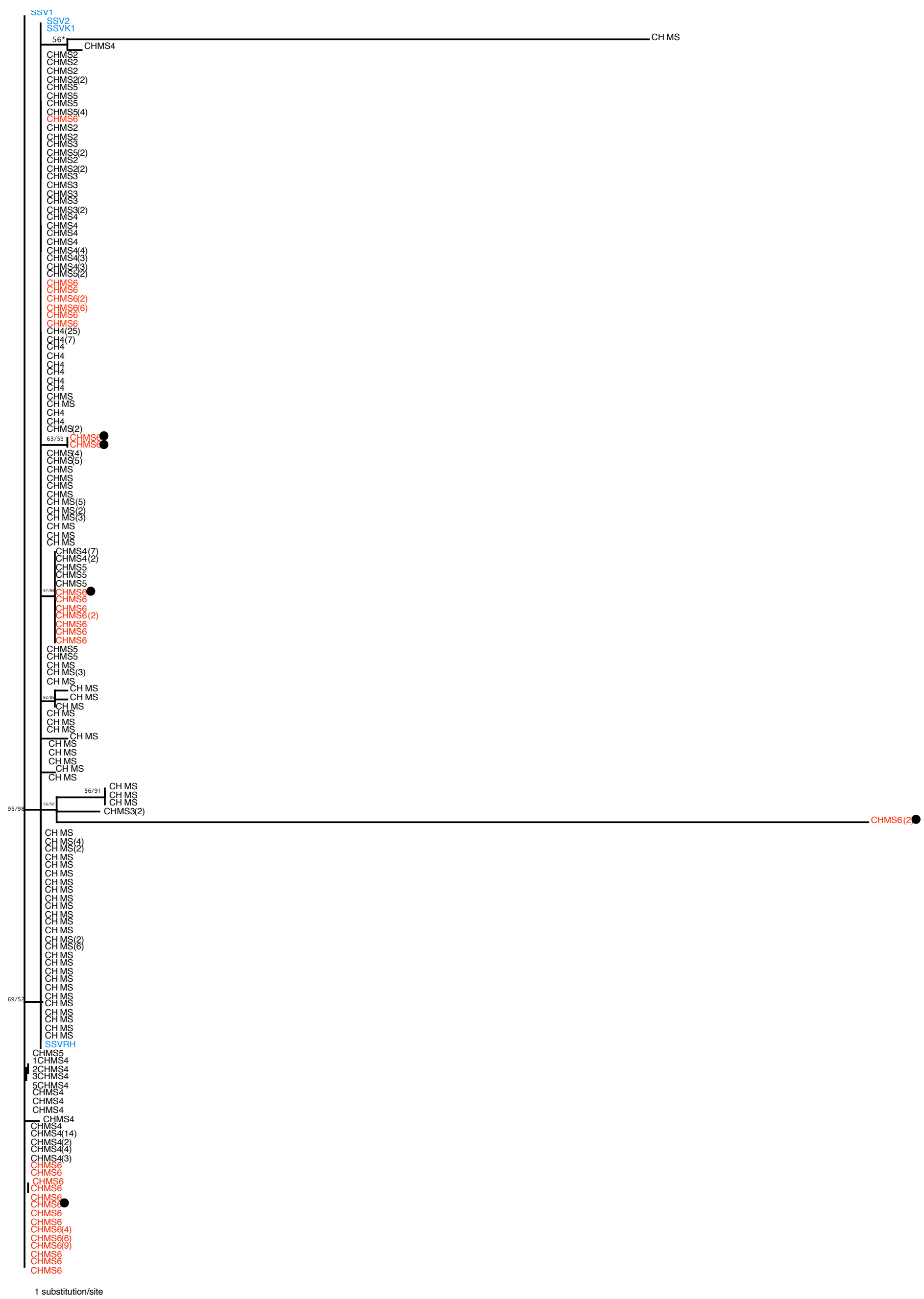


Figure 3.37. The SSV-like protein sequences detected in Crater Hills over time. The sequences in red were detected in October 2003. The * represents a group of identical protein sequences. The numbers in parentheses represents the times a sequence was detected.

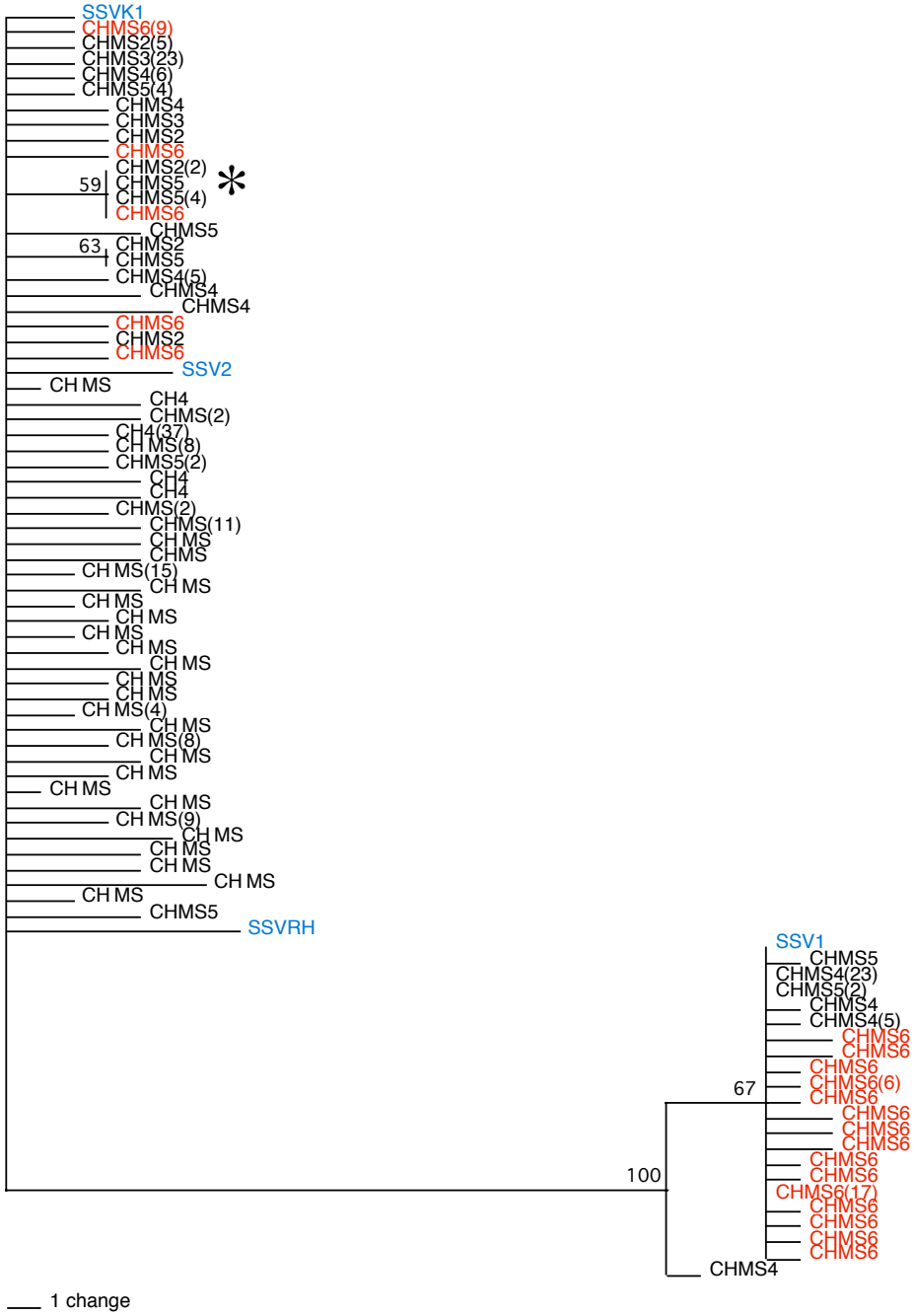


Figure 3.38. The SSV-like DNA sequences isolated from Crater Hills monitor site from August 2002 through October 2003. This analysis was conducted with non-functional and functional ORFs. The numbers in the parentheses represents how many times a particular sequence was detected.

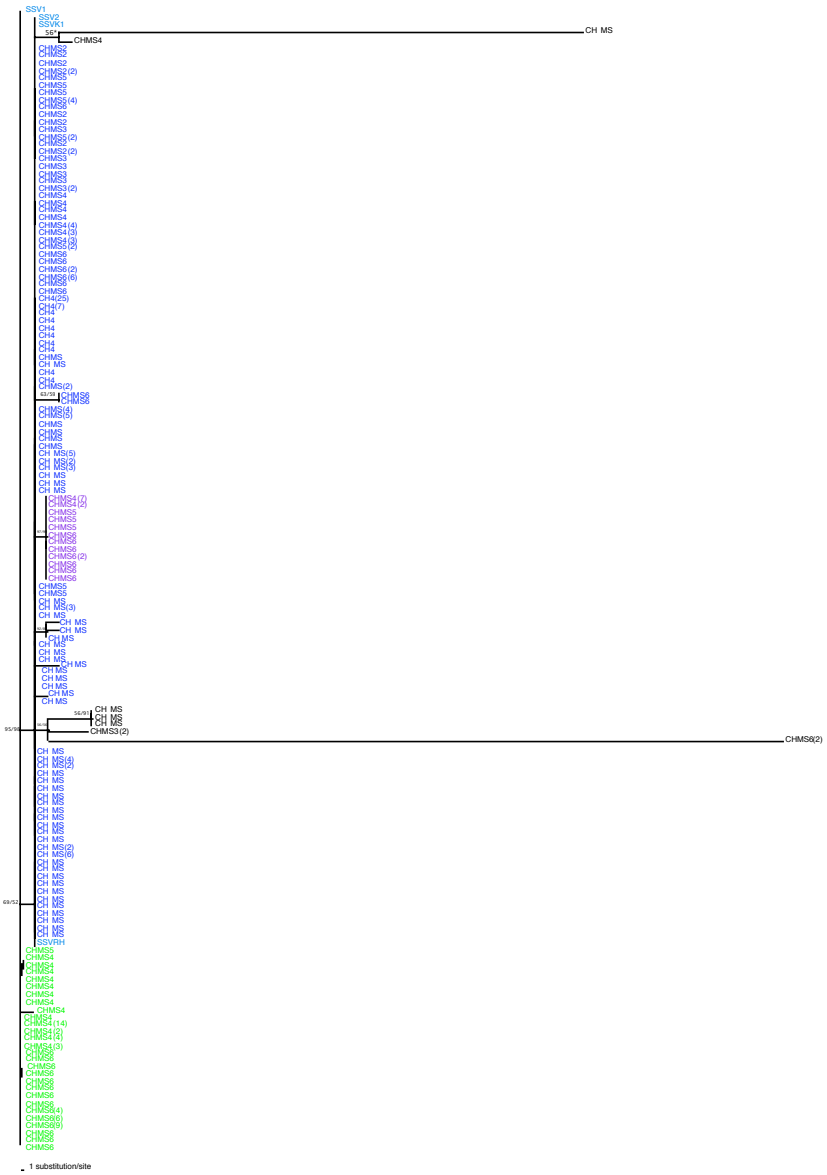


Figure 3.39. The SIRV-like sequences collected from Crater Hills. The red sequences are those sampled in October 2002. A. DNA sequences; B. Inferred protein sequences; the * represents identical protein sequences.

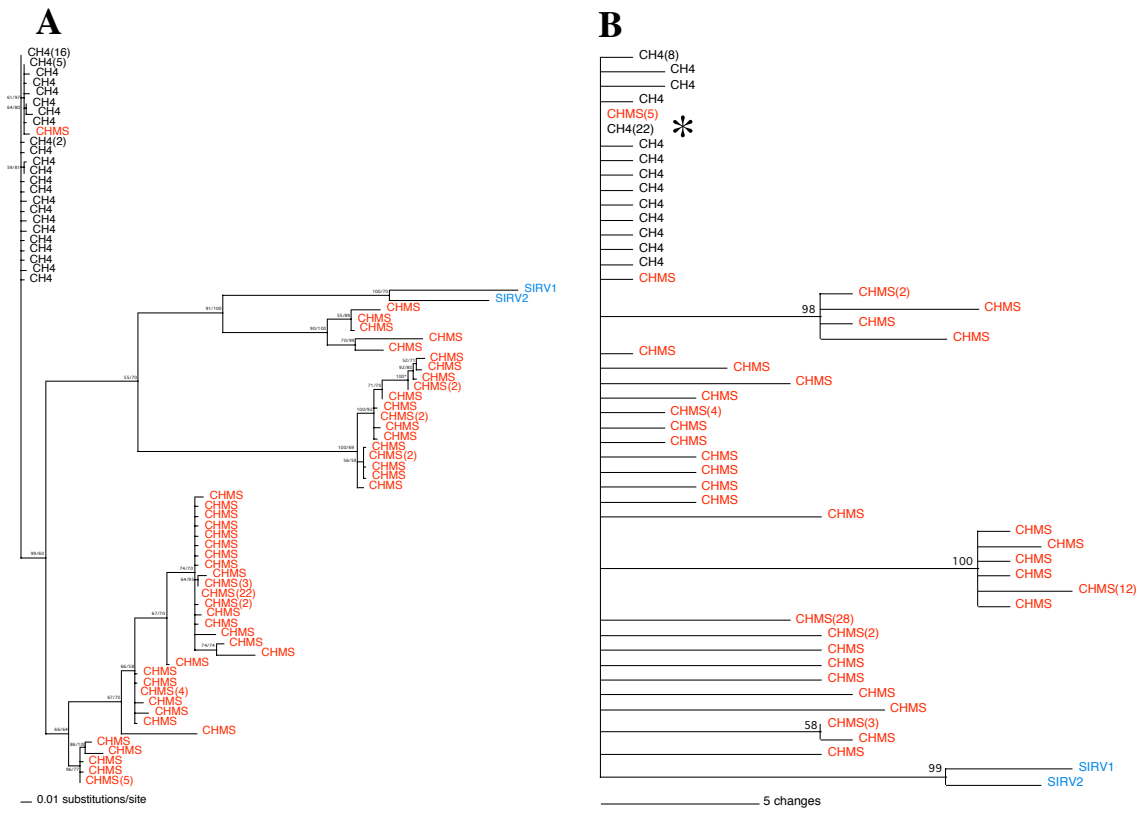


Figure 3.40. The SIRV-like DNA sequences collected from Crater Hills. The red sequences were collected in January 2003. The number in parentheses represents how many times a sequence was detected.

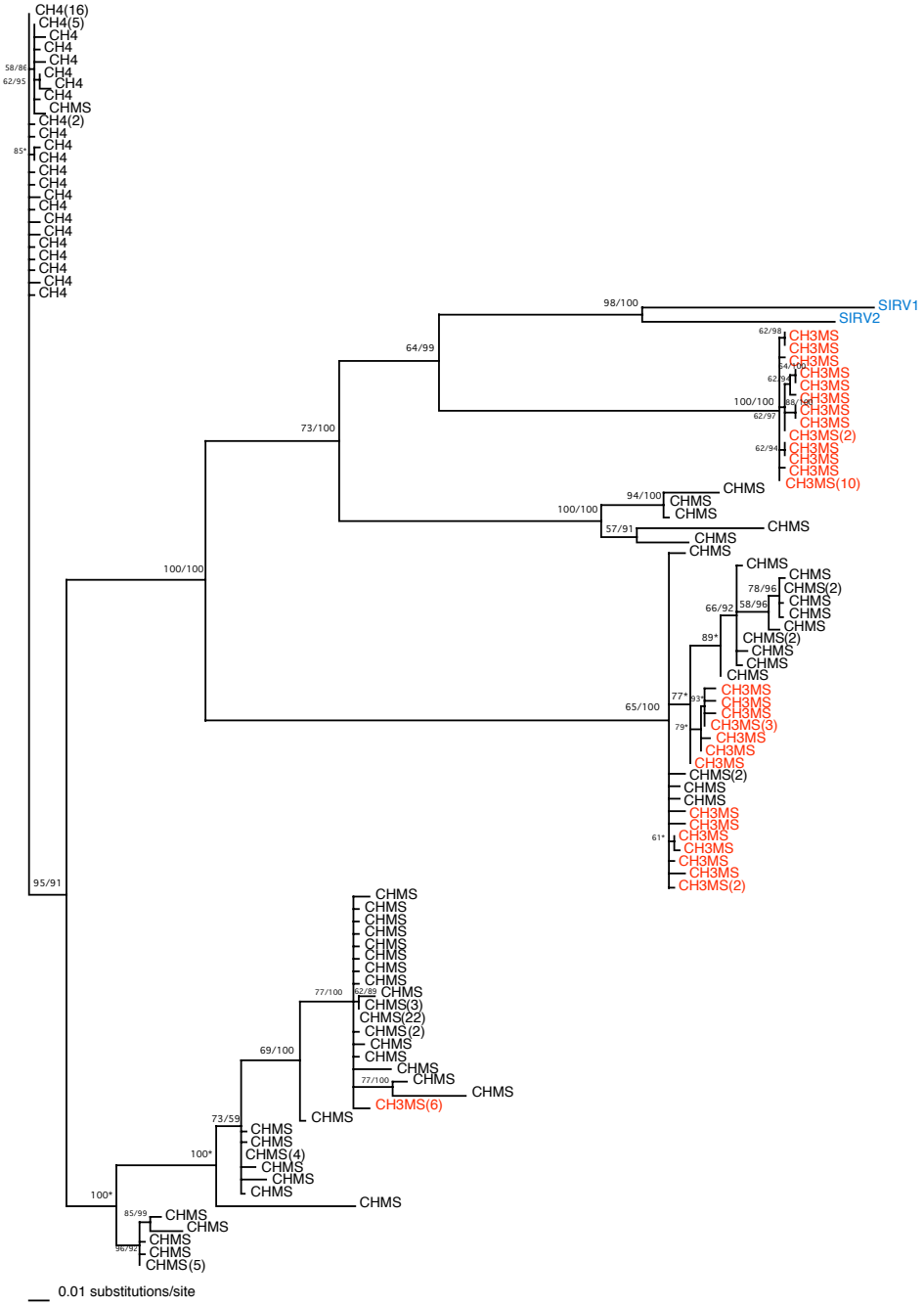


Figure 3.41. The SIRV-like protein sequences from Crater Hills. The red sequences are those collected in January 2003. The * represents identical protein sequences. The number in parentheses indicates how many times a sequence was detected.

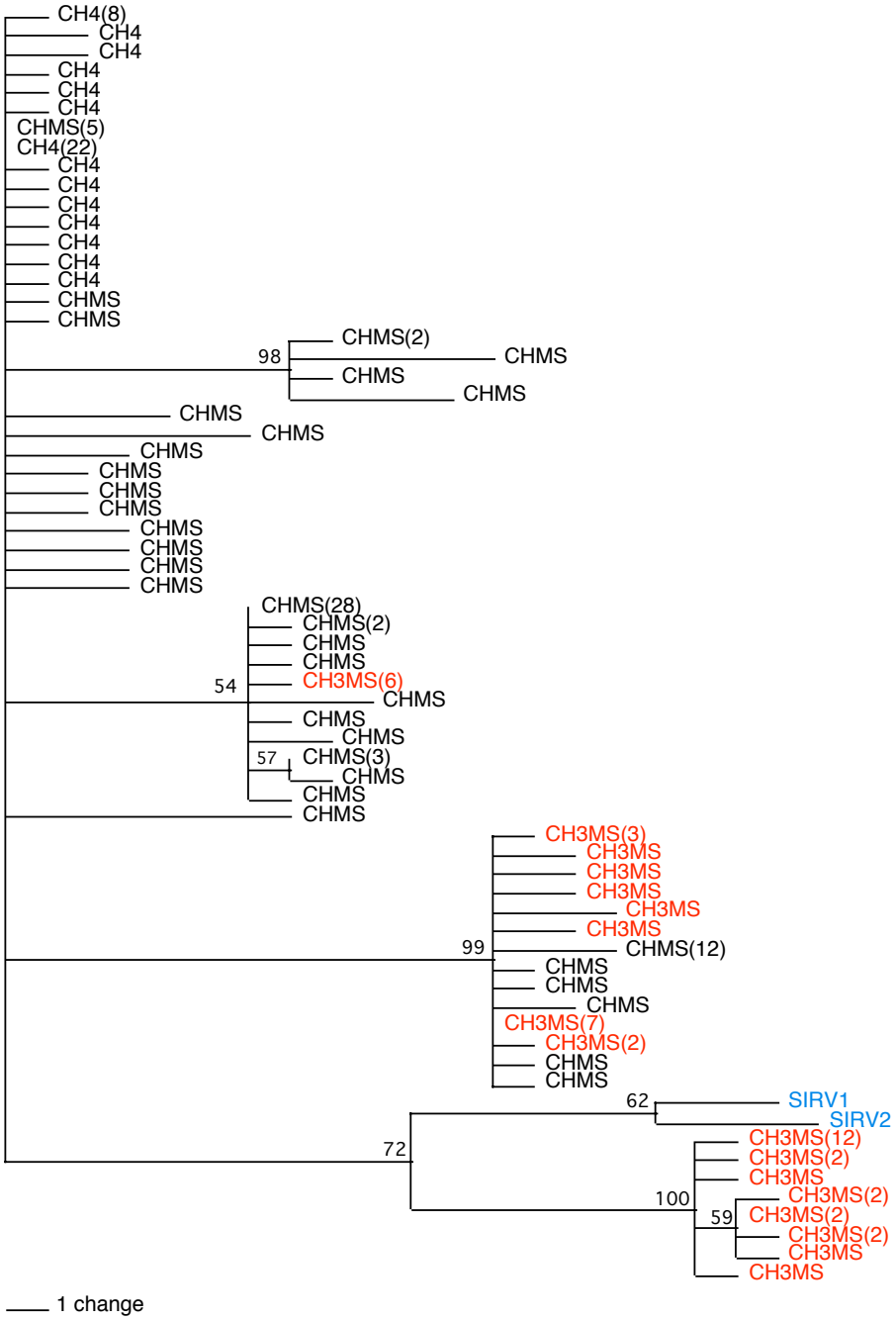


Figure 3.42. The SIRV-like DNA sequences collected from Crater Hills. The red samples were collected in February 2003. The number in parentheses indicates how many times a sequence was detected.

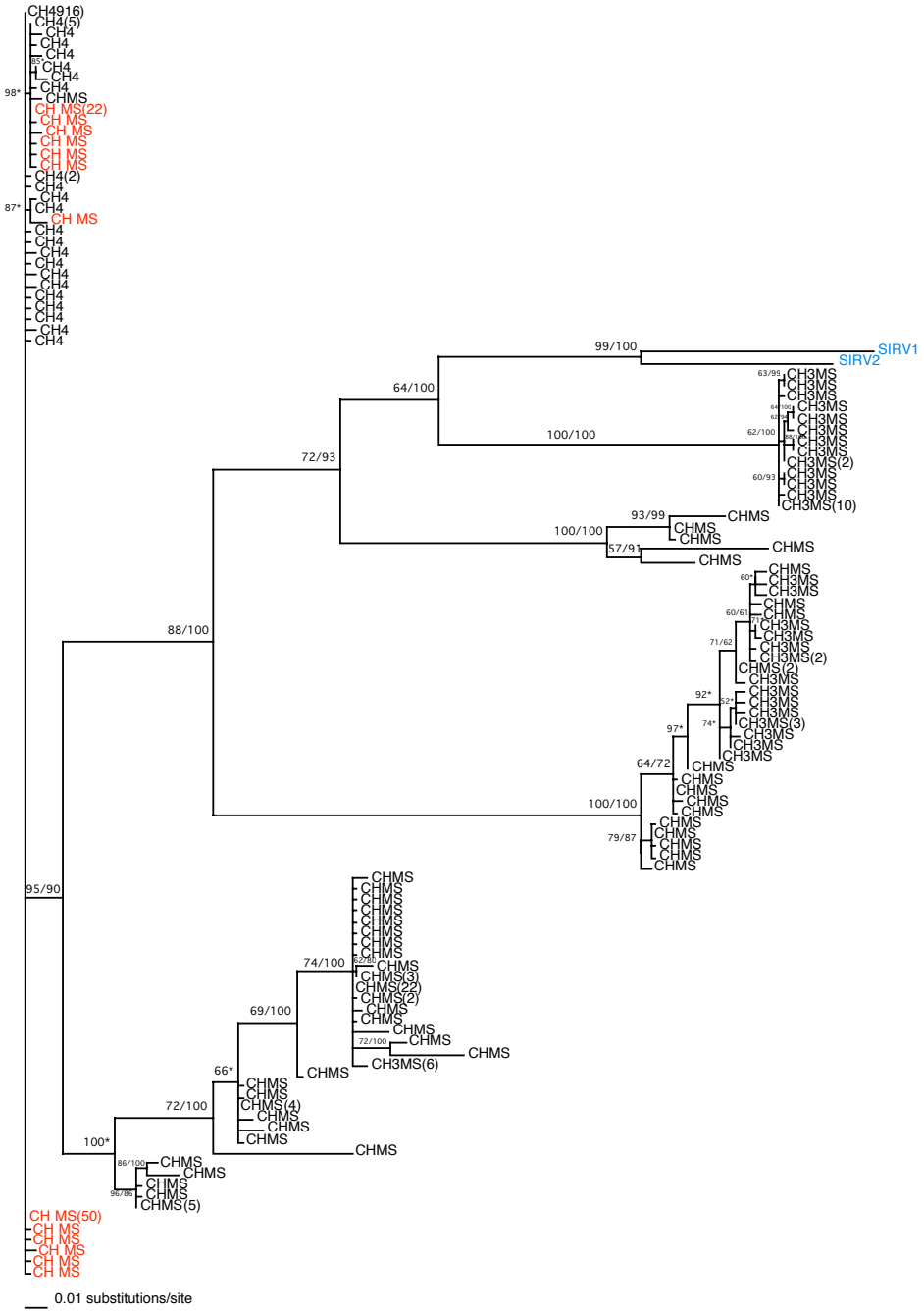


Figure 3.43. The SIRV-like protein sequences from Crater Hills. The sequences detected in February 2003 are colored red. The * represents identical protein sequences. The numbers in parentheses indicates how many times a sequence was detected.

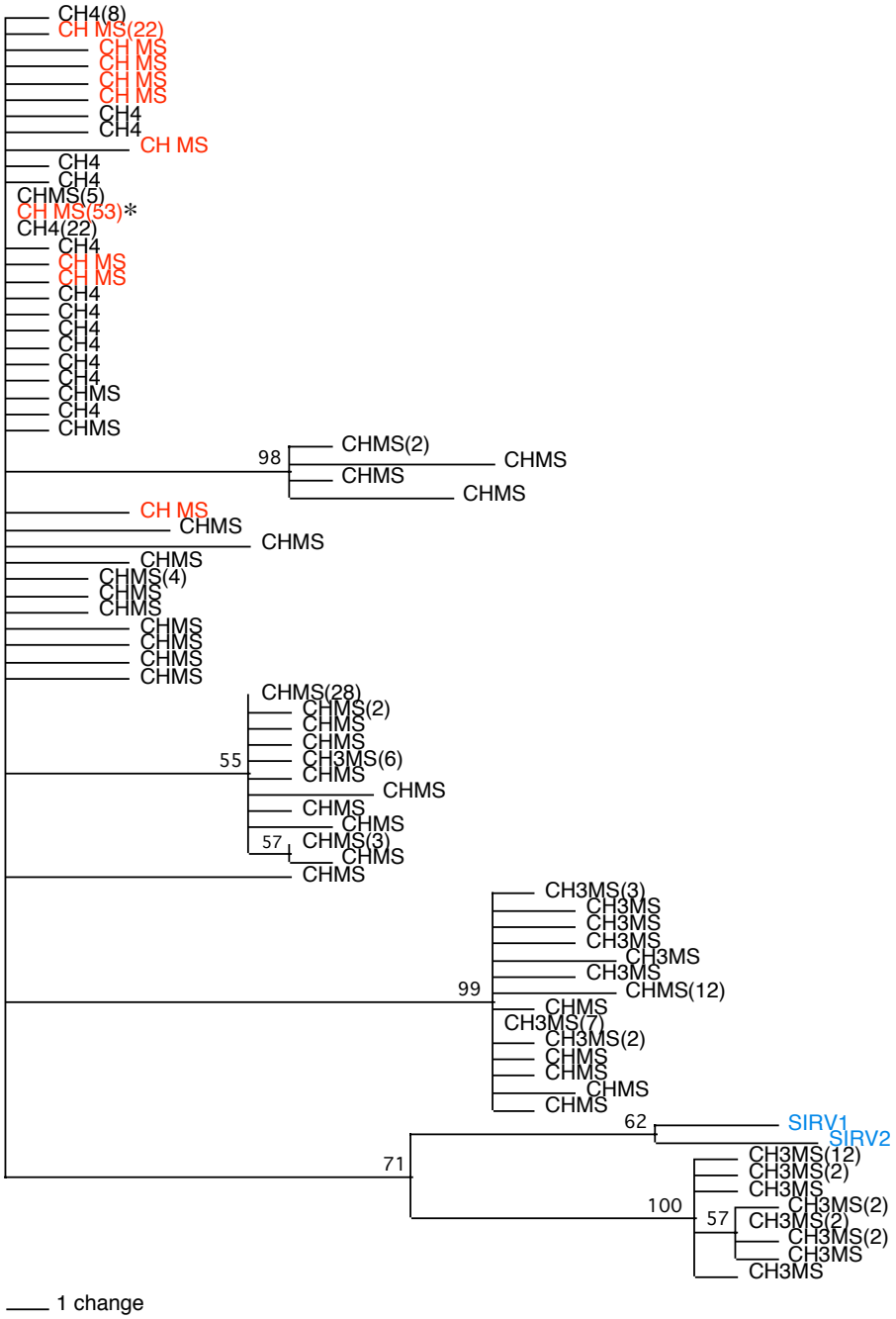


Figure 3.44. The SIRV-like DNA sequences collected from Crater Hills. The red sequences were detected in May 2003. The numbers in parentheses indicates how many times a sequence was detected.

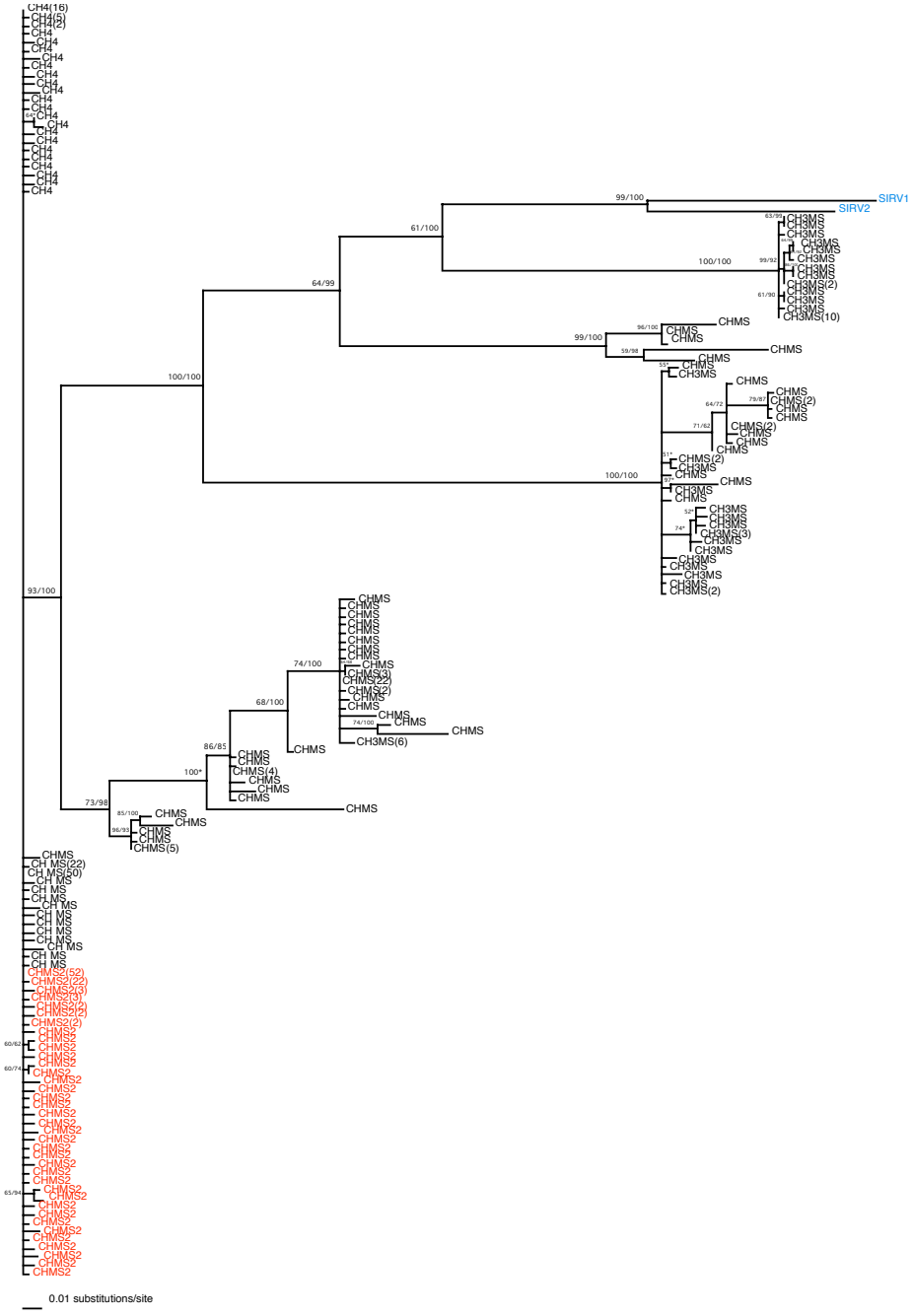


Figure 3.45. The SIRV-like protein sequences collected in Crater Hills. The sequences from May 2003 are colored red. The * represents identical protein sequences. The number in parentheses indicates how many times a sequence was detected.

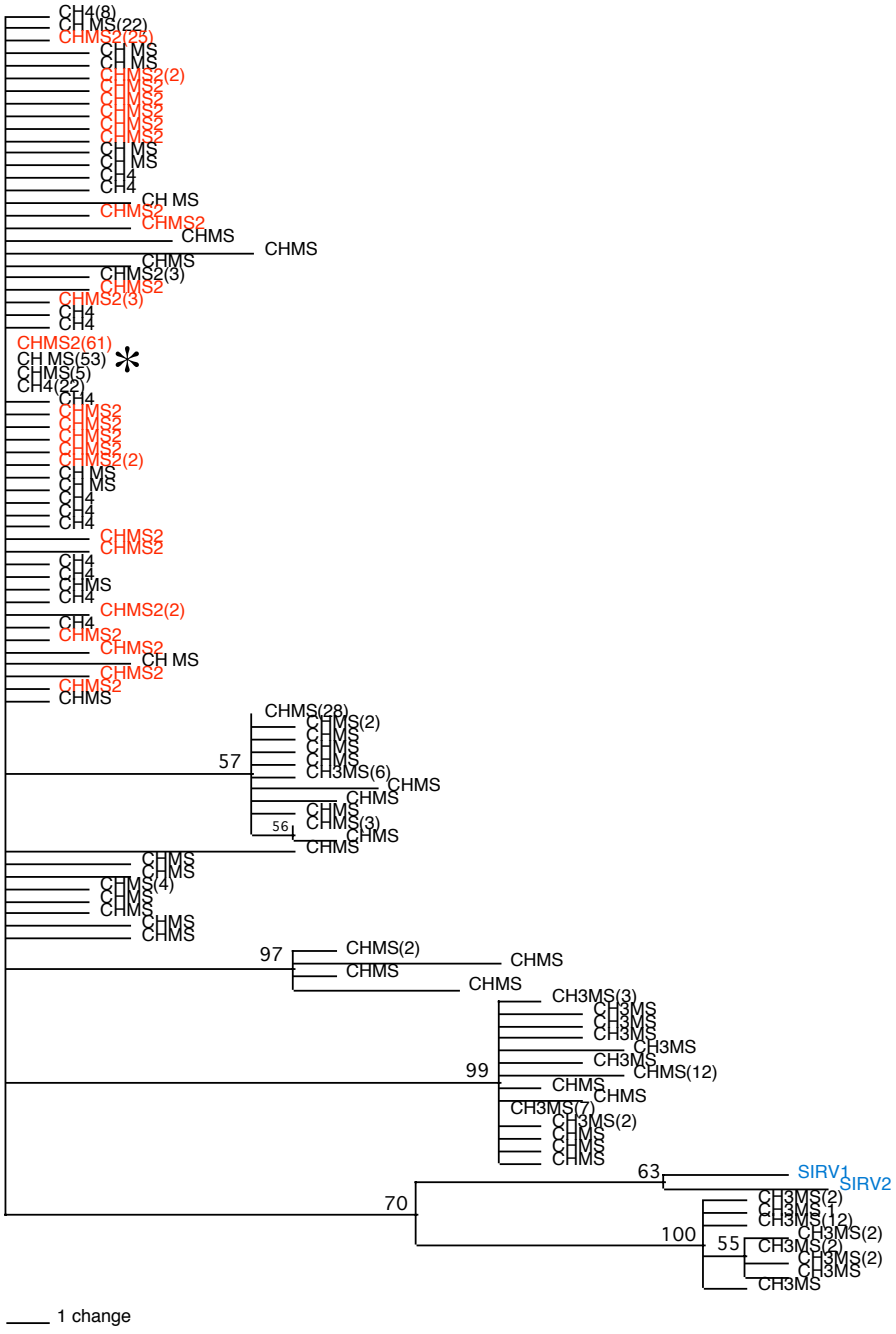


Figure 3.46. The SIRV-like DNA sequences detected in Crater Hills. The sequences detected in July 2003 are colored in red. The numbers in parentheses indicates how many times a sequence was detected.

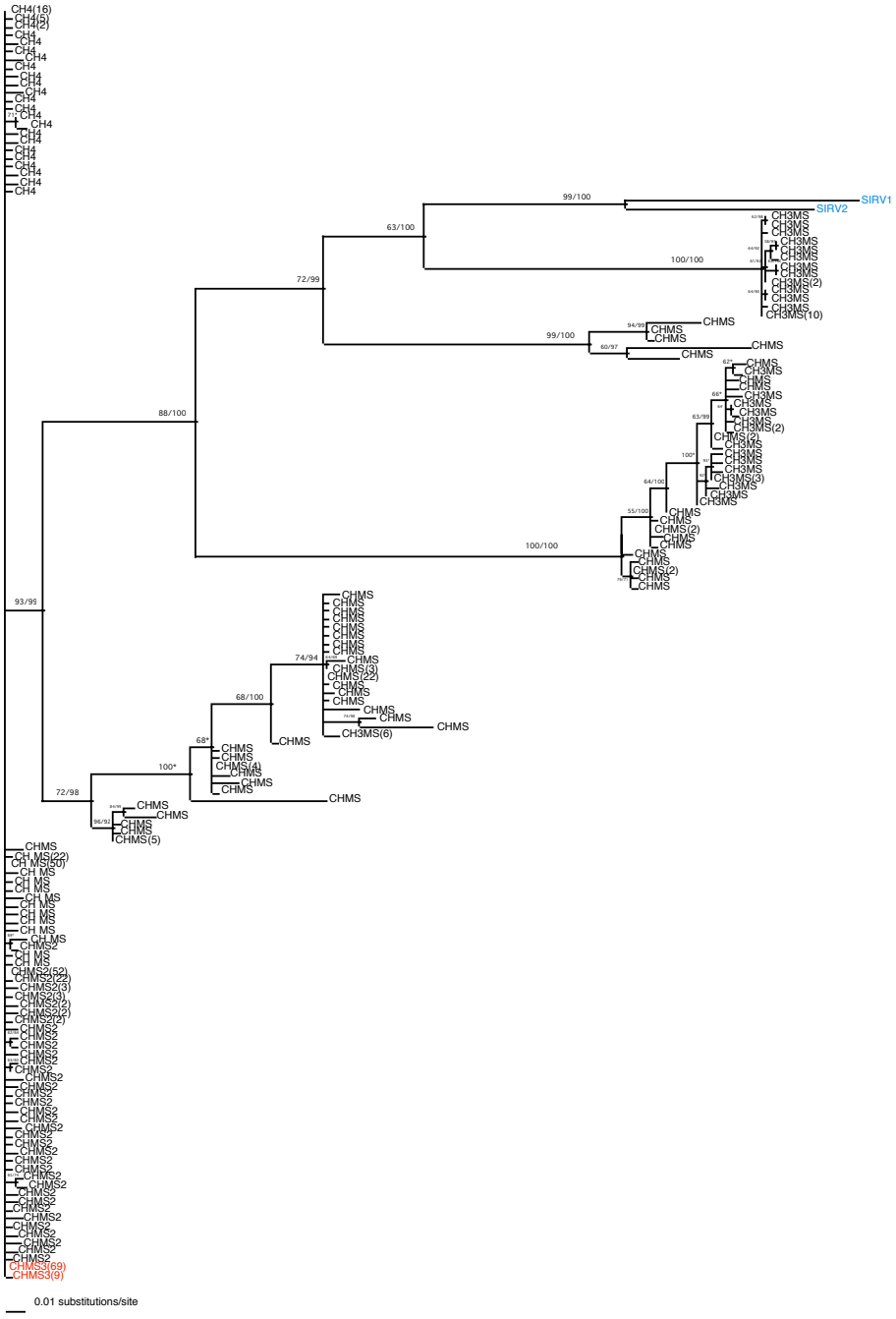


Figure 3.47. The SIRV-like protein sequences from Crater Hills. The sequences collected in July 2003 are colored red. The * represents identical protein sequences. The numbers in parentheses indicate how many times a sequence was detected.

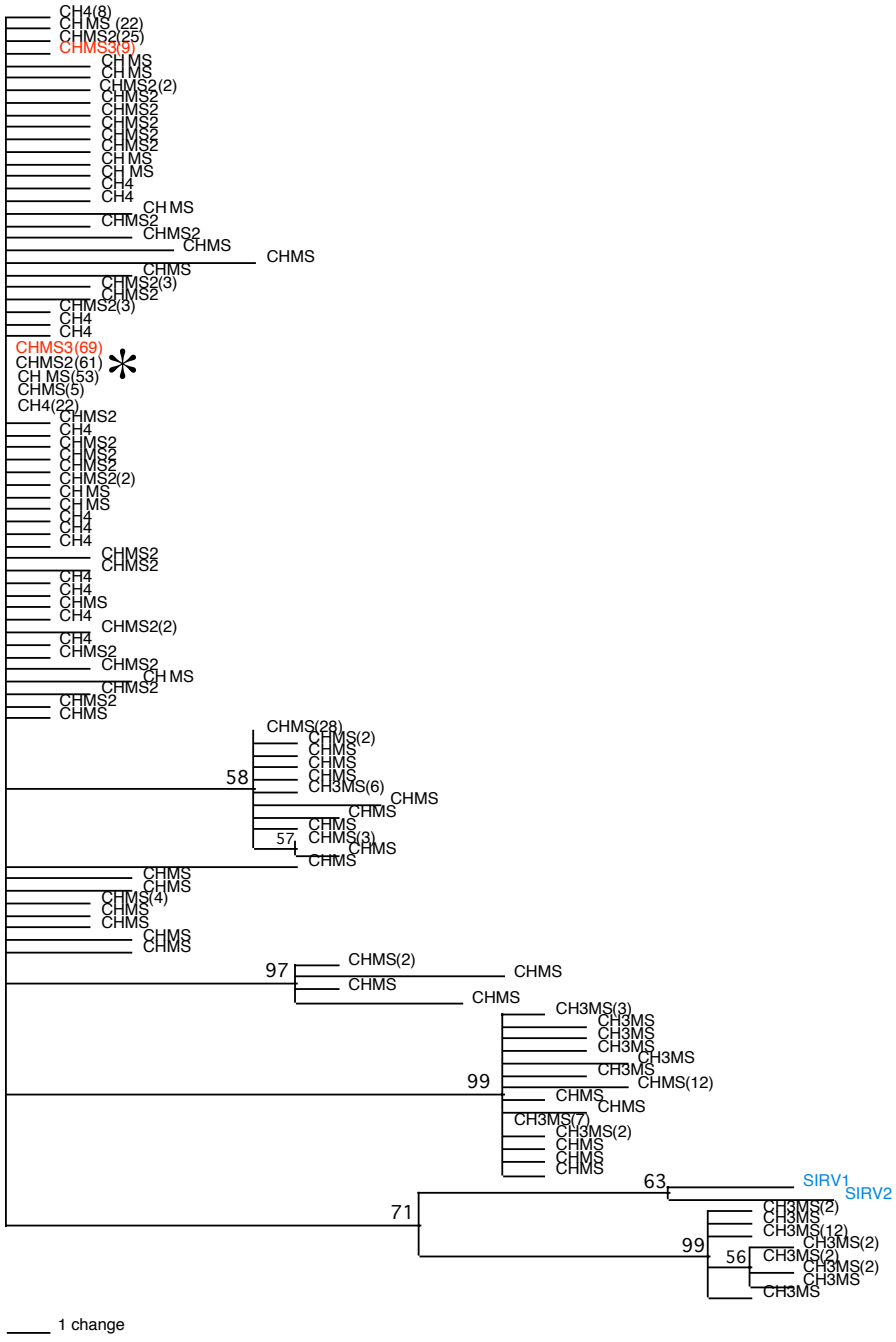


Figure 3.48. The SIRV-like DNA sequences from Crater Hills. The samples in red were collected in August 2003. The numbers in parentheses indicate how many times a sequence was detected.

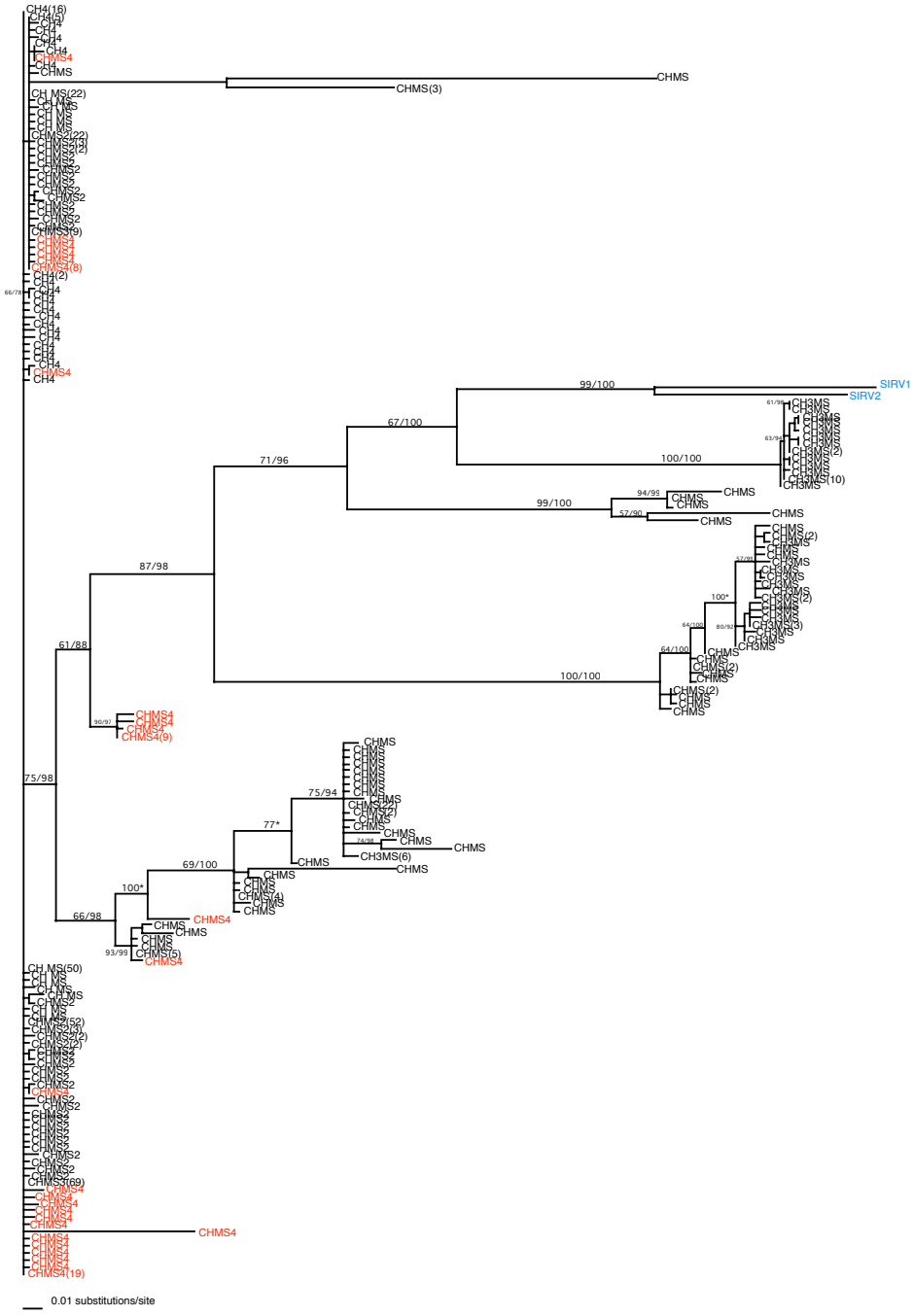


Figure 3.49. The SIRV-like protein sequences collected from Crater Hills. The samples in red were collected in August 2003. The * represents identical protein sequence. The numbers in parentheses indicate how many times a sequence was detected.

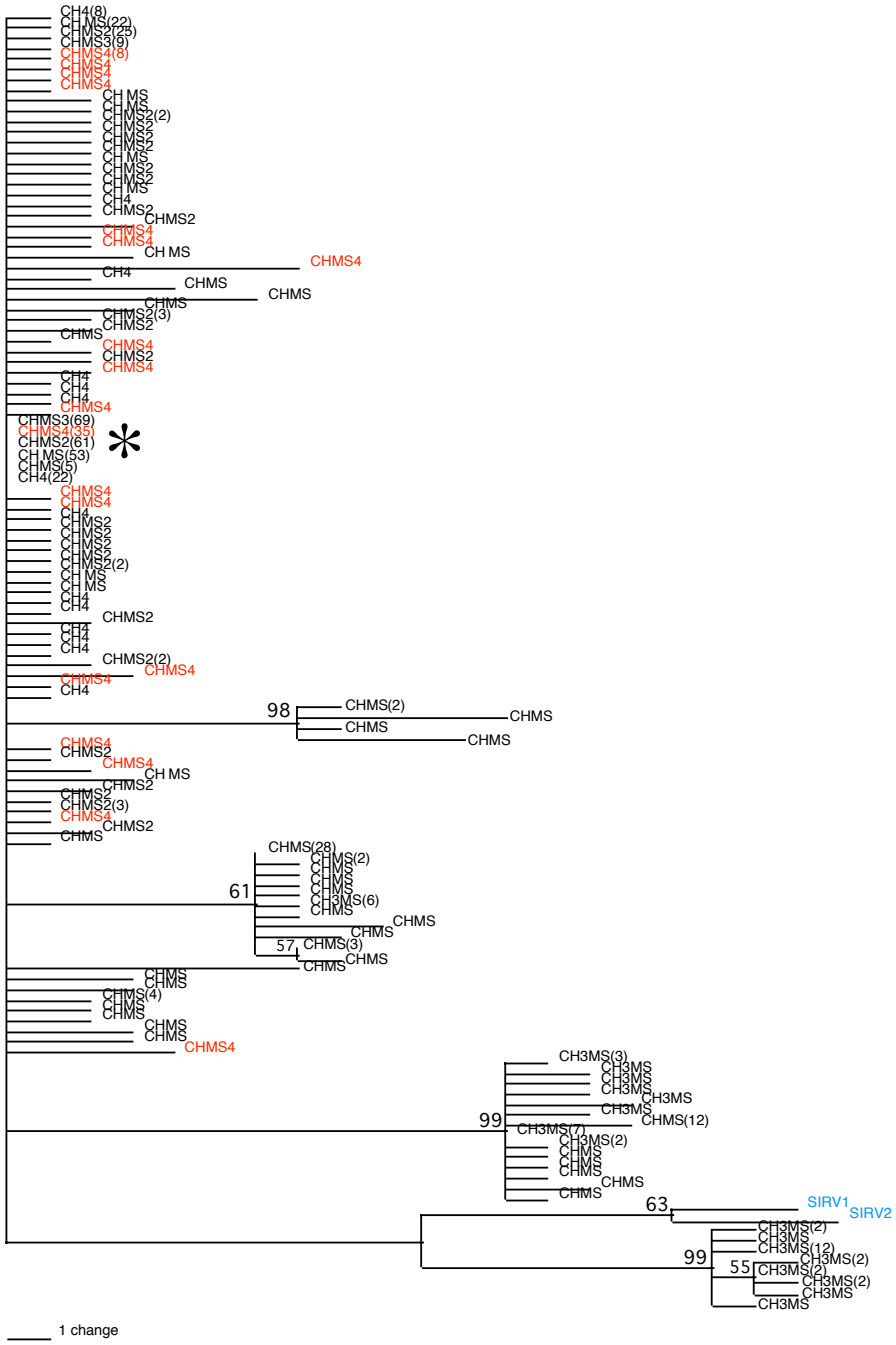


Figure 3.50. The SIRV-like DNA sequences from Crater Hills. The samples collected in September 2003 are colored red. The numbers in parentheses indicate how many times a sequence was detected.

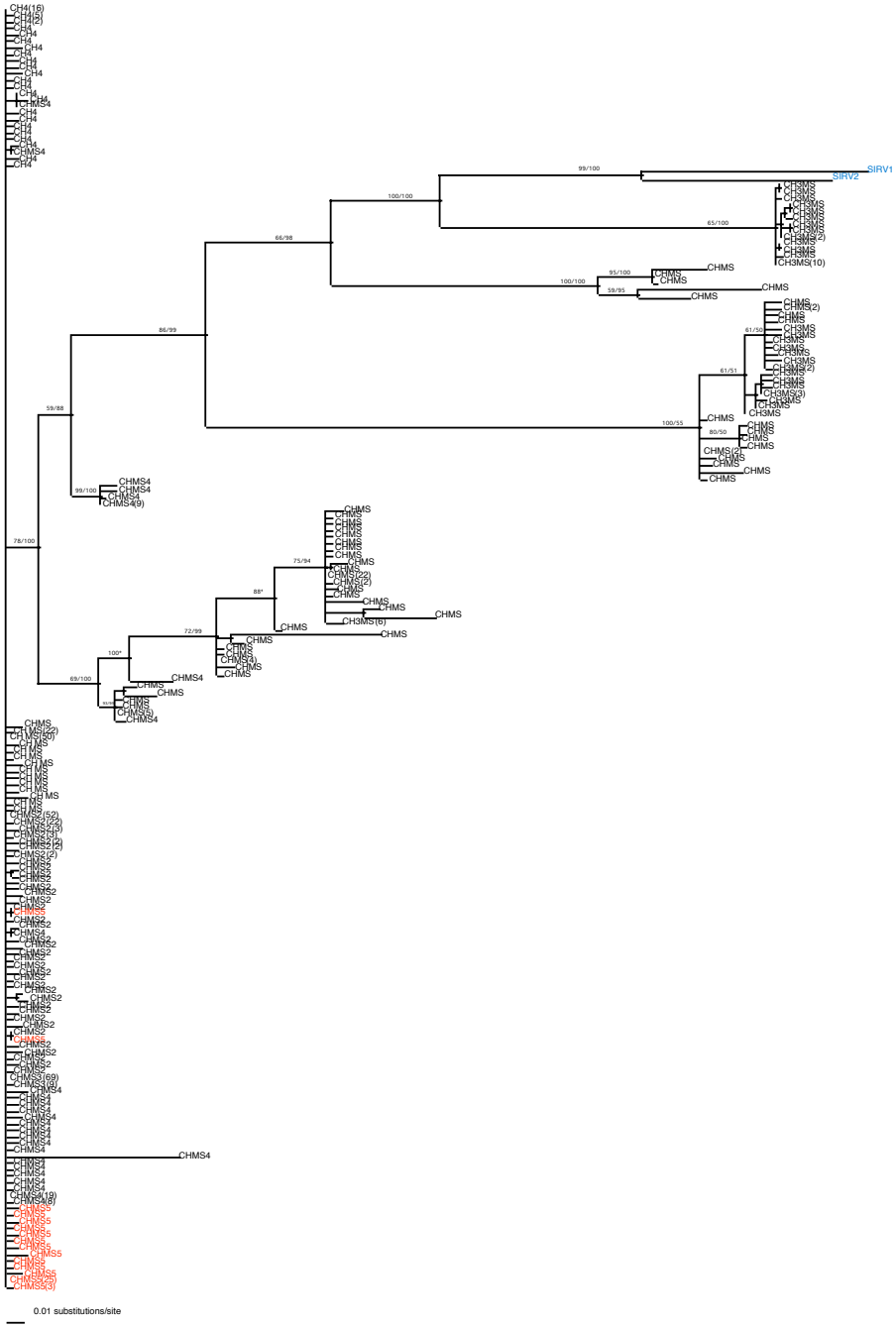


Figure 3.51. The SIRV-like protein sequences from Crater Hills over time. The samples in red represent those collected in September 2003. The * represents the identical protein sequences. The numbers in parentheses indicate how many times a sequence was detected.



Figure 3.52. The SIRV-like DNA sequences from Crater Hills. The sequences in red were collected in October 2003. The number in parentheses indicates how many times a sequence was detected.

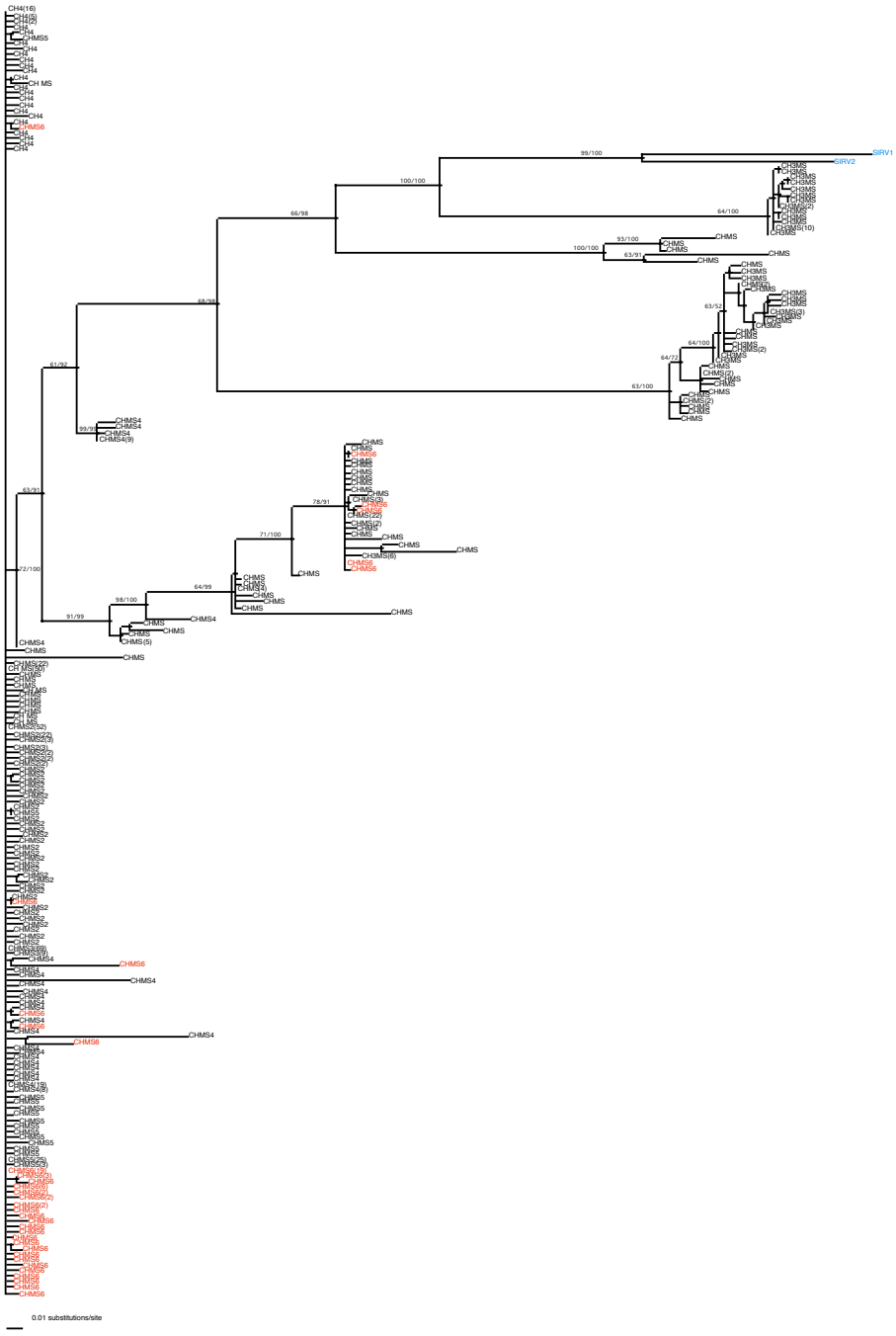


Figure 3.53. The SIRV-like protein sequences from Crater Hills. The sequences collected in October 2003 are colored red. The * represent sets of identical protein sequences. The numbers in parentheses indicate how many times a sequence was detected.

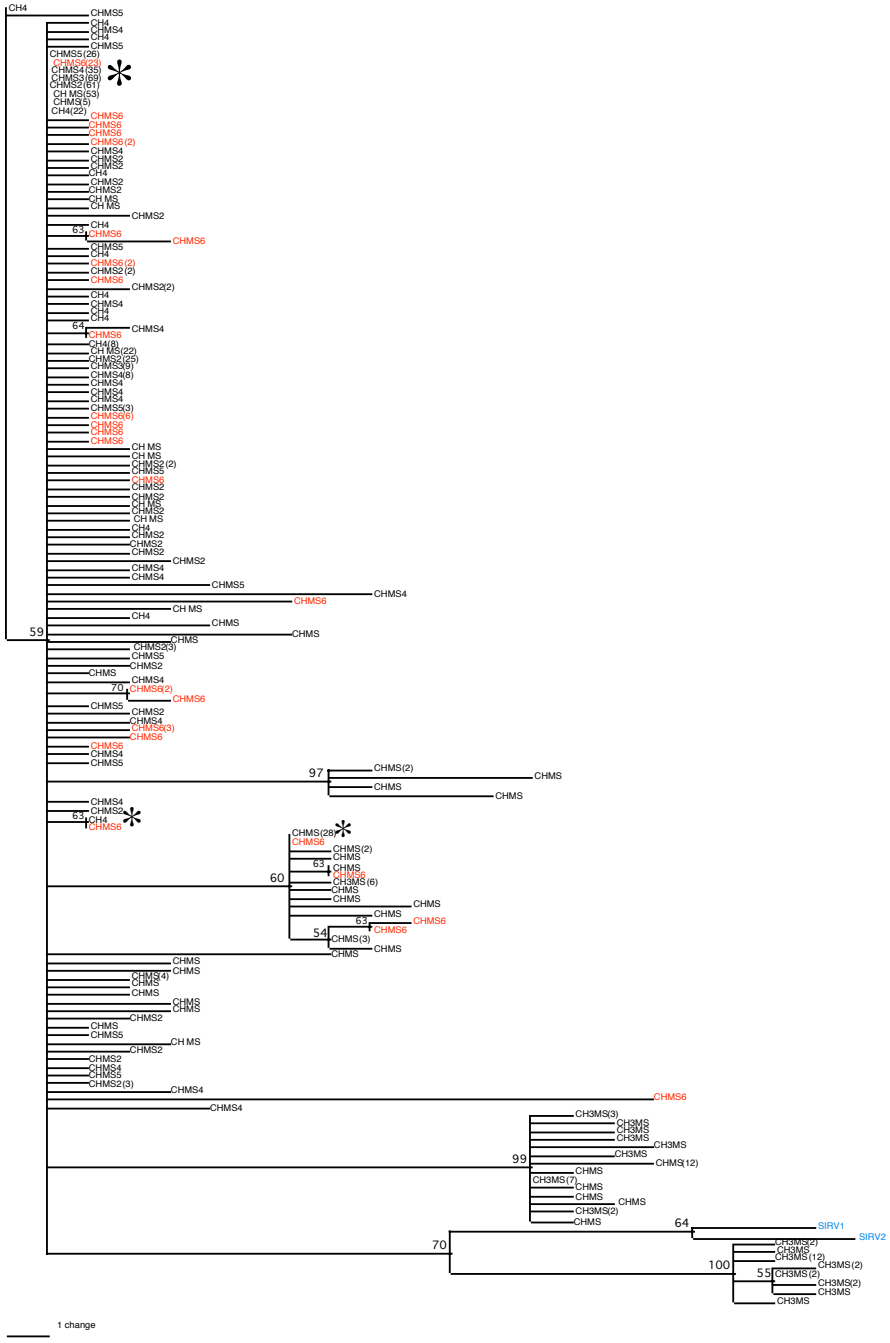


Figure 3.54. The SIRV-like DNA sequences from Crater Hills. The sequences collected in November 2003 are in red. The numbers in parentheses indicate how many times a sequence was detected.

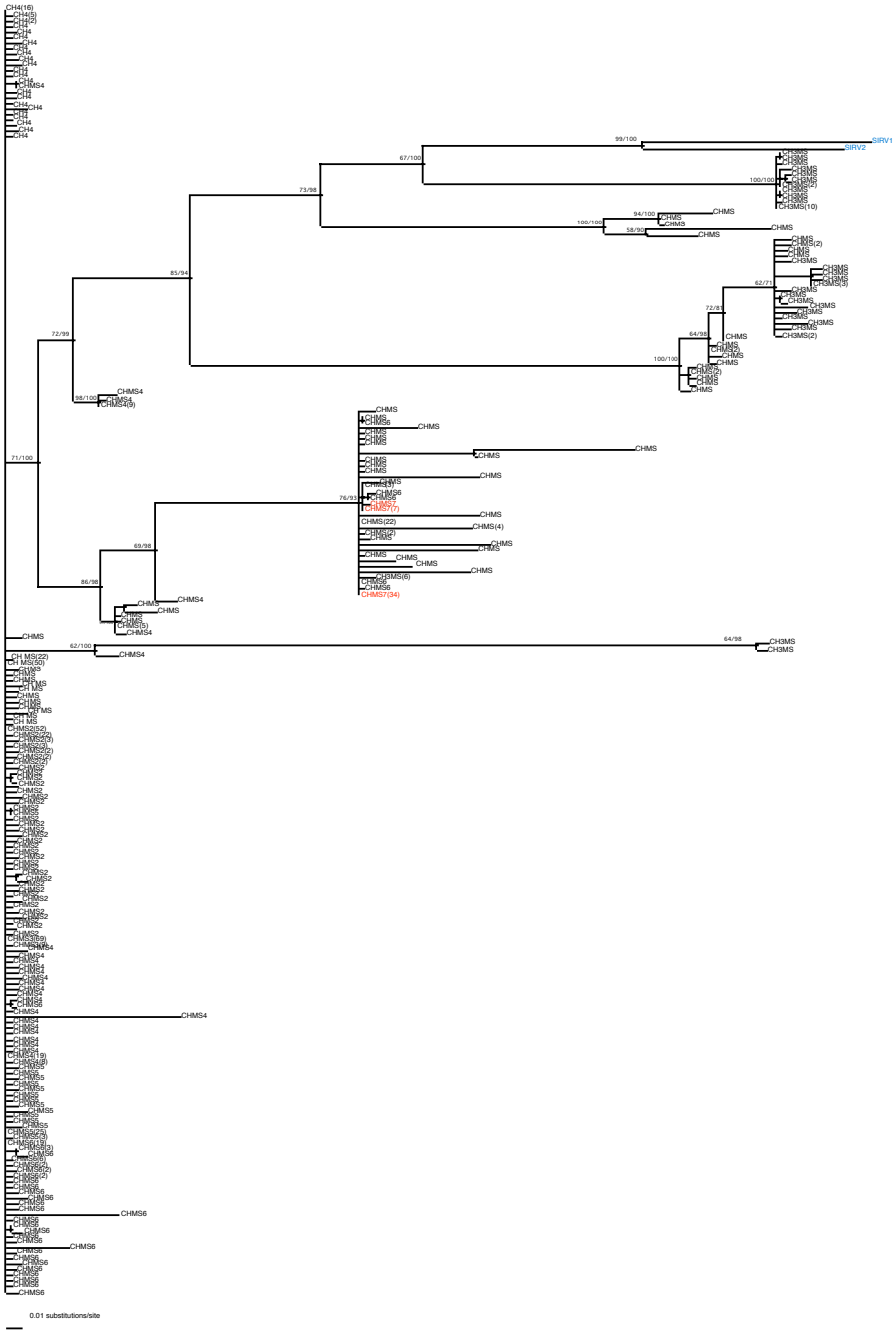


Figure 3.55. The SIRV-like protein sequences from Crater Hills. The sequences in red were detected in the November 2003 sample. The * represents groups of identical sequences. The numbers in parentheses indicates how many times a sequence was detected.

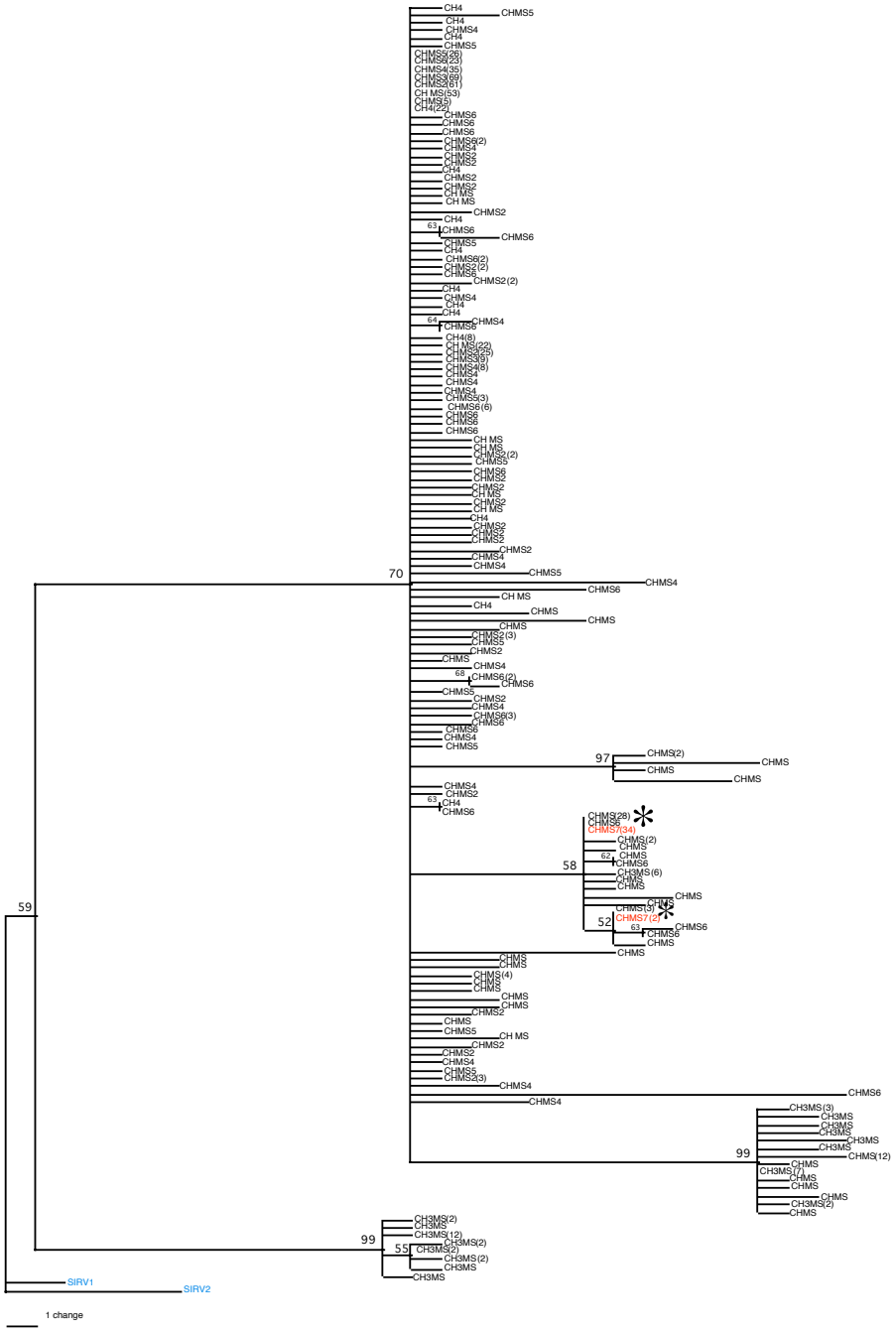


Figure 3.56. The SIRV-like DNA sequences obtained from Crater Hills. The sequences in red represent those collected in February 2004. The numbers in parentheses indicate how many times a sequence is detected.

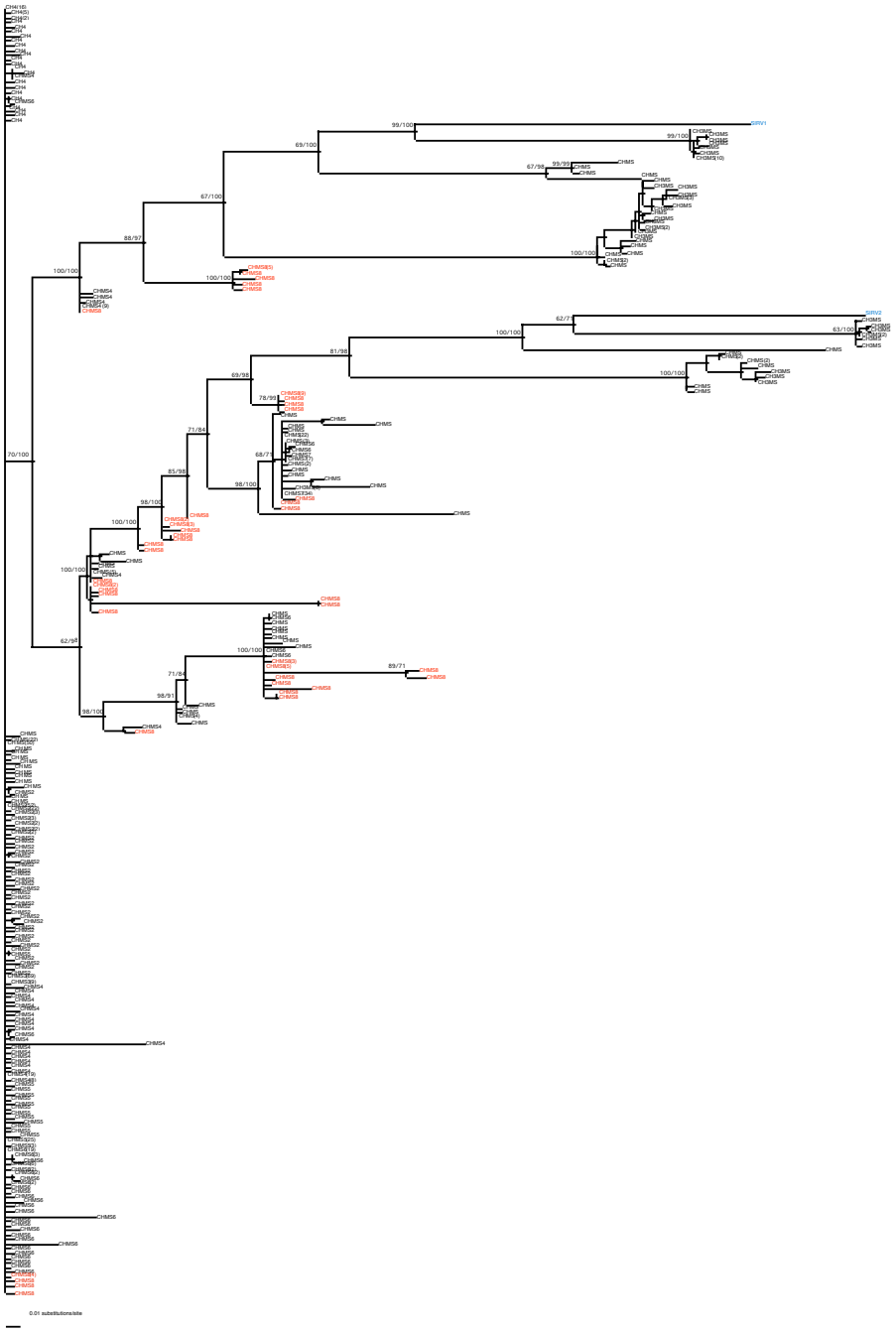


Figure 3.57. The SIRV-like protein sequences from Crater Hills. The sequences in red were collected in February 2004. The * represents groups of sequences with identical protein sequences. The numbers in parentheses indicate how many times a sequence was detected.

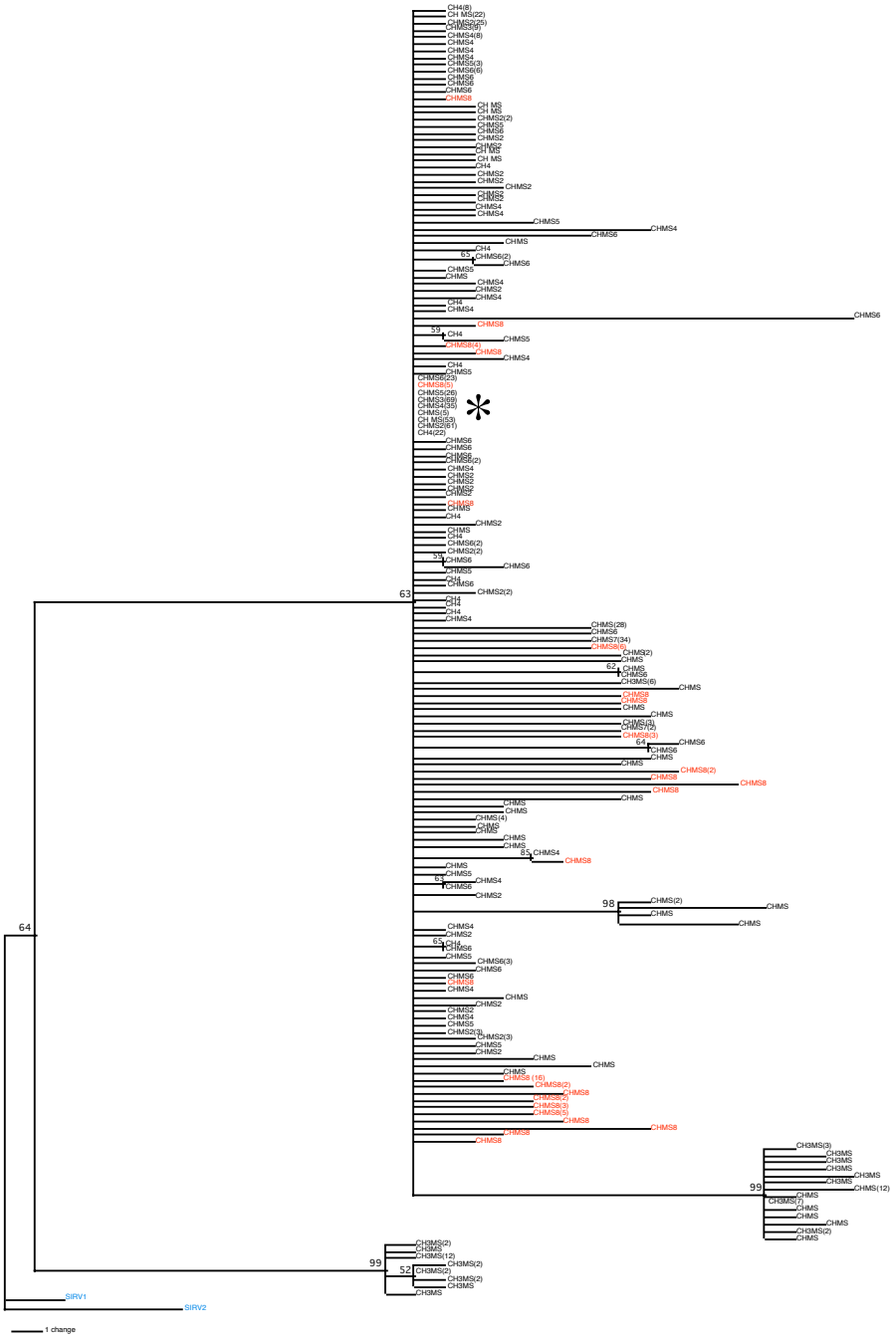


Figure 3.58. The SIRV-like DNA sequences from Crater Hills. The sequences collected in April 2004 are indicated in red. The numbers in parentheses indicate how many times a sequence was detected.

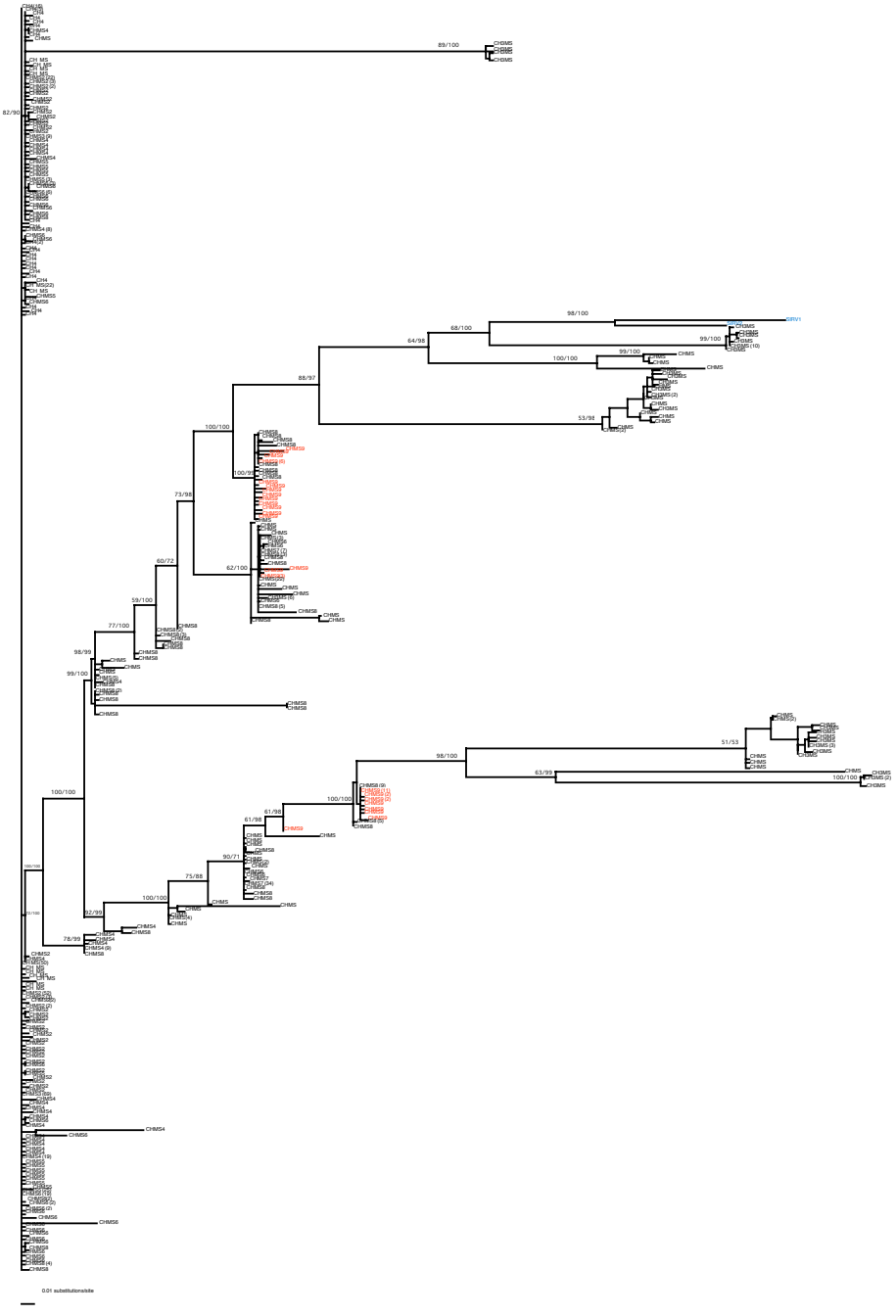


Figure 3.59. The SIRV-like protein sequences from Crater Hills. The protein sequences detected in April 2004 are colored red. The numbers in parentheses indicate how many times a sequence was detected.

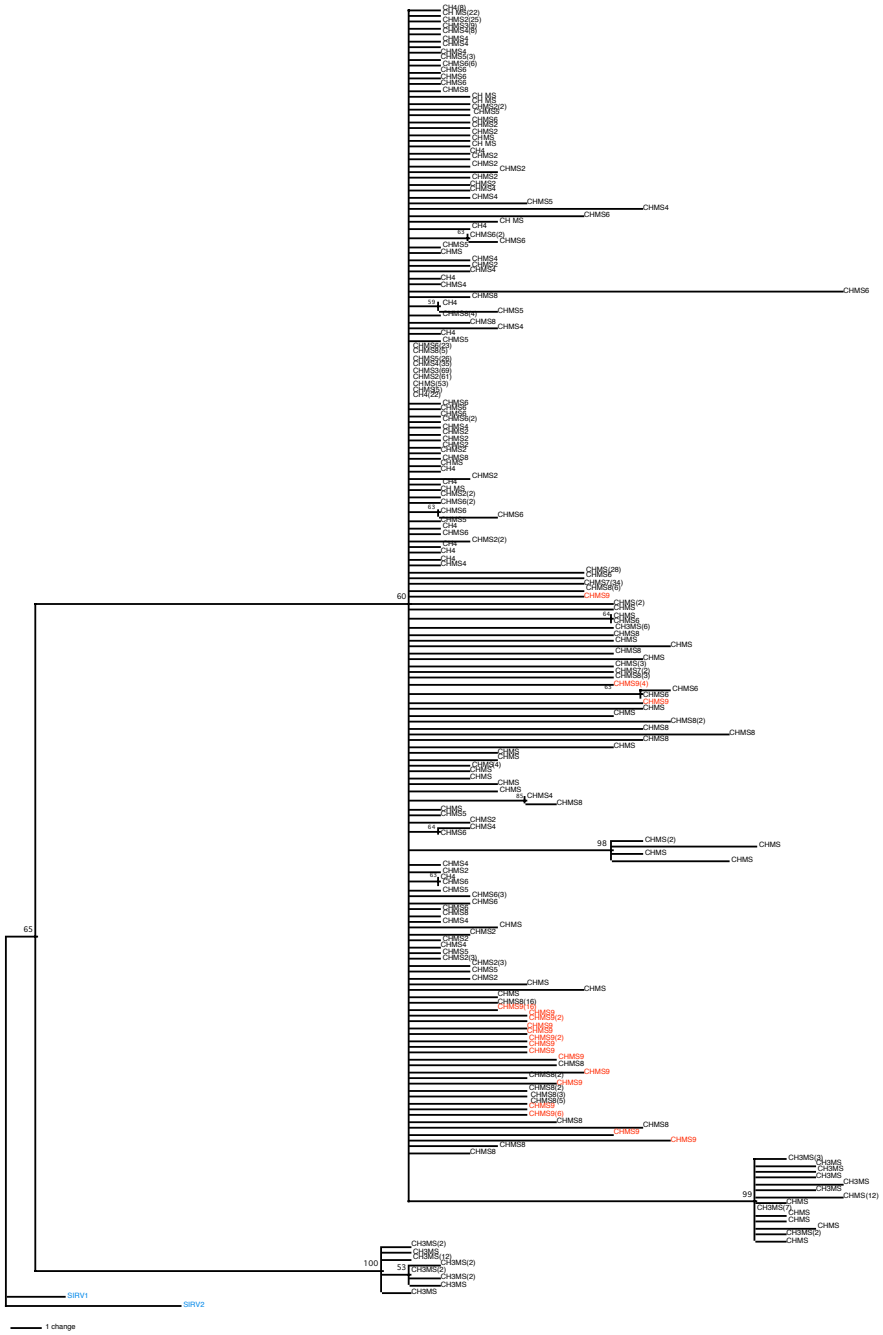


Figure 3.60. The SIRV-like DNA sequences collected from Crater Hills. The sequences collected in June 2004 are colored red. The numbers in parentheses indicate how many times a sequence was detected.

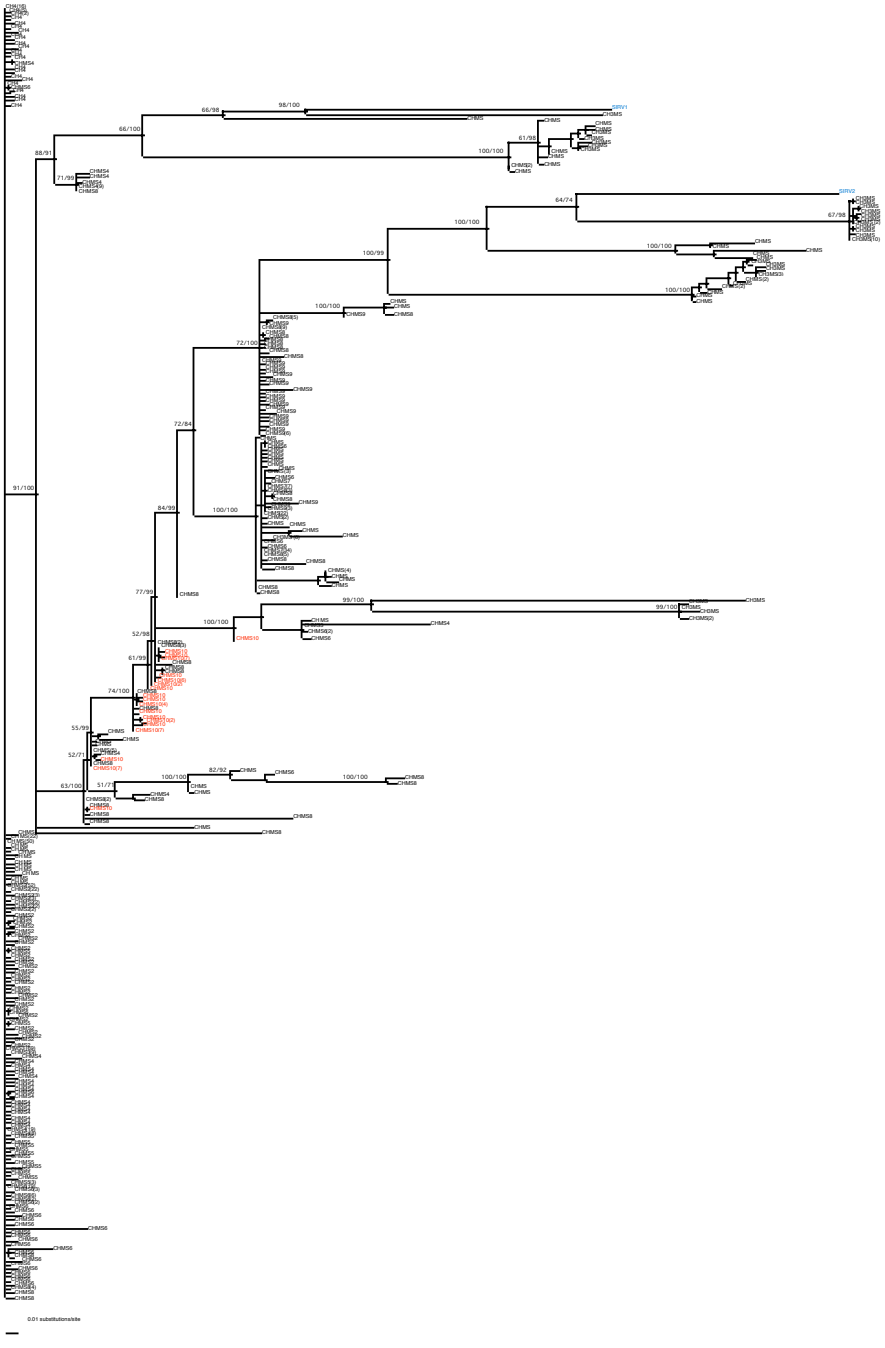
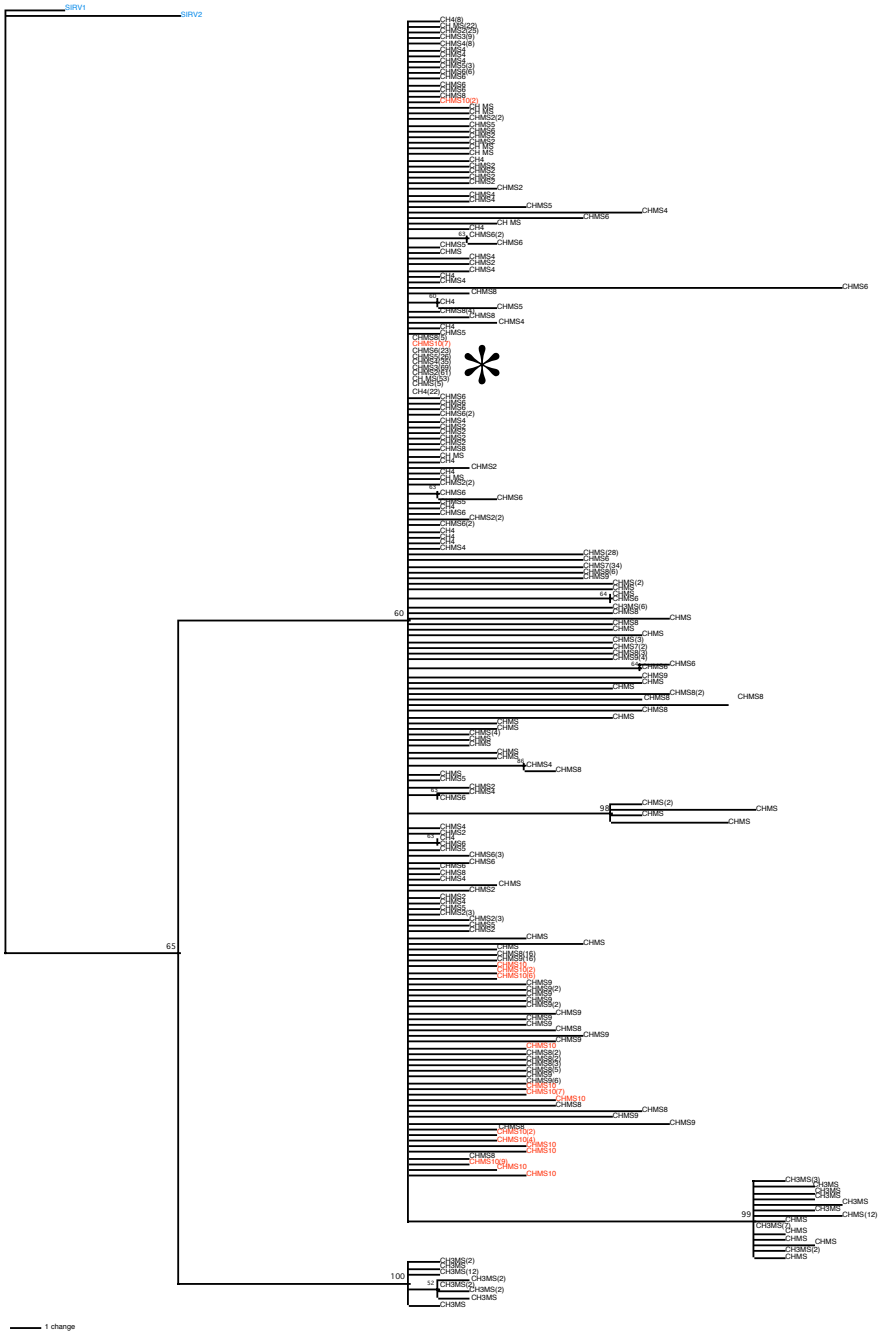


Figure 3.61. The SIRV-like protein sequences from Crater Hills. The sequences colored red were collected in June 2004. The * represents a group of identical protein sequences. The numbers in parentheses indicate how many times a sequence was detected.



APPENDIX B

SUPPLEMENTARY PHYLOGENETIC ANALYSES FROM THE RABBIT CREEK MONITOR SITE

Figure 4.29. SSV-like 3F/4R sequences collected from Rabbit Creek. The red sequences represent those collected in October 2002. A. DNA sequences; the • represent non-functional ORFs. B. Inferred protein sequences; the * represents identical protein sequences. The numbers in parentheses indicate how many times a sequence was detected.

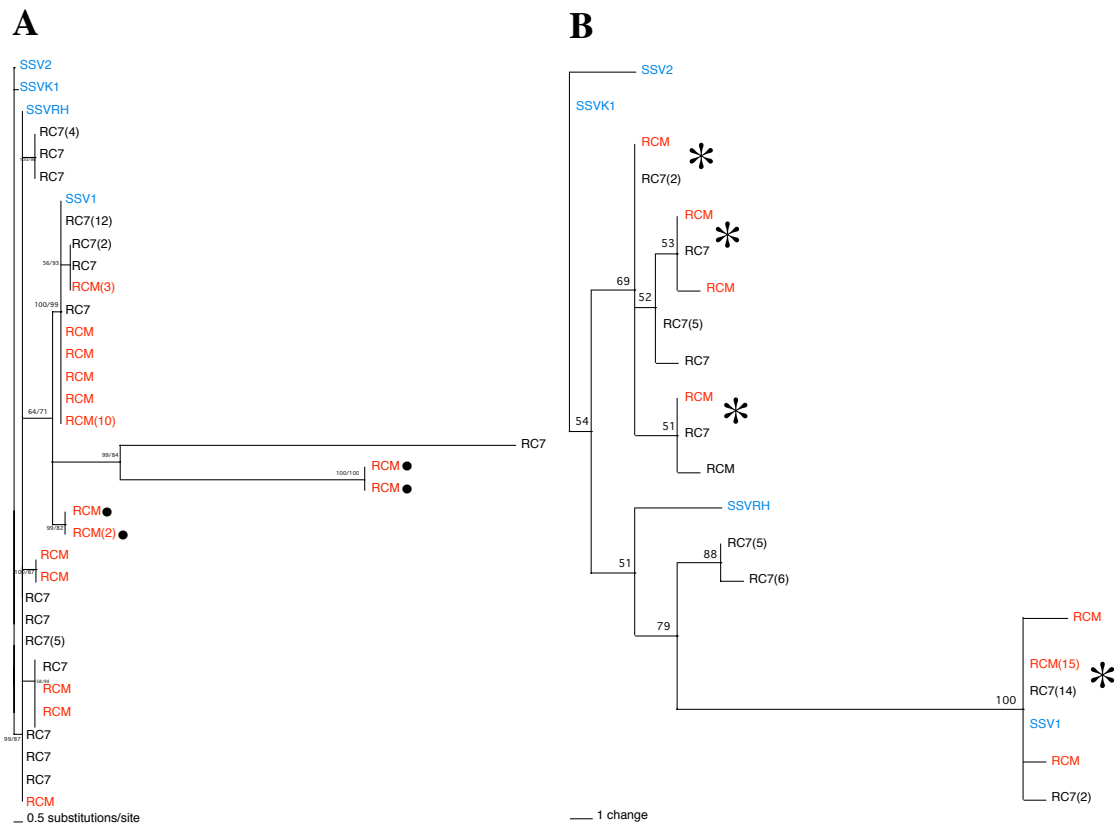


Figure 4.30. SSV-like 3F/4R sequences collected from Rabbit Creek. The red sequences represent those collected in January 2003. A. DNA sequences; the • represent non-functional ORFs. B. Inferred protein sequences; the * represents identical protein sequences. The numbers in parentheses indicate how many times a sequence was detected.

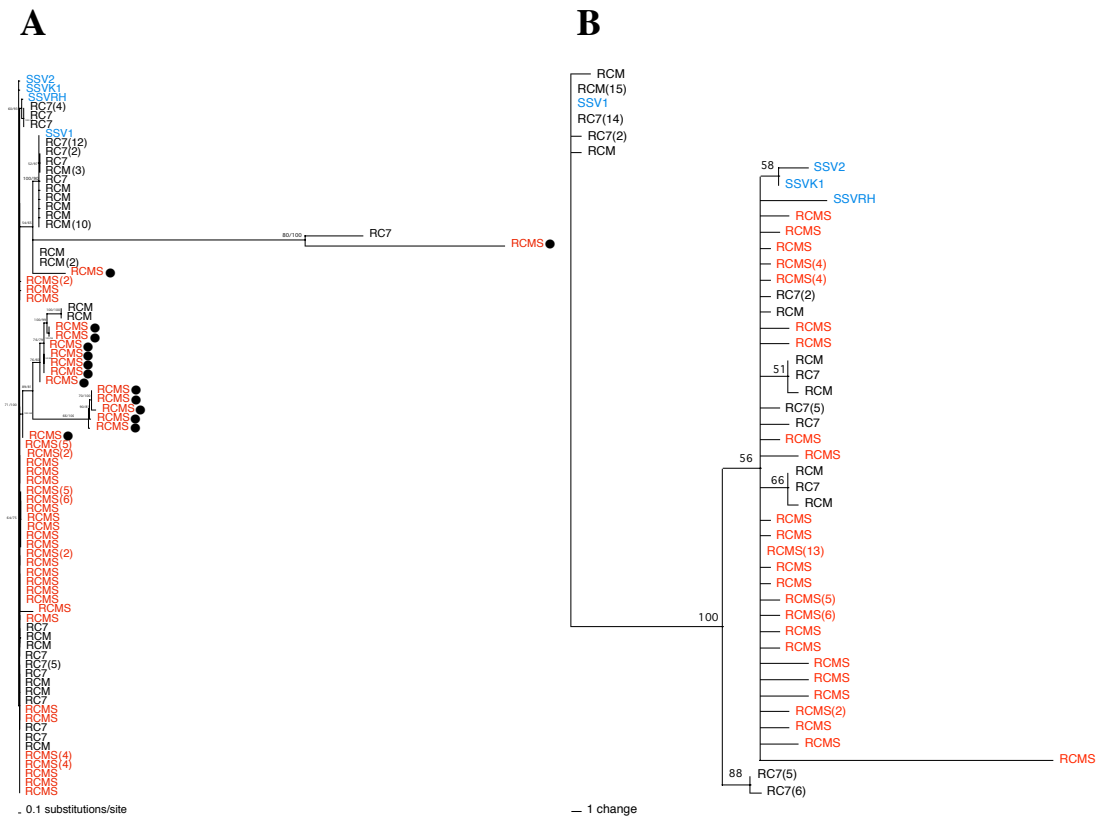


Figure 4.31. SSV-like 3F/4R DNA sequences collected from Rabbit Creek. The red sequences represent those collected in August 2003. The • represent non-functional ORFs. The numbers in parentheses indicate how many times a sequence was detected.

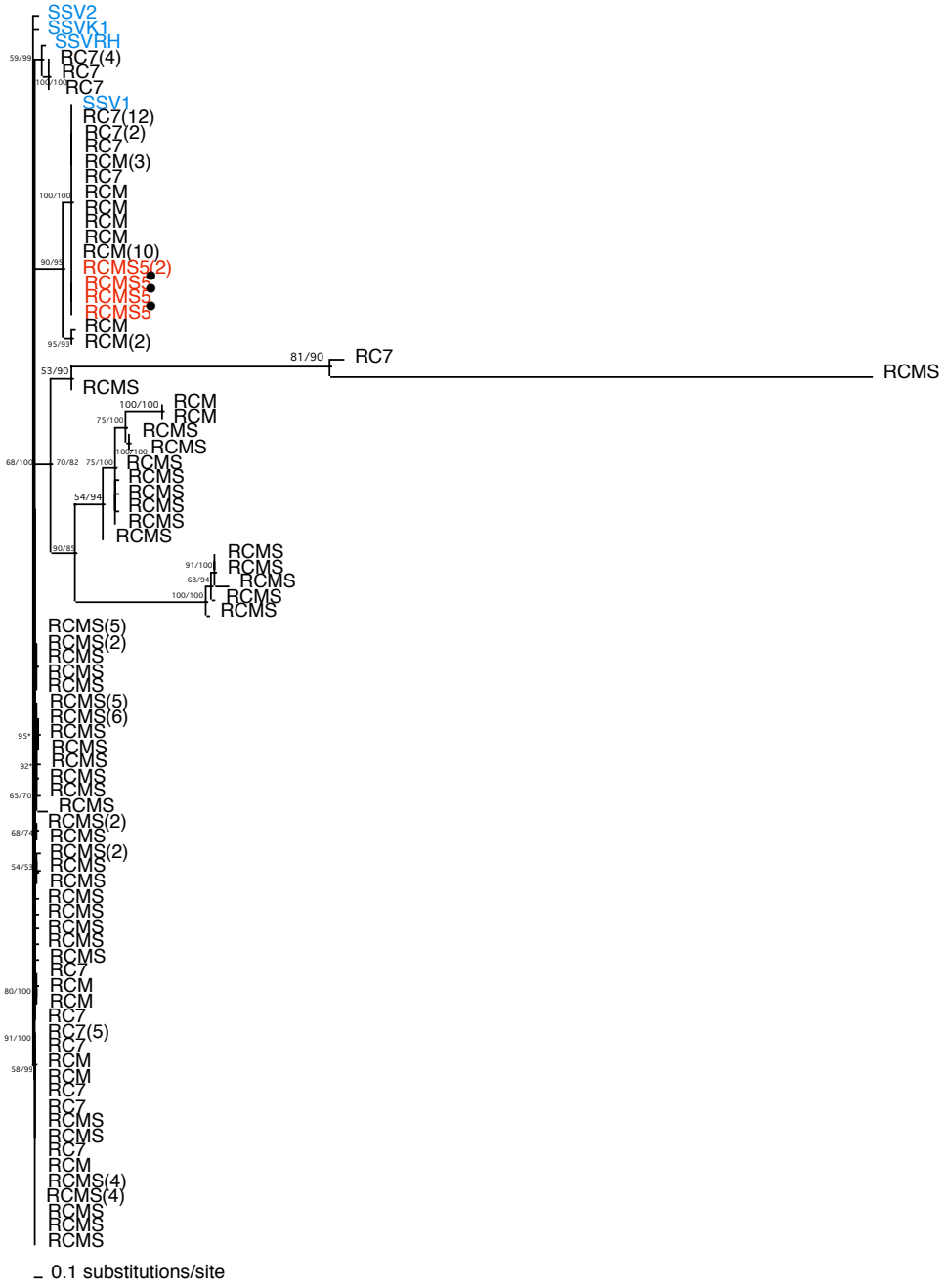


Figure 4.32. SSV-like 3F/4R protein sequences collected from Rabbit Creek. The red sequences represent those collected in August 2003. The * represents identical protein sequences. The numbers in parentheses indicate how many times a sequence was detected.

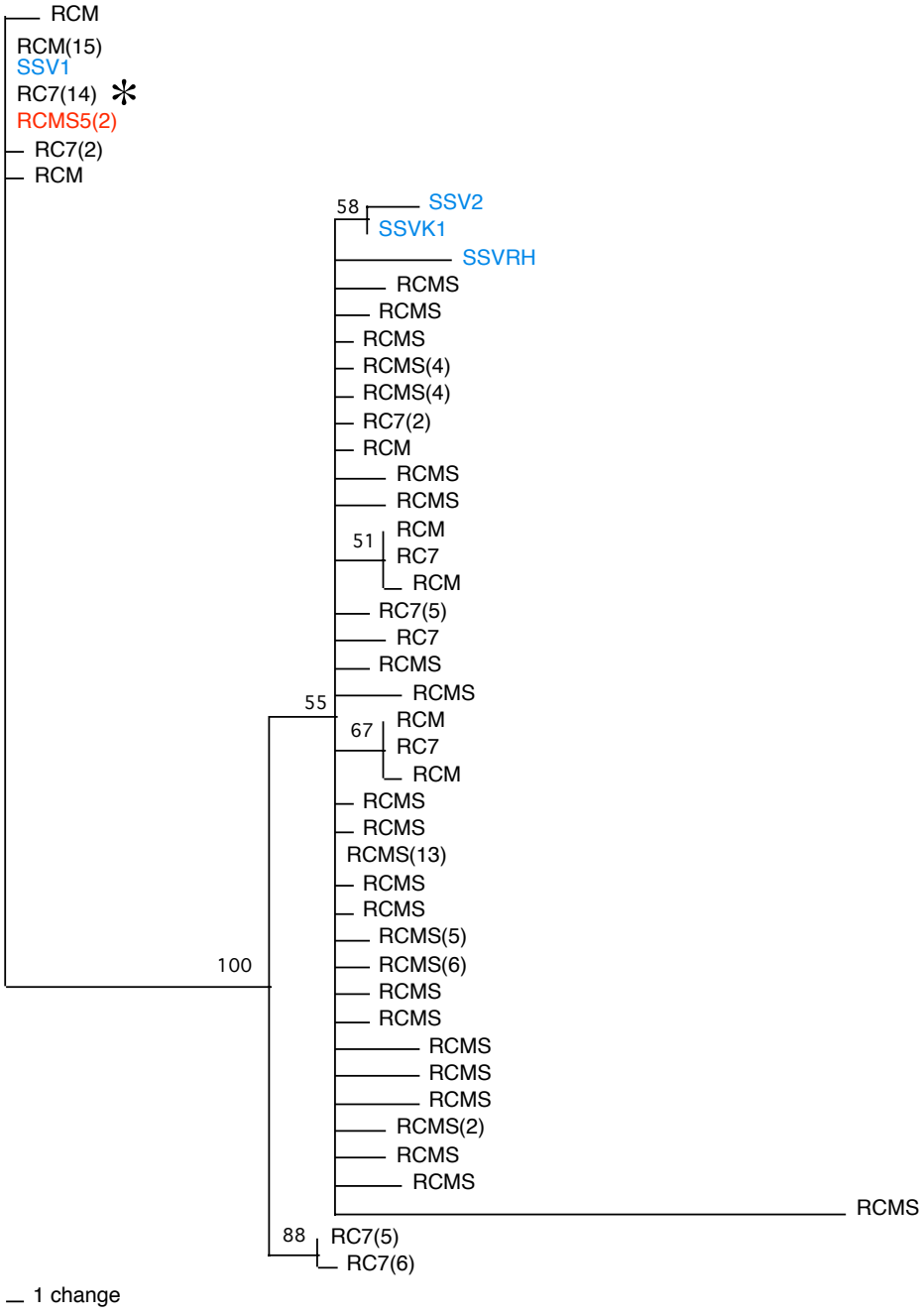


Figure 4.33. SSV-like 3F/4R DNA sequences collected from Rabbit Creek. The red sequences represent those collected in October 2003. The • represent non-functional ORFs. The numbers in parentheses indicate how many times a sequence was detected.

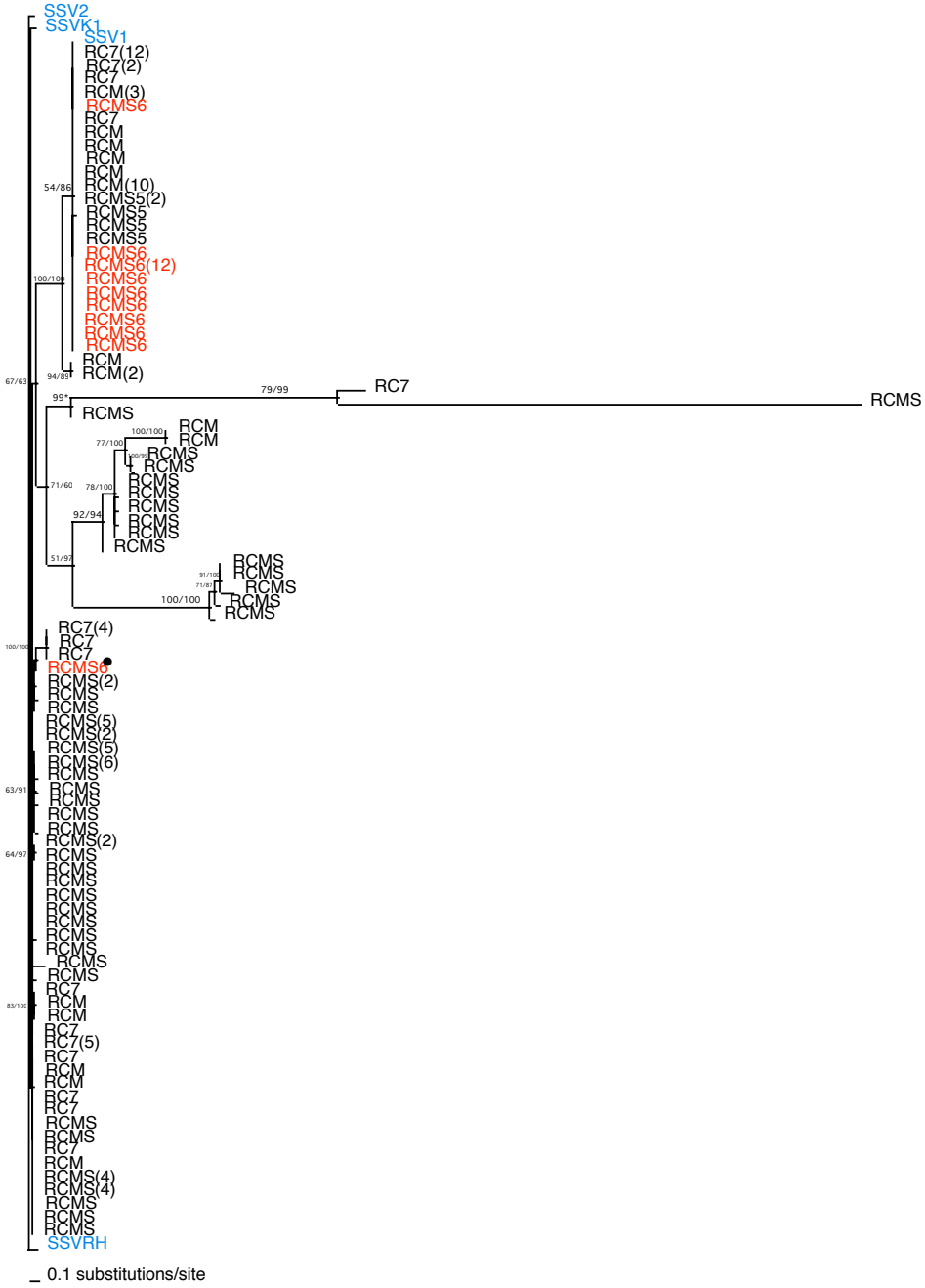


Figure 4.34. SSV-like 3F/4R protein sequences collected from Rabbit Creek. The red sequences represent those collected in October 2003. The * represents identical protein sequences. The numbers in parentheses indicate how many times a sequence was detected.

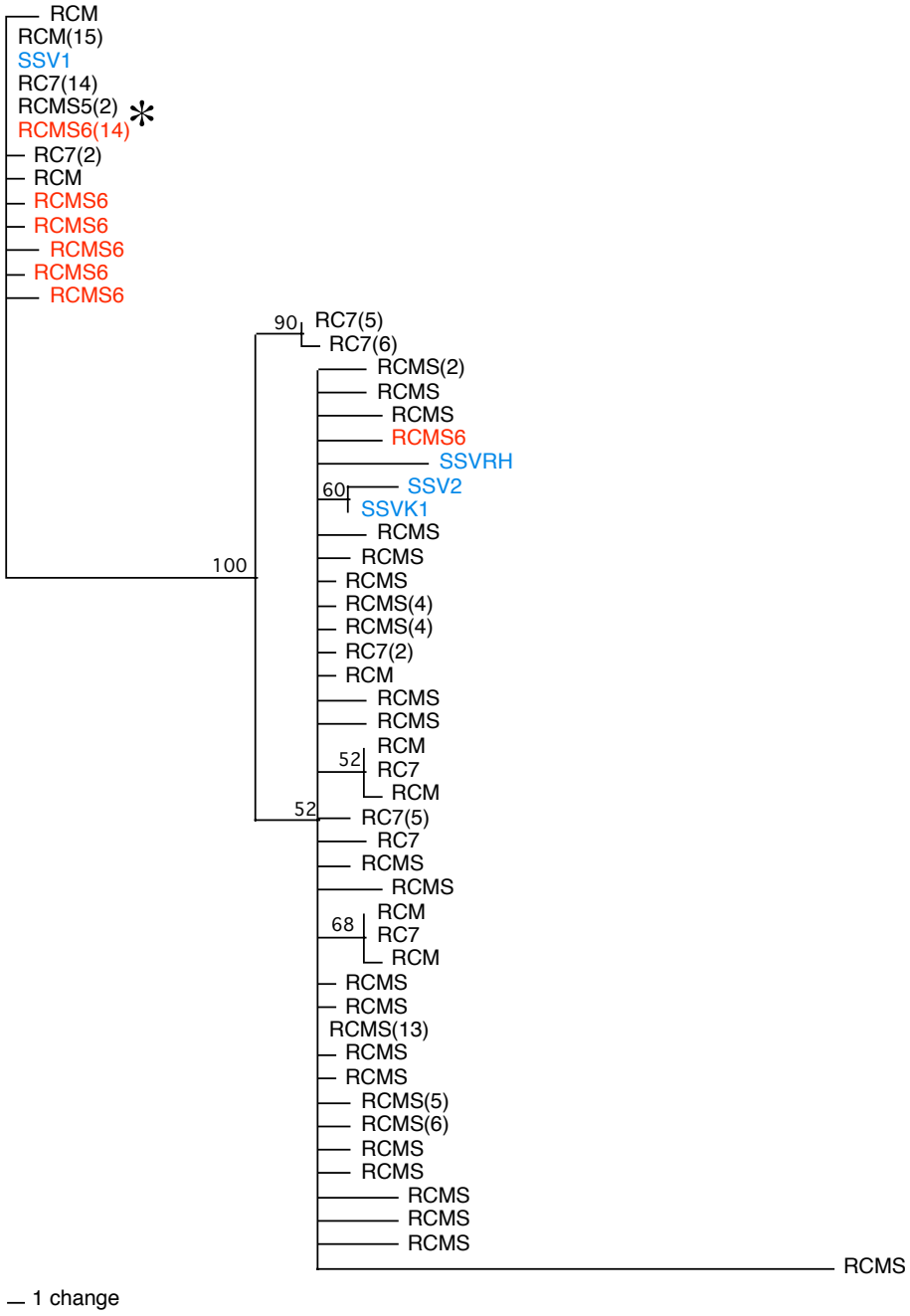


Figure 4.35. SSV-like 3F/4R DNA sequences collected from Rabbit Creek. The red sequences represent those collected in November 2003. The • represent non-functional ORFs. The numbers in parentheses indicate how many times a sequence was detected.

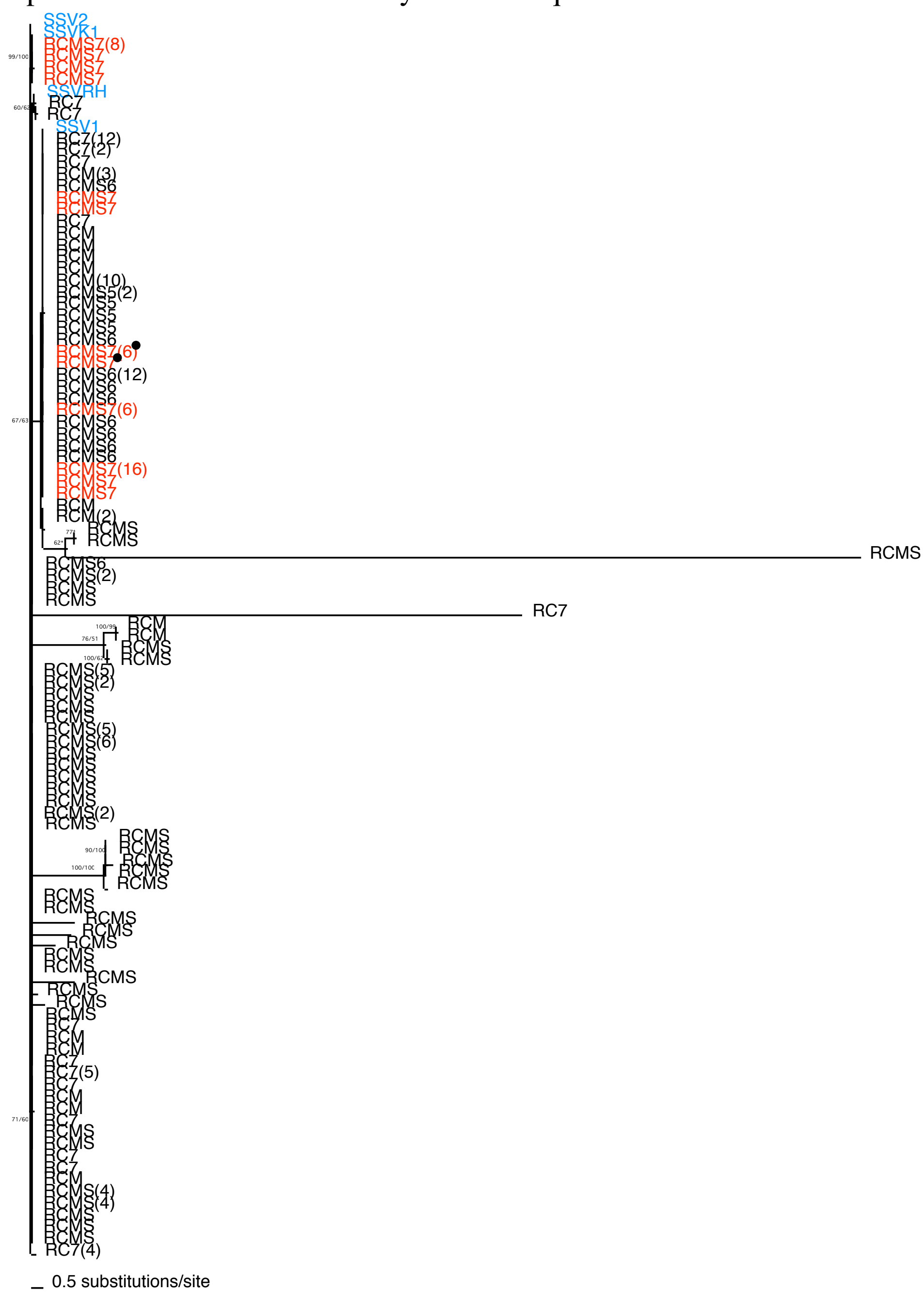


Figure 4.36. SSV-like 3F/4R protein sequences collected from Rabbit Creek. The red sequences represent those collected in November 2003. The * represents identical protein sequences. The numbers in parentheses indicate how many times a sequence was detected.

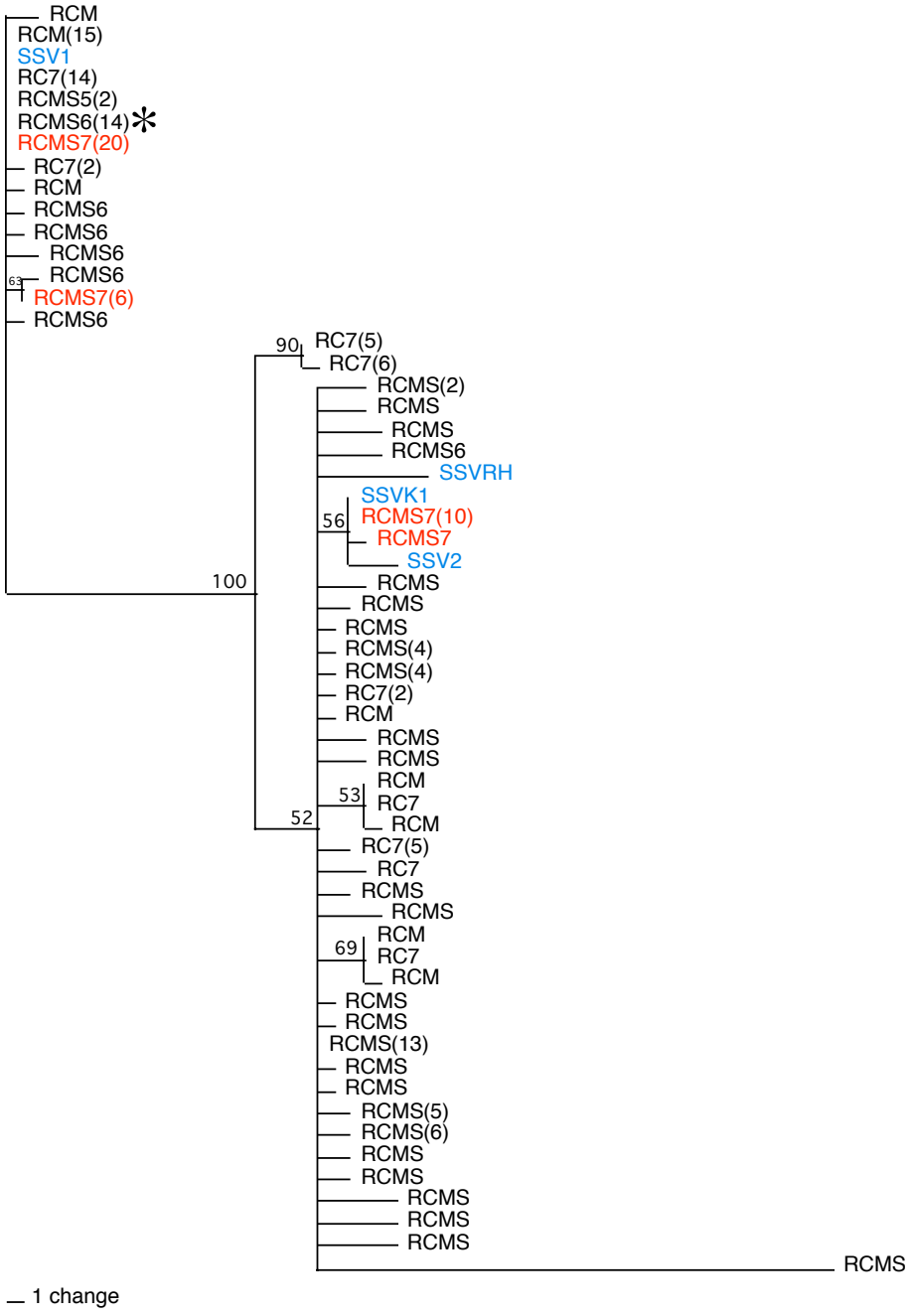


Figure 4.37. SSV-like 3F/4R DNA sequences collected from Rabbit Creek. The red sequences represent those collected in January 2004. The • represent non-functional ORFs. The numbers in parentheses indicate how many times a sequence was detected.

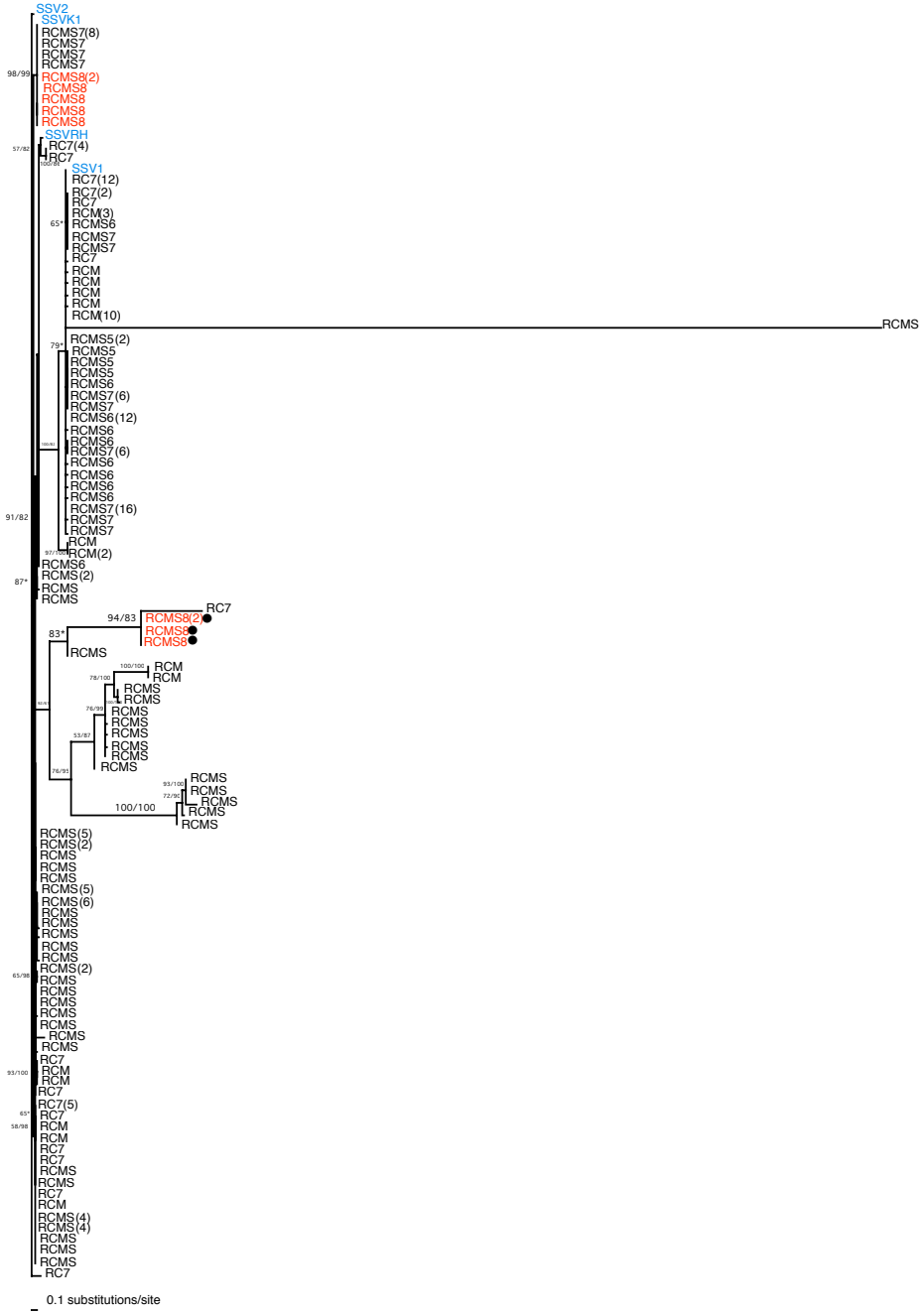


Figure 4.38. SSV-like 3F/4R protein sequences collected from Rabbit Creek. The red sequences represent those collected in January 2004. The * represents identical protein sequences. The numbers in parentheses indicate how many times a sequence was detected.

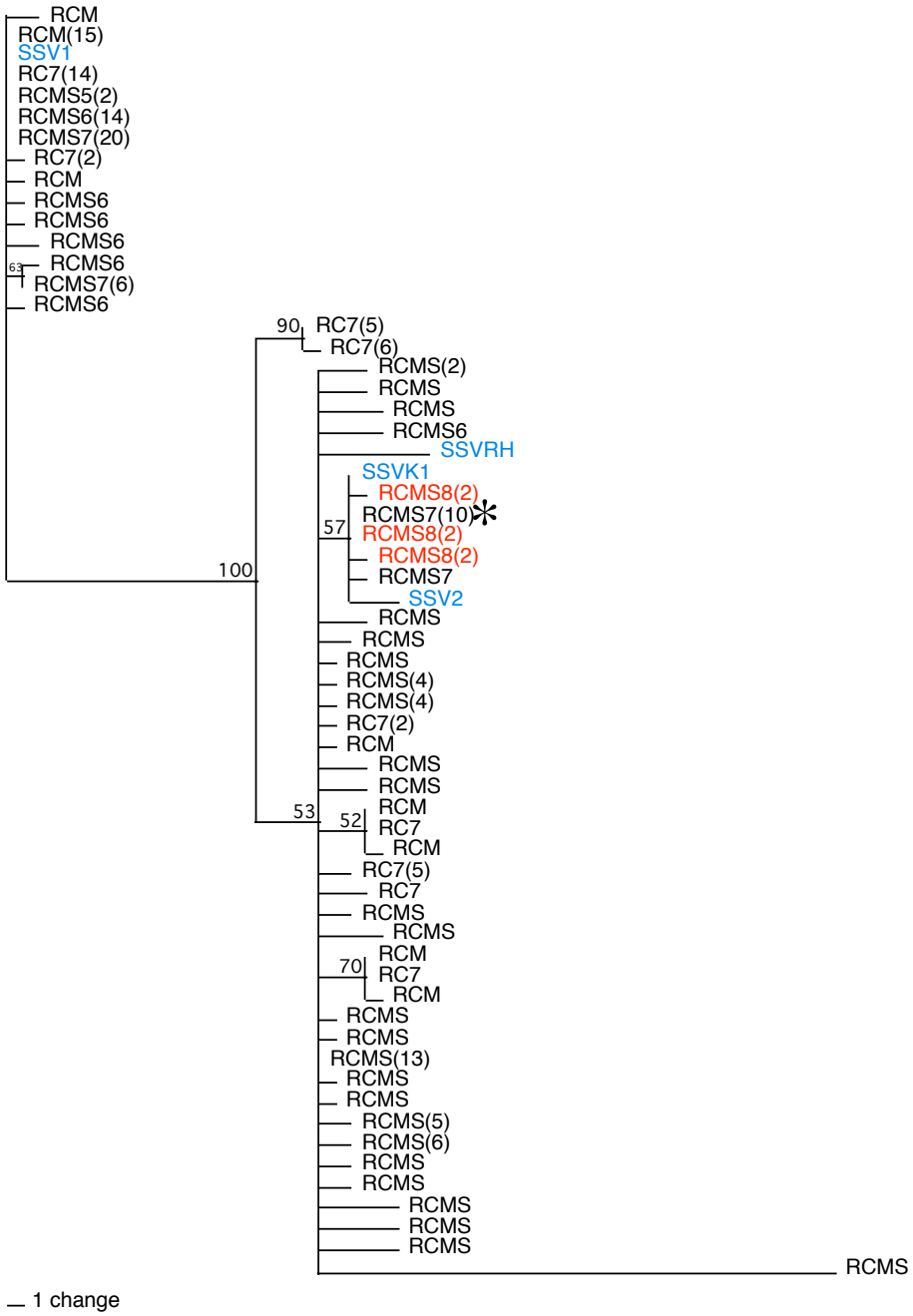


Figure 4.39. The SSV-like 3F/4R DNA sequence groups detected from July 2002 to January 2004 in Rabbit Creek. This analysis included functional and non-functional ORFs. The numbers in parentheses indicate how many times a sequence was detected.

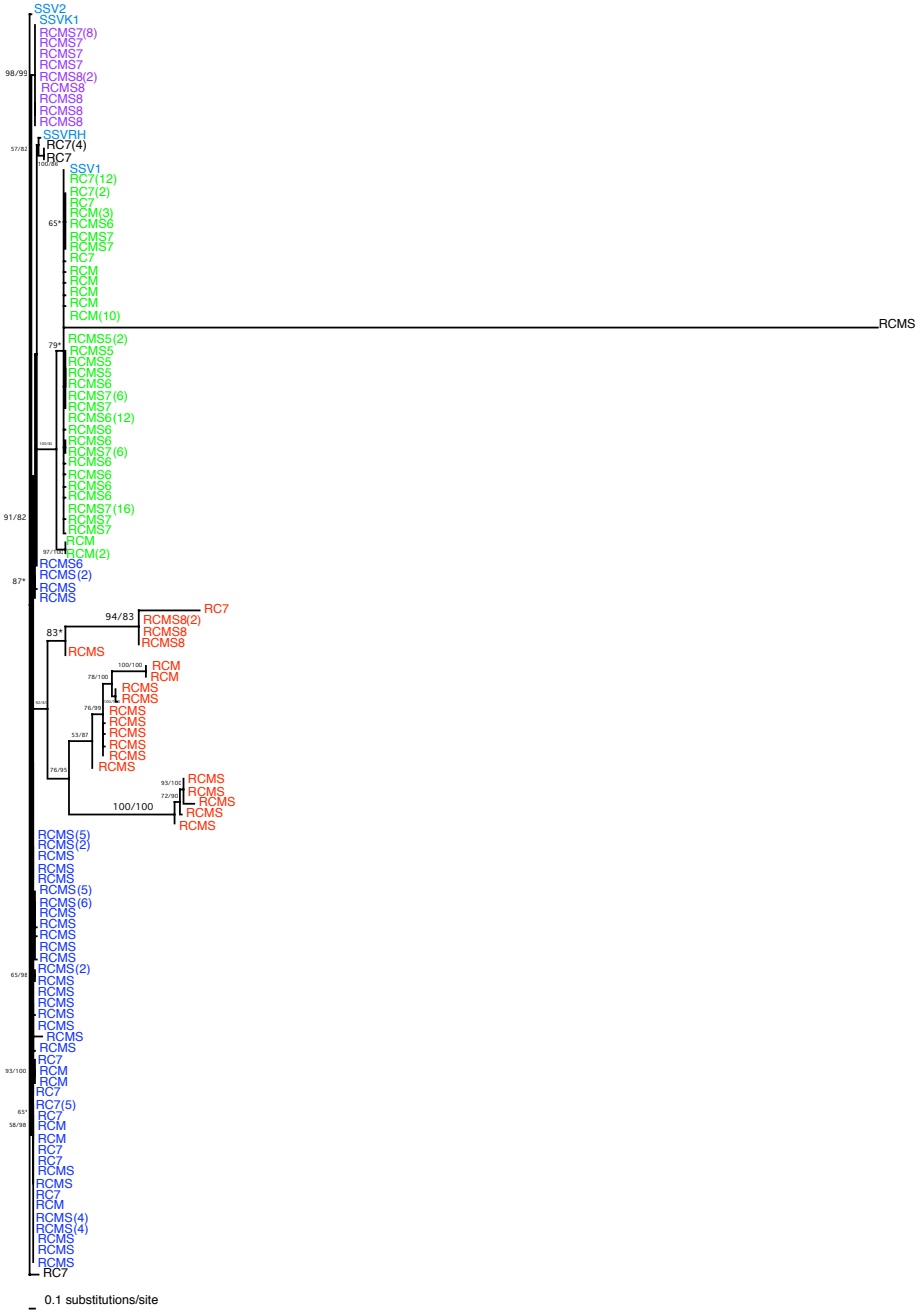


Figure 4.40. SIRV-like sequences collected from Rabbit Creek. The red sequences represent those collected in October 2002. The numbers in parentheses indicate how many times a sequence was detected. A. DNA sequences. B. Inferred protein sequences; the * represents identical protein sequences.

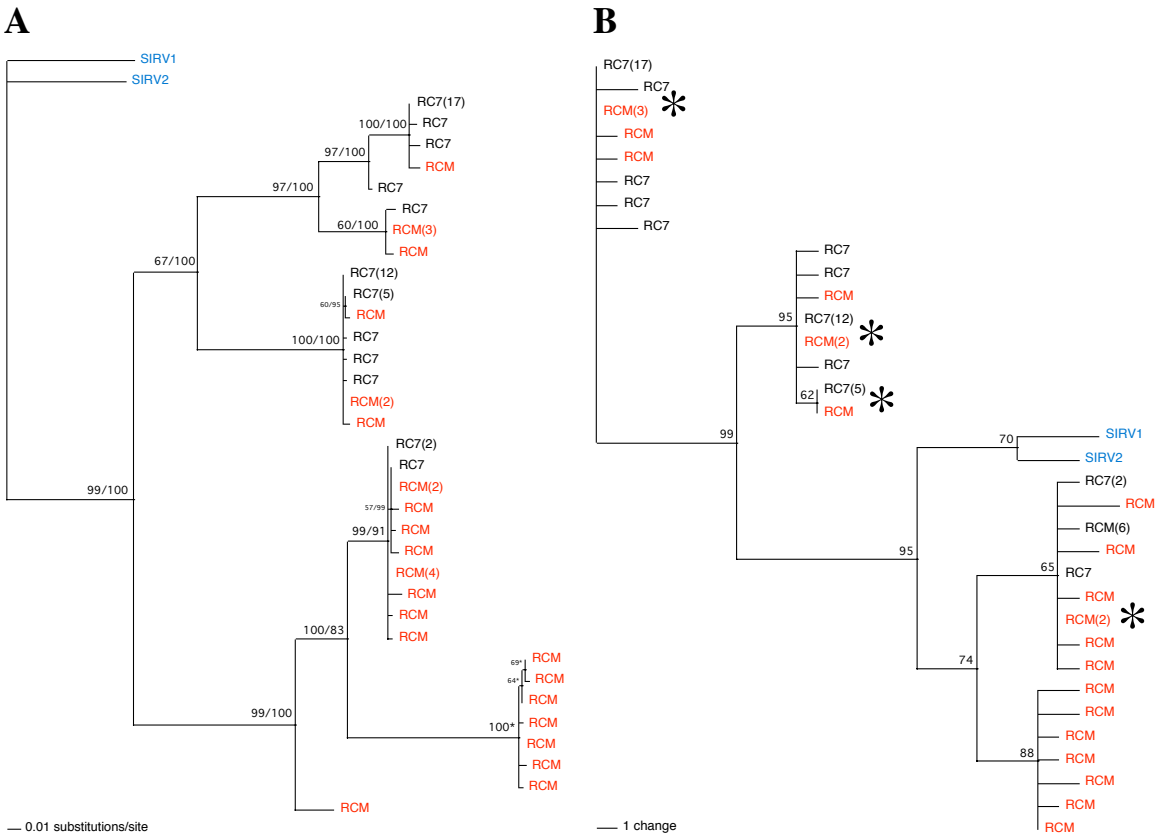


Figure 4.41. SIRV-like DNA sequences collected from Rabbit Creek. The red sequences represent those collected in January 2003. The numbers in parentheses indicate how many times a sequence was detected.

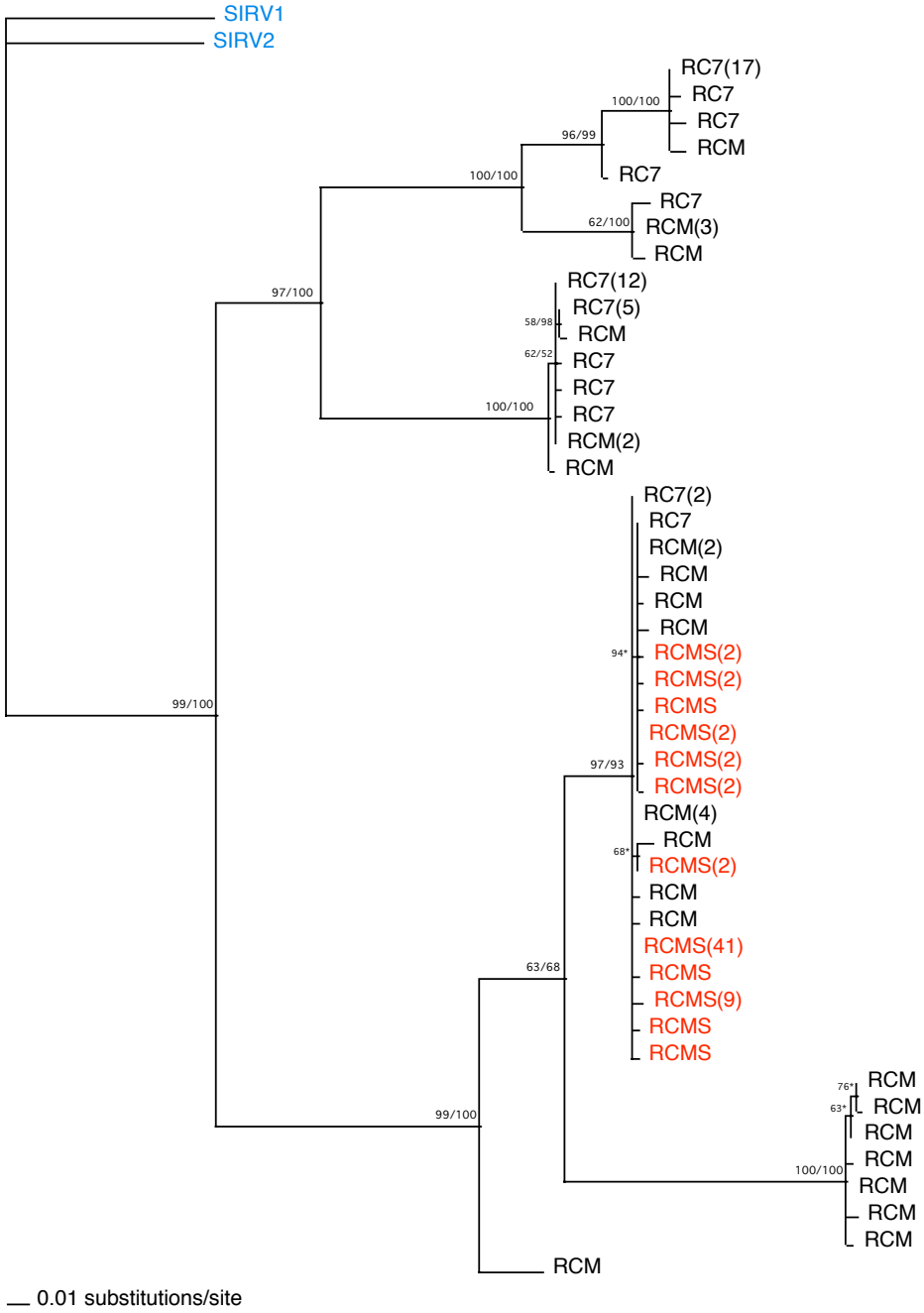


Figure 4.42. SIRV-like protein sequences collected from Rabbit Creek. The red sequences represent those collected in January 2003. The * represents identical protein sequences. The numbers in parentheses indicate how many times a sequence was detected.

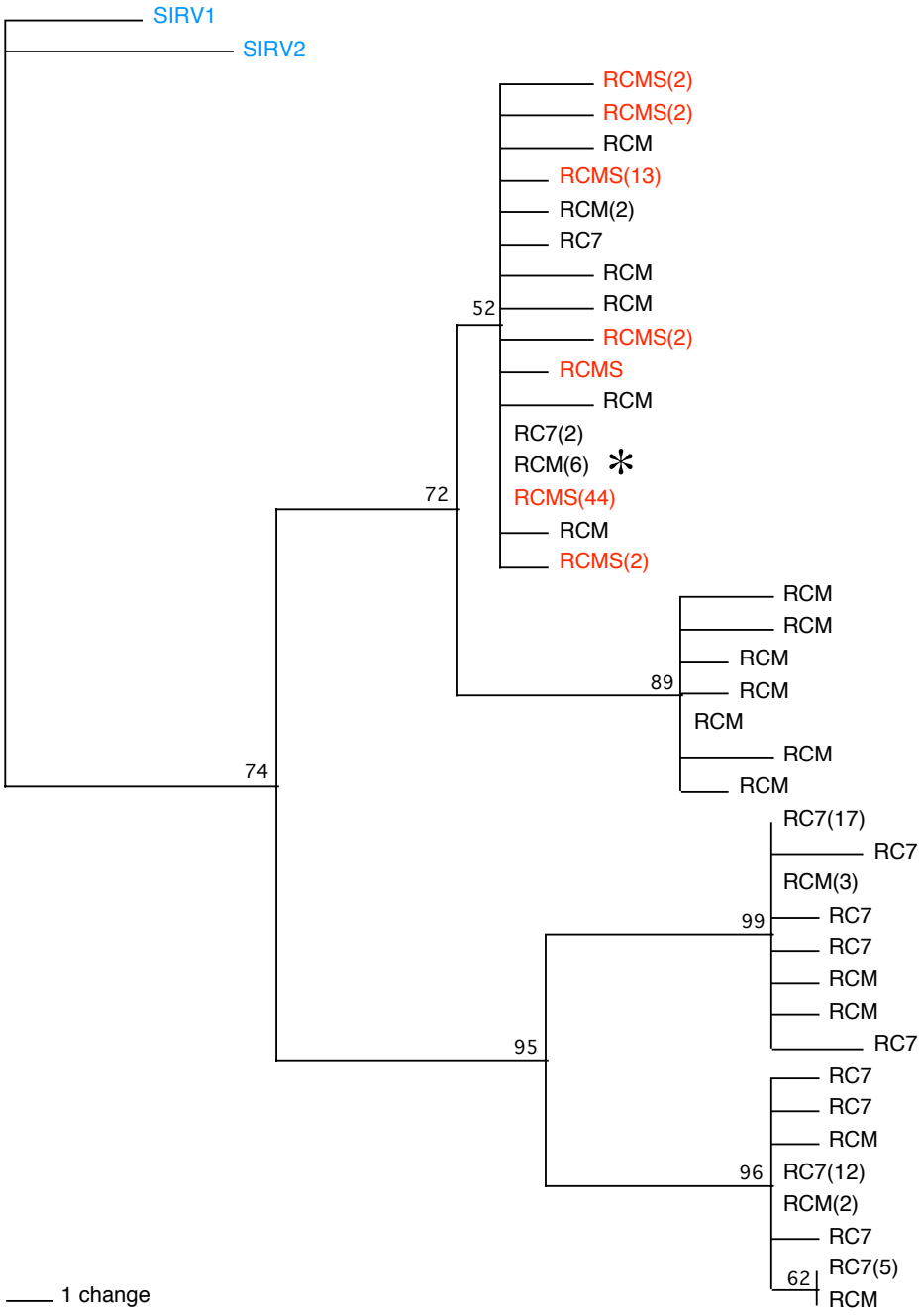


Figure 4.43. SIRV-like DNA sequences collected from Rabbit Creek. The red sequences represent those collected in May 2003. The numbers in parentheses indicate how many times a sequence was detected.

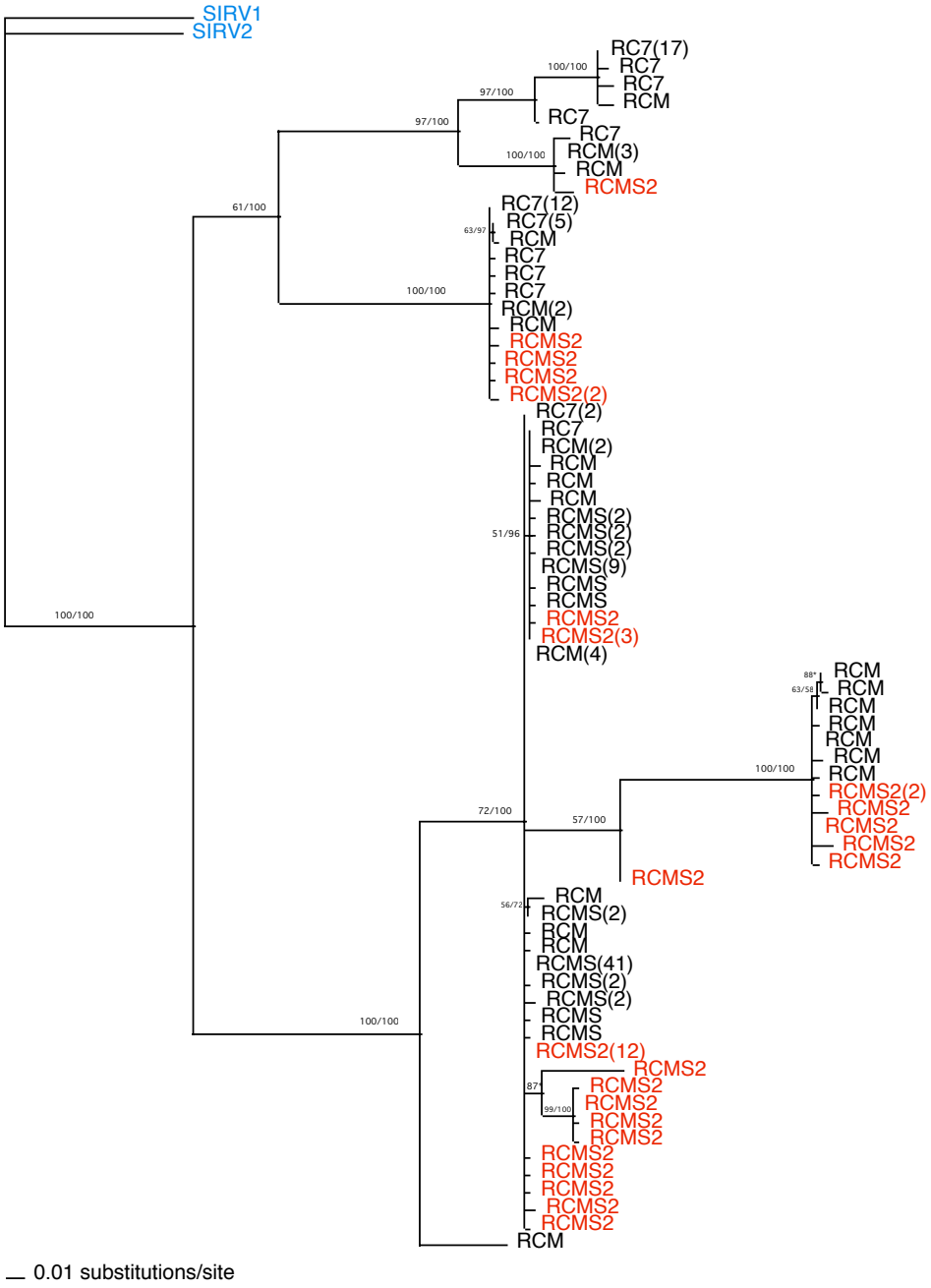


Figure 4.44. SIRV-like protein sequences collected from Rabbit Creek. The red sequences represent those collected in May 2003. The * represents identical protein sequences. The numbers in parentheses indicate how many times a sequence was detected.

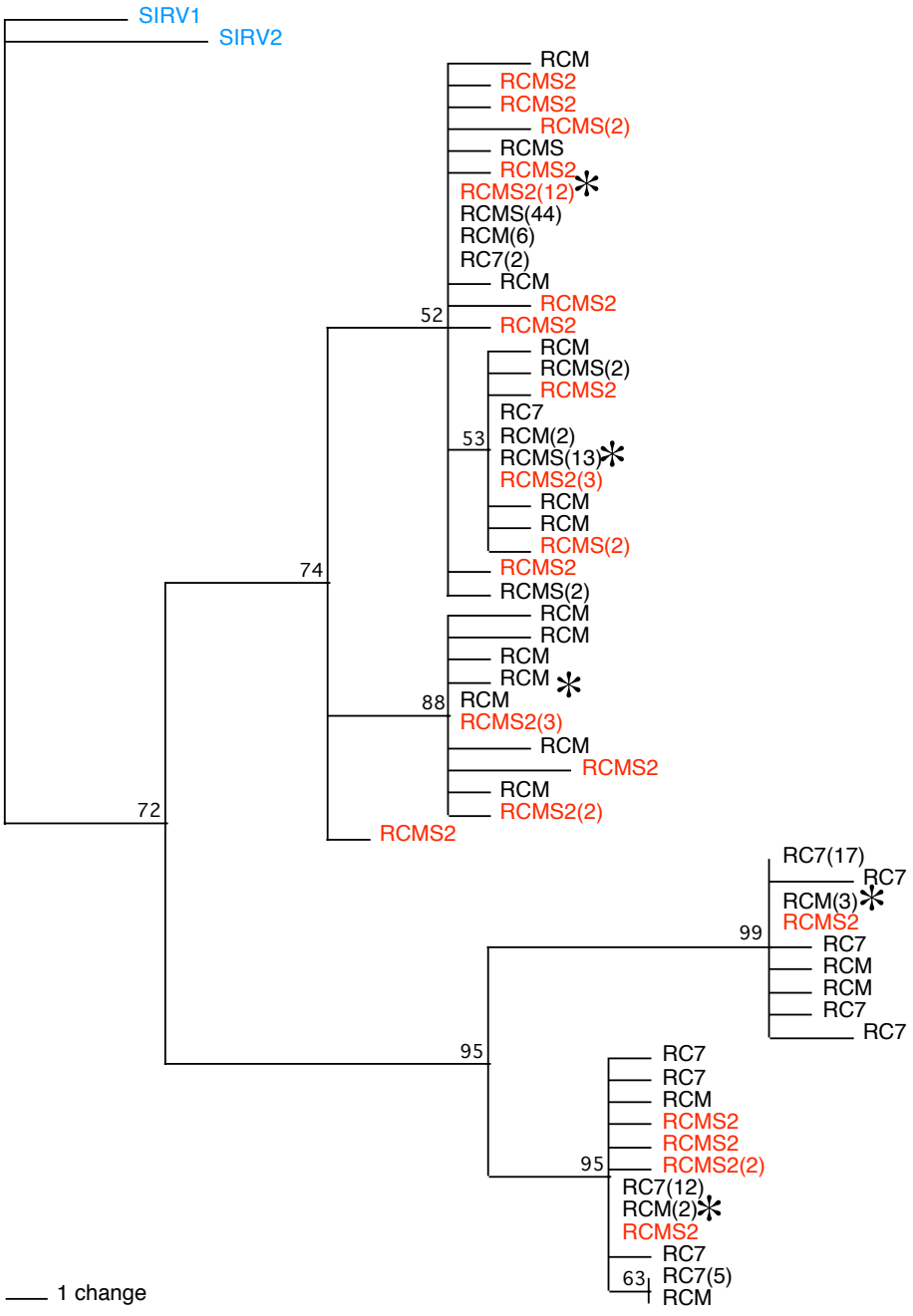


Figure 4.45. SIRV-like DNA sequences collected from Rabbit Creek. The red sequences represent those collected in June 2003. The numbers in parentheses indicate how many times a sequence was detected.

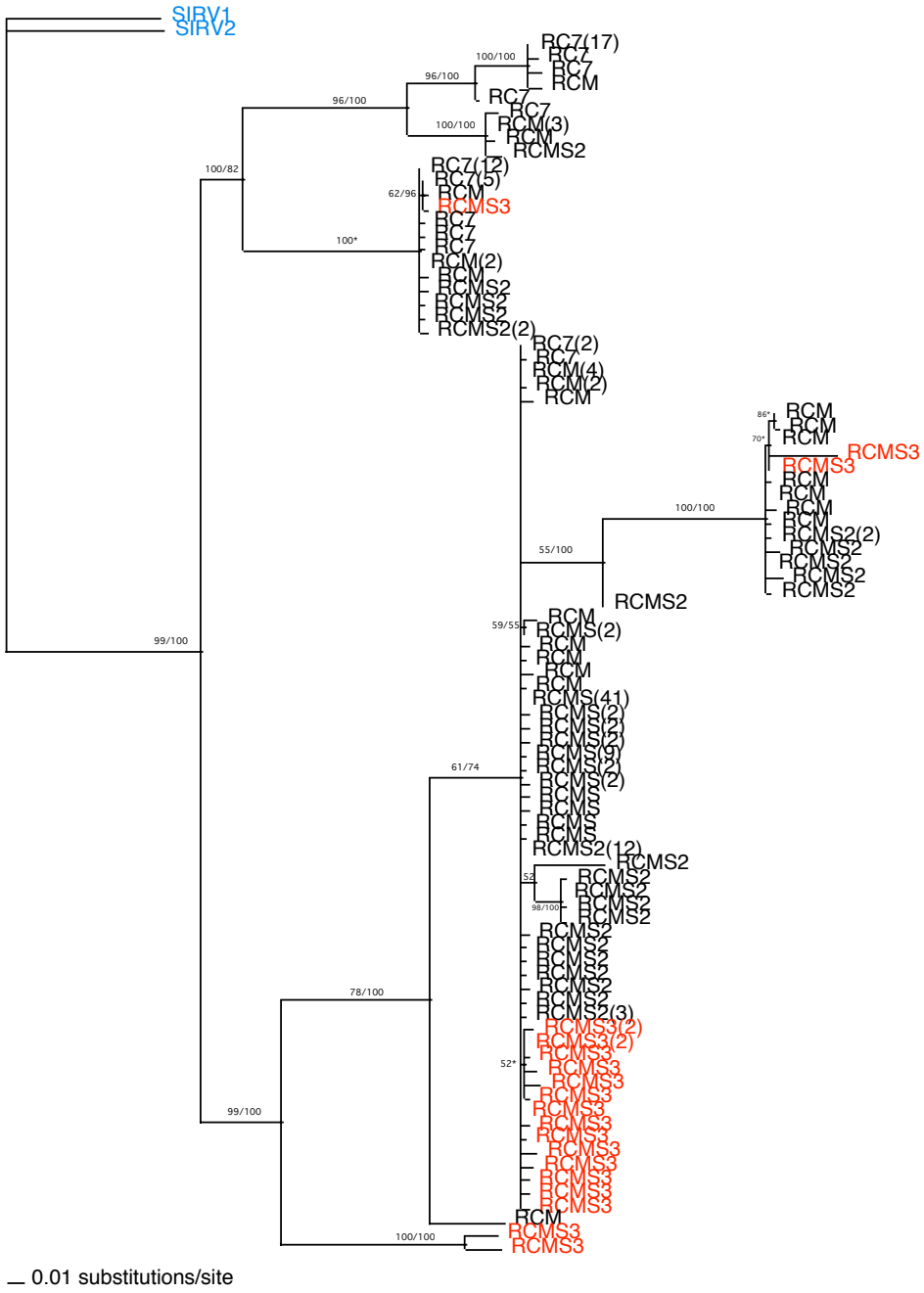


Figure 4.46. SIRV-like protein sequences collected from Rabbit Creek. The red sequences represent those collected in June 2003. The * represents identical protein sequences. The numbers in parentheses indicate how many times a sequence was detected.

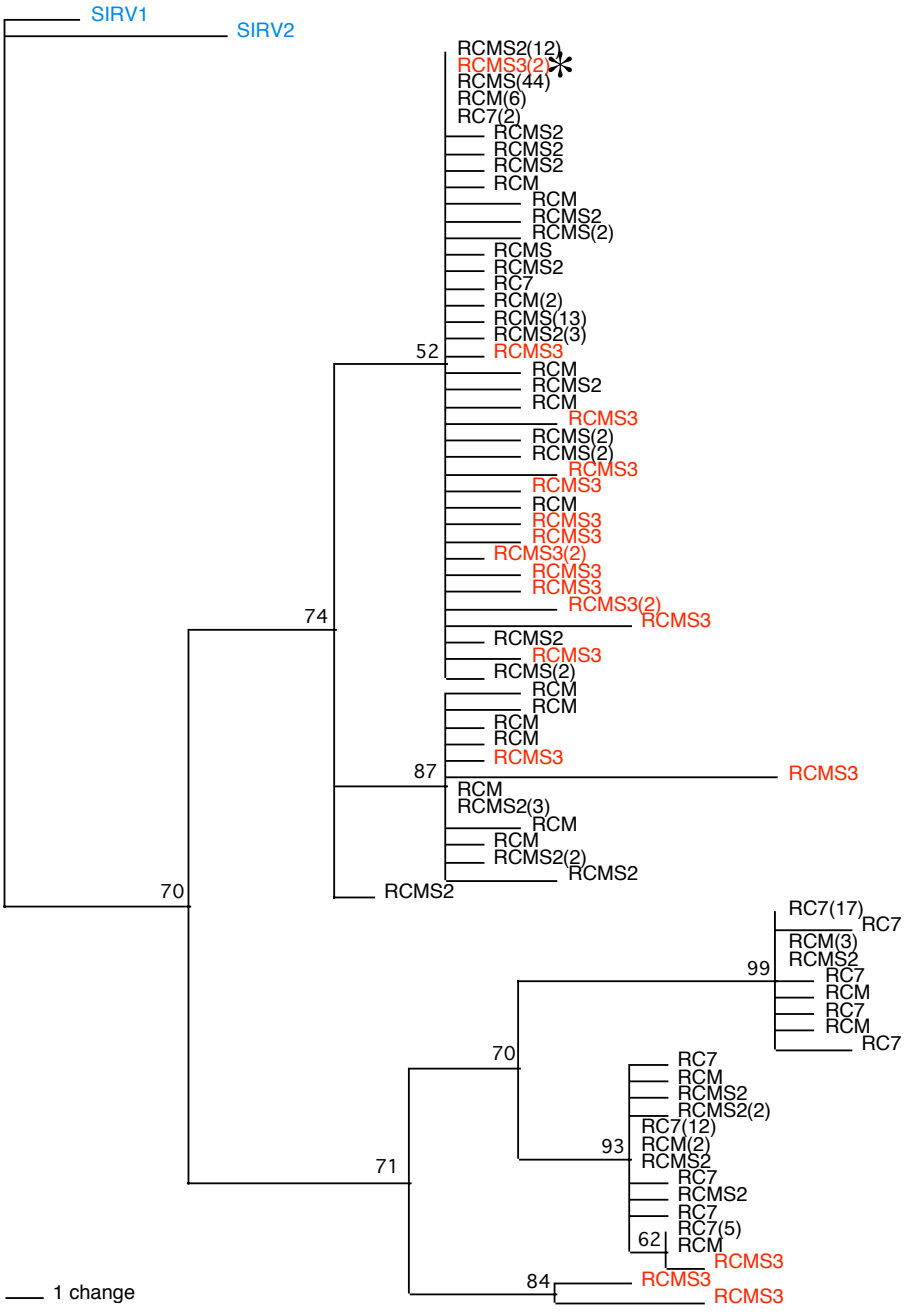


Figure 4.47. SIRV-like DNA sequences collected from Rabbit Creek. The red sequences represent those collected in July 2003. The numbers in parentheses indicate how many times a sequence was detected.

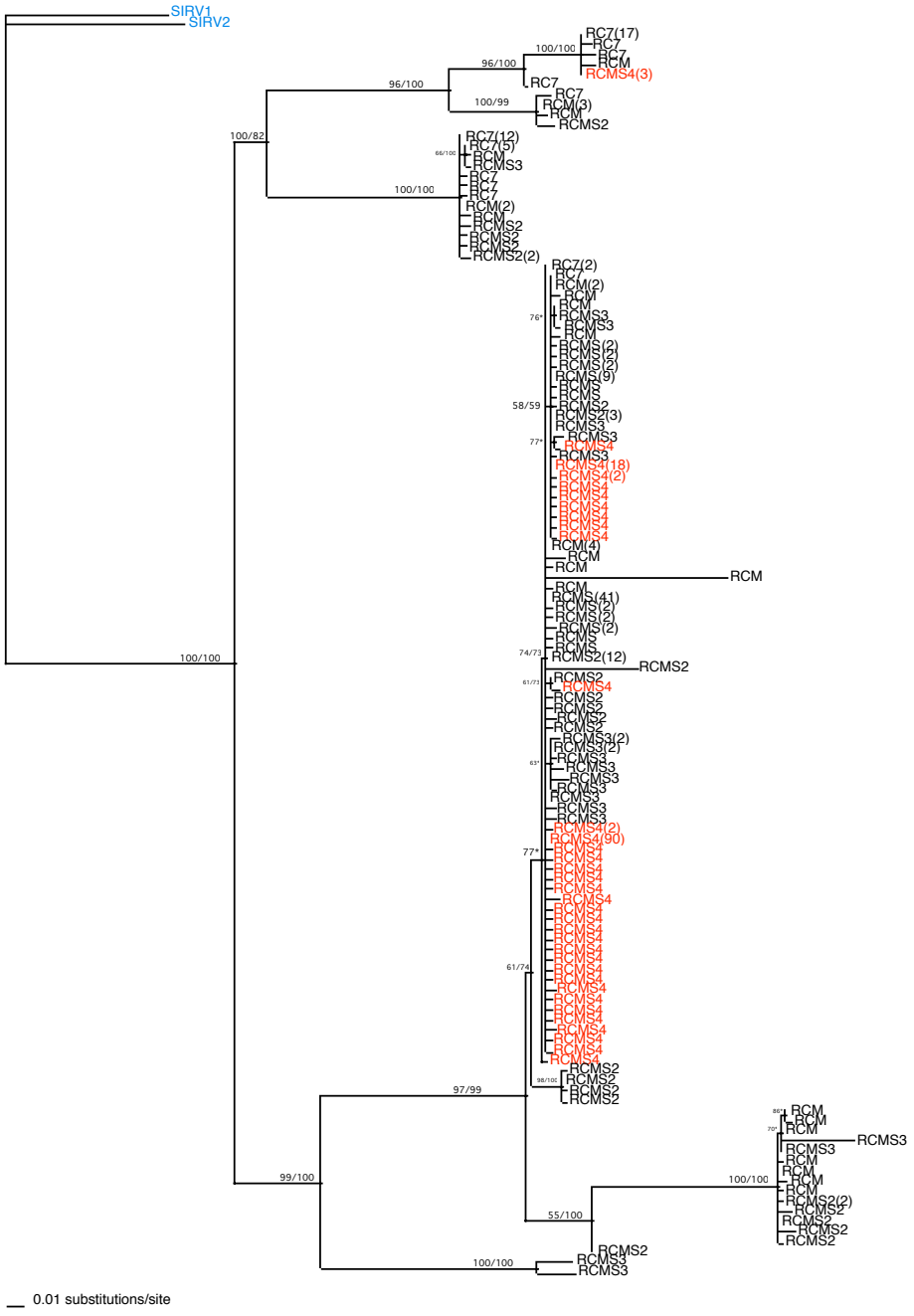


Figure 4.48. SIRV-like protein sequences collected from Rabbit Creek. The red sequences represent those collected in July 2003. The * represents identical protein sequences. The numbers in parentheses indicate how many times a sequence was detected.

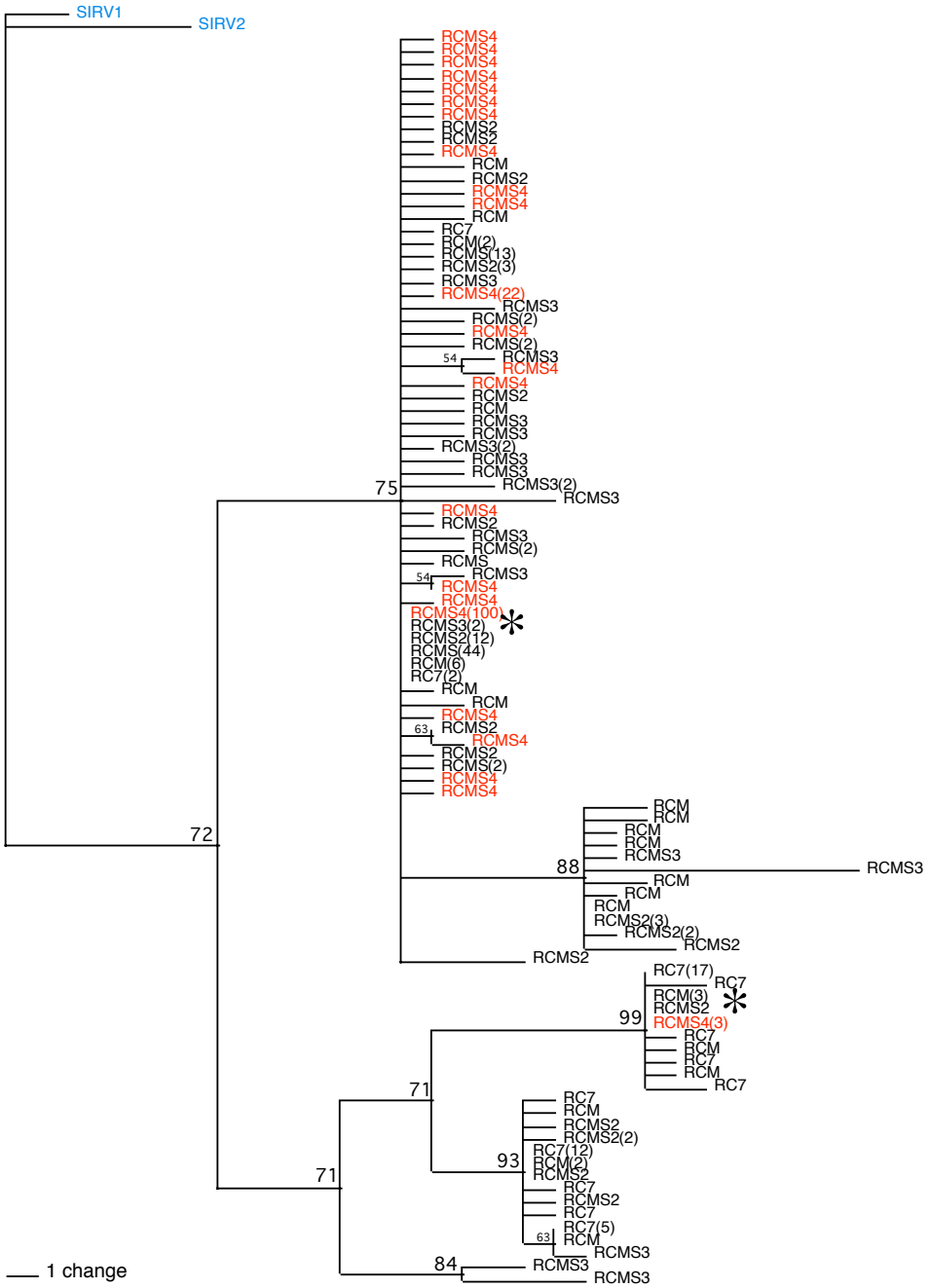


Figure 4.49. SIRV-like DNA sequences collected from Rabbit Creek. The red sequences represent those collected in August 2003. The numbers in parentheses indicate how many times a sequence was detected.

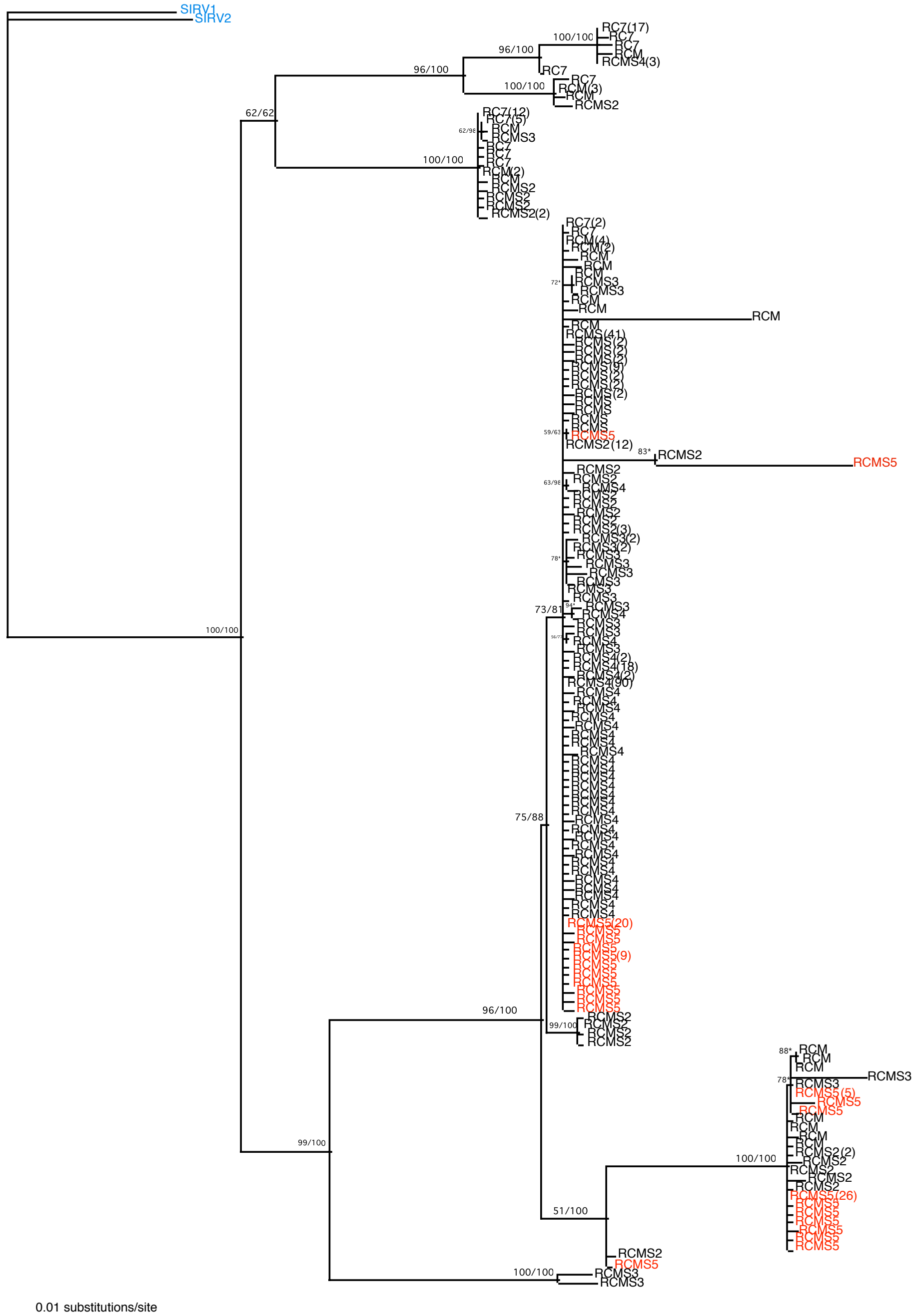


Figure 4.50. SIRV-like protein sequences collected from Rabbit Creek. The red sequences represent those collected in August 2003. The * represents identical protein sequences. The numbers in parentheses indicate how many times a sequence was detected.



Figure 4.51. SIRV-like DNA sequences collected from Rabbit Creek. The red sequences represent those collected in October 2003. The numbers in parentheses indicate how many times a sequence was detected.

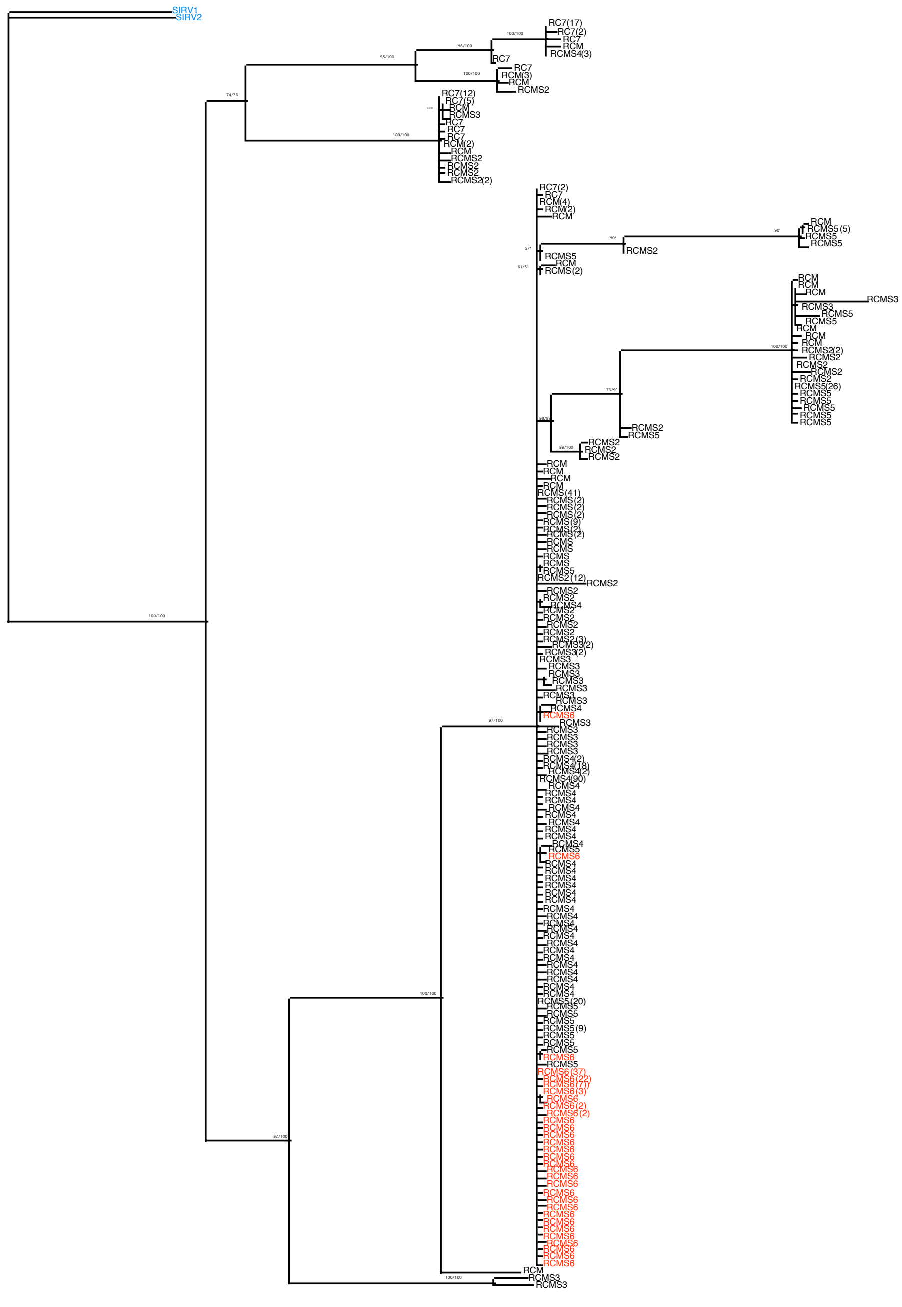


Figure 4.52. SIRV-like protein sequences collected from Rabbit Creek. The red sequences represent those collected in October 2003. The * represents identical protein sequences. The numbers in parentheses indicate how many times a sequence was detected.



Figure 4.53. SIRV-like DNA sequences collected from Rabbit Creek. The red sequences represent those collected in November 2003. The numbers in parentheses indicate how many times a sequence was detected.

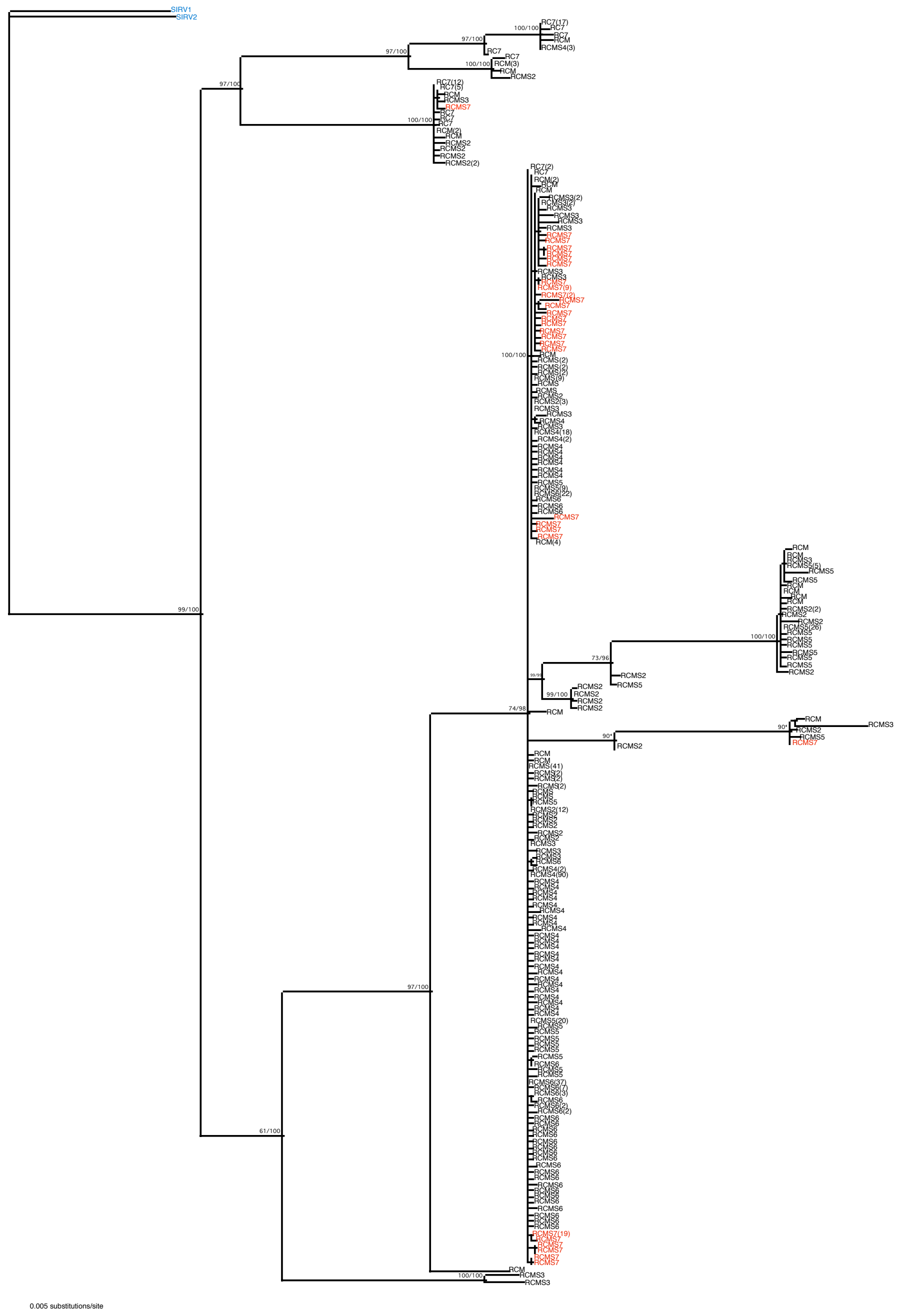


Figure 4.54. SIRV-like protein sequences collected from Rabbit Creek. The red sequences represent those collected in November 2003. The * represents identical protein sequences. The numbers in parentheses indicate how many times a sequence was detected.



Figure 4.55. SIRV-like DNA sequences collected from Rabbit Creek. The red sequences represent those collected in January 2004. The numbers in parentheses indicate how many times a sequence was detected.

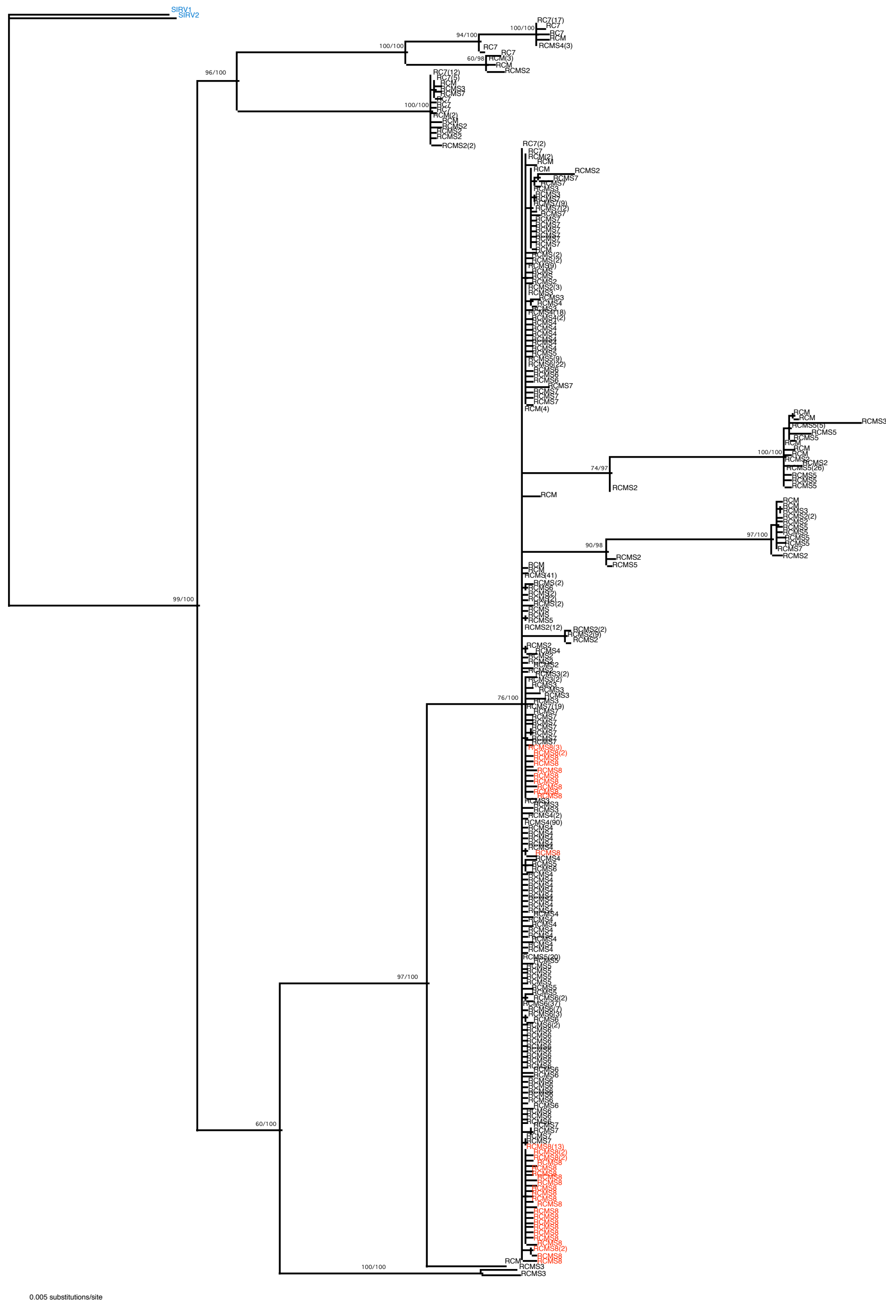


Figure 4.56. SIRV-like protein sequences collected from Rabbit Creek. The red sequences represent those collected in January 2004. The numbers in parentheses indicate how many times a sequence was detected.

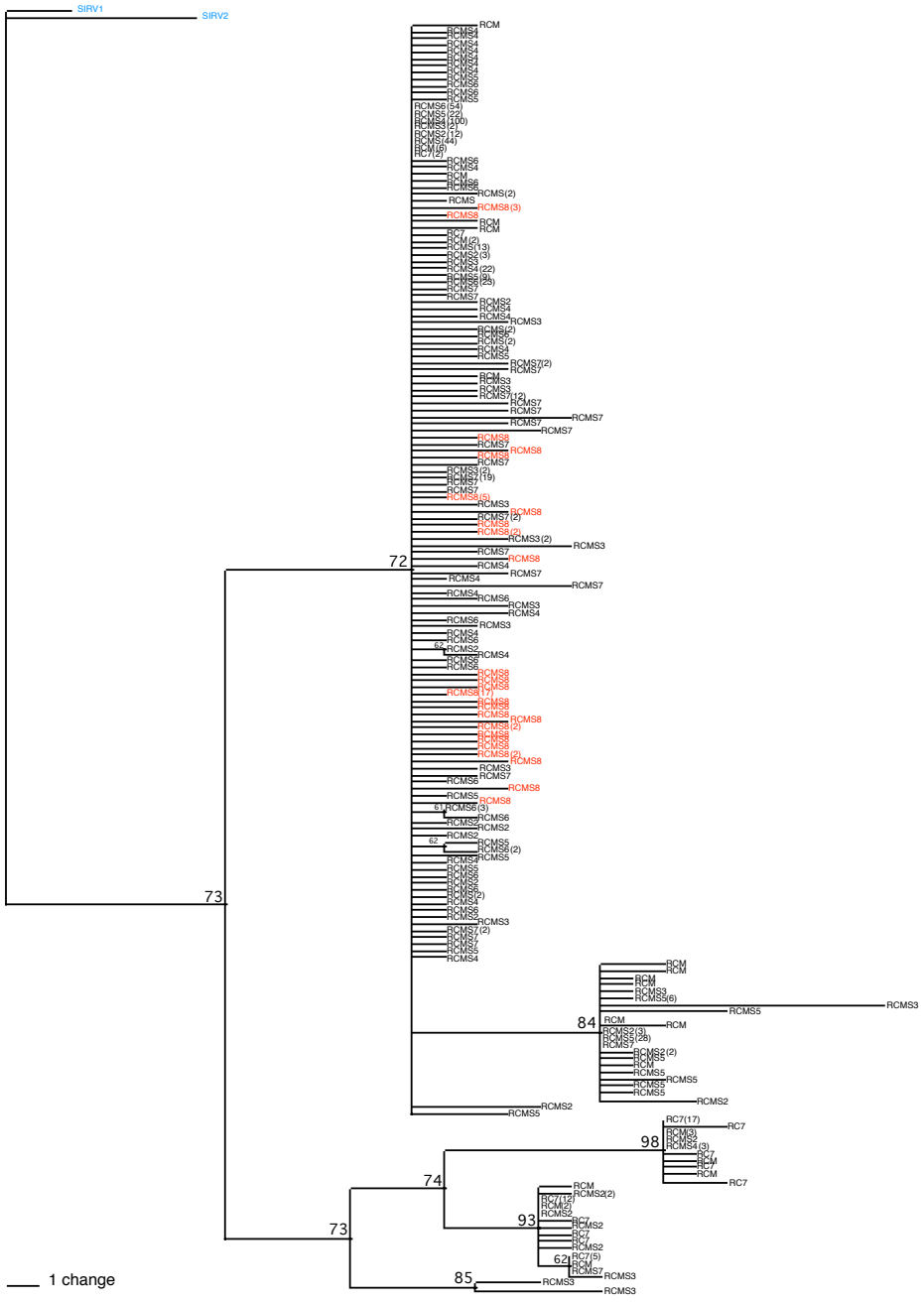


Figure 4.57. SIRV-like DNA sequences collected from Rabbit Creek. The red sequences represent those collected in March 2004. The numbers in parentheses indicate how many times a sequence was detected.

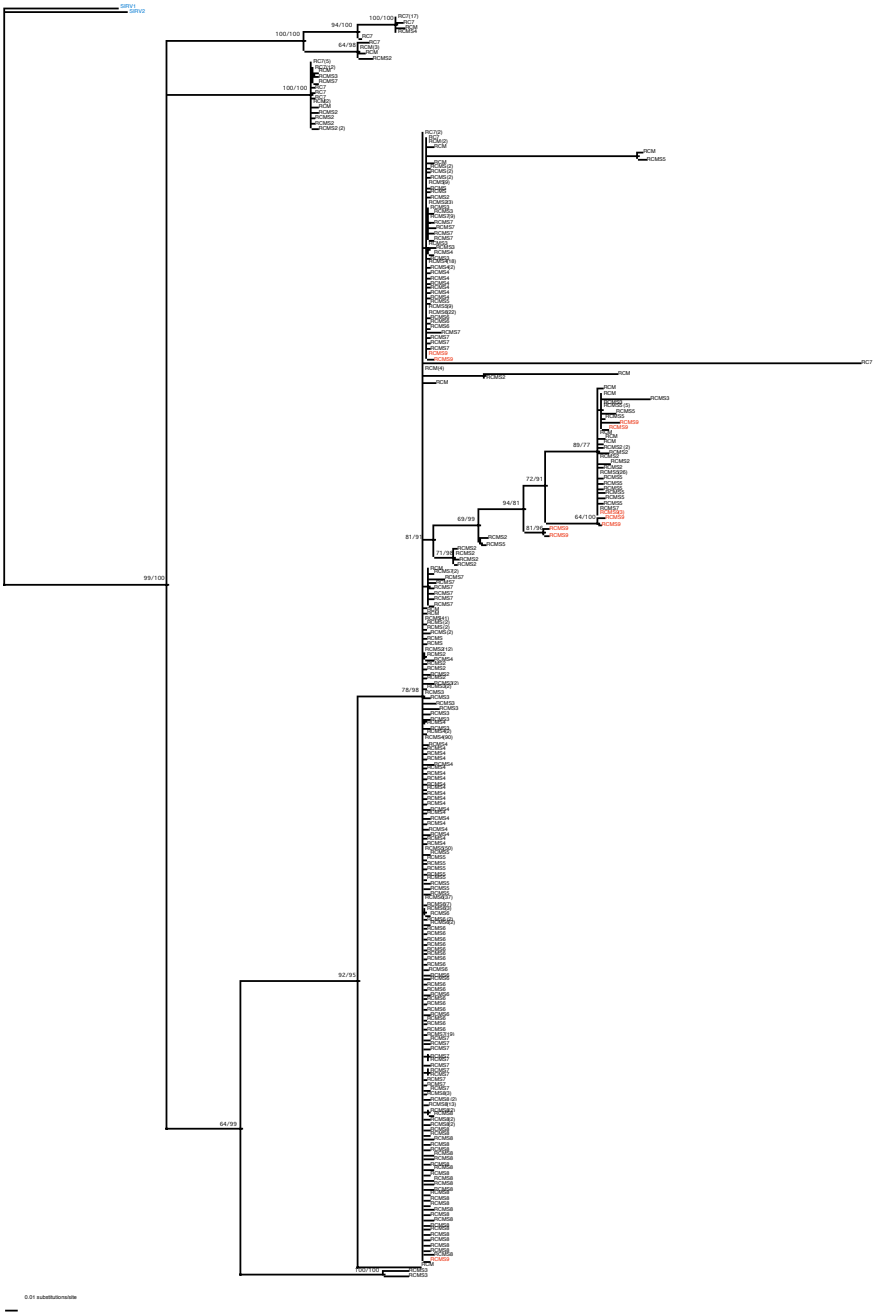


Figure 4.58. SIRV-like protein sequences collected from Rabbit Creek. The red sequences represent those collected in March 2004. The * represents identical protein sequences. The numbers in parentheses indicate how many times a sequence was detected.

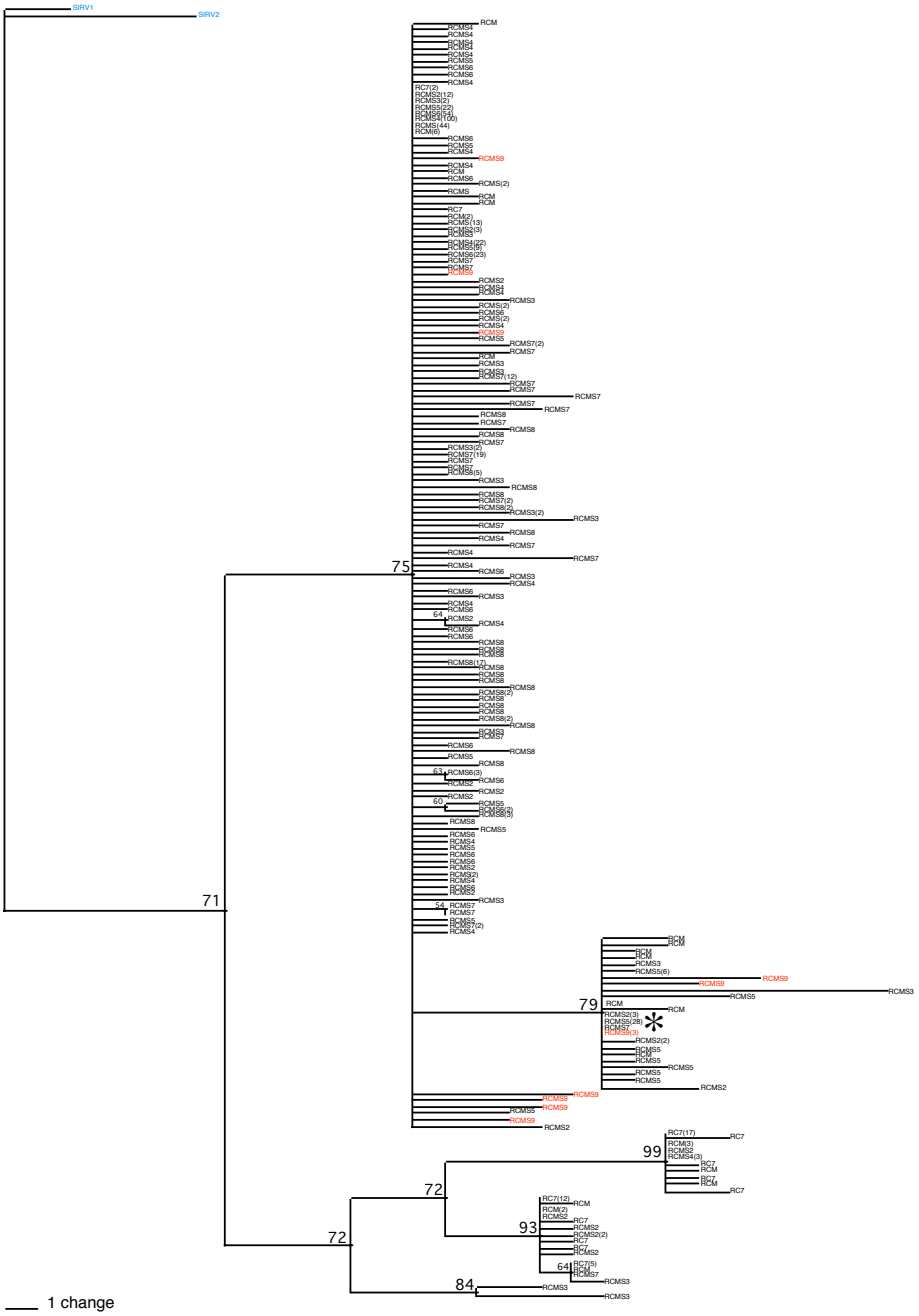


Figure 4.59. SIRV-like coat protein DNA sequences collected from Rabbit Creek. The red sequences represent those collected in June 2004. The numbers in parentheses indicate how many times a sequence was detected.

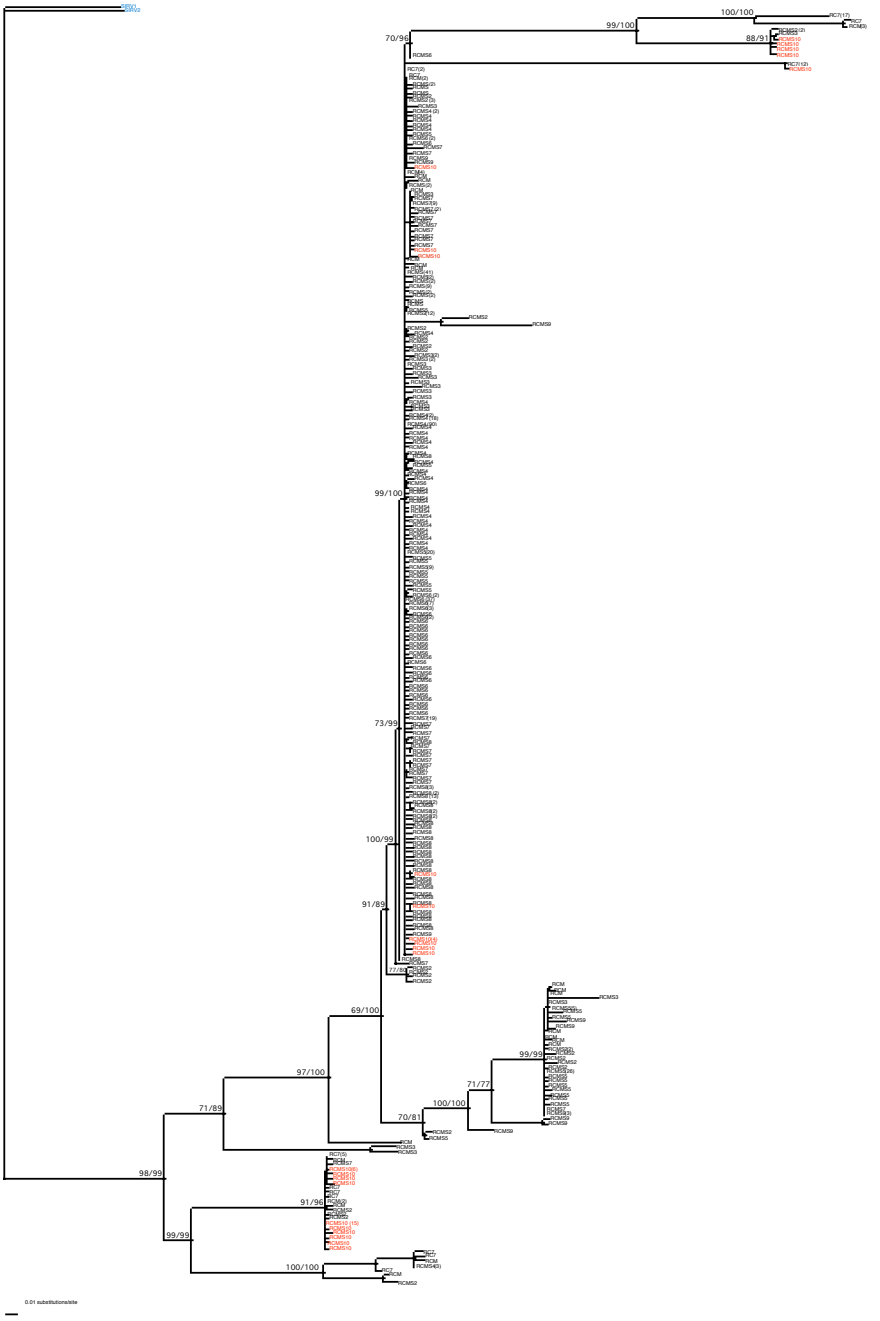
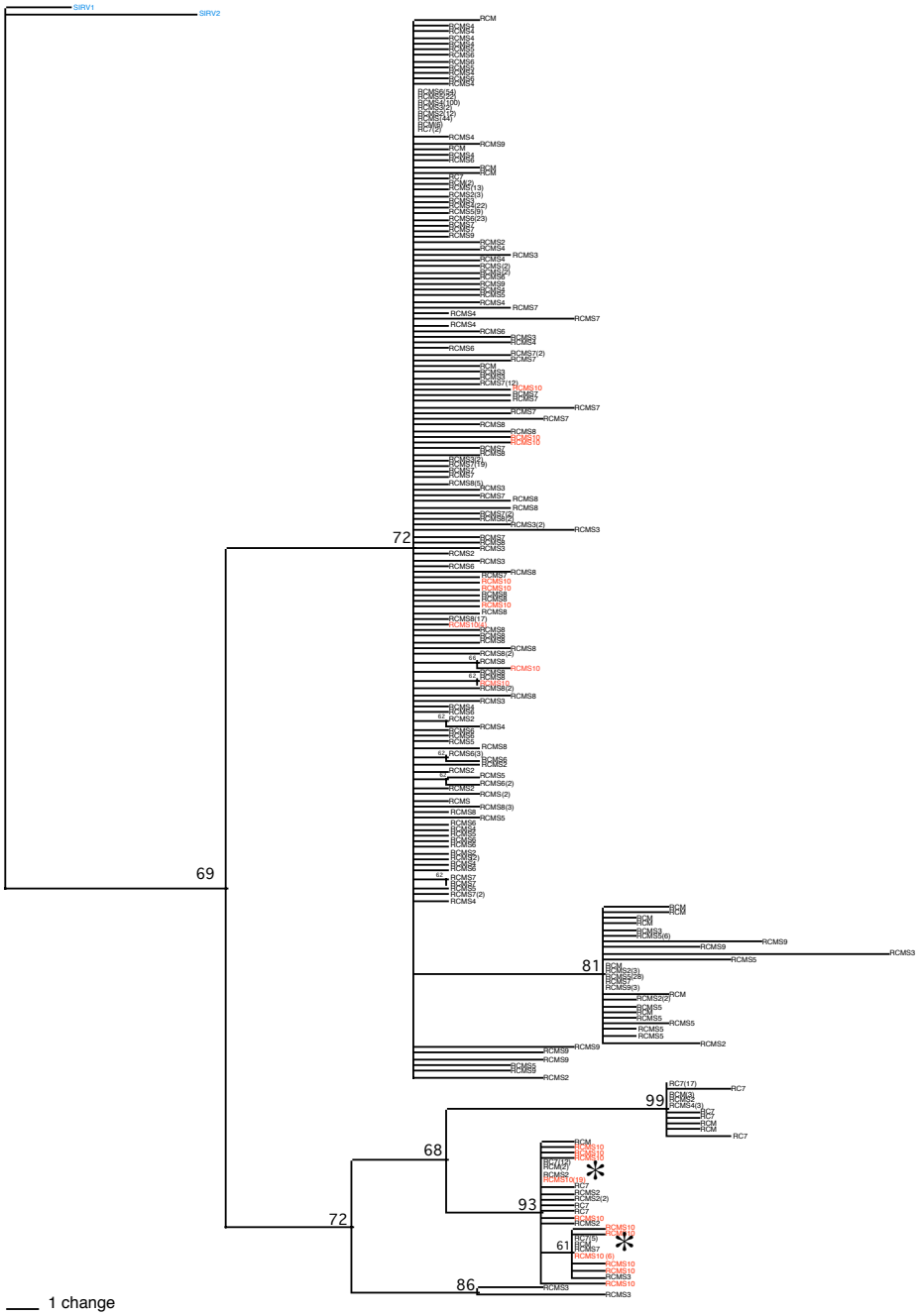


Figure 4.60. SIRV-like protein sequences collected from Rabbit Creek. The red sequences represent those collected in June 2004. The * represents identical protein sequences. The numbers in parentheses indicate how many times a sequence was detected.



APPENDIX C

SUPPLEMENTARY PHYLOGENTIC ANALYSES FROM THE RAGGED HILLS MONITOR SITE

Figure 5.25. SSV-like 3F/4R sequences from Ragged Hills. The red sequences represent those collected in November 2002. The numbers in parentheses indicate how many times a sequence was detected. A. DNA sequences; the • represents non-functional ORFs. B. Inferred protein sequences; the * represents identical sequences.

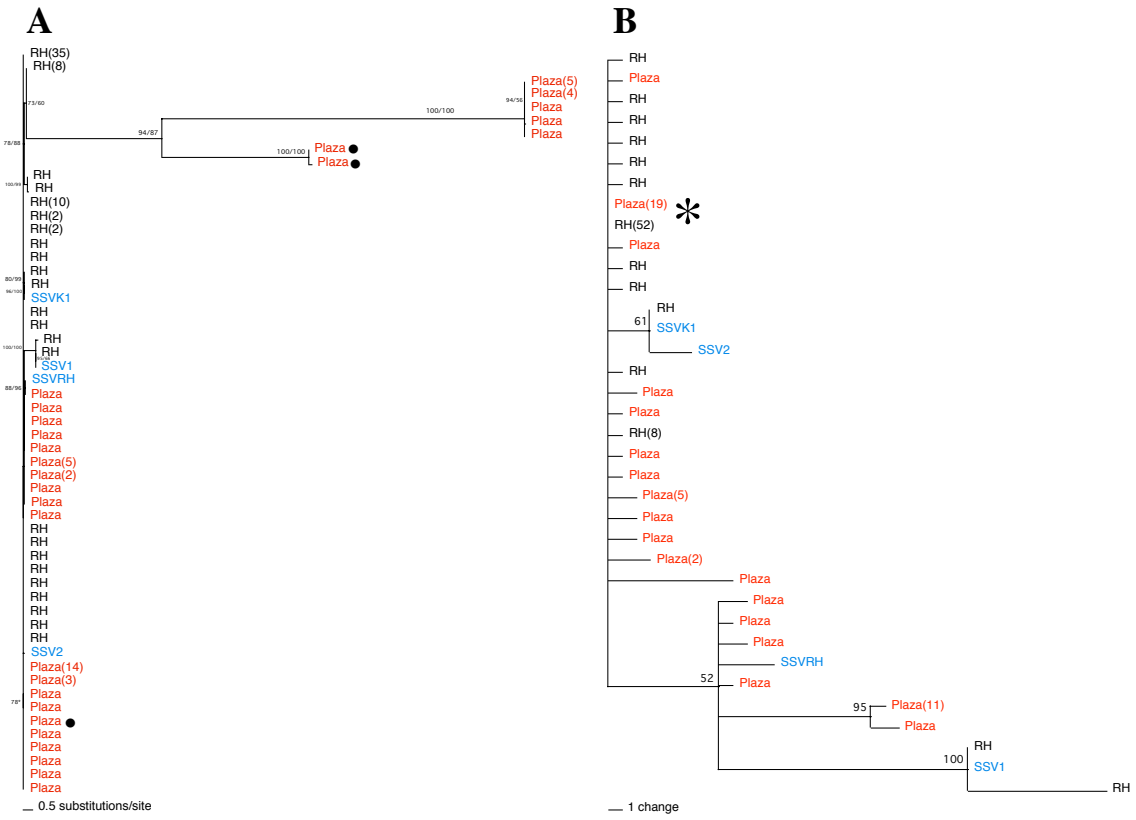


Figure 5.26. SSV-like 7F/8R sequences from Ragged Hills. The red sequences represent those collected in November 2002. The numbers in parentheses indicate how many times a sequence was detected. A. DNA sequences. B. Inferred protein sequences.

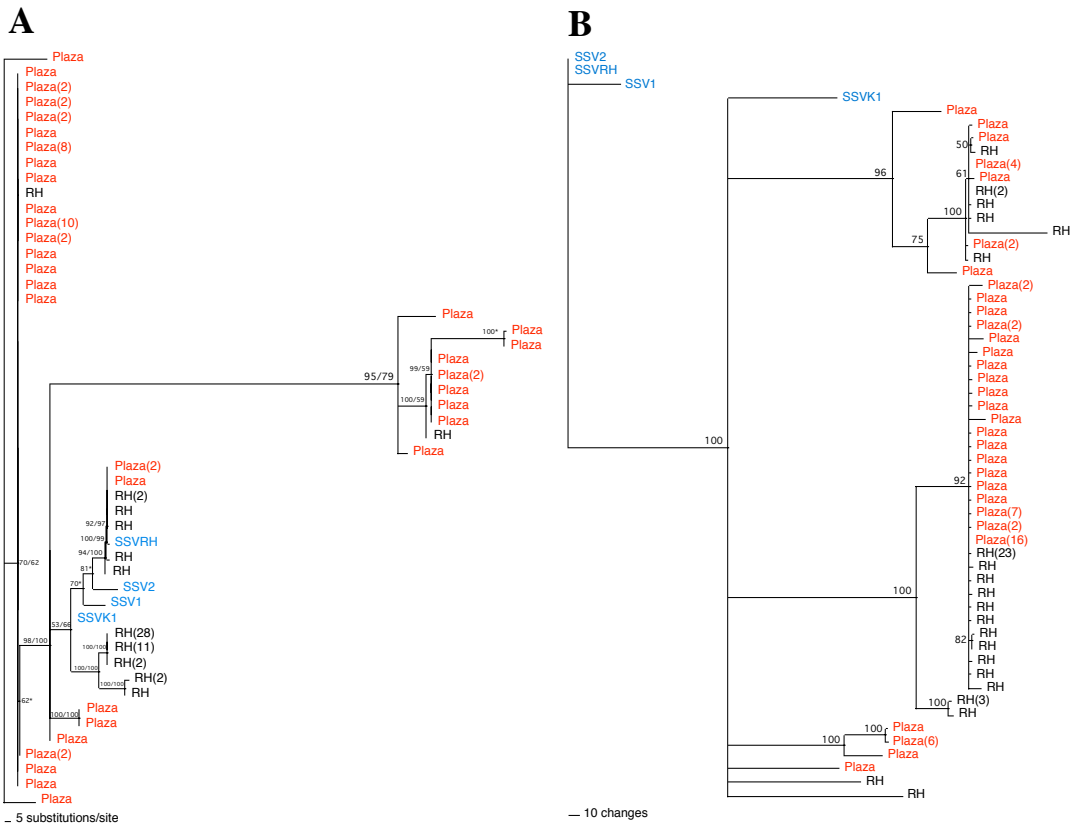


Figure 5.27. SSV-like 3F/4R DNA sequences from Ragged Hills. The red sequences represent those collected in February 2003. The numbers in parentheses indicate how many times a sequence was detected.

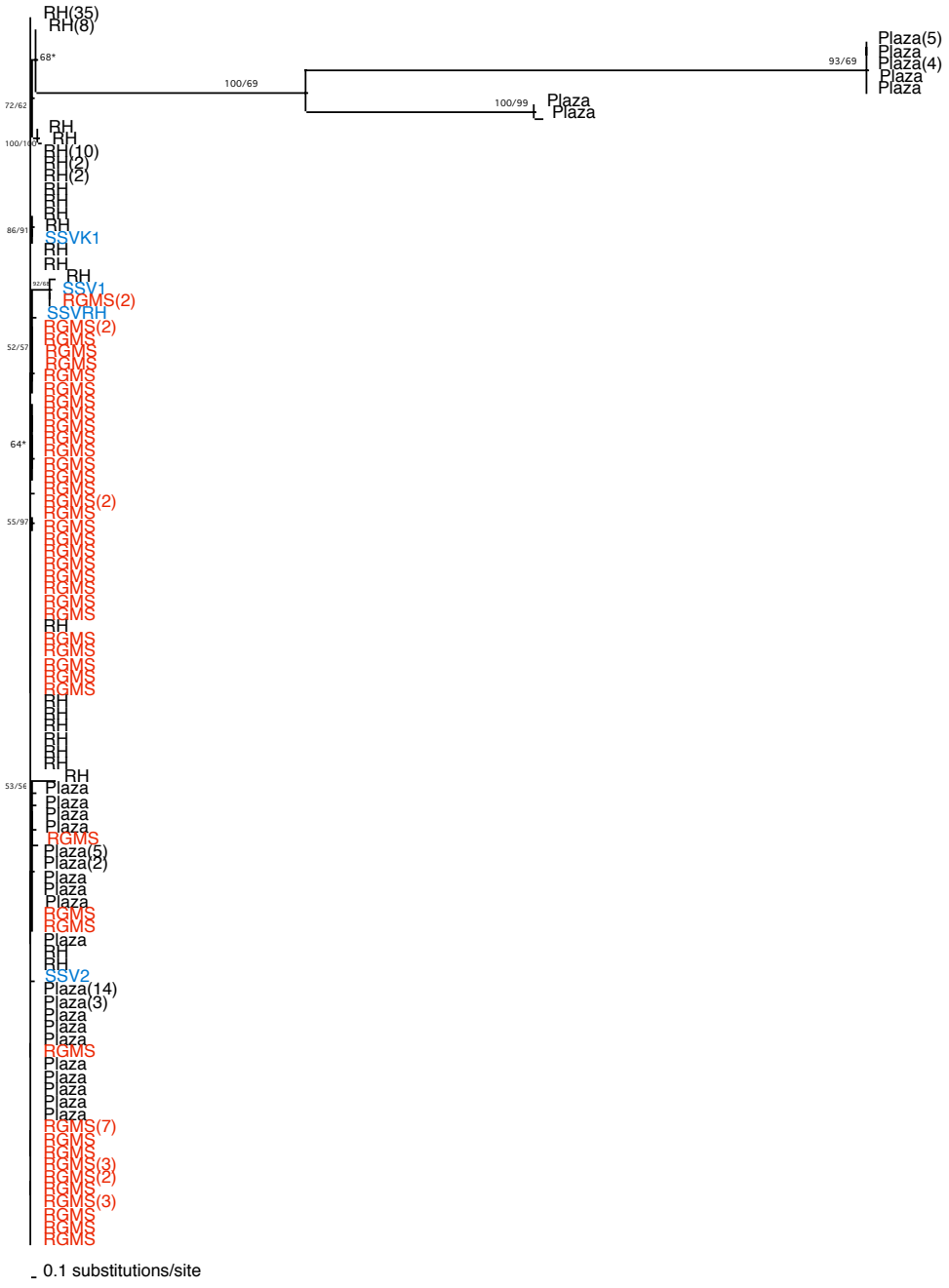


Figure 5.28. SSV-like 3F/4R protein sequences from Ragged Hills. The red sequences represent those collected in February 2003. The * represents identical sequences. The numbers in parentheses indicate how many times a sequence was detected.

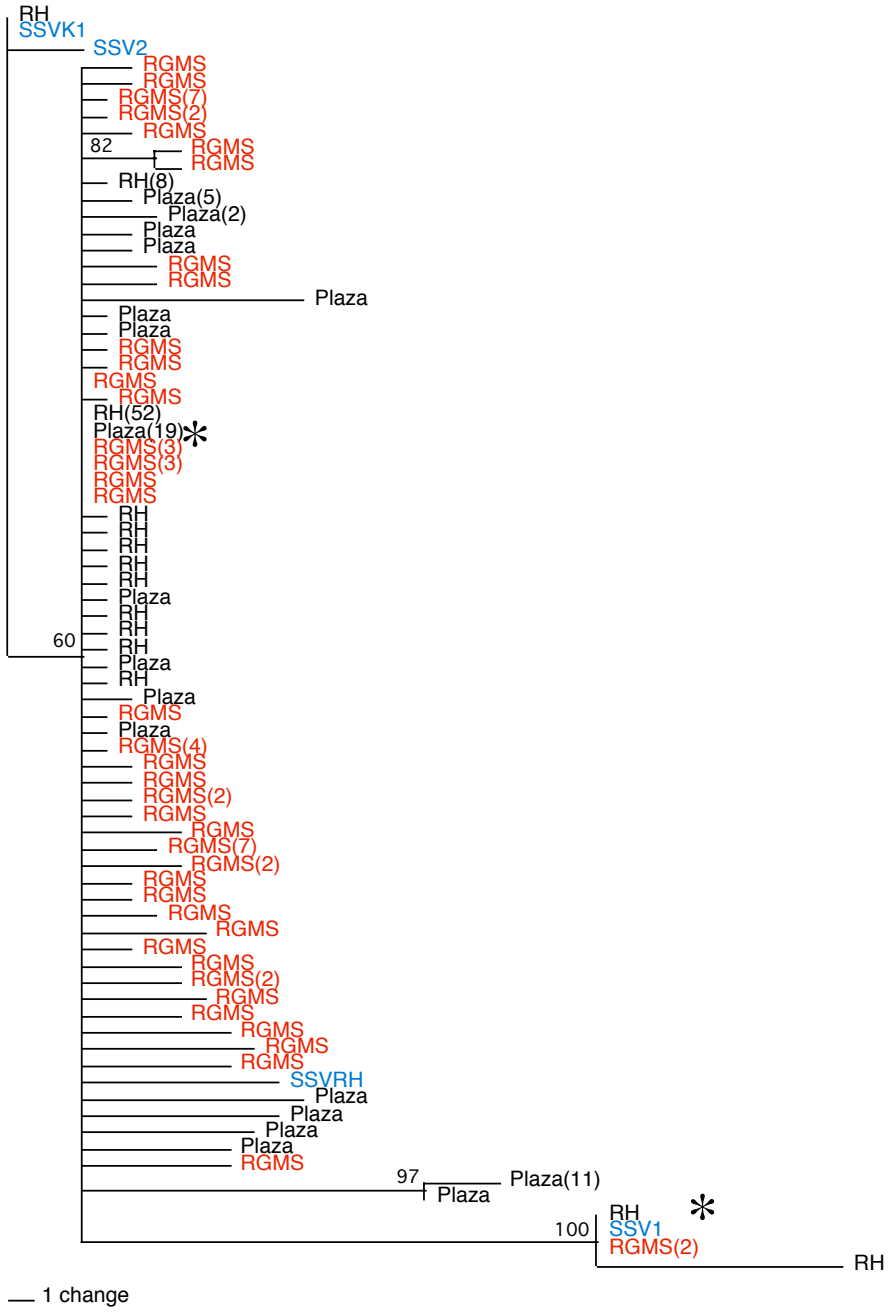


Figure 5.29. SSV-like 3F/4R DNA sequences from Ragged Hills. The red sequences represent those collected in April 2003. The numbers in parentheses indicate how many times a sequence was detected. The • represents a non-functional ORF.

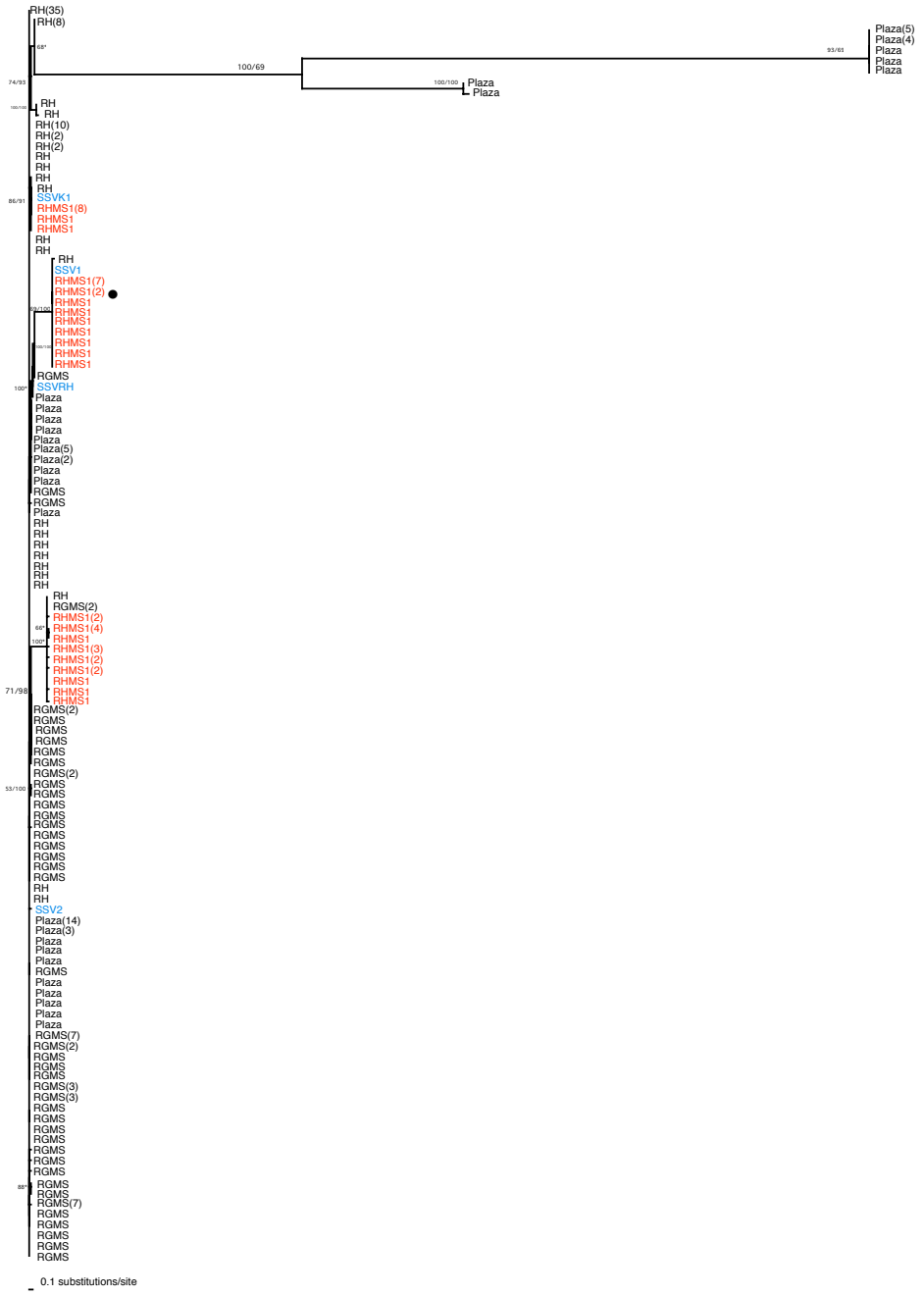


Figure 5.30. SSV-like 3F/4R protein sequences from Ragged Hills. The red sequences represent those collected in April 2003. The * represents identical sequences. The numbers in parentheses indicate how many times a sequence was detected.

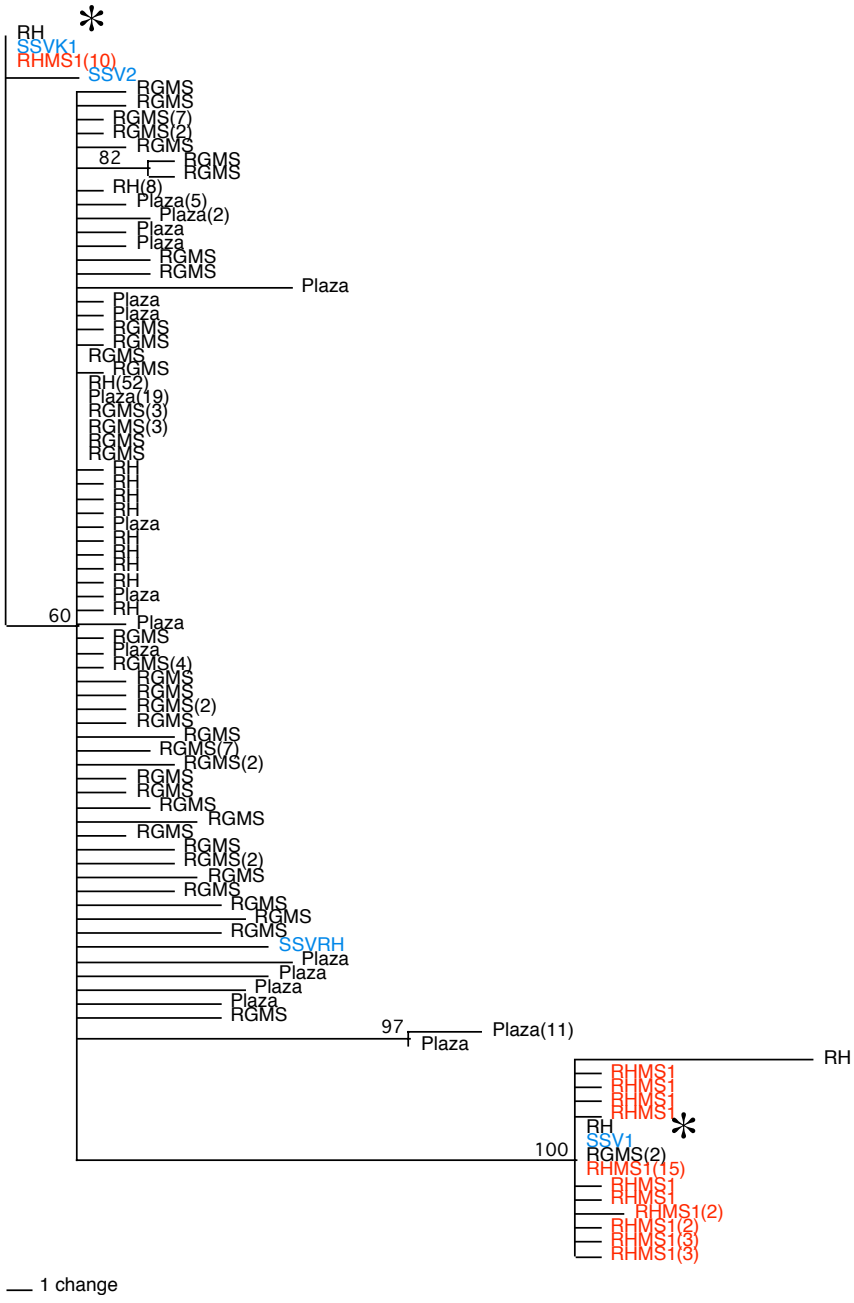


Figure 5.31. SSV-like 3F/4R DNA sequences from Ragged Hills. The red sequences represent those collected on May 6, 2003. The numbers in parentheses indicate how many times a sequence was detected. The • represents a non-functional ORF.

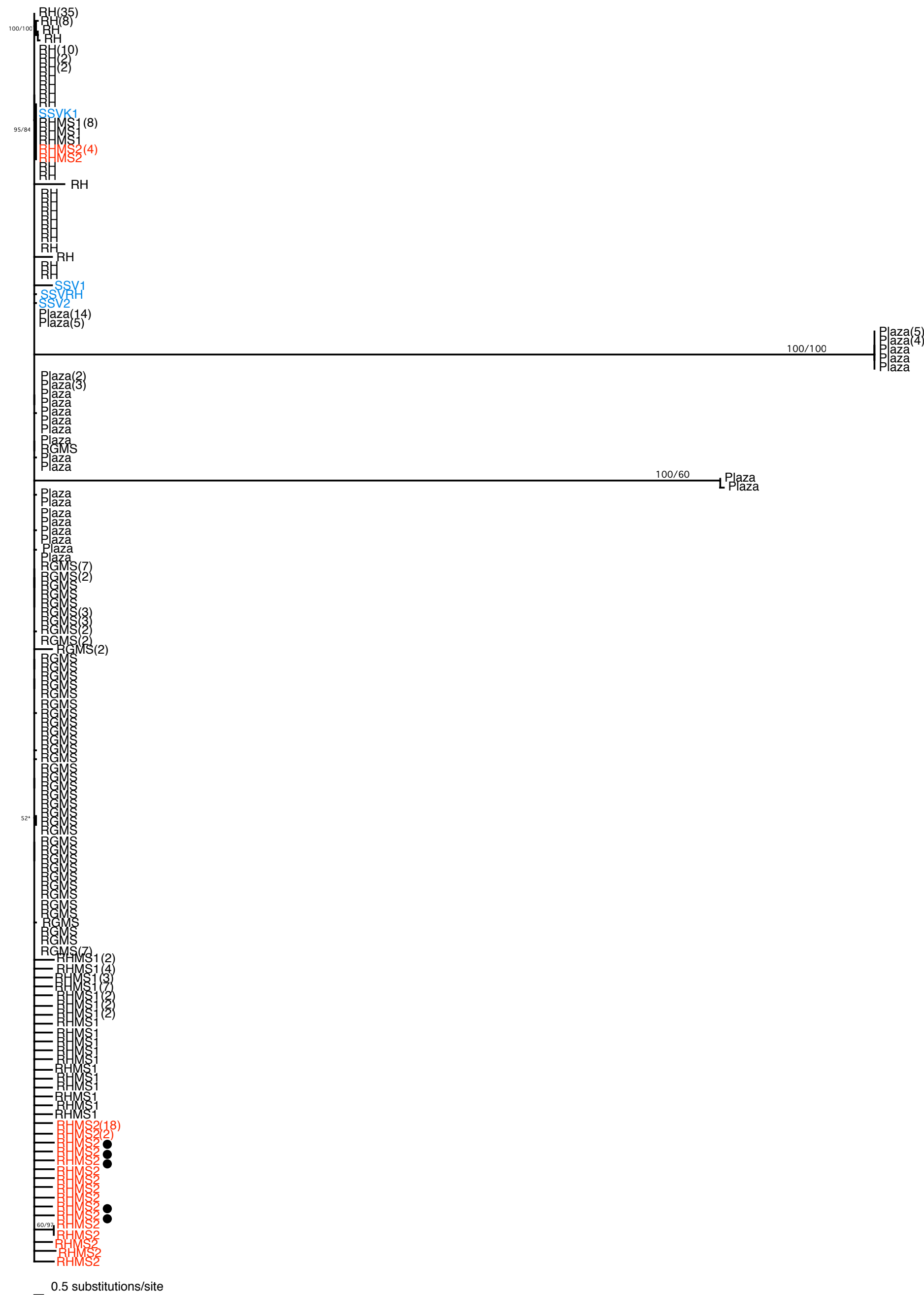


Figure 5.32. SSV-like 3F/4R protein sequences from Ragged Hills. The red sequences represent those collected on May 6, 2003. The * represents identical sequences. The numbers in parentheses indicate how many times a sequence was detected.

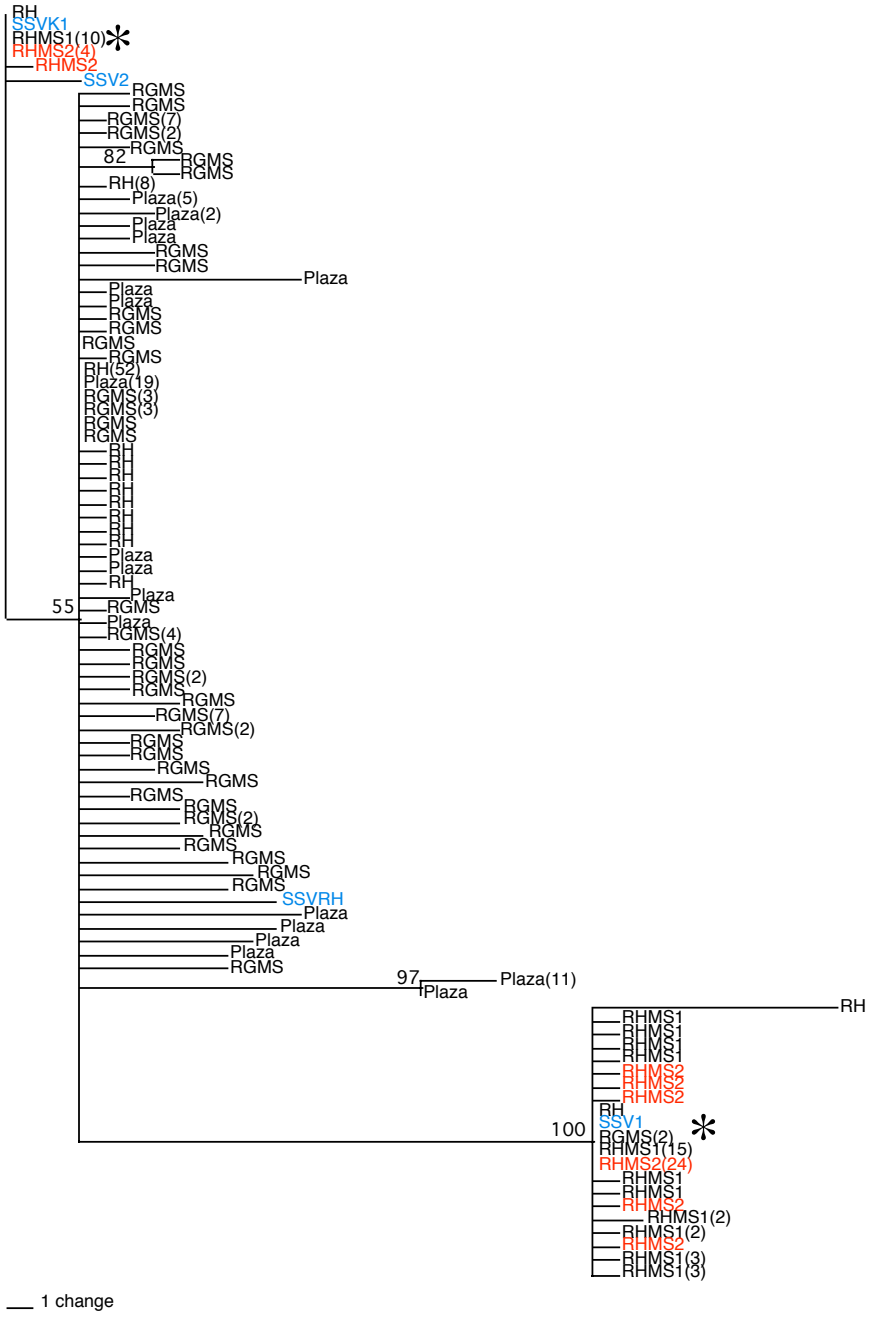


Figure 5.33. SSV-like 3F/4R DNA sequences from Ragged Hills. The red sequences represent those collected in November 2003. The numbers in parentheses indicate how many times a sequence was detected. The • represents a non-functional ORF.

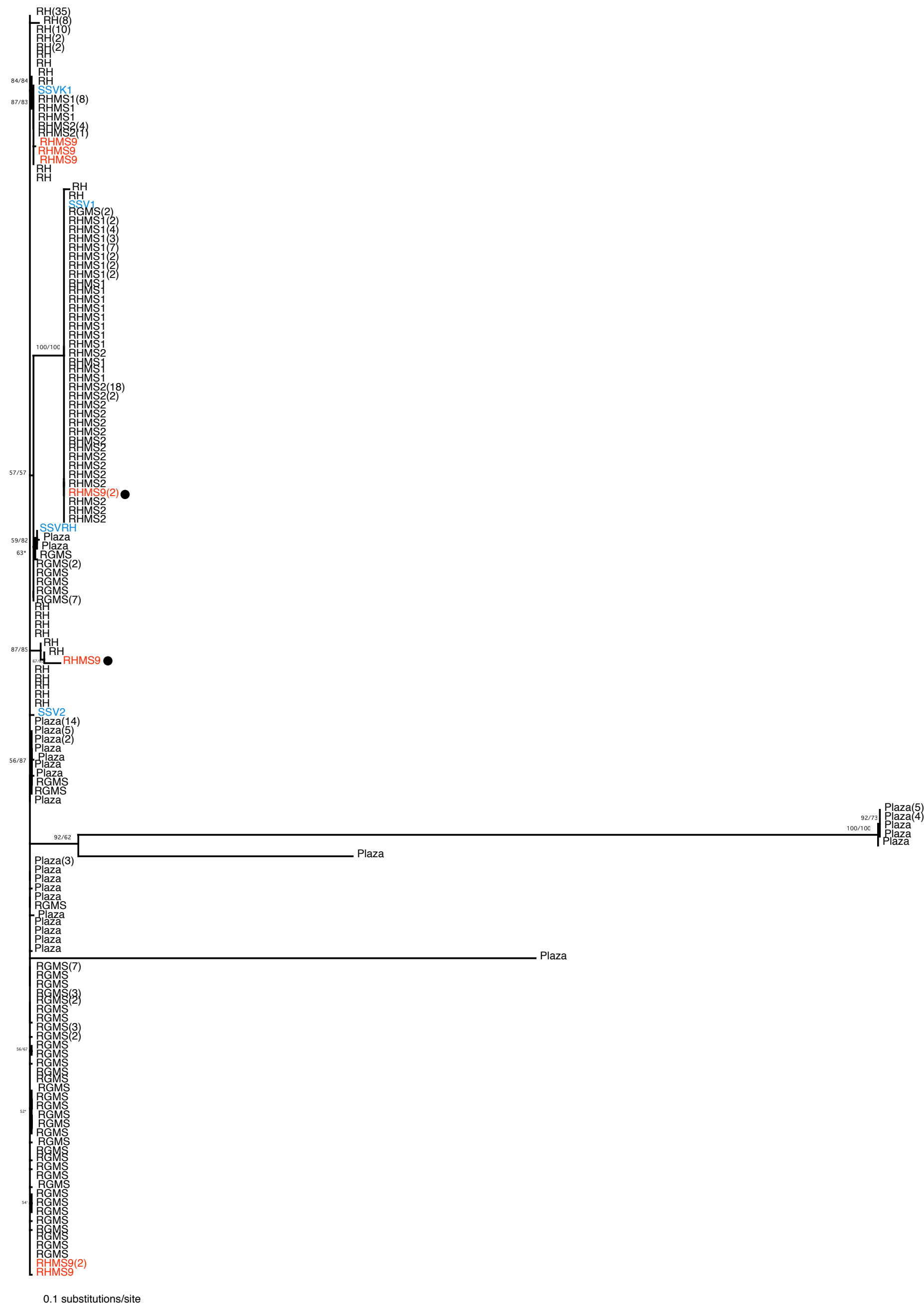


Figure 5.34. SSV-like 3F/4R protein sequences from Ragged Hills. The red sequences represent those collected in November 2003. The * represents identical sequences. The numbers in parentheses indicate how many times a sequence was detected.

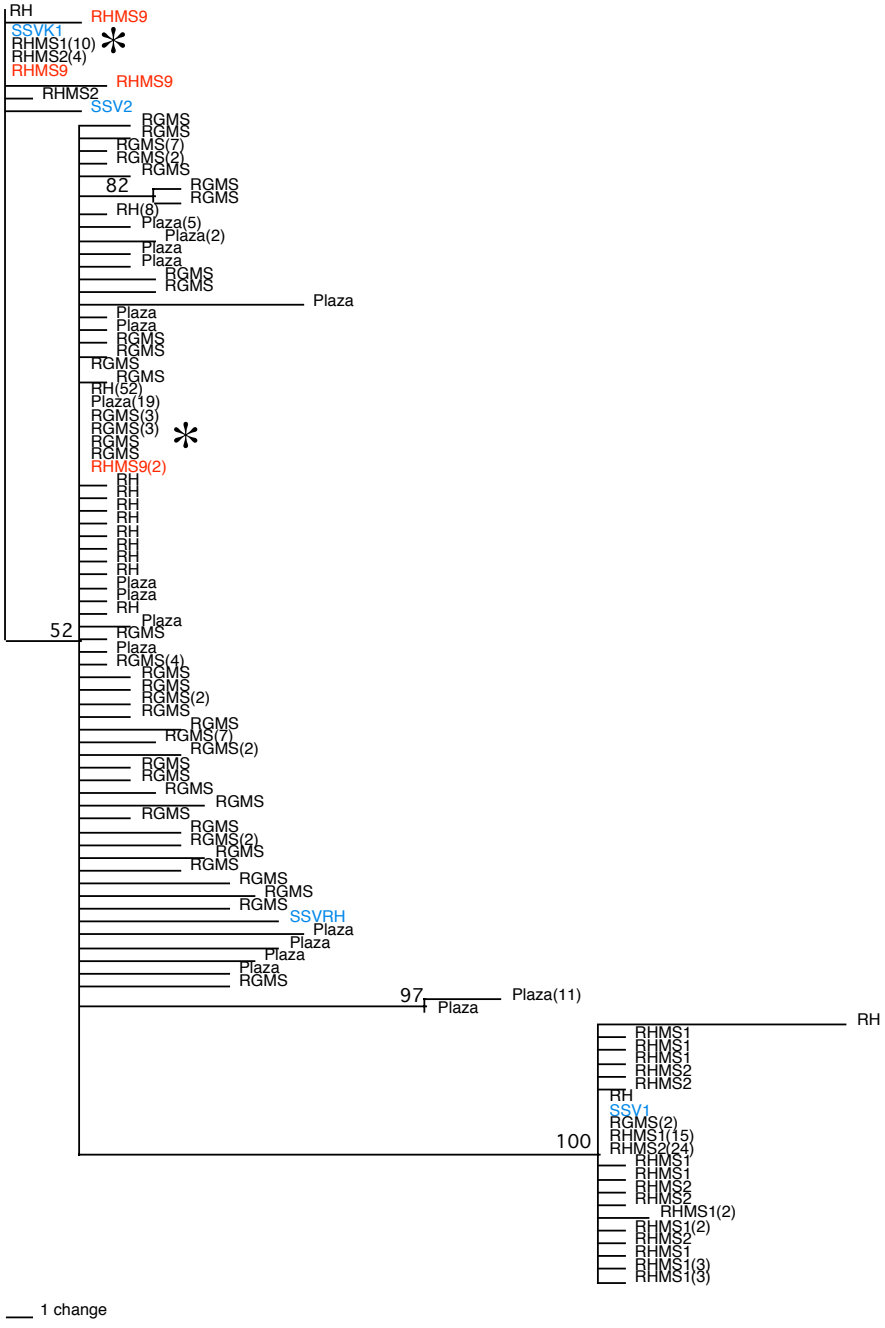


Figure 5.35. SSV-like 3F/4R DNA sequences from Ragged Hills. The different colors represent different groups of viruses detected over time. This analysis was conducted with both functional and non-functional ORFs. The numbers in parentheses indicate how many times a sequence was detected.

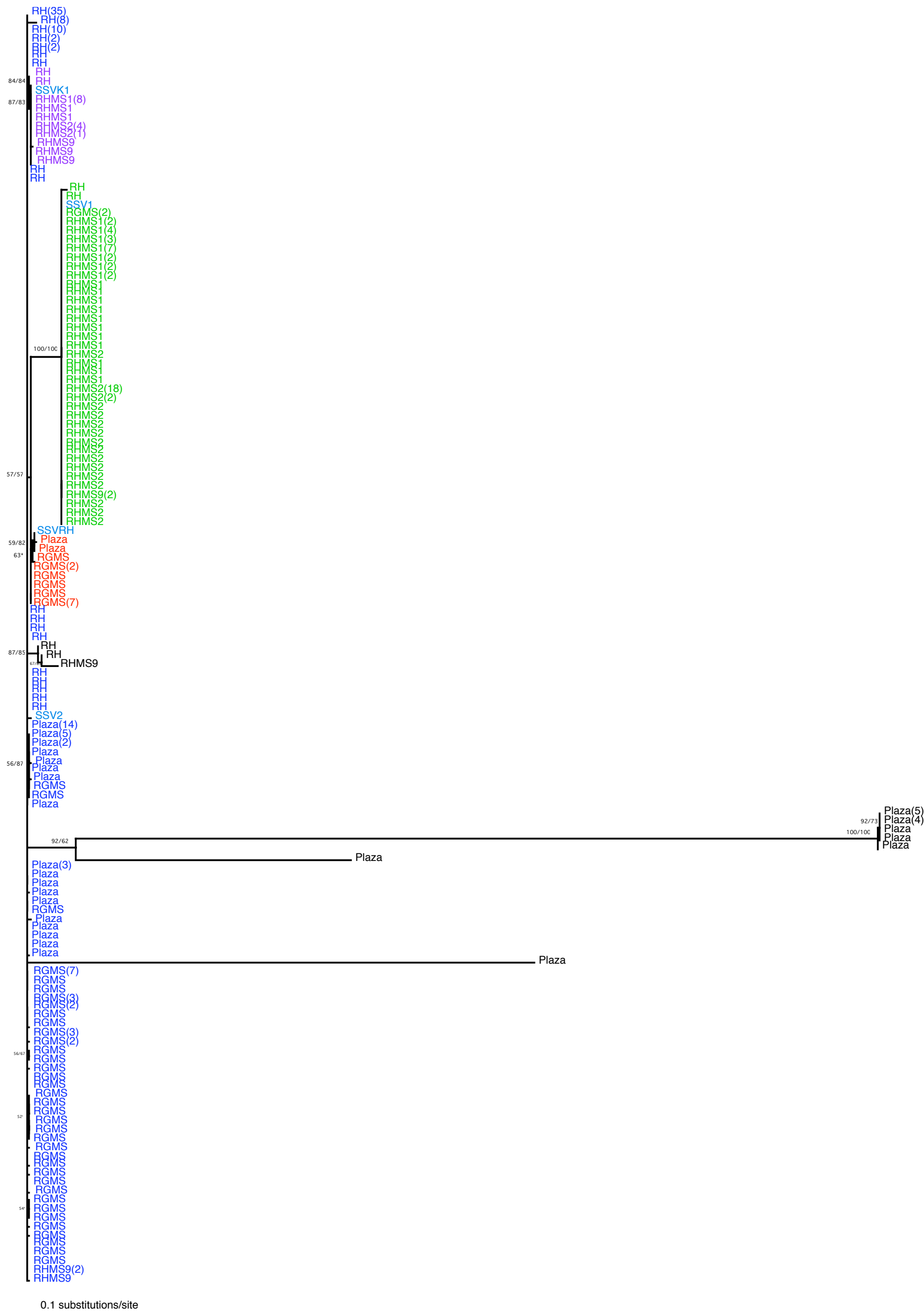


Figure 5.36. SIRV-like DNA sequences from Ragged Hills. The red represents sequences collected in November 2002. A. DNA sequences. B. Inferred protein sequences; the * represents identical protein sequences. The numbers in parentheses indicate how many times a sequence was detected.

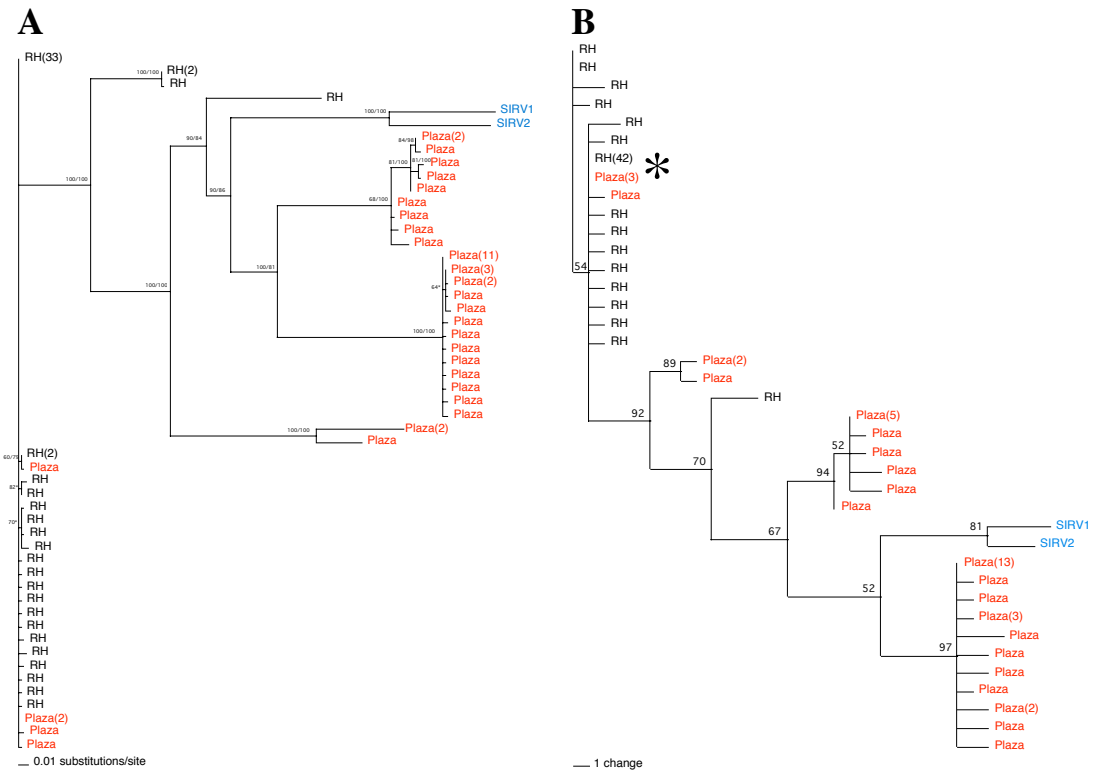


Figure 5.37. SIRV-like DNA sequences from Ragged Hills. The red represents the sequences collected in February 2003. The numbers in parentheses indicate how many times a sequence was detected.

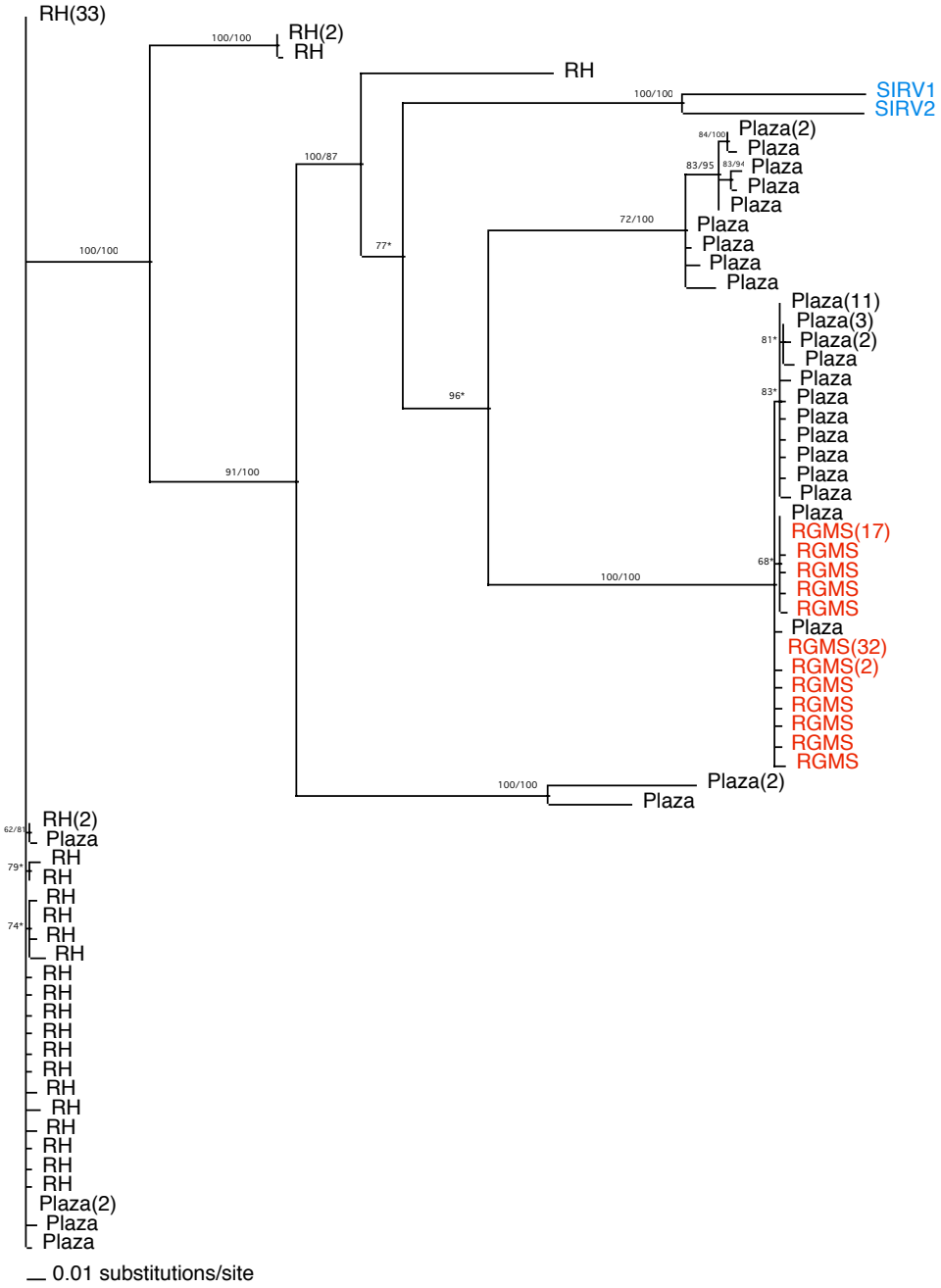


Figure 5.38. SIRV-like protein sequences from Ragged Hills. The red sequences represent those collected in February 2003. The numbers in parentheses indicate how many times a sequence was detected.

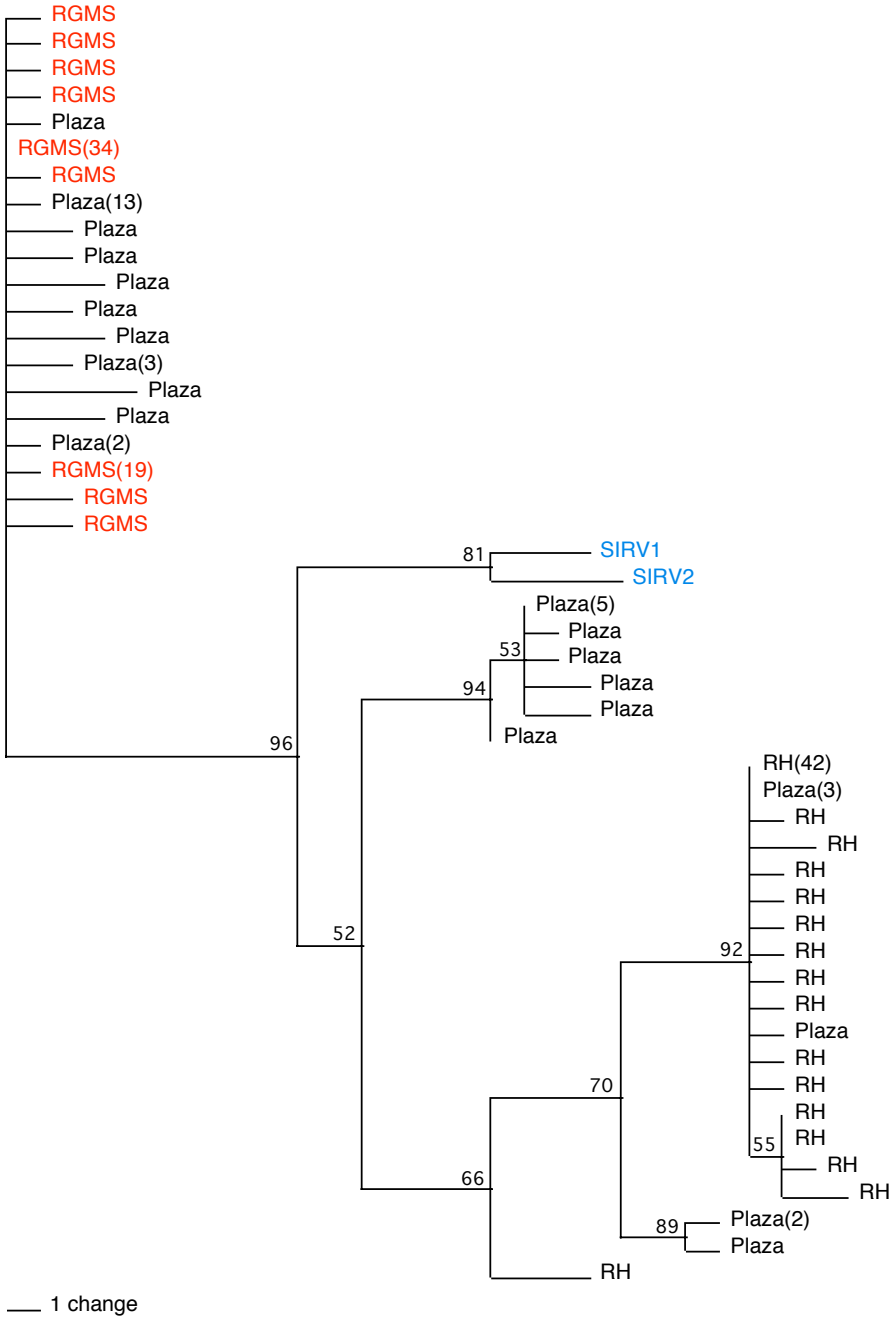


Figure 5.39. SIRV-like DNA sequences from Ragged Hills. The red sequences represent those collected in April 2003. The numbers in parentheses indicate how many times a sequence was detected.

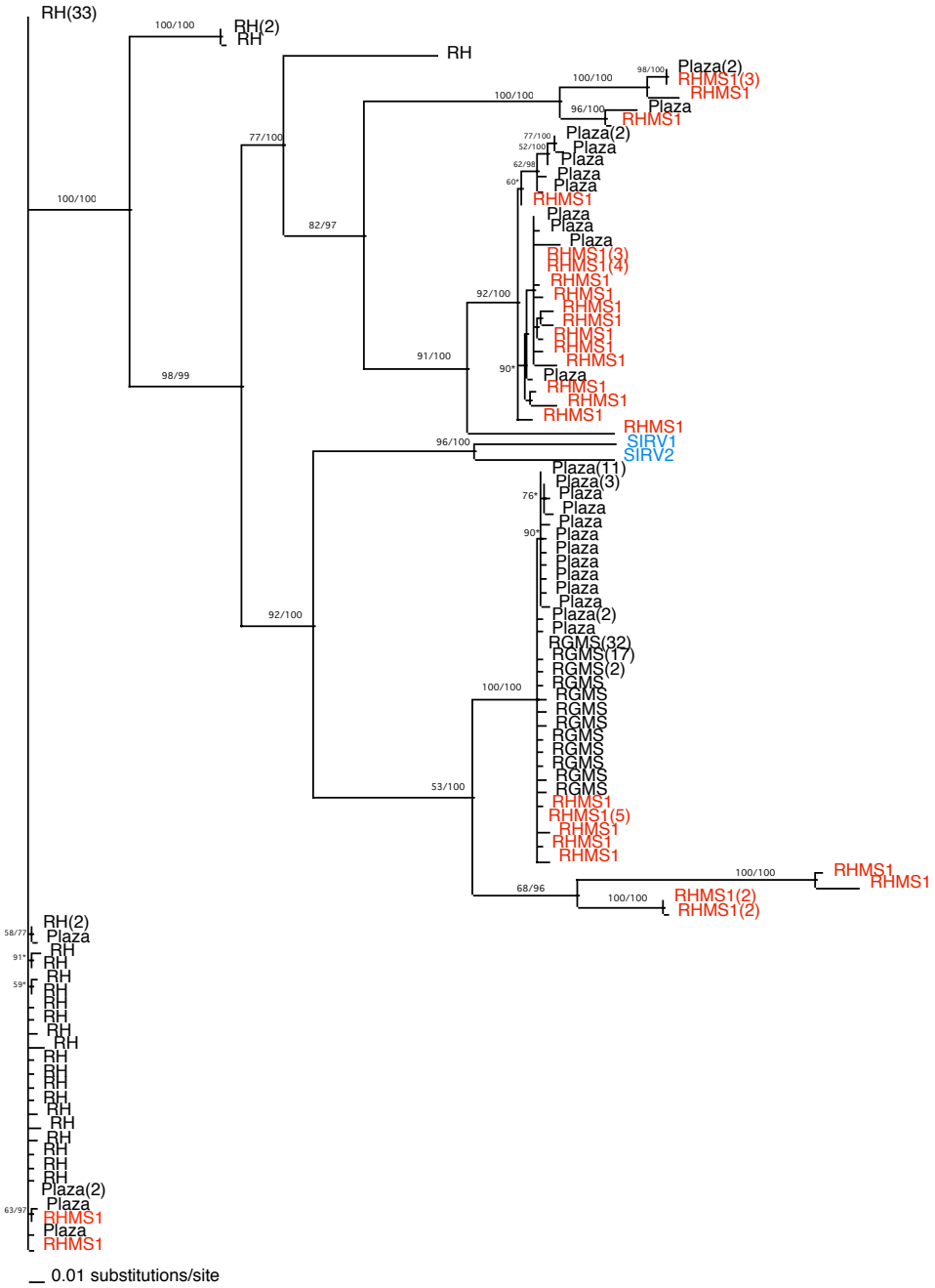


Figure 5.40. SIRV-like protein sequences from Ragged Hills. The red sequences represent those collected in April 2003. The * represents identical sequences. The numbers in parentheses indicate how many times a sequence was detected.

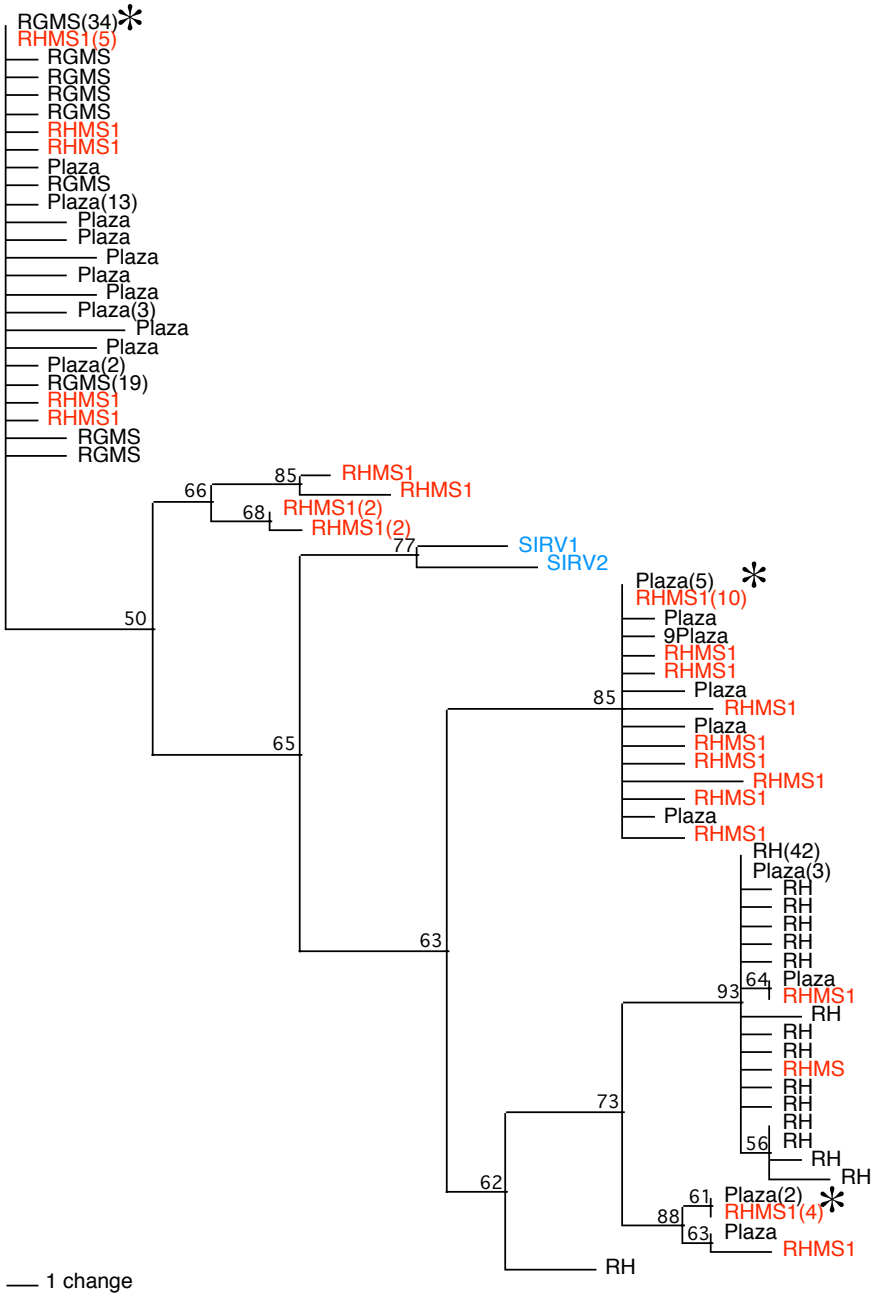


Figure 5.41. SIRV-like DNA sequences from Ragged Hills. The red sequences represent those collected on May 6, 2003. The numbers in parentheses indicate how many times a sequence was detected.

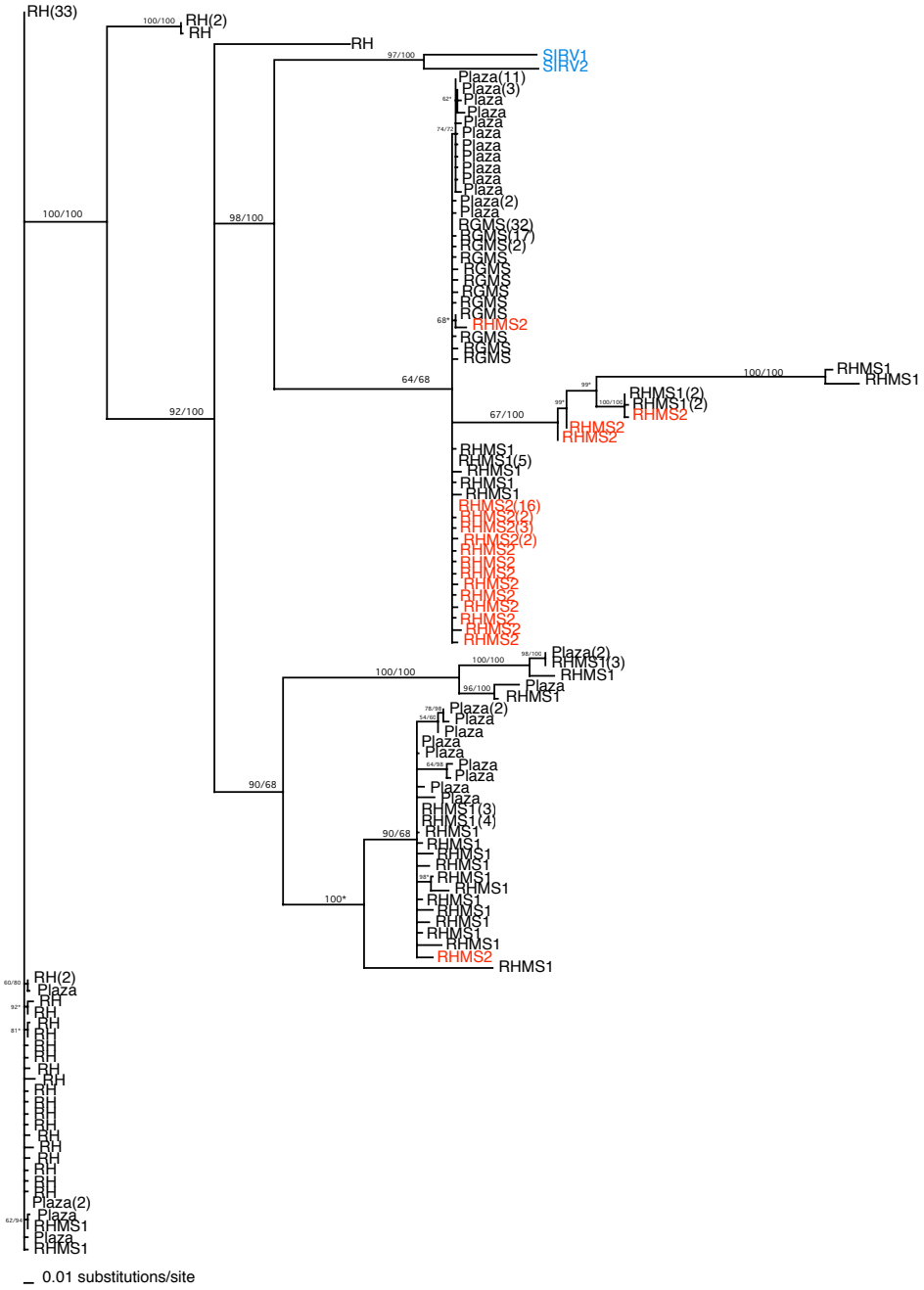


Figure 5.42. SIRV-like protein sequences from Ragged Hills. The red sequences represent those collected on May 6, 2003. The * represents identical sequences. The numbers in parentheses indicate how many times a sequence was detected.

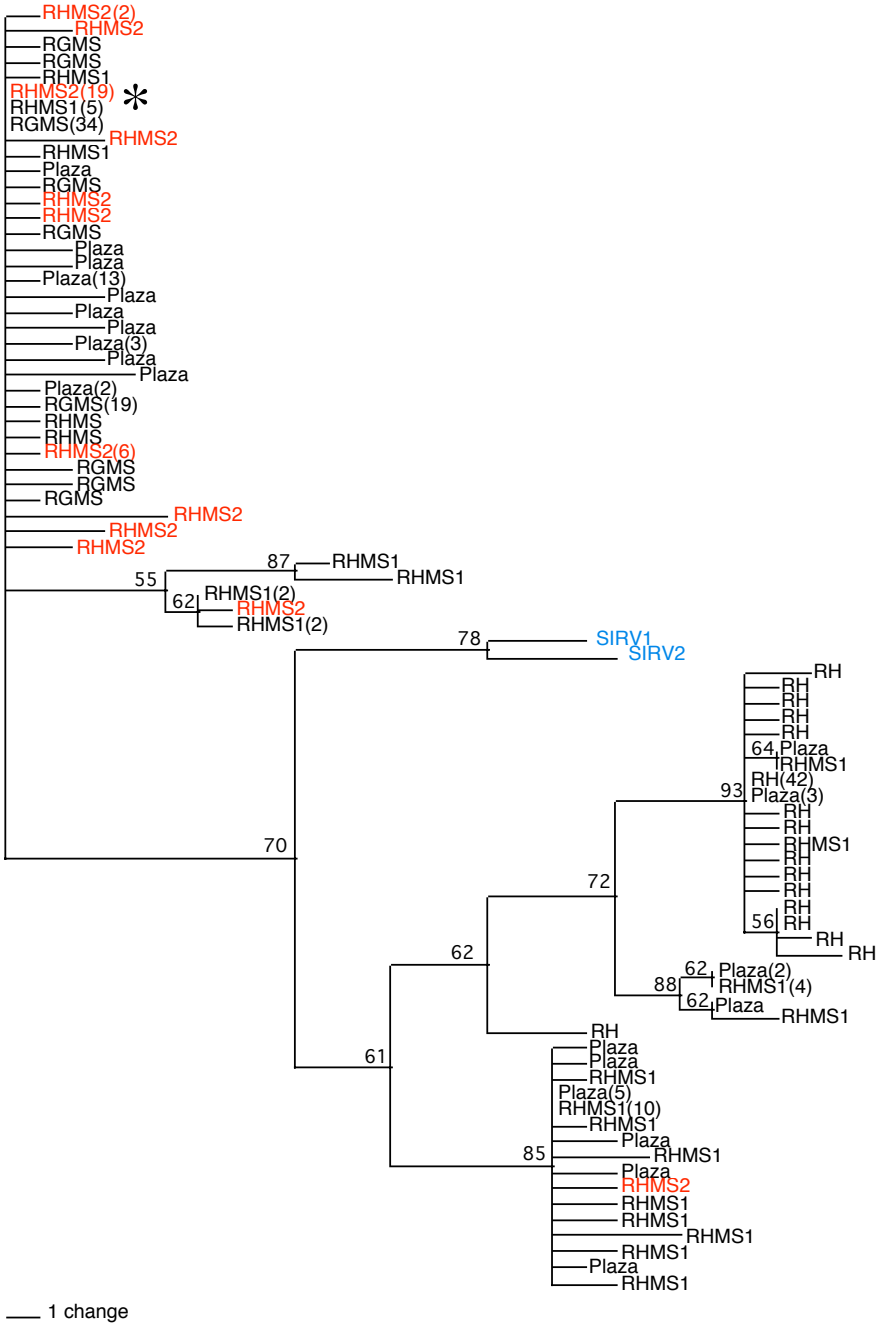


Figure 5.43. SIRV-like DNA sequences from Ragged Hills. The red sequences represent those collected on May 13, 2003. The numbers in parentheses indicate how many times a sequence was detected.

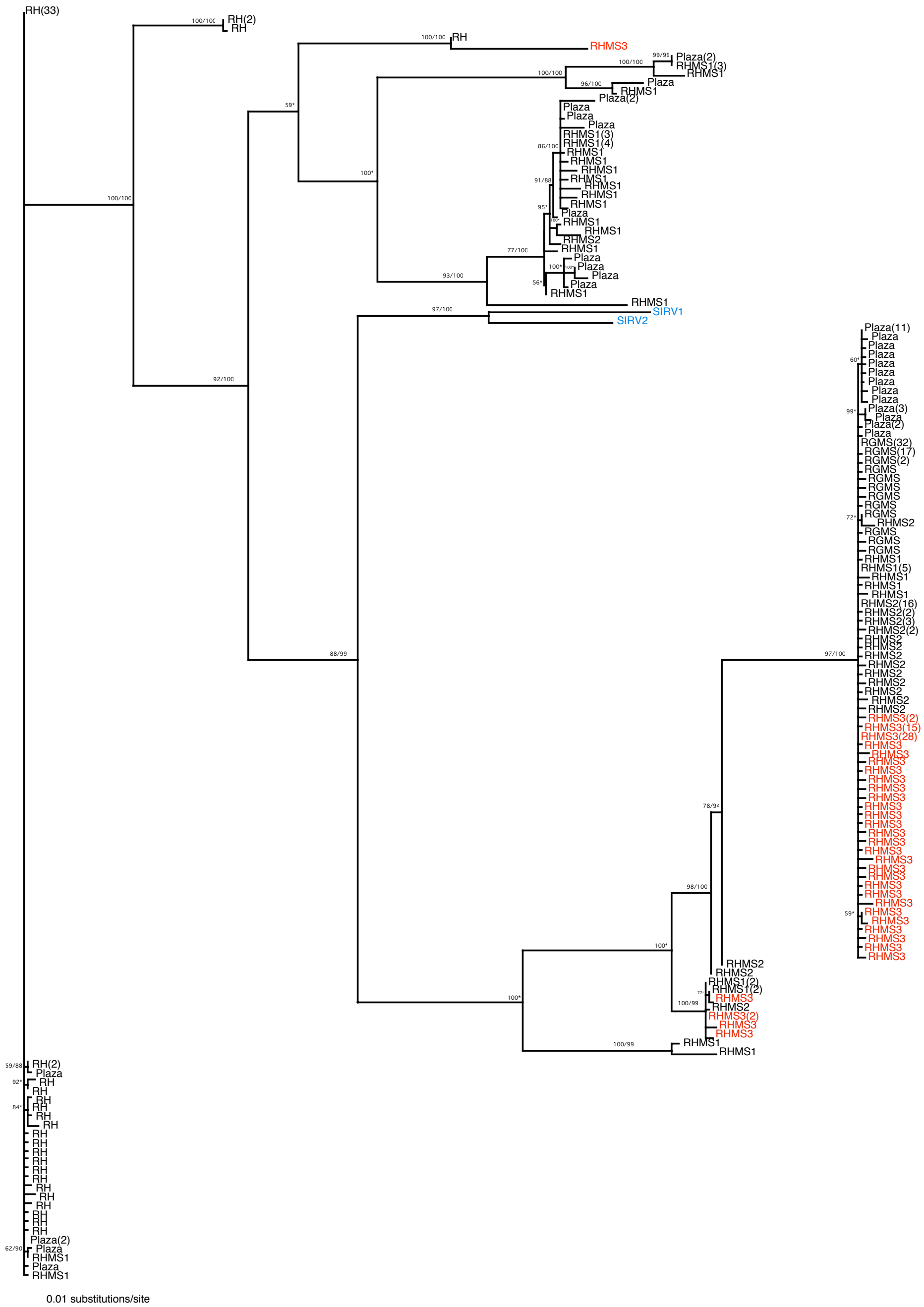


Figure 5.44. SIRV-like protein sequences from Ragged Hills. The red sequences represent those collected on May 13, 2003. The * represents identical sequences. The numbers in parentheses indicate how many times a sequence was detected.

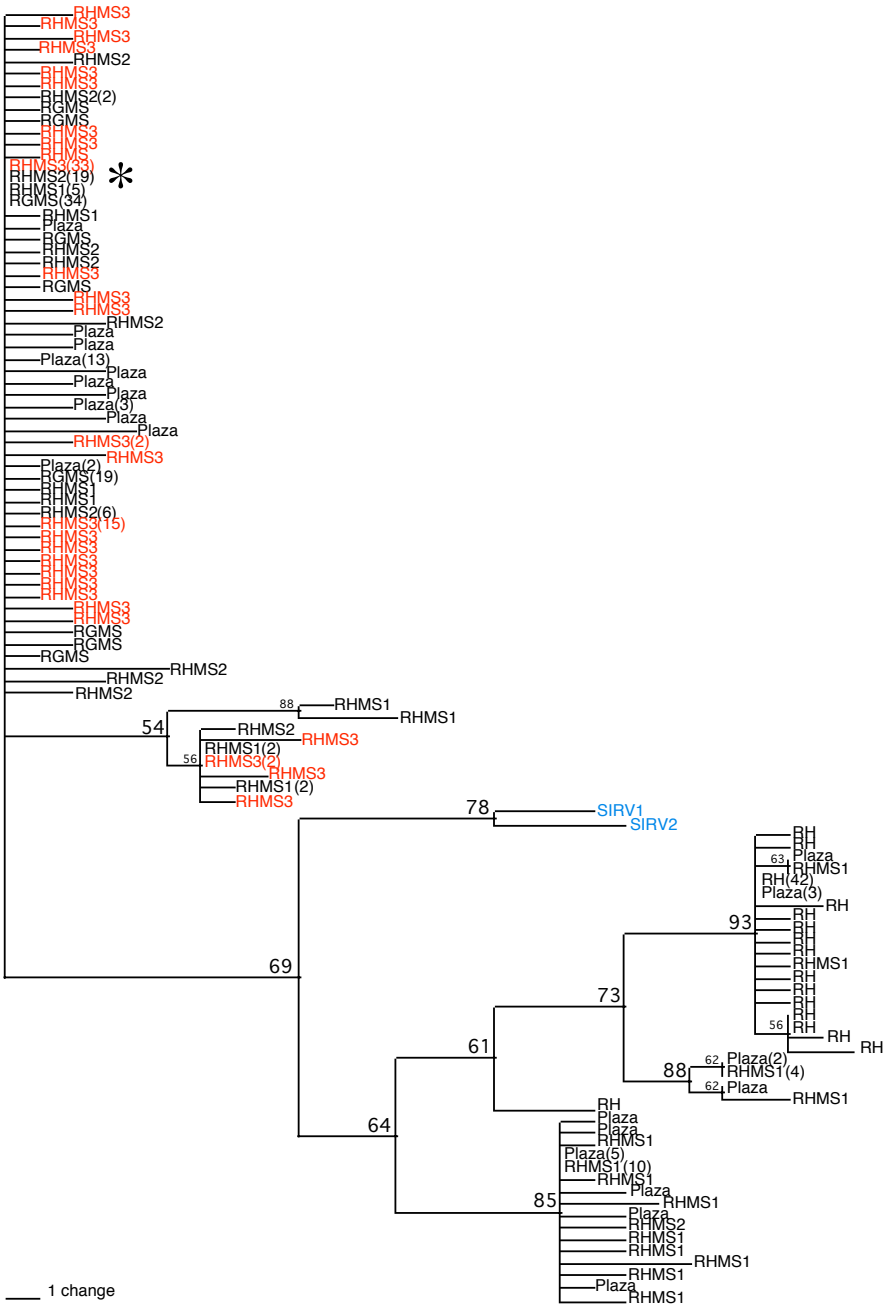


Figure 5.45. SIRV-like protein DNA from Ragged Hills. The red sequences represent those collected in June 2003. The numbers in parentheses indicate how many times a sequence was detected.

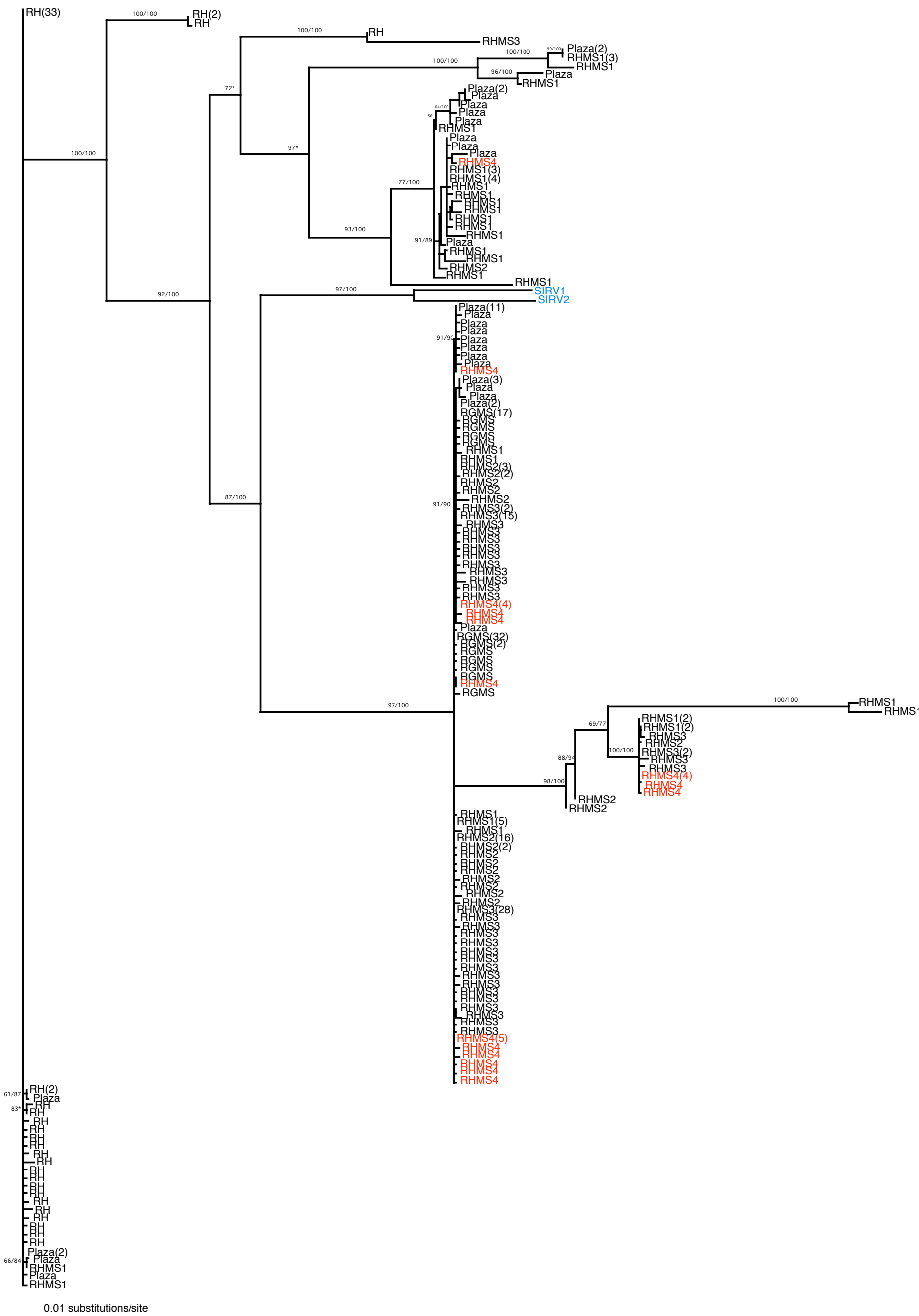


Figure 5.46. SIRV-like protein sequences from Ragged Hills. The red sequences represent those collected in June 2003. The * represents identical sequences. The numbers in parentheses indicate how many times a sequence was detected.

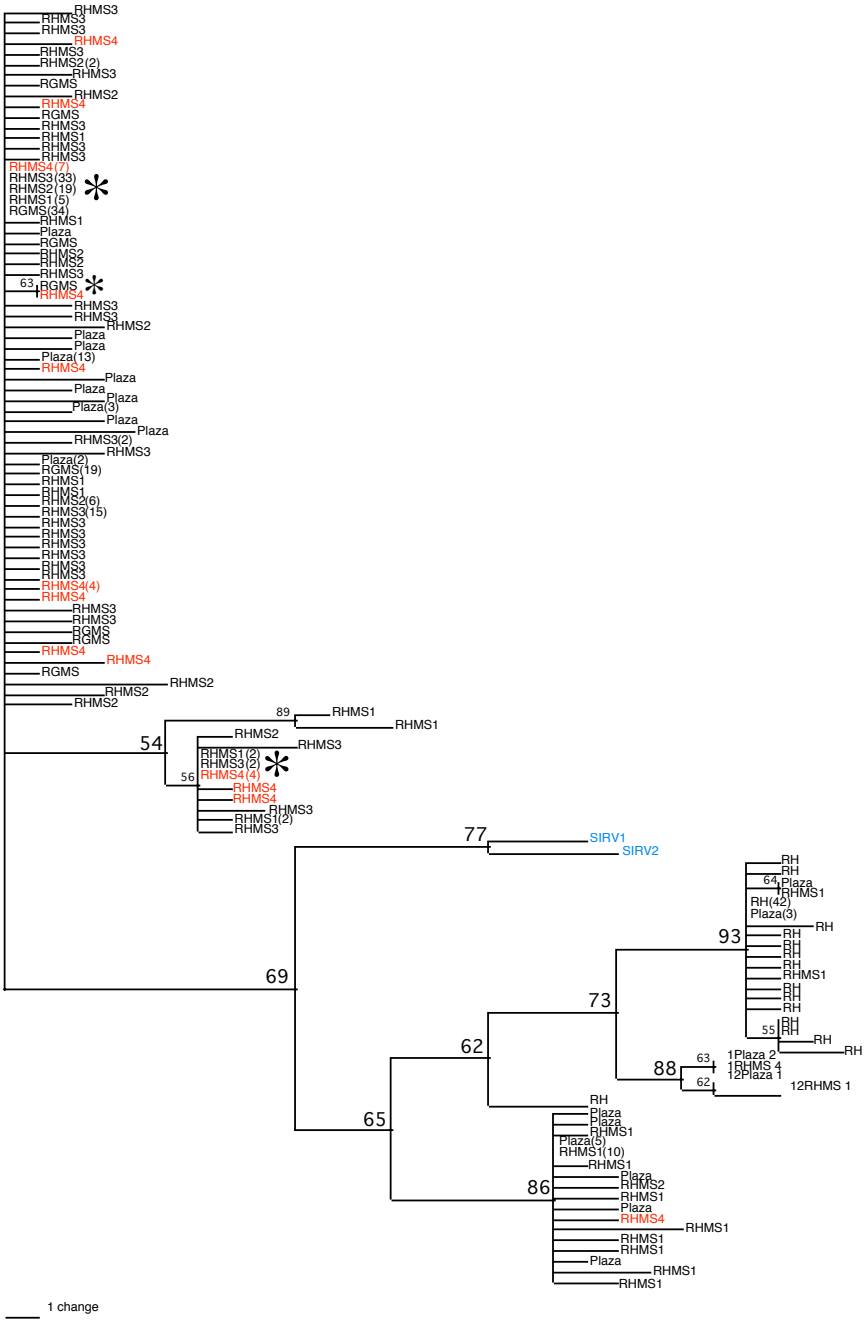


Figure 5.47. SIRV-like DNA sequences from Ragged Hills. The red sequences represent those collected in July 2003. The numbers in parentheses indicate how many times a sequence was detected.

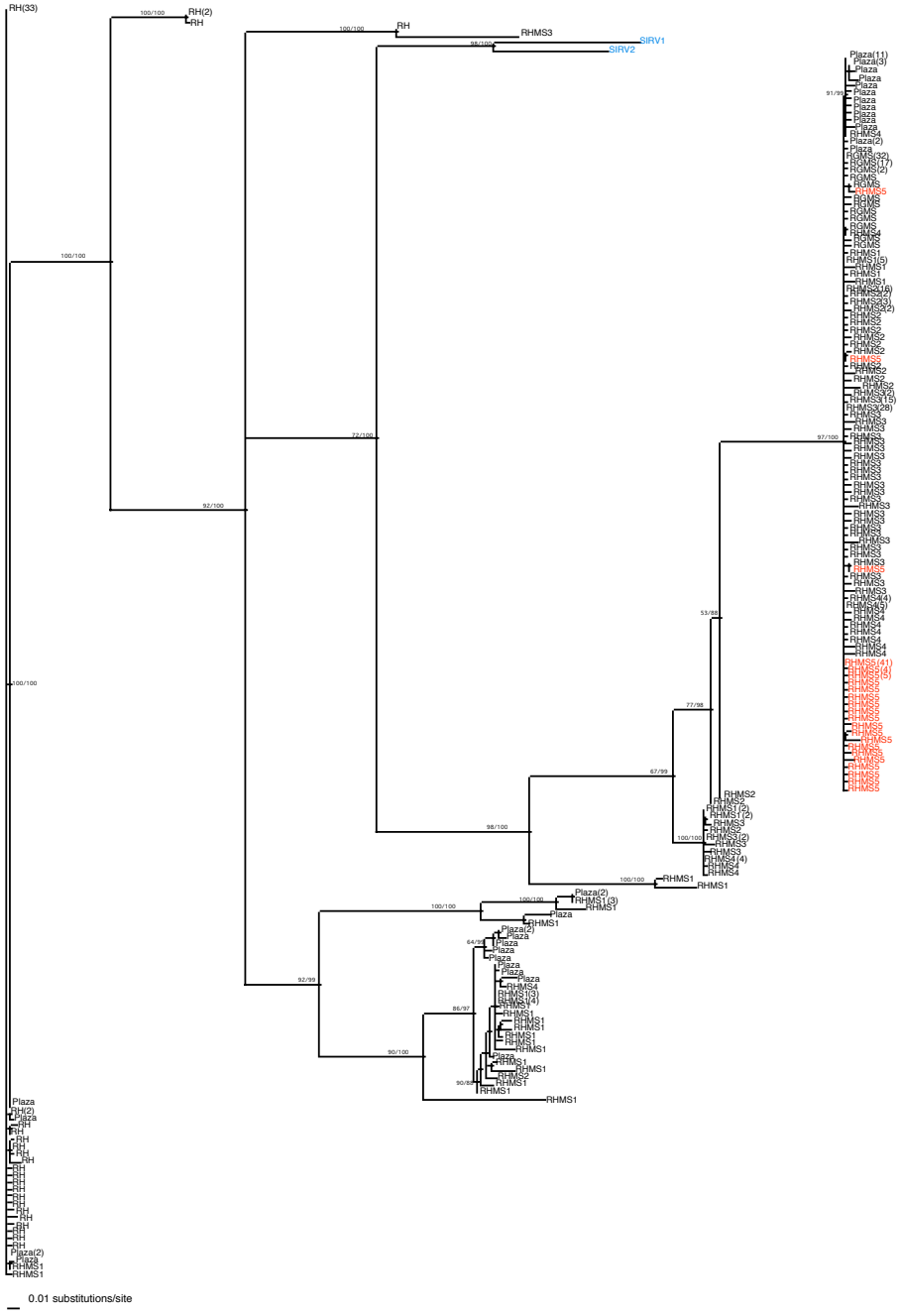


Figure 5.48. SIRV-like protein sequences from Ragged Hills. The red sequences represent those collected in July 2003. The * represents identical sequences. The numbers in parentheses indicate how many times a sequence was detected.

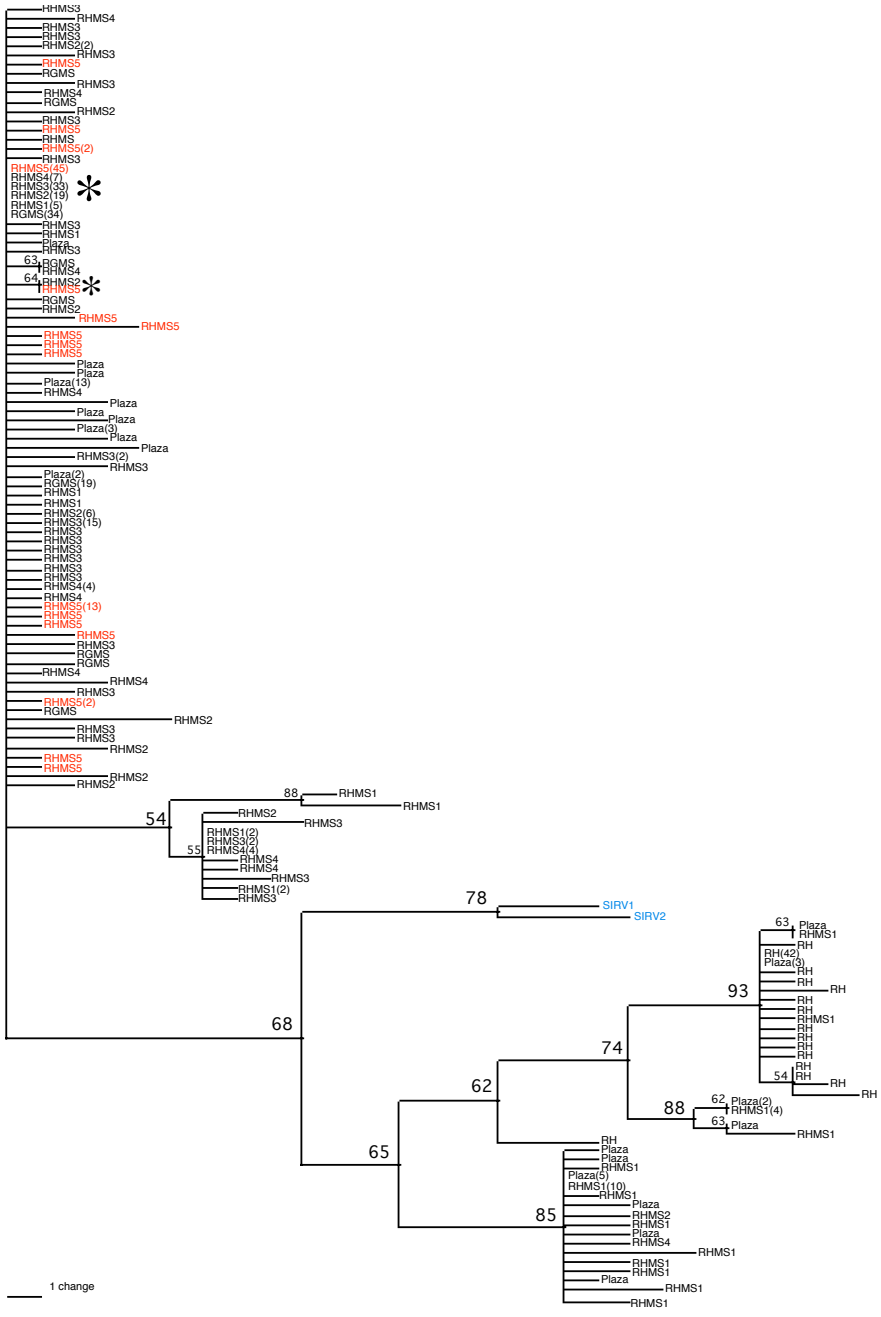


Figure 5.49. SIRV-like DNA sequences from Ragged Hills. The red sequences represent those collected on August 4, 2003. The numbers in parentheses indicate how many times a sequence was detected.

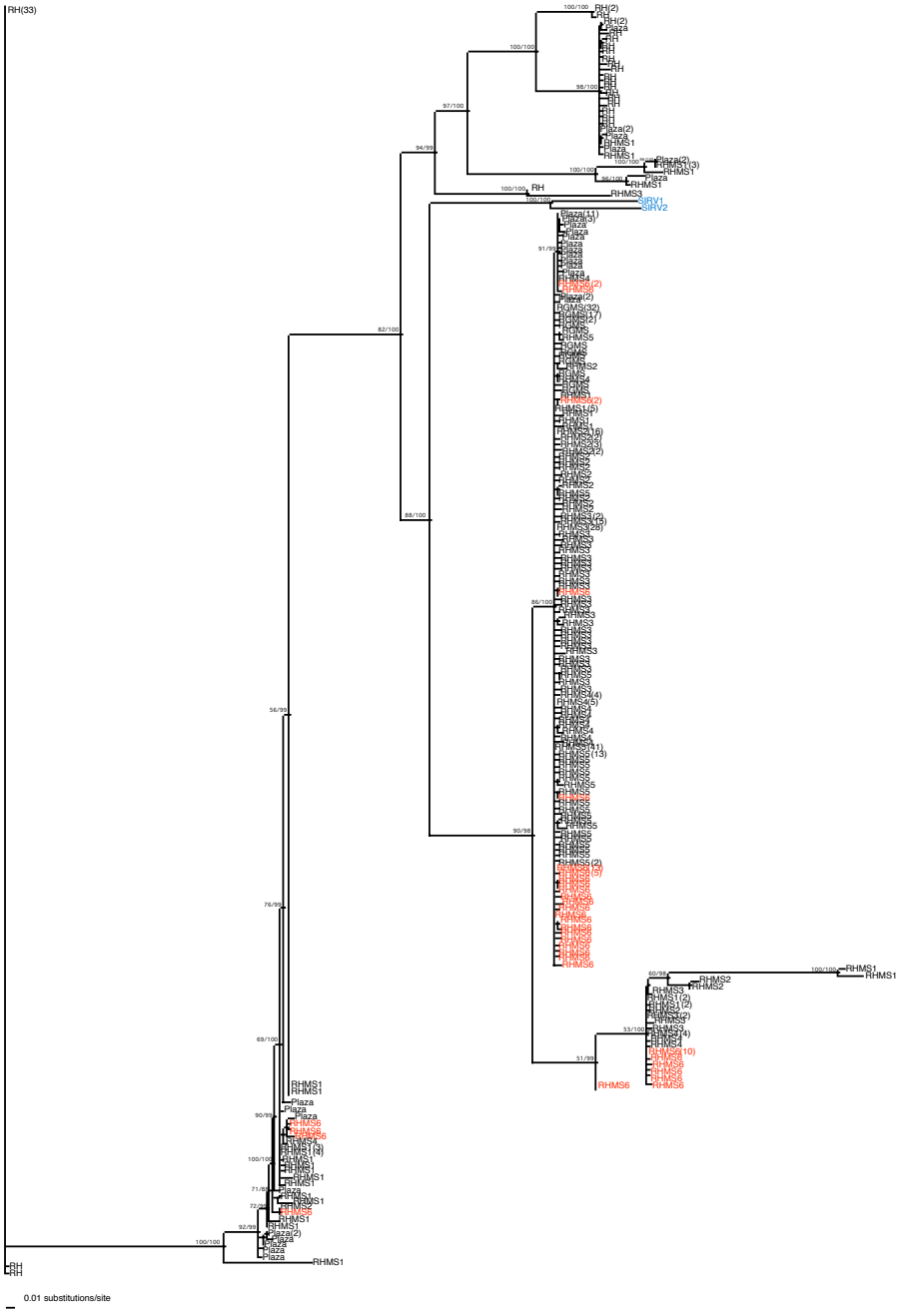


Figure 5.50. SIRV-like protein sequences from Ragged Hills. The red sequences represent those collected on August 4, 2003. The * represents identical sequences. The numbers in parentheses indicate how many times a sequence was detected.

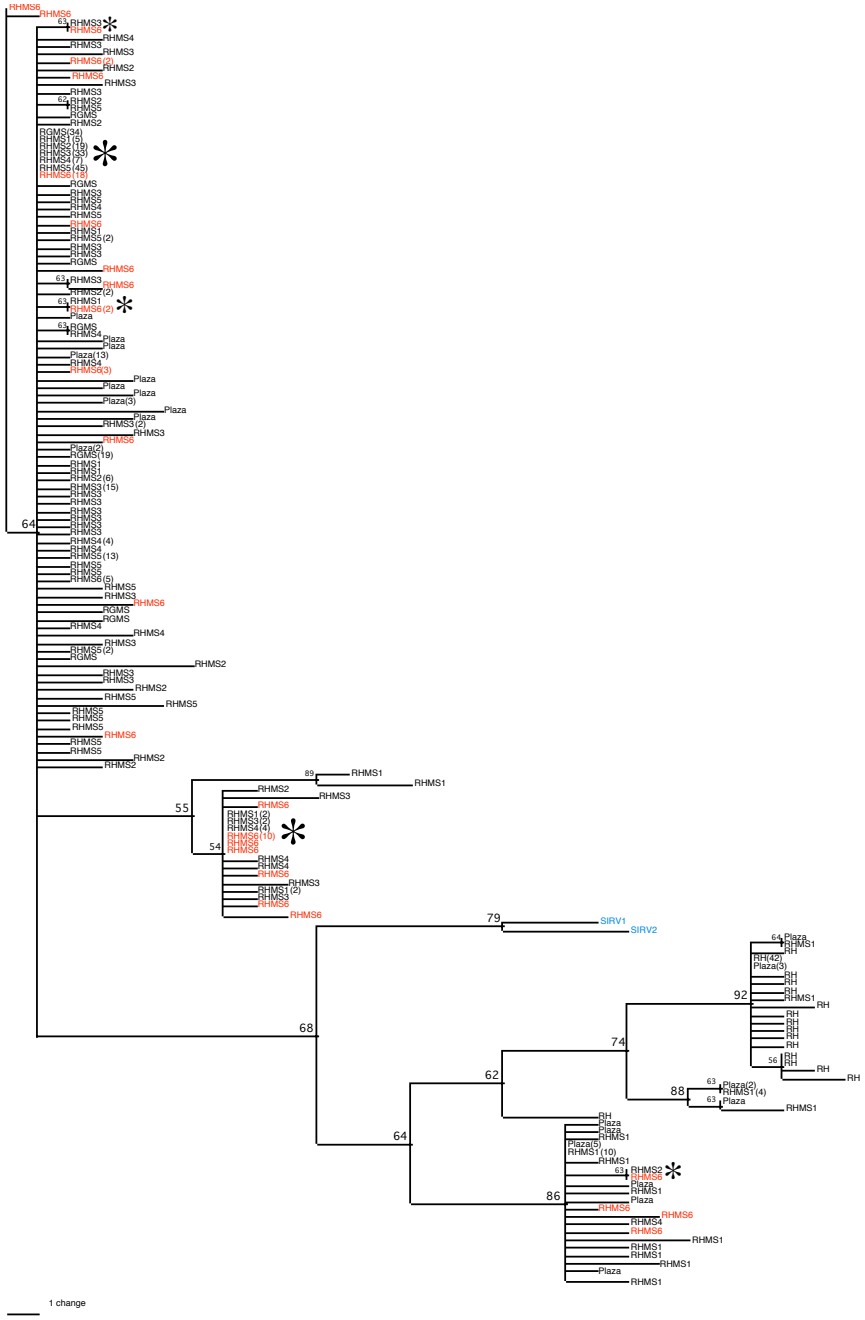


Figure 5.51. SIRV-like DNA sequences from Ragged Hills. The red sequences represent those collected on August 26, 2003. The numbers in parentheses indicate how many times a sequence was detected.

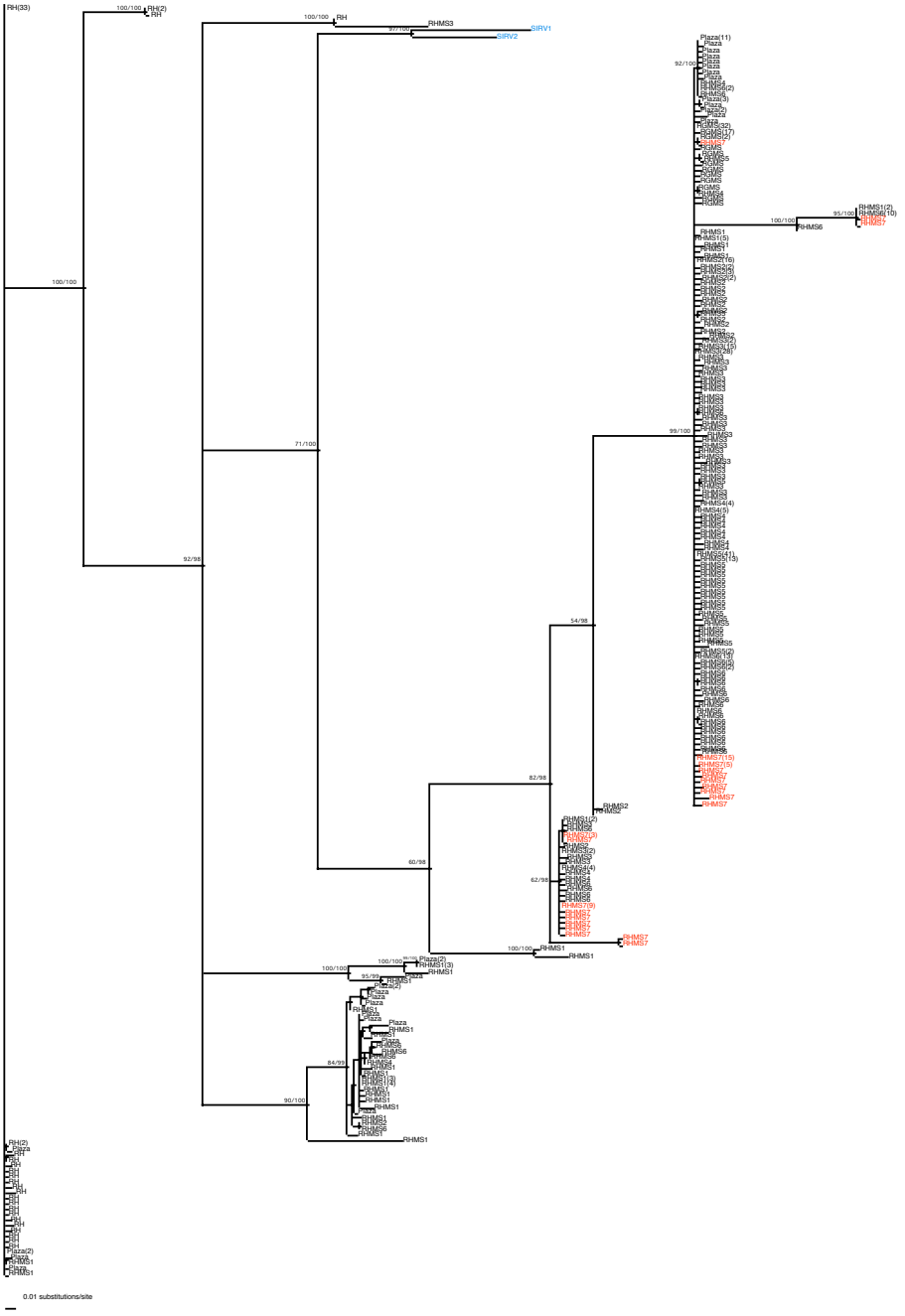


Figure 5.52. SIRV-like protein sequences from Ragged Hills. The red sequences represent those collected on August 26, 2003. The * represents identical sequences. The numbers in parentheses indicate how many times a sequence was detected.

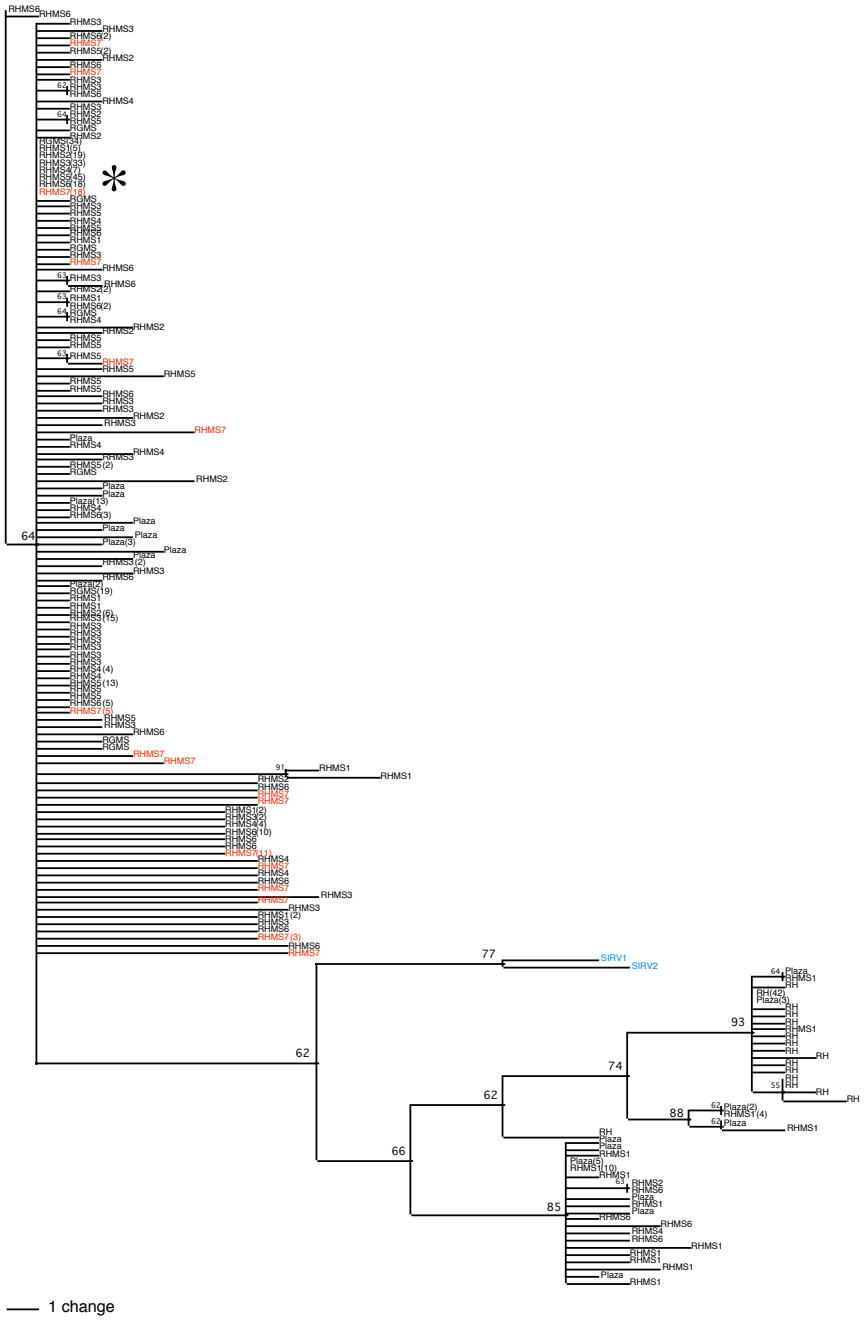


Figure 5.53. SIRV-like DNA sequences from Ragged Hills. The red sequences represent those collected in October 2003. The numbers in parentheses indicate how many times a sequence was detected.

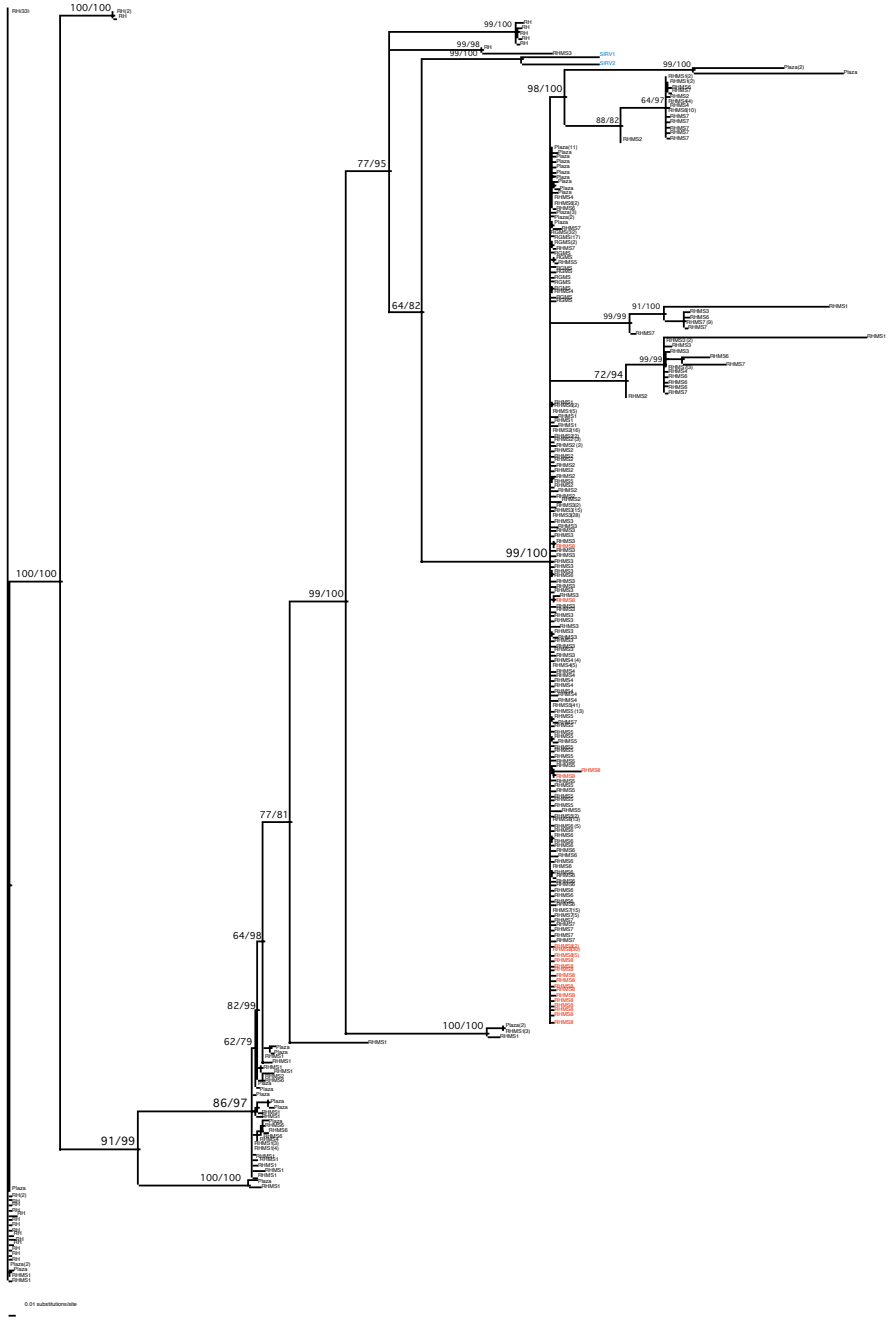


Figure 5.54. SIRV-like protein sequences from Ragged Hills. The red sequences represent those collected in October 2003. The * represents identical sequences. The numbers in parentheses indicate how many times a sequence was detected.

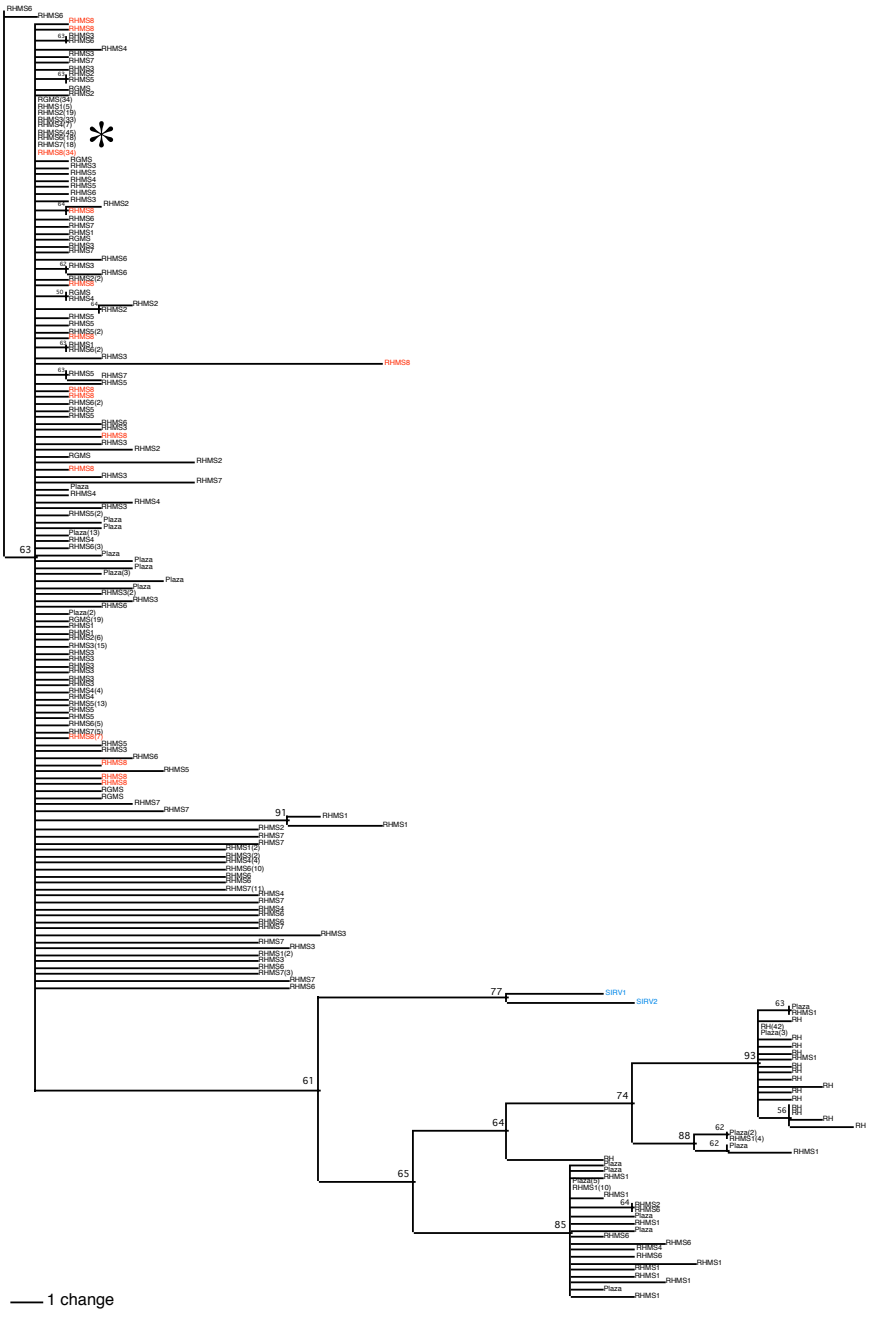


Figure 5.55. SIRV-like DNA sequences from Ragged Hills. The red sequences represent those collected in November 2003. The numbers in parentheses indicate how many times a sequence was detected.



Figure 5.56. SIRV-like protein sequences from Ragged Hills. The red sequences represent those collected in November 2003. The * represents identical sequences. The numbers in parentheses indicate how many times a sequence was detected.



Figure 5.57. SIRV-like DNA sequences from Ragged Hills. The red sequences represent those collected in April 2004. The numbers in parentheses indicate how many times a sequence was detected.

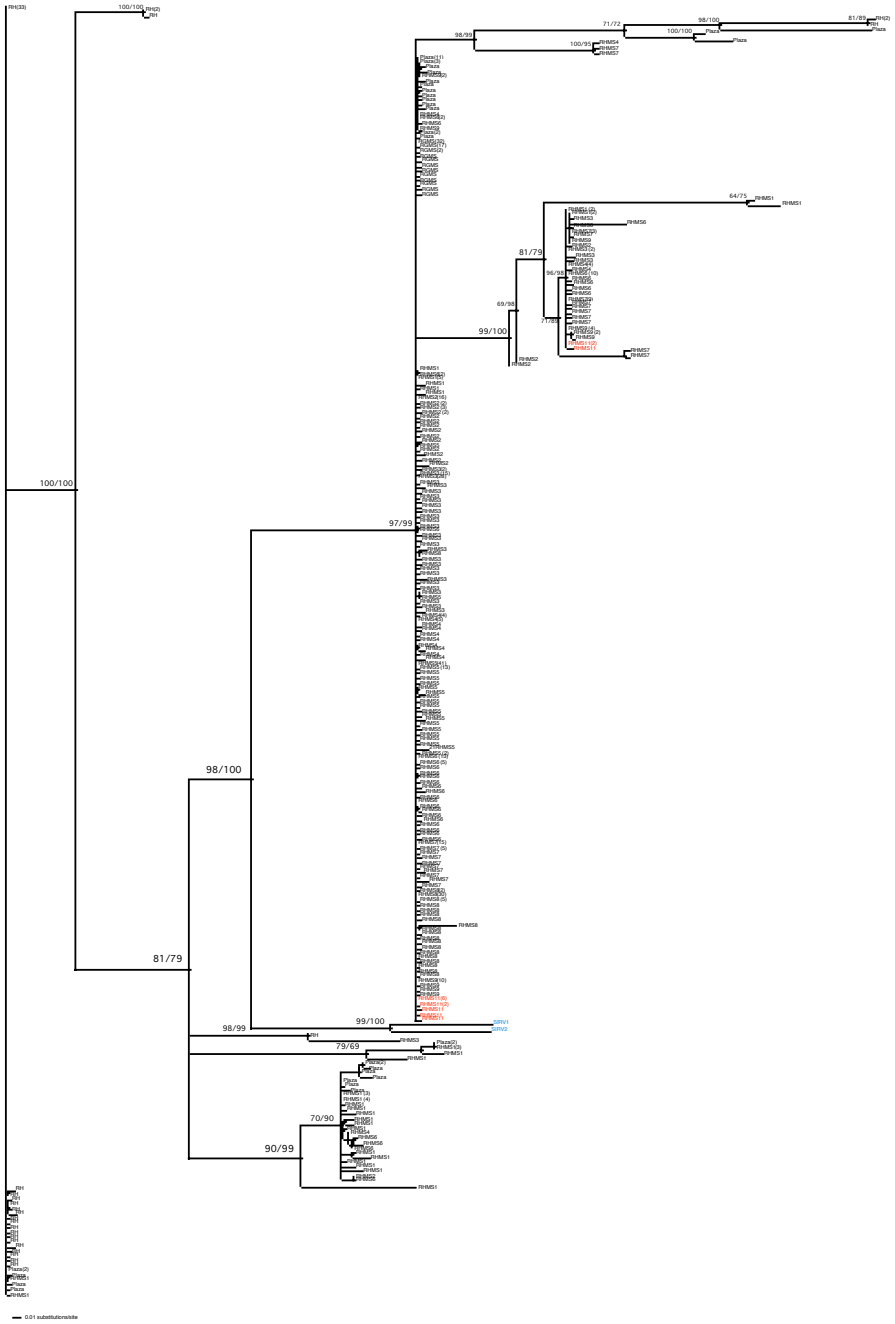


Figure 5.58. SIRV-like protein sequences from Ragged Hills. The red sequences represent those collected in April 2004. The * represents identical sequences. The numbers in parentheses indicate how many times a sequence was detected.

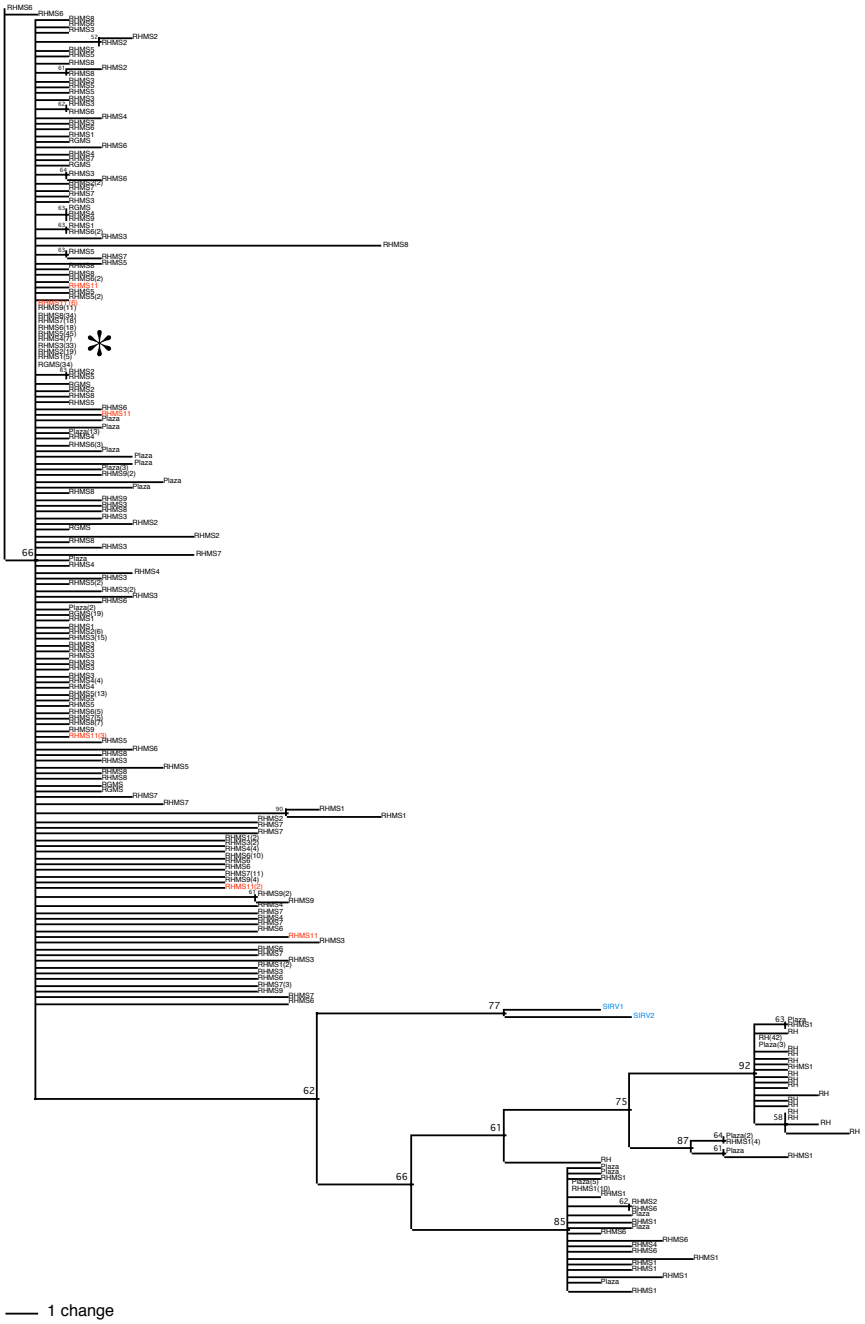


Figure 5.59. SIRV-like DNA sequences from Ragged Hills. The red sequences represent those collected in June 2004. The numbers in parentheses indicate how many times a sequence was detected.

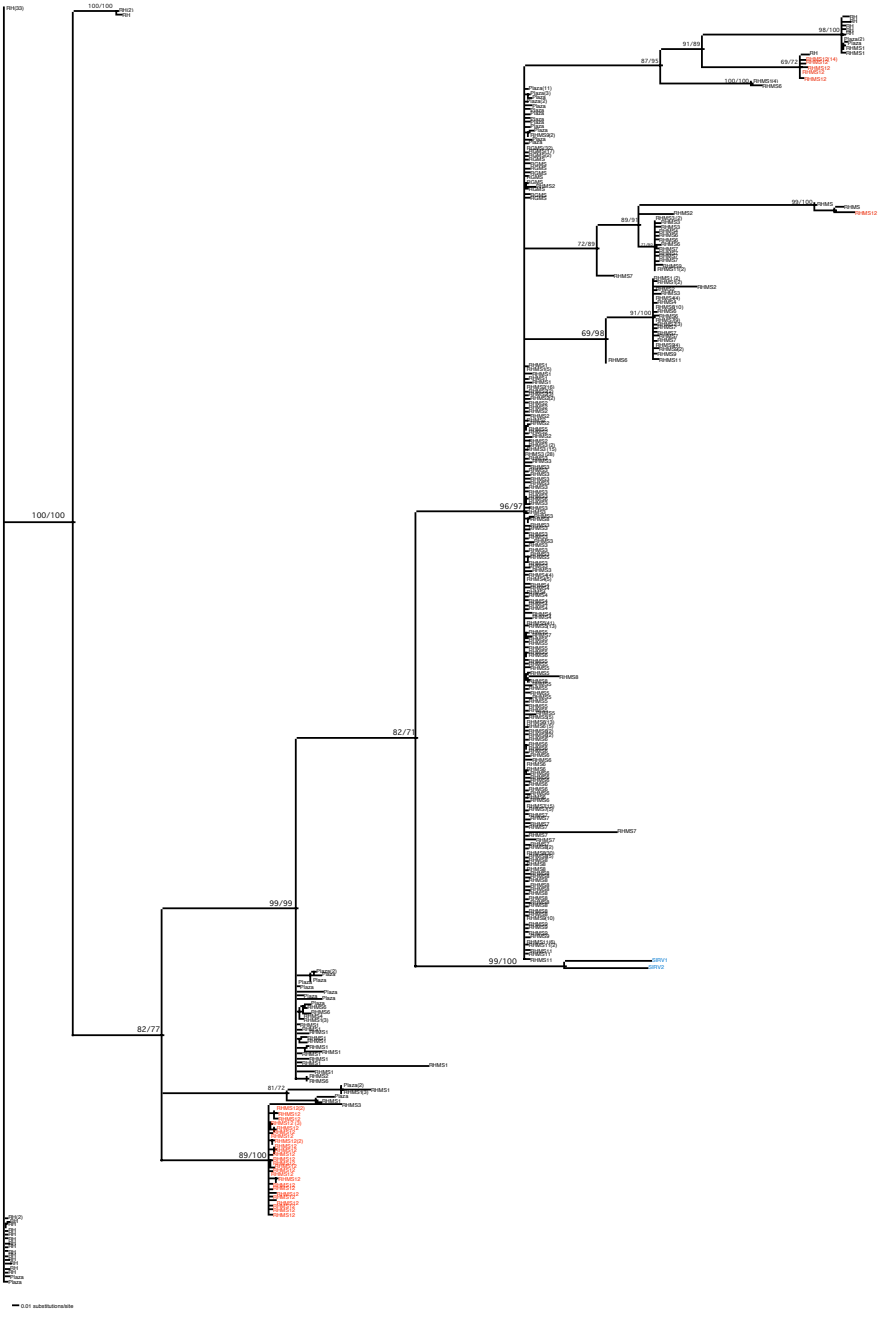


Figure 5.60. SIRV-like protein sequences from Ragged Hills. The red sequences represent those collected in June 2004. The * represents identical sequences. The numbers in parentheses indicate how many times a sequence was detected.

