

A metagenomic approach using next-generation  
sequencing for viral profiling of a vineyard  
and  
genetic characterization of Grapevine virus E

by  
Beatrix Coetzee



*Thesis presented in fulfilment of the requirements for the degree  
Master of Science in Genetics at Stellenbosch University*

Supervisor: Prof. Johan T. Burger  
Co-supervisor: Dr. Michael-John Freeborough

Department of Genetics  
Faculty of Science

December 2010

---

---

## Declaration

---

---

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the owner of the copyright thereof (unless to the extent explicitly otherwise stated) and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

---

B Coetzee

Date

---

---

## Abstract

---

---

Next-generation sequencing technologies are increasingly used in metagenomic studies, largely due to the high sequence data throughput capacity and unbiased approach in determining the genetic composition of an unknown environmental sample. This study investigated the applicability of the Illumina next-generation sequencing platform for metagenomic sequencing of grapevine viruses to provide the first complete viral profile, or virome, of a diseased vineyard.

Leaf material was harvested from 44 randomly selected vines in a leafroll-diseased vineyard in South Africa. Sample material was pooled and double-stranded RNA extracted. The dsRNA was sequenced as a paired-end sequencing run using the Illumina sequencing-by-synthesis technique, and more than 19 million sequence reads, equivalent to approximately 837 megabases of metagenomic sequence data, were obtained. Of these data, approximately 400 megabases could be assembled into 449 scaffolds, using the *de novo* assembler Velvet. These scaffolds were subjected to BLAST searches against the NCBI databases and top hit scores were used for virus identification. Based on the BLAST results, suitable sequences were selected from the NCBI database and used as reference sequence in MAQ mapping assemblies.

The bioinformatic analyses allowed for the determination of the virus species present, the most prominent variants, and the relative abundance of each. Four known grapevine viral pathogens were identified. Grapevine leafroll-associated virus 3, representing 59% of the analyzed short read sequence data, was identified as the most prominent virus species. Three variants of this virus were detected: GP18 was the most abundant, followed by a minor CI766/NY1 variant and a potential novel grapevine leafroll-associated ampelovirus. A single Grapevine rupestris stem pitting-associated virus variant, similar to SG1, and a *Grapevine virus A* variant, a member of molecular group III, were identified. This study is also the first to report the presence of Grapevine virus E (GVE) in South African vineyards.

Grapevine virus E was further genetically characterized and the genome sequence of GVE isolate SA94 determined. The GVE SA94 genome sequence, 7568 nucleotides in length, is the first complete genome sequence for the virus species. The genome organization of GVE SA94 is typical of vitiviruses, but in contrast to other RNA viruses, the AlkB domain is located within the helicase domain in open reading frame 1 (ORF 1). Grapevine virus E SA94 shares nearly 100% nucleotide identity with the Japanese TvP15 isolate and GVE 3404, a *de novo* scaffold generated from the metagenomic sequence data.

Bioinformatic analysis of metagenomic sequence data further revealed the presence of three fungus-infecting viral families, *Chrysoviridae*, *Totiviridae* and the unclassified dsRNA virus, *Fusarium graminearum* dsRNA mycovirus 4. A virus from the family *Chrysoviridae*, similar to *Penicillium chrysogenum* virus, was the second most abundant virus detected.

We demonstrated the successful application of a short read sequencing technology, such as the Illumina platform, for viral profiling of an infected vineyard. To our knowledge this is the first application of the Illumina technology for this purpose.

---

---

## Opsomming

---

---

Volgende-generasie tegnologie om basis volgordes van nukleinsure te bepaal, word al meer gebruik in metagenomiese studies. Dit is veral weens die hoë data-omset kapasiteit en onbevooroordeelde aanslag in die bepaling van die genetiese samestelling van onbekende omgewingsmonsters. Hierdie studie het die aanwending van die Illumina volgende-generasie volgorde-bepalingsplatform in 'n metagenomiese studie van wingerdvirusse, ondersoek. Dit het ten doel gehad om die eerste volledige virus profiel, of viroom, van 'n geïnfecteerde wingerd saam te stel.

Blaarmateriaal is verkry vanaf 44 lukraak-gekose wingerdstokke in 'n rolblad-geïnfecteerde wingerd in Suid-Afrika. Monster materiaal is saamgevoeg en dubbelstring-RNS geëkstraheer. Die dubbelstring-RNS is onderwerp aan gepaarde-ent volgorde-bepaling deur gebruik te maak van die Illumina volgorde-bepaling-deur-sintese tegniek. Meer as 19 miljoen volgorde reekse, ekwivalent aan ongeveer 837 megabasisse volgorde data, is verkry. Van hierdie data kon ongeveer 400 megabasisse saamgevoeg word in 449 konstruksie ("scaffolds"), deur gebruik te maak van die *de novo* samesteller Velvet. Hierdie konstruksie is onderwerp aan BLAST soektogte teen die NCBI databasisse en die hoogste trefslag-telling is gebruik vir virus identifikasie. Op grond van die "BLAST" resultate is geskikte volgordes geselekteer vanaf die NCBI databasis en gebruik as verwysingvolgordes in MAQ kartering-analises.

Met die bioinformatika analises kon die virus spesies teenwoordig, asook die mees prominente variante en relatiewe voorkoms van elk, bepaal word. Vier bekende virus wingerdpatogene is geïdentifiseer. Grapevine leafroll-associated virus 3, verteenwoordig deur 59% van die geanaliseerde kort-reeks volgorde data, is identifiseer as die mees prominente virus spesie. Drie variante van die virus is in die wingerdmonster opgespoor: GP18 kom die mees algemeen voor, gevolg deur 'n CL-766/NY1 variant en 'n potensiële nuwe wingerd rolblad-geassosieerde ampelovirus. 'n Enkele Grapevine rupestris stem pitting-associated virus variant, soortgelyk aan SG1, en 'n Grapevine virus A variant, 'n lid van molekuleêre groep III, is geïdentifiseer. Hierdie studie is ook die eerste om die teenwoordigheid van Grapevine virus E (GVE) in Suid-Afrikaanse wingerde te rapporteer.

Grapevine virus E is verder geneties gekarakteriseer en die genoomvolgorde van GVE isolaat SA94 is bepaal. Die GVE SA94 genoomvolgorde, 7568 nukleotiede lank, is die eerste volledige genoomvolgorde vir hierdie virus spesie. Die genoomorganisasie is tipies van vitivirusse, maar in kontras met ander RNA virusse is die AlkB domein binne-in die helikase domein van ooplesraam 1 (ORF 1) geleë. Grapevine virus E SA94 deel byna 100% nukleotied identiteit met die Japannese TvP15 isolaat en GVE 3404, 'n *de novo* konstruk gegenereer vanaf die metagenomiese volgorde data.

Bioinformatika analyses van die metagenomiese volgorde data het verder die teenwoordigheid van drie swam-infekterende virus families, die *Chrysoviridae*, *Totiviridae* en ongeklassifiseerde dubbelstring-RNS virus, *Fusarium graminearum* dsRNA mycovirus 4, aangetoon. 'n Virus van die *Chrysoviridae* familie, soortgelyk aan *Penicillium chrysogenum* virus, het die tweede meeste voorgekom in die wingerd monster.

Hierdie studie demonstreer die suksesvolle toepassing van 'n kort reeks volgordebepalingstegnologie soos die Illumina platform, vir die opstel van 'n virusprofiel van 'n geïnfekteerde wingerd. Sover ons kennis strek is hierdie die eerste aanwending van die Illumina tegnologie vir hierdie doel.

---

# Abbreviations

---

°C	Degrees Celsius
3'UTR	3' Untranslated Region
5'UTR	5' Untranslated Region
ABI	Applied Biosystems
AlkB	Alkylated DNA repair protein
APS	Adenosine Phosphosulphate
ATP	Adenosine Triphosphate
BLAST	Basic Local Alignment Search Tool
BLASTn	BLAST (search a nucleotide database using a nucleotide query)
BLASTx	BLAST (search protein database using a translated nucleotide query)
bp	base pairs
CDD	Conserved Domain Database
cDNA	complementary Deoxyribonucleic Acid
corp.	corporation
CP	Coat Protein
CRT	Cyclic Reversible Termination
CsCl	Cesiumchloride
CTAB	N-Cetyl-N,N,N-trimethyl Ammonium Bromide
cv.	cultivar
ddNTP	2',3'-dideoxynucleotide triphosphate
DNA	Deoxyribonucleic Acid
dsDNA	double-stranded Deoxyribonucleic Acid
dsRNA	double-stranded Ribonucleic Acid
eDNA	environmental Deoxyribonucleic Acid
ELISA	Enzyme-Linked Immunosorbent Assay
ESS	Environment Shotgun Sequencing
Gb	Gigabases
GLRaV-3	Grapevine leafroll associated virus-3
GOS	Global Ocean Sampling
GRSPaV	Grapevine rupestris stem pitting-associated virus
GRVfV	Grapevine rupestris vein-feathering virus
GSyV-1	Grapevine Syrah Virus-1
GVA	Grapevine virus A
GVB	Grapevine virus B
GVD	Grapevine virus D
GVE	Grapevine virus E
Hel	Helicase
LRS	Long Sequence Reads
MAQ	Mapping and Assembly with Quality
Mb	Megabases
min	minute
miRNA	micro Ribonucleic Acid
MP	Movement Protein
mRNA	messenger Ribonucleic Acid
Mtr	Methyltransferase

NB	Nucleic acid-Binding protein
NCBI	National Centre of Biotechnology Information
NGS	Next-Generation Sequencing
nr	non-redundant
nt	nucleotides
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
PcV	Penicillium chrysogenum virus
PE	Paired-End
pers. com.	personal communication
pM	picoMolar
PPi	Pyrophosphate
RdRp	RNA-dependant RNA polymerase
RLM-RACE	RNA Ligase-Mediated Rapid Amplification of cDNA Ends
RNA	Ribonucleic Acid
rRNA	Ribosomal Ribonucleic Acid
RT-PCR	Reverse Transcription - Polymerase Chain Reaction
SAFV	Scaffoldvirus
SAWIS	South African Wine Industry Information and Systems
SD	Shiraz Disease
SNP	Single Nucleotide Polymorphism
SOLiD	Sequencing by Oligo Ligation and Detection
SRS	Short Sequence Reads
ssDNA	single-stranded Deoxyribonucleic Acid
USA	United States of America
WOSA	Wines of South Africa



---

---

# Acknowledgements

---

---

I would like to express my sincerest gratitude and appreciation to the following people and institutions:

- My supervisor Prof. Johan Burger for his guidance and giving me the opportunity to do this study.
- Dr. Michael-John Freeborough and Dr. Hano Maree for their leadership and intellectual inputs.
- Dr. Dirk Stephan for his input and help with the GVE work.
- Prof. Jasper Rees and Dr. Jean-Marc Celton for allowing me to observe the sequencing procedure and help with the bioinformatic analysis.
- My colleagues in the Vitis lab for their friendship and input into this project.
- The Harry Crossley foundation and Stellenbosch University for personal financial assistance.
- Winetech and National Research Foundation (NRF) THRIP for the financial contribution towards this project. Opinions expressed and conclusions arrived at, are those of the authors and are not necessarily to be attributed to the NRF.
- My parents for their love, support and encouragement during this study.
- My Heavenly Father.

Dedicated to my loving parents.

---

---

## List of Figures

---

---

<b>Figure 2.1</b> Grapevine with typical leafroll symptoms <b>a)</b> Red cultivar displaying interveinal reddening <b>b)</b> White cultivar with leaves rolled downwards.....	7
<b>Figure 2.2</b> Grapevine with Shiraz disease symptoms <b>a)</b> Green shoots with a lack of lignification <b>b)</b> Typical Shiraz disease leaf discoloration patterns. Leave edges start to turn red progressing to completely red leaves (www.wynboer.co.za).....	8
<b>Figure 2.3</b> Typical Shiraz decline symptoms <b>a)</b> Reduced vigour and premature red discoloration of leaves <b>b)</b> Swelling at the graft union (www.wynboer.co.za). ....	8
<b>Figure 2.4</b> General comparison of the sequencing technologies from the three next-generation sequencing platforms: 454/Roche, Illumina and ABI SOLiD (Adapted from Hudson, 2008). ....	20
<b>Figure 2.5</b> Diagram illustrating the three steps of the Illumina Genome Analyzer sequencing technology (Adapted from: <a href="http://www.illumina.com">http://www.illumina.com</a> ). ....	22
<b>Figure 2.6</b> Diagram of a modified nucleotide used in Illumina sequencing (Adapted from Metzker, 2010).....	25
<b>Figure 2.7</b> Diagram illustrating the theory of De Bruijn graphs used in Velvet assembler <b>a)</b> Sequence read with possible k-mers <b>b)</b> De Bruijn graph featuring nodes and edges <b>c)</b> Eulerian paths showing two overlapping sequence reads. (Adapted from Pop, 2009). ....	30
<b>Figure 3.1</b> Comparative percentages for read counts utilized in scaffolds for each sequence classification according to best hit with BLASTn or BLASTx searches. <b>GLRaV-3</b> Grapevine leafroll-associated virus 3, <b>GRSPaV</b> Grapevine rupestris stem pitting-associated virus, <b>GVA</b> Grapevine virus A and <b>GVE</b> Grapevine virus E.....	43
<b>Figure 3.2</b> MAQ-reassembly of reads on four full-length genomes representing the dominant variants for <b>a)</b> GLRaV-3 (GP18), <b>b)</b> GRSPaV (SG1), <b>c)</b> GVA (GTR1-1) and <b>d)</b> GVA (P163-1). GVE was excluded due to the lack of a full-length genome. Schematic representations of virus genomes with numbered open reading frames are shown above graphs. Grey bars below graph highlight areas with no coverage. <b>GLRaV-3</b> Grapevine leafroll-associated virus 3, <b>GRSPaV</b> Grapevine rupestris stem pitting-associated virus, <b>GVA</b> Grapevine virus A.....	46

**Figure 3.3** Phylogenetic tree (bootstrap consensus tree) showing the relationship between the six complete genome sequences and the *de novo* generated scaffold (Node 192) for Grapevine rupestris stem pitting-associated virus (GRSPaV). Node 192 group with the SG1 (AY881626) strain. GenBank accession numbers are indicated in brackets. Bootstrap values (500 replicates) are indicated above the branches. The scale indicates number of substitutions per base position. .... 48

**Figure 3.4** Diagram to illustrate bioinformatics workflow used to analyze the Illumina short read sequence data. The various bioinformatic software tools and command used in the analyses are shown (**PE** Paired-end). .... 55

**Figure 4.1 a)** Schematic diagram of the genome organization of Grapevine virus E (SA94). **Mtr** methyltransferase, **Hel** helicase, **AlkB** AlkB conserved domain, **RdRp** RNA-dependant RNA polymerase, **MP** movement protein, **CP** coat protein, **NB** nucleic acid-binding protein, **?** protein with unknown function. **b)** MAQ-reassembly of metagenomic sequence reads on Grapevine virus E (SA94). Schematic representation of virus genome with numbered open reading frames is shown above graph. The four grey bars below graph highlight areas with no coverage. .... 62

---

---

## List of Tables

---

---

<b>Table 2.1</b>	Viruses reported to infect grapevine ( <i>Vitis</i> spp.).....	6
<b>Table 2.2</b>	Recent examples of viral metagenomic projects in different environments.....	13
<b>Table 2.3</b>	Comparison of the latest available next-generation sequencing platforms. Specifications for the Illumina Genome Analyzer II (used in this study) are included. (Data were obtained from the respective websites).....	21
<b>Table 2.4</b>	Examples of available <i>de novo</i> short read assemblers, mapping assemblers and alignment viewers. Websites are shown for more information on this software. ....	28
<b>Table 3.1</b>	Comparison of <i>de novo</i> and re-assembly data for the five dominant virus species identified in this study. <i>De novo</i> assembled scaffolds are classified according to best alignment (highest bit score) in the NCBI database found with BLASTn and BLASTx searches. MAQ re-assembly data are shown for the 23 representative variants identified after <i>de novo</i> assembly analysis. ....	45
<b>Table 4.1</b>	Genome position and size of open reading frames (ORFs) and untranslated regions (UTRs) of GVE SA94 and percentage nucleotide (amino acid in brackets) sequence identity to other members of the genus <i>Vitivirus</i> . ....	60

---

---

# Contents

---

---

<b>Declaration</b> .....	<b>i</b>
<b>Abstract</b> .....	<b>ii</b>
<b>Opsomming</b> .....	<b>iv</b>
<b>Abbreviations</b> .....	<b>vi</b>
<b>Acknowledgements</b> .....	<b>viii</b>
<b>List of Figures</b> .....	<b>x</b>
<b>List of Tables</b> .....	<b>xii</b>
<b>Contents</b> .....	<b>xiii</b>
<b>Chapter 1: Introduction</b> .....	<b>1</b>
1.1 Background and motivation for this study.....	1
1.2 Project proposal (Aims and Objectives).....	1
1.3 Chapter layout .....	2
1.4 References .....	4
<b>Chapter 2: Literature review</b> .....	<b>5</b>
2.1 Grapevine diseases and associated viruses in South Africa.....	5
2.1.1 Grapevine leafroll disease.....	7
2.1.2 Shiraz disease.....	7
2.1.3 Shiraz decline .....	8
2.1.4 Virus detection, prevention and novel virus discovery.....	9
2.2 Metagenomic sequencing.....	10
2.2.1 What is metagenomic sequencing? .....	10
2.2.2 Viral metagenomics .....	12
2.2.2.1 Viral enrichment .....	16
2.3. Next-generation sequencing.....	16
2.3.1 Introduction to next-generation sequencing.....	17
2.3.2 Comparison of three next-generation platforms.....	18
2.3.3 Technical overview of Illumina sequencing .....	21

2.3.4	Bioinformatic analysis of next-generation sequencing data .....	25
2.3.4.1	Bioinformatic challenges when dealing with next-generation sequencing data . .....	25
2.3.4.2	Bioinformatic tools for next-generation sequencing data .....	27
2.3.4.2.1	Velvet .....	29
2.3.4.2.2	MAQ.....	30
2.4	Conclusion.....	31
2.5	References .....	32
 <b>Chapter 3: Deep sequencing analysis of viruses infecting grapevines: Virome of a vineyard..</b>		<b>39</b>
3.1	Abstract.....	39
3.2	Introduction .....	39
3.3	Results.....	41
3.3.1	Sequencing.....	41
3.3.2	<i>De novo</i> sequence assembly and analysis.....	42
3.3.3	Re-assembly against reference sequences .....	44
3.4	Discussion .....	47
3.6	Conclusion.....	50
3.7	Materials and methods.....	52
3.7.1	Plant material.....	52
3.7.2	Sequencing.....	52
3.7.3	Sequence analysis .....	52
3.8	References .....	56
 <b>Chapter 4: The first complete nucleotide sequence of a Grapevine virus E variant.....</b>		<b>58</b>
 <b>Chapter 5: Conclusions .....</b>		<b>64</b>
 <b>Supplementary data 1 .....</b>		<b>67</b>
<b>Supplementary data 2 .....</b>		<b>68</b>
<b>Supplementary data 3 .....</b>		<b>120</b>
<b>Supplementary data 4 .....</b>		<b>122</b>

---

---

# Chapter 1: Introduction

---

---

## 1.1 Background and motivation for this study

Grapevine (*Vitis vinifera*) is one of the most widely grown crops in temperate climates (Martelli and Boudon-Padieu, 2006). In 2006 South Africa ranked as one of the ten largest wine producing countries in the world, producing 3% of the world's wine. More than 100 000 hectares of wine grape cultivars are under cultivation in South Africa and produced 1015,4 million liters of wine and grape juice in 2009 (WOSA:[http://www.wosa.co.za/sa/stats\\_worldwide.php](http://www.wosa.co.za/sa/stats_worldwide.php)). In 2008 the wine and related industries generated R26.2 billion of the country's gross domestic product and employed 275 000 people (SAWIS: <http://www.sawis.co.za/info/annualpublication.php>). Grapevine is therefore a valuable agricultural commodity and contributes significantly to the economy of the areas in which it is grown. This valuable crop plant is threatened by the 60 viruses known to infect grapevine (Martelli, 2009), and more suspected viral pathogens, reducing both crop yield and quality (Martelli and Boudon-Padieu, 2006). It is therefore an essential investment in the South African economy to study the viruses infecting grapevine.

The availability of next-generation sequencing platforms such as the Illumina, Roche/454 and ABI SOLiD, make it possible to study viral disease complexes using a metagenomic approach. These sequencing systems can sequence in parallel millions of DNA molecules, directly isolated from an environmental sample without the need for prior cloning. Recently, a number of papers reported on the use of next-generation sequencing analysis of viruses infecting crop plants (Adams et al., 2009; Al Rwahnih et al., 2009; Kreuze et al., 2009). These studies proved the use of next-generation sequencing technologies in metagenomic studies to identify the viral pathogens present and open the possibility to discover novel viruses.

## 1.2 Project proposal (Aims and Objectives)

This study aimed to evaluate the technique of metagenomic sequencing with next-generation sequencing technology using the Illumina Genome Analyzer II sequencing-by-synthesis technology to determine the viral profile of a diseased vineyard. The project focused on establishing the techniques for successful sequencing and acquiring the necessary skills and knowledge to perform bioinformatic analysis on the sequence data.



To achieve the proposed aim, the study was divided into several objectives:

- Identify diseased vineyard, harvest material from randomly select vines and extract dsRNA.
- Sequence dsRNA using the Illumina Genome Analyzer II (in collaboration with Prof. DJG Rees and Dr. J-M Celton at the University of Western Cape).
- Identify and implement suitable bioinformatic tools to analyze sequence data (in collaboration with Prof. DJG Rees at the University of Western Cape).
- Identify viruses present in the sample, determine prevalence and dominant variants of these viruses
- Identify novel viral pathogens.
- Further genetic characterization of novel viruses.

### **1.3 Chapter layout**

This thesis is divided into five chapters. Each of the chapters is separately introduced and a reference list provided.

#### **Chapter 1: Introduction**

This chapter provides a general introduction and motivation for the study. The aims and objectives of the study are stated.

#### **Chapter 2: Literature review**

In this chapter literature related to the project is reviewed. A brief overview of economical important viral diseases of grapevine and associated viruses in a South African context is presented. This is followed by a description of metagenomic sequencing, and specifically metagenomic projects studying viral communities. In the subsequent section next-generation sequencing is introduced, followed by a detailed description of the Illumina sequencing technology and the bioinformatic analysis of next-generation sequencing data.

### **Chapter 3: Deep sequencing analysis of viruses infecting grapevines: Virome of a vineyard**

This chapter describes the use of next-generation sequencing technology to elucidate disease etiology in grapevine and further extent the use for novel virus discovery. The results presented here highlight the applicability of Illumina short read sequencing to provide a comprehensive snapshot of the viral complement of a diseased vineyard.

The work described in this chapter is published as a peer-reviewed paper:

*Coetzee, B., Freeborough, M.-J., Maree, H.J., Celton, J.-M., Rees, D.J.G., Burger, J.T., 2010. Deep sequencing analysis of viruses infecting grapevines: Virome of a vineyard. Virology 400, 157-163.*

Additionally to the published paper, a diagram describing the bioinformatic workflow used to analyze the sequencing data, is provided.

### **Chapter 4: The first complete nucleotide sequence of a Grapevine virus E variant**

This chapter describes the genomic characterization of a South African variant of Grapevine virus E, a virus for the first time detected in South African vineyards by the metagenomic study (described in chapter 3).

The work described in this chapter is published as a peer-reviewed paper:

*Coetzee, B., Maree, H.J., Stephan, D., Freeborough, M.-J., Burger, J.T., 2010. The first complete nucleotide sequence of a Grapevine virus E variant. Arch. Virol. 155, 1357-1360.*

### **Chapter 5: Conclusions**

In this chapter the final conclusion and further prospects of this study are discussed.

## 1.4 References

- Adams, I.P., Glover, R.H., Monger, W.A., Mumford, R., Jackeviciene, E., Navalinskiene, M., et al., 2009. Next-generation sequencing and metagenomic analysis: a universal diagnostic tool in plant virology. *Mol. Plant Pathol.* 10, 537-545.
- Al Rwahnih, M., Daubert, S., Golino, D., Rowhani, A., 2009. Deep sequencing analysis of RNAs from a grapevine showing Syrah decline symptoms reveals a multiple virus infection that includes a novel virus. *Virology* 387, 395–401.
- Kreuze, J.F., Perez, A., Untiveros, M., Quispe, D., Fuentes, S., Barker, I., et al., 2009. Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: a generic method for diagnosis, discovery and sequencing of viruses. *Virology* 388, 1-7.
- Martelli, G.P., Boudon-Padieu, E., 2006. Directory of infectious diseases of grapevines and viroses and virus-like diseases of the grapevine: Bibliographic report 1998-2004. *Options Méditerran., Ser. B, Stud. Res.* 55, CIHEAM, 279.
- Martelli, G.P., 2009. Grapevine virology highlights 2006-2009. 16th meeting of the International Council for the study of virus and virus-like diseases of the grapevine, 15-23.

### Internet resources

- Wines of South Africa (WOSA):** [http://www.wosa.co.za/sa/stats\\_worldwide.php](http://www.wosa.co.za/sa/stats_worldwide.php) [accessed 30.03.2010]
- South African Wine Industry Information and Systems (SAWIS):**  
<http://www.sawis.co.za/info/annualpublication.php> [accessed 30.03.2010]

---

---

## Chapter 2: Literature review

---

---

This chapter presents a broad overview of the current literature relevant to this project. A brief overview is given of economically important grapevine disease complexes and associated viruses in South Africa and virus detection techniques. In the subsequent section, metagenomic sequencing is discussed with specific reference to viral metagenomic projects. This is followed by an introduction to next-generation sequencing technology and a comparison of the three main sequencing platforms. A more detailed description is given of the Illumina sequencing technology. The chapter is concluded with a discussion of the bioinformatic challenges analyzing next-generation sequencing data, and reference to specific bioinformatic software tools used in our analysis.

### **2.1 Grapevine diseases and associated viruses in South Africa**

Viruses pose a significant threat to grapevine and therefore to the wine industry. Grapevine is the perennial crop plant known to be infected by the highest number of viruses. Sixty viruses have been identified to date, with more viruses suspected to infect the plant (Martelli, 2009). Viruses negatively affect the physiology of grapevine, therefore reducing the vigour of the plant and shortening the productive life of the vineyard. Viral infection decreases both the quality and quantity of crop yield (Martelli and Boudon-Padieu, 2006). In the South African context, leafroll disease, Shiraz disease (SD) and Shiraz decline are the predominant virus-associated diseases observed in the fields. Table 2.1 presents a list of viruses known to infect grapevine.

**Table 2.1** Viruses reported to infect grapevine (*Vitis* spp.).

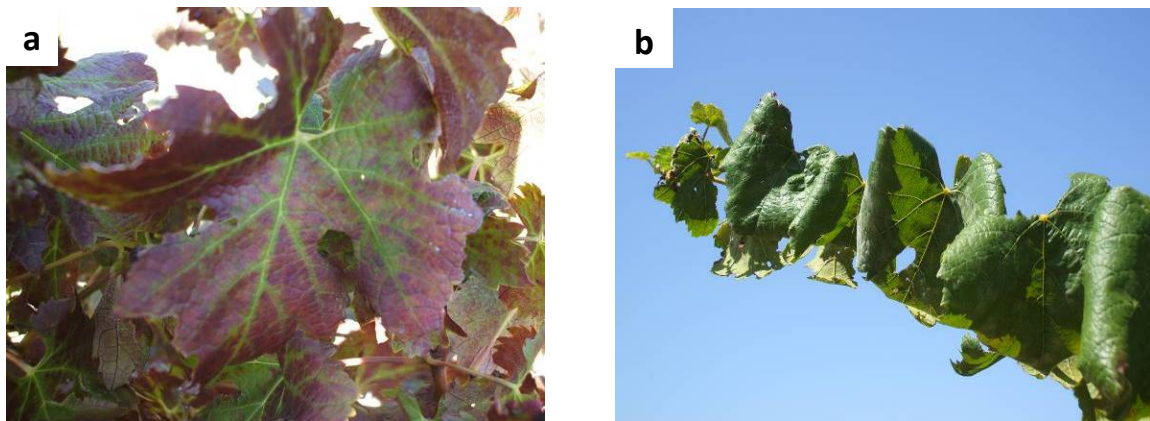
Family	Genus	Species <sup>a</sup>	
<b>Alfalexiviridae</b>	<i>Potexvirus</i>	<i>Potato virus X</i> (PVX)	
<b>Betaflexiviridae</b>	<i>Foveavirus</i>	<i>Grapevine rupestris stem pitting-associated virus</i> (GRSPaV)	
	<i>Trichovirus</i>	<i>Grapevine berry inner necrosis virus</i> (GINV)	
	<i>Vitivirus</i>	<i>Grapevine virus A</i> (GVA)	
		<i>Grapevine virus B</i> (GVB)	
		<i>Grapevine virus D</i> (GVD)	
		<i>Grapevine virus E</i> (GVE)	
<b>Bromoviridae</b>	<i>Alfamovirus</i>	<i>Alfalfa mosaic virus</i> (AMV)	
	<i>Cucumovirus</i>	<i>Cucumber mosaic virus</i> (CMV)	
	<i>Ilarvirus</i>	<i>Grapevine line pattern virus</i> (GLPV)	
		<i>Grapevine angular mosaic virus</i> (GAMoV)	
<b>Bunyaviridae</b>	<i>Tospovirus</i>	<i>Tomato spotted wilt virus</i> (TSWV)	
<b>Closteroviridae</b>		<i>Grapevine leafroll-associated virus 1</i> (GLRaV-1)	
		<i>Grapevine leafroll-associated virus 3</i> (GLRaV-3)	
		<i>Grapevine leafroll-associated virus 4</i> (GLRaV-4)	
		<i>Grapevine leafroll-associated virus 5</i> (GLRaV-5)	
		<i>Grapevine leafroll-associated virus 6</i> (GLRaV-6)	
		<i>Grapevine leafroll-associated virus 7</i> (GLRaV-7)	
		<i>Grapevine leafroll-associated virus 9</i> (GLRaV-9)	
		<i>Closterovirus</i>	<i>Grapevine leafroll-associated virus 2</i> (GLRaV-2)
	<b>Secoviridae</b>	<i>Fabavirus</i>	<i>Broad bean wilt virus</i> (BBWV)
<b>(Subfamily Comovirinae)</b>	<i>Nepovirus</i> : Subgroup A	<i>Arabidopsis mosaic virus</i> (ArMV)	
		<i>Grapevine deformation virus</i> (GDefV)	
		<i>Grapevine fanleaf virus</i> (GFLV)	
			<i>Raspberry ringspot virus</i> (RpRSV)
			<i>Tobacco ringspot virus</i> (TRSV)
	<i>Nepovirus</i> : Subgroup B	<i>Artichoke Italian latent virus</i> (AILV)	
		<i>Grapevine Anatolian ringspot virus</i> (GARSV)	
		<i>Grapevine chrome mosaic virus</i> (GCMV)	
			<i>Tomato black ring virus</i> (TBRV)
	<i>Nepovirus</i> : Subgroup C	<i>Blueberry leafmottle virus</i> (BLMoV)	
		<i>Cherry leafroll virus</i> (CLRV)	
		<i>Grapevine Tunisian ringspot virus</i> (GTRSV)	
		<i>Grapevine Bulgarian latent virus</i> (GBLV)	
		<i>Peach rosette mosaic virus</i> (PRMV)	
		<i>Tomato ringspot virus</i> (ToRSV)	
<i>Sadwavirus</i>	<i>Strawberry latent ringspot virus</i> (SLRSV)		
<b>Tombusviridae</b>	<i>Carmovirus</i>	<i>Carnation mottle virus</i> (CarMV)	
	<i>Necrovirus</i>	<i>Tobacco necrosis virus D</i> (TNV-D)	
	<i>Tombusvirus</i>	<i>Grapevine Algerian latent virus</i> (GALV)	
		<i>Petunia asteroid mosaic virus</i> (PAMV)	
<b>Tymoviridae</b>	<i>Marafivirus</i>	<i>Grapevine asteroid mosaic-associated virus</i> (GAMaV)	
		<i>Grapevine rupestris vein feathering virus</i> (GRVFV)	
		<i>Grapevine Syrah virus 1</i> (GSyV-1)	
	<i>Maculavirus</i>	<i>Grapevine fleck virus</i> (GFkV)	
		<i>Grapevine redglobe virus</i> (GRGV)	
<b>Virgviridae</b>	<i>Tobamovirus</i>	<i>Tobacco mosaic virus</i> (TMV)	
		<i>Tomato mosaic virus</i> (ToMV)	
<b>Unassigned genera</b>	<i>Idaeovirus</i>	<i>Raspberry bushy dwarf virus</i> (RBDV)	
	<i>Sobemovirus</i>	<i>Sowbane mosaic virus</i> (SoMV)	
<b>Unassigned viruses</b>		<i>Grapevine Ajinashika virus</i> (GAGV)	
		<i>Grapevine stunt virus</i> (GSV)	
		<i>Grapevine labile rod-shaped virus</i> (GLRSV)	

<sup>a</sup> Scientific names of definite viruses species are written in italics, names of tentative species are written in Roman characters. Adapted from: Martelli and Boudon-Padieu, 2006.

Updated virus taxonomy: International Committee on Taxonomy of Viruses -Virus Taxonomy: 2009 Release (<http://www.ictvonline.org/virusTaxonomy.asp?version=2009&bhcp=1>)

### 2.1.1 Grapevine leafroll disease

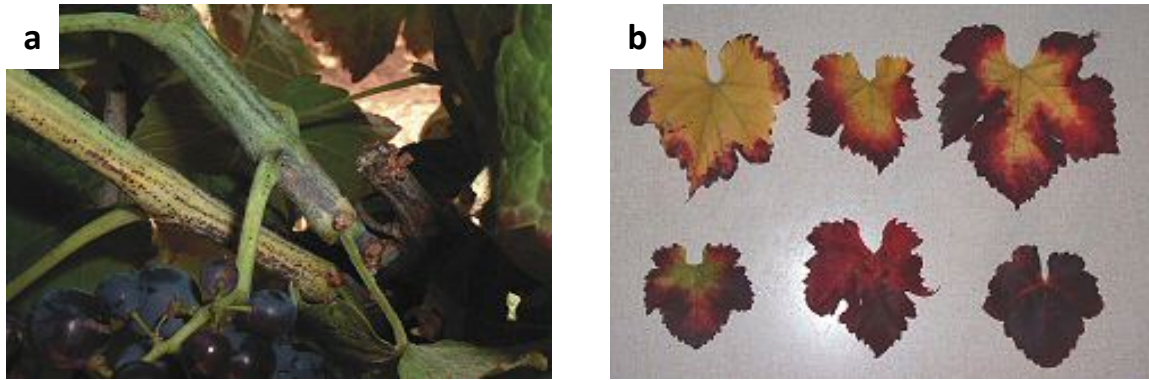
Grapevine leafroll disease is recognized as the most commonly occurring viral disease in South African vineyards (Pietersen, 2000) and worldwide (Martelli and Boudon-Padieu, 2006). There are currently up to 10 viruses recognized to be associated with grapevine leafroll disease, with Grapevine leafroll associated virus-3 (GLRaV-3) regarded as the most important (Pietersen, 2000). Most of the grapevine leafroll associated viruses are classified in the family *Closteroviridae*, genus *Ampelovirus*. Grapevine leafroll associated virus-3 is a phloem-limited virus causing the degradation of the vascular tissue (Karasev, 2000) and resulting in typical leafroll disease symptoms (Figure 2.1). In both red and white cultivars the leaf margins roll downwards. The leaves of red cultivars turn prematurely red, while the veins remain green. In white cultivars the interveinal regions turn yellow. The berry quality is also negatively affected with delayed ripening and lower sugar concentrations (Pietersen, 2000).



**Figure 2.1** Grapevine with typical leafroll symptoms **a)** Red cultivar displaying interveinal reddening **b)** White cultivar with leaves rolled downwards.

### 2.1.2 Shiraz disease

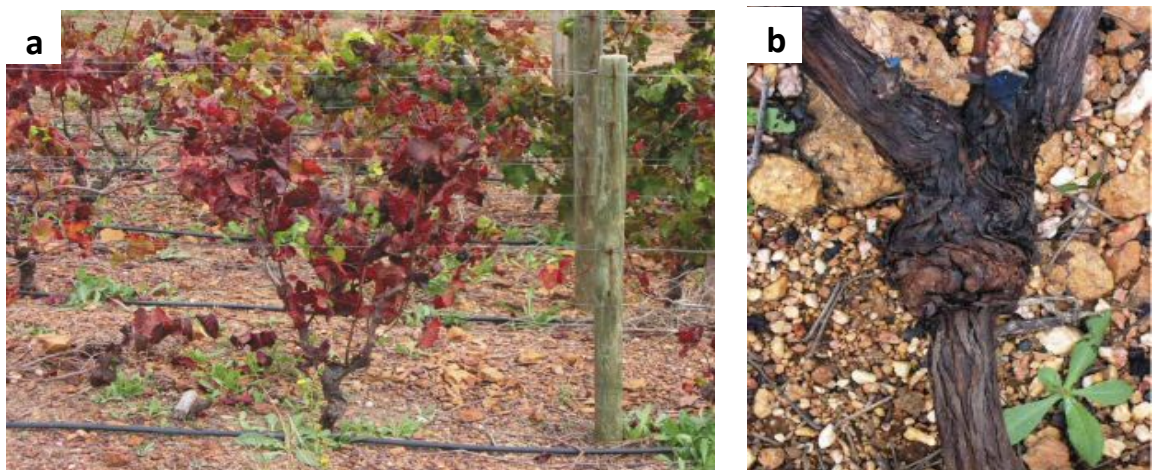
To date, Shiraz disease occurs only in South Africa. More susceptible cultivars, Shiraz, Merlot, Gamay, Malbec and Viognier develop typical symptoms (Figure 2.2), whereas in other cultivars the disease remains latent (Goszczyński et al., 2008). Infected vines display a lack of lignification, giving them their characteristic rubbery appearance. These vines show a reduction in vigour, never fully mature and usually die within 3 to 5 years. Leaves have a typical discoloration pattern, turning red from the outside edges to a complete discoloration, and leaf-fall is severely delayed. Infected vines have small bunches with reduced berry set, resulting in yield loss. Sugar concentration in these berries is lower (Goussard and Bakker, 2000). Three divergent molecular groups of Grapevine virus A (GVA) were identified in South Africa (Goszczyński and Jooste, 2003), of which variants of molecular group II were shown to be associated with Shiraz disease (Goszczyński, 2007b; Goszczyński et al., 2008).



**Figure 2.2** Grapevine with Shiraz disease symptoms **a)** Green shoots with a lack of lignification **b)** Typical Shiraz disease leaf discoloration patterns. Leaf edges start to turn red progressing to completely red leaves ([www.wynboer.co.za](http://www.wynboer.co.za)).

### 2.1.3 Shiraz decline

Symptoms of this disease include swelling of the graft union with thickened bark on and above the union. Deep grooving of the stems and premature red discoloration of the leaves from middle to late summer can be observed (Figure 2.3). Infected vines have reduced vigour and usually die within 5 to 10 years. Due to the reduced vigour the fruit yield from these vines is negatively affected (Al Rwahnih et al., 2009; Battany et al., 2004; Goszczynski, 2007a). Grapevine rupestris stem pitting-associated virus (GRSPaV) has been associated with the disease in other parts of the world (Habibi et al., 2006; Lima et al., 2006) and in South Africa (Goszczynski, 2007a). Al Rwahnih et al. (2009) suggested that three viruses might be the causal agents of Shiraz decline, Grapevine rupestris stem pitting-associated virus (GRSPaV), Grapevine rupestris vein-feathering virus (GRVfV) and the recently described Grapevine Syrah virus-1 (GSyV-1).



**Figure 2.3** Typical Shiraz decline symptoms **a)** Reduced vigour and premature red discoloration of leaves **b)** Swelling at the graft union ([www.wynboer.co.za](http://www.wynboer.co.za)).

#### **2.1.4 Virus detection, prevention and novel virus discovery**

To date, the grapevine plant has no known natural resistance to viruses and there is no known cure for virus infection. Currently, more tolerant cultivars or clones are used to limit the impact of viral diseases. The use of transgenic plants would be a further step to reduce the harmful effects of viral infection.

Grapevine viruses mainly are transmitted by insect vectors such as mealybugs, aphids and nematodes, but also mechanically by workers and implements. It is therefore essential to maintain proper vineyard sanitation to limit the spread of viral diseases. Insecticides are commonly used to control insects in the vineyard. Virus infected plants must be removed from the vineyard and proper quarantine methods maintained to prevent planting of infected propagation material. It is therefore essential to have sensitive and rapid detection methods to test for commonly occurring viruses (Martelli and Boudon-Padieu, 2006), but also to have techniques available to detect new emerging viruses.

Presently, the routine methods used to screen for viruses are enzyme-linked immunosorbent assay (ELISA) and reverse transcription polymerase chain reaction (RT-PCR). ELISA is a serological detection method relying on the interaction of the viral antigen and specific antibodies. Molecular techniques such as RT-PCR target the genetic material of the virus and rely on the amplification of a region of the viral genome using specific primers. Both these detection methods have the limitation that prior knowledge of the virus(es) present is necessary. Furthermore, they target viruses historically associated with the different grapevine diseases, therefore limiting their scope to discover novel viruses involved in the etiology of these diseases (Adams et al., 2009).

Conventional viral discovery rely on physical and/or biological characterization with techniques such as electron microscopy and indicator plants or nucleic acid based detection of novel viral pathogens (Kreuze et al., 2009), the non-specific amplification of viral nucleic acid, cloning and sequencing of a number of these clones. These techniques might be time-consuming and labour intensive. Virus detection is further complicated by mixed infections. Although a number of viruses have been shown to be associated with the respective diseases discussed above, viral diseases are often caused by virus complexes with more than one virus infecting a single plant. Mixed infections can play a role in increased disease severity and enhanced symptom



expression (Prosser et al., 2007), limiting the applicability of conventional techniques to reveal the complete etiology of a virus disease complex.

## **2.2 Metagenomic sequencing**

In the light of the above mentioned limitations of current grapevine virus diagnostic and detection techniques, a metagenomic approach by sequencing the total viral complement, or virome, of a diseased vineyard might circumvent those limitations.

### **2.2.1 What is metagenomic sequencing?**

Traditionally, microbiology focussed on studying single organisms. Single organisms are isolated from the environment and cultured *in vitro* to obtain pure cultures. Genetic material from these pure cultures can be isolated, sequenced and analyzed. Besides the obvious disadvantages, it is laborious and time-consuming; this technique has the added drawback of being culture-dependant. A large percentage of microbes in environmental samples cannot be cultured using standard culturing techniques, limiting the portion of genetic diversity present in an environment that can be studied and exploited (Handelsman, 2004; Hugenholtz and Tyson, 2008; Riesenfeld et al., 2004).

The term “metagenomics” was first used in 1998 (Handelsman et al., 1998) to describe the study of the collective genetic material from all microbes in a specific environment. Since metagenomics involve the cloning of genetic material isolated directly from the environment, it circumvents the need to isolate and cultivate organisms and is thus more time and cost-effective. Metagenomics also allows for the study of microorganisms in their natural environment and is not biased towards culturable organisms, therefore the total genetic diversity of microorganisms can be studied (Jones, 2010; Streit and Schmitz, 2004; Wooley et al., 2010). This collective genetic pool of microorganisms in an environment is called the metagenome (Kowalchuk et al., 2007; Schloss and Handelsman, 2005). Since the study field of metagenomics became popular, other synonymous terms have also been used in literature: environmental genomics, community genomics, population genomics (Handelsman, 2004) and ecological genomics (Xu, 2006).

Traditionally, metagenomic studies were conducted by doing environmental shotgun sequencing (ESS) or random shotgun sequencing (Wooley et al., 2010). The first step is to extract the genetic material, usually DNA, directly from an environmental sample (e.g. soil or water). The genetic material is then sheared into random fragments, cloned into vectors and used to transform suitable host cells to produce metagenomic libraries consisting of clones containing inserts of the environmental DNA. These libraries are either used for sequence-based or functional analysis. In sequence-based analysis, the clones are sequenced using Sanger sequencing. Clones can either be sequenced at random and computer software used to assemble the sequenced fragments into whole genomes, or clones containing a phylogenetic “signature” region such as 16S rRNA genes are sequenced to give an indication of the species present in the sample. In functional analysis, the transformed libraries are screened for the expression of specific proteins (Handelsman, 2004; Streit and Schmitz, 2004). For more details the reader is referred to a number of papers discussing metagenomics and metagenomic sequencing: Cardenas and Tiedje, 2008; Deutschbauer et al., 2006; Green and Keller, 2006; Guazzaroni et al., 2009; Handelsman, 2004; Kowalchuk et al., 2007; Raes et al., 2007; Riesenfeld et al., 2004; Schloss and Handelsman, 2005; Simon and Daniel, 2009; Snyder et al., 2009; Streit and Schmitz, 2004; Tringe and Rubin, 2005; Whitaker and Banfield, 2006; Wooley and Ye, 2009; Xu, 2006.

While the metagenomic approaches described above are highly effective in characterizing the microbial diversity present in a sample, the laborious and costly process of cloning is still necessary. Currently, next-generation sequencing technologies (discussed in section 2.3) opens the possibility to study microbial communities through direct sequencing of the environmental genetic material (Hall, 2007), circumventing the need for an initial cloning step. This sequencing technology is a fast high-throughput technique for sequencing DNA and thus more suitable for metagenomic sequencing than conventional Sanger sequencing (Cardenas and Tiedje, 2008). It is not biased towards any specific microbial group and does not rely on known sequence information and therefore has the potential to discover new organisms that are highly divergent from those already known (Snyder et al., 2009).

### **2.2.2 Viral metagenomics**

Traditionally, discovering novel viruses was dependant on the ability to culture the viruses in cell culture systems and to isolate pure virus particles for characterization. This is hampered by the fact that many microorganisms, and by extension their viruses, cannot be cultured using standard cell lines and techniques and further complicated by the low nucleic acid content of viruses. Additionally, viruses do not have conserved genetic elements that can be used to design sequencing primers and assess the diversity of a viral population (Bench et al., 2007; Thurber et al., 2009; Zhang et al., 2006).

Recent advances in sequencing and other molecular technologies facilitated viral metagenomic studies in a broad range of natural environments. Assessing the viral community through metagenomic techniques can provide insights to the community structure and diversity of viruses in a natural environment. These techniques have already been exploited by several projects. The largest metagenomics project to date surely is the global ocean sampling (GOS) expedition; collecting and sequencing material from many different oceans (Rusch et al., 2007; Venter et al., 2004) resulting in the characterization of the marine viruses of these oceans (Williamson et al., 2008). Table 2.2 presents a list of viral metagenomic projects most cited in recent literature. Viral metagenomics are reviewed in a number of papers: Allen and Wilson, 2008; Delwart, 2007; Edwards and Rohwer, 2005; Kristensen et al., 2010; Schoenfeld et al., 2010; Suttle, 2005; Thurber et al., 2009, and human viruses specifically: Tang and Chiu, 2010.

These studies prove that viral metagenomics can be an effective method for direct characterization of the virome of an environmental sample, providing valuable information on viral community structure and diversity and enabling the discovery of novel viruses with little or no sequence similarity to known viruses. Furthermore, this approach can be applied to a wide range of environmental samples. However, what was evident from these studies is that a large portion of the metagenomic sequences did not show significant similarity to sequences in databases, and therefore remained unassigned, showing our limited knowledge of the total scope of viral diversity present on earth (Edwards and Rohwer, 2005).

**Table 2.2** Recent examples of viral metagenomic projects in different environments.

Sample type	Sampling location	Viral enrichment	Nucleic acid extracted	Sequencing process <sup>a</sup>	Major findings/ novel viruses detected	Reference
Marine water	La Jolla, California and Mission Bay, San Diego	Filtration, density-dependent centrifugation	DNA	Sanger sequencing	> 65% of sequences not significantly similar to database sequences; high viral diversity.	Breitbart et al., 2002
Human faeces	Data not available	Filtration, CsCl gradient centrifugation	DNA	Sanger sequencing	Most sequences unrelated to sequences in databases; siphophages most common.	Breitbart et al., 2003
Marine sediment	Mission Bay, San Diego, USA	Filtration, CsCl gradient centrifugation	DNA	Sanger sequencing	75% of sequences not related to database sequences; high viral diversity found, dsDNA phages most abundant.	Breitbart et al., 2004
Equine faeces	Data not available	Filtration, nuclease treatment	DNA	Sanger sequencing	Only 32% of sequences could be classified; hundreds of uncharacterized viruses detected.	Cann et al., 2005
Human blood	San Diego, California, USA	CsCl gradient centrifugation, nuclease treatment	DNA	Sanger sequencing	Both ssDNA and dsDNA viruses could be recovered from blood sample; presence of novel anellovirus.	Breitbart and Rohwer, 2005
Marine water	Sargasso sea; Gulf of Mexico; British Columbia coast and Artic ocean	Filtration, density-dependant centrifugation	DNA	Pyrosequencing	Novel single -stranded DNA chp1-like microphage found.	Angly et al., 2006
Marine water	English Bay and Strait of Georgia, British Columbia, Canada	Filtration	RNA	Sanger sequencing	High viral diversity detected; genomes assembled of several previously unknown RNA viruses.	Culley et al., 2006
Human faeces	San Diego, California, USA	Filtration	RNA	Sanger sequencing	Plant pathogenic viruses were most abundant.	Zhang et al., 2006
Marine water	Chesapeake Bay	Filtration	DNA	Sanger sequencing	High portion of unknown and novel sequences, cyanophages most abundant.	Bench et al., 2007
Soil from desert, prairie and rainforest	Peru; California; Kansas	Filtration, CsCl gradient centrifugation	DNA	Sanger sequencing	Soil viruses are taxonomicly diverse and distinct from viral communities in other environments.	Fierer et al., 2007
Human faeces (infant)	Data not available	Filtration, CsCl gradient centrifugation	DNA	Sanger sequencing	Environment dominated by phages; most sequences not similar to database sequences.	Breitbart et al., 2008
Stromatolites and thrombolites	Mexico and Bahamas	Filtration	DNA	Pyrosequencing	>97% of recovered sequences remained unknown; phage genotypes are geographically restricted.	Desnues et al., 2008
Human faeces	Melbourne, Australia and Seattle, USA	Filtration	RNA	Micro-mass sequencing	Known entric viruses and putative novel viruses detected.	Finkbeiner et al., 2008
Human faeces	South Asia	Filtration, nuclease treatment	Total nucleic acid	Sanger sequencing	A previously unreported genus of the <i>Picornaviridae</i> family was detected.	Kapoor et al., 2008
Soil from rice paddy	Deajeon, Korea	Filtration, nuclease treatment	DNA	Sanger sequencing	More than 60% of sequences did not show significant similarity to database sequences; putative novel ssDNA virus.	Kim et al., 2008

<sup>a</sup> Pyrosequencing refers to next-generation sequencing  
CsCl cesiumchloride

**Table 2.2 continued** Recent examples of viral metagenomic projects in different environments.

Sample type	Sampling location	Viral enrichment	Nucleic acid extracted	Sequencing process <sup>a</sup>	Major findings/ novel viruses detected	Reference
<i>Diploria strigosa</i> (coral)	Mout Irvine Bay, Brucchoo, Tobago	CsCl gradient centrifugation, nuclease treatment	DNA	Sanger sequencing	Herpes-like sequences detected; cyanophages were most abundant phages.	Marhaver et al., 2008
Marine water	Tampa Bay	Filtration, CsCl gradient centrifugation	Total nucleic acid	Pyrosequencing	6.6% of sequence reads were identifiable; virusintergrase genes are present.	McDaniel et al., 2008
<i>Ambrosia psilotachya</i> (western ragweed)	Tallgrass Prairie Preserve, Oklahoma, USA	Ultracentrifugation	Total nucleic acid	Sanger sequencing	Evidence for novel viruses belonging to the families <i>Caulimoviridae</i> and <i>Flexiviridae</i> .	Melcher et al., 2008
Cerebrospinal fluid (organ transplant patients)	Australia	Nuclease treatment	RNA	Pyrosequencing	Presence of novel Arenavirus found.	Palacios et al., 2008
Hot springs	Yellowstone, USA	Filtration	DNA	Sanger sequencing	High viral diversity found.	Schoenfeld et al., 2008
<i>Porites compressa</i> (finger coral)	Hawaii	Data not available	Data not available	Pyrosequencing	Stressors induce production of herpes-like viruses in coral.	Vega Thurber et al., 2008
Marine water	37 sites along a transect from Halifax, Nova Scotia through the South Pacific Gyre	Filtration	DNA	Sanger sequencing	High viral diversity, most abundant bacteriophage is related to the cyanomyovirus P-SSM4.	Williamson et al., 2008
Tomato, <i>Liatrix spicata</i>	Poland	Data not available	RNA	Pyrosequencing/ Sanger sequencing	Novel cucumovirus (Gayfeather mild mottle virus) detected.	Adams et al., 2009
<i>Vitis vinifera</i> (Grapevine)	California, USA	Data not available	dsRNA and total nucleic acid	Pyrosequencing	Novel marafivirus (Grapevine Syrah virus-1) detected.	Al Rwahnih et al., 2009
Human faeces	Pakistan and Afghanistan	Filtration, nuclease treatment	RNA and DNA	Sanger sequencing	Divergent strains of Saffoldvirus (SAFV) detected.	Blinkova et al., 2009
Human liver and serum (hemorrhagic fever patients)	Lusaka, Zambia and Johannesburg, South Africa	Nuclease treatment	RNA	Pyrosequencing	A novel hemorrhagic fever-associated Arenavirus (Lujo Virus) detected and characterized.	Briese et al., 2009
Fresh water lake	Maryland, USA	Filtration and nuclease treatment	RNA	Pyrosequencing and Sanger sequencing	Majority of sequences did not show significant similarity to database sequences; 30 viral families and previously unknown dsRNA virus (related to Banna virus) detected.	Djikeng et al., 2009
Sweetpotato	Lima, Peru	Data not available	RNA	Illumina/Solexa sequencing	Establish the use of deep sequencing for virus detection and diagnosis in plants.	Kreuze et al., 2009

<sup>a</sup> Pyrosequencing refers to next-generation sequencing  
CsCl cesiumchloride

**Table 2.2 continued** Recent examples of viral metagenomic projects in different environments.

Sample type	Sampling location	Viral enrichment	Nucleic acid extracted	Sequencing process <sup>a</sup>	Major findings/ novel viruses detected	Reference
Antarctic lake	Livingston Island, Antarctica	Data not available	Data not available	Pyrosequencing	Environment dominated by eukaryotic infecting viruses; previously undescribed ssDNA viruses detected; high taxonomic diversity found.	Lopez-Bueno et al., 2009
Various plant species	Tallgrass Prairie Preserve, Oklahoma, USA	Virus-like particles prepared by ultracentrifugation	Total nucleic acid	Sanger sequencing	Evidence of a novel member of the genus <i>Tymovirus</i> .	Muthukumar et al., 2009
Human nasal and faecal specimens	Osaka, Japan	Data not available	RNA	Pyrosequencing	Establish the use of high-throughput sequencing for direct metagenomic detection of pathogens.	Nakamura et al., 2009
Sea turtle fibropapilloma	Florida, USA	Filtration, CsCl gradient centrifugation, nuclease treatment	DNA	Sanger sequencing	Novel single-stranded DNA virus (sea turtle tornovirus 1) was discovered.	Ng et al., 2009
Reclaimed water	Florida, USA	Filtration, density-dependant centrifugation, nuclease treatment	DNA and RNA	Pyrosequencing	Phages most common; novel eukaryotic viruses and plant pathogens detected; no established human pathogens detected.	Rosario et al., 2009
Human faeces (children with acute flaccid paralysis)	South Asia	Filtration, nuclease treatment	DNA and RNA	Pyrosequencing and Sanger sequencing	Presence of bacteriophages, plant viruses and potential novel Enterovirus detected.	Victoria et al., 2009
Human respiratory tract specimens (cystic fibrosis patients)	San Diego, California, USA	Filtration, CsCl gradient centrifugation, nuclease treatment	DNA	Pyrosequencing	High viral diversity, majority of which were previously uncharacterized.	Willner et al., 2009
Chimpanzee faeces	Africa and USA	Filtration, nuclease treatment	Total nucleic acid	Sanger sequencing	Novel circular DNA virus (Chimpanzee stool-associated circular virus (ChiSSV)) detected.	Blinkova et al., 2010
Human and chimpanzee faeces and farm animals	Africa; Middle east; USA	Filtration, nuclease treatment	Total nucleic acid	Pyrosequencing	Members of novel genus <i>Cyclovirus</i> (in <i>Circoviridae</i> family) detected.	Li et al., 2010a
Bat guano	San Saba, Texas; Northern California	Filtration, nuclease treatment	DNA and RNA	Pyrosequencing	Presence of dietary viruses from plants and insects; novel mammal infecting viruses detected, no human pathogens.	Li et al., 2010b
Activated sludge	Auburn, Alabama, USA	Beef buffer treatment, CsCl gradient centrifugation, nuclease treatment	DNA	Sanger sequencing	Phages dominated the viral component of sequences, genes involved in macromolecule metabolism are most abundant	Parsley et al., 2010
Plants	Oklahoma and Costa Rica	Data not available	dsRNA	Pyrosequencing	Putative new plant viruses detected.	Roossinck et al., 2010

<sup>a</sup> Pyrosequencing refers to next-generation sequencing  
CsCl cesiumchloride

### **2.2.2.1 Viral enrichment**

A disadvantage of shotgun sequencing is that it cannot distinguish between target, host and other contaminating nucleic acids (Delwart, 2007), complicating the subsequent bioinformatics analyses and lowering the number of target sequences obtained. Therefore, a pre-enrichment step is usually performed to increase the concentration of target organism nucleic acids in the sample (Kreuze et al., 2009). In viral metagenomics, filtration is often used to select viral particles based on size and exclude bacteria and other larger organisms and reduce host contamination (see Table 2).

When working with grapevine viruses, double-stranded RNA (dsRNA) extraction is an effective way to enrich the sample for virus-specific nucleic acids (see Al Rwahnih et al., 2009 and Roossinck et al., 2010 for plant virus metagenomic studies performing dsRNA extraction). All known grapevine infecting viruses are single-stranded RNA viruses, having either positive or negative RNA strands as genomes. Upon entering the host, the virus' RNA is translated into proteins by the host's enzymes. The virus replicates by forming a complementary RNA strand on the existing template by means of a virus encoded RNA-dependant RNA polymerase. The virus replicative cycle therefore involves a dsRNA intermediate molecule. The presence of dsRNA is regarded as an indication of a viral infection, as RNA from the host plant normally does not exist in a double-stranded form (Dodds et al., 1984). Thus, by applying a procedure for dsRNA enrichment, a large proportion of extracted nucleic acids will be virus-derived. However, this will exclude all DNA viruses, of which there are to date none known to infect grapevine. Additionally, RNA viruses with a low dsRNA concentration at the time of extraction might be overlooked.

## **2.3. Next-generation sequencing**

The now common technique of DNA sequencing, determining the sequence of the four nucleotide types in a given DNA string, was first developed in 1977 by Frederick Sanger (Sanger et al., 1977), for which he later received the Nobel Prize. Another, but less known, sequencing technique was also developed in the seventies, and is described as the "chemical cleavage" method by Allan Maxam and Walter Gilbert (Maxam and Gilbert, 1977).

Sanger sequencing, also known as dideoxy- or chain terminating sequencing, currently marketed by Applied Biosystems, is still the most widely used sequencing technique and is described as the “gold standard” for sequencing in terms of accuracy and read length (Harismendy et al., 2009; Morozova and Marra, 2008). It is based on the use of labelled dideoxynucleotides. Since the development of Sanger sequencing more than 30 years ago, the sequencing technique underwent major improvements. Initially, the ddNTPs were radioactively labelled, electrophoresis done on polyacrylamide slab-gels, and the sequence read by hand from exposed film. Today the four ddNTPs are differentially fluorescently labelled and detected by an automated detector. Microcapillaries are used for electrophoretic sizing and base calling is automated (Harismendy et al., 2009; Morozova and Marra, 2008; Shendure and Ji, 2008; Strausberg et al., 2008).

Although Sanger sequencing is still the most commonly used sequencing technique, it has a drawback: pre-sequencing amplification and electrophoretic sizing of DNA fragments is needed (Millar et al., 2008; Varshney et al., 2009), limiting data throughput of this technology. In recent years, so called “second generation” or “next-generation” sequencing (NGS) instruments have been developed, breaking these limitations of Sanger sequencing.

### **2.3.1 Introduction to next-generation sequencing**

The first NGS platform became commercially available in 2005, the FLX Genome Sequencer by 454 Life Sciences (Margulies et al., 2005; <http://454.com>), now owned by the F. Hoffmann- La Roche group (Switzerland). In January 2007 Illumina Inc. (San Diego, California, USA) launched their Genome Analyzer I (Bentley, 2006; <http://www.illumina.com>) and the Applied Biosystems (ABI) SOLiD (<http://www3.appliedbiosystems.com>) became available in October of the same year. These three platforms currently dominate the sequencing arena, although third generation sequencing technologies, for example the HelicoScope from Helicos BioSciences (<http://www.helicosbio.com>), the Polonator instrument ([www.polonator.org](http://www.polonator.org)), “single molecule real time” sequencing from Pacific Biosciences ([www.pacificbiosciences.com](http://www.pacificbiosciences.com)) and nanopore sequencing (<http://mcb.harvard.edu/branton/projects-NanoporeSequencing.htm>), are in the developing phase or have already been commercially released.



Next-generation sequencing reduced dramatically the time and cost of sequencing (Harismendy et al., 2009; Mardis, 2008), a single instrument like the Illumina HiSeq 2000 can now generate 200 gigabases of sequencing data per week ([www.illumina.com](http://www.illumina.com)) for a fraction of the price of Sanger sequencing. Besides the parallel sequencing, increasing the data output, DNA is sequenced directly and there is no requirement for an initial cloning step, adding to the time and cost savings (Jones, 2010).

The use of universal adaptors, rather than sequence-specific primers, makes next-generation sequencing specifically suitable for metagenomic sequencing. Random shotgun sequencing using high-throughput next-generation sequencing has the ability to reveal the sequence of all the genetic material in a sample without prior knowledge on the organisms present in the sample and can facilitate large-scale studies to determine population structure and epidemiology of biological important samples (Hall, 2007; Mardis, 2008; Tucker et al., 2009).

### **2.3.2 Comparison of three next-generation platforms**

Next-generation sequencing technologies can be distinguished from Sanger sequencing in that they do not use chain termination chemistry and electrophoresis. Rather they rely on amplification of single DNA molecules to generate clusters of DNA templates held at defined locations on a solid support. This procedure is called solid-phase amplification. These clusters of identical molecules are then sequenced in parallel by cyclic incorporation and measurement of fluorescently labelled nucleotides (Illumina) or short oligonucleotides (ABI SOLiD), or the detection of by-products (454/Roche). Because of the parallel sequencing of the amplified clusters, this technology is also called massive parallel sequencing or high-throughput sequencing. For the purpose of this review, brief descriptions of each next-generation platform are given, however for further details, readers are advised to consult the following review papers: Ansorge, 2009; De Magalhães et al., 2010; Kato, 2009; Kircher and Kelso, 2010; Mardis, 2008; Metzker, 2010; Pettersson et al., 2009; Shendure and Ji, 2008; Strausberg et al., 2008; Tucker et al., 2009; Zhou et al., 2010. Figure 2.4 is a graphical comparison of the sequencing technologies of these three next-generation sequencing platforms.

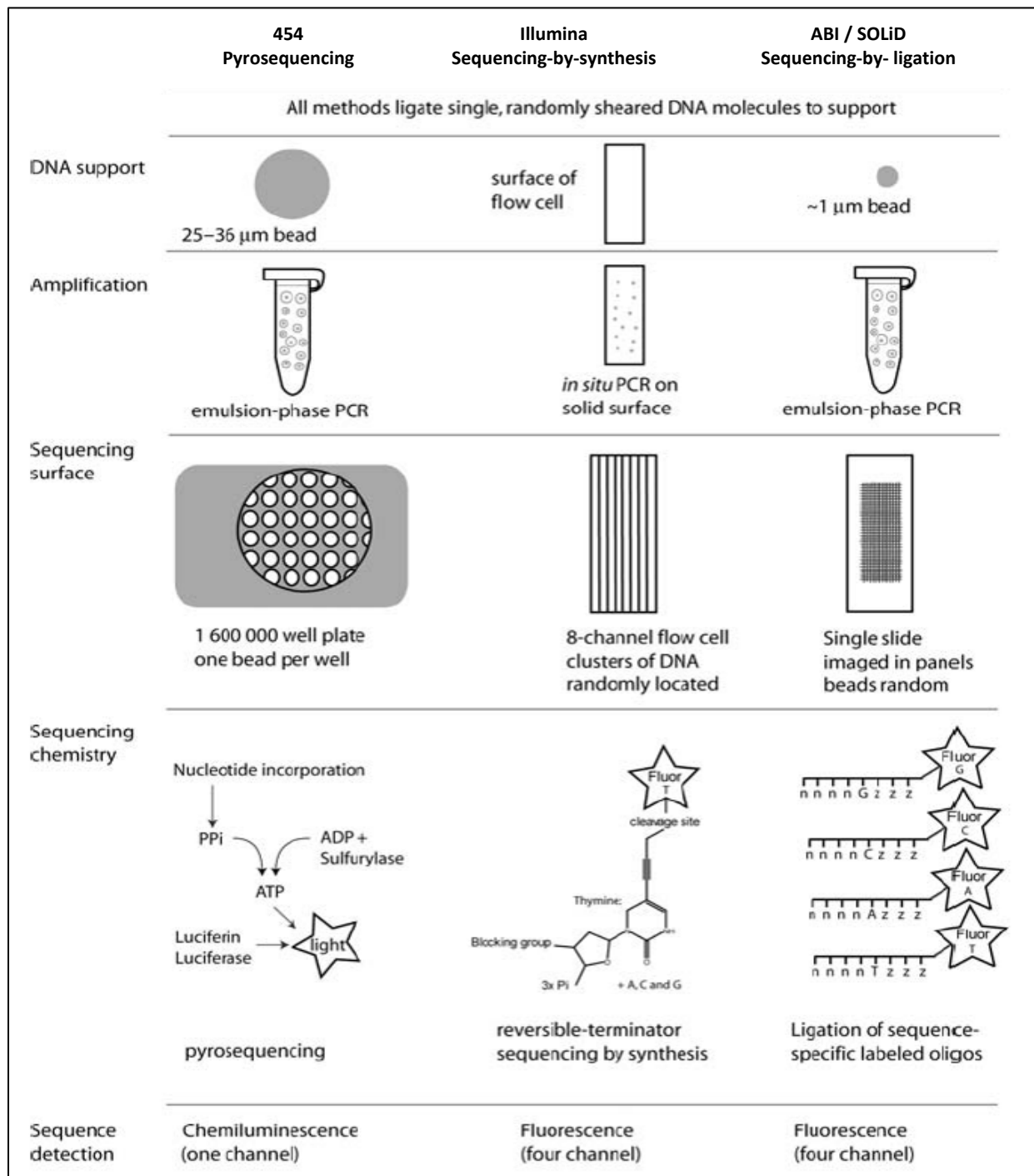
Both 454 and SOLiD sequencing technologies use DNA fragments ligated to specific adaptors and bound to beads. During emulsion PCR the fragments are amplified. Water droplets, each

containing one bead and the PCR reagents, are immersed in oil and put through the PCR thermal cycling. The beads with the amplified templates are then placed on the sequencing surface (Ansorge, 2009; Mardis, 2008; Tucker et al., 2009).

The 454 technology use a picotiter plate as solid support, one bead is deposited per well. To start the sequencing reaction, latex beads containing the enzymes sulphurylase and luciferase are added. The sequencing reactants with only one type of nucleotide are added. DNA polymerase incorporates the nucleotide, building a complementary DNA chain on the templates. Upon incorporation of a nucleotide, a pyrophosphate group (PPi) is released, forming adenosine triphosphate (ATP) in the presence of adenosine phosphosulphate (APS), catalyzed by sulphurylase. Luciferase catalyzes the reaction of luciferin and ATP to produce light. The emitted light can be detected and the beads incorporating the specific nucleotide recorded. The reactants are then washed away and the sequencing reaction repeated sequentially with the other three nucleotides. Since this technique relies on the release of pyrophosphate, it is called pyrosequencing (Ansorge, 2009; Mardis, 2008; Shendure and Ji, 2008).

The ABI SOLiD (Sequencing by *Oligo Ligation and Detection*) technology, also called sequencing by ligation, uses the same emulsion PCR procedure as in 454. The beads are then bound to a glass support surface, and primers hybridized to the adapters. A mixture of 3' degenerate octamers, each with a fluorescent label at the 5' end, and ligase are added to the beads. In each degenerate octamer, the fluorescent label corresponds to the fixed nucleotides. A hybridized octamer is ligated to the primer by ligase and identified by the fluorescent label, thereby determining the sequence of the template corresponding to the fixed nucleotides in the octamer. The label and rest of the octamer are cleaved, and more cycles of incorporation and detection follow (Mardis, 2008; Metzker, 2010; Shendure and Ji, 2008).

Illumina sequencing technology rely on amplification of single DNA molecules hybridized to the surface of a flow cell, followed by sequencing-by-synthesis, the detection of every fluorescently labelled nucleotide built into the growing DNA strand (Ansorge, 2009; Tucker et al., 2009). Illumina sequencing is reviewed in detail in section 2.3.3.



**Figure 2.4** General comparison of the sequencing technologies from the three next-generation sequencing platforms: 454/Roche, Illumina and ABI SOLiD (Adapted from Hudson, 2008).

Currently, all three sequencing platforms provide a good service and each renders a slightly different output dependent on the strong point of that system. Comparative figures for the latest platforms are shown in Table 2.3. In recent years these platforms have seen a dramatic increase in sequence quality, read length and total data output, therefore the figures quoted here will soon be improved on. In 2009 the instrument costs itself were similar (~\$500,000) (De

Magalhães et al., 2010). The choice of sequencing platform will depend on the available input material, nature of the experimental design and expected outcome.

**Table 2.3** Comparison of the latest available next-generation sequencing platforms. Specifications for the Illumina Genome Analyzer II (used in this study) are included. (Data were obtained from the respective websites).

Sequencing platforms				
	454 GS FLX Titanium	Illumina HiSeq 2000 <sup>b</sup>	ABI SOLiD 4 System <sup>c</sup>	Illumina Genome Analyzer II <sup>b</sup>
<b>Output/ run<sup>a</sup></b>	400-600 Mb	150-200 Gb	80-100 Gb	3 Gb
<b>Run Time</b>	10 hours	8 days	12- 16 days	5 days
<b>Read Length</b>	400 bases	2x100 bases	2x50 bases	2x36 bases
<b>Raw Accuracy</b>	99%	98.50%	99.94%	98.50%

<sup>a</sup> **Mb** Megabases, **Gb** Gigabases

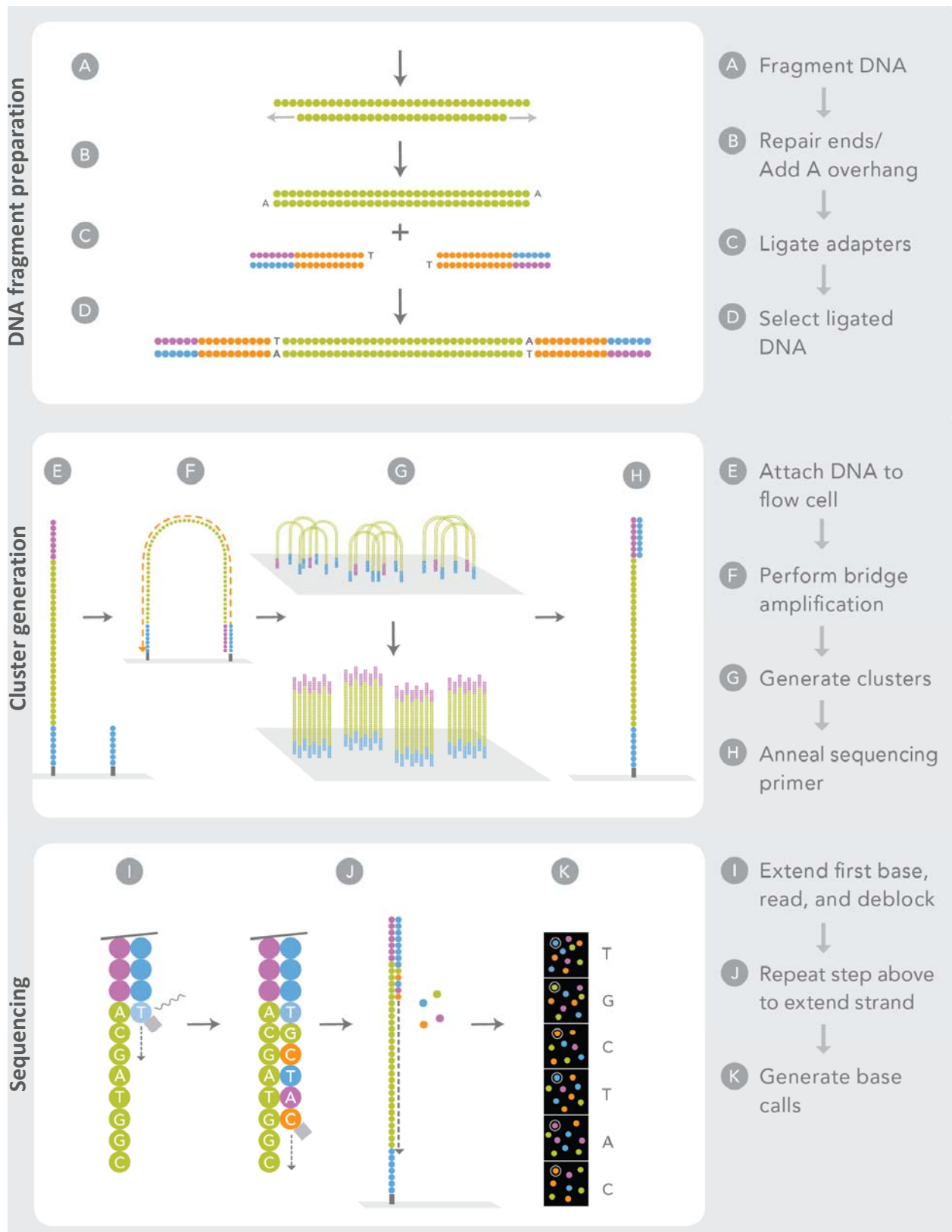
<sup>b</sup> Information based on paired-end libraries

<sup>c</sup> Information based on mate-paired libraries

For the purpose of this study we chose to use the Illumina Genome Analyzer II, at the time of our sequencing the latest sequencing platform from Illumina. In comparison with the FLX Genome Sequencer, then the flagship sequencer from 454/Roche, the Illumina Genome Analyzer can generate DNA sequencing data at substantially higher throughput and lower cost, requiring less starting material. Moreover, the observation of the sequencing procedure was possible, as an Illumina Genome Analyzer II instrument was available at a nearby location.

### 2.3.3 Technical overview of Illumina sequencing

Recently Illumina launched their latest sequencing platform, the HiSeq 2000. The HiSeq 2000 provides an improvement in terms of read length and run time and boost a 150 to 200 Gb data output capacity per run (Table 2.3, <http://www.illumina.com>), more than any other current sequencing platform. The technology used in Illumina's sequencing platforms, first developed by Solexa (Bennett, 2004) and acquired by Illumina in 2007, is based on the work of Turcatti and colleagues (Fedurco et al., 2006; Turcatti et al., 2008). Figure 2.5 is a graphical representation of the three main steps in the Illumina sequencing protocol: preparation of DNA, cluster generation and sequencing.



**Figure 2.5** Diagram illustrating the three steps of the Illumina Genome Analyzer sequencing technology (Adapted from: <http://www.illumina.com>).

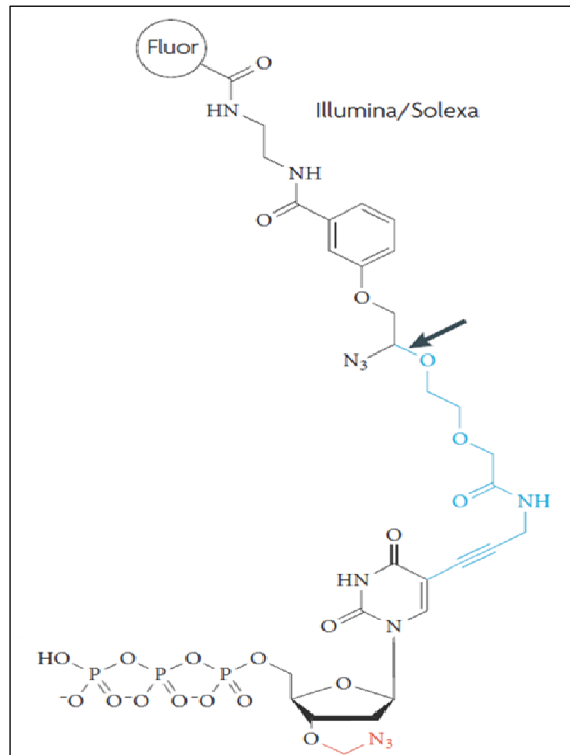
During preparation, DNA is fragmented by mechanical force, such as nebulization, or enzymatic digestion (Figure 2.5A). The ends of the double-stranded DNA are blunt-ended using the exonuclease activity of Klenow and T4 DNA polymerase and a 3'-adenine overhang is added by Klenow (exo-) polymerase (Figure 2.5B). Adapters are ligated to the ends of the fragmented strands (Figure 2.5C). The Illumina system uses Y-shaped adapters to ensure that each fragment has distinct up- and downstream adapters (Linnarsson, 2010). DNA fragments with added adapters are size fractionated on agarose gel (Figure 2.5D) and purified. The specific application will determine which size fraction is selected. The purified DNA fragments are amplified by PCR using primers complementary to the adapters, to enrich the input material for properly adapted fragments. The quality of the DNA library is validated by ensuring that a fragment of the selected size is visible on an agarose gel and optional by cloning and Sanger sequencing. Absorbance at 260nm is measured to quantify the DNA library and to dilute it to the optimum concentration (Linnarsson, 2010; Wilhelm and Landry, 2009). Besides DNA, other nucleic acids including mRNA and small RNAs, such as miRNAs can also be used, with modifications to the preparation. In this study we used dsRNA as starting material (discussed in section 3.3.1) The dsRNA was denatured by heating, flash cooled, converted to cDNA (discussed in section 3.7.2) and further treated as described above.

In the cluster generation step the single DNA fragments are clonally amplified to form clusters of identical molecules. A cluster of identical molecules is necessary during sequencing to enhance the fluorescence sufficiently for the imaging systems to detect it. A glass flow cell is prepared by dividing it into eight lanes and coating the interior surface with oligonucleotides complementary to the adapters used (Mardis, 2008). The oligonucleotides are linked to the glass surface via a flexible linker (Shendure and Ji, 2008). The glass flow cell acts as a solid support on which amplification takes place, allowing the clusters to be spatially separated. Currently the flow cell can retain up to 1 billion clusters, each with ~1000 template copies (www.illumina.com; Shendure and Ji, 2008).

The adapter-ligated DNA fragments are linked to the surface of the flow cell by a heating and cooling step (Figure 2.5E) (Mardis, 2008). Clusters are generated by incubation with free nucleotides and *Bst* polymerase. The free ends of the fragments hybridize with nearby complementary adapters. The adapters act as a primer, and a complementary DNA strand is

synthesized, a procedure named bridge amplification (Figure 2.5F). The double stranded DNA fragments are denatured with formamide (Shendure and Ji, 2008) and the process repeated to generate clusters of identical DNA fragments representative of a single DNA fragment in the original library (Figure 2.5G). In the last cycle the DNA fragments are denatured and a sequencing primer annealed to each DNA strand (Figure 2.5H) (Shendure and Ji, 2008; Tucker et al., 2009).

The flow cell with the DNA clusters is subjected to sequencing. The Illumina system applies the cyclic reversible termination (CRT) method, a cyclic method that involves three steps: nucleotide incorporation, fluorescence detection and cleavage. The nucleotides are modified to have their 3'-OH group blocked with reversible terminators, to ensure that only one nucleotide is incorporated per cycle, and to contain a cleavable fluorophore (Figure 2.6). Polymerase and the four differentially fluorescent labelled, modified nucleotides are added to each lane of the flow cell. The sequencing primers prime DNA synthesis and the first nucleotide is incorporated in the complementary DNA strand, building on the template of the original molecule. After each incorporation cycle the excess reactants are washed away and an image is taken of the clusters to identify the incorporated differentially fluorescent labelled nucleotide. The 3'-ends of the nucleotides are unblocked, the fluorescent labels cleaved, washed away and the nucleotide incorporation cycle is repeated (Figure 2.5I and J). Each lane of the flow cell is divided into 300 tiles that are individually photographed. The images are captured by a camera and the four fluorescent colours are detected using two lasers. At the end of the sequencing run, the images are analyzed to identify the nucleotides incorporated and hence the sequence of the DNA fragment in each cluster (Figure 2.5K). As one base is determined per sequencing cycle, the number of cycles equals the read length (Ansorge, 2009; Metzker, 2010; Pettersson et al., 2009).



**Figure 2.6** Diagram of a modified nucleotide used in Illumina sequencing (Adapted from Metzker, 2010).

If a paired-end sequencing run is performed, the DNA templates are inverted, the end that was sequenced is now attached to the glass flow cell, allowing for the other end to be sequenced ([www.illumina.com](http://www.illumina.com)). Paired-end reads are generated by size-selection of DNA fragments (usually only 200bp fragments are selected) and sequencing of the first number of nucleotides, depending on the read length used, from both ends of the fragments (Medvedev et al., 2009).

### 2.3.4 Bioinformatic analysis of next-generation sequencing data

The advent of NGS platforms led to a significant increase in sequencing throughput, allowing discoveries not previously accessible. But this progress in sequencing technology has not been matched by an equal progress in bioinformatic tools necessary for analyzing and interpreting the wealth of sequencing data (McPherson, 2009).

#### 2.3.4.1 Bioinformatic challenges when dealing with next-generation sequencing data

The characteristics of NGS data, shorter read lengths and higher coverage, introduce major bioinformatic challenges when analyzing these data. Sequence reads are produced in large



quantities and place different demands on assemblers than Sanger reads. Alignments using NGS data typically take longer to complete, it may consume hundreds of central processing unit (CPU) hours to align and assemble millions of short DNA reads to create contigs. Additionally, the shorter read length of NGS sequence data complicates the assembly of repetitive regions or gene duplications (Pop, 2009).

Next-generation sequencing data are generally considered to have a lower raw accuracy in comparison to Sanger sequence data (De Magalhães et al., 2010; Tucker et al., 2009). The quality of the sequence data are influenced by the sample preparation and sequencing protocols (Trapnell and Salzberg, 2009). Furthermore, each of the respective NGS platforms introduces its own specific type of sequencing errors and error profiles (De Magalhães et al., 2010; Strausberg et al., 2008). In Illumina sequencing data, substitutions are the most frequent error type, especially after incorporation of a “G” base. Homopolymeric repeats are the cause of most errors in 454 data, causing insertions/deletions (Metzker, 2010). Because sequence quality typically deteriorates with increase in read length, it is often needed to trim error-prone read ends, reducing the data set size. However, the greater depth of coverage of sequenced fragments, allow for bioinformatic assembly of good quality contigs (Imelfort and Edwards, 2009).

The added challenge of metagenomic sequencing with high-throughput sequencing technology is to assemble the sequences in contigs, to construct genomes and assign an identity to sequence fragments. In conventional sequencing projects, all sequenced fragments originate from a single genome. When using a metagenomic sequencing approach on genetic material from an environmental sample, however, it is unknown which sequences originate from which species (Wooley and Ye, 2009). A further obstacle is the genetic variation of a population within species, complicating the assembly of full genome sequences (Miller et al., 2010; Tyson et al., 2004). It is also difficult to distinguish between true sequence variation and sequencing errors (Valdivia-Granda, 2008). To enable quantitative analysis, the sequencing data must reflect the original abundance of RNA molecules in the sample. Some RNA molecules can form secondary structures, preventing it from denaturation and efficient sequencing, resulting in a skewed data set (Beaume et al., 2010).

#### **2.3.4.2 Bioinformatic tools for next-generation sequencing data**

Since 2005 several *in silico* software tools and algorithms have been developed specifically for analysis of NGS data in a broad range of application fields. The reader are referred to several papers discussing a number of these bioinformatic tools: Horner et al., 2010; Imelfort and Edwards, 2009; Li and Homer, 2010; Miller et al., 2010; Nielsen et al., 2010; Scheibye-Alsing et al., 2009; Trapnell and Salzberg, 2009. The performance of the analysis software depends greatly on the sequence length, depth of sequence coverage (fold coverage) and relative error-free sequencing data (DiGiustini et al., 2009). No single bioinformatic tool will be suitable for analysis of all NGS data, as each was designed having a specific application in mind and each has its own inherent weaknesses and strengths.

The “first step” analysis tools are typically categorized in three groups: *de novo* sequence assemblers, used to create contigs from short read sequence data purely based on overlapping sequence information; mapping assemblers, short read sequence data are mapped to a reference genome and a consensus generated; and alignment viewers, used to visualize the alignments. In Table 2.4 a list of such of NGS data analysis software is presented. In section 2.3.3.4.1 and 2.3.3.4.2 software used in this study are discussed: Velvet, a *de novo* short read assembler and MAQ, a mapping assembler.

Besides this alignment and assembly software, software for variant, single nucleotide polymorphism and insertion/deletion detection, are available. Integrated analysis software, able to perform all or some of the above mention functions, are also offered. The respective sequencing platforms developed their own analysis tools: Consensus Assessment of Sequence and Variation (CASAVA) developed by Illumina ([http://www.illumina.com/software/genome\\_analyzer\\_software.ilmn#third\\_party\\_tools](http://www.illumina.com/software/genome_analyzer_software.ilmn#third_party_tools)) and the GS De Novo Assembler, Reference mapper and Variant Analyzer by 454/Roche (<http://454.com/products-solutions/analysis-tools/index.asp>), are provided for free with the respective instruments.

**Table 2.4** Examples of available *de novo* short read assemblers, mapping assemblers and alignment viewers. Websites are shown for more information on this software.

<b>De novo short read assemblers</b>	
ABYSS	<a href="http://www.bcgsc.ca/platform/bioinfo/software/abyss">http://www.bcgsc.ca/platform/bioinfo/software/abyss</a>
ALLPATHS	<a href="http://www.broadinstitute.org/science/programs/genome-biology/computational-rd/computational-research-and-development">http://www.broadinstitute.org/science/programs/genome-biology/computational-rd/computational-research-and-development</a>
EDENA	<a href="http://www.genomic.ch/edena.php">http://www.genomic.ch/edena.php</a>
EULER-SR	<a href="http://euler-assembler.ucsd.edu/portal/">http://euler-assembler.ucsd.edu/portal/</a>
MIRA3	<a href="http://chevreux.org/projects_mira.html">http://chevreux.org/projects_mira.html</a>
SHARCGS	<a href="http://sharcgs.molgen.mpg.de/download.shtml">http://sharcgs.molgen.mpg.de/download.shtml</a>
SSAKE	<a href="http://www.bcgsc.ca/platform/bioinfo/software/ssake/releases/3.2">http://www.bcgsc.ca/platform/bioinfo/software/ssake/releases/3.2</a>
VCAKE	<a href="http://mac.softpedia.com/get/Math-Scientific/VCAKE.shtml">http://mac.softpedia.com/get/Math-Scientific/VCAKE.shtml</a>
Velvet	<a href="http://www.ebi.ac.uk/~zerbino/velvet/">http://www.ebi.ac.uk/~zerbino/velvet/</a>
<b>Mapping assemblers</b>	
AB mapreads	<a href="http://solidsoftwaretools.com/gf/project/mapreads/">http://solidsoftwaretools.com/gf/project/mapreads/</a>
BFAST	<a href="http://sourceforge.net/apps/mediawiki/bfast/index.php?title=Main_Page">http://sourceforge.net/apps/mediawiki/bfast/index.php?title=Main_Page</a>
Bowtie	<a href="http://bowtie.cbcb.umd.edu">http://bowtie.cbcb.umd.edu</a>
BWA	<a href="http://maq.sourceforge.net/bwa-man.shtml">http://maq.sourceforge.net/bwa-man.shtml</a>
CloudBurst	<a href="http://cloudburst-bio.sourceforge.net/">http://cloudburst-bio.sourceforge.net/</a>
GMAP	<a href="http://www.gene.com/share/gmap/">http://www.gene.com/share/gmap/</a>
MAQ	<a href="http://maq.sourceforge.net">http://maq.sourceforge.net</a>
MOM	<a href="http://mom.csbc.vcu.edu">http://mom.csbc.vcu.edu</a>
MOSAİK	<a href="http://bioinformatics.bc.edu/marthlab/Mosaik">http://bioinformatics.bc.edu/marthlab/Mosaik</a>
MuMRRescueLite	<a href="http://genome.gsc.riken.jp/osc/english/dataresource/">http://genome.gsc.riken.jp/osc/english/dataresource/</a>
PASS	<a href="http://pass.cribi.unipd.it">http://pass.cribi.unipd.it</a>
RMAP	<a href="http://rulai.cshl.edu/rmap/">http://rulai.cshl.edu/rmap/</a>
SeqMap	<a href="http://biogibbs.stanford.edu/~jiangh/SeqMap/">http://biogibbs.stanford.edu/~jiangh/SeqMap/</a>
SHRIMP	<a href="http://compbio.cs.toronto.edu/shrimp/">http://compbio.cs.toronto.edu/shrimp/</a>
Soap	<a href="http://soap.genomics.org.cn">http://soap.genomics.org.cn</a>
Vmatch	<a href="http://www.vmatch.de/">http://www.vmatch.de/</a>
ZOOM	<a href="http://www.bioinform.com">http://www.bioinform.com</a>
<b>Genome Browsers/Alignment Viewers</b>	
EagleView	<a href="http://bioinformatics.bc.edu/marthlab/EagleView">http://bioinformatics.bc.edu/marthlab/EagleView</a>
MapView	<a href="http://evolution.sysu.edu.cn/mapview/">http://evolution.sysu.edu.cn/mapview/</a>
SAMtools	<a href="http://www.bcgsc.ca/platform/bioinfo/software/sam">http://www.bcgsc.ca/platform/bioinfo/software/sam</a>

Compiled from:

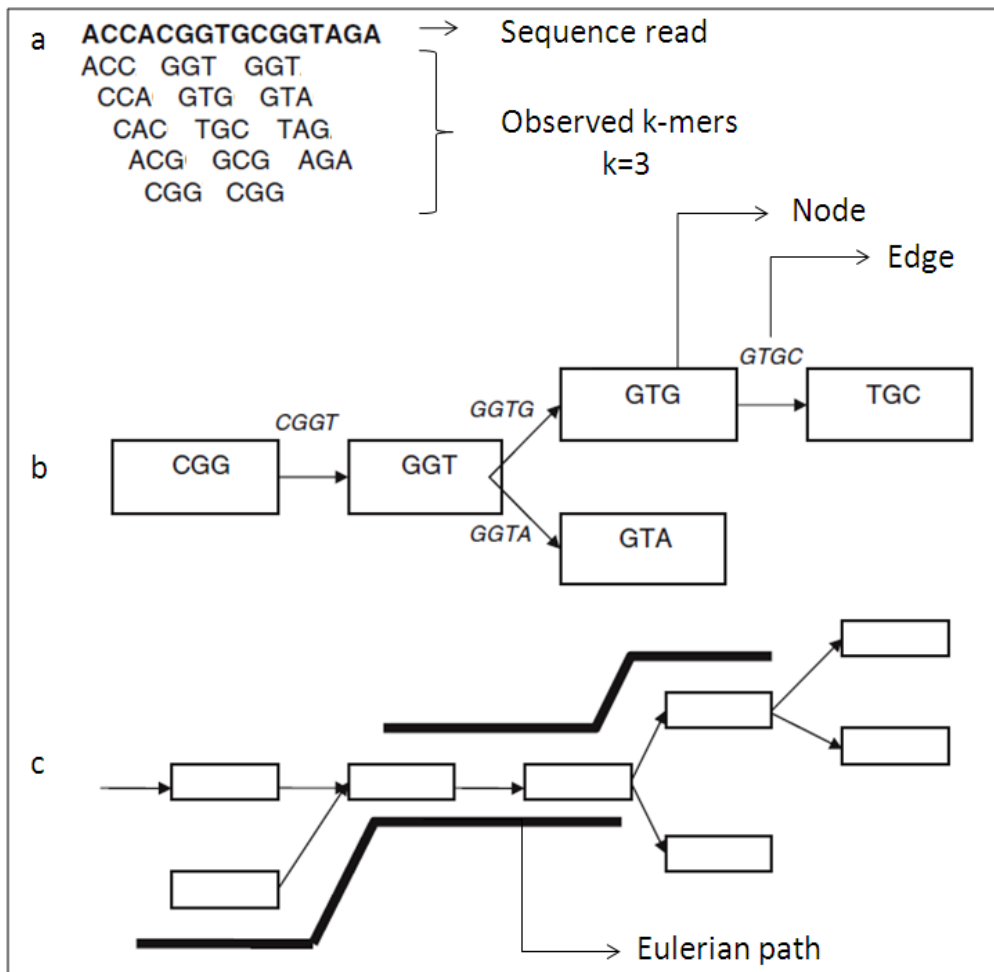
Imelfort and Edwards, 2009; Horner et al., 2010; Li and Homer, 2010; Miller et al., 2010; Nielsen et al., 2010; Scheibye-Alsing et al., 2009; Trapnell and Salzberg, 2009; Varshney et al., 2009  
<http://www.genomeweb.com/sequencing/moving-simulations-real-data-short-read-assemblers-startfacing>  
[http://www.illumina.com/software/genome\\_analyzer\\_software.ilmn#third\\_party\\_tools](http://www.illumina.com/software/genome_analyzer_software.ilmn#third_party_tools)  
<http://seqanswers.com/forums/showthread.php?t=43>,  
[http://mi.caspuir.it/ngs/software\\_review.php](http://mi.caspuir.it/ngs/software_review.php)  
<http://sequenceassembly.com/>

#### 2.3.4.2.1 Velvet

Illumina sequence data from this study were assembled using Velvet (Zerbino and Birney, 2008; <http://www.ebi.ac.uk/~zerbino/velvet>), a *de novo* short read assembler recommended by Illumina. Velvet is a set of algorithms specifically developed to assemble short sequence reads produced by next-generation sequencing technologies and is based on manipulating de Bruijn graphs for time-efficient assembly of short sequence reads (Zerbino and Birney, 2008).

The sequence reads are first divided into a subset of shorter fragments of a fixed length, called the hash length ( $k$ ), and these fragments are called  $k$ -mers. Each  $k$  sequence observed is represented as a “node” in the De Bruijn graph, with an “edge” or connection between them if they occurred adjacently in the original sequence read. A correct path through these nodes, representing an original read or genome present is called an Eulerian path (Pop and Salzberg, 2008; Pop, 2009). Figure 2.7 is a graphical illustration of the De Bruijn graph theory. Velvet uses “hash table” indexing to assist rapid searching of the de Bruijn graph to construct contigs from the overlapping sequence reads. In the hash tables the identity and position of the first occurrence of each  $k$ -mer in the de Bruijn graph are recorded (Flicek and Birney, 2009).

Contigs are assembled into scaffolds using paired-end information. Paired-end reads are created by fragmenting DNA followed by selection of a specific fragment size. These fragments are sequenced from both ends. Forward and reverse reads from a specific fragment remain linked with the known insert length between them. This data can be used to link contigs to create scaffolds, although no overlapping sequence data is available (Zerbino and Birney, 2008).



**Figure 2.7** Diagram illustrating the theory of De Bruijn graphs used in Velvet assembler **a)** Sequence read with possible k-mers **b)** De Bruijn graph featuring nodes and edges **c)** Eulerian paths showing two overlapping sequence reads. (Adapted from Pop, 2009).

### 2.3.4.2.2 MAQ

Mapping and Assembly with Quality (MAQ) assembler (Li et al., 2008; <http://MAQ.sourceforge.net/>) is a freely available software. MAQ is a mapping assembler, so called because the original short sequence reads are reassembled or “mapped” to a reference sequence. A coverage profile for the reference sequence can be generated, showing the percentage sequence covered and depth of sequence, by calculating the number of times each nucleotide is covered by a sequence read. MAQ can also be used for single nucleotide polymorphism (SNP) detection and to determine nucleotide variation among sequence reads in the raw data and the final consensus sequence. MAQview is included in the MAQ package and can be used to view the alignments (Li et al., 2008).

MAQ uses “seed indexing” for fast sequence alignment (Li et al., 2008). The first 28bp of the sequence read, regarded as the most accurate part (Trapnell and Salzberg, 2009), are divided into four segments, called “seeds”. When set to default mode, MAQ will report on reads aligning with two or less mismatches. These two mismatches can be in two seeds at most, with the two other seeds having perfect matches. By aligning all six the possible combinations of seeds to the reference sequence it is possible to create a list of candidate positions where the read can align with two or less mismatches in the first 28bp. This information is stored in six hash tables for fast indexing (Li and Homer, 2010). Alignment of seeds is much less computational intensive than aligning the complete sequence read. The mismatches for each alignment are then weighted using the quality score for nucleotides in the sequence read. Quality scores are provided with the raw data to give an indication of the confidence that each base is called correctly. Mismatches caused by sequencing errors should rather occur at nucleotide positions with lower quality score, while true SNP’s is at position with higher quality scores (Flicek and Birney, 2009; Trapnell and Salzberg, 2009).

## **2.4 Conclusion**

Next-generation sequencing technologies are now used in an increasingly diverse range of applications to answer biological problems not previously thought to be possible. Improvements to the current sequencing technologies are being developed at a rapid tempo, and new techniques such as multiplexing and sample tagging, are constantly added. The high sequence throughput capacity enables large-scale metagenomic studies, including studies to investigate disease pathogens. Although, at present the cost of the technique is still limiting, in future next-generation sequencing can become a routine diagnostic technique for detection of pathogens.

## 2.5 References

- Adams, I.P., Glover, R.H., Monger, W.A., Mumford, R., Jackeviciene, E., Navalinskiene, M., et al., 2009. Next-generation sequencing and metagenomic analysis: a universal diagnostic tool in plant virology. *Mol. Plant Pathol.* 10, 537-545.
- Al Rwahnih, M., Daubert, S., Golino, D., Rowhani, A., 2009. Deep sequencing analysis of RNAs from a grapevine showing Syrah decline symptoms reveals a multiple virus infection that includes a novel virus. *Virology* 387, 395-401.
- Allen, M.J., Wilson, W.H., 2008. Aquatic virus diversity accessed through omic techniques: a route map to function. *Curr. Opin. Microbiol.* 11, 226-232.
- Angly, F.E., Felts, B., Breitbart, M., Salamon, P., Edwards, R.A., Carlson, C., et al., 2006. The marine viromes of four oceanic regions. *PLoS Biol.* 4, e368.
- Ansorge, W.J., 2009. Next-generation DNA sequencing techniques. *N. Biotechnol.* 25, 195-203.
- Battany, M., Rowhani, A., Golino, D., 2004. Syrah in California: decline or disorder?. *Pract. Winery Vineyards* 2004, 1-7.
- Beaume, M., Hernandez, D., Francois, P., Schrenzel, J., 2010. New approaches for functional genomic studies in staphylococci. *Int. J. Med. Microbiol.* 300, 88-97.
- Bench, S.R., Hanson, T.E., Williamson, K.E., Ghosh, D., Radosovich, M., Wang, K., et al., 2007. Metagenomic characterization of Chesapeake Bay virioplankton. *Appl. Environ. Microbiol.* 73, 7629-7641.
- Bennett, S., 2004. Solexa Ltd. *Pharmacogenomics* 5, 433-438.
- Bentley, D.R., 2006. Whole-genome re-sequencing. *Curr. Opin. Genet. Dev.* 16, 545-552.
- Blinkova, O., Kapoor, A., Victoria, J., Jones, M., Wolfe, N., Naeem, A., et al., 2009. Cardioviruses are genetically diverse and cause common enteric infections in South Asian children. *J. Virol.* 83, 4631-4641.
- Blinkova, O., Victoria, J., Li, Y., Keele, B.F., Sanz, C., Ndjanga, J.-N., et al., 2010. Novel circular DNA viruses in stool samples of wild-living chimpanzees. *J. Gen. Virol.* 91, 74-86.
- Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J.M., Segall, A.M., Mead, D., et al., 2002. Genomic analysis of uncultured marine viral communities. *Proc. Natl. Acad. Sci. U.S.A.* 99, 14250-14255.
- Breitbart, M., Hewson, I., Felts, B., Mahaffy, J.M., Nulton, J., Salamon, P., et al., 2003. Metagenomic analyses of an uncultured viral community from human feces. *J. Bacteriol.* 185, 6220-6223.
- Breitbart, M., Felts, B., Kelley, S., Mahaffy, J.M., Nulton, J., Salamon, P., et al., 2004. Diversity and population structure of a near-shore marine-sediment viral community. *Proc. R. Soc. Lond. B Biol. Sci.* 271, 565-574.
- Breitbart, M., Rohwer, F., 2005. Method for discovering novel DNA viruses in blood using viral particle selection and shotgun sequencing. *Bio Techniques* 39, 729-736.
- Breitbart, M., Haynes, M., Kelley, S., Angly, F., Edwards, R.A., Felts, B., et al., 2008. Viral diversity and dynamics in an infant gut. *Res. Microbiol.* 159, 367-373.
- Briese, T., Paweska, J.T., McMullan, L.K., Hutchison, S.K., Street, C., Palacios, G., et al., 2009. Genetic detection and characterization of Lujo virus, a new hemorrhagic fever-associated arenavirus from southern Africa. *PLoS Pathog.* 5, e1000455.
- Cann, A.J., Fandrich, S.E., Heaphy, S., 2005. Analysis of the virus population present in equine faeces indicates the presence of hundreds of uncharacterized virus genomes. *Virus Genes* 30, 151-156.

- Cardenas, E., Tiedje, J.M., 2008. New tools for discovering and characterizing microbial diversity. *Curr. Opin. Biotechnol.* 19, 544-549.
- Culley, A.I., Lang, A.S., Suttla, C.A., 2006. Metagenomic analysis of coastal RNA virus communities. *Science* 312, 1795-1798.
- De Magalhães, J.P., Finch, C.E., Janssens, G., 2010. Next-generation sequencing in aging research: Emerging applications, problems, pitfalls and possible solutions. *Ageing Res. Rev.* 9, 315-323.
- Delwart, E.L., 2007. Viral metagenomics. *Rev. Med. Virol.* 17, 115-131.
- Desnues, C., Rodriguez-Brito, B., Rayhawk, S., Kelley, S., Tran, T., Haynes, M., et al., 2008. Biodiversity and biogeography of phages in modern stromatolites and thrombolites. *Nature* 452, 340-343.
- Deutschbauer, A.M., Chivian, D., Arkin, A.P., 2006. Genomics for environmental microbiology. *Curr. Opin. Biotechnol.* 17, 229-235.
- DiGiustini, S., Liao, N.Y., Platt, D., Robertson, G., Seidel, M., Chan, S.K., et al., 2009. De novo genome sequence assembly of a filamentous fungus using Sanger, 454 and Illumina sequence data. *Genome Biol.* 10, R94.
- Djikeng, A., Kuzmickas, R., Anderson, N.G., Spiro, D.J., 2009. Metagenomic analysis of RNA viruses in a fresh water lake. *PLoS One* 4, e7264.
- Dodds, J.A., Morris, T.J., Jordan, R.L., 1984. Plant viral double-stranded RNA. *Annu. Rev. Phytopathol.* 22, 151-168.
- Edwards, R.A., Rohwer, F., 2005. Viral metagenomics. *Nature Rev. Microbiol.* 3, 504-510.
- Fedurco, M., Romieu, A., Williams, S., Lawrence, I., Turcatti, G., 2006. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. *Nucleic Acids Res.* 34, e22.
- Fierer, N., Breitbart, M., Nulton, J., Salamon, P., Lozupone, C., Jones, R., et al., 2007. Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. *Appl. Environ. Microbiol.* 73, 7059-7066.
- Finkbeiner, S.R., Allred, A.F., Tarr, P.I., Klein, E.J., Kirkwood, C.D., Wang, D., 2008. Metagenomic analysis of human diarrhea: viral detection and discovery. *PLoS Pathog.* 4, e1000011.
- Flicek, P., Birney, E., 2009. Sense from sequence reads: methods for alignment and assembly. *Nat. Methods* 6, S6-S12.
- Goszczynski, D.E., Jooste, A.E.C., 2003. Identification of divergent variants of grapevine virus A. *Eur. J. Plant Pathol.* 109, 397-403.
- Goszczynski, D.E., 2007a. Investigation of association of viruses of *Viti*- and *Foveavirus* genera with Shiraz disease and Shiraz (syn. Syrah) decline in South Africa. *Syrah Vine Health Symposium*, 8-11.
- Goszczynski, D.E., 2007b. Single-strand conformation polymorphism (SSCP), cloning and sequencing reveal a close association between related molecular variants of grapevine virus A (GVA) and Shiraz disease in South Africa. *Plant pathol.* 56, 755-762.
- Goszczynski, D.E., Du Preez, J., Burger, J.T., 2008. Molecular divergence of grapevine virus A (GVA) variants associated with Shiraz disease in South Africa. *Virus Res.* 138, 105-110.
- Goussard, P., Bakker, H., 2000. Characterisation of grapevines visually infected with Shiraz disease associated viruses. *Wynboer* [online] <http://www.wynboer.co.za/recentarticles/200612shiraz.php3>.
- Green, B.D., Keller, M., 2006. Capturing the uncultivated majority. *Curr. Opin. Biotechnol.* 17, 236-240.



- Guazzaroni, M., Beloqui, A., Golyshin, P.N., Ferrer, M., 2009. Metagenomics as a new technological tool to gain scientific knowledge. *World J. Microbiol. Biotechnol.* 25, 945-954.
- Habili, N., Farrokhi, N., Lima, M.F., Nicholas, P., Randles, J.W., 2006. Distribution of Rupestris stem-pitting-associated virus variants in two Australian vineyards showing different symptoms. *Ann. Appl. Biol.* 148, 91-96.
- Hall, N., 2007. Advanced sequencing technologies and their wider impact in microbiology. *J. Exp. Biol.* 210, 1518-1525.
- Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J., Goodman, R.M., 1998. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem. Biol.* 5, R245-R249.
- Handelsman, J., 2004. Metagenomics: application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.* 68, 669-685.
- Harismendy, O., Ng, P., Strausberg, R., Wang, X., Stockwell, T., Beeson, K., et al., 2009. Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biol.* 10, R32.
- Horner, D.S., Pavesi, G., Castrignano, T., De Meo, P.D., Liuni, S., Sammeth, M., et al., 2010. Bioinformatics approaches for genomics and post genomics applications of next-generation sequencing. *Brief. Bioinforma.* 11, 181-197.
- Hudson, M.E., 2008. Sequencing breakthroughs for genomic ecology and evolutionary biology. *Mol. Ecol. Res.* 8, 3-17.
- Hugenholtz, P., Tyson, G.W., 2008. Microbiology: Metagenomics. *Nature* 455, 481-483.
- Imelfort, M., Edwards, D., 2009. De novo sequencing of plant genomes using second-generation technologies. *Brief. Bioinforma.* 10, 609-618.
- Jones, W.J., 2010. High-throughput sequencing and metagenomics. *Estuaries Coasts* 33, 944-952.
- Kapoor, A., Victoria, J., Simmonds, P., Slikas, E., Chieochansin, T., Naeem, A., et al., 2008. A highly prevalent and genetically diversified *Picornaviridae* genus in South Asian children. *Proc. Natl. Acad. Sci. U.S.A.* 105, 20482-20487.
- Karasev, A.V., 2000. Genetic diversity and evolution of closteroviruses. *Annu. Rev. Phytopathol.* 38, 293-324.
- Kato, K., 2009. Impact of the next generation DNA sequencers. *Int. J. Clin. Exp. Med.* 2, 193-202.
- Kim, K., Chang, H., Nam, Y., Roh, S.W., Kim, M., Sung, Y., et al., 2008. Amplification of uncultured single-stranded DNA viruses from rice paddy soil. *Appl. Environ. Microbiol.* 74, 5975-5985.
- Kircher, M., Kelso, J., 2010. High-throughput DNA sequencing - concepts and limitations. *BioEssays* 32, 524-536.
- Kowalchuk, G.A., Speksnijder, A.G.C.L., Zhang, K., Goodman, R.M., Van Veen, J.A., 2007. Finding the needles in the metagenome haystack. *Microb. Ecol.* 53, 475-485.
- Kreuze, J.F., Perez, A., Untiveros, M., Quispe, D., Fuentes, S., Barker, I., et al., 2009. Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: a generic method for diagnosis, discovery and sequencing of viruses. *Virology* 388, 1-7.
- Kristensen, D.M., Mushegian, A.R., Dolja, V.V., Koonin, E.V., 2010. New dimensions of the virus world discovered through metagenomics. *Trends Microbiol.* 18, 11-19.
- Li, H., Ruan, J., Durbin, R., 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* 18, 1851-1858.

- Li, H., Homer, N., 2010. A survey of sequence alignment algorithms for next-generation sequencing. *Brief. Bioinforma.* doi:10.1093/bib/bbq015.
- Li, L., Kapoor, A., Slikas, B., Bamidele, O.S., Wang, C., Shaukat, S., et al., 2010a. Multiple diverse circoviruses infect farm animals and are commonly found in human and chimpanzee feces. *J. Virol.* 84, 1674-1682.
- Li, L., Victoria, J.G., Wang, C., Jones, M., Fellers, G.M., Kunz, T.H., et al., 2010b. Bat guano virome: predominance of dietary viruses from insects and plants plus novel mammalian viruses. *J. Virol.* 84, 6955-6965.
- Lima, M.F., Alkowni, R., Uyemoto, J.K., Golino, D., Osman, F., Rowhani, A., 2006. Molecular analysis of a California strain of Rupestris stem pitting-associated virus isolated from declining Syrah grapevines. *Arch. Virol.* 151, 1889-1894.
- Linnarsson, S., 2010. Recent advances in DNA sequencing methods – general principles of sample preparation. *Exp. Cell Res.* 316, 1339-1343.
- Lopez-Bueno, A., Tamames, J., Velazquez, D., Moya, A., Quesada, A., Alcamí, A., 2009. High diversity of the viral community from an Antarctic lake. *Science* 326, 858-861.
- Mardis, E.R., 2008. The impact of next-generation sequencing technology on genetics. *Trends Genet.* 4, 133-141.
- Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., et al., 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376-380.
- Marhaver, K.L., Edwards, R.A., Rohwer, F., 2008. Viral communities associated with healthy and bleaching corals. *Environ. Microbiol.* 10, 2277-2286.
- Martelli, G.P., Boudon-Padieu, E., 2006. Directory of infectious diseases of grapevines and viroses and virus-like diseases of the grapevine: Bibliographic report 1998-2004. *Options Méditerran., Ser. B, Stud. Res.* 55, CIHEAM, 279.
- Martelli, G.P., 2009. Grapevine virology highlights 2006-2009. 16th meeting of the International Council for the study of virus and virus-like diseases of the grapevine, 15-23.
- Maxam, A., Gilbert, W., 1977. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. U.S.A.* 74, 560-564.
- McDaniel, L., Breitbart, M., Mobberley, J., Long, A., Haynes, M., Rohwer, F., et al., 2008. Metagenomic analysis of lysogeny in Tampa Bay: implications for prophage gene expression. *PLoS One* 3, e3263.
- McPherson, J.D., 2009. Next-generation gap. *Nat. Methods* 6, S2-S5.
- Medvedev, P., Stanciu, M., Brudno, M., 2009. Computational methods for discovering structural variation with next-generation sequencing. *Nat. Methods* 6, S13-S20.
- Melcher, U., Muthukumar, V., Wiley, G.B., Min, B.E., Palmer, M.W., Verchot-Lubicz, J., et al., 2008. Evidence for novel viruses by analysis of nucleic acids in virus-like particle fractions from *Ambrosia psilostachya*. *J. Virol. Methods* 152, 49-55.
- Metzker, M.L., 2010. Sequencing technologies the next generation. *Nat. Rev. Genet.* 11, 31-46.
- Millar, C.D., Huynen, L., Subramanian, S., Mohandesan, E., Lambert, D.M., 2008. New developments in ancient genomics. *Trends Ecol. Evol.* 23, 386-393.
- Miller, J.R., Koren, S., Sutton, G., 2010. Assembly algorithms for next-generation sequencing data. *Genomics* 95, 315-327.
- Morozova, O., Marra, M.A., 2008. Applications of next-generation sequencing technologies in functional genomics. *Genomics* 92, 255-264.

- Muthukumar, V., Melcher, U., Pierce, M., Wiley, G.B., Roe, B.A., Palmer, M.W., et al., 2009. Non-cultivated plants of the Tallgrass Prairie Preserve of northeastern Oklahoma frequently contain virus-like sequences in particulate fractions. *Virus Res.* 141, 169-173.
- Nakamura, S., Yang, C., Sakon, N., Ueda, M., Tougan, T., Yamashita, A., et al., 2009. Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach. *PLoS One* 4, e4219.
- Ng, T.F.F., Manire, C., Borrowman, K., Langer, T., Ehrhart, L., Breitbart, M., 2009. Discovery of a novel single-stranded DNA virus from a sea turtle fibropapilloma by using viral metagenomics. *J. Virol.* 83, 2500-2509.
- Nielsen, C.B., Cantor, M., Dubchak, I., Gordon, D., Wang, T., 2010. Visualizing genomes: techniques and challenges. *Nat. Methods* 7, S5-S15.
- Palacios, G., Druce, J., Du, L., Tran, T., Birch, C., Briese, T., et al., 2008. A new arenavirus in a cluster of fatal transplant-associated diseases. *N. Engl. J. Med.* 358, 991-998.
- Parsley, L.C., Consuegra, E.J., Thomas, S.J., Bhavsar, J., Land, A.M., Bhuiyan, N.N., et al., 2010. Census of the viral metagenome within an activated sludge microbial assemblage. *Appl. Environ. Microbiol.* 76, 2673-2677.
- Pettersson, E., Lundeberg, J., Ahmadian, A., 2009. Generations of sequencing technologies. *Genomics* 93, 105-111.
- Pietersen, G., 2000. Spread of grapevine leafroll disease in South Africa - a difficult, but not insurmountable problem. Wynboer [online] <http://www.wynboer.co.za/recentarticles/0406leaf.php3>.
- Pop, M., Salzberg, S.L., 2008. Bioinformatics challenges of new sequencing technology. *Trends Genet.* 24, 142-149.
- Pop, M., 2009. Genome assembly reborn: recent computational challenges. *Brief. Bioinforma.* 10, 354-366.
- Prosser, S.W., Goszczynski, D.E., Meng, B., 2007. Molecular analysis of double-stranded RNAs reveals complex infection of grapevines with multiple viruses. *Virus Res.* 124, 151-159.
- Raes, J., Foerstner, K.U., Bork, P., 2007. Get the most out of your metagenome: computational analysis of environmental sequence data. *Curr. Opin. Microbiol.* 10, 490-498.
- Riesenfeld, C.S., Schloss, P.D., Handelsman, J., 2004. Metagenomics: genome analysis of microbial communities. *Microbiol. Mol. Biol. Rev.* 38, 525-552.
- Roossinck, M.J., Saha, P., Wiley, G.B., Quan, J., White, J.D., Lai, H., et al., 2010. Ecogenomics: using massively parallel pyrosequencing to understand virus ecology. *Mol. Ecol.* 19, 81-88.
- Rosario, K., Nilsson, C., Lim, Y.W., Ruan, Y., Breitbart, M., 2009. Metagenomic analysis of viruses in reclaimed water. *Environ. Microbiol.* 11, 2806-2820.
- Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S., et al., 2007. The sorcerer II global ocean sampling expedition: Northwest Atlantic through eastern tropical Pacific. *PLoS Biol.* 5, 0398-0431.
- Sanger, F., Nicklen, S., Coulson, A., 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* 74, 5463.
- Scheibye-Alsing, K., Hoffmann, S., Frankel, A., Jensen, P., Stadler, P.F., Mang, Y., et al., 2009. Sequence assembly. *Comput. Biol. Chem.* 33, 121-136.
- Schloss, P.D., Handelsman, J., 2005. Metagenomics for studying unculturable microorganisms: cutting the Gordian knot. *Genome Biol.* 6, 229.

- Schoenfeld, T., Patterson, M., Richardson, P.M., Wommack, K.E., Young, M., Mead, D., 2008. Assembly of viral metagenomes from Yellowstone hot springs. *Appl. Environ. Microbiol.* 74, 4164-4174.
- Schoenfeld, T., Liles, M., Wommack, K.E., Polson, S.W., Godiska, R., Mead, D., 2010. Functional viral metagenomics and the next generation of molecular tools. *Trends Microbiol.* 18, 20-29.
- Shendure, J., Ji, H., 2008. Next-generation DNA sequencing. *Nat. Biotechnol.* 26, 1135-1145.
- Simon, C., Daniel, R., 2009. Achievements and new knowledge unraveled by metagenomic approaches. *Appl. Microbiol. Biotechnol.* 85, 265-276.
- Snyder, L.A.S., Loman, N., Pallen, M.J., Penn, C.W., 2009. Next-generation Sequencing—the promise and perils of charting the great microbial unknown. *Microb. Ecol.* 57, 1-3.
- Strausberg, R.L., Levy, S., Rogers, Y., 2008. Emerging DNA sequencing technologies for human genomic medicine. *Drug Discov. Today* 13, 569-577.
- Streit, W.R., Schmitz, R.A., 2004. Metagenomics – the key to the uncultured microbes. *Curr. Opin. Microbiol.* 7, 492-498.
- Suttle, C.A., 2005. Viruses in the sea. *Nature* 437, 356-361.
- Tang, P., Chiu, C., 2010. Metagenomics for the discovery of novel human viruses. *Future Microbiol.* 5, 177-189.
- Thurber, R.V., Haynes, M., Breitbart, M., Wegley, L., Rohwer, F., 2009. Laboratory procedures to generate viral metagenomes. *Nat. Protoc.* 4, 470-483.
- Trapnell, C., Salzberg, S.L., 2009. How to map billions of short reads onto genomes. *Nat. Biotechnol.* 27, 455-457.
- Tringe, S.G., Rubin, E.M., 2005. Metagenomics: DNA sequencing of environmental samples. *Nat. Rev. Genet.* 6, 805-814.
- Tucker, T., Marra, M., Friedman, J.M., 2009. Massively parallel sequencing: the next big thing in genetic medicine. *Am. J. Hum. Genet.* 85, 142-154.
- Turcatti, G., Romieu, A., Fedurco, M., Tairi, A.P., 2008. A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. *Nucleic Acids Res.* 36, e25.
- Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., et al., 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 428, 37-43.
- Valdivia-Granda, W., 2008. The next meta-challenge for bioinformatics. *Bioinformatics.* 2, 358-362.
- Varshney, R.K., Nayak, S.N., May, G.D., Jackson, S.A., 2009. Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends Biotechnol.* 27, 522-530.
- Vega Thurber, R.L., Barott, K.L., Hall, D., Liu, H., Rodriguez-Mueller, B., Desnues, C., et al., 2008. Metagenomic analysis indicates that stressors induce production of herpes-like viruses in the coral *porites compressa*. *Proc. Natl. Acad. Sci. U.S.A.* 105, 18413-18418.
- Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., et al., 2004. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304, 66-74.
- Victoria, J.G., Kapoor, A., Li, L., Blinkova, O., Slikas, B., Wang, C., et al., 2009. Metagenomic analyses of viruses in stool samples from children with acute flaccid paralysis. *J. Virol.* 83, 4642-4651.

- Whitaker, R.J., Banfield, J.F., 2006. Population genomics in natural microbial communities. *Trends Ecol. Evol.* 21, 508-516.
- Wilhelm, B.T., Landry, J., 2009. RNA-Seq—quantitative measurement of expression through massively parallel RNA-sequencing. *Methods* 48, 249-257.
- Williamson, S.J., Rusch, D.B., Yooseph, S., Halpern, A.L., Heidelberg, K.B., Glass, J.I., et al., 2008. The sorcerer II global ocean sampling expedition: Metagenomic characterization of viruses within aquatic microbial samples. *PLoS One* 3, e1456.
- Willner, D., Furlan, M., Haynes, M., Schmieder, R., Angly, F.E., 2009. Metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis. *PLoS One* 4, e7370.
- Wooley, J.C., Godzik, A., Friedberg, I., 2010. A primer on metagenomics. *PLoS Comput. Biol.* 6, e1000667.
- Wooley, J.C., Ye, Y., 2009. Metagenomics: facts and artifacts, and computational challenges. *J. Comp. Sci. Technol.* 25, 71-81.
- Xu, J., 2006. Microbial ecology in the age of genomics and metagenomics: concepts, tools, and recent advances. *Mol. Ecol.* 15, 1713-1731.
- Zerbino, D.R., Birney, E., 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* 18, 821-829.
- Zhang, T., Breitbart, M., Lee, W.H., Run, J., Wei, C.L., Soh, S.W.L., et al., 2006. RNA viral community in human feces: prevalence of plant pathogenic viruses. *PLoS Biol.* 4, e3 0108-0118.
- Zhou, X.G., Ren, L.F., Li, Y.T., Zhang, M., Yu, Y.D., Yu, J., 2010. The next-generation sequencing technology: a technology review and future perspective. *Sci. China Life Sci.* 53, 44-57.

#### Internet resources

##### International Committee on Taxonomy of Viruses (2009):

<http://www.ictvonline.org/virusTaxonomy.asp?version=2009&bhcp=1> [accessed 15.06.2010]

##### Sequencing platforms:

**454/Roche:** <http://454.com> [accessed 12.04.2010]

**Illumina:** <http://www.illumina.com> [accessed 12.04.2010]

**ABI/SOLID:** <http://www3.appliedbiosystems.com> [accessed 12.04.2010]

**HelicoScope:** <http://www.helicosbio.com> [accessed 20.04.2010]

**Polonator:** [www.polonator.org](http://www.polonator.org) [accessed 20.04.2010]

**Pacific Biosciences:** [www.pacificbiosciences.com](http://www.pacificbiosciences.com) [accessed 20.04.2010]

**Nanopore sequencing:** <http://mcb.harvard.edu/branton/projects-NanoporeSequencing.htm>  
[accessed 20.04.2010]

##### Specific bioinformatic software tools:

**Illumina:** [http://www.illumina.com/software/genome\\_analyzer\\_software.ilmn#third\\_party\\_tools](http://www.illumina.com/software/genome_analyzer_software.ilmn#third_party_tools)  
[accessed 03.05.2010]

**454/Roche:** <http://454.com/products-solutions/analysis-tools/index.asp> [accessed 03.05.2010]

**Velvet:** <http://www.ebi.ac.uk/~zerbino/velvet> [accessed 10.07.2010]

**MAQ:** <http://MAQ.sourceforge.net/> [accessed 10.07.2010]

##### Lists of bioinformatic software tools:

<http://www.genomeweb.com/sequencing/moving-simulations-real-data-short-read-assemblers-startfacing>  
[accessed 20.06.2010]

[http://www.illumina.com/software/genome\\_analyzer\\_software.ilmn#third\\_party\\_tools](http://www.illumina.com/software/genome_analyzer_software.ilmn#third_party_tools) [accessed 20.06.2010]

<http://seqanswers.com/forums/showthread.php?t=43> [accessed 21.06.2010]

[http://mi.caspur.it/ngs/software\\_review.php](http://mi.caspur.it/ngs/software_review.php) [accessed 21.06.2010]

<http://sequenceassembly.com/> [accessed 21.06.2010]

---

## Chapter 3: Deep sequencing analysis of viruses infecting grapevines: Virome of a vineyard

---

Beatrix COETZEE , Michael-John FREEBOROUGH, Hans J MAREE, Jean-Marc CELTON,  
D Jasper G REES and Johan T BURGER (2010) *Virology* 400(2):157-163.

### 3.1 Abstract

Double-stranded RNA, isolated from 44 pooled randomly selected vines from a diseased South African vineyard, has been used in a deep sequencing analysis to build a census of the viral population. The dsRNA was sequenced in an unbiased manner using the sequencing-by-synthesis technology offered by the Illumina Genome Analyzer II and yielded 837 megabases of metagenomic sequence data. Four known viral pathogens were identified. It was found that Grapevine leafroll associated-virus 3 (GLRaV-3) is the most prevalent species, constituting 59% of the total reads, followed by Grapevine rupestris stem pitting-associated virus and Grapevine virus A. Grapevine virus E, a virus not previously reported in South African vineyards, was identified in the census. Viruses not previously identified in grapevine were also detected. The second most prevalent virus detected was a member of the *Chrysoviridae* family similar to *Penicillium chrysogenum* virus. Sequences aligning to two other mycoviruses were also detected.

### 3.2 Introduction

Grapevine is an important global commodity crop which is widely planted throughout temperate regions. Viruses, however, are a significant factor in reducing the quality and quantity of the yield and are known to reduce the productive life of vineyards (Martelli and Boudon-Padieu, 2006). Grapevine is subject to infection by more than 60 different viruses, the most known for any crop plant (Martelli and Boudon-Padieu, 2006). Most important grapevine virus diseases are caused by complexes of viruses, with up to nine different viruses having been identified in a single vine (Prosser et al., 2007). In South Africa, as in most grape-growing regions of the world, grapevine leafroll is regarded to be the most significant virus disease affecting grapevine, with Shiraz disease and Shiraz decline becoming more prominent as emerging diseases in the industry. Leafroll disease is well characterized and is associated with

up to 10 different viruses of the *Ampelovirus* genus, with Grapevine leafroll-associated virus 3 (GLRaV-3) being the most prominent and widely distributed virus associated with the disease. In South African vineyards it appears that Grapevine rupestris stem pitting-associated virus (GRSPaV) is most frequently associated with Shiraz decline (Goszczyński, pers. com.), whilst Grapevine virus A (GVA) seems to be most frequently associated with Shiraz disease (Goszczyński, 2007). The etiologies of these and most other grapevine virus diseases are not resolved. While these viruses have certainly been shown to be closely associated with the respective diseases, it is generally accepted that additional viruses are associated with these diseases. Moreover, anecdotal evidence exists for differing etiologies for the same disease, depending on geographical region, the stage of the growing season and the grapevine genotype.

Present disease diagnostics rely on ELISA or RT-PCR and target the viruses that have historically been associated with these diseases. While these tests are highly specific, they may not result in an accurate reflection of the etiological status of the tested plant, or of the particular disease, since none of the current diagnostic techniques address the potential contribution of other known or unknown viruses that may be involved in the etiology of a particular disease. Moreover, the error prone replication of RNA viruses leads to quasispecies, which can further complicate PCR-based detection assays as not all variants of the virus may be detected.

New and powerful technologies which are able to sequence viruses from environmental samples without the need for laborious and costly purification, cloning and screening techniques can result in the generation of sequence information for the complete virome in an unbiased fashion (Al Rwahnih et al., 2009; Cann et al., 2005; Kreuze et al., 2009; Williamson et al., 2008). These data could help to identify new viruses in disease complexes, identify the dominant variants of the viral species and give an indication of the frequency of viruses found in the infected material. Consequently, the data can be used to develop more accurate diagnostic assays. Next generation sequence data has been used successfully to evaluate the viruses present in a single grapevine plant displaying typical Syrah decline symptoms (Al Rwahnih et al., 2009), to identify an unknown virus in *Liatris spicata* (Adams et al., 2009) and to provide deep sequence analysis of virus infected sweetpotato plants (Kreuze et al., 2009). In these three studies, potential new viruses were discovered after analysis of the generated

sequence data. Al Rwahnih et al. (2009) generated long read sequences (LRSs) of ~200 bp and used BLAST analysis to confirm the expected presence of GRSPaV, to show the presence of rupestris vein-feathering virus and to identify a new marafivirus, grapevine syrah virus-1. Similarly, Adams et al. (2009) determined the sequence of Pepino mosaic virus from tomato using LRSs. These researchers also identified a new member of the *Cucumovirus* genus, Gayfeather mild mottle virus, from a diseased *Liatris* plant of unknown etiology. In a study by Kreuze et al. (2009) sweetpotato plants were inoculated with two known viruses singly or in combination. Sequence data obtained from short read sequences (SRSs) of ~24 bp yielded the expected full-length sequences of the inoculated viruses (Sweetpotato feathery mottle virus and Sweetpotato chlorotic stunt virus), as well as that of unknown viruses belonging to the *Badnavirus* and *Mastrevirus* genera. These three studies indicate the usefulness of next generation sequencing technologies and a metagenomics approach to identify known viruses and to discover possible new viruses from diseased plant material.

Adams et al. (2009) and Kreuze et al. (2009) concluded that next generation sequencing technologies can be used as a diagnostic tool to identify a plant virus when no prior knowledge of the virus is available. Therefore, it can be used as an investigative technique to generate sequences of multiple viruses, if present, in an unbiased fashion due to the use of non sequence specific primers. This could lead to the identification of the viruses and a better understanding of the disease especially when its etiology is currently unknown.

In this paper we describe the use of sequencing-by-synthesis technology ([http://www.illumina.com/technology/sequencing\\_technology.ilmn](http://www.illumina.com/technology/sequencing_technology.ilmn)) on the massively parallel Illumina Genome Analyzer II, to sequence an environmental sample composed of 44 randomly selected vines, to determine the viral profile of a severely diseased vineyard.

### **3.3 Results**

#### **3.3.1 Sequencing**

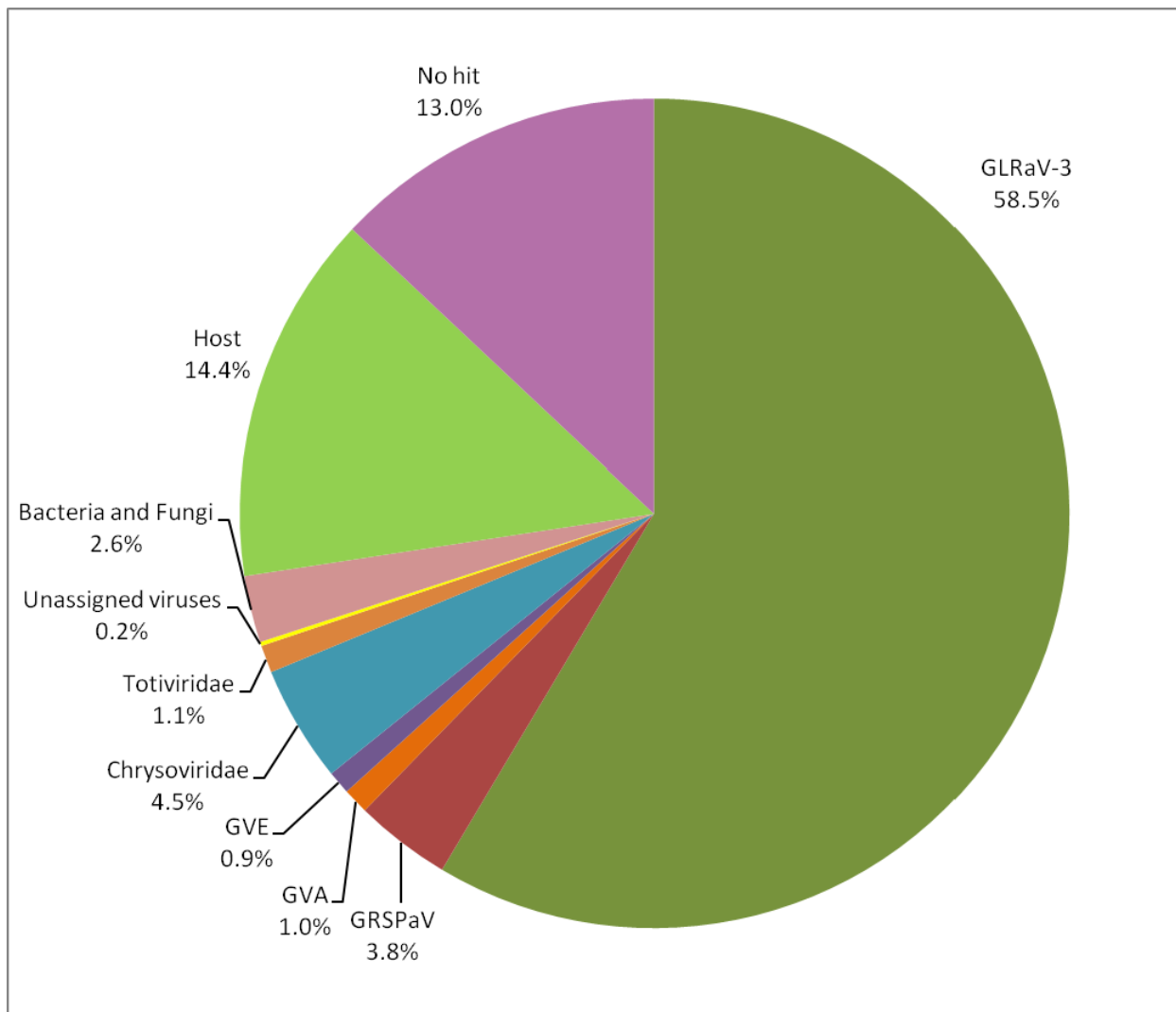
Double-stranded RNA was isolated from lignified cane material pooled from 44 vines that were randomly selected in a *Vitis vinifera* cv. Merlot vineyard. Integrity of the dsRNA was confirmed by diagnostic RT-PCR amplification of GLRaV-3, GRSPaV and GVA (data not shown).



The dsRNA was prepared for cDNA synthesis followed by sequencing. cDNA fragments of ~200 bp were selected for PCR enrichment. Quantitative PCR was used to determine the optimal amplification of the ~200 bp cDNA fragments (Supplementary data 1) and 33 PCR cycles were used for subsequent cDNA amplification. Sequencing-by-synthesis was performed using the Illumina Genome Analyzer II with 51 cycles. Approximately 13 million clusters were obtained on a single sequencing lane of the Illumina flow cell. Approximately 74% of these clusters could be analyzed and yielded quality sequence data. The paired-end sequence data yielded 19,247,026 reads, which translates to more than 837 megabases of sequence data from 1/8<sup>th</sup> of an Illumina flow cell.

### **3.3.2 *De novo* sequence assembly and analysis**

Reads were assembled into scaffolds using the Velvet 0.7.31 *de novo* assembly algorithm (Zerbino and Birney, 2008; <http://www.ebi.ac.uk/~zerbino/velvet>). A variety of parameters were tested, yielding scaffolds that varied between 100 and 8,624 nt in length. Parameters for optimal assembly were selected based on number and length of the scaffolds obtained. The following parameters were used for all further analyses: hash length of 23, coverage cut-off of 50, expected coverage of 1,000 and a minimum scaffold length of 100. In this assembly, 7,895,103 reads (41%) assembled into 449 scaffolds (Supplementary data 2). Forty-eight of the scaffolds were larger than 1,000 nt and the largest scaffold was 8,624 nt in length. These scaffolds were subjected to BLAST (Basic Local Alignment Search Tool) (Altschul et al., 1997) searches against the NCBI non-redundant (nr) DNA and protein databases and classified according to the sequences they aligned to with the highest bit score. The scaffolds aligned to GLRaV-3 (124), GRSPaV (1) and GVA (7), which represent 59%, 4% and 1% respectively of the analyzed read data. Grapevine virus E (GVE) was also identified, with 2 scaffolds aligning, accounting for 1% of the read data (Figure 3.1). A single scaffold aligned to GVB. However, due to the low homology (47% amino acid identity) and read content of this scaffold, it was omitted from Figure 3.1.



**Figure 3.1** Comparative percentages for read counts utilized in scaffolds for each sequence classification according to best hit with BLASTn or BLASTx searches. **GLRaV-3** Grapevine leafroll-associated virus 3, **GRSPaV** Grapevine rupestris stem pitting-associated virus, **GVA** Grapevine virus A and **GVE** Grapevine virus E

Furthermore, BLAST searches against the non-redundant databases detected the presence of three mycoviral families. Twenty-six scaffolds, accounting for 5% of the assembled read data aligned to the fungi-infecting dsRNA viruses from the family *Chrysoviridae*: *Penicillium chrysogenum* virus (PcV), *Helminthosporium victoriae* 145S virus and *Cryphonectria nitschkei* chrysovirus 1. Two scaffolds (1.1% of read data) aligned to *Aspergillus mycovirus* 1816, a member of the family *Totiviridae* and three to the unassigned *Fusarium graminearum* dsRNA mycovirus 4 (0.2% of read data). Twenty-two of the 31 scaffolds aligning to these viruses aligned at a low homology and were only identified when searching against the NCBI (nr) protein database. Bacterial, fungal and host-derived scaffolds were also identified but will not be discussed in further detail. A further 188 scaffolds representing 13% of analyzed read data

did not align to any known sequences in the NCBI databases. Sequence classifications, the number of scaffolds and read counts incorporated in scaffolds of the dominant viruses, are listed in Table 3.1 (Velvet *de novo* assembly). To give an indication of relative abundance, the percentage read counts in each sequence classification is given in Figure 3.1.

### **3.3.3 Re-assembly against reference sequences**

Twenty three representative variants for the five dominant viral species identified in BLAST searches, were extracted from the NCBI database, and used as reference sequences in further re-assembly studies using the Mapping and Assembly with Quality (MAQ) assembler (Li et al., 2008; <http://MAQ.sourceforge.net/>). Complete genome sequences of GLRaV-3, GRSPaV, GVA and PcV were used, and a partial sequence of GVE, since the genome for the specific strain with the highest sequence identity has not been fully sequenced yet. The extracted sequences for the 23 representative variants were used simultaneously as reference sequences in a MAQ re-assembly analysis to determine the dominant variant for each virus species based on the read count. A total of 3,860,942 reads accounting for approximately 20% of the total read data aligned to the reference sequences (Table 3.1). To verify the presence of these variants, reference sequences were used individually for re-assembly analysis to calculate the read count, average depth of sequence and the average coverage of the respective genomes (Table 3.1). The most prominent variants for each virus species could be selected based on these criteria (Figure 3.2). However, the reference strains used in the assemblies are only representative of the viral groups and not identical to the variants detected in this study, therefore the random distribution of reads on the genomes with higher read counts are observed in areas with high sequence homology between the different viral variants. Data from individual re-assemblies were used for further analyses.

**Table 3.1** Comparison of *de novo* and re-assembly data for the five dominant virus species identified in this study. *De novo* assembled scaffolds are classified according to best alignment (highest bit score) in the NCBI database found with BLASTn and BLASTx searches. MAQ re-assembly data are shown for the 23 representative variants identified after *de novo* assembly analysis.

	Velvet <i>de novo</i> assembly					MAQ re-assembly					
	GenBank accession number	Strain <sup>a</sup>	No of scaffolds	Read count <sup>b</sup>	% read count	Simultaneous re-assembly		Individual re-assemblies			
						Read count	% read count	Read count	% read count	Average depth of sequence	Genome coverage
<b>GLRaV-3</b>	EU259806 <sup>c</sup>	GP18	66	2,608,947	56%	3,616,585	98%	4,242,321	54%	10,009	100%
	EU344893 <sup>c</sup>	CI-766	25	1,048,903	23%	4,673	0%	1,678,187	21%	3,951	90%
	EF508151 <sup>d</sup>		11	487,297	11%	-	-	-	-	-	-
	AF037268 <sup>c</sup>	NY1	6	120,333	3%	66,943	2%	1,922,950	25%	4,558	92%
	AJ606358 <sup>d</sup>		2	132,454	3%	-	-	-	-	-	-
	DQ314610 <sup>d</sup>		2	32,232	1%	-	-	-	-	-	-
	AJ748516 <sup>d</sup>		1	47,944	1%	-	-	-	-	-	-
	EF103903 <sup>d</sup>		1	26,255	1%	-	-	-	-	-	-
	AY704412 <sup>d</sup>		1	25,796	1%	-	-	-	-	-	-
	AY495340 <sup>d</sup>		1	20,226	0%	-	-	-	-	-	-
	AJ748536 <sup>d</sup>		1	17,824	0%	-	-	-	-	-	-
	EU344896 <sup>d</sup>		1	17,542	0%	-	-	-	-	-	-
	DQ911148 <sup>d</sup>		1	11,343	0%	-	-	-	-	-	-
	EF445656 <sup>d</sup>		1	1,728	0%	-	-	-	-	-	-
	AJ748520 <sup>d</sup>		1	1,189	0%	-	-	-	-	-	-
	FJ786016 <sup>d</sup>		1	801	0%	-	-	-	-	-	-
	DQ780887 <sup>d</sup>		1	740	0%	-	-	-	-	-	-
	ABY87025 <sup>e</sup>		1	20,397	0%	-	-	-	-	-	-
			<b>124</b>	<b>4,621,195</b>	<b>100%</b>	<b>3,688,201</b>		<b>7,843,458</b>	<b>100%</b>		
<b>GRSPaV</b>	AY881626 <sup>c</sup>	SG1	1	296,352	100%	120,654	94%	135,421	61%	653	90%
	AY881627 <sup>c</sup>	BS	ND	ND	ND	4,251	3%	16,696	8%	76	35%
	AF057136 <sup>c</sup>	GRSPaV-1	ND	ND	ND	1,605	1%	32,075	14%	152	56%
	AF026278 <sup>c</sup>	GRSPaV-1	ND	ND	ND	1,462	1%	31,929	14%	150	55%
	AY368590 <sup>c</sup>	SY	ND	ND	ND	163	0%	3,071	1%	14	10%
	AY368172 <sup>c</sup>	PN	ND	ND	ND	151	0%	2,582	1%	12	10%
			<b>1</b>	<b>296,352</b>	<b>100%</b>	<b>128,286</b>		<b>221,774</b>			
<b>GVA</b>	X75433 <sup>c</sup>	Is151	4	33,672	43%	1,230	4%	2,533	4%	14	22%
	DQ855088 <sup>c</sup>	P163-1	3	45,444	57%	12,773	41%	26,190	38%	149	94%
	DQ787959 <sup>c</sup>	GTR1-1	ND	ND	ND	16,217	52%	28,892	42%	166	93%
	DQ855084 <sup>c</sup>	GTG11-1	ND	ND	ND	685	2%	2,606	4%	14	24%
	DQ855086 <sup>c</sup>	GTR1-2	ND	ND	ND	196	1%	1,169	2%	6	10%
	DQ855082 <sup>c</sup>	P163-M5	ND	ND	ND	83	0%	1,122	2%	6	10%
	DQ855081 <sup>c</sup>	GTR1SD-1	ND	ND	ND	12	0%	1,007	1%	5	12%
	DQ855083 <sup>c</sup>	KWVMo4-1	ND	ND	ND	12	0%	1,152	2%	6	9%
	DQ855087 <sup>c</sup>	BMO32-1	ND	ND	ND	7	0%	1,313	2%	7	11%
	AF007415 <sup>c</sup>	PA3	ND	ND	ND	0	0%	1,084	2%	6	10%
	AY244516 <sup>c</sup>		ND	ND	ND	0	0%	1,084	2%	6	10%
			<b>7</b>	<b>79,116</b>	<b>100%</b>	<b>31,215</b>		<b>68,152</b>			
<b>GVE</b>	AB432910 <sup>c</sup>	TvAQ7	1	69,780	97%	18	0%	160	1%	1	3%
	AB432911 <sup>d</sup>	TvP15	1	1,898	3%	13,204	100%	13,207	99%	178	99%
			<b>2</b>	<b>71,678</b>	<b>100%</b>	<b>13,222</b>		<b>13,367</b>			
<b>PcV</b>	AF296439 <sup>d</sup>		6	84,403	23%	-	-	-	-	-	-
	AF296442 <sup>d</sup>		1	751	0%	-	-	-	-	-	-
	AAM95601 <sup>e</sup>		6	9,025	3%	-	-	-	-	-	-
	AAM95604 <sup>e</sup>		5	122,873	34%	-	-	-	-	-	-
	AAM95602 <sup>e</sup>		3	70,909	20%	-	-	-	-	-	-
	AAM95603 <sup>e</sup>		2	54,690	15%	-	-	-	-	-	-
	AAM68955 <sup>e</sup>		1	14,017	4%	-	-	-	-	-	-
	AAM68953 <sup>e</sup>		1	1,973	1%	-	-	-	-	-	-
	ACT79258 <sup>e</sup>		1	537	0%	-	-	-	-	-	-
	NC_007539 <sup>c</sup>	Seg 1 ATCC 9480	ND	ND	ND	0	0%	0	0%	0	0%
	NC_007540 <sup>c</sup>	Seg 2 ATCC 9480	ND	ND	ND	0	0%	0	0%	0	0%
	NC_007541 <sup>c</sup>	Seg 3 ATCC 9480	ND	ND	ND	0	0%	0	0%	0	0%
	NC_007542 <sup>c</sup>	Seg 4 ATCC 9480	ND	ND	ND	0	0%	0	0%	0	0%
			<b>26</b>	<b>359,178</b>	<b>100%</b>	<b>0</b>		<b>0</b>			
<b>Total read count</b>						<b>3,860,924</b>		<b>8,146,751</b>			

**GLRaV-3** Grapevine leafroll-associated virus 3

**GRSPaV** Grapevine rupestris stem pitting-associated virus

**GVA** Grapevine virus A

**GVE** Grapevine virus E

**PcV** Penicillium chrysogenum virus

ND – not detected

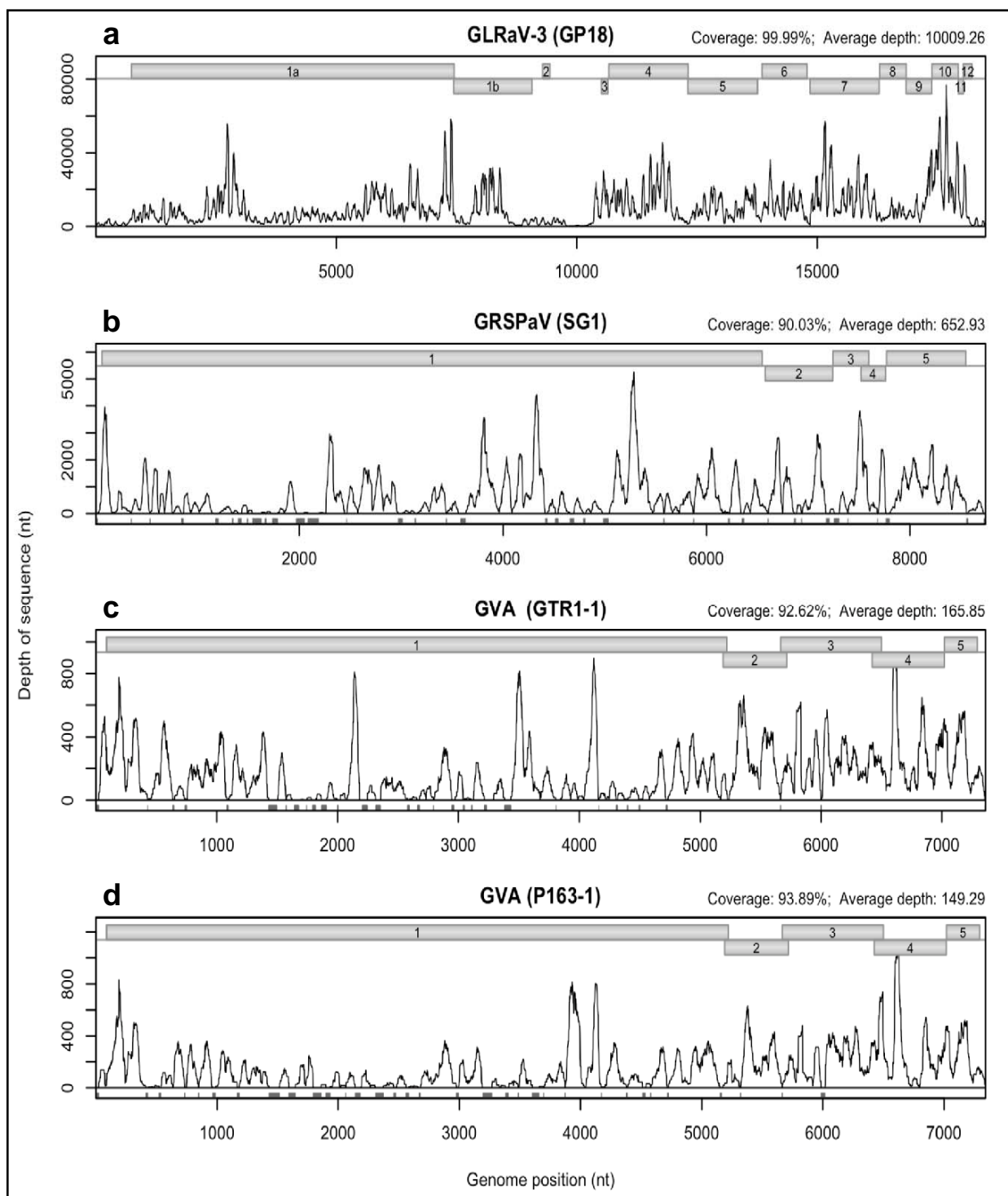
<sup>a</sup> Strain names only shown for sequences used in MAQ.

<sup>b</sup> Number of reads incorporated in scaffolds

<sup>c</sup> Complete genome sequence

<sup>d</sup> Partial genome sequence

<sup>e</sup> Protein sequence



**Figure 3.2** MAQ-reassembly of reads on four full-length genomes representing the dominant variants for **a)** GLRaV-3 (GP18), **b)** GRSPaV (SG1), **c)** GVA (GTR1-1) and **d)** GVA (P163-1). GVE was excluded due to the lack of a full-length genome. Schematic representations of virus genomes with numbered open reading frames are shown above graphs. Grey bars below graph highlight areas with no coverage. **GLRaV-3** Grapevine leafroll-associated virus 3, **GRSPaV** Grapevine rupestris stem pitting-associated virus, **GVA** Grapevine virus A

The GLRaV-3 variant GP18 had the highest average depth of sequence (10,009 nt coverage) followed by GRSPaV SG1 (653 nt), GVE TVP15 (178 nt) and GVA variants GTR1-1 (165 nt) and P163-1 (149 nt) (Table 3.1, Figure 3.2). The genome coverage for the four full-length genomes of the dominant variants were 100% for GLRaV-3 (GP18), 90% for GRSPaV (SG1), 93% for GVA variant GTR1-1 and 94% for P163-1 (Table 3.1, Figure 3.2). GVE was excluded due to the lack of a full length reference genome. The average depth of sequence and genome coverage of PcV was low due to the poor sequence similarity to the reference sequence.

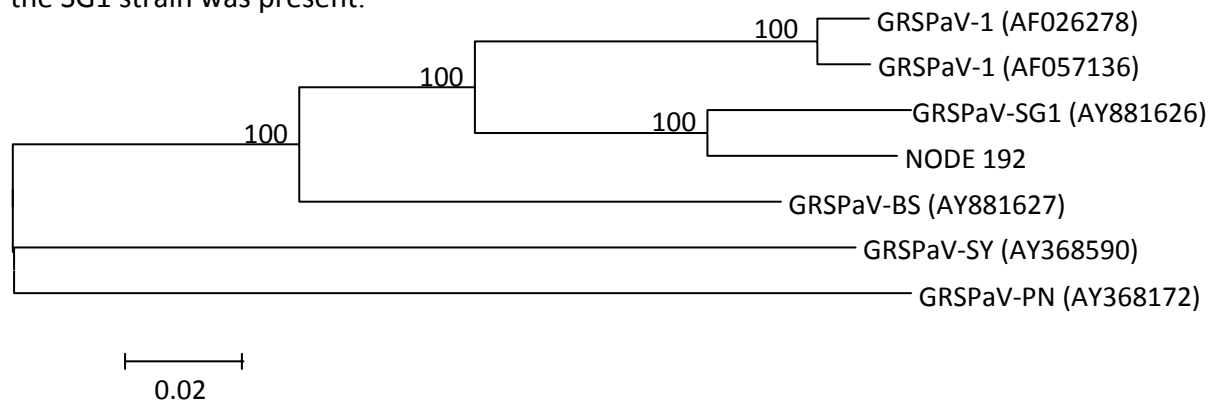
Despite the use of a dsRNA-specific extraction procedure, host nucleic acids were also sequenced. Ten percent of the total reads aligned to the 19 *Vitis vinifera* chromosomes (GenBank accession no: NC\_012007 – NC\_012025), chloroplast (GenBank accession no: NC\_007957) and mitochondrial sequences (GenBank accession no: NC\_012119) with an average depth of sequence of 0.4 nt and genome coverage of 0.4%. This implies that the dsRNA purification protocol is an efficient enrichment procedure, but does not totally exclude all host single-stranded RNA.

### **3.4 Discussion**

Both *de novo* and re-assembly analyses showed the GP18 variant of GLRaV-3 (Maree et al., 2008) to be the most abundant isolate of all the viruses identified in this sample. Sixty-six of the 124 *de novo* assembled scaffolds (56% of the GLRaV-3 read data) aligned preferentially to the GP18 variant. The 58 remaining scaffolds aligned to 17 different GLRaV-3 sequences available in the NCBI database (Table 3.1). The lack of a full genome length scaffold for GLRaV-3 could be due to the diversity observed within the scaffolds and the possibility of two or more variants of GLRaV-3 being present in the sample. Re-assembly analysis confirmed the dominance of a GP18-like variant, as the highest number of reads could be re-assembled against this reference sequence, with 100% genome coverage and an average depth of sequence of 10,009 nt (Table 3.1, Figure 3.2A). After re-assembly analysis, data suggested that two additional GLRaV-3 variants (CI-766 and NY1) were present based on the read count, average depth of sequence and genome coverage (Table 3.1). These two variants share 97.6% identity, which suggested that one GLRaV-3 variant similar to the NY1/CI-766 variants might be present in the vineyard. Moreover, the high variability noted in BLASTn pairwise analysis suggests that the identification

of new grapevine leafroll-associated ampeloviruses from this sequence data cannot be excluded (Supplementary data 3).

A near full-length GRSPaV scaffold (8,624 nt in length, accounting for 296,352 reads) was generated in the *de novo* assembly, BLASTn alignments indicated that this viral variant shares 92.3% sequence identity with the SG1 variant (Figure 3.3, Supplementary data 3) (Meng et al., 2005). Re-assembly of the reads against GRSPaV full length genomes confirmed that the GRSPaV is similar to the SG1 variant due to the read count (135,421 reads), average depth of sequence (653 nt) and 90% genome coverage (Table 3.1, Figure 3.2B). The GRSPaV scaffold, Node 192, had complete coverage, showing that a GRSPaV variant similar but not identical to the SG1 strain was present.



**Figure 3.3** Phylogenetic tree (bootstrap consensus tree) showing the relationship between the six complete genome sequences and the *de novo* generated scaffold (Node 192) for Grapevine rupestris stem pitting-associated virus (GRSPaV). Node 192 group with the SG1 (AY881626) strain. GenBank accession numbers are indicated in brackets. Bootstrap values (500 replicates) are indicated above the branches. The scale indicates number of substitutions per base position.

Analysis of the GVA scaffolds indicated that a variant in the molecular group III (Goszczyński et al., 2008) is present in this vineyard. Seven *de novo* assembled scaffolds, accounting for 79,116 reads, aligned to GVA variants. Three scaffolds, including a scaffold of 5,826 nt, aligned to the P163-1 variant in molecular group III, and four scaffolds aligned to the Is151 strain in molecular group I. Re-assemblies against the molecular group III variants, P163-1 and GTR1-1 (Goszczyński et al., 2008) were comparable with respect to read count, average depth of sequence, genome coverage (Table 3.1, Figure 3.2C and D) and BLASTn pairwise alignment of scaffolds (Supplementary data 3). This suggests that a molecular group III variant of GVA is present. The four *de novo* assembled scaffolds which aligned preferentially to a GVA variant in molecular group I were not confirmed by re-assembly analysis as the genome coverage, depth of

sequence and number of reads aligning to variants from this molecular group were poor (Table 3.1). These results lead us to propose that only a GVA variant of molecular group III is present in the diseased vineyard.

The sequence data identified a fourth virus, GVE, which has not previously been detected in South Africa. Subsequent PCR amplification with diagnostic primers, confirmed the presence of GVE in the samples from this vineyard. Two scaffolds were obtained for GVE, the largest being 5,172 nt in length, which aligned to and extends the partial sequence of the TvP15 strain available in the NCBI database (Nakaune et al., 2008). Although one of the scaffolds aligned with a higher bit score to the TvAQ7 strain, probably because the TvAQ7 sequence is essentially complete and allows for a larger part of the scaffold to align, the scaffold had higher homology to the TvP15 partial sequence (Supplementary data 3). This suggests that read data can be assembled into scaffolds that can align to known virus sequences but can also extend the sequences of partially sequenced viral genomes, expanding existing sequence data. The presence of the TvP15 isolate was confirmed with re-assembly analysis as this GVE isolate had significantly higher read count, average depth of sequence and genome coverage than the GVE TvAQ7 strain (Table 3.1). This also shows that care must be taken when assigning the data to specific strains depending on already sequenced complete genomes.

A total of 5% of the analyzed read data aligned to members of the *Chrysoviridae* family, making it the second most abundant virus group isolated from this vineyard after GLRaV-3. Homology to known members of the *Chrysoviridae* was too low to align at the nucleotide level apart from seven scaffolds, 19 scaffolds were only revealed when searching the NCBI (nr) protein database. The yet unidentified chrysovirus that we detected could not be classified beyond family level with either *de novo* assembly or reassembly analysis. These viruses are divergent from all other members of the family that have been sequenced to date. The PcV reference sequences used in the analysis did not correspond at the nucleotide level to the member(s) of the *Chrysoviridae* detected in this sample (Supplementary data 3) and thus no reads aligned to PcV in MAQ re-assembly analysis (Table 3.1). The role of these viruses in plant disease is unknown at present. It is also not clear whether members of this family use plants or fungi as the primary host. We propose that these viruses can be isolated from grapevine phloem tissue, implying that plants may act as host to the isolated virus. However, the possibility that an as yet unidentified fungal



host may exist within the grapevine cannot be excluded, we therefore propose that the possible role that these dsRNA viruses play in plant diseases be investigated further. Similar ambiguous results were obtained with Amasya cherry disease (Covelli et al., 2004).

To estimate the relative abundance of the four dominant virus species in the diseased vineyard, the average depth of sequence and the number of re-assembled reads were normalized against the genome length of the respective reference sequences. When considering only these four virus species, results indicated that GLRaV-3 accounts for 90%, GRSPaV for 7% and GVA and GVE for 1.5% each of the known virus infection of this severely diseased vineyard. These percentages are similar to the relative abundance of the seven viruses when using the read counts in the *de novo* assemblies for the virus-associated scaffolds only. From this analysis GLRaV-3 accounts for 84%, unidentified *Chrysoviridae* for 7%, GRSPaV for 5%, GVA for 1%, GVE for 1%, *Totiviridae* for 2% and *Fusarium graminearum* dsRNA mycovirus 4 for less than 1% of the total virus population within the vineyard sample.

In total 188 scaffolds did not align to any known GenBank sequences in either of the non-redundant DNA or protein databases. Sequencing and/or assembly errors may have contributed to the unknown status of these scaffolds. Fifty-three of these scaffolds have more than 10% unassigned nucleotides (N's) in the sequence. The remaining 135 scaffolds could be of viral, bacterial or fungal origin and need to be investigated further. In total, 41% of read data could be analyzed. This is comparable to the 50% of data with no known homology to other submitted sequences found by another metagenomic study (Kerkhof and Goodman, 2009)

### **3.6 Conclusion**

We were able to show that sequencing-by-synthesis, using a massively parallel sequencing platform, generated read data of high quantity and quality from low amounts of viral dsRNA, extracted from infected grapevine material, to provide the first virome of a diseased vineyard. Analysis of read data identified four different viruses that infected these grapevines, as well as their most prominent variants. The data further showed the presence of at least three unknown viruses that could be assigned to different virus families.

Both *de novo* assembly of reads and re-assembly on reference sequences indicated that GLRaV-3 was the dominant virus present in the sample, accounting for greater than 84% of the viral population of the diseased vineyard based on read count in *de novo* assemblies. Assembly analysis was able to identify the dominant variant of GLRaV-3 and the presence of at least one more GLRaV-3 minor variant in this vineyard. Similarly, a dominant GVA variant within molecular group III was identified. With *de novo* assembly a single scaffold spanning 99% of the GRSPaV genome was generated with high sequence identity to the SG1 strain. We also report the presence of GVE in the vineyard and have extended the incomplete sequence of the TvP15 strain of this virus (Node 3404, supplementary data 2).

This is the first report of sequence data for members of the dsRNA *Chrysoviridae* family isolated from grapevine. The high read counts that aligned with the *Chrysoviridae* isolate(s) suggest that these viruses were isolated directly from the phloem material of grapevine. The current analysis of sequence data also detected the presence of viruses similar to a member of the *Totiviridae* and the unclassified dsRNA virus, *Fusarium graminearum* dsRNA mycovirus 4. However, classifications and identification of these mycoviruses might become more specific as new sequence data became available on the NCBI database.

The use of LRSs generated by 454 technology has been the *de facto* choice for metagenomics and *de novo* sequencing of viral genomes. This is the first report that indicates that *de novo* and re-assembly analysis of short sequence reads generated from dsRNA can successfully distinguish between four different known viral species and gives an indication of the dominant variants of these species. Furthermore, read data analysis helped to detect at least three unknown viruses from viral families that have not previously been reported to infect grapevine, resulting in the first virome of a severely diseased vineyard being sequenced. Our results prove the feasibility of next generation high-throughput sequencing technology using Illumina technology and dsRNA as starting material in a metagenomics approach to determine the virome of a severely diseased vineyard and suggest that this approach can be used to elucidate the etiologies of the world's notorious grapevine virus diseases.

## **3.7 Materials and methods**

### **3.7.1 Plant material**

Plant material was sourced from a severely diseased vineyard (cv. Merlot) in the Stellenbosch region of South Africa. Plants were randomly selected during winter dormancy when no apparent symptoms were visible. Phloem scrapings were collected from 44 vines during August 2008. During the following growth season, symptoms were observed of these selected vines (Supplementary data 4). Vines in this severely diseased vineyard displayed typical and atypical leafroll and Shiraz disease symptoms. These symptoms include reddening of leaves while veins remain green, downwards rolling of the leaves, canes with a lack of lignification, swelling at the graft union and reduced vigor. A number of asymptomatic vines were also observed. Double-stranded RNA was isolated from the samples using a cellulose extraction protocol adapted from Valverde et al. (1990). Reverse transcriptase PCR screening for the most prevalent viruses (GLRaV-3, GRSPaV and GVA) was performed to confirm the quality of the dsRNA.

### **3.7.2 Sequencing**

Sequencing was performed on the Illumina Genome Analyzer II. Fifty nanograms of dsRNA was subjected to a heat denaturation step at 95°C for 10 min and flash cooled in ice water. Fragmentation, conversion to cDNA and preparation for sequencing was performed using the Illumina mRNA Sequencing v2 kit. The 15 PCR cycles recommended by Illumina were insufficient for cDNA amplification. Quantitative-PCR was used to determine an optimal number of amplification cycles. After comparison to positive control samples, 33 cycles of PCR amplification was found to be the midpoint of the log phase of amplification for the cDNA (Supplementary data 1). PCR enrichment of the adapter-ligated cDNA fragments was therefore conducted with 33 cycles of amplification for further sequence determination. The enriched DNA template was quantified, diluted to 2.5pM and hybridized to 1/8<sup>th</sup> of a flow cell. Clusters were generated on the flow cell using the Illumina Paired-end Cluster Generation v2 kit. The flow cell with the DNA clusters was subjected to sequencing on the Illumina Genome Analyzer II using the SBS Sequencing v3 kit.

### **3.7.3 Sequence analysis**

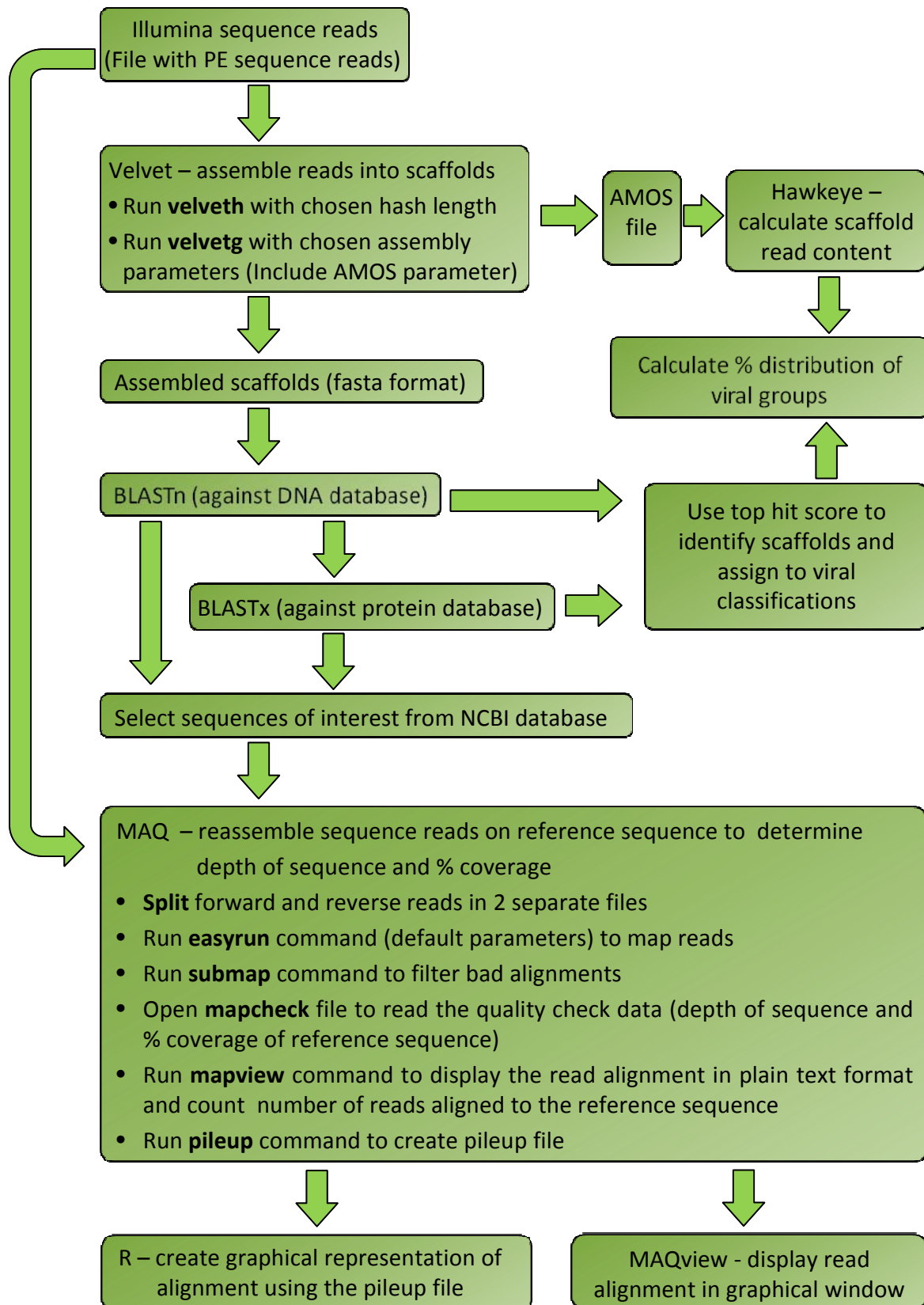
Paired-end (PE) sequence data were assembled into scaffolds using the short read *de novo* assembler Velvet 0.7.31 (Zerbino and Birney, 2008; <http://www.ebi.ac.uk/~zerbino/velvet>).

Parameters used were: hash length of 23, coverage cut-off of 50, expected coverage of 1,000 and a minimum scaffold length of 100. First a hash table, depending on the hash length, is created using the executable “velveth”. Scaffolds are then assembled with the “velvetg” executable according to the chosen parameters. Velvet generates an AMOS file containing the statistical information on the scaffolds. This file was viewed in with the assembly viewer Hawkeye (Schatz et al., 2007; <http://sourceforge.net/apps/mediawiki/amos/index.php?title=Hawkeye>) to calculate the number of reads assembled in each scaffold.

The assembled scaffolds were used to search the NCBI DNA database using the BLAST algorithm (Basic Local Alignment Search Tool) (Altschul et al., 1997). With BLASTn the nucleotide database is searched with a nucleotide query. Scaffolds were classified according to the GenBank entry found with the highest hit score. Default parameters were used and an expect value of less than  $10^{-5}$  was regarded as a significant hit. Scaffolds with no significant hit were used in BLASTx searches. BLASTx searches the NCBI protein database, translating the query sequence in all six possible open reading frames (ORF). This search is therefore able to align sequences on amino acid level, which are too diverse to align on nucleotide level. Percentage distribution of viral and other classifications were made based on the BLAST results and the read content of each scaffold, as calculated from the AMOS file. However, this annotation is known to be incomplete as many viruses are not yet identified or their genome sequence not determined and therefore not made available in public databases such as GenBank.

Based on the BLAST results, sequences of interest were selected from the NCBI database and used as reference sequences in subsequent analyses. Mapping and Assembly with Quality (MAQ) reference assembler 0.7.1 (Li et al., 2008; <http://MAQ.sourceforge.net/>) was used in the re-assembly analysis. Default parameters were used with the “easyrun” command to map the sequence reads to the reference sequences. The “submap” command was used to filter the alignment data and only reads aligning to the references with a high confidence are reported on. The number of reads aligning to the reference sequences, coverage and average depth were determined by viewing the “mapview” and “mapcheck” files for each alignment. The assembly viewer MAQview, included in the MAQ package, was used to display the MAQ generated alignment. A script run in the R statistical software environment ([53](http://www.r-</a></p></div><div data-bbox=)

project.org/) was used to create a graphical representation from the “pileup” files generated for each alignment. The graphs show the coverage and depth of sequence for selected reference sequences (Figure 3.2 and Figure 4.1). Figure 3.4 is a diagram illustrating the bioinformatics workflow used to analyze the short read sequence data.



**Figure 3.4** Diagram to illustrate bioinformatics workflow used to analyze the Illumina short read sequence data. The various bioinformatic software tools and commands used in the analyses are shown (**PE** Paired-end).

## Acknowledgments

The financial assistance of Winetech and the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the authors and are not necessarily to be attributed to the NRF. The authors also thank Paul Krige (Kanonkop Wine Estate, Stellenbosch, South Africa) for sample material. Anthony la Grange's assistance with graphical presentation of data is appreciated.

## 3.8 References

- Adams, I.P, Glover, R.H, Monger, W.A, Mumford, R., Jackeviciene, E., Navalinskiene, M., et al., 2009. Next-generation sequencing and metagenomic analysis: a universal diagnostic tool in plant virology. *Mol. Plant Pathol.* 10, 537-545.
- Al Rwahnih, M., Daubert, S., Golino, D., Rowhani, A., 2009. Deep sequencing analysis of RNAs from a grapevine showing Syrah decline symptoms reveals a multiple virus infection that includes a novel virus. *Virology* 387, 395-401.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., et al., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389-3402.
- Cann, A.J., Fandrich, S.E., Heaphy, S., 2005. Analysis of the virus population present in equine faeces indicates the presence of hundreds of uncharacterized virus genomes. *Virus Genes* 30, 151-156.
- Covelli, L., Coutts, R.H.A., Serio, F., Citir, A., Açıkgöz, S., Hernández, C., et al., 2004. Cherry chlorotic rusty spot and Amasya cherry diseases are associated with a complex pattern of mycoviral-like double-stranded RNAs. I. Characterization of a new species in the genus *Chrysovirus*. *J. Gen. Virol.* 85, 3389-3397.
- Goszczynski, D.E., 2007. Single-strand conformation polymorphism (SSCP), cloning and sequencing reveal a close association between related molecular variants of Grapevine virus A (GVA) and Shiraz disease in South Africa. *Plant Pathol.* 56, 755-762.
- Goszczynski, D.E., du Preez, J., Burger, J.T., 2008. Molecular divergence of Grapevine virus A (GVA) variants associated with Shiraz disease in South Africa. *Virus Res.* 138, 105-110.
- Kerkhof, L.J., Goodman, R.M., 2009. Ocean microbial metagenomics. *Deep-Sea Res. II.* 56, 1824-1829.
- Kreuze, J.F, Perez, A., Untiveros, M., Quispe, D., Fuentes, S., Barker, I., et al., 2009. Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: a generic method for diagnosis, discovery and sequencing of viruses. *Virology* 388, 1-7.
- Li, H., Ruan, J., Durbin, R., 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* 18, 1851-1858.
- Maree, H.J, Freeborough, M.-J., Burger, J.T, 2008. Complete nucleotide sequence of a South African isolate of Grapevine leafroll-associated virus 3 reveals a 5' UTR of 737 nucleotides. *Arch. Virol.* 153, 755-757.
- Martelli, G.P, Boudon-Padieu, E., 2006. Directory of infectious diseases of grapevines and viroses and virus-like diseases of the grapevine: Bibliographic report 1998-2004. *Options Méditerran., Ser. B, Stud. Res.* 55, CIHEAM, 279.
- Meng, B., Li, C., Wang, W., Goszczynski, D., Gonsalves, D., 2005. Complete genome sequences of two new variants of Grapevine rupestris stem pitting-associated virus and comparative analyses. *J. Gen. Virol.* 86, 1555-1560.

- Nakaune, R., Toda, S., Mochizuki, M., Nakano, M., 2008. Identification and characterization of a new vitivirus from grapevine. *Arch. Virol.* 153, 1827-1832.
- Prosser, S.W., Goszczynski, D.E, Meng, B., 2007. Molecular analysis of double-stranded RNAs reveals complex infection of grapevines with multiple viruses. *Virus Res.* 124, 151-159.
- Schatz, M.C., Phillippy, A.M., Shneiderman, B., Salzberg, S.L., 2007. Hawkeye: An interactive visual analytics tool for genome assemblies. *Genome Biol.* 8, R34.1-R34.12.
- Valverde, R.A., Nameth, S.T., Jordan, R.L., 1990. Analysis of double-stranded RNA for plant virus diagnosis. *Plant Dis.* 74, 255-258.
- Williamson, S.J., Rusch, D.B., Yooseph, S., Halpern, A.L., Heidelberg, K.B., Glass, J.I., et al., 2008. The sorcerer II global ocean sampling expedition: metagenomic characterization of viruses within aquatic microbial samples. *PLoS One*, 3.
- Zerbino, D.R., Birney, E., 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* 18, 821-829.

### Internet resources

- Illumina sequencing technology:** [http://www.illumina.com/technology/sequencing\\_technology.ilmn](http://www.illumina.com/technology/sequencing_technology.ilmn)  
[accessed 17.05.2010]
- Velvet:** <http://www.ebi.ac.uk/~zerbino/velvet> [accessed 10.07.2010]
- MAQ:** <http://MAQ.sourceforge.net/> [accessed 10.07.2010]
- Hawkeye:** <http://sourceforge.net/apps/mediawiki/amos/index.php?title=Hawkeye> [accessed 17.05.2010]
- R:** <http://www.r-project.org/> [accessed 10.05.2010]

### Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.virol.2010.01.023](https://doi.org/10.1016/j.virol.2010.01.023).

<b>Supplementary data 1</b> .....	<b>67</b>
<b>Supplementary data 2</b> .....	<b>68</b>
<b>Supplementary data 3</b> .....	<b>120</b>
<b>Supplementary data 4</b> .....	<b>122</b>



---

## Chapter 4: The first complete nucleotide sequence of a Grapevine virus E variant

---

**Beatrix COETZEE, Hans J MAREE, Dirk STEPHAN, Michael-John FREEBOROUGH and Johan T BURGER (2010) Archives of Virology 155:1357-1360**

Grapevine virus E (GVE) was first isolated from a Japanese table grape cultivar (*Vitis labrusca* cv. Aki Queen and Pione) and classified as a new member of the genus *Vitivirus* in 2008 (Nakaune et al., 2008). Although one of the plants used as viral source, had stem pitting disease, no relationship between GVE and any disease symptoms could be established. Grapevine virus E is a positive sense single-stranded RNA virus with a genome organization resembling that of Grapevine virus A (GVA). Currently, the only sequence data available are the incomplete sequences for GVE isolates TvAQ7 (AB432910) and TvP15 (AB432911). Both isolates are transmissible by the mealybug *Pseudococcus comstocki* (Nakaune et al., 2008).

Recently, a metagenomic sequencing study reported the presence of GVE in South African vineyards for the first time (Coetzee et al., 2010). The virus was detected in a severely diseased vineyard (cv. Merlot) in the Stellenbosch region of South Africa. Sequence data from the Illumina Genome Analyzer II were subjected to *de novo* assembly and two scaffolds were found to preferentially align to GVE when doing BLAST searches against the NCBI's non-redundant nucleotide databases. The GVE variant present in the metagenomic sample was homologous to the incomplete GVE TvP15 sequence (Nakaune et al., 2008). A scaffold assembled from metagenomic sequencing data (Coetzee et al., 2010), GVE Node 3404, was edited to 4216 nt to contain only high quality sequence data and submitted to GenBank (accession number GU903011). This scaffold aligned to TvAQ7 at nt positions 373 to 4589.

Based on the available sequence information (GVE Node 3404, TvP15 and TvAQ7 sequence data) the following set of diagnostic primers was designed: GVE-1-For 5'-AATGGAGTCAAAAGCGATCC-3' and GVE-Rev 5'-GTAGGGTCAATCAACCAACA-3'. To identify a GVE-infected plant, total RNA was isolated from individual plants using a CTAB method adapted for grapevine (White et al., 2008). The extracted RNA was used as template in RT-PCR using the diagnostic primers. A single grapevine plant, SA94 (*Vitis vinifera* cv. Shiraz), from a different

vineyard and cultivar than the metagenomic study, was identified. The vine displayed typical Shiraz disease symptoms (*e.g.* canes with a lack of lignification, delayed leaf fall and reduced vigor) and tested positive for GVE. Sequencing of the PCR product confirmed the GVE infection status. The vine also tested positive for Grapevine rupestris stem pitting-associated virus (GRSPaV), Grapevine leafroll-associated virus 3 (GLRaV-3) and GVA infections. RNA extracted from this vine, was used to determine the complete nucleotide sequence of the South African GVE isolate SA94.

A set of nine primer pairs was designed to amplify overlapping regions spanning the GVE genome. Primer pairs were used to synthesize cDNA from total RNA using AMV reverse transcriptase. The cDNA was then used as template in a PCR with a high fidelity DNA polymerase. Annealing temperatures and extension times were adapted according to the primer pairs and the expected sizes of the amplified fragments. PCR amplicons of the expected sizes were excised from agarose gels, purified and directly sequenced using the primers used for amplification. Sequence data from the 9 overlapping amplicons generated a contig of 7042 nt using BioEdit (Hall, 1999).

The 3'-terminal nucleotide sequence of the virus genome was determined by cDNA synthesis from total RNA with the oligo(dT) primer 5'-TACGATGGCTGCAG(T)<sub>17</sub>-3' (Meng et al., 2005). Similar to other members of the genus *Vitivirus*, GVE harbours a 3'-polyadenylated genome, allowing the oligo d(T) primer to anneal and prime cDNA synthesis. The cDNA was used as template in a PCR using the oligo d(T) primer and a GVE specific primer designed to anneal ~200 nt upstream of the expected 3'-terminal end. The amplicon was excised from the gel, purified and cloned into a standard cloning vector. Purified plasmid DNA from five clones was sequenced and completed the 3'-terminal end of the genome sequence by adding an additional 70 nt to the previously generated contig.

RNA Ligase Mediated Rapid Amplification of cDNA Ends (RLM-RACE) was used to determine an additional 456 nt towards the 5'-terminal end and thereby completing the GVE SA94 genomic sequence. Two GVE specific nested reverse primers were designed and used with total RNA as template in the FirstChoice<sup>®</sup> RLM-RACE kit (Ambion, USA) according to the manufacturer's protocol. Amplicons were purified, cloned and sequenced as described above.

The complete sequence of GVE SA94 is 7568 nt in length, excluding the poly(A)-tail, and extends the 5'-terminal end of GVE TvAQ7 by 8 nt. This is the first report of a complete nucleotide sequence of a GVE isolate. The complete sequence was deposited in the GenBank database (accession GU903012). Open reading frames (ORFs) were identified using the NCBI ORF finder and protein domains were identified with BLASTP and searching the NCBI Conserved Domain Database (CDD) (Marchler-Bauer, 2008). Nucleotide sequences for GVE TvAQ7 and TvP15 isolates, GVA, Grapevine virus B (GVB) and respective protein sequences were extracted from NCBI and compared to GVE SA94 to determine sequence similarity. Multiple sequence alignments were performed and nucleotide and amino acid sequence identities calculated using Vector NTI 10 software package (Invitrogen Corp.). Positions and sizes of the ORF's and untranslated regions in the GVE SA94 genome and percentage nucleotide and amino acid sequence identities to other members of the genus *Vitivirus* are shown in Table 4.1.

**Table 4.1** Genome position and size of open reading frames (ORFs) and untranslated regions (UTRs) of GVE SA94 and percentage nucleotide (amino acid in brackets) sequence identity to other members of the genus *Vitivirus*.

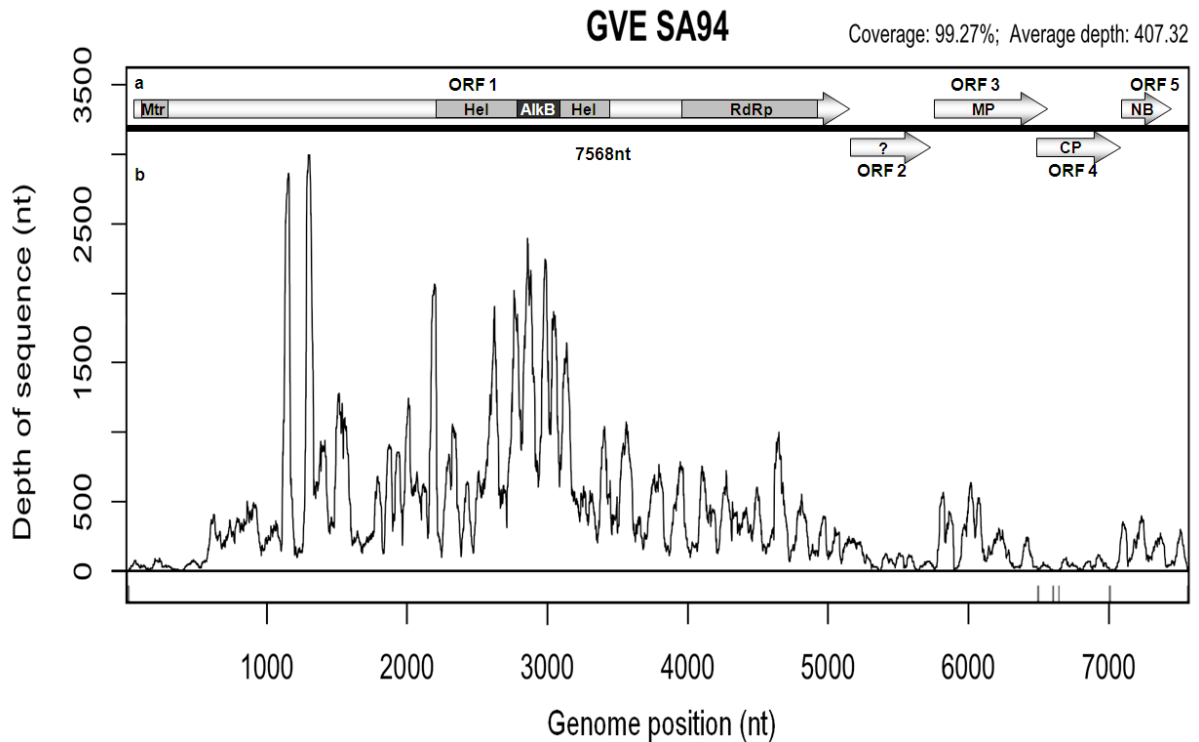
UTR or ORF	Genome position	Size (nt)	GVE TvAQ7	GVE TvP15	GVA	GVB
5'-UTR	1-65	65	73.8 (-)	- (-)	27.6 (-)	43.1 (-)
1 replicase	66-5165	5100	69.4 (72.9)	- (-)	50.5 (29.6)	51.2 (31.7)
2 hypothetical protein	5172-5747	576	52.9 (44.6)	96.5 (92.1)	44.5 (13.5)	43.4 (13.1)
3 movement protein	5768-6574	807	72.4 (76.1)	98.3 (100)	47.5 (26.4)	45.3 (18.3)
4 coat protein	6495-7094	600	78.4 (86.9)	99 (99.5)	50.8 (42.6)	50.7 (35.5)
5 nucleic acid binding	7112-7462	351	75.9 (75.9)	98.6 (96.6)	41.9 (16.1)	60.5 (64.0)
3'-UTR	7463-7568	106	58.4 (-)	99.1 (-)	33.0 (-)	38.7 (-)

GenBank accession numbers for nucleotide sequences for GVE TvAQ7 (AB432910), TvP15 (AB432911), Grapevine virus A (GVA; X75433) and Grapevine virus B (GVB; X75448) and protein sequences for these viruses (AQ7: BAG68224 - BAG682248, TvP15: BAG68229 - BAG68233, GVA: CAA53182 - CAA531826 and GVB: CAA53196 - CAA53200 ).

The genome organization and protein domains predicted *in silico* for GVE SA94 were identical to those previously described for the TvP15 and TvAQ7 isolates (Nakaune et al., 2008). The genome consists of five open reading frames (Figure 4.1a) and is polyadenylated at the 3'-end. The 5'-terminus most likely contains a 5'-methylated cap structure due to the presence of the methyl-transferase domain in the first ORF. This was confirmed by the ability of the RLM-RACE to determine the 5' end of the genome. Grapevine virus E SA94 ORF1 (nt 66-5165) encodes a replication-associated polypeptide of 1699aa, sharing 72.9% aa identity to the corresponding polypeptide of TvAQ7 (no sequence data covering this region is available for TvP15). The

replicase polypeptide contains conserved domains identified for methyl-transferases, helicases and RNA-dependant RNA polymerases (Figure 4.1a). An AlkB domain, belonging to the 2OG-Fe(II) oxygenase superfamily, was identified at aa 949-1035, located within the helicase domain (Figure 4.1a). The AlkB domain and its position was found to be the same for the TvAQ7 sequence. This domain, serving a role in repair of methylation damage, is implicated in counteracting the host defence mechanisms (Aravind and Koonin, 2001). Flexiviruses are thought to be a major source of AlkB domains for other RNA viruses which acquire the domain in mixed infections via recombination (Van den Born et al., 2008). Grapevine virus E is the first virus reported to have an AlkB domain located within the helicase domain. In other viruses, including all flexiviruses, where this domain has been identified, the AlkB domain is located upstream of a helicase domain (Martelli et al., 2007). The impact of this altered arrangement on the function of the AlkB domain is unknown. The GVE SA94 ORF1 does not overlap ORF2, as it is the case for the isolate TvAQ7. As was reported earlier for GVE TvAQ7 (Nakaune et al., 2008), GVE SA94 ORF2 encodes a protein with unknown function and ORF3, ORF4 and ORF5 encodes the movement protein, coat protein and a protein with nucleic acid binding activity, respectively as determined by sequence similarity to other vitiviruses.

Grapevine virus E SA94 shares a higher nucleotide sequence identity with isolate TvP15 (98.1%) than with isolate TvAQ7 (69.6 %). Grapevine virus E SA94 is therefore an isolate of the same putative strain as TvP15 with aa identities ranging from 92.1 % for the hypothetical protein (ORF2) to 100% for the movement protein (ORF3). The only other vitiviruses with the complete genome sequence available are GVA and GVB, displaying 49.6% and 51.2% nucleotide sequence identity to GVE SA94, respectively.



**Figure 4.1 a)** Schematic diagram of the genome organization of Grapevine virus E (SA94). **Mtr** methyltransferase, **Hel** helicase, **AlkB** AlkB conserved domain, **RdRp** RNA-dependant RNA polymerase, **MP** movement protein, **CP** coat protein, **NB** nucleic acid-binding protein, **?** protein with unknown function. **b)** MAQ-reassembly of metagenomic sequence reads on Grapevine virus E (SA94). Schematic representation of virus genome with numbered open reading frames is shown above graph. The four grey bars below graph highlight areas with no coverage.

Grapevine virus E Node 3404 from the metagenomic sequence data, aligned to GVE SA94 at nt position 382-4597, covering 56% of the genome. These GVE sequences share 98.2% nucleotide identity. To verify the similarity of the South African GVE variants, GVE SA94 was used with the Mapping and Assembly with Quality (MAQ) assembler (Li and Durbin, 2008). MAQ is a reference assembler that will map the pool of metagenomic sequence reads to a reference sequence, calculating the number of reads mapping, average depth of sequence and the average coverage of the reference sequence. Only reads aligning with a high confidence to the reference sequence are taken into account, thereby giving an indication of the sequence similarity between the metagenomic data and the reference sequence. Although, no scaffolds were obtained aligning to the remaining GVE SA94 genome with high confidence, when re-assembling the metagenomic sequence reads on GVE SA94, 99.3% coverage was obtained with an average sequence read depth of 407.32 nt (Figure 4.1b). This confirms the high degree of homology between GVE Node 3404 and isolate GVE SA94. The homology between these GVE sequences is in contrast to the variation observed for South African GVA variants in the same

metagenomic sequence data pool (Coetzee et al., 2010), suggesting a more homogeneous population structure for GVE.

We determined the first complete nucleotide sequence of a GVE isolate. The complete GVE SA94 sequence, 7568 nt in length, shares 98.1% and 98.2% sequence identity with TvP15 and GVE Node 3404, respectively. This is indicative of a high sequence similarity between South African GVE variants. The genome organization of GVE SA94 resembles that of the TvAQ7 and TvP15 isolates, consisting of five ORFs, capped at the 5'-end and 3'-polyadenylated. Grapevine virus E is the only plant virus described so far where the AlkB domain is situated within the helicase domain.

### Acknowledgements

Anthony la Grange's assistance with graphical presentation of data is appreciated.

### References

- Aravind, L., Koonin, E.V., 2001. The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate-and iron-dependent dioxygenases. *Genome Biol.* 2, 0007.1-0007.8.
- Coetzee, B., Freeborough, M.-J., Maree, H.J., Celton, J.-M., Rees, D.J.G., Burger, J.T., 2010. Deep sequencing analysis of viruses infecting grapevines: Virome of a vineyard. *Virology* 400, 157-163.
- Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids Symp. Ser.* 41, 95-98.
- Li, H., Ruan, J., Durbin, R., 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* 18, 1851-1858.
- Nakaune, R., Toda, S., Mochizuki, M., Nakano, M., 2008. Identification and characterization of a new vitivirus from grapevine. *Arch. Virol.* 153, 1827-1832.
- Marchler-Bauer, A., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C., Fong, J.H., et al., 2008. CDD: specific functional annotation with the Conserved Domain Database. *Nucleic Acids Res.* 37, 205-210.
- Martelli, G.P., Adams, M.J., Kreuze, J.F., Dolja, V.V., 2007. Family *Flexiviridae*: a case study in virion and genome plasticity. *Annu. Rev. Phytopathol.* 45, 4.1-4.28.
- Meng, B., Li, C., Goszczynski, D.E., Gonsalves, D., 2005. Genome sequences and structures of two biologically distinct strains of Grapevine leafroll-associated virus 2 and sequence analysis. *Virus Genes* 31, 31-41.
- Van den Born, E., Omelchenko, M.V., Bekkelund, A., Leihne, V., Koonin, E.V., Dolja, V.V., et al., 2008. Viral AlkB proteins repair RNA damage by oxidative demethylation. *Nucleic Acids Res.* 36, 5451-5461.
- White, E.J., Venter, M., Hiten, N.F., Burger, J.T., 2008. Modified Cetyltrimethylammonium bromide method improves robustness and versatility: the benchmark for plant RNA extraction. *Biotechnol. J.* 3, 1424-1428.

---

## Chapter 5: Conclusions

---

In this study we demonstrated the feasibility of a metagenomic sequencing approach using the Illumina next-generation sequencing technology to provide the viral profile and discover novel viruses in a diseased vineyard. Using the Illumina sequencing-by-synthesis technique, we obtained high quality and quantity sequence data, allowing us to identify known and previously unknown viral pathogens in grapevine. This is the first report of a metagenomic study using short read sequencing data from the Illumina sequencing platform for viral profiling of diseased grapevine.

Bioinformatic tools, Velvet and MAQ were successfully implemented to allow the analysis of the 837 megabases of metagenomic sequence data. Four viral pathogens known to infect grapevine and the dominant variants of these viruses could be identified. Grapevine leafroll-associated virus 3 was the dominant virus present in the vineyard, representative of 58.5% of analyzed read data, with GP18 as the most prevalent variant. A second minor variant and a potential novel grapevine leafroll-associated ampelovirus were also detected. Grapevine rupestris stem pitting-associated virus was represented by a single scaffold covering 99% of the genome and shared 92.3% nucleotide identity with the SG1 variant. The dominant Grapevine virus A variant was identified as a member of molecular group III.

The fourth virus detected in the metagenomic study, Grapevine virus E, was most similar to the incomplete sequence of the Japanese TvP15 isolate. Grapevine virus E was not previously detected in South African vineyards. Primers designed using the *in silico* assembled scaffold as template, could detect GVE in grapevine plants from a different vineyard than the metagenomic study vineyard, proving that the *in silico* assembled scaffolds represent true viral templates present in the sample. The metagenomic study provided the basis for the further genetic characterization of a South African GVE isolate, GVE SA94. The GVE SA94 nucleotide sequence is 7,568 nt in length and represents the first complete nucleotide sequence of a GVE isolate. This GVE isolate shares 98.2% identity with the GVE scaffold from the metagenomic study. Genomic organization resembles that of the previously characterized Japanese GVE isolates and other vitiviruses, consisting of five ORFs, a 5'-cap and 3'-polyadenylated genome.

To our knowledge, GVE is the first plant virus reported to have an AlkB domain located within the helicase domain.

Furthermore, analysis of the metagenomic sequence data revealed the presence of viruses from three fungal virus families. With 4.5% of sequence read data, the dsRNA *Chrysoviridae* family represented the second most abundant viral group. The other two families present are the *Totiviridae* and the unclassified dsRNA virus, *Fusarium graminearum* dsRNA mycovirus 4. These viral families have not been previously reported to be isolated from grapevine.

This study serves as a pilot study for future research using metagenomic sequencing for the viral profiling of vineyards. This project also opened interesting avenues to further pursue in the genetic characterization of grapevine pathogens. Follow-on research projects resulting from this study are already in progress:

- Development of an efficient diagnostic test and investigation into the incidence and presence of strain variants of GVE in South African vineyards, possible association of symptoms and the presence of GVE, molecular characterization of GVE (MSc project).
- Further genetic characterization and phylogenetic analysis of the novel *grapevine leafroll-associated virus 3* variant (Post Doctoral project).
- Polymerase chain reaction based detection of dsRNA fungal virus families, *Chrysoviridae*, *Totiviridae* and *Fusarium graminearum* in grapevine and detection of possible fungal hosts (Two BScHons projects).
- Metagenomic sequencing using Illumina sequencing technology of viruses from grapevine plants displaying Shiraz disease symptoms (Post Doctoral project).
- Investigation into the accurateness of Illumina short read sequence data using the near-complete Grapevine rupestris stem pitting-associated virus scaffolds as assessment measurement (Peer-reviewed article).

Research questions that warrants further investigation:

- The sequence dataset generated in this study can be used in the evaluation of the suitability of available bioinformatic software tools for metagenomic studies and how parameters should be adjusted for optimal assembly.
- Fungal, bacterial and host (*Vitis vinifera*) sequence data, scaffolds which remained unassigned and sequence read data not assembled into scaffolds.



We conclude that metagenomic sequence analysis using short read sequences from the Illumina sequencing platform can be used for virus profiling, detection and identification from grapevine samples. The identification of all the viruses present in the leafroll-diseased vineyard will assist in the elucidation of the viral disease etiology. Moreover, metagenomic sequencing and the bioinformatic sequence analysis approach described here, is not limited to viral diseases and can be applied to investigate the etiology of diseases caused by other grapevine pathogens or pathogens of other crop plants, animals, humans or environmental samples.

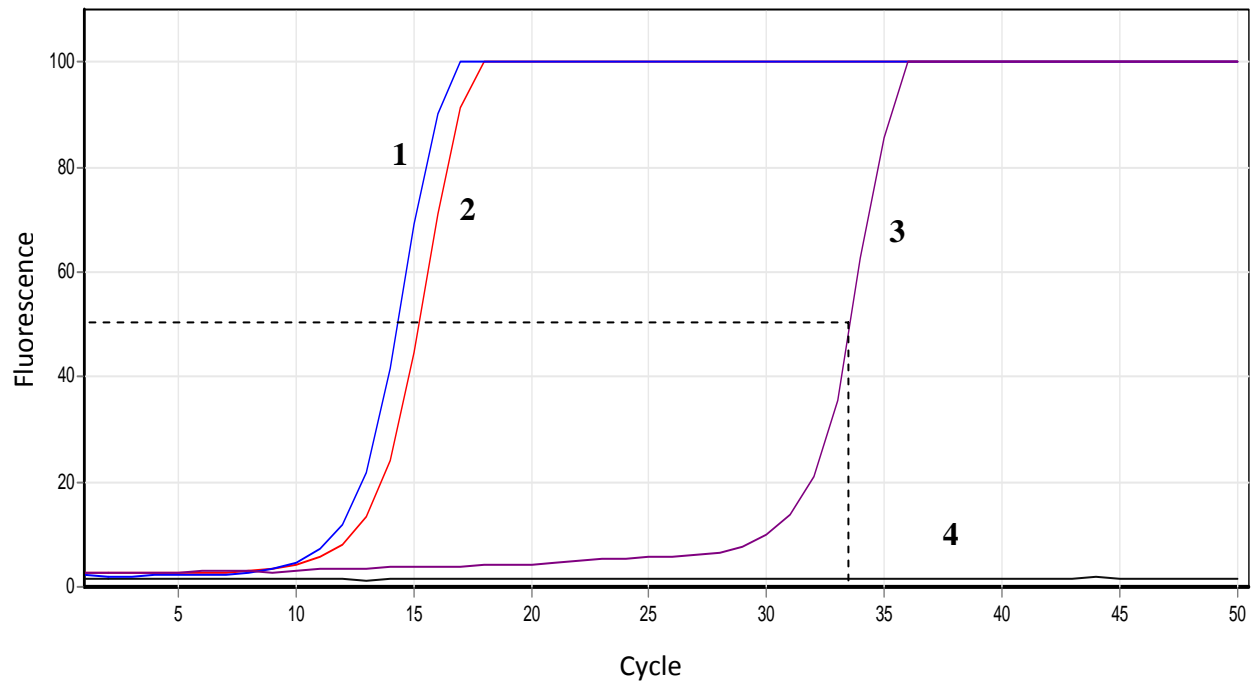
---

---

## Supplementary data 1

---

---



Quantitative-PCR to determine the optimal number of PCR cycles for amplification. Curve 3 indicates that 33 cycles are the midpoint of the log phase for the sample. Curves 1 and 2 are positive controls and curve 4 the negative control.

---

## Supplementary data 2

---

Supplementary data 2: Scaffolds assembled with Velvet *de novo* assembly(Parameters: hash length 23, coverage cut-off 50, expected coverage 1,000, minimum length 100)

```
>NODE_26_length_175_cov_10474.411133AGGATGTTGATCTTGAGGCTTTGTTGCGGAGCGCTGTAGAA
ATGTTGTGTTGGTTAAGCCGTCGTCAAATCTTCAAACGGGTTACAGTTAACGTTACTTATTTACGCATGTTTGTAG
CGGTGCTAATTGTCAGCTTTTGTAGAAGGCGATGAGGCACCTTAGAAAAACCCATCAGAGTAGCGGTACACTAATTG
CG
>NODE_38_length_703_cov_4877.066895AGTATGTGGACGTAACGCAGAATTTAGAAAAGTACGATCGTGC
CGTTAAGAGTTATGGAAATAAAGAAGAGGCGAGTATCGGCGGCACGTTGCTTTACCGAAGGTTGGTATCTGCATACG
TAGATCTTTTATACGAACTTGCAGGAGTTGCTGTCAGATGAAGTAACTAAGGCCAGAACCGATACCAGTTTCTGCCTA
CGCTACTGACTCTATGGCTTTCTTAGTTAAGATGTTGCCCTGACCAGCCCGTGAGCAGTGGTTGAGAGACATGTTA
GGATATCTGCTAGTACGAAGACGACCCGGCAAATTTTTCTACGACATAAGAGTAGCTTGGGTATATGACGTGATCGC
TACGCTCAAGCTGGTTATAAGGTCGTTTTTCAACAAGGACACACCCGGGGTATTAAAAGACTTAAAACCATATGTGC
CTATAGAGTTCCTTCGACCCCTTTACGAGCTTTTCATCCTATTTCTCTAGGTTGAGTTATGAGATGACGACAGGTAA
AGGGGAAAGATATGCCCGGAGATCGCCGAGAAGTTGGTGCGCCGTCTAATGGAGGAAAAATATAGGTTAAAAGATTA
ACCCAGTGGCGGCCTTAATAATTATACTGATTTATTACTCTATTTACGGCACAAACGCTACCAGGATTAACGACG
CCCAGATTTCTCAATGTGAGGATAAAAGGGAAAAGTCAAAAAAGTTTCGTTACGGGGGTAGAAGA
>NODE_61_length_124_cov_5647.378906CCAGTGATGGCTCTCAAACCCGCTACACCCCAATGACGGTT
TGAAGTGGTAAAAACCGGACCTCTGAAGGATGCCGTCGTGGAGTACTTGAAGTACCTGGCTATAGGCATTGAGAGG
ACAATACAGGGCGTTGCTTATGGCTAG
>NODE_70_length_130_cov_9078.715820TCAAGGATCAACATCCTAGCGATCCACAGTTTGAAATATAC
CCATTACACTGACTAGATAATTCGTTTCGTAGTATGTCTGCTTCCCTCAACTGCGGTCAGTCCGGGTCTGAGTTAGCA
TTATGCGTTTGCCAGAGCATTGATCACTTAAGT
>NODE_71_length_90_cov_12555.222656AGTAGCGTAATGGTGGTGTGAGTAAAATACCTCATAAAGGAT
CGCACAGTATTGGTTTGGAGACGCGTCGTTTTTCTCCTAATATAGCTTTCCCTTACTTCCGAATAACT
>NODE_72_length_89_cov_4795.775391CTTCCCTTACTTCCGAAATAACTAAGTAGACCTTCTTTTTGAC
CTAATTCGGACTCCACCCTATACCCATCTTTGACGCTTATATTTTTACAGTAGAGTAGAGTACGAAAA
>NODE_75_length_2151_cov_1319.034424CAATACCCTGTTCTTACACGCATACGTACGACGTGCAAGGT
GGTGACGTTGTTTATGGTATATTTGAAGGAGAGTACAATAGGGCCTTTTTGAATGAACCGACTTTCGCGGGCTGTG
GACGCACAGAGGAGATCCTAAAGAAAAGGACGTAGCTCAGGTTTAAATTTGTCGACTGATGGTACTGTGCTTTAATA
GTGAATGGGGTGGAAATTAAGAACGAGTACCTGGTGCCGGGTACCACCGACGTCTAAACTCGCTAACATACAAATC
AGGAAGAGACGGATTGGAAGCTAAAGCAATACCTGAGTATTTGACCACGCTGAATGTTTTACACACAAAAGATTTCA
CAAGGAAGAATTTGGACGTAAGGATAAAGAATTTTCTGATTTGCGCATTGAAGATAATTTTATTAATCGAAGATT
AATCCGCAATCGCTTCTAGATGGATAAGTATTGTTATATAACCGGATTTCTAGATCCTAACGCTGCTAGGTGCGAAG
TATTTAAAGTTGTAAACCAAGGAATATATAGGACCGAGTGGTCTGATTTTTCAAAGGGCACAAGTAAAGTACACCGTA
TTATGGGACGGTAGCCCGGAGAGTGGTGGTTTCGTTCTGCGCCGCGAGGCTACGGTGGATGCTTTTCGCGTACT
TCTTGTTAAAGGGTGGTTTAAACGACAACCCTCTCGACCCAGGTAATTTGTGAGAGGTACATTAGGACCTCGGAGAT
TTAACTAGCCTATTTAAGAGGTTAGAACGGGGAAAGATTTACGCGTCGCGATCTATAGATAGCGGTCTTCCATATGA
GGATAGGAAGAAAATCGAGGAAGCATCGCGTAATTTATCGCCGTCGGACGCCGCTTTCTGCAAGGCCGCTCCGTTT
AGGTTGGAAGGTTTGTAGACCCGACGAAAACAGTTAGAGGTTGCCGAAAATACCGTTACGAAATAATGAAGCTAAACAAA
AGAGTCCGGTCAAATCAGGCGTCTTTACCTAAGGTGGCTTCTGCGTTTTGTCGACTTTTTATAACCGATTTACAGGAAC
TTTGTCTGATGAGGTATTAAGGCCAGAACCGACACTGTGTCCGCTTATGCGACGGAATCCATGGAGTTCATGGTTA
GGATGTTGCCCTTAACCTCTAAAGAACAATGGTTGAGGGATCTACTGGATACTCTTGGTTCGGAGGCGTCTCTGTT
AACTTTTCTACGATTCAGAGTTTCTTGGGTCTACGACGTCGTGGCCACGATTAAGGTGCTTTGAAACTTTTCTT
TAATAAGGACATGTTAGGTGGCATCAAAGAGTTAAAACCTATGTACCAATCGAGTCATTTGATCCATCCATGAGC
TATCGTCTTATTTTTCGAGGTTAAGCTACCAACTAAGCGCCGGGAAAGGCGCTAAGATATGCCCTGAGCTCGCAGAG
AAATTAGTTCCCGTCTTATGGAAGATAATCACAGGTTGAGATTAACCTCCAGTGGTCTCGTTGATTATAATATTAGT
TTATTATTCCATTTACGGGACAAACGCCACTAGGATAATGAGACGCCCGTGCCTTGAAGGTCATAAAGGGGTA
AGCTCGAGGAAGTTTTCGCTGTTAGGAGTCAAGATCGAGCTCACAGATTGTCGGAGAAGCGCGGAATAAACGCTCAG
CGTATCGTATGTAGGTAATATAGCGATTTGGCGTACTTAGCCGACGTCGTTATGACATTCACAGAAAACCTGGAA
GACGTTAAACTATGTCGACGGTAGGTTAGCGTATGACACCCGCGACTGCATAACGTCCTAGGGTGAAGAACATGGTGG
ATCTAGCTGATCACGCTAGGATTATACAGTACATAAAGACGAACGACAATCAAGTCACTGGGACTACGATGCCACAT
CAGCAATAATGCCGATTAGCGATGGTACGATGGTTCTCGTATTTAATTTATAAAGAAGTTTTAATTTATTTTGTAG
TAAACGCATTGCTTTGAGTTACACTAATGGCTTTTCGATTTAAAGTTGGGGCAGATATATGAGGTAATTTCCCGAGAAC
```

AATTTGAGAGTAAGGATAGGCGATGCCGCAACAAGGGAAGTTTCAGCAAACCAGAACTCTTAAAGTATGTTAAGGACGG  
TACTCAAGCAGAATTGACGACCATAACCGGTGGTGCCGAAAAAGTACGTGTTTCG  
>NODE\_82\_length\_92\_cov\_7448.358887TTTTCCGGTTAACATATAAAAACCCCTCTACACCGTCTCGCGA  
TAAGCTTCCCCCTGCTTTACCAATTTAAATAAGTAGAGTTACCTACAGGTATGAAAATTTGGGTTTCAAC  
>NODE\_84\_length\_91\_cov\_6711.000000AAAATTTGGGTTTCAACGTAAACACTACCAGCTACAGGAGTAC  
CACACCACTGCCCCGATGGGAAAAGCAGATTGTGCTGGAAGTGGTTCCAAAATCTATACCTACTTCCA  
>NODE\_91\_length\_171\_cov\_1929.116943TAGCATTGAAATTTAGCTAGACACAGGCCGTTTAGAAATAAG  
ACGGTGACGGCCGATAACACCTTACACAACGTCTTCCAACCTATTGCAGAAGAAATAGTTACGATCGATGTCCATAAA  
TCGGTGAAAAATTTAGAAATATTTACCTTTTATTGATAATTCATGGGGGCTTATACACACATAAACTTTCATGA  
>NODE\_98\_length\_214\_cov\_10218.125977TTAAGATATCTGCGGCAAATTTGGCAACATCACTCCTGACCT  
TAACTAGAGCACTTGACCCCCCAGTCATAACAGCAATAACCGGGTCTGGATAAAAAGTTGTCAAGACCGGTTTTGAAC  
ACTTCCACAGCCCTAGCGCTGAATGGCGCAACAATTGCGTCTTAGTTTTGGTTCGCTAGTCCAAAATCCACACCCCTCCAC  
ACCCCTTCTACGGGGATAAGGTGTTGCGTTATAGTGTTAT  
>NODE\_109\_length\_900\_cov\_888.768860CCGCTCTTCCGATCTCGCACGTGCCGACGTCTGGATGTCAGC  
CGACCGCTCTGCAAACCAATAATGGCGCGGGCCGTAGCATCAGGACTACGCAACGGCACGTTCTGCATCACAGGGA  
GCACCTCCACATAGCAGTGAATGAATCCTCTTCCCAGTCGTAGCGAGAAAACACGCCTGCTTAGCCCCACAGCGGCC  
TGCTTATCACAAGCTGCTGCGAGCAGAGCTTTTGTGGTTGCGTTCGAGCTATTGACATGTATGCGAACACCGCTCAA  
CGACTTTGCGGCTTACATTCCTCCTCCGGAGTGGCTGCAGTGGCAACCCAGCGTGTAGACGGACTGGGTCTAACA  
ACACTGACACGTGACGCGCCAGCCACCACGTCCAAATGCACTAACCACATCAACACCAGCAGTTTTTCTCCACATGCTT  
GTGCATAACACTGACGTGATAATGATAGTGCTCATTTCCACACCAGGAGACACATTGTGACTACGCTTGACGCGCT  
AACCAAGCAGCGACCGATGCGCGTTGTCCCAGGTATAAGGGCTCACAGCGCTCGAAAACACCAGACACTAACACAGGA  
CACCGATCACGCTTAGTTCGGTGCAGGATAAGCTTGGTAGTGATTGCGCAAATTCGGGGACGTTAGTTGCGCACCCGA  
ATGCTCAATCGCGTACTGAAAGCTCACCCAGCGATATCATCCGGCGTGAGACTCTTGCAAAAAGTTCCCAATCCCC  
GGTGATAACATAAGCTCTGCCACAGACTTCCGGCTCGACGAGGACCCAGAAAAATCAAAGGCCCTTGCCTCGGC  
TCCACTTTGCTGTGCATCACAGACCTCAACGCGGCAGGCTGTCCGAATCCCGGAGCGACCGACAACATTGACTCAA  
ATAGTGCATCAGCCATGGCGAAGTTGATAAATT  
>NODE\_112\_length\_234\_cov\_2845.593994GTCCGGCTGTGGCCGTCACCAAGGCGACCGTAGCTAAACGTT  
CAAGTCCCCTTAGCGCTTGTATATCGTTTTTCTACCTCATAACGCCATTTCCGGCTGCAGTATGCTAGGCATTTGGGCAC  
ACGCTCTTCCGCGGCACCTAATGTTTTTCTTCCGGCTTAGGGACATTGCTTGGGGCGAGGGCTAGCGCGAACACTTGG  
AAGTTTGGGGGTTCTCCAATAATTGGTTCGCCGTTCCCGAGGTTGTTTTGGCGAGGGAGGAG  
>NODE\_113\_length\_1984\_cov\_653.768127TGTCGAGAGGGAAACAGCCCGGCTCACAGCTAAGCCCCCT  
AAATGACCGCTCAGTGATAAAGGAGGTAGGGGTGCAGAGACAGCAGGAGGTTTGCCTAGAAGCAGCCACCCTTGAA  
AGATGCGTAATAGCTCACTGGTTCGAGCGCTCTTCCGCGAAGATGAACCGGGGCTAAGCGATCTGCCAAGCTGTGG  
GATGTA AAAAGTAAAAATGCATCGGTAGGGGAGCGTTCCGCTTAGAGGGAAGCACCCCGCGCAGCGGCTGTGGA  
CGAAGCGGAAGCGAGAATGTCCGCTTGAGTAACGCAAACATTGGTGAGAATCCAATGCCCCGAAAACTAAGGGTTC  
CTCCGCAAGGTTTCGTCCACGGAGGGTGTAGTACGGCCTAAGATCAGGCCGAAAAGGCGTAGTTCGATGGGCAACAGGTG  
AATATTCCTGTACTACCCCTTGTGGTCCCAGGACGGAGGAGGGCTAGGTTAGCCGAAAAGATGGTTATCGGTTCAA  
GGACGCAAGGTGACCCTGCTTTTTTCCAGGGTAAGAAGGGGTAGAGAAAATGCCTCGAGCCAATGTTTCGAGTACCAGGC  
GCTACGGCGCTGAAGTAACCCATGCCATACTCCCAGGAAAAGCTCGAACGACCTTCAACAAAAGGGTACCTGTACCC  
GAAACCGACACAGGTGGGTAGGTAGAGAATACCTAGGGGCGCGAGACAACCTCTCTAAGGAACTCGGCAAAAATAGC  
CCCGTAACTTCCGGGAGAAGGGGTGCCTCCTCACAAGGGGGTTCGAGTGACCAGGCCCGGGCGACTGTTTACCAAAA  
ACACAGGTCTCCGCAAAGTTCGTAAGACCATGTATGGGGGCTGACGCCTGCCAGTGCCGGAAGGTCAAGGAAGTTGG  
TGACCTGATGACAGGGGAGCCGGCGACCGAAGCCCCGTTGAACGGCGGGCCGTAACAATAACGGTCTTAAGGTAGCGA  
AATTCCTTGTCCGGTAAGTTCCGACCCGCACGAAAGGCGTAACGATCTGGGCACTGTCTCGGAGAGAGGCTCGGTGA  
AATAGACATGTCTGTGAAGATGCGGACTACCTGCACCTGGACAGAAAAGACCTATGAAGCTTCACTGTTCCCTGGGA  
TTGGCTTTGGGCCTTCTTCTGCGCAGCTTAGGTGGAGGGCGAAGAAGGCTCCTTCCGGGGGGCCCGAGCCATCAGT  
GAGATAACACTCTGGAAGAGCTAGAATTTCTAACCTTGTGTGAGGACCTACGGGCAAGGGACAGTCTCAGGTAGACA  
GTTTCTATGGGGCGTAGGCCTCCCAAAAGGTAACGGAGGCGTCAAAGGTTTCTTCCGGGCCGACGGAGATTGGCC  
TCGAGTGCAAAGGCAGGAAGGGAGCTTACTGCAAGACCCACCCGCTCGAGCAGGGACGAAAAGTCCGCTTAGTGATC  
CGACGGTCCGAGTGGAAGGGCCGCTCGCTCAACGGATAAAAAGTTACTCTAGGGATAACAGGCTGATCTTCCCAAGA  
GCTCACATCGACGGGAAGGTTTGGCACCTCGATGTCGGCTCTTCCGACCTGGGGCTGTAGTATGTTCCAAGGGTTG  
GGCTGTTCCGCCATTAAGCGGTACGTGAGTTGGGTTTAGAACGTCGTGAGACATTTCCGGTCCATATCCGGTGTGGG  
CGTTAGAGCATTGAGAGGACCTTTCCCTAGTACGAGAGGACCCGGGAAGGACGCACCTCTGGTGTACCAGTTATCGTG  
CCCACGGTAAACGCTGGGTAGCCAAGTGCAGGAGCGGATAACTGCTGAAAAGCATCTAAGTAGTAAGCCCACCCCAAGA  
TGAGTGCTCTCCTACTCCGACTTCCCTCAGAGCCTCCGGTAGCACAGCCGAGACAGCGAGGGTTCTCTGCCCTTCCG  
GGGATGGAGCGACAGAAGTTTTGAGAATTCAAGAGAAGGTACGGCGAGACGAGCCGTTTATCATTACGATAGGTGT  
CAAGTGAAGTGCAGTGATGTATGCAGCTGAGGCATCCTA  
>NODE\_115\_length\_532\_cov\_1011.278198GACTGCGTGTCTCTCAGTGAATGTGGTTTCGCCAACCCCTT  
TCACTGTCTGGCATCATCCATCACATGGCTCCTGACCCTACCTCGACAACCTCTCCTTGATTCGACGCCGCTCTCTC  
AATAGCCACCTGGCTCACAGCGGCCTCTCCCTTTCTATGGCCATTCAACCTCTGCCGGAGTCTGTGCCTTCGG  
ACGTTGCCAGGCTCGTAGCATCCATGCCCCACGGTGAAGGCTCCAATCTCGCCTCCCTCGTCAATGCAATGGCGACC  
GCCAATGGGGAGTGGGATCGCCTCACACGAAACGATACTTCTCCCGTGTGCTCTCCATGCTGCGGGCGCTGCCCG  
AGTTCGATCACAATCCGCTGTTGAGTTGCTGCGCTCGATGCCATCGCGTCATGTGGTTTTGATGGTGTGAATGGTG

CGGTTTCAGCCTTCGCTGACGTTTCGCCTTGCTTACAAAAACGTTACGAAAAGTTGTTGCGCACCAGTTCGCACGATCT  
AGACATGCCGTCTTTTTCTCGGATTGAACTCCCGGGCGGTGATCTCCGAAAGT  
>NODE\_116\_length\_90\_cov\_2982.088867CGACTGGGTACCAACCCTCCGAAAGTGTCTACGAAACCCTCTC  
GTTCTCACCTTACCGACTCCCCCTTGTACTCCCCTGGATGCCAACACCACGTCGCTGGTGCGGTTG  
>NODE\_117\_length\_575\_cov\_1100.130493ACCCAGTCGCTGGTGCGTTGTGCGCTCTCGTGCGAAACAC  
CCTTAGCGTCGGAATAATTTCCATCCAGCATAACGTTTGTGCTTGCCAATCCAGTCTCTGGTGTTTATGAGAGCCATG  
TCATTTCTGGCGTTACGACTGGGTACCACCCCTTGTGCGGTTCCATCCGGCAGTGCCAAGGCCCTTCCGAGTCAA  
ACCACTCGCCGATGGTGGCATCATCGCCGCTCTTAGGGCTTACGACCAGCCCCGGCCGAAGCCTGGATCTCACCTTC  
GTGAGTTCTGGCGGTGCTCCGATCGCCTTGCTGCCGCCACGAAAAGGCCCTCGAAGATGGAGAACCCTCCATGGTA  
ATATCTGAGTACCAGATACCCAAACCGTTTTCTTTTTGGTGGGGGGACTGTTACCAGACGAAACGCCCTTTCGTGTTCC  
CGTCATGGGTGCGATCGGTGATGCTGCGCGCTACGTCGGTGGTGACTACGCAACTCGTATGCATGATGCCGCTGAAC  
TGGCCGACTATACTACGCATGACCCTGAGACGCTCCATCCCCGCTTTCTCAATTACGTAATGAGCGCCTCACCGTT  
CTCGACCCAAACACCGC  
>NODE\_118\_length\_1182\_cov\_192.102371TTTGCCGATTTTCGCTCAAAGGGTTGAAGGGAGATAGTGCA  
TCAAGCTGTTTCGCAAGGGCCAACCTTGATCCTCTTCCCCAGGGATCCGAGATGAGGGAACCCTAGGAGAGCCGCCGAC  
TCCAACACTCCGTCATGTACGATCCATACTAGATCTGACCAACTGCCCATCCTCTACGTTCTTGACAGCC  
CATCTTTGTCTCAGTAGAGTCTTTTCAGTGGCATGTTTCGGTCTCTTCCCCATTACTTAGAAAAAGTGAGCCACCGG  
TTCAGGTACAAGATACTATCATTACCGCCTGGACAATTAGACATCCAACCCGTAATCGCAACGACCCAATTGCAAGA  
GCGGAGCTCTACCAACTGAGCTATATCCCCGAGCCAAGTGGAGCATGCATGAAGGAGTTAGATGCTTCTTCTATTTC  
TTTTCCCTGGCACAGCTGGGCCATCCTGGACTTGAACCAGAGACCTCGCCCGTGAAGTAAATCATCGCACCTACGGT  
CCAACCAATTGGGAGAGAATCAATAGATTCTTTTTCGGGAGCGATTTCATCCTTCCGAAACGCAGCATACAACTCTCC  
GTTGTACTGCGCTCTCCAAGTGTGCTTGTTCCTTCTTACCATGGCAAGTCTTTGTGAAATAACTCCGAT  
GAGAAGAAAAAGAAGGCGTTAAGAGACCTCCTGGCCCAACCCTAGACACTCTAAGATCCTTTTTCAAACCTGCTC  
CCATTTTCGAGTCAAGAGATAGATAAATAGACACATCCCATTGCACTGATCGGGGGCGTTCGTAGTGACTGAGGGGGT  
CGAAGACCAAGAAGTGAGTTATTTATACCAAGCATTCTTCTTACGGCTAGATCCAATCTCCTGGTCCCTGTGAAAAG  
GAAAAAGAATTTACGTTCTTCTTTTCGGGAAGGGGAGGATTAGGAAAAATCCTATTGATTGCAGCTTTCTCCAGACCT  
CCGGGAAAAGCATGAAAAAAGGCTCGAATGGTACGATCCCTCCGTCACCCCAGAATGAAAGGGGTGATCTCG  
TAGTTCTTGGTCTGTGAAGATGCGTTGTTAGGTGCTCCATTTTATTTTCCATTGAGGCCGAACCTAAACCTGTGCT  
CGAGAGATAGCTGTCCATACTGATAAGGGATGTATGGATTCTCGAGAAGAGAGGAGCCGTGGTGGTCCCCCCCCGG  
ACCGCCCG  
>NODE\_125\_length\_281\_cov\_2963.658447TGCTTCTGCCTATTACGCTAGGGCTTTCTTTGGTCATAAG  
GGTTTTGCTTAAACGATAACCATACTCAACTTGCTTATGTTCCACCAGTTGAGGGGAAGGAATGTCTAATGACGAGAC  
GCTTAGGTATTACCGGACTTCGACTATGCAAGGTTGCGGGTCCATCTGGGACTCGGCATGATACCGGTTCTTGGGA  
GATGACAGTGTGATCCTCCTCTGGTATCTCAAGCTATGATGGTGGTCACAAACGTGCGTGATGTGGGGATTAGCA  
CCAACGGGGAAGTTGTTGGTGAAGAAGAGAC  
>NODE\_139\_length\_153\_cov\_4930.457520GTACCAGATATTCACTTTTACTTCTCACCCTTAAACAATA  
ACAGATACCGTACCGTCTGTGGATAAAATTAATTTGCGCTACGTCGGTCTCTTTTTGGGTCTCCCCCTACGTTTCGATA  
GCCCCGGAATGTCGGTTCATTTAGAAAAGCCCTGTTATTCTCCCCTTCCAATTATAC  
>NODE\_144\_length\_152\_cov\_1633.401367AAGCCTGATGATCAAAGGGAAGATCACTGGCTTTGGCGGGA  
AGCCCATGATGCGTGTGTTCACTTGTGCTGGATTACACCAACCTTTAATGAGACGCATGCCGTCAAGCACTTGCAA  
CTAATGATGACGTCAATGCGCAATGTGTACGCTCAGCATAATGCGCTATCCACTGA  
>NODE\_145\_length\_584\_cov\_2668.339111GAATACTCTTCGCCCTATCCTTTATGGCTTGAGCCATGATG  
GAGACGACTAACGCCACATGTCATCGGGAGAACTTCGACTTTATTATGTACAACGCGACTTCCCGTGATAACATG  
CGTTTTCTGCGACCTTACCGCTCCACAGCGTACTCAAAAAGTTCTGTTTTTCGAAAAGTACCAATTTGTGTTGAGAGCAT  
ACGAATAAACCTTATCGACAAAATCCGAATCTAGGACGATCGATTTAAAGCTGACGCTAGTCCCCTCAACAATTTGGT  
AAAAAAACCACCGACTTACCTACGAACGTTGGGACCAAAGACCTCTGGGTTGTTTTCCGATCCTTGCGCCTTCTGGGC  
CCTGCACATTGTGAAATGGTGATAACCGCATCTATAACCTTCGTACTCCAGTTTATATGCCGCACCAGATGAGACGC  
GGATGTATGCGTTTCGTAAGAAACCACTCACCTTTGGTATAGGAATGTTTCGTAGACTTCCCACACTGGCCCACGTT  
ATACGTGACAGCTTACCTTTCCCTCTTCACTTAAAGGTATCTAGCTCCTCAAACAGACCTCGCCATTTCCATACAG  
AAGTTCAATTGGGAACATAAGAGCTA  
>NODE\_146\_length\_269\_cov\_2808.308594CTGGTTTAAATGGGATTTCTCCAGGAAGGTGCGGCGGGTTA  
ACTTCTTCGACATTGGAGGAAGTTTTACTTACCATGTCAAAGCCGGGCACGTTAATTGTACGCTCTGTAATCCGGTT  
TTAGACGTTAAGGATGTGAAGCGGAGGATTAATGAACTCCTCTTTCTTTCCACTGCAGGCGGCGAATCGTTTGTATC  
TAGCGACCTCCTTACTGAAGCTGCTTCGAAATCCGTGTCTTATTGTTCTATGGAATCGCAGAGATGTAAGCATAGAG  
CCGACGCTGGGTTTTATGGT  
>NODE\_151\_length\_202\_cov\_4781.361328AGGGAGTTAAAGGAAAGACGAGTGAGGGTTTCGGGGGACAA  
GTGCTGAGGTTAAAGACGAAACAGTTTCTGATCCTTCTAGCTCTACATCACAAGTACGAGAATCCAAAACTCCGC  
AGATGTCGTTGAAACCGTACGCGTAGAAGAGAATGAGGAGGAAAAACAAGGTAAGGCTGAAGCTTATTGGACAAGAG  
GTCGCACAGCAGGAGGTTCTAATCGGGA  
>NODE\_152\_length\_951\_cov\_710.782349ATGGAGCATGATGGAAATCTTATTGCGCTAAGTACGCGGTTG  
TTCTCTGGAATGCGGAATACTACTGACTCTAACAGCCGTGCCAATCCGGCTTATTTGGATGTCATCGCTCAGCAGGT  
AGAGGTATTCACCGCGTGCCTCCATTGCTGTCAAAAACAAACCGGGGTGATGATGTGGCGGCTGAGGCGAAAAGCT  
TGTATGAGGCTGCATTGCTTATGCGTGTGCGCGGGGTTATGGGTTTTGGTTTTAAGGCTATAAAAACAACTCATTAGC

AGCAAGTACAGAGAATTTTTTCGGCTGTTTCGTGACGAAACTGGTGTGTATGGGAGTCTTTGCCGTATGCTAGCTTC  
CGCCTTATCGGGGCAATGGTTCGAACAGTGTGCTTCCCAAATTCATAGAGCCAGCGGCTAAGCTGCAGTGGTGG  
AAGTTGCTCGTAAGGCTGGGCGTGTGCGCGCAGCTTAAGCTTTATGGATAAAATTTGTTGCTGTTGCGTTTGCCAAG  
TGGGCTACTGATGGGGAACACCAAATGGCCGAGGCGGTGATCCATGGTACGCGGAAAAACAGGAGTCTGGGTATACC  
TACAGCAGGCGGAGACATCTTGGAGCTTGATTGGTGTGCTGAACCTGAGTCGCGGCTAGAGGTCGTGGGTGCTCCA  
AACAATGCGAGCATGAAGTTAGCGCAGGACACTGTCAATGATGCGGTGATCTGCTTGGCAGTGAGCGGTGGTGCC  
AGTGAAACTCTTGCAAATAGGTTGGCTGAAGGAGCTTTTATGGGGCAGTGGCTAGGAATCGAGGGCCAGGTGTCT  
TGCGAGTTGAAATAGGCGCACGGCTCGCAGAATTATTCGTAGAAAGAGCATACTGAGGACGCTTTTCCTGGCATG  
ACAAGTAAGGACTTTTTATGCCTCTTTGGTCCGTTTCGCGGACGCAACTGCAAGCTCTGAACCGTGCCGGCGCACGTTA  
TGCGAGC

>NODE\_168\_length\_341\_cov\_1062.296143ATGAGCGTGTAGATTTTTGGACCTGCCTTTGAAGGGGAACTA  
GTCCGAAAGTTACCCACTAGCCATTTTCGTTGCTGTGAACGGCTTTCTGGAAGATCTGTTAGATGGGTGCCAGCTTT  
CGATTACGACTTCTTCGAGGATGATTTTTGAAACGTCTGATCAATCTTTCTCGTGAAGACGCGCGCATTTTCGGACA  
ATTTTTCTCATTTTACTTCTAGAATAGAAGATAAGTATTACAGTTTTATTAGGTCTAGTGTAGGTTTACCTAAGCGC  
AACACCTTGAAGTGAACCTAGTCACTTTTGAACAGGAATTTCAATGCCGATCGCGGTGTAACTTGGATGTGA  
TGATGTGGCGGCCG

>NODE\_169\_length\_246\_cov\_4475.503906AGCGCCACTAGGTTCTCAAATGCGTTTACCTTTGTCGATAG  
CTTGAAAGGGAGGAGTGTCTTTTTCTCAAAGCTGGGTGAAGGGTATACCTATAATGGTGGTAGTCACGTTTTCAT  
CAGGGTGGGCCCCGTGTCTTAGAAGATATCTTGACGGCAATTAAGTACCCGAGCGTCTTTGACCATTGCTTAGTGCA  
GAAGTACAAGATGGGTGGGGGTGTACCATTCCACGGGATGATGAGGAATGTTATCCATCGGACAACCCCATC  
>NODE\_170\_length\_80\_cov\_13524.337891GAGCTCAGTGCGGCATAACCCAGTGAACGTGTATGTCTTGA  
CAGACCAACAAGTATATGCGGTTGTTTTAGTGAACAATGAGTCATCCGCTTCTTTGTCTT

>NODE\_171\_length\_251\_cov\_4155.115723TGTTTGTCTCAATGACCGTAAGGGGAACGGAAAAAGTGTTA  
CCCTGAAAAGGTCGTTAGTGGTAAGGATAAGCCGGTAGTGAGATCACTGTCCCAAAGGGCCAATTGGAACACTGAT  
GACGTGGCTGAAATAAACGCTGACGTGTATTTGTGCATGACCCAGTTTGGAGAAGTCGGATATGAAGAGTCTGTTGA  
AGGGAAAAGGGAAAGAAACAACAAGTGTGACAGTACATGAAGCACAGGGAAAAACATTCGGTGATGTTGGTGTGT  
T

>NODE\_172\_length\_535\_cov\_2093.725342CATCTAGAGCCTCCACCCGAATGCTCACCAGGGGCGTTTCC  
CTTGCGACATTTGTTGATGTGTCCAGCCGGCGCTCACATCTCTCGCCACCACATCAATTCGACCCGAATTGATGC  
GCGTGTGATGTTGACTATGTTTGTGCTGATGCCTGGGATTTCTCTCCACCTGTGAGCCTGACCAGTATGATGTGGT  
TGTGCGGACCAAGTGTGTTGGGCTTATACTCCGCGGGTCTGGCCCCCGTGACGAGGCCGCTTTTCTCGAGCTCGTT  
GTCGATGTCCTAGCCCCGTGATGGTTCTTTCTTCACTGATCATCATGTGCCTTCGCCATCTGTGATTCGTTAAA  
ACGCGTAGCCGATCTTGCCGGCCAGCACTTTGACCTCGCAACCAATTCCTGGGCGCTATTCCGATGACCTCTGCTACT  
CATGGACTGCTATGGTAAATGAGGCGGTCTCTTTGATTTGGCTCTTCTCGTACGCATAGGTTCGGAATCTATGGTCCAG  
TGTTGGGAGATGTTTCACTTCAGAACGCGTGCCTTCCCCAGCCACACGTCAG

>NODE\_173\_length\_156\_cov\_7377.679688TCTTCCGCCCCGATCGGCGCGCTGGGCGCGCCTTGGGTCCA  
AAAGGAGGGGATGATCCCCGCTCCGATTCACGGAATAAGTAAAAATAACGTTAAAAAGTAGTGGTATTTCACTTTCCG  
CGTTTCCGGCTCCCACTTATCTACACCTCTCAAGTCATTTCAAAAAGTCGGACTAGAGT

>NODE\_174\_length\_118\_cov\_3126.677979ACCGCTCTTCCGATCTCTGCAAAGGATTCTACCCGCGCTC  
GGTCCGAATGATACTTCAAGGCGGCCCGTGCGGCTCGTCCGCGCGGGCTTGACCAAGGACAGTGCCTCTGGGG  
GCCGGGTGGCCCCCTACTAGGGT

>NODE\_192\_length\_8602\_cov\_707.504211TTTGAGTGAAAACCTATAAAAAACACACCAAAGCTTCTTTAAC  
TTTCCACACCTTATATATTAATAGGATTTAGTATGGAAAGGGAATACTATTAGTACGGTATTCAGCGAACAGGCT  
TAACCCAGCCTTGAAATCGGGGAATTATTGAACTGCGGAATTGCCATTAGCTCCATCTCCTCATTCATGAGTAACA  
TTTGAGAAGGGGTTGTTGATGAGTGCACCAGCCGTTCCACCACTAATCTCTCCAAGGTTGGAACCTCCGCTTGCCCTC  
AGCCATAGCTTGTCTGAGCACTTGAACCTCAAAAAGAGGCGATATTAGCGACCATCTCAGCTTTTGTGGAGCTCTTT  
TAATTCCACCCTTTGGTTCAAGCGCACTCTCATCTGTGACTCCAGGAAGAAGTCAAAGGCTGCAAAATTTCTCATTC  
TCATTAACCCCTTTCTTGGCCCAATTAGCTGGTGGTATTCCCGTCTCCAGATGGATATTCACACGATTTTTGCATA  
GTACATGCAAAATCTGCGCAATGTGGTTACCTCCCTTATAATCCCGCAATTTCTGATCTATTAGATCACAGAATG  
GACTTTTGCCAATCAAGGTTGACCTTTGGGAGCTTCCAACATCAGAGCAGTGACGTGCTATATCTATGGCCACATCA  
AATTGTGCGCTCTCAGTATGCCTGATTCTCCAATGCAGCAATCAGCTCATCTATAGCACCTCATCCATGACATT  
TGGAGAAATTACAACCGTTTTCTGTTTTCAACACCTCCCTCATGTCCACTGTTTTTGGAGTGCATTTTCAATAACCC  
TCTTACGTTTTAGCAAGCACACCACTGAGAGTAGGCGGTGTGTTTTGAGCCCTCTGGCTGCTTTTTGAGCTCTAGCCAAC  
TCCAGTGATTTAGCCGGGTTTTCAAATGCCTCATTTGACTCTCCAGGGAGTTTGCCAATCTGACTTGCCATCCAGCC  
AATTTAACAATGATAATCAATTTCTTCAAGTAGTATGGCTAATCCCTAAAAGCTTCAAACCTCCCAACACCTCCTTCA  
CGGTTAATGCAAGGGCCTCACTTGGGCGCAATTTCTCAATCAAAATCGATTCTCCAGTGGCCCTGATAAAGCAACCT  
TGTTACTACTACTACTGCTGTCTACATGAGCAAGGATTTGGATGGCAATTGCAATTCCTACAAAACAGAAATCC  
AAACAGACAATACATGTATGAATGCAATTAGACCTATGACAAGCATTAAAGGGCACGAAATTTCTCAGACTGACCAAAT  
ATTCCTGAACTGGGGAAGGACTGCTTAGGACCACAGTAAGTTATACTTTTTGTGCCGTCCCGTAAACGCCCCCGTG  
AGGAAATTGATGGACGTTGTCTCCTGAATATGGTAGATTTGACTTTCTAAGAAAATGCAGCACAAAGTCCAATCCCA  
AGCCGACAGTCAATGGAGTTATGGTTTTTGCCTAATTAGCAGGTTGCTGAAAAGGCATCTTACTGCATAATGACCAA  
TTTTCTTCTATGCCTAGTGGCAGCGATGTAACACAGTGTCTGTCTCTGGGGTTGGATCCCTTGATTTCAAGAGGG  
TTACCACATTGAAAGTTGAACCTCGAACTTCAACCGCAACTTGAAATCAGAGTTGTGCGCAACAGCGAGATCTACC

GCCTCCTTGCCAAAACATATCAGCACCCCTCAACATCTAGAGTGAAAGGGTGACCCACGATAACTTCGTCCAACCC  
GTCGCTGCTAACTTTGAAACCGAAAGAGGAAAGGTATTTGCAAGTGTTTGCACCAAAACCTGCAGGTCTTGGACTTAA  
TGAAATGCGCCTCTTTGGGAATACAGGTATTCTGGTACGGATCTGAAAAAAGTACATCAAAAACCGAAATGTCTTGG  
ACGGACAAATACTCATCAAGTATATTCACCTTTCCCTGCACTGCGCATCTGGAGAGTAAGCCTTGATGTACCTTCC  
TGTGAGATTAGGATTGTCTGGAACACCTGCCGTGAAAGCCACGAAACTCTCGTCTCTGAGATTAACTCCCTTATTA  
AACTGCTCTTACCCTGCCAGGCACACTATGTATAACAACCTGGAAATTTGAGCACACTGAAAACACCTATAAACTCA  
AAAGCTGATAATGCTTTAACTAAATTATTTCATAATACTCAAACAACGCCAAATACTTAAGGCCTAATCAACTGCTTC  
AAATAGGCCTGAAATTTGAGCACCTCATTATTTATGCTTGTTCCTCACTAGGAACCTCACACAATTTGTAGTGTGCCTCCA  
CCTCTTCTCTGTGCATCATTTCATGATAGCTCACCCAATCTGTATGCATAGGACACCTCTATCGCATAAATGTCA  
ATACAGTTCTCCAAATTGCCATTTCTCTGGCTATGCATATTTCTCTCCAGATTAGTTGAGGTTTCTTGAAGATTCC  
CTCCTTAAATAAGCACCATCCACAGAATGTGGGGTTAGAAAACAACTGAACTTTGGCCTTAAGGCAGATCATGTTTA  
GAAAGCCCTCATGCTCAGTTTTAAGCCGCAATCTTCTATTAGCACACATGTGCATCCCTGCAAAATGCTAATTTGACTC  
CGAACCCGTCAGTTCATATCTCAGAAAAGTGAACAACATGTTAGCCATAGTGTGAAACAGAAATGTGCTGGCTTCCC  
CAGTGAACCTCATTATAGAAAATGAACCAAGCTTTGACCCAAAGTGAATTTTGATGAACTCATAATCCAAAATTAGA  
TCTTCAGGCAATCCCAAAAACCTTATTATCTGCAACTCGAATGCTAGGATAAAAGTGGTCTTGGGAGGCATCAAAGGC  
TTCTGACTCTGACTCTGTGCAAAATGCCAGTGAACCCCTGCTGCTAACCCTCACTCATCTAGACTTTTAC  
CTGAATGGATGTAGAAGTTAGGCTTTCAGAGCTTGCATCAGCTTCTTCTCGATGTATCTCATATAAGGTGCAAAACGG  
ACCAAAACAGCGTGTGAAAGCAAGCGAGGGTCTGCCAGCTTTAGCCACCCTAAGCCTATTGTCAAAATTTAGTGCA  
ATGCTGCGACTTAATAAAAATTAGAGCAACATCCAACAGCCAATCCCTGCATGACCTGCCTGAATGGTTCTCTATAG  
TCGCTGCACTCTTACTAGCTTTCTTTGCTTCAAACCTCCCGCACTGCCTCATTGAACATATTGGAATTTGACTGGCA  
TTTGTAGTGGCACCTTAGACAGGAAAGAATCTAGTAAGAACCTCCCATATGGCTTTGCTCGCCTTAGTTTGGCTGCTTC  
CACCTGGGGCTTTGAAAATTTGAGACGTTTTTCGAACAGCCATGAGGAAGGGTTATGGAATCATTTCCTTTATGCCTT  
GGGTAATGCTTTCAAATCTTTACATGCATTGCTCAGTTGCTCACCAGCACCACTCTGTGCTCATCAAAGAACTG  
CTGAGTTACCTCATAGCCACAGCGCACCTTCCCGATACTCTTTTAAAGGCCATCTTACCACCACCTGGCTCTGGTT  
GACTCCAAATTTGCTTAGGGGAATATGGGTTTTGAACCATTCTTCATTTGAGCATTCCATTTCCATCTCTTCAACTTC  
ACAATCCTCATCTTGACCCAGGAAAAGCATAAATTTAGCCATGGATCACCACACAACCTTTCTTCTTAACCTCCCT  
CATTTTTCCCAATGCTGACATTGAAACCCCTTAAAGAAGTTAGGCTCGCCAGGAAAGTAGGTCTCGTATAACATTTCTCT  
CCACTTTTAGCCGTCAAAAAGTGGTAAAGAGGTTTACCAGCAAAATGCGAGAGAAAACTTTCCACCCTCAAACCTGT  
GATGTTTACCAAATCCAAATTATGGCTGAAGCGAGATAAAGCAGTGAGCCACCTTCGATCGTCAGTATGAAAGGAGT  
CATGTGATATAAGAATTCCTCCATTTTTGAAGTTCAACCCAGTTGATTACCAAATGTAAGAGTGAGCTGTTTCATC  
CCAAAGTAGGACTGGACTATTTTCTTCTCCTCAAAACTGGAACTAGAACAGCTTCCAAGGGACCTTTTGACCTCGC  
CAGGAGGAAGTATGCATGTTGTGGAAGATTGCATAATCAGCAAAGGCAGCAGTCAATTGAGCCCTTTTGAAGTGAC  
ATGGTAGTCTAAATTCAAACATTTGGGTTACAAAACGCTGCTTCTATGTTGTACCTATACGCTCTTTCTTCAAGT  
ATTTCTGAGTATGTCGCTCTGCTCAGCTCCCAAAATAGCCCGATCCTTTTGTGAATCATATGTCGTCTGGCATGGATC  
ACCAGCAAGTGAATGAAAGCGTCATTGCGTATGACAAGCAGACATAGATCAATGTATCCAGGGGAAACCAACTGAA  
CCTCATCGAAAATTACACAATGGCCAGCCTTTCAGACCAGAAAATTTTCTAAGGAAAACTCAAAGTCACTACATCC  
CACCTTCTCTTTGATTTGCCTGCTTTGGCAACTTTGCAGCCATCATCCATTTCAAGATCATTCTTGATTGAATTT  
TGCCAAGGCTCGTCTAGGGGAAACAAAAGTTACAAATTTTCTTTCAGACCTTTCATGAAACTTTTATAACAAGTAAC  
TCTTTCCAGCCCCAAATGTGCCAAGCAGCAGCTTAGCTCTCTTGGCTTAATATCCGCAGTTGATATAACCAGAAAGC  
CAGGGTTTTGCATCACTGAACTTTTTCATCACTTATGACACCAGTTAGGCCCCCTAGCAAGCACCTTGCAGCTTTTC  
AGCCCGCTCAAAGTTCCGGAAATAGGTTATCTTGGACCCATTCAATTGAAAGAAGATCAAGGGCGGATCGCTTAACCT  
TCCTCAATTCATCTACTTTTTATGCTCTGAGGGCCATTGCCAATCTTCTCAACATGCCTTATGTGCTCATTGGTTATA  
TCAAACAAGGCGGAAATCTTACCATTTGGGTTAATAACCTCAATTTCCCCATTTATGTACACTTCGCGCAAATGTC  
AAAAGCCTCAAAGCTTCAGTCAAGTACATCATGTTAAGCCCTCCTCCCTGGCACAGCTCCTCAAATAATGTGGTTAG  
TGCTAGGCCTACATAGATAATTTAACACATCAACTTCACGCCTCCCAAGAGCTGAGGCGATGGCTGTGATCATGCAA  
CCATTCATGAGGGCGCATGGCTGGAAGTGATTGGACTTCAGGCAAGAATTCACATAGAGGCCACTCACTTTCCCC  
AAATTTCAAACCTGACTTCATATTCTGGTGTACCCTATCAGCCTTACGTGGTGCCGGACACAAAATAATGCAATCA  
TTTCACTCTCAGCATAGGCATTCTCTTCCAATTGTTTAGAAATAGTTGGAGCTGAAAGATCTGAACTAACCAGGCGC  
TCACTTTCCACAAAAGATCGAATGCCGGCCCTTCAGAGCAAGACCATCAGTACTGAGTAAGAAACCAACTGAATGCCA  
AAAGCAATTGCCATCACCAGGTACGGAGAAAGTTAAAATCATCAATATCATAATCAGCACATGCCTCAATCTGGC  
AACCTAGCAATACCTCCTTGGATCCGGAGAATGAGCACTGTATTTGTGGACCCTTAGTTTACCAGTGGATCTAACT  
GGGGCTTTGTATGTTATTACAAAATGGAGCCCTCCTCCATTTTTTTAATGTGGAAGTCCCTTTGAAAACCTTCAAAA  
AGACCCAGAACTTTTGGACCAAAGTCTGTTTTCTATGGATAGAGCACCTCAAGAACAAGGATTGACTGTTTTTTTGC  
TGCGGGGCATCTCTAAAACCTTAGATAGACCAGGGCCTGTTGTGCGAAAACACCACCCTATCACATTGACCTAGCCCT  
TGTGCTCCTAAAGCCAGTTTTCAAATTTATTGATAAGATCCTTGAACCTTATGAAAATTTGTGACAAAGAGAAAACAAGAGCA  
GAGGTCTGAGGTAACACTTGTGCCATTATAGAAAATTGATTTCTTGGATTCCATGAAGACCAGTTGCCACATGCTTC  
GGTCAAGCTTTTTCGTGATTGAGTTCAATAGGTACCTGTGGATTTACAAATATTAGATCATCTTCTTCCAGTTTTTGG  
CGCCTGAAGGTTAGGGACACCCGCCCTTGTGTGCAATTGGAAACCCCATGGAAGTAATTACATTGCATTTCCCTTAGG  
CATTTCAAAAACCTCATTAGGCTCAATTGTGAGGCAACAATGATTACCCTCAACCAGGTTTCGAGACTTACTCACAG  
TGAAAACCTGCGGAGCCAATTAATAAACAGTTAAAACCTTGTGCCAGTTTCGTAGCAGCTCTCATTTGCTTTATGA  
AGGGCAACTTTGAGTTTTTTTTCATAGACCTGAGCCAGGCACGAATTTGTAGTAGTTACGTCCACCAGTAGAGTTTAG  
AATTTGGTCCATGAACTTTGGCCACCCTAAGCTTGTGTGGCTTCCACCATTGTAGCTATAACATGAAATCCCTGGTT  
TTGAGAAGAATGATGCTCCCCTGCCTTTTAACTGTCTTTAAAGTCAATGTTTTTAAACCACCAATTTGAGATCCTCC

GGTAGGACTTTAACTGCAATCTTTAGCTCTTCCAAGTCAAAAACTTTTTTCAGCAGCTTTGATTGGTGACTCGACATC  
GCTAACAGAGAGGTTGTGAGCAGTTGACTGGACGCTAGAAAAGCTCACTGGAAACCCGGCTAACCTCACAATTGACTG  
ACTGTGACCTCTCAGAGGTTGAACTAGCTGAATAAACAGTCCGATTGACTTATCATTTGAGGTTGACTCGGAGGTG  
AAGGCATCATTCCGATCATAATCTCTTATGAAAACCTTTGGTCAAGCTTAGCAGAGCTTCAGGGAGTGTTCCTGTAA  
TACAACAGACTTTGTAGGCCTTGATAACCTCACAATAAGATCAGCAAAGGACGATGCTCTATCCTTCAGAAACTTGA  
GATTGGACCTTCGCTTAAGGAGGAACCATTTGGAGACCACAAGAATTGAAAACAATGGGTGAGAGCTTTTATGTACTTG  
AAAAGACCCACTTTGTAAAGAATTCTGGACCTCACGCATCATTATTGTCCACCAAAAACTGAATAGTTTCGACTG  
TCCAGTGGCTTTTAGCGCAGCCTTAACCTTTGATAGGGCAAGATTGACCTCCATGGGAGACGGTCTAAAAGTAGGAT  
CGTCAGACAAAGGGATTAGGGATAAAAACCGCTGCATGATTTGAAAATTAAGTGCACCTGAAAATTTATGCTCAACT  
GAGAATTTGGAATTAAGCCAATATGAGGACCCAAGGAAAAACGGGGCCCTCTGACGATTGAATTCAGATCCCCAAG  
TCCAAACCAAGAATTATCCATGACCACCTCAACGTTCTCTCTATCAACTCATTGAGGCAAGAGAAAAAGAAATAAGT  
GCTCAAAGAAATCAGGAGTGATGTCCACCAATTAACCGAGAAGTTAAATGGCTCAAGTGAGTAAACAAAATGAATCT  
AGTGAACACTCTCGAAAATTTGAAGAAAGCCTAGCAAGAGCATTGGCATCATTCGGCAGAAGAATGATATGACTAA  
CTTAGAAATGTTAGGCATTAGTAAAGAGCCACTCCTTCCACAACCTTATAACCAATGATGAAAAATCCTCAATGAATT  
TTATCTCCCTCCCTGTGGGCTTATCAATGAGTTGACGAAGCTTTGCCTCAGCAGACTCCTTGTCTGGCTTCTTAAAG  
GTCTTTAAGTAAAGGTATATCTTTGGTAACGACATCATGGTGCACCTGGAAAGCAACTGTGCATACCACGGCTTAGAGG  
TTCAAGATCCTTTATACCTATTGCATCAAAACCAGAAAAACACCTCATGTTTGAAGTCAAAAAGATTGCCCATGGGGG  
TGAAGCTAAGCAAGTGGTGGCAGAATTTTGGTGAATGATATCAACTGAATAGCGTGAGCCATCAGGCAGAACCACA  
GATCTGGCCTTTAGTAAGTATCTGCATGACAAAAGTTGTGAATACATCTCATTCCAGTTACCATCTGGAGCGAAGAG  
CAACTGATCGTTACTAATTTTGTACTGGTAGGCCAAGGGTTAAGGCTCTCAGGAGAACCAACCAGCACTTCTGGGG  
GAATCACTGCTGTGCGCAAAAGCATACTTGGCTTAGCCACATCCAAGAAATGATTAGATCACTAATAGACCAGTAA  
TGAATTTTATCATGCAAAAAAGATTCTTAGCACCTGTGGAATCACGCGCGGGACCAACCTTCGGAGTGTGCTAGA  
ATCACCATGCCTTTCTTGTGATGATTTGGCTTGACTTGTCACTAGAACTTATCACAATTCAGATCCATACCTGC  
TAACATCAGACTAGTTACAAATCTATTTATGAGTGCTACTAAGCTGAGATCCTTGTCTCTTTTTTTTAAAACTGC  
AACTTAAAATCCTTGATTCTACAAAATAAAATGAGTTATTAACATAACTAGGTAACAATGTACAGGATGTGATT  
TTCTAAAGTCTTGACAGACTGGGTGTGAGTGAGGTCTAAAGCTGTAAGGGCTTAGATAAATGCCAGAGGATATCAACC  
TTTTTTTACTTCTATCGGACAATGCGAAAGAGAACAATTTGTGCTGACTTTTCTCTGTATCCACCAATGCCTTGAGA  
GCTGTAGCAGAACTCTGGATTGTTTCATCAGAGGTG  
>NODE\_197\_length\_294\_cov\_2308.466064TTTATGAATCGCTACCAGGCTTTCACGTGTATAAGACCGCC  
AATGATCTTATCTATCACACTACGGTTGATGGTTTGC GCGTGAGGGATTTACCATATGTTTTTCATCGCTGAAAGGGG  
CATATTCGTGAAGGGCAAGGATGTGCGAGCTGCTCACTTCTCCATGTGATAATCTGTTTGTATGCGACGATATTATTG  
TTTTCCATGACGCCATCAATTTAATCGGCGCTTGAAGGTTTCTCGGTGCGGAATGGTTGGGGAGTCGTTTAAAGTC  
TTTTGAGTATAAATGTTATAACGCCCCCCAGGTGGTGGTAAAGA  
>NODE\_209\_length\_78\_cov\_6084.063965TGACGCTTGGTGTGATAGGGAGATTTCTGTCTACCTTCTTTT  
GTGGTGAGTACGATAGCGTGGTAAGGACAATGACCAAGAGCTAGTGTGTCCGTCGG  
>NODE\_211\_length\_107\_cov\_3491.224365TAAGGCTATCATCGCCACTAACGACGACGTAATCGTAATCT  
AGACCCCTAAGTACTACGCTTAAAAGAGTGACAAAACTATGCTGTTCCCAACCAAGTATTGGCTCCACCCTACG  
TCTCTGAGAGC  
>NODE\_231\_length\_496\_cov\_3082.310547TCGAAAAGGAAGATGAGAAGCAACGACGCTGTGTAATTTCT  
CCCCTCAGCCTCTCATTGTTAAACCAACACCACCTTTGGTAGTGCTCACGAACGCCCTTCTCGCATGCTCCACAA  
CTCTATACGTGGTAGGGCAACATCACACCCTGGGGCCGTATTACAATGAAATCTTCAATGTAGCCCTACCCCGCA  
CGATGGGGAAGAACTTAACAATGCTATCCTTGACATAAAAAATATCCTCTCCTACGAGAGGCTTCATCCAGATTTTG  
GCCAAACAAAGGTGTTCTTAACTTTGATAGGAACGTTAAGAGCCCTGTTGACAATTTCCCTTTCCGCAAGATGACG  
GCGTCGCTCTCATTGCCCCACCTTTACTGAAGAAAAGGCCAAACGTCAGAGTGGCCTTCTCCTCCAACGACTGTT  
GTGGAACAGATTGCGCTGGTTTTGGTGTCTTCTGCTTATTTCGGAAGTGGTACATGGACGCCAACAGCTGTTGGCTTA  
ACAGCCCATCGTGAC  
>NODE\_232\_length\_106\_cov\_4692.660156GAAGCTGTAACACATGCGTTGATTGTTGTACTACTACCACCG  
TGATGTCCGTTTACCCCGTCAAGCGCGAGCAGCCTAAACCGGCTGTCAAACAAGATGAGCAGAAGCCCAAACGG  
CAAGTGTAC  
>NODE\_234\_length\_106\_cov\_11011.660156TGAATACAAACCATGAAAGCGTGGCCTATCGATCCTTTAG  
ACCTTCGGAATTTGAAGCTAGAGGTGTCAGAAAAGTTACCACAGGGATAACTGGCTTGTGGCAGCCAAGCGTTTATA  
GCACGTTGGCT  
>NODE\_239\_length\_125\_cov\_8892.591797CGTTGCTGGTGGTGTTCAGCATGCTCTTGGAACGCCCGTTT  
CCATCAGTTCCCGATGCTGTTGGCCAGGCATGCCGCCCTCGCTGATTTACGTGATGATCTTGACGTCGCTTTTTCT  
GCGCCGAGATTCTCCTCCCTCCTGTCTGC  
>NODE\_242\_length\_273\_cov\_2249.256348TTAAACCGTGAGGTAGAGTTTGTATCTTACCTCGAAATA  
AAACCTATTCTACAAAAACGAAAACCGAACAACAAAAGACTAGCCTAATTTTCAATTTTCTTACTTTTCTTTATTTT  
AAATTTATTGAAGGATTTTTTACCCATTACTGGTTGTTAAACCTTCTCTTTTTCTACATATGCCAATTAGAGCATA  
AATAGTAACTAATAAACATAGCGATTAACGTTACGCCTATAAATACATCCACCCGTCACAAACGTCACTTCGCACG  
ACGCAATAGTGTACCGCTACTCT  
>NODE\_245\_length\_110\_cov\_10014.099609CTTACGTAACCTACTACC GCGACGCTGAGCTGCATAGAAC  
TATTCTTATTGATAGTATAGACTCCGCTTCGCTTACGACCAAGTGGGGTGTGTTGGTGGGCATAGCTAGAGGTTTGC  
TTTAAACATTCAGAA



>NODE\_248\_length\_146\_cov\_1695.520508ATCCCAACCATATTCACCAGCTAGGGAAATTATACTGCCCAA  
CTATGTGCCCTAGCTCAGCAATGTCTGATTCTTACCTCGTTGGCGCCGAAAAAGTCTAGTTGCCTTAGTGTACGTA  
CTGGTGAGCAATTTCGCGCTCAGCATTGCACCCTTTAGCATACTACTCTCAC  
>NODE\_249\_length\_626\_cov\_3854.255615CTTCTGTAGCAAGATTGACTTTTCGGCAAAATGTCTCAGAC  
GTTTCCAACGGAGCGGGGTTCCACCGCTGGTTAGCTTTGTCTTAAACCACCTGCGCTTCTTGTCTCATGATA  
AACAACTCGACGCTATGCTGCGATGCTTCTCAACGGTTGGAATGAGGATCAGCTTCGACGCGCTCCCAACACCA  
CTACCTTCAACTGGTAAGCCATCGTCAGAAAAATCCGAAGAAGCGTCTGCAGTACTGGTCTCGTTTTCCCTAAACC  
ACCTATCTCAGAAGGACCAGCGCCATCATCATAATCAAATCACGATAAAAGTTCAACGCTTGGTCATAAACGTTGC  
GACCCTCAATCGGCGGTACATACGTCAACTGTGGGACGGTTTCACTCATCAAGGCCCGAATGATTAGCGATATACC  
CAGAGCTACTTGGGGTAACATAGATTTATAACTTCTCCCTTCCCACATCACCTCGGGTACGGAAAAAAGCCGTTAG  
AGAAACCCCGAGCGTCCAAGAGTGAGTGCTTGCCTTAAACCCGAACAGCGTCCCAATCCAAAAGAAAAACATCAA  
TGCCTGGGGAGAGCATTGGCCCCAAATTCACAACGTACAGCCACTCACTACATTAGGACGTAGGAA  
>NODE\_258\_length\_1010\_cov\_910.492065ACGCAGCGGTGGGGAAACGCATTACTACGGGCGTCGGCCT  
CCTGCGTGGCTCGCGACCACCCTCCAACCAACCCTCACGTTTGGGAAGGATGAACCGAAAGCTGCAAAGCGCTTGGC  
TGATGGGACAATTGAGCCGCAATTCCCCGATTCATCTTCAACCTCTCTCAAATCAACTATGCTCTTGCGGTTTTTC  
TCCACGGCGATGTCTCAAATCTCTCCAGGACCTCGACCCCACTCATGGGCCAGGATTTGGGCCAGGATTCGACGCGGACGC  
TCGGGAAAATTCATTAACGTCGTTGAGCGCGCCTTTTCTGGTGGCCTCACCGTTCCGTTGACGAGAGAAATGGTCAT  
GTCTGATATTGAAAAGTGGGATGCGTTTTATTTCGCGAGGTTCTCTTAGAGCTTGGTGCTGACAATCTCGAGCTTGCCG  
TTCACAAAGGCAACCTCGACTCCCCCGCACTTGCAGCTCGTGCTGCGATGTATGGTGTTCCTCCGCCCGCAGCGCTTG  
AGAAAACCTCGTCGAACATCCATCAGGTTATTTTGTGATCTATATGGGACCATGCCAAGTGGTTCTTATTACACGTC  
GCTCATAAATACGAATTGCAACAATCTCCTTCTCCTCGGGCACATAATTGATCGTGTGTCATCTGAGACCCCGCTA  
CGCCTGCCGGTGCCGCTGAGGACGTCGCTCGGGTTGCGTACGGGCGCATGGTCTCTATGGCGACAATCAACTGTTT  
TCTGAAAAGATATTCAAGGAGTTTGGTCTTGCCTATGATCCCGCAAGCACGCTTCTTCTCGCCGGTTTGGAAAT  
GAAGCTGAAAATTGATGAGACCGAAGTGACCACCCACTTGGGACGTGTGAGATTTTGTTCGCGCTGTGCTCCGTA  
CTCCTGAGGGTCTTCTCATTACCCGCGCCACGCTGCACTTACCAGAAACCTTGCCGCCCGCCTGAGCATGATC  
CCGTAAGCGACAAGCTTTACGTGCGTGCAATAATGGCTGACTACATGGGGACTGACCCCATCGCATA  
>NODE\_259\_length\_91\_cov\_3670.395508TCAATTGGGAACATAAGAGCTAAATCGAAAACCAAACACCCC  
TTAAGATCCATCGCTTCAGCAACCTGTCTGGGAGACATATCGTAGACGTCCACCATAAAACCCAGCGTCGGC  
>NODE\_271\_length\_79\_cov\_4408.025391GACCTTATAAGCATTACGTTTTCAGTTTGTAGCTTGAGAAACAA  
CCAGGGTAGAGGTTCTGATCCACAGACACTATCTTTGTTATCTTGTAGATACGCGTCCA  
>NODE\_280\_length\_111\_cov\_7422.531738CTGTGGTAGCACTGTTGAGACCTTCCAAATTCGGATCTCTC  
TTCTTCACTGCCATATTTATGTCTCAGAACTTCCCACATTAGCCGTAATGGTGGGCGGTGCTATTGGGTGACTTGA  
CAAATTCGTCCACTA  
>NODE\_283\_length\_354\_cov\_8623.019531GAATGCTCTAACCGTCTCCACCTTTCATCGTTCTAAATGCGT  
GGGTTGATTATAGCCTTTAAGTATCCCGCATGACTTTTGTATCCCTAACCTCTCAACAAAAACCTCGCGATACTTTC  
TTTGGCATCTACGGATATCGCGTCTTCCACTCACAAACGATTGGTTGTAGTAATTTAATTTAACGCTTTCTTTTCGTA  
AGTATGAGGTCTACGTGGGCGTTAGTCTTAATTGACTTCGATAAATTTAAGATGAGAAAAGTGCAGTGAATACAGCTT  
CATTTGAACCCTTGGGTAGTCCGAGGACAAAACGTAAGTATCGAGTATATCTCTCCCTTTAACCTTCTCTTTGGTC  
AGTTTCTTCCCACTTCATGTCATTTAA  
>NODE\_291\_length\_175\_cov\_9019.673828AATGGCTCATTTAAATCAGTTATAGTTTGTGTTGATGGTATCT  
GCTACTCGGATAACCGTAGTAATTCTAGAGCTAATACGTGCAACAAACCCCGACTTCTGGAAGGGATGCATTTATTA  
GATAAAAGGTCGACGCGGGCTTTGCCCGTTGCTCTGATGATTATGATAACTCGACGGATCGCACGGCCATCGTGCC  
GG  
>NODE\_292\_length\_159\_cov\_3980.635254TTTACGCCGGGATCACGACGAACCGTAATTTAAAACCTGGTCT  
GAATTCATCTCTAGATTATCAAATACTTTCTAAGATGTTATACGTCTCCGAACCTTTACGCCTCAAATTAGCCGCT  
TTAATCGCTACAGCGCAAGCCATTGCGTTAGCCTTTAAACACGTCTGCCGGGTAATATCTATT  
>NODE\_301\_length\_154\_cov\_1537.980469CGGCGATTGGTTACTGCAGTCATTCGGTCTACAGCGTTGAT  
GTTGCTTGTCAAGTTGTACGTCAAATCACTAGCGTGGCTGACATCTCCGACTTCGACGTTACACCTAGATTATAAT  
CTTGTAGGATAAGATAACATTTTCATCATCATGTGTGCGATGTCTGACGGGACCGAAAT  
>NODE\_302\_length\_185\_cov\_3660.129639TACTACAATTAGGCGTCTTCGCGACTGGCGGTCTGAAGGC  
CTCCTTTTTTGGAGACATAAATCAGATACCATTATAAACAGTGAGAAGGTGTTAGGATGGATTGCGCTGTATTTG  
TTCCAAAGAAGGAAAGTGTGATACACTTCTAAATCATAACAGATGTCCGGTAAGACGTTTGTACTTGTGTCTCTC  
AATGACCGTAAA  
>NODE\_303\_length\_517\_cov\_1788.413940AACTTAAAATAAACCGTGAATTGGTAGTGTACTCTTTTTACG  
ACAGTGAATCCACTTTTCAGCATCACGAAAGGATTTCACTTCGAGTTGTGCAAAAATACTATCATATTGTAATTCGTA  
GGCTAAAACCTTTCAAATTTCCCCGAGTTGCGTTTGGTTTATAACACACCGTTATCAGAACTGGCCTAAAACCTCCATC  
CTTTTACCTTTTATATATTTTCTAAAATGTTATACGTCTCCGAACCTTACGTCTCAAATTCGCCGACTTAATCGCTA  
TCGCACAAGCCATCGAATTTGCTTTCAAACGTCTGTTGGGTAAAAACCTGTTATCTAACAGCGCAAAGTCGAAGCAA  
TAAGCGGTAACCTTTTGGGTACTCCGTGTTTAGCTAAAGCAGTATACGCTGGTTGGACCTTCTTTTCAATTAGAAG  
CGTTACGGCGGTGTGCGTAAAGTATCTCATAAATGATCTTACCGTATTCTGTTGGAGACGCATCAAGTTTTCCACCTA  
ATATAGCTTCTCTCACTTCTTAAATAACTTATCGAG  
>NODE\_304\_length\_80\_cov\_3473.462402CACGTGGTTAGGAAATAGTTTGGTCTTGTGCACCTTGTGTGTC  
CGTAGTACTTAGGGGATTAGATTATAGCTATATCGTAGTTAGCGGGCATGATAGCCCTTA

>NODE\_305\_length\_88\_cov\_3788.477295GAAACTTAAAGGAATTGACGGAAAGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACCCAACACGGGGAAACTTACCAGGTCCAGACATAGTAAGGATTGACAGGA

>NODE\_312\_length\_1944\_cov\_353.311218CAGTACAAGTTGTGTGTGGAATTAAGTACCGTTCCGGAAAT TACTGGTCATTTGCATCGTGTCTCTACTTGTCTTCTGAGCTTTGCCAGAACAAAGTTAGGTTTACTCGTCTGTGACTTA TC GTTTCGAGCGCTTCCGACGTTATGGCTGATTACCACTCGCCCCGCATGCATACGGAGTATGAAGCTCTTCTGGCGC AGAGTTCGAGGATTGCAAGCTGGGCAGATGAAGTTGAGGATGCTATGGGCACACCTCAGGGGCTGTTCGGCTCAGGG GTGCTCAACAATGGGGATGAGGGGGCTTATACTCGTGTAGATTATCACTGGGGACGCAAAGGTGTACCGGAGGCTGA TTTACTACGTGCGTCTCGTGGATTGGGGCAAACCTCTATCTTCAGACGTGCTGGGGCTCCTGTCCCAGAGTTGGCCC CGCGTGATGCTGAGGCGTGCAGATATTGCTGCGGCTTGCCAAGCGGGCCATCGGTTCGTGTCCGCTGTGTGGACTGAT GCTGATCTAAAAGGAACGGCTTCCCGTGAAGCTCTCCAAGGACTAGAAGGCAAGCTACTCCGCTTGTTC AAGGATGG AGAAGCGCGGTTGGCACTCCCTAGTGGACACATAGCCCAAGGATATCCTACTGAGGGCAACCCGGGCTATGAAGGTG CTGCAGAGCCTGACACATGGTACACTATACTCTTTCGCGGACAATGTGGAGTTGTATGTGCTAGTCAATGCTACTCTT GAGATCGTTGAGGTGGCAACTGGTGTGATCTGCTAGCGTTACGGGCTGGGTATTTTTTCAGCGTACACTGATTTTTCT CGCGGTGTGTTCTGCTATTCAGTGTGTGCTGCTAGATGACGAATGTCCAACAAGTTACTCTCTCCTGCAAGACGCGA AGTTGTCTGGCATGTTGCAGTTTGCATCTCGCCGAGCTGGAGGTTTCGTCCGGTGTACCGCATTATGGCGGAGGTGGA CCGTCCATATTATGTCACCCCAAGGTGAATATTTATCGATAGAGCTCATGACTGATCGGCGCGTACCTGTGACCCA TCTCATGACGCAGGGAGAATCGCCAGTGGTGGCTCAGGGTATTATGAGCGCAGGGGACGTACCGCCAACCTTAGTCCG TCGCTTTGGGTACAGATGGTGGGAAATCCGTCACAGGTTCCGTTACTGGTTATACGGAACTCACATACTGGTGGCTTC CAGTTAGTGACAACCAAGCCACGGCTAATATAGTGCTCCCCGGTCTTAGGAGTGCGCAGAAAGCGGCTCCCATGGT GCGTGATGGGCTATTTTACCAGCTGGAACAGCGTCTTTGTTTCGGGGACGGACATGGGCGGGTGTTCACGATAGCGG ATGAGACTGAGGTAGTGGTGCCTTACTGCTTAAACGACTCTCGACCCGAATAGACCGTTGCGGGTTAATAGTCGAGAT GTGTATCGTCTCTACGAGTACCTGGTGGATCTAGGTGGCTACAAGGCTGGGAGGTCAAACCTGAAGTTGACTTGAC CCTAGCCTATTTTGACAAAGCGGGTTCGCTGCTGGGAGAGTGGAGGTGCGCCATGGACCCTACAACGCCCAAGGT GGGCGGACGTCTCCCGCAATGGTGGTTAAAAGCGAACATGGGCTCTGACTGTATCATGGATTTTGGCTGCTCTGTT CTGGGGCGGTTCTTATTGGCGCCGCACGTTTCGAGGTGGTTCGATTCTATGGGATTGAGGCCCTTGCCTTAAATACTGG CGGATACTTCCGTGACATAGCATGTGCTGCGAGGGATGGGACTGCTCTGGCTAAGATGGCGCAGATGCCGGTGTCCCT ACGTCATGGATACTGAGTTTTTGAACGAAAGAACTCGTCCGGGAGAGCTTGAGCGTTATGTGTTTCGCTGCTGGGGTT GCTAAATTCTGTGCTGGGAAACATATAGTTAGTTATAGTGTGCGTGATGACTCTGCGTTCCTCCGGCAGTTTCATTGA

>NODE\_314\_length\_78\_cov\_4961.115234CAGCCGAGTTGTAACCGGTACCGCTGATTATAACTTCCCTT AAACGAAGCTTTCGTTGTTGAGGCACTGAAAGGTCTTGGTATACCGGTTAGAGGTGTTT

>NODE\_318\_length\_233\_cov\_1149.806885TGAAAGTCTACGTGCGTATAAGCACCCATTCAAGAAATATA AAGCAAAAACCTCACTGCTATTTTCATATAGCCGTCGACATAGATCGCCTCTACTTCTTTTGCAAAAAGTTGAAAAAC GTTATGTAGGGTATTATTCGGCGGCTCACTGACTTGTTCCTAAACGCTTGTGTCTAGCTAAGTTCCATGCTAGTATAG CGTCGTTGTTAAACAAGTCAACGACGGTCAACACAATCGATCGTGTACGGAAAGAATF

>NODE\_319\_length\_104\_cov\_1672.730713CGGGTTATGAAAACCCAGTGAGGCAGTACCTGGCGTACTTT ACGCCAACATTTATAACAGCAACGCTGAACGGTAAGTTAGCTATGAACGAGAAAAGTTATGGCGCAGCATGGAGTACC TCCAAAAT

>NODE\_324\_length\_84\_cov\_8535.440430AACGGGGCAGGAATATAGTTTTACATGATAAGTTGTGTGACCG CGCTTTTCTCCTGCATCTTTACTGCGTGCCTAGAGCGTTTTGAAATATGTAGTTGGATGAAAGGT

>NODE\_327\_length\_907\_cov\_6541.166504AGAAAAATTAGTATTACGCTCACCGTTACTATAGCAGCGAT CGCCTCACCGTTAATATTCTCCTTACGTAAGCGCAACACAACATTTCTTAAAGTGTGCTCCGTTACACAATGTAAGC ATCAGCATCCCCTCGCTAACCCACAACACTACTTACGTACCCATTTTCTCACTCACCGAAATATGCGTGCAGGAGTAGG TCCGCTTCCCTCAATGGCAGTTAGTCCAGCTTGGGTTATAATCATATTGCTTGCTAGAGCAAAGATCGCTCAAATGTT TGAGATATCTTTTGCTTCGCCCTTTTTCCGTTGCACACCTCGAACAAATTCATCTCTTTAATGGATTTCTATAACGG CCTCAGAGTGCTTAGTTAGGTTCTTAGCTATATTTGCTAAACACCCACCGCGTCATAGTTAGCTGATATATCTATG CTCTGGATTAATATCGACTTATGTAGTTCCGTTGTTGAGCACGAAGCTACGTACGACTTTAAAAACATTTACTTGGCG TGTGTGGATCATCAAACCCCTTACCGAGTGTATCAACTATATTCTTATAACTATTAATTAAGAGATACAGTGCCG ATAGATCATTGTTGAACAAAGCGGAAAGTATTTGGACGATTATGAATGATAGATCCATAGCGATTTAACCGAGAGAA GAAAAGACAATATACTGCAGTAAACACCCGCTCTGAAGTTTTCCCTTCGCAGTGAATTTCAAGGGAGTGTCCGTAAT CCTTACCTCGAGATACTCCCAACCCCAACGATATACCAACTTCCGCAATGTCCGAAACCCATCTACCACTGTGGAATAC CTCCTTTCTTTTTCAAAGGCGTCCGAGAGGTTTCGATACCTTGACCTCCTCCAACGCCGTTATCGAAGCATCGGCT ACGAGTATTAATCGCCGACGAATGCTCTAACCGTCTCCAC

>NODE\_335\_length\_112\_cov\_323.553558GTTAAAAAGCTCGTAGTTGAACTTCAGGCGGGGTTGGGCGGT CCGCTTAACGGCGTGTACTGTCCGACCCTGTCTTACTTCTTGGTGTGAGCCGGCGTGGCTTTTATTGGTTGCGTCCGGG AACCAGGACTTTTAC

>NODE\_341\_length\_84\_cov\_16262.619141GGGCGAACGTGTGAAGCAAAGCATCGGATACTGATTGAGGT GATACATCACTAACATTATGTGCCGACTTTATCGTCCAACCTTCGAGCTCAGTGCAGGCATACACCA

>NODE\_343\_length\_257\_cov\_80.856033CCCTCAGATAGAGCCTTGCCCCCTGATTTTTGCACAATTGTAC GGGCTGCACCTTGGCTTTGTGGTGCCTGGTTTGGAAACACTTGGTGCCTCTCATCCGACACGCCCGCGCTTACCTG CAACATCCTGCCTACGCCGAAGACACGCCGAGTGCCTTGAACGTCAAGCTAATACGCCCTTCGATCAATAGAACTT ACAGAATGTAAGTGCCTCTTTGGAAACCACAGGGCATAAGAAAAGTAGTCACCTGAAGCTACATTCATAACCATGAC CTTAC

>NODE\_345\_length\_79\_cov\_2976.202637TGTTGGACGTGTAGTTCTTACCTGTGCGTAATCTCGATACCTC  
GCTTATTGATGTACTTCTTAGCAAAAATAAATTCCCTCCCCAAATACGGTTTTGTTGTGA  
>NODE\_350\_length\_337\_cov\_1174.596436CCATGGAGTCGTGCAACCCACATTACAACCGCGATCGGGC  
TTGAAATTCCTATTTTTCAAAGTGACTAGGTTACACTTCAAGGTGTTGCGCTTCGGTAAACCTACGCTAGACCTAAT  
AAAACGAAAAACCTATCTTCTATTTTTGACGTAAAATGAGAAAAGGAATCGGAAATGCGCACATCTTCTATAAGGAA  
GGACTGATCTGAAGTTTCAAATCATCTCAAAGGTGATAATCTAAAGCCGACAAACCGTACGAGTAAGTCTCGA  
GAAACCCATTACGGCTACGAAATGACTTGTGGTATCTTCCGTACCAATTCCTTCGAAGGTGCGTCCAAAATTC  
ATACGCTCAT  
>NODE\_351\_length\_80\_cov\_13650.912109TATAACCGGGCCGTCGAGGCAAGGCCAGGCCTCGATGAGT  
AGGAGGGCGGGCGGGCGCCGCAAAACCCGGGGCGCGAGCCCGGGCGGAGCGGCCGTCGG  
>NODE\_363\_length\_145\_cov\_7365.379395GGTGAAGTCTCTTCTGATCCTCGGGTACGTTCTCGCCCA  
GATAGTTATACGAGGGGATTATTTGATACAAAGAGCAGACTTTCCTCGAATTTAAAAACTTATCATTTACTTTATCG  
ATAGCCAGTAATAAGCTAGCTTCGGGTCTTATGGGAAGTGACGCAGTAT  
>NODE\_391\_length\_969\_cov\_314.233215AATTACCGCCGTATGGCTGATCGGCGATTACTAGCGATTCC  
AACTTCATGTTCCCGAGTTGCAGAGAACAATCCGAATCGAGGCAATCTTTCGGATTTCGCTCCGCTTACAGCCTTG  
CTTCCCATTGTAATTGCCATTGTAGCACGTGTGTGGCCCCACCTATAAGGGCCATGAGACTTGACGTCATCCCCACC  
TTCCTCCAGTATATCACTGGCAGTCCCTCGTGAGTGCGGCACGCACCTTTTTGTTGTTTCGGAGCCGTTTTGGCGG  
GGCGTACTAAACCCACTACGTACCACACCACCGGGCGGCTCGCCTGAATGCCGAGTCTTTCCTCCGCCGCAACTCGA  
CGTCGTCGTCACCTGCAAGATAAGGCCAAAAAAGGACTTCTTCTTTTAGGAAAAAGCCCTTCTATGTTATTAGTAA  
AGCGCGCTAGCTGCAAGAAAAGGGCTGCGCTAACCGCTTTAGTTTTGATTGCAGCTCGCGCAAGGCGCTCACATTTT  
GTGGCTCGTCACGACATGTTGTTGATATTGATTGAGTTTGGGGACTTTTCGCTGACTGAACCCATATTACGTTACTC  
ACCGTCACTGGTACTTTGACTAGACTATAGGTAGGTATCAGTGGGGCTTGTAGACTCGTGACGCTGAACGACACATT  
CTAAAACGAAAGCGCACACTAGAAAGTGCTTCGAAAGGCGCCGGCTCCCTTCTTACTGACAGCACAGTACGTGCTG  
GCACTCAATTAGTAGCGCTGGCAGCTCACTCGGCTCCTCGGCTCACTTCGGTTGCAAAGACTTCTCCTTAGGCGCA  
TGTCTCAGCAACACAAAACGAGGGTTTTGCTCGTTATAGGACTTGACCAAACATCTCAGCACAGCTGACGACAG  
CCATGCAGCACCTGTATGAAAGTCAGTACCATCCCGTTAAGGACAGTTTTTGTGTTTCATATGTCAAGGGCTGGTAA  
GGTTTTGCGCGTAGTATCGAATTAA  
>NODE\_396\_length\_122\_cov\_2203.778809TGATAGAAGGGCCCTGCGATCAATCCGAGAAGTCCGAAGCG  
ACTCACAAGAACGTTGAGAAAAGTGCTTGGTCATCGTCACTTCCGTAAAAACTTTTAGCGACCCGTTGAAAGCGTCTTC  
TTGATCTTTGGGGTGACGATGCTGAC  
>NODE\_399\_length\_90\_cov\_1043.377808ACATCGCACCTATGTGAGGGTTTTTCGCGAATCATTGCGTCC  
CTTTGAGAAAACATATCGCACAATTTTCGTGTCCAAATGGGAAAGCAATGCCGATTGGTCCCTTAATACCA  
>NODE\_400\_length\_376\_cov\_204.021271GATCGCTTTGCGACACGTGGCATCGTCACTTCCAGGATATCTC  
GGAAATTCACATCCGGAATTCGTCCGGCCAACATTTCCACCTTCTCCGGATAATCTCAGATCCGGCCGCTGTGCGC  
GGATGGGAGAGGAGGACGATTCAACTTCCCCAGGAAGACATATCTGGATCCTTTCGGTAACGCTTATCCGGGAGA  
CCGTGGCCAACGATTACGATTCCCCAGATAGCTTGGTTCGTCAGATACTCGCTATTTCGCCCGATCCATTTCCGCCCA  
CAATAAAAAAGCAAACCTCCGACAATACTGACGACCAAGTCTCCTGAACTGTCAATCATCTTTAATGCAAGATACACT  
CGACAAGACATTTCCCAAATCTTCTCAGGTTGCCTGACGCATCCCCGAT  
>NODE\_403\_length\_367\_cov\_2457.509521TTGGCGTGGTGGCCGAAACAGTCGCCAATGGTACTTCTTGT  
TTCACTTCTTTCTGAGCATGTACCTCGGACTGGGTGACCTTTGGAATATCAGCCTGTGGCGAACTCACAAGACTGGA  
CTGTACTGGTTCCCTGTGCGCTTCGTTGACCGCGTTTTGTGCTGTGGTTACCATCTGCAGCTTTCCTTTCTTGT  
GTCGGCGCCTCACCGACGACGCTTTCTGGGAGCAACCTTTGCGCTCCGCTCGACCTCGCAAACCGTCTCACCCAC  
TAGATTGTTGGTGAAGACATCGTCGACACCTTTTTGCACACTTTCAACAAAGGCCATGGCATCAGCGGTATTAGAA  
CTCGACGGATCACTTTGTCTCTCCACCACAGCCGCTGATG  
>NODE\_405\_length\_316\_cov\_1514.737305GTAGGGCTTCCGTAATCGGTAAGTAACGTTTCATCGTAGATA  
TCCTTTTCTCATTAATAGACAAAATTGAGTCTGGAAAGCCACTCTACCTTGGCCCTTAGCATTTATTAACCTCCGCT  
ATATCGGACGTCAATTCTACCGTGTCTAACTGATCTGGTATATTGGATGAGTCCGACGGTGAAGCCGTCGTCGATGT  
TGGCGTTAGCAACGAAGAGTCGTCTACGCTCATTACGTCTTCTAGTCTAGCAAGGAAAGATTTCTCATCGGTTTTA  
ATATAATCTTTCTAAAGTCTATATCTCTGCCTTGTAAAGAAAATCGACCTTATAGACATTACGTTCA  
>NODE\_406\_length\_132\_cov\_1616.909058GACCTTATAGACATTACGTTCAATTTAGCCGACGAAAAGCA  
GCCCCGGTAGAGTTTTATATCTACTGAAACTATCTTAGTTATCTTATATATACGAGTCCAGTCTTAGGCAACCCTG  
TGACATCTGTCTGAGTGTAGGATGAATTGACAGCAG  
>NODE\_407\_length\_136\_cov\_3959.073486TACAACATCTTTTTCGACAGATGAGTAAGGATATCAGCGAGAT  
CAGCGAGATACAACGTGCTGGTTACCTAGAAAATACTTAGGAGACGGGCAAGCCGACACCAGACATATTTTTCGACG  
TCTTAACCAATAACAAAACGAAGATAAGGTGGTTAGTTAA  
>NODE\_415\_length\_94\_cov\_12330.010742TCATCTCTAGGTTAAAAGCCATCGGACCTATCGGTTTTAATA  
ATCGTGCAGATCCCTTTCCGCTTGTATAACAATGACGTGACAGCGCTGTACATCTTGATCAACGCCGTAAAAAA  
>NODE\_431\_length\_267\_cov\_2313.000000GCTTAGCTGAGGTGTCGAAAGTCGTTTGTCCAGTAACTCA  
ATGTTACTTTCTGATTGGCTGGACAAGAGAGCTCCTAATGCTTATAAGTCCCTCAAGCGAGCGCTAGGCTCCTTTGT  
TTTTCACCCGTCGATGATGACATCCTACACTCTCATGGTGAAGTCTGACGTAAAGCCGAAATTTGGACAATACGCCTT  
TGTCGAAATACGTACAGGGCAGAATATTGTTTATCATGATAGGTGTGTTACTGCGCTCTTTTCTTGAGTTTTTACG  
GCTTGTGTAGAACGTTT

>NODE\_432\_length\_132\_cov\_3317.386475GAGATTTGTCGTACTIONACTGATGTCCAACCTCGTAGGTGTAG  
TACTGGACGAATATCGCCAGATCATTAAAAAATTTCTTGCCAAAATCCATTAGTGTCCATCCCATGGTAGAACAGC  
CATTATCACTAACTACGTACTTTAAACGTTCTACA  
>NODE\_437\_length\_144\_cov\_3408.687500GCAGATTGGTCCCTTAAAAACCAACCTATTCGTTATGCTCGA  
TACCACTGAAGAATATACCTCCTCTTGATTGAGAGCGTTAGCGACCTACGCTAAGATCTCCCTCCAACCCGCGTTC  
TGTCGCCAGCGATGACGACATCTTTAGGCACATTGGGGAGATATACTG  
>NODE\_439\_length\_91\_cov\_797.362610CCCCGAAAGGGAAGCCGGTTAATATTCGGCACCTGGATGTGG  
ATTCTCCACGGCAACGTAAGTAAAGCGAAGACGACGGCGGGGGCCCTGGGAAGAGTTCTCTTTTCTTCT  
>NODE\_440\_length\_86\_cov\_5926.686035ATATACCTCGAAAAGCAACCAGATGGCGCGAAAAGTATGTAGTA  
GAGCGTCCAGACGGACTGAGACGACACATCCCTAATATATGACCCAATCTTTGTCATCCAGCTTCGA  
>NODE\_475\_length\_193\_cov\_212.528503CGGTGCAATCTGCTATGACGCCTGAGCAGCAGGCGCTCATGG  
AAGCTGCTGGAGTGCCTGGCGGGGTTGATAACTTACAGGTGACCAGCTCGCTGCGGGGCTTGAATGCTTGGAGAC  
TATAGGCTTGGCATAATCACGTCCACCAATGGTACTGGTACTGTGCAAGTACAGGGACCAGGAGATGGGGAAATTTG  
GAAAAAAAAAAAAAAAAAAAA  
>NODE\_485\_length\_110\_cov\_5807.745605TTCTCCCTGACAAAATATACCTAACTTTTCGCAACGTCGGA  
TCCACAATCACCACATAAAAAATACTTGCCTCTTCTTTGTCTTTAGGCCGAAAATATCTTCTAATTGATTTTCGCT  
TCTTCCAACGCTGA  
>NODE\_486\_length\_247\_cov\_3738.340088TCTTGGAGGTTGTCAAAGTCGCTAAAAACGAACGATCACCT  
GGTTTTGATACTTATAAAGGAAGCGCTTATAAACTATTACAACGCTCTTTTAAACCGATGAGAGTGCCGTACTAAGC  
TACTCTCGCGAAAGTATAGGGAATTTTCTCATAGCCAGGTGTAGTTTCGCAAACTCCTGCCGCTCATGAAGGGCT  
TGATCACTAATATAGGTCTGTAAGGCGACGATAGAAACAGCTAGAAGCTTTATCGGAGACTTAATACTCGTCGC  
>NODE\_490\_length\_250\_cov\_5982.388184GCGGCACAGGAGCCACCTAAGCAGCCGACAACGCAAGTGGG  
GGAACCACCAGAAGCCGATATAGGGTGGTGCCGAATCTGAGACTCTTACACCAAATAAGCTGGTTTTTCGAGAAA  
TCCAGACAATTTTAAAGACTATGGGTAAGGGGATAGCTTTGGACGTAACGGGGTTACCCATAAACCGAAAATTT  
AACGAGCCGGGAAAAGTATCAGTAGAGGTAGCAATGAAGATTAATACCGCATTGGTAGAGCTGTGCAAGAAGTTAT  
>NODE\_497\_length\_115\_cov\_1451.469604ACATGCATGGCTTAATCTTTGAGACAAGCATATGACTACTG  
GCAGGATCAACCAGGTAGCATTCTGTGCGACGCCGGCACCGCATGCGCCTGACCGACGCCCCCGAGGGGAGCGGG  
ACGGGAGGGCAAAGACGGG  
>NODE\_500\_length\_83\_cov\_2214.144531ATATCTGCCCTAATTTTTCAGTTCAAACGCCATCGCGTCCACTC  
AACAAAAACCACACTCAAAGCAATAAAAAATTTTATAAACTAATACGAGAAAACATCGTCGC  
>NODE\_503\_length\_464\_cov\_2426.655273TTTTTGTCTGAACAACACGACGTCCCAAAGGTTTTACCT  
TGAGCTTATGGACCGTACTTACTGGTGTTCCTTCCCTCCCTAAGCGTCTCTTATATCGGATTTCTCCAA  
TTGGGTCATGCACAAATAGATTTCCGCGTCGATACCCGCCACGTCATCGGTCTACCTATCGGTCTTTAGACAAAG  
ACCTGACCACTGGCCTCTCTTTACCCTAATAACCTTCTCTGGGTAACACTTCTCCGTCCTTTCGAACTTCTCCG  
GAAAGCAGGTAACAGACGTCCAATTGGACACCTATAAGACTTAGACGTGAAGATAATGTTCTCTTTTGGGGTAA  
ATACGGCACAATCCATTCTGAACACCTTTTCCCTGTTTATGAATGGAATTTGATTGATATCCCCGAAAAGAACCT  
TCTGAAGCGCCGGTCGCGAAGACACCCAACCTGAAGAAGCCCTTGGTGCATCATGTACT  
>NODE\_509\_length\_147\_cov\_6060.775391GAACTGAAATTAGGGCAGATATATGAAGTTGTCCCCGAAA  
CAATCTGAGAGTTAGAGTTAGGGGATGCGGCACAAGGAAAGTTTAGTAAGGCGAGTTTCTTAAAATACGTTAAGGAC  
GGGACGCAGGCGGAGTTAACGGGAATCGTCGTAGTGCCGAAAAGTATGTA  
>NODE\_522\_length\_125\_cov\_3433.632080CTGTGGTAGCACTGTTGACACCCTCAAATTTGGGATCACGC  
TTCTTACGGCTGCGTTGATGTCTCTGAACCTTCTACATTGCGCGTTATCACCGCGGTACTATTGGGCGCCCTAAC  
GAACTCATCCACCAGCATCGTCTCTTAC  
>NODE\_559\_length\_272\_cov\_2877.103027TACTGTCCACTGAACGCGACGCGTAAATCTTACCTTTAAC  
CAGGGTCTTGAAAACGCGCTTAAAGTCTTTAGATGATCTGATCCAATTTCTCACAGTTTATTGGGTTGGAGAGTGTGG  
TAGTCAATCCGCTTTCAACAAAAAATACGCGAAGCGTCTATTGTAGATTGCGAAGTTGATGTAAACCCACTTATCCT  
CGCAGCGGAGGTTTTCCAGACGACGGTGTACTTACTATCACGATTCAAAGGAGCGCCCTCCCGGTCCAATGTATT  
CCTTATCACTACCGAGAACAC  
>NODE\_565\_length\_85\_cov\_2789.282471TTTTTGGGTTAACGTACAACCTGCCGGACGTACCTTACGAA  
ACGCCTTTCCGGCTGCTTTACCTATGAAGTAATTCGAACCTCCGTGTGGTATAAAAAATCTGGGTT  
>NODE\_566\_length\_275\_cov\_2276.399902GACTTTTTGAGTAGAGAGAGTGGTGCATGGACGATTTTCAGA  
CAAGCTATACTGATACTAGTGGTCGACTTCGTTTTCGTTATTATCCTTCTGTTTGTATTCTTTTGTAGTCCCAAG  
GTTACAACAGGGCGTAGCCTTAATACCGGTATTAGGACCGTATAGTTAAGTGAAAAGGTCGCGATGGAGGTGGGTAT  
AGACTTTGGTACGACTTTTACGTAATCTGCTTTTTTACCTACGGGTGTAGCGGGGTGTACGCCGGTAGCAGGTAG  
CGTTTTCTGTCGAAACCCAGATTTTT  
>NODE\_589\_length\_225\_cov\_4266.759766CCCCGAAAGCTAAGTTAAAGGTCATAACCTTGCCTTTTCC  
TACACCTAATTGGAAAAGCTGGCCTTTCCAAACAAGTTTAAACCGTCAATATTGGATTATCGTGGGGATAACATTCT  
TCGTATCGGCGTGAATGGTACTGCCCCACCTTGCCTATACTTCTGGACGAGACAGTGTATGAAAACACTTGGGAA  
TTTCAAGACCGACAAAATGTCTGTCAAAATGCGCGGCCAACACAGCAGAACTA  
>NODE\_593\_length\_101\_cov\_3186.326660GTCCGTACCAGTTCTGAGTCTGACTGTTTCGACGCCCCGGGA  
AGGCCCCGAGGGAGCCGTTCCAGTCCGTCCCCCGGCCGACGCGGGGACCCGCTCTCGCCGCGGGAGCAGCTCG  
AGCAG

>NODE\_611\_length\_1743\_cov\_824.629944GACGTAGCGTCGCAAAACGACAACACTAAGGAAAGAAAATGA  
ACCTAAATAGAAGAAAAGAAAAGCGCCGCTAAAAGGCAAAACATAAAAAAAAAGAAACCTACGTAAAAAGACTAAAAAG  
AAAAAGGAAAAATCTTTGGCGTAGCGCCAAAGTAACCTACGTAATTATGTACATTTAAACTAATAAAACAATAAAAAA  
TGAAAGAATATCTACAAAGAAATTCTAACTGTACATCACGCCTAAACCTAAATAAAGAGACCCTATTTAAACACGGG  
TCACCTAACTACGGACTGTGGCGCCTTTTCACGCCGACCCGCTTAGTGGCATACTAACCACCACACTACTACGGGGAA  
CGGGCCTCCCAGCCTACCTTAGGCTGCTGCCACTAAGCTATGGAGATATATCCCAGCCCCGACATTTTGGGCTCCAT  
ATTATGGTGTTTAAGAAAGCCGTACCTAGCCCAGATACTACTACGAGACGCCCCGACGTCCCAGCGCTGGCTACGGCC  
CGAGTCTCCGCCGCTGGTCCCCGCGGTCAACTCGTTCAATCATGTCTGCTTAAACCAACGGTACCTTCAGGACCT  
AGCACTTTTCAGCGCTCCGTAGAGCCGGCTGCTAGACACCTGACCGATGCTGCCCTAGGGCTTCCGCCCAAATGCACG  
CACGTCTTACGCACAATGCCGCATGCACGACTTACATGGCAGCTACAGTCTTAAACCTAGTCTGTTCCCGTTCATCCCC  
CAAAGACACACTAAACTGAACACCTGTACATCACCTAAAAGGATATTTGGCACACTAAGACCAAATCCAACACTTCC  
AGCGCACTTTAAGTGCAGGCTGTTGGTACCTGTTGGCCCTATCTACCTTGACACAAGGCTAAAAGTCCGCGACAC  
CTGGCGCGCTATAGCTGATGTGTCTTGTCCGTAAGATAACAGGAGTATAAACTCCCGCATATGCGCTCCGACAAGA  
GTAGGAAGGGTAACCAAGACTTAAAGAAAAACGCTCTACTGTACATGTTCCACCGGCTTACGCTCACGGTGTGAA  
CAAAACAAGAAAAACACACAAAAAAGAAAAATGGGTCTTTAATACCCACTTCAAGAAGGAGTGCACCCGGTATTAAC  
CTCACGGATGGCACGCCTAAGAGAAGATGACAAACTGGAAGACTGAGATAAGTACGACGACCTTCGGTATTAACCT  
CACGAAGAACCAGGCACAGAAACACTATAGCGTGGCTAACGCGATAAACCCCTCTTTATCCGAAAAGGCATAACGA  
GAAGCTTTCCACCTTAATATTCTCGCCTTCAAGGCCAGCAAGGAAAGCGACTCCGTCTCTGAATAGGGCGCACGTC  
AAGATTCACGCTATTGTGGTAATATAACCTACAGAACTGCGAAAAATTTGCACTTAAAAACGTGCAAGACACACAAG  
CCCGAGTAAGTCCATCCCGAATGCTTATATTTCCGCGTACCAGCTTAGCTAATTCCTGGATGACGTCCTCTCTATT  
GAAACCTTTCGAAAGATCGACGAAAGATTGAAAAATCTCATGTAAAAGGTCGATATCTGAAGTTTTGGAAAGTCCAA  
ACTTAACGAAGAGTTTTAAGTGGATCGGGAACAAAAAGAGACTATCCTCGACTTGAACAAAACTTAGAACAAAA  
TATGGGAGCAGCTTGGTTAAATATCTTTACGTCAAAACCAAATTTTCACTCAGAACCAGCTATCAATATCTAACG  
GCTGCCGACTAAATATAAGGCTATCATCGC

>NODE\_626\_length\_79\_cov\_2981.860840GGAGAGCAGTGCTAACGGCACTTGACGCTTAGCTACTGTTGC  
CCTAGGTTACCGCTACCGCTGAAGATAACCCACCACGGGTTACAGCCATAACCACAACAT

>NODE\_627\_length\_893\_cov\_1674.977661GGGCGAAGAGTATTCGCTCTTACAATTTTATAAGAGCTAGT  
GATGGTGGGTTTTGGCAAACATCTTCAGGCTGTTCTTTCAAGCGCTGGGCGATTGTTTTTCGGGTGTGGTCTCGAGTT  
ATGCGAAAAGTATCGTGCACGATAATTTCAACATTTCTTGAGACATTGATAGCCATGCCTAAGTCTTTATCGTTAAA  
GTGCCAAGTCTGTGTTGGTCACAATTCATACTTCCGGGGCTTCCGAGAAGCTAGAAGTGAAGTGGAGCTTTCAATGT  
TTCTAGGGCAAATTTCAACAAGAGATTGCGTCTAGTTCGCTTTCGGGGTCTTCTCTAGAGCAATCGTCGAAGATTGCA  
TAAAGTCAATGAAGAGATTGAAAACGGAGGATGGGAAACCCGCTCCCAATAACTGAAGATTCTGTGTACGCCTTTATA  
ACGTCGAACATGTCGAAGCTGCACCTGCACCAGGCTGGTTGCTTGGAGGTTCTAAGGCCACCCGCTCCCTCTTTAGC  
TATGAAAGGTGCTGCCTCACGTACCTTCCGGACGAAGGCCCTTTCTGGTTTTAACCTCTTTCTTTCCGCTGGCAGCC  
TTTTTTATGATGAAGACCTGACGCCAGATGAGAGGCTTGAAGCATTAGCCAGACGTGAACATGTTTGTGAATTCACCG  
ATGGGTCTTTTGGAGTCCAAGTCGCCAGTGGTTAAAAAGGTTGTTTCCGGCTCAAAAACTTTCTTTCCGGAATTGTC  
TATGGAAGATTTTACCAGTTTCGTCTTGAAAAAATAAGAAAGTTATTGGTTTTCTTCGCTTTAAGCATGGCTCTTGCTC  
CCGTCGCTTGAAGTATAGAAGGAACATTCAGCGGGTAAGCCGGGAAGTTTTCTTACGGTTTTTGTTCGGGGTTGGG  
GATGTTGGTTTTGAGTGTCTCGTGGCTG

>NODE\_637\_length\_125\_cov\_4690.855957TCCAACCTTGGGTTTTACGTCTGCTTTACCATAAAGCGTATA  
AGAAGTTAACATAGACGGATGAAAGACAAACGAACCTAAAGCCCCGCTTGAAGAGACTTGTAAAGCTTAGTGGCCCTT  
TTGTCCAACCAATCAGACTAACAACATCG

>NODE\_641\_length\_209\_cov\_2539.583740TGGTTTTAAGGGACCAATCTGCGCTGCTTTCCACCTCGAT  
ACTAAGCTGTGCGATATGTTCTCTCAGAGGGACGCCATGATCAGGGAGAAAGCAATCAACATAAATGCGACGTATTT  
CTGAAACCGAGTAAAAGGAGAGATTGAGGGAGCTTTTTCTGAATTGCTTTTGGTTTTCTCCGATTCCGGTGAGGAG  
CAGTCAACCCGTTTGCCAACGCTATACGTGGCTGTTT

>NODE\_643\_length\_160\_cov\_2274.706299AACCAACAAAATATTTTTGTCTTTTTTCATGTGTTTGAACCT  
GCCCTCTACTTAGGATCGTCCCTGCTTGCATAGCCCAACTAAAATCGCTAAAATAAAAGGTTATTAACATGCTCACAA  
TGTTTATTCTATGAACACGTCCATTGGGTCCGACACTTCGCTCCTCGCTATACAGTAGTGTAT

>NODE\_646\_length\_166\_cov\_2690.608398AGAAACTAAAACCCCTAACACGTGAAAATCGACAGTTGG  
AGGAGCTTTACTAATGGGGGGTTTAGGTTCCGGACCAGATGTTTTGCTACGGGTAGTAGATGTTATATGTTTATTCT  
TAAAGGCGTTAATACTAGAGTGTGAAAGACAGACGTCGTCGACCGTGGCCCGCCGCTTGTAAACGGTACC

>NODE\_648\_length\_123\_cov\_4041.398438CATCGCGTTTTGAGTGGACAGTAACTTCCAAACGAAAGACAG  
GGAACGATATCAAGGAAGCGAGTGCAGACTATCACCATCAGACGCCGCTTTTTGCAAAGCCGTGTCGGTTACAGT  
GGGAAGTATGTGGACGTAACGCAGAA

>NODE\_654\_length\_154\_cov\_2556.889648ACGAAAACCTCTTCGCACAACAAATGACCCGCCGGATACT  
GTGCTAGCGGTTATACCGCGTAAACCACCAGCGGCTTGAGAACTCAATGAGCTTAGAGTTTCGTAACACTACTGCCCGC  
ACATCTTCTGGCGTGAGCCGCAATTCGCGCCTGTACTTTCCATGCAACAGGGGCCGC

>NODE\_655\_length\_225\_cov\_817.182251TCGCGTGTAGCTGTGAGCAAGGTAGTAGCTGGATCGAAAGAA  
TTTTGGAGTGAAGTTTTCTTGAATGACTTTACTACCTTTGTATTGCGGAATAAGGTGCTTATTGGGTTCTTTGTGGC  
GACTTTGGGTGCGGTTCCAATCGCATGGAAGTATAGGCGCGGACTTGTGGCTAATGCTAGGAGACATGCGGGCAGTA  
GTTATGAAACTCTAAGTTCGTTGTCTTCAAGCCGCTGGCGGTTTTACGCG

>NODE\_658\_length\_110\_cov\_89.236366CGTTGGGGGCAGGAGTTGACGGCAATGGAGTTCCCGTTGGGTA  
TGACAGCTTTGTTGCATCAACTCTTTTGCCAAGGCAGCATGGCATCAACCCTGATGGCGTGCCGTACGACAATGCAG  
CCAAAAGACGCG  
>NODE\_661\_length\_166\_cov\_9482.699219GACCGGACACGCGTCTCGTCAAGCCGCGTGCCGCTCATGGG  
CACAATGCCTATTTCTCATCGGAGACGAATTTTCGGGAGAACATAATGAATGACACCCTTCGCACTCTTGACGCTG  
CCGTGTCTAAGGAGTCATGCGCGCAAGGGTGAGTATTGGGAATGATAATGACCCCGCCTTGGGGCGGGC  
>NODE\_668\_length\_303\_cov\_508.135315CGCTAATGAACTTCAAAGGCGTTGTTGTACCCATGGGTCATG  
GCAAGACAACCTCTGGCTAAGGAGGAAGGATGGGTGGATGTGGATGCGCTGTTGCAACCAAGAGTGCTAGCTGATCTA  
AAGGAAGAGTTTTATGACCTGATAGCCTCAGGCACATCAGTGAGGAGGCATCAGTAGTGCTTGCCAAGGAGATATC  
ACCTGGGCTGAACTACTGAACCCATCGGAGACTGTTGTAATAATGTCCCCAACCTTCGAACTTCTAGATGTCTTAG  
GGGTGGAGTGTGTGGGGCGATTGTCGTTAAACCAGAGACAGTGCTTGCAGC  
>NODE\_670\_length\_1035\_cov\_228.138168ATTACGGCGATCATTACGGTGGTCACGTTTTTATAACGAAA  
TCTCGAGCAGTGTGGTTGATTTCGTTCAACGATTGGTTGAGGTTTTAAGGGTATGCTCTTCGAAGCCTTTTCGAAGG  
ATAAGGGGGGTATGATCTGCGCTCTTCGGACGCGCAGATCGCGCAATGGGCACGCGGCTGTAACCTGGATGGTTATGG  
TTGGAGTGTATCGCAGTTGCGCAGCATGTAGCGTAGTTAGGGGGTTCGCATCCTGCGTTCGGATGCCCCGCATGTCA  
CTCTATGGGTCTTCGGATCCCACGAACGGGCGACTAGTGACAGATTACGTGAGGTGTAATCCGTGACCCAGAAATG  
GTTTCGCGGTCTTGGTCTGAGCAGAGGTAAGGAGCTCAGGGCTAAACCCGGCCGTAAGGGGGAGGCGTCTGCGCGT  
ACATTTACAGTACGAGGCCATCGATCAGGTTGTGTCAACGTTTCCCGCGCTCGGGGGCGAAGGAACCTTGGTTCCGTT  
GACACTTCCGTGCGATAATAACCAGACAAGCTGGCACCATAGGCTAGGGCTGGGAGTGTGATGCGGTAGAGGTAGAAA  
TCCTCAGGTTACGTGTGCAGCACGCATTGGGGCGTACAACCTGAATATTCTCATATTCGATCATCTCCGCGTGATCGT  
CTAAGTGGGATGTCCCGCCACCTATTGATACGGTGGAGCTCGGGGAGAACATAGCTGTTGGGAGGGCGTACATAGGG  
GTTCCGAGTCCCCTACGTCAAATTTGGCTGCACCCGATTGGAGGAAAGAGTGGGCTTTGTCACCGTAATTTACGGATT  
GCTGTGACCGAGCGGCTTAAGCCGGCGGACAGGTGACCACTCTTGCCTAAGGATAAGGCTATGCCTTGCCTTGACTA  
CCGGCCAGATGGCCAAGTCACACAAAGGGTACAGACCTACGTGCCAGACTGGTGTCTGGGACCAGGGATCCTTCTT  
TTGAGTTCGATTCTCAAAGCCCTGGGAGTTGAACCCGCGTCTCGATACGCGGCTCTTCGGCCCTTAAGGACCAT  
TATGCTGCGCAACTG  
>NODE\_676\_length\_189\_cov\_4003.169434AGCGGTGAAATGCGTAGAGATCGGAAAGAACACCAACGGCG  
AAAGCACTCTGCTGGGCCGACACTGACACTGAAGAGAACGAAAGCTAGGGGAGCGAATGGGATTAGATACCCAGTA  
GTCCTAGCCGTAAACGATGGATACTAGGCGCTGTGCGTATCGACCCGTGCAAGTGTGATGCTGTAGCTAACCGGTTAAGTATC  
CCGCCTGGGGAGTACG  
>NODE\_677\_length\_165\_cov\_3177.236328CGCTATATTTCTAGCCATCAATAGAGCCCTATACGTCCCTC  
TCAATTCGAATTGCTAGGTACTTTCATGTATTCTACCATGCATCTTCACTAGTCCCTGTTTTACCACCTTCGAACGA  
CATTGGAGTGTAGTCCGGTTTCAGAGCCATGACGGGGTTTTTCAGATCTGAGCCATGGCCACTTCTTAA  
>NODE\_682\_length\_482\_cov\_1015.373474GTCAGGTGTTATCGCTCCGCCCATCTACGAGCCCCCGGA  
TCATTCGCTTTTTAAACAGTAAATTCGTTTCGATAAAGTTGCCGATGCAATGATTCCGATAAAACAATAAGGGTGTCAAG  
CGTGTCCCCGATCTTTTCGGACATTAATGCACATGATGTCTCAACCAGTGTGTAAGCTCGATGGCGTCCCTGGTGT  
TCTCCAGCTCGACGGTCTGTTGGGGTTTTTCGCTCCCCCTCCGCTGTCTTCTCTCTCCACTTGCGAAAAGCGTCC  
ACCCTCCCCTAACACTATCGAGCGAGGTGCTCCTCGATGATGCCGGGGCTGCGTCCGTTTTTTGTTACCGGTGCCATA  
TGCGTGGGTGATGCGCCTGCTGATCCGCTGGACTATGAGGCGTTGCGTCCCTCGTCCCAATGCTATCCCTCCTCGG  
CCCGTCCGGGATAGGCGTCAACACGCCGATCTCGTGCATGTTGTCGCGGCGATATGCTCGTCTTTGAGGGTGATG  
C  
>NODE\_687\_length\_415\_cov\_3580.684326ACCCTCACATAGGTGCGATGTGTTTCTGAAGCCGCAGGAAA  
AGGAGAAGCTGAGGGAACTTTTCCAGAACTTTTCGATACAGTTCTCCGACTCGGTGAGGAGTAGTACCCATTCGCTA  
ATGCCATGCGGAGCTGTTTCAATGGAATCTTTTCAGGAGGTGTGGTAGTGTATGCTTCTTCGATATTGGGGGGAGC  
TTCACGTATCATGTCAAAGCTGGCCATGTGAACTGTGATGCAATCCAGTCCTAGACGTTAAAGATGTGAAGCG  
GAGAATCAATGAGATCCTCTTTCTTTCCACAGCTGGGGAGATTTCGTACGTGTCCAGTGACATTCTAACTGAAGCGGC  
TTCAAAGTCTGTGCTTACTGTAGTCCGAGAATCGCAGAACTGCGATTCTAGAGCCGATGCGGGTTTTATGGTGGAT  
GTGTCCGATAT  
>NODE\_708\_length\_375\_cov\_2095.818604ATGGCTACCACCAGTATACGTATAGCCTTCACCCAGCTTCG  
AAAAGAAAACAGCGCGGCCCTCTTAACTTATCGACAAAATGAAAAGATTCCGAGAACCAGGAGAGATTGTCGAG  
TCGTTGAAGACATCTAGCTGAACACCACAGGAACACTTCTAAACATGCACAGTCTTTGTGCCCCACGTTCTCGTT  
AATTCTCAAGACCTTTTGTGTTGTTGCATGACCTTGGGCTGTTTGACCTCTTCAGAACGCATAACCGCTTTACCCT  
TATCCATAAACCGTTTTGACTTGGCCTTTTTCTTTCAGGCACAGTAAGTCTTTCTTTAACGCTCGGGATTGGGTTTGC  
GGTTGTCCACGACAACGCCTCCAACAACCTGGACACACGCCTCCCTT  
>NODE\_712\_length\_115\_cov\_2155.878174CCGGTATTTGTTGTTTCGTCAGGTTGCCTATCACCATGCCA  
ATTCCGCTTCTGCCCTGTGATACGTCCGTCAGTATGGTTCCTGCATCGAAATGTGGACCGGTCTTGACTGCCGCGA  
TGGAGTGAGCGGCAGGCCC  
>NODE\_713\_length\_144\_cov\_95.298615TCTGGGGTCCCCCGACAGACTGTATCAGATGATGTGGACAAT  
TGCCGCAAACTGAACTCGGTGAGGCGCTGATCTCAGCCTTCGAGAACATGCGGGGAAAATGTAAGCTGATGGCC  
ACATCGTCAAACGCCTTCCCCATTGCAATAGAGTCAACAGCATGGT  
>NODE\_723\_length\_371\_cov\_2784.571533GCATGTATCAACCGGATTGAGGAGGTGATAGCCCTCCTTCG  
CGGGCGTGGGCATCAAGCCGAGGATCTCCTCGTTGTGGTTGAGTGGGGTGGGGCCGTTGATCATGCGCTGTTCTCTC  
CGCTGCCGCTTCCGCTGGGATCGATATTGCTCTTGATGTTGCCGCTCTGGTATAGATCTTCCGGGAAAATGATGTTT



TCTTTTAATTCTTTTATTTGTTGGTTTTAGTAGTCGTAGTATAGTCTTAGTAGGTTTTCCCTCGTTGGGTGTAGTAGC  
AACTCAACGTGTTTTCTTTTTGTTTTCTTTATTTTTCTTTTCGTGTTTCATACGCGTTTTTCGTGGCGCGTGTGTTG  
CTTTCCTTAACGTATATTAAGTAGTGTAACGCGGTACGCGGGTTTTTCCTCCTCGCTATTCTACATTTGTATTCA  
GAACGCGCGTTTTGTTTTCTTTATTGGCGGGTTGAAATTTGCCCTTTTGTTCATGTTTTACTTTCACTCAGTTT  
CAACGAGCTAGTTAGTCATCGCATCGTCGTGTATGTACCAAACGACGTTACGCTGATAGTCTTTCCGTTGAAAGG  
AGTTTCGACGTAAGAGTGTCTTGTTCGATGTTTGTACGATCAGGAGTTAGCACGTATGGTTGACGTGCCCTTCCGT  
GCTGTGCGGTGCTTAGCTAAGATTTGAATTTTTGAG  
>NODE\_782\_length\_131\_cov\_3180.870117ACGCCCCGAGTCGATCGGCGGACCGGCTCCTCGCCGTTCCA  
CATCCGACCGGGGCGCATCGCCGGCCCCATCCGCTTCCCTCCCACAATTTCAAGCACTCTTTGACTCTTTTTCA  
AAGTCCTTTTTCATCTTTCCCTCGCGGTACTTGT  
>NODE\_790\_length\_84\_cov\_1507.047607GAGGGGAGATCTTGGCCTCCGTCGCCAATGCCCTTAATCAAG  
AGGAGGTCTATTCATCGGTGTTTTGAGTATACCAATAGACTGGTATTAAGGGACCAATCGGC  
>NODE\_808\_length\_354\_cov\_511.861572CCGGCTCGTCGCGGTTTTCGGGGGCGCTCCGGTCCCTCCATCTC  
GCAACTCTTTTTCTGACGGTACCGTCGTACCGTCGCGGAGCGCCCGCTCTTCTGCCAGCACCATGAAAAACGG  
CGCGAGCCGATTCTGACGTCTCTCTCTCCCCCTTACTCTTTACGAGTATGGGTTTCTCGTGCCGATCGTGTCTG  
GTTTTGGCGATACGGAGCGCTTTCTGCTGGTCCGCGGTCGACTTCATGAATAGTGACGCCGGTGCCTACCTCCGCG  
CGATGGAGTCGGAGTACAATATGGTCTTTTCCCAGTTTCTCGCGCGCTCTGTTCCCTCGCTGGATCGCGTGTGCTC  
GTCCTCGGCTCGGGTTCATCTAGAGC  
>NODE\_809\_length\_81\_cov\_2086.901123ACCATCTTTCCGGTCCCAACAGCTATGCTCTTACTCAAATCC  
ATCCGAAGACATCAGGATCGGTGATAGTGCCCTTGTGAAGGGTTCCACCTCCGTTTAC  
>NODE\_810\_length\_146\_cov\_149.815063TTCCGCGGTCTTCTGACCGGTCTACTCGCCGCGTTGCAGGC  
CAGCATCTGTGGTCCGCTGGATAAGACTTGAGGAATGTAGCTCCCTCGGGAGTGTATAGCTCTTGTGATGAC  
CGAGCGCCGGGCGAGGTCCGCGCTTCCGGCTAGGATGCTGGCGTAATGGT  
>NODE\_819\_length\_201\_cov\_103.432838GTACGACTTCGAACAACCTCTGAGGCTGCTCAAAGTAGCACA  
GGTAGAAACCAAGAATTGCGCCAGCAGGAGACTAGCCACAAATGGGCTGAGTTGTGCAGATATAATAGAGCAGATGT  
GCGACNNNNNNNNNNNNNNNNNNNCGATACTGAAGCTGTGAACCGGTATCACCGGGTGGGGGTGTA AAAACACAACTAA  
ATCAGGAACACTTAAGCTTACTTTTTAG  
>NODE\_820\_length\_86\_cov\_3854.162842TAACGTGTTGTGCGTTATTTTTTCCGTTGGTGATAATTTTTTT  
AGGCGGTAGGGGTATAGATAGAGCTATAGTTGAAGTCATTA AAACAAAAGATTA AAAGGAAAAAGTGTC  
>NODE\_821\_length\_400\_cov\_2351.135010TGTAAGAACAGGGTATTGGACTGCCCTTTGGGAAAAATCACG  
ACTGGTGTGAAGTCTGACACCTCATCTGTAAGTGTGTTTGTCAAAGTGTCTACTATCGTAAATCGCTGCTACCAGCT  
AAAATATCGCAGTAAACTTTTGCGCCACACGCTACAAGAACATCTAAAATCGGAGCTGTCAAACACCACCTTAGATA  
TCTGAGGTAAACGCCACCTTTTACGCACTTTTCGAGAGCGCGCTCGCACCACCGGTATCACAGCGATAACTGGA  
TCAGGGTAAAAATTGCTTAAACCCGTTTTGAATACGTCACCCGCCCTTTTACTAAATGGTCCGACTATGGCGTCCAA  
TTTTCTTCTGAGGTGAGATCCACGACCTCAACGCCACCTTCCACTGGCAGGAGTGTGCGTTATAGCGTTGGT  
>NODE\_827\_length\_84\_cov\_3883.583252AACAACAGTATTGCTGTTTTAAAAATCATCCATAACGATTACG  
CTTACTTCAATATAAAAATCACGATACTACCTTCGCTCAGCACGAAATCATTTCGCGTGGCGCCTT  
>NODE\_828\_length\_134\_cov\_4871.507324TTGTGCTTTGCGACGCTACGTCGAAATAACGAATTTTCCCTT  
CGTGTAGGCGTCTTGGTGTGCTTTATCGGCGCTAGGTGAGGTTTAGTAGCGACATAAATAGGTTTTCTGTCGAGAT  
TGGGATAGAACAAGTTCGCCTTAAAAAGAGAAATAGGG  
>NODE\_838\_length\_112\_cov\_975.955383ATTCCAACCAAAAAAGAGCGTCGGGATGATTGAGATGGTTTA  
GCCGCGTTCGCCTGGATGATGGGCGAGCGGATGTCAGGCCGAAGTGTCTACAACAGAACAGGGAAGAAACCAGT  
ACAGCTAGGACCACG  
>NODE\_846\_length\_120\_cov\_291.799988CATACATTGCATCTCTCATTTCGACACCGCTTGAGAAAATGG  
GGCTGAGGCGTTCGACGCGCTCCGTAGACACTGGCCTTGCTGGATGAAGCATGTACGAAATGCACTTACAGCGCGC  
ATGAACCCATCAGCGACATCGCC  
>NODE\_850\_length\_511\_cov\_282.093933CCTGAGGAGAATCGAAACCACCCGTGTGGTAGGAGCGGGCGA  
TTATAACAAACAAATACATATATTACATTACAATGCATTACTCAGACACTGAAGATACAGGCGGGGTATGGACGGCA  
GATGAAATATTGAAATTATAGAGCACGACTTTTGGTATGGTTTTAGGTACAGGTAGGCTGGAGTCCGGGACGACC  
CGAAATCACAATGGTCAATGCCAACACTACCTTGTGTCTGGTTCATCCAGTAGCATAACGGTTCATGCCAGGGAGC  
TCGGAACCTTCAATTTTATCATCGAGATCAGGAGTAGTTCCGTGTACAGTAGGCAAGAACTGCACAATTTATCTTATC  
GGCCTTAGTGAACGACAATACCGCACGAGATTGAGAGATGATACTAAACGACAGGGTGTGTGACAATAAATGC  
ATTCTATACTTGGTGCAGGACCTGAGGAGAAACGAAACCCCCGGTGGGTAGGCTCGTCCGGTCAATTATTACATACAG  
TACAATTTACAATCACACACACACAAAA  
>NODE\_851\_length\_87\_cov\_1854.252930ATACTCCAATCTGCCCTGACGTTCCAGTATCTATGTCACAG  
CTTAGCATCGAGCTGATGGCATGGGCTAAGTTCAATTTCCAGCCGGTTGGTTTTCTAGTATTCGGCAGC  
>NODE\_852\_length\_295\_cov\_2063.830566ATAACCTTCTTACACAGTTCCATAAGAGCGGCGTTTTATTTT  
CATGGCCACCTCCACGGACACTTTTTCTGGTTCATTAATAACTTTTTGGTTTTATGAGTGATCCCGTTAGATCCAACG  
CTATGCCCTTACCCATTGTCTTCAAAAACCTTGTGAGGATCCTTTTTCAAAAACGAGCTTATTTGGGGTTAGAGTTTCC  
GACTCCGGTACTACTCCCACTTCCGTTTTCGGGCGTTTTCCACTACTGGTGTGGTGTGCGGTTAACTGGTACCGGTGC  
CGCAGTTGCCAAAGCGGCTGTAGAGAACACGTACTTTTTCCGGCAC



>NODE\_855\_length\_143\_cov\_2158.097900TTCGTCGTGATCCCCGGCGTAAACACTGGTAGTTTGAAAGTG  
ATGGCTTATGACTTACACTACGACAACATATTCGACAATTGTGTGGTAAAGTCATTTTCGAGACACTGACACTGGATT  
CACCGTTATGAAAGAGTATTCGACGAATTCAGCGTTTATACTAAGTC  
>NODE\_858\_length\_123\_cov\_114.227646TGCGCGCACATTTGCAGCATTTTCCACCTAGGAGGATGAGT  
GGCGACTTCGATTACAAGGTAAGGGAGCCGGCGATGAGTCACCCATTCTACGGTAATACCAGACAAGCTGGCACCAT  
AGGCTAGGGCTGGGAGTGTGATGCGG  
>NODE\_864\_length\_273\_cov\_2962.256348AACGCATGTGTTACAGCTTCGAATACGGCGTCATAACTAGA  
CGCGTCGAAACTGCAGAAGTTGAGGTCCTCGCCTCCATCAAGACCCCTCTTACCATCCCCCTCAATGCACTAACTG  
TGCCGCTCACAACGGCGTATTGCCGCGCAAACCCCTTAACTGGTGGGGGGACGAACGTCCGCTTCACCCCTTTGGT  
GAGCTTACCAATATATATGAGAGCGGCACACTTGGGTGTGAAGTCAGATAACTTCATATTTCCCTGTGAACTTCATGG  
CTGTGAGGAAAGCTCCTACATCG  
>NODE\_867\_length\_1786\_cov\_137.621506AGAGGTATTTCCATGGGTTTGCCTTGGTATCGTGTTCATAC  
TGTCGTATTGAATGATCCCGGTCGATTGATTTCTGTCCATATAATGCATACAGCTCTGGTTGCTGGTTGGCCGGTT  
CGATGGCTCTATATGAATTAGCAGTTTTTGTATCCCTCTGACCCCGTTCTTGATCCAATGTGGAGACAAGGTATGTT  
GTTATACCTTTTCATGACTCGTTTTAGGAATAACTAATTCATGGGGCGGTTGGAGTATCACAGGCGGGACTATAACGAA  
TCCGGGTATTTGGAGTTACGAAGGTGTGGCCGGGCGACATATTGTGTTTTCTGGCTTCTTTCAGCATATCT  
GGCATTGGGTGTATTGGGATCTAGAAATATTTTGTGATGAACGTACAGGAAAACCTTCTTTGGATTGGCCAAAGATC  
TTTGGAAATTCATTTATTTCTATCAGGGTTGGCTTGTCTTTGGTTTTGGCGCCTTTTCATGTAAACAGGCTTGTATGGTCC  
TGGAATCTGGGTGTCCGACCCTTATGGACTAACTGGAAAAGTACAATCTGTAAAATCCAGCGTGGGGCGTGGAAGGTT  
TTGATCCTTTTGTTCGGGAGGAATAGCCTCTCATCATATTGCAGCAGGTACATTGGGCATATTAGCAGGCCATTC  
CATCTTAGTGTCCGTCGCTCAACGTCTATACAAAGGATTACGTATGGGAAAATTTGAAACCGTCTGTCCAGTAG  
TATTGCTGTCTTTTTTTGACGCTTTTTGTTGTTGCTGGAATATGTGGTATGGTTCAGCAACTACCCGATCGAAT  
TATTTGGTCCCCTCGTTATCAATGGGATCAGGGATACTTCCAACAAGAAATATATCGAAGAGTTGGTGTGGGCTA  
GCCGAAAATCAAAGTTTATCAGAAGCTTGGTCTAAAATTCCTGAAAAATTAGCCTTTTATGATTACATCGGCAATAA  
TCCGGCAAAGGGGGATTATTAGAGCGGGTCAATGGACAATGGGGATGGAATAGCTGTTGGATGGTTAGGACACC  
CTATCTTTAGAGATAAAGAAGGTGCTGAACTTTTTGTACGTGCTATGCCTACCTTTTTTTGAAACATTTCCAGTCGTT  
TTGGTAGACGGAGACGGAATTTAGAGCCGATGTTCTTTTTCAGAAAGGGCAGAATCGAAGTATAGTGTGCAACAAGT  
AGGTGTAACCTGTTGAGTTCTATGGCGGCGAACTCAATGGAGTCAGTTATAGTGATCCTGTACTGTGAAAAAATATG  
CTAGACGTGCTCAATTGGGTGAAATTTTTGAATTAGATCGTGCTACTTTGAAATCCGATGGTGTTTTTTCGTAGCAGT  
CCAAGGGGTTGGTTTACTTTTTGGACATGCTTCATTTGCTCTGCTCTTCTTCTTCGGACACATTTGGCATGGTGTAG  
AACCTTATTCAGAGACGTTTTTGTGTTGATTGACCCGGATTTGGATGCTCAAGTGGAAATTTGGAACATTTCAAAAAAC  
TTGGGGATCCAATAACAAGAAGACAAGTAGTCTGATACAACAACAACATTTCTTTGGTATCTTTCCCTCTATTTGCTT  
TTTTGATTTGACATAAGGTACCAGATAAATCGTGATTTCAATCACCCTTTTCTTTGACTCTTTGCTCTTTCTTTA  
TCCGGGAGATGGTCCCAAATAAACAACAAAACAGGTATGGAAGCTATAATTGTAAACCAGATCGAATCTATGGAAGCA  
TTGGTTTATACATTCCTCTTAGTCTCGACTCTAGGGATAAATTTTTTTCGCTATCTTTTTTTCGAGAACCCTG  
>NODE\_882\_length\_79\_cov\_528.848083GAGAACTCTTCCCGGGGCCCCACCAGTGTCAACCAACTTCATT  
TACGTTGCCGTTAAAGATCCACATCCTGGTTCCGGAATTTTAAACCGGATTTCCCTTTTCG  
>NODE\_886\_length\_219\_cov\_53.392693ATAATCGGGCTGGTGACAAAAGATATCTGGAGTAGCACATTTCA  
TGACAGATGAGAGTACTATAAAAGTACAGAAGNNNTTC  
AAGTCTGCTGGATCGATGAAAGTAACAGCTGTTGATTATGTCCCAAAGCCAGAATTGGTGCCGAAATGGAGGGTAT  
CTACACACAGTGGCTTGTATCAACAAAGATGAGAATAGGACAA  
>NODE\_888\_length\_653\_cov\_386.415009CCAACGAGTAGGCAGGCGTGGAGGTCAGTGACGAAGCCTTGG  
GGGTGACCCCGGGTAGAACGGCCTCTAGTGCAGATCTTGGTGGTAGTAGNNNNNNNNNNNNNNNNNNNNNNNNNTTC  
GGGCGCCGCTGGCGAAAGGGAAACCGGTTAATATTCCGGTACCTCGATGTGGATTATCCGCGGCAACGCAACTGAA  
GGTGGAGACGTCGGCGGGGGCCCCGGGAAGAGTTCTCTTTTTCTTCTTAACTGTCTATCACCCCTGAAATCGGTTTGT  
CCGGAGCTAGGGTTAATGGCAGGTAGAGCAGCACACCTTTGTGCTGTCCGGTGCCTCTCGACGACCCTTGAAAAAT  
CCGNACTTGTGCGTCGCCCGGCGAAAGGGAAAGCCGGTTAATTTCCGGCACCTGGATGTGGATNTAGGGTAATTTCCG  
TTTTAATGTAGGCGCTTGCCTCGCCCTCGAAAGGGAAAGCCGGTTAAAATTTCCGGCATCTGGATGTGGATTCTCCG  
CGGCAACGCAACTGAGAGCGGAGACCTTTGGCGGGAGCCCCAAGAAGAGTTCTCTTTTCTTCTTAACTGTCTATCACC  
CTGAAATCGGTTTGTCCGGAGCTAGGGTTCAATGGCCGGAAGAGCGCTGCACTTTTGGCGGCTTTGGTGCCTCCCG  
ACGAGCCTTGAAATCC  
>NODE\_891\_length\_292\_cov\_1888.342407CCTTATTCACTACCGAGAACACCTCATCCCTAGCCTCGTTA  
GGGTTTTAATATCCCCGTTATATAAATATATTTATCCATTCAAATTCGTATCTGTATCTACAGCAGATTTCAAAAAAT  
TTTTCTTCTATCCTTAAATCCGAGAACCCTTATCTTTGTTACTTAGGTTTTTCTCGTGAAAAGCCTTATTGTGCAAA  
ATATTCAGTGTGTCAAGTACTCTGGTATTGCCTTAGCCTCTAAATCTTCTCTCCCGGATTTATACACCAACGAATCT  
AGTACGTTTTGTTGTCCCGGTACCAGATATTCATTCCTTTAC  
>NODE\_903\_length\_345\_cov\_81.759422CTTTCTATGAAATGGCAGAAGGTTTGGATAAGGAGCATGTGT  
TAGTTGGTTATGATAGTTATGAAGAACCAGGACGACGTGTAATTGTTTATGAACGCCACCAAAGCCTACAAGATTGG  
CTCCAACAGCATCAAGTTGTGATTGGAGGGAGACTTAACCAAGAGTTGATATCCATTATATGGGACCTGCTTAAAGC  
CATTCAATACTTGGTGCCATTAAGAAAGCTGGAAGGTGGCTTTGACATGCATAATGTAGTAGTGGTGAAGGGAAGGG  
GAAAATTACAGGATTATGCGATGACTATGAAGGAGAAAATCTTCAGGTTATCCGACTTGAAATTCGAAGTGTGTTA  
GAAAACATTTTCAGCC

>NODE\_911\_length\_161\_cov\_1236.639771CTATAGGGACGTTTCGAGAATAGGACGTTTCGAGTATGCCGTC  
GGGGCGGTTAGGTCACAAAAGACCCACGTTATCACGGGGAGTCGTGTCTCCACAGCAAGGTTGATATTTCTCCTGA  
TGATATGTGGGGTTTAGTTGTGCTGTTATGGCTCAGGCAATTAAGGATAGGGCGAAGAGTATTC  
>NODE\_922\_length\_157\_cov\_806.554138GCCATTATATCCCAATTACGAAGTAAGATATTGTCTTCAGCG  
GAAGTATGCTGTAGCCATAATCCCTCATAATACCATTGATTACACTGCAGGTTTTTGGGGTCTTCGAGGACAACAA  
TGTTCCATGAGCACCAAGTACTCCCCAACGCACGTGGCTGCTGCTTCGAATGTCATACC  
>NODE\_936\_length\_277\_cov\_139.064987GGACTTTGCCGACATAGCGACTCTGGCCAACGCCATGGCAAC  
CCCTGACTACGAGTGGGACAGGGCTTCTAGGAACGAGTTTCGCCGAGCGACTGGTGAACCACCAGCCAAACGCCGCTA  
GGGCTTCCGCCAAAGCTCCGCTGAGTTGCTGCCATCGACGCACTCGTAGACTCTGGGTACGCTCTCTCCAACAT  
GCCGTCAATGAGGTCAATCTTGTCCGTGCGCTAAGGGCAGTGCGGTTTCAGGGACTTCACGCGACAGTTTTCCAGGTC  
CCGCCATGCCCTTGCCGTTTCGCATCG  
>NODE\_937\_length\_192\_cov\_70.651039TGAACCATTTCACTCTTAAATGAAGAAGGAAAACAGGCGGTGAG  
AACATGTTGCGACTGGAACCTTTACGTTTTGCTGCAAAACACAATTTTTATAAAGCTTAGCCAGTGGATTTAGAATA  
GAGGATTCAATGTAGACCGTCAAAAAATAACCGCATCTAAACCAGTAGGGGGATCGTATGAATAATGAGCAGCCCA  
TGAATTTCTTTTGCTAG  
>NODE\_944\_length\_101\_cov\_2164.970215CCCACACTGCGTATATCTTCTTCCCATCATCCGTCTCTATG  
TAATCAAATTTCTTAAATTCAGTAGAAGACGATGTGAAGAACGTACACGCTATCTGCATTACGTACAAGAATAATTT  
AGTCT  
>NODE\_948\_length\_114\_cov\_895.298218ATGTGCTCAGCCTCCATGTTGCATCTGGTATACATGTTGTCA  
TACGTGTACCATGTGCTTGGGTATTGCTGCGACGCAAGCGACAAGAGCCCTAGTTTCAGAGGCTGTAAAAACATCTCC  
ATCTATGACAATGCGGC  
>NODE\_967\_length\_1941\_cov\_125.461617TTTTCCACGCTTGAAAAAAAAAAAAACCTGGGGCGTATCGCTCA  
AATCATTGGTCCGGTCTGGATGTAGCTTTTTCCCCGGCAAGATGCCTAATATTTACAATGCGCTGGTAGTTAAGG  
GTCGAGATACTGTCCGTCAACAAATTAATGTGACTTGTGAGGTACAGCAATTATTAGGAAATAATCGAGTTAGAGCT  
GTGGCTATGAGTGTACAGATGGTCTAATGAGAGGAATGGAAGTGAATTGACACGGGAGCTCCTCTAAGTGTCCAGT  
CGGTGGAGCGACTCTCGGACGAATTTTCAACGTGCTTGGAGAGCCTGTTGATAATTTAGGTCCGTGTAGATACTCGCA  
CAACATCTCCTATTCATAGATCTGCGCCCGCTTTTATACAGTTAGATACAAAATTTATCCATTTTTGAAACAGGAATT  
AAAGTAGTGGATCTTTTAGCTCCTTATCGCCGTGGAGGAAAAATTTGGACTATTCGGGGGAGCTGGAGTGGGTAAAAC  
AGTACTCATTATGGAATTGATCAACAACATTGCCAAAGCTCATGGGGGTGTATCTGTATTTGGCGGAGTAGGTGAAC  
GTACTCGTGAAGGAAATGATCTTTACATGGAATGAAAGAATCTGGAGTGATTAATGAAAAAATATTTTCAGAAATCA  
AAAGTGGCTCTAGTCTACGGTCAGATGAATGAACCGCCGGGAGCTCGTATGAGAGTTGGTTTACTGCCCTAACTAT  
GGCGGAATATTTCCGAGATGTTAATGAACAAGACGTACTTCTATTTATCGACAATATCTTCCGTTTCGTCCAAGCAG  
GATCCGAAGTATCTGCCTTATTGGGCAGAATGCCTTCCGCTGTGGGTATCAACCTACCCTTAGTACCAGAAATGGGT  
ACTTTACAAGAAAGAATTACTTCTACCAAGGAGGGTCCATAACTTCTATTCAAGCAGTTATGTACTCGCGGACGA  
TTTGACCGACCCTGCCCTGCCACAACATTTGCACATTTAGATGCTACTACCGTACTATCAAGAGGATTAGCTGCCN  
NNNNNCCAGTAGATCCTTTAGATTCAACGTCAACTATGCTCCAACCTCGTATCGTTGGTGAGGAACATTATGAAACT  
GCGCAAAGAGTTAAGCAAACCTTTACAACGTTACAAGAATTTCAAGACATTATAGCTATCCTTGGGTTGGACGAATT  
ATCCGAAGAGGATCGTTTAAACCGTAGCAAGAGCACGAAAAATTTAGCGTTTTCTTATCACAACCTTTTTTCGTAGCAG  
AAGTATTTACCGTTTCTCCAGGGAAATATGTTGGTCTAGCAGAAACAATTAGAGGGTTTCAATTGATCCTTTCCGGA  
GAATTAGATGGTCTTCCCGAACAGGCCTTTTATTTGGTAGGTAACATCGATGAAGCTACCGCGAAGGCTATGAACTT  
AGAAATGGAGAGCAAATTTGAAGAAATGACCTTAAATCTTTGTGTACTGACCCCTAATCGAATTGTTTGGGATTCAAA  
AGTAAAAGAAATCATTTTATCTACTAATAGTGGACAAATTTGGCATATTACCAAATCACGCCCTATTGCCACAGCTG  
TAGATATAGGTATTTTGAAGAATACGCCTTAAACGACCAATGGTTAACGATGGCTCTGATGGGCGGTTTTGCTAGAATA  
GGCAATAATGAGATCACTGTTTTAGTAAATGATGCGGAGAAAGGGTAGTGACATTTAGCCACAAGAGCTCAGCAAAC  
TCTTGAAATAGCGGAAGCTAACTTGGAGAAAGCTGAAGGCAAGAGACAAAATAATTGAGGCAAAATCTAGCTCTTAGAC  
GAGCTAGGACACGAGTAGAGGCTATCAATGTGACTTCGTAACCTAGTCATGTGCCGAAAAATCAAAAAGAGTTCTGT  
TTATACACCCTATTTTGAATTCTGCCGATTGAATACAATCAAATACAATCAAATGGAATGGGATTTCTGATGCAAC  
>NODE\_972\_length\_151\_cov\_173.357620ATGGGGTGTCTATTCTATACAACATATGATTTTCTATCATTTTT  
GCTTATGGGGTGAGCATGTCTTTGGAAGGAGTGCACGCTGAATTCATACCATATCCAGTTAGTATCTCTCTGACTG  
TGTCTGGGATGAAATAATCAAACAAGCCGTCCAATCCAAGTTTCGGTCTCCAATA  
>NODE\_982\_length\_114\_cov\_366.350891AGCCGCTGGTAGTGAATATCACCTTGAATTTCTATTGCAGTG  
AGTGATACCTCCCCGTGAGACATGAGGTGAACCAATTGTTCTTCGATATGGCATGATCATCAACGGTCAAGATGTT  
AATGTATTCCTGATATC  
>NODE\_989\_length\_129\_cov\_2242.527100CGCTTATATTCTTACCCTGAGAATAAAGCAAGAAGAGCTGA  
ATCATAGCCACAATAAAGTTTCCCATATCCTGTGGCGATTGATTTTTCCACCTTTTTGAAAAGAACCAACCTACGCAACC  
TAGCATTATTTGTAGGGCTTCCGTAATCGGTA  
>NODE\_993\_length\_194\_cov\_104.953606CGATCGAACACGCAATGGCCGCTATGCTCAGCTGTGACATAG  
ACACAGGAACATCAGGCTTGGACTGGGGCTCAGGGTANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGGTGTGCGGAATG  
GTGTTGGACAAGGACGGTGAATAGTGATGCAGACAATACGGGGCCAAGCTCATCCGTGTATATTGGGTTTATCGGA  
AGCAGAAATATGGGGGCGAA  
>NODE\_995\_length\_212\_cov\_76.716980AGAGGCGAAAGGCCATGGGGAGATAACCGTCTGTGATCCATGGA  
TCTCCGATCGGGAAACCGTATCCAAGCTCCGTGGCTAGTCTGCGCTCTTTGGACTTTTTCAAACTTNNNNNNNNNNNNNN





GGATCGACTATGGCAGGAAGCCGAGCTCGAGGCCCTGGGGGATGATAAATTGACGGGCTTGCCCGAATGGTTGAGCT  
GGATAACAGCAAGGATGTACGGTGAGAGTGTGTTGAAAATGACGGGCATCAATTGGTTAGGGGATAAGAGTTTTGGC  
ACCACGATTCTGGAAGATGACGATCGCATTGAGCGTCAGGATACAGAAGAGGGGTGCTTTAAAGGGTTTCAGTATTT  
GGG  
>NODE\_1190\_length\_425\_cov\_484.362366CGCTCTTCCGATCTCGGCAACGGATATCTAGGCTCTCGCAT  
CGATGAAGAACGTAGCAAAATGCGATACTTGGTGTGAATTGCAGAATCCCGTGAATCATCGAGTCTTTGAACGCAAG  
TTGCGCCCCGAAGCCATTAGGCCGAGGGCACGCCTGCCCTGGGCGTCACGCACCCGTCGCCCCCCACACCTCCCTCC  
CCCCGGACGGGAGGCGGAGAGGGGGCGGACATTGGCCTCCCGTGGGCGCCCCAGCCCGGTTGGCCGAAAATCGGT  
CCCCGCGGACGTACGCCACGACGAGCGGTGGATTTGCGACGGCTCGGCGTCGCGCGCTAACGTCGCTCAGGGGC  
CCCCACGAGAACGGCCCCGAGACCCTCGATCGCGACCCAGGTCAGGCGGGACCACCCGCTGAGTTTAAAGCATATCA  
ATAAGCGGAGGAAAAAGAAAC  
>NODE\_1191\_length\_370\_cov\_145.343246AGAGGGGAAGCAAAAAATATCTCGGGTATTTAAGCGATCAA  
TGCTCTAGTAGACACATAATGCTAACTCAGGCCGGGCTAGCCGCGGTTGATGAAGCGGAAGCACTACGAACGAATTA  
TCTAGTTAGTAGTAACGGTTTTGTTTTCAAATTGCGGGATCGCTAGGATGTTGCTCTTGACACTTTGTAACGGAGCAC  
TGTGAAAATGTTGCGATGTTCAACCAGTCTCGAACATTGAAACGCGTTACAATTAGCACTACTAGTTTGTGCGTGT  
TGTAGCGGTGACAATCGCTCTTTGTTGTAGGAGACGATGAGGCACTTAGAAAAACCCATCAGGGTAGCGATACATT  
ATTGCTTAGTGCGAAGCGATGTTTGCACGGGTGGGATGTATT  
>NODE\_1198\_length\_229\_cov\_359.231445ATCGTTTTTTTTACTTAAATCAATGAAGCGGAACGGTCTTCA  
TCGACCATTTTCTTGCAATAAAGTCTTTCGCGGGCTGATCCGGGTTGATGACATTGTCAGGTGGGGAGTTTGGCTGG  
GGCGGCACATCTGTTAAACCATAACGCAGGTGTCCTAAGGGGGACTCATGGAGAACAGAAAATCTCCAGTAGAGCAA  
AGGGCAAAAGTCCCCTTGATTTTGTATTTTTCAGTGTGAATACAAACCATGAAAGTGT  
>NODE\_1202\_length\_1899\_cov\_353.832031CAGGCTTAGCAGGATATCGGCTACCTGGGACCTGACCGCC  
TTTTCTCCACCAAGTACAGAATATGGCTAACCACACATAATGCATTAGGTGGCAAGGGTATAATGCACGACCCCTT  
CTTTTCCCGGTTCCATTTTCCAAAAGTACTTGGACCAGGCATTAGGAACATAATCTGCCAGTGTGCTGCGCCAAT  
TTCACGAATTCAGGGAATTCAAATAGCGCAGACTTATTCAAGCGGACAGCCTTGAGCACGCGTACAGTCAATCCAGC  
CATCTCTTCAGTCACTTCTCCAGCTCCTCCAGCCGTTCTCAGGTACCTTCAAGAACACATCAGTCTTAGGAGAAC  
CAGTGACAGCACCGGCTCTCACCCACTTCCGCGGTTAGCTAGAAAATGTTCAAAGGTATCCACTTTCATCTCCGAGC  
CGTTGAGCACTGAGCCCATCATTGATGCTTGA AAAATCAGCTTGTTTCAAGATGGCTGCCTCAAAAACGCTTGTCAA  
AGCTTCCCTCCGACCAGCGCCATTCTCAAAGAATCTCTTGGGACTCACAGGAGCTACTCGGTGCGATATCTCGACAT  
CGATATCGTTAGTACCGATAATGCGACCGGCCAGATTGTCAAAGTACGTGAAATATGTTAGATCCTTGGCTTGCAAA  
GGCCGACCCAGGAAGTTAATACCTATCCGTACATTATTATGGATCAATTTGGCAACCACAATGAAATCTGAATCGGT  
GAGTGCCCCAAGCGGAGACGATTTAAGAGCTCACTGTAATGAGGAATAACATAAGCTCCACATCACCGTAAGCCA  
AGTGGCAACAATACCGTACATACCGTGGCCCAACTGACAGATAAACCCTGTTTGAACCATTCGTACTGACCGGTTGC  
TTAATAGCCTCCACTCCAGTCATGCGGACCCAGTCTGATATGTTAGCTACAGCGAAGCCAGCATGTCCAGCCGTTCC  
AGGGAAGAGCTCGCGCCATCGATCCTCATCCCAATGTTCAAACGCAATGGGAGGAGTAGCTTGTTCGCGCTCGCGA  
TAGCGTGAACCTACGCTGGCCAGACCGGATGATTATGTGCTGATGAGCCTCACTGGGCGCGGCATCAGCCAAGCTG  
TGTTTTAAGTACCGGCGACAAAACAGCACGGAACCAACCCACCGTAGCCGCATCGCGTAGTGCGCTATACGCCAACAG  
CATCAGCTCTTTCTCAGGCGCACGCGGTCCAGACGGGACCAAAAATATGTAGCCGAGCATTAGGCGGAGTAACCC  
CTCGAAGCGAGAACAATGCTCCGACACGGGCCGTCGAGCGCAGAGTACCCATTACAAGAGTACAGTTCAAGACCATGC  
CTACGAGAATACGTAACATTCGCGGAACGCTGGTACAGGCCACTCGCAGCACTGCCTCACGGTTGGCTGCAGACAG  
GGCTTGCCAACGCGGTGTGGCCGTGATCGCTTCGTCGGAACGCTCAGTACAGCCAAGATACGTAGCCCAAGAGCGT  
CCGCCAACTCAGCAGTGTGGCAGTACACCCTCGCGGTTAGCATCCGGCTCTACTTGCAGAAAAATCGCTTCGCC  
CGGAGCAACAACAGCTGGTTATGCACATGCATATGAGCAGCTGAGTTACCCAACAGACCATCTTCGCGCGAGCCTAG  
CATATCGTCCAATTCGCTGTCCAATGATAAATCGGACCCATCAGCGACAGATCGGAAGAGCGGTTCCANACCGATGTG  
CAAGGGTTGATTTTCTGAGCACATCGGAAGGACCACAGCCTGCAGCGCTGCCACCCTGGTAACTTTTGCAGATGA  
GGAGCGGAGTCGGCAAGAAGGCGAGTACCCTCTGAACCGACAGGTGTGTGATGTAAGACGCTCCAAAACGTGGCAGC  
CATGGCGTTGACTTGTCTGGCAAAGCTCAGA  
>NODE\_1204\_length\_580\_cov\_522.118958AGTGATTGCTTCCGTGCCGCTCAGGCGGTTAGCAGTGAGGC  
TGGGCTTCGTGAGGATGTCGATCCCCGTCCCTCCCTACCCCGGATCGCCTAGTGGCTCCTGGGCTGGCAGCAGCGCTC  
TGCTGTCACTGTTGCCCTGGCGACTCAGTTTCTACCGCTGGGTCGTCAGGACGTCATCGCTCCAGGCGCTCATCG  
AATGAGCGTCTGGTAGTGTGAGTACCGCACCGTCAACGGTCTCACTTACGGATACGTCGTGTCGACCGCTCGCGAG  
GCGAATCTGTGAAATGGGTGTTGGCCCGCAACAGCGACGAAAAGGTCTCACCTCCCGACCTTCTCGAGGGCGTTA  
GCACCAGTGTTCCTCGGACTCGCGCCGAGGCACGCTTTGTGCCCCGATCTCGCGGCGAAAAGGGTGGTTGGTGTCTT  
GGTTTGGCAATAGGTGGGCAAAAAGTCGTTCTGTGACAAAACGTGGCAACTTGGTTGTTATCAGACCAAAAGTAGAACAG  
ACACTGATTTTAAATCAGTGGTCTGCAGACGGGCGAGCAAAAGGACTTCAAGTTATGGTGGTGTTTTTGAGTGGATCATTA  
GTGGTTCTCTGGTAAGCGATTG  
>NODE\_1205\_length\_105\_cov\_126.371429GAGCATGTGGTTTAAATTCGATGATACGCGAGGAACCTTACC  
TAGGCTAGAATGCGCGTGACCGGCTCAGAGATGAGCCTTTCCTTCGGGACACAAAGCAAGGTGCTGCATGGCCGTCG  
TCAGCTCGT  
>NODE\_1206\_length\_90\_cov\_54.455555TTTACACTGTGTTGGCCCCACCATCGTTGCTGTTTAAAGTGTG  
TAACGGCCACTTGTCTTTACTTCCAGGCTGAGACGAAGCCTTACTGCTCACCAGTCTTCAGTTCTCATT  
>NODE\_1207\_length\_183\_cov\_712.863403CCAACCTGGGGGGGGGCGAAAAAATCAAGCTGTCAAGAGC  
AGCGGCGCGCAGATCAGCCAGAGTCATGAGGGGGTCCCCCGAGCCACGAGTGAAAACGTGAGCGAAGGCAACGGCCA

CCGGGAGACTACGAGTTGCGGTAGTCTGGATGACAGCATCATTTCGCTGGGTTGACGAAAGCCATACGCGTGGAGTCC  
TAGGAACACC

>NODE\_1219\_length\_3222\_cov\_246.119186GGAGAATCGAAACCACTGTGTGGTAGGGGCGGATTATTTA  
CATACTTACAACATACATAGCAAATATACATCAAAGATGACANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
NNAGCTGACCAAACCTCACGTTGTCGAGATAGCAAACAATATGCTCG  
ATTGCGTGCAGCTATGTGGTTATGACGTTATGGAATTAGAGCAAATGAGCATAACGAGTAATTAACAACATGACTATA  
GGTTGTGGGCCAATGTGGTTCTCATATGTAGGGCCAATGTCGTCCTGATCTTAGCTATAAAGCAAGTTGGCTTACA  
CGTCGAGTGTGGTCCGAGTCGACCTAGGAAATAATATGATACGCGCAGTTTTCGAGATGGCCATGTAACCCTGAG  
GAGAATCGAAACCCCGTGGGGTAGGTGAATTGGCCATATATTTATAGTATCTACACAATATACATAACCACAAATGG  
AAAACAAAATCATGANCTTGGCCATATTTATATTATTTATAGTATTTATACAGTATACATAACCTACAATGGAAAACA  
AAATCATGAGCTTGTTCCTTGGGCGGAGGTTCTCCAAAACAACAACGTCGAAAACACCTGAGCAACAAGAACTATGAC  
AAAATAGGAGTACCAAAAACAACCCACGCCGTAGGGAAAATACAACATCAAAGTACAAGAGTAAATCATTGCGGTT  
CTCTAAGATGCCATCAAGNACTTGTTCCTGGGCGGAAGCACTTCAGACAACAACGTCGAAAATATTTGAGCAACAG  
AAACTATGACAAAATCGGAGTACCCAAACAGAGTCAACGCCGAAAGGAAACGACAACATCAAAGCATAAGAGTAAAT  
CACTGCGGTTCTCTAAAATGCCATCAAGTTCAGCAGCCTCAGTGCCGTATACTTCGCTAATGAACTTCAAAGGCGTC  
GTTGTACCCATGGGTCAAGTAAAGACAACCCCTAGCTAAAGAAGAAGGGTGGGTAGATGTGGACGCGTTGTTACAACC  
GAGAGTGTAGCTGATTTGAAGGAAGAATTCTAGACTTGATAGCCTCAGGTACGTCAGTAGAGGAGGCATCAGTCG  
TGCTTGCCAAAGAGATATCACCTGGGCTCAAGTTACTGAACCCATCGGAAAACCGTTGTGCTGATGTCGCCAACATTC  
GAGCTCTTAGACGCTTGGGGGTAGAATGTGTGCGGGCGATTGTGCTCAAACCAGAGACAGTGCTTGCAGCTAATTC  
GTCGAGAGACGCCGGCGAACGCCTCTTAATTGAAAAAACAATGGAGGAGGTTTTGGCAAAGGATGGTCAGGAGTTCA  
TTGATGTGTTACCTGATAGTCTGGATGATGTGCGATGGTACATATACTGTCTGTGCAACCGCTATCACTNCAGTC  
TGGATGACGTGCGATGGTATATATACTGTTTGTGCAACCGCTATCANAAGCCTCGAGAGTACGGGATAGATGACCC  
TTCAGTGTACACACTACCTACGGAACATCAAAGTCAAAAACATACATGAAGTGGTTGCAGCCTATGATCAGGGCC  
TAGTCTCACGCGAGTTAGTGAACATCAAGTGTTCGCAATGGCCTCAAAGCTATAGAGGATTTGGTTTACCTGG  
AACGATTGGGCGAGAACCATGAGCTATGCCGCCATACACGGGGTGTATCTGCATTTGATGACCAGGACTGGGTAGG  
ATGGCCAGTTACACTAAAGAAAATATCTGAGGGAATACCGCTGGAGGAGCATGAAGACATACAGCTACTAGTGGAAAG  
CCCACAAAGGTGAACACGAGAGATTTCGTCTAGCACTAATACTGCATTTGAAAACGTTAGGCTTACCATCAGCAATT  
GCAAAAAGATTTTTCCCTTTGTAATACGCCGGGTGCACTGGGGGGGTGTGTINTATGGCTAAGCCACGTGAA  
TACGGAATAGATGACCCTTCAATTGCTACACACTACATACGGGACGTCGAAAAGTCAAAAACATACATGAAGTGGTCCG  
AGCCTACGATCAAGGCCTGGTCTCACGTGAATTAGTGAACATCAAGTGTTCGCAATGGCCTCAAAGCTACAGAG  
GATTTGGGTTACCTGGAATGATTGGGCAAGAATATGAGCTATGCTGCCATACGCGGGGTACATATGCATTCGAT  
GACCAGGACTGGTGGGTCAGTTACACTAAAGAGAAATATCTGAGGGAATACCGCTGGAAGAGCACGAGGACAT  
CCAGTATTAGTTGAAGCCCAAGGCGAACAGAGAGATTCGTGCTGGCATACTGCATTTGGAAGTGGTAAAGCTAG  
GCCTACCATCAGCAATTGCGAAAAGATCTTTCCCTTTGTAATCACTGCACTCGGGTGCATTTGGGAGGTTTGTGAC  
AAAATAAGAGAGGGTGTATATCGAGCAACAGCTTTGCGGGGACACAGTTGACCATGGAAGAACGTGAAATGATACT  
AAGCATGAGACTGCTAGCTGCTGGCTCGGTGGACCAGTTAAGGACAATCATGGGATCGCAAGGGGGATCATACCCCC  
GACAAATCCCTGATCAAAGCAGGAGGACCACATCATCGCCTCCTTGGCTAAAAGTCAAGTATGCTGTAAGTGACATA  
GACAAGAAAGAAGCCATGTTTGGATCATTAAATCAAAGGGACCGTAGTCCGTGAGATCAAATCGATTAATTTGGAACGG  
GAAGTTATTGCTTCGAGAGCAGCTGATAAAGGCGATAGGTATGGAGCTACTGAGCCGATGGGAGGACGAACCAGGTG  
TTGTCCAAGGGGTGAGCCGTCTGCTGAGGAGCATTGCAGTCAGATGGTATAAAGCATCAATAATCCGTGATGAATNN  
NNNTCGGCTGGAGATGAACCTAGCGCATGCCATCAGCTCAATGTTAAGCTG  
TGACATAGATACTGGAACGTCAGGGGCAGATTGGAGTATTAGGATCTTCGACTCACTGAAGAGTTTTCATGGTATGTG  
GCTTAGTGATAGGGAATGATGCGCACATCGTCATGCAGGATTCTCCAAACGGGAATCGGCCCTGCATACCTAGGCATC  
AGTGAGGCGGAGATCTGGAGTAGAATAGCCAAGCGGAATGTTCCAAGGAGCGCCATGGGCTATTTTTGCAACAGGAGT  
TGGGAATGTACAAAACATAATGGAAATAGGAGTCTGGTCAAACAGCA

>NODE\_1221\_length\_104\_cov\_195.692307TAAGTCTGCTGCAAAGGATTCATCCCAGCGCGCTGTCAAAT  
TGCTATCCGCCAATTACCTCTTCCGACTTCTACGAAAGAGGAGTTATGGCCTGCTAAGGTCCGGGGCCGAAGCCTAT  
TCGTATCC

>NODE\_1236\_length\_1673\_cov\_538.625793ATGGAAGAACGTGAAATGATACTAAGCATGAGACTACTTG  
CTGCGGGCTCCGTAGACCAATTAAGAACAATCATGGGGTCGCAAGGGGGCTCATACCCAGGCAAATCCCTGACCAA  
AAGCAGGAGGATCACATCATTGCCCTCCTTAGCCAAAGTCAAGTACGCTGTAAGCGACATCGACAAAAAAGAAGCCAT  
GTTTGAGTCACTAATCAAAGGGACCGTCGTCGCTGAGATCAAATCGATTAATTGGAATGGGAAGTTGCTGCTTCGAG  
AGCAGCTAATAAAAGCTATAGGTATGGAACACTAAGCCGATGGGAGGATGAGCCTGGTGTGGTGAAGGGGTAAGC  
CGTCTGCTAAGGAGCATCGCCGTTAGGTGGTATAAAGCCTCAATAATCCGTGACGAATGGTCAGATCTGGCGTGCCG  
AATACTAGAAACCAACCGGCTGGAATGAACTTAGCCCATGCCATCAGCTCGATGCTAAGCTGCGACATAGATACGG  
GAACGTCAGGGGCAGATTGGAGTATCAGGATTTTTGACTCCCTAAAGAGTTTTCATGGTGTGCGGCCCTAGTGATAGGG  
AATGATGCACACATCGTCATGCAGGACTCGCCAATGGGAACCGACCATGTATACTAGGTATCAGTGAAGCTGAGAT  
CTGGAGTAGGATAGCCAAGCGAAACGTTCCAAGAAGTGTATGGGCTACTTTGCAACAGGAGTTGGGAATGTACAAA  
AATAATGGAATAGGAGTCTGGTCAAACAGCAAAACGGTGTATGATAATGGAGATGATCAACGCAAAGAGCTGGATG  
CCTAAAGCCAGTGACAGAATGTTACTAAGTTGTATAGTAAGATGGGGGGAGCATTTCACAAAAGCCGGAAGAACTATA  
CATGTTTTCAAAGTTGGCAAACGTTACACAGAAAAAATATNCGACAGGATGTTACTGAGCTGTATAGTAAGATGGG  
GAGAGCATTTCACAAAACCGAAGAAGCTATATATGTTTTCAAAGTTAGCGAACTGTTACACAGAAAAAATATTCGGA  
AGGCAATACAAGACGGTGAAGGATAGACTAAGCAGCTTAAGCTTGATAGCAGCATCAGACGGGGGGCTAGAGTGTAC

GAGTGCACCTGTACGCGGAAAACCTTGATACAAAAGACAGGATGTGGACGGGGCACGGCAGTNGAGACAATACAAGAC  
GGTAAAGGATAGACTGAGTAGCTTGAGCTTGATAGCAGCATCGGACGGGGGGCTAGAATGTACGAGTGCACCTGTGC  
GCGGGAAGCTTGATACAAAGGATACGATGTGGACGGGGCACGGCAGTGTCACTGCGACTACGAAAGACTGTTAAAAAC  
AGGGCTTTGCGGTCAATAGACAAGATGTTAACTGACTATGAAGAGGATAATGGAAACAGGTGAGAAAAAGAATAAGNAG  
GGCACGGCAGGGTTACGGCAACAACAAAACTGTCAAAAACAGAGCCTTACGATCAATAGACAAAATGCTGACTGAC  
TATGAAGAAGATAATGGGACAGGTGAGAAAAGAATAAGTANAGGATCAGTACCTATGGAATCCATAGGGCCGGCTGA  
CGGCTGTTCTGCTACTACNCAGCTGTTCTGTTACTACAAAGAGAAAAGAAATCGAGAATAAATCTACATTGCGAATT  
AATCGATAGGTTTTTCGAAATAAATGAGTTATTTTTTAA

>NODE\_1243\_length\_88\_cov\_102.943184CACATATATAAGTCTAGCTCCGATCTCATCTATCACCCAACG  
ATGGACGGCTTGCGCGTAAGAGATTTACCATTTCGTGTTTCTCGCGGAAAAGGGGATCTTCATCAAGGG  
>NODE\_1248\_length\_87\_cov\_712.402283CGAGGCAGAATAAAAATTGTCCCATCCCACAAAACCCAGGCATA  
CAAGGGTAACGAGTGACGGTCGTCCGGCCACGCAACCTTATGTCTTAAATGGCAACGAGGAATCGCA  
>NODE\_1249\_length\_1087\_cov\_125.435143GAAAAGTAAGAGGGGGTAAAGTGCTCTTGTGTTGATTGTA  
CTCTAAATGTAAAGTTTCTTCTTCCGCATGTAGAAAAGGAATAAATAAATCAATCAAATTACGGGAGGCTTCATGAA  
GTGCTTCTTTAGGAGTTAAACTCCCATTGTCCATATTTGAGAAAAGTATCTCTTGTCTTTCATTCCCATTCCCA  
TAAGAATGAATACTATGATTCGCATTTGCAACAGGCATGAATACAGCATCTATAGGATAACTTCCGTCTTGAAAGTT  
ATTTGGCCTTTTTATACTATATCCGCGATTCTCTCGATTTGTAATCCAATACACAAATCAATCGGTTCCATCAGTC  
TAGCTATATGTTGTGATTATCAACGATTTCCACAGAAGGCGGTAATAATGATGTCTTGAGCAGTTACATATCCAGGA  
CCCCTGGCACAATAAACGCGCCGCGAGTTCCATACAGATTACTTCTCAATACAATTTCTTTCAAATTCATTAAAAAT  
TTCATGTACTGATTCTTGAATACCTACTATGGTAGAATATTCATGGGGTATTTTCTCAAATTTTGCACGTGTGATAC  
ATGTTCTTCTATTTCTCCAAGCAAAGCTCTTCGCATCGCAATGCCTATTGTGTGCGCTTGACCTTTAATAAGTGGGA  
GACAGAATAAAGCGTCCATAACAAGACGCTTATTGTCTACTCTTGATTCAACACACTTCCACTGCAGTGTCCGAGT  
AGATACTGTTACTTTCTCTCGAACCATAGTAATATTACTTGATCAGATCATTGAATCATTATTTCTCTTGAATCT  
CTTCAATGTTTATTTCTACACAGTTTATTCTTATAACAGCTCTTTTTTTTAGGAGGTCTACAGCCGTTATGTGGCATAG  
GAGTTACATCCCGTACGAAACTTAATAGTATACCCTTCTACGAATAGCTCGTAATGCTGCATCTCTTCCGAGACCA  
GGACCCTTTATCATGACTTCTGCTCGTTGCATACCTTGATCCACTACTGTACGAATAGCATTTCTGCTGCGGTTTTG  
AGCAGCAAATGGCGTCCCTCTTCTTGTACCCCTGAATCCACAAGTACCGGCCGAGGACCAAGAAACCA  
>NODE\_1254\_length\_248\_cov\_119.483871CAATTGGAACGAATCGTCTACTTAATACAGGTGGATATAG  
TAAATGGGTTAGCACGGACCACTGTTAATTCAAAATACGCAAAACGGACGTTAGTAGTCTACCCCAGAACTGGACG  
CGTATCTATAAGATAACGAAGATAGTGTCCGTGGATCAGAACCTCTACCCTGGTTGTTTCTCAAGCTCGAAAACCTGAA  
CGTAATGCTTATAAGGTCACTGACGGTTTCCCCAGTGGCAGCTTTCTTTAGGAACGTTTATGAAACCTTTGAG  
>NODE\_1273\_length\_662\_cov\_182.561935TGCGACCGTACTCCCAGGCGGAGTGTTCACGCGTTAGCT  
GGGCCCCGTGATCCGCGTAGACCAAGGGCGAAGACTCATCGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
NNNNNNNNNNCCAGCGTCCGTAGGGACCCAGAGAGCTGCCTTCGCTTTTGGCGTTCCTTCGTAGATCTCCGGATTTT  
ACCCCTACACACGAAATTCACCTCTCCTCTGTCTCACTCAAGTGAATGGTTTTCGAGAGCATTTCCGCCACTTTTTGG  
CGACTTTCACTTTCAACCCGATTACCCGCTACGTGCCCTTACGCCCAGTCATTCCGAAGAACAACCTTGCCCCCCC  
GTCTTACCAGCGGCTGCTGGCACGGAGTTAGCCGGGGCTTCTTCTCGAGTCTGTGCATGATCGCGCACTCGACGAAA  
GAGCTTTACAAGCGGCATTGCCCTTCTTCACTCACGCGATATTGCTGGATCGGGCTTTCCGCCATTGTCCAAGATTC  
CCCCTGCTGCCCCCGTGGGAGTCCGGGCCGTGTCTCAGTCCAGTGTGGCCGATCATCTAAAAGACCAGCTAAG  
CATCATTGGCTTGGTCCAGCTTTACCTGACCAACTACCTAATACTACGAGGCTCATCAAACAGCGCTTTTTTAGCTT  
TCTTCAGGATTTGGCCGAACCTGTTCCG

>NODE\_1280\_length\_242\_cov\_73.119835GAGCGGTGCGCAGAAAATATTGTGGCACTCCGTGGTCCAGCAA  
AGAATCAGGGCGCATGAGTCTAACATTACGGAATCACATCGTGCACGAAACAAAATCTGTTTCAGATGAGATCAGCG  
AGTACGNNNNNNNNNNNNNNNNNNNNNGCTTCTGAACGACAACAAAAGAGTTTGTGCGCCTTGACTGCTTCGCTGA  
ACATATGCATGTGACAGAGAATTCTGCATATCTACTATCAATGCACACAACCCGAGGCTCGTGCAAG  
>NODE\_1301\_length\_134\_cov\_398.865662TAGCAATGATAGTGAAGAGGTAGGCCTGACCTACAAAAGCA  
TCAAAGCCAAGAATTTGATGATCTTGTGGCCACAATGGAACTAATGGATCATTCATTAACTATGTGGGTGATGG  
GATGACAGGACAGTCCAACATAGCCAGATACATCAACG  
>NODE\_1316\_length\_93\_cov\_1792.268799TCATGTACATCCTGCGTGAGTAATTTCTTGTACTACGGCT  
GAGGATGCTGCAAGCACTTGGTTTCATAGTTTACCGGAATCGACATCCTTTTCGTCGCGATCTACATCGTGAC  
>NODE\_1322\_length\_269\_cov\_578.524170GCCTGGCTGGGTTGGTAGGGCTCAGCAGCGGAGCTTGTAA  
CTCAAGCGAGGGCATTAGTGTGGCTGGAGATTGGCCTGGCTGGGTTGGTAGGGCTTGGTGGGGTGATATTTAACTGG  
CTTGGCCGTTCCAAGTAGGGAGTTGTGCCCCGATATTCCGGCGAAGGCCTGGGTGTTTGGGTTCTTAAAGCTGAAAA  
AAAAAAAANGGTTGGAGATTGGTCTGGCTGGGTTGGTAAGGCCTGATGGGGTAATTGACCTGGTTCCGCCATTCCA  
AGTGGGGAGTTGAGCCCCG

>NODE\_1325\_length\_156\_cov\_256.589752AATGATCCTTCCGAGGTTACCTACGAAACCTTGTTTACG  
ACTTCTCCTTCTCTAAATGATAAGGTTTCAGTGGACTTCTCGCGACGTGCGGGCGGGACCGCCACGTCGCCCC  
GATCCGAACACTTACCAGGACCAATCAATCGGTAGGAGCGACGGGCGGTGTGTACAAAGG  
>NODE\_1337\_length\_882\_cov\_169.244904CACAAGTGGGACGCCAACATGCCGGAGGCACTCATTGGGCC  
AACCTTTGACCTAATTGAGTCTTTCGTCGACAAGTCTGAGCTCTCCGACCATGCGAAGGCCACCCGGGCCCTGATGC  
TCAAGGTGTGTAGGCGCAACTCATGGAGAAGCTCATCGAGCATCCTTCCGGGTACCTCGTCTGGCTTTTTGGGTGC  
ATGCCAAGTGGCTCTTTTCACTTCTTTCGGAACACCGTCCGCAATGACCTTTTGGTCACTGGGCACATCCTTCA  
TCGGGTTAGCACGGAGGCGCCTGATATGCTTGACGGAGCGGTGGGGAACTGGCAGCTCTTCTTCCGGGAAGGCTCG

CCTCGTACGGTGACAACCAACTCTTCACCGAGTCCCTGTTCTTTCGAGTTCGGTCTGCGGTATGACCCCCACAAGCAC  
GCCGACTTCTTGAGGTGCTTCGGCATGAAGCTCAAGGTTTCGGGAGACTGAGGTAAGCCCCAAGCTGTGCGAGAGCCAG  
GTTCTGCTCTCGCGGCGTCTGACGACTCCCTACGGGCTCCTCGTCACGCGGTACACACAGTCCCTTTACGCAAAGC  
TCGCCGGACGCCCTCTCCACGACGTTGTTCTTGACAAACTCTACGTCCGGGCGATGATGGTGGACCACCTTGAACC  
GATCCCCTGTTCTTTGAGCTGCTTTCAACCGTTGACGACTCGCTACAGGTGATCTCTCGAAAGTGAGCCTCACAGG  
TGCCGTGAGGGCAGCGCTGGAAAACACCGCGCAGCACCTTCTCGGAAAGATGATGACGAGTCTCTCATGCGGGTCC  
TCGCAACGCTCTCTTGTGAAGTCTTCGACAGACGCGCCCTCCTCTCCCTTACACCCCTTCCGTGGTTGGGGCAGG  
GCGGACAAGATGGGGC

>NODE\_1351\_length\_88\_cov\_461.784088GGTCTAGGCGTCCCAGATGCTGCTAACAAAGTATGGATATTG  
GACAAACAAGTACCAGAGGTAAATGAAGAATGGTATAAGGTGCTTGTACCTGATTACAAAAGCTAGTAA  
>NODE\_1353\_length\_86\_cov\_265.476746CCACCTGACAATGTCTTCAACCTGGATCGGCCCGTAAAGGAC  
CTTAAAGCCAAAATGTGACTCGATAGAGTCAGCTCCACCTCATTGAATAAGTAAAAAGACGATAAG  
>NODE\_1367\_length\_80\_cov\_52.537498GGGGGGACTACCGCGTGGTAGTATCTAGGGTGGGTGGTAACGG  
AAAAACCAAACGGGCTGTAGAGTTGCGAGACAGTCCAAAATGGAAGAGGATGTTCAACC  
>NODE\_1380\_length\_690\_cov\_182.459427GGCGACCGAAAGGGCATGGGCGGGGTGGTTGGAAGTGTA  
GAGGGCCTCTGGCGACTTGGGTAACGGAGGAGGTCGAGCTCGATAATGGAAGCGGTCCAGGCCTGGACCCCGTGGC  
CAGCGTCATCCGACCGAAGCGATGCCACTTGGTCAAAAATCCGGGTGCATGACGTCAGCGGAGTAGCAAACATCCGCC  
GTGTATCGGCCTGCGATCGTTGCCAGCTCAAAGAGTTCTGGAGCTGCGACGTCAGAGACTTTCAGCCCGTGTGCGCC  
GGAGGAAAAGGCATTGTGGTGGTTCGATGACAACGGAGCCGTCCGGCGAGAGGCACCTATGGGCNNNNNNNNNNNNNN  
NNATTGCGCCCATGAGGGGAAGTGAATAGACGAGGCCAAGGCACCTTGGTGGCCATGACGAGGTCAAAAAGTTCCCC  
CTCGTAGTCGCGCAAATAGGTCCATGCGTCCTCACAGGCAAAGATGCATCTACTGAGGCGGTGAGCCCGCGGAAT  
CGAGATGGTCTGAAAGAGTACTAAGGGCGGGTTGCGAAACGTCAACGAACACGACGCTGGCGGGCATCCTCCTCAGG  
AGGCGAATCATATTCCTTGACGAGCCGGAGCCTAGGATGAGTACGCGAGAGCCGGTCACTCGCACCCGCTCGTAGGAT  
ATACTGTGCGCAGGACGGCATTGTACTTCCGATCGAGCATGCTCAGTTGGAGGCCCT  
>NODE\_1417\_length\_571\_cov\_133.274963CGGTGATCTTTAAGTACTCGAGCCATGATTGAGGATCTGG  
CCAAGTGCCCATACCGCAGGCTCTGTGAGGCGGTACTCGAAAACCTGTGTGCGAGTGACAAATGCATCACCAGGCG  
AGAGCCTTTGGGGAGTTTATCAAAGGTGTATGAAGGCTCAGTAACGCCTCGCGTGAAACCTCGGTTGTGCAAAGGA  
ATTTGAGCAACTCGTCAATGCGTTTGTGGTTCTTCTCCCCGAACAATGCCTGGCACGTTCCCTCAAGAGCAGCTATG  
ACTGACGGGTTGTGCGGCTGACACCTTTACGTCTACGTCAATGGATTTGTCTATCTTGTGAGGATTGAGTAAACGAT  
CGGGTCAGTACCAAGGTGATCAACCATCATAGCGGAATGTACAACCTTGTCAACGGCGGGATCGTGGTATGGGCGTC  
CAGAGAGCTTTTGATACATTGAGGTGTGCGTTCTCAAAAGGAGAAAACCCATTGGCGTTTTACCACACCCGCGGGAG  
CAGAAGCGAAGCGACCGAGATCGGGAGACACCTCTGTTTCATCAACCTTTAGCGTCATTCCAAAGTTCTTGAGGAA  
GTCAGCGTGCGAG

>NODE\_1428\_length\_492\_cov\_76.697151GTTTAAAGCTGAAGGATAAACTACAGTTGTTAATTTCAAGTTT  
GAAGGCAGTACCCGACATTGACATGAACGCAAATCCCGACTCAGTTCAAGTAGCTCACTACNAAATGGACACAAAAG  
AAGAGCTCATTGACGAGTATTTAAGCTGAAAAGACAAACTACAATTTGTTGATTCCAAGTTTGAAAGCAGTCCCCGAC  
ATAGACATGAATGCAAATCCCGATTGAGTTCAAGTAGCGCATTACTTCTACGAAACCCACAATGATCAACATTTACAA  
GCGAGTGTATATAGCAAATGTGACAGCTTTAAGTGGTATGCATCTCCTGGAAACAAAGCACACGCTGATGAGCAGAC  
TTGAATCAGGCACTATTGACGTACCGAAGAAAGTGTGTCAAAAGCTCCATTATAAGAACCTTTTCAGATGCGCCCA  
GGTACCCTAATGATGTACATGCTCACCTCTGGTCTACTTTCTGCTATAAATTCGGCTCCAAAAATGATCATTTCCAA  
AGCGCAACGA

>NODE\_1432\_length\_290\_cov\_64.599998TTCCAGGTGCTCTGATAATGCCTGTAGAAGAAGGCGGCGGGT  
TGCACGGAAATGTGTTGAACTCGGGGCGGTTACGCTCGCGCACAGCGGAGCGCTGATTGGCTTTCTGGTTCCGTGCG  
CCAGACTTTCTTTCTCCTGGTAGTGAATACTACCGCCTGAATTGGACACCGTTCCGAAACCCAATTCGACATATCCA  
GGTCATCCACAGAATCCTCCTCCATTGCCTCTTCCGCGTGAGGGCTCGTTCAACTGCATCGCTCGCGCCGGGATAG  
GCGGGAAACTCCATCGTTAGATCGGGGGGATAAATACCT

>NODE\_1438\_length\_138\_cov\_61.289856ATTCTGTGCAGTGACGGGAACGAGTGGTGGATGTCCTTGAGTT  
GATGTCTGTATACGCGGTGTACGCGAGCGTGAAGATAACATTCTTGTGCGAGTTGAAGACCGAAATTTCTCAGAGAAAT  
TACTCACAGCAAGTCGTTCTGCTTTACCGAACAAGGATTGG

>NODE\_1459\_length\_93\_cov\_52.655914CGATCCTACTTTGGAGTGGTATGAGCCGGCTTTGATGGTGGTT  
GAACCAGACCCCGGATGTTCCGATTGCGGACAAATCTTCTGATCCTTTGATAGAGTTCTTGGCTGGAGCG

>NODE\_1467\_length\_348\_cov\_283.281616ATTTATCTATGGATTGCGGATGCTACACCAGGTCACCTTCC  
AATGGACATCACAAGCGAAGAATAAGAGAGCTCGGGGCCCCACAACGTTTACCATTCAAGCAGAACCCCTGCAAGTT  
NTGATGCCACCCCGGGTCACTTTCCCATGGACATCACGAGCGAAGAATTGAGGGAGCTCGGAGCCCCACAACGCTG  
GCGTTCAAGCAGAACCCTGCTAGCTTTCGACAACGCAGAGTTAATGAAGAAGATAGTCGGTGGGTGGTTCATGGCATG  
CAAAGTCAACCGTCCCATATTCTCCAAGGAATTACTTATCAAACCTAACCAAAGAATTCCTACTTACACCGCAAATGA  
TGAACAACCTCAGACGAACCT

>NODE\_1478\_length\_112\_cov\_73.857140ACGAAGTCCGGCGAAATAGAAGATCTACGTGACCAAATTTTC  
CTGTTCAAAGAAGATCTCTCTTCTTCTTATTCTCAACAGGAATGCATCAACAAAACCTGCCCTTCCATATAGATCG  
TCGTGGCATGAACTC

>NODE\_1485\_length\_2330\_cov\_164.300003GCCCTGTAGATTGTTCCCATGGTTCGATCCTTCCCAGTA  
AAACGCGGCGTGTTCGAATTCTGATCGCTTTTACGCGAGAAAGGGGGACCACCTCTAAGCCTAAGTATTCCTCAAT  
GACCGATAGCGTACAAGTACCGTGNNNNNNNNNNNNNTATTTAGGGAGTGCAATAGAGAACCTGAGATCCGATGCGA



ACAATCAGTCGAAGGAGCGGAGCTTAGAGCCTTTACTTTAGATTCTATTAGTAAAGCGCACTCACTCTAACGGCGTA  
CCTTTTGCATGATGGGTGAGCGAGGAAATGGGAACAGCGGCTTAAGCCATTAGGTGTAGGCGCTTTCCAGAGGTGGA  
ATCTTCTAGTTCTTCTATTTGACCCGAAACCGATCGATCTAGCCATGAGCAGGTTGAAGAGAGCTCTTACTGGCCT  
TGGAGGACCGAACCCACGTATGTGGCAAAATACGGGGATGACTTGTGGCTAGGGGTGAAAAGGCCAACCAAGATCGGA  
TATAGCTGGTTTTCCGCGAAATCTATTTTCAGTAGAGCGTATGATGTGATGGCCGAGGTAGAGCACTCAATGGGCT  
AGGGTGGCCCCCATTTCGCTTACCAACCCAGGGAACTCCGAATACAGGCCTAGATCGTTTGTACAGACAGACTT  
TTTGGGTGCTAAGATCCAAAGTCGAGAGGGAAACAGCCCGATCGTACGCTAAGGTCCCTAAGCAATCACTTAGTGG  
AAAAGGAAGTGATCGAGCGATGACAACCAGGAGGTGGGCTTGGAAAGCAGCCATCCTTTGAAGAAAGCGTAATAGCTC  
ACTGATCTAGCTCCATAGCACCCGAAAATGTATCAGGGCTCAAGTGATTACCGAAGCGACGAGACCTTGAAGCTGC  
TTTTTCAAGTGTCAGTAGCGGGACGTTTCTGTCAATCGGGGAAGGTTTTTTGGTGACAACACCTGGAGATATCAGAAGT  
GAGAATGCTGACATGAGTAACGATAAAATCCTGTGAAAAACACGATCGCCTGCCAGTGGAAAGGCTTTCTGCGTTCCAGT  
CAATCTACGCAGAGTGAATCGGTCCCTAAGGAACCCCGAAAGGGCTGCCGTCCGATGGGTACACGAAAGTGACGAA  
GTTGCTTTGACTACAGAACCATGCCTGTCTGTTGGAGCGAATTGGATGATCGGGCCGAGGGCTGCCCTCTTCCCC  
TCACTCTCCTTTCCCTAATATGAACCTTGAGTCATCAAAGCCTTTCAGACTCGGCCTGGCCCGGTGCCCTACGCAA  
CTGGCGCTTCAAAGGCGAAACTCTCGTCTGATGTTTGGCGACCTAAGGGGCGGACACTTTTTGTTATCTAAAGGCGA  
AGACTCTGAAGTGGGCGAACACTCGGCCAGCGGGCGAACATGCCGGGCTCCGGCTCCGGCTCGCGCACCTGTGACACACT  
TCGAAGCACTTTCTAGTGTGCGCTTTGTTTTAGAAATGTGTGCTTCAACGTACAGAGCTACAAGCCTTTCTCATT  
CAGTGCTCGCCTTTTGAATCTTCAGTAGGGCCTTTAGTCTTTTTGATTAGAGTAGGGGTGCGGAGAGAGCAGAGCGT  
ACCGCCTGCCATAGTCGCGAGTCTGTTTATAGTCGCGACTGTTGTCATAGTCAACAAGGTTGAAACTTCCAGGAAA  
AAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGATCCTCCCGGTGAACTGACCGTACCCCAAACCGACACAGGTGAACAAGTA  
GAGTATACTAGGGCGCTTGAGAGAACCATGTGCAAGGAACTCGGCAAAATGACCCCGTAACTTCCGGGAGAAGGGGTG  
CTCTCCTATCTTTTGAATTAGGAAAGCGGCACATAACAGGGGTAGCGACTGTTTATTAACACACAGGACTCTGCTA  
AGTGGAACACGATGTATAGAGTCTGACACCTGCCAGTGCCGGAAGGTCANTGAAGGTTAAGTGGAGGGGTGCAAG  
CTCTGAAATGAAGCCAGGTAAACGGCGCCGTAACAATAACGGTCCCTAAGGTAGCGAAATTCCTTGTGCGATAAGT  
AGCGACCTGCACGAATGGTGTAAACGACTGCCCCGCTGTCTCCGACATGGACCCGGTGAATTTGAATTCCTCGTGAAG  
ATGCGGAGTACCAACGGCTAGACGGTAAGACCCCGTGCACCTTAACTATAGCTTCGCAGTGACAACCTTGATCGAAT  
GTGTAGGATAGGTGGGAGGTGGTGACACACAACGACNCTTCCAGCTGTCTCAACCGCGAACTCGGCGAAATTCAC  
TACGAGTAAAGATGCTCGTTACGCGCAGCAGGACGGAAAGACCCCGGGACCTTTACTATAGTTTTGGTATTGGTGTT  
GG

>NODE\_1494\_length\_332\_cov\_369.596375GTGTGGCAGGACCGGACCGAGCATCTTGACACACCATTCCCT  
ACCCATCCGGGTACGGGGCGACATCCCACAGTGGAGAGAGAACATGTGATAGAAGACCTCACGGACGCTGTGGGA  
TAGGCCCCAAGCCTAAACTCGGGAACGCTGAAATTTGTGTGGCAGGACCGGACCGAGTATCTTGACACACCGCGAACG  
CTTGTAACTCAACGACCCCGACCTCCGCACGACTGGCTTTGGGGGGCCTCCCTTGGGCAATTCGAGCCCAAGATTCT  
GAGCATCGAGAGGGTTGTGCGCGACGGCAACAAGGTCATTCCAAAGAAGGCAACGTAGGAGGGATCGGTATCAGGA  
TCAAC

>NODE\_1498\_length\_126\_cov\_97.261902ACTCCATATTACAACCTCACAGATTTGACAACAGTAGTACCG  
ACGCATACGCCAAATCGAAAAGGATGCCAGCGTACCATTGAGGAACTCCTCGTAATAAAGCGCATTTTAAACATG  
TCATTCCATGAAGGAAAATACCGTTTCCA

>NODE\_1503\_length\_1066\_cov\_125.832085TTTTTTTAAACATCGTAAAAGGAGATCCGTTTTGTCTGGC  
ATTTATCCGAATAAAATCTTATGATTATTAAGAACTCAACGGGACCTTCCCTCTCGAATCAGACAACTTGCCAAA  
GAAGGGGTGGGTCCCATTGAGTTCTTACGCTTTTATGTCGACAACCTCAATTCATCCGATTACTACAGGGATGAACC  
CAATCCGGAATATGAACCATAAAAGAAAATGCCTATTAACCGATCACAGGAATACCAGTTACAGTACCTATTAGCC  
AAAGAGGAATCCTTCCAGTAGTATCGGCCATTTACCCCACTTCCCTCCACATTCATCAAGTGGTCATGCTAGAGAC  
ATAAACAGTCATGGATAATTATGAGATGAGATCCTTCCGAATGGGATAAGAGAATTCCTACTATTATTTATTTT  
ATTAGTCTCTTTGCTTCTTAAATTCTTAATTGAAAAATAAATTGGAAAAATAAACAGCAAGTACAAAAATGAGTAA  
TAACCCCAAGTAAAGACTGGTACGATTCAATTCAACATTTTGTTCGTTCCGGTGGTGGTGGTGGTGGTGGTGGTGGT  
TGAACAGCCAACCATCGCACTGTAAAAATTGGATAGGTTGATCTATGGTCATTGGGGCTCCTAAAAAGATCTACT  
AAATTCATCGAGTTGTTCCAAAGAATCAAAACGGCCGGTTATTAATGGAATTCCTTGTGCGCTCTCTGTAAAAACT  
CGTTTGGCCGAGGACTTCAAACACATCGTAAGCTAAACCCGTGACGAATAACCAACCCGCAATGAATAGGGAA  
GGTATAGTAATGCTATGAATGACCCAGTATCGAATACTGGTAATAATATCAGCAAAAGAACGTTCTCTGTGCTTCC  
AGACATGCCGAGCTCCACATATTCTTGTACAGTCAAAGGACGATCGATTTGCTAAAAGATGAGATCAGTAAATAGAA  
ATTCAGTAAAATGGATATTTGTGAGATCGTCAATATTGTACCGAGG

>NODE\_1511\_length\_80\_cov\_61.400002CGCGTGCCACGAATCTCACCCCCCAATTCGGCAAGATATAGTA  
CATGGCTGGTAACAATATAGTGCAAGAGCGTGCCTGGGAATAGTGCTCTCAACTTAGCG

>NODE\_1558\_length\_1220\_cov\_93.832787TTCCAAAATAGAAAACATTTGGGATTTTTTTGGAGATTGGA  
TGCAGTTACTAATTCATGATCTGGCATATACAGAATGAAAACCTTCAATTCGATTTCTACGAGAAATTTTATGAAAGC  
CTTTCAATTTGCTTCTTTCGATGGAAGTTTTATTTTCCAGAATGTATCCTAATTTTGGCCTAATTTCTTCTCTGA  
TGATCGATTCAACCTCTGATCAAAAAGATATACCTTGGTTATATTTTCACTCTTCAACAAGTTTGTAGTAATGAGCATA  
ACGGCCCTATTGTTCCGATGGAGAGAAGAACCTATGATTAGCTTTTCCGGAAATTTCAAACGAACAAATTCACGAA  
AATCTTTCAATTTCTTATTTTACTATGTTCAACTCTATGATTCTCTATCCGTAGAGTACATTTGAATGTACAGAAA  
TGGCTATAACAGAGTTTCTTATTTCGTATTAACAGCTACTCTAGGAGGAATGTTTTTATGCGGTGCTAACGATTTA  
ATAACTATCTTTGTAGCTCCAGAATGTTTTCAGTTTATGCTCCTACCTATTATCTGGATATACCAAGAAAGATGTACC

GTCTAATGAGGCTACTACGAAATATTTACTCATGGGTGGGGCAAGTTCTTCTATTCTGGTTCATGGTTTCTCTTGGC  
TATATGGTTCATCCGGGGGAGAGATCGAGCTTCAAGAAATAGTGAATGGTCTTATCAATACACAAAATGTATAACTCC  
CCAGGAATTTCAATTGCGCTTATATTCATCACTGTAGGAATTGGGTTCAAGCTTTCCCGAGCCCTTCTCATCAATG  
GACTCCTGACGTATACGAAGGAGTGCAGGTTGTTGATAAAATTCCTACCTCTCTATCTATCTCTGAGATGTTTGGAT  
TTTTCAAACCTCCATGGACATGCAGAAGAGAAATGCTATCCCCACTCGGACCAAGACATAACTTTTACTTGTTCAAA  
TAACAATTAAGGTGAAGCAGGGTCAGGAACAACGAATCTCTTTATGATAAACAGATCCATTTTGCAAGTTCGTTATT  
ACGGGTAGTTCTTACAAAGGATCGGACTAATGGCGTATAACAATACTTGAATTCCTGATGTAGATGCTACATAGTTGG  
TTCTCATCCTTCAGAGACTACGAGTGTAAATAGGAGCATCCGTCGACAAAAGGATCACCCCTAAGATGATCATCTCATG  
GCTATTGAGAACGAATCAAATCAGATGGTCTATTCTCAATCTTT  
>NODE\_1566\_length\_95\_cov\_65.368423CTTAGGTGCTACATACACAGTTCCCGGAGAGTCCGAAATGCCA  
GGAAGACAAATTGAGCCGAGAATGTCAGAATTCCTCATCAAGATGTCAATGCGAACAGTTACCGAGATGACGAA  
>NODE\_1568\_length\_800\_cov\_495.757507TGCTTCGCCATCGGTGTTCTCTCTGATATCTGCGCATTC  
ACCGCTACACCAGGAATTCCAATCTCCCCTACCGCACTCTAGTCNCCACCGGAAATTCCTCTGCCCTACCGTACT  
CCAGCTTGGTAGTTTCCACCGCCTGTCCAGGGTTGAGCCCTGGGATTTGACGGCGGACTTAAAAAGCCACCTACAGA  
CGCTTTACGCCCAATCATTCCGGATAACGCTTGCATCTCTGTATTACCGCGGCTGCTGGCACGAAGTTAGCCGGGG  
CTTATTTCCCGCATACCGTCAATTGCTTCTTCTCCGGGAAAAGAAGTTCACGACCCGTGGGCTTCTACCTCCAGCG  
GCATTGCTCCGTCAGGCTTTTCGCCATTGCGGAAAATTCCTACTGCTGCTGCCCTCCCGTAGGATTTGGACCGTCTC  
AGTTCCAGTGTGGCCGATCACCCCTCTCAGGCCGGCTACCCGTCGTCGCTTGGTAGGCCATTACCTCACCAACTAGC  
TAATCAGACGCGAGCCCTCCTCGGGCGGATTCTCCTTTTGGCTCCTCAGCCTACGGGGTATTAGCAGCCGTTTCCA  
GCTGTTGTTCCCTCCCAAGGGCAGGTTCTTACGCGTACTCACCCGTCGCCACTGGAAAACACCCTTCCCGTCCG  
ACTTGCATGTGTTAAGCATGCCGCCAGCGTTCATCCTGAGCCAGGATCGAACTCTCCATGAGATTCATAGTTGCATT  
ACTTATAGCTTCTTGTTCGTAGACAAAGCGGATTCCGAATTTGCTTTTCAATTCGAAGGCATAACTTGTATCCATGCG  
CTTCATATTCG  
>NODE\_1570\_length\_300\_cov\_113.886665TGGGATAGAAAAGGAGAAAACACCATTTCATGCATGTGGGGT  
GCATAGTCATAGTTTTCATGAGTCTGGCAAGATCTTCATCAGGCACATAACAAGGTGGTCATACAGGATGACAGGCTC  
ANNN  
TCGCCGTTGAAGATCTTTGTTTCAGGTTTCTCCATTAGGGTGTGATTGAGAGTGAGGATACTGGGTTTCGAGGACAA  
GCCACCCTTTGGCCATAAATATCATTGGTATATGCAAGTTTCTTAGATGA  
>NODE\_1571\_length\_1081\_cov\_176.346909ACCCAGATTCCAAATGAGCAAATTCAAACTTGAAAAGGAT  
CTTTCTGATTCTCGGGCAAAGGGATTGATCGAGAAAGATCTCTTGTCTTATTATAAGATCGTGATTTGATCCGCAT  
ATGTTTGGNTCTTCTCCTTTGATTTGAGAATAATCAAAAATCGAAAGTGTCAATTGGAACATGAAAACGTGACTGA  
ATTGGTCTAGTTACTTTCGGGACGGAGTGGAAAGGGGAGGAGATTCTCGAACGAGGAAAAGGATCCAATGACTT  
CAAAAAGAATTGAACGAGGAGCCGTATGAGTGAAGAAATCTCATGATCCGGTCTGTAGAGTGGCAGTAAGGGTACTTA  
TCTGTCAACTTTTCCACTATCACCCCCAAAAACCAAACTCTGCCTTACGTAAAGTTGCCAGATACGATTAACCTC  
TGGATTTGAAATCACTGCTTATATACCTGGTATTGGCCATAAATTTACAAGAACATTCTGTAGTCTTAGTAAAGGGG  
GAAGGGTTAAGGATTTACCCGGTGTGAGATATCACATTGTTTCGAGGAACTCTAGACGCTGTGCGGAGTAAAGGATCGT  
CAACAAGGGCGTTCTAGTGCCTTGTAGATTCTTATCCAAGACTTGTATCATTTGATGATACCATGTGAATCGCTAGA  
AACAAACATGTGAAGTGTATGGCTAACCCAATAACGAAAGTTTCGTAAGGGGACTGGAGCAGGCTACCATGAGACAA  
AAGATCTTCTTTCTAAAGGGATTTCGATTCCGAACTATTATATGTCCAAGGTCCAATATTGAAATAAATTCAGAGGTT  
TTCCCTGACTTTGTCCGTGTCAATAAACAATTCGAAATGCCTCGACTTTTTTTAGAACAGGTCCGAGTCAAATAGCAA  
TGATTGCAAGCACTTCTTTTTTACACTATTTTCGAAACCCAAGGACTCAATCGTATGGATATGTAATAACAGGATTT  
CCAATCCTAGCAGGAAAAGGAGGGAAACGGATACTCAATGTAAAGTGTAGTAAACAGAATTCATACTCGATCTCATG  
GATACATATAGAATTCTGTGAAAGCCGTATTTCGATGAAAGTCGTATGTACGGCTTGGAGGG  
>NODE\_1574\_length\_79\_cov\_57.303799TTTCTTTTCTCATTATGTACGTCTCAGCAACCTTCTTTGCAGT  
ATCATGCGTTAACGTAACAAACAGCATGTGTTCAAATGTCATCACTTCATCTTGGCTG  
>NODE\_1575\_length\_248\_cov\_107.068550TGAAGCGCTTGAGGCTTAGAATGAAGAGGGGGGTGCGGCGA  
CAGTTGCTTGACGAGCAAGGAGCCTGCCGACCCACGACCCAAATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
NNNNNNNNNNNNNNNNNNNNACCGGACTCGTCCGGAGGCGACTCCTGACCAGCACACGGGAGCAACGAGTGAAGCCTT  
GTAGGACCGATATAGGCATGGCGACGATGCGTCTGGCTCCTTTTCGAGGCAACACGCTCGCTCGAGGGCTTGAGG  
>NODE\_1589\_length\_93\_cov\_240.225800AAGCCTGATCCAGCCATGCCGCGTGAAGGCTTAAAGGCTT  
GGTTGTAAAGCTCTTTCGTCGAGTGCAGCATCATGACAGGACTCGAGGAAGAAGCCCGGCTAACTCCGTGCC  
>NODE\_1601\_length\_88\_cov\_50.204544GCAGCCAATCTTCTCATAGTTCTCGGCACTCTTATTTCATCC  
CTAAGAAAGAGGGATCCTTGTTCATAGGACGCTAGCCGTATCAGCCCTTTTCTTGTAGCTTGGCGCAA  
>NODE\_1610\_length\_328\_cov\_66.137192GCAGCGTGAAGCACGCGCTTCAAAAAACAGGTGCTGCTTTAAC  
GGCAGCGCTTGCCTTACAACGGACAGATTGCCTCCTGAGGCGAAGCGGGCTTTTTCAGCTTCTGCTTCTGCTACTG  
GCAAACCACCGTCTGTTGAGGNN  
NN  
CGGCAGTATGGATCAGATGTCAGCGTGCAGCTCTTAAGGGTGGTCAACATCAACCACGCATCACTTCTGACGCACC  
>NODE\_1639\_length\_487\_cov\_172.681732GGGTGGTAAGGATCGTGGACAACCCCCCAAGCCTGCATC  
AGAGCGCGCATGTTGATTGGGGTGACAAAGTCAACACCGGGGATACCGGGGGCCCGATCAACCACCGCGGAATGCA  
AGGAGGTAACAGTCCGAATGTCTCCTCTAGGGCCTTGGCTCAGGAGCGGGAGCCTATCATCAGGCGAGACACCAAGG  
ATAGCAGAGGACGCTCAGTCTCAAGGAGAAGAACAGTGGGGTGCATGACAGGAGGGATGGTGACGACGGAGGAAAA  
GTAATGATCCATGACCTCCGGAACGACACCATCGGAGATGGCCGCCATGACAGCCCTAGGGTTCAGGGAACCCACGAG



CCGACATGCATTGACGCGATTTTCGTAGGCATCATCATGGGAGAAAGATGTCAAAGGTTTGAATTAACCTGCTGACTGT  
AACATTAGCTCGGTGCAGTTCCTTTTGAACCCCTAGGCGCGAAAACGTCAACCATGTGTCGTGAACATTGAATTTTT  
GGGGTTTATTTAGTTTTGGGGTATTTTTCAGCCATTGCGACTGACCATTGCAACATTGACACACCCGAAGTCAAATTGA  
GTTGGTACGTAGCTGTGCCGCACATAGTAGTCAACACATTCTTTGGAAACCATCCCGTTTTCGAACATCCGTACGAG  
ATCATGTATGTTGACATCCTTATCGCCGGCTCTGTCCCCTCGCAACACCCACTTTGGAACGGTCTATCATAGCACT  
CGTTCC

>NODE\_1723\_length\_477\_cov\_80.215935GATCTATGTGTCTATATAGATCCTGTTTCATGGATTAACGAAA  
ATGTGCAAAAGCTCTATTTGCCTCTGCCATTCTATGAGTCTCTTCTTTTTGCGTATGGCATCGCCACTCCCTTTGG  
CAGCATCCACTAATTCGGAACCTAATTTGAAAGCCATATTTTCGACCCGGACGTTTTTCGGGATGCCCTAATAACCAA  
CGAATGGCAAGTGCTTTTTCTTTGTGTGGATCCTATTTCAATGGGAACCTTGATGAGTTCGATCCGCCTACACGCTTGC  
TTTTACTGCTATATCGGGAGTTACTCCACGATTGCTTTCAGCTAAAAACAGATAGTGGATTTGTTTCTGTCTTTTTGTT  
GAATCTTTTTTACGGATCGATAGATAATTTGATAAGCCAATGATTTTTTTTTCCGTGTTTCAGAATACGGTTAACCAAC  
ATGTTAACTAATCGATTACGATAAATTTGGATCGGATTTTTCAGTTTTTTTCTTCTGCAGTACCTCGACGTGAC

>NODE\_1724\_length\_102\_cov\_50.039215CGACATTTGAGGTAGTAGGGTGTACATGTCATCAAGTCTTAC  
CGCCCGTTTCGGTTGGCAGTGCCTCGCCCAACCCCTGCTTCAATCCCGAACAAACACGCATTCCTTAGCAGGATTATGAG  
ATGGG

>NODE\_1735\_length\_244\_cov\_58.069672AATGACGTATGTTTACAGTCCAGGCCGTCACAAGCTTGAAACA  
GTATACCTTTGTGCCCTACAAACTTGACGTATATTGCCTCCTCCATACTTAAACACCCGCGACATCATGGCGTGCACC  
TTGTAGTTGCCAGGGNN  
NNNNNNNNNNNNNCTCATTTCATACTCCAGATTGACAATACCCGAGTCTCTGTCCCCATCGCTCTCACTAT

>NODE\_1740\_length\_176\_cov\_88.448860GGGCTTCCACGTTTACAAGTCTAGCACTGATCTCATCTTTCA  
TTCGACACAAGACGGTTTTGCGCGTGAGAGACTTACCGTACGTGTTTCATAGCCGAAAAAGGGATCTTCATTAAGGGCA  
AAGATGTCGATGCAGTGATTGCTTTGGGCGATAATCTATTCGTATGCGACGACATACTGGTCTTCCATGATGCTATT  
AA

>NODE\_1757\_length\_768\_cov\_232.673172TTTCATAATGGAATGAAAGGCCCTCCGGCTCCCTCGACGAG  
AAAACCTCCGTGTTACTACACTTGTCTACTGAACAGACTGCGCGACGGACTCGAGGTCGTCAGGGTCNNNTGCGAGAG  
CAGACCATAACCCGTGGTTATCATTGACCGCACGGCATCACTAAGTTTCAGGGTTCGTTGGGTTGCCGCCGCCGGCAG  
GCACTTGAGGAGGGCTGGAATACCCGCATGTTGCGAAGAACATTGCACGCGATCATGGAGCTGAAGGTCTGGACCGTG  
TCGTCGTAGGACCCGGAGTACCCCGGGAGAAGCTCCAAGCTCACGTGCCCGTTGTGGCCGTAATTTGACCTTGGCTT  
GACGAACCTCACCCCCCTGCTGACAACTGAGGAGAGCAACGCGACAGCGCGTAATGTGCTCGCAGCCTGGGC  
ACCCCGGTTCCACTTGGGAGACGCACTCAAGATAGGGCTCGATGTCAGNNNNNNNNNNNNNNNNNNNNNNNGCGCTT  
GGCCGTTCCCATCGCAAGCTGTGCCCTAGCCCTGTGAGCGCAACGCGTAAACGCGAGCAGCGGCATCGACGCAGA  
GCTTAACAGGTGTGCCCTGCTCTCAGCACTCCCCACTCAGTAGGCAATCGAGGATCGAGAGTCTGTTGCCGTAG  
TATTGGAGGGCCNNCGGAGCGTTGGCATGCGGGGCGATC  
GCCGCTCTGGGGCGGAAGCGAGGAGCATCTGCAAGGTACACGTGGCATCGTCATCC

>NODE\_1759\_length\_78\_cov\_65.333336GCTGAATTGGCGTGTGCTTCCAATGCGATACCTCTCAAGAAGC  
ATATATTTATGAGATATGCAACTTACGTTGAGTTGAGAGATTGAAAGGACGTTTCGAG

>NODE\_1765\_length\_123\_cov\_95.959351ATCAGGGTGTCTACTACGATCTTAGGCAGGCTGATTGAAGA  
TCAGAAGTGTCTACGATCAGGGTGGTTCGTAGATCAGGCTGATCTTTGATCAGAAAGTGTCTGTAGATCAGGCTGATT  
GTAAATCAGGGTGTCTACTACGATC

>NODE\_1766\_length\_210\_cov\_70.790474GTACCAGATATTGATTCTTTACCTCCTCACCATTAACGATGA  
CAGATACCGTACCGTCCGTGGAAAGATTAACCTGCGCCACATCGCTTTCTTTTGGTTCTCCCTACGTTTCGATACG  
CCCCGGAATGTGCGTTCAATTTAGAAAAGCCCTGTTATTCTCCCTTCCAATTATACCGTATACCACTTCATCACCGC  
TCACCGTGTACCTATGAGTGTATGAGCAGGGTATTG

>NODE\_1767\_length\_84\_cov\_77.428574TGGGCGATTCCGTTCCGCCATCACTGACTGGCGAAGGATATGC  
GCAGTATTTACTGCACCGTCCGACCTTACGTTGGCTGGCGGAGGACAAAACAAAATCCGTGCT

>NODE\_1778\_length\_214\_cov\_100.126167CTTGGCGTTTAGCTCTTTCGCTGGTTAAGTCCACGACCTCTA  
CGCCCCCTTCAACTGGGATTAGATGTTGAGTTATCGCACTATTATTGGATAGATCCTCCTTCATCGAAGATACAAAT  
ATCCCTAACTTAGCGTCGGATACCTTCCCCTTAATTCTTTGTTTAAATAAACTCCACGATGGCTCTATCGATGTCTCT  
GCCGCTAAGAAATTATCACCCACTGAAAAGATGACGCATA

>NODE\_1779\_length\_191\_cov\_51.094242CTGAGCGGAGAGTCAAGTTCTTTAGGGCTTCAATGTGTCCG  
GCAAGTGAAGTAAACTGCGTGAGAAGTTGTGCAACTTCTCGGGGTGGTAATCGCCAGGGTATTTATCAAAGGCTTC  
TTTGATGCCTGGATCGATGTACGTTGTGAACTCTTGGCGAACTGGTTTTGGGAATTTGTGTGCGGCCAGGTAAGCTA  
GAGAGTTTTGTTGTTG

>NODE\_1797\_length\_1944\_cov\_263.051453TTTTTTTTTTTTGATCATAATAATTGCCAACAAAGAAATTTG  
GTTCTTTTTTACGAACTTGATGTTGATAAATCCGCGGATCTAGAAATTCATTTTCGGACAATTGAACCTTCTCAAACCTG  
TTTTTTTTTAAAGAACCAAAAAGATTTGTGCCAAAATAACAGAAGACAAGAAGAACAAGGCTTGGATACGTAATG  
GATCTTGAAGTACTATTTCTGCATCTCCCTGACCAAATCCACCCACATTAGGATTAATGTTTAAATGTTGATCAAGT  
TTAATAGATTCACCCCTCTGAAACAAGAAGTTCTGGTCTGAGGGGATAATATCAACCACTGACGTCCATCCGATGC  
ATCCGCTATGGTTATTTTCGTACCCCTTTTTCTTTTTCGTATGATTTTGTCTACTATACTGCGCTGTAGCATTAT  
AAACTGTATTGTTACTCTTGGCTCCCGTACGATAAATCTGACCCCTTCCCCTGTTCCCCTACGTATATGGGATAT  
TTTAAAGAAGTGAACATCTTTCTTAGTAGCGGGTTCGGGAGAAAAGAAATAGGAAAAGGTGATTTCACTATATTTTTGACC  
AGGAACAGGACCTATCACAAGAATATTTTTTTTTTAGTAGGGCGATAGTTCTGAAAAGATAGATTGCCATATCTTTTTCT

TTCATCTCGGGGAAATACGATCGGGTGGGGCTAATTCAAACCCCTCAGGTAAAAATAAGAACAGCCCCACATTCAA  
AGCCCTCTTTTACCATTAGCAAGAACTTGTTCATTTGCATATCATAAGGAATTCGAACAACTGCTTCAAATACAG  
TATCAGGAAGTACCGCTTGTGGAACCTCAATATCCACGGGCTTATTAGCTAAATGACAATTTGGCACATACAATACGT  
CCAGTTCGCTTCTCGTGGATTTTCATAACCCTGCTGTGCAAAAATGGGATATGCATTTGAAATGGATGTCCGAGTTAT  
TATATATATCATAAGCGATACGGAAATGGATCGAGTAATCTGTTCTTTATCCAAGAAAAGGTATTTCTAGTTTGA  
TGGTCCAATCATTGATCCCGAAAATTTCTACAATAAATTAGGTAGGTGCTAGTATAGTTCCCTATCCACGATTCTG  
CTATTTACTGAAATAATTTTACTGGAATATAGGAGGAATCCCTTGGATGGGTAGAAAATATAAAAAGTAAGCACGATT  
TTGAACACTTTGTTTATGTACAAAGTAACAAACATTCTAATTGGATTAATATCAACAATGGATTGGATCATTTTTCA  
GTCATTCAATTGAATGATAAATCACTACAAGTGACGGAGATACACGATTTAAATACCGGAAGATCCAATATTTAAAA  
TTGTATCTAGAATGACCGGAAAGGTGGAAACAAGACCAGATATAATTTGATCGTTTATGAGCAAAATCCAAAATCTTTG  
TAGACAGAGCCAATCATCAGTTCCCAACCATGGGGCGAGTATCCGATACATAAAATCAGTTAATAAAAAGAAATAGAAA  
AAGCTTTTATTGTGTGCTTAAAGTTATATAGGAATTGCTGAATCCAAGAGTTAAGAATAAGAAGTTCTTCATTACCC  
AGAATAGAATAACCACTTAGAATAACGAAACAGATTATATTTGTGAGAAAGTGCAAAAATCGTATGGATAGGATCCTC  
ATTGTACATCTTGATCAATTGGATCGTTTCTTTGTGGATTCCCATACGAAGCTTTTGTAGATGTGTTTCCGGGCATT  
CCTTTATCATTTCGTCCAACAAGAAGAGTTCTCTAATTCTATGAATTTTCTAGAATACTCTTTTCTTGAATATCA  
TTCAAAAAGTTTTCAGATTGCCTAGTATTCCACCAATTAGTAACCCAAGATTCCAGACTTTTATTAAGTGATAGAGA  
GATCCACCAGGGCAAAAATACTATAGATGCAAGATATAGAAGGGGAGTGAATGCTTCTTTTGTTCATTTTTTTTTT

>NODE\_1798\_length\_3936\_cov\_305.209106CGCAATCAACGAGGAAAAGGGCTGGTAAAGTTGAGTCGCTA  
CGCCTCCCATGATTCACAGTTTCTTGGGGCTAATAAAGCAGGTGCACAATATGAGCGCAAACTTGTAACTCCATGG  
CAAGGTACTCATCTCTTTCAGGAGAGTTCGAGGTAATCTGTTTGCATTCATAATGCCTGCAGGCCATGGCAAGACAACA  
ATTGCCATAAGTATGGTTTTGTTGACGTGCACATGCTTGTGACTCGTAAAGAGCATGATGACTTGGCTGAGGACAG  
GGCCTCCATCTTGGCCGGCCAACAGAGCTGGATGGCGCACAACACAAAATGGTATGCAAGGCTAAATCAGACGCTGG  
ACCTGCTGGACTACTCAATGCCGGTAATTCTGCTAGTGCACACCGAGGAATTAGCTCTGGAAGTGGAGCCAAAGCCG  
ATAGGATGTTTCAAGCTAAAAGACAAACTTTGAAACAAACATCCGTGACCGTGACGACCTGGGCAAAACATTCTC  
AAGGTTGAACTACACAACGTTACATACATCGAGACAAGTACCAAATCAGTTCAATGAACTCACGAACCGAGAGCTTG  
AAGGCATGATACTAAAATCTTGAACCTGTCAAACATGCCAATTGCTTGTCCACACAAGTTTCGATGAGTCGGTATGG  
AACGAGTGCTATGACAGAACAGTCCCAAAAGTGGGTATTGCGAGGGGACAAAAGCTGGAGATAAGGATGTGAATATACA  
TGATCTCGTTAGGATGTTTCGAAAACGGGATGGTTCCCAAAGAATGCGTTGATTATTATGTGCGACACAGCTACGTGC  
CAACCAATTTGACTTTGGTGTATCAATGTTTGAATGGTCCGTCGCAATGGCTGAGATAACCACCGAAAATAAATGCA  
CCCCAGAAATTCACGTCACGACGATATGGTTGATGTTTTCGCGCCGAGGGGATCGAAGGAACTGCACCGAGCCAA  
TGTCACGATCAGGCAGTTAATTCAGACTTTTGCATCTTCTCCACGATGATGCCTACGAAATTCGCTCAATGCATG  
TTGGCGAGCCTCAGGTGTTTGTGGCCAGCATACTAGCGCATGGAAGGGGGTGTACAGAATACGTCTGTAAATTTT  
ATTGTCAAGAAGTGGATGGGAGTGAACCTTACCACCTGGTCAGATGATGAAGAGCTTGCACTACATACATCAGATG  
TAGCCGTTTTCTAATGAACACAGAGATTACTGAGCAAGACAGACAAAAGTTTGGATGTATATGGACTTACTTGTGGGTA  
GAGGTGAGTATCAAATCGATGAAATGGCTGAAAGTCAAGCTCAGCCTCAGGCAATCCGACACTTACTCATCGAAGCATCTATCT  
TACAATCCTGACTTAAAGGCTTTCACTAACGAGCAATACATTAAGACTTCTCCCGGTCTGTTCAAATAGCTCATAT  
GCGCATCAAGCACAACCCAAAACCGCTACAGTTTACAACATTTACAGATTTTCTACAAAAGCACGGCACACATGGATCA  
CCAAAGGAGGGTTAGTGAGCAATACACTGCCCAAAGGTATGAAACAGTTTCTCATGCCAAATCTTTGACGCTGTGCGG  
GCCACCGTTGAGGAAATTTGAGGGTAGACACAACAAAAAATCGTTGTTTCGAGATGAACGAACTTACGACATATTGGT  
CGGTGTCAACAAACACAATTTCAACGTCACAAAACGATGATCAAGTACGAGGTGGGAAAAAAGACAGAACACTGC  
TCCCTGGAAGTCTAGCTCACTTCATCGTTTTCTGCTACATACTAATGTTGGCTGAAAAACAGGATCAGATTGGGAGT  
GTTTCGATTAACGCCATGGGTGATGGTGACATCCGGTACTTTGACAAGAAAATGTGACGGGAACCTTCCATGTATT  
GTACGACTGGGCAGACTTTAACGAACAACNGGGTATGAAACAATTTTCATGCCAAATCTTCGATGCTGTTGCGGCCA  
CCGTTGAAGAGATTGAGGGTAGGCATAATAAGAAGTCGCTGTTTCGAGATGAATGAACTTACGACATACTGGTCCGGT  
GTCAATAAACACAATTTCAATGTCAAAAAACAATGATTAAGTACGAGGTGGGTAAAAAAGACAGAACACTGCTTCC  
TGGAAGTTTAGCCACTTCATCGTTTTCTGTTACATATTGATGCTGGCTGAGAAAACAGGATCAAATCGGGAGTGTTC  
GCTTAAACGCCATGGGTGACGGCGATATCCGGTACTTCGACAAGAAAATGTCAACGGGAACGTTCCACGTCCTATAT  
GATTGGGCAGACTTTAATGAACAACACTCAGCCTGGGAGATGGCTCAAGTATAGGCGAATTTGAAAACGTCATGCC  
TGACACCCAGATTATCACCTATTTCTGCAGCGCTATAATCGAAGGAATGTATGAAATGGGCTACAGGACAAGGAGG  
GTCATTTACATAAGCTGTGGCGAGGGTTGACTCGGGTTGGCGTGGGACCACATGGATTAACACGGTGCNTGCTATA  
ATCGAGGGGATGTATGAAATGGGACTACAGGATAAGGAAGGCCATTTACATAAGTTGTGGAGAGGGCTGTACTCAGG  
CTGGCGTGGCACACATGGATTAACACGGTGTCAACTTCTGTTACGTTTTCAATTGCCCTACTAAAATATGGAGAGAT  
TAACGGGGGTGAGTGTGTGCTTATGGTCGACCACGGTGGTGNCGCCCTGCTAAAATATGGAAAAGGCTGACGGGGGTA  
AGTGTGTTGACTTATGGTAGACCACGGTGGTGTGATCTGGATCTGGTATTAAGTGAAGCCAGCGTAATGCCGACTTT  
TCTTGAATAATGGACAACATGCTGTTCAAGGCAACAAATGGAAGCAGATGTTTCGGTAAGCGGTGAGAAATTTTCA  
GAAACACGATCTCTCGTGCCAGAGTCTACGCTAGTCCGACTCGAGCGCTAGCGAGTTTCAATTGCCGGTGTATTGGGAG  
GGTGTGGGCGTGAACAGTCAAGGAAAGGGTTGTTAGTTTGTAGATCAGATAGGCAAGCTGCGAAGGAGAGGCGT  
CAGCAACGAGATGTGCCAGGGATTTACCATGTGTCATATCGCATTGGTGCAAGATCAAGGATGGTGAAGAGTGGGA  
TGAACCTTCCAGAAGCGGTTATTCATTGGAGTCAAGAGCAGGGTGGTCTAGGCGTCCAGATGCTGCTAACAAAGT  
ATGGATATTGGACAACAGGTACCTGAAGTAAATGAGGAATGGTACAAAAGTCTGTTGATCTGATTACAAAAGCTAGTA  
AGGACTATGTGAAAGTGTATCAGATGAAATGGCCCCGTTTCAGTATCGTAATAGAACGCCAAGAAGATTTGGCACAA  
ANAGACTACGTGAAAGTGTCTCTGATGAAATGGCCCCGATTTAGTATCGTAATAGAGCGCCAAGAAGATTTGGCACA

AAAAATGGCAGTTGACTCCTACGATCTTGAAAAGAGCATCGACAGAATGGCATGGAAAAAGCTACTAAAATTCGATG  
GCAAACCTGTGGGTTTTCAAAGACGTCATCGAGCCTATGGTGAATGACGATGTCATGAGCAGCTTCGTTGATTTCCNA  
CTACTGAAATTCGATGGCAAACCTGTAGCTTTTCAGAAAAGTTATCGAACCCATGGTGAATGACGATATTATGAGCAG  
CTTCGTTGATTTCCAAATCGATGAGCAGATGACCAGAAAAGTTTGACAAGGCATCACGATATCAGGAATACATTAAC  
ACCTGACCCTGATGACCATGCTATATCAAAGGAGCAATTAGTTCATCTCATGTCTGATGGGGAGGTATCACTCACT  
GCAATAGAATTTCAAGGTGATATTTACTACCAGCGGCTGGTGCCAGAATTCATCAGTACGAGAGCCGTTATATTTTG  
CAAAGCTGCGCTAAACGCTGAAGCAATAGACACTGAAAGTGCGGCTATAATATTTAAAACCATATGTT  
>NODE\_1799\_length\_78\_cov\_140.730774CGCGATGTTCCCTGAATACATCATCGACGACAAGCAAATCTC  
CGCATATCGCGCACGCGCTCTATGGGAAGATGGCAACTACGAGTCTTCGATTCTAGGC  
>NODE\_1805\_length\_83\_cov\_59.542168GTGGTCCACAGCTGGCGCGTTATGACTGGGTGCAAGTACCTCT  
CACTTTTTACAGCGCCACGCACACGCAGATATATCGAGTGCAGCTTGATCGCACGGTATTAC  
>NODE\_1806\_length\_120\_cov\_54.108334GTCATACAGGCGTATTCCTTCTGAGGGACAAGCAGCAGTTC  
GGACACTACAACATTTCTCTGTCACTGCCACTGAGCTTCTTGATCTTCTCGAGCGTGACCCTGACGCTGCCAAAA  
GGGCAATTCACGACATTGTCCAGG  
>NODE\_1814\_length\_110\_cov\_107.438713CGTAGGTATGGTGTATTGATATTTGATGACTTGGTCACCCT  
TGACGAGCACTACTCTCTGGTGAGCGAGAGGAAACGCATCATTGAGCAAGACGGTGAGTGGATAACAGCACAATGATA  
AGTGGTTTTGCCGAGCTAGGAAAACATTGAGTTTTCTGATGCTGGAAGAACACACTGTCGTTCTCGCACANCGGCTC  
TGTCATAGGAGCGAACATAATCGGAGGAATCAACTTGACAGAGGACGGTTTTAGCAAAGAACGTGAAGGACAGGGCA  
GCATTGCAAAGACTTGAAGCTCAAGGTAACCTTGATGCGGAACAAGGTGTACCATTTCATGT  
>NODE\_1831\_length\_125\_cov\_66.008003ATCAACTCGCCGATACTCTCGGTCCAGTTACGATTGATCGAG  
CGCCGACCCTATTTTTATACACTCGACTACTGATGCGTACCAAACCTTTGGGGGCGGAGATCATCTATCTCGAACCC  
GTACTAGATTTCAGTTCTGTCAATTCAAC  
>NODE\_1840\_length\_155\_cov\_121.000000TAACACATGCAAGTCGAGCGGGCCCTTCGGGGTCAGCGGCA  
GACGGGTGAGTAACACGTGGGAACGTACCCTTTGGTTCCGAATAACGCTGGGAAACTAGCGCTAATACCGGATACGC  
CCTTTTGGGGAAAGGCTTGCTGCCGAAGGATCGGCCCGCGTCTGATTAGCTAGTTGGTG  
>NODE\_1843\_length\_337\_cov\_72.863503TCTAGCGGCGACTGGTGTGTGTCGAGGCAAGCAGATTTCAGG  
AACCGCGATGTCAGTGGCGGAGCTGCACGTTACAAGCAAACTATTGTTATCAACANNNNNNNNNNNNNNNNNNNNN  
NNNNNNNNNNNNNNNGCACTGGAAGATCTTAAGCTCATGAATGATTATGAAATTAAGCGTCAAAATGTGGCGATACTTCA  
GAAGACCCAACATACATGTTAAGTTATACATGGACTATCACGTGTGAGGGGCTGACGAGTACTCTCAGGGAGACG  
AAACGCGTAAGCGTAGCCCTGGGCACAATTATAACAAATCAAATAAATAACAACAACAGAACAAACAGACAACAACAAA  
CAACTAAAA  
>NODE\_1845\_length\_83\_cov\_51.457832AGGCTTGCCATCTCTGCTCAGATATCTCGGTATTCATCAAGAA  
CCGGCTACAGCGCACATACGAATGCAGACTTTTCATCAAATCAGGCCACGCTTTGAAGTTCA  
>NODE\_1846\_length\_78\_cov\_143.256409GAGACCTTAACCTGCTAAATAGACAGGCGCCGGCTTTGGCTGGT  
CGCCGCTTCTTAGAGGGACTGTCAGCGTCTAGCTGACGGAAGTTTTGAGGCAATAACA  
>NODE\_1852\_length\_442\_cov\_85.368782GGTGTCTACACCTCTCGCCTATCGAAGTTCTGTTCAAAAACC  
TCGTCCGAGAACTTGTATAGAGAAGGATTTCCCGCGGCGCAGCAGTTCTTCCATACCAACTTAGCTGCCCGGCGCTG  
CTATTGGCATAACAACCGGTACACCATAGGTTGGCNN  
NNNGAATCGGCGA  
ACAACCGAACCTTTGGGACCTTCTTCAACCCAGGATGTGATGAGTCGACATCGAGGTGCCAAACGACTCCGTCTGAT  
AAGAGCTCTTGGGAGTCATCAGCCTGTTATCCCCGGCGTACCTTTGATCCGTTGAGCGAGAGCCCTTCCACACGGGA  
CTCCCGGATCACTATGGCCGACTTTTCGTCTCTGCTCG  
>NODE\_1861\_length\_476\_cov\_75.180672AAGCTCCGAACTTTACGAAGAGCTTCAGCGGGTCTGGAACGA  
AGTACAAACCGTCATCCACCTGGACCAAGAACTTAGAGCAAAAATATGGCGCCGCTTGATCGAAGACCTTACATCG  
AAACCAAATATCGCTCAATACCGTTGTGTGATCTCCAGAGGCACCCGACTAAAATATTAGACTATCGTCTCCACT  
AACTACTATATAATCATAATCTAGACCTCTAAGTACTANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
NN  
TATTACCGCAAAAAGAAAGACGACAAAACCTCCTTATCAACGCCAAGAGTCAGGAGAATCATCTCTCAACTTGCTT  
CATTAGAGCGCTTTGAGACTTGTCACTTGTCTATGTCCAATTCGTAAGTGTAGTATTGCCGAATGTCCC  
>NODE\_1862\_length\_211\_cov\_47.331753CTTGTGAGAACCTCATACCCATTTAGGGCGGAGGACACTCCC  
TTTAGGAGTGCTTGGTGGTCTACCTTTGGGCGGTACCATAGTCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
NN  
GTCATTGTCCAGCTCGCCTTCATTTATTGCATGCCTG  
>NODE\_1903\_length\_643\_cov\_95.740280GTCTCCTTGATATCCCAACGAAGCGAACCAGGCTCCGTAGTC  
TTTAAAGAGAAATGGCAGACTTGACACGTTTGGAGAGTTATTGCTTCTTCGGGTGACCTTGATAACGTTGCCCTAGCT  
GCAGTTCTTTACCACCTTTCTCCAGACCCGACGAAATCCCCTGACCTGGACTATTGCTTGTGAAATCCGCACTTCA  
TTGGTTATCCAGTCCCACCTCCTTTTGTTTTCCAGCGTCTTTCAAATCTTTTTCGCAGTCCCAGGAGTTTTCTG  
CTTTTTGTCTCTCAGTTCCCACCGCACGTTACGACGACATAGCATATTTTGCTAATGCCATGGCAACACCAGATCAT  
ACGTGGGATAGGACCGCCAGGAATGACCTCANNN  
GTCGTGACTGAATTCGCCGTTCTTGATCTAGTCGCCACGTCTCAATATGATTCAATAACGCATGCGACTACTGCCG  
TTGCCACGATGCGCCAGACGATAANNNNNNNNNNNNNNNNNNNNNTTAAAGTCCATAACAGGCAATTCAGCGGTACGC  
GTTGCAATCTCTTGGCTCATTAAATCTTCTGGCACTGATGAGACGCGGTACATATTGACGCCCTGGTGCCATCAAA  
AGAATCA



>NODE\_1958\_length\_295\_cov\_58.152542TCTCCTCAGGGTGTATCACCAACCTATTATCTGAATCGACTA  
ACTACTATGGCCCATCGGTGCGAGATGCGCTGCTTCTATAAAATGAACTCGCTTAGCCTGCGTGTATCCGCATCGAG  
TGTGTTGATTTCTCTCAATCTTCGGTACATTCTGGCTTCCACATGTGTGAGACTTCTAGACAGGCTGTCTTTTCGTTA  
CAGTTTTTTCTGCCATCTGCAATTCGAATGTGAGTAGTTTCAGCGGTTATAGCCTCCTCTACTACCCGCACCCTTTCG  
GCTGAAAGAAGGGCAGTTGGGTTGTCGTTTTCTAGTCGCGGCTTC  
>NODE\_1959\_length\_1477\_cov\_99.710220GGTGGGTAGGCTGCGGCTATTATTACAAAACACAGTTATAC  
TATATTACAAAACAAACCAACAACACACCAACTACAACAAATACAAAACAATAACAATGAGTCAAATCAGCTTTAACAA  
CGAGAGATACAATGAGAGGAACGCTGCGTTTTAACCTTCTTAGGTCGAGGCGCCAGCCCTCAGGAATGCAAAGGCTG  
CAAGAGACAAGATCACTTTCTGGAACCCCGGGACCAGGGAAGGAACAGGCAACCACAAGGAGAAGATCCTGACGTTG  
GGGAAGAAGATAGGGGCTACAGCAGCTTACTTTCGACAACAAAAGAAGCAGTGCACCTTGAGATAATTTGCGATTATCG  
CTTCACACTGCACTACCAGCTGTACGGAGANCTGGGAAAAGAAGATTGGGGCTACGGCAGCATACTTTGACAACAAAA  
GGAGCAGTGCACCTTGAGATAATCTGCGACTACCGTTTTACATTGCACTACCAGCTGTACGGAGACGTGGGGCGGCAA  
GCCGTGGAAGGGGAAAACACACTTTCCATATTTACGAAGTGGTTTTGGACAGGTGTGCTGTTGATGTTTTGTTGTT  
CGCAGATAAGCTAGAGCCCGTCCAGCNGGGGAAAATACGCTTTCCATCTTCCACGAAGTGGTGTGGGACAGGTGTG  
CCGTGATGTCTCGTTGTTTGTCTGATAAGCTAGAGCCCGTTTCAGCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
GCATGTA  
GCATGTTCAAACAATGCTGAACGAGGTGTCCACCGGGCAACACAAGCTAACCCGTTTGGTGAAGGGTATCTGATA  
CTATTAGAGGCGATGGACAGGGGGTCAATCGACGTAGTCGCTGACGTTTCAGAACGAAGTCTGTACACGCGCGCAGA  
CGTGGTCAATTCGTTTCAGGGTGGGAAATCGAGCATACTGAAACAACCAATGCCAGAGAGCGGCCCCACACGGCNT  
GATGTGGTTCGCTGATGTTTCAAGACGAGGTCTTGTACACGCCAGCAGACGTGGTCAATTCATTCAGGGTAGGAAACCG  
GGCATATGTGAACAACCAATGCCAGAAAGCGGTCTCACACGGCCGTTTTGTGGAAAATGTGCGGCGCTACCCGC  
CTCCCGAGCTCGCTAACTCGCACGTACAATCCCTGCGGATGCTGAGCATGTATTGTTGTTACACGGCGGGCTACTG  
GACCGGGGGCGTCAGATCAACCTCACGCCGGGCGTCATCCATTCTCGTTAATGGCTTACGCAATGGATAACCAGTTG  
TACAGCTTCTTGAACCCGNTCACTCCTCGCTAATGGCCTACGCAATGGATAACCAGCTGTACGGCATCCTTACAAC  
CGNN  
AGGTAGATCTGATGCTTCTTAGCATGATGAAGATGAACACGAGGTTAGACAAGCCTATGATAAGTAGGACCA  
>NODE\_1974\_length\_373\_cov\_54.329758TGGCGCAACACAAGCCATAGCTTTTCGGGAGATTTACGATTGA  
GTACCAACTGTACAGTGACTTGACGAGAGAGCAGGTGCTGGGTGAAGACACACTTNNNNNNNNNNNNNNNNNNNNNN  
NNNGGCCGAGTG  
GGCAATGAAGTTCACGCGCTACAGGCAACAGAAAGTATCTGAGCAAAGGAACAGGATGGGAGAGGGAGGTGTGCA  
ATCGCATATCTCCAGTGGTGGTTGAGTCGTTGCAAACTGTTTGC GCGTGGCAACTTCTGGGCAACACAAGTTAACG  
AGACTGGTTAAGGCATTCTTGATTACGATGATCTACTGAAGAAT  
>NODE\_1987\_length\_506\_cov\_76.897232TCGGTGCAGGGCTAAGGGAGGTGGATGGTCTGAAATTTGTCT  
GGGAAAGAAATAGGATCGCTTGCCGAGATAATCGATTACACGGGTACCACGACACTGCAGCTAGAGCCAAGTCGAT  
GAAGGAGCTCATTGAGAAGATGGATGACGACGGAGCTTGGGTCAATTACGAGCGCGATGAAGGGATGCTGGGACAGT  
CCAACCTAGCCAAACACATTAATCTGATGATTGATCGCGGGTCAACACGTGATACGTTTGGAAAGCGAGCTGAGGAGG  
ATGAGAGAAAGGAAGAATGACTGGACCGGAGATGTAGAGGACGCACCAATGGAAGGTTTACAATATGTGGCAAACAC  
CGCATCGTGGATCGNN  
CTGTTACCGCGGTTGGACGAGTCTNGACTTGCATTGCCAGACTGCTGCCGCGGTTGGACGAATCTTTACCATCCA  
CGATGTTGAGTTGATCTGGAAGTG  
>NODE\_1997\_length\_376\_cov\_58.800533TCGCGAGCAGTGGTTGAAGGATATGTTGGGTTATCTGCTGGT  
GAGGAGACGTCCAACAAACTTTTCTATGATTGAGAGTTCGCTTGGGTATTTCGACGTGCTTGTACTGTCAAGACTG  
CTTTGAGATTGTTCTTTAATAAGGAAGTACCTGGGGTATCAAGGAGATAAAGCCGATATGTCGGATAGAGTCTGTT  
GATCCCTTTTCATGAGCTATCTTCTTACTTTTCCAGGTTGAGCTACGAGCTGAGCAAGGGGGGAAAGATTTG  
CCCAGAAATCNN  
TAATAATATTAGTCTACTACTCTATTTATGGGACTAACGCCACTAGAA  
>NODE\_2005\_length\_106\_cov\_67.698112GTGTGCGCAGTTTTCTTTGCTCTGTACACCATCTGCTAAGTTT  
TGTGCGGCTATCCTTGAGTTCGCTTCTTCCCTGGTACTTCCACACTACTTGGTACTCGCTCTCTATCCTATC  
AACAAATTC  
>NODE\_2006\_length\_411\_cov\_63.408760ACCTCCTTGCTTCTCTCATGGTACAATCCTCTTCCCCTGAG  
CCCCCTTTCTCCTCGGTCCACAGAGACAAAATGTAGGACTGGTGCCAACAGTTCATCACGTAAGAAAGGGCTCACTG  
AGCCGGGATCACTAACTAANN  
NNNGCAATCGATCGGATCATATAGATATCCCTTCAACACAACATAGGTCATCGAAAGGGTCTCGGAGACCCACAAA  
GCACGAAAGTCAGGATCTTTTCAGAAAATGGATTCTTATTCGAAGAGTGCATAACCGCATGGATAAGCTCACACTAAC  
CCGTCAATTTGGGATCCACTTCCGGATTTTCTTGGGAGGTATCGGGAAGGAATTGGAATGAAAAAAAAAAAAATCA  
TACAGA  
>NODE\_2016\_length\_3023\_cov\_303.683105ATATTGCTTAAACAGCTCATAAATACAAAACAACAACACACA  
AACAGCTACCAATACCCACAATCATGTCTCAAATCAGTTACGATAACGCCAAATATGCCGAAAAGAAATTTGGCATT  
AACCAAATTAGGGAGGGGGCAAGTGCACCTGCGCAGGGCCAAAGCGCATAGGACAAAGGTGCAGCTATGGAACCTGC  
AACCCGGGATGGCACCCAAACACACAAGGAGAAAATCGTTACTCCTGCTGATGTGCGTGTACAGCTGCATATT  
TCGACAACAACGCTAGCTCAGCGTTGGAGGTTCATATGCGATGACAAATTCACCATCCACTATCAGCTATATGGGGAT  
ATATCTCGTGAAGCAGTTCACGGTCAAACACTATCTCCATATTCAGTGAGGTGCTTTGGGACAGGTGTGAAGTGGC  
GGCCACATTGTATCCAGGCCACTGGATGAAGTTCAAGGCCGCGAGAAATTAGTGGCGGCAGCTAGGGAAGGNCCCT  
GGCCCGCTGGATGAGGTTCAAGGTGCGGAAAACTAGTGGCGGCAGCTAGGGAAGGAATCCCAAACCGAGATGATGT



CAGCAAAGTTACTAGCTGGAATCGTGACAACGTTTCGCCAGCTACAAGATGAGGACCTTGGGATGTTCAAGGTACTAC  
TGGAAAATGTAGTAGTTCGGACAGTCAAAAATAACAAGATTAGTCAAAGGTTTCTCATGCTACTGGAATGCATGGAT  
CGCGGTTCAATCGATGTGGTGTAGATGTCCAAAACCTCTGTTGTGTACAATCCAAGAGATGTTGTAACACAGCTTTAG  
ACATGGCGACAGGTATACGTCTTCAACTCGCGGCCACATCAAACCCGCACACTGCAGTTTTGTGGCGTATGTGCG  
GTGCTTATCCACCACCTGAACNCGTCGTGAACAGTTTCAGACATGGGGACAGGTGTACGTCTTTAACTCTCGACCA  
ACATCCAACCCACACACGGCAGTCTTGTGGCGCATGTGTGGTGCATATCCACCACCTGAACTACAGGNTCTGGGTGC  
ACAGGGAGCCGAAGTCCACTTGACCCAGGCCTTATTTATGCCAGTTTGTATGACATNTGTTTTCAACTCTCGACCGA  
CATCTAACCCACACACGGCGGTTTTGTGGCGCATGTGTGGTGCCTTATCCACCACCTGAACTACAAGGCTCGCACGTC  
ACAGTTCAGCTGATGCTGTACATAATTACATGGTGGTGCAGGGAGTCTAGGTGCGCAGGGAGCTGAAGTCCACCT  
GACCCAGGCCTCATTTATGCCAGTTTGTATGACATACGCCCTTGTATGATGGTGCACCCGAACACCTCCAGCAGGCCT  
TGATAATTGCGTGTCTTTTACAACAAAATCGTTACTTCTCACGTGTTAAGCTACCAGCCGTTGTATCGTCTGTTGAT  
CTGATGGTACCGGGCTTTCTAAAACCAACACAAGGCTCGACAAGCCTATTGTTTCAAGAGAGATGGCTGTATCCAT  
CGGCCGCTTACACCAAATGCTTTTTATTTCATGAACATCAAAGACATTCTCACCTCGGCGGAGCTTAGTACTGCAGCTG  
GGTTTGACCCATCCGTGAGTATGAGGGCTTATTTGAGCTCGCAGGCAACGTTGATCAAACAAATGGCAGGAGATATA  
TCAAGCCTGAATCTCTGGAAGCTACACTGAAGATGAAAGTGCACGAACNACGTTGATCAAGCAGATGGCAGGAGAT  
ATATCAAGTTTAAATCTTTTGAAGCCAGCTGAAGATGAAAGTGCATGAACTACTGAGAGAAGATGACTTCAGTGA  
TCTGCTGAACATATCTGTTTTTGAAGGGCTCTGGTTATGTGATGAAAACACTCGTAGCGTTGACAACGGCGTGTAG  
CAGCTCTATCAAATGGCATTTCGACATGAGTGACAATGTAANTAACGGCGTGTGGCAGCTTTATCCAATGGTATC  
TCCGACATGAGTGACAATGTAACATCATTTCGATGTGGTCAAAGGGGAAATGGACTTAGGCAAAAATAGTCTACAAGGA  
AAGGGACATACCGCATGGTTTCAATTCAGGTTAGGTGGGGGATGTGTGACCAAAACGGACTGTAGTCCGGGAGGCCCC  
AGCACGTAAACGTGTTTTGAAGCCCATGACTTTGGTCCACCCTTGTAAACAAGGTTCTGGATCCGCCTTAAAGGGCCA  
CAAGGAAGAAAAGATTTGGACCGTCAACCAATGAAGCCCCGAAGTACATGCTACAAGTCCGGTCCCTCCGATAAGG  
AAAATTACCCCTCTAGACCAGTGGCCATTGCTAAATTAACCTGTCTGATAGGGCAACCCAAATCCACATCTGAAGAAAT  
GTCCAGCGAATCAAGCTACAGGGAGGTAATTGAACCACATTCCCCCTTGGNCAAATCACCCGTATAGAGGCAACTC  
AGTCAACAACCGAGGAAGCATCTAGCGAATCAAGTTACAGGGAGGTAATTGAACCGCATTCCCCCTCTGGAGCAATCT  
CCACCTCCCAACTACGAAAGTCCCAGAGCAGCAAAAGGCTGTACAGCAGCAGTGAAGATCTGTTTACAACCGAGC  
CAAAANTTACGAAAGTCCAGAAAGCAGCAAAAGGTCAGACAGCAGTAGTGAAAAGTCTGTTTTACAACCGGGCCAAAA  
TAGCGCACGAAGCAATGGCAATGTTTACGCAAGCGGTTGGAACAGCAGCTGACTCAAGAAAAATAGACAAAGTCGTC  
GGTGACACAGCTGGCCAAGTGTGAGCCATTGATTCAATAAAGAAAAGTGAAGGCTCGGTGAAAAGTGAACGCA  
GAAGGAAATGGGCTCATTAGCAATGATAGTCAAGAGGTTGGGATTGACTTACAAAGCATCAAAAAGCCAAGGATTTCCG  
ATGATCTCGTGTGACACAATGGAAACAAACGGATCGTTTTATTAATATGTGGGTGATGGGANCGATCTCATGACCACA  
ATGGAAAACAACGGATCATTCAATTAATATGTAGGTGATGGAATGACAGGACAATCGAACATAGCCAGATACATCAA  
CG

>NODE\_2018\_length\_104\_cov\_65.826920GGTTCGCGCTTCAGGACCTTCAACGCTTCAGCGCGCTCTGCCA  
AGCCCATGCCATACTTGTTCGACGCAAATAGTTGTGTGCTAAGAAAAGCTATAAGCTTGGATCGTCAAGGGTTTAGGC  
CCTGGGG

>NODE\_2019\_length\_570\_cov\_122.894737AAGGAGGGATCTTTTATAAACTACACAGGCGAGGGCATGAC  
CGGCCAATCATCCATTGCTAGGCATATCAATGATATGATTGACAAGGGTTCTTCAAGGAATACCTTTGGATCTATCA  
GGAGGGCCATGCTAAAGGACAAAGTGAATGGGAGGGCAATGTTGAAGATGCTCCGAAAAGATGCCTTGCGGTATATC  
GTCAACTCGCCAACATGGTTGAGGTATCTAGAGGGCCACGACGTTGACAGAGAAGAAAAGGATGTTGAATCAAACCG  
CCCAATGAACCAAATACTAGCCGAGNNTCGCAGCAGCAAGGAATTGCTGACAAGCGACGTCAATAAGGTCTACAGGT  
GGATCAGAGATGCAACTCCTAGTTACTTCCCGCTCAACATAACAAGTACGAACTCGACAACGCCGGTCCGCCAGAC  
ACGCTGAGCTTACGGAACAACCCAGCCAGCTTCGGCCAACGGGAAGTCAAGTGAAGGAGACAGTAGGTGGATAATACT  
GGCATGCAGAGTGAATAGGGCCATTTTCTCAAAAGACCTACTCATCAAACTAAACAAAGGAATTTCCCTTACGCCTC  
AGATGATGTACC

>NODE\_2020\_length\_380\_cov\_112.839470CCCGATCGTTGAGAGGTGGCGTTTTCCGCACGTGACCATGTC  
ATGTCAGTTAGGGGCATTAAGATGGTGGTCCGGTTGACATCGTCAATTAAGTCTGTGCCTGATGGCCATCTCCCTGT  
CGTTGCGACACAATTTGAGCGTGCAGCATGTTCAAGGTCGTTCCACGTGCAGACGGAGACATTATCAATGCACTGA  
GAACTATGACGGCCAGGCGAATGCGAGTGTGCAAAACAGGAGAAAATTCGTTAGAGATTCTCCCGCCTCACTGAC  
GCAGTCAACAAAGCCAAGACACTCAAAAATCCATCTTTTCAAGTGTCCGTTTTACCAACTTCCGCATCCATTCTCATT  
TGGTGGTGGGCCACCGCTCCTGAGCGCCCTCAAAAGCACAGGCGCTTGCTG

>NODE\_2041\_length\_294\_cov\_102.731293GTCATGCCTGCTGGAGAAGATTATTTCACTATTCTGCTCGGC  
TATAATTGAGGGTATGTTTCAAGGATGGGGCTGGAGGACAGAGACGGAAATTTGACACAAAAGTGGAGGGGGCTCTACT  
CCGGGTGGCGAGGTACGACGTGGATTAACACGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCTT  
ATAACCGGACAGAGTGTGTTGACTCATGGTGGATCATGGTGGTGTGACCTTGTGCTGCTGAGTCAAGG  
GATGCCTAGGTTTCAAGGATGGGACAAATGCTGTTCAAGG

>NODE\_2047\_length\_200\_cov\_50.025002GACAGGGGTAGGAAACCTTGAAGAAAAGTTTGAAGGCTTGAAG  
GGCTTACATGATGAGAGAGAAAAGAGAGAAGGGAGAGCATAAGCTTGAAGAGAAAACAGAACTAAAATGAAGCGT  
GCTTACAAATGAGGACCGAGCCTCCTTAAATAGGCAAAAAGGAATCTTAAATAGTAACCATGAACAGTAGACCCGT  
GCTTAAATAGTCTTGAATAGTAAAAG

>NODE\_2049\_length\_83\_cov\_92.024094GGGCGAACGTGTGAAGCAAAGCGTCCGATACAGACTGAGGTGA  
CGCGTGCCTAATATATGTGCCTACCTTATCATCCAACCTTTGAGCTCAGAGCGGCATACACCA

>NODE\_2067\_length\_490\_cov\_99.265305GCGCTCTTTTCGGGAGCCGTGCCCTAGGAACATGTTTCATGATT  
TTGTTCTCCATATTGTGTGTTTTGTAAAGTAAAGTTATATTTGTAATGGCAATTCACCTACCCNNNNNNNNNNNNNN  
NNNNNNNNNNNNNNNNATCAGAAAGCCATTTACAGACAGCACGAATCATGTTGTTTCCAGGTCTATGGGTGACCAG  
CACTCTACAGTCAAACCCACCTGCTTTATCGCCATGATCATTGATGACATAGGCCCCACATACGAGAACCACATTGG  
CCCGCACCCAACAGTCAATATTATTGAGCACCTAATGTTTCATCTGCTCCAGTTCCATAACCCCGTACCACATAGTT  
CCACACAATCGAGCATGTTATTGGCAATTTCAACTATGTGACTCTCATCCGCTGCTGATCAAAATGTGGGTGGTC  
AGATCGGTGTCCCAAGCTTCGGTGGCCCCAGCGTACTGACTCGATCTTTGCCCTGGCGGCATTGTACCGGAGTTTGA  
GATGTAGA

>NODE\_2070\_length\_537\_cov\_48.824955TGATTGCCAGGGTGTCTAATATAGAAGAGGCGATTACGTCA  
AGTTCGTAGGGCACAAAGGAATATTATTCCAGGCCTGTGACGGGCTGGATGTAATATAAGACAATGGGACGTTTAC  
GAGCTATCGGACGGGTTTTGTGANN  
NN  
ATATGGAGGCTGCACTAAAAGGTTGATGCTGGGAGATAGTAAGACGAAGAAGTCGAGTGTGGTTGGCAAGCATGGG  
GCTAGCNN  
NNNNNNNNNNNTCTCAAAGCTTCCTTCTCGGATCGAAGCCCGACTCCTTACGCTTCGCTTCAAATTTCAAAGCGTA  
GTGGTGCCTATGGGTACGGGAAAACGACACTTGCCAGGGAAGAAGGCTGGGTGG

>NODE\_2077\_length\_2602\_cov\_727.339722GAAAAAGTAAGGAATTAGAACTTGAAAAACAAAGATTGG  
GTTGCGCCATACATATGACAGAGTATAAAATAATAATGATTTGGCGAATTAAATACCATGATCTAATAATGAACCA  
TTCTGATTAGTTGATAATATTAGTTGAGAATTTTGTGAAAAGATTCTGTGAAAAGTTTTCATTAACACCTAATTCATG  
TCGAGTAGACCTTGTGTTTGTGTTGTGAAAATTTAATTTCGTGAGTTGTAGGGAGGGACTTATGTCACCACAAACA  
GAGACTAAAGCAAGTGTGGATTCAAAGCCGGTGTAAAGATTACAAATTGACTTATTATACTCCTGAATATGAGAC  
CAAACCTACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAACCTGGAGTTCCACCCGAAGAAGCAGGGGCCGCGG  
TAGCTGCTGAATCTTCTACTGGTACATGGACAACCTGTGTGAACTGATGGACTTACCAGCCTTGATCGTTACAAAGG  
ACGATGCTACCACATCGAGCCCCTTGGCTGGAGAAGAAAGTCAATTTATTGCTTATGTAGCTTACCCTTTAGACCTTT  
TTGAAGAAGGCTCTGTTACTAACATGTTTACTTCCATTGTGGGTAATGTGTTGGGTTCAAAGCTCTTGC GCGCTCT  
ACGTCTAGAGGATCTGCGAATCCCCCTGCTTATTCTAAAACCTTTCCAAGGCCCGCTCATGGCATCCAAGTTGAGA  
GAGATAAATTGAACAAGTATGGTTCGTCCTTATTGGGATGTAATTAACCTAAAATTGGGGTTATCCGCTAAGAAC  
TATGGTAGAGCAGTTTATGAATGTCTCCGCGGTGGACTTGATTTTACTAAAAGATGATGAGAACGTGAACTCCCAACC  
ATTTATGCGTTGGAGAGACCGTTTTCTTATTTTGTGCCGAAGCCATTTTTTAAATCACAGGCTGAAAACAGGTGAAATCA  
AAGGGCATTACTTGAATGCTACTGCAGGTACATGCGAAGAAATGATCAAAAAGGGCTGTATTTGCCAGAGAAATGGGA  
GTTTCTATCGTAATGCATGACTACTTAAACAGGGGGATTACCAGCAATACTAGCTTGGCTCATTATTGCCGAGATAA  
TGGCCTACTTCTTACATCCATCGTGCATGCAATGCAGTTATTGATAGACAGAAGAATCATGGTATGCACTTTCGCTG  
TACTAGATAAAGCCTTACGCTCTGCTGGAGGAGATCATATTACGCCGGTACCGTAGTAACTTGAAGGAGAA  
AGAGAGATCACTTTGGGCTTTGTTGATTATTACGTGATGATTTTGTGAAAAAGACCGAAGTCGCGGTATTTATTT  
CACTCAAGATTGGGTCTCTTACCAGGTGTTCTGCCAGTGGCTTCTGGGGGTATTACGTTTTGGCATAATGCCCTGCTC  
TGACCGAGATCTTTGGAGATGATTCCGTACTACAGTTCCGTGGAGGAACTTTAGGACACCCTTGGGGAAATGCACCG  
GGTGCCGTAGCTAATCGAGTAGCTCTTGAAGCATGTGTACAAGCTCGTAATGAGGGACGTGATCTTGCTCGTGAGGG  
TAATGAAATTATCCGTGCAGCTAGCAAATGGAGTCTGAACTAGCTGCTGCTTGTGAAAGTATNGAAAGGAAATCAA  
TTTGAATTTCCAGCAATGGATACTTTGTAATCCAGTAATTTATTGTTTCCTAAATGAAATGCAATTAACCTCG  
GCCAATCTTTTACTAAAAAAAAGGATTGAGCCGAATACAAAAAGACTATTGTATGTATATATTTTGGATAGATA  
TACTTATCTAGATATACAANACAAAATCTAAGACTAAACAACCTCAAATTTCTATTGTTGTGTGTAGATCCACAATTA  
ATCCTATGGATCCTTAGGATTGGCGTATTCTTTTATATCTTGTAGTTTTCGGCCATATATCGAGCCAAGTATCACAAC  
TCCTTCTACCCCATCTGTATATTGTCCTTTTCACTCCGTGTTGGAATAGAACTTATTATTAGATGATATTTTACG  
AAAAAATTGTTTCATAGGAGAGAACAATACTTCTTTTTTTTCGATGCGAATTTGACACGCCATGGGAGAAACCACTC  
TTTATATTTATAATTGAAAGGGGTTCCGCCGTATCGTATAGTGAAGTTATATCCCAGATTCCCACAAAAATAGAATC  
ATTTCTTTCAATACTCACTCGTTATTAGTTAATAATCCTACTGATTGGATCTATATGCTTATTGTGATAGGAAATGA  
AATATTCAAATGATTTTTTCATCGAATGACTATTCAATTTATTGATTTTAAATGTAATAGGGGGCAAGAAAGCTCTAT  
GGAAAAATGGTGGTTCAATTCGATGTTGTCTAACGAGAAGTTAGAATACAGGTGTGGGCTAAGTAAATCAACGGACA  
GTCCTGATCCTATTGAAAATAACAGTGAAGTGAAGACCGGGTTATAAATAATACGGATAAAAAACATTAATAATTGG  
AGTCAAATTAAGAGTTCTAGTTATAGTAATGTTGATCATTATTTCGGCATCAGGGACATTCGGAATTTTCATCTCTGA  
TGACACTTTTTTAGTTAGGGATAGGAATGGGGTCAGTTATTTTC

>NODE\_2088\_length\_210\_cov\_82.709526GGGCAGCGAGAGCGGGAGCAGAGGTGTCGACGAATGTGGCAG  
AAAGAACGCCGCGATTGAGGAGGGAAGCCGTTGATTTTGAAGAGCCGTACCAAGGACGAGAACCGGGCACCAGCA  
AGACGCACGGACCTGCGTATGTAAGTGGTGGATCAGACGGGTTTGGCTCACGCTCTAGGGCGCGGAGGTGAATTCAGC  
GGCGAGTTTTAAATCTCGGTACCCCGCTCGTCCGA

>NODE\_2093\_length\_507\_cov\_109.660751TAAGGCGTTGAAGACGCTCGAGGGCCCTTATATCAGTGTCA  
ACAAAGGTAACAGCACGGGACCGCTCTGAGGAGAGGAATCATTTCTCTCGACGCGCCCGACCCGAGTATAACGAT  
TGATGCGCCAGTAATGCGAAGTGCACGGAACAACAATTCCTTGTCTACACGGGCATAAGCACCTTGGTCTGACGCA  
GAACCGTGCCGCTGGAGACGCCATGAACTCGTGTCTGTCATCATTTCCAAATTTGGGTGTACCTNNNNNNNNNNNN  
NN  
GTTATCAGATGATGACAAGGCTGAAGAACGGCTGCGGTAAGAATAGGCGAGTGAACCACCTCGACGAGACTTCTG  
ACGTCTGCGCTGAAAAGCAGGAGAGTTAAGTCGCGAATCAATTTGATGGGATCGCCCTCGTTCCATGCCTCACAAATC  
TGACACCTTCTCTGCATTATAGGTTTC

>NODE\_2102\_length\_355\_cov\_75.839439TTTTACAATGGCGTCAGGTGATGCCACCACCCCTCGGGT  
TGGTAGGTTTTGCGTTGGTTAATTCCTGGTCAATTCCATTCCCGGACCTCTATTTACAACAGTCATGAACTTTTTGT  
ACATCGTTGGCGACTCGGAAGTGATGACGGAGCAGATTGAGCGTGGCGACGTTCCAAAAGAACTGGCGGCCATGTTT  
GATTTATCCGGTCCACTTCTCGACTCATGGACAATGACCTCCTGTTTACCAGGGCAGAGGCTATGAATTCGGACT  
GTCTAAGGGGCCAGCACACGCAAGGAATGTGGNTCTCTTGTTCACCAGAGCGGAGGCTTATGAGTTCGGACTGTCCA  
AGGGCCAGCACACGCAAGGAATGTGG  
>NODE\_2103\_length\_617\_cov\_53.435982TGTGGGCATAAGCTGAACGCGGCTTGACAAAATCGGACACCCG  
GCATGCTCACGAGCTCAACGACATCACGTAGGCCNN  
NNNNNNNNNNNNNNNNCGCCTTGTACGGTTTGACCTACGCCCGTGAATTCCACACAGTGAAGGAAGAAATGAAGTCC  
CAAACACCAAGCACAGCAGTGAAGCGCGGACANNN  
NNNNNNNGCCGATGGCTTCCCAGACATAGTTGATCCACCCCAACGTACACAACCTGAACCAGGCCCTGATGACGTGT  
AGACCTCGGAGAGNN  
NN  
NNNNNNNNNNNNNNNNNNNNCAACATCAACGGCGACATCAATCTTTGCTACTGCCAGGCCGGCAATGACGAGGTCAG  
GATCAAGGTCGCCTTCCCAGTGCACGGCAAAAAGCCGTTGACGTGGGTCTTACCACG  
>NODE\_2105\_length\_1182\_cov\_60.683586TGCGGTCCAGAATCAAAAAAATAAGAGTTTGGGGCATTAAA  
TAGCTTGTATCCATAGAACATCTGGCGTAACATAGATAATAAAATAAGGAGTTAATATCATTCCAATTGCCATTA  
CAAAAGTAATTAGTATTTTTGGCATTAAAAGAAATTTTTGGCTGGTAATTATTCCAAAAAAGACTATCAATTTCTGCA  
ACAAAACCGCTCATGCCCGGTAATGCAAGGGAAGCCATCGATAAGATACTGAACATTGTGAATATTTTTGGCATTGG  
GATAGCTATTTCCGCCATCTCGTCGAGATAAATAAGACGTATTCTATCATAACTGGTTCCTGCCAAGAAAAAAGCG  
CAGCACCAAGAAATCCATGAGAGATTATTTGTAATAAGGCTCCGTTGAGTCCCGTATCGGTTATAGAACCAATTCCT  
ATAATTATGAAACCCATATGAGATACAGAGGAATAGGCTATTCTTTTTTTTAAATTACGTGACCAGAAGATGTTGA  
AGCTGCATAGATTATTTGCATTGTACCTACTACTACCAACCAGGGAGAAAATATAGAATGAGCGTGGGGTAATAATT  
CCATATTGATCCGAACCAACCATACGCTCCCATTTTTAATAAGATTCCAGCTAGAAGCATACAAGTACTATAATGT  
GCTTCTCCATGGGTGTCTGGTAACCATGTATGTAGTGGGATAATCGGTTNNNNNNNNNNNNNATAAAAAATCCAATATA  
GAATATTATTTCCAATGCCACAGGATACGATTGATTAGCCGATGTTTCAAAATTTAATGTTGGTTTCGTTGGAACCAT  
ATAAGCCGATACCCAGAACCCCCAGTAAGAGAAAAATGGAACCTCCTGCAGTGTACAAAATAAACTTTGTAGCTGAG  
TACAGACGTTTTCTTTCCCCCCCCACATAGATAGAAGTAGATAAAACCGGAATTAATTCTAACTCCCACATGATGAAAA  
AAGTAAAAGGTCCCAGAGAAGAAAATGATCCTATTTGACCGCTGTACATTGCTAACATCAGGAAAATGGAATAATCGGG  
AATCTCGAGTAACTGGCCAAGCCGCTAAAGTAGCTAAAGTGGTGATAAATCCCGTCAGTAAAATGGGTCCATAGAC  
AGTCCATCTATTCTAATCTCCAGTCAAAATCAAAAAAATTGATCCATTTATAATCCTCCACCAGTTGGATTAATGG  
ATCATCTA  
>NODE\_2108\_length\_223\_cov\_41.125561TTTTTTTTTTTTTTTTTTTTTTTTTTGGTATCGCGCTTCTTTACCATG  
GCTGGTCGTGGGCTATACCAATATGTGCAACGTGGCTACGAGACAGTGTATTCTAGGGTCCGGGCACAGAGTGAGGT  
NN  
CGATGTGTTGTCCAACCCTGAATATTGCCTCGCCGCTTCTATGCATAC  
>NODE\_2109\_length\_239\_cov\_113.841003ACTCGCACCAAGCACGTGTTGGTGGCGTTTTTCTACGTGCA  
TTCCAGGGGTTTCGTGACAGCTGCTCATGGAGTTGGCCCGTGAAGGCTAGCACTGCAGGCGCCGGACATAGTGGCTN  
NNNNNNCCACGATTTTCGAGGAAGCGGAAGATGTGACGAAACTGGCATGTTAGTGCTGTTTGCAAGAACGCTCCACG  
ACACGCGAGATCAGCGCGACCTCATAGGGCATGGTATGGGTTAAACAAGCCACCTTGATCGGTT  
>NODE\_2111\_length\_186\_cov\_65.188171GTGCAAGTCTTTTCGACCTCTAGGAGGAAACACTTCAATCAT  
GTCAGCCGATACATCAAACCTTCTCCGGTTTGTTCAGCTTTGGAGGTATCTCTGCCATGGCAACTGACCACTCAAACA  
TGGATACACCAAAGTCAAGTGAAGTGGCACGTAGCTATGCCTAACGTAATAGTCGACGCACTCTTTAGGCACCATT  
CCGTTCTCAAAC  
>NODE\_2112\_length\_163\_cov\_198.306747ACCCAGACGCGCCTACCAAGAGGGTTGCCCGCATCGAAACC  
CCGCCGAGTCGTAACCACGACGAGTACGCGGTTATAGCGGTCAACGCAATAACATCACCATTTTCAGCCTGTTTCTT  
CGCCAAGCTGCCATAGGCTTCCACACGATACACACCTACCGTGACCGAGTAGGATGGGCAAAATGCAC  
>NODE\_2116\_length\_91\_cov\_136.857147TCCGGAACTTGTACACGGGTGAACCAAAGTCATGGGCTT  
TGAAACGCGTTTGCATCGGGGCTTCCCAGTGCAGTCTGTTTTGGTTACACATACCACCGAACTGTAA  
>NODE\_2119\_length\_97\_cov\_84.288658AAGAATACAGTACATGTCAACCGGATGTACCGGTTGGTGCA  
GAGGACAAGAATCAATCGAGTCCAAGGGCGCCTTTGGCGTACAACACTACGGAGAAAACCTCGAGACAGTAAGACGCA  
>NODE\_2127\_length\_205\_cov\_52.624390CTCAGCCAGTATTATTCTATCAGGGGGCGCCAGCGGCTAGT  
AGATCGACCCCGAGACAGTGAAGAGCTACCTATTAAGCATAACATGTACCCTCTGGGTGCTGTCCANNNNNNNNNNN  
NN  
ATCCTGCACCGTATCCTCCATCCAATCCGTC  
>NODE\_2128\_length\_349\_cov\_37.802292GCGATTTTAAGCGTATCCATACCCTTCGTGCGGGAGCAGTCA  
CCCCTGTGGGTGATGCGACCAGGTGGAACCGACCATCGGATGATANNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
NN  
TGAAGGAAGTGACCGCAAGTTTGTGAGACTCAAGGGTCAAGGACGCATAACACCTGGGCTCTCTCGGTTTCGAT  
GCCGGGGCACAGTGATACGCTATTACCCAAAATAAGTACCCAGCGGAGGGGATGCTGCCCTGCCCAAGGGGGCGAG  
GGCAGATACTAGCCTTGACTA  
>NODE\_2137\_length\_213\_cov\_51.586853TAGGGGACGGGCAAGCCGACACAGATTTATTTTTTCGACGTGT  
TGACTAACAACAAGGCGAGGATAAGGTGGTTAGTTAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN









>NODE\_2333\_length\_88\_cov\_50.977272GGGGAATACCGTTCTCTCGGGGCGACTTCACGCAATGATGAGA  
GATTGGTTCCAACGCTATGCGATTCAATACTGAGGTTTCGGAGAGGCCGGACGACAAATGATCCGCGAG  
>NODE\_2336\_length\_209\_cov\_69.133972GCAACTCATAACCATATTTTCGAGCGTTACCTTTCCATCGTT  
TGGGATCCACAACTTGGAGTAATCCTCCTTATCGTATTTGACCATACTAGGATCATCGTCAGGACCGACAGTA  
AGTTTGCCTGCGGAATGAGTGTAGCTCCAGTGACACGAACCGGCTACGACATCCAGCCCTTAACCGCAGCAACATG  
ATAAGGAAAGACACCGTTCGACGACCTTGATGCCGG  
>NODE\_2340\_length\_285\_cov\_57.431580ATTAATAAAGCGTCCCCGCTGTCGATTTTAAACGGTGTAGGTT  
GGCTTTAACTTCTCGATGTAGCGCTCAACGTTATCTCTCGTCACACCTACCCACCCCTTTTCGGGTTTAAACATACCAA  
TCTCCCCTGCACGCCCTCGCGATAGGCCTTTCCCGCAGCCTTACCTATTAAGTAAGTGTGCTTCTGCTGGTACAA  
ATATCTGGGTTTTGACATAGACTGCCAGCCACTGGAGTACAACCACCGACACCAGACGGGGAAAAACAAATTTGTG  
CTAAATGTGGTGCCAAATCTATACCGACTTCCA  
>NODE\_2355\_length\_241\_cov\_115.199173AAGGCGGGTGAGAAACCCGTCCTCCGAAAGACCAAGGGTTC  
CAGGGCCAGGTTAATCCGCCCTGGGTAAGTCGGGACCTAAGGCGAGGCCGACAGGCGTAGTCGATGGACAACGGGTT  
GATATTTCCCGTACCGGCGAACAACCGCCCAAGCTAATCCAGTGGTGCTAAGAGTCTAACC CGGCTGAGTGGATCCC  
TTCGNNNNNNNNNNNNNNTGGTTGGTGGAGCGTGTGATCCGATCCGGTAGTCAAGCGTATTAACAG  
>NODE\_2366\_length\_214\_cov\_76.929909ACCTTCCCAGCGTGAATCTAAACGCCCTCGTCGATCCAAAGA  
ATGTTGCACCTCGAAACAGGCCTGTGCATCAAGTAGNN  
NN  
AATGGCATACTGACTTGCCCGAACCCGCTATGCCGTCGACG  
>NODE\_2371\_length\_244\_cov\_89.213112CAGCCTGAGTGATATTGGGCATCTCGGCCGTAACCAATGACC  
GCATACCTTGCCTCCCCATAGCGGGAGTTACCTCGGCACGACGTGGGTGTGGCGCNNNNNNNNNNNNNNNNNNNN  
NNNGGTTAT  
ACCCCAAAGCGCCCCCACCACCTTGTGACACTAGCGAGGACACTATGCGAGGTAGCCGAACATCAGCTGC  
>NODE\_2389\_length\_392\_cov\_74.512756AGGCAGTAGCTCAAAGGCTGCCTCTCTGGTCAAGGTTCTCG  
GCCGGCGCCACCAGGTGCTTGTCAAGTACGTGCGCTTGTACGTGCTGGCAAACCCCTTGCGGCGTTCGAGCAAAAT  
CCGGAGTCAGCATTCAATGGCGTCATCGCTCCGGATGAGAAACCACCAGCCGACAGCAACGGCAACAATGGCAGTAC  
TGCCGGCTAGGGCATAGGCCGGCAACTCGCTGAGCTTTGCTGCATACAAGTGTGACGTGCCTAGCTTGCACGTGCCG  
TTGGCGGGGAGTTTATCCGCTAGTTGCCGAGGCCAGCCCCCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
NNNNNNNNNNNTCAGCGTTCGCGCTCCAGATTCTAGTGTCTGATCGCGCTCGGCACATTAGTTCGC  
>NODE\_2406\_length\_90\_cov\_53.133335TTGTCCGCAAAAAGCATAAGTTGAATCCCTTTTGAATTTGACG  
AATTGGGCGGGAAGATCGCGATACATAGCAGCAAGGTGGAATTCGTATATGAAACAAAGGGCACTGGA  
>NODE\_2420\_length\_294\_cov\_54.860546CAACTTCCCTGTTGGCACCAATCTCCGCATCACGCTCGTCCG  
TGACACCATTGTAGCTTTAGATACTAGAGGAGGATCCATCACTGTCATCTCCAGGAACCGTATCATGCCGAGTCCCA  
GATTGGACCCGCACCTTCGTATAGTCAAAGTACGGTAGTACCTAAGCGTTTCATCATAACATTTCTCCCTCTA  
CTGGTGAACATAAGCCAATTGAGGTATGGTGTGTTAAGTAAGCCCTTACGATCAGAGAAACACCTAACGCAATT  
GGTGGTAACAACGAGCTAATGCTCCTCCCTTGCCAAACAACCT  
>NODE\_2426\_length\_273\_cov\_49.578754CTCATAACCAGCTGGAATCATCCCAACACGAGGATAGATTTG  
ACGAATTTCAATTGGTTGGATTGATGCATGGCAGAAATTCTGTAGTTCCGTTGGGNNNNNNNNNNNNNNNNNNNN  
NN  
NN  
TGGAAATGGATTGACACCAAAT  
>NODE\_2430\_length\_260\_cov\_93.484619AAAGGCAATCAAAGGACTAAGCGCAATCAGATGGATTGGCT  
GAGATCGTAGACCCAGTTTGTGTAGGTCCTGGGGTTGGAGTCTGTTGATGGCAATGTGGTCATAGTAATGGTAGACGA  
TTCTGTACTGATTAATCGTATCATTGATGACTGTGGTAGCATCGGGACTTCCGGGAGCGTAAGTCCGGCTCATGGTCA  
ATAGGAGAAGGATCTCCAAGACGGCGACGAACCGGACGTTCCGGGGCGGCCTGTGACGGATTCCAGTTAGCGAAGTA  
ATCGGAAAG  
>NODE\_2435\_length\_92\_cov\_80.206520CCTCAACCTTTGAGATCCTGACTTCTCTCAGGTAAGGGTGCCC  
CAGCTTCAAAGAAACAGTGCTTGAATTTGCGGGCGGAGACACCGAAGGGATCAGTGTCCCGGTCAAGCC  
>NODE\_2443\_length\_898\_cov\_116.067932ATTATTGACACCGTAGACAAGTCCAAATACTACTATTTCAGG  
CATAACGATCACACAGCATAGAGATGTCAAACCACACAACAAAATAAATCAAATTCAAAAACAAAATATCATCA  
ACTACAGACCCACCAACAGCAGAANAACACAGCATTACCCCCGAAACAAGATGAGAAGAAAGAGATCAAACAGAGC  
AGTTCCGGTAAACGAGATGTCACAAGGGAAACAGAAACAGGAAGAACACCAGGCGATTGCAGAGAAGATCGAGAAATCA  
AAAGCAAGTGAACAAGTTGACACTGCATCTTTTCAATTTAGCACAACCTTTAGCGTTAGCAACAGCGATGGAAAAGAT  
GATAGCACTGCAATGAGGCCAACACAGACAGCTGAAATAACTCCGCAACACATAACCAGTCCCGTGCACAAATCAC  
TAGTCTCCCAACTCAAACCGTCCGTCAGGTTGAAGTGCAGTATGTCACAACNNNNNNNNNNNNNNNNNNNNNN  
NN  
AATGGTAGCAGCACCTTACGTTGACATTAGAGAAACCGGATTCAGTGTACAAATTTGATTGCGCAGAAAATGAAAC  
TTCCGCTCACAACATTTGCCTCAACTCACTATCCTGACCCGTTGTTTCCAACCTTTTCAGAAAATCGCTGCAGAAATC  
AACCAGCCTTTGAATTACGGGTATCTCGCGATAGACACAGTCAATGCCGAATTTACAGATGAAGATCGACCGGTGCAA  
TCTAAGGGAATGGTGTGTTGAAGTGAATGGTCTAACTCCCGGTACGATAACCCAATTCACAGGAGGCGCAGTCCCAT  
TTCGAGAAGGCTCAATGAATGGTCCAATCGCA  
>NODE\_2444\_length\_1803\_cov\_705.532471CATAAATGCTCTAGTAGGTATCGAACACAGCATTTTTCAA  
TTTATTTACCAGAACTTTCTCTTGTACCTTGCTTTAACTTGTTTTAACTCTCGTTAAGTTTCGTTGTTCTGTTTT



TGCCTTTCCTAAAAATAACAAAAATCTTCTTTCTTTTCTTCTTTTCTGCTTCGTTNNNNNNNNNNNNNNNNNNNN  
NNNNTGGGCTTTCCTAATTTGAAAAGATAAAATATAAGTTTCTTTCTTTTCTTTTGTGCGTAGTGTGCGTCTTT  
CCTTTAAGTAGCAGTAAGTTCTCTTTTCTTTTGCACCTTTTCTTGTGTGGCGTTAGTTTTTCTAAAAATTA  
AAATCAAAAATATTTCTTTTTTAGATTTGTTTGCCTCCTTAATTTTTCTTTTTTAAATAGTTAAGTTAAC  
GCTTGTTTTAAATTGTCGGTAGTGGTTGTAGTGCCGCTTTTCTAATTTTAAAGTTACAGAGTCCACTTTTTCTTTAT  
TTAATTTTATTAGTATTTTCTTTGTCTTTCTTTCTTACGTAGGGTCTTTTCTTCCCTTTCTGTTGAGTAAGTT  
GCTCTAACGGAATTGTTCTTACCTAAGGTTTAAAGTCTAATTTTTGTGTTACTTACATACGCGGTTTAGTTCTTTT  
CTTTTCTAAGTACTTTTTCTTTTACCTAAGGTCATGGATTACATTGCCCCATTGCGCGTGTTTTCTTTCTCATG  
TTAACTCCACTTTGGAGTTCTTTAGGTACAATAAGAGTAATAGTGACGTCAACGCTTCTCTCAGCTTTTAAAGTTC  
ACCGGTAATGTTAAAGTTACCGAATTTTTCTCTAAAGGCGCAGTACTGATCTATGTAGGTAAGCTCACCAAGGGTGT  
GAAGCGGACCTTTGTCCCCCACCTATTAAGGGTTTTGGTCCGCGAGTACGCCGTTGTAAGCGGTAGCGTGGCCGTTA  
CACGCGCGGATGGTAAGAAGGTACTTTTAGAGGCGCGCACGTCTCTTCCGTTTGTTCGATTCTTCTGAATTCAG  
GCCGATTTTGGAGCACTACCAATGCTTTAATCGTGGTGCCTACTATAGAGTAGTGCCATTTGTGCCAAAGCGAAC  
CTCAAATGCTGTTAGTGAGGTTAAGAGCCAAAAGACGAGCGATGTCTCGGTGAAGAAAGGCTGGGCTCTAAAACCAA  
CATCTGTGGGGTGCATACCCCACTTCCGAAGAAGGCGGAGGTGGTAACCTCAAAGTGTGCTGAACCGGCTTGGCC  
GTGGAGAAAGGGCGAAGATCACGTTCCGGTTGTTCTTCCGTTAAAGGTGGCGGAGATGAAACCGCAGTGTACT  
TAGGAAAGGTGCACTTTTTAATAGAGCTCTAATGTTCCGATTATAGTGCGAAATACCTTTGTTTGGGCAAGGATTT  
GGGATGAAACCACGCGTAGGAGAAGCTACTTCTACGTTAAAGATGGGGTGTTCACCATTTCCCATTATACGGGGT  
AGAGCCACCATAGAGGATTTTATTGTGAACACTACCCAGGCGAGTACAGTTGCTTGGCCGATATTTATTTGTGGTC  
AATGCGGGAGAAGGCGTACGTCCACACCTCAAGAGGGTGGTGTGGTTCAACAACGAGCGCATCAGGGCGGAGATTT  
ATAGGCGCGTGTGTTTCTTCTTCTTCTTCCGTTAGGGTTTTTAAATGCATTTGGGTTTTGCGATGATCAGAGCGGTA  
AGATTTAAGGGCACTAACGTTCTGCACATCCCATCTTCAATGAAGAGCGAATCTATGGCTGGAAGACGGGAGCGT  
ATATCTCCCAATG

>NODE\_2462\_length\_997\_cov\_102.745239CTGGCTTTTGTGTGGACTTCGGCATTTTTACTGCACGCTAT  
TGAATGAACGACTGCTCCAGGTCCGTCAGGTTTCTTTTTGTCTGGCCTTGAGCGGAAAGCACGCCATAGATCGGTGN  
NNNATACCCCATGGCCGAGTAAGCAC  
TCTGATAACGTCCAGTGCAGATGATGGTGAAGGTCAAGTGAATGAAACGACGACGGTTTTGAAGAAAACGCCGNNN  
NNNNNNNNNNNNNNNNNNNNATTTGGAAGTGAAGATCCGGACGACCATCAGATGGTATAACGACAGAAATCACCAACTATGT  
AGCGCATCATGTGCGGGCGATGTAATAAGTATTCCAGCAGACGCAAGTGGCGCTAGTACTGACCTTAGCGGAGACAAT  
GACGAGTAATCCAATGGGTGAGCGGGAGCGTCCCAACGTTTCGTGACTCCAGTGATGACAATGACACCAACAGTGGT  
ACCGGGTGGGCCGGTTCGAAANNGTCAAAACAGTGTTTATCGTCCACCTC  
CGGCGTGGAGAAATAAATATCGCATAAGGGCCATCAAGGAGGAGAACGCCAGGAAACCGTCAACTTCGGCAATGC  
GGTCAAAATNN  
NN  
AGGGGTGTTATGTTTCTGGCCCTGTATAGAAACATTTCCCAAACCTGGACCATGTGCGAGGCAAGTGGACGAGAAAA  
GGTCAAGTCAAGCAAAATCAGGATGGTCAAGTGCACAAGAATAGCAAACTCCATGTCATACCTGCCACCGATGG  
TGGCAAGGTCAAAAGTTTCTTTGGCGTGTGTCTAATGATAGGGTCCGCTGTG

>NODE\_2463\_length\_542\_cov\_37.859779TCGTAGTGCCAAGACCAGTAGGAACCAACACGTAAGTGTGAGAG  
AACAGACTGCCATAAATAGCTCTGATTTTCCATTTTCTTTCATGCTTTGGTGTGCGCCCTGGCTAGATGAACAGGTGC  
CGTCTCAAGCACTTGCCAAATTTNN  
NNNNNNNCATTCAACCATATAGCCTTGTGCTTTTTGCAATGTTCTTCTCTCCAGACATCGATCTCTAATCATCCAN  
NN  
TGCCTCGAGAGTGTCCAAAAGAAGATTTCTCAACTCATTANNN  
NN  
NATATAGCGATTGCGTAAATCATCTTCACTGTGAGACCACAAGTTTTTCGGCCACCTGT

>NODE\_2466\_length\_318\_cov\_60.578617GTTTAGATGAAATGCAAAGGTTTCTAGGTAGTTTGTCCAAAG  
TGGATAGGGTAGTGTACAGCATCATATACCCATCAGAGGTAGAAGCAGGATATTCACAAAGCCTGTTCCNNNNNNNN  
NN  
CTTGGCTGCTGAGATGCTCAAAGACAGAGGATCAAAGGACGCCATGGACGATCACAAGCTCCAAACAATCGGG  
GCTCATCATCTTTCAGCGCCATAAAGGGTAGTTATCTAACGGAGGAATCGTACAAGTACGATAACT

>NODE\_2468\_length\_526\_cov\_207.193924GGATTGCCCTAGTAACGGCGAGTGAAGCGGCAACAGCTCAA  
ATTTGAAATCTGGCTCTTTTAGGGTCCGAGTTGTAATTTGCAGAGGNAGCAATAGCTCAAATTTGAAATCTGGCGCT  
TGCGTCCGAGTTGTAATTTGTAGAGGATGCTTCTGAGTAACCACCGACCTAAGTTTCTTTGGAACAGGACGNGTATGC  
GGTCCGAAAGGTGCTCTATACGTAGCTCCTTGCAGAGTGCAGTTGTTTNTCACAGAGGGTGAAGTCCCGTACGTG  
GTCGCTGGCTATTGCCGTGTAAGGCCCTTCGACGAGTGCAGTTGTTTGGGAATGCAGCTCAAAATGGGTGGTAAAT  
TTCTTCTAAAGCTAAATATCGGCCAGAGACCGATAGCGACAAGTAGAGTGATCGAAAGATGAAAAGCACTTTGGAA  
AGAGAGTCAAACAGCACGTGAAATTTGTTGAAAGGGAAGCGCTTGCAGCCAGACTTGCTTACAGTTGCTCATCCGGT  
TTCTACCCGGTGCCTTCTGTAGGCAGGCCAGCATCAGTTTGG

>NODE\_2491\_length\_254\_cov\_63.208660ACGTGCTGAGACAGTTTCGGTCCCTATCTGCCGTGGGTGTTGG  
AGTATTGAGAGGATTTGTCCCTAGTACGAGAGGACCGNATCACATCCTGGGGCTGGAGCAGGTCCCAAGGGTTTGGC  
TGTTCCGCAATTAAGTGGNN  
NNNNNNNNNNNNCCGTGGGCGTCAAAATTTGAGAGGAGTTGACCTTAGTACGAGAGGACCGGGTTGAACGTACCTCTG  
GTG

>NODE\_2502\_length\_239\_cov\_311.577393GATTACACCCCGCCCATATTGTAATTGCAATTCATAGACAA  
GCTGCCAAAGCATTTCGCAAAGACAACCAAGCCTATCAAATGTGGTTACAGACNNNAGCTCTATTTCAAACCAAG  
TAAAATGTGCGGGACCAACATCATCGTTTCTAGCACGGATTCTGGACTTAGAGGCGTTACCCGTAATCCAGCGCACG  
GTAGCTTCGCGTCATTGCCCGATCGGACAGACGCAAATACCAAATATCTGAATGAGCGGTTCTCTCT  
>NODE\_2505\_length\_79\_cov\_55.177216AGAGTGCAACGTGGACACAAACGAGCGGAAGTTGGTGTAAAG  
GAATGGGCAGAACACAAAGGTTATTTGAAACAATTATGCTTCCAAACAGTGACGAAAT  
>NODE\_2508\_length\_93\_cov\_69.634407CCCAATGAAATTTCTCCGGTGTGACGCTTCTCACGGTGGCCAGC  
TTGTCATCTGCCAACATGGCATCTTCAACGTCAGGGTGCAGCCCCCTCTCCAAACATGATGGCCGGTCCCATT  
>NODE\_2513\_length\_331\_cov\_82.549850TAGGGCGCGTAGTCAGTGGGGGTAGACCGGAAACCTAGTGAG  
CTACCCTTGAGCAGGATGAAGGAGCGGTAACACGCTGTGGAGGTCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
NN  
AGCGTCTGG  
ACTCCGAATGCCAGCAAATTCATACACGGCAGTGAGGCCTGGGGTGTAAAGTCCCCGGCCGAGAGGGAAAGAACC  
AGA  
>NODE\_2527\_length\_584\_cov\_94.361298TTGGCAAAGTACCGACATGTGAAGATAGAGGCATTTGGCGAG  
TGTAATTTTACGAGTAGTGCGAGGAAGTTCCGATTCTTACTAGCAGGACAGNNNNNNNNNNNNNNNNNNNNNNNN  
NN  
NTTGAAGGAAGGATTACCGGTTTTGAGGTTGGTCATGAAGGTGATGAGTCGAGCCATGAGCATGGGATCCAGTTTGT  
CAAGATTCGGTACGCCTTTCCGAAGATATGCGAGATTGANCGTGTGAACTTATCCAGATGAGTTGCTATCTTGGTC  
GTCGCGCTTGGATGGCATCAGAGTTTGGTCAATCCTGTGAGCTTTCTCAGCAATGAAACAATGAGCTCTTTTCGAT  
CTCTTTCACTGTGTCTTCAATGCAGCGTACGGTTTTGACGAATTCGTGAAAAAGCGTTGGAGTGAAGAAAAAGCT  
TTCGATAGACAGCAGGTTGAAGTCCGGAGAAGATCATCTTACAGTGGCGTGGCGTTTTGAATCATGACTGGATGGC  
AATTCATCGAGGATGTGCGGAACTT  
>NODE\_2528\_length\_1006\_cov\_84.276344TTGCTATGACTTTTTACTATCGTCCGACGGTTACTGAGGCT  
TTTGCTTCTGTTCAATACATAATGACTGAAGCGAACTTCGGTTGGTTAATCCGATCAGTTCATCGATGGTCCGCAAG  
TATGATGGTCTAATGATGATCCTTACGATTTTCGTGTATATCTCACCGGTGGATTTAAAAAACCTCCGCAATTGA  
CTTGGGTTACAGGTGTGGTTCTGGGTGTATTGACCGCATCTTTTGGTGTAACTGGTTATTCTTTACCTTGGGACCAA  
ATTGGTTATTGGGCAGTCAAAATTGTAACAGGTGTACCTGAAGCTATTCTGTAAATAGGATCACCTTTGGTAGAGTT  
ATTACGCGGAAGTGCTAGTGTGGGACAATCCACTTTGACCCGTTTTTATAGTTTTACACACTTTTGTATTACCTCTTC  
TTACGGCCGTATTTATGTTAATGCACTTCCCAATGATACGTAAGCAAGGTATTTCTGGTCTTTATAGAGAAGATAG  
ATCATAGATAATTTATAATCAATCATTTACACTACGGGTAGGAAGAATAGTATTTCAATGCTACAAAATATGGATTATT  
GAAAAGAATAAGACATGTATTTGGATCTTTCCCTTCAACTCCACAATATTGATTTTTTATAACGAATAGTTGAAGT  
GAATTTCTCCGAAGAGAAAATGGATTATGGGAGTGTGTGACTTGAACCTATTGATTGGGCTTGCAGATATATACCTTT  
ATCCGCCACATTGAAATTCACAACCAAATGTGTCTTTGTTCTAACCACCGCAAGCCCCATACAGAATAGGCTGGT  
TCGCTTGAAGAGAATCTTTTCTATGATCAGACCCGAATCATGTGCTGCATGAACAGGCTCCGTAAGATCCAGTAGAA  
TAAGTGATGCGTCATGATCCAGATTATGTTTTATCTATTTCACTTACTTAACTTAATAGTAAGTATGGAAATGCATT  
CATTTCTCTGCATCGACCCGATTTATGATACTATCGGAGTGAAAACAAGGGATCTAAAAGAAGA  
>NODE\_2533\_length\_494\_cov\_211.704453CTGGCCACGCTACCAAGATCGTTCATCACAAAACGCCCTCG  
TCATGCGCGGGTCCGGTAGTACCGGCACACCGTCCGTAATGCCATAACGAAATCAATNNNNNNNNNNNNNNNNNN  
NN  
NNNNNNNNNNNNNNNNNNNNNTGAAGGTGCACGGAAGGATAGCCGCGGGCGATGTGCAGAAGGTTTACAACATGTC  
GGAGGAAGTCTTAGCACGAAGAGTTTCCATGACGCGCAAGAGCTGGCGGGATTGGTTCAACACATGGGATTCAACA  
TGGACCTGTATGATCGTGATACGAACCGAGTGACCAGATACGGCACGCATGAAGGGCAGCACACGGTCCAGTTGTC  
AAGAGTGGGAACCACTTCGATGCCGGCGTATACGGGGGCGGTGAAAACAGAAATAACAGTAGCCCACCTAGAGGAGCA  
GAGCATAAGAAGC  
>NODE\_2536\_length\_110\_cov\_100.727272ATGTTGGCGGAGAAAAGTGAAGTTCTTCAGGAACACAATAAC  
AGCAGAAGGGACGTATGCCAGCCCAACAGGGGATTGGCCGATTTATCGCAGGTGATTGGGAAACGTTCTGGCAAGT  
TCACGATACGTGAG  
>NODE\_2540\_length\_100\_cov\_68.629997ACAAGGACCTTACTCGGATACGTGGTCCGCGCGATCAAGC  
GCATCGAGAGAAATACACAGGCCTGCATACAGCGGTGCTGGAGGTGGTGTCCGCGCGCGACAGGCGGATTGAC  
ATG  
>NODE\_2541\_length\_183\_cov\_82.032784TTCTGCAAGAAGGAGAATCAGCACCTGAAGAAGTGACACCT  
GTCAGAGACCCGCGCCAGTGAAGCATGAGAGAGAAGAGGAACCAACAGAGCAAAAATTCGGTAACTACCGTCCCAT  
GATCACCAGTTTAAATCGACCCAGCTTTGTTATCCAAAGTTGAATCTCCACCCATTCCAAAAGCCAACCTCCGCTGTT  
AGAGCCCTC  
>NODE\_2544\_length\_155\_cov\_70.929031AGAAGGTGTACGATGTTTCTCTTGGAGAAGCTGTTCCCTA  
AGAGCATCATAAAGGGTAACGTGCACGCTGAGATTATCGTAAAGAACGGGAAGGTGATGTGGAGCTTGACCTACTT  
GGTCCGGATGGGAGTGTGCTCTCGTGTCTTACGGTAAACCTTACGTGCATGTTGGTT  
>NODE\_2547\_length\_83\_cov\_100.819275GCGACGATGTTTCTCGTATTAGGTTTATAAAATTTTAAAT  
GCTTTTTGAGTGGGTTTTGTTGAGTGCAGCTGATGGCATTGAACTGAAATTAGGGCAGATAT  
>NODE\_2553\_length\_140\_cov\_53.292858GGAAGATGCTAGGAATGCAGTCCGGGCTGGCCAATAAGCTCT  
TTCCCTCTACTCGGTACGGCGCGTGCAGTGAACCGGAGTATTCGATAAAAATAAGGGAAAGCGTCATCGCAAGCAAC  
ACGTTACTGGGAATGCCACTAAGTGGAGAAGAACCGGATTTGG

>NODE\_2559\_length\_112\_cov\_80.919640CAAAGCTGCTCTGTCTCTGTTGCAGTAACCACATTGCACGCC  
TTCTGTACACCAACCCTTGTGCCTTGACACCCGCTGAACACTGCTTTCTCCTTGTCTGCCGACGTGATGCAATAT  
CACTAATAACGCAT  
>NODE\_2560\_length\_991\_cov\_140.327957TTGTAGTTACGCAGGGCATTGGAGGCAACGATTGAATCAAT  
CGTGTCAAGCGTATCTTCCCTCCGCATTGTCATGTTGTGAGATTTCCGCTGATAAATGTGAGTTGTGTCCGTACATGC  
TCCGGGGCTTAACGAGGCGAACATCACCATTGCCAGCAATGCGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
NN  
NN  
GCCGGCGGCAGTAGCAGTTCCGAGCAACCGAAAACAAGTGACGTTCCGTAAGATCGCAACTCAGACAGTCTGACCGA  
CGCNN  
AAATGACGGGTGAGTATGTTGTGCGAGCTACATTAATGCGCTGTAGGAGACCGGCTGAAGTCGGATATTCACGACA  
GGCATTCTGGGGAGACCACGCGTAGGAGCGGATGCCAACAAACATGAGAAAATGTGCGAGACATCATCATGAGATGGTGC  
GTTCTTGTGTTGAACCTGGAAGGTCAATCCCTGAAAGGCCGACGTCTATGCACACATCAATTTTGGCAACTGCAACTG  
CAGCAAGAACTGTGGNAATACTCAGTTCAGTCGCGTTCTGCGCCGCTGCCATAGCACTCTCATTTCAATTTAACTGG  
AGGAAATAGCGACGTTGCGNN  
GGAAACAGATTCAGTCGAATCTGCGTTGCTCTTGGCTCTTCCGGTCTTGGAGCCGCAAGACATCAGCGGGACACGT  
CAATTTATGGGCCAGGCTAAACAAGACACCATGTTGGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
GGGAGGAGAAAGGCTGAACGTAACCGGCTGCATGCATTAATCGTGTC  
>NODE\_2564\_length\_380\_cov\_95.634209ATGCCTGCAGGCCACGGGAAGACGACAATTGCACACAAGTAC  
GGATTTGTGGATGTAGACATGCTAGTAACACGCAAGGAGCATGACGATCTAGCCGAAGACAGGGCGTCGATTCTGGC  
GGGGCAGCAGACATGGATGGAGCACAATACTAAGTGGTATGCCCGGCTCAATCAGACGCTTGATTTGCTAGACTACT  
CTATGCCCGTAATAATCTTGGTTACACGGAAGAAATTGCCCTCAGACTCGGCGGAGGCCCATCGGTNNNNNNNN  
NN  
CATCCAAGCAGGTACCAAACCAGTTTGCATTAACAACAACAGGGACCTGGA  
>NODE\_2583\_length\_289\_cov\_63.487888GGTGTCTCGCACAGATTACCGCGGAAACGCACGTTGATGAC  
GGGGCAATGGCAATGATGGCAGACAACTTGGATAACCACCTCATCGTGGTAGAGCAAGACAGTGTGATAGC  
NN  
NN  
CTCCTGTGCCTGTTCTTTTGTCTTCTCAGGAGTGGAGTGTCTGGAGCTGTAGTACTCACGTCGCGTGTCTATG  
AACTCGTCTATGTCTCCAAAAGGTTTCTTACCCCTATC  
>NODE\_2585\_length\_97\_cov\_60.319588TTTCATCAAGGCCAAGAACGATGGCGTCATGAGCGGATTTTGT  
CTGGGAAACGCAGGGACCGAAGAGTGGGAAAGGTGAGCTGAGAAACCAATCGCGGGCGAGTGGTGGAACTTGTAA  
>NODE\_2604\_length\_83\_cov\_68.590363GCAACAGTCACGCCAATCCCAACCATATCAGAAGTATGACGG  
AAAGAACAGTATGTATCGTGGCCCCGAGAGTATCGCTTCTCGTCTTATCACGCCGTCTGTCGG  
>NODE\_2609\_length\_142\_cov\_74.507042CTCCTCTCATGTTCTGTTCTCCTCTTTGGTCTCACAAGCTGTGGT  
TCAGGTAAGTTACCCATAGACGTAGACTTATCACCGTAATATGTTGTTCCACTTTTGTATGTCTCGAGTTCGGCTTCCCT  
GTTTATGTGTATACATGTGCCTTCAATCACAACAGTTGACTTAGG  
>NODE\_2612\_length\_5804\_cov\_126.984497GTCTCATAAACATCCCCGCGGTGCAAGGCCTTGTTCAGAG  
TCTCCAAATCCTTGGTGTTCGCTTTACGTTAAAGACCTTTATCTCATGAAGAGCGTGTGATGATCCCTTTGAATCC  
TGAGAATCTCCGTTGGACATACTGCGGATCCCCTGCTGAGTACTTTATTCTTGAGCACACCTTCTTATTATTAACAT  
ACAGATCTACACTTAACCTAGAACGCTCGAAATTGACGACTAGCTCCCCGTTACAGCACCCAAATCAACCGGTAGA  
AACGCTGCACCTCTTGCTATGCGATCTAGCAAGTATCCAGTATCAGACTAAGTGGACGAGGGCTGAAACTAANN  
NN  
NN  
CAACACGTGGAGCACGTTATCAATCGACGAGTAACCCAAACTACGCAAACCTCTAACAACCTAATACACTTATCTA  
AAGGCAATGACTCTATGGACTCGCTACCTAACGTAATCAAGACCAAAACGAAGGAAATCAAGCAATTCGGTA  
CTCTCGCCGACATCATTCAACAGAAAGTATCGTCTCCATCGCTATGGGCTCACACTCATCATCACTGAACAACC  
CGGCTATCTTCTTCCGGGAGTTTATGTTTGTGCTCCACAATCTCTCTGATCAACTCCTGCTGAGCATCAACATCAATA  
TTCACATCATATAGGTACTCACTTAAACGGTACCCGAAGCTCGCCTCGATAGCGTAATTCACGATGCATAGGTCCAG  
ATTCCCTGATCTTTGCGCTATCTTCCACCTATCCAGGATCAGATTTGGTCTTTAATGATCCCGTACGGGCTCATCC  
TCCNN  
NN  
ATGCAACTAAAACCATATTGCTCATGTTGTTGAAGAAGAAGGTTCCGAATTCACCGGAGAACCTCATTATGGCCAA  
ATCACCCAAAGCATGCCATCATCAGTTTGGAGGTTTACTACTCCTGTATCATTTCTCGGGCCAGAGGAAGAATC  
TGAGTGCCTCTATCTCCATATCCAGTATCTTCTCATCTGGGATCTATCAAATGCCTCATAATCCGAATCTGTTCCC  
AGCATAACCCTCGACCCAGGTCTTTGACCACTTGTCCAAATCCATGTAGTTCTTCTGGGAGTAGATCATCACATNGCC  
CTCGACCAAGTCTTTGACCACTTGTCCAAAGTCCATGTAGTTTTTCTGGGAGTAGATCATCACATTTTTCAGGGAGGA  
GCTCTCTGAATGCTTTTTCTGTTTGAAGCAGCAAGGGCCCAAACCTGCATAGTACAGAGTGAAGCAAGCTATG  
GTCTGTCCGGCCTTTGCATCCACTCCTCTTTTCTCCATTTTGGTGCAGGTCTGTTGCTNTTGTCCAGCCTTTGCATC  
CACCCCTCTTTTCTCCATTTTCTGTCAGGTCTGCTGCTTGGAGGAAGATCTTGGGTAATTTGCTGGGCCAATCGGGAT  
CACTTCTTCCCGAGTGTGTTCTATCAGCGCGGCGCTCTTTTCCACCCGCTTCTCATGAATTCCTTCTCGCATTTT  
TCAAGCTCGGGCAATGACTCTATATCCTTCCACTGATAAGTCTCTTTGAGCACGGAGAACAATTTGGCGCCCGATCCC  
ATGACAGCTAAGGTACTTTCTCCTATTCCCTTCATAGCTTCTATAACCTCAAACGCTTCTTCCACCGACATCATGAAGG  
TGGCTGTATCATCTGATGATGCCTGAGATATAAAGCTTTGTGGGTATCGGGTTGGCCCCCATTTGTCCGTACCCC  
TCCTCATCTATTTGTTTCAAGTGCGCCAGCCTCGGTGATTTGCTCTCTGTCTCCTTTGCTTCAAGTCAAACGGTAT  
CAACTCGTTTCAAGTCTTCACTTAAACGGTAAGTGAAGTCTTATCGGCTCAAGGGCGACTGGCTCCTCAATTTCCACCT



>NODE\_2650\_length\_226\_cov\_82.805313CCTCTGTGCTTTTCTCTCTCATTTCTCATCCGCCAACACTGCC  
AGATTGATCTTTGTTGCAAGTAAGGTAGCTGCCGAATCTAGGTGCAACGCTGGCTCGTAATGGCTAACAGATCCAT  
TGCTGTGCCAGCATCCCTTAACCTGTGAGCATTGGCAGTACTGTCCGGTTTCATACCGAAGTAGACGAGCTTTGGCT  
CACCTCTCAAGTTCAATATGTTCTCTGCACGCTGGCAACCACATCGCGTTT  
>NODE\_2655\_length\_107\_cov\_69.102806CCCGTACCAATAACAACCTCGCGCTATTGCGCGATTTTAATG  
CCTTTATTACTGAGACACGTGAGGCTTTTGGCATCGCCGATAACAACATCGATGCGCTCTGCGACGCGAACTAGAC  
CAGCTGACCT  
>NODE\_2656\_length\_83\_cov\_61.698795GGGTAAAAAGGTGACCAGCAAGAAAGCGAAGCCACGATGTGGC  
TCGTGCGGAAAGAAAGGACATCAGGCGACCGACTGCAAGTCGAAAAACTAGCAACGCACCGG  
>NODE\_2658\_length\_195\_cov\_53.384617CTTGCCCTCTTGTACTCTCTAAGGAAACAAACCCGCCCCGTAC  
ACTTCTTGAATCAATCTTAACGACCTTACGACAACATGAGCAACTGCAGCGACCTNNNNNNNNNNNNNNNNNNNN  
NNNNNNNNNNNNNNNNNNNNCCGCTCCTGATCGTGACAACGATTTCAATTCAGGACTGGAGTGAAAAATGTCTATGGG  
AATTTATCAGACATCAACGAA  
>NODE\_2660\_length\_409\_cov\_38.474327GAGGCAGCAGTAGGGAATATTGGGCAATGGGCGAGAGCCTGA  
CCCAGCCATGCCGCGTGCAGGAAGAAGGCTTTCTGAGTCGTAACCTGCTTTTGACAGGGANNNNNNNNNNNNNNN  
NN  
GAGGGTGCAGCGTTGTCCGGATTTATTGGGTTTAAAGGGTGCCTAGGCGGTTCTTTTAGTCCNNNNNNNNNNNN  
NN  
NNNNTGGAGGAATACCGATGGCGAAGGCAGCCATCTGGCCTAATACTGACGCTGAGGTACGAAAGCATGGGGAGCAA  
ACAG  
>NODE\_2668\_length\_1027\_cov\_415.327179GAGGAGAATCGAAACCCTGGTTGGTAGGAGCAGGCAGG  
ATTATAATTACGATATAACATAACAACATATATTACACACAANNNNNNNNNNNNNNNNNNNNNNNNNAAGATGAAGTG  
GAAACCTGGACAGAAGAGGAAATTAAGAAACCATAGAACATGATTTCTGGTACGGTTTCAGGTCACGCATAGGCTG  
GTGTCCCACGAAGCCCACATCGTGATGGTCAGCAGTCACACGAGTATGCTCGCGGGCTCCGGCTCACCAATACGGT  
TCATGCCCAAGGAGCTTGGGAGTGGCAACTTCTTCGTCGAGGTTTCGCAATAGTGGAATATAACAACAGACAGCAGCTA  
GTGGACTATCTAACTGGTGAGAACAACCTTAGTTATACCACTTGACGTACGTAACATGCTGCTCAACAGTGATGTTGC  
CGTCATAATAGATGCATACTATTTACTTTTAGTCTATGCAAGACTAGAGCCCTGAGGAGAATCGAAACCCCGTGGG  
GTAGGCTCGTTGGTCATTTATATAAATAGCTTTTGTCTTACATTAACACATACAAACCCATATCCAAACACATACAGA  
ACACAATTTGCAACAAACATGTGAGCAGTCGACAGAGTGGTGGACTACATGTCCGGGTTACCTCAGGGTGCCGAGCA  
TGCGAAAACCTATGCTACCGAGTTTGTGAGAGAACAATGAGAAGCGGAGTGGCCATTTGTGAATGGTGTGACGAAG  
AAAAGAAGATGATCTCGAAAGGGGAGAAGGAGATAATCGGGAGGTTGAACTGCCCGGCTAAGTTACCGGCCACAAT  
TTCGTCAATGACGTCAACCTCGGTGTCTCGGTTTTCGGCTGAGAAGAAGAGATACGTACGGCGTGAAAAACAGCCAA  
GCTCGACTTCTTGGGCTCCAGGGATTGGTTCTTAGAACCCGCAAAATCAAACGCAATGTACTCGGCCAAAGAATCAA  
CAGTGACCTGTTGGACAAGGTTCAAAAACAGAAGATTCTCGCCAAGTGTCCGACAATCTCTGGTGAGTTGGCAAGT  
GATATATT  
>NODE\_2669\_length\_971\_cov\_131.069000TAGATCTAGAGGGAAAATTATGAGCATTACGTTTCATGCATAA  
CTTCCATACCAAGGTTAGCACGGTTGATAATATCAGCCCAAGTATTAATTACACGACCTTGGCTATCAACTACAGAT  
TGTTTGAATTTGAAACCATTTAGGTTGAAAGCCATAGTGCTGATACCTAAAGCAGTGAAACAGATACCTACTACAGG  
CCAAGCAGCTAGGAAGAAATGTAAAGAACGAGAATTATTGAAACTAGCATATTGGAAGATCAATCGGCCAAAATAAC  
CATGAGCAGCCACAATATTATAAGTTTCTTCTCTTGACCGAATCTGTAACTGCATTAGCAGATTCATTTTCTGTG  
GTTTCCCTGATCAAACCTAGANN  
ATGGGTGCATAAGGATGTTGTGCTCAGCCTGGAATACAATCATGAAGTTGAAAGTACCAGAGATTCCTAGAGGCATA  
CCATCAGAAAAGCTTCTTGACCAATTGGGTAGATCAAGAAAACAGCAGCAGCAGCTGCAACAGGAGCTGAATATGC  
AACAGCAATCCAAGGACGCATACCCAGACGGAAACTAAGTTCCCCTCAGCAGCCATGTAACAAGCTACACCAAGTA  
AGAAGTGTAGAACAATTAGCTCATAAGGACCACCATTTGTATAACCATTCATCAACGGATGCTGCTTCCCATATTGGG  
TAAAAGTGCAACCTATAGCTGCAGAAGTAGGAATAATGGCACAGAGATAATATTGTTTCCATAAAAGAAAAA  
AAAAAANNNTCAATAAGGTC  
GGGATCATCAAAACACCAACCATCCAATGTAAAGACGGTTTTTCAAGTGTGCTGGTTATCCAGTTACAGAAGCGACCCCA  
TAGGCTTTTCGCTTTTCGCGTCTCTCTAAA  
>NODE\_2675\_length\_671\_cov\_44.181820CACCCCGCAACCTAACGAACGCTTTAACGTAAGCTTTAGCAT  
CACCGGAATCGATCCGCTCTGCAGTAACTGATGATGCTTTCAGCANNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
NN  
NNNNNNNNNNNNNNNNNNNNNTAAATTGGGTCTAGTGCAGTGGTAATCTTTCTAGTGGTGAGAAATGTCCAGT  
CANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTATAAGTGATCATTATGTCAAGCAAACCTCTAGTATCAGCCTCTC  
TTTCAACCATTTGTCCGCGTTTGTTCAACTACTCTTGCAGCAGTTTCATGGCANNNNNNNNNNNNNNNNNNNN  
NNNTTGTCTATACCTTGA  
ATCTCTTTTGTCTCGATGCGAAAACCTTAGAACTAAAGGATGCTTCATCAGATCTTCAATCTTTCTCATGTCGGGGAA  
TAGGTGAATGTGATCATATCCAATCACGAATCAAACCTCGATTAACTTATCACTGTCTATTTCGGGAGGCAATCGGCG  
TATTCAAATAAGGAAGTATTCGACGTTTCAGATTG  
>NODE\_2679\_length\_295\_cov\_40.837288AACGCTGCTAAAATACCTCCACCGTACGCTTCACTTGCTGAT  
CGAGTGTCAATTCGGGCATCTTCCAATCTCGTCAAGATCATTGCGTTTCGTTTTCGCTTTTCTCCAATGGTGGGTA  
GACTCTAATCTTGAATTTGGTACTCTTTTTCANN

NNGGTGATTCCCATGATTGAAATTTTTGAC  
GGTATTGTTCTAGATTAGTGGGGACTGGTATTGAAAAGTTGGCCG  
>NODE\_2682\_length\_115\_cov\_69.052177TTGATGGTAAGCTCCACTTCGTTGCAGACACACCACCTCGTG  
GCTCTCCAGCAAGTTCAAAGACTGTAGCCCATGGAAGATCAGGTTAGGACAGCGGATGCCCAAGCAAGCCTCCCC  
CCCCGAGAGGAGAGACCT  
>NODE\_2701\_length\_521\_cov\_43.579655AAACACCCATGCCCTAGCACCTAGTGCAACGGCTCCTGGTGC  
AACAACAAGCACCAGTTTACTTGGGGGGGTGGTGTATTAGTAGCAGTGGGTGGCAAAGTAGCTTTGTTGCCTATTC  
CGTTANNN  
ATTTGCTCGTAGCTCTCGTTTGATACCGGATAAAGCAAACCTTGGTTTTCGTTTCCCATGTGATGGACCTGGAAGAG  
GGGGTACATGTCAAGTATCAGCTTGGGATCACGTCTTCTTAGGGCTNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
NN  
ACTTTGCGCAAAGTTCCATTACTATTAATGGGTGGCTCCGCGATTTCTTATGGGCACAGGCATCCCAGGTAATTCAG  
TCTTATGGTTCTTCAATATCCGCATATGGTCTTTTTTTTC  
>NODE\_2702\_length\_673\_cov\_95.799408CGGAGGCCAGAGCACGATGGACAAAGATTGAGCTGGTTACAA  
AGCGTATGAAATGGCACACGCGAACATTGATGTTCAAGCTCACNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
CGAAACTGAACGTAGACAGTCTGTATGGTTTCTTACACTCTGTTCAAGTCCCCTTAATCAACATGCCAGCGCACGA  
TGACATGACGCAAATGGACATACTGGTAGAGCCTCACGACGAAGAAGCGTTTCATGCAACATACACTTTAGCTCCTT  
GGATAGTGGACGAAGTAACGACTAACGCTCTCCCCACAGCTTTCATTTCTTCAAACAAAATCCGAAGGGGAATATTC  
GAAGTCATTGCAAATCATGCATTACATTCTTGCAAAAATCTGCTTTAGATGACTCAGCCACATTGACGGGTGACT  
ATCAAGAAGGGAAGACGCCACATCAACTGCTTTGAGTGCAATGGAACAAAATCGTCACTGAAAATCGCACGCGTTAGGT  
CACCAATGTCGTTGACCAACTCTCAACATNN  
TCCACATGAGATGTTTGTGGTAGACTGGCTCAAATTTTGTCTTTCGCCCTGATTTGTCTGAGAATGCGACGCTGA  
TGTTTCCCCTTACCTCAAGCTACAACCTCACGTGAT  
>NODE\_2709\_length\_393\_cov\_62.432571CCAACCCAGTGGGCCCTGACACCGGACTAATCTCTGCCACGC  
TCTCGATTCTACCAGTTTCCATAGCGCTAGCAGCAAGGTCAGTCCGGACGACGTACGGCTGCTACCCGAGACAAAG  
TCCCCAAGTGCGCCCGCACTATCGCCNN  
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCCAGCGAGCGCAAGCATGTTTCGGATCCTTGGCAAAAATCATGATGAACG  
AAGTCACAGAAGCACAGCCAAGTGANNN  
ATGATGACAGGTGCTCGTGAATTGTAACCGTAGGGGTATTTTAAATCACAAGGCATCCGCCAC  
>NODE\_2726\_length\_231\_cov\_55.948051CAATGAAGGGAGGTAGAGCAGGAAAACATGCAATTACGTGCA  
ACGAGATCGGTTCAAAGGCGATTACGTTGAGTCTANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
NNNNNNNNNNNNNNNNNNNNNTCCAGTAACGGAATGGTACAAGTTTACACCTGATGAGGCAATGAAGTGGGCACACAGT  
GAANCTTGACAATGCGCAAAGGCCTAGAACCAATAATTTTGTGGTGCATTGCTGG  
>NODE\_2732\_length\_80\_cov\_66.037498CTGGGGGTGATAACTTCTCGTCTCTGACCTAACCCAGGTGAATT  
GTAAATAACCTATACACTCATAGGAAGTGGCACAATATTATGCCAGCCCTCCCATTTGGC  
>NODE\_2733\_length\_85\_cov\_55.529411ATTTGTTTCGCCAGATCATTGTTTCATCTGTTTGTGCACCACCA  
GCCACCAATCACCAGTACTATCACAACCACACCCCCGATCACCAGTGCCCGCTGTTCTTGAT  
>NODE\_2743\_length\_207\_cov\_79.086960TGGGGGTGAATGTTTCTTAGTGAATATGATGACCTAGGAATC  
TTACTTTAGTTTGGAAAGAGGTTAATCTTAGTGGCTGACAAGCTTAAACCATTGCTCTTAACAGTCTCAATAAATCTA  
TTGNTTTCGAATGCTGCTCAACAGAATCTGAGTAAATCAAACATCATCAATGTAAACAATAATGAAATCAGAAACA  
AGCCTTTTATGTCTAAATCCCCTTGGAGTTGTT  
>NODE\_2774\_length\_259\_cov\_48.683399CCATAACCTTCTTACACAACCTCCATCAATGCGGCGTTAATCT  
TCATTGCCACCTCAACCGATACTTTCCCTGGCTCGTTAATAACTTTCGGTTNNNNNNNNNNNNNNNNNNNN  
NN  
NNNNNNNNNNNNNNNNNNNNNTACCCCTATTTAGCTTCCAGGTGGTTCCGCCACTTGCCTTGTGGCAGCTTAGGTGGCTCC  
TGTGCCGC  
>NODE\_2812\_length\_85\_cov\_98.129410CCATATGATCACTGGATATGACCCATTTGAATTTGTCTGACCTG  
ACACAGCGACTTTGGGGCTATAACACCAAGTCAATCCCTCCACACGTTGCCACAATCGTGATTC  
>NODE\_2823\_length\_101\_cov\_69.346535CTAGACATGGTTATGAAAATTGAACTCATTTTAGGGTTTTTG  
AACTAGAGCCTAATAGCAGATTGACAGACCTTATGAGACGTCACCCCAACCTCTTACAGCTAACAAACGCCTATC  
CAG  
>NODE\_2824\_length\_206\_cov\_63.116505GCAGATATCTAATGTCTGTTCGACGATACCGATTTTAGATG  
TTCGGTGGCTTGGCGGCTAAAGTGTATTGTGATATACTGTCTGGTAATAGTGGCTACGATTNNNNNNNNNNNN  
NN  
TAATCTTCCGAAGGGTAGTCTTATACTTGC  
>NODE\_2828\_length\_470\_cov\_62.146809TGATATTTTCAACCCGGCTCGTGCCGTAAGTGGTGTGAAGC  
AATGACGCGTCATCGATTCCGTGATCTCTAGGCATTGCCATGGGTATGCAAAGTCGGTTACAAAGGCAGTATATATA  
CCAACGAACGTCATCGAGTGTTCGGAAGTGAAGANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
NNNNNNNNNNNNNNNNNNNNCTCAATGAGTACTCTCTCGCCTGGCTCCCGTGAAGCGTTGGCAGCCAAGACTACTT  
TCGCACCGACACACTCGACGCACAGCACATCTAGTAATTCGAACNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
NNNNNNNNNNNNAGGTGCAATTTCTTAGCCATGACTACTGACGCCTCCTCAACGGATGTACCAGATGTAATCAGATC  
ATAAACTCCTCCTTCAAATCAGCAAGCACACGGCTGCAGCAACGCGTCGACGTCGATCCAGC

>NODE\_2836\_length\_88\_cov\_59.000000TCTTCTACTTCTTTTCAGCGTCTTATCGATTTCGCCAGTTTCTA  
GAATTTCTTCAAGCAATCTCTCATAACATCTCTTCTGAGCGGCTTCGTAATCTATCTATCTACCCA  
>NODE\_2851\_length\_94\_cov\_56.053192AAAAATCGTGAGTCTAGACGAAAGATGCTGTACTCGTTTGAATT  
CTCTGCTATGGGGTAAGAGCGCCACCCGGGAGAAAACGGATGTGAGCAATTTCTGCTAGACTCACGATTTTT  
>NODE\_2862\_length\_371\_cov\_55.204853CCCCGAAGCGTACAATACACTTATGTTGAGGAAGAAGTAGCC  
CCGTGAGCGGCTGTGATGGAAAGACAAAGCGATCCGTGCGGGTCTAATGCCGCCGGCACACGGCTTCTGATGAGAG  
CGTACAAAGAGGCGTTCGATGGTGTCTGTTTCATCAGCAATTTAGTGGGGAATTTGGCTTGTGAAGTTNAGGCAGACGAG  
AGGGGTTGTGTCCGAAAAGCGACGTCCGTGAGCACTGGTTTCAGGAAAGGGGAAGAACCCTTGACGACACGCAGCG  
CAAGCTGCGGTTAACGAAAGTGACGGGGAACCTGCTCAGTCTAGTTTTGAGAGTGCACCACCAGCAGATGTTCCGAA  
GGTCCGCCGGCTGAAGTGCCTGCTCAGAAAAGAGTGAAGAA  
>NODE\_2866\_length\_494\_cov\_74.949394GAGTGACGTCCAAATCGGAGAGCGTTGACCACGCTGATCTGA  
CGTACCGAATAACACAGAACATACCCCTGCCGACCGAATGCCGGCCGTGACTGAGCACCGAATTTGTCATCGACGCT  
GACGGGTTTACTCCAATGAGCGGGNACTAGTGCCATGCTACTTGAGAAGAATCCTGGCGCTTTGTCACACTCATA  
CGGCTTGTGTGCGCCACCGAACGCCGTATCCATGCAATAGCTCAGCGGTACCATGCTATTAACCTATTGCAGTGGGG  
CAGTCGCTTGACAATGTCCGCCATAAGTTTTGCACTTGCCTCATGTTTTGAAAGGCTGATATCAGCGCTCGACTG  
AGTTGAGTTTGGCTGCGATTGTCCACATCATCTGGTACAGCGGTCAGGAGTACCCAGTACGTGGTATACTGTGAA  
CAACATGTACACCAACTGCAACATGGAGGCCGATGACCTGGTTCATGATATCGAGCGGAAAGTTTACAGTTGACACGA  
GTATGGTCTGGG  
>NODE\_2875\_length\_231\_cov\_33.896103ACGCTATAAAGACCTAGAGGCGAAAAACAAGACCCGTAGGAT  
GGAACTACGATTAACGAATAGAACAGTACGTTTAAAGTCCGAATGACCATGAACANNNNNNNNNNNNNNNNNNN  
NN  
GACAATGAACCGTGCAGGCACATCCAGCGCAAGTCATTTAAACCCGTTGCGACCCCG  
>NODE\_2884\_length\_209\_cov\_58.282295ACTAGCTAGTAAAGGATGTCTTTGGTTTACTAGAGACATCGA  
AATATTGTTTTAGCTCGGTGGAAACAAAATGCTTTTCTCAGGATTTCTCTCAAATAGAAAATAGAGAACGAAGTAACT  
AGAAAGATTCTTATAATCACCTGTTCTAGAGGGATCATCTAGAAAGCCAATCGTTTTGAATGCATTCAGGCAGAAA  
AGCTGACATAGATGTTATGGATCGGATTTTTTTTT  
>NODE\_2888\_length\_154\_cov\_51.636364TACCGACCGAAACCAACGTTTTCGAGAATGGACGTTTGGATG  
CTATCGCCCAAGCTGGGACTGGGTCTTTCCACCTTTATTATTCTTCTTCAGCAATTTTGGTCTCGCTTGGATTGTCC  
AGTGGGATTATTTAATCATAGCAGATGAATCAGCAGCCTTTCCATCTCTGGGAAGGA  
>NODE\_2890\_length\_93\_cov\_64.978493TTTATATTTTAGTTCGGGTACGTGTTGCCTCTTTCCGATTGC  
GTGAAGTTTGTTCCTGATGTGCTGAAAGTGGCTGAGCTTCTGGGAGAACCGGGTCAAGACCCTTTAACCTC  
>NODE\_2896\_length\_339\_cov\_71.383484AGGCTACGGGCCAGGTGCGGGACGATCAGGCAAACTCGTTGA  
TCTTGTCTCATCCTCTCTTGGCTCTGGGTTTAAAGCAAATGCTGGCAACTTCATCATGACCGACATTTNNNNNNNN  
NN  
NNNTCGCTTGCTGCCCGTGGCCTAATGCTCAGTGTGTCACGCAGGCAGCTAATGGAGAAACTTCTTGAGCATCCATC  
GGGCTATCTTGTACGACTATTTGGCTGTATGCCAAGTGGGTGCTATTACACATCACTCCTCAATACGATTGCGAATG  
ACCTCCTCGCA  
>NODE\_2903\_length\_1194\_cov\_56.968174TGGAATGGATTGATCTTTTTTCAGACAACGAATAGCAACCCAA  
AAGAGTAAGCACGTTATAGAGTCAAAGGCGCAACAACAACAACAATCAACAACAACATCAACACAAACCAACAA  
CAATGTCAAACCACTACCCACCAATTTACCCTTCTCAGAAGATGAAACTCAAACACACATACCAGGGAAAANNNNNN  
NN  
AGAATACGAGGATACACCCACGCCACATGAGCAAAATTTCTCGCTCAGCACAGTATTTGCATGGCACCGTAGCTA  
CCGCATTAATTGCAACCTTACCATAACCACGGGTACGAATTTTCATTTCCGGATCAACTACTGAAGGCCGAACCTGTA  
GCTATCGAAGTTCCAGAAGAATTGTTGAATCCGCCAGAAATAAACAATTTACACCTCGACATAGATACCGCCACGGG  
GAGAGTCTTTTTAAACCAAAACGNNN  
NN  
TCAACGAGAGCGAAGTTACGAGGAATGGGCGAATTTGCTTGTGGTGTGGGCGATCATGTATATAAGTGCGGGAA  
ATGTGTTGAAATTGAAAATTAAGAACAACAGAGGTAATGGATTTGAAACTAGACATGTATGCGCATTTTCATTTCGC  
CTTCATTTACGGTATCTCAAGTTGTTGATCCTTATTATGATATGCCATTTTTCNNNNNNNNNNNNNNNNNNNN  
NN  
CGTTGGCACAATCACCACTCACTTTGAAGCACTGGCAACCAACTACTGGCTGTACGCTGCGAAGGAATCATGNNNN  
NN  
NN  
AGATTATCAACCTGTAGTAA  
>NODE\_2908\_length\_372\_cov\_80.134407CGAACTGAAACATCTCAGTAGCTGGAGGAAAAGACATCAACC  
GAGATTCGGTTAGTAGTGGCGAGCGAACCGGACCAGGCCAGNTCAGGCCAGTGCTGGCTTTAGTTTACCGGAACG  
GTCTGGAATGGCCGGCGAGATGGGTGACAGCCCCGTACGGGACGGACGATTGCCAGGACTCGAGTAGGGCGGGACAC  
GTGAAATCCTGTCTGAACATGGGGGGACACCCTCCAAGCCTAAATACTCGTACATGACCGATAGCGAACTAGTACC  
GTGAGGGAAAGGTGAAAAGAACCCCGGGAGGGGAGTGAACAGTTTCTGAAACCCGATGCTTACAAACAGTGGGAGC  
TCAAGGTTCTGCTGAGTGACCGCGTACCTTTTGTATAATGGGT  
>NODE\_2925\_length\_79\_cov\_126.392403GCGTCCGAGTTCTGTGACGAGAAAAGACCCCAATGCATGGGCCA  
GGGTATGGGCCCAATCGCGGACCGCCTGGAAGTACACCAACGTGTTGGACAAAAGCGTT







TCACATAATGTGCAAGCAGGCACAAGCAGGCTTGTGTTTTGAAAAGAAACCAAGGTGACAANNNNNNNNNNCGCAAGTGC  
CTTGCGTGACCTGCGCAGTTGGATCGTGTCTCTCCATTCTGTNTGCCAGTGCTTTGCGTGACCTGCGCAATTGAA  
TCGTGTCTCTCCCATTCTGTCTATCCCCTTGGACTTGATAAGAAGTTCAATAGCTGTAGTGGCATTGGTGACCGTA  
GCGTAACCGGCCGCAACGATACTCTCGAGCGCGACGAACTCGAGCGCACACTGACCTGACAATTGGGCTGAGTTTGA  
ATNAATACTCTCTAGCGCGACAACTCGAGAGCAGATTGACCCGACAATTGGGCTGAATTTGAATTGTTGGACACAA  
CGTCATTGATTGCTGCCGCGCGTCCCCGTCTGTCCCAAGCAGCATCAGCTGTTGCCATCGCGTTAACGAGATCTGCC  
ACATCCCTTGAAGAGGCAACGTCAAATCTGGCCAACAAATCGAACAGCTCTTCCCTCGACTTGAACGCTTTGCCCGC  
CAAGCGCCCATACGACGGATATGGGCTCCTCACCATCCACTCAACGTACGACTCAACCTCCGGTGCAGTGATGGTCT  
GCTGGACAGTTGGATCGGGAAACCCATGGAAAATCAATGTTGCAAGAACCAACGAGTCAGCGTCGACGCCAAGACA  
AGTGACCGTAAACGTGCTAGGTTTTCCAGCCATTGTATGATTGACTCGAGGATGCGTTACAAGCACCAGCGAGTCAGC  
GTCAGACGCCAAG

>NODE\_3209\_length\_3210\_cov\_105.625237TATACACATATAGACTTTTCATGAGTCAACTTTGCTGCGAG  
ACAAGCAGGCTTACTTTTTCTTTTAAAACAGCAAACGATGCCCTCCCGATCCTCCTGGGTACGTACGCCCAGACAGT  
TACACGAGGGCTTACTTAATACAAAGGGCCGATTTTCTCAAGTCCGAGATTTATCGGTGACCTTATCAGTAGCCAG  
CAATAAACTAGCAACTGGCCTTATGGGAAGCGACGNNN  
NNNNNNNNNNNNNNNNNNNGTATGCCACAATAAGCCGTACCTGGGGAGGGAGGTCATCTTTGTGAAGAAATATAGGC  
GGGAGAGGGGTGCAAAATAACCACTGGAAAAAATTACTCCTCAAACAATTGGAATGAGGCANNNNNNNNNNNNNNNNN  
NN  
NNNNNNNNNNNNNNNNNNNGTATCAGTCGACCAGAACCTCTATCCCCTGTTGTTTCTCAACACAGAAGCTGAATGT  
GATGTTGATAAGGTGCTCTTGGTTTTCCCTATTAGCGTTTTTTATAGAAAACGTGTTGAAAGCCTTTGAAGAAGT  
CATTCAATGCTAGACTCGAAGACGTGATGAGCATTGACGATTGATGTTGTCAGCACCAGCGCTATAACATCCGCG  
CCGACGGACGACCAAGGTCCGTTAGAGCAACTCGACACAGTGGCATTGACGTCAGACGTAGCGGAGTTGGTTGGCGC  
TAGAGGGAGAGGTAGGATAAGTTTTCCAGACTCGACGATATCGATTAATGAGGCAGATATTTACGACGAGCGTTACC  
TACCGATTACTGAAGCCTTACAGATAAACGCAAGATTGCGTAGATTGGTCTTTTCGAAAGGGGTAGTCAGACACCG  
CGGGATATGGGAAATTTTATAGTGGCTATGATACAACCTCTTTGTACTCTACTCAACCGTAAAGAACATAAGTGTCAA  
GGATGGTTACAGGGTAGAGACGGAGTTGGGTAATAAAAGAGTTTTCGTTGAGTTATTCGGAGGTAAGGGAGGCTATTT  
TGGATGGGACTTTGGACGCGTCCCCAACCAATACCGTGCATCCCTTAATGAGGTATTTTACCCACACCGCTATAAC  
GCTACTGATAGAAAAGAAGATTCAACCAGCGTACACAGCTCTCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
TTTTGACTTTGCGTTGCTAGACAACAGGTATTACCCAGCCGACATATTA AAAAGCCAATGCNNNNNNNNNNNNNNNN  
NN  
NNNNNGGAAAGCGTGAATAATCGAAGGATGGAGTTTACAGCAGTTCTAGTGTGAGTGCAGGTAATAATCCGAGTGAGAG  
TACCAGTGACTTGAAGCTCACAACTTACGACTTACATTATGATAAATATCTTTGATAAATGTGAGGTGAGGTGCTTTT  
GAGACCGCAGTGGGTTTACCGGTCGTCAGGAGTACTCAACAACTCGGCTTTTATGTTGAGTCTTTATAAGCTG  
TTTTCCGCGGTTTTTAATAAGGAGGGTGAATGATAAGCAACGATGAGGGTCTAGTTTTCAGAGTATATAATGTATT  
CCCACAAATGTGCAAGATATTAACGGGATTAGTGAAGTACAACGCACCGGTTATGTAGAGAACTACTTTGGGCGATG  
GACAAGCCGACACCGATTTATTTTTTTGAGGTTTTGACGATCAATAAATCAAAAAATAAGGTGGCTAGTTGCTAGAGAC  
CATCACGCGTGAATGGGATGTTGAATGATTTAAAGTGGGAGGAGAGTAACAAGGAAAAAGTTCAAGGGGAGAGATAT  
ACNN  
NN  
NNNNGAAAGCTGCTGTGCTCACTGACGCTCGTGAAGTATAAGTAAAGTTTTTGGTGGATAGGTGTGGGTCACC  
GAGGTGCTGCTGCGTTATGAAGCCTTTGATGACTAAGTGTGCAACGCATCAGTAGAAAACAGTCAGGAAAATCG  
TTGGGACTTAATACTCACCGCGGACGCTTCTGTTTTGAGCATTGGAAGAAGCCAAATACCTTCAAGATAGCTTCCGC  
CTTAGAAAAGAGAAAAGGGGAAGTACTATTACAGAGGTAGTTGTGGTTCCGACATAGCGAAAAGTTAGGTATACATTGGT  
GGGGGAGCGCAGAGCAAAAGAGTGCCTGGGCTCTTTGAAACTGCATTGTGAAAGGTAGAGCAGAGGGCGGTGACGTGG  
TGCAGCTTCTCCTTATTTCTTTACTTTGAAGTATAATGGACCTGTGCTTTATCATCGTACAGATACTTTCCGCT  
TTGTATAATAACGACGTCACGGCTCTGTACATCTTGATTAACGCATTTAAGAATGTGTGCGATTCCGCGGGCGAAGC  
GGCGATGAATAATCCGCTGGCGGAGATCAGCTTAATAAAGGCTTACGTGAACACGACCGCAACGACCGATTACATC  
GGACGATACTCACAGAGAGCATAGACTCCGCGTTTGCATGACCAAGTAGGGTGCATTGTGGACATCGCTAGAAAAC  
CTGGTTAAGCATTGAGAAAACGTTTTGGAGATTATTA AAACTATGGAATTGTTTCGAGGTGTGTAGTGGAAAAAGAGG  
GAGTAGAAGGTATCTTAAACACTTGGAGCATCAATGCTTTGGTAAACATATAATGATGACTCAAGTCGGGCTAACAG  
CGATCGACGAGGCGGACGTTGCGGACCAACTATTTGTCAGTAGTAATGGTTACGTAGCGAATTCGCGGTTGGCT  
AGGATGTTGATTTTAACTTTGTGTAACGGAGCACTGTAGTGGCGTTTTGTTGCGAGTCCCTTCCGCCACAGACGTT  
ACCTCTGTCTTTATTAGTGTGTGCGTGTCTACTAGTATTGGTAGTAATTTACCATCGCTAAAAGAGAAATGAGACATTT  
AGAGAAGCCTATTAGAGTAGCAGTACACTATTGCG

>NODE\_3211\_length\_5053\_cov\_91.307739CTCTCCCTCTTTTACGAACCGTAATCTGTCAAGTTTGTGA  
CCAGAGATACTACTGGATTTACAATCCTTATTATTGTGCATTATAGCACCACACTTACACACANTACCATTAAGTAG  
TTTTTTACTATAGCTAGGATAGATGATTATTCCTCATCATCTGAGGTTTCTAATATATTTATATGATACTCGGTCT  
CCACCCACTTGACCCAAGTTCGATAAACAGGAGTCTCGCCCTCCAGGGCTACTCTTCCCTCTTTTACGAACCGGAGT  
CGATCAAGTTTATGACCCGAGATTGTACTAGACCTACAATCTTTATTATTGTGCAATATAGCACCACACTTACACAC  
ANGCGCCTAGCGCGCTACGCTTAGCATACTGACCTACCCGAGAGAAAACGATGGGTCAATCCATCTATATCTCGAC  
AGCTTGCTCCCCGACGCTGGACTGAGCGTTGAACACCCCTTTGCCCTTCAGTGCAGAAAGAGACGAGAGTGCATGG  
ATTGTATAACAGTAGCCTCCTGCAGAGTTAAGGTCTTCAAGTCCATGCCCGACGCAAAAGTCAAAACATAACCTGTGGC  
TCCTTGAAGCCTGACCTAGTCATCTTNN



>NODE\_3224\_length\_865\_cov\_99.323700CAGGGTCCAATACCGGGGTCGATAAAGAGGACGAGCAGTGAAG  
AGAAGTGTCTTTGAGCGAATGCTCGATCTCCTCGAACTTGAAAGGCGAGAAAACTTTCTTTACGTAAAAGNNNNNN  
NN  
GGGAGGAGGCGACTCCGTTGGATTCAACATTGTTGATCAGGAACTTCGAGCGTTGGACTTTCTCTTTGAAGAGGAG  
GGGAAGTCTAACTTCGTAGTTGTTGGGGCGGGTGTTCGACGTCATCATCATTGAAGATGGGGATGTGTAGACGGC  
GAGAAGGGACGTCATTTCTTTGGACGGGTCCAATCGAATTAATCAAAGCGGTGATTAGGACCGGAAAAGTGGTGAAA  
GAAATGGAGGGGACGTTAAGACCACGAGTAATGTCGGGGTCTGACGGCGTGCNNNNNNNNNNNNNNNNNNNNNNNN  
NN  
NGCATTTGGTGACGTCGTTG  
AAGCCGAGTTCACGGGCGATAGGCCAAACGATGAAGCACATCATGCGAAAAGAACATCCAATAGGCGTTCCTAGCGTT  
GATAGTGTGAAGAGTGTAAAGGATAGGTGTCCTCGTTCTCATCACGGACGATGTGACTGTTGGCGGCGAGGTTCTGTG  
CATAAGAATTGAGGACGGAACGGCCAGTGGGGACGTTATCCCAGGGTATGGCGGGATTGGCGTGTCTGCGTTAGCA  
GGTGTGATGTCCGGGTGATCAGACATGGTGTGTTGTTAAGGATAAAAAGTGGGAAAACTTTGCAAAAACAGGGGCG  
>NODE\_3225\_length\_129\_cov\_52.906979CATCCAGGAACCAATGGCGTGAACCGAGGAAAGTCGAGCTCGA  
CTGTTTTTTCCCGGCGCACACGATTGTTGGCCCGCAAGACACGCCCATTCCTAGCTTCAACTCTCTGTCAAACCCG  
TCACGTGTAACTTTGCGGAGCAGTTGAGCCT  
>NODE\_3227\_length\_204\_cov\_27.357843GACGAAACACTTAGGTGTAGCCGTGAGGTAACAATAACAAA  
GAAACCAGAACAGCAAAACAAATAACAAGCCAAACACANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
NN  
GGGAGGGTTCACACAGGCGAAGTAGTATGAA  
>NODE\_3234\_length\_100\_cov\_53.779999AGACTGGGGTCTTGGTGTGCCTATGGAAGATGGTGTCTATCT  
TGGTGTAGATGGTAGCCACAAAAACAGCAGCCCCACCTCGTATTGTTAATGTACCAAAAAGCAGCTAGCATGGGT  
GAG  
>NODE\_3238\_length\_85\_cov\_56.670589CTCGAAGAATATCTCCATGAACCTTGTGTAGTATCTTCCCCA  
TTGTTGTCTTCTCAAATTAGCGCGTTGTACAAATGTGAGAGTGGATTTGTTGAAACCAGCAGTG  
>NODE\_3253\_length\_98\_cov\_53.867348AGAAGTGCAGTGGCTCAGGGCAGAAAAATCCGGTTCATGGCTTT  
AAAACCGGACTATACTCAGATGACGTTTGGAGTGCATAAAAACCGGGACTTCAGAGGACGCAGTAATTGAGTACATGA  
>NODE\_3260\_length\_1230\_cov\_109.134956AATTACAAAATCGAACCAACGCAAAAAAACGAAAAAAGCGT  
TTTCTTCCCGCGCCGGTGGTGTAGAGGACAGCCACTTTGATCGAAAAANTTACACTATTGAACGACACAGTCGCAC  
GCTTGAGTTTCGATACAATTTGACAAAAATTTAAGATTCAACAAGCATAAAGTGCAGAACATAAAAACATATCAGTAGT  
GACAGACAATTACAAAACCGAACACAACAACAATAACACACACACAAAACACGAACAAGATACTAAGCAACACATCA  
GCCATTATGCTATCAAGAGCAGCATTAGAGTGTAAACCTGACTTCGATGGGGGTGACAGCCNTTCGTATGTGTCTGA  
CCCTTTGATGGGTGGAACGGCAACCAAGCTACATATTGATANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
NN  
AGATATGGTGGCTATGATTTNACCTGGCATAAGCAGATGTGCTGAGTGGCACCAGGATGGCAGGATGGCATTGAGGTA  
TCGTACCTGACGAGTATACTTACAGGATACCTGCTCTGGTACCTCCTAGCCGATTATACCGATTGAGCACTATACC  
CATAGGCATAGGCATTATGAGGAACGGTCTACACATGGCAACCCGACCCTGGCCTTGATCTCTTAGTGTGGTATGT  
CGATGAACCGGTGGTACTAACGAGGTTACCTGGGATGCAGCGTCCATCTGTCTCATCAATANNNNNNNNNNNNNNN  
NN  
AGGACTTTACATGCAAAGGCCAATTGTTGGAACCTGGTGTGCTGGCCATGCATTGGGGAGTATACTACATCATCTA  
ACCGGTGAGTTCCATCCTGGTAGAAATTCGATGTTCAAGCACCCAGAGAAAGTTAGGCATCTGGCAGAAAATCCTCTG  
CTATAGGCTGCATCGGCACCTTAGTGTGTGCCGATCCAACAACGCTGGATCCTGCGTTTTCAATATGTGGTATTGCAA  
TCCTAATGGAGGTATTCTTTCTGGCGATTGGTGTGAGCGGCTTCGGCACCTGCCAATGGATGTATATAACCTTCAG  
ATAGCGATGCCGCTCATAGCAGGCCACAAGATTGGTGAAGCCTGGGCTCAGGAGACG  
>NODE\_3262\_length\_271\_cov\_75.538742CGTCGACGATATTTACTCAGCTAATAGTACAACAGTTGTCCA  
TCAGGAGACAAGGTGGATAGGGGGTGTAGACGGCAAGTTTTGCGCGCACTTCAAACCTCACGACCCACTCANATTCTG  
CGCGCATTTCAAACCTCACAACTCATTGTCCAAGAGCGCTGAAGCGTNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
AATGGTTAATGGCATTCTGTGCAATCAAGGGCGAAGTTGGGACTTCGCTCTTTACAGTAGGCGAAGACCGGTGCCCTG  
AGTGGCGACGTGCCAAGCTG  
>NODE\_3325\_length\_96\_cov\_52.500000CGCGTGTGACTGAACTATGACTAGCTCGCAAGCGCTACGCGA  
AGTGCTCGCCGAGGACAGACTGACTCTGTCTGTTGCGCCCTTTGCTGCTCATCTCTATGGCCAGTCCGCGC  
>NODE\_3363\_length\_87\_cov\_55.850574GCATCACCGACGGTGTTCCTGTGCTGCCACTCGAGCGTGGCG  
GGGTGTTTGTGATGGCACAATCTTTGTAGCATGTGCCAGCGTGCCTATGTCCCTTAAAGGGATTATG  
>NODE\_3383\_length\_221\_cov\_31.162895CCACCCAGCCTTCTTCCCTGGCCAGAGTGGTTTTTTCCGTGGC  
CCATTGGAACCAACACCCTTGAAATTTTGCAACGATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
NN  
CTCGCGCATGTCTGACCCATGCCGGTCTTTCACTCCCCCGATTGTC  
>NODE\_3404\_length\_5150\_cov\_263.661163GAAAAATTAGTTGTTTTAAAAATCAAAGAAAAATAGATTAG  
ATGGTAATATGGCTATTATTAATAGAATAATTACGCAAAGGACGCTTTAAGATAACAAGGATCCACTCAGAAACATA  
TGTTTCGATGAGGATCCAGTATCTTTAAGAATTTGCGAGATAGTACTAGAGTAGTAATACATGATGAGGTACATTA  
TTGGGGTTTTGTATGATTTCCAAAATTTTATGCCTCATTAATTGCCCACTAGTGTATAGTATTAATTTACCCGCTG  
AGCTCCATAATGGATACACTTTCTCTGTATCCAGATTTGTACGAGTTTCATTTAAGTCCAGACGGGAAAACATTC  
ACTTGAAGCCAGATGGGAAAGCTTCAGGGTCATACAGGCAACCAGTGAATTCTGGCTCTTGAGCACAAAACAGAC  
CATGGATCTGAATGGGAGGACATGGACCATAACCAAATTTGAAAGCATAGGGTCCCATCACCTCTTTTTGTGACCGC

TTGGAGATCTTGTGACTGAATCTGAGGCCATATATGATGACTATTCTTTGGTGGATCCTCGATTATTCAGCAATCTT  
GAACAAAGAATGCCAAAATTGAGAGCGTCTACATGCATAAGGTTATTCTACTTTGATGGCGCTCAAGAAGCCCGA  
TGCCGCTTCAGCTGTGAGCAAACCTCAGGCAATTAGCTAGAGGGGATGAAACCACTGATGAAATGATATTTAGCGGGG  
TGTTGCAACCCAGATAGCTGAGTTGAAATACTTCAGCGAAGTTGGAGGTTAATAGACCTGAGCAAAATTCATGAAA  
GTTGGGTTTTGCCAGATTCTTTGGCCAACAGGTTGAGTATATTCTGAACAAGAAGTGGTATCATTTAGACAGTTTTAC  
TCAAATGGTAAAACCTGATGGTTAAAGCTGAAGTGGTGATAAAGATGGGTTACAAGCCAGTCGTGGGGGGCAATCAGA  
ATATACTAGTAGAACCACTGGTTAATGGCTTAAGCAACAGATTAATAATCAGCTGGATTACAAGCTCTGGAGGATCTG  
GCCGAACCCGGTGTTCCTGAACGTTCAAGAAGCCCTTACTCTTTTCAATTTATAATGATGGGGAAAGGATGTGTGATGC  
AGGTGATCCGGTCTTGGAAATTGGCTGAATTCAGAAGATGCTTGGGCACCCAAAGAACATTTTTGAACAGAAATGGT  
GTTTTCTACCCCAAAAGCCAGCCGATGGGAGCAACTTCACGTATCAGGAAAGTGGCGAAGAATATATATATCCGTGTG  
TACCTCATAGCAAACCCAGAACCTCATGTTCTACTTTTTAGAGGGGTTGATAGATCTAGAGGAGTACAAACGGATGGT  
GTCTGCACAACCAACTAGGCAGACAGTTGTGACTATGGTGGAAAGAGGTAACCTCAGAAAACCTGAAGAAGAAGAAAAGT  
TTGAGCAAAGCATGGAGAAATTGAGAGAGGTGAGGAACAATTTGTGCCTTATCAAACCAATAGCTGATCACTTTGGT  
TTACCTGTGCCCGTGGTAATAAGCAAAGCAACAGTTGAGATGCCAAACTTCTCAAGGTACCTCAATGAGGAAGGTCT  
GAGCATGCCCGCTTATACATGCTATGTAAGAACAATGAATCTTACACTGAGTATACTGGCTGATTCCGGCTATCTCC  
ATGTGCAGGGAGCTTACAAGCCATTAGGGCTCAGAATAACAGGTGATCACGCCGTTCTTACAGTACATAAAGTCC  
AAAAATGATCCCTCAGATGCATTGATGGTGAATCCTGGTGTGGGTTCAATGGAGATAGAAGTATCATCCACCTATGC  
AGCAAAGCTCCAAGCATCTTTGCAAAAAGGGTTTTACAGGGTTAAAGCTGAACGATTTTTGGGGGTAAATGGACTAATA  
AGAAGAAGCTTGAAGAGTTGAGAATCTGACTGTGAGCACATGTTTTGGGTTTTGCGGGCTCAGGTAAAACCAAGTGGC  
ATTACTCAGATGCTCAAACCTCGGCCATGGAATGGAGGTTGTGGTGGTAAGCCACGAAGAGCTCTCAGTGATGAATG  
GAAGAAAGAGCTTAAAGGATACGGACGTGAAAGTTTTACGTACGAGAATTTCTTCTGAAACATAATCATAAAGCAG  
ATCTCTTGATACTGGATGAAGTGCCACTTTTTACCCCGGGGTACATTGACCTGGTTCATTTTTCAAACCAAGTGGG  
CATATAATGATACTTGGGGACCCTTTTGCAAAACAAGTTACCATGCAGATAGTGATGCCATAACTTTGGCAGAGGTAGA  
GGGTGACATCTTTAAAAGGTTACATCCAACATCTAAGGGTAAATGCGCTTGTGGGTTAACCATAGTACCTTGCAAAT  
TCAATGGGGAGGTACCAAAGTTTTGATTTTTTCAGCAGCCGATCAATTTGAAAGGGGAGAAAGGGGCTCTTCTATAGTAA  
GGGGGTGAAGGTTACAAGTATAATGGGGCCAACATGAATCAAGAGGATGGCTGGACACTTTGGATACTATAGTTGAA  
TCTTGCGGCTACTGTGCAAATGATTTTCGATCATTGTTTTGGTGAAGAATATGTGGCTGGAGGATCAATTTGGTGC  
TTCTGATGATGAACCCATTTATCCATTGGACAACCCAAATATTGACAGTGCAACTGAGCGGTTCTTCCACTTTCCAT  
TGAGCTGCAAGAAAGGTGACACTGCACTAGAGCTGGATGGGGCACAATTTCTTCTTGATGCCAAAATGGGTGCCAAAGA  
AGTCACAAACATGCAGTGAAGGCACATGAACACAGAATCAGCCTAACCTTCAGGAGCACTAGACCATTAGAGTTGAT  
GTCAGTGAAGGAATTAGGTGTGCCATACTCTTTTTAACAAATCGTCTGTGCTGAGCTGAACAAAATATTTTAGGTGTC  
CTTCATTCCGGTCTTGGCCACTTTGAAGTGAAGAGGTTGAAAAAACTGAGCAAAGAGCTTTTAAACATTATGTTTTAGC  
AGGGCCACAGTGGAGGAGGAGCAGAAAAATCTTGATATATGACCACTTTGGCAGGCTCAAGTCTGAGCAGATGT  
GATCCAACGTACTTTGATTCTGGAGCAATGAAATGTTTCAGATGAGGTGGTGATCACAGCCCTAACAGAGCTAGGA  
AAGGGATACACATGTTCTTCAAGGTAAGCAGGAATGATTTGAAGAAATGTGCGTCAGAATTTCTCAGGGAAATCATA  
ATCAGTGGTAAGATCAAGGAAGGGAACATGCTGGCGAAGATTGAGGCCAAGTTGGGGAACTGTAAGTTCACAAGGAA  
AAATGTGCATATAGGATCAGGAAGTGTGAAATTTGAGGAACATCTAAAAGGGGATCCTGGTCTAAAAGCAATGCTCT  
TGATACTAGAGGCCGAGGAGATGGAGCCAGAGTTTCATGCAGCAGGAGGTAGTGCTGAAACCATCCGCACTCATTTG  
GCCCTGAGCACCTTTTACAATGAACAGTTTCGCATCTGAATTTAAAAGCAAAGGAGAACAGGGAGCATTATCTGTATGG  
TACCGGTTATTCCACACAAATTTAGGGATGACAAGGATTCAGAGTACCACCCAGGTCCATCAGCTCCAAGTTCATCT  
ATCTCCATCATACTGCTGATGATGATGTGCTCTTCTATCTGTCAATCAAGAAGCGGTTGAGATTTGCGGACTATGAA  
AAGAACGCAAGATCGTTCAGGCTAAAAGAAAAGCTGGGCCAATCCATATTCAGTGAATTCCTGAAGAGGGCAGATTT  
CATGAACCTCACCTACCCACCAGCAGTGGATGAGACTCAGATGGCTCTTGATTTACAATGAAGAGAATTCAGAAAA  
GTGCAAGCATATTGGAGGCTCATAGTTATCGGTCCGATGCTGATTGGCCAAGCAATTTATCTCAAGCTGTTTCATCAA  
AATCAAGATTGCACAAAGATGGAGAAAAGGGGTTCTGATGCAAAGGCTGGCCAGACCATTGCCTGCTTCTCCCATGC  
TGTTTTATGCAAATTCGGGCCAATTTCTGAGAAAAACTGAGGCTCAATTTGAGGGCGATCTTGCCACCACATGTGTTGA  
TATTTTCTCAAAGAATTATGATGATCTCAACACCTGGAGTAAGAGGTAATTTGATGACTTTAGTGGCACAGATAGC  
GATTATGAAGCTTTTGACAGATCACAAGATGGTGCTATCTTATCTTTTCGAGATATGTCTGCTGAAGCACTTTTTGTG  
GCCTGAAGAGCTCATAGANNN  
NN  
NNNNNATCAACCCATAGCATTGCTGGGGATGATATGGTGGCACCAGGCCAACTCATAGTAAACAATACCATGAGCT  
CCATTTTGAACCAACTTGAGCTTAAAGCAAAGGTGAATTACAGTGATTCACCTTTATTCTGTGGTTGGAGAAATGAGC  
CCATATGGTATTGTTAAGNNNAAGGTAAGCTGG  
AAGACTGTTTGGCCAACATGCCCTTGAGGCAAGTTACGGTTACAGGTTGAGCGAACACCTTTGTGAGCTGAANNNTCT  
TGACGCCTTCCAGGAATTGATTAGGAAGATAGTGATGCTCAAAGGAAAATTTGCCAAAAAAAATAGCATCACTATTCGAT  
GATTTCTGATGATATAGTAAGCTCTGATGAGGATTTAGAATAGCAGTAGATGCAGGTAAGGCAGCTAGTTAGGAGTAA  
GTTTATAGTATTAACCTTTAGCCAATAAATTTGAATAGNGTTGGTGAAGCTCGGTGGAAGTCTAGACTAAGTCAATTAAG  
TATGGGTAGTGCTTATCTAGGTGAGAGTCGTAGTGCTGCAAAACGTAGAGCGGTAAGGTATGTTAGGTGTTATTTGTT  
GTGGTAGATTAGAATGTAATGGTAATGTGCGCACAACTATATACAAGATCAAGTTAAGCAGGCCATAAGAACCCCG  
GCGACTAGGTTTTTAACCAAAAACGAGGGTAATTACAAAAATGCTGCAATCAACTTGGCTCTTAGTGTATCTAGTAA  
AATGGGGAAGTCTGAGTACCCATATAAGTTAAAAA  
>NODE\_3408\_length\_596\_cov\_57.151009GAATGTGGGGTTGCTTGGTGAATAGAGTATCGTCCGGCCTTCT  
TCGTCTGAACAAAACACGTCGCCAAATGTTTTTCCCTGAGCTTCGTGAACGGTCATGACTGGTACTTCCCTTACCC



## Supplementary data 3

BLASTn (bl2seq) pairwise alignment scores for scaffolds aligning to reference sequences from GenBank with an E-value of less than 1e-5. Only dominant strains for each virus species were selected as reference sequences.

Virus	GenBank accession	Strain	Blast (bl2seq) score						
			Scaffold number	Scaffold length (nt)	Max score	Total score	Query coverage	E-value	Max ident
GLRaV-3	EU259806	GP18	399	112	203	203	0%	1.00E-55	100%
			500	105	190	190	0%	7.00E-52	100%
			611	1765	3110	3110	9%	0	99%
			734	705	1236	1236	3%	0	99%
			350	359	632	632	1%	0	99%
			231	518	888	888	2%	0	98%
			687	437	755	755	2%	0	98%
			403	389	663	663	2%	0	98%
			242	295	518	518	1%	3.00E-150	98%
			112	256	435	435	1%	3.00E-125	98%
			292	181	315	315	0%	4.00E-89	98%
			341	106	179	179	0%	1.00E-48	98%
			170	102	172	172	0%	2.00E-46	98%
			38	725	1202	1202	3%	0	97%
			738	698	1189	1189	3%	0	97%
			125	303	486	486	1%	2.00E-140	97%
			169	268	450	450	1%	1.00E-129	97%
			490	272	450	450	1%	1.00E-129	97%
			171	273	446	446	1%	2.00E-128	97%
			486	269	441	441	1%	7.00E-127	97%
			302	207	338	338	1%	4.00E-96	97%
			654	176	297	297	0%	1.00E-83	97%
			407	158	268	268	0%	5.00E-75	97%
			724	155	255	255	0%	3.00E-71	97%
			70	152	250	250	0%	1.00E-69	97%
			280	133	223	223	0%	2.00E-61	97%
			232	128	214	214	0%	7.00E-59	97%
			304	102	167	167	0%	8.00E-45	97%
			139	175	286	286	0%	2.00E-80	96%
			509	169	275	275	0%	4.00E-77	96%
			245	132	212	212	0%	3.00E-58	96%
			725	115	181	181	0%	4.00E-49	96%
			26	197	309	309	1%	2.00E-87	95%
			637	147	226	226	0%	1.00E-62	95%
			648	145	219	219	0%	2.00E-60	95%
			271	101	159	159	0%	1.00E-42	95%
			324	106	156	156	0%	1.00E-41	95%
			314	100	147	147	0%	7.00E-39	95%
			559	294	453	453	1%	1.00E-130	94%
			485	132	197	197	0%	6.00E-54	94%
			61	146	212	212	0%	3.00E-58	93%
			71	112	163	163	0%	1.00E-43	93%
			2215	298	178	237	0%	1.00E-47	91%
			363	167	217	217	0%	8.00E-60	90%
			655	247	333	333	1%	2.00E-94	89%
			1766	232	307	307	1%	8.00E-87	89%
			2774	281	123	209	0%	2.00E-31	89%
1254	270	347	347	1%	1.00E-98	88%			
2420	316	381	381	1%	6.00E-109	87%			
2049	105	131	131	0%	6.00E-34	87%			
415	116	131	131	0%	6.00E-34	87%			
2547	105	127	127	0%	7.00E-33	86%			
2137	235	93.3	175	0%	3.00E-22	86%			
1948	301	331	331	1%	1.00E-93	85%			
343	279	307	307	1%	1.00E-86	84%			
1740	198	217	217	1%	1.00E-59	84%			
209	100	109	109	0%	2.00E-27	84%			
2824	228	98.7	166	0%	8.00E-24	83%			
1191	392	374	374	2%	1.00E-106	81%			
168	363	347	347	1%	2.00E-98	81%			
2340	307	282	282	1%	5.00E-79	81%			
522	147	134	134	0%	7.00E-35	80%			
319	126	120	120	0%	1.00E-30	80%			
3253	120	109	109	0%	2.00E-27	80%			
1861	498	165	327	2%	1.00E-43	79%			
3209	3232	569	1750	14%	2.00E-164	78%			
503	486	383	383	2%	3.00E-109	78%			
431	289	242	242	1%	4.00E-67	78%			

GLRaV-3 Grapevine leafroll-associated virus 3

Virus	GenBank accession	Strain	Scaffold number	Blast (bl2seq) score							
				Scaffold length (nt)	Max score	Total score	Query coverage	E-value	Max ident		
GLRaV-3	EU259806	GP18	1778	236	197	197	1%	1.00E-53	78%		
			677	187	150	150	0%	1.00E-39	78%		
			437	166	141	141	0%	5.00E-37	78%		
			3152	141	120	120	0%	1.00E-30	78%		
			3408	618	152	253	1%	1.00E-39	77%		
			211	129	102	102	0%	3.00E-25	77%		
			2947	112	86	86	0%	2.00E-20	77%		
			259	113	80.6	80.6	0%	1.00E-18	77%		
			146	291	206	206	1%	3.00E-56	76%		
			2862	393	277	277	2%	2.00E-77	75%		
			318	255	178	178	1%	1.00E-47	75%		
			641	231	158	158	1%	1.00E-41	75%		
			708	397	123	123	0%	4.00E-31	75%		
			2444	1825	452	536	6%	3.00E-129	74%		
			303	539	333	333	2%	5.00E-94	74%		
			197	316	201	201	1%	1.00E-54	74%		
			646	188	120	120	1%	2.00E-30	74%		
			440	108	68	68	0%	6.00E-15	74%		
			345	101	66.2	66.2	0%	2.00E-14	74%		
			1082	237	134	134	1%	1.00E-34	73%		
			432	154	89.7	89.7	0%	3.00E-21	73%		
			406	154	87.8	87.8	0%	9.00E-21	73%		
			145	606	320	320	3%	4.00E-90	72%		
			627	915	271	271	2%	3.00E-75	72%		
1243	110	64.4	64.4	0%	7.00E-14	72%					
1997	398	210	210	2%	3.00E-57	71%					
283	376	96.9	96.9	1%	5.00E-23	71%					
565	107	50	50	0%	2.00E-09	70%					
589	247	100	100	1%	2.00E-24	69%					
3465	633	192	192	2%	1.00E-51	68%					
405	338	82.4	82.4	1%	9.00E-19	67%					
249	648	75.2	75.2	1%	3.00E-16	65%					
GRSPaV	AY881626	SG1	192	8624	13040	1.30E+04	98%	0	93%		
GVA	DQ855088	P163-1	972	173	295	295	2%	2.00E-83	97%		
			2612	5826	2248	7540	68%	0	96%		
			3211	5075	964	3211	48%	0	96%		
			2199	103	127	127	1%	3.00E-33	92%		
			2254	235	73.4	141	2%	1.00E-16	80%		
			2466	340	113	113	2%	2.00E-28	74%		
			2259	1005	89.7	166	4%	7.00E-21	73%		
			DQ787959	GTR1-1	2612	5826	2233	7774	67%	0	97%
			972	173	289	289	2%	7.00E-82	97%		
			3211	5075	966	3331	52%	0	96%		
2199	103	113	113	1%	6.00E-29	89%					
2254	235	82.4	146	2%	2.00E-19	79%					
2466	340	104	159	3%	1.00E-25	73%					
2259	1005	62.6	62.6	1%	1.00E-12	69%					
GVE	AB432911	TVP15	3404	5172	673	1657	33%	0	99%		
			1570	322	443	443	10%	4.00E-128	87%		
PcV	NC_007539	Seg 1 ATCC 9480	1351	110	55.4	55.4	1%	7.00E-12	86%		
			2999	226	46.4	46.4	1%	8.00E-09	83%		
			2041	316	87.8	150	5%	4.00E-21	79%		
			1798	3958	206	650	34%	7.00E-56	77%		
			1706	741	73.4	132	7%	2.00E-16	75%		
			2111	208	68	68	2%	2.00E-15	74%		
			3031	475	57.2	57.2	2%	1.00E-11	72%		
			982	136	50	50	2%	4.00E-10	72%		
			2323	667	89.7	89.7	6%	2.00E-21	70%		
			2626	572	109	109	10%	2.00E-27	64%		
			22								
			22								
	NC_007540	Seg 2 ATCC 9480	2299	1040	50	50	3%	3.00E-09	71%		
	NC_007541	Seg 3 ATCC 9480					No hit				
	NC_007542	Seg 4 ATCC 9480	1093	4214	53.6	99	5%	9.00E-10	78%		
			713	166	68	68	4%	1.00E-15	73%		
			922	179	37.4	37.4	4%	3.00E-06	67%		

**GLRaV-3** Grapevine leafroll-associated virus 3, **GRSPaV** Grapevine rupestris stem pitting-associated virus, **GVA** Grapevine virus A, **GVE** Grapevine virus E and **PcV** Penicillium chrysogenum virus



## Supplementary data 4

Symptoms observed in the study vineyard at different observation dates

		Observation date							
		20/11/08*	15/01/09			11/02/09			
Row	Vine	Growth	Growth	Berries	Symptoms	Growth	Berries	Symptoms	Atypical symptoms
1	29	3	3	-	1	3	-	-	Complete red leaves
1	47	3	2	US	1	2	-	1	
2	3	3	3	UR	1	3	-	1	Bubbly leaves
2	8	3	3	-	-	3	-	-	
2	22	1	1	-	1	1	-	3	Complete red leaves
2	34	3	3	-	-	3	-	-	
2	44	3	3	-	-	3	-	-	
3	6	3	3	-	1	3	-	3	
3	17	3	3	-	1	3	-	-	
3	25	1	1	-	1	1	-	1	
3	34	1	1	-	-	1	UR	1	
3	41	3	3	-	1	3	-	-	
4	5	3	3	UR	1	3	-	1	
4	14	3	3	-	-	3	-	-	
4	27	3	3	UR	1	3	-	2	Complete red leaves
4	31	3	3	-	1	1	US	1	
4	39	3	2	UR	-	3	-	1	
5	3	3	1	-	-	1	-	-	
5	11	3	3	-	1	3	-	1	
5	17	3	2	-	2	1	-	2	
5	30	3	3	-	1	3	-	1	
5	37	3	3	-	-	3	-	-	
6	2	3	3	-	1	3	-	1	Complete red leaves/ Bubbly leaves
6	8	3	3	-	1	3	-	1	Only leaf margins roll downwards
6	15	3	2	-	-	1	UR	1	Yellow leaves tips
6	28	3	2	-	2	2	-	2	
6	37	2	2	-	1	2	-	1	
6	41	3	2	-	1	2	-	1	
7	6	3	2	-	1	2	-	1	
7	31	2	1	-	1	2	-	2	
7	42	3	1	-	1	2	-	1	Only red veins
8	1	3	3	UR	2	3	-	2	
8	8	3	3	UR	1	3	-	1	
8	13	3	3	-	1	2	-	1	
8	37	2	2	-	1	2	-	-	
9	8	2	1	-	2	1	US/UR	3	
9	21	3	3	-	1	3	-	1	
9	28	3	2	-	1	2	-	2	
9	29	3	2	-	1	3	-	1	Bubbly leaves
10	3	3	3	-	1	3	-	1	
10	20	3	3	-	1	3	-	1	
10	29	3	2	-	1	2	-	1	
10	42	3	2	-	1	1	-	2	
10	46	3	3	-	2	3	-	2	

\* At 20/11/08 no abnormal berries or disease symptoms were observed

<b>Growth</b>	1	Severely stunted
	2	Mildly Stunted
	3	Vigorous
<b>Berries</b>	US	Uneven berry set
	UR	Uneven ripening
	UN	Unripe
<b>Symptoms</b>	1	Very little symptoms
	2	Mild symptoms
	3	Severe symptoms (red of leaves, downwards rolling of the leaves)
	-	no observed symptoms