ASSESSING THE DIVERSITY OF *CRYPTOSPORIDIUM* SPECIES USING GENOMICS

by

FIIFI AGYABENG-DADZIE

(Under the Direction of Jessica C Kissinger and Travis C Glenn)

ABSTRACT

Cryptosporidium is a diverse genus of parasites that causes severe diarrhea, particularly in infants and immunocompromised individuals in low- to middle-income countries. Due to the challenges of obtaining sufficient pure parasites, the genomic diversity of *Cryptosporidium* species remains poorly understood. Furthermore, there is evidence of multiple subtypes within a single population. This increases the likelihood of genomic recombination events, further contributing to this parasite's complex diversity. This dissertation explores the challenges hindering research on *Cryptosporidium* diversity and proposes potential solutions.

The first chapter reviews the current knowledge, advances, and applications of genomics in *Cryptosporidium* research. Although progress has been made in genomic research, approximately 50% of the *Cryptosporidium* genomic assemblies in the National Center for Biotechnology Information (NCBI) database consist of only two species. The second chapter evaluates Multiple Displacement Amplification (MDA) in association with whole-genome Oxford Nanopore Technology (ONT) sequencing. I assessed its efficiency with varying DNA inputs and sequencing output. ONT rapid libraries were successfully generated from starting DNA inputs as low as 0.025 ng using MDA.

Considering the evidence of mixed infections, the third chapter assesses recombination events within samples from a single host. Single oocysts from a mouse model infected with *C. parvum* and *C. tyzzeri* were sequenced and analyzed. Despite the limited breadth of coverage in the single oocyst assembly, recombination events were identified, indicating genetic exchange between species or subtypes during mixed infection.

In Chapter 4, I conducted a comparative genomic analysis of whole-genome SNP data from *Cryptosporidium parvum* SRA datasets, revealing two major clades: Africa–China and Europe–USA, with minimal differentiation between the latter. Chinese samples (subtype IId) were highly clonal, while African samples showed divergent subgroups and possible gene flow. Samples from the USA showed signatures suggesting recent introduction from Europe, whereas European populations were diverse and recombining. Positive selection signals on chromosome 6, in Africa–USA comparisons, identified genes linked to host and environmental adaptation. These findings highlight the role of geography and evolution in shaping *C. parvum* diversity and provide a framework for surveillance and control.

INDEX WORDS: *Cryptosporidium parvum*, *Cryptosporidium* diversity, Genomic recombination, Multiple displacement amplification, Whole genome sequencing, genomic assembly, Next-generation sequencing, Hybridization capture, Single-nucleotide polymorphism

ASSESSING THE DIVERSITY OF *CRYPTOSPORIDIUM* SPECIES USING GENOMICS

by

FIIFI AGYABENG-DADZIE

BS, Kwame Nkrumah University of Science and Technology, Ghana, 2008

MS, Valdosta State University, 2015

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2025

ASSESSING THE DIVERSITY OF *CRYPTOSPORIDIUM* SPECIES USING GENOMICS

by

FIIFI AGYABENG-DADZIE

Major Professors:	Jessica C Kissinger
	Travis C Glenn
Committee:	Casey Bergman
	Vasant Muralidharan
	John P Wares

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
December 2025

DEDICATION

To my Dad in heaven, thank you for instilling in me the drive to always do my best and to live a life worth living. Although I wish you were here, I know your values, love and wisdom will continue to guide me every day.

To my Mom, for always supporting and believing in me. Thank you for your continued prayers for me. I couldn't have done it without you.

ACKNOWLEDGEMENTS

I want to start by thanking God for His guidance, wisdom, and protection throughout my life and career. To my family and friends, thank you for being there and supporting me through both good times and tough times.

To Dr. John Elder, thank you for believing in me and motivating me to embark on this journey. Dr. Kissinger and Dr. Glenn, I say “Ayikoo”. You had the time and patience to teach me more than I thought I could learn. I will forever be grateful. To the Department of Genetics, thank you for giving me the chance to prove to myself that I can do it. You created an environment for success.

To my future wife, you already know.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ..... **V**

LIST OF TABLES ..... **IX**

LIST OF FIGURES ..... **XI**

CHAPTER 1 ..... **1**

CRYPTOSPORIDIUM* GENOMICS – CURRENT UNDERSTANDING, ADVANCES AND APPLICATIONS **1**

ABSTRACT 2

INTRODUCTION 3

TECHNOLOGICAL ADVANCES HAVE FACILITATED *CRYPTOSPORIDIUM* GENOMICS 4

GENOMIC AND TRANSCRIPTOMIC DATA ARE ABUNDANT YET INCOMPLETE 5

COMPARATIVE GENOMICS OF *CRYPTOSPORIDIUM* SPECIES YIELDS INFORMATIVE INSIGHTS 7

SUBTELOMERIC CHROMOSOMAL REGIONS CONTAIN GENE FAMILIES AND APPEAR TO BE HIGHLY DYNAMIC 9

POPULATION GENOMIC STUDIES PROVIDE INSIGHTS INTO VARIATION, EVOLUTION, AND TRANSMISSION 9

TRANSCRIPTOMICS IN *CRYPTOSPORIDIUM* 10

MAJOR ADVANCES THAT ARE GENERATING ABUNDANT, INFORMATIVE GENOMIC AND TRANSCRIPTOMIC DATA 12

APPLICATIONS TO CONTROL AND PREVENTION 14

GENETIC MARKERS, DIAGNOSTICS, AND SURVEILLANCE 15

CONCLUSIONS 16

ACKNOWLEDGMENTS 17

REFERENCES 18

TABLES 32

FIGURES 36

CHAPTER 2 ..... **37**

EVALUATING THE BENEFITS AND LIMITS OF MULTIPLE-DISPLACEMENT AMPLIFICATION WITH WHOLE-GENOME OXFORD NANOPORE SEQUENCING* **37**

ABSTRACT 37

INTRODUCTION 39

MATERIALS AND METHODS 42

RESULTS 46

DISCUSSION 49

ACKNOWLEDGEMENTS 53

REFERENCES 54

TABLES 57

FIGURES 59

CHAPTER 3 ..... **63**

SINGLE-OOCYSTS SEQUENCING ENABLES DETECTION OF RECOMBINATION IN DRUG-SELECTED *CRYPTOSPORIDIUM* HYBRIDS **63**

ABSTACT 64

INTRODUCTION 65

MATERIALS AND METHODS 66

RESULTS 68

DISCUSSION 70

CONCLUSION 73

REFERENCES 75

TABLES 80

FIGURES 84

CHAPTER 4 ..... **90**

GLOBAL POPULATION STRUCTURE AND EVOLUTIONARY INSIGHTS INTO *CRYPTOSPORIDIUM PARVUM* REVEALED BY WHOLE-GENOME SNP ANALYSIS **90**

ABSTRACT 91
INTRODUCTION 93
METHODS 94
RESULTS 97
DISCUSSION 103
CONCLUSIONS 111
ACKNOWLEDGEMENTS 112
REFERENCES 113
TABLES 125
FIGURES 145

CHAPTER 5 165

CONCLUSION 165

APPENDIX A: SUPPLEMENTAL MATERIAL OF CHAPTER 2 170

APPENDIX B: SUPPLEMENTAL MATERIAL OF CHAPTER 3 178

APPENDIX C: SUPPLEMENTAL MATERIAL OF CHAPTER 4 182

LIST OF TABLES

Page

Table 1.1: Current available *Cryptosporidium* genomic and transcriptomic data 32

Table 2.1: Observed amplification yield increase by sample type..... 57

Table 2.2: Estimated DNA concentration in a single cell of the organisms studied in this project 58

Table 3.1: Whole-genome amplification and T7 endonuclease reaction yield..... 80

Table 3.2: A summary of ONT sequencing read metrics for putative recombinant single oocysts and Illumina sequencing metrics for parental reads..... 81

Table 3.3: Genome coverage summary at 1× and 10× depth thresholds for recombinant single-oocyst samples, bulk-sequenced progeny, and parental strains. 82

Table 3.4: Summary of variant types detected in recombinant single-oocyst genomes. 83

Table 4.1: A summary sequencing coverage metrics for all *Cryptosporidium parvum* SRA dataset..... 125

Table 4.2: A summary of genes located at regions of the chromosomes identified as hotspots for high SNP density (variants per kb > 10) with gene description and pN/pS ratios..... 139

Table 4.3: A summary of genes with a pN/pS ratio greater than 1.5 across all populations, including a functional description of the genes..... 141

Table 4.4: Genes located within the top 5% Fst regions on Chromosome 6 from the Africa–USA population comparison 143

LIST OF FIGURES

Page

Figure 1.1: Genome assembly impacts annotation quality, gene family member estimates, and genetic variation analyses 36

Figure 2.1: Serial dilution test with *E.faecium* sample DNA 69

Figure 2.2: DNA fragment and read size range pre- and post-multiple displacement amplification using size-controlled fragmented DNA..... 60

Figure 2.3: Impact of MDA-generated concatemers on the genome assembly 61

Figure 2.4: Circos plot illustrating a synteny comparison between the reference S. aureus ATCC-29213 genome sequence and pre- and post-amplification genome assemblies 62

Figure 3.1: Fluorescence-activated cell sorting (FACS) of putative recombinant single oocysts based on dual fluorescence..... 84

Figure 3.2: Mapping and genome coverage across recombinant single oocysts, bulk progeny, and parental samples. 85

Figure 3.3: Chromosomal inheritance in recombinant progeny visualized using IGV..... 86

Figure 3.4: A nucleotide-level evidence of a recombination event..... 88

Figure 4.1: Parsing of *Cryptosporidium parvum* samples from NCBI 145

Figure 4.2: A heatmap of SNP density per chromosome..... 148

Figure 4.3: Global population structure of *Cryptosporidium parvum* illustrating the clustering of samples..... 149

Figure 4.4: A pairwise comparison of *Cryptosporidium parvum* samples from around the world 152
Figure 4.5: Diversity within and between populations 154
Figure 4.6: Selection pressures within the various populations 156
Figure 4.7: Linkage Disequilibrium coefficient r^2 as an indicator of association between SNPs 158
Figure 4.8: Distribution of genes and their potential association with population differentiation 160

CHAPTER 1

INTRODUCTION AND REVIEW OF *CRYPTOSPORIDIUM* GENOMICS – CURRENT UNDERSTANDING, ADVANCES AND APPLICATIONS*

Reprinted here with permission from the publisher.

*Agyabeng-Dadzie, F., Xiao, R., & Kissinger, J. C. (2024). *Cryptosporidium* Genomics — Current Understanding, Advances, and Applications. *Current Tropical Medicine Reports*, *11*(2), 92-103.

Author contributions: F.A-D., R.X. and J.C.K wrote and edited the review. F.A-D prepared Table 1. J.C.K made Fig. 1

Abstract

Purpose of Review Here: We highlight the significant contribution that genomics-based approaches have made to *Cryptosporidium* research and the insights these approaches have generated into *Cryptosporidium* biology and transmission.

Recent Findings: There are advances in genomics, genetic manipulation, gene expression, and single-cell technologies. New and better genome sequences have revealed variable sub-telomeric gene families and genes under selection. RNA expression data now include single-cell and post-infection time points. These data have provided insights into the *Cryptosporidium* life cycle and host–pathogen interactions. Antisense and ncRNA transcripts are abundant. The critical role of the dsRNA virus is becoming apparent.

Summary: The community's ability to identify genomic targets in the abundant, yet still lacking, collection of genomic data, combined with their increased ability to assess function via gene knock-out, is revolutionizing the field. Advances in the detection of virulence genes, surveillance, population genomics, recombination studies, and epigenetics are upon us.

Keywords: *Cryptosporidium* · Telomere-to-telomere, T2T · Single-oocyst sequencing · Hybrid capture · Transcriptomics · Population genomics

Introduction

Cryptosporidiosis is a neglected disease caused by apicomplexan parasites in the genus *Cryptosporidium*. It has devastating impacts on the most vulnerable, especially infants and the immunosuppressed [1]. As the scope and significance of this infectious disease has become apparent [2, 3••], the research community has responded. Over the last decade, there have been significant advances in genetics, genomics, the ability to culture the parasite and perform high-throughput screening, host–pathogen interactions, surveillance, and therapeutics. The advances in these areas have yielded large quantities of associated genomic data (Table 1) that are fueling advances in our understanding of *Cryptosporidium*, its life cycle [4, 5], its evolution [6], transmission [7••], and host–pathogen interactions [8, 9, 10••]. These data will also facilitate the design of better surveillance tools for local, regional, and hopefully, global use.

Genomic data, however, are not without their challenges. Historically, *Cryptosporidium* genome sequences have been very hard to generate due to a lack of pure parasite material. This challenge has recently been overcome for genome sequence generation, but not for transcriptomics of postinfection *Cryptosporidium* life cycle stages. Another challenge arises from the fact that most available data were created with short-read sequencing approaches. While Illumina sequencing is highly accurate, it cannot yield complete genome assemblies and poses significant challenges for the analysis of gene families and repetitive sequences. Currently, cloning is also impossible for *Cryptosporidium*. Thus, nearly all genomic and transcriptomic data have been generated using populations of parasites rather than purified isogenic clones. This fact creates considerable challenges for genome assembly, data analysis, and interpretation. Here, we highlight key advances, remaining challenges and future prospects.

Technological Advances Have Facilitated *Cryptosporidium* Genomics

Whole Genome Sequencing

With the advent of high-accuracy second-generation sequencing (Illumina short reads) and the large fragment sequencing capabilities of third-generation sequencing (Pacific Biosciences and Oxford Nanopore Technologies, ONT), *Cryptosporidium* genome sequences are being generated at an increased rate [11]. Currently, there are 74 *Cryptosporidium* genome sequence assemblies located in the NCBI GenBank and more than half have been submitted since 2018. Whole-genome sequences are needed to facilitate the research community's ability to design and interpret their experiments. As additional genome sequences become available for new species and strains, a framework for a more holistic genomic comparative analysis is being constructed. The power of comparative insights is significant [12, 13••, 14••, 15]. For example, the addition of a small number of genome sequences shed considerable insight into the diversity and evolution of a species, *C. parvum*, which revealed the existence of an anthroponotic subclade that was likely shaped via introgression of DNA from other *Cryptosporidium* species and subtypes [6].

Single-Oocyst Sequencing

Genome sequence generation for Cryptosporidium has historically been quite difficult due to the large number of oocysts required for DNA preparation. Oocysts, which contain four haploid sporozoites, have ~ 40 fg of genomic DNA. Most clinical samples do not contain a sufficient number of oocysts to reach the minimum DNA requirements for sequencing library preparation. Thus, important isolates have historically been propagated in immunosuppressed mice or gnotobiotic pigs. This process is difficult, expensive, and time-consuming. However, with the advent of single-oocyst sequencing, the possibility of generating genomic sequences

from a single oocyst is a reality [16]. The protocol involves oocyst sorting, lysis, genome amplification with multiple displacement amplification (MDA), and sequencing with short-read Illumina [16]. This technique has recently been modified to utilize long-read ONT sequencing [17••, 18]. Given the obligately sexual nature of *Cryptosporidium* and the existence of four related haploid sporozoites within an oocyst, single-oocyst sequencing is also a promising technique for studies of diversity within a single infection [16]. Single-oocyst sequencing also creates an avenue for studying diversity and recombination events within a single oocyst [17••].

Hybrid Capture from Fecal DNA Samples

Hybrid capture, i.e., the selective enrichment of particular DNA sequences via hybridization to long, single-stranded RNA probes representing the target genome sequence of interest [19], is an ideal approach for isolating *Cryptosporidium* genomic DNA from fecal DNA samples. Most clinical *Cryptosporidium* fecal DNA samples contain abundant microbial, food, and human DNA content. Recently, hybrid capture has proven tractable for fecal DNA samples with a *Cryptosporidium* qPCR Ct score of < 20 and for much higher values if a double enrichment is performed (Bayona et al., in prep). The use of hybridization capture has made fecal DNA samples accessible for genome sequence generation. This development will permit samples from numerous studies, sitting in freezers to be analyzed. Importantly, the hybrid capture baits can be customized to have a wider sequence divergence range to facilitate detection of less common human-infecting species. Smaller subsets of probes can be tailored for specific regions of the genome to provide a multi-locus approach to quickly screen large numbers of samples and facilitate outbreak investigations.

Genomic and Transcriptomic Data Are Abundant yet Incomplete

As we can see in Table 1, there are over 1000 genomic data sets for *C. parvum* and *C. hominis*; yet, there are only nine assembled and annotated genome sequences. Outside of these two prominent human-infecting species, the situation is bleak. A few dozen genome sequence data sets exist for all other species, and more than half of all named species have no genomic sequence data and are missing from Table 1.

Most existing *Cryptosporidium* genome sequences also present several challenges for the community. Most were generated using only short-read technologies that produce assemblies that contain gaps and compressed sequence regions (Fig. 1). A complete telomere-to-telomere, T2T, chromosomal assembly would contain eight chromosomes. Although the karyotype is unknown for most species, current assemblies contain dozens to hundreds of contigs, few telomere sequences, and many unassembled reads. Genome sequence assembly gaps most often arise in genome regions that contain repetitive sequences, making short reads difficult to place. Long stretches of repetitive sequence also generate gaps, as does the merger of recent gene duplications that reside in multiple locations in the genome (Fig. 1). Diversity within the population of parasites being sequenced can also create gaps because some parasites may possess structural variants like indels and inversions or differences in gene family or repeat copy numbers. When genome sequences contain gaps, it is difficult to know if genes are actually missing, thus posing significant challenges for comparative genomics.

The inclusion of long-read sequencing approaches and hybrid genome assemblies utilizing both long- and short-read approaches is the answer. Long reads, which can reach 100 + kb in length, can cover large genomic regions permitting an exact determination of repeat or gene copy numbers (Fig. 1) and provide proof of genome rearrangements. Depending on

population numbers, contigs for differing genotypes within a population can be obtained, i.e., evidence for parasites that have two vs three copies of a particular gene in the same isolate.

When looking at RNA sequence data, the landscape is barren (Table 1), and most existing data sets are a combination of host and parasite transcripts since purification of post-infection parasites remains tenuous at best [20]. There are 145 RNA data sets for *C. parvum*, four for *C. hominis*, and one for *C. baileyi*. This paucity of data has significant consequences for the community. It means that the few genome sequences that have annotation have had to rely on orthology and de novo gene prediction alone. This means that species-specific genes are very difficult to discover, ncRNA genes will be missed, and untranslated regions (UTRs) will be unannotated making it difficult to know where promoters are located since transcription initiation sites are unknown and studies of post-transcriptional regulation, which often involve sequences in the 5′ and 3′ UTRs, are impossible.

Comparative genomics of *Cryptosporidium* species yields informative insights

Many Human-Infecting Species Are Closely Related

Genome sequences for the species most often observed in humans revealed that the genome sequences are highly similar and highly syntenic [6, 21, 22]. Several recent papers have also demonstrated the complex population genomic structure of *C. parvum* and *C. hominis* in natural infections [6, 7••, 15, 23] and highlighted the role that recombination and introgression have played during evolution [6, 14••, 15]. These and other works also show the impact that recombination can have with respect to the generation of novelty with proven impacts on transmission [7••]. Comparative genomics has also revealed the strikingly close relationship between the genome sequences of *C. cuniculus*, which infects rabbits and humans, and *C.*

hominis [6]. Interestingly, an even closer genomic relationship was observed between the genome sequences of *C. parvum* and *C. tyzzeri*, yet *C. tyzzeri* only infects mice [24].

Recently, a genetic cross was reported not only within *C. parvum* but also between *C. parvum* and *C. tyzzeri* [17••]. In addition to being a major genetics breakthrough for *Cryptosporidium* research, this finding, combined with the recombination and introgression observations above, raises questions regarding the definition of what constitutes a species in *Cryptosporidium*. For many reasons, we are not advocating changes, only the recognition of just how similar some subclades and species are to each other at the genomic sequence level and how little we know about their host range [25]. *C. parvum* shares 96.8% identity with *C. hominis*, 97.2% identity with *C. tyzzeri*, 97% identity with *C. cuniculus*, and 91.3% identity with *C. meleagridis* (% identity is average nucleotide identity). These species also share almost complete synteny (gene order and orientation) and appear to differ by only a few sub-telomeric genes, if any [24, 26••] with the exception of *C. meleagridis* that appears, on the basis of long reads, to have a few dozen small intra- and inter-chromosomal rearrangements relative to the other species [27]. Thus, the genetic basis of host preference and pathogenicity may extend from gene content differences to also include single nucleotide variants, small indels, and possible differences in gene regulation. It is worth noting that significant differences in gene content between these species are found with short read analyses, including one in which some of us have participated [22, 28], highlighting the impact of technology and assembly quality on downstream analyses. Notably, synteny with species outside of this group, for which we have genome sequences, no longer extends for the full length of the chromosome and instead is broken down into smaller units of recognizable synteny [11, 29].

Subtelomeric Chromosomal Regions Contain Gene Families and Appear to be Highly Dynamic

In general, the subtelomeric regions of eukaryotic chromosomes are more dynamic in terms of gene copy numbers and levels of observed variation than the rest of the chromosome, and this is particularly the case in pathogenic organisms [30, 31]. Genes that encode proteins involved in host–pathogen interactions and environmental responses are often, but not uniquely, located in sub-telomeric locations [30, 31]. As a result, these regions of the genome are notoriously difficult to assemble. They also represent some of the fastest evolving regions of the genome and, thus, are interesting from the perspectives of host–pathogen biology, evolution, and diagnostics/surveillance.

The first T2T genome sequences for *C. parvum* revealed surprises regarding higher than expected (based on previous short-read assemblies) gene copy number for a number of genes located in subtelomeric regions, e.g., MEDLE genes, tryptophan synthase beta, and rRNA genes among others [24, 26••]. They also revealed that three different chromosomes shared a total of four highly similar subtelomeric chromosome ends, indicating that recombination had occurred between chromosomes [24, 26••]. Better assembly and identification of genes in subtelomeric regions are likely to be crucial for our understanding of important aspects of *Cryptosporidium* biology. For example, MEDLE proteins, most of which are encoded in subtelomeric regions, are important secreted pathogenesis determinants [9, 32, 33] that appear to be differentially present across a number of species [29].

Population Genomic Studies Provide Insights into Variation, Evolution, and Transmission

Short-read sequencing technology permitted the generation of nearly 700 genomic data sets for *C. parvum* and nearly 400 for *C. hominis*. These highly accurate reads have been used to

detect variants that exist among and between the different populations of *Cryptosporidium* parasites that have been sequenced [6, 7••, 13••, 14••, 15, 34, 35•]. The results have been illuminating. They have revealed a discordance in some cases between gp60 single locus typing and genome ancestry, mixed infections with the same or different species, recombination events within species and hybridization between species, discovery of novel subclades, and in general demonstrated the role that admixture has had on shaping population structure [13••, 15].

These studies have also revealed how little we know about the global population structure of *Cryptosporidium* species and the forces driving their evolution in differing environments and out-break scenarios [6, 7••, 13••, 14••, 15, 34, 35•]. These studies also reveal the critical role that the reference genome has in the determination of differences in gene content and polymorphisms. Figure 1 highlights the theoretical outcome of determining single-nucleotide variants (SNVs) in two different scenarios, uncompressed gap free and compressed gapped genome sequences. Thus, a degree of caution is warranted for the interpretation of variant calling until the community has more complete reference genome sequences. The community would greatly benefit from a more diverse set of reference genome sequences and methods for capturing novel genomic content that may not be present in any given reference genome sequence.

Transcriptomics in *Cryptosporidium*

Annotation, Antisense, and ncRNA Transcripts

Utilization of small RNA-seq and PacBio long-read Iso-seq and ONT Direct RNAseq has significantly advanced our understanding of the *Cryptosporidium* transcriptome. These technologies have enabled the identification of untranslated regions (UTRs), as well as a variety

of long and short non-coding RNAs (ncRNAs), including anti-sense transcripts of unknown function [26••, 36, 37]. Furthermore, single-molecule long-read RNAseq has been instrumental in demonstrating that approximately 10% of *C. parvum* genes have polycistronic transcripts, offering new insights into gene expression biology and regulation in this important pathogen (Xiao et al., in prep).

Differential Gene Expression

There is a significant amount of RNA-seq data for *C. parvum* (Table 1). Studies involving oxidative and heat stress on *C. parvum* oocysts have identified genes responsive to environ-mental cues [38]. Transcriptomic analyses of *C. parvum*-infected HCT-8 cells have revealed gene regulation patterns during early stages of infection [39•]. Comparative RNA-seq studies of both *C. parvum* and *C. hominis* have provided novel data sets on host-parasite RNA interactions during infection [40]. Research on AP-2 transcription factor deletion and its impact on sex differentiation and oocyst shedding, RNA m6A-immunoprecipitation in infected host cells, and differential gene expression in distinct *Cryptosporidium* species using enteroids have all contributed significantly to our understanding of life cycle progression and interactions with the host [41••, 42, 43••]. Additionally, RNA-seq in drug screening assays has suggested potential inhibition of translation during parasite sexual differentiation [44].

RNA expression data sets are desperately needed for additional species, especially *C. hominis* (Table 1). These data will permit comparative transcriptomic analyses and provide additional insight into gene regulatory differences that may exist between species. Additionally, RNA data can also be used to greatly improve the annotation of reference genome sequences for the community by providing evidence for UTRs and alternative isoforms, if present.

Epigenetics

Recent studies have shown that *C. parvum* has enzymatically functional histone methyltransferases, indicating developmentally dependent histone modifications. These findings also suggest that *C. parvum* infection can alter the epigenetic landscapes of host cells [45••]. Additionally, ATAC-seq in *C. parvum* sporozoites has provided the first glimpse into the parasite's accessible chromatin landscape, prior to invasion, enabling new research into the regulation of parasite gene expression (Xiao et al., in prep). The community will benefit greatly from additional research into this exciting layer of gene regulation in *Cryptosporidium* especially during it developmental life cycle.

Major Advances That Are Generating Abundant, Informative Genomic and Transcriptomic Data

Transgenics

CRISPR/Cas9 mediated genetic modifications work in *Cryptosporidium* [46, 47]. Genetically modified *C. tyzerri* and *C. parvum* are being used to study host protective immunity and *Cryptosporidium* biology [47–49]. Dihydrofolate reductase-thymidylate synthase (DHFR-TS) and inosine monophosphate dehydrogenase (IMPDH) have been knocked out to study nucleotide synthesis [50]. *C. parvum* parasites were still viable after gene knockouts, suggesting alternative pathways for sequestering nucleotides from the host. This knowledge will help in drug development as some drugs are designed to target nucleotide synthesis [50], and these pathways have been successfully targeted in other apicomplexans.

Several selectable genetic markers are now available [17••, 46, 51••], and this development has permitted genetic crosses [17••, 51••] in immunocompromised mice. This

advance is of considerable significance. The model, combined with genomic sequences, opens up reverse genetic research into *Cryptosporidium* biology, development, and the important question of host specificity.

As cloning of individual parasites is not yet possible, many phenotypic effects like changes in gene regulation or recombinant progeny are assessed via analyses of genome and transcriptome sequencing [17••], thus generating many new, important data sets.

RNA Host-Pathogen Interactions

Recent advances looking at host-*Cryptosporidium* interactions have revealed that a wide variety of novel RNA forms are involved. For example, host circular RNA ciRS-7 is upregulated during *C. parvum* infection in HCT-8 cells. This upregulation influences the NF-κB signaling pathway by sponging miR-1270, which, in turn, significantly impacts the propagation of *C. parvum* [52]. A recent scRNA study of *C. parvum*-infected intestinal epithelial cells has led to the development of a model to explain the role of IFN-gamma in controlling *C. parvum* infection in intestinal epithelial cells [53].

On the parasite side, recent research has shown that some *C. parvum* long noncoding RNAs can localize to the host cell [10••] and manipulate host cell gene expression by suppressing the expression of CDH3 and LOXL4 [8]. One group has also explored RNA-based therapy, in which a single-stranded antisense RNA designed to parasite protein coding genes can silence parasite genes when loaded with argonaute protein [54].

The *Cryptosporidium* RNA Virus and Host-Pathogen Interactions

Previous studies have discovered two unique extrachromosomal linear double-stranded RNAs in *C. parvum*, encoding an RNA-dependent RNA polymerase and a protein kinase [55].

The *Cryptosporidium* dsRNA virus has been detected in isolates of *C. parvum*, *C. hominis*, *C. meleagridis*, and *C. felis* [56]. Recent research has shown that the dsRNA virus can hijack a host long noncoding RNA, U90926 [10••]. Additional research has also shown that the *Cryptosporidium* dsRNA virus can trigger the host type I IFN antiviral pathway to dampen the hosts' antiparasitic response, thus facilitating parasite success [57••]. All of these studies have generated numerous exciting host-pathogen gene expression data sets.

Applications to Control and Prevention

The Influence of Genomics on *Cryptosporidium* Surveillance

The natural environments that serve as reservoirs for *Cryptosporidium* are waterbodies and hosts. Several detection techniques have been developed and improved significantly, from conventional microscopy to immunological assays, flow cytometry, and nucleic acid-based methods [58]. Nucleic acid–based detection methods are highly efficient in detecting mixed infections and mixed populations with low abundance subpopulations present [59]. With nucleic acid detection–based approaches such as PCR, qPCR, and DNA sequencing, tiny amounts of *Cryptosporidium* DNA can be detected, and new bioinformatics tools make it easier to type Sanger sequences of *Cryptosporidium* species based on gp60 and SSU rRNA [60]. Though DNA sequencing as a detection method is relatively expensive, the ability to multiplex samples makes it more affordable, and its high accuracy makes it a better choice than microscopy [61, 62]. Accurately detecting and identifying *Cryptosporidium* sp. in an outbreak, based on genomic data, facilitate inquiries into the source of the outbreak and, on a broader scale, the epidemiology of the disease [63].

Clinical samples are essential in the study of cryptosporidiosis. Unfortunately, clinical isolates do not contain enough oocysts for traditional sequencing as discussed above. However, multiple displacement amplification (MDA), a type of whole genome amplification (WGA) technique, can be used to amplify the amount of DNA. MDA introduces very few errors, amplification is random, and the entire genome can be amplified for sequencing [64, 65]. WGA provides a means to obtain sufficient DNA from samples for use in analyses as well as long-read sequencing [66]. Studies of parasite diversity within and between samples can inform upon the source of the infection and the diversity of *Cryptosporidium* circulating within the population.

Genetic Markers, Diagnostics, and Surveillance

Phenotypic and antibody diagnostic identification are unable to distinguish *Cryptosporidium* species and subtypes; therefore, the genetic markers used, such as 18SrRNA and gp60, were developed as useful single genetic markers. Due to the sexual nature of *Cryptosporidium*, and the fact that it has eight independently segregating chromosomes, multi-locus typing is a more ideal approach to typing [67]. Although there is no consensus on the specific markers to use, there is agreement that a multi-locus approach is needed [68] to better inform on *Cryptosporidium* epidemiology. Also, multi-locus genotyping is better because more data are available for identification and it adds the possibility of species subtype identification [69].

The increasing availability of full genome sequence data from increasing numbers of isolates should make the determination of appropriate loci for typing easier, but challenges remain. First, the community still lacks genomic sequence data from isolates circulating in many important regions of the world with a high incidence of *Cryptosporidium* infection. Just compare the burden reported in Gilbert et al. [3••] with the source of available genomes sequences in Fan

et al. [11]. This situation is beginning to change, and sequences from isolates in other countries are emerging [15]. Hopefully, the genomic advances described above will facilitate this process and unlock the potential of existing samples and lead to the strategic collection of others. Additional genome sequences from new geographic locations and environments will allow the community to survey the extent of the genomic diversity that exists globally and design markers to account for it. One can also imagine the need for specialized markers to very quickly evolving regions of the genome that can be utilized in outbreak scenarios to detect variants as they arise.

Second, in order to appropriately assess genomic variation and rapidly evolving genomic regions, complete T2T reference genomes for the species most commonly infecting humans should be established and adopted.

Finally, markers for routine surveillance are also needed. RT-PCR tests for the *Cryptosporidium* dsRNA virus are very sensitive due to viral abundance [70]. The method has been used to successfully identify *Cryptosporidium* infection in calves, lambs, goats, and environmental water samples across the world [71–73]. However, we do not yet know the full extent to which the dsRNA virus is present in different species [56].

Conclusions

Advances in genomics and transcriptomics are impacting all arenas of *Cryptosporidium* research [74••], from evolution to the life cycle to host-pathogen interactions and surveillance. The *Cryptosporidium* research community has come far, very quickly with many new technologies, approaches, and data sets. Much of this new data is available for use and mining in the NCBI GenBank [75] and CryptoDB.org [76•].

The *Cryptosporidium* community is also struggling a bit with the difficult challenges posed by this important pathogen and the state of genomics technology. The lack of transcriptomic data for species other than *C. parvum* and the lack of genome sequences for more than half of the named species are real challenges. Complete genome sequences are still too hard to generate and even harder to consistently annotate, especially in the absence of RNA data. This reality impacts their utility and application to important needs like global surveillance and determination of complete gene repertoires. Complete, annotated reference genome sequences greatly facilitate experimental design, e.g., gene knock-outs, and pathway analyses. Likewise, analyses of data that require a reference genome sequence for interpretation, e.g., transcriptome and proteome data analyses and comparative genomics and evolution studies, will also benefit.

The community is also struggling with appropriate geographic representation of genomic data sets from many of the countries most affected by this pathogen. This lack of representation impacts the development of more representative, multi-locus diagnostics and impacts our knowledge base for epidemiological studies and outbreak investigations. Given how difficult *Cryptosporidium* is to work with, genomics advances have come far, but more is needed.

Acknowledgments

The authors would like to thank R.P. Baptista for useful discussions and L.R. Penumarthi for assistance with Table 1.

References

1. Checkley W, White AC Jr, Jaganath D, Arrowood MJ, Chalmers RM, Chen XM, et al. A review of the global burden, novel diagnostics, therapeutics, and vaccine targets for *Cryptosporidium*. Lancet Infect Dis. 2015;15(1):85–94. https://doi.org/10.1016/ S1473- 3099(14) 70772-8.

2. Kotloff KL, Nataro JP, Blackwelder WC, Nasrin D, Farag TH, Panchalingam S, et al. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case control study. Lancet. 2013;382(9888):209–22. https://doi.org/10.1016/ S0140-6736(13) 60844-2.

3.•• Gilbert IH, Vinayak S, Striepen B, Manjunatha UH, Khalil IA, Van Voorhis WC, et al. Safe and effective treatments are needed for cryptosporidiosis, a truly neglected tropical disease. BMJ Glob Health. 2023;8(8). https://doi. org/10.1136/bmjgh-2023- 012540. This work is the most recent collection of statistics, deaths, and DALYs on the global burden of *Cryptosporidium* and places it in the context of other neglected tropical tropical diseases. It also addresses the need for safe and effective therapeutic treatments.

4. English ED, Guerin A, Tandel J, Striepen B. Live imaging of the *Cryptosporidium* parvum life cycle reveals direct development of male and female gametes from type I meronts. PLoS Biol. 2022;20(4):e3001604. https://doi.org/10. 1371/journ al. pbio.30016 04.

5. Tandel J, English ED, Sateriale A, Gullicksrud JA, Beiting DP, Sullivan MC, et al. Life cycle progression and sexual development of the apicomplexan parasite *Cryptosporidium*

parvum. Nat Microbiol. 2019;4(12):2226–36. https://doi.org/10. 1038/s41564-019-0539-x.

6. Nader JL, Mathers TC, Ward BJ, Pachebat JA, Swain MT, Robinson G, et al. Evolutionary genomics of anthroponosis in *Cryptosporidium*. Nat Microbiol. 2019;4(5):826–36. https://doi.org/10.1038/s41564-019-0377-x.

7.•• Huang W, Guo Y, Lysen C, Wang Y, Tang K, Seabolt MH, et al. Multiple introductions and recombination events underlie the emergence of a hyper-transmissible *Cryptosporidium hominis* subtype in the USA. Cell Host Microbe. 2023;31(1):112- 23 e4. https://doi.org/10.1016/j.chom.2022.11.013. This work demonstrates how multiple different recombination events involving *C. hominis* isolates from within the USA and other countries generated the hyper-transmissiable *C. hominis* IfA12G1R5 subtype that is increasing in incidence in the USA.

8. Ming Z, Gong AY, Wang Y, Zhang XT, Li M, Li Y, et al. Transsuppression of host CDH3 and LOXL4 genes during *Cryptosporidium* parvum infection involves nuclear delivery of parasite Cdg7_FLc_1000 RNA. Int J Parasitol. 2018;48(6):423–31. https://doi.org/10.1016/j. ijpara.2017.10.008.

9. Dumaine JE, Sateriale A, Gibson AR, Reddy AG, Gullicksrud JA, Hunter EN, et al. The enteric pathogen *Cryptosporidium* parvum exports proteins into the cytosol of the infected host cell. Elife. 2021;10. https://doi.org/10.7554/eLife.70451.

10.•• Graham ML, Li M, Gong AY, Deng S, Jin K, Wang S, et al. *Cryptosporidium parvum* hijacks a host's long noncoding RNA U90926 to evade intestinal epithelial cell-autonomous antiparasitic defense. Front Immunol. 2023;14:1205468. https://doi.org/10.3389/fimmu.2023.12054 68. This work demonstrates the role that the

Cryptosporidium virus plays in affecting regulation of a host lncRNA to epigenetically alter the host cell anti-parastitic response

11. Fan Y, Feng Y, Xiao L. Comparative genomics: how has it advanced our knowledge of cryptosporidiosis epidemiology? Parasitol Res. 2019;118(12):3195–204. https://doi.org/10.1007/s00436-019-06537-x.

12. Abrahamsen MS, Templeton TJ, Enomoto S, Abrahante JE, Zhu G, Lancto CA, et al. Complete genome sequence of the apicomplexan *Cryptosporidium parvum*. Science. 2004;304(5669):441–5. https://doi.org/10.1126/science.10947 86.

13.•• Corsi GI, Tichkule S, Sannella AR, Vatta P, Asnicar F, Segata N, et al. Recent genetic exchanges and admixture shape the genome and population structure of the zoonotic pathogen *Cryptosporidium parvum*. Mol Ecol. 2023;32(10):2633–45. https://doi.org/10.1111/ mec.16556. This work identifies and traces the gene flow that resulted from several recombination events in *C. parvum* between ruminant and human isolates. They also age the recomination events and show that 50% have occurred in the last ~200 years.

14.•• Tichkule S, Caccio SM, Robinson G, Chalmers RM, Mueller I, Emery-Corbin SJ, et al. Global population genomics of two subspecies of *Cryptosporidium hominis* during 500 years of evolution. Mol Biol Evol. 2022;39(4). https://doi.org/10.1093/molbev/ msac0 56. This work demonstrates the evolution of two proposed subspecies of *C. hominis* that differ by environment and transmission, i.e., between low-income and high-income countries. These subtypes differ in putative resistance genes, effective population sizes, and there is a bias in the direction of gene flow between them.

15. Tichkule S, Jex AR, van Oosterhout C, Sannella AR, Krumkamp R, Aldrich C, et al. Comparative genomics revealed adaptive admixture in *Cryptosporidium hominis* in Africa. Microb Genom. 2021;7(1). https://doi.org/10.1099/mgen.0.000493.

16. Troell K, Hallstrom B, Divne AM, Alsmark C, Arrighi R, Huss M, et al. *Cryptosporidium* as a testbed for single cell genome characterization of unicellular eukaryotes. BMC Genomics. 2016;17:471. https://doi.org/10.1186/ s12864-016- 2815-y.

17.•• Shaw S, Cohn IS, Baptista RP, Xia G, Melillo B, Agyabeng-Dadzie F, et al. Genetic crosses within and between species of *Cryptosporidium*. Proc Natl Acad Sci U S A. 2024;121(1):e2313210120. https:// doi.org/ 10.1073/ pnas. 2313210120. This work highlights the development of a new selectable marker for genetic studies in *Cryptosporidium*. They use this and an existing marker to perform genetic crosses between *C. parvum* parasites and between *C. parvum* and *C. tyzzeri*. This work also reports on the successful use of genome amplification and long-read sequencing of single oocysts.

18. Agyabeng-Dadzie F, Beaudry M, Deyanov A, Slanis H, Duong MQ, Turner R, et al. Evaluating the benefits and limits of multiple displacement amplification with whole-genome Oxford Nanopore Sequencing. bioRxiv. 2024. https://doi.org/10. 1101/2024.02.09.579537.

19. Gaudin M, Desnues C. Hybrid capture-based next generation sequencing and its application to human infectious diseases. Front Microbiol. 2018;9:2924. https:// doi.org/10.3389/ fmicb.2018.02924.

20. Kissinger JC, Hermetz KE, Woods KM, Upton SJ. Enrichment of *Cryptosporidium parvum* from in vitro culture as measured by total RNA and subsequent sequence

analysis. Mol Biochem Parasitol. 2018;220:5–9. https://doi.org/10.1016/j.molbi opara. 2017.12.004.

21. Xu P, Widmer G, Wang Y, Ozaki LS, Alves JM, Serrano MG, et al. The genome of *Cryptosporidium hominis*. Nature. 2004;431(7012):1107–12. https:// doi.org/10. 1038/nature02977.

22. Ifeonu OO, Chibucos MC, Orvis J, Su Q, Elwin K, Guo F, et al. Annotated draft genome sequences of three species of *Cryptosporidium*: *Cryptosporidium meleagridis* isolate UKMEL1, *C. baileyi* isolate TAMU-09Q1 and *C. hominis* isolates TU502_2012 and UKH1. Pathog Dis. 2016;74(7). https:// doi.org/10.1093/ femspd/ ftw080.

23. Gilchrist CA, Cotton JA, Burkey C, Arju T, Gilmartin A, Lin Y, et al. Genetic diversity of *Cryptosporidium hominis* in a Bangladeshi community as revealed by whole-genome sequencing. J Infect Dis. 2018;218(2):259–64. https://doi.org/10.1093/ infdis/jiy121.

24. Baptista RP, Li Y, Sateriale A, Sanders MJ, Brooks KL, Tracey A, et al. Long-read assembly and comparative evidence-based reanalysis of *Cryptosporidium* genome sequences reveal expanded transporter repertoire and duplication of entire chromosome ends including subtelomeric regions. Genome Res. 2022;32(1):203–13. https://doi.org/10.1101/gr.275325.121.

25. Widmer G, Koster PC, Carmena D. *Cryptosporidium hominis* infections in non-human animal species: revisiting the concept of host specificity. Int J Parasitol. 2020;50(4):253–62. https://doi.org/10.1016/j.ijpara. 2020.01.005.

26.•• Baptista RP, Xiao R, Li Y, Glenn TC, Kissinger JC. New T2T assembly of *Cryptosporidium parvum* IOWA annotated with reference genome gene identifiers.

bioRxiv. 2023. https:// doi.org/10.1101/ 2023.06.13.544219. This pre-print reports on the first telomer to telomere genomic assembly for *C. parvum* that contains all 16 telomeres. This genome sequence is annotated using gene IDs from the current reference IOWA genome assembly when possible and extensive annotation of non-coding RNAs.

27. Penumarthi LR, Baptista RP, Beaudry MS, Glenn TC, Kissinger JC. A new chromosome-level genome assembly and annotation of *Cryptosporidium meleagridis* bioRxiv. 2024. https:// doi. org/10.1101/2024.02.16.580748

28. Arias-Agudelo LM, Garcia-Montoya G, Cabarcas F, Galvan-Diaz AL, Alzate JF. Comparative genomic analysis of the principal *Cryptosporidium* species that infect humans. PeerJ.2020;8:e10478. https://doi.org/10.7717/peerj.10478.

29. Xu Z, Li N, Guo Y, Feng Y, Xiao L. Comparative genomic analysis of three intestinal species reveals reductions in secreted pathogenesis determinants in bovine-specific and non-pathogenic *Cryptosporidium* species. Microb Genom. 2020;6(6). https://doi.org/10.1099/mgen.0.000379.

30. Otto TD, Bohme U, Sanders M, Reid A, Bruske EI, Duffy CW, et al. Long read assemblies of geographically dispersed Plasmodium falciparum isolates reveal highly structured subtelomeres. Wellcome Open Res. 2018;3:52. https://doi.org/10. 12688/ wellcomeopenres.14571.1.

31. Dunn MJ, Shazib SUA, Simonton E, Slot JC, Anderson MZ. Architectural groups of a subtelomeric gene family evolve along distinct paths in Candida albicans. G3 (Bethesda). 2022;12(12). https://doi.org/10.1093/g3journal/jkac283.

32. Fei J, Wu H, Su J, Jin C, Li N, Guo Y, et al. Characterization of MEDLE-1, a protein in early development of *Cryptosporidium parvum*. Parasite Vectors. 2018;11(1):312. https://doi.org/10.1186/s13071- 018-2889-2.

33. Xu Z, Guo Y, Roellig DM, Feng Y, Xiao L. Comparative analysis reveals conservation in genome organization among intestinal *Cryptosporidium* species and sequence divergence in potential secreted pathogenesis determinants among major human-infecting species. BMC Genomics. 2019;20(1):406. https://doi. org/10.1186/s12864-019- 5788-9.

34. Baptista RP, Cooper GW, Kissinger JC. Challenges for *Cryptosporidium* population studies. Genes (Basel). 2021;12(6). https://doi. org/ 10. 3390/ genes 12060 894.

35.• Wang T, Guo Y, Roellig DM, Li N, Santin M, Lombard J, et al. Sympatric recombination in zoonotic *Cryptosporidium* leads to emergence of populations with modified host preference. Mol Biol Evol. 2022;39(7). https://doi.org/10.1093/molbev/msac150. This comparative genomic analysis reveals how past and ongoing recombination events, especially between parasite strains isolated from humans and animals, have affected the population structure of *C. parvum* and affected host preference.

36. Li Y, Baptista RP, Mei X, Kissinger JC. Small and intermediate size structural RNAs in the unicellular parasite *Cryptosporidium parvum* as revealed by sRNA-seq and comparative genomics. Microb Genom. 2022;8(5). https://doi.org/10.1099/ mgen.0. 000821.

37. Li Y, Baptista RP, Sateriale A, Striepen B, Kissinger JC. Analysis of long non-coding RNA in *Cryptosporidium parvum* reveals significant stage-specific antisense

transcription. Front Cell Infect Microbiol. 2020;10:608298. https://doi.org/10.3389/fcimb.2020.608298.

38. Temesgen TT, Tysnes KR, Robertson LJ. Use of oxidative stress responses to determine the efficacy of inactivation treatments on *Cryptosporidium* oocysts. Microorganisms. 2021;9(7). https://doi.org/10.3390/micro organisms9071463.

39.• Sun L, Li J, Xie F, Wu S, Shao T, Li X, et al. Whole transcriptome analysis of HCT-8 cells infected by *Cryptosporidium parvum*. Parasit Vectors. 2022;15(1):441. https://doi.org/ 0.1186/s13071- 022- 05565-4. This work examines host gene expression 3 and 12 h post-infection and highlights a number of coding and non-coding gene expression differences.

40. Greigert V, Saraav I, Son J, Zhu Y, Dayao D, Antia A, et al. *Cryptosporidium* infection of human small intestinal epithelial cells induces type III interferon and impairs infectivity of Rotavirus. Gut Microbes. 2024;16(1):2297897. https://doi.org/10.1080/19490 976. 2023. 22978 97.

41.•• Tandel J, Walzer KA, Byerly JH, Pinkston B, Beiting DP, Striepen B. Genetic ablation of a female-specific apetala 2 transcription factor blocks oocyst shedding in *Cryptosporidium parvum*. mBio. 2023;14(2):e0326122. https://doi.org/10.1128/mbio.03261-22. This work identifies male- and female specific AP2 transcription factors and develops conditional gene knock-out strategies to study the female AP2. Transcriptomics revealed a role for the female AP2 in crystalloid protein expression and transmission.

42. Xia Z, Xu J, Lu E, He W, Deng S, Gong AY, et al. m6A mRNA methylation regulates epithelial innate antimicrobial defense against cryptosporidial infection. Front Immunol. 2021;12:705232. https://doi.org/10.3389/fimmu.2021.705232.

43.•• Deng M, Hou T, Mao X, Zhang J, Yang F, Wei Y, et al. Cultivation of host-adapted *Cryptosporidium parvum* and *Cryptosporidium hominis* using enteroids for cryopreservation of isolates and transcriptomic studies of infection. bioRxiv: Cold Spring Harbor Laboratory; 2023. https://doi.org/10.1101/2023. 12. 06. 570384. This preprint reports on a new murine enteriod system that works for particular isolates of both of the major human-infecting *Cryptosporidium* species in which the life cycle can be completed and parasites can be cryopreserved. The system has been assessed with transciptomic analyses.

44. Hasan M, Mattice E, Teixeira JE, Jumani RS, Stebbins EE, Klopfer C, et al. *Cryptosporidium* life cycle small molecule probing implicates translational repression and an apetala 2 transcription factor in sexual differentiation. bioRxiv. Cold Spring Harbor Laboratory; 2023. https://doi.org/10.1101/2023. 12. 18. 572108.

45. Sawant M, Benamrouz-Vanneste S, Meloni D, Gantois N, Even G, Guyot K, et al. Putative SET-domain methyltransferases in *Cryptosporidium parvum* and histone methylation during infection. Virulence. 2022;13(1):1632–50. https://doi.org/10.1080/ 21505 594. 2022. 21233 63.

46. Vinayak S, Pawlowic MC, Sateriale A, Brooks CF, Studstill CJ, Bar-Peled Y, et al. Genetic modification of the diarrhoeal pathogen *Cryptosporidium parvum*. Nature. 2015;523(7561):477–80. https://doi.org/10.1038/nature14651.

47. Sateriale A, Pawlowic M, Vinayak S, Brooks C, Striepen B. Genetic manipulation of *Cryptosporidium parvum* with CRISPR/Cas9. Methods Mol Biol. 2020;2052:219–28. https://doi.org/10.1007/978-1-4939- 9748-0_13.

48. Sateriale A, Slapeta J, Baptista R, Engiles JB, Gullicksrud JA, Herbert GT, et al. A genetically tractable, natural mouse model of cryptosporidiosis offers insights into host protective immunity. Cell Host Microbe. 2019;26(1):135-46 e5. https://doi.org/10.1016/j.chom.2019.05.006.

49. Vinayak S. Recent advances in genetic manipulation of *Cryptosporidium*. Curr Opin Microbiol. 2020;58:146–52.https:// doi. org/ 10. 1016/j. mib. 2020. 09. 010.

50. Pawlowic MC, Somepalli M, Sateriale A, Herbert GT, Gibson AR, Cuny GD, et al. Genetic ablation of purine salvage in *Cryptosporidium parvum* reveals nucleotide uptake from the host cell. Proc Natl Acad Sci U S A. 2019;116(42):21160–5. https://doi.org/10.1073/pnas.19082 39116.

51.•• Hanna JC, Corpas-Lopez V, Seizova S, Colon BL, Bacchetti R, Hall GMJ, et al. Mode of action studies confirm on-target engagement of lysyl-tRNA synthetase inhibitor and lead to new selection marker for *Cryptosporidium*. Front Cell Infect Microbiol. 2023;13:1236814. https://doi.org/10.3389/fcimb.2023.1236814. This work described the development of a new selectable genetic marker for *Cryptosporidium* genetics studies and reports on its use in a genetic cross in *C. parvum*.

52. Yin YL, Liu TL, Yao Q, Wang YX, Wu XM, Wang XT, et al. Circular RNA ciRS-7 affects the propagation of *Cryptosporidium parvum* in HCT-8 cells by sponging miR-1270 to

activate the NF-kappaB signaling pathway. Parasit Vectors. 2021;14(1):238. https://doi.org/10.1186/s13071-021-04739-w.

53. Pardy RD, Walzer KA, Wallbank BA, Byerly JH, O'Dea KM, Cohn IS, et al. Analysis of intestinal epithelial cell responses to *Cryptosporidium* highlights the temporal effects of IFN-gamma on parasite restriction. bioRxiv. 2023. https://doi.org/10.1101/2023. 11. 14. 567008.

54. Castellanos-Gonzalez A, Sadiqova A, Ortega-Mendez J, White AC Jr. RNA-based therapy for *Cryptosporidium parvum* infection: proof-of-concept studies. Infect Immun. 2022;90(7):e0019622. https://doi.org/10.1128/iai.00196-22.

55. Khramtsov NV, Woods KM, Nesterenko MV, Dykstra CC, Upton SJ. Virus-like, double-stranded RNAs in the parasitic protozoan *Cryptosporidium parvum*. Mol Microbiol. 1997;26(2):289–300. https://doi.org/10.1046/j.1365- 2958.1997.5721933.x.

56. Leoni F, Gallimore CI, Green J, McLauchlin J. Characterisation of small double stranded RNA molecule in *Cryptosporidium hominis*, *Cryptosporidium felis* and *Cryptosporidium meleagridis*. Parasitol Int. 2006;55(4):299–306. https://doi.org/10.1016/j.parint.2006.06.006.

57.•• Deng S, He W, Gong AY, Li M, Wang Y, Xia Z, et al. *Cryptosporidium* uses CSpV1 to activate host type I interferon and attenuate antiparasitic defenses. Nat Commun. 2023;14(1):1456. https://doi.org/10.1038/s41467- 023-37129-0. This work highlights the role of the type1 IFN receptor in *C. parvum* Tinfection and demonstrates a role for the *Cryptosporidium* virus in altering host cell gene expression and triggering a type 1 IFN response in infected cells. This response allows the parasite to evade the epithelial antiparasitic response.

58. Jex AR, Smith HV, Monis PT, Campbell BE, Gasser RB. *Cryptosporidium*–biotechnological advances in the detection, diagnosis and analysis of genetic variation. Biotechnol Adv. 2008;26(4):304–17. https://doi.org/10.1016/j.biotechadv.2008.02.003.

59. Luka G, Samiei E, Tasnim N, Dalili A, Najjaran H, Hoorfar M. Comprehensive review of conventional and state-of-the-art detection methods of *Cryptosporidium*. J Hazard Mater. 2022;421:126714. https://doi.org/10.1016/j.jhazmat.2021.126714.

60. Yanta CA, Bessonov K, Robinson G, Troell K, Guy RA. CryptoGenotyper: a new bioinformatics tool for rapid *Cryptosporidium* identification. Food Waterborne Parasitol. 2021;23:e00115. https://doi.org/10.1016/j.fawpar.2021.e00115.

61. Ryan U, Paparini A, Oskam C. New technologies for detection of enteric parasites. Trends Parasitol. 2017;33(7):532–46. https://doi.org/10.1016/j. pt. 2017.03.005.

62. Xiao L, Feng Y. Molecular epidemiologic tools for waterborne pathogens *Cryptosporidium* spp. and Giardia duodenalis. Food Waterborne Parasitol. 2017;8–9:14–32. https://doi.org/10.1016/j.fawpar.2017.09.002.

63. Alderisio KA, Mergen K, Moessner H, Madison-Antenucci S. Identification and evaluation of *Cryptosporidium* species from New York City cases of cryptosporidiosis (2015 to 2018): a watershed perspective. Microbiol Spectr. 2023;11(1):e0392122. https://doi.org/10.1128/spectrum.03921-22.

64. Dean FB, Hosono S, Fang L, Wu X, Faruqi AF, Bray-Ward P, et al. Comprehensive human genome amplification using multiple displacement amplification. Proc Natl Acad Sci US A. 2002;99(8):5261–6. https://doi.org/10.1073/pnas.082089499.

65. Hou Y, Wu K, Shi X, Li F, Song L, Wu H, et al. Comparison of variations detection between whole-genome amplification methods used in single-cell resequencing. Gigascience. 2015;4:37. https://doi.org/10.1186/s13742-015-0068-3.

66. Guo Y, Li N, Lysen C, Frace M, Tang K, Sammons S, et al. Isolation and enrichment of *Cryptosporidium* DNA and verification of DNA purity for whole-genome sequencing. J Clin Microbiol. 2015;53(2):641–7. https://doi.org/10.1128/JCM. 02962-14.

67. Widmer G, Lee Y. Comparison of single- and multilocus genetic diversity in the protozoan parasites *Cryptosporidium parvum* and *C. hominis*. Appl Environ Microbiol. 2010;76(19):6639–44. https://doi.org/10.1128/AEM.01268-10.

68. Chalmers RM, Perez-Cordon G, Caccio SM, Klotz C, Robertson LJ. participants of the *Cryptosporidium* genotyping w. *Cryptosporidium* genotyping in Europe: the current status and processes for a harmonised multi-locus genotyping scheme. Exp Parasitol. 2018;191:25–30. https://doi.org/10.1016/j.exppara.2018.06.004.

69. Uran-Velasquez J, Alzate JF, Farfan-Garcia AE, Gomez-Duarte OG, Martinez-Rosado LL, Dominguez-Hernandez DD, et al. Multilocus sequence typing helps understand the genetic diversity of *Cryptosporidium hominis* and *Cryptosporidium parvum* isolated from Colombian patients. PLoS ONE. 2022; 17 (7): e0270995. https://doi.org/10.1371/journal.pone.02709 95.

70. de Souza MS, O'Brien C, Santin M, Jenkins M. A highly sensitive method for detecting *Cryptosporidium parvum* oocysts recovered from source and finished water using RT-PCR directed to Cryspovirus RNA. J Microbiol Methods. 019;156:77–80. https://doi.org/10.1016/j. mimet. 2018.11.022.

71. Adjou KT, Chevillot A, Lucas P, Blanchard Y, Louifi H, Arab R, et al. First identification of *Cryptosporidium parvum* virus 1(CSpV1) in various subtypes of *Cryptosporidium parvum* from diarrheic calves, lambs and goat kids from France. Vet Res. 2023;54(1):66. https://doi.org/10.1186/s13567-023-01196-4 .

72. Berber E, Simsek E, Canakoglu N, Sursal N, Gencay GA. Newly identified *Cryptosporidium parvum* virus-1 from newborn calf diarrhoea in Turkey. Transbound Emerg Dis. 2021;68(4):2571–80. https://doi.org/10.1111/tbed.13929 .

73. Chae JB, Shin SU, Kim S, Jo YM, Roh H, Chae H, et al. The first identification of *Cryptosporidium parvum* virus-1 (CSpV1) in Hanwoo (Bos taurus coreanae) calves in Korea. Vet Sci. 2023;10(11). https://doi.org/10.3390/vetsci10110633 .

74.•• Dabrowska J, Sroka J, Cencek T. Investigating *Cryptosporidium* spp. using genomic, proteomic and transcriptomic techniques: current progress and future directions. Int J Mol Sci. 2023;24(16). https:// doi. org/ 10. 3390/ ijms2 41612 867. This review is a detailed history with timelines of omics research in *Cryptosporidium*.

75. Sayers EW, Cavanaugh M, Clark K, Pruitt KD, Sherry ST, Yankie L, et al. GenBank 2024 update. Nucleic Acids Res. 2024;52(D1):D134–7. https://doi.org/10.1093 nar/gkad903 .

76.• Warrenfeltz S, Kissinger JC, EuPath DBT. Accessing *Cryptosporidium* omic and isolate data via CryptoDB.org. Methods Mol Biol. 2020;2052:139–92. https:// doi. org/ 10. 1007/ 978-1-4939- 9748-0_ 10. This work describes a database where existing *Cryptosporidium* omic data can be searched, visualized, and downloaded.

Papers of particular interest, published recently, have been highlighted as:

• Of importance
•• Of major importance

Tables

Table 1.1 Current available *Cryptosporidium* genomic and transcriptomic data

Organism	**Strain / Isolate**	**Sequencing techniques**	**Reference Genome GeneBank ID**	**Genome sequences**	**Assembly Status**	**size range (Mb)**	**genomes annotated**	**SRA genome sequence available**	**SRA RNA data sets available**
Cryptosporidium andersoni	37034 30847 31729	Illumina	GCA_001865355.1	3	Scaffold	8.966 - 9.089	1	5	0
Cryptosporidium baileyi	TAMU-09Q1	Illumina	GCA_001593455.1	1	Contig	8.494	0	5	1
Cryptosporidium bovis	45015 SCAU24463 SCAU42270 SCAU24365 SCAU42290 SCAU24174 42482	Illumina	GCA_009768925.2	7	Scaffold Contig	9.021 - 9.111	1	7	0
Cryptosporidium canis	33844 45460 25894	Illumina	GCA_027243985.1	3	Scaffold Contig	8.553 - 8.746	3	4	0
Cryptosporidium cuniculus	UKCU2	Illumina	GCA_004337835.1	1	Scaffold	9.184	0	3	0
Cryptosporidium felis	44884	Illumina	GCA_014529505.1	1	Scaffold	8.551	1	6	0

Cryptosporidium hominis	TU502 IbA9G3 TKD180701022-N703-AK415 TKD180701022-AK1581-AK428 37999 30976 TU502_2012 UKH1 UKH3 30974 UKH5 UKH4 33537 TU502	Illumina Ion torrent	GCA_000006425.1	15	Chromosomes Scaffold Contig	8.692 - 9.180	5	391	4
Cryptosporidium meleagridis	UKMEL1 UKMEL3 UKMEL4 TU1867	Illumina	GCA_001593445.1	3	Scaffold Contig	8.707 - 8.973	1	8	0
Cryptosporidium muris	RN66	Hybrid sequencing Sanger	GCA_000006515.1	1	Scaffold	9.245	1	5	0

Cryptosporidium parvum	Iowa type II IOWA-ATCC IOWA UKP1 TU114 UKP6 HLJ 11730 Waterborne O18 UKP2 HB 12536 HN O20 GD 22971 UKP4 IIdA19G1 UKP3 UKP8 UKP7 UKP5 UKP16 UKP15 IIdA20G1 UKP14 IaA15G1R1 UKP12 UKP13 Iowa	Sanger Illumina Ion torrent Oxford Nanopore PacBio	GCA_000165345.1	27	Complete T2T Chromosome Scaffold Contig	8.882 - 9.402	4	698	145
Cryptosporidium ryanae	45019	Illumina	GCA_009792415.2	1	Scaffold	9.059	1	1	0
Cryptosporidium tyzzeri	UGA55	Illumina	GCA_007210665.1	1	Scaffold	9.016	1	4	0
Cryptosporidium ubiquitum	39726 39668 39725 UKUB2 UKUB1	*Illumina*	GCA_001865345.1	5	Scaffold	8.966 - 9.101	1	4	0

Cryptosporidium viatorum	UKVIA1	Illumina	GCA_004337795.1	1	Scaffold	9.264	0	1	0
Cryptosporidium sp. Chipmunk genotype I	37763	Illumina	GCA_004936735.2	1	Scaffold	9.052	1	1	0
Cryptosporidium sp. Chipmunk LX-2015	chipmunk LX-2015	Illumina	GCA_000831705.1	1	Contig	9.51	0	1	0
Cryptosporidium sp.	SCAU42500	Illumina	GCA_029747635.1	1	Scaffold	9.094	0	1	0

T2T telomere to telomere
Data were accessed at the NCBI GenBank in January 2024
*Unconfirmed *Cryptosporidium* species

Figures

A. Reads mapped to Reference Genome **without** Gene Family 1 compression

B. Reads mapped to Reference Genome with Gene Family 1 compression

Figure. 1. Genome assembly impacts annotation quality, gene family member estimates, and genetic variation analyses. A. Long-read assembly can clearly identify all three copies of gene 1, and the average read depth is uniform suggesting the assembly does not contain compressed, i.e., merged, assembled sequence in this area. B. Short-read assembly cannot separate the three closely related gene 1 family members, as evidenced by the read pile-up. The phenomenon is called compression since three genes are merged and annotated as only 1 gene. Also, since the ends of the gene 1 reads are different for each gene copy, the contig cannot be extended, and a gap in the assembly is generated. Compressions do not only affect the determination of gene number; they also affect estimates of genetic variation. Reads that were generated from different family members are all mapped to one locus; thus, the estimate of variation is artificially high. This image was created with BioRender.com

CHAPTER 2

EVALUATING THE BENEFITS AND LIMITS OF MULTIPLE-DISPLACEMENT AMPLIFICATION WITH WHOLE-GENOME OXFORD NANOPORE SEQUENCING*

Reprinted here with permission from the publisher.

*Agyabeng-Dadzie F, Beaudry MS, Deyanov A, Slanis H, Duong MQ, Turner R, Khan A, Arias CA, Kissinger JC, Glenn TC, de Paula Baptista R. Evaluating the Benefits and Limits of Multiple Displacement Amplification with Whole-Genome Oxford Nanopore Sequencing. Mol Ecol Resour. 2025 Feb 28:e14094. doi: 10.1111/1755-0998.14094. Epub ahead of print. PMID: 40018988.

Author contributions**:** F.A.-D., M.S.B., A.D., H.S., M.Q.D. and R.P.B. performed the research. M.S.B., J.C.K., T.C.G. and R.P.B. conceived the research. F.A.-D., M.S.B., A.D., M.Q.D. and R.P.B. analysed the results. F.A.-D., M.S.B., A.D., R.T., A.K., C.A.A., J.C.K., T.C.G. and R.P.B. contributed to writing the manuscript. C.A.A., J.C.K., T.C.G. and R.P.B. obtained funding. All authors reviewed and approved the manuscript.

Abstract

Multiple displacement amplification (MDA) outperforms conventional PCR in long-fragment and whole-genome amplification, making it attractive to couple MDA with long-read sequencing of samples with limited quantities of DNA to obtain improved genome assemblies. Here, we explore the efficacy and limits of MDA for efficient, low-cost genome sequence assembly using Oxford Nanopore Technologies (ONTs) rapid library preparations and minION sequencing. We successfully generated almost complete genome sequences for all organisms examined, including Gram-positive (*Staphylococcus aureus*, *Enterococcus faecium*) and Gram-negative (*Escherichia coli*) prokaryotes, as well as one challenging eukaryotic pathogen (*Cryptosporidium* spp), representing a broad spectrum of critical infectious disease pathogens. High-quality data from those samples were generated starting with only 0.025 ng of total DNA. Controlled sheared DNA samples exhibited a distinct pattern of size increase after MDA, which may be associated with the amplification of long, low-abundance fragments present in the assay, along with generating concatemeric sequences during amplification. To address concatemers, we developed a computational pipeline (CADECT: Concatemer Detection Tool) to identify and remove putative concatemeric sequences. This study highlights the efficacy of MDA in generating high-quality genome assemblies from limited amounts of input DNA. Additionally, the CADECT pipeline effectively mitigated the impact of concatemeric sequences, enabling the assembly of contiguous sequences even in cases where the input genomic DNA was degraded. These results have significant implications for the study of organisms that are challenging to culture *in vitro*, such as *Cryptosporidium*, and for expediting critical results in clinical settings with limited quantities of available genomic DNA.

Introduction

The advent of next-generation sequencing technologies has revolutionized genomics research by enabling the rapid and cost-effective generation of vast amounts of sequencing data [1, 2]. Among these technologies, Oxford Nanopore Technology (ONT) stands out due to its ability to provide long-read sequencing data in real-time, with lower instrument costs compared to other major commercial long-read sequencing platforms, such as PacBio [3]. ONT sequencing has been utilized for numerous applications, including *de novo* genome assembly, metagenomics, and pathogen detection. However, ONT sequencing library preparations typically still require higher quality and greater quantities of DNA inputs than may be available for many projects. ONT rapid library preparations usually need at least 50–200 ng of input DNA for both the older and newer kits per sample. However, more input DNA is required when pooling fewer than eight barcoded samples, as a minimum of 400 ng is recommended for loading onto a MinION flow cell. Many samples also have degraded DNA with fragments that are too short for ONT rapid library preparation or have limited utility in assembly. Removing these small fragments to optimize read length and assembly processes further reduces the available DNA quantity. This poses challenges when working with samples containing limited quantity and/or degraded DNA [4]. For this reason, alternative library preparation or sequencing techniques, including short-read sequencers (e.g., Illumina, Element Biosciences AVITI), are often preferred for handling samples with low molecular weight and/or low quantities of DNA.

To overcome these limitations, multiple displacement amplification (MDA) has emerged as a valuable and highly efficient method for amplifying small quantities of DNA. MDA has significant advantages over conventional PCR-based techniques such as random or degenerate oligonucleotide-primed PCR (DOP-PCR) and primer extension preamplification (PEP) [5].

MDA advantages include reduced waste of rare samples since we can use lower amounts of DNA, isothermal amplification for efficiency, heightened sensitivity in detecting low amounts of DNA inputs, minimized coverage bias and error rates when compared to the other PCR-based methods and amplification of longer DNA fragments [6, 7]). MDA uses Phi29 DNA polymerase, which has a displacement activity that enables the isothermal amplification of DNA with high fidelity and exponential amplification of DNA molecules [6]. This technique has been successfully applied in various genomic studies, including single-cell sequencing [8], ancient DNA analysis, and microbiome studies [9, 10]. While a protocol for MDA with ONT ligation sequencing kits is available using the repli-G kit (Qiagen, Germany), the potential of MDA's application with ONT rapid kits, which offer faster processing times and less cost (formalin-fixed paraffin-embedded, FFPE, DNA repair is not needed in the rapid kit protocol), though it yields relatively smaller fragments compared to ligation kits, has not been investigated. Consequently, MDA's potential limitations and impacts on whole-genome assembly in this context remain relatively unexplored.

The use of MDA combined with ONT sequencing has the potential to unlock genomic insights for organisms that are small to microscopic, especially those that are difficult or impossible to culture *in vitro* (e.g., *Cryptosporidium* species, *Mycobacterium leprae* and *Treponema pallidum*). Furthermore, clinical samples and isolates with limiting amounts of DNA pose a challenge for rapid and accurate genome sequence analysis, especially in urgent clinical situations where timely results are crucial. Working with degraded DNA samples becomes an issue since it could limit the sequence genomic coverage and assembly [11, 12]. While MDA effectively amplifies low DNA inputs, it presents significant challenges when applied to severely degraded DNA samples [12]. Specifically, MDA can negatively impact genomic analysis

through (i) reduced efficiency due to DNA breaks or lesions, leading to incomplete amplification; (ii) coverage bias resulting in uneven genomic representation; (iii) potential contaminant interference with the amplification reaction [13]; and (iv) the yield from MDA decreases as the length of the input DNA decreases [14].

Beyond these initial limitations, MDA introduces additional complexities in genomic sequencing, including (i) nonspecific amplification from primer dimer formation and template switching, (ii) chimeric DNA rearrangements and (iii) inherent genomic representation amplification bias that can compromise the accuracy and comprehensiveness of molecular analysis [9].

Some studies show that chimeric reads are usually inverted chimeras or direct chimeras, but it was previously observed that most of the detected MDA chimeric sequences (85%) are inverted chimeras, such as inverted sequences with intervening deletions, which can be caused by template switching [15, 16] These chimeric sequences are known to affect genome sequencing because they can be considered amplification artefacts, which cannot be used for genome assembly [16]. Studies suggest that chimerism in MDA sequencing data is a significant concern that is gaining attention, particularly with the rise of single-cell studies [17].

To address the challenges associated with artifactual concatemeric sequences generated during MDA, we developed a novel bioinformatic tool called CADECT (Concatemer Detection Tool), which is made available at https://github.com/rpbap/ CADECT. This tool enabled the identification and removal of putative inverted chimeric concatemers, thus improving the accuracy and contiguity of the genome assembly.

Our study aims to provide valuable insights into the use of MDA for whole-genome ONT sequencing, particularly for low molecular weight and/or low quantities of DNA samples,

highlighting its potential as a powerful method to obtain high-quality long-read sequencing data. We assessed the MDA advantages, constraints and effectiveness for whole-genome assembly in microbial organisms with genome sizes < 10 Mb. This is especially significant for infectious disease agents, where obtaining enough DNA can be challenging. Overall, our study underscores the potential of MDA in enabling high-quality long-read sequencing from challenging low-concentration DNA samples, emphasizing its importance in various genomic research and clinical applications.

Materials and Methods

Sample Collection and Preparation

A 100 ng DNA sample of *Cryptosporidium meleagridis* isolate TU1867 was obtained from BEI Resources (NR-2521) (Manassas, VA). DNA samples from cultured *Staphylococcus aureus* ATCC-29213, *Enterococcus faecium* TX-1330, and *Escherichia coli* strain K12, which were available in our laboratory, were used for testing. The bacterial DNA samples were prepared for downstream processing using the QIAGEN QIAamp DNA Mini Kit with lysozyme for Gram-positive samples and buffer ATL (tissue lysis buffer) for Gram-negatives. To assess sequence integrity, an *S. aureus* DNA sample aliquot was sheared using NEBNext dsDNA Fragmentase for 15min to generate fragments approximately 1000 bp in size. All DNA samples were quality controlled using a TapeStation (Agilent Technologies, Santa Clara, CA) and Qubit (ThermoFisher Scientific, Waltham, MA). In addition, we conducted serial dilutions on all samples to assess the limit of detection for amplification in the assay. The dilutions ranged from 2.5E-5 ng to 2.5 ng, allowing us to determine the minimum concentration at which successful amplification could be achieved.

Multiple Displacement Amplification (MDA)

Prior to MDA, the concentration of DNA was obtained using a Qubit fluorometer dsDNA high-sensitivity assay kit (ThermoFisher, Waltham, MA). For the *C. meleagridis* DNA, three different amounts were used as input for whole-genome amplification (i.e., 2, 5 and 10 ng) in a final volume of 5 µL. For the bacterial samples, MDA was performed on 2.5 ng of fragmented intact S. aureus DNA as well as serial dilutions of DNA from *E. faecium* ranging from 2.5 ng to 2.5E-5 ng. 400 ng of non-amplified DNA was used as an input control for the ONT rapid kit library preparation (Oxford Nanopore Technologies, Oxford, United Kingdom).
MDA was performed using the Qiagen repli-G kit (CAT #150023, Qiagen, Hilden, Germany), following the manufacturer's instructions. Following this, concentrations of DNA were obtained using a Qubit fluorometer dsDNA high-sensitivity assay kit (ThermoFisher, Waltham, MA).

For T7 Endonuclease I debranching, up to 42µL (i.e., all product from the MDA reaction) of MDA DNA was used as input (Catalogue #M0302, New England BioLabs, Ipswich, MA). After scaling up the reaction to accommodate a 42 µL input, all reaction components were added following the manufacturer's guidelines and incubated for 15 min at 37°C in a BioRad T100 thermocycler (BioRad, Hercules, CA). The incubated reaction was brought up to a final volume of 50 µL using TE buffer pH 8. AMPure XP beads (CAT# A63880) were prepared ahead of time following the manufacturer's instructions, and 35µL of beads were added to the reaction and mixed thoroughly. The bead-reaction mixture was placed on a rotator mixer (e.g., Hula mixer) for 10min at room temperature. Following this, the bead-reaction mixture was spun down and placed on a magnet until the eluate was clear and colourless. With the bead-reaction mixture on the magnet, the clear supernatant was pipetted off and 200 µL of freshly prepared 70% ethanol

was carefully added not to disturb the pellet (i.e., wash step). The wash step was repeated one time, for a total of two washes. After removing the supernatant from the second wash, 49µL of water was used to resuspend the pellet, which was immediately incubated for 1 min at 50°C in a BioRad T100 thermo cycler, followed by 5 min at room temperature. The bead-reaction mixture was placed back on the magnet, and 49µL of the elute was transferred to a sterile 1.5 mL tube. Concentrations of DNA were obtained using a Qubit fluorometer dsDNA high-sensitivity assay kit (ThermoFisher, Waltham, MA). A summarised diagram of the protocol is provided (Appendix A: Supplementary Figure 2).

Whole-Genome Sequencing and Assembly

Sequencing of the amplified DNA samples was performed using the ONT SQK-RBK110.96 kit for library preparation R9.4.1 MinION flow cells (Oxford Nanopore Technologies, Oxford, UK). The amplified DNA samples were prepared according to the kit instructions and loaded onto the flow cell for sequencing following the manufacturer's instructions. Sequencing was carried out in Mk1B and GridION MK1 devices for 72h, and the resulted fast5 files were base-called using guppy v6.3.7 using the high accuracy model (dna_r9.4.1_450bps_hac). By sequencing for 72h, we aim to recover as much data as possible. The real-time sequencing ca- pability of ONT allowed us to access data as it is generated, which is particularly advantageous in urgent cases like clinical assays. When it is necessary to pellet cultured bacteria to obtain sufficient DNA, this approach enables sequencing without requiring prolonged cultivation periods. Furthermore, we are able to commence data analysis as soon as the sequencing reaches the desired depth, as monitored by the MinKNOW ONT software. Flye 2.9 [18] was used for assembly. For samples with > 100× coverage, the ‘-asm -coverage 100’ parameter was used to improve assembly and facilitate the assembler's performance. We

then used Nextpolish 1.4.1 [19] to increase the overall base call quality of the genome and facilitate further quality control analysis such as Benchmarking Universal Single-Copy Orthologs (BUSCO) scores and better gene annotation. Illumina sequencing was not used here because the objective of this research was to determine how MDA would affect long-read generation.

Putative Concatemer Detection in Intact Versus Fragmented Amplified Samples

To examine the impact of fragmentation on MDA products, we treated DNA aliquots with NEBNext dsDNA Fragmentase (CAT# M0348S) at 37°C for 16 min, creating fragments between 500 and 1000bp. This enzyme-based method induces DNA shearing, generating fragments of specified sizes in a time- dependent manner. The process provides random fragmentation similar to mechanical methods. Both fragmented and non-fragmented (high molecular weight DNA) samples were sequenced as described above.

The CADECT tool (https://github.com/rpbap/CADECT) was developed in-house and was used for the detection and removal of putative concatemeric chimeric sequences in the ONT amplified reads. CADECT splits all reads into separate files and performs sliding windows with a user-defined preferred size and gap between windows. For ONT amplified reads, a window size of ≥500bp with no overlaps was used (e.g., −w 500 and -s 500). Reads generating less than one window (<1kb in size in the 500bp window example) were skipped, and their IDs were stored in the short.txt output file. Fragment windows from reads with more than two windows were aligned using nucmer from mummer 4 [20], and reads with overlaps were reported in the stats file, with their IDs stored in the concat_ID output file. Statistics including the total number of reads, number of putative concatemers, number of reads with no concatemer detection and

overlap frequency were recorded in the stats.txt output file. Fastq/Fasta files containing the characterized reads were generated for further analysis.

These methods were employed to investigate the benefits and limitations of MDA in whole-genome ONT, focusing on low-concentration DNA samples.

Results

MDA Efficiency Across Different Species

Our whole-genome amplification (WGA), using MDA, results reveal that in each sample type tested, we find an overall fold change of >500× in comparison to the input sample DNA (Table 1). Following amplification, approximately 1.5µg of the product was debranched using T7 endonuclease prior to library preparation for ONT sequencing. Usually, in our assays, we obtained an ~45% recovery after this step, with the loss mostly attributed to the bead purification process (Table 1). Although a large amount of DNA is lost during the DNA purification step post-T7 endonuclease reaction, an overall fold change of around 100× is observed compared to the MDA DNA input. The variation observed in the T7 recovery is affected by pipetting and the Ampure XP magnetic bead ratio that is set based on the DNA integrity from the QC. In this manuscript, we used a fixed ratio for all samples.

We successfully obtained contiguous and sometimes even chromosomal-level assemblies from the samples analyzed in this study, starting with MDA DNA inputs much lower than Oxford Nanopore's recommended minimum of 50 ng and 200 ng for the rapid barcode kit (RBK) for old and new library prep kits, respectively (Appendix A: Supplementary Table 1).

For certain samples, such as *E. faecium*, we observed that achieving improved contiguity required generating higher depth coverage during the sequencing. Our results indicate that, for this organism, reaching depths beyond 70× allowed us to attain a chromosomal-level assembly

with only 2.5ng of starting total DNA (Appendix A: Supplementary Table 2). In comparison, increased sequencing depth on samples that started with less than 0.001ng of input into the MDA did not enhance contiguity. Combining separate MDA amplifications of the same limited input samples did improve the final genome coverage because the random nature of the initial templates and amplification process. Thus, multiple independent MDAs appears to be advantageous because it could randomly amplify by chance different regions that are beneficial for the genome assembly.

MDA Using Serially Diluted Samples

Serial dilutions of a single *E. faecium* sample reveal successful DNA amplification even at a low initial DNA amount of 2.5E-5 ng (Figure 1A). The MDA technique imposes a size limit on its amplified products, with an average product length of 10–12kb [6]. The debranching step prior to library prep also leads to a reduction in the mean sequence read sizes (Appendix A: Supplementary Table 3). Post MDA, the average size of the reads observed in our study was within the range of 2–3kb. In contrast, standard Oxford Nanopore Technologies (ONT) assays without amplification (i.e., ONT RBK), which include a transposase step that simultaneously cleaves template molecules and attaches tags to the cleaved ends, are expected to generate DNA fragments ranging from 5 to 20kb. When assessing genome coverage from the tested samples, such as *E. faecium*, we observed that DNA inputs below 0.025 ng result in incomplete coverage of certain genomic regions that are distributed randomly across the genomes tested (Figure 1B). The same is expected for the other samples [8, 21, 22].

MDA Results in an Unexpected Size Increase From Fragmented DNA Samples

Following controlled enzymatic fragmentation using a dsDNA fragmentase (to simulate degraded input DNA) and MDA according to the repli-G manufacturer protocol, we observed an

unexpected size-increase distribution of fragments (Figure 2A). Indeed, for all samples, except *Cryptosporidium*, the size distribution post-MDA was nearly identical for intact and fragmented input DNA. Subsequent analysis of the ONT sequencing results revealed the existence of longer read lengths than those present in the same sample without MDA amplification (Figure 2B). When analyzing the sequence content of the run (all reads generated), two distinct types of reads were identified. Some represented potentially low-abundance longer reads that escaped fragmentation during the enzyme incubation and were subsequently amplified. The other reads were primarily chimeric concatemers, likely generated through template switching of short fragments during MDA (Figure 3A). While the occurrence of concatemers in MDA assays has been reported previously [16, 23], they are typically present in low amounts after sequencing. In our case, the fragmentation process seemed to enhance the prevalence of these chimeric reads in our ONT sequencing. As expected, assembly of the data revealed that the presence of these chimeric/concatemer regions significantly impacted genome assembly, resulting in a bubble fragmentation effect across the entire genome and affecting contiguity (Figure 3B).

Concatemer Detection Tool

To identify and eliminate potential concatemers generated by MDA, we designed a concatemer detection tool specifically tailored for raw ONT reads called CADECT. This tool enables the differentiation of putative concatemeric chimeric reads from non-concatemeric ones. To achieve this, the process involves dividing each long-read sequence into multiple fragments using a sliding window approach and then aligning these fragments with one another using Biopython global pairwise2 (pairwise2.align.globalxx). The underlying hypothesis is that the presence of a concatemer would result in certain windows aligning with each other (presence of

alignments with global alignment coverage greater than 0.7), thereby confirming the existence of a potential concatemer or tandem repeat within the sequenced read. No overlapping detection under the same parameters as mentioned above would be identified as non-concatemeric. Reads with lengths less than twice the given window size are categorized and stored as short reads. Additionally, it incorporates a size selection mechanism to isolate longer reads, thereby streamlining the genome assembly process and enhancing contiguity.

Following evaluation of the CADECT pipeline on fragmented DNA and a comparative analysis of results pre- and post-amplification assays (Appendix A: Supplementary Table 1), we confirmed that the final genome assembly exhibited significantly reduced fragmentation. The integration of MDA and CADECT proved to be effective, particularly in handling challenging, low-quantity DNA samples. This combination facilitated the generation of nearly complete genome assemblies with depths above 70× (Figure 4; Appendix A: Supplementary table 4).

Overall, when comparing the data before and after CADECT using default parameters with a 500 base window size, we observe that it is a stringent process, which separates putative concatemers and shorter reads and tends to affect the average final depth of the final input. Specifically, in the case of *S. aureus*, we note that for high-quality intact amplified DNA, the detection of putative chimeras concatemers and size selection decreases coverage by 40%, whereas for amplified fragmented samples, it decreases coverage by 50% (Appendix A: Supplementary Table 5). Due to size selection, the effect on depth is more pronounced for fragmented samples.

Discussion

Our study demonstrates that MDA offers a promising solution for amplifying low amounts of DNA from precious samples for ONT sequence generation with the ONT rapid

barcode sequencing Library kit (RBK). In this study, we demonstrated key points: (A) Using our method, we can successfully sequence samples with DNA inputs as low as 0.025 pg, corresponding to approximately 1.5 times the DNA content of a single *Cryptosporidium* oocyst (genome size ~9Mb; Table 2). For bacteria with genome sizes ranging from 3 to 7 Mb, this DNA amount would represent contributions from more cells. However, as shown in Figure 1 for bacteria, achieving whole genome and plasmid assemblies with whole-genome coverage typically requires approximately 2.5 pg. of input DNA, equivalent to more haploid genome equivalents. This demonstrates that while our approach has high sensitivity, generating complete genome assemblies is more reliable when working with DNA from more cells. (B) Single-oocyst sequencing minimizes the variability associated with bulk sequencing approaches, providing more consistent genomic insights. Nonetheless, for applications requiring complete assemblies, starting with higher DNA concentrations is advisable to ensure robust results. (C) MDA using ONT RBK kits shows a good option to decrease the library prep time and cost, generating closed bacterial genomes and chromosomal-level genomes with a limited amount of DNA input. Here, we have only examined organisms with genome sizes < 10 Mb. Larger genome sizes will require additional starting material, and smaller genome sizes should have success with even less input DNA.

We were able to obtain whole-genome sequences at the chromosomal level for almost all tested organisms when generating depth coverages > 70×. This indicates that we can overcome the challenge of the relatively shorter reads that MDA with T7 debranching generates compared to reads generated from higher DNA input without amplification. Although the V14 RBK kits for R10.4.1 flow cells were not available at the time experiments were conducted, we have subsequently used the updated ONT V14 chemistry and confirmed that the protocol yields repro-

ducible results, consistent with our earlier findings (data not shown). It is essential to highlight that highly complex regions, such as repetitive regions with tandem repeats larger than the window size used for CADECT detection, might be identified as concatemeric reads. This occurs because the tool detects repeat overlaps, which can lead to their exclusion before the assembly process, potentially causing some regions of the genome to remain fragmented. Thus, the judicious use of CADECT, including optimizing the window size, is warranted depending on the size and structure of repeats in the genome of the organisms being sequenced and assembled. At a lower amount of initial DNA, we observed that the amplification appears to be random, exhibiting no apparent coverage amplification bias across the genome (Figure 1). At extremely low DNA amounts, achieving full coverage of the target genome may be challenging, but the method remains valuable for potential taxon identification and may prove effective for the identification of plasmids as well (Figure 1). While the effectiveness in metagenomic samples requires further evaluation, there is promise in using this approach for taxon identification. Extremely low-abundance samples tend to produce patchy sequence information. Thus, although extremely low-concentration samples provide valuable sequence information, they also lack coverage in many regions, which impacts the assembly process and the ability to produce full genomic information. Combining multiple MDA replicates is likely to increase the chances of amplifying more regions and thus will be more likely to enhance genome coverage versus deeper sequencing.

In the case of sheared DNA, the higher impact on the depth after CADECT is primarily related to loss in the size selection pipeline (Appendix A: Supplementary Table 5). However, our concatemer detection tool, CADECT, effectively identified and removed several concatemers, facilitating the assembly and yielding good results. This highlights the importance of

bioinformatic tools in overcoming challenges associated with amplification artefacts, thus improving the accuracy of genome assembly.

It is worth noting that the CADECT pipeline will remove a significant number of reads, which will impact depth for an optimal genome assembly. If there is not sufficient coverage obtained post-CADECT run, an alternative is to merge the reads identified as short by the program with the non-concatemeric reads. As observed previously, the chimeric rate produced by MDA is positively associated with the mean read length [16], indicating a decreased likelihood of chimeric reads in this short dataset. Consequently, this dataset is less likely to negatively impact the assembly process. In more complex genome sequences that are rich in repeats, further investigation is required to address these regions effectively and be able to distinguish concatemers from genuine repetitive patterns within the genome. As a solution, the CADECT pipeline generates a separate concatemer fastq file. This file includes putative concatemeric regions as well as true repeats. For example, in highly repetitive genomes such as trypanosomatids with an ~50% genomic repeat content [24], CADECT would detect a high number of reads containing real tandem repeats in the genome as putative concatemers, which would result in a higher impact on coverage depth loss and also impact the genome content of the organisms used for assembly. To mitigate this, we recommend incorporating a repeat identification step into the pipeline, such as using RepeatModeler [25] trained with the organism of interest on the putative concatemer generated sequence file from CADECT. This additional step would enhance the recovery of information and data for the subsequent assembly process. Moving forward, it is crucial to continue exploring the potential of MDA in various biological contexts and optimize the amplification protocol to minimize potential amplification biases and errors. Additionally, considering the clinical applications of MDA, further research and

development of rapid and reliable sequencing approaches are necessary to unlock its full potential in diagnosing and monitoring infectious diseases and other clinical applications.

Acknowledgements

This work was funded in part by the National Institute for Allergy and Infectious Diseases, NIAID, R01 AI148667 to T.C.G. and J.C.K., NIH T32GM142623 to F.A.-D. The funders played no role in the study design, data collection and analysis, decision to publish or preparation of the manuscript.

References

1. Slatko, B.E., A.F. Gardner, and F.M. Ausubel, *Overview of next-generation sequencing technologies.* Current protocols in molecular biology, 2018. **122**(1): p. e59.
2. Hu, T., et al., *Next-generation sequencing technologies: An overview.* Human immunology, 2021. **82**(11): p. 801-811.
3. Pacbio, <*Technical-Note-Preparing-DNA-for-PacBio-HiFi-Sequencing-Extraction-and-Quality-Control.pdf*>. 2022.
4. Delahaye, C. and J. Nicolas, *Sequencing DNA with nanopores: Troubles and biases.* PloS one, 2021. **16**(10): p. e0257521.
5. Hou, Y., et al., *Comparison of variations detection between whole-genome amplification methods used in single-cell resequencing.* Gigascience, 2015. **4**(1): p. s13742-015-0068-3.
6. Dean, F.B., et al., *Comprehensive human genome amplification using multiple displacement amplification.* Proceedings of the National Academy of Sciences, 2002. **99**(8): p. 5261-5266.
7. Fullwood, M.J., et al., *The use of multiple displacement amplification to amplify complex DNA libraries.* Nucleic acids research, 2008. **36**(5): p. e32.
8. Troell, K., et al., *Cryptosporidium as a testbed for single cell genome characterization of unicellular eukaryotes.* BMC Genomics, 2016. **17**: p. 471.
9. Binga, E.K., R.S. Lasken, and J.D. Neufeld, *Something from (almost) nothing: the impact of multiple displacement amplification on microbial ecology.* The ISME journal, 2008. **2**(3): p. 233-241.

10. Lasken, R.S., *Genomic DNA amplification by the multiple displacement amplification (MDA) method.* Biochemical Society Transactions, 2009. **37**(2): p. 450-453.

11. Ceccherini, M., et al., *Degradation and transformability of DNA from transgenic leaves.* Applied and environmental microbiology, 2003. **69**(1): p. 673-678.

12. Dabney, J. and M. Meyer, *Extraction of highly degraded DNA from ancient bones and teeth*, in *Ancient DNA: methods and protocols*. 2019, Springer. p. 25-29.

13. Wang, G., et al., *DNA amplification method tolerant to sample degradation.* Genome research, 2004. **14**(11): p. 2357-2366.

14. Lage, J.M., et al., *Whole genome analysis of genetic alterations in small DNA samples using hyperbranched strand displacement amplification and array–CGH.* Genome research, 2003. **13**(2): p. 294-307.

15. Lasken, R.S. and T.B. Stockwell, *Mechanism of chimera formation during the Multiple Displacement Amplification reaction.* BMC biotechnology, 2007. **7**(1): p. 19.

16. Lu, N., et al., *Chimera: The spoiler in multiple displacement amplification.* Computational and Structural Biotechnology Journal, 2023. **21**: p. 1688-1696.

17. Hård, J., et al., *Long-read whole-genome analysis of human single cells.* Nature Communications, 2023. **14**(1): p. 5164.

18. Kolmogorov, M., et al., *Assembly of long, error-prone reads using repeat graphs.* Nature biotechnology, 2019. **37**(5): p. 540-546.

19. Hu, J., et al., *NextPolish: a fast and efficient genome polishing tool for long-read assembly.* Bioinformatics, 2020. **36**(7): p. 2253-2255.

20. Marçais, G., et al., *MUMmer4: A fast and versatile genome alignment system.* PLoS computational biology, 2018. **14**(1): p. e1005944.

21. Chen, Y., et al., *Chromosome-level genome assembly of Cryptosporidium parvum by long-read sequencing of ten oocysts.* Sci Data, 2024. **11**(1): p. 1287.

22. Penumarthi, L.R., et al., *A new chromosome-level genome assembly and annotation of Cryptosporidium meleagridis.* Scientific Data, 2024. **11**(1): p. 1388.

23. Paul, P. and J. Apgar, *Single-molecule dilution and multiple displacement amplification for molecular haplotyping.* Biotechniques, 2005. **38**(4): p. 553-559.

24. El-Sayed, N.M., et al., *The genome sequence of Trypanosoma cruzi, etiologic agent of Chagas disease.* Science, 2005. **309**(5733): p. 409-415.

25. Flynn, J.M., et al., *RepeatModeler2 for automated genomic discovery of transposable element families.* Proceedings of the National Academy of Sciences, 2020. **117**(17): p. 9451-9457.

Tables

Table 1: Observed amplification yield increase by sample type

		MDA input (ng)	MDA output (ng)	T7 output (ng)	T7 recovery (%)	Estimated DNA output Fold Increase
Gram-positive	*S. aureus*	2.5	1500	894	59.6	357.6×
	E. faecium	2.5	1430	749	52.4	299.6×
Gram-negative	*E. coli*	5.0	8360	552	36.8	110.0×
Eukaryotic Pathogen (*Cryptosporidium* ssp.)		2.5	1976	555	44.1	222.0×
Positive control (Calf thymus) (*Bos taurus)*		5.0	3800	620	41.33	124.0×

SAMPLE	Estimated genome size (Mb)	Amt of DNA/cell (fg)
E. coli	5.00	5.43
S. aureus	2.81	3.03
E. faecium	2.91	3.14
Cryptosporidium*. ssp. (oocysts)	9.2	39.70
***Cryptosporidium* ssp. (sporozoite)**		9.90

Table 2: Estimated DNA concentration in a single cell of the organisms studied in this project.

*One *Cryptosporidium* ssp. oocyst contains four haploid oocytes.

Figures

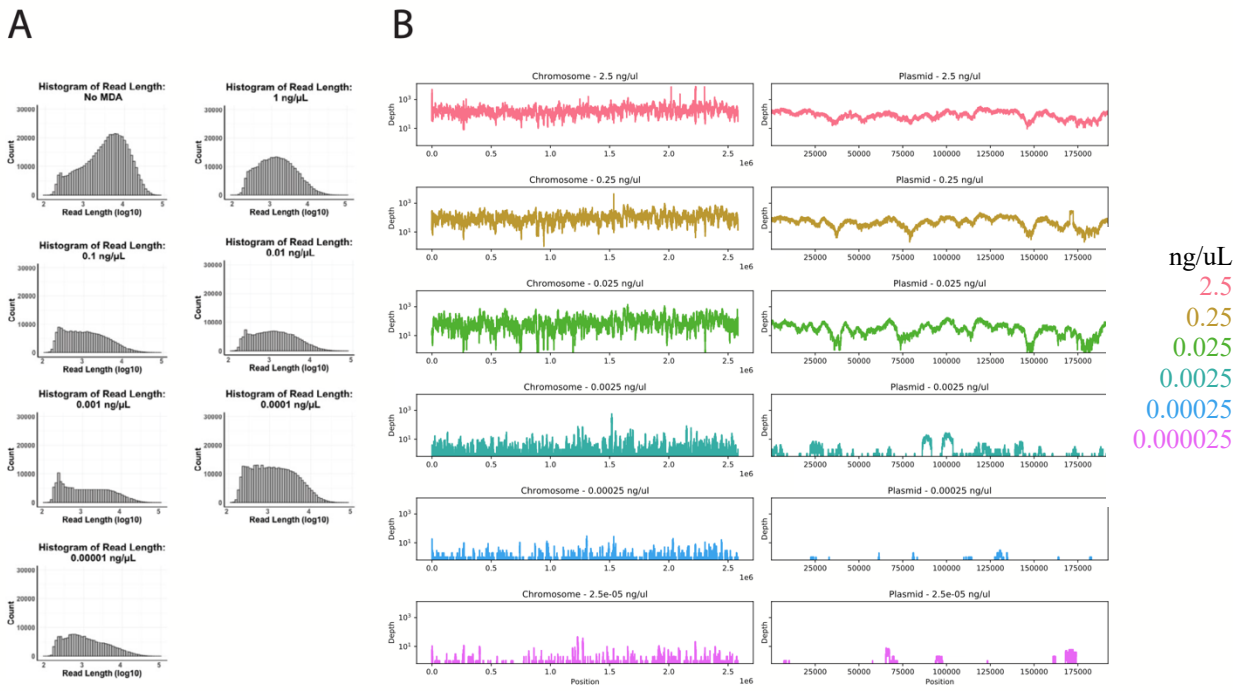


Figure 1. Serial dilution test with *E. faecium* sample DNA. Panel (A) depicts the distribution and counts (y-axis) of read lengths (x-axis) across all diluted samples sequenced after MDA. The control sample had 400 ng used to see how a normal run in a sample without MDA would look like on read length distribution for comparison with the MDA diluted samples. Panel (B) illustrates the breadth of coverage of the chromosomal and plasmid regions across all diluted and subsequently amplified samples, with read depth on the y-axis and genome position on the x-axis.

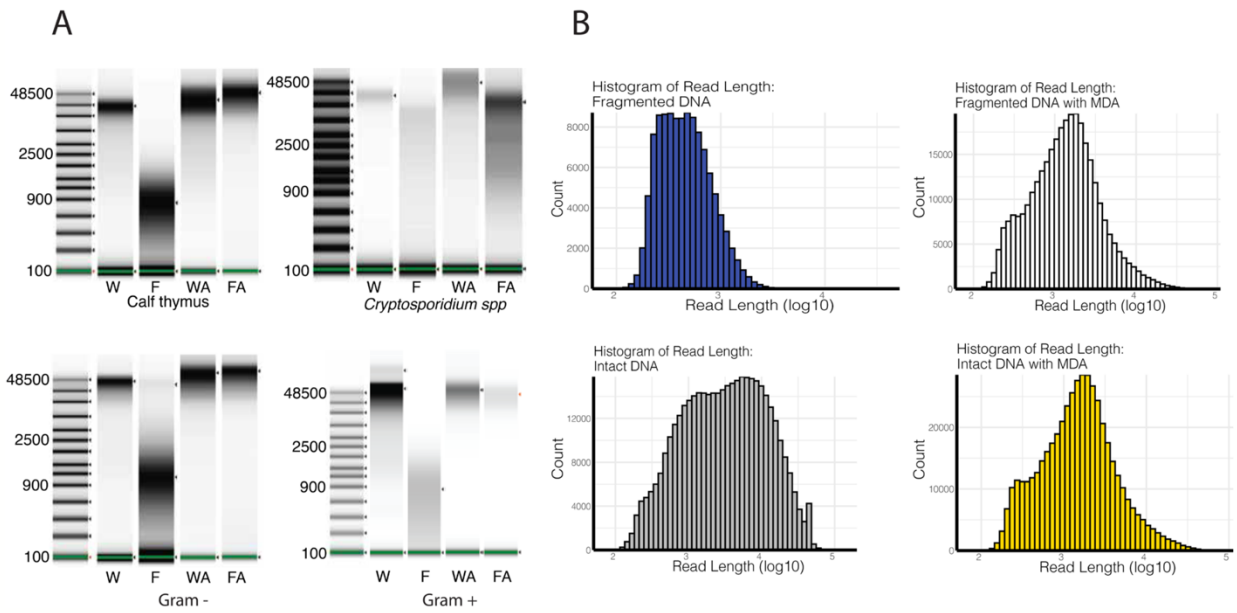


Figure 2. DNA fragment and read size range pre- and post-multiple displacement amplification using size-controlled fragmented DNA. (A) TapeStation results for different organism sets. F=Fragmented DNA; W=whole intact DNA; WA and FA=after amplification. Uncropped TapeStation results are in Appendix A: Supplementary Figure 1; (B) ONT sequencing read length results obtained before and after amplification for *S. aureus*.

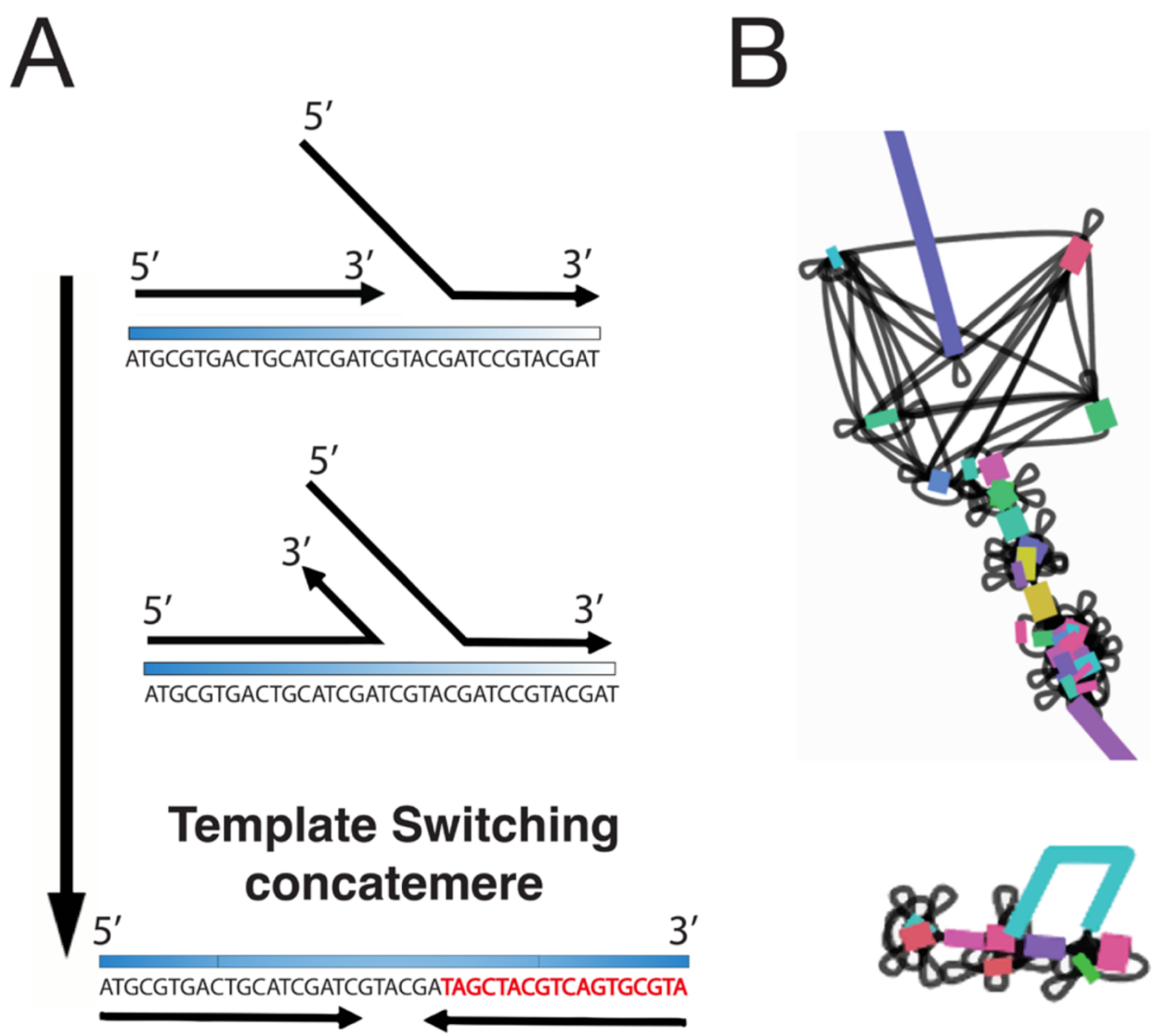


Figure 3. Impact of MDA-generated concatemers on the genome assembly. (A) Concatemers generated by template switching; (B) Graph representation of the effect of concatemers on genome assembly (bubble fragmentation effect). Visualization made using BANDAGE v.9.0.

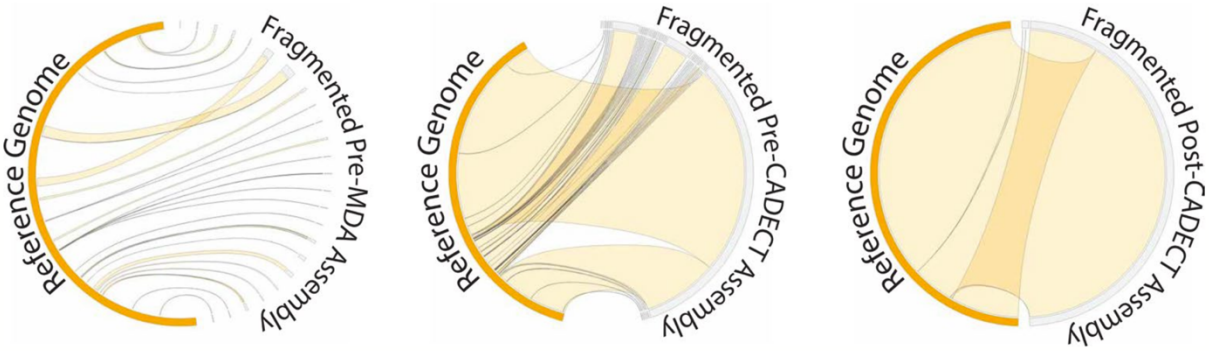


Figure 4. Circos plot illustrating a synteny comparison between the reference *S. aureus* ATCC-29213 genome sequence and pre- and post- amplification genome assemblies. The Circos plots contrast the assemblies resulting from the amplified fragmented DNA before and after CADECT processing. A comparison between the reference genome and (A) genomic assembly of fragmented DNA sample without MDA, (B) genomic assembly of fragmented DNA sample with MDA before CADECT and (C) genomic assembly of fragmented DNA sample with MDA after CADECT. Grey lines are potential secondary alignments (multicopy region, repeats or concatemers).

CHAPTER 3

SINGLE-OOCYSTS SEQUENCING ENABLES DETECTION OF RECOMBINATION IN DRUG-SELECTED *CRYPTOSPORIDIUM* HYBRIDS

Although I am not the first author of the publication, I have received permission to discuss my contribution.

*Shaw, Sebastian, Ian S. Cohn, Rodrigo P. Baptista, Guoqin Xia, Bruno Melillo, **Fiifi Agyabeng-Dadzie**, Jessica C. Kissinger, and Boris Striepen. "Genetic crosses within and between species of *Cryptosporidium*." *Proceedings of the National Academy of Sciences* 121, no. 1 (2024): e2313210120.

Author contributions: S.S., G.X., B.M., and B.S. designed research; S.S., I.S.C., R.P.B., and F.A.-D. performed research; R.P.B., G.X., B.M., F.A.-D., and J.C.K. contributed new reagents/analytic tools; S.S., R.P.B., F.A.-D., J.C.K., and B.S. analyzed data; and S.S. and B.S. wrote the paper.

Abstact

Cryptosporidium is an obligate apicomplexan parasite that undergoes sexual reproduction, making it an excellent model for studying recombination. Although co-infections with multiple *Cryptosporidium* strains have been observed, direct evidence of interspecies recombination has been lacking. This study provides the first experimental confirmation of recombination between two *Cryptosporidium* species, *C. parvum* and *C. tyzzeri,* using a controlled genetic cross. By using a novel dual-drug-selectable marker system developed by Shaw *et al.* (2024), we were able to identify recombinant progeny.

To validate the occurrence of recombination and not only at the selectable markers, I applied single-oocyst Oxford Nanopore long-read sequencing, which enabled precise detection of recombinant haplotypes and crossover regions. Despite challenges with low DNA input and unequal amplification of genetic material across the genome, high-accuracy long reads revealed that hybrid progeny inherited large, mosaic chromosomal regions from both parents, confirming genome-wide recombination.

These findings demonstrate that species boundaries in *Cryptosporidium* are more blurred than previously thought and highlight sexual recombination as one of the key drivers of genetic diversity and adaptation. This approach lays the groundwork for future functional studies and advances genetic tools for investigating virulence, host specificity, and therapeutic targets.

Introduction

Cryptosporidium is an obligate intracellular apicomplexan parasite and a major cause of diarrheal disease, particularly in immunocompromised individuals and children under two years of age living in low- and middle-income countries [26, 27]. The World Health Organization ranks *Cryptosporidium* as the fifth most important foodborne parasite [28, 29]. With a host range spanning more than 40 animal species [30-32], *Cryptosporidium* is responsible for over 133,000 human deaths annually and more than 8 million disability-adjusted life years (DALYs), surpassing the global burdens of schistosomiasis and dengue [33].

The parasite's life cycle includes both asexual and sexual stages. Asexual reproduction begins with the excystation of sporozoites, followed by three rounds of mitotic division, leading to the production of gametes. Sexual reproduction involves fertilization and meiosis, resulting in the formation of an oocyst containing four haploid sporozoites [34, 35]. Recombination, a critical step during meiosis, is a fundamental evolutionary mechanism that reshuffles genetic material, increases diversity, and enables populations to adapt to environmental pressures and immune responses [36]. For *Cryptosporidium*, which possesses a compact genome and limited metabolic plasticity [37, 38], generating diversity through recombination is especially vital. Genetic exchange allows for the emergence of novel phenotypes, altered host specificity [39, 40], and increased transmissibility [41, 42]. While population genomic studies have suggested that recombination plays a major role in shaping the parasite's diversity [40, 43], direct experimental evidence has been lacking. This is mainly due to experimental challenges, most notably, the parasite's inability to complete its life cycle in vitro and the historical limitation of having only one drug-selectable marker (neomycin phosphotransferase, neo) for isolating recombinant progeny.

To address this knowledge gap, Dr. S. Shaw and colleagues at the Striepen Lab (University of Pennsylvania) developed a novel dual-drug selection system by introducing phenylalanyl-tRNA synthetase (*pheRS*) as a second selectable marker. A single-nucleotide mutation in *pheRS* confers resistance to the antimalarial compound BRD7929 [44, 45]. Combined with the existing neo marker for paromomycin resistance, this dual-marker system enables robust selection of recombinants, even in interspecies crosses.

In this chapter, I will elaborate on a groundbreaking study that explores recombination in *Cryptosporidium* using genetic crosses and Single-Oocyst Oxford Nanopore (ONT) Sequencing (Shaw *et al.* 2024). I collaborated with the Striepen Lab as part of the Kissinger Lab to investigate recombination between *Cryptosporidium parvum* and *Cryptosporidium tyzzeri*. We performed a genetic cross and used dual-drug selection to isolate recombinant progeny. To confirm recombination and characterize the genomic structure of individual parasites, I employed single-oocyst sequencing, a technique that enables precise mapping of parental contributions at the individual level. Despite the typical limitation of low genome coverage associated with this method, it provided strong evidence for interspecies recombination. Together with bulk sequencing and molecular assays, our work establishes a powerful platform for studying genetic exchange, hybrid viability, and adaptation in *Cryptosporidium*.

Materials and methods

To evaluate the efficiency of the proposed dual drug-selectable marker system in facilitating genetic crosses and isolating recombinant progeny in *Cryptosporidium*, *C. parvum* and *C. tyzzeri* were used as parental strains. Each parent was transformed with a unique drug-resistant marker, *i.e., C. parvum* expressed the BRD7929-resistant mutated *phreRS* gene while *C. tyzzeri* expressed the paromomycin-resistant *neo* gene. The transgenic strain generation, drug

treatment and selection protocols were performed by the Striepen Lab as described in the Shaw *et al.*, 2024 manuscript [46]. Following the isolation of putative recombinant oocysts, I performed single-oocyst sequencing to confirm the recombinant identity by assessing the genomic contribution from both parental species.

DNA preparation and single-oocyst sequencing

To confirm the identity of recombinant progenies, I sorted individual putative recombinant oocysts into 96-well plates containing 1 µL of TE buffer using the MoFlo XDP-Astrios EQ cell sorter (Beckman Coulter, CA, USA) with a 0.5-drop envelope [8]. *C. parvum* parental strains expressed mNeonGreen, while *C. tyzzeri* strains expressed tdTomato. Sorting was based on the detection of both fluorescent markers, enabling the identification of double-positive oocysts representing putative recombinants. Fluorescence-based gating was visualized and managed using Summit 6.3.1, MoFlo Astrios software (Beckman Coulter). Given the expected low DNA concentrations from single-oocyst *Cryptosporidium* isolates, I performed whole genome amplification (WGA) [47, 48] on the extracted DNA using a modified REPLI-g Mini/Midi protocol (Qiagen). This employed multiple displacement amplification [48] (MDA) optimized for Oxford Nanopore Technologies (ONT) rapid kits. This approach was performed on 15 individual putative recombinant progeny oocysts (Table 1). To ensure sequencing compatibility, the amplified DNA was debranched using T7 endonuclease (New England Biolabs) prior to ONT long-read sequencing. Sequencing of the putative recombinant single oocysts was conducted using ONT GridION R10 flow cells with the Rapid Barcoding Kit SQK-RBK114.24**.** We also processed and sequenced a bulk sample of putative recombinant progeny oocysts to serve as a higher-coverage reference, allowing for validation of recombination patterns and comparison against single-oocyst data.

Processing of single-oocyst sequence reads

To assess the quality of the reads from the progeny, I used Nanoplot v1.42 [49]. Reads were then mapped to a modified *C. parvum* BGFT2T genome assembly as reference [50] using Minimap v2.24 [51]. The modified reference included the sequences of the selectable markers from each parent at the orthologous loci (chromosomes 3 and 5). Short-read sequences of the parent species were also assessed using seqkit v2.9 [52] and mapped to the modified *C. parvum* BGFT2T genome assembly with BWA v0.7 [53, 54].

To detect single-nucleotide polymorphisms (SNPs) in both parental species and putative recombinant progeny, I used BCFtools v1.10 [55] with parameters appropriate for a haploid organism (--ploidy 1), a minimum base quality score (QUAL> 40), mapping quality (MQ> 60), and read depth > 10. The distribution of SNPs across recombinant regions was visualized using the Integrative Genomics Viewer (IGV) [56] with allele frequency set to 0.5. To further investigate recombination, BAM files from the parental and progeny samples were filtered to extract reads mapping to recombinant regions and subsequently converted to FASTQ format using SAMtools [57]. The long reads from single-oocyst progeny were then aligned to the parental contigs, and multiple sequence alignments were generated using Clustal Omega within Geneious Prime 2022.2.2 (https://www.geneious.com) using default settings.

Results

Processing of putative recombinant oocysts

To confirm the identity of the putative recombinant progeny, I employed a two-step approach: (1) sorting individual oocysts based on the presence of dual fluorescence markers (mNeonGreen and tdTomato), and (2) validating the recombinant state at the nucleotide level. Sorting of double-positive single oocysts into 96-well plates was successful (Fig. 1), yielding

over 30,000 recombinant oocysts confirming the isolation of potential hybrid progeny. DNA extraction followed by whole-genome amplification (WGA) was performed on each oocyst with a success rate of 11 out of 15 (73%) samples yielding measurable DNA (Table 1). Although DNA amplification was successful, amplification was not uniform across the genome (Appendix B: Supplementary Figure 1). Sequencing was conducted on the seven highest-yielding individual oocyst DNA samples using Oxford Nanopore Technologies (ONT), generating long reads with an average length of 4,297 bases (Table 2), which enabled downstream alignment and analysis of recombination events. Additionally, Illumina whole-genome sequencing was performed on bulk DNA from each parental line, and WGA-ONT sequencing was also conducted on bulk DNA from the putative recombinant progeny. The bulk sequencing of progeny DNA provided essential positive control for identifying parental haplotypes and validating recombination patterns observed in the single oocyst data.

Calling of variants

Following alignment to the modified BGFT2T reference genome, I observed that both parental samples and the bulk Oxford Nanopore (ONT) sequencing data from putative recombinant progeny exhibited near-complete genome coverage. Specifically, each achieved approximately 99% genome coverage at a minimum sequencing depth of 10×, indicating high-quality and uniform mapping across the genome. In contrast, coverage among the single-oocyst sequenced progeny was more variable. Although all seven individual samples had an average genome coverage exceeding 20% at 1×, only three displayed >15% genome coverage at a 10× read depth threshold (Fig. 2; Table 3) and were used in subsequent analysis (namely, progeny 1a, 2a, and 7b). The *C. tyzzeri* parental sample yielded a total of 204,860 variants, including 197,256 single-nucleotide polymorphisms (SNPs), whereas the *C. parvum* parent contained 1,301

variants, of which 1,037 were SNPs. Across all samples, over 90% of SNPs were biallelic, with the exception of the *C. parvum* parent, which had a lower percentage (75%) of biallelic SNPs (TableTable 4).

Recombinant detection in IGV

To visually identify recombination events based the SNP data from selected progeny (progeny 1a, 2a and 7b), I used IGV. The observed patterns confirmed both parental inheritance and recombination. Some chromosomes in the progeny matched one parent almost entirely, while others resembled the second parent. In some cases, individual chromosomes showed a mix of regions from both parents, forming mosaic patterns that suggest recombination between the two species. Notably, we found evidence of genetic exchange between two closely related species, *C. parvum* and *C. tyzzeri*. For example, one progeny inherited most of chromosome 2 from *C. parvum*, while chromosome 5 was almost entirely from *C. tyzzeri*. Chromosome 4, in particular, showed a clear mosaic structure, with alternating segments from both species (Fig. 3). These findings support the occurrence of genome-wide recombination beyond just the regions targeted by selectable markers. On certain occasions, the data from the bulk ONT sequences of the putative recombinant progeny were not in agreement with the single oocyst sequences. Multi-sequence alignment of possible recombinant region in Geneious showed evidence of crossing over occurring (Fig 4).

Discussion

The fact that *Cryptosporidium* is a single-host parasite that undergoes obligatory sexual reproduction makes it an ideal model for studying recombination, especially given that its meiosis follows the Mendelian inheritance pattern [58]. Interestingly, multiple studies have reported that different strains of *Cryptosporidium* can be found within a single host [59]. When

you consider these two facts together, it raises several important questions. First, can a single host be infected with more than one *Cryptosporidium* species at the same time? And if so, could recombination occur between them? Second, in regions where *Cryptosporidium* is endemic and reinfection is common, are people being infected repeatedly by the same strain, or is the parasite evolving alongside the host's immune system? In such areas, many different strains may circulate, and recombination could allow traits that promote the parasite's survival in the environment or increased virulence to be shared among them. These questions are even more important when we consider that some *Cryptosporidium* species, like *C. parvum, C. hominis*, and *C. tyzzeri,* are very closely related [60]. Third, to truly understand whether recombination is recent or ancient (introgression), and to test whether it occurs between species, controlled genetic crosses are necessary. All these questions can now be explored efficiently due to the development of the dual drug-selectable marker system [46]. By using two different drug-resistance genes (like *phreSR* and *neo*), researchers can accurately identify and isolate recombinant progeny that have inherited genetic material from both parents. This dual selection approach greatly reduces the chance of false positives and improves the precision of genetic crossing experiments.

While dual drug-selectable markers such as *neo* and *phreSR* are effective for isolating recombinant *Cryptosporidium* parasites, relying solely on fluorescent labels associated with these markers is not sufficient to confirm recombination [61]. To confirm that recombination occurred between the two parental species, *C. parvum* and *C. tyzzeri*, I performed single-oocyst sequencing using Oxford Nanopore Technology (ONT). Genomic sequencing was essential because it provided base-pair level evidence that the progeny inherited genetic material from both parents. While bulk Illumina sequencing is widely used due to its high base accuracy, high

throughput, and relatively lower cost [2], it is not ideal for detecting recombination in individual parasites. Bulk sequencing captures an average signal [62] across many oocysts, which can mask the genetic contributions of each parent and obscure individual recombination events. This is evident in our study as changing the allele frequency threshold in IGV drastically changes the variant output, suggesting that the heterogenous nature of bulk sequencing affects the detection of recombination events. Additionally, Illumina short reads (typically 150 bp) require high depth of coverage for reliable assemblies, and even then, they often leave gaps or unresolved regions, especially in repetitive or complex parts of the genome. Even when using long reads in bulk sequencing, the mixture of genomes can still make it difficult to clearly identify recombination breakpoints. As observed in study, single-oocyst sequencing analyzes the genome of a single parasite (oocyst), which is the direct result of a single meiotic event. This approach allows for clear identification of recombinant haplotypes and precise mapping of crossover regions. ONT long reads, often spanning over 5 kb, help resolve the structure of recombined chromosomes more clearly. The use of high-accuracy R10 flow cells further enhances confidence in base calls, achieving Phred scores of 30 or higher. However, some major limitations of single-oocyst sequencing is the extremely small amount of DNA extracted from a single oocyst (~39.4×10^{-6} ng) and the high amount of DNA required for ONT sequencing (≥1 μg). Although whole-genome amplification is used to increase DNA yield, it introduces amplification variability, leading to some regions being heavily amplified than others. Also, any loss of DNA material during library preparation cannot be recovered. This can result in incomplete genome coverage and missing data in some regions. Despite these limitations, single-oocyst long-read sequencing remains a powerful method for detecting recombination in *Cryptosporidium* and providing the resolution necessary to distinguish parental contributions across the genome.

Our study provides compelling evidence that *C. parvum* and *C. tyzzeri*, two known distinct species, are capable of undergoing successful genetic crosses that result in viable hybrid progeny. Single-oocyst sequencing revealed that progeny inherited large chromosomal segments (> 20 kb) and mosaic structures from both parental species, indicative of recombination and meiotic crossover. These observations highlight that recombination in *Cryptosporidium* is not restricted to the engineered selectable markers but occurs broadly across the genome, underscoring the remarkable plasticity of its genetic makeup. Recombinant events were determined and validated with single oocyst sequencing; however, the exact breakpoint for crossing over to occur is not known. Although these crosses were performed under controlled laboratory conditions, the findings suggest that different *Cryptosporidium* species may naturally co-infect a single host and undergo genetic exchange. This is particularly relevant in the context of immunocompromised individuals, where opportunistic pathogens such as *Cryptosporidium* could thrive. The ability of *C. parvum* to infect immunocompromised mice and recombine with another species supports the idea that species boundaries in *Cryptosporidium* may be more permeable than previously assumed. This could explain sporadic reports of atypical species in unexpected hosts [63, 64] and raises important questions about host range expansion, zoonotic potential, and the role of recombination in driving phenotypic diversity. Altogether, these findings highlight the evolutionary flexibility of *Cryptosporidium* and position sexual recombination as a key mechanism underlying its rapid adaptation and diversification.

Conclusion

In summary, this chapter presents the first direct experimental evidence of interspecies recombination between *C. parvum* and *C. tyzzeri* through a controlled genetic cross using a novel dual drug-selectable marker system. By applying single-oocyst Oxford Nanopore long-read

sequencing, I confirmed that viable hybrid progeny inherited large chromosomal segments and mosaic regions from both parental species, which clearly indicates genome-wide recombination and meiotic crossover. While bulk Illumina sequencing offered a broad view of population-level variation, it lacked the resolution to resolve individual recombination events occasionally. In contrast, single-oocyst sequencing enabled precise detection of recombinant haplotypes and crossover points at the individual level even though genome coverage was low.

These findings not only validate the feasibility of genetic exchange between closely related *Cryptosporidium* species, but also highlight the fluidity of species boundaries and the parasite's capacity for adaptation through sexual recombination. The dual drug-selection system marks a significant advancement in *Cryptosporidium* genetics, enabling efficient and accurate identification of recombinants and supporting future functional studies. Collectively, this research provides a robust framework for exploring recombination-driven traits including virulence, host specificity, and immune evasion while also enabling advanced genetic manipulation and accelerating therapeutic discovery in this globally significant pathogen.

References

1. Kotloff, K.L., et al., *Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study.* Lancet, 2013. **382**(9888): p. 209-22.
2. Checkley, W., et al., *A review of the global burden, novel diagnostics, therapeutics, and vaccine targets for cryptosporidium.* Lancet Infect Dis, 2015. **15**(1): p. 85-94.
3. Blackburn, B.G., et al., *Cryptosporidiosis associated with ozonated apple cider.* Emerging Infectious Diseases, 2006. **12**(4): p. 684.
4. Deng, M.Q. and D.O. Cliver, *Comparative detection of Cryptosporidium parvum oocysts from apple juice.* Int J Food Microbiol, 2000. **54**(3): p. 155-62.
5. Slapeta, J., *Cryptosporidiosis and Cryptosporidium species in animals and humans: a thirty colour rainbow?* Int J Parasitol, 2013. **43**(12-13): p. 957-70.
6. Chalmers, R.M., et al., *Comparison of diagnostic sensitivity and specificity of seven Cryptosporidium assays used in the UK.* J Med Microbiol, 2011. **60**(Pt 11): p. 1598-1604.
7. Khan, A., J.S. Shaik, and M.E. Grigg, *Genomics and molecular epidemiology of Cryptosporidium species.* Acta Trop, 2018. **184**: p. 1-14.
8. Gilbert, I.H., et al., *Safe and effective treatments are needed for cryptosporidiosis, a truly neglected tropical disease*. 2023, BMJ Specialist Journals. p. e012540.
9. English, E.D., et al., *Live imaging of the Cryptosporidium parvum life cycle reveals direct development of male and female gametes from type I meronts.* PLoS Biol, 2022. **20**(4): p. e3001604.
10. Tandel, J., et al., *Life cycle progression and sexual development of the apicomplexan parasite Cryptosporidium parvum.* Nat Microbiol, 2019. **4**(12): p. 2226-2236.

11. Stapley, J., et al., *Recombination: the good, the bad and the variable*. 2017, The Royal Society. p. 20170279.

12. Yu, Y., et al., *A unique hexokinase in Cryptosporidium parvum, an apicomplexan pathogen lacking the Krebs cycle and oxidative phosphorylation.* Protist, 2014. **165**(5): p. 701-14.

13. Alcock, F., et al., *A small Tim homohexamer in the relict mitochondrion of Cryptosporidium.* Mol Biol Evol, 2012. **29**(1): p. 113-22.

14. Kissinger, J.C., *Evolution of Cryptosporidium.* Nat Microbiol, 2019. **4**(5): p. 730-731.

15. Nader, J.L., et al., *Evolutionary genomics of anthroponosis in Cryptosporidium.* Nat Microbiol, 2019. **4**(5): p. 826-836.

16. Huang, W., et al., *Multiple introductions and recombination events underlie the emergence of a hyper-transmissible Cryptosporidium hominis subtype in the USA.* Cell Host Microbe, 2023. **31**(1): p. 112-123 e4.

17. Guo, Y., et al., *Comparative genomic analysis reveals occurrence of genetic recombination in virulent Cryptosporidium hominis subtypes and telomeric gene duplications in Cryptosporidium parvum.* BMC genomics, 2015. **16**(1): p. 320.

18. Tichkule, S., et al., *Global Population Genomics of Two Subspecies of Cryptosporidium hominis during 500 Years of Evolution.* Mol Biol Evol, 2022. **39**(4).

19. Kato, S., et al., *Molecular identification of the Cryptosporidium deer genotype in the Hokkaido sika deer (Cervus nippon yesoensis) in Hokkaido, Japan.* Parasitol Res, 2016. **115**(4): p. 1463-71.

20. Vinayak, S., et al., *Bicyclic azetidines kill the diarrheal pathogen Cryptosporidium in mice by inhibiting parasite phenylalanyl-tRNA synthetase.* Sci Transl Med, 2020. **12**(563).

21. Shaw, S., et al., *Genetic crosses within and between species of Cryptosporidium.* Proc Natl Acad Sci U S A, 2024. **121**(1): p. e2313210120.

22. Troell, K., et al., *Cryptosporidium as a testbed for single cell genome characterization of unicellular eukaryotes.* BMC Genomics, 2016. **17**: p. 471.

23. Agyabeng-Dadzie, F., et al., *Evaluating the Benefits and Limits of Multiple Displacement Amplification With Whole-Genome Oxford Nanopore Sequencing.* Molecular Ecology Resources, 2025: p. e14094.

24. Guo, Y., et al., *Isolation and enrichment of Cryptosporidium DNA and verification of DNA purity for whole-genome sequencing.* J Clin Microbiol, 2015. **53**(2): p. 641-7.

25. De Coster, W. and R. Rademakers, *NanoPack2: population-scale evaluation of long-read sequencing data.* Bioinformatics, 2023. **39**(5): p. btad311.

26. Baptista, R.P., et al., *New T2T assembly of Cryptosporidium parvum IOWA annotated with reference genome gene identifiers.* bioRxiv, 2023.

27. Li, H., *New strategies to improve minimap2 alignment accuracy.* Bioinformatics, 2021. **37**(23): p. 4572-4574.

28. Shen, W., et al., *SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation.* PloS one, 2016. **11**(10): p. e0163962.

29. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows–Wheeler transform.* bioinformatics, 2009. **25**(14): p. 1754-1760.

30. Danecek, P., *Bon eld JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM et al: Twelve years of SAMtools and BCFtools.* Gigascience2021, 2021. **10**(2): p. 2.

31. Li, H., *Improving SNP discovery by base alignment quality.* Bioinformatics, 2011. **27**(8): p. 1157-1158.

32. Robinson, J.T., et al., *igv. js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV).* Bioinformatics, 2023. **39**(1): p. btac830.

33. Danecek, P., et al., *Twelve years of SAMtools and BCFtools.* Gigascience, 2021. **10**(2): p. giab008.

34. Kimball, A., et al., *Mendelian segregation and high recombination rates facilitate genetic analyses in Cryptosporidium parvum.* PLoS Genet, 2024. **20**(6): p. e1011162.

35. Gilchrist, C.A. *Cryptosporidium infection in Bangladesh children*. in *Eukaryome Impact on Human Intestine Homeostasis and Mucosal Immunology: Overview of the First Eukaryome Congress at Institut Pasteur. Paris, October 16–18, 2019.* 2020. Springer.

36. Baptista, R.P., et al., *Long-read assembly and comparative evidence-based reanalysis of Cryptosporidium genome sequences reveal expanded transporter repertoire and duplication of entire chromosome ends including subtelomeric regions.* Genome Res, 2022. **32**(1): p. 203-213.

37. Jensen, E.C., *Use of fluorescent probes: their effect on cell biology and limitations.* The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology, 2012. **295**(12): p. 2031-2036.

38. Hu, T., et al., *Next-generation sequencing technologies: An overview.* Human immunology, 2021. **82**(11): p. 801-811.

39. Kulkarni, P., et al., *Capture of water-borne colloids in granular beds using external electric fields: improving removal of Cryptosporidium parvum.* Water Res, 2005. **39**(6): p. 1047-60.

40. Ryan, U.M., et al., *Taxonomy and molecular epidemiology of Cryptosporidium and Giardia - a 50 year perspective (1971-2021).* Int J Parasitol, 2021. **51**(13-14): p. 1099-1119.

41. Ryan, U., et al., *An Update on Zoonotic Cryptosporidium Species and Genotypes in Humans.* Animals (Basel), 2021. **11**(11).

Tables

Table 1. Whole-genome amplification and T7 endonuclease reaction yield. A summary of WGA and T7 reaction results for putative single-oocyst recombinant progeny. The initial raw DNA in a single oocycst is approximately

	WGA Products		After T7 reaction	
Sample_Name	**Conc. (ng/uL)**	**Yield in 23uL (ng/uL)**	**Conc. (ng/uL)**	**Yield in 23uL (ng/uL)**
Progeny 1a	12.60	289.8	5.18	119.14
Progeny 2a	13.40	308.2	6.88	158.24
Progeny 3a	NA	NA	NA	NA
Progeny 4a	NA	NA	NA	NA
Progeny 5a	0.93	37	NA	NA
Progeny 6a	0.07	2.8	NA	NA
Progeny 7a	NA	NA	NA	NA
Progeny 1b	NA	NA	NA	NA
Progeny 2b	6.41	147.43	5.27	121.21
Progeny 3b	4.65	106.95	5.37	123.51
Progeny 4b	5.62	129.26	6.38	146.74
Progeny 5a	2.51	57.73	NA	NA
Progeny 6a	14.7	338.1	6.65	152.95
Progeny 7a	8.36	192.28	4.64	106.72
Progeny 8a	1.91	43.93	NA	NA

(a) Represent possible F1 progeny

(b) Represents possible F2 progeny

Only WGA yields above 4 ng/uL were subjected to T7 reaction

Table 2. A summary of ONT sequencing read metrics for putative recombinant single oocysts and Illumina sequencing metrics for parental reads.

Progeny_reads	Number of reads	Sequence length	Avg length	N50	Q15 %
Progeny 1a	52,586	179,134,692	3406.50	6,155	30.20
Progeny 2a	17,435	75,208,930	4313.70	7,754	30.00
Progeny 2b	4,020	17,599,041	4377.90	8,103	30.10
Progeny 3b	5,665	28,182,438	4974.80	10,506	40.40
Progeny 4b	5,328	22,825,298	4284.00	8,408	44.10
Progeny 6b	11,075	50,812,268	4588.00	8,919	40.90
Progeny 7b	35,677	147,577,426	4136.50	7,536	29.10

Parental_reads	Number of reads	Sequence length	Avg length	N50	Q30 %
CP_R1_Parent	2,604,907	379,220,248	145.60	149	89.36
CP_R2_Parent	2,604,907	379,364,500	145.60	149	85.20
CT_R1_Parent	13,080,195	1,867,231,427	145.60	149	93.82
CT_R2_Parent	13,080,195	1,867,562,463	145.60	149	90.80

Table 3. Genome coverage for ONT sequences data at 1× and 10× depth thresholds for recombinant single-oocyst samples, bulk-sequenced progeny, and parental strains.

File name	Avg depth	Genome coverage at 10x	Genome coverage at 1x
Bulk_CpCt_ Progeny*	113.51	100%	100%
Cp_Parent	77.76	99%	100%
Ct_Parent	393.80	98%	99%
Progeny 1a*	18.68	18%	46%
Progeny 2a*	7.89	18%	59%
Progeny 2b	1.78	5%	36%
Progeny 3b	2.97	4%	23%
Progeny 4b	2.39	5%	28%
Progeny 6b	5.35	8%	30%
Progeny 7b*	15.49	25%	68%

*Samples analyzed for evidence of recombination
Bulk_CpCt_progeny is the ONT sequencing of the bulk putative recombinant progenies

Table 4. Summary of variant types detected in recombinant single-oocyst genomes. Counts of single-nucleotide polymorphisms (SNPs), insertions, and deletions identified in each sample.

Progeny_reads	SNP count	Indel_count	Biallelic_SNPs
Bulk_CpCt_Progeny*	61,734	3,781	57,953
Cp_Parent	1,037	264	773
Ct_Parent	197,256	7,604	189652
Progeny 1a*	15,675	528	15,147
Progeny 2a*	15,741	454	15,287
Progeny 2b	4,835	167	4,668
Progeny 3b	333	20	313
Progeny 4b	3,260	65	3,195
Progeny 6b	9,348	192	9,156
Progeny 7b*	24,387	728	23,659

*Samples analyzed for evidence of recombination

Figures

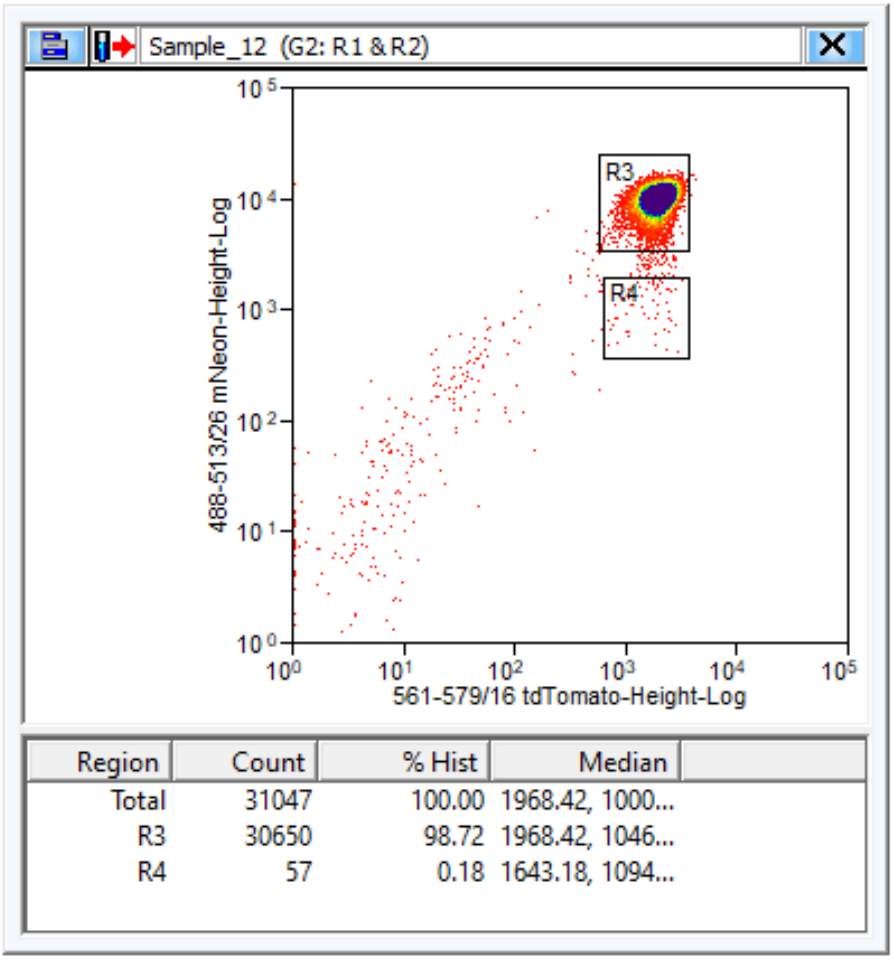


Figure 1. **Fluorescence-activated cell sorting (FACS) of putative recombinant single oocysts based on dual fluorescence**. *C. parvum* oocysts express mNeonGreen and *C. tyzzeri* express tdTomato; double-positive events represent potential recombinants sorted into 96-well plates.

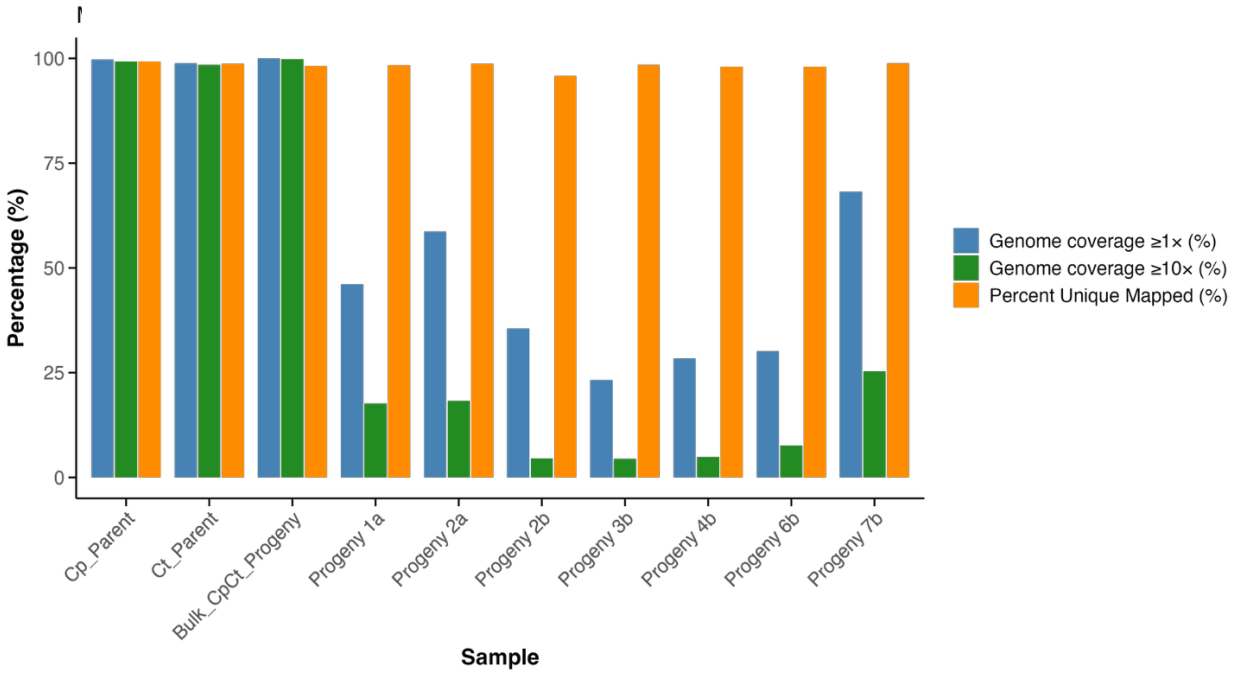
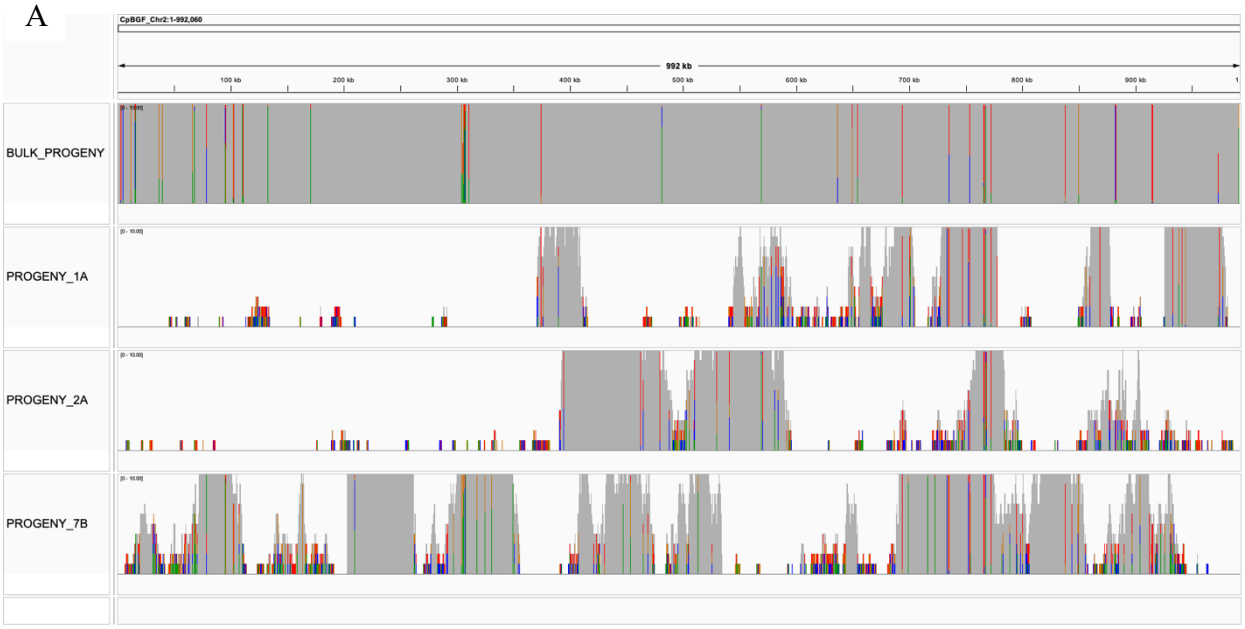


Figure 2. Mapping and genome coverage across recombinant single oocysts, bulk progeny, and parental samples. This bar plot summarizes three sequencing metrics for each sample: the percentage of reads successfully mapped to the reference genome, genome coverage at 1× depth, and genome coverage at 10× depth. Samples include single recombinant oocysts, the bulk-sequenced recombinant pool, and the two parental species (*C. parvum* and *C. tyzzeri*).

A


CpBGF_Chr2:1-992,060
992 kb
100 kb
200 kb
300 kb
400 kb
500 kb
600 kb
700 kb
800 kb
900 kb
BULK_PROGENY
PROGENY_1A
PROGENY_2A
PROGENY_7B

B


CpBGF_Chr5:264,472-917,869
653 kb
300 kb
400 kb
500 kb
600 kb
700 kb
800 kb
900 kb
BULK_PROGENY
PROGENY_1A
PROGENY_2A
PROGENY_7B

C

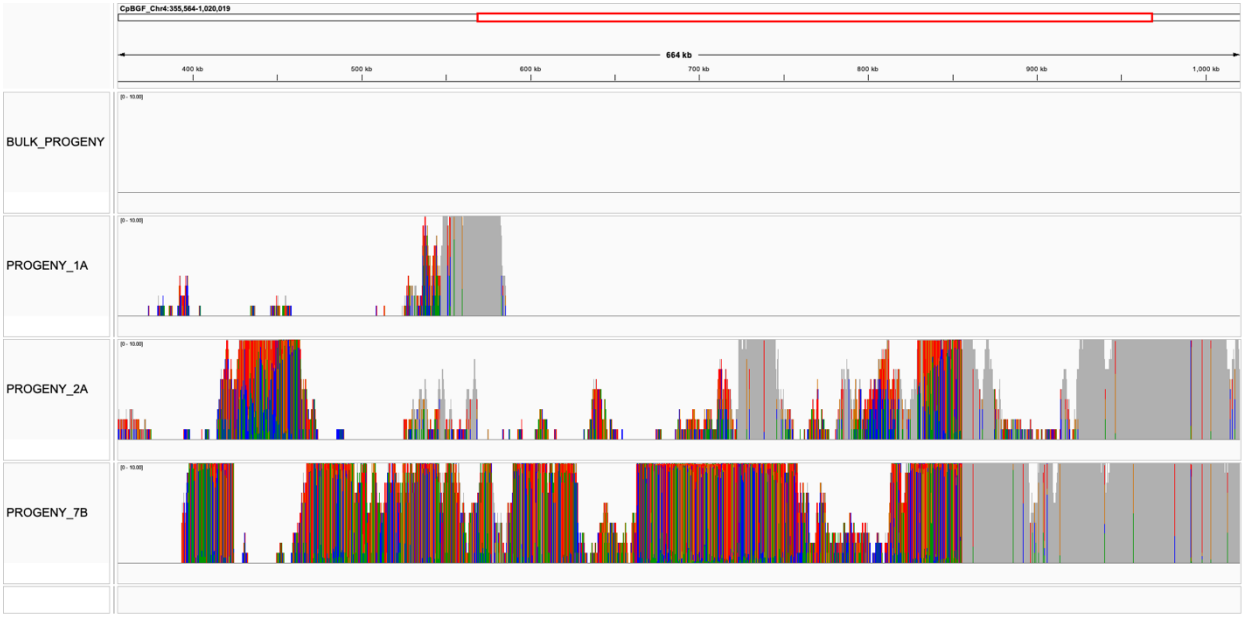


Figure 3. Chromosomal inheritance in recombinant progeny visualized using IGV. Coloured bars represent SNPs, with *C. parvum* BGFT2T as reference genome, therefore SNPs shown are likely derived from *C. tyzzeri*. (A) Chromosome 2 is predominantly inherited from *C. parvum*, while (B) chromosome 5 is primarily inherited from *C. tyzzeri*. (C) Chromosome 4 displays a mosaic pattern with alternating segments from both parental species, illustrating genome-wide recombination and crossover events.

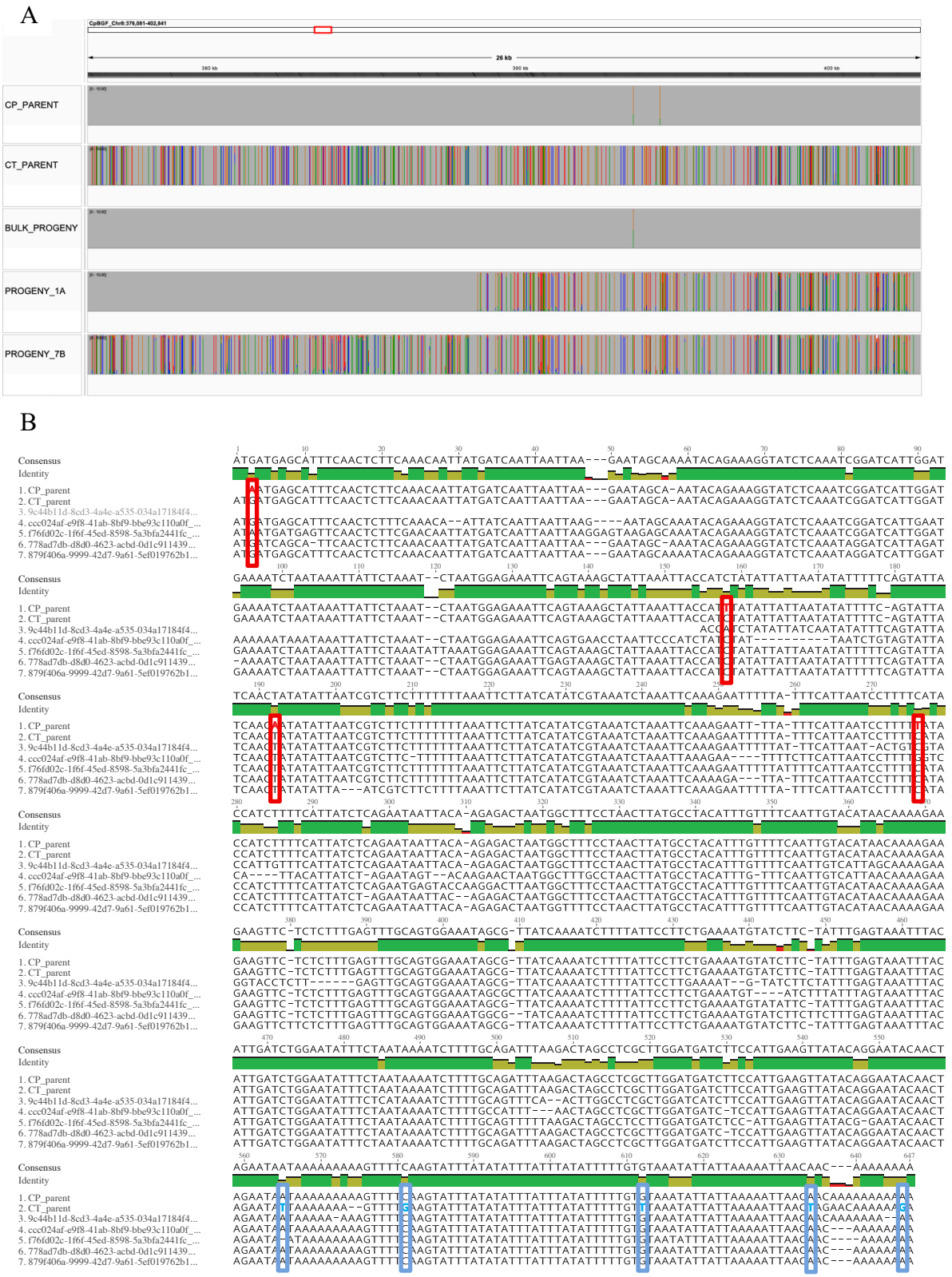


Figure 4. **A nucleotide-level evidence of a recombination event**. (A) IGV visualization of progeny 1a shows a transition on chromosome 8 from a region lacking SNPs (matching *C.*

parvum) to a region rich in SNPs (matching *C. tyzzeri*), suggesting a recombination crossover. (B) Multiple sequence alignment from the same region illustrates progeny 1a reads initially aligning with *C. tyzzeri* (red SNP) and then switching to *C. parvum* (blue SNP). Although the recombination is evident, the precise breakpoint remains unresolved.

CHAPTER 4

GLOBAL POPULATION STRUCTURE AND EVOLUTIONARY INSIGHTS INTO *CRYPTOSPORIDIUM PARVUM* REVEALED BY WHOLE-GENOME SNP ANALYSIS

Abstract

Cryptosporidium parvum is a globally distributed apicomplexan parasite and one of the leading causes of diarrheal disease in humans and animals. Although historically considered genetically homogeneous, this view has been increasingly challenged due to its broad host range and complex transmission dynamics. Using whole-genome SNP data from 408 high-quality *C. parvum* samples spanning four continents, I conducted a population genomic analysis to explore evolutionary divergence, population structure, and selection pressures.

Phylogenetic analysis identified two major global lineages: a European-USA clade and a China-Africa clade, consistent with previous studies. Despite low overall nucleotide diversity ($\pi < 0.002$), the African and European populations exhibited greater within-group variation, while the USA and Chinese populations showed signatures of recent clonal expansion. Fixation index (Fst) values exceeded 0.25 across most intercontinental comparisons, except between Europe and the USA, with Fst approximately 0.1, indicating strong global differentiation alongside evidence of recent gene flow within the Europe–USA clade.

Genes under positive selection ($pN/pS > 1.5$) were detected across all populations, including conserved loci implicated in host invasion and immune evasion. Additionally, each population displayed unique candidate genes under selection, pointing to region-specific adaptive pressures. Genotyping of the *gp60* locus revealed three dominant subtypes (IIa, IIc and IId) that broadly reflected the phylogenetic structure while also uncovering outliers and admixed lineages.

Together, these whole-genome analyses provide a high-resolution view of *C. parvum* population dynamics and evolution, underscoring the roles of clonal expansion, geographic

isolation, and host-driven selection. These insights have important implications for global surveillance and control efforts.

Introduction

Cryptosporidium is a globally significant apicomplexan parasite and the second leading cause of the diarrheal disease known as cryptosporidiosis [26] in infants under two years of age and immunocompromised individuals. Between 2011 and 2016, *Cryptosporidium* was responsible for 63% of protozoan waterborne outbreaks, primarily infecting immunocompromised individuals and children under the age of two in low- and middle-income countries [27]. These infections are associated with malnutrition, stunted growth, and premature death. The 1993 Milwaukee, WI, USA outbreak, which affected over 400,000 people and resulted in 69 deaths [65, 66], shows the parasite's devastating public health impact. The incidence and environmental persistence of *Cryptosporidium* oocysts, *i.e*., resistant to chlorination [67, 68], evasion of filtration [69], and commonly found on produce and dairy products [28, 70] has led the World Health Organization to rank *Cryptosporidium* fifth among the most significant foodborne parasites [71, 72].

Despite this global burden, no vaccine exists, and the only FDA-approved treatment, nitazoxanide, has limited efficacy in immunocompromised individuals and infants under two years of age [26, 27]. Several technical challenges have further hindered genomic epidemiology efforts in *Cryptosporidium*, including difficulties in obtaining enough clinical samples, isolating oocysts from fecal samples, inefficient long-term *in vitro* culturing systems, genome fragmentation from short-read sequencing, frequent mixed infections, and limited phylogenetic resolution from commonly used single markers, such as 18S rRNA and *gp60* [73, 74]. Although there are 74 C*ryptosporidium* genome assemblies in the NCBI (National Center for Biotechnology Information), approximately 55% of them are *C. parvum* and *C. hominis* and only 28% of the *Cryptosporidium* genomic assemblies have been annotated [75]. Additionally, there

are currently over 40 identified species and many other unclassified genotypes with varying host specificities [32, 74, 76]. There is evidence of some species having several hosts, emphasizing their anthroponotic or zoonotic nature, which makes understanding the mechanisms involved in transmission and host specificity more relevant [74, 77]. All these factors highlight a significant gap that remains in our knowledge regarding the population structure and evolution of *Cryptosporidium* species on a global scale.

To aid in filling the gap, I performed a whole-genome SNP analysis of *C. parvum* to identify subtle but relevant biological variations from 408 *C. parvum* samples found in NCBI's Sequence Read Archive (SRA), even though *C. parvum* is known to be genetically very similar [60] . SNPs from whole-genome sequencing (WGS) data provide high-resolution insights and comprehensive coverage of genomic diversity, including both coding and non-coding regions. This approach enables fine-scale differentiation of closely related isolates, thereby facilitating accurate resolution of population structure, quantifying genetic diversity, and detecting signals of recombination or selection across geographic populations [78-82]. Due to the financial and technical challenges involved in whole genome sequencing, I also assessed whether a subset of SNPs derived from conserved single-copy genes could faithfully replicate the population structure and evolutionary signals detected through genome-wide SNP analysis.

Methods

SRA Sample Collection and Quality Control

To obtain the most comprehensive dataset from across the globe, I downloaded paired-end *C. parvum* Illumina whole-genome sequencing reads (n = 540) and their associated metadata from the NCBI SRA. Trimmomatic v0.39 [83], was used to check for and remove all possible adapters as well as low-quality sequences with a sliding window of 5 bases, Phred score <20 and

a length of <50 bp. Read quality and sample composition were assessed using Seqkit v2.5.1 [52] and Kraken2 v2.1.3 - EupathDb [84], respectively. To verify taxonomic identity and quantify genome coverage, samples were mapped to the *C. parvum* BGFT2T reference genome [85] using BWA v0.7.17 [53, 54], followed by genome depth and coverage calculations with SAMtools v1.16.1 [57]. Samples not classified as *C. parvum,* exhibiting signs of contamination or mixed infection, or having < 90% genome coverage at 10× depth were excluded from subsequent analysis.

SNP Calling and Filtering

Population variation was examined through variant calling with BCFtools v1.10.2 [86] using the "-c and –ploidy 1" options on uniquely mapped reads. A limitation of calling variants with a ploidy of 1 is that heterogeneous sites (mixed samples) are missed, as variants are called based on only the major allele rather than accounting for within-sample diversity. Individual VCF files were merged into a single multi-sample VCF file. Only high-confidence, biallelic SNPs were retained with depth >30, mapping quality >30, fraction of MQ=0 <0.1, variant distance bias >0.2, and minor allele frequency >2.5%.

Population Structure Analysis

Genetic structure within and between populations was assessed by processing the filtered VCF file using PLINK v2 [87] and vcf2phylip.py [88], incorporating *C. hominis* as an outgroup for phylogenetic analysis. ADMIXTURE v1.3.0 [89] was used to infer population ancestry proportions, testing K values from 1 to 13. Phylogenetic analysis was performed using IQ-TREE v2 [90], employing the maximum likelihood method, 1000 bootstrap replicates, and ascertainment bias correction (ASC) optimized for SNP-only datasets. A genomic relationship

matrix (GRM) was computed using PLINK2 and a pairwise distance matrix was derived for principal component analysis (PCA) and hierarchical clustering.

Population Diversity Metrics

Evolutionary forces acting on populations were evaluated using VCFtools v0.1.16 [91]. Tajima's D [92, 93] was calculated with sliding windows of 10 kb with a 5 kb overlap. Also, genetic differentiation between and within populations was assessed using the fixation index (Fst) and nucleotide diversity (π) with a window of 10 kb, respectively. To examine recombination signatures, the linkage disequilibrium coefficient (r^2) within each population was estimated using PLINK v1.9, and r^2 was calculated between each SNP. SnpEff v5 [94] was used to annotate the SNPs and classify them as synonymous or nonsynonymous. Genes in the top 5 percent Fst regions were selected for further analysis. To assess selection pressure acting on the genes of interest, I estimated the ratio of non-synonymous to synonymous polymorphisms (pN/pS). Single-nucleotide polymorphisms (SNPs) were classified based on their predicted functional effects. Non-synonymous polymorphisms were defined as variants annotated with the following effects: missense mutations. Synonymous polymorphism included only synonymous variants. The pN/pS ratio was calculated as the proportion of non-synonymous polymorphism divided by the proportion of synonymous polymorphism, providing an estimate of the selective pressure acting on each gene. This estimate is an oversimplified version of pN/pS, as normalization is not done with the synonymous and nonsynonymous sites. All mutations that led to no change in CDS were regarded as synonymous mutations, while those that led to changes in the coding regions were referred to as nonsynonymous mutations. Genes with a dS = 0 were excluded from the analysis as it results in an infinite pN/pS ratio. Also, genes with a dN and a dS value of zero were excluded, as it implies that the SNP is outside the protein-coding region. In

order to identify genes with a potential high association with population dynamics, I conducted a comparative analysis across all populations, focusing on genes exhibiting high pN/pS ratios with genomic regions of high population differentiation.

SNPs from single-copy genes

Conserved single-copy genes were obtained from 29 publicly available *C. parvum* genome assemblies in NCBI. I evaluated the completeness of the assemblies using BUSCO v5.5.0 [95] and excluded assemblies with less than 90% completeness. Gene annotation for each genome assembly was performed using Liftoff v1.6.3, [96] with the *Cryptosporidium parvum* BGFT2T genome as a reference and its GFF. Coding and protein sequences were extracted using AGAT v1.1 [97, 98] for each file. Orthologous gene relationships were inferred with OrthoFinder v2.5.5 [99] using the acquired protein sequences. I selected single-copy genes present in all remaining genome assemblies. SNPs located within single-copy genes were extracted from the whole-genome SNP dataset, enabling a focused subset analysis within a genome-wide context. I also excluded single-copy genes with fewer than two SNPs, which comprised over 50% of the single-copy genes, from further analysis in order to focus on highly variable genes.

Results

Dataset Composition and Sample Filtering

I began with an initial dataset of 540 *C. parvum* genomic SRA read dataset samples downloaded from NCBI. Twenty-eight samples were identified as technical duplicates, and five samples containing only single-end reads were excluded based on the Seqkit v2.5.1 outputs (Appendix C: Supplementary Table 1). Upon analyzing the geographic distribution of the remaining 507 samples, based on the metadata, approximately 80% originated from Europe and

the United States, indicating a pronounced sampling bias toward Western populations (Figure 1a). A taxonomic verification using Kraken2 identified eight samples belonging to non-target organisms, namely *Acanthamoeba palestinensis*, *Candida albicans*, *Giardia intestinalis*, and *Hammondia hammondi (*Appendix C*:* Supplementary Table *2).* I also observed that some samples had a substantial proportion of the reads assigned to *C. hominis*, *C. baileyi*, and *C. ubiquitum*. To maintain the integrity of the species in question (*C. parvum*) and avoid confounding effects from inter-species variation, I excluded all non-*C. parvum* samples and samples with less than 90% genome coverage from my downstream analyses (Table 1). Finally, a total of 408 samples, including a *C. hominis* sample as an outgroup, remained for use in the phylogenomics and population diversity analysis (Fig 1bc).

Identification of Genetic Variants

The comprehensive variant analysis process identified 272,508 raw single-nucleotide polymorphisms (SNPs) across the retained sample set. Following stringent quality control measures that included filtering for only biallelic variants and applying established quality thresholds, the final dataset consisted of 9,701 high-quality SNPs with transition and transversion percentages of 71% and 25% respectively. Genomic distribution analysis revealed pronounced heterogeneity in SNP density across the *C. parvum* chromosomes. Notably, chromosomes 1, 4, 5, and 6 exhibited SNP hotspots with an SNP density above 10/kb at a few specific loci (Figure 2). The highest SNP density is at the end of chromosome 6. The majority of genes in these highly variable regions were uncharacterized proteins and signal peptides (Table 2).

Comparative Phylogenomics Across Global *C. parvum* Isolates

I conducted a maximum-likelihood phylogenetic analysis using SNP data from whole-genome sequencing (WGS) to investigate evolutionary relationships across samples and

populations. The phylogenetic tree constructed from genome-wide SNPs revealed strong geographic clustering of samples (Figure 3a). The geographical location assigned to each sample is based on the metadata information from NCBI, which indicates the area where the samples were collected. Two major evolutionary clades emerged: a China-Africa clade, where these populations share a more recent common ancestor, and a European-USA clade, where samples cluster together with USA samples consistently descending from European lineages. Glycoprotein subunit, *gp60*, is an established genetic marker used to subtype *Cryptosporidium* and is also known to play a pivotal role in transmission [100-103]). Three *gp60* subtypes were identified, ie, IIa, IIc and IId. Though not all the samples had the *gp60* gene identified, the population structure was independently validated by *gp60* genotypes from metadata, which clustered samples based on their geographic origins, confirming the phylogenetic assignments. Notably, USA samples exhibited extremely short branch lengths, indicating minimal genetic divergence among them.

Population Structure Inferred from Principal Component and Admixture Analyses

I performed ADMIXTURE analysis using genome-wide SNP data to explore the underlying population structure. Cross-validation analysis across 13 replicates identified the optimal number of ancestral populations at K = 5. All five populations showed some form of admixture. Four of the ancestral populations had a clear geographical identity but the remaining one consisted of samples from several geographical locations (Figure 3b). Population 5 was predominantly composed of individuals from the USA, while Population 2 represented the majority of samples from China. Populations 3 and 5 were primarily associated with European samples, suggesting the possibility of having two diverging lineages. Population 1 displayed a distribution pattern that appeared to consist of all individuals from all other regions, suggesting it

may represent a more ancient or shared ancestry component, including Africa (Figure 3b). To validate the population structure, I conducted Principal Component Analysis (PCA) using a genomic relationship matrix (GRM) derived from whole-genome sequencing (WGS) SNPs. Most samples from Europe (blue) and USA (red) populations clustered closely together along the positive axis of PC1. Conversely, samples from China (purple), Africa (green), and the remaining Europe occupied the negative PC1 space, forming a clearly separated cluster. This highlights substantial genetic divergence from the European-USA lineage as seen in the phylogenetic tree (Figure 3c).

Comparative Analysis of Genomic Variation Among Samples

The hierarchical clustering based on the genomic distance matrix demonstrated strong geographical grouping (Figure 4a). Two main clades are observed, *i.e*., Europe and USA, as well as the Africa and China clades. This is consistent with the results of the phylogenetic and ADMIXTURE analyses, thereby validating these results. The dendrogram revealed distinct population-specific clusters with USA samples (red) and European samples (blue) forming tightly consolidated groups, while Chinese samples (purple) and African samples (green) clustered separately, showing fewer diverging branches. This shows that USA and European populations exhibit greater within-population diversity compared to the China samples, which formed more compact clusters. African samples appeared at distinct branches, highlighting their unique genetic profiles and deepest divergence from other populations.

The heatmap of the genomic distance matrix reveals distinct patterns of genetic similarity and clustering among global samples (Figure 4b). Generally, samples from China, Africa, and some isolated European populations are genetically identical within their groups, whereas samples from the USA–Europe clade show more genetic differences. Interestingly, the analysis

also revealed some unexpected genetic relationships among samples. Some samples from China and Europe show high similarity to samples from Africa (Figure 4 b). This cross-continental similarity suggests either shared ancestral lineages or more recent genetic exchange between these geographically distant populations.

Comparative Genomic Diversity Among Populations

To assess within-population genetic variation, I calculated nucleotide diversity (π) for each sample, grouping them by geographical location. Although overall π values were low across all populations, median < 0.0005, significant differences were detected when comparing nucleotide diversity between populations (Figure 5a). A Kruskal-Wallis test confirmed that nucleotide diversity varies significantly across the geographical populations ($p < 2.2e\text{-}16$), with pairwise comparisons revealing strong separation in most cases.

While nucleotide diversity measures variability within a population, I also evaluated between-population divergence by calculating the fixation index (Fst). Pairwise comparisons revealed substantial genetic differentiation, with Fst values exceeding 0.4 in most cases (Figure 5b). The only exception was observed between the "European-USA" populations, which had an Fst value slightly less than 0.2. "Africa-China" comparisons exhibited two prominent Fst peaks, 0.25 and 0.085. Additionally, I observed that any population compared with China had an Fst greater than 0.7.

I performed a Tajima's D analysis to investigate the types of selection and population events influencing population structure (Figure 6a). Most populations showed positive Tajima's D scores across the genome, indicating balancing selection or population contraction, except for the USA population (Red), which was largely skewed toward negative values, suggesting purifying selection or more likely population expansion (Figure 6b). The African population

(green) showed a distribution with a peak just below zero, with most of the genome displaying a positive Tajima's D score. The Chinese population (purple) had three peaks, one below zero and the others on the positive side, while the European population (blue) showed broad, right-skewed distributions.

To determine the exchange rate of genetic material between populations, I assessed the distribution of linkage disequilibrium (LD), measured as r^2 over a window size of 10kb, across four populations based on the whole-genome sequencing SNP analysis (Figure 7a and b). While chapter 3 demonstrates that recombination between species is possible under laboratory conditions, these LD patterns can reveal whether genetic exchange occurs in natural populations. Africa exhibited moderate to high LD, with a median r^2 of ~0.5-0.55, with the majority of SNPs in the genome falling within that range. The USA population had a similar pattern as the Africa population, but with a lower r^2 median of 0.4-0.5. Although China had a median r^2 similar to that of Africa, its r^2 values varied widely across the entire genome. Europe populations had the lowest median r^2 <0.2, with most of the values concentrated in that range, showing a consistently narrow distribution across the genome. These observations collectively highlight distinct LD patterns among the populations, with Africa and USA showing elevated LD, China displaying high variability and Europe exhibing consistently low LD across most genomic windows reflecting their unique demographic histories and evolutionary forces.

To determine the potential type of selection acting on genes within a population, I calculated the ratio of nonsynonymous to synonymous mutations (pN/pS) per gene across all populations. The majority of the genes, approximately 700 globally, had a pN/pS ratio of zero, indicating purifying selection. The highest pN/pS ratio I observed was 11, corresponding to gene *cpbgf_6003050*, an uncharacterized protein detected in the African and Chinese populations

(Figure 8a and b). Also, I observed that some genes were present in some populations but absent in others at specific pN/pS thresholds (Appendix C: Supplementary Figure 1). By focusing on the regions of elevated population differentiation (top 5% of Fst) and pN/pS ratios and selection, I identified an average of 96.5 genes per comparison (Figure 8c). The genome-wide survey of Fst revealed an interesting pattern of hotspots for population divergence across the genome. Hence, focusing on chromosome 6 of the Africa–USA comparison resulted in 47 genes, with 6 of them having pN/pS ratios >1 (Figure 8d and e)

Reference-Based Single-Copy Gene Identification

To support comparative genomic analysis, I curated a high-confidence set of single-copy genes (SCGs) from available *C. parvum* assemblies. One genome assembly (GCA_000209695.1_ASM20969v1) was excluded due to a BUSCO score of less than 15%. The remaining assemblies had BUSCO completeness scores above 95% and reference genome coverage greater than 75%. Gene content analysis across these assemblies identified 4,226 genes, of which approximately 60% were present in all of the assemblies. I used this threshold to define a set of 2,511 high-confidence single-copy genes. The population structure and phylogenetic analyses yielded results similar to those of the WGS SNP analysis.

Discussion

In this study, I analyzed the population structure and genetic diversity of 407 *C. parvum* SRA read datasets from 4 continents using whole-genome SNP (WGS SNP) data. There has been extensive research into the population dynamics of outbreaks of *C. parvum* and *C. hominis* [43, 74, 104] ; however, some of the studies are usually localized to a specific region or an outbreak and may lack the diversity required to make substantial predictions and observations. My goal was to perform the most extensive comparative genomic analyses so far, to investigate

the population structure and phylogenetic relationships among various *C. parvum* populations using genome-wide variant data from all submitted samples in the NCBI across the globe.

An analysis of 407 high-quality *C. parvum* genomic read data sets from multiple countries on 4 continents revealed a strong geographic sampling bias toward Europe and the United States of America, which underscores the need for more balanced global sampling to fully characterize *C. parvum* diversity. The amount of information available in the metadata of these samples limits research and impacts our ability to track the transmission or origin of the parasite. This means that even though the metadata may include the geographical location (origin) of the sample, it does not indicate whether the host acquired the parasite from that location or contracted the infection before arriving there.

From the initial 272,508 raw SNPs, a stringent filtering pipeline reduced this to 9,701 high-quality biallelic SNPs, resulting in an average of 1.05 SNPs per kb. This dramatic reduction highlights the importance of rigorous quality control in distinguishing true biological variation from sequencing noise or low-quality variant calls. Additionally, the decline reflects the impact of limited sample sizes, which can lead to insufficient allele representation and reduce the statistical power to confidently call SNPs across all populations. Smaller sample sizes are usually prone to high levels of missing data and reduced minor allele frequency support. Also, some rare alleles can also be lost due to this stringency, hence the emphasis on the need for adequate sampling. Genomic coverage thresholds (≥90% at 10×) ensured that only samples suitable for confident variant calling were included in downstream analyses. Interestingly, SNP density varied across the *C. parvum* genome, with chromosomes 1, 4, 5, 6, 7, and 8 having regions of increased SNP density, >5 SNPs per kb, usually near the ends of the chromosomes. These patterns suggest that the subtelomeric regions may serve as evolutionary hotspots with genes

such as the *SKR1* gene family, *FLGN* gene families among others influencing virulence and host specificity [105].

Although historically considered genetically homogeneous by some [60], the current population structure tells a more complex story. Whole-genome SNP phylogenetic analysis reveals distinct population structure shaped by geographical location, as two major clades are identified for *C. hominis* and *C. parvum* [43] [104]. This was also evident in my analysis, which identified two major clades, *i.e*., one composed primarily of European and U.S. isolates, and another containing samples from China and Africa. This broad division reflects both historical spread and evolutionary divergence of subtype groups. The *gp60* gene (encoding the 60-kDa glycoprotein) is the most widely used single-locus marker for *Cryptosporidium* subtyping [101]. Beyond its role in molecular typing, recent studies have demonstrated its involvement in the parasite's infectivity [106]. In the dataset, three major subtype families were identified: IIa, IIc, and IId. These subtypes generally clustered together, though some outliers were observed. Further research is needed to clarify how subtype diversity, particularly within the gene's repetitive region, relates to host specificity and transmission dynamics. Admixture analysis (K=5) complements these findings, revealing overlapping genetic components between the USA and Europe, while African and Chinese populations remain more distinct. Unexpected outlier lineages, mainly composed of subtypes IIc and IId, do not conform to continental boundaries and likely reflect ancient recombination, retention of ancestral polymorphisms, or independent introductions [39-41]. These admixed lineages demonstrate that while the overall population structure is shaped by geography, gene flow and historical recombination events have blurred the lines in some regions.

The Chinese population stands out as the most genetically uniform group, with an extremely low nucleotide diversity (π). These samples all belonged to subtype IId and formed a distinct, compact clade, consistent with a recent clonal expansion after a founder's effect. Hierarchical clustering and genomic distance heatmaps further confirm this homogeneity, suggesting that the current Chinese population likely originated from a small founder event or has undergone limited recombination since its emergence. In the African population, two main clades have been identified with different *gp60* subtypes, namely IId and IIa. The IId subtypes (SRR10363596, SRR15694514, and SRR15694514) are genetically identical to the China clonal population, so I propose two hypotheses. First, the China population may have emerged due to a founder's effect from the IId African sample. The second hypothesis is that the IId subtype is actually a China population that might have been introduced into Africa through migration (gene flow), which is why it is also called an African sample. Both hypotheses are supported by the relatively high nucleotide diversity (π) of the designated African sample, despite its clonal nature, as the population comprises two diverging subpopulations.

The U.S. and European populations show different patterns compared to the more structured and clonal populations in Africa and China. The U.S. population likely originated from Europe, as supported by the shared IIa subtype and the short branch length from the common ancestor [104]. However, while both regions carry this subtype, their population structures differ. In the U.S., the population is not clonal, but overall diversity is low. The phylogenetic tree shows short branches without strong separation between lineages, suggesting a recent introduction followed by early stages of diversification. This pattern points to a founder event or bottleneck, where a small number of genotypes were introduced from Europe and began to spread. Since then, the U.S. population has expanded across different hosts or environments,

but there hasn't been enough time or pressure for clear subgroups to form. Instead, multiple closely related genotypes appear to be circulating together. This highlights how founder events and recent expansions can lead to shallow genetic structure while still allowing the population to adapt to new settings [107].

The European population is genetically diverse and not clonal, exhibiting a level of nucleotide diversity similar to that of Africa. Its phylogenetic structure shows two distinct subgroups: one made up entirely of European samples, and another that includes U.S. samples branching from it. This supports the idea that the U.S. population originated from a subset of the European population, rather than a recent common ancestor. The presence of both a deeply rooted European clade and a separate lineage leading to U.S. samples indicates that Europe has been a long-standing reservoir of genetic diversity. Unlike populations shaped by recent clonal expansion, the structure in Europe reflects a more stable and historically diverse transmission pattern, likely maintained through repeated introductions and circulation across hosts and regions.

Understanding the selection pressures acting on populations and genes is very important, as it influences the organism's adaptation to the environment and response to host immunity. In this study, the isolates were derived from two main hosts, *i.e*., human and cattle. However, *C. parvum* has been documented in other non-human primates and livestock. The broad host range exposes the parasite to a varied and overlapping selection pressures ranging from host-specific immune responses to varying intestinal and geographical environments. Combining these factors with the differences in water infrastructure and sanitation contributes significantly to the genetic diversity and population structure of the parasite, Population genomic metrics including Tajima's D, linkage disequilibrium (LD), rate of nonsynonymous to synonymous ratios (pN/pS) and

Fixation index (Fst) shed light on the selective pressures acting on these populations. The clonal nature of the China population accompanied with its highly positive Tajima's D values across many genomic regions, suggests mostly neutral and balancing selection. This type of selection helps maintain intermediate alleles, thereby broadening the host range of the pathogen, *i.e.*, cattle, humans, sheep and goats. Also, although China samples show a varying r^2 across the genome, with areas of low and high LD, limited recombination is assumed due to the presence of large, intact haplotype blocks, as genetic material exchange occurs between identical isolates. These combined signals point to clonal expansion with recombination occurring between identical isolates, most likely from an outbreak.

Though the African population is clonal in structure, like the China population, it is made up of two main clades, which didn't arise from a single outbreak. The split reflects differences in host, specifically between *Bos taurus* (cattle) and humans, as each clade adapts to a distinct host environment. This kind of structure suggests that while the population shares a common origin, it has since diversified in a meaningful way. Even within a clonal background, the genome appears to be holding onto variation in regions likely tied to host interaction or immune pressure [42]. Balancing selection may be acting to maintain that variation, especially in genes that matter for survival in different host contexts. Altogether, the Africa population shows how clonal populations can still evolve, conserving most of the genome while allowing key regions to remain variable.

The USA population shows a different evolutionary pattern. While its overall nucleotide diversity is low, the phylogenetic structure is highly branched, suggesting the presence of multiple lineages that likely originated from Europe. The combination of strong population structure and low within-lineage variation implies the action of purifying or directional selection

possibly as the population adapts to local environmental pressures or host immune responses. LD patterns suggest limited recombination, allowing distinct lineages to persist over time. The USA population, therefore, reflects a scenario where multiple introductions have occurred, but recombination remains restricted, and selection acts to maintain locally advantageous variants. This configuration is epidemiologically important, as it implies that even genetically restricted populations can harbor considerable structural diversity, potentially giving rise to novel strains.

Samples within the Europe population are highly diverse and recombining. This is supported by its relatively high π values, a widespread distribution of Tajima's D scores, and a low median LD, all indicative of frequent recombination and heterogeneous selection pressures. Some genomic regions in Europe show signatures of purifying selection (negative Tajima's D), while others suggest balancing selection (positive D), likely reflecting adaptation to diverse hosts or environments. This is also significant as the European population has a subpopulation divergent from the Europe–USA clade. The European population thus appears dynamic and fluid, with gene flow and recombination continually reshaping its structure. This genetic fluidity increases the risk of new strain emergence and complicates efforts to track and control transmission, underscoring the importance of targeted molecular surveillance.

Pairwise FST analysis further confirms the strong genetic structuring among populations. Africa, China, and the USA samples are highly divergent from one another, with Fst values exceeding 0.8, consistent with long-term isolation or independent clonal expansions. This implies that about 80% of the genetic variation between these samples is due to differences between the populations. Europe shows intermediate Fst values, suggesting that it may serve as a hub of admixture or gene flow between populations. The lowest mean Fst is observed between

the USA and Europe (Fst = 0.1), indicating recent divergence or ongoing gene exchange, which aligns with their shared clustering patterns.

To understand how *Cryptosporidium parvum* adapts to different hosts and environments, I analyzed genes with pN/pS ratios greater than 1.5, which is a signal of positive selection. Several genes, including *cgd1_140* (CPLSP gene family) [108], *cgd1_150* (SKSR gene family) [105], and other secretory proteins, were under positive selection in all four populations: Africa, Europe, USA, and China (Table 3, Appendix C: Supplementary Figure 1). These shared genes likely play essential roles in parasite survival, helping it evade host immunity, invade cells, and persist in different environments. In addition to these common genes, each population also showed unique sets of positively selected genes. For example, *cgd2_1240* (EH domain-containing protein), *cgd4_2110* (papain family cysteine protease) [109], and *cgd6_1140* (AP2/ERF domain-containing protein) [110] were only under high selection in Africa populations. These differences likely reflect local adaptations to specific host immune systems, ecological conditions, or transmission settings. Note that the absence of a gene in one population doesn't necessarily mean it's not present, it may simply not have passed SNP filtering thresholds. To explore diversification and selection more closely, I focused on chromosome 6, specifically in the context of Africa–USA populations. This region has previously been linked to selective sweeps in *C. hominis* from Africa and China [43]. My analysis found both characterized and uncharacterized genes in this region under selection (Table 4 and Appendix C: Supplemental table 3). Among these, *cgd6_1570* (Apyrase)[111], *cgd6_2280* (Glycosylphosphatidylinositol-mannosyltransferase I) [112], and *cgd6_1300* (a TSP1-domain-containing uncharacterized protein) [113] are examples of genes found in regions of high Fst. These genes have orthologs in other parasites that are known to play roles in host cell attachment and invasion**,** suggesting their

involvement in host-pathogen interactions. Their presence in both high Fst and high pN/pS regions suggests they are evolving under host-driven selection pressures. These genes can be considered strong candidates for functional studies and may serve as future therapeutic targets**.**

Although the well-known *gp60* gene (pN/pS = 2) was not among the top-selected genes in this analysis, its importance in virulence and subtyping is well established [100, 101]. To further explore its role, Dr. Rodrigo de Paula Baptista, a research scientist from Houston Methodist Research Institute, and I performed *gp60* genotyping using the genome assemblies of all samples, which helped resolve previously missing metadata and confirmed the geographic clustering of isolates. The availability of *gp60* subtype data enhances the potential for future research to explore associations between *gp60* subtypes, geographic distribution, transmission routes, and host infectivity [114]. Interestingly, a recombinant *C. hominis* subtype identified in the USA, originating from African and European lineages, was found to be hyper-transmissible [41]. This raises the possibility that U.S. *C. parvum* strains may also exhibit increased transmissibility, potentially driven by rapid evolution linked to variation within the *gp60* gene.

The density subset analysis of SNPs produced promising results, as all three SNP density categories, high**,** moderate, and low, yielded consistent phylogenetic trees and population structure patterns. This consistency suggests that the loci within these subsets capture meaningful genetic signals and are likely to reflect true evolutionary relationships. As such, SNPs in these regions represent strong candidates for in-depth population analyses and may also serve as effective targets for hybridization-based sequencing approaches**.**

Conclusions

In this study, I performed the most comprehensive population genomic analysis of *Cryptosporidium parvum* to date, using whole-genome SNP data from over 400 globally

distributed samples. My goal was to gain a deeper understanding of the population structure and evolutionary mechanisms of *Cryptosporidium parvum* on a global scale.

The results show that the current population structure of *C. parvum* is shaped by a mix of geographic separation, clonal expansion, founder events, and gene flow. I show that *C. parvum* evolves through a combination of mutation, selection and adaptation to its environment and host. These forces act differently across regions and populations, leading to both genetically diverse and highly clonal lineages. Understanding this variation is essential for tracking transmission, detecting emerging variants, and designing control strategies. Importantly, this study lays the groundwork for a reproducible workflow to study the global population diversity and structure of *Cryptosporidium*. However, several key questions remain. What are the best tools and metrics to use when comparing population structure across such a complex parasite? Should each sample be treated as an individual or as a representative of a population? And perhaps most fundamentally, what defines a true species in *Cryptosporidium*, especially given its high rate of recombination, gene flow, and host switching? Addressing these questions will require continued global surveillance, expanded sampling, and thoughtful interpretation of genomic data. However, this study represents a significant step forward in understanding the evolution and spread of this important parasite.

Acknowledgements

We acknowledge the guidance and contribution of Rodrigo Baptista, Aziz Khan and Natalia Bayona.

References

1. Slatko, B.E., A.F. Gardner, and F.M. Ausubel, *Overview of next-generation sequencing technologies.* Current protocols in molecular biology, 2018. **122**(1): p. e59.
2. Hu, T., et al., *Next-generation sequencing technologies: An overview.* Human immunology, 2021. **82**(11): p. 801-811.
3. Pacbio, <*Technical-Note-Preparing-DNA-for-PacBio-HiFi-Sequencing-Extraction-and-Quality-Control.pdf*>. 2022.
4. Delahaye, C. and J. Nicolas, *Sequencing DNA with nanopores: Troubles and biases.* PloS one, 2021. **16**(10): p. e0257521.
5. Hou, Y., et al., *Comparison of variations detection between whole-genome amplification methods used in single-cell resequencing.* Gigascience, 2015. **4**(1): p. s13742-015-0068-3.
6. Dean, F.B., et al., *Comprehensive human genome amplification using multiple displacement amplification.* Proceedings of the National Academy of Sciences, 2002. **99**(8): p. 5261-5266.
7. Fullwood, M.J., et al., *The use of multiple displacement amplification to amplify complex DNA libraries.* Nucleic acids research, 2008. **36**(5): p. e32.
8. Troell, K., et al., *Cryptosporidium as a testbed for single cell genome characterization of unicellular eukaryotes.* BMC Genomics, 2016. **17**: p. 471.
9. Binga, E.K., R.S. Lasken, and J.D. Neufeld, *Something from (almost) nothing: the impact of multiple displacement amplification on microbial ecology.* The ISME journal, 2008. **2**(3): p. 233-241.

10. Lasken, R.S., *Genomic DNA amplification by the multiple displacement amplification (MDA) method.* Biochemical Society Transactions, 2009. **37**(2): p. 450-453.

11. Ceccherini, M., et al., *Degradation and transformability of DNA from transgenic leaves.* Applied and environmental microbiology, 2003. **69**(1): p. 673-678.

12. Dabney, J. and M. Meyer, *Extraction of highly degraded DNA from ancient bones and teeth*, in *Ancient DNA: methods and protocols*. 2019, Springer. p. 25-29.

13. Wang, G., et al., *DNA amplification method tolerant to sample degradation.* Genome research, 2004. **14**(11): p. 2357-2366.

14. Lage, J.M., et al., *Whole genome analysis of genetic alterations in small DNA samples using hyperbranched strand displacement amplification and array–CGH.* Genome research, 2003. **13**(2): p. 294-307.

15. Lasken, R.S. and T.B. Stockwell, *Mechanism of chimera formation during the Multiple Displacement Amplification reaction.* BMC biotechnology, 2007. **7**(1): p. 19.

16. Lu, N., et al., *Chimera: The spoiler in multiple displacement amplification.* Computational and Structural Biotechnology Journal, 2023. **21**: p. 1688-1696.

17. Hård, J., et al., *Long-read whole-genome analysis of human single cells.* Nature Communications, 2023. **14**(1): p. 5164.

18. Kolmogorov, M., et al., *Assembly of long, error-prone reads using repeat graphs.* Nature biotechnology, 2019. **37**(5): p. 540-546.

19. Hu, J., et al., *NextPolish: a fast and efficient genome polishing tool for long-read assembly.* Bioinformatics, 2020. **36**(7): p. 2253-2255.

20. Marçais, G., et al., *MUMmer4: A fast and versatile genome alignment system.* PLoS computational biology, 2018. **14**(1): p. e1005944.

21. Chen, Y., et al., *Chromosome-level genome assembly of Cryptosporidium parvum by long-read sequencing of ten oocysts.* Sci Data, 2024. **11**(1): p. 1287.

22. Penumarthi, L.R., et al., *A new chromosome-level genome assembly and annotation of Cryptosporidium meleagridis.* Scientific Data, 2024. **11**(1): p. 1388.

23. Paul, P. and J. Apgar, *Single-molecule dilution and multiple displacement amplification for molecular haplotyping.* Biotechniques, 2005. **38**(4): p. 553-559.

24. El-Sayed, N.M., et al., *The genome sequence of Trypanosoma cruzi, etiologic agent of Chagas disease.* Science, 2005. **309**(5733): p. 409-415.

25. Flynn, J.M., et al., *RepeatModeler2 for automated genomic discovery of transposable element families.* Proceedings of the National Academy of Sciences, 2020. **117**(17): p. 9451-9457.

26. Kotloff, K.L., et al., *Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study.* Lancet, 2013. **382**(9888): p. 209-22.

27. Checkley, W., et al., *A review of the global burden, novel diagnostics, therapeutics, and vaccine targets for cryptosporidium.* Lancet Infect Dis, 2015. **15**(1): p. 85-94.

28. Blackburn, B.G., et al., *Cryptosporidiosis associated with ozonated apple cider.* Emerging Infectious Diseases, 2006. **12**(4): p. 684.

29. Deng, M.Q. and D.O. Cliver, *Comparative detection of Cryptosporidium parvum oocysts from apple juice.* Int J Food Microbiol, 2000. **54**(3): p. 155-62.

30. Slapeta, J., *Cryptosporidiosis and Cryptosporidium species in animals and humans: a thirty colour rainbow?* Int J Parasitol, 2013. **43**(12-13): p. 957-70.

31. Chalmers, R.M., et al., *Comparison of diagnostic sensitivity and specificity of seven Cryptosporidium assays used in the UK.* J Med Microbiol, 2011. **60**(Pt 11): p. 1598-1604.

32. Khan, A., J.S. Shaik, and M.E. Grigg, *Genomics and molecular epidemiology of Cryptosporidium species.* Acta Trop, 2018. **184**: p. 1-14.

33. Gilbert, I.H., et al., *Safe and effective treatments are needed for cryptosporidiosis, a truly neglected tropical disease*. 2023, BMJ Specialist Journals. p. e012540.

34. English, E.D., et al., *Live imaging of the Cryptosporidium parvum life cycle reveals direct development of male and female gametes from type I meronts.* PLoS Biol, 2022. **20**(4): p. e3001604.

35. Tandel, J., et al., *Life cycle progression and sexual development of the apicomplexan parasite Cryptosporidium parvum.* Nat Microbiol, 2019. **4**(12): p. 2226-2236.

36. Stapley, J., et al., *Recombination: the good, the bad and the variable*. 2017, The Royal Society. p. 20170279.

37. Yu, Y., et al., *A unique hexokinase in Cryptosporidium parvum, an apicomplexan pathogen lacking the Krebs cycle and oxidative phosphorylation.* Protist, 2014. **165**(5): p. 701-14.

38. Alcock, F., et al., *A small Tim homohexamer in the relict mitochondrion of Cryptosporidium.* Mol Biol Evol, 2012. **29**(1): p. 113-22.

39. Kissinger, J.C., *Evolution of Cryptosporidium.* Nat Microbiol, 2019. **4**(5): p. 730-731.

40. Nader, J.L., et al., *Evolutionary genomics of anthroponosis in Cryptosporidium.* Nat Microbiol, 2019. **4**(5): p. 826-836.

41. Huang, W., et al., *Multiple introductions and recombination events underlie the emergence of a hyper-transmissible Cryptosporidium hominis subtype in the USA.* Cell Host Microbe, 2023. **31**(1): p. 112-123 e4.

42. Guo, Y., et al., *Comparative genomic analysis reveals occurrence of genetic recombination in virulent Cryptosporidium hominis subtypes and telomeric gene duplications in Cryptosporidium parvum.* BMC genomics, 2015. **16**(1): p. 320.

43. Tichkule, S., et al., *Global Population Genomics of Two Subspecies of Cryptosporidium hominis during 500 Years of Evolution.* Mol Biol Evol, 2022. **39**(4).

44. Kato, S., et al., *Molecular identification of the Cryptosporidium deer genotype in the Hokkaido sika deer (Cervus nippon yesoensis) in Hokkaido, Japan.* Parasitol Res, 2016. **115**(4): p. 1463-71.

45. Vinayak, S., et al., *Bicyclic azetidines kill the diarrheal pathogen Cryptosporidium in mice by inhibiting parasite phenylalanyl-tRNA synthetase.* Sci Transl Med, 2020. **12**(563).

46. Shaw, S., et al., *Genetic crosses within and between species of Cryptosporidium.* Proc Natl Acad Sci U S A, 2024. **121**(1): p. e2313210120.

47. Agyabeng-Dadzie, F., et al., *Evaluating the Benefits and Limits of Multiple Displacement Amplification With Whole-Genome Oxford Nanopore Sequencing.* Molecular Ecology Resources, 2025: p. e14094.

48. Guo, Y., et al., *Isolation and enrichment of Cryptosporidium DNA and verification of DNA purity for whole-genome sequencing.* J Clin Microbiol, 2015. **53**(2): p. 641-7.

49. De Coster, W. and R. Rademakers, *NanoPack2: population-scale evaluation of long-read sequencing data.* Bioinformatics, 2023. **39**(5): p. btad311.

50. Baptista, R.P., et al., *New T2T assembly of Cryptosporidium parvum IOWA annotated with reference genome gene identifiers.* bioRxiv, 2023.

51. Li, H., *New strategies to improve minimap2 alignment accuracy.* Bioinformatics, 2021. **37**(23): p. 4572-4574.

52. Shen, W., et al., *SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation.* PloS one, 2016. **11**(10): p. e0163962.

53. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows–Wheeler transform.* bioinformatics, 2009. **25**(14): p. 1754-1760.

54. Danecek, P., *Bon eld JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM et al: Twelve years of SAMtools and BCFtools.* Gigascience2021, 2021. **10**(2): p. 2.

55. Li, H., *Improving SNP discovery by base alignment quality.* Bioinformatics, 2011. **27**(8): p. 1157-1158.

56. Robinson, J.T., et al., *igv. js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV).* Bioinformatics, 2023. **39**(1): p. btac830.

57. Danecek, P., et al., *Twelve years of SAMtools and BCFtools.* Gigascience, 2021. **10**(2): p. giab008.

58. Kimball, A., et al., *Mendelian segregation and high recombination rates facilitate genetic analyses in Cryptosporidium parvum.* PLoS Genet, 2024. **20**(6): p. e1011162.

59. Gilchrist, C.A. *Cryptosporidium infection in Bangladesh children*. in *Eukaryome Impact on Human Intestine Homeostasis and Mucosal Immunology: Overview of the First Eukaryome Congress at Institut Pasteur. Paris, October 16–18, 2019*. 2020. Springer.

60. Baptista, R.P., et al., *Long-read assembly and comparative evidence-based reanalysis of Cryptosporidium genome sequences reveal expanded transporter repertoire and duplication of entire chromosome ends including subtelomeric regions.* Genome Res, 2022. **32**(1): p. 203-213.

61. Jensen, E.C., *Use of fluorescent probes: their effect on cell biology and limitations.* The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology, 2012. **295**(12): p. 2031-2036.

62. Kulkarni, P., et al., *Capture of water-borne colloids in granular beds using external electric fields: improving removal of Cryptosporidium parvum.* Water Res, 2005. **39**(6): p. 1047-60.

63. Ryan, U.M., et al., *Taxonomy and molecular epidemiology of Cryptosporidium and Giardia - a 50 year perspective (1971-2021).* Int J Parasitol, 2021. **51**(13-14): p. 1099-1119.

64. Ryan, U., et al., *An Update on Zoonotic Cryptosporidium Species and Genotypes in Humans.* Animals (Basel), 2021. **11**(11).

65. Lefebvre, M., et al., *Cryptosporidium-Biofilm Interactions: a Review.* Appl Environ Microbiol, 2021. **87**(3).

66. MacKenzie, W.R., et al., *Massive outbreak of waterborne cryptosporidium infection in Milwaukee, Wisconsin: recurrence of illness and risk of secondary transmission.* Clin Infect Dis, 1995. **21**(1): p. 57-62.

67. Korich, D.G., et al., *Effects of ozone, chlorine dioxide, chlorine, and monochloramine on Cryptosporidium parvum oocyst viability.* Appl Environ Microbiol, 1990. **56**(5): p. 1423-8.

68. Temesgen, T.T., K.R. Tysnes, and L.J. Robertson, *Use of Oxidative Stress Responses to Determine the Efficacy of Inactivation Treatments on Cryptosporidium Oocysts.* Microorganisms, 2021. **9**(7).

69. Wood, M., et al., *Role of filtration in managing the risk from Cryptosporidium in commercial swimming pools - a review.* J Water Health, 2019. **17**(3): p. 357-370.

70. Deng, M.Q. and D.O. Cliver, *Comparative detection of Cryptosporidium parvum oocysts from apple juice.* International Journal of Food Microbiology, 2000. **54**(3): p. 155-162.

71. Devleesschauwer, B., et al., *Risk ranking of foodborne parasites: state of the art.* Food and waterborne parasitology, 2017. **8**: p. 1-13.

72. Ali, M., et al., *Food and waterborne cryptosporidiosis from a One Health perspective: A comprehensive review.* Animals: an Open Access Journal from MDPI, 2024. **14**(22): p. 3287.

73. Baptista, R.P., G.W. Cooper, and J.C. Kissinger, *Challenges for Cryptosporidium Population Studies.* Genes (Basel), 2021. **12**(6).

74. Feng, Y., U.M. Ryan, and L. Xiao, *Genetic Diversity and Population Structure of Cryptosporidium.* Trends Parasitol, 2018. **34**(11): p. 997-1011.

75. Agyabeng-Dadzie, F., R. Xiao, and J.C. Kissinger, *Cryptosporidium Genomics - Current Understanding, Advances, and Applications.* Curr Trop Med Rep, 2024. **11**(2): p. 92-103.

76. Bouzid, M., et al., *Cryptosporidium pathogenicity and virulence.* Clin Microbiol Rev, 2013. **26**(1): p. 115-34.

77. Xiao, L. and Y. Feng, *Molecular epidemiologic tools for waterborne pathogens Cryptosporidium spp. and Giardia duodenalis.* Food Waterborne Parasitol, 2017. **8-9**: p. 14-32.

78. Arias-Agudelo, L.M., et al., *Comparative genomic analysis of the principal Cryptosporidium species that infect humans.* PeerJ, 2020. **8**: p. e10478.

79. Wojcik, G.L., et al., *Genome-wide association study of cryptosporidiosis in infants implicates PRKCA.* MBio, 2020. **11**(1): p. 10.1128/mbio. 03343-19.

80. Khan, A., et al., *Phylogenomic reconstruction of Cryptosporidium spp. captured directly from clinical samples reveals extensive genetic diversity.* bioRxiv, 2024.

81. Innes, E.A., et al., *A one health approach to tackle cryptosporidiosis.* Trends in parasitology, 2020. **36**(3): p. 290-303.

82. Morris, A.V., et al., *Parapipe: A Pipeline for Parasite Next-Generation Sequencing Data Analysis Applied to Cryptosporidium.* Access Microbiology, 2025: p. 000993. v1.

83. Bolger, A.M., M. Lohse, and B. Usadel, *Trimmomatic: a flexible trimmer for Illumina sequence data.* Bioinformatics, 2014. **30**(15): p. 2114-2120.

84. Wood, D.E., J. Lu, and B. Langmead, *Improved metagenomic analysis with Kraken 2.* Genome biology, 2019. **20**: p. 1-13.

85. Baptista, R.d.P., et al., *New T2T assembly of Cryptosporidium parvum IOWA II annotated with Legacy-Compatible Gene identifiers.* Scientific Data, 2025. **12**(1): p. 1039.

86. Li, H., *A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data.* Bioinformatics, 2011. **27**(21): p. 2987-2993.

87. Purcell, S., et al., *PLINK: a tool set for whole-genome association and population-based linkage analyses.* The American journal of human genetics, 2007. **81**(3): p. 559-575.

88. Ortiz, E.M., *vcf2phylip v2. 0: convert a VCF matrix into several matrix formats for phylogenetic analysis*. 2019.

89. Alexander, D.H., J. Novembre, and K. Lange, *Fast model-based estimation of ancestry in unrelated individuals.* Genome research, 2009. **19**(9): p. 1655-1664.

90. Minh, B.Q., et al., *IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era.* Molecular biology and evolution, 2020. **37**(5): p. 1530-1534.

91. Danecek, P., et al., *The variant call format and VCFtools.* Bioinformatics, 2011. **27**(15): p. 2156-2158.

92. Nielsen, R., *Statistical tests of selective neutrality in the age of genomics.* Heredity, 2001. **86**(6): p. 641-647.

93. Tajima, F., *Statistical method for testing the neutral mutation hypothesis by DNA polymorphism.* Genetics, 1989. **123**(3): p. 585-595.

94. Cingolani, P., et al., *A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3.* fly, 2012. **6**(2): p. 80-92.

95. Simão, F.A., et al., *BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs.* Bioinformatics, 2015. **31**(19): p. 3210-3212.

96. Shumate, A. and S.L. Salzberg, *Liftoff: accurate mapping of gene annotations.* Bioinformatics, 2021. **37**(12): p. 1639-1643.

97. Dainat, J., D. Hereñú, and P. Pucholt, *AGAT: Another Gff Analysis Toolkit to handle annotations in any GTF.* GFF format, 2020. **10**.

98. Gonzales-Irribarren, A. and A. Fu, *gtfsort: a tool to efficiently sort GTF files.* bioRxiv, 2023: p. 2023.10. 21.563454.

99. Emms, D.M. and S. Kelly, *OrthoFinder: phylogenetic orthology inference for comparative genomics.* Genome biology, 2019. **20**: p. 1-14.

100. Feng, Y., U.M. Ryan, and L. Xiao, *Genetic diversity and population structure of Cryptosporidium.* Trends in parasitology, 2018. **34**(11): p. 997-1011.

101. Robinson, G., et al., *Deciphering a cryptic minefield: A guide to Cryptosporidium gp60 subtyping.* Current Research in Parasitology & Vector-Borne Diseases, 2025. **7**: p. 100257.

102. Chalmers, R.M., et al., *Analysis of the Cryptosporidium spp. and gp60 subtypes linked to human outbreaks of cryptosporidiosis in England and Wales, 2009 to 2017.* Parasit Vectors, 2019. **12**(1): p. 95.

103. Tichkule, S., et al., *Comparative genomics revealed adaptive admixture in Cryptosporidium hominis in Africa.* Microb Genom, 2021. **7**(1).

104. Bellinzona, G., et al., *Comparative genomics of Cryptosporidium parvum reveals the emergence of an outbreak-associated population in Europe and its spread to the United States.* Genome Res, 2024. **34**(6): p. 877-887.

105. He, W., et al., *SKSR1 identified as key virulence factor in Cryptosporidium by genetic crossing.* Nature communications, 2025. **16**(1): p. 1-15.

106. Li, M., et al., *Variant surface protein GP60 contributes to host infectivity of Cryptosporidium parvum.* Commun Biol, 2024. **7**(1): p. 1175.

107. Szűcs, M., et al., *Rapid adaptive evolution in novel environments acts as an architect of population range expansion.* Proceedings of the National Academy of Sciences, 2017. **114**(51): p. 13501-13506.

108. Sanderson, S.J., et al., *Determining the protein repertoire of Cryptosporidium parvum sporozoites.* Proteomics, 2008. **8**(7): p. 1398-1414.

109. Na, B.K., et al., *Cryptopain-1, a cysteine protease of Cryptosporidium parvum, does not require the pro-domain for folding.* Parasitology, 2009. **136**(2): p. 149-57.

110. Oberstaller, J., et al., *The Cryptosporidium parvum ApiAP2 gene family: insights into the evolution of apicomplexan AP2 regulatory systems.* Nucleic Acids Res, 2014. **42**(13): p. 8271-84.

111. Manque, P.A., et al., *Identification and immunological characterization of three potential vaccinogens against Cryptosporidium species.* Clinical and Vaccine Immunology, 2011. **18**(11): p. 1796-1802.

112. Wang, D., C. Wang, and G. Zhu, *Genomic reconstruction and features of glycosylation pathways in the apicomplexan Cryptosporidium parasites.* Front Mol Biosci, 2022. **9**: p. 1051072.

113. Deng, M., et al., *Cryptosporidium parvum genes containing thrombospondin type 1 domains.* Infect Immun, 2002. **70**(12): p. 6987-95.

114. Oladele, D.B., et al., *A review of recent Cryptosporidium hominis and Cryptosporidium parvum gp60 subtypes.* Current Research in Parasitology & Vector-Borne Diseases, 2025: p. 100292.

Tables

Table 1. A summary sequencing coverage metrics for all *Cryptosporidium parvum* SRA dataset. The average depth and breadth of coverage are shown at different minimum threshold of 5x and 10x. Samples that passed the quality control criteria were retained for downstream analysis.

File	Sample_ID	Avg_ Depth	Breadth _5x	Breadth _10x	Sample excluded
ERR1760143.fastq	ERR1760143--2013-	7.55	0.63	0.27	YES[a]
ERR2889313.fastq	ERR2889313-UK-2015-	168.27	0.99	0.99	NO
ERR2889314.fastq	ERR2889314-UK-2016-	143.81	0.94	0.89	YES[a]
ERR2889315.fastq	ERR2889315-UK-2016-	136.94	0.99	0.98	NO
ERR2889325.fastq	ERR2889325-UK-2015-	85.73	0.98	0.94	NO
ERR2889326.fastq	ERR2889326-UK-2015-	158.42	0.99	0.98	NO
ERR2889327.fastq	ERR2889327-UK-2016-	98.01	0.97	0.93	NO
SRR10336407.fastq	SRR10336407-US-2013-Hs	49.92	0.99	0.98	NO
SRR10336411.fastq	SRR10336411-US-2009-Bt	0.77	0.01	0.00	YES[a]
SRR10336412.fastq	SRR10336412-US-2008-Bt	454.17	1.00	1.00	NO
SRR10336413.fastq	SRR10336413-US--Bt	1.15	0.02	0.00	YES[a]
SRR10336414.fastq	SRR10336414-US--Bt	0.54	0.00	0.00	YES[a]
SRR10336415.fastq	SRR10336415-US--Bt	0.24	0.00	0.00	YES[ab]
SRR10363299.fastq	SRR10363299-US-2013-Hs	81.77	0.99	0.99	NO
SRR10363300.fastq	SRR10363300-US-2016-Hs	317.10	0.99	0.99	NO
SRR10363301.fastq	SRR10363301-US-2016-Hs	314.49	0.99	0.99	NO
SRR10363302.fastq	SRR10363302-US-2016-Hs	352.90	0.99	0.99	NO
SRR10363303.fastq	SRR10363303-US-2016-Hs	247.76	0.99	0.99	NO
SRR10363304.fastq	SRR10363304-US-2016-Hs	315.80	0.99	0.99	NO
SRR10363305.fastq	SRR10363305-US-2016-Hs	341.60	1.00	0.99	NO
SRR10363306.fastq	SRR10363306-US-2016-Hs	286.59	0.99	0.99	NO
SRR10363308.fastq	SRR10363308-US-2016-Hs	357.71	0.99	0.99	NO
SRR10363309.fastq	SRR10363309-US-2015-Hs	84.86	0.98	0.95	YES[b]
SRR10363310.fastq	SRR10363310-CN-2015-Bt	297.54	0.99	0.99	NO
SRR10363311.fastq	SRR10363311-US-2013-Hs	30.52	0.98	0.93	NO
SRR10363315.fastq	SRR10363315-CN-2015-Bt	175.29	0.99	0.99	NO
SRR10363316.fastq	SRR10363316-US-2015-Bt	555.09	1.00	0.99	NO
SRR10363317.fastq	SRR10363317-US-2015-Bt	451.06	1.00	0.99	NO
SRR10363319.fastq	SRR10363319-US-2015-Hs	188.08	0.97	0.95	NO
SRR10363320.fastq	SRR10363320-US-2015-Hs	448.63	0.99	0.99	NO
SRR10363321.fastq	SRR10363321-US-2015-Hs	475.99	1.00	0.99	NO

SRR10363322.fastq	SRR10363322-US-2013-Hs	22.60	0.80	0.61	YES[a]
SRR10363323.fastq	SRR10363323-US-2015-Bt	16.64	0.94	0.76	YES[a]
SRR10363324.fastq	SRR10363324-US-2015-Bt	21.74	0.92	0.79	YES[a]
SRR10363325.fastq	SRR10363325-US-2015-Bt	18.85	0.93	0.78	YES[a]
SRR10363326.fastq	SRR10363326-US-2015-Bt	20.50	0.94	0.80	YES[a]
SRR10363327.fastq	SRR10363327-US-2015-Bt	19.25	0.92	0.76	YES[a]
SRR10363328.fastq	SRR10363328-US-2014-Bt	66.36	0.99	0.99	NO
SRR10363333.fastq	SRR10363333-US-2016-Hs	227.92	0.99	0.99	NO
SRR10363335.fastq	SRR10363335-US-2013-Hs	96.23	0.99	0.99	NO
SRR10363336.fastq	SRR10363336-US-2016-Hs	277.64	0.99	0.99	NO
SRR10363337.fastq	SRR10363337-US-2016-Hs	117.98	0.99	0.99	NO
SRR10363338.fastq	SRR10363338-US-2016-Hs	198.83	0.99	0.99	NO
SRR10363339.fastq	SRR10363339-US-2016-Hs	313.71	0.99	0.99	NO
SRR10363340.fastq	SRR10363340-US-2016-Hs	297.35	0.99	0.99	NO
SRR10363341.fastq	SRR10363341-US-2016-Hs	67.14	0.98	0.94	NO
SRR10363342.fastq	SRR10363342-US-2016-Hs	69.33	0.99	0.97	NO
SRR10363343.fastq	SRR10363343-US-2016-Hs	222.75	0.99	0.99	NO
SRR10363344.fastq	SRR10363344-US-2016-Hs	301.57	0.99	0.99	NO
SRR10363345.fastq	SRR10363345-US-2016-Hs	264.11	0.99	0.99	NO
SRR10363346.fastq	SRR10363346-US-2013-Hs	54.08	0.99	0.98	NO
SRR10363347.fastq	SRR10363347-US-2016-Hs	237.82	0.99	0.99	NO
SRR10363348.fastq	SRR10363348-US-2016-Hs	243.58	0.99	0.99	NO
SRR10363354.fastq	SRR10363354-US-2016-Hs	186.41	0.99	0.99	NO
SRR10363355.fastq	SRR10363355-US-2016-Hs	229.69	0.99	0.99	NO
SRR10363357.fastq	SRR10363357-US-2016-Hs	388.28	0.99	0.99	NO
SRR10363358.fastq	SRR10363358-US-2013-Hs	95.29	0.99	0.98	NO
SRR10363362.fastq	SRR10363362-US-2019-Hs	33.43	0.88	0.75	YES[a]
SRR10363363.fastq	SRR10363363-US-2019-Hs	149.47	0.98	0.96	NO
SRR10363364.fastq	SRR10363364-US-2019-Hs	63.56	0.96	0.92	NO
SRR10363365.fastq	SRR10363365-US-2019-Hs	137.31	0.98	0.96	NO
SRR10363369.fastq	SRR10363369-SP-2014-Hs	67.85	0.60	0.51	YES[a]
SRR10363371.fastq	SRR10363371-US-2019-Hs	1.52	0.08	0.01	YES[a]
SRR10363372.fastq	SRR10363372-US-2019-Hs	14.27	0.76	0.55	YES[a]
SRR10363373.fastq	SRR10363373-US-2019-Hs	1.03	0.04	0.00	YES[a]
SRR10363374.fastq	SRR10363374-US-2018-Hs	5.85	0.45	0.20	YES[a]
SRR10363376.fastq	SRR10363376-US-2018-Hs	7.22	0.47	0.27	YES[a]
SRR10363378.fastq	SRR10363378-US-2018-Hs	83.00	0.98	0.96	NO
SRR10363379.fastq	SRR10363379-US-2018-Hs	77.24	0.97	0.94	NO
SRR10363380.fastq	SRR10363380-US-2014-Bt	54.88	0.99	0.98	NO

SRR10363381.fastq	SRR10363381-US-2018-Hs	3.83	0.31	0.10	YES[a]
SRR10363382.fastq	SRR10363382-US-2018-Hs	53.84	0.86	0.76	YES[a]
SRR10363383.fastq	SRR10363383-US-2018-Hs	41.95	0.92	0.83	YES[a]
SRR10363384.fastq	SRR10363384-US-2018-Hs	71.03	0.98	0.95	NO
SRR10363385.fastq	SRR10363385-US-2018-Hs	62.95	0.96	0.91	NO
SRR10363386.fastq	SRR10363386-US-2018-Bt	19.96	0.71	0.53	YES[a]
SRR10363387.fastq	SRR10363387-US-2018-Bt	32.16	0.91	0.82	YES[a]
SRR10363391.fastq	SRR10363391-US-2014-Bt	82.77	1.00	0.99	NO
SRR10363395.fastq	SRR10363395-US-2017-Hs	419.05	0.99	0.99	NO
SRR10363397.fastq	SRR10363397-US-2017-Hs	175.05	0.99	0.99	NO
SRR10363401.fastq	SRR10363401-US-2017-Hs	112.97	0.99	0.99	NO
SRR10363402.fastq	SRR10363402-US-2014-Bt	78.05	1.00	0.99	NO
SRR10363403.fastq	SRR10363403-US-2017-Hs	143.13	0.99	0.99	NO
SRR10363404.fastq	SRR10363404-US-2017-Hs	293.15	1.00	0.99	NO
SRR10363405.fastq	SRR10363405-US-2017-Hs	198.36	0.99	0.99	NO
SRR10363406.fastq	SRR10363406-US-2017-Hs	139.55	0.99	0.99	NO
SRR10363409.fastq	SRR10363409-US-2017-Hs	181.95	0.99	0.99	NO
SRR10363410.fastq	SRR10363410-US-2017-Hs	179.21	0.99	0.99	NO
SRR10363411.fastq	SRR10363411-US-2017-Hs	138.88	0.99	0.99	NO
SRR10363412.fastq	SRR10363412-US-2017-Hs	231.87	0.99	0.99	NO
SRR10363413.fastq	SRR10363413-US-2014-Bt	90.75	1.00	0.99	NO
SRR10363414.fastq	SRR10363414-US-2017-Hs	277.90	0.99	0.99	NO
SRR10363415.fastq	SRR10363415-US-2017-Bt	3.02	0.06	0.04	YES[ac]
SRR10363416.fastq	SRR10363416-US-2017-Hs	226.82	0.99	0.99	NO
SRR10363417.fastq	SRR10363417-US-2017-Hs	305.33	0.97	0.96	NO
SRR10363419.fastq	SRR10363419-US-2017-Hs	205.78	0.99	0.99	NO
SRR10363421.fastq	SRR10363421-US-2017-Hs	169.18	0.99	0.99	NO
SRR10363422.fastq	SRR10363422-US-2017-Hs	220.29	0.99	0.99	NO
SRR10363423.fastq	SRR10363423-US-2017-Hs	126.27	0.99	0.99	NO
SRR10363424.fastq	SRR10363424-US-2014-Bt	123.42	0.99	0.99	NO
SRR10363425.fastq	SRR10363425-US-2017-Hs	142.43	0.99	0.99	NO
SRR10363428.fastq	SRR10363428-US-2017-Hs	95.45	0.99	0.99	NO
SRR10363429.fastq	SRR10363429-US-2017-Hs	437.98	1.00	0.99	NO
SRR10363430.fastq	SRR10363430-US-2017-Hs	88.19	0.99	0.97	NO
SRR10363431.fastq	SRR10363431-US-2017-Hs	269.82	0.99	0.99	NO
SRR10363432.fastq	SRR10363432-US-2017-Hs	2.29	0.06	0.05	YES[ac]
SRR10363433.fastq	SRR10363433-US-2017-Hs	156.37	0.99	0.99	NO
SRR10363434.fastq	SRR10363434-US-2017-Hs	132.36	0.99	0.99	NO
SRR10363435.fastq	SRR10363435-US-2014-Hs	50.38	0.97	0.92	NO

SRR10363436.fastq	SRR10363436-US-2017-Hs	45.06	0.50	0.42	YES[a]
SRR10363437.fastq	SRR10363437-US-2017-Hs	175.13	0.99	0.99	NO
SRR10363438.fastq	SRR10363438-US-2017-Hs	103.34	0.99	0.97	NO
SRR10363441.fastq	SRR10363441-US-2017-Hs	194.13	0.98	0.96	NO
SRR10363443.fastq	SRR10363443-US-2017-Hs	97.07	0.99	0.99	NO
SRR10363444.fastq	SRR10363444-US-2017-Hs	114.21	0.99	0.99	NO
SRR10363445.fastq	SRR10363445-US-2017-Hs	106.04	0.99	0.99	NO
SRR10363446.fastq	SRR10363446-US-2014-Hs	19.12	0.92	0.69	YES[a]
SRR10363447.fastq	SRR10363447-US-2017-Hs	111.70	0.99	0.99	NO
SRR10363448.fastq	SRR10363448-US-2017-Hs	111.47	0.99	0.99	NO
SRR10363449.fastq	SRR10363449-US-2017-Hs	86.07	0.99	0.98	NO
SRR10363450.fastq	SRR10363450-US-2017-Hs	134.83	0.99	0.99	NO
SRR10363451.fastq	SRR10363451-US-2017-Hs	61.40	0.84	0.74	YES[a]
SRR10363453.fastq	SRR10363453-US-2017-Hs	90.05	0.90	0.82	YES[a]
SRR10363457.fastq	SRR10363457-US-2014-Bt	14.58	0.92	0.68	YES[a]
SRR10363460.fastq	SRR10363460-US-2017-Hs	65.94	0.96	0.91	NO
SRR10363461.fastq	SRR10363461-US-2017-Hs	78.24	0.99	0.99	NO
SRR10363462.fastq	SRR10363462-US-2017-Hs	115.71	0.99	0.99	NO
SRR10363463.fastq	SRR10363463-US-2017-Hs	102.32	0.99	0.98	NO
SRR10363465.fastq	SRR10363465-US-2017-Hs	105.38	0.99	0.99	NO
SRR10363466.fastq	SRR10363466-US-2017-Hs	101.19	0.99	0.98	NO
SRR10363467.fastq	SRR10363467-US-2017-Hs	100.56	0.94	0.89	YES[a]
SRR10363468.fastq	SRR10363468-US-2014-Hs	1.51	0.05	0.00	YES[a]
SRR10363469.fastq	SRR10363469-US-2013-Hs	759.90	1.00	1.00	NO
SRR10363470.fastq	SRR10363470-US-2017-Hs	75.27	0.96	0.91	NO
SRR10363473.fastq	SRR10363473-US-2017-Hs	98.26	0.99	0.99	NO
SRR10363474.fastq	SRR10363474-US-2017-Hs	88.54	0.99	0.98	NO
SRR10363475.fastq	SRR10363475-US-2017-Hs	111.34	0.99	0.99	NO
SRR10363476.fastq	SRR10363476-US-2017-Hs	113.58	0.99	0.99	NO
SRR10363477.fastq	SRR10363477-US-2017-Hs	81.80	0.99	0.99	NO
SRR10363478.fastq	SRR10363478-US-2017-Hs	81.12	0.99	0.98	NO
SRR10363479.fastq	SRR10363479-US-2017-Hs	139.18	0.98	0.95	NO
SRR10363480.fastq	SRR10363480-US-2014-Hs	57.42	0.99	0.98	NO
SRR10363481.fastq	SRR10363481-US-2017-Hs	125.12	0.99	0.98	NO
SRR10363483.fastq	SRR10363483-US-2017-Hs	114.24	0.99	0.99	NO
SRR10363484.fastq	SRR10363484-US-2016-Hs	186.10	0.99	0.99	NO
SRR10363485.fastq	SRR10363485-US-2016-Hs	129.62	0.98	0.94	NO
SRR10363486.fastq	SRR10363486-US-2017-Hs	78.05	0.82	0.73	YES[a]
SRR10363487.fastq	SRR10363487-US-2017-Hs	300.66	0.99	0.99	NO

SRR10363488.fastq	SRR10363488-US-2017-Hs	183.09	0.99	0.99	NO
SRR10363491.fastq	SRR10363491-US-2017-Hs	150.26	0.99	0.99	NO
SRR10363492.fastq	SRR10363492-EG-2017-Bt	14.67	0.61	0.41	YES[a]
SRR10363493.fastq	SRR10363493-EG-2017-Bt	133.43	0.98	0.95	NO
SRR10363494.fastq	SRR10363494-US-2014-Hs	6.63	0.66	0.21	YES[a]
SRR10363495.fastq	SRR10363495-EG-2017-Bt	216.95	0.99	0.99	NO
SRR10363496.fastq	SRR10363496-EG-2017-Bt	179.67	0.79	0.70	YES[a]
SRR10363497.fastq	SRR10363497-EG-2017-Bt	136.83	0.84	0.76	YES[a]
SRR10363498.fastq	SRR10363498-EG-2017-Bt	199.43	0.98	0.95	NO
SRR10363499.fastq	SRR10363499-EG-2017-Bt	195.12	0.99	0.99	NO
SRR10363500.fastq	SRR10363500-EG-2017-Bt	189.24	0.99	0.99	NO
SRR10363501.fastq	SRR10363501-EG-2017-Bt	142.26	0.99	0.99	NO
SRR10363504.fastq	SRR10363504-US-2016-Hs	47.98	0.99	0.97	NO
SRR10363505.fastq	SRR10363505-US-2014-Hs	3.15	0.24	0.02	YES[a]
SRR10363506.fastq	SRR10363506-US-2016-Hs	339.36	0.99	0.99	NO
SRR10363509.fastq	SRR10363509-US-2016-Hs	531.08	0.99	0.99	NO
SRR10363511.fastq	SRR10363511-US-2016-Hs	176.44	0.99	0.99	NO
SRR10363512.fastq	SRR10363512-US-2016-Hs	660.89	1.00	0.99	NO
SRR10363514.fastq	SRR10363514-US-2016-Hs	392.42	0.99	0.99	NO
SRR10363515.fastq	SRR10363515-US-2016-Hs	516.76	1.00	0.99	NO
SRR10363516.fastq	SRR10363516-US-2016-Hs	372.04	0.99	0.99	NO
SRR10363517.fastq	SRR10363517-US-2013-Hs	46.20	0.99	0.98	YESb
SRR10363518.fastq	SRR10363518-US-2016-Hs	445.75	0.99	0.99	NO
SRR10363519.fastq	SRR10363519-US-2016-Hs	357.51	0.99	0.99	NO
SRR10363520.fastq	SRR10363520-US-2016-Hs	528.21	1.00	0.99	NO
SRR10363521.fastq	SRR10363521-US-2016-Hs	289.77	0.99	0.99	NO
SRR10363522.fastq	SRR10363522-US-2016-Hs	244.35	0.99	0.99	NO
SRR10363523.fastq	SRR10363523-US-2016-Hs	526.61	1.00	0.99	NO
SRR10363525.fastq	SRR10363525-US-2016-Hs	467.86	0.99	0.99	NO
SRR10363526.fastq	SRR10363526-US-2016-Hs	297.98	0.99	0.99	NO
SRR10363527.fastq	SRR10363527-US-2016-Hs	160.52	0.98	0.97	NO
SRR10363528.fastq	SRR10363528-US-2013-Hs	2.83	0.18	0.02	YES[a]
SRR10363529.fastq	SRR10363529-US-2016-Hs	594.16	1.00	0.99	NO
SRR10363530.fastq	SRR10363530-US-2016-Hs	428.35	1.00	0.99	NO
SRR10363531.fastq	SRR10363531-US-2016-Hs	345.20	0.99	0.99	NO
SRR10363532.fastq	SRR10363532-US-2016-Hs	464.16	0.99	0.99	NO
SRR10363533.fastq	SRR10363533-US-2016-Hs	295.73	0.99	0.99	NO
SRR10363535.fastq	SRR10363535-CN-2016-Bt	273.49	0.98	0.97	NO
SRR10363536.fastq	SRR10363536-CN-2016-Bt	284.68	0.99	0.99	NO
SRR10363537.fastq	SRR10363537-CN-2016-Bt	292.45	0.99	0.99	NO

SRR10363538.fastq	SRR10363538-US-2016-Hs	251.54	0.99	0.99	NO
SRR10363539.fastq	SRR10363539-US-2016-Hs	64.60	0.93	0.86	YES[a]
SRR10363540.fastq	SRR10363540-US-2013-Hs	98.22	1.00	0.99	NO
SRR10363547.fastq	SRR10363547-US-2016-Hs	165.43	0.99	0.99	NO
SRR10363551.fastq	SRR10363551-US-2013-Hs	14.92	0.88	0.64	YES[a]
SRR10363552.fastq	SRR10363552-US-2016-Hs	52.98	0.98	0.98	Noh
SRR10363554.fastq	SRR10363554-US-2016-Hs	222.21	0.99	0.99	NO
SRR10363556.fastq	SRR10363556-US-2016-Hs	120.14	0.99	0.99	NO
SRR10363557.fastq	SRR10363557-US-2016-Hs	362.81	0.99	0.99	NO
SRR10363558.fastq	SRR10363558-US-2013-Hs	32.09	0.98	0.96	NO
SRR10363559.fastq	SRR10363559-US-2012-Hs	779.64	1.00	1.00	NO
SRR10363561.fastq	SRR10363561-US-2016-Hs	457.42	0.99	0.99	NO
SRR10363562.fastq	SRR10363562-US-2016-Hs	466.26	0.99	0.99	NO
SRR10363563.fastq	SRR10363563-CN-2016-Bt	253.34	0.99	0.99	NO
SRR10363564.fastq	SRR10363564-CN-2016-Bt	229.80	0.99	0.99	NO
SRR10363565.fastq	SRR10363565-CN-2016-Bt	290.48	0.99	0.99	NO
SRR10363566.fastq	SRR10363566-US-2013-Hs	57.14	0.94	0.86	YES[a]
SRR10363567.fastq	SRR10363567-CN-2016-Bt	204.72	0.99	0.99	NO
SRR10363568.fastq	SRR10363568-CN-2016-Bt	298.54	1.00	0.99	NO
SRR10363569.fastq	SRR10363569-CN-2016-Bt	252.17	0.99	0.99	NO
SRR10363570.fastq	SRR10363570-CN-2016-Bt	243.00	0.99	0.99	NO
SRR10363571.fastq	SRR10363571-CN-2016-Bt	256.63	0.99	0.99	NO
SRR10363572.fastq	SRR10363572-CN-2016-Bt	255.56	0.99	0.99	NO
SRR10363573.fastq	SRR10363573-CN-2016-Bt	265.22	0.99	0.99	NO
SRR10363574.fastq	SRR10363574-CN-2016-Bt	275.40	0.99	0.99	NO
SRR10363575.fastq	SRR10363575-CN-2016-Bt	255.61	0.99	0.99	NO
SRR10363576.fastq	SRR10363576-CN-2016-Bt	301.52	1.00	0.99	NO
SRR10363577.fastq	SRR10363577-US-2013-Hs	13.58	0.87	0.61	YES[a]
SRR10363578.fastq	SRR10363578-CN-2016-Bt	233.76	0.99	0.99	NO
SRR10363579.fastq	SRR10363579-CN-2016-Bt	217.12	0.99	0.99	NO
SRR10363580.fastq	SRR10363580-CN-2016-Bt	287.89	1.00	0.99	NO
SRR10363581.fastq	SRR10363581-CN-2016-Bt	300.35	0.99	0.99	NO
SRR10363582.fastq	SRR10363582-CN-2016-Bt	106.24	0.99	0.99	NO
SRR10363583.fastq	SRR10363583-US-2016-Hs	483.64	0.99	0.99	NO
SRR10363584.fastq	SRR10363584-US-2016-Hs	410.09	0.99	0.99	NO
SRR10363585.fastq	SRR10363585-US-2016-Hs	74.32	0.55	0.53	YES[ac]
SRR10363586.fastq	SRR10363586-US-2016-Hs	111.30	0.99	0.99	NO
SRR10363587.fastq	SRR10363587-US-2016-Hs	241.45	0.99	0.99	NO
SRR10363589.fastq	SRR10363589-US-2013-Hs	91.51	0.99	0.99	NO
SRR10363592.fastq	SRR10363592-US-2016-Hs	185.28	0.99	0.99	NO

SRR10363594.fastq	SRR10363594-US-2016-Hs	265.31	0.99	0.99	NO
SRR10363596.fastq	SRR10363596-EG-2016-Hs	127.14	0.99	0.99	NO
SRR10363597.fastq	SRR10363597-US-2016-Hs	127.41	0.99	0.98	NO
SRR10363598.fastq	SRR10363598-US-2016-Hs	379.49	0.99	0.99	NO
SRR10363599.fastq	SRR10363599-US-2016-Hs	226.11	0.99	0.99	NO
SRR10996883.fastq	SRR10996883-UG-Date-Hs	190.40	1.00	0.99	NO
SRR10997270.fastq	SRR10997270-UG-2003-Hs	8.80	0.65	0.32	YES[a]
SRR10997286.fastq	SRR10997286-UG-2003-Hs	3.67	0.28	0.06	YES[a]
SRR10997324.fastq	SRR10997324-UG-2003-Hs	4.58	0.37	0.10	YES[a]
SRR11516700.fastq	SRR11516700-FR-2017-Bt	127.22	0.99	0.99	NO
SRR11516701.fastq	SRR11516701-US-2017-Bt	193.97	0.98	0.97	NO
SRR11516702.fastq	SRR11516702-US-2017-Bt	204.89	0.98	0.97	NO
SRR11516703.fastq	SRR11516703-US-2017-Bt	187.45	0.98	0.97	NO
SRR11817809.fastq	SRR11817809-IT-2016-Bt	424.36	1.00	0.99	NO
SRR11817810.fastq	SRR11817810-IT-2016-Oa	305.77	1.00	0.99	NO
SRR11817811.fastq	SRR11817811-IT-2016-Bt	384.28	1.00	0.99	NO
SRR11817812.fastq	SRR11817812-IT-2016-Oa	357.47	1.00	0.99	NO
SRR11817813.fastq	SRR11817813-IT-2015-Oa	621.21	1.00	1.00	NO
SRR11817814.fastq	SRR11817814-IT-2015-Oa	589.06	1.00	1.00	NO
SRR11817815.fastq	SRR11817815-IT-2015-Ch	538.83	1.00	1.00	NO
SRR11817816.fastq	SRR11817816-IT-2015-Ch	601.45	1.00	1.00	NO
SRR11817817.fastq	SRR11817817-IT-2015-Bt	981.71	1.00	1.00	NO
SRR11817818.fastq	SRR11817818-IT-1990-Bt	862.52	1.00	1.00	NO
SRR11817819.fastq	SRR11817819-DM-1996-Bt	982.26	1.00	1.00	NO
SRR11817820.fastq	SRR11817820-IT-2016-Ch	501.94	0.99	0.99	NO
SRR11817821.fastq	SRR11817821-IT-2016-Oa	367.06	1.00	0.99	NO
SRR11817822.fastq	SRR11817822-IT-2014-Ch	1098.12	1.00	1.00	NO
SRR11817823.fastq	SRR11817823-IT-2013-Ch	652.13	1.00	1.00	NO
SRR11818073.fastq	SRR11818073-SP-2015-Hs	533.58	1.00	0.99	NO
SRR11818074.fastq	SRR11818074-SL-2016-Hs	508.11	1.00	0.99	NO
SRR11818075.fastq	SRR11818075-SL-2016-Hs	473.68	1.00	0.99	NO
SRR11818076.fastq	SRR11818076-SL-2016-Hs	328.50	1.00	0.99	NO
SRR11818077.fastq	SRR11818077-SL-2016-Hs	269.05	0.99	0.99	NO
SRR11818078.fastq	SRR11818078-SL-2016-Hs	543.97	1.00	1.00	NO
SRR11818079.fastq	SRR11818079-SL-2016-Hs	464.14	1.00	1.00	NO
SRR15694483.fastq	SRR15694483-US-2017-Hs	183.09	0.99	0.99	NO
SRR15694484.fastq	SRR15694484-CN-2019-Ct	124.27	0.99	0.98	NO
SRR15694485.fastq	SRR15694485-US-2014-Ct	94.99	0.99	0.99	NO
SRR15694486.fastq	SRR15694486-US-2015-Ct	445.58	0.99	0.99	NO
SRR15694487.fastq	SRR15694487-US-2015-Ct	452.41	1.00	0.99	NO

SRR15694488.fastq	SRR15694488-US-2017-Ct	300.66	0.99	0.99	NO
SRR15694489.fastq	SRR15694489-US-2014-Ct	145.65	1.00	0.99	NO
SRR15694490.fastq	SRR15694490-US-2016-Hs	306.43	0.99	0.99	NO
SRR15694491.fastq	SRR15694491-US-2015-Ct	429.83	1.00	0.99	NO
SRR15694492.fastq	SRR15694492-US--Ct	86.96	0.99	0.99	NO
SRR15694493.fastq	SRR15694493-US-2014-Ct	101.26	1.00	0.99	NO
SRR15694494.fastq	SRR15694494-US-2018-Hs	245.31	0.99	0.99	NO
SRR15694495.fastq	SRR15694495-CN-2019-Ct	103.06	0.98	0.97	NO
SRR15694496.fastq	SRR15694496-US-2015-Ct	475.51	1.00	1.00	NO
SRR15694497.fastq	SRR15694497-US-2014-Ct	66.15	0.99	0.98	NO
SRR15694498.fastq	SRR15694498-US--Ct	76.64	1.00	0.99	NO
SRR15694499.fastq	SRR15694499-US--Ct	255.80	1.00	1.00	NO
SRR15694500.fastq	SRR15694500-US-2015-Ct	350.97	0.99	0.99	NO
SRR15694501.fastq	SRR15694501-US-2018-Hs	241.44	0.99	0.99	NO
SRR15694502.fastq	SRR15694502-US-2015-Ct	64.28	0.98	0.97	NO
SRR15694503.fastq	SRR15694503-US-2015-Ct	422.62	1.00	0.99	NO
SRR15694504.fastq	SRR15694504-US-2015-Ct	348.37	1.00	0.99	NO
SRR15694505.fastq	SRR15694505-US-2015-Ct	488.91	1.00	0.99	NO
SRR15694506.fastq	SRR15694506-CN-2019-Ct	96.73	0.98	0.97	NO
SRR15694507.fastq	SRR15694507-EG-2017-Ct	133.43	0.98	0.95	NO
SRR15694508.fastq	SRR15694508-EG-2017-Ct	216.95	0.99	0.99	NO
SRR15694509.fastq	SRR15694509-EG-2017-Ct	199.43	0.98	0.95	NO
SRR15694510.fastq	SRR15694510-EG-2017-Ct	195.12	0.99	0.99	NO
SRR15694511.fastq	SRR15694511-EG-2017-Ct	189.24	0.99	0.98	NO
SRR15694512.fastq	SRR15694512-EG-2017-Ct	142.26	0.99	0.99	NO
SRR15694513.fastq	SRR15694513-EG-2017-Hs	104.61	0.99	0.98	NO
SRR15694514.fastq	SRR15694514-EG-2017-Hs	127.14	0.99	0.99	NO
SRR15694515.fastq	SRR15694515-EG-2013-Ct	17.14	0.90	0.68	YES[a]
SRR15694516.fastq	SRR15694516-EG-2013-Ct	99.50	0.99	0.99	NO
SRR15694517.fastq	SRR15694517-CN-2019-Ct	92.88	0.98	0.97	NO
SRR15694518.fastq	SRR15694518-CR-2017-Hs	83.75	0.99	0.99	NO
SRR15694519.fastq	SRR15694519-CR-2017-Ct	101.23	0.99	0.99	NO
SRR15694520.fastq	SRR15694520-CN-2020-Ct	0.95	0.04	0.03	YES[ac]
SRR15694521.fastq	SRR15694521-CN-2015-Ct	50.15	0.41	0.35	YES[a]
SRR15694522.fastq	SRR15694522-CN-2016-Ct	253.34	0.99	0.99	NO
SRR15694523.fastq	SRR15694523-CN-2016-Ct	229.80	0.99	0.99	NO
SRR15694524.fastq	SRR15694524-CN-2016-Ct	290.48	0.99	0.99	NO
SRR15694525.fastq	SRR15694525-CN-2016-Ct	204.73	0.99	0.99	NO
SRR15694526.fastq	SRR15694526-CN-2016-Ct	298.54	1.00	0.99	NO
SRR15694527.fastq	SRR15694527-CN-2016-Ct	252.17	0.99	0.99	NO

SRR15694528.fastq	SRR15694528-CN-2018-Ct	189.98	0.98	0.98	NO
SRR15694529.fastq	SRR15694529-CN-2016-Ct	243.00	0.99	0.99	NO
SRR15694530.fastq	SRR15694530-CN-2016-Ct	256.63	0.99	0.99	NO
SRR15694531.fastq	SRR15694531-CN-2016-Ct	255.56	0.99	0.99	NO
SRR15694532.fastq	SRR15694532-CN-2016-Ct	265.22	0.99	0.99	NO
SRR15694533.fastq	SRR15694533-CN-2016-Ct	275.40	0.99	0.99	NO
SRR15694534.fastq	SRR15694534-CN-2016-Ct	255.61	0.99	0.99	NO
SRR15694535.fastq	SRR15694535-CN-2016-Ct	301.52	1.00	0.99	NO
SRR15694536.fastq	SRR15694536-CN-2016-Ct	233.76	0.99	0.99	NO
SRR15694537.fastq	SRR15694537-CN-2016-Ct	217.12	0.99	0.99	NO
SRR15694538.fastq	SRR15694538-CN-2016-Ct	287.89	1.00	0.99	NO
SRR15694539.fastq	SRR15694539-CN-2017-Ct	187.83	0.99	0.99	NO
SRR15694540.fastq	SRR15694540-CN-2016-Ct	300.35	0.99	0.99	NO
SRR15694541.fastq	SRR15694541-CN-2016-Ct	106.24	0.99	0.99	NO
SRR15694542.fastq	SRR15694542-CN-2015-Ct	297.54	0.99	0.99	NO
SRR15694543.fastq	SRR15694543-CN-2015-Ct	330.43	0.99	0.99	NO
SRR15694544.fastq	SRR15694544-CN-2015-Ct	318.02	0.99	0.99	NO
SRR15694545.fastq	SRR15694545-CN-2015-Ct	175.29	0.99	0.99	NO
SRR15694546.fastq	SRR15694546-CN-2016-Ct	273.49	0.98	0.97	NO
SRR15694547.fastq	SRR15694547-CN-2016-Ct	284.68	0.99	0.99	NO
SRR15694548.fastq	SRR15694548-CN-2016-Ct	292.45	0.99	0.99	NO
SRR15694549.fastq	SRR15694549-CN-2019-Bc	166.02	0.98	0.98	NO
SRR15694550.fastq	SRR15694550-CN-2017-Ct	179.46	0.99	0.98	NO
SRR15694551.fastq	SRR15694551-CN-2019-Ct	110.26	0.98	0.97	NO
SRR15694552.fastq	SRR15694552-CN--M-GKO	596.53	0.99	0.99	NO
SRR15694553.fastq	SRR15694553-CN-2019-Ct	1.27	0.07	0.01	YES[a]
SRR15694554.fastq	SRR15694554-CN-2019-Ct	15.30	0.49	0.41	YES[ac]
SRR15694555.fastq	SRR15694555-CN-2019-Ct	103.61	0.99	0.98	NO
SRR15694556.fastq	SRR15694556-CN-2019-Ct	90.15	0.98	0.97	NO
SRR15694557.fastq	SRR15694557-CN-2019-Ct	86.60	0.98	0.97	NO
SRR15694558.fastq	SRR15694558-CN-2019-Ct	106.24	0.99	0.98	NO
SRR15694559.fastq	SRR15694559-CN-2019-Ct	91.76	0.98	0.97	NO
SRR15694560.fastq	SRR15694560-CN-2019-Ct	93.14	0.99	0.98	NO
SRR15694561.fastq	SRR15694561-CN-2017-Ct	171.66	0.99	0.99	NO
SRR15694562.fastq	SRR15694562-CN-2019-Ct	125.10	0.99	0.98	NO
SRR15694563.fastq	SRR15694563-CN-2019-Ct	153.94	0.98	0.98	NO
SRR15694564.fastq	SRR15694564-CN-2019-Ct	177.99	0.98	0.98	NO
SRR15694565.fastq	SRR15694565-CN-2019-Ct	99.72	0.99	0.98	NO
SRR15694566.fastq	SRR15694566-CN-2019-Ct	93.33	0.98	0.98	NO

SRR15694567.fastq	SRR15694567-CN-2019-Ct	116.52	0.98	0.97	NO
SRR15694568.fastq	SRR15694568-CN-2019-Ct	83.80	0.98	0.97	NO
SRR15694569.fastq	SRR15694569-CN-2019-Ct	84.95	0.98	0.96	NO
SRR15694570.fastq	SRR15694570-CN-2019-Ct	84.40	0.98	0.97	NO
SRR15694571.fastq	SRR15694571-CN-2017-Ad	36.87	0.79	0.66	YES[a]
SRR15694572.fastq	SRR15694572-CN-2017-Ct	198.95	0.99	0.99	NO
SRR15694573.fastq	SRR15694573-CN-2017-Ct	201.90	0.99	0.99	NO
SRR16770664.fastq	SRR16770664-US-2020-V	671.87	1.00	1.00	NO
SRR16990213.fastq	SRR16990213-US-2020-Ct	278.47	1.00	1.00	NO
SRR18395901.fastq	SRR18395901-FR-2022-Hs	142.76	0.98	0.97	NO
SRR21763624.fastq	SRR21763624-IT-2009-Bt	204.74	1.00	0.99	NO
SRR21763625.fastq	SRR21763625-IT-2020-Bt	173.74	0.99	0.99	NO
SRR21763626.fastq	SRR21763626-IT-2020-Bt	198.04	0.99	0.99	NO
SRR21763627.fastq	SRR21763627-IT-2020-Bt	135.55	0.99	0.99	NO
SRR21763628.fastq	SRR21763628-IT-2020-Bt	162.92	0.99	0.99	NO
SRR21763629.fastq	SRR21763629-IT-2020-Oa	161.98	1.00	0.99	NO
SRR21763630.fastq	SRR21763630-IT-2020-Oa	39.54	0.98	0.95	NO
SRR21763631.fastq	SRR21763631-IT-2020-Oa	197.05	0.99	0.99	NO
SRR21763632.fastq	SRR21763632-IT-2020-Bt	173.17	1.00	0.99	NO
SRR21763633.fastq	SRR21763633-IT-2020-Bt	160.38	0.99	0.99	NO
SRR21763634.fastq	SRR21763634-IT-2020-Bt	203.59	1.00	0.99	NO
SRR21763635.fastq	SRR21763635-IT-2020-Oa	85.91	0.99	0.99	NO
SRR21763636.fastq	SRR21763636-IT-2020-Oa	147.57	0.99	0.99	NO
SRR25512130.fastq	SRR25512130-US--Mm	78.54	1.00	0.99	NO
SRR26320644.fastq	SRR26320644-UK-2016-Hs	126.87	0.99	0.99	NO
SRR26320645.fastq	SRR26320645-UK-2016-Hs	64.11	0.88	0.79	YES[a]
SRR26320646.fastq	SRR26320646-UK-2016-Hs	19.50	0.39	0.28	YES[a]
SRR26320647.fastq	SRR26320647-UK-2016-Hs	147.66	0.99	0.98	NO
SRR26320648.fastq	SRR26320648-UK-2016-Hs	170.47	0.99	0.99	NO
SRR26320649.fastq	SRR26320649-UK-2016-Hs	153.16	0.99	0.99	NO
SRR26320650.fastq	SRR26320650-UK-2016-Hs	0.76	0.02	0.01	YES[a]
SRR26320651.fastq	SRR26320651-UK-2016-Hs	186.46	0.99	0.99	NO
SRR26320652.fastq	SRR26320652-UK-2016-Hs	147.37	0.99	0.97	NO
SRR26320653.fastq	SRR26320653-UK-2016-Hs	179.63	0.99	0.99	NO
SRR26320654.fastq	SRR26320654-DM-2020-Bt	30.53	0.88	0.76	YES[a]
SRR26320655.fastq	SRR26320655-UK-2016-Hs	200.16	0.99	0.99	NO
SRR26320656.fastq	SRR26320656-UK-2016-Hs	161.19	0.99	0.99	NO
SRR26320657.fastq	SRR26320657-UK-2016-Hs	173.78	0.99	0.99	NO
SRR26320658.fastq	SRR26320658-SW-2020-Hs	73.06	0.98	0.96	NO
SRR26320659.fastq	SRR26320659-SW-2020-Hs	0.53	0.02	0.01	YES[ad]

SRR26320660.fastq	SRR26320660-SW-2020-Hs	40.73	0.99	0.97	NO
SRR26320661.fastq	SRR26320661-SW-2020-Hs	8.74	0.71	0.35	YES[a]
SRR26320662.fastq	SRR26320662-SW-2020-Bt	507.18	0.99	0.99	NO
SRR26320663.fastq	SRR26320663-SW-2020-Bt	650.76	1.00	0.99	NO
SRR26320664.fastq	SRR26320664-SW-2020-Bt	537.83	0.99	0.99	NO
SRR26320665.fastq	SRR26320665-DM-2020-Bt	856.99	1.00	0.99	NO
SRR26320666.fastq	SRR26320666-PG-2020-Bt	481.57	0.99	0.99	NO
SRR26320667.fastq	SRR26320667-PO-2020-Bt	330.21	0.99	0.99	NO
SRR26320669.fastq	SRR26320669-NW-2020-Bt	346.34	0.99	0.98	NO
SRR26320670.fastq	SRR26320670-NW-2020-Bt	538.26	0.99	0.99	NO
SRR26320671.fastq	SRR26320671-NW-2020-Bt	468.99	0.99	0.99	NO
SRR26320672.fastq	SRR26320672-NW-2020-Bt	434.38	0.99	0.99	NO
SRR26320673.fastq	SRR26320673-HU-2020-Bt	559.51	0.99	0.99	NO
SRR26320674.fastq	SRR26320674-HU-2020-Bt	427.87	0.99	0.99	NO
SRR26320675.fastq	SRR26320675-HU-2020-Bt	419.83	0.99	0.99	NO
SRR26320676.fastq	SRR26320676-DM-2020-Bt	45.75	0.90	0.81	YES[a]
SRR26320677.fastq	SRR26320677-HU-2020-Bt	588.85	0.99	0.99	NO
SRR26320678.fastq	SRR26320678-HU-2020-Bt	481.76	0.99	0.99	NO
SRR26320679.fastq	SRR26320679-HU-2020-Bt	880.33	0.99	0.99	NO
SRR26320680.fastq	SRR26320680-HU-2020-Bt	557.83	0.99	0.99	NO
SRR26320681.fastq	SRR26320681-HU-2020-Bt	506.70	0.99	0.99	NO
SRR26320682.fastq	SRR26320682-HU-2020-Bt	480.18	0.99	0.99	NO
SRR26320683.fastq	SRR26320683-HU-2020-Bt	559.81	0.99	0.99	NO
SRR26320684.fastq	SRR26320684-HU-2020-Bt	658.24	0.99	0.99	NO
SRR26320685.fastq	SRR26320685-GE-2020-Bt	186.33	0.99	0.99	NO
SRR26320686.fastq	SRR26320686-GE-2020-Bt	440.27	1.00	1.00	NO
SRR26320687.fastq	SRR26320687-DM-2020-Bt	482.18	0.99	0.99	NO
SRR26320688.fastq	SRR26320688-GE-2020-Bt	337.77	1.00	0.99	NO
SRR26320689.fastq	SRR26320689-GE-2020-Bt	381.39	1.00	1.00	NO
SRR26320690.fastq	SRR26320690-GE-2020-Bt	387.91	1.00	1.00	NO
SRR26320691.fastq	SRR26320691-GE-2020-Bt	351.15	1.00	1.00	NO
SRR26320692.fastq	SRR26320692-GE-2020-Bt	434.56	1.00	1.00	NO
SRR26320693.fastq	SRR26320693-GE-2020-Bt	440.88	1.00	1.00	NO
SRR26320694.fastq	SRR26320694-GE-2020-Bt	399.91	1.00	1.00	NO
SRR26320695.fastq	SRR26320695-GE-2020-Bt	240.90	0.99	0.99	NO
SRR26320696.fastq	SRR26320696-GE-2020-Bt	216.63	0.99	0.99	NO
SRR26320697.fastq	SRR26320697-GE-2020-Bt	224.72	0.99	0.99	NO
SRR26320698.fastq	SRR26320698-DM-2020-Bt	0.06	0.00	0.00	YES[ad]
SRR26320699.fastq	SRR26320699-GE-2020-Bt	176.13	0.99	0.98	NO
SRR26320700.fastq	SRR26320700-GE-2020-Bt	206.64	0.99	0.99	NO

SRR26320701.fastq	SRR26320701-GE-2020-Bt	204.73	0.99	0.99	NO
SRR26320702.fastq	SRR26320702-GE-2020-Bt	219.90	0.99	0.99	NO
SRR26320703.fastq	SRR26320703-FR-2019-Ca	0.52	0.01	0.00	YES[a]
SRR26320704.fastq	SRR26320704-DM-2020-Bt	12.54	0.62	0.42	YES[a]
SRR26320705.fastq	SRR26320705-FL-2019-Bt	162.15	0.99	0.99	NO
SRR26320706.fastq	SRR26320706-FL-2019-Hs	207.86	0.99	0.99	NO
SRR26320707.fastq	SRR26320707-FL-2019-Hs	4.38	0.21	0.11	YES[a]
SRR26320708.fastq	SRR26320708-FL-2019-Hs	4.11	0.20	0.10	YES[a]
SRR26320709.fastq	SRR26320709-FL-2019-Hs	145.24	0.99	0.99	YESb
SRR26320710.fastq	SRR26320710-FL-2019-Hs	0.31	0.00	0.00	YES[ad]
SRR26320711.fastq	SRR26320711-FL-2019-Hs	0.04	0.00	0.00	YES[a]
SRR26320712.fastq	SRR26320712-FL-2019-Bt	184.07	1.00	0.99	NO
SRR26320713.fastq	SRR26320713-FL-2019-Bt	206.14	1.00	0.99	NO
SRR26320714.fastq	SRR26320714-FL-2019-Bt	163.12	1.00	0.99	NO
SRR26320715.fastq	SRR26320715-DM-2020-Bt	0.10	0.00	0.00	YES[ad]
SRR26320716.fastq	SRR26320716-FL-2019-Hs	12.59	0.31	0.19	YES[a]
SRR26320717.fastq	SRR26320717-FL-2019-Bt	153.96	0.99	0.99	NO
SRR26320718.fastq	SRR26320718-FL-2019-Hs	115.30	0.99	0.99	NO
SRR26320719.fastq	SRR26320719-FL-2020-Bt	202.65	1.00	1.00	NO
SRR26320720.fastq	SRR26320720-FR-2019-Ca	230.13	0.99	0.98	NO
SRR26320721.fastq	SRR26320721-FR-2019-Ca	342.39	0.99	0.98	NO
SRR26320722.fastq	SRR26320722-FR-2019-Oa	390.43	0.99	0.99	NO
SRR26320723.fastq	SRR26320723-FR-2019-Oa	344.15	0.99	0.99	NO
SRR26320724.fastq	SRR26320724-FR-2019-Oa	59.95	0.94	0.89	YES[a]
SRR26320725.fastq	SRR26320725-DM-2020-Bt	10.68	0.62	0.39	YES[a]
SRR26320726.fastq	SRR26320726-FR-2018-Bt	430.15	0.99	0.99	NO
SRR26320727.fastq	SRR26320727-FR-2018-Bt	466.89	0.99	0.99	NO
SRR26320728.fastq	SRR26320728-FL-2019-Hs	1.32	0.05	0.02	YES[ad]
SRR26320729.fastq	SRR26320729-FL-2019-Hs	112.12	0.99	0.99	NO
SRR26320730.fastq	SRR26320730-FL-2019-Hs	1.43	0.00	0.00	YES[a]
SRR26320731.fastq	SRR26320731-FL-2020-Bt	236.48	1.00	1.00	NO
SRR26320732.fastq	SRR26320732-FL-2020-Bt	281.97	1.00	0.99	NO
SRR26320733.fastq	SRR26320733-FL-2020-Bt	192.19	1.00	0.99	NO
SRR26320734.fastq	SRR26320734-FL-2019-Hs	187.29	0.99	0.99	NO
SRR26320735.fastq	SRR26320735-FL-2019-Hs	148.94	0.99	0.99	NO
SRR26320737.fastq	SRR26320737-FL-2019-Hs	87.04	0.99	0.96	YESd
SRR26320738.fastq	SRR26320738-DM-2020-Bt	0.60	0.01	0.00	YES[ad]
SRR26320739.fastq	SRR26320739-DM-2020-Bt	72.32	0.96	0.91	NO
SRR26320740.fastq	SRR26320740-DM-2020-Bt	262.11	0.99	0.99	NO

SRR26320741.fastq	SRR26320741-UK-2016-Hs	143.17	0.99	0.99	NO
SRR26320742.fastq	SRR26320742-UK-2016-Hs	180.71	0.99	0.99	NO
SRR26320743.fastq	SRR26320743-UK-2016-Hs	198.28	0.99	0.99	NO
SRR26320744.fastq	SRR26320744-UK-2015-Hs	82.21	0.67	0.58	YES[a]
SRR26320745.fastq	SRR26320745-UK-2015-Hs	111.19	0.99	0.99	NO
SRR26320746.fastq	SRR26320746-UK-2015-Hs	60.46	0.90	0.82	YES[a]
SRR26320747.fastq	SRR26320747-UK-2015-Hs	5.95	0.19	0.12	YES[a]
SRR26320748.fastq	SRR26320748-UK-2017-Bt	205.32	1.00	0.99	NO
SRR26320749.fastq	SRR26320749-UK-2017-Bt	196.86	0.99	0.99	NO
SRR26320750.fastq	SRR26320750-UK-2015-Hs	15.87	0.49	0.36	YES[a]
SRR26320751.fastq	SRR26320751-DM-2020-Bt	447.66	0.99	0.99	NO
SRR26320752.fastq	SRR26320752-DM-2020-Bt	0.50	0.01	0.00	YES[ad]
SRR26320753.fastq	SRR26320753-DM-2020-Bt	330.34	0.99	0.99	NO
SRR3091751.fastq	SRR3091751-SW-2012-Bt	340.11	0.93	0.89	YES[a]
SRR3091752.fastq	SRR3091752-SW--Bt	72.67	0.83	0.72	YES[a]
SRR3091753.fastq	SRR3091753-SW--Bt	87.49	0.43	0.35	YES[a]
SRR3091754.fastq	SRR3091754-SW--Bt	106.16	0.85	0.78	YES[a]
SRR3091755.fastq	SRR3091755-SW--Bt	136.30	0.73	0.63	YES[a]
SRR3091756.fastq	SRR3091756-SW--Bt	25.40	0.47	0.35	YES[a]
SRR3091757.fastq	SRR3091757-SW--Bt	86.06	0.56	0.46	YES[a]
SRR3091758.fastq	SRR3091758-SW--Bt	87.65	0.48	0.39	YES[a]
SRR3091759.fastq	SRR3091759-SW--Bt	79.68	0.60	0.52	YES[a]
SRR3091760.fastq	SRR3091760-SW--Bt	94.17	0.69	0.59	YES[a]
SRR3091761.fastq	SRR3091761-SW--Bt	99.45	0.86	0.79	YES[a]
SRR3473975.fastq	SRR3473975-CN-2009-Bt	125.62	0.99	0.98	NO
SRR6117460.fastq	SRR6117460-UK-2012-Hs	43.22	0.99	0.98	NO
SRR6147472.fastq	SRR6147472-UK-2020-Hs	201.82	0.99	0.97	NO
SRR6147581.fastq	SRR6147581-UK-2013-Hs	223.63	0.99	0.97	NO
SRR6147587.fastq	SRR6147587-UK-2013-Hs	19.29	0.97	0.87	YES[a]
SRR6147945.fastq	SRR6147945-UK-2013-Hs	81.04	0.99	0.99	NO
SRR6147964.fastq	SRR6147964-UK-2013-Hs	97.91	0.96	0.91	NO
SRR6148259.fastq	SRR6148259-UK-2013-Hs	244.83	0.99	0.97	NO
SRR6813716.fastq	SRR6813716-UK-2003-Hs	53.35	0.83	0.77	YES[a]
SRR6813717.fastq	SRR6813717-UK-2013-Hs	30.90	0.84	0.75	YES[a]
SRR6813718.fastq	SRR6813718-UK-2013-Hs	60.63	0.87	0.78	YES[a]
SRR6813719.fastq	SRR6813719-UK-2014-Hs	106.32	0.98	0.95	NO
SRR6871415.fastq	SRR6871415-UK-2012-Hs	910.77	1.00	0.99	NO
SRR7898459.fastq	SRR7898459-UK-2013-Hs	122.91	0.98	0.96	NO
SRR9070312.fastq	SRR9070312-US-2017-Bt	192.76	0.99	0.99	NO

SRR9070313.fastq	SRR9070313-US-2017-Bt	172.40	0.99	0.99	NO
SRR9070314.fastq	SRR9070314-US-2017-Bt	169.47	0.99	0.99	NO

Sample ID: sample name format to include some metadata information (Accession_number - geographical_origin - year_of_collection - host)

[a] genome coverage < 90% at 10x

[b] A potential mixed *Cryptosporidium* infection

[c] The most abundant component is not *C. parvum*

[d] *Cryptosporidium* is not in the top 3 most abundant species present.

Table 2. A summary of genes located at regions of the chromosomes identified as hotspots for high SNP density (variants per kb > 10) with gene description and pN/pS ratios

CHROM	START	END	GENE_ID	Ortholog_ID	Gene description	dN	dS	pN/pS
Chr1	139828	1420 30	cpbgf_100455	cgd1_455	Uncharacterized protein	10.00	8.00	1.25
Chr1	142197	144821	cpbgf_100460	cgd1_460	Uncharacterized protein	14.00	12.00	1.17
Chr1	144503	149975	cpbgf_100470	cgd1_470	Armadillo-type fold/Signal peptide region containing protein	54.00	19.00	2.84
Chr1	148749	150026	cpbgf_100490	cgd1_490	Uncharacterized protein	7.00	12.00	0.58
Chr1	148749	150026	cpbgf_100490	cgd1_490	Uncharacterized protein	7.00	12.00	0.58
Chr1	150370	151175	cpbgf_100493	cgd1_493	Uncharacterized protein	0.00	2.00	0.00
Chr1	151403	153217	cpbgf_100500	cgd1_500	Exonuclease V	4.00	8.00	0.50
Chr1	153151	155131	cpbgf_100510	cgd1_510	Uncharacterized protein	3.00	17.00	0.18
Chr1	155272	156546	cpbgf_100520	cgd1_520	V-type proton ATPase proteolipid subunit	0.00	1.00	0.00
Chr1	157013	157501	cpbgf_100530	cgd1_530	H/ACA ribonucleoprotein complex subunit Nop10	0.00	2.00	0.00
Chr1	157688	158471	cpbgf_100540	cgd1_540	V-type proton ATPase proteolipid subunit	0.00	4.00	0.00
Chr1	159204	160166	cpbgf_100543	cgd1_543	C2H2-type Zinc finger containing protein	0.00	2.00	0.00
Chr4	1100520	1101041	cpbgf_40023	cgd4_23	unspecified product	2.00	0.00	NA
Chr4	1101799	1103191	cpbgf_40020	cgd4_20	Uncharacterized protein	13.00	16.00	0.81
Chr4	1105532	1106182	cpbgf_40010	cgd4_10	Uncharacterized protein	10.00	7.00	1.43
Chr5	598	1483	cpbgf_5005500	cgd5_5500	Uncharacterized protein	11.00	10.00	1.10
Chr5	2092	2797	cpbgf_5005490	cgd5_5490	Uncharacterized protein	1.00	0.00	NA
Chr5	3323	4013	cpbgf_5005480	cgd5_5480	Uncharacterized protein	2.00	0.00	NA
Chr5	5918	6737	cpbgf_5005470	cgd5_5470	Uncharacterized protein	19.00	3.00	6.33
Chr5	16717	18007	cpbgf_5005460	cgd5_5460	Uncharacterized protein	9.00	5.00	1.80
Chr6	4656	5865	cpbgf_60010	cgd6_10	Uncharacterized protein	17.00	3.00	5.67

Chr6	8372	9695	cpbgf_60020	cgd6_20	IMP dehydrogenase/GMP reductase	7.00	10.00	0.70
Chr6	10588	11610	cpbgf_60030	cgd6_30	Uncharacterized protein	2.00	9.00	0.22
Chr6	11996	12826	cpbgf_60040	cgd6_40	Uncharacterized protein	9.00	6.00	1.50
Chr6	13143	15466	cpbgf_60050	cgd6_50	Uncharacterized protein	3.00	0.00	NA
Chr6	16799	22404	cpbgf_60060	cgd6_60	Uncharacterized protein	38.00	21.00	1.81

Table 3. A summary of genes with a pN/pS ratio greater than 1.5 across all populations, including a functional description of the genes.

Gene_ID	Product Description	pN/pS
cpbgf_100120	Signal peptide region containing protein	5.00
cpbgf_100140	putative Secreted Protein (CpLSP gene family)	7.00
cpbgf_100150	putative Secreted Protein (SKSR gene family)	2.20
cpbgf_100470	Armadillo-type fold/Signal peptide region containing protein	2.84
cpbgf_2002950	RING/Armadillo-type fold domain containing protein	2.75
cpbgf_3001160	Uncharacterized protein	1.86
cpbgf_3001170	Uncharacterized protein	3.00
cpbgf_3001880	RNA recognition motif domain containing protein	3.00
cpbgf_3002180	Type I fatty acid synthase	2.20
cpbgf_3003390	RNA polymerase II-associated protein 1 C-terminal	2.00
cpbgf_3003560	CorA family mitochondrial membrane protein	3.00
cpbgf_3004190	Secreted insulinase like peptidase/M16 peptidase-like Metalloenzyme	2.00
cpbgf_3004240	Insulinase (Peptidase family M16)	3.00
cpbgf_4003690	Large glycine-rich repeat low complexity protein	7.00
cpbgf_4003970	Uncharacterized protein	3.50
cpbgf_5003390	Uncharacterized protein	2.00
cpbgf_5005470	Uncharacterized protein	6.33
cpbgf_6001060	Protein with spectrin repeats,CG12008-like	3.00
cpbgf_6001100	putative Secreted Protein	2.00
cpbgf_6001730	FAT/phospatidylinositol 3-4 kinase/Transcription-associated protein 1	3.00
cpbgf_6001970	Uncharacterized protein	3.00
cpbgf_6003050	Uncharacterized protein	6.00
cpbgf_6003940	Uncharacterized protein	4.50
cpbgf_6004070	EEIG1/EHBP1 N-terminal domain containing protein	7.00
cpbgf_6005520	Metalloenzyme,LuxS/M16 peptidase-like	4.40

cpbgf_7001130	Bromo domain-containing protein	7.00
cpbgf_7002340	Uncharacterized Protein	5.00
cpbgf_7004500	putative Secreted Protein	2.75
cpbgf_7004520	ABC transporter integral membrane type-1 fused domain containing protein	2.00

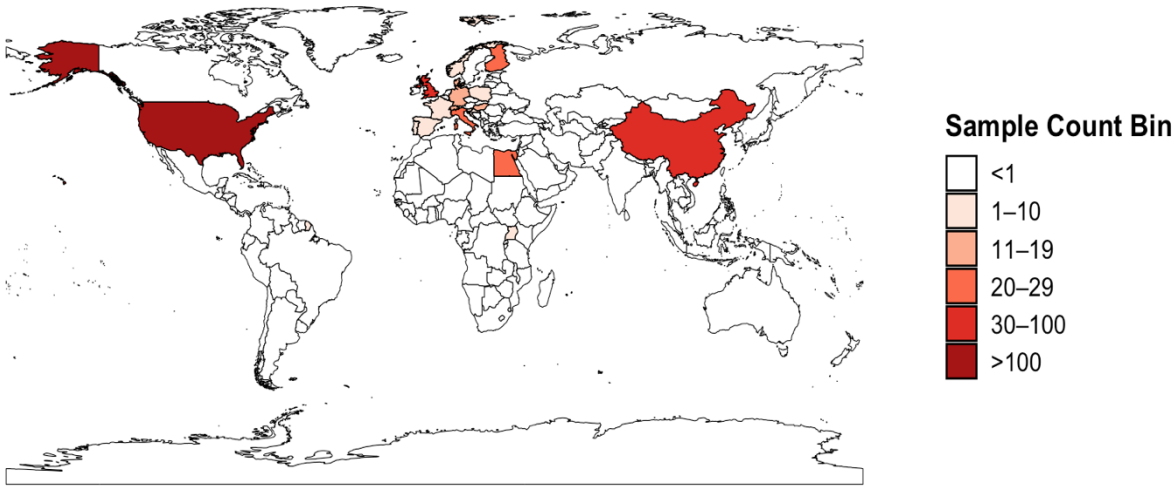
Table 4. Genes located within the top 5% Fst regions on Chromosome 6 from the Africa–USA population comparison. Synonymous and nonsynonymous variant effects were identified using SnpEff, and pN/pS ratios were calculated to evaluate selective pressures acting on these genes

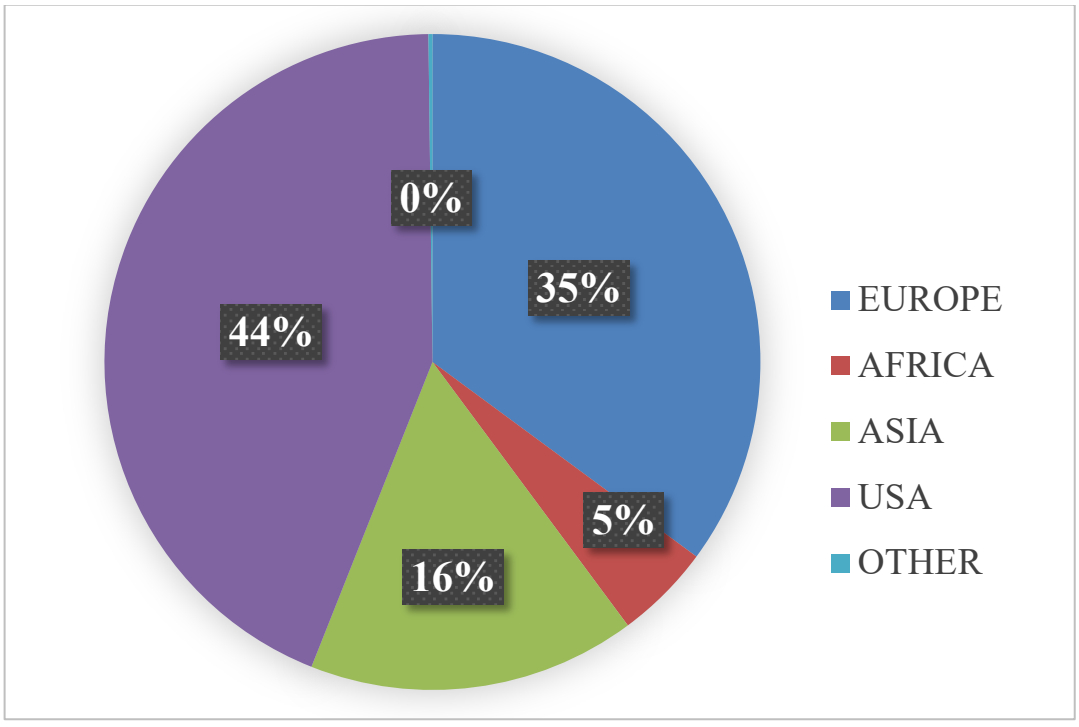
Gene_ID	Product Description	Global_pN/pS
cpbgf_6001340	Calycin/START domain containing protein	0.00
cpbgf_6001370	Hus1-like/Mec3 mitotic and DNA damage checkpoint protein	0.00
cpbgf_6001420	Cdc2-related protein kinase, putative	0.00
cpbgf_6001440	Uncharacterized protein	0.00
cpbgf_6001450	Polysaccharide pyruvyl transferase with Fringe-like domain	0.00
cpbgf_6001470	SET/Zinc finger, RING/FYVE/PHD-type domain containing protein	0.00
cpbgf_6001530	DNA repair protein Rad4	1.00
cpbgf_6001570	Apyrase	NA
cpbgf_6001580	Exonuclease 1/PIN/XPG/Rad2 endonuclease domain containing protein	0.00
cpbgf_6001590	Ataxin-2 C-terminal/MIF4G domain containing protein	NA
cpbgf_6001730	Cleavage and polyadenylation specificity factor 4-like 3x Zn C3H1 domains	3.00
cpbgf_6001750	Rab geranylgeranyl transferase beta / prenyltransferase alpha/alpha toroid fold containing protein	1.00
cpbgf_6001760	Phosphatidylinositol 3-/4-kinase	2.00
cpbgf_6001780	CG6539/Dhh1-like SF II RNA helicase	NA
cpbgf_6001930	Uncharacterized protein	0.00
cpbgf_6001960	Glycosyltransferase 2-like	0.33
cpbgf_6001970	Uncharacterized protein	3.00
cpbgf_6001990	Uncharacterized protein	0.00
cpbgf_6002050	AMP-activated protein kinase beta chain	0.00
cpbgf_6002060	Cdc39p protein-like C-terminal Not1, CCR4-Not complex component Not1	0.00
cpbgf_6002070	Clathrin/coatomer adaptor N-terminal	NA
cpbgf_6002080	Kem1p-like 5'-3' exonuclease	NA
cpbgf_6002100	Uncharacterized protein	2.00
cpbgf_6002140	Potassium channel domain containing protein	1.00
cpbgf_6002210	Uncharacterized transmembrane Protein	NA

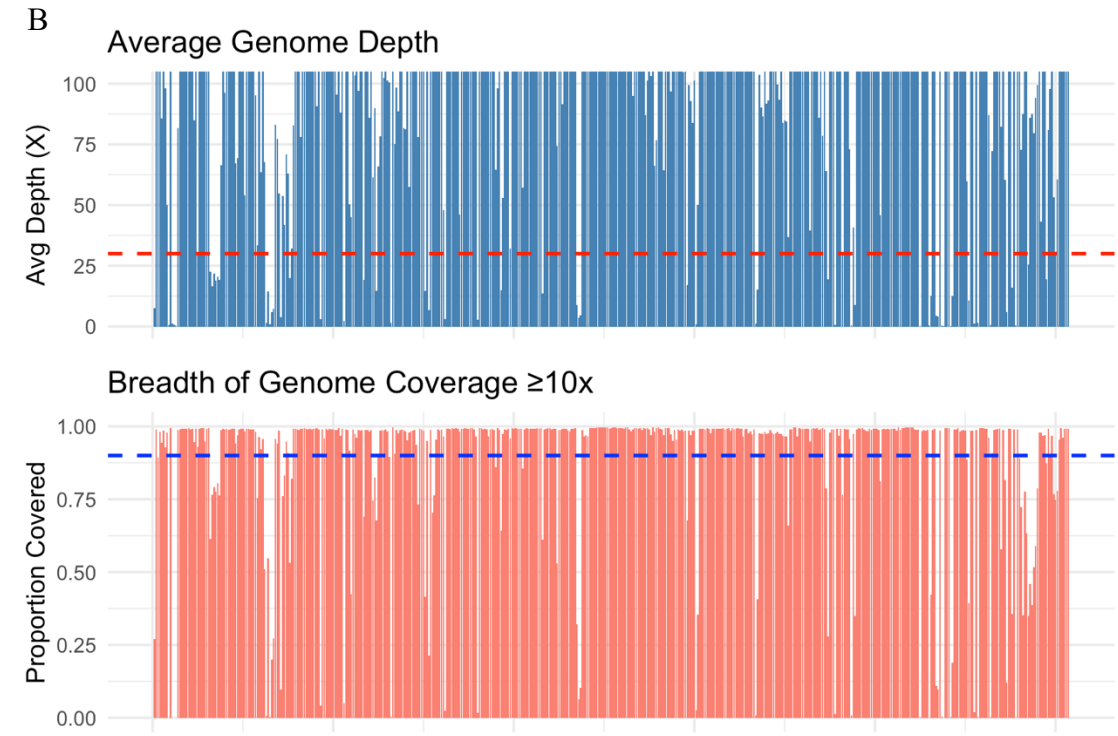
cpbgf_6002240	Zinc finger, RING/FYVE/PHD-type domain containg protein	0.00
cpbgf_6002250	Myb-like DNA-binding domain containing protein	NA
cpbgf_6002260	Hca4p helicase DBP4 (Helicase CA4). EIF4A-1-family RNA SFII helicase	NA
cpbgf_6002270	Uncharacterized protein	0.50
cpbgf_6002300	Anaphase-promoting complex subunit 1	NA
cpbgf_6002310	Signal Peptide Uncharacterized transmembrane Protein	NA
cpbgf_6002320	Bromodomain/HAS/DEXDc/SNF2 domain containing protein	0.00
cpbgf_6002810	RNA polymerase I specific transcription initiation factor RRN3	0.00
cpbgf_6002830	Pre-mRNA-splicing factor SF3a complex subunit 2 (Prp11)	NA
cpbgf_6002840	Sucrose-phosphatase-like protein /HAD superfamily hydrolase	NA
cpbgf_6002860	Uncharacterized protein	3.00
cpbgf_6003040	RNA-polymerase II-associated protein 3-like protein	0.75
cpbgf_6003050	Uncharacterized protein	6.00
cpbgf_6003080	Uncharacterized protein	NA
cpbgf_600830	Pleckstrin homology domain containing protein	0.83
cpbgf_600850	Thioredoxin	0.00
cpbgf_600860	Uncharacterized protein	NA
cpbgf_600880	Gdb1p glycogen debranching enzyme	NA
cpbgf_600900	Peptidase C78, ubiquitin fold modifier-specific peptidase 1/ 2	0.50

Figures

A


Sample Count Bin
<1
1–10
11–19
20–29
30–100
>100


0%
35%
44%
5%
16%
EUROPE
AFRICA
ASIA
USA
OTHER



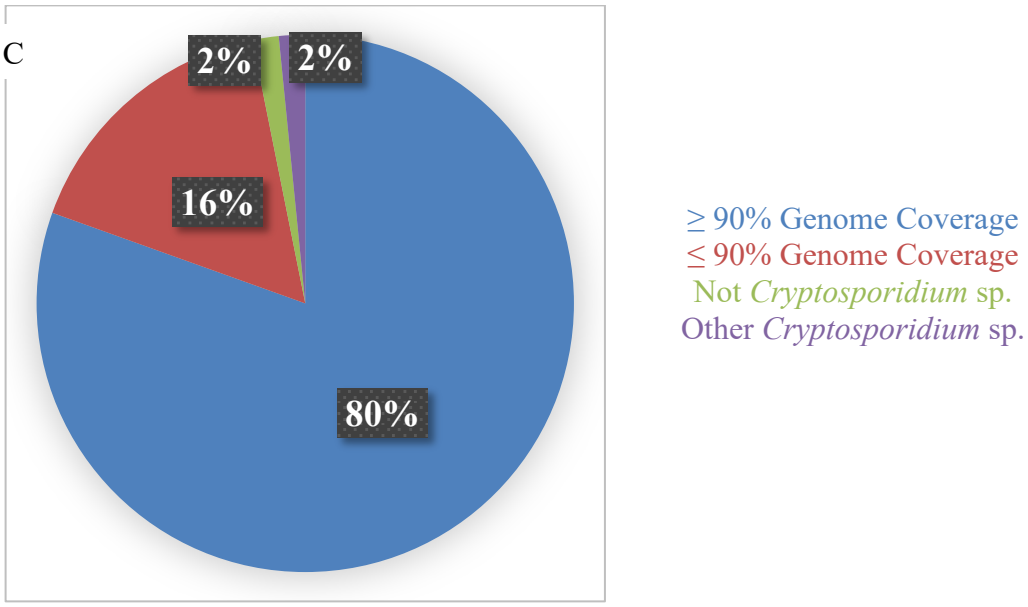


Figure 1: Parsing of *Cryptosporidium parvum* samples from NCBI. A) The geographical distribution of the SRA sample and the proportion from each location. B) The depth of coverage (blue) with red line indicating depth at 30x and the breadth of coverage (orange) at a depth of

10x with blue line indicating 90% genome coverage. C) the proportion of SRA sample left after filtering out possible contaminants, mixed populations and samples with less than 90% genome coverage at 10x.

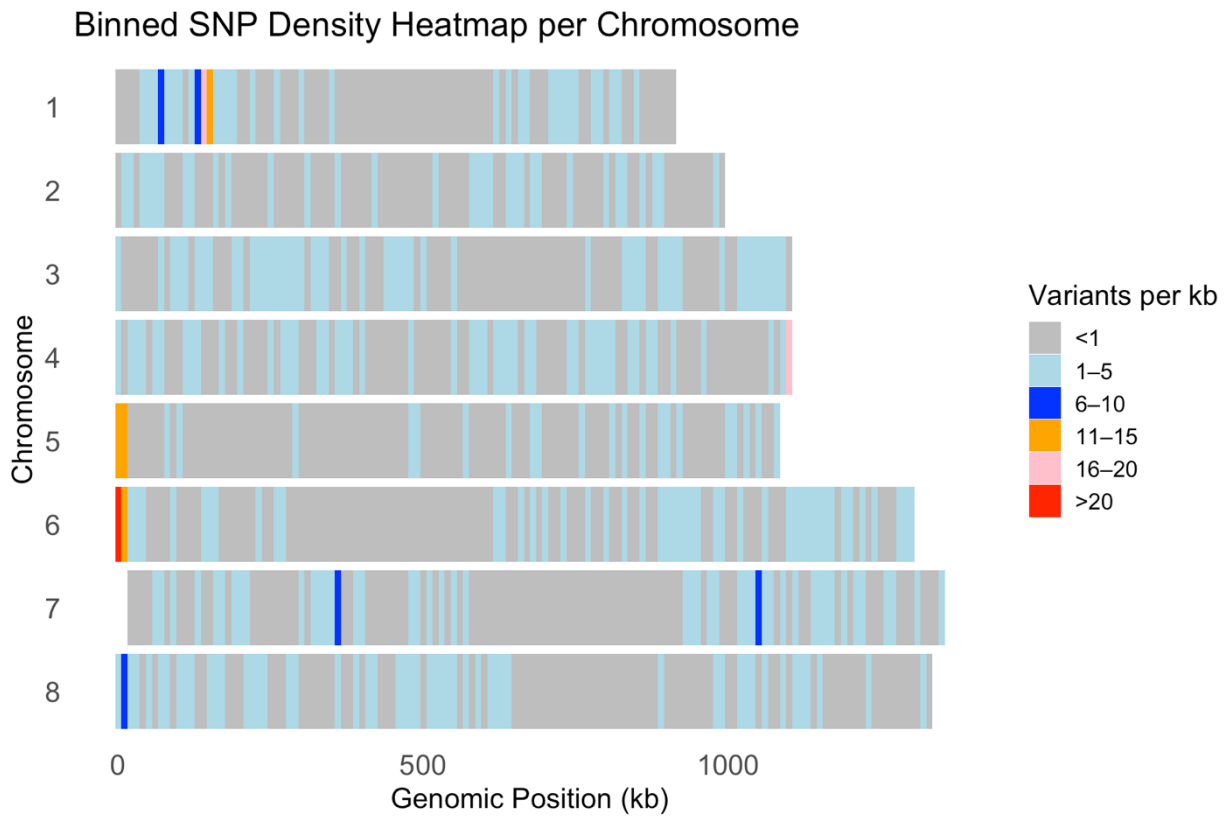
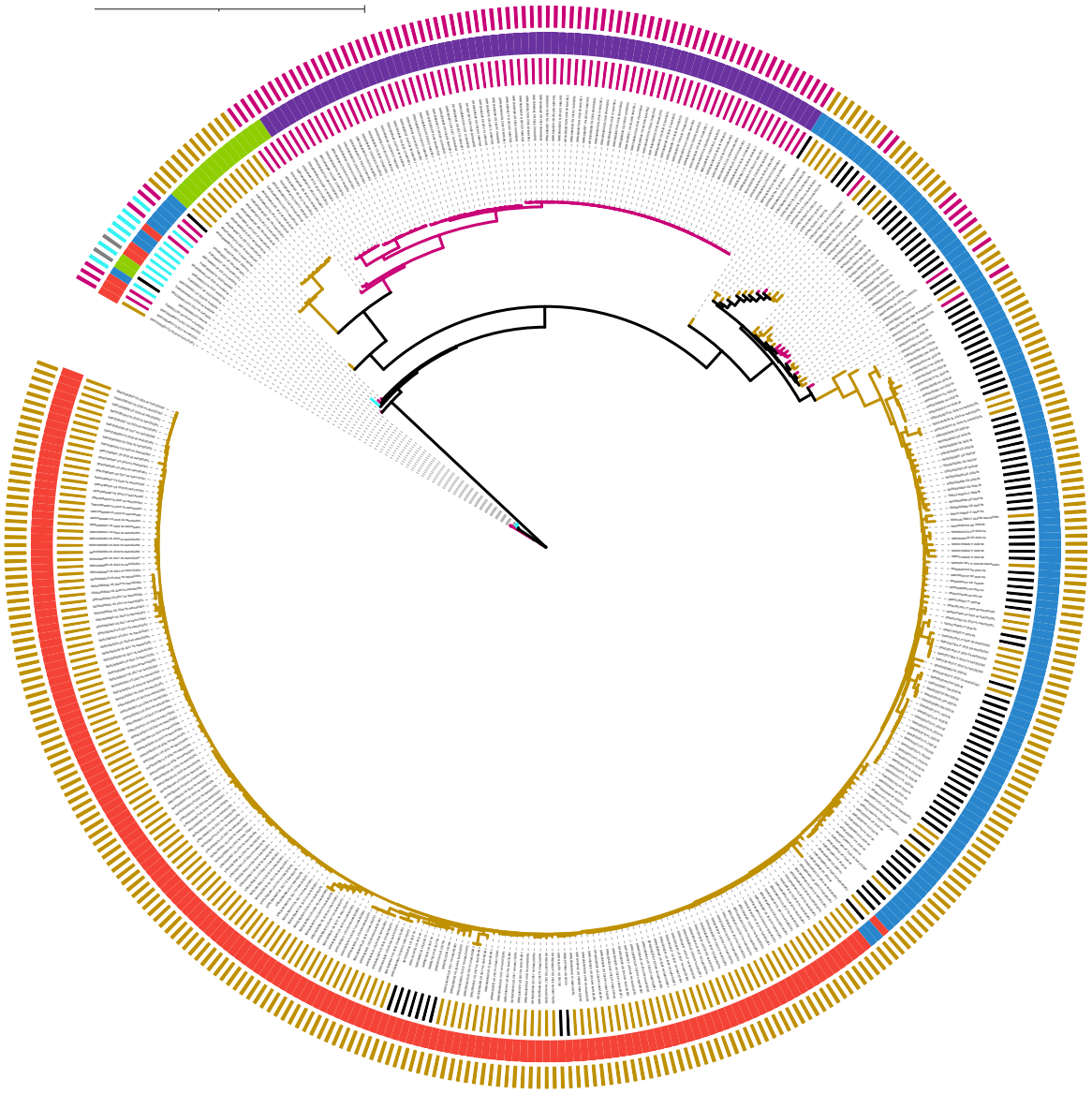


Figure 2: A heatmap of SNP density per chromosome. SNP density calculated across the genome in non-overlapping 10 kb windows. Each region is colored based on the number of SNPs at that location, with different colours indicating the SNP count in each region.

A


Tree scale 1
GP60 SUBTYPE
IIa gp60
IIc gp60
IId gp60
IIh gp60
GEOGRAPHICAL LOCATION
Africa
China
Europe
USA

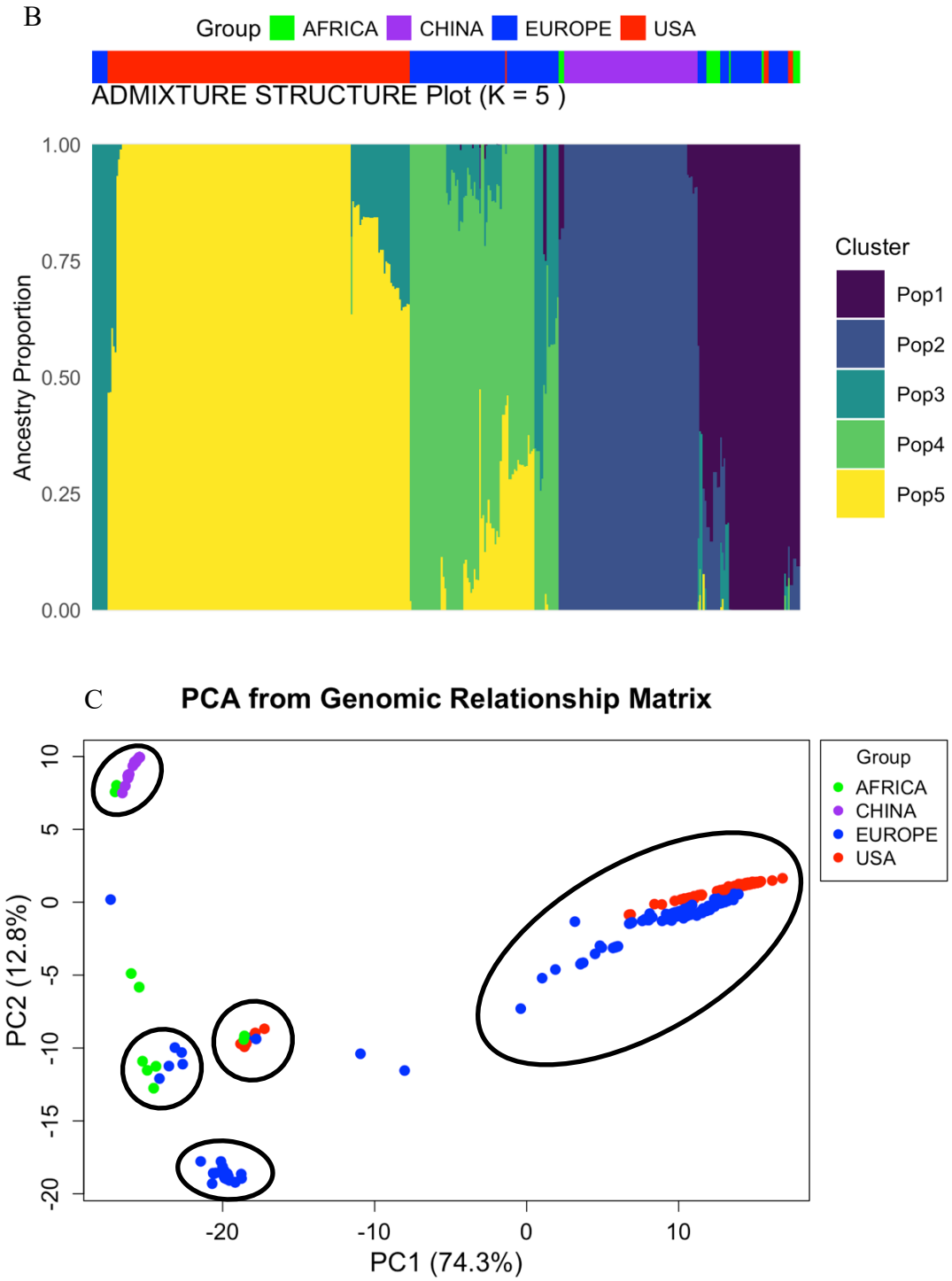


Figure 3: Global population structure of *Cryptosporidium parvum* illustrating the clustering of samples. A) A maximum likelihood-based tree. The colored inner layer represents *gp60* from metadata, the middle layer is the reported geographical origin of the sample, and the outer layer is the *gp60* according to our analysis. B) An admixture plot showing inferred proportions of

shared ancestry across the samples and the colors indicate ancestral components. The group bar on top indicates the geographical origin of the samples. C) A PCA plot showing clustering of samples. Points represent samples and colors depict the geographical origin of the samples.

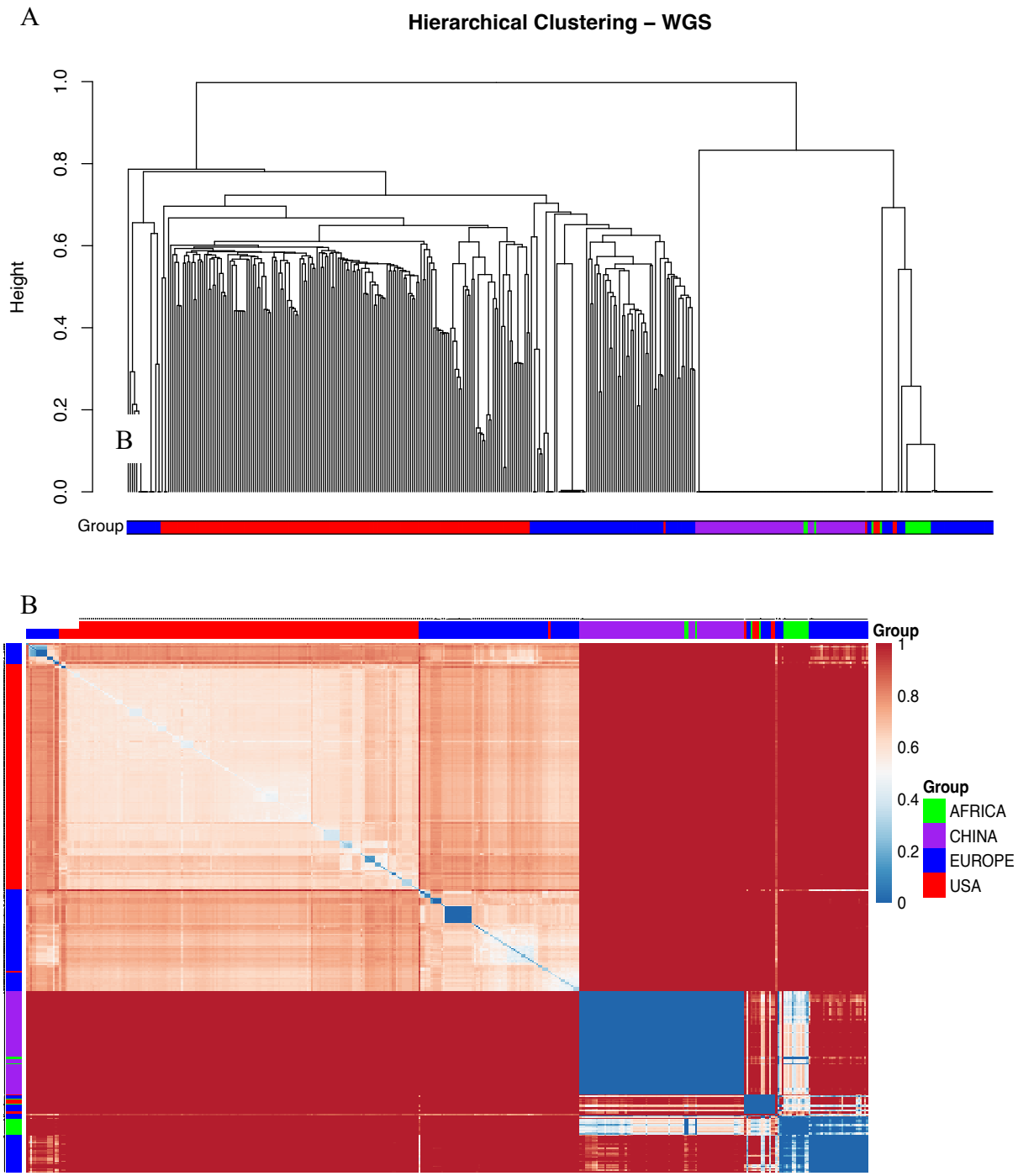
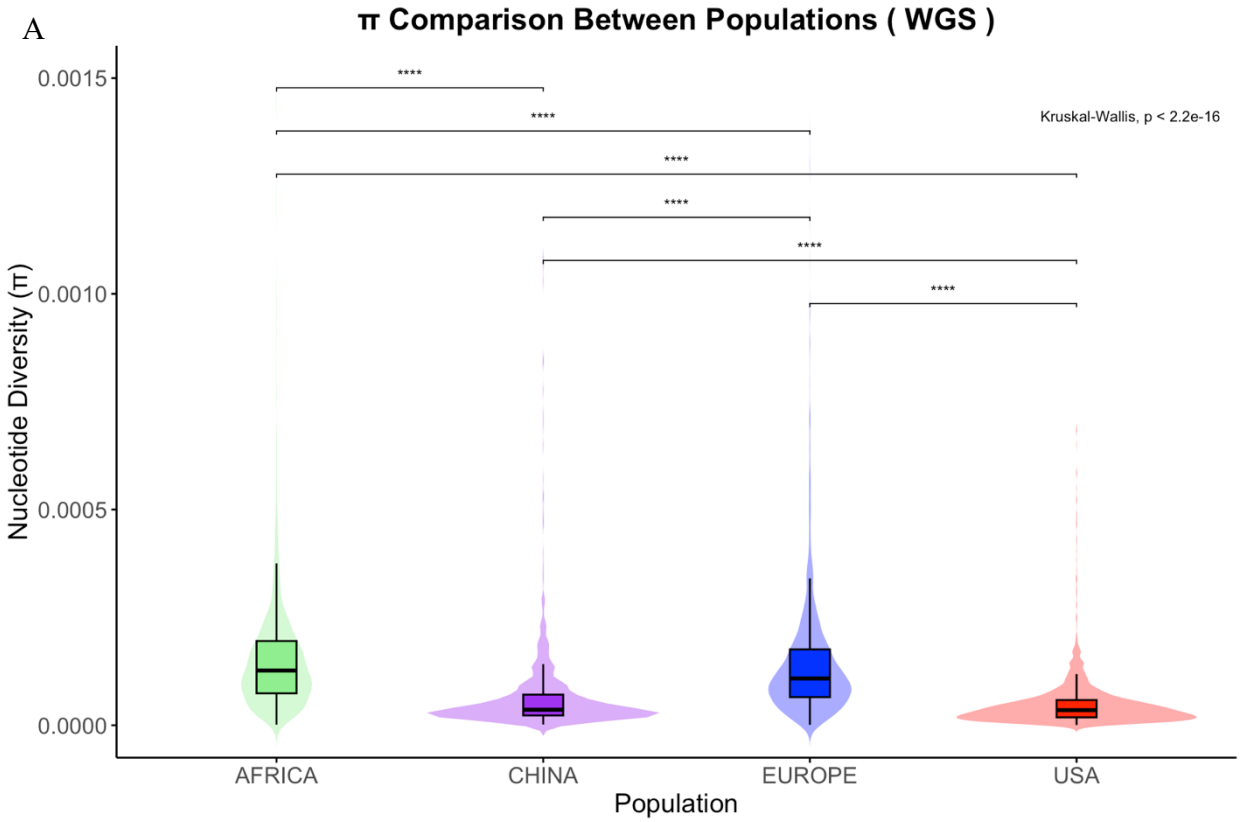


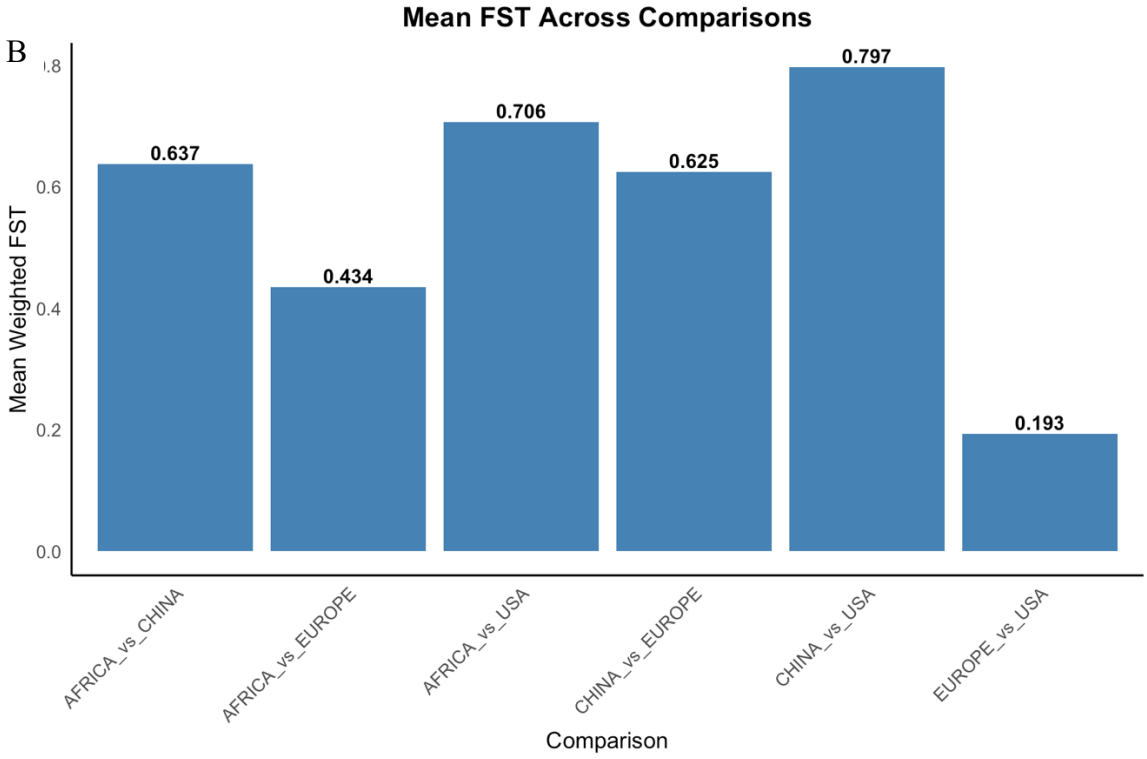
Figure 4. A pairwise comparison of *Cryptosporidium parvum* samples from around the world. A) Hierarchical clustering of samples based on pairwise distances illustrating the genetic relationships and population structure. Populations are color-coded as in the legend. B) A heat

map showing pairwise similarity between samples. The color scale depicts similarity, and the group represents the geographical origin of each sample.

A


π Comparison Between Populations (WGS)
Kruskal-Wallis, p < 2.2e-16
****
****
****
****
****
****
0.0015
0.0010
0.0005
0.0000
Nucleotide Diversity (π)
AFRICA
CHINA
EUROPE
USA
Population

B


Mean FST Across Comparisons
0.637
0.434
0.706
0.625
0.797
0.193
0.6
0.4
0.2
0.0
Mean Weighted FST
AFRICA_vs_CHINA
AFRICA_vs_EUROPE
AFRICA_vs_USA
CHINA_vs_EUROPE
CHINA_vs_USA
EUROPE_vs_USA
Comparison

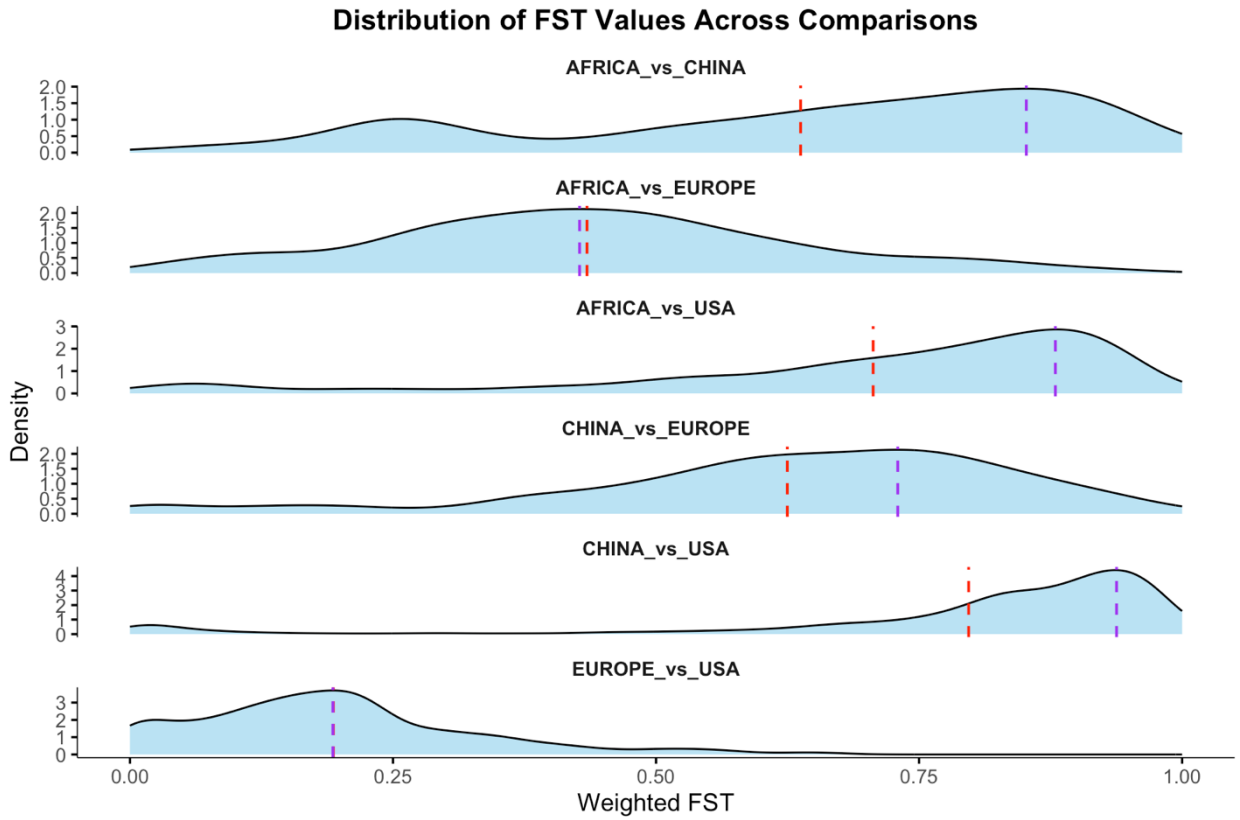
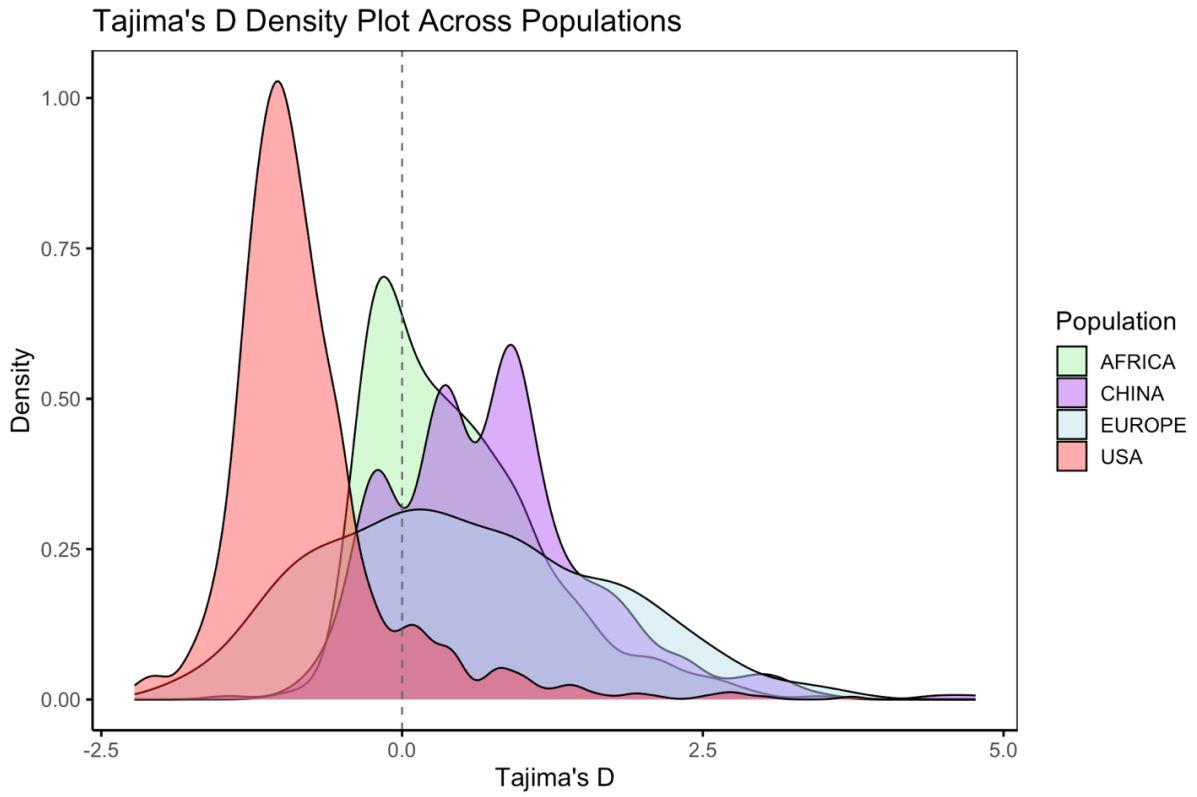


Figure 5: Diversity within and between populations. A) A combined box and violin plot showing the distribution of nucleotide diversity within a population. The violin plot illustrates the density of nucleotide diversity across the genome, while the box plot highlights the range and outliers. B) A summary of the differentiation between the populations. The Top panel shows the overall (mean) divergence between the populations. The bottom panel shows a density plot of FST values per comparison with the red dashed line indicating the mean FST, while the purple dashed line marks the FST mode.

A


Tajima's D Density Plot Across Populations
Density
1.00
0.75
0.50
0.25
0.00
-2.5
0.0
2.5
5.0
Tajima's D
Population
AFRICA
CHINA
EUROPE
USA

B

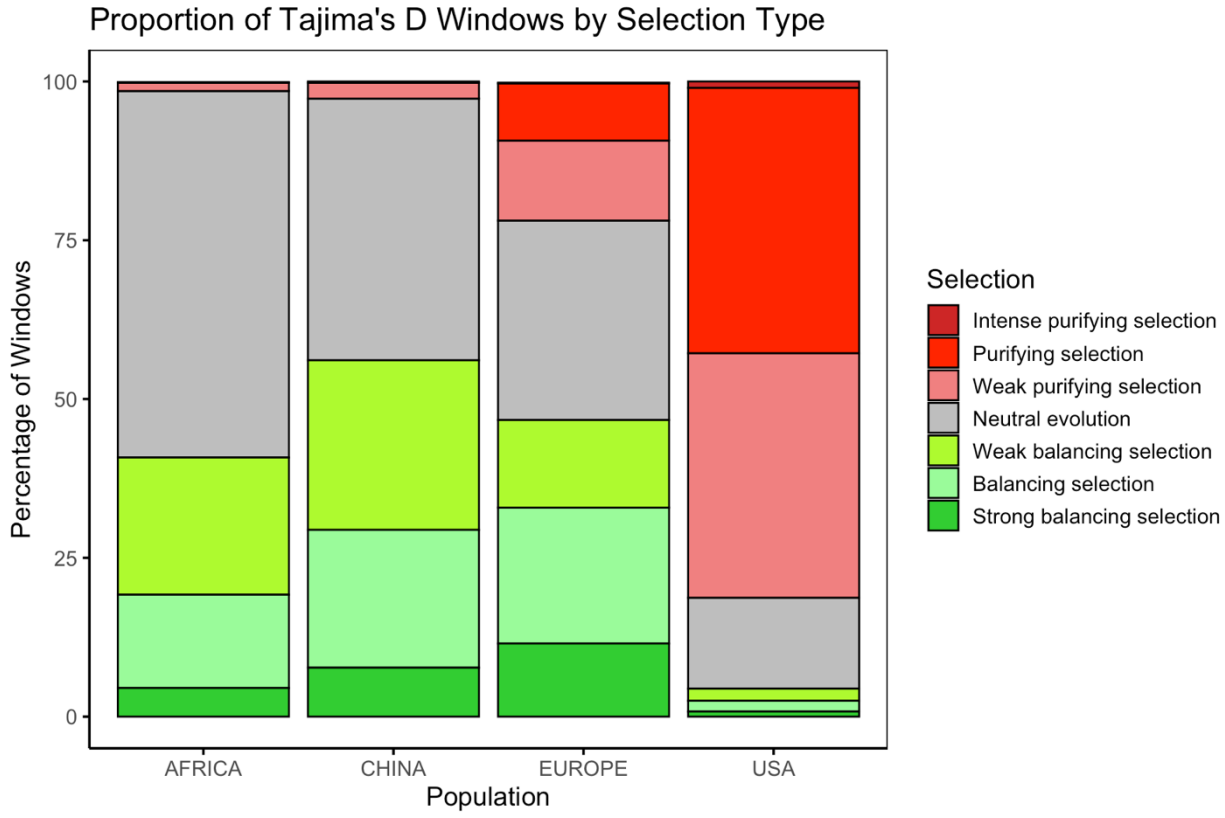
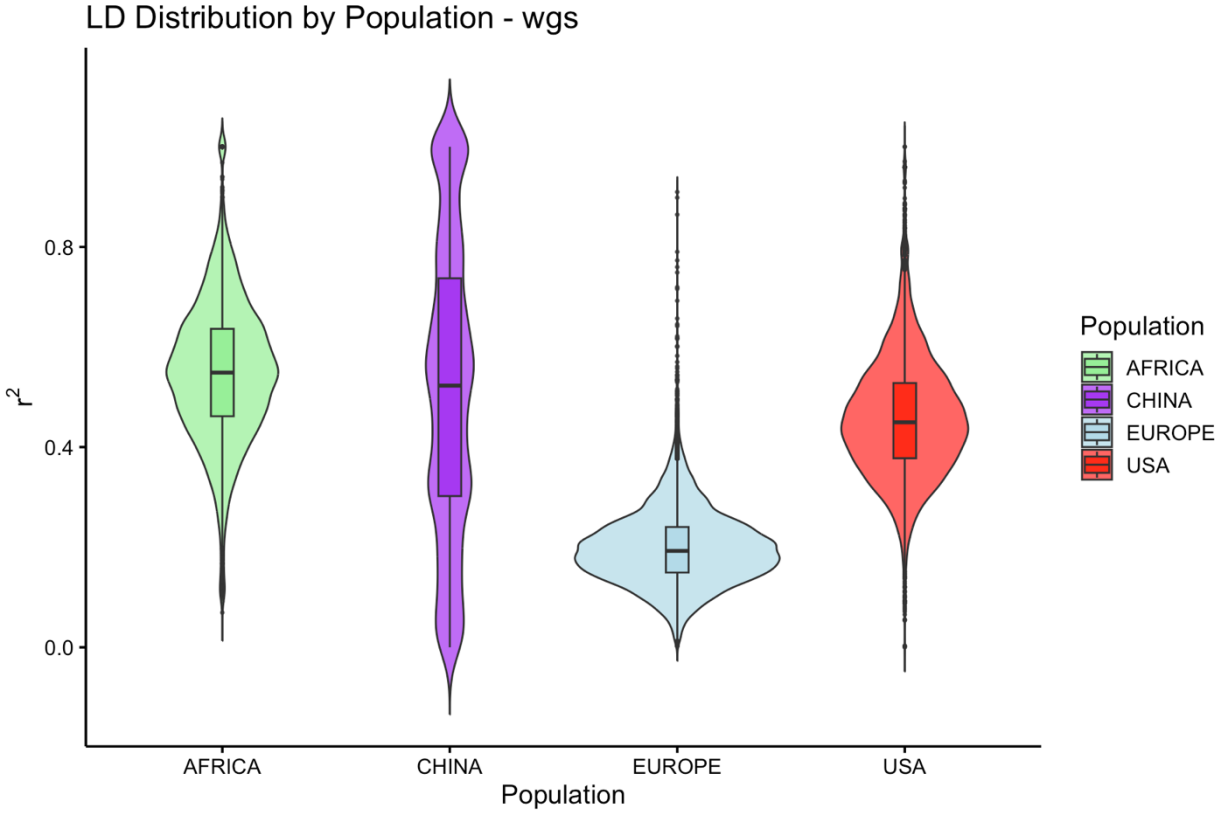

Proportion of Tajima's D Windows by Selection Type
Percentage of Windows
100
75
50
25
0
AFRICA
CHINA
EUROPE
USA
Population
Selection
Intense purifying selection
Purifying selection
Weak purifying selection
Neutral evolution
Weak balancing selection
Balancing selection
Strong balancing selection

Figure 6: Selection pressures within the various populations. A) The curve represents the density of Tajima's D values within a population. The dashed line at the zero mark on the x-axis signifies neutral evolution, with values below and above indicating signs of purifying and balancing selection, respectivelyor non equilibrium demography. B) A stacked bar plot showing the proportion of the 10kb windowed genome under different selection pressures within the population. Selection pressures were inferred based on Tajima's D thresholds ranging from intense purifying selection (D<-2) to Strong balancing selection (D>2).

A


LD Distribution by Population - wgs
0.8
0.4
0.0
r2
AFRICA
CHINA
EUROPE
USA
Population
Population
AFRICA
CHINA
EUROPE
USA

B

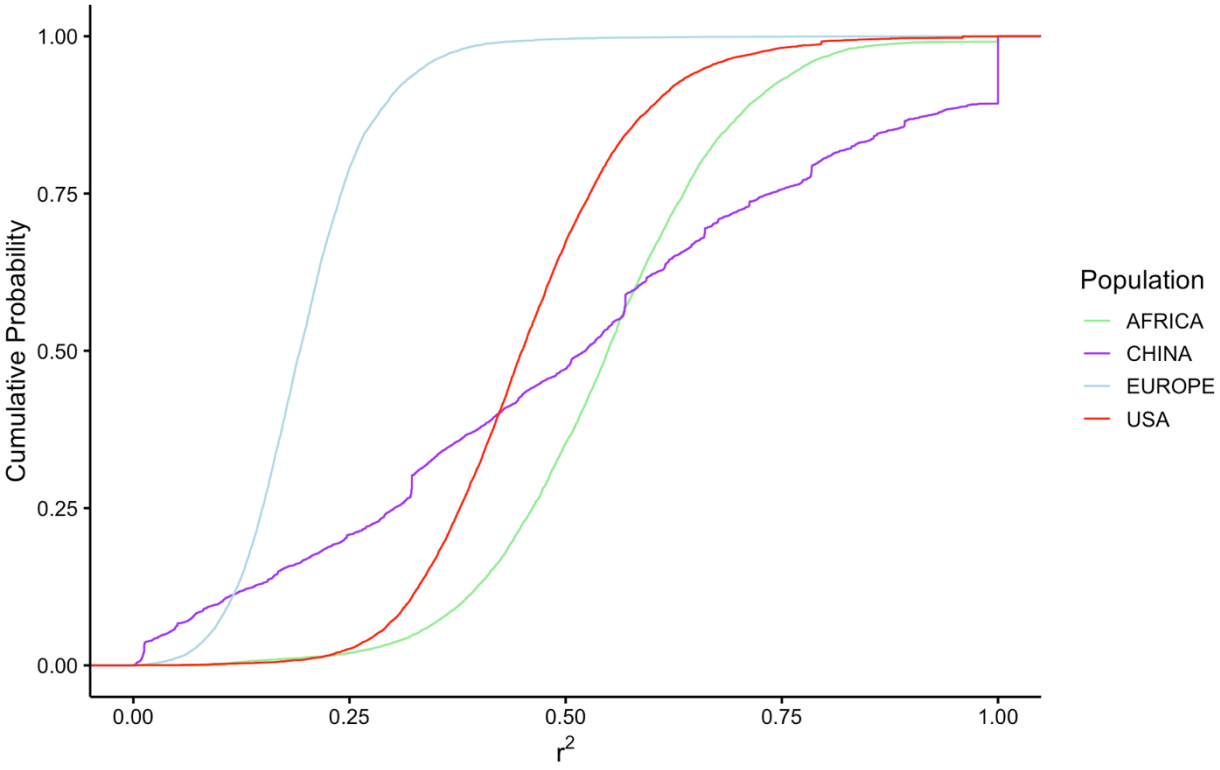
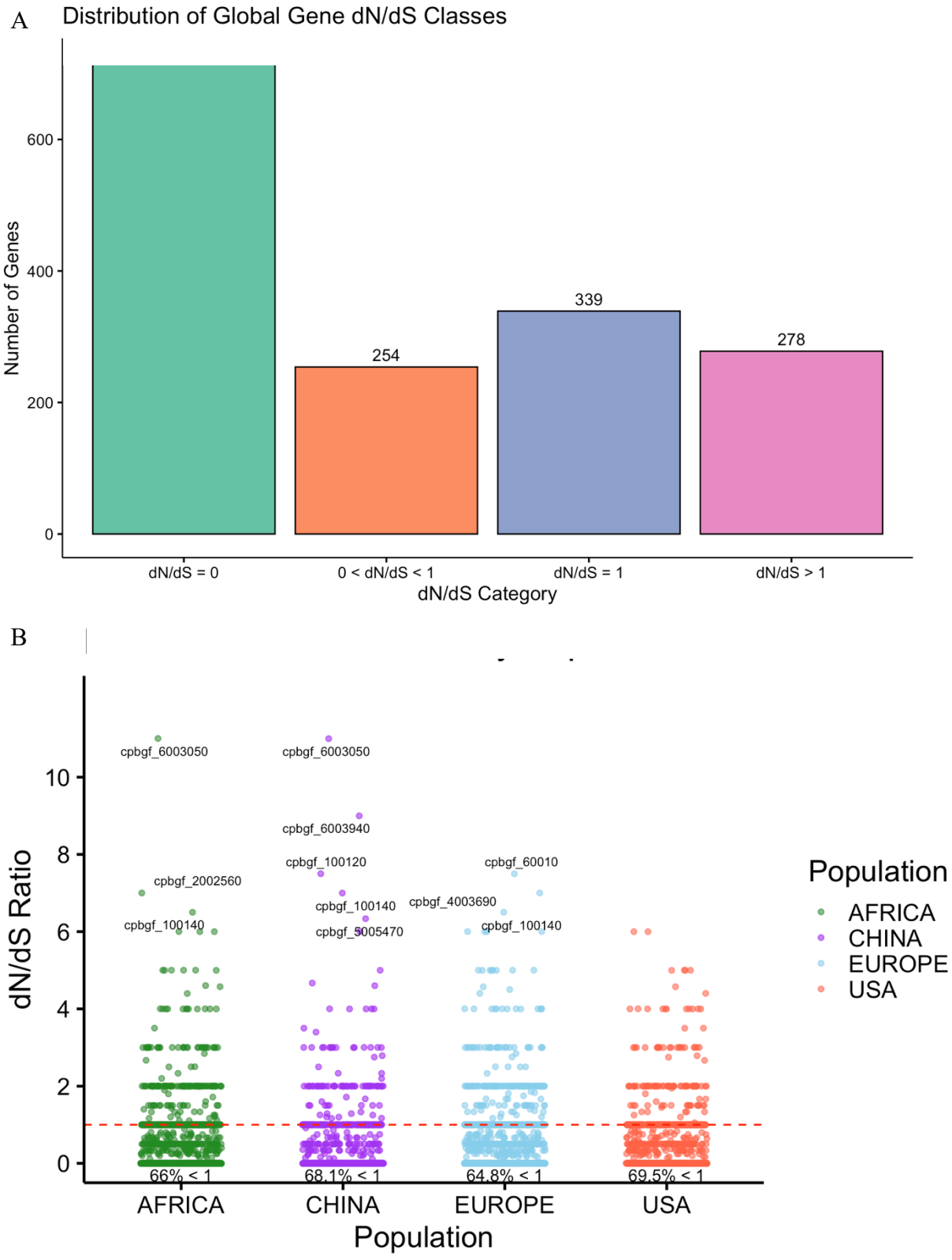
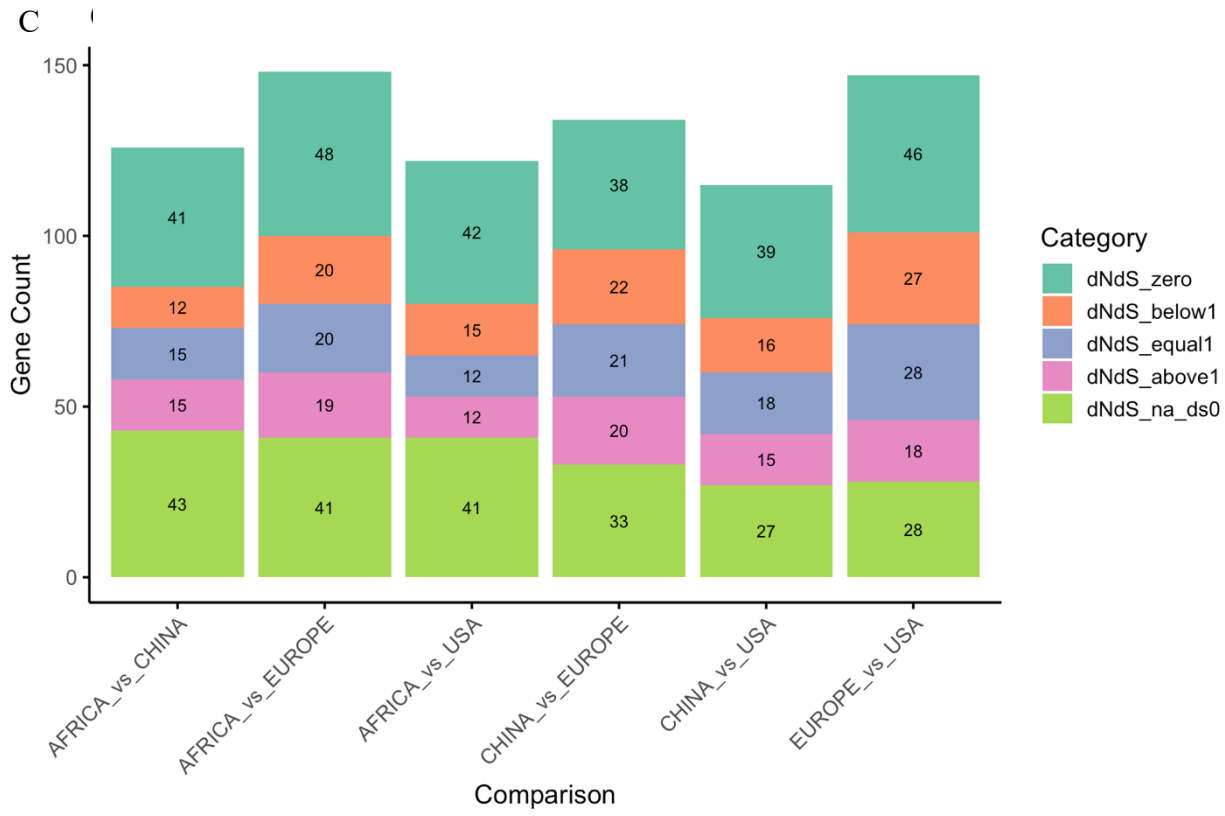

1.00
0.75
0.50
0.25
0.00
Cumulative Probability
0.00
0.25
0.50
0.75
1.00
r2
Population
AFRICA
CHINA
EUROPE
USA

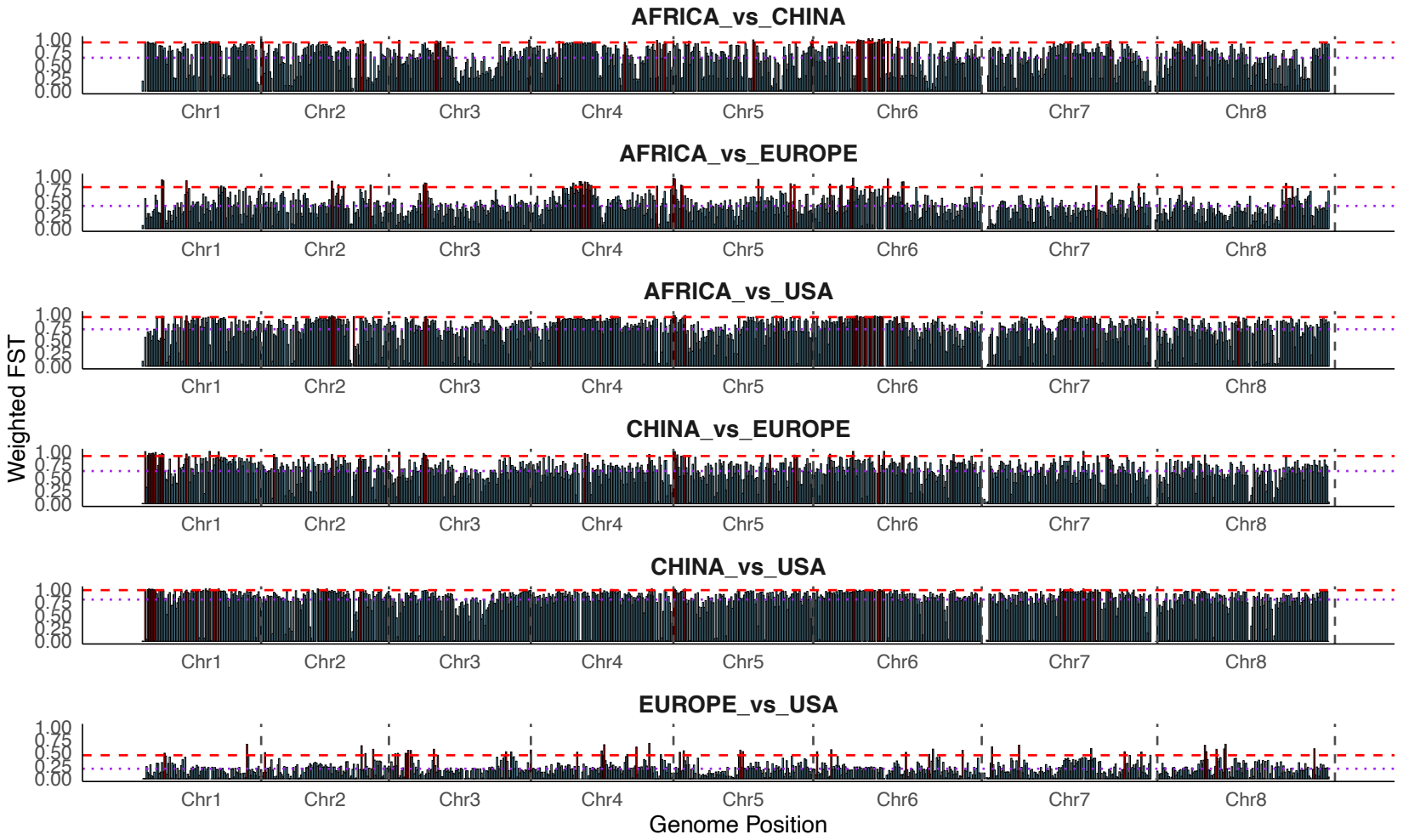
Figure 7: Linkage Disequilibrium coefficient r^2 as an indicator of association between SNPs. A) A violin-box plot showing the distribution and median of r^2 values across the genome with a non-overlapping 10kb window. The violin plot illustrates the density of r^2 values, while the box plot shows central tendencies and outliers reflecting variation in LD structure across the populations. B) a Cumulative Distribution Function (CDF) of r^2 showing the distribution of recombination across the genome. CDF shows the proportion of SNPs below a given r^2 threshold, providing insight into the decay of Linkage Disequilibrium.


A
Distribution of Global Gene dN/dS Classes
Number of Genes
600
400
200
0
254
339
278
dN/dS = 0
0 < dN/dS < 1
dN/dS = 1
dN/dS > 1
dN/dS Category
B
dN/dS Ratio
10
8
6
4
2
0
cpbgf_6003050
cpbgf_6003050
cpbgf_6003940
cpbgf_100120
cpbgf_60010
cpbgf_2002560
cpbgf_100140
cpbgf_4003690
cpbgf_100140
cpbgf_100140
cpbgf_5005470
66% < 1
68.1% < 1
64.8% < 1
69.5% < 1
AFRICA
CHINA
EUROPE
USA
Population
Population
AFRICA
CHINA
EUROPE
USA

C


150
100
50
0
Gene Count
41
12
15
15
43
48
20
20
19
41
42
15
12
12
41
38
22
21
20
33
39
16
18
15
27
46
27
28
18
28
AFRICA_vs_CHINA
AFRICA_vs_EUROPE
AFRICA_vs_USA
CHINA_vs_EUROPE
CHINA_vs_USA
EUROPE_vs_USA
Comparison
Category
dNdS_zero
dNdS_below1
dNdS_equal1
dNdS_above1
dNdS_na_ds0

D


AFRICA_vs_CHINA
AFRICA_vs_EUROPE
AFRICA_vs_USA
CHINA_vs_EUROPE
CHINA_vs_USA
EUROPE_vs_USA
Weighted FST
1.00
0.75
0.50
0.25
0.00
Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Genome Position

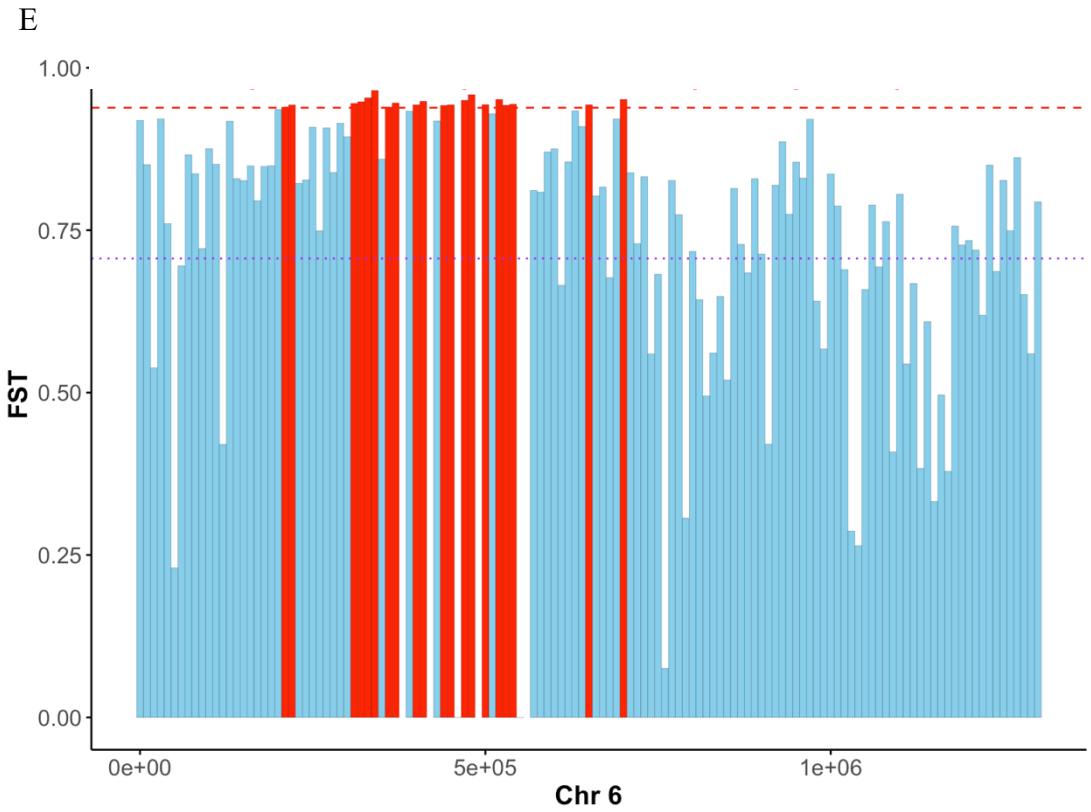


Figure 8: Distribution of genes and their potential association with population differentiation. A) A bar plot of the global distribution of genes characterized by pN/pS ratios, highlighting the prevalence of purifying, neutral, and positive selection. B) A scatter plot displaying per-gene pN/pS ratios across populations, highlighting genes under strong positive selection (pN/pS ratio > 6) and the proportion of genes with a pN/pS ratio < 1 (red dashed line) for each population. C) Genome-wide distribution of pN/pS ratios of genes located within the top 5% FST outlier regions, showing patterns of selection across the genome. D & E) Bar plots showing elevated regions of genetic differentiation (top 5% FST) and their overlap with genes of various pN/pS ratios at the genomic level (D) and on chromosome 6 (E). The red dashed line indicates the 95th percentile threshold, and the purple dotted line represents the mean FST of the population. Each bar represents and 10kb window with the red bars signifying regions with the top 5% Fst.

CHAPTER 5

CONCLUSION

In the race to prevent the transmission of infectious diseases and to discover therapeutics to treat these diseases, researchers from various disciplines, such as molecular biology, genomics, epidemiology, and bioinformatics, are collaborating toward this end goal. To be successful in this endeavor, we must fully understand not just the anatomy and physiology of the parasite but also its genomic makeup. *Cryptosporidium*, an apicomplexan parasite, is responsible for causing cryptosporidiosis, which is characterized by severe diarrhea. Its impact is felt globally as it causes significant harm to its victims and, in some cases, leads to death. With its anthroponotic and zoonotic nature, a holistic approach is needed to mitigate the spread of the disease. A reliable approach is to characterize the parasite at the genomic level, as that will enable the identification of critical genes involved in virulence, drug resistance, and host adaptation, paving the way for effective interventions and treatments.

In Chapter 1, I discuss the key challenges, recent breakthroughs, and future directions in *Cryptosporidium* genomics and transcriptomics, which are critical for understanding its biology, evolution, and epidemiology. Although 74 *Cryptosporidium* genome assemblies exist in NCBI (January 2024), over 50% are from *C. parvum* and *C. hominis*, highlighting limited representation. To overcome sample purity issues, hybrid capture enables enrichment of parasite DNA from fecal samples, while single-oocyst sequencing allows high-resolution analysis of intra-host diversity and recombination from minimal input. Comparative genomics reveals that *C. parvum*, *C. tyzzeri*, *C. hominis*, and *C. cuniculus* share ~96% genome identity and synteny, despite having different hosts, suggesting host specificity may be driven by subtle differences in

subtelomeric regions with variable gene families like MEDLE. Population genomics has uncovered recombination, admixture, and novel subclades with improved assemblies like the *C. parvum* telomere-to-telomere (T2T) assembly, improving variant detection. Transcriptomic data, though still limited, have advanced our understanding of gene regulation and host-parasite interactions. Tools such as CRISPR, long-read RNA-seq, and transgenics now enable functional studies, supporting efforts in surveillance, diagnostics, and outbreak response. However, with all these advances, a broader sampling and more complete reference genomes remain essential.

With knowledge of the current state of genomic research in *Cryptosporidium*, Chapter 2 evaluates the effectiveness of Multiple Displacement Amplification (MDA) combined with long-read Oxford Nanopore Technologies (ONT) sequencing under conditions of limited DNA input. This approach is particularly critical for *Cryptosporidium*, as ONT library preparation typically requires ~1 µg of input DNA, which far exceeds what is obtainable from most samples. In the case of single-oocyst sequencing, the challenge is even greater, with each oocyst containing only ~40 fg of DNA, which falls well below the minimum input required even for low-input Illumina short-read protocols (typically 1–10 ng). Although MDA introduces uneven amplification and variable genome coverage, it successfully amplified DNA inputs as low as 2.5×10^{-5} ng. With 0.025 ng of DNA, a complete genome coverage was achieved after sequencing. Also, the introduced chimeras can be removed using CADECT (developed by Baptista R.) before downstream analysis. The findings from this study are especially important because they demonstrate that MDA can be effectively used with DNA quantities similar to that of a single *Cryptosporidium* oocyst, which contains about 1.5 times more DNA than the minimum amount required for complete genome recovery.

In Chapter 3, I elaborate on my contribution to the collaborative study "Genetic crosses within and between species of *Cryptosporidium* – PMID 38147547." In this study, a dual drug-selectable marker system (phreSR and neo) was developed and used in the identification of hybrid progeny of *C. parvum* and *C. tyzzeri*. These hybrids contain both markers, indicating the occurrence of recombination. My role involved implementing MDA-based single-oocyst sequencing to validate recombination genome-wide, beyond the marker loci. This approach enabled direct confirmation of genetic crosses by sequencing individual oocysts and identifying specific crossover regions at the nucleotide level. There was evidence that some chromosomes of the recombinant progeny mostly came from a single parent, and in some cases, it showed a mosaic pattern from both parents. Importantly, successful amplification from low-input DNA highlights MDA and single oocyst sequencing as a robust method for capturing recombination. These findings confirm the possibility of interspecies recombination and underscore the permeability of species boundaries, highlighting *Cryptosporidium*'s capacity for genetic adaptation. The combination of dual-selection markers and single-oocyst sequencing provides a powerful platform for investigating traits shaped by recombination, including virulence, host specificity, and immune evasion.

Following the validation of MDA and the use of single-oocyst sequencing to detect recombination in *Cryptosporidium*, in Chapter 4, I shift my focus to the global diversity of *C. parvum*. In this study, I analyzed SNPs from 408 *C. parvum* SRA datasets from four continents. A critical limitation was the skewed sampling, with ~80% of the 507 total datasets originating from Europe and the USA, despite higher disease incidence in Africa and Asia. Analysis of high-quality biallelic SNPs revealed distinct population structures shaped by geography, separating isolates into two main clusters: Africa-China and Europe-USA. China samples formed a highly

clonal group dominated by subtype IId, likely reflecting recent clonal expansion. African samples contained two divergent clades, with evidence of possible gene flow from China. USA samples, primarily subtype IIa, clustered closely with Europe, suggesting a recent introduction and limited diversification. European populations were the most diverse and recombining. Diversification and selection analyses (pN/pS, Tajima's D, LD, and Fst) revealed population-specific evolutionary pressures, including balancing selection in China and Africa, and directional selection in the USA. Shared and population-specific positively selected genes, many linked to virulence and host interaction, were identified, particularly on chromosome 6. These findings not only enhance our understanding of *C. parvum* evolution but also provide a valuable genomic framework for the broader field, informing molecular surveillance, control strategies, functional research on host adaptation and pathogenicity, and creating a pipeline for investigating *Cryptosporidium* diversity on a broader scale.

Future steps will include stratifying the dataset by host (e.g., human, bovine) and by demography (geographic location and year) to separate host-driven adaptation from geographic or temporal structure. Despite the haploid nature of *Cryptosporidium* sporozoites, variants will be called in diploid mode to capture within-sample mixtures arising from multiple sporozoites per oocyst and occasional co-infection. Diploid calling preserves within-sample heterogeneity and yields allele fractions based on AD/DP. When needed, a strict haploid consensus can be derived by masking mixed sites or selecting the majority allele. Within each stratum, we will apply consistent analyses: PCA and ADMIXTURE to assess clustering, and windowed FST to measure differentiation, while including the other factor as a covariate in association tests. The host-focused analysis will identify loci and haplotypes enriched in specific hosts, compare recombination and linkage disequilibrium between host groups, and nominate a compact marker

set for surveillance. The demographic-focused analysis will map structure across regions and years, estimate gene flow, and pair FST with allele-frequency–based scans to distinguish neutral demography from selection, refining clusters and highlighting regions under local adaptation. These analyses will yield host- and region-specific summaries, updated genome-wide F_ST profiles, and a prioritized list of candidate genes for functional validation and surveillance.

APPENDIX A: SUPPLEMENTAL MATERIAL OF CHAPTER 2

Supplemental Tables

Supplemental Table 1. Whole genome assembly statistics for the data generated using MDA.

Assembly	Condition	Total length (nt)	# of contigs	Largest contig (nt)	GC%	N50	L50	# N's per 100 Kbp
C. meleagridis	MDA (5 ng)	9,171,013	13	1,363,785	30.92	1,103,979	4	0
E. coli	MDA-Intact DNA	6,702,699	252	479,759	50.49	157,769	12	0
	MDA-Fragmented DNA	157,593	12	25,757	49.23	20,338	4	0
E.faecium	400ng (no MDA)	2,775,595	2	2,583,377	38.27	2,583,377	1	0
	1 ng	3,603,364	216	257,348	38.3	64,352	14	0
	1E-1ng	4,287,796	570	83,451	38.66	27,070	49	0
	1E-2 ng	4,146,213	1,066	65,055	38.94	13,948	95	0
	1E-3 ng	1,174	2	670	53.15	670	1	0
	1E-4 ng	78,463	53	15,260	53.83	2,204	7	0
	1E-5 ng	54,253	32	6,962	56.49	3,816	6	0
S. aureus	Intact DNA	2,766,204	3	2,717,982	32.86	2,717,982	1	0
	Fragmented DNA	854,501	198	33,734	33.62	4,926	36	0

Intact DNA post-MDA	2,763,611	4	2,717,354	32.86	2,717,354	1	0
Fragmented DNA post-MDA	2,842,696	20	1,890,364	32.84	1,890,364	1	0

Supplemental Table 2. Sequencing Statistics for *E. faecium** WGS results at 2.5 ng and 0.2.5 ng starting DNA input.

Starting total DNA	**2.5 ng**	**0.25 ng**
Depth	73×	35×
Total length (bp)	2,778,112	2,918,507
Number of contigs	3	35
GC%	38.2	38.28
N50	2,589,111	233,792
L50	1	4
Number of N's per 100 Kbp	0	0

*The expected genome size for *E. faecium* species varies between 2.6 and 3.2 Mb

Supplemental Table 3. Read length distribution among the DNA dilutions for *E. faecium,* based on starting DNA concentration.

Starting DNA Amount (ng)	Mean Read Length (bp)	Median Read Length (bp)
Control 400	6,134	3,819
2.5	2,533	1,308
2.5E-01	2,305	1,002
2.5E-02	2,390	1,147
2.5E-03	3,121	1,086
2.5E-04	2,555	1,185
2.5E-05	2,641	957

Supplemental Table 4. Comparison between *S. aureus* ATCC-29213 genome sequences assembled before and after CADECT with assembly length, number of contigs, number of reads, and mean read length.

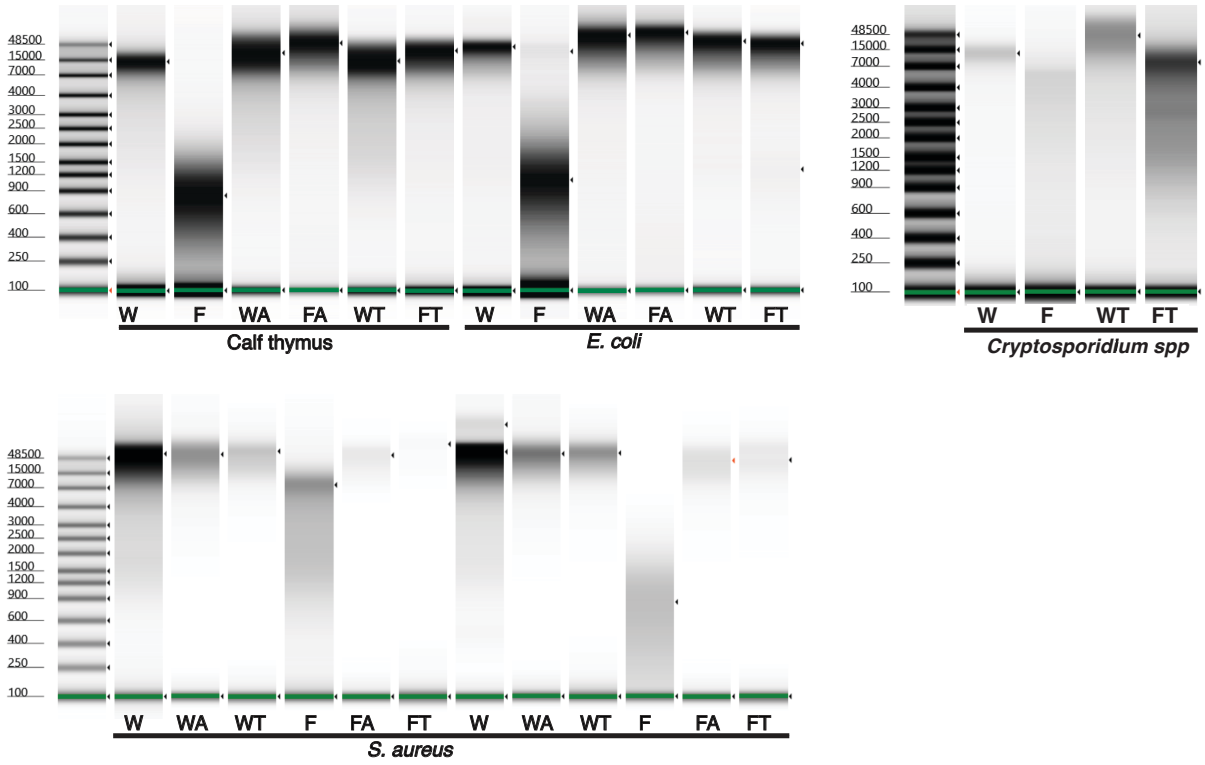
	Fragmented DNA with MDA before CADECT	**Fragmented DNA with MDA after CADECT**
Assembly length (bp)	2,842,696	2,752,482
Number of contigs	20	3
Number of reads	324,197	292,789
Mean read length (bp)	2,163	1,534
Median read length (bp)	1,352	1,202

Supplemental Table 5. Comparison between *S. aureus* results from CADECT using low DNA input samples.

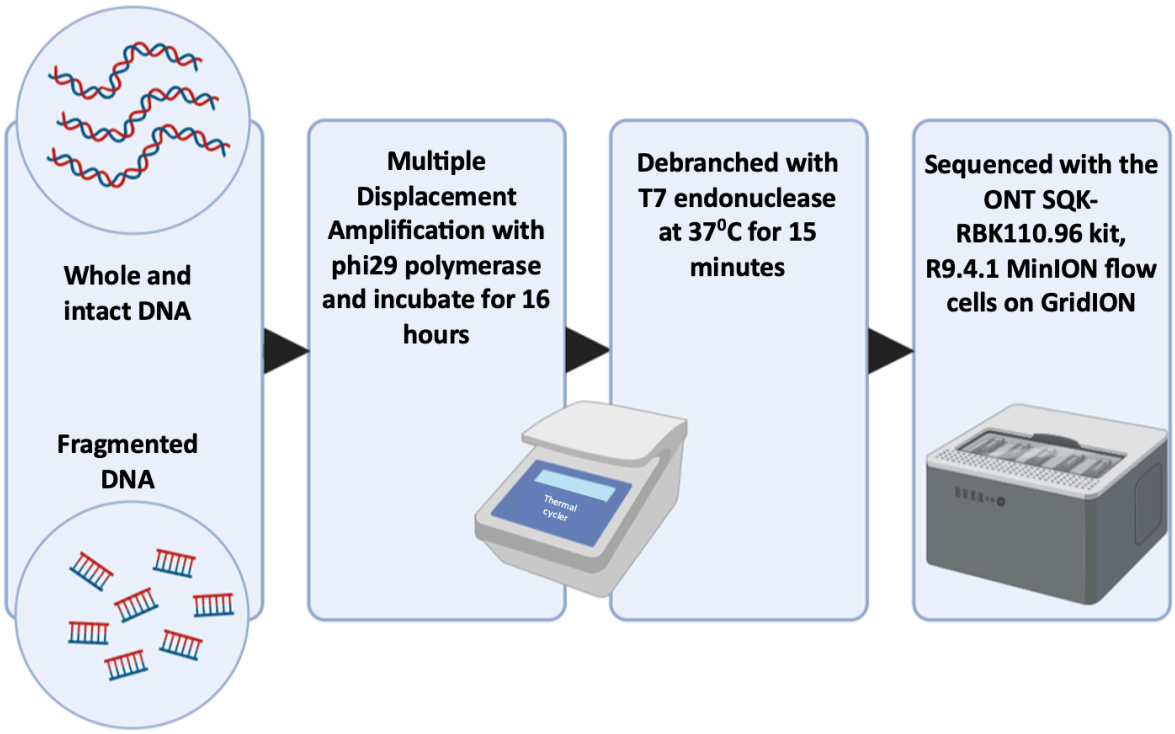
DNA input Condition	**Intact DNA**		**Fragmented DNA**	
MDA	No	yes	No	yes
Sequencing coverage input	46.8	92.3	21.7****	77.4
Total number of sequenced base pairs	131,012,134	258,412,289	60,893,145	216,785,548
Number of shorter reads detected*	6,397	49,746	109,830	54,268
Number of non-concatemer reads detected	12,556	44,440	2,548	35,986
Number of putative concatemer reads detected	1,047	5,814	26**	9,746**
Total number of Reads analyzed:	20,000	100,000	112,404	100,000
read loss (%)	37	56	98	64
Total number of non-Concatemer base pairs	103,146,375	153,374,280	5,304,400	96,993,865
Coverage loss (x)	10.0	37.5	19.9	42.8

*Shorter reads were reads detected below the default setting of 500 nt window size
**Due to size selection putative concatemers were classified as short reads
***Loss if using just the reads characterized as non-concatemeric
****Without the amplification ONT had a bad throughput for the fragmented samples at low input values to generate longer reads, resulting in a low sequencing coverage.

Supplemental figures



Supplemental Figure 1. Uncropped DNA Fragment and read Size Range pre- and post- Multiple Displacement Amplification using Size-Controlled Fragmented DNA. Uncropped TapeStation results for different organism sets; (W = whole intact DNA; F = Fragmented DNA; WA and FA = after amplification; and WT and FT = after T7 debranching). Two samples of *S. aureus* were tested and we kept both results info in the uncropped version. FA and WA for *Cryptosporidium* were not measured at the time that the experiment was done.

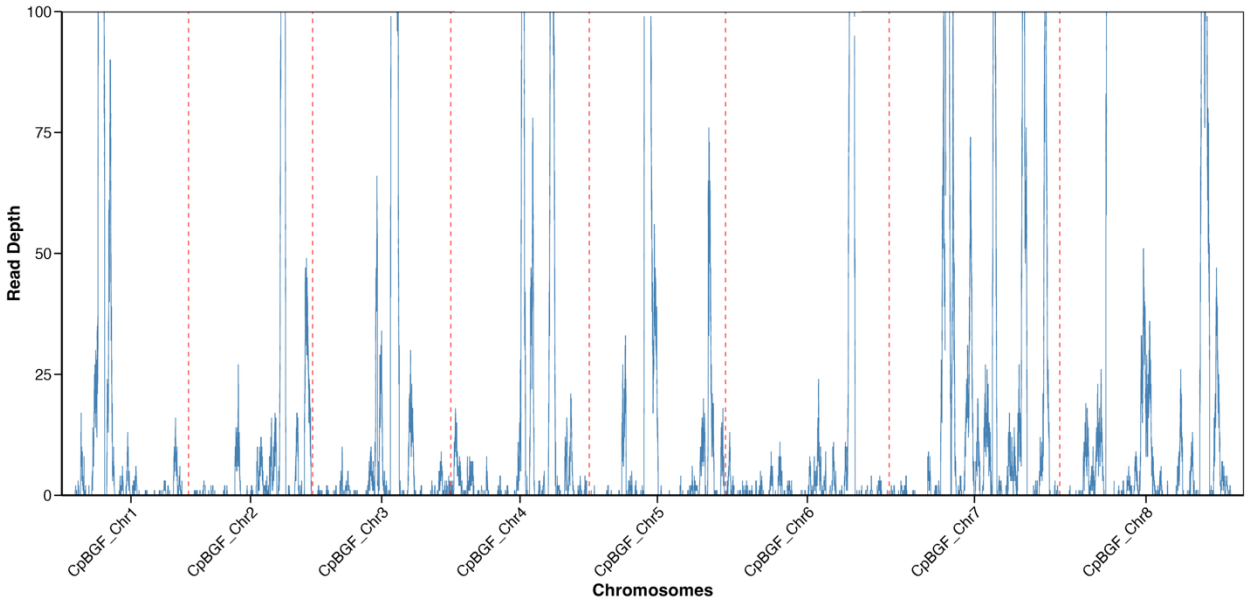


Supplemental Figure 2. Summarized Diagram of the **MDA protocol.** DNA from both long and short DNA fragment samples were amplified separately following the MDA manufacturer's protocol. T7 endonuclease was used to facilitate cleavage at points of mismatches. Sequencing was performed on the ONT GridION.

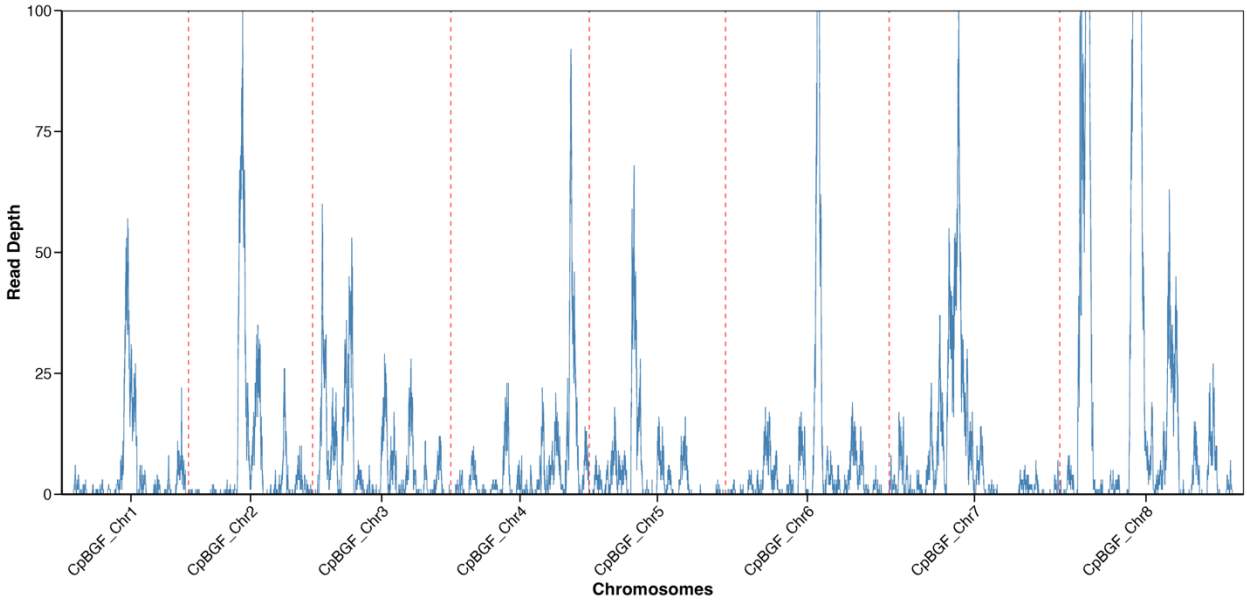
APPENDIX B: SUPPLEMENTAL MATERIAL OF CHAPTER 3

Supplemental Figures

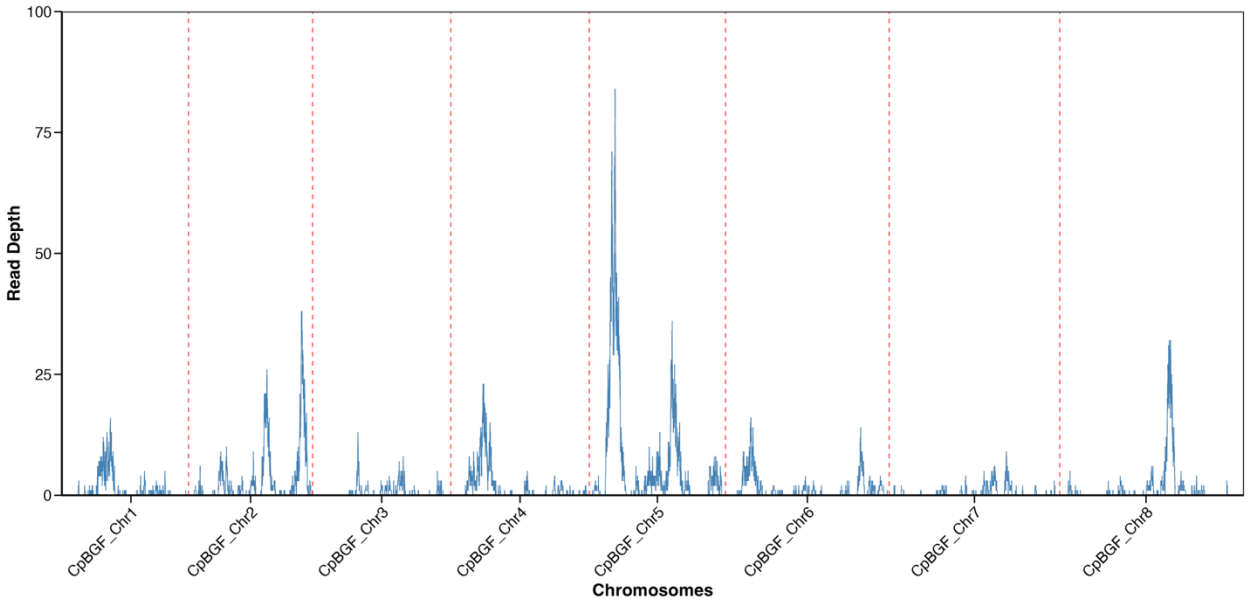
Progeny_1a



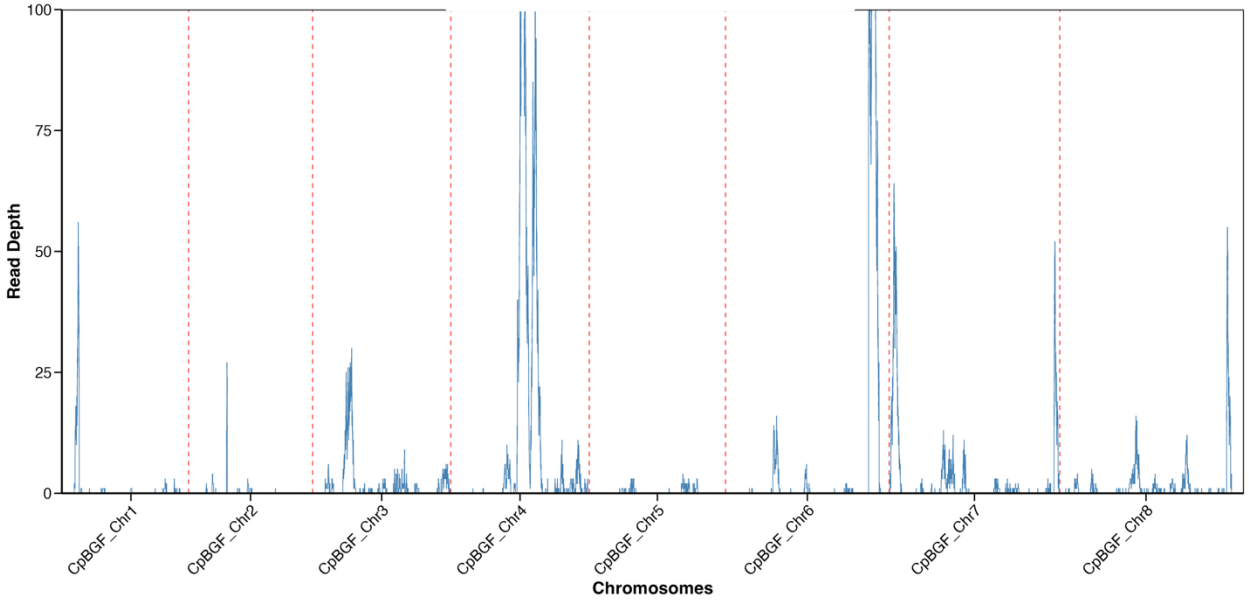
Progeny_2a



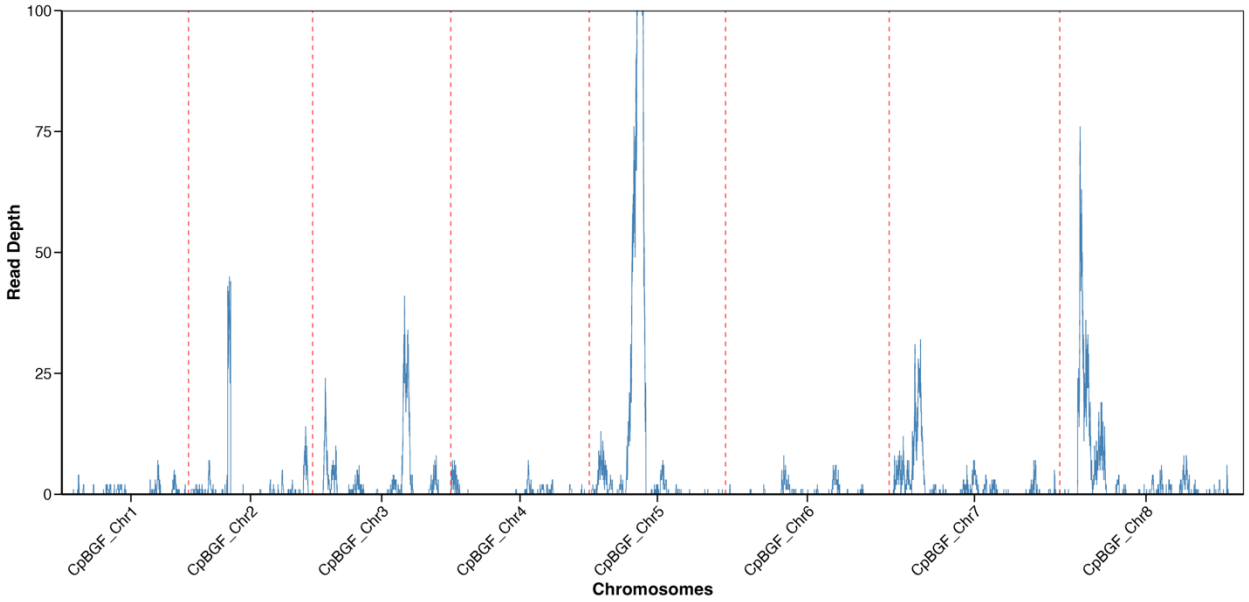
Progeny_2b



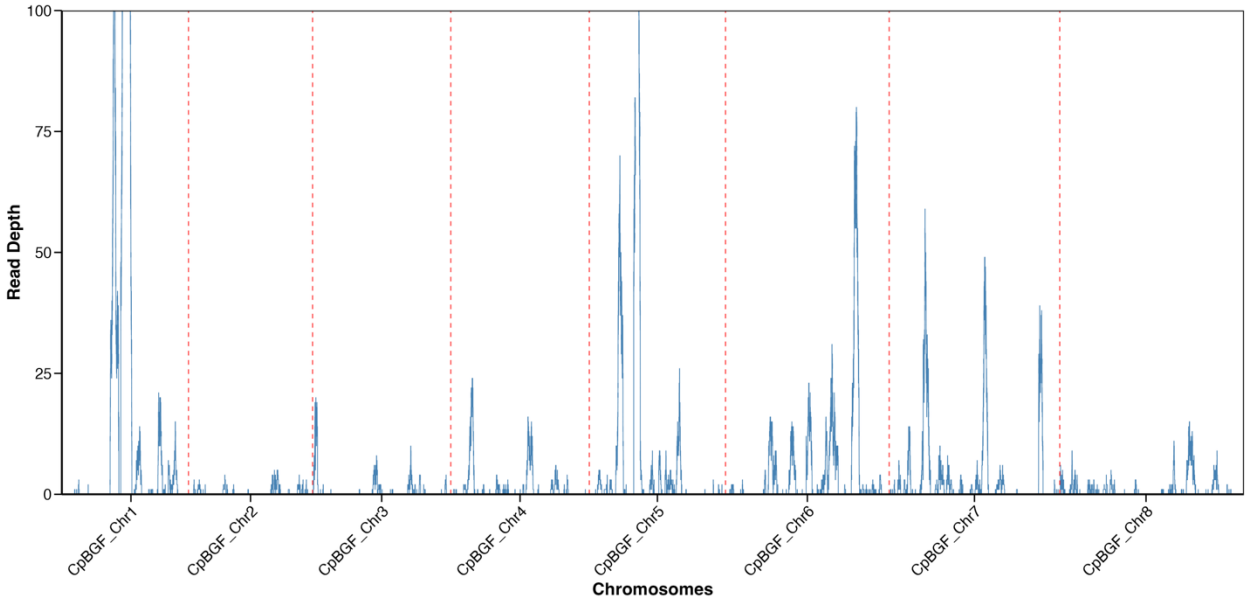
Progeny_3b

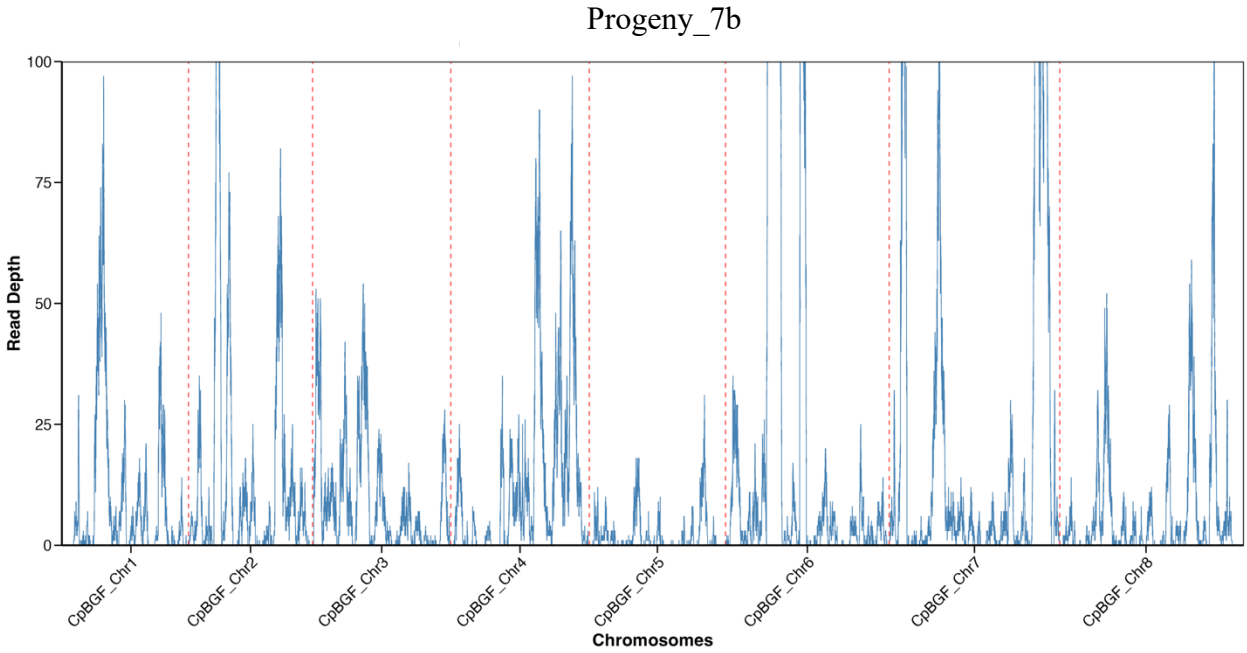


Progeny_4b



Progeny_6b





Supplementary Figure 1. Multiple Displacement Amplification (MDA) leads to random and uneven DNA amplification throughout the genome. The genome coverage plots for individual single-oocyst sequences display variable amplification patterns, highlighting the stochastic nature of MDA.

APPENDIX C: SUPPLEMENTAL MATERIAL OF CHAPTER 4

Supplemental Tables

Supplemental Table 1. Summary of sequencing metrics for all C. parvum SRA datasets obtained from NCBI. The table includes read parameters, quality scores, GC content, and duplicate read status.

File	Num_seqs	Sum_len	Min_len	Avg_len	N50	Q20(%)	Q30(%)	GC(%)	Duplicate
SRR10336414_1.fastq	14,307	3,591,057	251	251	251	94.91	92.47	32.3	No
SRR10336414_2.fastq	14,307	3,591,057	251	251	251	80.6	75.41	32.49	No
SRR10336415_1.fastq	14,560	3,654,560	251	251	251	92.36	88.3	32.5	No
SRR10336415_2.fastq	14,560	3,654,560	251	251	251	48.01	38.79	36.56	No
SRR10336411_1.fastq	18,324	4,599,324	251	251	251	95.95	94.07	31.24	No
SRR10336411_2.fastq	18,324	4,599,324	251	251	251	85.56	81.52	31.43	No
SRR10336413_1.fastq	26,969	6,769,219	251	251	251	96.3	94.52	30.77	No
SRR10336413_2.fastq	26,969	6,769,219	251	251	251	88.04	84.35	30.78	No
SRR10363371_1.fastq	385,018	79,194,890	35	205.7	250	95.07	92.64	50.91	No
SRR10363371_2.fastq	385,018	80,838,545	35	210	250	76.63	67.72	52.48	No
SRR10363325_1.fastq	400,395	100,098,750	250	250	250	96.52	95.42	31.63	No
SRR10363325_2.fastq	400,395	100,098,750	250	250	250	90.08	87.58	31.69	No
SRR10363327_1.fastq	416,624	104,156,000	250	250	250	96.71	95.67	32.15	No
SRR10363327_2.fastq	416,624	104,156,000	250	250	250	90.98	88.65	32.18	No
SRR10363323_1.fastq	426,308	106,577,000	250	250	250	96.62	95.46	32.94	No
SRR10363323_2.fastq	426,308	106,577,000	250	250	250	89.99	87.33	32.99	No
SRR10363326_1.fastq	434,744	108,686,000	250	250	250	96.88	95.86	31.35	No
SRR10363326_2.fastq	434,744	108,686,000	250	250	250	90.74	88.32	31.39	No
SRR10363324_1.fastq	466,285	116,571,250	250	250	250	96.1	94.98	32.51	No
SRR10363324_2.fastq	466,285	116,571,250	250	250	250	89.18	86.65	32.7	No
SRR10997286_1.fastq	498,065	65,246,515	131	131	131	97.2	92.93	40.34	No
SRR10997286_2.fastq	498,065	71,223,295	143	143	143	88.58	76.73	40.85	No

SRR3091756_1.fastq	568,701	170,610,300	300	300	300	86.15	76.09	34	No
SRR3091756_2.fastq	568,701	170,610,300	300	300	300	89.9	85.48	34.29	No
SRR10997324_1.fastq	595,002	77,945,262	131	131	131	96.97	92.53	38.23	No
SRR10997324_2.fastq	595,002	85,085,286	143	143	143	88.21	76.2	38.82	No
SRR10363362_1.fastq	1,051,060	212,113,200	35	201.8	250	95.38	93.84	36.23	No
SRR10363362_2.fastq	1,051,060	213,065,911	35	202.7	250	89.63	86.29	36.64	No
SRR10363552_1.fastq	1,143,017	286,897,267	251	251	251	96.06	93.51	30.27	No
SRR10363552_2.fastq	1,143,017	286,897,267	251	251	251	91.19	86.07	30.23	No
SRR10363386_1.fastq	1,188,327	268,646,219	35	226.1	251	94.43	92.14	40.43	No
SRR10363386_2.fastq	1,188,327	269,159,363	35	226.5	251	89.95	86.14	40.74	No
SRR10363373_1.fastq	1,205,725	233,938,569	35	194	250	89.34	85.34	47.3	No
SRR10363373_2.fastq	1,205,725	239,276,246	35	198.5	250	74.53	66.35	48.91	No
SRR6147587_1.fastq	1,321,400	185,250,097	35	140.2	151	99.17	97.3	32.4	No
SRR6147587_2.fastq	1,321,400	185,138,823	35	140.1	151	97.91	94.75	32.38	No
SRR10997270_1.fastq	1,342,155	175,822,305	131	131	131	96.46	91.74	37.98	No
SRR10997270_2.fastq	1,342,155	191,928,165	143	143	143	87.77	75.66	38.63	No
SRR10336407_1.fastq	1,392,370	348,092,500	250	250	250	90.68	89.2	33.98	No
SRR10336407_2.fastq	1,392,370	348,092,500	250	250	250	84.38	82.04	33.79	No
SRR10363539_1.fastq	1,413,878	354,883,378	251	251	251	96.4	94.06	31.79	No
SRR10363539_2.fastq	1,413,878	354,883,378	251	251	251	91.9	87.11	31.83	No
SRR10363381_1.fastq	1,427,579	322,642,673	35	226	250	94.7	92.27	44.76	No
SRR10363381_2.fastq	1,427,579	324,420,802	35	227.3	250	84.21	78.23	45.23	No
SRR10363460_1.fastq	1,522,994	382,271,494	251	251	251	95.76	93.85	32.77	No
SRR10363460_2.fastq	1,522,994	382,271,494	251	251	251	89.92	86.1	32.84	No
SRR3091752_1.fastq	1,533,000	459,900,000	300	300	300	93.78	88.71	33.12	No
SRR3091752_2.fastq	1,533,000	459,900,000	300	300	300	88.07	80.2	33.11	No
SRR10363376_1.fastq	1,541,886	360,369,525	35	233.7	251	94.27	91.45	44.93	No
SRR10363376_2.fastq	1,541,886	360,988,623	35	234.1	251	86.78	81.16	45.28	No
SRR10363432_1.fastq	1,550,293	389,123,543	251	251	251	96.46	93.98	38.93	No

SRR10363432_2.fastq	1,550,293	389,123,543	251	251	251	90.94	85.45	39.03	No
SRR6813717_1.fastq	1,597,322	241,195,622	151	151	151	95.47	92.18	38.92	No
SRR6813717_2.fastq	1,597,322	241,195,622	151	151	151	89.96	84.27	38.35	No
SRR10363451_1.fastq	1,605,291	402,928,041	251	251	251	95.94	94.01	33.8	No
SRR10363451_2.fastq	1,605,291	402,928,041	251	251	251	89.62	85.53	33.84	No
SRR10363372_1.fastq	1,663,885	343,445,840	35	206.4	250	93.74	91.22	44.44	No
SRR10363372_2.fastq	1,663,885	349,488,778	35	210	251	78.09	70.87	45.75	No
SRR10363486_1.fastq	1,670,048	419,182,048	251	251	251	96.4	94.03	30.78	No
SRR10363486_2.fastq	1,670,048	419,182,048	251	251	251	92.03	87.26	30.81	No
SRR10363461_1.fastq	1,678,879	421,398,629	251	251	251	95.78	93.92	31.04	No
SRR10363461_2.fastq	1,678,879	421,398,629	251	251	251	90.2	86.45	31.09	No
SRR10363385_1.fastq	1,684,027	374,306,800	35	222.3	250	96.3	94.95	34.98	No
SRR10363385_2.fastq	1,684,027	375,020,343	35	222.7	251	90.93	88.15	35.08	No
SRR6871414.fastq	1,690,102	472,586,837	16	279.6	335	90.64	74.55	34.09	No*
SRR10363477_1.fastq	1,732,626	434,889,126	251	251	251	96.09	94.31	31	No
SRR10363477_2.fastq	1,732,626	434,889,126	251	251	251	91.08	87.55	31.03	No
SRR10363478_1.fastq	1,766,294	443,339,794	251	251	251	96.05	94.19	31.44	No
SRR10363478_2.fastq	1,766,294	443,339,794	251	251	251	90.61	86.94	31.48	No
SRR10363384_1.fastq	1,774,848	369,555,786	35	208.2	250	97.03	95.98	33.64	No
SRR10363384_2.fastq	1,774,848	369,849,591	35	208.4	250	94.73	93.1	33.68	No
SRR10363383_1.fastq	1,781,309	372,772,921	35	209.3	250	96.05	94.47	39.07	No
SRR10363383_2.fastq	1,781,309	373,779,220	35	209.8	250	91.17	88.06	39.25	No
SRR10363470_1.fastq	1,782,152	447,320,152	251	251	251	95.72	93.75	32.98	No
SRR10363470_2.fastq	1,782,152	447,320,152	251	251	251	90.32	86.58	33.1	No
SRR10363364_1.fastq	1,822,246	358,574,389	35	196.8	250	95.95	94.65	33.93	No
SRR10363364_2.fastq	1,822,246	359,782,430	35	197.4	250	89.43	86.19	34.16	No
SRR10363551_1.fastq	1,849,878	462,469,500	250	250	250	89.17	86.81	46.02	No
SRR10363551_2.fastq	1,849,878	462,469,500	250	250	250	81.55	78.09	45.76	No
SRR10363528_1.fastq	1,850,821	462,705,250	250	250	250	89.53	87.43	43.57	No

SRR10363528_2.fastq	1,850,821	462,705,250	250	250	250	83.12	80.01	44.02	No
SRR15694518_1.fastq	1,858,138	466,392,638	251	251	251	95.17	93.23	30.98	No
SRR15694518_2.fastq	1,858,138	466,392,638	251	251	251	88.65	84.55	31.05	No
SRR10363449_1.fastq	1,859,321	466,689,571	251	251	251	95.94	94.04	30.92	No
SRR10363449_2.fastq	1,859,321	466,689,571	251	251	251	89.69	85.65	30.93	No
SRR10363379_1.fastq	1,885,133	412,925,167	35	219	250	96.79	95.65	33.89	No
SRR10363379_2.fastq	1,885,133	413,482,146	35	219.3	250	92.64	90.44	33.95	No
SRR10363566_1.fastq	1,899,613	474,903,250	250	250	250	95.48	94.13	37.36	No
SRR10363566_2.fastq	1,899,613	474,903,250	250	250	250	90.93	88.73	37.51	No
SRR10363382_1.fastq	1,911,621	436,764,560	35	228.5	251	95.69	93.98	38.6	No
SRR10363382_2.fastq	1,911,621	437,768,102	35	229	251	88.25	84.31	38.85	No
SRR10363346_1.fastq	1,913,419	478,354,750	250	250	250	90.91	89.21	37.44	No
SRR10363346_2.fastq	1,913,419	478,354,750	250	250	250	85.96	83.65	37.44	No
SRR10363558_1.fastq	1,923,837	480,959,250	250	250	250	88.52	86	45.39	No
SRR10363558_2.fastq	1,923,837	480,959,250	250	250	250	78.87	75.3	46.06	No
SRR15694557_1.fastq	1,962,910	490,727,500	250	250	250	94.41	86.57	31.53	No
SRR15694557_2.fastq	1,962,910	490,727,500	250	250	250	94.61	87.05	31.57	No
SRR10363378_1.fastq	1,967,450	430,368,059	35	218.7	250	96.32	95	33.74	No
SRR10363378_2.fastq	1,967,450	430,668,602	35	218.9	250	94.89	93.2	33.79	No
SRR10363415_1.fastq	1,968,357	494,057,607	251	251	251	96.2	93.63	39.05	No
SRR10363415_2.fastq	1,968,357	494,057,607	251	251	251	90.34	84.65	39.14	No
SRR15694568_1.fastq	1,979,027	494,756,750	250	250	250	93.23	84.76	31.76	No
SRR15694568_2.fastq	1,979,027	494,756,750	250	250	250	93.52	85.36	31.87	No
SRR15694570_1.fastq	1,997,075	499,268,750	250	250	250	93.31	84.86	31.62	No
SRR15694570_2.fastq	1,997,075	499,268,750	250	250	250	93.65	85.59	31.72	No
SRR15694569_1.fastq	2,005,909	501,477,250	250	250	250	93.37	84.97	31.71	No
SRR15694569_2.fastq	2,005,909	501,477,250	250	250	250	93.71	85.69	31.82	No
SRR6117460_1.fastq	2,015,213	259,666,403	35	128.9	150	97.22	93.58	33.45	No
SRR6117460_2.fastq	2,015,213	259,714,429	35	128.9	150	94.75	89.67	33.49	No

SRR3091757_1.fastq	2,030,676	609,202,800	300	300	300	90.16	86.14	34.4	No
SRR3091757_2.fastq	2,030,676	609,202,800	300	300	300	83.78	76.81	34.76	No
SRR10363428_1.fastq	2,031,533	509,914,783	251	251	251	96.37	94.03	30.75	No
SRR10363428_2.fastq	2,031,533	509,914,783	251	251	251	91.8	86.98	30.77	No
SRR10363589_1.fastq	2,037,803	509,450,750	250	250	250	96.98	96.08	32.67	No
SRR10363589_2.fastq	2,037,803	509,450,750	250	250	250	94.32	92.92	32.71	No
SRR10363474_1.fastq	2,058,876	516,777,876	251	251	251	96.01	94.11	32.72	No
SRR10363474_2.fastq	2,058,876	516,777,876	251	251	251	90.72	87.03	32.71	No
SRR10363335_1.fastq	2,073,251	518,312,750	250	250	250	97.08	96.22	32.04	No
SRR10363335_2.fastq	2,073,251	518,312,750	250	250	250	94.57	93.25	32.07	No
SRR10363473_1.fastq	2,080,406	522,181,906	251	251	251	96.17	94.35	30.82	No
SRR10363473_2.fastq	2,080,406	522,181,906	251	251	251	91.19	87.61	30.82	No
SRR10363387_1.fastq	2,082,884	295,420,286	35	141.8	177	96.53	95.22	40.92	No
SRR10363387_2.fastq	2,082,884	296,729,550	35	142.5	179	93.9	91.63	41.06	No
SRR15694559_1.fastq	2,087,574	521,893,500	250	250	250	94.29	86.43	31.79	No
SRR15694559_2.fastq	2,087,574	521,893,500	250	250	250	94.68	87.21	31.84	No
SRR10363443_1.fastq	2,098,769	526,791,019	251	251	251	95.23	93.33	30.97	No
SRR10363443_2.fastq	2,098,769	526,791,019	251	251	251	89.43	85.61	31.07	No
SRR15694560_1.fastq	2,125,241	531,310,250	250	250	250	94.19	86.32	31.86	No
SRR15694560_2.fastq	2,125,241	531,310,250	250	250	250	94.68	87.26	31.89	No
SRR10363374_1.fastq	2,133,570	422,402,075	35	198	250	95.8	93.41	50.06	No
SRR10363374_2.fastq	2,133,570	424,476,889	35	199	250	88.25	82.46	50.37	No
SRR10363467_1.fastq	2,148,341	539,233,591	251	251	251	96.16	94.32	31.45	No
SRR10363467_2.fastq	2,148,341	539,233,591	251	251	251	91	87.36	31.38	No
SRR15694556_1.fastq	2,151,197	537,799,250	250	250	250	93.32	84.91	31.61	No
SRR15694556_2.fastq	2,151,197	537,799,250	250	250	250	93.64	85.61	31.72	No
SRR15694519_1.fastq	2,168,783	544,364,533	251	251	251	95.79	93.98	30.72	No
SRR15694519_2.fastq	2,168,783	544,364,533	251	251	251	90.34	86.63	30.77	No
SRR15694566_1.fastq	2,184,342	546,085,500	250	250	250	93.53	85.19	31.49	No

SRR15694566_2.fastq	2,184,342	546,085,500	250	250	250	93.85	85.82	31.56	No
SRR10363463_1.fastq	2,190,253	549,753,503	251	251	251	96.24	94.45	31.24	No
SRR10363463_2.fastq	2,190,253	549,753,503	251	251	251	90.22	86.37	31.24	No
SRR10363466_1.fastq	2,194,751	550,882,501	251	251	251	96.1	94.25	31.49	No
SRR10363466_2.fastq	2,194,751	550,882,501	251	251	251	90.92	87.27	31.52	No
SRR10363435_1.fastq	2,201,449	550,362,250	250	250	250	90.3	88.54	38.97	No
SRR10363435_2.fastq	2,201,449	550,362,250	250	250	250	85.27	82.9	38.58	No
SRR10363358_1.fastq	2,230,001	557,500,250	250	250	250	96.42	95.44	32.98	No
SRR10363358_2.fastq	2,230,001	557,500,250	250	250	250	93.72	92.25	33.07	No
SRR10363453_1.fastq	2,233,361	560,573,611	251	251	251	96.03	94.19	31.97	No
SRR10363453_2.fastq	2,233,361	560,573,611	251	251	251	90.22	86.42	31.99	No
SRR10363465_1.fastq	2,259,286	567,080,786	251	251	251	95.96	94.04	30.42	No
SRR10363465_2.fastq	2,259,286	567,080,786	251	251	251	90.04	86.04	30.4	No
SRR10363299_1.fastq	2,261,087	565,271,750	250	250	250	91.44	90.02	33.98	No
SRR10363299_2.fastq	2,261,087	565,271,750	250	250	250	84.86	82.59	34.19	No
SRR3091761_1.fastq	2,267,313	680,193,900	300	300	300	89.9	85.01	33.97	No
SRR3091761_2.fastq	2,267,313	680,193,900	300	300	300	85.12	78.4	34.29	No
SRR10363445_1.fastq	2,267,615	569,171,365	251	251	251	96.08	94.3	30.88	No
SRR10363445_2.fastq	2,267,615	569,171,365	251	251	251	90.19	86.36	30.91	No
SRR15694506_1.fastq	2,276,443	569,110,750	250	250	250	93.44	85.05	31.87	No
SRR15694506_2.fastq	2,276,443	569,110,750	250	250	250	93.75	85.72	31.98	No
SRR3091754_1.fastq	2,277,029	683,108,700	300	300	300	92.45	87.58	33.83	No
SRR3091754_2.fastq	2,277,029	683,108,700	300	300	300	88.4	81.67	34	No
SRR15694513_1.fastq	2,282,002	572,782,502	251	251	251	95.08	93.25	31.59	No
SRR15694513_2.fastq	2,282,002	572,782,502	251	251	251	89.93	86.37	31.68	No
SRR10363311_1.fastq	2,354,516	588,629,000	250	250	250	94.3	92.35	41.54	No
SRR10363311_2.fastq	2,354,516	588,629,000	250	250	250	87.98	84.77	41.71	No
SRR15694565_1.fastq	2,369,801	592,450,250	250	250	250	93.36	84.92	31.59	No
SRR15694565_2.fastq	2,369,801	592,450,250	250	250	250	93.8	85.86	31.74	No

SRR10363475_1.fastq	2,374,915	596,103,665	251	251	251	96.01	94.2	31.24	No
SRR10363475_2.fastq	2,374,915	596,103,665	251	251	251	90.33	86.59	31.27	No
SRR6813718_1.fastq	2,378,388	359,136,588	151	151	151	94.89	90.83	37.31	No
SRR6813718_2.fastq	2,378,388	359,136,588	151	151	151	98.1	95.12	37.49	No
SRR10363438_1.fastq	2,390,209	599,942,459	251	251	251	96.38	94.01	31.9	No
SRR10363438_2.fastq	2,390,209	599,942,459	251	251	251	91.79	86.91	31.91	No
SRR15694553_1.fastq	2,393,313	598,328,250	250	250	250	95.04	88.24	50.04	No
SRR15694553_2.fastq	2,393,313	598,328,250	250	250	250	95.45	88.97	50.12	No
SRR10363369_1.fastq	2,404,569	603,546,819	251	251	251	96.04	94.33	32.85	No
SRR10363369_2.fastq	2,404,569	603,546,819	251	251	251	90.66	87.1	32.89	No
SRR10363447_1.fastq	2,407,548	604,294,548	251	251	251	95.68	93.83	31.01	No
SRR10363447_2.fastq	2,407,548	604,294,548	251	251	251	89.93	86.14	30.99	No
SRR10363448_1.fastq	2,411,101	605,186,351	251	251	251	95.91	94.02	31.06	No
SRR10363448_2.fastq	2,411,101	605,186,351	251	251	251	89.43	85.33	31.05	No
SRR10363444_1.fastq	2,415,779	606,360,529	251	251	251	96.29	94.56	31.26	No
SRR10363444_2.fastq	2,415,779	606,360,529	251	251	251	90.89	87.29	31.29	No
SRR10363476_1.fastq	2,420,988	607,667,988	251	251	251	96	94.13	30.38	No
SRR10363476_2.fastq	2,420,988	607,667,988	251	251	251	90.63	86.94	30.29	No
SRR15694555_1.fastq	2,423,821	605,955,250	250	250	250	93.46	85.08	31.46	No
SRR15694555_2.fastq	2,423,821	605,955,250	250	250	250	93.6	85.45	31.56	No
SRR10363401_1.fastq	2,427,746	609,364,246	251	251	251	96.36	93.99	30.84	No
SRR10363401_2.fastq	2,427,746	609,364,246	251	251	251	90.75	85.4	30.88	No
SRR10363517_1.fastq	2,427,983	606,995,750	250	250	250	95.76	94.46	39.43	No
SRR10363517_2.fastq	2,427,983	606,995,750	250	250	250	91.99	89.93	39.55	No
SRR10363483_1.fastq	2,430,739	610,115,489	251	251	251	96.3	94.54	31.45	No
SRR10363483_2.fastq	2,430,739	610,115,489	251	251	251	91.2	87.68	31.43	No
SRR3091758_1.fastq	2,431,933	729,579,900	300	300	300	85.77	82.04	33.95	No
SRR3091758_2.fastq	2,431,933	729,579,900	300	300	300	79.34	72.69	34.49	No
SRR15694517_1.fastq	2,432,112	608,028,000	250	250	250	93.31	85	34.03	No

SRR15694517_2.fastq	2,432,112	608,028,000	250	250	250	93.65	85.63	34.02	No
SRR3091760_1.fastq	2,433,874	730,162,200	300	300	300	86.26	82.13	34.79	No
SRR3091760_2.fastq	2,433,874	730,162,200	300	300	300	80.7	74.06	35.25	No
SRR10363322_1.fastq	2,441,438	610,359,500	250	250	250	94.02	92.66	40.17	No
SRR10363322_2.fastq	2,441,438	610,359,500	250	250	250	90.1	87.98	40.31	No
SRR15694495_1.fastq	2,446,269	611,567,250	250	250	250	93.43	85.09	31.92	No
SRR15694495_2.fastq	2,446,269	611,567,250	250	250	250	93.74	85.72	32.05	No
SRR3091759_1.fastq	2,449,224	734,767,200	300	300	300	78.62	74.25	36.35	No
SRR3091759_2.fastq	2,449,224	734,767,200	300	300	300	72.32	65.86	37.25	No
SRR10363462_1.fastq	2,472,867	620,689,617	251	251	251	96.13	94.33	31.31	No
SRR10363462_2.fastq	2,472,867	620,689,617	251	251	251	90.58	86.92	31.35	No
SRR15694558_1.fastq	2,548,837	637,209,250	250	250	250	94.26	86.44	32.37	No
SRR15694558_2.fastq	2,548,837	637,209,250	250	250	250	94.61	87.13	32.41	No
SRR10363556_1.fastq	2,579,651	647,492,401	251	251	251	96.15	93.64	30.82	No
SRR10363556_2.fastq	2,579,651	647,492,401	251	251	251	91.22	86.08	30.84	No
SRR15694551_1.fastq	2,642,318	660,579,500	250	250	250	93.06	84.57	32.13	No
SRR15694551_2.fastq	2,642,318	660,579,500	250	250	250	93.38	85.22	32.22	No
SRR10363481_1.fastq	2,649,765	665,091,015	251	251	251	96.31	94.56	30.66	No
SRR10363481_2.fastq	2,649,765	665,091,015	251	251	251	91.02	87.44	30.67	No
SRR10363596_1.fastq	2,706,576	679,350,576	251	251	251	95.95	94.25	31.6	No
SRR10363596_2.fastq	2,706,576	679,350,576	251	251	251	91.45	88.2	31.66	No
SRR15694514_1.fastq	2,706,576	679,350,576	251	251	251	95.95	94.25	31.6	No
SRR15694514_2.fastq	2,706,576	679,350,576	251	251	251	91.45	88.2	31.66	No
SRR10363423_1.fastq	2,708,391	679,806,141	251	251	251	96.03	93.51	30.68	No
SRR10363423_2.fastq	2,708,391	679,806,141	251	251	251	91.03	85.88	30.72	No
SRR10363436_1.fastq	2,732,132	685,765,132	251	251	251	95.76	93	37.36	No
SRR10363436_2.fastq	2,732,132	685,765,132	251	251	251	90.13	84.36	37.47	No
SRR15694567_1.fastq	2,792,090	698,022,500	250	250	250	93.41	85.06	31.83	No
SRR15694567_2.fastq	2,792,090	698,022,500	250	250	250	93.5	85.3	31.9	No

SRR10363577_1.fastq	2,815,539	703,884,750	250	250	250	95.4	93.93	42.21	No
SRR10363577_2.fastq	2,815,539	703,884,750	250	250	250	90.07	87.5	42.4	No
SRR15694562_1.fastq	2,869,123	717,280,750	250	250	250	94.78	87.22	32.05	No
SRR15694562_2.fastq	2,869,123	717,280,750	250	250	250	95.09	87.97	32.09	No
SRR10363434_1.fastq	2,902,546	728,539,046	251	251	251	96.27	93.86	30.82	No
SRR10363434_2.fastq	2,902,546	728,539,046	251	251	251	90.86	85.56	30.8	No
SRR10363450_1.fastq	2,919,174	732,712,674	251	251	251	96.07	94.23	31.04	No
SRR10363450_2.fastq	2,919,174	732,712,674	251	251	251	90.35	86.54	30.96	No
SRR10363480_1.fastq	2,920,301	730,075,250	250	250	250	96.08	94.85	38.49	No
SRR10363480_2.fastq	2,920,301	730,075,250	250	250	250	92.62	90.64	38.65	No
SRR18395902.fastq	2,921,126	828,362,555	8	283.6	363	83.03	45.38	34.29	No*
SRR15694484_1.fastq	2,925,121	731,280,250	250	250	250	93.44	85.06	31.4	No
SRR15694484_2.fastq	2,925,121	731,280,250	250	250	250	93.82	85.79	31.47	No
SRR10363479_1.fastq	2,933,805	736,385,055	251	251	251	96.28	94.55	31.2	No
SRR10363479_2.fastq	2,933,805	736,385,055	251	251	251	91.25	87.78	31.23	No
SRR15694554_1.fastq	2,966,912	741,728,000	250	250	250	96.85	95	45	No
SRR15694554_2.fastq	2,966,912	741,728,000	250	250	250	91.75	88.17	44.97	No
SRR11516700_1.fastq	2,973,265	697,591,430	35	234.6	299	96.55	90.74	32.45	No
SRR11516700_2.fastq	2,973,265	698,666,735	35	235	299	93.19	84.37	32.6	No
SRR25512130_1.fastq	2,987,227	443,224,246	35	148.4	151	96.68	92.11	30.1	No
SRR25512130_2.fastq	2,987,227	443,243,958	35	148.4	151	95.37	89.57	30.36	No
SRR10363406_1.fastq	3,008,413	755,111,663	251	251	251	96.35	93.97	31.02	No
SRR10363406_2.fastq	3,008,413	755,111,663	251	251	251	90.64	85.21	31.07	No
SRR18395903.fastq	3,016,632	822,002,410	8	272.5	358	82.92	45.74	31.01	No*
SRR6813716_1.fastq	3,045,081	459,807,231	151	151	151	95.49	91.6	39.34	No
SRR6813716_2.fastq	3,045,081	459,807,231	151	151	151	97.04	93.76	39.52	No
SRR10363403_1.fastq	3,051,594	765,950,094	251	251	251	96.08	93.54	30	No
SRR10363403_2.fastq	3,051,594	765,950,094	251	251	251	91.63	86.7	30.01	No
SRR10363411_1.fastq	3,058,483	767,679,233	251	251	251	95.62	92.88	30.1	No

SRR10363411_2.fastq	3,058,483	767,679,233	251	251	251	89.99	84.3	30.13	No
SRR9070314_1.fastq	3,068,351	913,331,556	35	297.7	301	96.01	90.99	31.12	No
SRR9070314_2.fastq	3,068,351	913,482,162	35	297.7	301	92.75	84.7	31.28	No
SRR15694563_1.fastq	3,083,551	770,887,750	250	250	250	97.76	96.46	32.2	No
SRR15694563_2.fastq	3,083,551	770,887,750	250	250	250	94.29	91.74	32.04	No
SRR10363425_1.fastq	3,105,827	779,562,577	251	251	251	95.81	93.21	31.02	No
SRR10363425_2.fastq	3,105,827	779,562,577	251	251	251	90.18	84.56	31.07	No
SRR10363492_1.fastq	3,127,114	784,905,614	251	251	251	93.86	89.94	40	No
SRR10363492_2.fastq	3,127,114	784,905,614	251	251	251	87.05	79.49	40.12	No
SRR3091753_1.fastq	3,129,065	938,719,500	300	300	300	79.25	74.95	36.03	No
SRR3091753_2.fastq	3,129,065	938,719,500	300	300	300	73.76	66.89	36.47	No
SRR9070313_1.fastq	3,174,175	944,406,315	35	297.5	301	95.36	90.19	31.21	No
SRR9070313_2.fastq	3,174,175	944,500,482	35	297.6	301	92.29	84.37	31.36	No
SRR10363491_1.fastq	3,197,312	802,525,312	251	251	251	96.25	93.86	30.94	No
SRR10363491_2.fastq	3,197,312	802,525,312	251	251	251	91.65	86.8	31	No
SRR15694549_1.fastq	3,330,588	832,647,000	250	250	250	97.63	96.33	32.25	No
SRR15694549_2.fastq	3,330,588	832,647,000	250	250	250	94.52	92.12	32.15	No
SRR10363501_1.fastq	3,331,081	836,101,331	251	251	251	94.79	91.38	30.31	No
SRR10363501_2.fastq	3,331,081	836,101,331	251	251	251	86.52	78.74	30.34	No
SRR15694512_1.fastq	3,331,081	836,101,331	251	251	251	94.79	91.38	30.31	No
SRR15694512_2.fastq	3,331,081	836,101,331	251	251	251	86.52	78.74	30.34	No
SRR10363494_1.fastq	3,374,592	337,459,200	100	100	100	96.57	91	50.79	No
SRR10363494_2.fastq	3,374,592	337,459,200	100	100	100	86.76	79.14	50.94	No
SRR3091755_1.fastq	3,506,102	1,051,830,600	300	300	300	87.51	82.7	34.34	No
SRR3091755_2.fastq	3,506,102	1,051,830,600	300	300	300	83.41	77.31	35	No
SRR10363433_1.fastq	3,515,883	882,486,633	251	251	251	95.51	92.69	30.56	No
SRR10363433_2.fastq	3,515,883	882,486,633	251	251	251	88.66	82.22	30.59	No
SRR10363421_1.fastq	3,551,381	891,396,631	251	251	251	96.39	94.07	31.14	No
SRR10363421_2.fastq	3,551,381	891,396,631	251	251	251	92.34	87.86	31.19	No

SRR15694564_1.fastq	3,558,212	889,553,000	250	250	250	97.74	96.42	32.39	No
SRR15694564_2.fastq	3,558,212	889,553,000	250	250	250	94.65	92.2	32.25	No
SRR10363547_1.fastq	3,640,715	913,819,465	251	251	251	96.04	93.56	31.9	No
SRR10363547_2.fastq	3,640,715	913,819,465	251	251	251	91.05	85.96	32.01	No
SRR9070312_1.fastq	3,651,879	1,086,750,636	35	297.6	301	94.74	89.46	31.55	No
SRR9070312_2.fastq	3,651,879	1,087,030,387	35	297.7	301	90.93	82.33	31.9	No
SRR10363315_1.fastq	3,701,690	929,124,190	251	251	251	97.05	95.64	30.76	No
SRR10363315_2.fastq	3,701,690	929,124,190	251	251	251	92.13	89.03	30.76	No
SRR15694545_1.fastq	3,701,690	929,124,190	251	251	251	97.05	95.64	30.76	No
SRR15694545_2.fastq	3,701,690	929,124,190	251	251	251	92.13	89.03	30.76	No
SRR10363437_1.fastq	3,702,319	929,282,069	251	251	251	96.42	94.11	30.92	No
SRR10363437_2.fastq	3,702,319	929,282,069	251	251	251	92.07	87.44	30.92	No
SRR6147964_1.fastq	3,746,781	516,028,156	35	137.7	151	99.12	97.04	36.28	No
SRR6147964_2.fastq	3,746,781	515,753,313	35	137.7	151	97.39	93.67	36.36	No
SRR10363586_1.fastq	3,753,873	942,222,123	251	251	251	95.74	92.97	33.67	No
SRR10363586_2.fastq	3,753,873	942,222,123	251	251	251	89.15	82.87	33.73	No
SRR15694571_1.fastq	3,767,733	945,700,983	251	251	251	95.1	91.85	33.56	No
SRR15694571_2.fastq	3,767,733	945,700,983	251	251	251	90.52	84.78	33.67	No
SRR10363397_1.fastq	3,799,267	953,616,017	251	251	251	96.28	93.87	30.67	No
SRR10363397_2.fastq	3,799,267	953,616,017	251	251	251	90.7	85.25	30.69	No
SRR15694528_1.fastq	3,805,006	951,251,500	250	250	250	97.79	96.53	32.68	No
SRR15694528_2.fastq	3,805,006	951,251,500	250	250	250	95.08	92.91	32.57	No
SRR10363504_1.fastq	3,830,286	961,401,786	251	251	251	95.51	92.66	37.98	No
SRR10363504_2.fastq	3,830,286	961,401,786	251	251	251	89.35	83.22	38.13	No
SRR10363410_1.fastq	3,853,566	967,245,066	251	251	251	96.04	93.53	30.68	No
SRR10363410_2.fastq	3,853,566	967,245,066	251	251	251	91.08	85.94	30.63	No
SRR10363409_1.fastq	3,904,691	980,077,441	251	251	251	96.13	93.66	30.41	No
SRR10363409_2.fastq	3,904,691	980,077,441	251	251	251	91.22	86.09	30.43	No
SRR10363585_1.fastq	3,950,047	991,461,797	251	251	251	96.38	94.05	31.35	No

SRR10363585_2.fastq	3,950,047	991,461,797	251	251	251	91.79	86.99	31.4	No
SRR10363365_1.fastq	3,953,391	730,249,629	35	184.7	249	96.75	95.71	34.53	No
SRR10363365_2.fastq	3,953,391	731,297,336	35	185	249	93.63	91.52	34.68	No
SRR11516703_1.fastq	3,960,182	1,190,228,484	141	300.5	301	92.87	84.85	35.19	No
SRR11516703_2.fastq	3,960,182	1,190,396,738	142	300.6	301	86.75	75.13	35.48	No
SRR10363488_1.fastq	3,972,769	997,165,019	251	251	251	96.19	93.76	30.64	No
SRR10363488_2.fastq	3,972,769	997,165,019	251	251	251	90.54	85.1	30.68	No
SRR15694483_1.fastq	3,972,769	997,165,019	251	251	251	96.19	93.76	30.64	No
SRR15694483_2.fastq	3,972,769	997,165,019	251	251	251	90.54	85.1	30.68	No
SRR15694561_1.fastq	4,042,349	1,014,629,599	251	251	251	94.1	90.55	32.61	No
SRR15694561_2.fastq	4,042,349	1,014,629,599	251	251	251	89.57	83.74	32.72	No
SRR15694550_1.fastq	4,062,656	1,019,726,656	251	251	251	94.83	91.56	32.06	No
SRR15694550_2.fastq	4,062,656	1,019,726,656	251	251	251	88.7	82.16	32.18	No
SRR11516702_1.fastq	4,154,894	1,175,290,897	35	282.9	301	96.84	91.62	33.93	No
SRR11516702_2.fastq	4,154,894	1,179,467,788	35	283.9	301	88.02	76.1	34.2	No
SRR10363592_1.fastq	4,159,256	1,043,973,256	251	251	251	95.91	93.31	30.95	No
SRR10363592_2.fastq	4,159,256	1,043,973,256	251	251	251	89.17	83.03	31.03	No
SRR15694539_1.fastq	4,205,031	1,055,462,781	251	251	251	94.75	91.44	32.42	No
SRR15694539_2.fastq	4,205,031	1,055,462,781	251	251	251	90.2	84.49	32.5	No
SRR10363441_1.fastq	4,211,270	1,057,028,770	251	251	251	96.03	93.57	31.06	No
SRR10363441_2.fastq	4,211,270	1,057,028,770	251	251	251	91.63	86.82	31.11	No
SRR10363405_1.fastq	4,235,458	1,063,099,958	251	251	251	96.23	93.87	30.92	No
SRR10363405_2.fastq	4,235,458	1,063,099,958	251	251	251	91.22	86.21	30.99	No
SRR10363500_1.fastq	4,240,425	1,064,346,675	251	251	251	95.19	92.05	31.51	No
SRR10363500_2.fastq	4,240,425	1,064,346,675	251	251	251	89.09	82.71	31.53	No
SRR15694511_1.fastq	4,240,425	1,064,346,675	251	251	251	95.19	92.05	31.51	No
SRR15694511_2.fastq	4,240,425	1,064,346,675	251	251	251	89.09	82.71	31.53	No
SRR11516701_1.fastq	4,275,456	1,284,884,905	145	300.5	301	92.17	85.02	35.83	No
SRR11516701_2.fastq	4,275,456	1,285,492,643	165	300.7	301	85.02	73.93	36.54	No

SRR15694572_1.fastq	4,288,469	1,076,405,719	251	251	251	95.03	91.81	31.25	No
SRR15694572_2.fastq	4,288,469	1,076,405,719	251	251	251	91.28	86.09	31.29	No
SRR10363497_1.fastq	4,294,126	1,077,825,626	251	251	251	95.28	92.15	33.49	No
SRR10363497_2.fastq	4,294,126	1,077,825,626	251	251	251	89.31	82.96	33.55	No
SRR10363419_1.fastq	4,301,281	1,079,621,531	251	251	251	96.47	94.16	30.38	No
SRR10363419_2.fastq	4,301,281	1,079,621,531	251	251	251	92.8	88.5	30.39	No
SRR10363468_1.fastq	4,333,953	1,087,822,203	251	251	251	96.96	95.22	43.06	No
SRR10363468_2.fastq	4,333,953	1,087,822,203	251	251	251	89.75	85.71	43.05	No
SRR10363430_1.fastq	4,357,912	1,093,835,912	251	251	251	95.74	92.99	38.15	No
SRR10363430_2.fastq	4,357,912	1,093,835,912	251	251	251	90.69	85.2	38.22	No
SRR15694573_1.fastq	4,398,170	1,103,940,670	251	251	251	95.15	91.94	30.92	No
SRR15694573_2.fastq	4,398,170	1,103,940,670	251	251	251	90.65	84.99	30.91	No
SRR10363499_1.fastq	4,431,330	1,112,263,830	251	251	251	94.75	91.4	30.49	No
SRR10363499_2.fastq	4,431,330	1,112,263,830	251	251	251	88.3	81.57	30.5	No
SRR15694510_1.fastq	4,431,330	1,112,263,830	251	251	251	94.75	91.4	30.49	No
SRR15694510_2.fastq	4,431,330	1,112,263,830	251	251	251	88.3	81.57	30.5	No
SRR10363342_1.fastq	4,464,112	1,120,492,112	251	251	251	94.44	92.23	42.37	No
SRR10363342_2.fastq	4,464,112	1,120,492,112	251	251	251	86.29	81.72	42.63	No
SRR10363328_1.fastq	4,481,980	448,198,000	100	100	100	98.85	96.09	32.04	No
SRR10363328_2.fastq	4,481,980	448,198,000	100	100	100	93.3	89.08	32.14	No
SRR10363496_1.fastq	4,516,909	1,133,744,159	251	251	251	95.18	91.99	32.89	No
SRR10363496_2.fastq	4,516,909	1,133,744,159	251	251	251	86.48	78.69	33.04	No
SRR10363493_1.fastq	4,532,570	1,137,675,070	251	251	251	94.99	91.72	34.06	No
SRR10363493_2.fastq	4,532,570	1,137,675,070	251	251	251	88.77	82.2	34.14	No
SRR15694507_1.fastq	4,532,570	1,137,675,070	251	251	251	94.99	91.72	34.06	No
SRR15694507_2.fastq	4,532,570	1,137,675,070	251	251	251	88.77	82.2	34.14	No
SRR15694515_1.fastq	4,550,421	1,142,155,671	251	251	251	95.74	93.78	44.58	No
SRR15694515_2.fastq	4,550,421	1,142,155,671	251	251	251	88.94	85.02	44.8	No
SRR10363579_1.fastq	4,573,537	1,147,957,787	251	251	251	96.98	95.61	31.13	No

SRR10363579_2.fastq	4,573,537	1,147,957,787	251	251	251	92.63	89.81	31.13	No
SRR15694537_1.fastq	4,573,537	1,147,957,787	251	251	251	96.98	95.61	31.13	No
SRR15694537_2.fastq	4,573,537	1,147,957,787	251	251	251	92.63	89.81	31.13	No
SRR10363422_1.fastq	4,651,283	1,167,472,033	251	251	251	96.3	93.94	30.45	No
SRR10363422_2.fastq	4,651,283	1,167,472,033	251	251	251	92.15	87.56	30.41	No
SRR10363554_1.fastq	4,693,431	1,178,051,181	251	251	251	96.4	94.07	31.02	No
SRR10363554_2.fastq	4,693,431	1,178,051,181	251	251	251	92.24	87.66	31	No
SRR10363446_1.fastq	4,693,600	469,360,000	100	100	100	98.49	94.79	39.97	No
SRR10363446_2.fastq	4,693,600	469,360,000	100	100	100	92.55	87.64	40.05	No
SRR10363416_1.fastq	4,735,063	1,188,500,813	251	251	251	96.32	93.97	30.56	No
SRR10363416_2.fastq	4,735,063	1,188,500,813	251	251	251	92.91	88.73	30.52	No
SRR10363457_1.fastq	4,781,901	1,200,257,151	251	251	251	97.24	95.72	39.49	No
SRR10363457_2.fastq	4,781,901	1,200,257,151	251	251	251	91.11	87.52	39.49	No
SRR10363309_1.fastq	4,832,425	1,212,938,675	251	251	251	96.55	94.83	41.9	No
SRR10363309_2.fastq	4,832,425	1,212,938,675	251	251	251	91.05	87.73	41.93	No
SRR10363599_1.fastq	4,860,143	1,219,895,893	251	251	251	95.89	94.07	30.54	No
SRR10363599_2.fastq	4,860,143	1,219,895,893	251	251	251	90.57	86.98	30.61	No
SRR10363564_1.fastq	4,869,705	1,222,295,955	251	251	251	96.86	95.4	30.65	No
SRR10363564_2.fastq	4,869,705	1,222,295,955	251	251	251	91.77	88.58	30.59	No
SRR15694523_1.fastq	4,869,705	1,222,295,955	251	251	251	96.86	95.4	30.65	No
SRR15694523_2.fastq	4,869,705	1,222,295,955	251	251	251	91.77	88.58	30.59	No
SRR10363338_1.fastq	4,878,290	1,224,450,790	251	251	251	95.89	93.9	33.05	No
SRR10363338_2.fastq	4,878,290	1,224,450,790	251	251	251	89.44	85.33	33.12	No
SRR10363337_1.fastq	4,882,049	1,225,394,299	251	251	251	96.09	94.16	35.37	No
SRR10363337_2.fastq	4,882,049	1,225,394,299	251	251	251	90.1	86.21	35.49	No
SRR10363567_1.fastq	4,909,852	1,232,372,852	251	251	251	96.4	94.8	33.48	No
SRR10363567_2.fastq	4,909,852	1,232,372,852	251	251	251	91.66	88.71	33.6	No
SRR15694525_1.fastq	4,909,852	1,232,372,852	251	251	251	96.4	94.8	33.48	No
SRR15694525_2.fastq	4,909,852	1,232,372,852	251	251	251	91.66	88.71	33.6	No

SRR10363495_1.fastq	4,920,039	1,234,929,789	251	251	251	94.97	91.69	31.25	No
SRR10363495_2.fastq	4,920,039	1,234,929,789	251	251	251	88.48	81.71	31.27	No
SRR15694508_1.fastq	4,920,039	1,234,929,789	251	251	251	94.97	91.69	31.25	No
SRR15694508_2.fastq	4,920,039	1,234,929,789	251	251	251	88.48	81.71	31.27	No
SRR10363498_1.fastq	4,929,637	1,237,338,887	251	251	251	94.92	91.6	32.59	No
SRR10363498_2.fastq	4,929,637	1,237,338,887	251	251	251	88.5	81.78	32.67	No
SRR15694509_1.fastq	4,929,637	1,237,338,887	251	251	251	94.92	91.6	32.59	No
SRR15694509_2.fastq	4,929,637	1,237,338,887	251	251	251	88.5	81.78	32.67	No
SRR10363597_1.fastq	4,944,832	1,241,152,832	251	251	251	95.88	93.8	38.68	No
SRR10363597_2.fastq	4,944,832	1,241,152,832	251	251	251	88.84	84.71	38.89	No
SRR10363578_1.fastq	5,012,354	1,258,100,854	251	251	251	97.06	95.73	31.74	No
SRR10363578_2.fastq	5,012,354	1,258,100,854	251	251	251	92.91	90.13	31.74	No
SRR15694536_1.fastq	5,012,354	1,258,100,854	251	251	251	97.06	95.73	31.74	No
SRR15694536_2.fastq	5,012,354	1,258,100,854	251	251	251	92.91	90.13	31.74	No
SRR10363582_1.fastq	5,015,225	1,258,821,475	251	251	251	95.2	93.34	39.58	No
SRR10363582_2.fastq	5,015,225	1,258,821,475	251	251	251	89.98	86.75	39.66	No
SRR15694541_1.fastq	5,015,225	1,258,821,475	251	251	251	95.2	93.34	39.58	No
SRR15694541_2.fastq	5,015,225	1,258,821,475	251	251	251	89.98	86.75	39.66	No
SRR15694516_1.fastq	5,016,775	1,259,210,525	251	251	251	96.83	95.21	39.64	No
SRR15694516_2.fastq	5,016,775	1,259,210,525	251	251	251	91.28	87.86	39.55	No
SRR10363412_1.fastq	5,027,170	1,261,819,670	251	251	251	95.49	92.82	30.75	No
SRR10363412_2.fastq	5,027,170	1,261,819,670	251	251	251	90.97	85.92	30.87	No
SRR10363354_1.fastq	5,078,205	1,274,629,455	251	251	251	95.84	93.91	34.72	No
SRR10363354_2.fastq	5,078,205	1,274,629,455	251	251	251	90.52	86.93	34.83	No
SRR10363363_1.fastq	5,096,568	797,712,007	35	156.5	207	96.25	95.1	34.65	No
SRR10363363_2.fastq	5,096,568	798,441,610	35	156.7	207	94.48	92.71	34.77	No
SRR15694501_1.fastq	5,121,367	1,285,463,117	251	251	251	96.68	94.47	32.03	No
SRR15694501_2.fastq	5,121,367	1,285,463,117	251	251	251	93.04	88.89	32.15	No
SRR10363341_1.fastq	5,122,901	1,285,848,151	251	251	251	96.09	94.1	43.35	No

SRR10363341_2.fastq	5,122,901	1,285,848,151	251	251	251	90.32	86.55	43.46	No
SRR10363570_1.fastq	5,134,566	1,288,776,066	251	251	251	96.94	95.48	30.51	No
SRR10363570_2.fastq	5,134,566	1,288,776,066	251	251	251	92.33	89.28	30.49	No
SRR15694529_1.fastq	5,134,566	1,288,776,066	251	251	251	96.94	95.48	30.51	No
SRR15694529_2.fastq	5,134,566	1,288,776,066	251	251	251	92.33	89.28	30.49	No
SRR15694498_1.fastq	5,209,459	520,945,900	100	100	100	98.89	96.17	30.31	No
SRR15694498_2.fastq	5,209,459	520,945,900	100	100	100	93.9	89.98	30.35	No
SRR15694494_1.fastq	5,230,259	1,312,795,009	251	251	251	96.3	93.96	31.9	No
SRR15694494_2.fastq	5,230,259	1,312,795,009	251	251	251	92.87	88.73	31.91	No
SRR26320658_1.fastq	5,249,996	792,591,141	1	151	151	97.4	92.61	35.61	No
SRR26320658_2.fastq	5,249,996	792,690,915	8	151	151	95.61	89.15	35.74	No
SRR18395901_1.fastq	5,256,966	771,265,043	35	146.7	151	94.6	91.08	33.1	No
SRR18395901_2.fastq	5,256,966	771,401,546	35	146.7	151	92.93	88.28	33.03	No
SRR10363569_1.fastq	5,278,703	1,324,954,453	251	251	251	97.15	95.77	30.76	No
SRR10363569_2.fastq	5,278,703	1,324,954,453	251	251	251	93.16	90.36	30.73	No
SRR15694527_1.fastq	5,278,703	1,324,954,453	251	251	251	97.15	95.77	30.76	No
SRR15694527_2.fastq	5,278,703	1,324,954,453	251	251	251	93.16	90.36	30.73	No
SRR10363347_1.fastq	5,280,075	1,325,298,825	251	251	251	96.12	94.22	31.19	No
SRR10363347_2.fastq	5,280,075	1,325,298,825	251	251	251	89.87	85.82	31.22	No
SRR10363587_1.fastq	5,311,355	1,333,150,105	251	251	251	96.32	94.03	32.36	No
SRR10363587_2.fastq	5,311,355	1,333,150,105	251	251	251	92.01	87.48	32.52	No
SRR10363575_1.fastq	5,317,635	1,334,726,385	251	251	251	97.11	95.82	30.85	No
SRR10363575_2.fastq	5,317,635	1,334,726,385	251	251	251	93.31	90.74	30.86	No
SRR15694534_1.fastq	5,317,635	1,334,726,385	251	251	251	97.11	95.82	30.85	No
SRR15694534_2.fastq	5,317,635	1,334,726,385	251	251	251	93.31	90.74	30.86	No
SRR10363563_1.fastq	5,322,130	1,335,854,630	251	251	251	97	95.68	30.79	No
SRR10363563_2.fastq	5,322,130	1,335,854,630	251	251	251	92.88	90.2	30.8	No
SRR15694522_1.fastq	5,322,130	1,335,854,630	251	251	251	97	95.68	30.79	No
SRR15694522_2.fastq	5,322,130	1,335,854,630	251	251	251	92.88	90.2	30.8	No

SRR10363571_1.fastq	5,364,662	1,346,530,162	251	251	251	97.07	95.7	30.95	No
SRR10363571_2.fastq	5,364,662	1,346,530,162	251	251	251	92.89	90.1	30.92	No
SRR15694530_1.fastq	5,364,662	1,346,530,162	251	251	251	97.07	95.7	30.95	No
SRR15694530_2.fastq	5,364,662	1,346,530,162	251	251	251	92.89	90.1	30.92	No
SRR15694502_1.fastq	5,369,422	536,942,200	100	100	100	98.88	95.93	34.43	No
SRR15694502_2.fastq	5,369,422	536,942,200	100	100	100	93.97	89.75	34.49	No
SRR10363303_1.fastq	5,439,610	1,365,342,110	251	251	251	96.28	94.49	31.19	No
SRR10363303_2.fastq	5,439,610	1,365,342,110	251	251	251	91.83	88.48	31.27	No
SRR15694492_1.fastq	5,490,333	549,033,300	100	100	100	98.9	96.2	30.86	No
SRR15694492_2.fastq	5,490,333	549,033,300	100	100	100	95.02	91.46	30.75	No
SRR10363573_1.fastq	5,504,164	1,381,545,164	251	251	251	97.08	95.75	30.65	No
SRR10363573_2.fastq	5,504,164	1,381,545,164	251	251	251	93.46	90.91	30.59	No
SRR15694532_1.fastq	5,504,164	1,381,545,164	251	251	251	97.08	95.75	30.65	No
SRR15694532_2.fastq	5,504,164	1,381,545,164	251	251	251	93.46	90.91	30.59	No
SRR10363538_1.fastq	5,506,947	1,382,243,697	251	251	251	96.54	94.27	31.31	No
SRR10363538_2.fastq	5,506,947	1,382,243,697	251	251	251	91.9	87.12	31.34	No
SRR10363333_1.fastq	5,544,235	1,391,602,985	251	251	251	95.11	93.26	33.59	No
SRR10363333_2.fastq	5,544,235	1,391,602,985	251	251	251	89.3	85.76	33.81	No
SRR6147945_1.fastq	5,598,532	775,283,772	35	138.5	151	99.13	97.16	32.89	No
SRR6147945_2.fastq	5,598,532	774,879,576	35	138.4	151	97.64	94.29	32.88	No
SRR26320661_1.fastq	5,647,572	852,562,881	2	151	151	97.02	91.79	35.02	No
SRR26320661_2.fastq	5,647,572	852,716,239	9	151	151	96.73	91.35	35.08	No
SRR10363594_1.fastq	5,647,580	1,417,542,580	251	251	251	96.06	93.56	30.6	No
SRR10363594_2.fastq	5,647,580	1,417,542,580	251	251	251	92	87.29	30.63	No
SRR10363572_1.fastq	5,649,748	1,418,086,748	251	251	251	97.22	95.97	32.27	No
SRR10363572_2.fastq	5,649,748	1,418,086,748	251	251	251	93.81	91.4	32.26	No
SRR15694531_1.fastq	5,649,748	1,418,086,748	251	251	251	97.22	95.97	32.27	No
SRR15694531_2.fastq	5,649,748	1,418,086,748	251	251	251	93.81	91.4	32.26	No
ERR2889320_1.fastq	5,667,506	855,793,406	151	151	151	96.39	90.97	31.53	Yes

ERR2889320_2.fastq	5,667,506	855,793,406	151	151	151	84.48	69.11	31.13	Yes
SRR26320745_1.fastq	5,667,506	855,793,406	151	151	151	96.39	90.97	31.53	Yes
SRR26320745_2.fastq	5,667,506	855,793,406	151	151	151	84.48	69.11	31.13	Yes
SRR10363485_1.fastq	5,685,948	1,427,172,948	251	251	251	96.04	94.2	40.06	No
SRR10363485_2.fastq	5,685,948	1,427,172,948	251	251	251	91.06	87.58	40.21	No
SRR10363348_1.fastq	5,700,691	1,430,873,441	251	251	251	96.01	94.09	32.32	No
SRR10363348_2.fastq	5,700,691	1,430,873,441	251	251	251	89.8	85.83	32.37	No
SRR21763636_1.fastq	5,760,660	869,859,660	151	151	151	97.39	92.94	32.15	No
SRR21763636_2.fastq	5,760,660	869,859,660	151	151	151	95.94	89.76	31.96	No
ERR2889316_1.fastq	5,769,453	871,187,403	151	151	151	97.53	93.51	31.7	Yes
ERR2889316_2.fastq	5,769,453	871,187,403	151	151	151	89.71	79.37	31.91	Yes
SRR26320741_1.fastq	5,769,453	871,187,403	151	151	151	97.53	93.51	31.7	Yes
SRR26320741_2.fastq	5,769,453	871,187,403	151	151	151	89.71	79.37	31.91	Yes
ERR2889321_1.fastq	5,785,760	873,649,760	151	151	151	95.83	89.58	43.54	Yes
ERR2889321_2.fastq	5,785,760	873,649,760	151	151	151	86.19	72.75	44.44	Yes
SRR26320746_1.fastq	5,785,760	873,649,760	151	151	151	95.83	89.58	43.54	Yes
SRR26320746_2.fastq	5,785,760	873,649,760	151	151	151	86.19	72.75	44.44	Yes
SRR10363391_1.fastq	5,797,963	579,796,300	100	100	100	98.88	96.24	29.79	No
SRR10363391_2.fastq	5,797,963	579,796,300	100	100	100	84.93	77.77	30.14	No
SRR10363343_1.fastq	5,803,158	1,456,592,658	251	251	251	95.68	93.82	34.68	No
SRR10363343_2.fastq	5,803,158	1,456,592,658	251	251	251	89.72	85.9	34.84	No
SRR10363574_1.fastq	5,823,175	1,461,616,925	251	251	251	96.25	94.9	30.88	No
SRR10363574_2.fastq	5,823,175	1,461,616,925	251	251	251	92.38	89.83	30.94	No
SRR15694533_1.fastq	5,823,175	1,461,616,925	251	251	251	96.25	94.9	30.88	No
SRR15694533_2.fastq	5,823,175	1,461,616,925	251	251	251	92.38	89.83	30.94	No
SRR10363431_1.fastq	5,838,320	1,465,418,320	251	251	251	96.35	93.98	30.37	No
SRR10363431_2.fastq	5,838,320	1,465,418,320	251	251	251	90.77	85.4	30.4	No
SRR10363414_1.fastq	5,856,668	1,470,023,668	251	251	251	96.45	94.16	31.02	No
SRR10363414_2.fastq	5,856,668	1,470,023,668	251	251	251	92.35	87.82	31.01	No

SRR10363345_1.fastq	5,872,469	1,473,989,719	251	251	251	96.13	94.31	31.73	No
SRR10363345_2.fastq	5,872,469	1,473,989,719	251	251	251	90.41	86.59	31.79	No
SRR26320659_1.fastq	5,879,275	887,473,642	4	150.9	151	95.49	88.13	35.25	No
SRR26320659_2.fastq	5,879,275	887,679,120	4	151	151	96.18	89.88	35.24	No
SRR10363380_1.fastq	5,951,571	595,157,100	100	100	100	98.13	94.67	38.31	No
SRR10363380_2.fastq	5,951,571	595,157,100	100	100	100	93.73	89.54	38.3	No
SRR21763633_1.fastq	5,970,058	901,478,758	151	151	151	97.41	93.15	31.61	No
SRR21763633_2.fastq	5,970,058	901,478,758	151	151	151	96.65	91.15	31.57	No
SRR21763629_1.fastq	5,997,850	905,675,350	151	151	151	95.6	88.73	31.3	No
SRR21763629_2.fastq	5,997,850	905,675,350	151	151	151	96.67	90.82	31.23	No
SRR10363565_1.fastq	6,008,242	1,508,068,742	251	251	251	96.92	95.64	31.19	No
SRR10363565_2.fastq	6,008,242	1,508,068,742	251	251	251	93.53	91.17	31.21	No
SRR15694524_1.fastq	6,008,242	1,508,068,742	251	251	251	96.92	95.64	31.19	No
SRR15694524_2.fastq	6,008,242	1,508,068,742	251	251	251	93.53	91.17	31.21	No
ERR2889354_1.fastq	6,022,359	909,376,209	151	151	151	93.73	86.55	46.23	Yes
ERR2889354_2.fastq	6,022,359	909,376,209	151	151	151	84.7	71.64	46.89	Yes
SRR26320650_1.fastq	6,022,359	909,376,209	151	151	151	93.73	86.55	46.23	Yes
SRR26320650_2.fastq	6,022,359	909,376,209	151	151	151	84.7	71.64	46.89	Yes
SRR10363580_1.fastq	6,028,321	1,513,108,571	251	251	251	96.78	95.4	30.58	No
SRR10363580_2.fastq	6,028,321	1,513,108,571	251	251	251	92.54	89.76	30.56	No
SRR15694538_1.fastq	6,028,321	1,513,108,571	251	251	251	96.78	95.4	30.58	No
SRR15694538_2.fastq	6,028,321	1,513,108,571	251	251	251	92.54	89.76	30.56	No
SRR10363511_1.fastq	6,029,791	1,513,477,541	251	251	251	95.78	93.91	37.49	No
SRR10363511_2.fastq	6,029,791	1,513,477,541	251	251	251	90.93	87.53	37.57	No
SRR10363535_1.fastq	6,032,291	1,514,105,041	251	251	251	95.17	93.27	31.98	No
SRR10363535_2.fastq	6,032,291	1,514,105,041	251	251	251	89.39	85.64	32.16	No
SRR15694546_1.fastq	6,032,291	1,514,105,041	251	251	251	95.17	93.27	31.98	No
SRR15694546_2.fastq	6,032,291	1,514,105,041	251	251	251	89.39	85.64	32.16	No
ERR2889325_1.fastq	6,078,976	917,925,376	151	151	151	96.22	90.9	41.12	No

ERR2889325_2.fastq	6,078,976	917,925,376	151	151	151	88.17	76.41	41.63	No
ERR2889359_1.fastq	6,084,579	918,771,429	151	151	151	95.54	89.17	43.75	Yes
ERR2889359_2.fastq	6,084,579	918,771,429	151	151	151	89.55	78.47	44.22	Yes
SRR26320645_1.fastq	6,084,579	918,771,429	151	151	151	95.54	89.17	43.75	Yes
SRR26320645_2.fastq	6,084,579	918,771,429	151	151	151	89.55	78.47	44.22	Yes
SRR26320718_1.fastq	6,091,163	919,765,613	151	151	151	95.72	90.4	37.73	No
SRR26320718_2.fastq	6,091,163	919,765,613	151	151	151	94.61	87.84	37.78	No
SRR10363540_1.fastq	6,137,376	613,737,600	100	100	100	98.91	96.24	30.14	No
SRR10363540_2.fastq	6,137,376	613,737,600	100	100	100	93.97	90.22	30.26	No
SRR10363402_1.fastq	6,142,875	614,287,500	100	100	100	98.79	95.9	30.41	No
SRR10363402_2.fastq	6,142,875	614,287,500	100	100	100	81.54	73.77	30.84	No
ERR2889355_1.fastq	6,155,254	929,443,354	151	151	151	96.93	92.12	31.72	Yes
ERR2889355_2.fastq	6,155,254	929,443,354	151	151	151	89.99	79.31	32.12	Yes
SRR26320649_1.fastq	6,155,254	929,443,354	151	151	151	96.93	92.12	31.72	Yes
SRR26320649_2.fastq	6,155,254	929,443,354	151	151	151	89.99	79.31	32.12	Yes
ERR2889358_1.fastq	6,173,111	932,139,761	151	151	151	94.16	87.11	50.02	Yes
ERR2889358_2.fastq	6,173,111	932,139,761	151	151	151	87.05	74.85	50.49	Yes
SRR26320646_1.fastq	6,173,111	932,139,761	151	151	151	94.16	87.11	50.02	Yes
SRR26320646_2.fastq	6,173,111	932,139,761	151	151	151	87.05	74.85	50.49	Yes
SRR26320717_1.fastq	6,199,515	936,126,765	151	151	151	97.1	92.65	33.28	No
SRR26320717_2.fastq	6,199,515	936,126,765	151	151	151	96.03	90.25	33.35	No
SRR10363505_1.fastq	6,204,638	620,463,800	100	100	100	98.24	94.09	44.08	No
SRR10363505_2.fastq	6,204,638	620,463,800	100	100	100	91.97	86.41	44.21	No
SRR10363568_1.fastq	6,206,323	1,557,787,073	251	251	251	96.97	95.57	30.31	No
SRR10363568_2.fastq	6,206,323	1,557,787,073	251	251	251	93.14	90.43	30.21	No
SRR15694526_1.fastq	6,206,323	1,557,787,073	251	251	251	96.97	95.57	30.31	No
SRR15694526_2.fastq	6,206,323	1,557,787,073	251	251	251	93.14	90.43	30.21	No
SRR26320730_1.fastq	6,243,261	942,732,411	151	151	151	96.56	91.56	33.36	No
SRR26320730_2.fastq	6,243,261	942,732,411	151	151	151	95.35	88.99	33.6	No

SRR10363310_1.fastq	6,247,025	1,568,003,275	251	251	251	97.08	95.79	31.32	No
SRR10363310_2.fastq	6,247,025	1,568,003,275	251	251	251	92.86	90.16	31.36	No
SRR15694542_1.fastq	6,247,025	1,568,003,275	251	251	251	97.08	95.79	31.32	No
SRR15694542_2.fastq	6,247,025	1,568,003,275	251	251	251	92.86	90.16	31.36	No
SRR21763630_1.fastq	6,254,042	944,360,342	151	151	151	96.24	90.43	44.92	No
SRR21763630_2.fastq	6,254,042	944,360,342	151	151	151	96.87	91.65	44.93	No
SRR10363536_1.fastq	6,263,922	1,572,244,422	251	251	251	95.11	93.23	31.69	No
SRR10363536_2.fastq	6,263,922	1,572,244,422	251	251	251	89.64	85.92	31.91	No
SRR15694547_1.fastq	6,263,922	1,572,244,422	251	251	251	95.11	93.23	31.69	No
SRR15694547_2.fastq	6,263,922	1,572,244,422	251	251	251	89.64	85.92	31.91	No
SRR10363413_1.fastq	6,290,630	629,063,000	100	100	100	98.97	96.42	30.09	No
SRR10363413_2.fastq	6,290,630	629,063,000	100	100	100	85.59	78.36	30.39	No
ERR2889312_1.fastq	6,296,377	950,752,927	151	151	151	96.6	91.37	31.24	Yes
ERR2889312_2.fastq	6,296,377	950,752,927	151	151	151	84.89	69.72	30.84	Yes
SRR26320644_1.fastq	6,296,377	950,752,927	151	151	151	96.6	91.37	31.24	Yes
SRR26320644_2.fastq	6,296,377	950,752,927	151	151	151	84.89	69.72	30.84	Yes
SRR10363537_1.fastq	6,298,131	1,580,830,881	251	251	251	95.7	93.87	31.01	No
SRR10363537_2.fastq	6,298,131	1,580,830,881	251	251	251	90.83	87.39	31.02	No
SRR15694548_1.fastq	6,298,131	1,580,830,881	251	251	251	95.7	93.87	31.01	No
SRR15694548_2.fastq	6,298,131	1,580,830,881	251	251	251	90.83	87.39	31.02	No
SRR10363581_1.fastq	6,307,037	1,583,066,287	251	251	251	97.15	95.84	30.68	No
SRR10363581_2.fastq	6,307,037	1,583,066,287	251	251	251	93.03	90.3	30.67	No
SRR15694540_1.fastq	6,307,037	1,583,066,287	251	251	251	97.15	95.84	30.68	No
SRR15694540_2.fastq	6,307,037	1,583,066,287	251	251	251	93.03	90.3	30.67	No
SRR10363576_1.fastq	6,310,781	1,584,006,031	251	251	251	96.94	95.58	30.76	No
SRR10363576_2.fastq	6,310,781	1,584,006,031	251	251	251	92.84	90.1	30.65	No
SRR15694535_1.fastq	6,310,781	1,584,006,031	251	251	251	96.94	95.58	30.76	No
SRR15694535_2.fastq	6,310,781	1,584,006,031	251	251	251	92.84	90.1	30.65	No
ERR2889328_1.fastq	6,331,440	956,047,440	151	151	151	96.15	90.22	49.62	Yes

ERR2889328_2.fastq	6,331,440	956,047,440	151	151	151	86.31	72.62	50.47	Yes
SRR26320750_1.fastq	6,331,440	956,047,440	151	151	151	96.15	90.22	49.62	Yes
SRR26320750_2.fastq	6,331,440	956,047,440	151	151	151	86.31	72.62	50.47	Yes
SRR10363404_1.fastq	6,340,746	1,591,527,246	251	251	251	96.14	93.69	30.2	No
SRR10363404_2.fastq	6,340,746	1,591,527,246	251	251	251	90.81	85.52	30.21	No
SRR10363355_1.fastq	6,346,936	1,593,080,936	251	251	251	95.98	94.18	35.05	No
SRR10363355_2.fastq	6,346,936	1,593,080,936	251	251	251	91.37	88.06	35.14	No
ERR2889322_1.fastq	6,386,463	964,355,913	151	151	151	94.84	88.14	50.43	Yes
ERR2889322_2.fastq	6,386,463	964,355,913	151	151	151	86.97	74.19	50.81	Yes
SRR26320747_1.fastq	6,386,463	964,355,913	151	151	151	94.84	88.14	50.43	Yes
SRR26320747_2.fastq	6,386,463	964,355,913	151	151	151	86.97	74.19	50.81	Yes
SRR26320714_1.fastq	6,390,720	964,998,720	151	151	151	97.12	92.44	32.4	No
SRR26320714_2.fastq	6,390,720	964,998,720	151	151	151	95.17	88.49	32.32	No
SRR10363487_1.fastq	6,432,187	1,614,478,937	251	251	251	96.14	93.65	30.52	No
SRR10363487_2.fastq	6,432,187	1,614,478,937	251	251	251	91.37	86.29	30.53	No
SRR15694488_1.fastq	6,432,187	1,614,478,937	251	251	251	96.14	93.65	30.52	No
SRR15694488_2.fastq	6,432,187	1,614,478,937	251	251	251	91.37	86.29	30.53	No
ERR2889343_1.fastq	6,463,616	976,006,016	151	151	151	96.67	91.59	32.2	Yes
ERR2889343_2.fastq	6,463,616	976,006,016	151	151	151	91.32	81.84	32.32	Yes
SRR26320656_1.fastq	6,463,616	976,006,016	151	151	151	96.67	91.59	32.2	Yes
SRR26320656_2.fastq	6,463,616	976,006,016	151	151	151	91.32	81.84	32.32	Yes
SRR26320660_1.fastq	6,531,458	986,079,581	1	151	151	97.47	92.83	32.96	No
SRR26320660_2.fastq	6,531,458	986,200,313	8	151	151	96.74	91.23	33.02	No
SRR15694490_1.fastq	6,531,719	1,639,461,469	251	251	251	95.98	94.16	31.09	No
SRR15694490_2.fastq	6,531,719	1,639,461,469	251	251	251	91.05	87.57	31.16	No
SRR21763625_1.fastq	6,533,136	986,503,536	151	151	151	97.54	93.16	31.67	No
SRR21763625_2.fastq	6,533,136	986,503,536	151	151	151	95.98	89.9	31.69	No
SRR3473975_1.fastq	6,537,248	660,262,048	101	101	101	95.8	89.87	32.12	No
SRR3473975_2.fastq	6,537,248	660,262,048	101	101	101	95.94	90.71	32.18	No

ERR2889326_1.fastq	6,545,791	988,414,441	151	151	151	96.85	92.03	32.95	No
ERR2889326_2.fastq	6,545,791	988,414,441	151	151	151	91.38	81.71	33.22	No
ERR2889313_1.fastq	6,553,989	989,652,339	151	151	151	97.46	93.38	31.05	No
ERR2889313_2.fastq	6,553,989	989,652,339	151	151	151	91.54	82.28	31.18	No
SRR21763628_1.fastq	6,601,088	996,764,288	151	151	151	96.97	91.65	32.53	No
SRR21763628_2.fastq	6,601,088	996,764,288	151	151	151	96.64	91.01	32.65	No
SRR6813719_1.fastq	6,627,377	1,000,733,927	151	151	151	90.13	85.42	38.18	No
SRR6813719_2.fastq	6,627,377	1,000,733,927	151	151	151	95.19	91.81	38.74	No
SRR10363340_1.fastq	6,667,410	1,673,519,910	251	251	251	96.36	94.57	31.74	No
SRR10363340_2.fastq	6,667,410	1,673,519,910	251	251	251	90.73	87.04	31.79	No
SRR7898459_1.fastq	6,711,982	739,605,581	40	110.2	143	99.81	98.23	34.81	No
SRR7898459_2.fastq	6,711,982	734,320,586	40	109.4	142	99.75	97.95	34.79	No
SRR26320705_1.fastq	6,716,643	1,014,213,093	151	151	151	96.99	92.29	33.49	No
SRR26320705_2.fastq	6,716,643	1,014,213,093	151	151	151	95.64	89.58	33.59	No
SRR10363336_1.fastq	6,717,693	1,686,140,943	251	251	251	95.84	93.88	32.9	No
SRR10363336_2.fastq	6,717,693	1,686,140,943	251	251	251	89.89	85.98	32.89	No
SRR21763635_1.fastq	6,725,209	1,015,506,559	151	151	151	97.51	93.09	37.26	No
SRR21763635_2.fastq	6,725,209	1,015,506,559	151	151	151	96.19	90.39	37.34	No
SRR10363339_1.fastq	6,763,402	1,697,613,902	251	251	251	96.05	94.26	31.35	No
SRR10363339_2.fastq	6,763,402	1,697,613,902	251	251	251	90.83	87.32	31.43	No
ERR2889327_1.fastq	6,805,145	1,027,576,895	151	151	151	96.16	90.73	40.65	No
ERR2889327_2.fastq	6,805,145	1,027,576,895	151	151	151	88.79	77.58	41.08	No
SRR10363313_1.fastq	6,811,725	1,709,742,975	251	251	251	97.17	95.81	31.1	Yes
SRR10363313_2.fastq	6,811,725	1,709,742,975	251	251	251	92.8	89.95	31.01	Yes
SRR15694544_1.fastq	6,811,725	1,709,742,975	251	251	251	97.17	95.81	31.1	Yes
SRR15694544_2.fastq	6,811,725	1,709,742,975	251	251	251	92.8	89.95	31.01	Yes
SRR15694497_1.fastq	6,843,548	684,354,800	100	100	100	98.74	95.68	35.17	No
SRR15694497_2.fastq	6,843,548	684,354,800	100	100	100	95.3	91.63	35.14	No
SRR10363344_1.fastq	6,863,473	1,722,731,723	251	251	251	96.15	94.34	32.37	No

SRR10363344_2.fastq	6,863,473	1,722,731,723	251	251	251	90.8	87.2	32.33	No
SRR26320712_1.fastq	6,896,605	1,041,387,355	151	151	151	97.29	92.54	31.31	No
SRR26320712_2.fastq	6,896,605	1,041,387,355	151	151	151	95.68	89.41	31.36	No
SRR10363312_1.fastq	6,927,735	1,738,861,485	251	251	251	97.14	95.85	30.88	Yes
SRR10363312_2.fastq	6,927,735	1,738,861,485	251	251	251	93.03	90.31	30.92	Yes
SRR15694543_1.fastq	6,927,735	1,738,861,485	251	251	251	97.14	95.85	30.88	Yes
SRR15694543_2.fastq	6,927,735	1,738,861,485	251	251	251	93.03	90.31	30.92	Yes
SRR10363304_1.fastq	7,020,482	1,762,140,982	251	251	251	96.16	94.33	31.25	No
SRR10363304_2.fastq	7,020,482	1,762,140,982	251	251	251	91.37	87.92	31.32	No
ERR2889356_1.fastq	7,021,541	1,060,252,691	151	151	151	96.88	92.47	32.83	Yes
ERR2889356_2.fastq	7,021,541	1,060,252,691	151	151	151	91.33	82.09	32.96	Yes
SRR26320648_1.fastq	7,021,541	1,060,252,691	151	151	151	96.88	92.47	32.83	Yes
SRR26320648_2.fastq	7,021,541	1,060,252,691	151	151	151	91.33	82.09	32.96	Yes
ERR2889357_1.fastq	7,034,562	1,062,218,862	151	151	151	95.77	90.55	35.98	Yes
ERR2889357_2.fastq	7,034,562	1,062,218,862	151	151	151	90.09	80.34	35.97	Yes
SRR26320647_1.fastq	7,034,562	1,062,218,862	151	151	151	95.77	90.55	35.98	Yes
SRR26320647_2.fastq	7,034,562	1,062,218,862	151	151	151	90.09	80.34	35.97	Yes
ERR2889346_1.fastq	7,051,385	1,064,759,135	151	151	151	95.97	90.42	35.68	Yes
ERR2889346_2.fastq	7,051,385	1,064,759,135	151	151	151	90.95	81.34	35.87	Yes
SRR26320652_1.fastq	7,051,385	1,064,759,135	151	151	151	95.97	90.42	35.68	Yes
SRR26320652_2.fastq	7,051,385	1,064,759,135	151	151	151	90.95	81.34	35.87	Yes
ERR2889347_1.fastq	7,142,829	1,078,567,179	151	151	151	97.56	93.56	31.62	Yes
ERR2889347_2.fastq	7,142,829	1,078,567,179	151	151	151	92.33	83.58	31.71	Yes
SRR26320651_1.fastq	7,142,829	1,078,567,179	151	151	151	97.56	93.56	31.62	Yes
SRR26320651_2.fastq	7,142,829	1,078,567,179	151	151	151	92.33	83.58	31.71	Yes
SRR10363506_1.fastq	7,148,969	1,794,391,219	251	251	251	96.37	94.03	30.67	No
SRR10363506_2.fastq	7,148,969	1,794,391,219	251	251	251	92.7	88.35	30.69	No
ERR2889345_1.fastq	7,161,905	1,081,447,655	151	151	151	97.44	93.32	31.18	Yes
ERR2889345_2.fastq	7,161,905	1,081,447,655	151	151	151	90.27	80.15	31.41	Yes

SRR26320653_1.fastq	7,161,905	1,081,447,655	151	151	151	97.44	93.32	31.18	Yes
SRR26320653_2.fastq	7,161,905	1,081,447,655	151	151	151	90.27	80.15	31.41	Yes
SRR10363424_1.fastq	7,181,749	718,174,900	100	100	100	99.01	96.49	30.45	No
SRR10363424_2.fastq	7,181,749	718,174,900	100	100	100	96.15	93.12	30.35	No
SRR15694493_1.fastq	7,210,238	721,023,800	100	100	100	98.92	96.24	30.66	No
SRR15694493_2.fastq	7,210,238	721,023,800	100	100	100	95.68	92.34	30.65	No
SRR26320733_1.fastq	7,217,667	1,089,867,717	151	151	151	96.87	91.43	31.51	No
SRR26320733_2.fastq	7,217,667	1,089,867,717	151	151	151	96.93	91.64	31.61	No
ERR2889317_1.fastq	7,229,384	1,091,636,984	151	151	151	97.47	93.44	31.42	Yes
ERR2889317_2.fastq	7,229,384	1,091,636,984	151	151	151	90.56	80.61	31.6	Yes
SRR26320742_1.fastq	7,229,384	1,091,636,984	151	151	151	97.47	93.44	31.42	Yes
SRR26320742_2.fastq	7,229,384	1,091,636,984	151	151	151	90.56	80.61	31.6	Yes
SRR10363300_1.fastq	7,232,625	1,815,388,875	251	251	251	95.89	94.14	32.48	No
SRR10363300_2.fastq	7,232,625	1,815,388,875	251	251	251	91.28	88.07	32.64	No
SRR10363306_1.fastq	7,299,676	1,832,218,676	251	251	251	96.61	94.82	34.09	No
SRR10363306_2.fastq	7,299,676	1,832,218,676	251	251	251	91.86	88.41	34.12	No
SRR10363417_1.fastq	7,331,951	1,840,319,701	251	251	251	96.38	94.02	32.44	No
SRR10363417_2.fastq	7,331,951	1,840,319,701	251	251	251	92.63	88.22	32.45	No
ERR2889342_1.fastq	7,344,184	1,108,971,784	151	151	151	96.8	92.34	33.25	Yes
ERR2889342_2.fastq	7,344,184	1,108,971,784	151	151	151	90.73	81.13	33.37	Yes
SRR26320657_1.fastq	7,344,184	1,108,971,784	151	151	151	96.8	92.34	33.25	Yes
SRR26320657_2.fastq	7,344,184	1,108,971,784	151	151	151	90.73	81.13	33.37	Yes
SRR26320734_1.fastq	7,413,253	1,119,401,203	151	151	151	97.51	93.22	32.54	No
SRR26320734_2.fastq	7,413,253	1,119,401,203	151	151	151	96.59	91.11	32.4	No
SRR26320708_1.fastq	7,473,499	1,128,498,349	151	151	151	90.59	83.84	50.07	No
SRR26320708_2.fastq	7,473,499	1,128,498,349	151	151	151	91.55	85.27	50.4	No
SRR15694485_1.fastq	7,510,819	751,081,900	100	100	100	98.64	95.49	33.57	No
SRR15694485_2.fastq	7,510,819	751,081,900	100	100	100	94.52	90.48	33.54	No
SRR10363314_1.fastq	7,528,877	1,889,748,127	251	251	251	85.05	81.3	48.26	Yes

SRR10363314_2.fastq	7,528,877	1,889,748,127	251	251	251	77.97	73.14	49.49	Yes
SRR15694521_1.fastq	7,528,877	1,889,748,127	251	251	251	85.05	81.3	48.26	
SRR15694521_2.fastq	7,528,877	1,889,748,127	251	251	251	77.97	73.14	49.49	
SRR10363308_1.fastq	7,572,778	1,900,767,278	251	251	251	96.48	94.8	31.22	No
SRR10363308_2.fastq	7,572,778	1,900,767,278	251	251	251	92.7	89.72	31.22	No
SRR26320719_1.fastq	7,579,774	1,144,545,874	151	151	151	94.59	86.44	31.53	No
SRR26320719_2.fastq	7,579,774	1,144,545,874	151	151	151	97.46	92.61	31.34	No
ERR2889324_1.fastq	7,581,839	1,144,857,689	151	151	151	96.82	91.92	31.03	Yes
ERR2889324_2.fastq	7,581,839	1,144,857,689	151	151	151	92.62	83.99	31.09	Yes
SRR26320749_1.fastq	7,581,839	1,144,857,689	151	151	151	96.82	91.92	31.03	Yes
SRR26320749_2.fastq	7,581,839	1,144,857,689	151	151	151	92.62	83.99	31.09	Yes
SRR10363305_1.fastq	7,600,050	1,907,612,550	251	251	251	96.2	94.38	31.55	No
SRR10363305_2.fastq	7,600,050	1,907,612,550	251	251	251	91.3	87.82	31.57	No
ERR2889318_1.fastq	7,648,626	1,154,942,526	151	151	151	97.53	93.5	31.95	Yes
ERR2889318_2.fastq	7,648,626	1,154,942,526	151	151	151	92.13	82.94	32.07	Yes
SRR26320743_1.fastq	7,648,626	1,154,942,526	151	151	151	97.53	93.5	31.95	Yes
SRR26320743_2.fastq	7,648,626	1,154,942,526	151	151	151	92.13	82.94	32.07	Yes
SRR21763632_1.fastq	7,652,926	1,155,591,826	151	151	151	97.37	92.72	31.49	No
SRR21763632_2.fastq	7,652,926	1,155,591,826	151	151	151	96.52	90.92	31.53	No
SRR10363557_1.fastq	7,685,020	1,928,940,020	251	251	251	96.47	94.24	32.12	No
SRR10363557_2.fastq	7,685,020	1,928,940,020	251	251	251	92.51	88.22	32.25	No
SRR26320713_1.fastq	7,737,259	1,168,326,109	151	151	151	97.49	93.04	31.44	No
SRR26320713_2.fastq	7,737,259	1,168,326,109	151	151	151	95.88	89.73	31.5	No
ERR2889344_1.fastq	7,743,499	1,169,268,349	151	151	151	97.54	93.56	30.93	Yes
ERR2889344_2.fastq	7,743,499	1,169,268,349	151	151	151	92.2	83.42	31	Yes
SRR26320655_1.fastq	7,743,499	1,169,268,349	151	151	151	97.54	93.56	30.93	Yes
SRR26320655_2.fastq	7,743,499	1,169,268,349	151	151	151	92.2	83.42	31	Yes
ERR2889315_1.fastq	7,751,256	1,170,439,656	151	151	151	96.64	92.03	35.45	No
ERR2889315_2.fastq	7,751,256	1,170,439,656	151	151	151	86.69	73	35.29	No

SRR26320699_1.fastq	7,790,810	1,176,412,310	151	151	151	96.97	91.93	36.9	No
SRR26320699_2.fastq	7,790,810	1,176,412,310	151	151	151	95.34	89.09	37.27	No
SRR21763634_1.fastq	7,816,964	1,180,361,564	151	151	151	97.55	93.17	31.31	No
SRR21763634_2.fastq	7,816,964	1,180,361,564	151	151	151	95.15	88.09	31.15	No
SRR21763624_1.fastq	7,852,503	1,185,727,953	151	151	151	97.28	92.62	31.76	No
SRR21763624_2.fastq	7,852,503	1,185,727,953	151	151	151	96.6	91.04	31.91	No
ERR2889323_1.fastq	7,927,596	1,197,066,996	151	151	151	97.56	93.6	31.19	Yes
ERR2889323_2.fastq	7,927,596	1,197,066,996	151	151	151	92.06	83.18	31.3	Yes
SRR26320748_1.fastq	7,927,596	1,197,066,996	151	151	151	97.56	93.6	31.19	Yes
SRR26320748_2.fastq	7,927,596	1,197,066,996	151	151	151	92.06	83.18	31.3	Yes
SRR10363302_1.fastq	8,014,576	2,011,658,576	251	251	251	96.27	94.51	32.17	No
SRR10363302_2.fastq	8,014,576	2,011,658,576	251	251	251	92.02	88.88	32.13	No
SRR10363527_1.fastq	8,054,739	2,021,739,489	251	251	251	94.55	91.88	40.99	No
SRR10363527_2.fastq	8,054,739	2,021,739,489	251	251	251	86.26	81.39	41.29	No
SRR15694504_1.fastq	8,079,040	2,027,839,040	251	251	251	93.26	90.15	30.63	No
SRR15694504_2.fastq	8,079,040	2,027,839,040	251	251	251	83.67	79.15	29.19	No
SRR26320685_1.fastq	8,103,743	1,223,665,193	151	151	151	97.65	93.61	36.7	No
SRR26320685_2.fastq	8,103,743	1,223,665,193	151	151	151	94.96	88.35	35.88	No
SRR21763626_1.fastq	8,188,554	1,236,471,654	151	151	151	97.4	92.72	33.82	No
SRR21763626_2.fastq	8,188,554	1,236,471,654	151	151	151	96.62	91.08	33.77	No
SRR26320710_1.fastq	8,219,199	1,241,099,049	151	151	151	95.75	90.97	43.14	No
SRR26320710_2.fastq	8,219,199	1,241,099,049	151	151	151	95.09	89.22	43.25	No
SRR26320729_1.fastq	8,221,829	1,241,496,179	151	151	151	96.27	91.65	39.59	No
SRR26320729_2.fastq	8,221,829	1,241,496,179	151	151	151	95.44	89.6	39.58	No
SRR26320728_1.fastq	8,235,357	1,243,538,907	151	151	151	94.15	88.91	49.65	No
SRR26320728_2.fastq	8,235,357	1,243,538,907	151	151	151	93.38	87.25	49.91	No
SRR10363301_1.fastq	8,272,583	2,076,418,333	251	251	251	95.84	94.15	34.48	No
SRR10363301_2.fastq	8,272,583	2,076,418,333	251	251	251	92	89.1	34.68	No
SRR10363519_1.fastq	8,299,160	2,083,089,160	251	251	251	94.98	92.54	31.4	No

SRR10363519_2.fastq	8,299,160	2,083,089,160	251	251	251	87.66	82.9	31.56	No
SRR26320706_1.fastq	8,305,085	1,254,067,835	151	151	151	97.21	92.69	32.88	No
SRR26320706_2.fastq	8,305,085	1,254,067,835	151	151	151	96.4	90.82	32.85	No
SRR21763627_1.fastq	8,324,349	1,256,976,699	151	151	151	97.66	93.45	36.29	No
SRR21763627_2.fastq	8,324,349	1,256,976,699	151	151	151	96.83	91.59	36.27	No
SRR10363598_1.fastq	8,337,429	2,092,694,679	251	251	251	96.48	94.84	30.99	No
SRR10363598_2.fastq	8,337,429	2,092,694,679	251	251	251	91.5	88.17	31.11	No
SRR10363526_1.fastq	8,398,126	2,107,929,626	251	251	251	95.26	92.84	34.91	No
SRR10363526_2.fastq	8,398,126	2,107,929,626	251	251	251	87.77	83.01	35.06	No
SRR6147472_1.fastq	8,427,141	1,118,069,738	35	132.7	151	99.07	96.99	36.35	No
SRR6147472_2.fastq	8,427,141	1,117,606,934	35	132.6	151	97.29	93.43	36.45	No
SRR26320711_1.fastq	8,513,705	1,285,569,455	151	151	151	87.49	79.86	51.76	No
SRR26320711_2.fastq	8,513,705	1,285,569,455	151	151	151	88.41	80.83	51.61	No
SRR26320716_1.fastq	8,519,657	1,286,468,207	151	151	151	94.29	88.89	45.28	No
SRR26320716_2.fastq	8,519,657	1,286,468,207	151	151	151	93.77	87.52	45.59	No
SRR26320735_1.fastq	8,811,943	1,330,603,393	151	151	151	96.66	91.86	38.85	No
SRR26320735_2.fastq	8,811,943	1,330,603,393	151	151	151	96.19	90.58	39	No
SRR10363516_1.fastq	8,826,695	2,215,500,445	251	251	251	94.71	92.28	32.5	No
SRR10363516_2.fastq	8,826,695	2,215,500,445	251	251	251	88.33	84.03	32.55	No
SRR15694489_1.fastq	8,892,307	889,230,700	100	100	100	98.93	96.33	30.02	No
SRR15694489_2.fastq	8,892,307	889,230,700	100	100	100	95.8	92.73	30.01	No
SRR26320707_1.fastq	8,940,009	1,349,941,359	151	151	151	92.22	86.03	46.04	No
SRR26320707_2.fastq	8,940,009	1,349,941,359	151	151	151	92.52	86.62	46.48	No
SRR10363395_1.fastq	8,944,317	2,245,023,567	251	251	251	96.22	93.85	31.43	No
SRR10363395_2.fastq	8,944,317	2,245,023,567	251	251	251	91.2	86.19	31.55	No
SRR10363531_1.fastq	8,975,160	2,252,765,160	251	251	251	94.92	92.57	34.47	No
SRR10363531_2.fastq	8,975,160	2,252,765,160	251	251	251	88.79	84.46	34.65	No
SRR10363533_1.fastq	9,033,767	2,267,475,517	251	251	251	95.48	93.12	36.54	No
SRR10363533_2.fastq	9,033,767	2,267,475,517	251	251	251	89.59	85.23	36.66	No

SRR10363484_1.fastq	9,049,279	2,271,369,029	251	251	251	95.05	92.39	43.5	No
SRR10363484_2.fastq	9,049,279	2,271,369,029	251	251	251	88.79	84.16	43.78	No
SRR26320731_1.fastq	9,133,609	1,379,174,959	151	151	151	97.13	92.27	31.84	No
SRR26320731_2.fastq	9,133,609	1,379,174,959	151	151	151	96.83	91.59	31.69	No
SRR21763631_1.fastq	9,137,329	1,379,736,679	151	151	151	97.6	93.38	34.82	No
SRR21763631_2.fastq	9,137,329	1,379,736,679	151	151	151	96.5	90.86	34.75	No
SRR26320737_1.fastq	9,217,412	1,391,829,212	151	151	151	96.16	91.35	43.47	No
SRR26320737_2.fastq	9,217,412	1,391,829,212	151	151	151	95.5	89.8	43.47	No
SRR10363514_1.fastq	9,226,550	2,315,864,050	251	251	251	94.62	92.19	31.92	No
SRR10363514_2.fastq	9,226,550	2,315,864,050	251	251	251	86.95	82.14	32.1	No
SRR26320701_1.fastq	9,241,274	1,395,432,374	151	151	151	96.64	90.94	37.91	No
SRR26320701_2.fastq	9,241,274	1,395,432,374	151	151	151	96.07	90.44	37.59	No
SRR26320709_1.fastq	9,290,504	1,402,866,104	151	151	151	96.26	91.52	39.91	No
SRR26320709_2.fastq	9,290,504	1,402,866,104	151	151	151	95.55	89.91	39.84	No
SRR26320700_1.fastq	9,308,323	1,405,556,773	151	151	151	95.99	90.75	37.25	No
SRR26320700_2.fastq	9,308,323	1,405,556,773	151	151	151	95.89	89.99	37.43	No
SRR6147581_1.fastq	9,342,641	1,288,405,261	35	137.9	151	99.16	97.15	36.31	No
SRR6147581_2.fastq	9,342,641	1,287,673,075	35	137.8	151	97.62	94.07	36.38	No
SRR26320715_1.fastq	9,393,758	1,418,457,458	151	151	151	98.32	95.08	58.96	No
SRR26320715_2.fastq	9,393,758	1,418,457,458	151	151	151	94.52	88.84	59.54	No
SRR10363429_1.fastq	9,526,359	2,391,116,109	251	251	251	96.24	93.83	30.55	No
SRR10363429_2.fastq	9,526,359	2,391,116,109	251	251	251	90.79	85.45	30.6	No
SRR10363530_1.fastq	9,552,706	2,397,729,206	251	251	251	95.47	93.28	31.19	No
SRR10363530_2.fastq	9,552,706	2,397,729,206	251	251	251	89.32	85.14	31.24	No
SRR10363521_1.fastq	9,607,779	2,411,552,529	251	251	251	94.91	92.57	37.01	No
SRR10363521_2.fastq	9,607,779	2,411,552,529	251	251	251	89.17	85.19	37.13	No
SRR10363584_1.fastq	9,820,359	2,464,910,109	251	251	251	95.42	93.28	33.29	No
SRR10363584_2.fastq	9,820,359	2,464,910,109	251	251	251	90.72	87.06	33.38	No
SRR26320696_1.fastq	9,827,125	1,483,895,875	151	151	151	97.46	93.35	37.29	No

SRR26320696_2.fastq	9,827,125	1,483,895,875	151	151	151	95.99	90.29	36.99	No
SRR26320702_1.fastq	9,891,461	1,493,610,611	151	151	151	97.29	92.87	37.53	No
SRR26320702_2.fastq	9,891,461	1,493,610,611	151	151	151	96.03	90.3	36.84	No
SRR15694500_1.fastq	10,099,921	2,535,080,171	251	251	251	93.06	89.88	31	No
SRR15694500_2.fastq	10,099,921	2,535,080,171	251	251	251	71.62	63.84	29.94	No
SRR10363561_1.fastq	10,130,341	2,542,715,591	251	251	251	95.5	93.25	31.51	No
SRR10363561_2.fastq	10,130,341	2,542,715,591	251	251	251	90.07	86	31.64	No
SRR10363518_1.fastq	10,151,259	2,547,966,009	251	251	251	94.89	92.51	31.47	No
SRR10363518_2.fastq	10,151,259	2,547,966,009	251	251	251	88.17	83.54	31.62	No
SRR26320697_1.fastq	10,290,343	1,553,841,793	151	151	151	96.7	91.08	37.02	No
SRR26320697_2.fastq	10,290,343	1,553,841,793	151	151	151	95.81	89.97	37.59	No
SRR15694503_1.fastq	10,382,186	2,605,928,686	251	251	251	92.99	89.85	30.76	No
SRR15694503_2.fastq	10,382,186	2,605,928,686	251	251	251	81.71	76.69	29.34	No
SRR16990213_1.fastq	10,492,905	1,584,428,655	151	151	151	98.36	94.94	30.82	No
SRR16990213_2.fastq	10,492,905	1,584,428,655	151	151	151	96.42	91.13	31	No
SRR10363562_1.fastq	10,497,667	2,634,914,417	251	251	251	95.44	93.25	31.92	No
SRR10363562_2.fastq	10,497,667	2,634,914,417	251	251	251	90.26	86.37	32.05	No
SRR10363320_1.fastq	10,568,769	2,652,761,019	251	251	251	93.28	90.18	30.57	No
SRR10363320_2.fastq	10,568,769	2,652,761,019	251	251	251	82.59	77.74	29.16	No
SRR10363525_1.fastq	10,590,120	2,658,120,120	251	251	251	95.19	92.88	31.07	No
SRR10363525_2.fastq	10,590,120	2,658,120,120	251	251	251	88.04	83.29	31.2	No
SRR10336412_1.fastq	10,606,628	2,662,263,628	251	251	251	96.43	95.1	30.63	No
SRR10336412_2.fastq	10,606,628	2,662,263,628	251	251	251	93	90.54	30.64	No
SRR15694486_1.fastq	10,620,834	2,665,829,334	251	251	251	93.73	90.86	31.29	No
SRR15694486_2.fastq	10,620,834	2,665,829,334	251	251	251	83.87	79.58	29.85	No
SRR10363583_1.fastq	10,627,080	2,667,397,080	251	251	251	95.73	93.6	31.45	No
SRR10363583_2.fastq	10,627,080	2,667,397,080	251	251	251	90.76	86.9	31.54	No
SRR26320725_1.fastq	10,646,785	1,607,664,535	151	151	151	98.25	94.81	59.18	No
SRR26320725_2.fastq	10,646,785	1,607,664,535	151	151	151	94.47	88.61	60.12	No

SRR10363532_1.fastq	10,746,062	2,697,261,562	251	251	251	95.51	93.33	32.41	No
SRR10363532_2.fastq	10,746,062	2,697,261,562	251	251	251	90.54	86.71	32.44	No
SRR26320732_1.fastq	10,749,815	1,623,222,065	151	151	151	96.39	90.25	31.31	No
SRR26320732_2.fastq	10,749,815	1,623,222,065	151	151	151	97.33	92.41	31.21	No
SRR15694487_1.fastq	10,776,043	2,704,786,793	251	251	251	93.17	90.1	30.68	No
SRR15694487_2.fastq	10,776,043	2,704,786,793	251	251	251	82.1	77.2	29.15	No
SRR26320695_1.fastq	10,995,504	1,660,321,104	151	151	151	97.07	92.07	37.38	No
SRR26320695_2.fastq	10,995,504	1,660,321,104	151	151	151	96	90.15	36.81	No
ERR1760143_1.fastq	11,062,724	564,198,924	51	51	51	79.94	45.74	32.98	No
ERR1760143_2.fastq	11,062,724	564,198,924	51	51	51	70.69	43.61	32.23	No
SRR6148259_1.fastq	11,161,344	1,393,341,808	35	124.8	150	99.07	97.08	36.93	No
SRR6148259_2.fastq	11,161,344	1,393,056,903	35	124.8	150	97.83	94.69	37.02	No
SRR15694496_1.fastq	11,287,905	2,833,264,155	251	251	251	92.88	89.74	30.5	No
SRR15694496_2.fastq	11,287,905	2,833,264,155	251	251	251	81.48	76.43	29.08	No
SRR10363321_1.fastq	11,336,308	2,845,413,308	251	251	251	93.19	90.05	30.84	No
SRR10363321_2.fastq	11,336,308	2,845,413,308	251	251	251	82.54	77.69	29.4	No
SRR10363515_1.fastq	11,427,486	2,868,298,986	251	251	251	95.22	92.98	31.07	No
SRR10363515_2.fastq	11,427,486	2,868,298,986	251	251	251	89.55	85.43	31.24	No
SRR26320654_1.fastq	11,627,345	1,755,729,095	151	151	151	97.66	93.41	51	No
SRR26320654_2.fastq	11,627,345	1,755,729,095	151	151	151	93.49	86.95	51.71	No
SRR10363520_1.fastq	11,833,310	2,970,160,810	251	251	251	95.19	92.94	31.11	No
SRR10363520_2.fastq	11,833,310	2,970,160,810	251	251	251	89.16	84.94	31.23	No
SRR10363318_1.fastq	11,847,764	2,973,788,764	251	251	251	92.83	89.65	30.18	Yes
SRR10363318_2.fastq	11,847,764	2,973,788,764	251	251	251	71.82	64.13	29.06	Yes
SRR15694491_1.fastq	11,847,764	2,973,788,764	251	251	251	92.83	89.65	30.18	Yes
SRR15694491_2.fastq	11,847,764	2,973,788,764	251	251	251	71.82	64.13	29.06	Yes
SRR26320703_1.fastq	11,872,705	1,792,778,455	151	151	151	98.17	94.79	56.93	No
SRR26320703_2.fastq	11,872,705	1,792,778,455	151	151	151	94.83	89.21	57.46	No
SRR26320698_1.fastq	11,877,761	1,793,541,911	151	151	151	97.95	93.96	59.75	No

SRR26320698_2.fastq	11,877,761	1,793,541,911	151	151	151	95.34	89.61	60.89	No
SRR10363523_1.fastq	11,882,513	2,982,510,763	251	251	251	95.08	92.82	31.29	No
SRR10363523_2.fastq	11,882,513	2,982,510,763	251	251	251	88.25	83.72	31.46	No
SRR26320704_1.fastq	11,918,167	1,799,643,217	151	151	151	98.25	94.97	59.86	No
SRR26320704_2.fastq	11,918,167	1,799,643,217	151	151	151	94.66	88.98	60.93	No
SRR10363357_1.fastq	11,943,462	2,997,808,962	251	251	251	95.53	93.17	36.65	No
SRR10363357_2.fastq	11,943,462	2,997,808,962	251	251	251	90.03	85.87	36.78	No
SRR10363522_1.fastq	11,955,988	3,000,952,988	251	251	251	93.68	91.49	32.47	No
SRR10363522_2.fastq	11,955,988	3,000,952,988	251	251	251	87.53	83.72	32.84	No
SRR10363509_1.fastq	12,068,805	3,029,270,055	251	251	251	95.1	92.77	30.57	No
SRR10363509_2.fastq	12,068,805	3,029,270,055	251	251	251	89.64	85.61	30.77	No
SRR10363317_1.fastq	12,134,128	3,045,666,128	251	251	251	93.38	90.41	30.86	No
SRR10363317_2.fastq	12,134,128	3,045,666,128	251	251	251	73.23	65.87	29.83	No
SRR11817810_1.fastq	12,174,705	1,826,205,750	150	150	150	98.13	95.07	30.66	No
SRR11817810_2.fastq	12,174,705	1,826,205,750	150	150	150	92.84	83.74	31.01	No
SRR26320738_1.fastq	12,211,503	1,843,936,953	151	151	151	97.96	94.31	61.63	No
SRR26320738_2.fastq	12,211,503	1,843,936,953	151	151	151	93.83	87.38	62.67	No
SRR10363560_1.fastq	12,436,036	1,256,039,636	101	101	101	99.43	97.9	30.6	Yes
SRR10363560_2.fastq	12,436,036	1,256,039,636	101	101	101	98.37	96.21	30.51	Yes
SRR15694499_1.fastq	12,436,036	1,256,039,636	101	101	101	99.43	97.9	30.6	Yes
SRR15694499_2.fastq	12,436,036	1,256,039,636	101	101	101	98.37	96.21	30.51	Yes
SRR26320752_1.fastq	12,713,974	1,919,810,074	151	151	151	96.3	89.6	60.8	No
SRR26320752_2.fastq	12,713,974	1,919,810,074	151	151	151	94.39	88.31	61.29	No
SRR26320688_1.fastq	12,815,743	1,935,177,193	151	151	151	97.61	93.28	31.22	No
SRR26320688_2.fastq	12,815,743	1,935,177,193	151	151	151	96.07	90.04	31.11	No
SRR11818076_1.fastq	12,914,021	1,937,103,150	150	150	150	97.8	94.28	30.99	No
SRR11818076_2.fastq	12,914,021	1,937,103,150	150	150	150	94.8	87.6	31.05	No
SRR10363529_1.fastq	13,368,702	3,355,544,202	251	251	251	95.58	93.4	31.19	No
SRR10363529_2.fastq	13,368,702	3,355,544,202	251	251	251	89.11	84.77	31.29	No

SRR15694505_1.fastq	13,401,087	3,363,672,837	251	251	251	93.24	90.17	30.84	No
SRR15694505_2.fastq	13,401,087	3,363,672,837	251	251	251	72.04	64.4	29.78	No
SRR26320691_1.fastq	13,540,163	2,044,564,613	151	151	151	97.33	92.89	31.03	No
SRR26320691_2.fastq	13,540,163	2,044,564,613	151	151	151	96.47	90.87	31.03	No
SRR15694520_1.fastq	13,585,650	2,037,847,500	150	150	150	97.62	93.38	40.78	No
SRR15694520_2.fastq	13,585,650	2,037,847,500	150	150	150	96.28	90.28	40.8	No
SRR15694552_1.fastq	13,940,559	3,485,139,750	250	250	250	96.25	89.9	32.12	No
SRR15694552_2.fastq	13,940,559	3,485,139,750	250	250	250	94.8	87.73	32.35	No
SRR11817812_1.fastq	13,977,210	2,096,581,500	150	150	150	98.18	95.17	30.74	No
SRR11817812_2.fastq	13,977,210	2,096,581,500	150	150	150	94.07	85.98	30.99	No
SRR3473977.fastq	14,188,762	1,418,876,200	100	100	100	96.58	90.99	31.62	No*
SRR11818077_1.fastq	14,220,558	2,133,083,700	150	150	150	98.12	94.99	33.48	No
SRR11818077_2.fastq	14,220,558	2,133,083,700	150	150	150	93.64	85.1	33.79	No
SRR11817821_1.fastq	14,421,155	2,163,173,250	150	150	150	98.17	95.15	31.22	No
SRR11817821_2.fastq	14,421,155	2,163,173,250	150	150	150	94.14	86.14	31.48	No
SRR10363512_1.fastq	14,502,831	3,640,210,581	251	251	251	95.59	93.47	31.55	No
SRR10363512_2.fastq	14,502,831	3,640,210,581	251	251	251	91.17	87.56	31.66	No
SRR26320689_1.fastq	14,516,843	2,192,043,293	151	151	151	95.79	88.81	31.43	No
SRR26320689_2.fastq	14,516,843	2,192,043,293	151	151	151	96.9	91.37	31.43	No
SRR10363316_1.fastq	14,734,476	3,698,353,476	251	251	251	93.43	90.48	31.03	No
SRR10363316_2.fastq	14,734,476	3,698,353,476	251	251	251	74.37	67.38	29.93	No
SRR10363319_1.fastq	14,833,191	3,723,130,941	251	251	251	93.5	90.49	29.86	No
SRR10363319_2.fastq	14,833,191	3,723,130,941	251	251	251	82.96	78.2	28.38	No
SRR11817811_1.fastq	15,252,354	2,287,853,100	150	150	150	98.19	95.19	30.65	No
SRR11817811_2.fastq	15,252,354	2,287,853,100	150	150	150	93.83	85.55	30.92	No
SRR26320667_1.fastq	15,272,747	2,306,184,797	151	151	151	97.94	94.04	36.61	No
SRR26320667_2.fastq	15,272,747	2,306,184,797	151	151	151	94	87.27	36.88	No
SRR26320690_1.fastq	15,393,153	2,324,366,103	151	151	151	97.11	92.04	30.9	No
SRR26320690_2.fastq	15,393,153	2,324,366,103	151	151	151	96.04	90.04	31.11	No

SRR26320694_1.fastq	15,460,192	2,334,488,992	151	151	151	97.02	91.99	31.3	No
SRR26320694_2.fastq	15,460,192	2,334,488,992	151	151	151	96.6	91.09	31.09	No
SRR26320676_1.fastq	15,872,378	2,396,729,078	151	151	151	98.44	95.4	53.92	No
SRR26320676_2.fastq	15,872,378	2,396,729,078	151	151	151	95.27	89.66	54.89	No
SRR26320739_1.fastq	16,192,386	2,445,050,286	151	151	151	97.44	93.73	56.07	No
SRR26320739_2.fastq	16,192,386	2,445,050,286	151	151	151	93.85	87.37	56.61	No
SRR11817809_1.fastq	16,555,164	2,483,274,600	150	150	150	98.13	95.08	30.34	No
SRR11817809_2.fastq	16,555,164	2,483,274,600	150	150	150	94.12	86.13	30.6	No
SRR26320692_1.fastq	16,931,892	2,556,715,692	151	151	151	97.37	92.7	31.16	No
SRR26320692_2.fastq	16,931,892	2,556,715,692	151	151	151	96.26	90.33	31.21	No
SRR26320693_1.fastq	17,134,071	2,587,244,721	151	151	151	97.48	92.91	31.05	No
SRR26320693_2.fastq	17,134,071	2,587,244,721	151	151	151	96.4	90.67	30.99	No
SRR26320686_1.fastq	17,211,332	2,598,911,132	151	151	151	97.46	92.93	31.31	No
SRR26320686_2.fastq	17,211,332	2,598,911,132	151	151	151	96.33	90.5	31.34	No
SRR26320721_1.fastq	17,450,131	2,634,969,781	151	151	151	97.71	93.54	38.99	No
SRR26320721_2.fastq	17,450,131	2,634,969,781	151	151	151	96.34	91.51	39.39	No
SRR26320669_1.fastq	17,512,095	2,644,326,345	151	151	151	98.51	95.77	39.12	No
SRR26320669_2.fastq	17,512,095	2,644,326,345	151	151	151	94.96	89.73	41.19	No
SRR26320724_1.fastq	17,513,300	2,644,508,300	151	151	151	98.31	94.73	43.04	No
SRR26320724_2.fastq	17,513,300	2,644,508,300	151	151	151	95.32	89.44	42.95	No
SRR26320720_1.fastq	17,946,193	2,709,875,143	151	151	151	97.96	93.65	40.8	No
SRR26320720_2.fastq	17,946,193	2,709,875,143	151	151	151	95.12	90.17	41.31	No
SRR11818075_1.fastq	18,284,283	2,742,642,450	150	150	150	97.85	94.4	30.94	No
SRR11818075_2.fastq	18,284,283	2,742,642,450	150	150	150	94.8	87.63	30.99	No
SRR3473976.fastq	18,907,631	1,890,763,100	100	100	100	94.95	87.5	31.99	No*
SRR26320675_1.fastq	19,219,587	2,902,157,637	151	151	151	98	94.17	36.76	No
SRR26320675_2.fastq	19,219,587	2,902,157,637	151	151	151	95.42	89.48	38.09	No
SRR26320722_1.fastq	19,321,594	2,917,560,694	151	151	151	97.88	93.33	38.81	No
SRR26320722_2.fastq	19,321,594	2,917,560,694	151	151	151	95.25	90.46	40.02	No

ERR2889314_1.fastq	19,441,124	2,935,609,724	151	151	151	95.62	90.51	45.74	No
ERR2889314_2.fastq	19,441,124	2,935,609,724	151	151	151	90.54	80.76	45.54	No
ERR2889319_1.fastq	19,565,261	2,954,354,411	151	151	151	96.61	92.07	47.75	Yes
ERR2889319_2.fastq	19,565,261	2,954,354,411	151	151	151	89.92	79.42	47.78	Yes
SRR26320744_1.fastq	19,565,261	2,954,354,411	151	151	151	96.61	92.07	47.75	Yes
SRR26320744_2.fastq	19,565,261	2,954,354,411	151	151	151	89.92	79.42	47.78	Yes
SRR11817820_1.fastq	19,709,731	2,956,459,650	150	150	150	98.12	95.07	31.91	No
SRR11817820_2.fastq	19,709,731	2,956,459,650	150	150	150	94.36	86.68	32.15	No
SRR26320674_1.fastq	19,784,818	2,987,507,518	151	151	151	98.12	94.44	36.65	No
SRR26320674_2.fastq	19,784,818	2,987,507,518	151	151	151	95.55	89.62	38.16	No
SRR11818074_1.fastq	19,901,755	2,985,263,250	150	150	150	97.82	94.34	30.77	No
SRR11818074_2.fastq	19,901,755	2,985,263,250	150	150	150	94.56	87.19	30.83	No
SRR11818078_1.fastq	20,016,478	3,022,488,178	151	151	151	98.23	95.44	30.58	No
SRR11818078_2.fastq	20,016,478	3,022,488,178	151	151	151	94.49	88.96	30.85	No
SRR11818073_1.fastq	20,961,997	3,144,299,550	150	150	150	98.18	95.18	30.93	No
SRR11818073_2.fastq	20,961,997	3,144,299,550	150	150	150	94.05	85.96	31.19	No
SRR3091751_1.fastq	21,037,979	3,122,219,698	35	148.4	185	94.96	93.13	42.38	No
SRR3091751_2.fastq	21,037,979	3,150,813,064	35	149.8	188	87.44	83.63	42.57	No
SRR11817815_1.fastq	21,071,357	3,181,774,907	151	151	151	98.18	95.34	31.93	No
SRR11817815_2.fastq	21,071,357	3,181,774,907	151	151	151	94.75	89.4	32.18	No
SRR26320672_1.fastq	21,115,871	3,188,496,521	151	151	151	96.47	92.16	37.8	No
SRR26320672_2.fastq	21,115,871	3,188,496,521	151	151	151	94.7	89.61	39.03	No
SRR26320726_1.fastq	21,205,856	3,202,084,256	151	151	151	97.23	93.59	38.45	No
SRR26320726_2.fastq	21,205,856	3,202,084,256	151	151	151	95.37	90.62	39.31	No
SRR26320723_1.fastq	21,338,987	3,222,187,037	151	151	151	96.3	93.12	40.22	No
SRR26320723_2.fastq	21,338,987	3,222,187,037	151	151	151	94.82	90.03	40.95	No
SRR26320753_1.fastq	21,714,989	3,278,963,339	151	151	151	97.95	94.38	44.27	No
SRR26320753_2.fastq	21,714,989	3,278,963,339	151	151	151	95.3	89.92	44.19	No
SRR11817814_1.fastq	21,800,766	3,291,915,666	151	151	151	98.22	95.45	30.39	No

SRR11817814_2.fastq	21,800,766	3,291,915,666	151	151	151	94.45	88.93	30.65	No
SRR11817816_1.fastq	21,990,980	3,320,637,980	151	151	151	98.23	95.46	30.44	No
SRR11817816_2.fastq	21,990,980	3,320,637,980	151	151	151	94.63	89.21	30.7	No
SRR26320671_1.fastq	22,580,015	3,409,582,265	151	151	151	97.91	94.09	38.16	No
SRR26320671_2.fastq	22,580,015	3,409,582,265	151	151	151	94.97	89.38	38.23	No
SRR26320678_1.fastq	22,630,267	3,417,170,317	151	151	151	98.04	94.45	37.46	No
SRR26320678_2.fastq	22,630,267	3,417,170,317	151	151	151	95.38	89.94	37.56	No
SRR26320740_1.fastq	22,785,112	3,440,551,912	151	151	151	97.52	93.72	47.77	No
SRR26320740_2.fastq	22,785,112	3,440,551,912	151	151	151	95.45	90.42	48.23	No
SRR11818079_1.fastq	22,899,339	3,457,800,189	151	151	151	98.02	95.1	34.25	No
SRR11818079_2.fastq	22,899,339	3,457,800,189	151	151	151	94.48	88.92	34.49	No
SRR11817813_1.fastq	22,924,088	3,461,537,288	151	151	151	98.22	95.44	30.52	No
SRR11817813_2.fastq	22,924,088	3,461,537,288	151	151	151	94.68	89.29	30.78	No
SRR26320682_1.fastq	22,968,504	3,468,244,104	151	151	151	97.52	92.8	37.64	No
SRR26320682_2.fastq	22,968,504	3,468,244,104	151	151	151	95.08	89.96	38.5	No
SRR26320751_1.fastq	23,233,203	3,508,213,653	151	151	151	97.94	94.16	38.46	No
SRR26320751_2.fastq	23,233,203	3,508,213,653	151	151	151	95.32	90.13	39.19	No
SRR26320727_1.fastq	23,408,887	3,534,741,937	151	151	151	98.48	95.4	38.7	No
SRR26320727_2.fastq	23,408,887	3,534,741,937	151	151	151	96.52	91.78	38.42	No
SRR26320666_1.fastq	23,941,726	3,615,200,626	151	151	151	97.99	94.05	38.75	No
SRR26320666_2.fastq	23,941,726	3,615,200,626	151	151	151	95.13	90.55	39.27	No
SRR11817823_1.fastq	24,235,010	3,659,486,510	151	151	151	98.2	95.42	30.92	No
SRR11817823_2.fastq	24,235,010	3,659,486,510	151	151	151	94.87	89.61	31.14	No
SRR26320681_1.fastq	24,517,262	3,702,106,562	151	151	151	98.04	94.38	37.7	No
SRR26320681_2.fastq	24,517,262	3,702,106,562	151	151	151	95.88	90.76	37.99	No
SRR26320687_1.fastq	24,716,940	3,732,257,940	151	151	151	96.23	92.39	38.21	No
SRR26320687_2.fastq	24,716,940	3,732,257,940	151	151	151	94.88	89.52	40.02	No
SRR26320662_1.fastq	24,795,861	3,744,175,011	151	151	151	97.87	93.98	38.57	No
SRR26320662_2.fastq	24,795,861	3,744,175,011	151	151	151	96.27	91.18	38.36	No

SRR26320670_1.fastq	25,653,845	3,873,730,595	151	151	151	98.21	94.9	37.42	No
SRR26320670_2.fastq	25,653,845	3,873,730,595	151	151	151	95.66	90.59	37.44	No
SRR6871415_1.fastq	26,447,619	6,373,374,908	35	241	299	96.29	91.58	35.04	No
SRR6871415_2.fastq	26,447,619	6,403,132,656	35	242.1	299	90.26	81.38	35.45	No
SRR26320673_1.fastq	27,219,642	4,110,165,942	151	151	151	97.29	93.2	37.29	No
SRR26320673_2.fastq	27,219,642	4,110,165,942	151	151	151	95.62	89.83	39.28	No
SRR26320680_1.fastq	27,254,170	4,115,379,670	151	151	151	97.64	93.19	37.19	No
SRR26320680_2.fastq	27,254,170	4,115,379,670	151	151	151	94.52	89.11	38.8	No
SRR10996883_1.fastq	27,565,702	4,162,421,002	151	151	151	98.61	96.17	42.36	No
SRR10996883_2.fastq	27,565,702	4,162,421,002	151	151	151	93.13	85.67	42.27	No
SRR26320664_1.fastq	28,001,911	4,228,288,561	151	151	151	98.07	94.67	39.07	No
SRR26320664_2.fastq	28,001,911	4,228,288,561	151	151	151	94.3	89.1	40.5	No
SRR26320683_1.fastq	28,173,809	4,254,245,159	151	151	151	98.13	94.75	38.28	No
SRR26320683_2.fastq	28,173,809	4,254,245,159	151	151	151	95.75	90.43	38.18	No
SRR16770664_1.fastq	28,349,914	4,280,837,014	151	151	151	97.42	93.16	36.39	No
SRR16770664_2.fastq	28,349,914	4,280,837,014	151	151	151	92.04	87.06	37.06	No
SRR26320677_1.fastq	29,396,222	4,438,829,522	151	151	151	96.5	92.64	37.52	No
SRR26320677_2.fastq	29,396,222	4,438,829,522	151	151	151	96.04	90.86	37.94	No
SRR26320684_1.fastq	33,031,364	4,987,735,964	151	151	151	97.09	92.96	37.59	No
SRR26320684_2.fastq	33,031,364	4,987,735,964	151	151	151	95.6	89.87	37.55	No
SRR26320663_1.fastq	36,403,301	5,496,898,451	151	151	151	98	94.44	38.62	No
SRR26320663_2.fastq	36,403,301	5,496,898,451	151	151	151	95.47	90.07	40.72	No
SRR10363469_1.fastq	39,239,272	3,923,927,200	100	100	100	98.49	94.86	31.53	No
SRR10363469_2.fastq	39,239,272	3,923,927,200	100	100	100	96.12	91.28	31.56	No
SRR10363559_1.fastq	40,395,527	4,039,552,700	100	100	100	98.55	94.99	31.64	No
SRR10363559_2.fastq	40,395,527	4,039,552,700	100	100	100	96.22	91.43	31.71	No
SRR26320665_1.fastq	42,549,921	6,425,038,071	151	151	151	98.36	95.14	37.49	No
SRR26320665_2.fastq	42,549,921	6,425,038,071	151	151	151	96.84	91.73	36.78	No
SRR11817818_1.fastq	46,232,746	4,669,507,346	101	101	101	97.88	96.57	31.68	No

SRR11817818_2.fastq	46,232,746	4,669,507,346	101	101	101	92.86	89.33	31.49	No
SRR26320679_1.fastq	47,282,557	7,139,666,107	151	151	151	97.93	94.08	38.25	No
SRR26320679_2.fastq	47,282,557	7,139,666,107	151	151	151	94.87	89.67	38.95	No
SRR11817819_1.fastq	50,760,234	5,126,783,634	101	101	101	97.94	96.67	30.9	No
SRR11817819_2.fastq	50,760,234	5,126,783,634	101	101	101	93.67	90.54	30.73	No
SRR11817817_1.fastq	54,380,295	5,492,409,795	101	101	101	97.18	95.74	32.17	No
SRR11817817_2.fastq	54,380,295	5,492,409,795	101	101	101	92.43	89.01	32.05	No
SRR11817822_1.fastq	56,863,190	5,743,182,190	101	101	101	98.01	96.77	31.39	No
SRR11817822_2.fastq	56,863,190	5,743,182,190	101	101	101	93.6	90.45	31.22	No

Supplemental Table 2. Top three predicted species for each SRA dataset sample obtained from NCBI. Reads were mapped to the Cryptosporidium parvum BGFT-T genome assembly and species identification was performed using Kraken2 with the EupathDB database

Sample	Taxon	Taxon_ID	Rank	Percentage
report_kraken2_Trim_ERR1760143	Cryptosporidium parvum	5807	species	50.4
report_kraken2_Trim_ERR1760143	Cryptosporidium hominis	237895	species	2.66
report_kraken2_Trim_ERR1760143	Cryptosporidium tyzzeri	756076	species	2.37
report_kraken2_Trim_ERR2889313	Cryptosporidium parvum	5807	species	84.87
report_kraken2_Trim_ERR2889313	Cryptosporidium hominis	237895	species	3.11
report_kraken2_Trim_ERR2889313	Cryptosporidium tyzzeri	756076	species	2.67
report_kraken2_Trim_ERR2889314	Cryptosporidium parvum	5807	species	26.64
report_kraken2_Trim_ERR2889314	Cryptosporidium hominis	237895	species	0.63
report_kraken2_Trim_ERR2889314	Cryptosporidium tyzzeri	756076	species	0.48
report_kraken2_Trim_ERR2889315	Cryptosporidium parvum	5807	species	68.54
report_kraken2_Trim_ERR2889315	Cryptosporidium hominis	237895	species	1.71
report_kraken2_Trim_ERR2889315	Cryptosporidium tyzzeri	756076	species	1.42
report_kraken2_Trim_ERR2889325	Cryptosporidium parvum	5807	species	48.64
report_kraken2_Trim_ERR2889325	Cryptosporidium hominis	237895	species	1.74
report_kraken2_Trim_ERR2889325	Cryptosporidium tyzzeri	756076	species	1.38
report_kraken2_Trim_ERR2889326	Cryptosporidium parvum	5807	species	82.17
report_kraken2_Trim_ERR2889326	Cryptosporidium hominis	237895	species	1.94
report_kraken2_Trim_ERR2889326	Cryptosporidium tyzzeri	756076	species	1.61
report_kraken2_Trim_ERR2889327	Cryptosporidium parvum	5807	species	50.72
report_kraken2_Trim_ERR2889327	Cryptosporidium hominis	237895	species	1.19
report_kraken2_Trim_ERR2889327	Cryptosporidium tyzzeri	756076	species	1.02
report_kraken2_Trim_SRR10336407	Cryptosporidium parvum	5807	species	79.95
report_kraken2_Trim_SRR10336407	Cryptosporidium hominis	237895	species	2.58
report_kraken2_Trim_SRR10336407	Cryptosporidium tyzzeri	756076	species	1.92
report_kraken2_Trim_SRR10336411	Cryptosporidium parvum	5807	species	87.84
report_kraken2_Trim_SRR10336411	Cryptosporidium hominis	237895	species	4.55
report_kraken2_Trim_SRR10336411	Cryptosporidium tyzzeri	756076	species	2.55
report_kraken2_Trim_SRR10336412	Cryptosporidium parvum	5807	species	91.8
report_kraken2_Trim_SRR10336412	Cryptosporidium hominis	237895	species	1.94
report_kraken2_Trim_SRR10336412	Cryptosporidium tyzzeri	756076	species	1.14
report_kraken2_Trim_SRR10336413	Cryptosporidium parvum	5807	species	89.96
report_kraken2_Trim_SRR10336413	Cryptosporidium hominis	237895	species	3.51

report_kraken2_Trim_SRR10336413	Cryptosporidium tyzzeri	756076	species	1.42
report_kraken2_Trim_SRR10336414	Cryptosporidium parvum	5807	species	84.75
report_kraken2_Trim_SRR10336414	Cryptosporidium hominis	237895	species	6.13
report_kraken2_Trim_SRR10336414	Cryptosporidium tyzzeri	756076	species	1.33
report_kraken2_Trim_SRR10336415	Cryptosporidium parvum	5807	species	46.49
report_kraken2_Trim_SRR10336415	Cryptosporidium hominis	237895	species	19.68
report_kraken2_Trim_SRR10336415	Cryptosporidium tyzzeri	756076	species	14.31
report_kraken2_Trim_SRR10363299	Cryptosporidium parvum	5807	species	79.86
report_kraken2_Trim_SRR10363299	Cryptosporidium hominis	237895	species	2.5
report_kraken2_Trim_SRR10363299	Cryptosporidium tyzzeri	756076	species	1.88
report_kraken2_Trim_SRR10363300	Cryptosporidium parvum	5807	species	86.8
report_kraken2_Trim_SRR10363300	Cryptosporidium hominis	237895	species	1.64
report_kraken2_Trim_SRR10363300	Cryptosporidium tyzzeri	756076	species	1.13
report_kraken2_Trim_SRR10363301	Cryptosporidium parvum	5807	species	80.33
report_kraken2_Trim_SRR10363301	Cryptosporidium hominis	237895	species	1.41
report_kraken2_Trim_SRR10363301	Trypanosoma cruzi	5693	species	1.26
report_kraken2_Trim_SRR10363302	Cryptosporidium parvum	5807	species	88.39
report_kraken2_Trim_SRR10363302	Cryptosporidium hominis	237895	species	1.62
report_kraken2_Trim_SRR10363302	Cryptosporidium tyzzeri	756076	species	1.04
report_kraken2_Trim_SRR10363303	Cryptosporidium parvum	5807	species	92.49
report_kraken2_Trim_SRR10363303	Cryptosporidium hominis	237895	species	1.81
report_kraken2_Trim_SRR10363303	Cryptosporidium tyzzeri	756076	species	1.12
report_kraken2_Trim_SRR10363304	Cryptosporidium parvum	5807	species	91.82
report_kraken2_Trim_SRR10363304	Cryptosporidium hominis	237895	species	1.8
report_kraken2_Trim_SRR10363304	Cryptosporidium tyzzeri	756076	species	1.1
report_kraken2_Trim_SRR10363305	Cryptosporidium parvum	5807	species	89.42
report_kraken2_Trim_SRR10363305	Cryptosporidium hominis	237895	species	1.73
report_kraken2_Trim_SRR10363305	Cryptosporidium tyzzeri	756076	species	1.1
report_kraken2_Trim_SRR10363306	Cryptosporidium parvum	5807	species	77.75
report_kraken2_Trim_SRR10363306	Trypanosoma cruzi	5693	species	1.35
report_kraken2_Trim_SRR10363306	Cryptosporidium hominis	237895	species	1.31
report_kraken2_Trim_SRR10363308	Cryptosporidium parvum	5807	species	92.75
report_kraken2_Trim_SRR10363308	Cryptosporidium hominis	237895	species	1.76
report_kraken2_Trim_SRR10363308	Cryptosporidium tyzzeri	756076	species	1.09
report_kraken2_Trim_SRR10363309	Cryptosporidium parvum	5807	species	34.6
report_kraken2_Trim_SRR10363309	Trypanosoma cruzi	5693	species	4.76
report_kraken2_Trim_SRR10363309	Cryptosporidium hominis	237895	species	0.71
report_kraken2_Trim_SRR10363310	Cryptosporidium parvum	5807	species	92.1
report_kraken2_Trim_SRR10363310	Cryptosporidium hominis	237895	species	1.87
report_kraken2_Trim_SRR10363310	Cryptosporidium tyzzeri	756076	species	1.45

report_kraken2_Trim_SRR10363311	Cryptosporidium parvum	5807	species	26.14
report_kraken2_Trim_SRR10363311	Cryptosporidium hominis	237895	species	0.82
report_kraken2_Trim_SRR10363311	Cryptosporidium tyzzeri	756076	species	0.56
report_kraken2_Trim_SRR10363315	Cryptosporidium parvum	5807	species	92.62
report_kraken2_Trim_SRR10363315	Cryptosporidium hominis	237895	species	2.09
report_kraken2_Trim_SRR10363315	Cryptosporidium tyzzeri	756076	species	1.43
report_kraken2_Trim_SRR10363316	Cryptosporidium parvum	5807	species	90.26
report_kraken2_Trim_SRR10363316	Cryptosporidium hominis	237895	species	2.11
report_kraken2_Trim_SRR10363316	Cryptosporidium tyzzeri	756076	species	1.46
report_kraken2_Trim_SRR10363317	Cryptosporidium parvum	5807	species	91.52
report_kraken2_Trim_SRR10363317	Cryptosporidium hominis	237895	species	2.13
report_kraken2_Trim_SRR10363317	Cryptosporidium tyzzeri	756076	species	1.49
report_kraken2_Trim_SRR10363319	Cryptosporidium parvum	5807	species	27.27
report_kraken2_Trim_SRR10363319	Cryptosporidium hominis	237895	species	1.03
report_kraken2_Trim_SRR10363319	Cryptosporidium tyzzeri	756076	species	0.39
report_kraken2_Trim_SRR10363320	Cryptosporidium parvum	5807	species	92.17
report_kraken2_Trim_SRR10363320	Cryptosporidium hominis	237895	species	2.07
report_kraken2_Trim_SRR10363320	Cryptosporidium tyzzeri	756076	species	1.49
report_kraken2_Trim_SRR10363321	Cryptosporidium parvum	5807	species	92.75
report_kraken2_Trim_SRR10363321	Cryptosporidium hominis	237895	species	1.84
report_kraken2_Trim_SRR10363321	Cryptosporidium tyzzeri	756076	species	1.27
report_kraken2_Trim_SRR10363322	Cryptosporidium parvum	5807	species	18.12
report_kraken2_Trim_SRR10363322	Cryptosporidium hominis	237895	species	1.04
report_kraken2_Trim_SRR10363322	Cryptosporidium tyzzeri	756076	species	0.54
report_kraken2_Trim_SRR10363323	Cryptosporidium parvum	5807	species	76.53
report_kraken2_Trim_SRR10363323	Cryptosporidium hominis	237895	species	1.32
report_kraken2_Trim_SRR10363323	Cryptosporidium tyzzeri	756076	species	0.95
report_kraken2_Trim_SRR10363324	Cryptosporidium parvum	5807	species	92.56
report_kraken2_Trim_SRR10363324	Cryptosporidium hominis	237895	species	1.45
report_kraken2_Trim_SRR10363324	Cryptosporidium tyzzeri	756076	species	1.08
report_kraken2_Trim_SRR10363325	Cryptosporidium parvum	5807	species	93
report_kraken2_Trim_SRR10363325	Cryptosporidium hominis	237895	species	1.51
report_kraken2_Trim_SRR10363325	Cryptosporidium tyzzeri	756076	species	1.08
report_kraken2_Trim_SRR10363326	Cryptosporidium parvum	5807	species	91.65
report_kraken2_Trim_SRR10363326	Cryptosporidium hominis	237895	species	1.62
report_kraken2_Trim_SRR10363326	Cryptosporidium tyzzeri	756076	species	1.18
report_kraken2_Trim_SRR10363327	Cryptosporidium parvum	5807	species	90.08
report_kraken2_Trim_SRR10363327	Cryptosporidium hominis	237895	species	1.39
report_kraken2_Trim_SRR10363327	Cryptosporidium tyzzeri	756076	species	1
report_kraken2_Trim_SRR10363328	Cryptosporidium parvum	5807	species	66.17

report_kraken2_Trim_SRR10363328	Cryptosporidium hominis	237895	species	2.1
report_kraken2_Trim_SRR10363328	Cryptosporidium tyzzeri	756076	species	1.57
report_kraken2_Trim_SRR10363333	Cryptosporidium parvum	5807	species	84.27
report_kraken2_Trim_SRR10363333	Cryptosporidium hominis	237895	species	1.46
report_kraken2_Trim_SRR10363333	Cryptosporidium tyzzeri	756076	species	0.97
report_kraken2_Trim_SRR10363335	Cryptosporidium parvum	5807	species	84.86
report_kraken2_Trim_SRR10363335	Cryptosporidium hominis	237895	species	2.64
report_kraken2_Trim_SRR10363335	Cryptosporidium tyzzeri	756076	species	1.93
report_kraken2_Trim_SRR10363336	Cryptosporidium parvum	5807	species	83.74
report_kraken2_Trim_SRR10363336	Cryptosporidium hominis	237895	species	1.61
report_kraken2_Trim_SRR10363336	Cryptosporidium tyzzeri	756076	species	0.97
report_kraken2_Trim_SRR10363337	Cryptosporidium parvum	5807	species	48.81
report_kraken2_Trim_SRR10363337	Cryptosporidium hominis	237895	species	2.09
report_kraken2_Trim_SRR10363337	Trypanosoma cruzi	5693	species	1.63
report_kraken2_Trim_SRR10363338	Cryptosporidium parvum	5807	species	83.43
report_kraken2_Trim_SRR10363338	Cryptosporidium hominis	237895	species	1.47
report_kraken2_Trim_SRR10363338	Cryptosporidium tyzzeri	756076	species	0.93
report_kraken2_Trim_SRR10363339	Cryptosporidium parvum	5807	species	92.36
report_kraken2_Trim_SRR10363339	Cryptosporidium hominis	237895	species	1.7
report_kraken2_Trim_SRR10363339	Cryptosporidium tyzzeri	756076	species	1.11
report_kraken2_Trim_SRR10363340	Cryptosporidium parvum	5807	species	89.22
report_kraken2_Trim_SRR10363340	Cryptosporidium hominis	237895	species	1.71
report_kraken2_Trim_SRR10363340	Cryptosporidium tyzzeri	756076	species	1.08
report_kraken2_Trim_SRR10363341	Cryptosporidium parvum	5807	species	26.24
report_kraken2_Trim_SRR10363341	Trypanosoma cruzi	5693	species	6.62
report_kraken2_Trim_SRR10363341	Cryptosporidium hominis	237895	species	0.53
report_kraken2_Trim_SRR10363342	Cryptosporidium parvum	5807	species	32.94
report_kraken2_Trim_SRR10363342	Trypanosoma cruzi	5693	species	5.3
report_kraken2_Trim_SRR10363342	Cryptosporidium hominis	237895	species	0.54
report_kraken2_Trim_SRR10363343	Cryptosporidium parvum	5807	species	78.19
report_kraken2_Trim_SRR10363343	Cryptosporidium hominis	237895	species	1.27
report_kraken2_Trim_SRR10363343	Trypanosoma cruzi	5693	species	1.25
report_kraken2_Trim_SRR10363344	Cryptosporidium parvum	5807	species	88.15
report_kraken2_Trim_SRR10363344	Cryptosporidium hominis	237895	species	1.53
report_kraken2_Trim_SRR10363344	Cryptosporidium tyzzeri	756076	species	1
report_kraken2_Trim_SRR10363345	Cryptosporidium parvum	5807	species	90.81
report_kraken2_Trim_SRR10363345	Cryptosporidium hominis	237895	species	1.71
report_kraken2_Trim_SRR10363345	Cryptosporidium tyzzeri	756076	species	1.06
report_kraken2_Trim_SRR10363346	Cryptosporidium parvum	5807	species	61.62
report_kraken2_Trim_SRR10363346	Cryptosporidium hominis	237895	species	1.81

report_kraken2_Trim_SRR10363346	Cryptosporidium tyzzeri	756076	species	1.28
report_kraken2_Trim_SRR10363347	Cryptosporidium parvum	5807	species	90.92
report_kraken2_Trim_SRR10363347	Cryptosporidium hominis	237895	species	1.87
report_kraken2_Trim_SRR10363347	Cryptosporidium tyzzeri	756076	species	1.16
report_kraken2_Trim_SRR10363348	Cryptosporidium parvum	5807	species	86.23
report_kraken2_Trim_SRR10363348	Cryptosporidium hominis	237895	species	1.52
report_kraken2_Trim_SRR10363348	Cryptosporidium tyzzeri	756076	species	1.02
report_kraken2_Trim_SRR10363354	Cryptosporidium parvum	5807	species	73.38
report_kraken2_Trim_SRR10363354	Trypanosoma cruzi	5693	species	1.7
report_kraken2_Trim_SRR10363354	Cryptosporidium hominis	237895	species	1.63
report_kraken2_Trim_SRR10363355	Cryptosporidium parvum	5807	species	70.86
report_kraken2_Trim_SRR10363355	Trypanosoma cruzi	5693	species	1.92
report_kraken2_Trim_SRR10363355	Cryptosporidium hominis	237895	species	1.54
report_kraken2_Trim_SRR10363357	Cryptosporidium parvum	5807	species	66.88
report_kraken2_Trim_SRR10363357	Trypanosoma cruzi	5693	species	2.34
report_kraken2_Trim_SRR10363357	Cryptosporidium hominis	237895	species	1.07
report_kraken2_Trim_SRR10363358	Cryptosporidium parvum	5807	species	78.15
report_kraken2_Trim_SRR10363358	Cryptosporidium hominis	237895	species	2.39
report_kraken2_Trim_SRR10363358	Cryptosporidium tyzzeri	756076	species	1.91
report_kraken2_Trim_SRR10363362	Cryptosporidium parvum	5807	species	71.32
report_kraken2_Trim_SRR10363362	Cryptosporidium hominis	237895	species	1.22
report_kraken2_Trim_SRR10363362	Cryptosporidium tyzzeri	756076	species	0.82
report_kraken2_Trim_SRR10363363	Cryptosporidium parvum	5807	species	73.17
report_kraken2_Trim_SRR10363363	Cryptosporidium hominis	237895	species	1.35
report_kraken2_Trim_SRR10363363	Cryptosporidium tyzzeri	756076	species	0.95
report_kraken2_Trim_SRR10363364	Cryptosporidium parvum	5807	species	80.68
report_kraken2_Trim_SRR10363364	Cryptosporidium hominis	237895	species	1.54
report_kraken2_Trim_SRR10363364	Cryptosporidium tyzzeri	756076	species	1.03
report_kraken2_Trim_SRR10363365	Cryptosporidium parvum	5807	species	78.85
report_kraken2_Trim_SRR10363365	Cryptosporidium hominis	237895	species	1.39
report_kraken2_Trim_SRR10363365	Cryptosporidium tyzzeri	756076	species	0.95
report_kraken2_Trim_SRR10363369	Cryptosporidium parvum	5807	species	56.38
report_kraken2_Trim_SRR10363369	Cryptosporidium hominis	237895	species	1.09
report_kraken2_Trim_SRR10363369	Cryptosporidium tyzzeri	756076	species	0.55
report_kraken2_Trim_SRR10363371	Cryptosporidium parvum	5807	species	9.86
report_kraken2_Trim_SRR10363371	Cryptosporidium hominis	237895	species	0.18
report_kraken2_Trim_SRR10363371	Cryptosporidium tyzzeri	756076	species	0.12
report_kraken2_Trim_SRR10363372	Cryptosporidium parvum	5807	species	21.47
report_kraken2_Trim_SRR10363372	Cryptosporidium hominis	237895	species	0.38
report_kraken2_Trim_SRR10363372	Cryptosporidium tyzzeri	756076	species	0.24

report_kraken2_Trim_SRR10363373	Cryptosporidium parvum	5807	species	2.02
report_kraken2_Trim_SRR10363373	Talaromyces stipitatus	28564	species	0.05
report_kraken2_Trim_SRR10363373	Cryptosporidium hominis	237895	species	0.03
report_kraken2_Trim_SRR10363374	Cryptosporidium parvum	5807	species	6.13
report_kraken2_Trim_SRR10363374	Cryptosporidium hominis	237895	species	0.11
report_kraken2_Trim_SRR10363374	Cryptosporidium tyzzeri	756076	species	0.07
report_kraken2_Trim_SRR10363376	Cryptosporidium parvum	5807	species	9.82
report_kraken2_Trim_SRR10363376	Cryptosporidium hominis	237895	species	0.13
report_kraken2_Trim_SRR10363376	Cryptosporidium tyzzeri	756076	species	0.09
report_kraken2_Trim_SRR10363378	Cryptosporidium parvum	5807	species	85.52
report_kraken2_Trim_SRR10363378	Cryptosporidium hominis	237895	species	1.56
report_kraken2_Trim_SRR10363378	Cryptosporidium tyzzeri	756076	species	1.02
report_kraken2_Trim_SRR10363379	Cryptosporidium parvum	5807	species	85.86
report_kraken2_Trim_SRR10363379	Cryptosporidium hominis	237895	species	1.48
report_kraken2_Trim_SRR10363379	Cryptosporidium tyzzeri	756076	species	0.99
report_kraken2_Trim_SRR10363380	Cryptosporidium parvum	5807	species	41.69
report_kraken2_Trim_SRR10363380	Monocercomonoides exilis	2049356	species	1.38
report_kraken2_Trim_SRR10363380	Cryptosporidium hominis	237895	species	1.26
report_kraken2_Trim_SRR10363381	Cryptosporidium parvum	5807	species	5.74
report_kraken2_Trim_SRR10363381	Leishmania braziliensis	5660	species	0.1
report_kraken2_Trim_SRR10363381	Cryptosporidium hominis	237895	species	0.1
report_kraken2_Trim_SRR10363382	Cryptosporidium parvum	5807	species	60.82
report_kraken2_Trim_SRR10363382	Cryptosporidium hominis	237895	species	0.91
report_kraken2_Trim_SRR10363382	Cryptosporidium tyzzeri	756076	species	0.57
report_kraken2_Trim_SRR10363383	Cryptosporidium parvum	5807	species	50.63
report_kraken2_Trim_SRR10363383	Cryptosporidium hominis	237895	species	0.86
report_kraken2_Trim_SRR10363383	Cryptosporidium tyzzeri	756076	species	0.57
report_kraken2_Trim_SRR10363384	Cryptosporidium parvum	5807	species	83.61
report_kraken2_Trim_SRR10363384	Cryptosporidium hominis	237895	species	1.62
report_kraken2_Trim_SRR10363384	Cryptosporidium tyzzeri	756076	species	1.05
report_kraken2_Trim_SRR10363385	Cryptosporidium parvum	5807	species	79.82
report_kraken2_Trim_SRR10363385	Cryptosporidium hominis	237895	species	1.3
report_kraken2_Trim_SRR10363385	Cryptosporidium tyzzeri	756076	species	0.87
report_kraken2_Trim_SRR10363386	Cryptosporidium parvum	5807	species	33.82
report_kraken2_Trim_SRR10363386	Penicillium rubens	1108849	species	1.21
report_kraken2_Trim_SRR10363386	Cryptosporidium hominis	237895	species	0.49
report_kraken2_Trim_SRR10363387	Cryptosporidium parvum	5807	species	40.93
report_kraken2_Trim_SRR10363387	Cryptosporidium hominis	237895	species	0.87
report_kraken2_Trim_SRR10363387	Cryptosporidium tyzzeri	756076	species	0.62
report_kraken2_Trim_SRR10363391	Cryptosporidium parvum	5807	species	79.53

report_kraken2_Trim_SRR10363391	Cryptosporidium hominis	237895	species	2.86
report_kraken2_Trim_SRR10363391	Cryptosporidium tyzzeri	756076	species	2.13
report_kraken2_Trim_SRR10363395	Cryptosporidium parvum	5807	species	92.92
report_kraken2_Trim_SRR10363395	Cryptosporidium hominis	237895	species	1.96
report_kraken2_Trim_SRR10363395	Cryptosporidium tyzzeri	756076	species	1.17
report_kraken2_Trim_SRR10363397	Cryptosporidium parvum	5807	species	92.97
report_kraken2_Trim_SRR10363397	Cryptosporidium hominis	237895	species	1.87
report_kraken2_Trim_SRR10363397	Cryptosporidium tyzzeri	756076	species	1.26
report_kraken2_Trim_SRR10363401	Cryptosporidium parvum	5807	species	93.33
report_kraken2_Trim_SRR10363401	Cryptosporidium hominis	237895	species	1.82
report_kraken2_Trim_SRR10363401	Cryptosporidium tyzzeri	756076	species	1.2
report_kraken2_Trim_SRR10363402	Cryptosporidium parvum	5807	species	75.5
report_kraken2_Trim_SRR10363402	Cryptosporidium hominis	237895	species	2.78
report_kraken2_Trim_SRR10363402	Cryptosporidium tyzzeri	756076	species	2.06
report_kraken2_Trim_SRR10363403	Cryptosporidium parvum	5807	species	92.31
report_kraken2_Trim_SRR10363403	Cryptosporidium hominis	237895	species	2.22
report_kraken2_Trim_SRR10363403	Cryptosporidium tyzzeri	756076	species	1.41
report_kraken2_Trim_SRR10363404	Cryptosporidium parvum	5807	species	92.45
report_kraken2_Trim_SRR10363404	Cryptosporidium hominis	237895	species	2.11
report_kraken2_Trim_SRR10363404	Cryptosporidium tyzzeri	756076	species	1.36
report_kraken2_Trim_SRR10363405	Cryptosporidium parvum	5807	species	91.32
report_kraken2_Trim_SRR10363405	Cryptosporidium hominis	237895	species	3.46
report_kraken2_Trim_SRR10363405	Cryptosporidium tyzzeri	756076	species	1.18
report_kraken2_Trim_SRR10363406	Cryptosporidium parvum	5807	species	93.1
report_kraken2_Trim_SRR10363406	Cryptosporidium hominis	237895	species	1.8
report_kraken2_Trim_SRR10363406	Cryptosporidium tyzzeri	756076	species	1.18
report_kraken2_Trim_SRR10363409	Cryptosporidium parvum	5807	species	92.67
report_kraken2_Trim_SRR10363409	Cryptosporidium hominis	237895	species	2.08
report_kraken2_Trim_SRR10363409	Cryptosporidium tyzzeri	756076	species	1.35
report_kraken2_Trim_SRR10363410	Cryptosporidium parvum	5807	species	92.87
report_kraken2_Trim_SRR10363410	Cryptosporidium hominis	237895	species	1.92
report_kraken2_Trim_SRR10363410	Cryptosporidium tyzzeri	756076	species	1.28
report_kraken2_Trim_SRR10363411	Cryptosporidium parvum	5807	species	91.74
report_kraken2_Trim_SRR10363411	Cryptosporidium hominis	237895	species	2.34
report_kraken2_Trim_SRR10363411	Cryptosporidium tyzzeri	756076	species	1.51
report_kraken2_Trim_SRR10363412	Cryptosporidium parvum	5807	species	92.67
report_kraken2_Trim_SRR10363412	Cryptosporidium hominis	237895	species	1.96
report_kraken2_Trim_SRR10363412	Cryptosporidium tyzzeri	756076	species	1.29
report_kraken2_Trim_SRR10363413	Cryptosporidium parvum	5807	species	76.89
report_kraken2_Trim_SRR10363413	Cryptosporidium hominis	237895	species	2.85

report_kraken2_Trim_SRR10363413	Cryptosporidium tyzzeri	756076	species	2.14
report_kraken2_Trim_SRR10363414	Cryptosporidium parvum	5807	species	93.52
report_kraken2_Trim_SRR10363414	Cryptosporidium hominis	237895	species	1.64
report_kraken2_Trim_SRR10363414	Cryptosporidium tyzzeri	756076	species	1.15
report_kraken2_Trim_SRR10363415	Cryptosporidium ubiquitum	857276	species	1.15
report_kraken2_Trim_SRR10363415	Cryptosporidium sp. chipmunk LX-2015	1603071	species	0.87
report_kraken2_Trim_SRR10363415	Cryptosporidium baileyi	27987	species	0.58
report_kraken2_Trim_SRR10363416	Cryptosporidium parvum	5807	species	92.95
report_kraken2_Trim_SRR10363416	Cryptosporidium hominis	237895	species	1.99
report_kraken2_Trim_SRR10363416	Cryptosporidium tyzzeri	756076	species	1.29
report_kraken2_Trim_SRR10363417	Cryptosporidium parvum	5807	species	81.98
report_kraken2_Trim_SRR10363417	Cryptosporidium hominis	237895	species	1.44
report_kraken2_Trim_SRR10363417	Cryptosporidium tyzzeri	756076	species	1.01
report_kraken2_Trim_SRR10363419	Cryptosporidium parvum	5807	species	92.7
report_kraken2_Trim_SRR10363419	Cryptosporidium hominis	237895	species	2.03
report_kraken2_Trim_SRR10363419	Cryptosporidium tyzzeri	756076	species	1.34
report_kraken2_Trim_SRR10363421	Cryptosporidium parvum	5807	species	93.47
report_kraken2_Trim_SRR10363421	Cryptosporidium hominis	237895	species	1.7
report_kraken2_Trim_SRR10363421	Cryptosporidium tyzzeri	756076	species	1.15
report_kraken2_Trim_SRR10363422	Cryptosporidium parvum	5807	species	92.79
report_kraken2_Trim_SRR10363422	Cryptosporidium hominis	237895	species	2.01
report_kraken2_Trim_SRR10363422	Cryptosporidium tyzzeri	756076	species	1.34
report_kraken2_Trim_SRR10363423	Cryptosporidium parvum	5807	species	93
report_kraken2_Trim_SRR10363423	Cryptosporidium hominis	237895	species	1.93
report_kraken2_Trim_SRR10363423	Cryptosporidium tyzzeri	756076	species	1.29
report_kraken2_Trim_SRR10363424	Cryptosporidium parvum	5807	species	81.13
report_kraken2_Trim_SRR10363424	Cryptosporidium hominis	237895	species	2.34
report_kraken2_Trim_SRR10363424	Cryptosporidium tyzzeri	756076	species	1.77
report_kraken2_Trim_SRR10363425	Cryptosporidium parvum	5807	species	93.41
report_kraken2_Trim_SRR10363425	Cryptosporidium hominis	237895	species	1.79
report_kraken2_Trim_SRR10363425	Cryptosporidium tyzzeri	756076	species	1.17
report_kraken2_Trim_SRR10363428	Cryptosporidium parvum	5807	species	92.73
report_kraken2_Trim_SRR10363428	Cryptosporidium hominis	237895	species	1.89
report_kraken2_Trim_SRR10363428	Cryptosporidium tyzzeri	756076	species	1.26
report_kraken2_Trim_SRR10363429	Cryptosporidium parvum	5807	species	92.87
report_kraken2_Trim_SRR10363429	Cryptosporidium hominis	237895	species	1.91
report_kraken2_Trim_SRR10363429	Cryptosporidium tyzzeri	756076	species	1.28
report_kraken2_Trim_SRR10363430	Cryptosporidium parvum	5807	species	40.91
report_kraken2_Trim_SRR10363430	Cryptosporidium hominis	237895	species	0.82

report_kraken2_Trim_SRR10363430	Cryptosporidium tyzzeri	756076	species	0.51
report_kraken2_Trim_SRR10363431	Cryptosporidium parvum	5807	species	92.58
report_kraken2_Trim_SRR10363431	Cryptosporidium hominis	237895	species	2.04
report_kraken2_Trim_SRR10363431	Cryptosporidium tyzzeri	756076	species	1.34
report_kraken2_Trim_SRR10363432	Cryptosporidium ubiquitum	857276	species	1.2
report_kraken2_Trim_SRR10363432	Cryptosporidium sp. chipmunk LX-2015	1603071	species	0.77
report_kraken2_Trim_SRR10363432	Cryptosporidium baileyi	27987	species	0.56
report_kraken2_Trim_SRR10363433	Cryptosporidium parvum	5807	species	92.84
report_kraken2_Trim_SRR10363433	Cryptosporidium hominis	237895	species	1.93
report_kraken2_Trim_SRR10363433	Cryptosporidium tyzzeri	756076	species	1.27
report_kraken2_Trim_SRR10363434	Cryptosporidium parvum	5807	species	91.2
report_kraken2_Trim_SRR10363434	Cryptosporidium hominis	237895	species	1.89
report_kraken2_Trim_SRR10363434	Cryptosporidium tyzzeri	756076	species	1.25
report_kraken2_Trim_SRR10363435	Cryptosporidium parvum	5807	species	49.9
report_kraken2_Trim_SRR10363435	Cryptosporidium hominis	237895	species	1.46
report_kraken2_Trim_SRR10363435	Cryptosporidium tyzzeri	756076	species	1.06
report_kraken2_Trim_SRR10363436	Cryptosporidium parvum	5807	species	33.3
report_kraken2_Trim_SRR10363436	Cryptosporidium hominis	237895	species	0.71
report_kraken2_Trim_SRR10363436	Cryptosporidium tyzzeri	756076	species	0.43
report_kraken2_Trim_SRR10363437	Cryptosporidium parvum	5807	species	93.2
report_kraken2_Trim_SRR10363437	Cryptosporidium hominis	237895	species	1.73
report_kraken2_Trim_SRR10363437	Cryptosporidium tyzzeri	756076	species	1.21
report_kraken2_Trim_SRR10363438	Cryptosporidium parvum	5807	species	85.7
report_kraken2_Trim_SRR10363438	Cryptosporidium hominis	237895	species	1.58
report_kraken2_Trim_SRR10363438	Cryptosporidium tyzzeri	756076	species	1.11
report_kraken2_Trim_SRR10363441	Cryptosporidium parvum	5807	species	91.98
report_kraken2_Trim_SRR10363441	Cryptosporidium hominis	237895	species	1.72
report_kraken2_Trim_SRR10363441	Cryptosporidium tyzzeri	756076	species	1.13
report_kraken2_Trim_SRR10363443	Cryptosporidium parvum	5807	species	92.98
report_kraken2_Trim_SRR10363443	Cryptosporidium hominis	237895	species	1.99
report_kraken2_Trim_SRR10363443	Cryptosporidium tyzzeri	756076	species	1.27
report_kraken2_Trim_SRR10363444	Cryptosporidium parvum	5807	species	93.58
report_kraken2_Trim_SRR10363444	Cryptosporidium hominis	237895	species	1.74
report_kraken2_Trim_SRR10363444	Cryptosporidium tyzzeri	756076	species	1.11
report_kraken2_Trim_SRR10363445	Cryptosporidium parvum	5807	species	93.19
report_kraken2_Trim_SRR10363445	Cryptosporidium hominis	237895	species	1.84
report_kraken2_Trim_SRR10363445	Cryptosporidium tyzzeri	756076	species	1.21
report_kraken2_Trim_SRR10363446	Cryptosporidium parvum	5807	species	21.86
report_kraken2_Trim_SRR10363446	Cryptosporidium hominis	237895	species	0.81

report_kraken2_Trim_SRR10363446	Cryptosporidium tyzzeri	756076	species	0.61
report_kraken2_Trim_SRR10363447	Cryptosporidium parvum	5807	species	93.34
report_kraken2_Trim_SRR10363447	Cryptosporidium hominis	237895	species	1.8
report_kraken2_Trim_SRR10363447	Cryptosporidium tyzzeri	756076	species	1.17
report_kraken2_Trim_SRR10363448	Cryptosporidium parvum	5807	species	93.42
report_kraken2_Trim_SRR10363448	Cryptosporidium hominis	237895	species	1.76
report_kraken2_Trim_SRR10363448	Cryptosporidium tyzzeri	756076	species	1.15
report_kraken2_Trim_SRR10363449	Cryptosporidium parvum	5807	species	93.28
report_kraken2_Trim_SRR10363449	Cryptosporidium hominis	237895	species	1.84
report_kraken2_Trim_SRR10363449	Cryptosporidium tyzzeri	756076	species	1.2
report_kraken2_Trim_SRR10363450	Cryptosporidium parvum	5807	species	91.89
report_kraken2_Trim_SRR10363450	Cryptosporidium hominis	237895	species	2.04
report_kraken2_Trim_SRR10363450	Cryptosporidium tyzzeri	756076	species	1.16
report_kraken2_Trim_SRR10363451	Cryptosporidium parvum	5807	species	76.79
report_kraken2_Trim_SRR10363451	Cryptosporidium hominis	237895	species	1.55
report_kraken2_Trim_SRR10363451	Cryptosporidium tyzzeri	756076	species	0.89
report_kraken2_Trim_SRR10363453	Cryptosporidium parvum	5807	species	74.87
report_kraken2_Trim_SRR10363453	Cryptosporidium hominis	237895	species	7.21
report_kraken2_Trim_SRR10363453	Cryptosporidium tyzzeri	756076	species	0.98
report_kraken2_Trim_SRR10363457	Cryptosporidium parvum	5807	species	5.9
report_kraken2_Trim_SRR10363457	Cryptosporidium hominis	237895	species	0.15
report_kraken2_Trim_SRR10363457	Trypanosoma congolense	5692	species	0.09
report_kraken2_Trim_SRR10363460	Cryptosporidium parvum	5807	species	86.91
report_kraken2_Trim_SRR10363460	Cryptosporidium hominis	237895	species	1.68
report_kraken2_Trim_SRR10363460	Cryptosporidium tyzzeri	756076	species	0.97
report_kraken2_Trim_SRR10363461	Cryptosporidium parvum	5807	species	93.08
report_kraken2_Trim_SRR10363461	Cryptosporidium hominis	237895	species	1.89
report_kraken2_Trim_SRR10363461	Cryptosporidium tyzzeri	756076	species	1.19
report_kraken2_Trim_SRR10363462	Cryptosporidium parvum	5807	species	92.64
report_kraken2_Trim_SRR10363462	Cryptosporidium hominis	237895	species	2.08
report_kraken2_Trim_SRR10363462	Cryptosporidium tyzzeri	756076	species	1.11
report_kraken2_Trim_SRR10363463	Cryptosporidium parvum	5807	species	93.44
report_kraken2_Trim_SRR10363463	Cryptosporidium hominis	237895	species	1.7
report_kraken2_Trim_SRR10363463	Cryptosporidium tyzzeri	756076	species	1.08
report_kraken2_Trim_SRR10363465	Cryptosporidium parvum	5807	species	93.09
report_kraken2_Trim_SRR10363465	Cryptosporidium hominis	237895	species	2.02
report_kraken2_Trim_SRR10363465	Cryptosporidium tyzzeri	756076	species	1.26
report_kraken2_Trim_SRR10363466	Cryptosporidium parvum	5807	species	91.38
report_kraken2_Trim_SRR10363466	Cryptosporidium hominis	237895	species	1.72
report_kraken2_Trim_SRR10363466	Cryptosporidium tyzzeri	756076	species	1.12

report_kraken2_Trim_SRR10363467	Cryptosporidium parvum	5807	species	92.96
report_kraken2_Trim_SRR10363467	Cryptosporidium hominis	237895	species	1.63
report_kraken2_Trim_SRR10363467	Cryptosporidium tyzzeri	756076	species	1.02
report_kraken2_Trim_SRR10363468	Cryptosporidium parvum	5807	species	0.62
report_kraken2_Trim_SRR10363468	Phytophthora ramorum	164328	species	0.09
report_kraken2_Trim_SRR10363468	Cryptosporidium hominis	237895	species	0.05
report_kraken2_Trim_SRR10363469	Cryptosporidium parvum	5807	species	79.45
report_kraken2_Trim_SRR10363469	Cryptosporidium hominis	237895	species	2.11
report_kraken2_Trim_SRR10363469	Cryptosporidium tyzzeri	756076	species	1.55
report_kraken2_Trim_SRR10363470	Cryptosporidium parvum	5807	species	83.31
report_kraken2_Trim_SRR10363470	Cryptosporidium hominis	237895	species	1.73
report_kraken2_Trim_SRR10363470	Cryptosporidium tyzzeri	756076	species	1.02
report_kraken2_Trim_SRR10363473	Cryptosporidium parvum	5807	species	93.02
report_kraken2_Trim_SRR10363473	Cryptosporidium hominis	237895	species	1.87
report_kraken2_Trim_SRR10363473	Cryptosporidium tyzzeri	756076	species	1.18
report_kraken2_Trim_SRR10363474	Cryptosporidium parvum	5807	species	85.4
report_kraken2_Trim_SRR10363474	Cryptosporidium hominis	237895	species	1.51
report_kraken2_Trim_SRR10363474	Cryptosporidium tyzzeri	756076	species	0.96
report_kraken2_Trim_SRR10363475	Cryptosporidium parvum	5807	species	93.6
report_kraken2_Trim_SRR10363475	Cryptosporidium hominis	237895	species	1.71
report_kraken2_Trim_SRR10363475	Cryptosporidium tyzzeri	756076	species	1.12
report_kraken2_Trim_SRR10363476	Cryptosporidium parvum	5807	species	92.79
report_kraken2_Trim_SRR10363476	Cryptosporidium hominis	237895	species	2.21
report_kraken2_Trim_SRR10363476	Cryptosporidium tyzzeri	756076	species	1.24
report_kraken2_Trim_SRR10363477	Cryptosporidium parvum	5807	species	92.93
report_kraken2_Trim_SRR10363477	Cryptosporidium hominis	237895	species	2.01
report_kraken2_Trim_SRR10363477	Cryptosporidium tyzzeri	756076	species	1.15
report_kraken2_Trim_SRR10363478	Cryptosporidium parvum	5807	species	91.09
report_kraken2_Trim_SRR10363478	Cryptosporidium hominis	237895	species	1.94
report_kraken2_Trim_SRR10363478	Cryptosporidium tyzzeri	756076	species	1.08
report_kraken2_Trim_SRR10363479	Cryptosporidium parvum	5807	species	92.39
report_kraken2_Trim_SRR10363479	Cryptosporidium hominis	237895	species	2.17
report_kraken2_Trim_SRR10363479	Cryptosporidium tyzzeri	756076	species	1.43
report_kraken2_Trim_SRR10363480	Cryptosporidium parvum	5807	species	36.55
report_kraken2_Trim_SRR10363480	Cryptosporidium hominis	237895	species	1.19
report_kraken2_Trim_SRR10363480	Cryptosporidium tyzzeri	756076	species	0.88
report_kraken2_Trim_SRR10363481	Cryptosporidium parvum	5807	species	92.95
report_kraken2_Trim_SRR10363481	Cryptosporidium hominis	237895	species	1.94
report_kraken2_Trim_SRR10363481	Cryptosporidium tyzzeri	756076	species	1.31
report_kraken2_Trim_SRR10363483	Cryptosporidium parvum	5807	species	93.06

report_kraken2_Trim_SRR10363483	Cryptosporidium hominis	237895	species	1.6
report_kraken2_Trim_SRR10363483	Cryptosporidium tyzzeri	756076	species	1.04
report_kraken2_Trim_SRR10363484	Cryptosporidium parvum	5807	species	42.63
report_kraken2_Trim_SRR10363484	Trypanosoma cruzi	5693	species	2.79
report_kraken2_Trim_SRR10363484	Cryptosporidium hominis	237895	species	0.68
report_kraken2_Trim_SRR10363485	Cryptosporidium parvum	5807	species	45.3
report_kraken2_Trim_SRR10363485	Trypanosoma cruzi	5693	species	4.43
report_kraken2_Trim_SRR10363485	Cryptosporidium hominis	237895	species	0.96
report_kraken2_Trim_SRR10363486	Cryptosporidium parvum	5807	species	91.13
report_kraken2_Trim_SRR10363486	Cryptosporidium hominis	237895	species	2.57
report_kraken2_Trim_SRR10363486	Cryptosporidium tyzzeri	756076	species	1.21
report_kraken2_Trim_SRR10363487	Cryptosporidium parvum	5807	species	93.11
report_kraken2_Trim_SRR10363487	Cryptosporidium hominis	237895	species	1.9
report_kraken2_Trim_SRR10363487	Cryptosporidium tyzzeri	756076	species	1.34
report_kraken2_Trim_SRR10363488	Cryptosporidium parvum	5807	species	90.35
report_kraken2_Trim_SRR10363488	Cryptosporidium hominis	237895	species	3.14
report_kraken2_Trim_SRR10363488	Cryptosporidium tyzzeri	756076	species	1.98
report_kraken2_Trim_SRR10363491	Cryptosporidium parvum	5807	species	92.98
report_kraken2_Trim_SRR10363491	Cryptosporidium hominis	237895	species	1.83
report_kraken2_Trim_SRR10363491	Cryptosporidium tyzzeri	756076	species	1.27
report_kraken2_Trim_SRR10363492	Cryptosporidium parvum	5807	species	9.44
report_kraken2_Trim_SRR10363492	Cryptosporidium hominis	237895	species	0.41
report_kraken2_Trim_SRR10363492	Cryptosporidium tyzzeri	756076	species	0.14
report_kraken2_Trim_SRR10363493	Cryptosporidium parvum	5807	species	60.75
report_kraken2_Trim_SRR10363493	Cryptosporidium hominis	237895	species	1.12
report_kraken2_Trim_SRR10363493	Cryptosporidium tyzzeri	756076	species	0.84
report_kraken2_Trim_SRR10363494	Cryptosporidium parvum	5807	species	9.65
report_kraken2_Trim_SRR10363494	Cryptosporidium hominis	237895	species	0.26
report_kraken2_Trim_SRR10363494	Cryptosporidium tyzzeri	756076	species	0.19
report_kraken2_Trim_SRR10363495	Cryptosporidium parvum	5807	species	92.32
report_kraken2_Trim_SRR10363495	Cryptosporidium hominis	237895	species	1.83
report_kraken2_Trim_SRR10363495	Cryptosporidium tyzzeri	756076	species	1.42
report_kraken2_Trim_SRR10363496	Cryptosporidium parvum	5807	species	86.23
report_kraken2_Trim_SRR10363496	Cryptosporidium hominis	237895	species	1.43
report_kraken2_Trim_SRR10363496	Cryptosporidium tyzzeri	756076	species	1.14
report_kraken2_Trim_SRR10363497	Cryptosporidium parvum	5807	species	65.02
report_kraken2_Trim_SRR10363497	Cryptosporidium hominis	237895	species	1.33
report_kraken2_Trim_SRR10363497	Cryptosporidium tyzzeri	756076	species	1.06
report_kraken2_Trim_SRR10363498	Cryptosporidium parvum	5807	species	84.31
report_kraken2_Trim_SRR10363498	Cryptosporidium hominis	237895	species	1.6

report_kraken2_Trim_SRR10363498	Cryptosporidium tyzzeri	756076	species	1.29
report_kraken2_Trim_SRR10363499	Cryptosporidium parvum	5807	species	91.72
report_kraken2_Trim_SRR10363499	Cryptosporidium hominis	237895	species	2.17
report_kraken2_Trim_SRR10363499	Cryptosporidium tyzzeri	756076	species	1.57
report_kraken2_Trim_SRR10363500	Cryptosporidium parvum	5807	species	92.33
report_kraken2_Trim_SRR10363500	Cryptosporidium hominis	237895	species	1.76
report_kraken2_Trim_SRR10363500	Cryptosporidium tyzzeri	756076	species	1.36
report_kraken2_Trim_SRR10363501	Cryptosporidium parvum	5807	species	91.85
report_kraken2_Trim_SRR10363501	Cryptosporidium hominis	237895	species	2.22
report_kraken2_Trim_SRR10363501	Cryptosporidium tyzzeri	756076	species	1.58
report_kraken2_Trim_SRR10363504	Cryptosporidium parvum	5807	species	25.25
report_kraken2_Trim_SRR10363504	Cryptosporidium hominis	237895	species	0.56
report_kraken2_Trim_SRR10363504	Cryptosporidium tyzzeri	756076	species	0.33
report_kraken2_Trim_SRR10363505	Cryptosporidium parvum	5807	species	2.24
report_kraken2_Trim_SRR10363505	Cryptosporidium hominis	237895	species	0.12
report_kraken2_Trim_SRR10363505	Cyclospora cayetanensis	88456	species	0.08
report_kraken2_Trim_SRR10363506	Cryptosporidium parvum	5807	species	92.94
report_kraken2_Trim_SRR10363506	Cryptosporidium hominis	237895	species	1.9
report_kraken2_Trim_SRR10363506	Cryptosporidium tyzzeri	756076	species	1.25
report_kraken2_Trim_SRR10363509	Cryptosporidium parvum	5807	species	89.9
report_kraken2_Trim_SRR10363509	Cryptosporidium hominis	237895	species	2.49
report_kraken2_Trim_SRR10363509	Cryptosporidium tyzzeri	756076	species	1.98
report_kraken2_Trim_SRR10363511	Cryptosporidium parvum	5807	species	49.76
report_kraken2_Trim_SRR10363511	Cryptosporidium hominis	237895	species	9.9
report_kraken2_Trim_SRR10363511	Trypanosoma cruzi	5693	species	3.03
report_kraken2_Trim_SRR10363512	Cryptosporidium parvum	5807	species	92.33
report_kraken2_Trim_SRR10363512	Cryptosporidium hominis	237895	species	1.7
report_kraken2_Trim_SRR10363512	Cryptosporidium tyzzeri	756076	species	1.14
report_kraken2_Trim_SRR10363514	Cryptosporidium parvum	5807	species	91.41
report_kraken2_Trim_SRR10363514	Cryptosporidium hominis	237895	species	1.58
report_kraken2_Trim_SRR10363514	Cryptosporidium tyzzeri	756076	species	1.09
report_kraken2_Trim_SRR10363515	Cryptosporidium parvum	5807	species	92.32
report_kraken2_Trim_SRR10363515	Cryptosporidium hominis	237895	species	1.9
report_kraken2_Trim_SRR10363515	Cryptosporidium tyzzeri	756076	species	1.23
report_kraken2_Trim_SRR10363516	Cryptosporidium parvum	5807	species	87.61
report_kraken2_Trim_SRR10363516	Cryptosporidium hominis	237895	species	1.58
report_kraken2_Trim_SRR10363516	Cryptosporidium tyzzeri	756076	species	1.07
report_kraken2_Trim_SRR10363517	Cryptosporidium parvum	5807	species	35.48
report_kraken2_Trim_SRR10363517	Cryptosporidium hominis	237895	species	15.67
report_kraken2_Trim_SRR10363517	Cryptosporidium tyzzeri	756076	species	0.73

report_kraken2_Trim_SRR10363518	Cryptosporidium parvum	5807	species	92.39
report_kraken2_Trim_SRR10363518	Cryptosporidium hominis	237895	species	1.66
report_kraken2_Trim_SRR10363518	Cryptosporidium tyzzeri	756076	species	1.14
report_kraken2_Trim_SRR10363519	Cryptosporidium parvum	5807	species	92.27
report_kraken2_Trim_SRR10363519	Cryptosporidium hominis	237895	species	1.71
report_kraken2_Trim_SRR10363519	Cryptosporidium tyzzeri	756076	species	1.15
report_kraken2_Trim_SRR10363520	Cryptosporidium parvum	5807	species	90.98
report_kraken2_Trim_SRR10363520	Cryptosporidium hominis	237895	species	2.18
report_kraken2_Trim_SRR10363520	Cryptosporidium tyzzeri	756076	species	1.51
report_kraken2_Trim_SRR10363521	Cryptosporidium parvum	5807	species	61.72
report_kraken2_Trim_SRR10363521	Trypanosoma cruzi	5693	species	3.21
report_kraken2_Trim_SRR10363521	Cryptosporidium hominis	237895	species	1.59
report_kraken2_Trim_SRR10363522	Cryptosporidium parvum	5807	species	43.65
report_kraken2_Trim_SRR10363522	Cryptosporidium hominis	237895	species	1.14
report_kraken2_Trim_SRR10363522	Cryptosporidium tyzzeri	756076	species	0.79
report_kraken2_Trim_SRR10363523	Cryptosporidium parvum	5807	species	92.19
report_kraken2_Trim_SRR10363523	Cryptosporidium hominis	237895	species	1.88
report_kraken2_Trim_SRR10363523	Cryptosporidium tyzzeri	756076	species	1.24
report_kraken2_Trim_SRR10363525	Cryptosporidium parvum	5807	species	92.3
report_kraken2_Trim_SRR10363525	Cryptosporidium hominis	237895	species	1.89
report_kraken2_Trim_SRR10363525	Cryptosporidium tyzzeri	756076	species	1.28
report_kraken2_Trim_SRR10363526	Cryptosporidium parvum	5807	species	75.12
report_kraken2_Trim_SRR10363526	Trypanosoma cruzi	5693	species	1.75
report_kraken2_Trim_SRR10363526	Cryptosporidium hominis	237895	species	1.25
report_kraken2_Trim_SRR10363527	Cryptosporidium parvum	5807	species	42.17
report_kraken2_Trim_SRR10363527	Trypanosoma cruzi	5693	species	4.31
report_kraken2_Trim_SRR10363527	Cryptosporidium hominis	237895	species	0.86
report_kraken2_Trim_SRR10363528	Cryptosporidium parvum	5807	species	3.08
report_kraken2_Trim_SRR10363528	Pneumocystis jirovecii	42068	species	1.59
report_kraken2_Trim_SRR10363528	Cryptosporidium hominis	237895	species	0.09
report_kraken2_Trim_SRR10363529	Cryptosporidium parvum	5807	species	92.51
report_kraken2_Trim_SRR10363529	Cryptosporidium hominis	237895	species	1.8
report_kraken2_Trim_SRR10363529	Cryptosporidium tyzzeri	756076	species	1.2
report_kraken2_Trim_SRR10363530	Cryptosporidium parvum	5807	species	92.26
report_kraken2_Trim_SRR10363530	Cryptosporidium hominis	237895	species	1.95
report_kraken2_Trim_SRR10363530	Cryptosporidium tyzzeri	756076	species	1.23
report_kraken2_Trim_SRR10363531	Cryptosporidium parvum	5807	species	80.05
report_kraken2_Trim_SRR10363531	Cryptosporidium hominis	237895	species	1.19
report_kraken2_Trim_SRR10363531	Trypanosoma cruzi	5693	species	0.95
report_kraken2_Trim_SRR10363532	Cryptosporidium parvum	5807	species	87.35

report_kraken2_Trim_SRR10363532	Cryptosporidium hominis	237895	species	1.52
report_kraken2_Trim_SRR10363532	Cryptosporidium tyzzeri	756076	species	1.02
report_kraken2_Trim_SRR10363533	Cryptosporidium parvum	5807	species	67.39
report_kraken2_Trim_SRR10363533	Trypanosoma cruzi	5693	species	2.29
report_kraken2_Trim_SRR10363533	Cryptosporidium hominis	237895	species	1.01
report_kraken2_Trim_SRR10363535	Cryptosporidium parvum	5807	species	92.84
report_kraken2_Trim_SRR10363535	Cryptosporidium hominis	237895	species	1.7
report_kraken2_Trim_SRR10363535	Cryptosporidium tyzzeri	756076	species	1.22
report_kraken2_Trim_SRR10363536	Cryptosporidium parvum	5807	species	92.42
report_kraken2_Trim_SRR10363536	Cryptosporidium hominis	237895	species	1.89
report_kraken2_Trim_SRR10363536	Cryptosporidium tyzzeri	756076	species	1.36
report_kraken2_Trim_SRR10363537	Cryptosporidium parvum	5807	species	92.38
report_kraken2_Trim_SRR10363537	Cryptosporidium hominis	237895	species	2.04
report_kraken2_Trim_SRR10363537	Cryptosporidium tyzzeri	756076	species	1.47
report_kraken2_Trim_SRR10363538	Cryptosporidium parvum	5807	species	90.32
report_kraken2_Trim_SRR10363538	Cryptosporidium hominis	237895	species	1.66
report_kraken2_Trim_SRR10363538	Cryptosporidium tyzzeri	756076	species	1.1
report_kraken2_Trim_SRR10363539	Cryptosporidium parvum	5807	species	90.81
report_kraken2_Trim_SRR10363539	Cryptosporidium hominis	237895	species	1.59
report_kraken2_Trim_SRR10363539	Cryptosporidium tyzzeri	756076	species	0.97
report_kraken2_Trim_SRR10363540	Cryptosporidium parvum	5807	species	79.95
report_kraken2_Trim_SRR10363540	Cryptosporidium hominis	237895	species	2.78
report_kraken2_Trim_SRR10363540	Cryptosporidium tyzzeri	756076	species	2.06
report_kraken2_Trim_SRR10363547	Cryptosporidium parvum	5807	species	90.55
report_kraken2_Trim_SRR10363547	Cryptosporidium hominis	237895	species	2.01
report_kraken2_Trim_SRR10363547	Cryptosporidium tyzzeri	756076	species	1.08
report_kraken2_Trim_SRR10363551	Cryptosporidium parvum	5807	species	18.28
report_kraken2_Trim_SRR10363551	Cryptosporidium hominis	237895	species	0.57
report_kraken2_Trim_SRR10363551	Cryptosporidium tyzzeri	756076	species	0.42
report_kraken2_Trim_SRR10363552	Cryptosporidium hominis	237895	species	93.25
report_kraken2_Trim_SRR10363552	Cryptosporidium parvum	5807	species	1.18
report_kraken2_Trim_SRR10363552	Cryptosporidium tyzzeri	756076	species	0.89
report_kraken2_Trim_SRR10363554	Cryptosporidium parvum	5807	species	93.22
report_kraken2_Trim_SRR10363554	Cryptosporidium hominis	237895	species	1.7
report_kraken2_Trim_SRR10363554	Cryptosporidium tyzzeri	756076	species	1.15
report_kraken2_Trim_SRR10363556	Cryptosporidium parvum	5807	species	93.29
report_kraken2_Trim_SRR10363556	Cryptosporidium hominis	237895	species	1.82
report_kraken2_Trim_SRR10363556	Cryptosporidium tyzzeri	756076	species	1.2
report_kraken2_Trim_SRR10363557	Cryptosporidium parvum	5807	species	92.07
report_kraken2_Trim_SRR10363557	Cryptosporidium hominis	237895	species	1.65

report_kraken2_Trim_SRR10363557	Cryptosporidium tyzzeri	756076	species	1.08
report_kraken2_Trim_SRR10363558	Cryptosporidium parvum	5807	species	38.12
report_kraken2_Trim_SRR10363558	Cryptosporidium hominis	237895	species	1.16
report_kraken2_Trim_SRR10363558	Cryptosporidium tyzzeri	756076	species	0.82
report_kraken2_Trim_SRR10363559	Cryptosporidium parvum	5807	species	79.17
report_kraken2_Trim_SRR10363559	Cryptosporidium hominis	237895	species	2.11
report_kraken2_Trim_SRR10363559	Cryptosporidium tyzzeri	756076	species	1.56
report_kraken2_Trim_SRR10363561	Cryptosporidium parvum	5807	species	91.84
report_kraken2_Trim_SRR10363561	Cryptosporidium hominis	237895	species	1.63
report_kraken2_Trim_SRR10363561	Cryptosporidium tyzzeri	756076	species	1.12
report_kraken2_Trim_SRR10363562	Cryptosporidium parvum	5807	species	89.86
report_kraken2_Trim_SRR10363562	Cryptosporidium hominis	237895	species	1.83
report_kraken2_Trim_SRR10363562	Cryptosporidium tyzzeri	756076	species	1.1
report_kraken2_Trim_SRR10363563	Cryptosporidium parvum	5807	species	91.14
report_kraken2_Trim_SRR10363563	Cryptosporidium hominis	237895	species	2.36
report_kraken2_Trim_SRR10363563	Cryptosporidium tyzzeri	756076	species	1.68
report_kraken2_Trim_SRR10363564	Cryptosporidium parvum	5807	species	92.15
report_kraken2_Trim_SRR10363564	Cryptosporidium hominis	237895	species	2.24
report_kraken2_Trim_SRR10363564	Cryptosporidium tyzzeri	756076	species	1.59
report_kraken2_Trim_SRR10363565	Cryptosporidium parvum	5807	species	91.78
report_kraken2_Trim_SRR10363565	Cryptosporidium hominis	237895	species	2.13
report_kraken2_Trim_SRR10363565	Cryptosporidium tyzzeri	756076	species	1.59
report_kraken2_Trim_SRR10363566	Cryptosporidium parvum	5807	species	56.18
report_kraken2_Trim_SRR10363566	Cryptosporidium hominis	237895	species	1.9
report_kraken2_Trim_SRR10363566	Cryptosporidium tyzzeri	756076	species	1.54
report_kraken2_Trim_SRR10363567	Cryptosporidium parvum	5807	species	81.17
report_kraken2_Trim_SRR10363567	Cryptosporidium hominis	237895	species	1.73
report_kraken2_Trim_SRR10363567	Cryptosporidium tyzzeri	756076	species	1.26
report_kraken2_Trim_SRR10363568	Cryptosporidium parvum	5807	species	91.56
report_kraken2_Trim_SRR10363568	Cryptosporidium hominis	237895	species	2.43
report_kraken2_Trim_SRR10363568	Cryptosporidium tyzzeri	756076	species	1.67
report_kraken2_Trim_SRR10363569	Cryptosporidium parvum	5807	species	92.32
report_kraken2_Trim_SRR10363569	Cryptosporidium hominis	237895	species	2.14
report_kraken2_Trim_SRR10363569	Cryptosporidium tyzzeri	756076	species	1.55
report_kraken2_Trim_SRR10363570	Cryptosporidium parvum	5807	species	92.04
report_kraken2_Trim_SRR10363570	Cryptosporidium hominis	237895	species	2.26
report_kraken2_Trim_SRR10363570	Cryptosporidium tyzzeri	756076	species	1.61
report_kraken2_Trim_SRR10363571	Cryptosporidium parvum	5807	species	92.08
report_kraken2_Trim_SRR10363571	Cryptosporidium hominis	237895	species	2.09
report_kraken2_Trim_SRR10363571	Cryptosporidium tyzzeri	756076	species	1.49

report_kraken2_Trim_SRR10363572	Cryptosporidium parvum	5807	species	87.9
report_kraken2_Trim_SRR10363572	Cryptosporidium hominis	237895	species	1.79
report_kraken2_Trim_SRR10363572	Cryptosporidium tyzzeri	756076	species	1.37
report_kraken2_Trim_SRR10363573	Cryptosporidium parvum	5807	species	91.36
report_kraken2_Trim_SRR10363573	Cryptosporidium hominis	237895	species	2.38
report_kraken2_Trim_SRR10363573	Cryptosporidium tyzzeri	756076	species	1.69
report_kraken2_Trim_SRR10363574	Cryptosporidium parvum	5807	species	90.83
report_kraken2_Trim_SRR10363574	Cryptosporidium hominis	237895	species	2.42
report_kraken2_Trim_SRR10363574	Cryptosporidium tyzzeri	756076	species	1.73
report_kraken2_Trim_SRR10363575	Cryptosporidium parvum	5807	species	91.78
report_kraken2_Trim_SRR10363575	Cryptosporidium hominis	237895	species	2.23
report_kraken2_Trim_SRR10363575	Cryptosporidium tyzzeri	756076	species	1.6
report_kraken2_Trim_SRR10363576	Cryptosporidium parvum	5807	species	92.21
report_kraken2_Trim_SRR10363576	Cryptosporidium hominis	237895	species	2.1
report_kraken2_Trim_SRR10363576	Cryptosporidium tyzzeri	756076	species	1.48
report_kraken2_Trim_SRR10363577	Cryptosporidium parvum	5807	species	9.09
report_kraken2_Trim_SRR10363577	Edhazardia aedis	70536	species	0.42
report_kraken2_Trim_SRR10363577	Cryptosporidium hominis	237895	species	0.26
report_kraken2_Trim_SRR10363578	Cryptosporidium parvum	5807	species	90.71
report_kraken2_Trim_SRR10363578	Cryptosporidium hominis	237895	species	1.81
report_kraken2_Trim_SRR10363578	Cryptosporidium tyzzeri	756076	species	1.34
report_kraken2_Trim_SRR10363579	Cryptosporidium parvum	5807	species	92.33
report_kraken2_Trim_SRR10363579	Cryptosporidium hominis	237895	species	2.1
report_kraken2_Trim_SRR10363579	Cryptosporidium tyzzeri	756076	species	1.4
report_kraken2_Trim_SRR10363580	Cryptosporidium parvum	5807	species	91.86
report_kraken2_Trim_SRR10363580	Cryptosporidium hominis	237895	species	2.42
report_kraken2_Trim_SRR10363580	Cryptosporidium tyzzeri	756076	species	1.57
report_kraken2_Trim_SRR10363581	Cryptosporidium parvum	5807	species	92.06
report_kraken2_Trim_SRR10363581	Cryptosporidium hominis	237895	species	2.17
report_kraken2_Trim_SRR10363581	Cryptosporidium tyzzeri	756076	species	1.55
report_kraken2_Trim_SRR10363582	Cryptosporidium parvum	5807	species	41.59
report_kraken2_Trim_SRR10363582	Trypanosoma cruzi	5693	species	1.02
report_kraken2_Trim_SRR10363582	Cryptosporidium hominis	237895	species	0.99
report_kraken2_Trim_SRR10363583	Cryptosporidium parvum	5807	species	91.69
report_kraken2_Trim_SRR10363583	Cryptosporidium hominis	237895	species	1.66
report_kraken2_Trim_SRR10363583	Cryptosporidium tyzzeri	756076	species	1.13
report_kraken2_Trim_SRR10363584	Cryptosporidium parvum	5807	species	84.24
report_kraken2_Trim_SRR10363584	Cryptosporidium hominis	237895	species	1.52
report_kraken2_Trim_SRR10363584	Cryptosporidium tyzzeri	756076	species	1.04
report_kraken2_Trim_SRR10363585	Cryptosporidium ubiquitum	857276	species	99.25

report_kraken2_Trim_SRR10363585	Cryptosporidium parvum	5807	species	0.14
report_kraken2_Trim_SRR10363585	Cryptosporidium hominis	237895	species	0.06
report_kraken2_Trim_SRR10363586	Cryptosporidium parvum	5807	species	57.67
report_kraken2_Trim_SRR10363586	Cryptosporidium hominis	237895	species	4.54
report_kraken2_Trim_SRR10363586	Cryptosporidium tyzzeri	756076	species	0.77
report_kraken2_Trim_SRR10363587	Cryptosporidium parvum	5807	species	89.19
report_kraken2_Trim_SRR10363587	Cryptosporidium hominis	237895	species	1.59
report_kraken2_Trim_SRR10363587	Cryptosporidium tyzzeri	756076	species	1.16
report_kraken2_Trim_SRR10363589	Cryptosporidium parvum	5807	species	81.9
report_kraken2_Trim_SRR10363589	Cryptosporidium hominis	237895	species	2.66
report_kraken2_Trim_SRR10363589	Cryptosporidium tyzzeri	756076	species	1.87
report_kraken2_Trim_SRR10363592	Cryptosporidium parvum	5807	species	91.95
report_kraken2_Trim_SRR10363592	Cryptosporidium hominis	237895	species	1.89
report_kraken2_Trim_SRR10363592	Cryptosporidium tyzzeri	756076	species	1.19
report_kraken2_Trim_SRR10363594	Cryptosporidium parvum	5807	species	93.18
report_kraken2_Trim_SRR10363594	Cryptosporidium hominis	237895	species	1.92
report_kraken2_Trim_SRR10363594	Cryptosporidium tyzzeri	756076	species	1.26
report_kraken2_Trim_SRR10363596	Cryptosporidium parvum	5807	species	91.89
report_kraken2_Trim_SRR10363596	Cryptosporidium hominis	237895	species	1.78
report_kraken2_Trim_SRR10363596	Cryptosporidium tyzzeri	756076	species	1.44
report_kraken2_Trim_SRR10363597	Cryptosporidium parvum	5807	species	51.54
report_kraken2_Trim_SRR10363597	Trypanosoma cruzi	5693	species	2.86
report_kraken2_Trim_SRR10363597	Cryptosporidium hominis	237895	species	1.11
report_kraken2_Trim_SRR10363598	Cryptosporidium parvum	5807	species	91.92
report_kraken2_Trim_SRR10363598	Cryptosporidium hominis	237895	species	2.1
report_kraken2_Trim_SRR10363598	Cryptosporidium tyzzeri	756076	species	1.33
report_kraken2_Trim_SRR10363599	Cryptosporidium parvum	5807	species	92.64
report_kraken2_Trim_SRR10363599	Cryptosporidium hominis	237895	species	2.08
report_kraken2_Trim_SRR10363599	Cryptosporidium tyzzeri	756076	species	1.27
report_kraken2_Trim_SRR10996883	Cryptosporidium parvum	5807	species	26.96
report_kraken2_Trim_SRR10996883	Cryptosporidium hominis	237895	species	1.12
report_kraken2_Trim_SRR10996883	Cryptosporidium tyzzeri	756076	species	0.91
report_kraken2_Trim_SRR10997270	Cryptosporidium parvum	5807	species	27.57
report_kraken2_Trim_SRR10997270	Aspergillus fumigatus	746128	species	6.16
report_kraken2_Trim_SRR10997270	Plasmodium falciparum	5833	species	3.88
report_kraken2_Trim_SRR10997286	Cryptosporidium parvum	5807	species	30.23
report_kraken2_Trim_SRR10997286	Aspergillus fumigatus	746128	species	9.44
report_kraken2_Trim_SRR10997286	Cryptosporidium hominis	237895	species	1.1
report_kraken2_Trim_SRR10997324	Cryptosporidium parvum	5807	species	31.7
report_kraken2_Trim_SRR10997324	Aspergillus fumigatus	746128	species	2.11

report_kraken2_Trim_SRR10997324	Plasmodium falciparum	5833	species	1.3
report_kraken2_Trim_SRR11516700	Cryptosporidium parvum	5807	species	84.59
report_kraken2_Trim_SRR11516700	Cryptosporidium hominis	237895	species	1.7
report_kraken2_Trim_SRR11516700	Cryptosporidium tyzzeri	756076	species	1.11
report_kraken2_Trim_SRR11516701	Cryptosporidium parvum	5807	species	88.66
report_kraken2_Trim_SRR11516701	Cryptosporidium hominis	237895	species	1.21
report_kraken2_Trim_SRR11516701	Cryptosporidium tyzzeri	756076	species	0.79
report_kraken2_Trim_SRR11516702	Cryptosporidium parvum	5807	species	92.46
report_kraken2_Trim_SRR11516702	Cryptosporidium hominis	237895	species	1.08
report_kraken2_Trim_SRR11516702	Cryptosporidium tyzzeri	756076	species	0.7
report_kraken2_Trim_SRR11516703	Cryptosporidium parvum	5807	species	91.04
report_kraken2_Trim_SRR11516703	Cryptosporidium hominis	237895	species	1.14
report_kraken2_Trim_SRR11516703	Cryptosporidium tyzzeri	756076	species	0.73
report_kraken2_Trim_SRR11817809	Cryptosporidium parvum	5807	species	85.24
report_kraken2_Trim_SRR11817809	Cryptosporidium hominis	237895	species	2.83
report_kraken2_Trim_SRR11817809	Cryptosporidium tyzzeri	756076	species	2.06
report_kraken2_Trim_SRR11817810	Cryptosporidium parvum	5807	species	86.05
report_kraken2_Trim_SRR11817810	Cryptosporidium hominis	237895	species	2.65
report_kraken2_Trim_SRR11817810	Cryptosporidium tyzzeri	756076	species	2.01
report_kraken2_Trim_SRR11817811	Cryptosporidium parvum	5807	species	85.18
report_kraken2_Trim_SRR11817811	Cryptosporidium hominis	237895	species	2.58
report_kraken2_Trim_SRR11817811	Cryptosporidium tyzzeri	756076	species	1.82
report_kraken2_Trim_SRR11817812	Cryptosporidium parvum	5807	species	86.57
report_kraken2_Trim_SRR11817812	Cryptosporidium hominis	237895	species	2.46
report_kraken2_Trim_SRR11817812	Cryptosporidium tyzzeri	756076	species	1.76
report_kraken2_Trim_SRR11817813	Cryptosporidium parvum	5807	species	86.5
report_kraken2_Trim_SRR11817813	Cryptosporidium hominis	237895	species	2.62
report_kraken2_Trim_SRR11817813	Cryptosporidium tyzzeri	756076	species	1.78
report_kraken2_Trim_SRR11817814	Cryptosporidium parvum	5807	species	86.16
report_kraken2_Trim_SRR11817814	Cryptosporidium hominis	237895	species	2.7
report_kraken2_Trim_SRR11817814	Cryptosporidium tyzzeri	756076	species	1.89
report_kraken2_Trim_SRR11817815	Cryptosporidium parvum	5807	species	81.79
report_kraken2_Trim_SRR11817815	Cryptosporidium hominis	237895	species	2.33
report_kraken2_Trim_SRR11817815	Cryptosporidium tyzzeri	756076	species	1.62
report_kraken2_Trim_SRR11817816	Cryptosporidium parvum	5807	species	87.26
report_kraken2_Trim_SRR11817816	Cryptosporidium hominis	237895	species	2.64
report_kraken2_Trim_SRR11817816	Cryptosporidium tyzzeri	756076	species	1.81
report_kraken2_Trim_SRR11817817	Cryptosporidium parvum	5807	species	76.61
report_kraken2_Trim_SRR11817817	Cryptosporidium hominis	237895	species	2.21
report_kraken2_Trim_SRR11817817	Cryptosporidium tyzzeri	756076	species	1.64

report_kraken2_Trim_SRR11817818	Cryptosporidium parvum	5807	species	78.09
report_kraken2_Trim_SRR11817818	Cryptosporidium hominis	237895	species	2.33
report_kraken2_Trim_SRR11817818	Cryptosporidium tyzzeri	756076	species	1.83
report_kraken2_Trim_SRR11817819	Cryptosporidium parvum	5807	species	79.35
report_kraken2_Trim_SRR11817819	Cryptosporidium hominis	237895	species	2.64
report_kraken2_Trim_SRR11817819	Cryptosporidium tyzzeri	756076	species	2.06
report_kraken2_Trim_SRR11817820	Cryptosporidium parvum	5807	species	85.27
report_kraken2_Trim_SRR11817820	Cryptosporidium hominis	237895	species	2.21
report_kraken2_Trim_SRR11817820	Cryptosporidium tyzzeri	756076	species	1.75
report_kraken2_Trim_SRR11817821	Cryptosporidium parvum	5807	species	85.17
report_kraken2_Trim_SRR11817821	Cryptosporidium hominis	237895	species	2.44
report_kraken2_Trim_SRR11817821	Cryptosporidium tyzzeri	756076	species	1.86
report_kraken2_Trim_SRR11817822	Cryptosporidium parvum	5807	species	80.13
report_kraken2_Trim_SRR11817822	Cryptosporidium hominis	237895	species	2.37
report_kraken2_Trim_SRR11817822	Cryptosporidium tyzzeri	756076	species	1.92
report_kraken2_Trim_SRR11817823	Cryptosporidium parvum	5807	species	85.94
report_kraken2_Trim_SRR11817823	Cryptosporidium hominis	237895	species	2.46
report_kraken2_Trim_SRR11817823	Cryptosporidium tyzzeri	756076	species	1.76
report_kraken2_Trim_SRR11818073	Cryptosporidium parvum	5807	species	86.47
report_kraken2_Trim_SRR11818073	Cryptosporidium hominis	237895	species	2.42
report_kraken2_Trim_SRR11818073	Cryptosporidium tyzzeri	756076	species	1.75
report_kraken2_Trim_SRR11818074	Cryptosporidium parvum	5807	species	85.51
report_kraken2_Trim_SRR11818074	Cryptosporidium hominis	237895	species	2.6
report_kraken2_Trim_SRR11818074	Cryptosporidium tyzzeri	756076	species	1.87
report_kraken2_Trim_SRR11818075	Cryptosporidium parvum	5807	species	85.96
report_kraken2_Trim_SRR11818075	Cryptosporidium hominis	237895	species	2.53
report_kraken2_Trim_SRR11818075	Cryptosporidium tyzzeri	756076	species	1.92
report_kraken2_Trim_SRR11818076	Cryptosporidium parvum	5807	species	84.14
report_kraken2_Trim_SRR11818076	Cryptosporidium hominis	237895	species	2.48
report_kraken2_Trim_SRR11818076	Cryptosporidium tyzzeri	756076	species	1.77
report_kraken2_Trim_SRR11818077	Cryptosporidium parvum	5807	species	63.76
report_kraken2_Trim_SRR11818077	Cryptosporidium hominis	237895	species	2.01
report_kraken2_Trim_SRR11818077	Cryptosporidium tyzzeri	756076	species	1.55
report_kraken2_Trim_SRR11818078	Cryptosporidium parvum	5807	species	87.09
report_kraken2_Trim_SRR11818078	Cryptosporidium hominis	237895	species	2.66
report_kraken2_Trim_SRR11818078	Cryptosporidium tyzzeri	756076	species	1.97
report_kraken2_Trim_SRR11818079	Cryptosporidium parvum	5807	species	64.91
report_kraken2_Trim_SRR11818079	Cryptosporidium hominis	237895	species	1.87
report_kraken2_Trim_SRR11818079	Cryptosporidium tyzzeri	756076	species	1.31
report_kraken2_Trim_SRR15694483	Cryptosporidium parvum	5807	species	90.35

report_kraken2_Trim_SRR15694483	Cryptosporidium hominis	237895	species	3.14
report_kraken2_Trim_SRR15694483	Cryptosporidium tyzzeri	756076	species	1.98
report_kraken2_Trim_SRR15694484	Cryptosporidium parvum	5807	species	91.02
report_kraken2_Trim_SRR15694484	Cryptosporidium hominis	237895	species	1.78
report_kraken2_Trim_SRR15694484	Cryptosporidium tyzzeri	756076	species	1.46
report_kraken2_Trim_SRR15694485	Cryptosporidium parvum	5807	species	63.23
report_kraken2_Trim_SRR15694485	Cryptosporidium hominis	237895	species	1.83
report_kraken2_Trim_SRR15694485	Cryptosporidium tyzzeri	756076	species	1.4
report_kraken2_Trim_SRR15694486	Cryptosporidium parvum	5807	species	91.83
report_kraken2_Trim_SRR15694486	Cryptosporidium hominis	237895	species	1.94
report_kraken2_Trim_SRR15694486	Cryptosporidium tyzzeri	756076	species	1.35
report_kraken2_Trim_SRR15694487	Cryptosporidium parvum	5807	species	92.18
report_kraken2_Trim_SRR15694487	Cryptosporidium hominis	237895	species	2.04
report_kraken2_Trim_SRR15694487	Cryptosporidium tyzzeri	756076	species	1.38
report_kraken2_Trim_SRR15694488	Cryptosporidium parvum	5807	species	93.11
report_kraken2_Trim_SRR15694488	Cryptosporidium hominis	237895	species	1.9
report_kraken2_Trim_SRR15694488	Cryptosporidium tyzzeri	756076	species	1.34
report_kraken2_Trim_SRR15694489	Cryptosporidium parvum	5807	species	80.77
report_kraken2_Trim_SRR15694489	Cryptosporidium hominis	237895	species	2.66
report_kraken2_Trim_SRR15694489	Cryptosporidium tyzzeri	756076	species	1.99
report_kraken2_Trim_SRR15694490	Cryptosporidium parvum	5807	species	91.1
report_kraken2_Trim_SRR15694490	Cryptosporidium hominis	237895	species	2.55
report_kraken2_Trim_SRR15694490	Cryptosporidium tyzzeri	756076	species	2.09
report_kraken2_Trim_SRR15694491	Cryptosporidium parvum	5807	species	90.95
report_kraken2_Trim_SRR15694491	Cryptosporidium hominis	237895	species	2.37
report_kraken2_Trim_SRR15694491	Cryptosporidium tyzzeri	756076	species	1.6
report_kraken2_Trim_SRR15694492	Cryptosporidium parvum	5807	species	72.54
report_kraken2_Trim_SRR15694492	Cryptosporidium hominis	237895	species	2.52
report_kraken2_Trim_SRR15694492	Cryptosporidium tyzzeri	756076	species	1.91
report_kraken2_Trim_SRR15694493	Cryptosporidium parvum	5807	species	75.25
report_kraken2_Trim_SRR15694493	Cryptosporidium hominis	237895	species	2.59
report_kraken2_Trim_SRR15694493	Cryptosporidium tyzzeri	756076	species	1.89
report_kraken2_Trim_SRR15694494	Cryptosporidium parvum	5807	species	90.42
report_kraken2_Trim_SRR15694494	Cryptosporidium hominis	237895	species	2.6
report_kraken2_Trim_SRR15694494	Cryptosporidium tyzzeri	756076	species	2.17
report_kraken2_Trim_SRR15694495	Cryptosporidium parvum	5807	species	91.21
report_kraken2_Trim_SRR15694495	Cryptosporidium hominis	237895	species	1.62
report_kraken2_Trim_SRR15694495	Cryptosporidium tyzzeri	756076	species	1.36
report_kraken2_Trim_SRR15694496	Cryptosporidium parvum	5807	species	92.33
report_kraken2_Trim_SRR15694496	Cryptosporidium hominis	237895	species	2.01

report_kraken2_Trim_SRR15694496	Cryptosporidium tyzzeri	756076	species	1.3
report_kraken2_Trim_SRR15694497	Cryptosporidium parvum	5807	species	41.52
report_kraken2_Trim_SRR15694497	Cryptosporidium hominis	237895	species	1.16
report_kraken2_Trim_SRR15694497	Cryptosporidium tyzzeri	756076	species	0.91
report_kraken2_Trim_SRR15694498	Cryptosporidium parvum	5807	species	76.25
report_kraken2_Trim_SRR15694498	Cryptosporidium hominis	237895	species	2.66
report_kraken2_Trim_SRR15694498	Cryptosporidium tyzzeri	756076	species	1.94
report_kraken2_Trim_SRR15694499	Cryptosporidium parvum	5807	species	77.52
report_kraken2_Trim_SRR15694499	Cryptosporidium hominis	237895	species	2.57
report_kraken2_Trim_SRR15694499	Cryptosporidium tyzzeri	756076	species	1.72
report_kraken2_Trim_SRR15694500	Cryptosporidium parvum	5807	species	91.84
report_kraken2_Trim_SRR15694500	Cryptosporidium hominis	237895	species	2.05
report_kraken2_Trim_SRR15694500	Cryptosporidium tyzzeri	756076	species	1.44
report_kraken2_Trim_SRR15694501	Cryptosporidium parvum	5807	species	90.94
report_kraken2_Trim_SRR15694501	Cryptosporidium hominis	237895	species	2.45
report_kraken2_Trim_SRR15694501	Cryptosporidium tyzzeri	756076	species	1.99
report_kraken2_Trim_SRR15694502	Cryptosporidium parvum	5807	species	52.97
report_kraken2_Trim_SRR15694502	Cryptosporidium hominis	237895	species	1.29
report_kraken2_Trim_SRR15694502	Cryptosporidium tyzzeri	756076	species	0.96
report_kraken2_Trim_SRR15694503	Cryptosporidium parvum	5807	species	92.49
report_kraken2_Trim_SRR15694503	Cryptosporidium hominis	237895	species	2.03
report_kraken2_Trim_SRR15694503	Cryptosporidium tyzzeri	756076	species	1.38
report_kraken2_Trim_SRR15694504	Cryptosporidium parvum	5807	species	92.68
report_kraken2_Trim_SRR15694504	Cryptosporidium hominis	237895	species	1.91
report_kraken2_Trim_SRR15694504	Cryptosporidium tyzzeri	756076	species	1.23
report_kraken2_Trim_SRR15694505	Cryptosporidium parvum	5807	species	91.67
report_kraken2_Trim_SRR15694505	Cryptosporidium hominis	237895	species	2.07
report_kraken2_Trim_SRR15694505	Cryptosporidium tyzzeri	756076	species	1.41
report_kraken2_Trim_SRR15694506	Cryptosporidium parvum	5807	species	91.36
report_kraken2_Trim_SRR15694506	Cryptosporidium hominis	237895	species	1.65
report_kraken2_Trim_SRR15694506	Cryptosporidium tyzzeri	756076	species	1.37
report_kraken2_Trim_SRR15694507	Cryptosporidium parvum	5807	species	60.75
report_kraken2_Trim_SRR15694507	Cryptosporidium hominis	237895	species	1.12
report_kraken2_Trim_SRR15694507	Cryptosporidium tyzzeri	756076	species	0.84
report_kraken2_Trim_SRR15694508	Cryptosporidium parvum	5807	species	92.32
report_kraken2_Trim_SRR15694508	Cryptosporidium hominis	237895	species	1.83
report_kraken2_Trim_SRR15694508	Cryptosporidium tyzzeri	756076	species	1.42
report_kraken2_Trim_SRR15694509	Cryptosporidium parvum	5807	species	84.31
report_kraken2_Trim_SRR15694509	Cryptosporidium hominis	237895	species	1.6
report_kraken2_Trim_SRR15694509	Cryptosporidium tyzzeri	756076	species	1.29

report_kraken2_Trim_SRR15694510	Cryptosporidium parvum	5807	species	91.72
report_kraken2_Trim_SRR15694510	Cryptosporidium hominis	237895	species	2.17
report_kraken2_Trim_SRR15694510	Cryptosporidium tyzzeri	756076	species	1.57
report_kraken2_Trim_SRR15694511	Cryptosporidium parvum	5807	species	92.33
report_kraken2_Trim_SRR15694511	Cryptosporidium hominis	237895	species	1.76
report_kraken2_Trim_SRR15694511	Cryptosporidium tyzzeri	756076	species	1.36
report_kraken2_Trim_SRR15694512	Cryptosporidium parvum	5807	species	91.85
report_kraken2_Trim_SRR15694512	Cryptosporidium hominis	237895	species	2.22
report_kraken2_Trim_SRR15694512	Cryptosporidium tyzzeri	756076	species	1.58
report_kraken2_Trim_SRR15694513	Cryptosporidium parvum	5807	species	90.2
report_kraken2_Trim_SRR15694513	Cryptosporidium hominis	237895	species	2.52
report_kraken2_Trim_SRR15694513	Cryptosporidium tyzzeri	756076	species	2.04
report_kraken2_Trim_SRR15694514	Cryptosporidium parvum	5807	species	91.89
report_kraken2_Trim_SRR15694514	Cryptosporidium hominis	237895	species	1.78
report_kraken2_Trim_SRR15694514	Cryptosporidium tyzzeri	756076	species	1.44
report_kraken2_Trim_SRR15694515	Cryptosporidium parvum	5807	species	7.14
report_kraken2_Trim_SRR15694515	Monocercomonoides exilis	2049356	species	6.94
report_kraken2_Trim_SRR15694515	Cryptosporidium hominis	237895	species	0.17
report_kraken2_Trim_SRR15694516	Cryptosporidium parvum	5807	species	39.16
report_kraken2_Trim_SRR15694516	Cryptosporidium hominis	237895	species	0.82
report_kraken2_Trim_SRR15694516	Cryptosporidium tyzzeri	756076	species	0.6
report_kraken2_Trim_SRR15694517	Cryptosporidium parvum	5807	species	83.08
report_kraken2_Trim_SRR15694517	Cryptosporidium hominis	237895	species	1.46
report_kraken2_Trim_SRR15694517	Cryptosporidium tyzzeri	756076	species	1.27
report_kraken2_Trim_SRR15694518	Cryptosporidium parvum	5807	species	90.4
report_kraken2_Trim_SRR15694518	Cryptosporidium tyzzeri	756076	species	2.29
report_kraken2_Trim_SRR15694518	Cryptosporidium hominis	237895	species	2.15
report_kraken2_Trim_SRR15694519	Cryptosporidium parvum	5807	species	92.21
report_kraken2_Trim_SRR15694519	Cryptosporidium hominis	237895	species	2.13
report_kraken2_Trim_SRR15694519	Cryptosporidium tyzzeri	756076	species	1.5
report_kraken2_Trim_SRR15694520	Cryptosporidium baileyi	27987	species	0.12
report_kraken2_Trim_SRR15694520	Sordaria macrospora	5147	species	0.09
report_kraken2_Trim_SRR15694520	Cryptosporidium ubiquitum	857276	species	0.05
report_kraken2_Trim_SRR15694521	Cryptosporidium parvum	5807	species	14.53
report_kraken2_Trim_SRR15694521	Cryptosporidium hominis	237895	species	0.37
report_kraken2_Trim_SRR15694521	Cryptosporidium tyzzeri	756076	species	0.18
report_kraken2_Trim_SRR15694522	Cryptosporidium parvum	5807	species	91.14
report_kraken2_Trim_SRR15694522	Cryptosporidium hominis	237895	species	2.36
report_kraken2_Trim_SRR15694522	Cryptosporidium tyzzeri	756076	species	1.68
report_kraken2_Trim_SRR15694523	Cryptosporidium parvum	5807	species	92.15

report_kraken2_Trim_SRR15694523	Cryptosporidium hominis	237895	species	2.24
report_kraken2_Trim_SRR15694523	Cryptosporidium tyzzeri	756076	species	1.59
report_kraken2_Trim_SRR15694524	Cryptosporidium parvum	5807	species	91.78
report_kraken2_Trim_SRR15694524	Cryptosporidium hominis	237895	species	2.13
report_kraken2_Trim_SRR15694524	Cryptosporidium tyzzeri	756076	species	1.59
report_kraken2_Trim_SRR15694525	Cryptosporidium parvum	5807	species	81.17
report_kraken2_Trim_SRR15694525	Cryptosporidium hominis	237895	species	1.73
report_kraken2_Trim_SRR15694525	Cryptosporidium tyzzeri	756076	species	1.26
report_kraken2_Trim_SRR15694526	Cryptosporidium parvum	5807	species	91.56
report_kraken2_Trim_SRR15694526	Cryptosporidium hominis	237895	species	2.43
report_kraken2_Trim_SRR15694526	Cryptosporidium tyzzeri	756076	species	1.67
report_kraken2_Trim_SRR15694527	Cryptosporidium parvum	5807	species	92.32
report_kraken2_Trim_SRR15694527	Cryptosporidium hominis	237895	species	2.14
report_kraken2_Trim_SRR15694527	Cryptosporidium tyzzeri	756076	species	1.55
report_kraken2_Trim_SRR15694528	Cryptosporidium parvum	5807	species	92.12
report_kraken2_Trim_SRR15694528	Cryptosporidium hominis	237895	species	1.39
report_kraken2_Trim_SRR15694528	Cryptosporidium tyzzeri	756076	species	1.21
report_kraken2_Trim_SRR15694529	Cryptosporidium parvum	5807	species	92.04
report_kraken2_Trim_SRR15694529	Cryptosporidium hominis	237895	species	2.26
report_kraken2_Trim_SRR15694529	Cryptosporidium tyzzeri	756076	species	1.61
report_kraken2_Trim_SRR15694530	Cryptosporidium parvum	5807	species	92.08
report_kraken2_Trim_SRR15694530	Cryptosporidium hominis	237895	species	2.09
report_kraken2_Trim_SRR15694530	Cryptosporidium tyzzeri	756076	species	1.49
report_kraken2_Trim_SRR15694531	Cryptosporidium parvum	5807	species	87.9
report_kraken2_Trim_SRR15694531	Cryptosporidium hominis	237895	species	1.79
report_kraken2_Trim_SRR15694531	Cryptosporidium tyzzeri	756076	species	1.37
report_kraken2_Trim_SRR15694532	Cryptosporidium parvum	5807	species	91.36
report_kraken2_Trim_SRR15694532	Cryptosporidium hominis	237895	species	2.38
report_kraken2_Trim_SRR15694532	Cryptosporidium tyzzeri	756076	species	1.69
report_kraken2_Trim_SRR15694533	Cryptosporidium parvum	5807	species	90.83
report_kraken2_Trim_SRR15694533	Cryptosporidium hominis	237895	species	2.42
report_kraken2_Trim_SRR15694533	Cryptosporidium tyzzeri	756076	species	1.73
report_kraken2_Trim_SRR15694534	Cryptosporidium parvum	5807	species	91.78
report_kraken2_Trim_SRR15694534	Cryptosporidium hominis	237895	species	2.23
report_kraken2_Trim_SRR15694534	Cryptosporidium tyzzeri	756076	species	1.6
report_kraken2_Trim_SRR15694535	Cryptosporidium parvum	5807	species	92.21
report_kraken2_Trim_SRR15694535	Cryptosporidium hominis	237895	species	2.1
report_kraken2_Trim_SRR15694535	Cryptosporidium tyzzeri	756076	species	1.48
report_kraken2_Trim_SRR15694536	Cryptosporidium parvum	5807	species	90.71
report_kraken2_Trim_SRR15694536	Cryptosporidium hominis	237895	species	1.81

report_kraken2_Trim_SRR15694536	Cryptosporidium tyzzeri	756076	species	1.34
report_kraken2_Trim_SRR15694537	Cryptosporidium parvum	5807	species	92.33
report_kraken2_Trim_SRR15694537	Cryptosporidium hominis	237895	species	2.1
report_kraken2_Trim_SRR15694537	Cryptosporidium tyzzeri	756076	species	1.4
report_kraken2_Trim_SRR15694538	Cryptosporidium parvum	5807	species	91.86
report_kraken2_Trim_SRR15694538	Cryptosporidium hominis	237895	species	2.42
report_kraken2_Trim_SRR15694538	Cryptosporidium tyzzeri	756076	species	1.57
report_kraken2_Trim_SRR15694539	Cryptosporidium parvum	5807	species	90.68
report_kraken2_Trim_SRR15694539	Cryptosporidium hominis	237895	species	1.7
report_kraken2_Trim_SRR15694539	Cryptosporidium tyzzeri	756076	species	1.38
report_kraken2_Trim_SRR15694540	Cryptosporidium parvum	5807	species	92.06
report_kraken2_Trim_SRR15694540	Cryptosporidium hominis	237895	species	2.17
report_kraken2_Trim_SRR15694540	Cryptosporidium tyzzeri	756076	species	1.55
report_kraken2_Trim_SRR15694541	Cryptosporidium parvum	5807	species	41.59
report_kraken2_Trim_SRR15694541	Trypanosoma cruzi	5693	species	1.02
report_kraken2_Trim_SRR15694541	Cryptosporidium hominis	237895	species	0.99
report_kraken2_Trim_SRR15694542	Cryptosporidium parvum	5807	species	92.1
report_kraken2_Trim_SRR15694542	Cryptosporidium hominis	237895	species	1.87
report_kraken2_Trim_SRR15694542	Cryptosporidium tyzzeri	756076	species	1.45
report_kraken2_Trim_SRR15694543	Cryptosporidium parvum	5807	species	92.02
report_kraken2_Trim_SRR15694543	Cryptosporidium hominis	237895	species	2.11
report_kraken2_Trim_SRR15694543	Cryptosporidium tyzzeri	756076	species	1.48
report_kraken2_Trim_SRR15694544	Cryptosporidium parvum	5807	species	90.97
report_kraken2_Trim_SRR15694544	Cryptosporidium hominis	237895	species	1.9
report_kraken2_Trim_SRR15694544	Cryptosporidium tyzzeri	756076	species	1.32
report_kraken2_Trim_SRR15694545	Cryptosporidium parvum	5807	species	92.62
report_kraken2_Trim_SRR15694545	Cryptosporidium hominis	237895	species	2.09
report_kraken2_Trim_SRR15694545	Cryptosporidium tyzzeri	756076	species	1.43
report_kraken2_Trim_SRR15694546	Cryptosporidium parvum	5807	species	92.84
report_kraken2_Trim_SRR15694546	Cryptosporidium hominis	237895	species	1.7
report_kraken2_Trim_SRR15694546	Cryptosporidium tyzzeri	756076	species	1.22
report_kraken2_Trim_SRR15694547	Cryptosporidium parvum	5807	species	92.42
report_kraken2_Trim_SRR15694547	Cryptosporidium hominis	237895	species	1.89
report_kraken2_Trim_SRR15694547	Cryptosporidium tyzzeri	756076	species	1.36
report_kraken2_Trim_SRR15694548	Cryptosporidium parvum	5807	species	92.38
report_kraken2_Trim_SRR15694548	Cryptosporidium hominis	237895	species	2.04
report_kraken2_Trim_SRR15694548	Cryptosporidium tyzzeri	756076	species	1.47
report_kraken2_Trim_SRR15694549	Cryptosporidium parvum	5807	species	92.51
report_kraken2_Trim_SRR15694549	Cryptosporidium hominis	237895	species	1.51
report_kraken2_Trim_SRR15694549	Cryptosporidium tyzzeri	756076	species	1.33

report_kraken2_Trim_SRR15694550	Cryptosporidium parvum	5807	species	91.61
report_kraken2_Trim_SRR15694550	Cryptosporidium hominis	237895	species	1.88
report_kraken2_Trim_SRR15694550	Cryptosporidium tyzzeri	756076	species	1.4
report_kraken2_Trim_SRR15694551	Cryptosporidium parvum	5807	species	91.34
report_kraken2_Trim_SRR15694551	Cryptosporidium hominis	237895	species	1.5
report_kraken2_Trim_SRR15694551	Cryptosporidium tyzzeri	756076	species	1.28
report_kraken2_Trim_SRR15694552	Cryptosporidium parvum	5807	species	92.98
report_kraken2_Trim_SRR15694552	Cryptosporidium hominis	237895	species	1.6
report_kraken2_Trim_SRR15694552	Cryptosporidium tyzzeri	756076	species	1.22
report_kraken2_Trim_SRR15694553	Cryptosporidium parvum	5807	species	0.97
report_kraken2_Trim_SRR15694553	Acanthamoeba palestinensis	28015	species	0.37
report_kraken2_Trim_SRR15694553	Enterocytozoon bieneusi	31281	species	0.15
report_kraken2_Trim_SRR15694554	Cryptosporidium ubiquitum	857276	species	22.02
report_kraken2_Trim_SRR15694554	Cryptosporidium parvum	5807	species	0.49
report_kraken2_Trim_SRR15694554	Phytophthora sojae	67593	species	0.04
report_kraken2_Trim_SRR15694555	Cryptosporidium parvum	5807	species	91.57
report_kraken2_Trim_SRR15694555	Cryptosporidium hominis	237895	species	1.71
report_kraken2_Trim_SRR15694555	Cryptosporidium tyzzeri	756076	species	1.47
report_kraken2_Trim_SRR15694556	Cryptosporidium parvum	5807	species	90.99
report_kraken2_Trim_SRR15694556	Cryptosporidium hominis	237895	species	1.72
report_kraken2_Trim_SRR15694556	Cryptosporidium tyzzeri	756076	species	1.47
report_kraken2_Trim_SRR15694557	Cryptosporidium parvum	5807	species	91.2
report_kraken2_Trim_SRR15694557	Cryptosporidium hominis	237895	species	1.71
report_kraken2_Trim_SRR15694557	Cryptosporidium tyzzeri	756076	species	1.43
report_kraken2_Trim_SRR15694558	Cryptosporidium parvum	5807	species	86.59
report_kraken2_Trim_SRR15694558	Cryptosporidium hominis	237895	species	1.61
report_kraken2_Trim_SRR15694558	Cryptosporidium tyzzeri	756076	species	1.32
report_kraken2_Trim_SRR15694559	Cryptosporidium parvum	5807	species	90.67
report_kraken2_Trim_SRR15694559	Cryptosporidium hominis	237895	species	1.72
report_kraken2_Trim_SRR15694559	Cryptosporidium tyzzeri	756076	species	1.49
report_kraken2_Trim_SRR15694560	Cryptosporidium parvum	5807	species	90.34
report_kraken2_Trim_SRR15694560	Cryptosporidium hominis	237895	species	1.76
report_kraken2_Trim_SRR15694560	Cryptosporidium tyzzeri	756076	species	1.51
report_kraken2_Trim_SRR15694561	Cryptosporidium parvum	5807	species	87.28
report_kraken2_Trim_SRR15694561	Cryptosporidium hominis	237895	species	1.75
report_kraken2_Trim_SRR15694561	Cryptosporidium tyzzeri	756076	species	1.29
report_kraken2_Trim_SRR15694562	Cryptosporidium parvum	5807	species	88.85
report_kraken2_Trim_SRR15694562	Cryptosporidium hominis	237895	species	1.75
report_kraken2_Trim_SRR15694562	Cryptosporidium tyzzeri	756076	species	1.48
report_kraken2_Trim_SRR15694563	Cryptosporidium parvum	5807	species	92.63

report_kraken2_Trim_SRR15694563	Cryptosporidium hominis	237895	species	1.52
report_kraken2_Trim_SRR15694563	Cryptosporidium tyzzeri	756076	species	1.33
report_kraken2_Trim_SRR15694564	Cryptosporidium parvum	5807	species	93.23
report_kraken2_Trim_SRR15694564	Cryptosporidium hominis	237895	species	1.39
report_kraken2_Trim_SRR15694564	Cryptosporidium tyzzeri	756076	species	1.2
report_kraken2_Trim_SRR15694565	Cryptosporidium parvum	5807	species	89.63
report_kraken2_Trim_SRR15694565	Cryptosporidium hominis	237895	species	1.96
report_kraken2_Trim_SRR15694565	Cryptosporidium tyzzeri	756076	species	1.63
report_kraken2_Trim_SRR15694566	Cryptosporidium parvum	5807	species	91.62
report_kraken2_Trim_SRR15694566	Cryptosporidium hominis	237895	species	1.7
report_kraken2_Trim_SRR15694566	Cryptosporidium tyzzeri	756076	species	1.42
report_kraken2_Trim_SRR15694567	Cryptosporidium parvum	5807	species	90.99
report_kraken2_Trim_SRR15694567	Cryptosporidium hominis	237895	species	1.52
report_kraken2_Trim_SRR15694567	Cryptosporidium tyzzeri	756076	species	1.27
report_kraken2_Trim_SRR15694568	Cryptosporidium parvum	5807	species	91.8
report_kraken2_Trim_SRR15694568	Cryptosporidium hominis	237895	species	1.55
report_kraken2_Trim_SRR15694568	Cryptosporidium tyzzeri	756076	species	1.31
report_kraken2_Trim_SRR15694569	Cryptosporidium parvum	5807	species	91.84
report_kraken2_Trim_SRR15694569	Cryptosporidium hominis	237895	species	1.59
report_kraken2_Trim_SRR15694569	Cryptosporidium tyzzeri	756076	species	1.36
report_kraken2_Trim_SRR15694570	Cryptosporidium parvum	5807	species	91.63
report_kraken2_Trim_SRR15694570	Cryptosporidium hominis	237895	species	1.66
report_kraken2_Trim_SRR15694570	Cryptosporidium tyzzeri	756076	species	1.36
report_kraken2_Trim_SRR15694571	Cryptosporidium parvum	5807	species	19.23
report_kraken2_Trim_SRR15694571	Cryptosporidium hominis	237895	species	0.53
report_kraken2_Trim_SRR15694571	Cryptosporidium tyzzeri	756076	species	0.32
report_kraken2_Trim_SRR15694572	Cryptosporidium parvum	5807	species	92.17
report_kraken2_Trim_SRR15694572	Cryptosporidium hominis	237895	species	2.09
report_kraken2_Trim_SRR15694572	Cryptosporidium tyzzeri	756076	species	1.5
report_kraken2_Trim_SRR15694573	Cryptosporidium parvum	5807	species	92.36
report_kraken2_Trim_SRR15694573	Cryptosporidium hominis	237895	species	2.03
report_kraken2_Trim_SRR15694573	Cryptosporidium tyzzeri	756076	species	1.52
report_kraken2_Trim_SRR16770664	Cryptosporidium parvum	5807	species	87.27
report_kraken2_Trim_SRR16770664	Cryptosporidium hominis	237895	species	2.61
report_kraken2_Trim_SRR16770664	Cryptosporidium tyzzeri	756076	species	1.62
report_kraken2_Trim_SRR16990213	Cryptosporidium parvum	5807	species	87.33
report_kraken2_Trim_SRR16990213	Cryptosporidium hominis	237895	species	2.35
report_kraken2_Trim_SRR16990213	Cryptosporidium tyzzeri	756076	species	1.46
report_kraken2_Trim_SRR18395901	Cryptosporidium parvum	5807	species	84.66
report_kraken2_Trim_SRR18395901	Cryptosporidium hominis	237895	species	1.44

report_kraken2_Trim_SRR18395901	Cryptosporidium tyzzeri	756076	species	0.95
report_kraken2_Trim_SRR21763624	Cryptosporidium parvum	5807	species	82.15
report_kraken2_Trim_SRR21763624	Cryptosporidium hominis	237895	species	2.52
report_kraken2_Trim_SRR21763624	Cryptosporidium tyzzeri	756076	species	1.76
report_kraken2_Trim_SRR21763625	Cryptosporidium parvum	5807	species	84.86
report_kraken2_Trim_SRR21763625	Cryptosporidium hominis	237895	species	2.33
report_kraken2_Trim_SRR21763625	Cryptosporidium tyzzeri	756076	species	1.65
report_kraken2_Trim_SRR21763626	Cryptosporidium parvum	5807	species	76.27
report_kraken2_Trim_SRR21763626	Cryptosporidium hominis	237895	species	2.31
report_kraken2_Trim_SRR21763626	Cryptosporidium tyzzeri	756076	species	1.72
report_kraken2_Trim_SRR21763627	Cryptosporidium parvum	5807	species	50.78
report_kraken2_Trim_SRR21763627	Cryptosporidium hominis	237895	species	1.47
report_kraken2_Trim_SRR21763627	Cryptosporidium tyzzeri	756076	species	1.02
report_kraken2_Trim_SRR21763628	Cryptosporidium parvum	5807	species	77.23
report_kraken2_Trim_SRR21763628	Cryptosporidium hominis	237895	species	2.22
report_kraken2_Trim_SRR21763628	Cryptosporidium tyzzeri	756076	species	1.55
report_kraken2_Trim_SRR21763629	Cryptosporidium parvum	5807	species	83.09
report_kraken2_Trim_SRR21763629	Cryptosporidium hominis	237895	species	2.88
report_kraken2_Trim_SRR21763629	Cryptosporidium tyzzeri	756076	species	2.11
report_kraken2_Trim_SRR21763630	Cryptosporidium parvum	5807	species	19.27
report_kraken2_Trim_SRR21763630	Cryptosporidium hominis	237895	species	0.55
report_kraken2_Trim_SRR21763630	Cryptosporidium tyzzeri	756076	species	0.42
report_kraken2_Trim_SRR21763631	Cryptosporidium parvum	5807	species	67.53
report_kraken2_Trim_SRR21763631	Cryptosporidium hominis	237895	species	2.05
report_kraken2_Trim_SRR21763631	Cryptosporidium tyzzeri	756076	species	1.54
report_kraken2_Trim_SRR21763632	Cryptosporidium parvum	5807	species	71.67
report_kraken2_Trim_SRR21763632	Cryptosporidium hominis	237895	species	2.24
report_kraken2_Trim_SRR21763632	Cryptosporidium tyzzeri	756076	species	1.57
report_kraken2_Trim_SRR21763633	Cryptosporidium parvum	5807	species	83.95
report_kraken2_Trim_SRR21763633	Cryptosporidium hominis	237895	species	2.41
report_kraken2_Trim_SRR21763633	Cryptosporidium tyzzeri	756076	species	1.71
report_kraken2_Trim_SRR21763634	Cryptosporidium parvum	5807	species	83.42
report_kraken2_Trim_SRR21763634	Cryptosporidium hominis	237895	species	2.63
report_kraken2_Trim_SRR21763634	Cryptosporidium tyzzeri	756076	species	1.83
report_kraken2_Trim_SRR21763635	Cryptosporidium parvum	5807	species	39.84
report_kraken2_Trim_SRR21763635	Cryptosporidium hominis	237895	species	1.15
report_kraken2_Trim_SRR21763635	Cryptosporidium tyzzeri	756076	species	0.88
report_kraken2_Trim_SRR21763636	Cryptosporidium parvum	5807	species	80.56
report_kraken2_Trim_SRR21763636	Cryptosporidium hominis	237895	species	2.34
report_kraken2_Trim_SRR21763636	Cryptosporidium tyzzeri	756076	species	1.65

report_kraken2_Trim_SRR25512130	Cryptosporidium parvum	5807	species	85.21
report_kraken2_Trim_SRR25512130	Cryptosporidium hominis	237895	species	2.73
report_kraken2_Trim_SRR25512130	Cryptosporidium tyzzeri	756076	species	1.77
report_kraken2_Trim_SRR26320644	Cryptosporidium parvum	5807	species	86.03
report_kraken2_Trim_SRR26320644	Cryptosporidium hominis	237895	species	2.06
report_kraken2_Trim_SRR26320644	Cryptosporidium tyzzeri	756076	species	1.5
report_kraken2_Trim_SRR26320645	Cryptosporidium parvum	5807	species	36.23
report_kraken2_Trim_SRR26320645	Cryptosporidium hominis	237895	species	0.91
report_kraken2_Trim_SRR26320645	Cryptosporidium tyzzeri	756076	species	0.64
report_kraken2_Trim_SRR26320646	Cryptosporidium parvum	5807	species	11.1
report_kraken2_Trim_SRR26320646	Kwoniella heveanensis	89924	species	0.3
report_kraken2_Trim_SRR26320646	Cryptosporidium hominis	237895	species	0.28
report_kraken2_Trim_SRR26320647	Cryptosporidium parvum	5807	species	73.74
report_kraken2_Trim_SRR26320647	Cryptosporidium hominis	237895	species	1.5
report_kraken2_Trim_SRR26320647	Cryptosporidium tyzzeri	756076	species	1.11
report_kraken2_Trim_SRR26320648	Cryptosporidium parvum	5807	species	83.64
report_kraken2_Trim_SRR26320648	Cryptosporidium hominis	237895	species	1.83
report_kraken2_Trim_SRR26320648	Cryptosporidium tyzzeri	756076	species	1.3
report_kraken2_Trim_SRR26320649	Cryptosporidium parvum	5807	species	87.67
report_kraken2_Trim_SRR26320649	Cryptosporidium hominis	237895	species	1.94
report_kraken2_Trim_SRR26320649	Cryptosporidium tyzzeri	756076	species	1.4
report_kraken2_Trim_SRR26320650	Cryptosporidium parvum	5807	species	0.31
report_kraken2_Trim_SRR26320650	Cryptosporidium hominis	237895	species	0.07
report_kraken2_Trim_SRR26320650	Sarcocystis neurona	42890	species	0.04
report_kraken2_Trim_SRR26320651	Cryptosporidium parvum	5807	species	87.82
report_kraken2_Trim_SRR26320651	Cryptosporidium hominis	237895	species	2
report_kraken2_Trim_SRR26320651	Cryptosporidium tyzzeri	756076	species	1.42
report_kraken2_Trim_SRR26320652	Cryptosporidium parvum	5807	species	72.27
report_kraken2_Trim_SRR26320652	Cryptosporidium hominis	237895	species	1.58
report_kraken2_Trim_SRR26320652	Cryptosporidium tyzzeri	756076	species	1.27
report_kraken2_Trim_SRR26320653	Cryptosporidium parvum	5807	species	87.37
report_kraken2_Trim_SRR26320653	Cryptosporidium hominis	237895	species	2.22
report_kraken2_Trim_SRR26320653	Cryptosporidium tyzzeri	756076	species	1.69
report_kraken2_Trim_SRR26320654	Cryptosporidium parvum	5807	species	8.69
report_kraken2_Trim_SRR26320654	Acanthamoeba mauritaniensis	196912	species	2.08
report_kraken2_Trim_SRR26320654	Acanthamoeba palestinensis	28015	species	0.92
report_kraken2_Trim_SRR26320655	Cryptosporidium parvum	5807	species	87.64
report_kraken2_Trim_SRR26320655	Cryptosporidium hominis	237895	species	2.22
report_kraken2_Trim_SRR26320655	Cryptosporidium tyzzeri	756076	species	1.56
report_kraken2_Trim_SRR26320656	Cryptosporidium parvum	5807	species	86.37

report_kraken2_Trim_SRR26320656	Cryptosporidium hominis	237895	species	1.83
report_kraken2_Trim_SRR26320656	Cryptosporidium tyzzeri	756076	species	1.32
report_kraken2_Trim_SRR26320657	Cryptosporidium parvum	5807	species	82.86
report_kraken2_Trim_SRR26320657	Cryptosporidium hominis	237895	species	1.78
report_kraken2_Trim_SRR26320657	Cryptosporidium tyzzeri	756076	species	1.25
report_kraken2_Trim_SRR26320658	Cryptosporidium parvum	5807	species	43.18
report_kraken2_Trim_SRR26320658	Cryptosporidium hominis	237895	species	1.21
report_kraken2_Trim_SRR26320658	Cryptosporidium tyzzeri	756076	species	0.85
report_kraken2_Trim_SRR26320659	Hammondia hammondi	99158	species	0.15
report_kraken2_Trim_SRR26320659	Cryptosporidium parvum	5807	species	0.03
report_kraken2_Trim_SRR26320659	Sarcocystis neurona	42890	species	0.02
report_kraken2_Trim_SRR26320660	Cryptosporidium parvum	5807	species	18.54
report_kraken2_Trim_SRR26320660	Cryptosporidium hominis	237895	species	0.59
report_kraken2_Trim_SRR26320660	Cryptosporidium tyzzeri	756076	species	0.43
report_kraken2_Trim_SRR26320661	Cryptosporidium parvum	5807	species	4.39
report_kraken2_Trim_SRR26320661	Hammondia hammondi	99158	species	0.22
report_kraken2_Trim_SRR26320661	Cryptosporidium hominis	237895	species	0.13
report_kraken2_Trim_SRR26320662	Cryptosporidium parvum	5807	species	74.09
report_kraken2_Trim_SRR26320662	Cryptosporidium hominis	237895	species	1.49
report_kraken2_Trim_SRR26320662	Cryptosporidium tyzzeri	756076	species	1.07
report_kraken2_Trim_SRR26320663	Cryptosporidium parvum	5807	species	71.39
report_kraken2_Trim_SRR26320663	Cryptosporidium hominis	237895	species	1.5
report_kraken2_Trim_SRR26320663	Cryptosporidium tyzzeri	756076	species	1.09
report_kraken2_Trim_SRR26320664	Cryptosporidium parvum	5807	species	72.14
report_kraken2_Trim_SRR26320664	Cryptosporidium hominis	237895	species	1.51
report_kraken2_Trim_SRR26320664	Cryptosporidium tyzzeri	756076	species	1.08
report_kraken2_Trim_SRR26320665	Cryptosporidium parvum	5807	species	76.59
report_kraken2_Trim_SRR26320665	Cryptosporidium hominis	237895	species	1.51
report_kraken2_Trim_SRR26320665	Cryptosporidium tyzzeri	756076	species	1.07
report_kraken2_Trim_SRR26320666	Cryptosporidium parvum	5807	species	73.8
report_kraken2_Trim_SRR26320666	Cryptosporidium hominis	237895	species	1.46
report_kraken2_Trim_SRR26320666	Cryptosporidium tyzzeri	756076	species	1.06
report_kraken2_Trim_SRR26320667	Cryptosporidium parvum	5807	species	77.39
report_kraken2_Trim_SRR26320667	Cryptosporidium hominis	237895	species	1.57
report_kraken2_Trim_SRR26320667	Cryptosporidium tyzzeri	756076	species	1.25
report_kraken2_Trim_SRR26320669	Cryptosporidium parvum	5807	species	71.86
report_kraken2_Trim_SRR26320669	Cryptosporidium hominis	237895	species	1.46
report_kraken2_Trim_SRR26320669	Cryptosporidium tyzzeri	756076	species	1.15
report_kraken2_Trim_SRR26320670	Cryptosporidium parvum	5807	species	75.82
report_kraken2_Trim_SRR26320670	Cryptosporidium hominis	237895	species	1.64

report_kraken2_Trim_SRR26320670	Cryptosporidium tyzzeri	756076	species	1.27
report_kraken2_Trim_SRR26320671	Cryptosporidium parvum	5807	species	75.04
report_kraken2_Trim_SRR26320671	Cryptosporidium hominis	237895	species	1.46
report_kraken2_Trim_SRR26320671	Cryptosporidium tyzzeri	756076	species	1.17
report_kraken2_Trim_SRR26320672	Cryptosporidium parvum	5807	species	74.62
report_kraken2_Trim_SRR26320672	Cryptosporidium hominis	237895	species	1.62
report_kraken2_Trim_SRR26320672	Cryptosporidium tyzzeri	756076	species	1.28
report_kraken2_Trim_SRR26320673	Cryptosporidium parvum	5807	species	75.79
report_kraken2_Trim_SRR26320673	Cryptosporidium hominis	237895	species	1.66
report_kraken2_Trim_SRR26320673	Cryptosporidium tyzzeri	756076	species	1.18
report_kraken2_Trim_SRR26320674	Cryptosporidium parvum	5807	species	77.15
report_kraken2_Trim_SRR26320674	Cryptosporidium hominis	237895	species	1.62
report_kraken2_Trim_SRR26320674	Cryptosporidium tyzzeri	756076	species	1.17
report_kraken2_Trim_SRR26320675	Cryptosporidium parvum	5807	species	77.68
report_kraken2_Trim_SRR26320675	Cryptosporidium hominis	237895	species	1.54
report_kraken2_Trim_SRR26320675	Cryptosporidium tyzzeri	756076	species	1.09
report_kraken2_Trim_SRR26320676	Cryptosporidium parvum	5807	species	9.53
report_kraken2_Trim_SRR26320676	Acanthamoeba mauritaniensis	196912	species	5.24
report_kraken2_Trim_SRR26320676	Acanthamoeba palestinensis	28015	species	2.12
report_kraken2_Trim_SRR26320677	Cryptosporidium parvum	5807	species	75.33
report_kraken2_Trim_SRR26320677	Cryptosporidium hominis	237895	species	1.52
report_kraken2_Trim_SRR26320677	Cryptosporidium tyzzeri	756076	species	1.13
report_kraken2_Trim_SRR26320678	Cryptosporidium parvum	5807	species	76.38
report_kraken2_Trim_SRR26320678	Cryptosporidium hominis	237895	species	1.6
report_kraken2_Trim_SRR26320678	Cryptosporidium tyzzeri	756076	species	1.17
report_kraken2_Trim_SRR26320679	Cryptosporidium parvum	5807	species	73.8
report_kraken2_Trim_SRR26320679	Cryptosporidium hominis	237895	species	1.48
report_kraken2_Trim_SRR26320679	Cryptosporidium tyzzeri	756076	species	1.17
report_kraken2_Trim_SRR26320680	Cryptosporidium parvum	5807	species	75.64
report_kraken2_Trim_SRR26320680	Cryptosporidium hominis	237895	species	1.59
report_kraken2_Trim_SRR26320680	Cryptosporidium tyzzeri	756076	species	1.14
report_kraken2_Trim_SRR26320681	Cryptosporidium parvum	5807	species	76.06
report_kraken2_Trim_SRR26320681	Cryptosporidium hominis	237895	species	1.48
report_kraken2_Trim_SRR26320681	Cryptosporidium tyzzeri	756076	species	1.05
report_kraken2_Trim_SRR26320682	Cryptosporidium parvum	5807	species	75.4
report_kraken2_Trim_SRR26320682	Cryptosporidium hominis	237895	species	1.68
report_kraken2_Trim_SRR26320682	Cryptosporidium tyzzeri	756076	species	1.3
report_kraken2_Trim_SRR26320683	Cryptosporidium parvum	5807	species	74.55
report_kraken2_Trim_SRR26320683	Cryptosporidium hominis	237895	species	1.59
report_kraken2_Trim_SRR26320683	Cryptosporidium tyzzeri	756076	species	1.15

report_kraken2_Trim_SRR26320684	Cryptosporidium parvum	5807	species	76.17
report_kraken2_Trim_SRR26320684	Cryptosporidium hominis	237895	species	1.45
report_kraken2_Trim_SRR26320684	Cryptosporidium tyzzeri	756076	species	1.03
report_kraken2_Trim_SRR26320685	Cryptosporidium parvum	5807	species	77.46
report_kraken2_Trim_SRR26320685	Cryptosporidium hominis	237895	species	1.69
report_kraken2_Trim_SRR26320685	Cryptosporidium tyzzeri	756076	species	1.17
report_kraken2_Trim_SRR26320686	Cryptosporidium parvum	5807	species	85.04
report_kraken2_Trim_SRR26320686	Cryptosporidium hominis	237895	species	2.45
report_kraken2_Trim_SRR26320686	Cryptosporidium tyzzeri	756076	species	1.7
report_kraken2_Trim_SRR26320687	Cryptosporidium parvum	5807	species	71.93
report_kraken2_Trim_SRR26320687	Cryptosporidium hominis	237895	species	1.36
report_kraken2_Trim_SRR26320687	Cryptosporidium tyzzeri	756076	species	0.97
report_kraken2_Trim_SRR26320688	Cryptosporidium parvum	5807	species	85.1
report_kraken2_Trim_SRR26320688	Cryptosporidium hominis	237895	species	2.61
report_kraken2_Trim_SRR26320688	Cryptosporidium tyzzeri	756076	species	1.92
report_kraken2_Trim_SRR26320689	Cryptosporidium parvum	5807	species	85.7
report_kraken2_Trim_SRR26320689	Cryptosporidium hominis	237895	species	2.36
report_kraken2_Trim_SRR26320689	Cryptosporidium tyzzeri	756076	species	1.68
report_kraken2_Trim_SRR26320690	Cryptosporidium parvum	5807	species	83.17
report_kraken2_Trim_SRR26320690	Cryptosporidium hominis	237895	species	2.86
report_kraken2_Trim_SRR26320690	Cryptosporidium tyzzeri	756076	species	1.99
report_kraken2_Trim_SRR26320691	Cryptosporidium parvum	5807	species	84.14
report_kraken2_Trim_SRR26320691	Cryptosporidium hominis	237895	species	2.64
report_kraken2_Trim_SRR26320691	Cryptosporidium tyzzeri	756076	species	1.81
report_kraken2_Trim_SRR26320692	Cryptosporidium parvum	5807	species	84.93
report_kraken2_Trim_SRR26320692	Cryptosporidium hominis	237895	species	2.52
report_kraken2_Trim_SRR26320692	Cryptosporidium tyzzeri	756076	species	1.75
report_kraken2_Trim_SRR26320693	Cryptosporidium parvum	5807	species	84.93
report_kraken2_Trim_SRR26320693	Cryptosporidium hominis	237895	species	2.58
report_kraken2_Trim_SRR26320693	Cryptosporidium tyzzeri	756076	species	1.78
report_kraken2_Trim_SRR26320694	Cryptosporidium parvum	5807	species	84.4
report_kraken2_Trim_SRR26320694	Cryptosporidium hominis	237895	species	2.58
report_kraken2_Trim_SRR26320694	Cryptosporidium tyzzeri	756076	species	1.78
report_kraken2_Trim_SRR26320695	Cryptosporidium parvum	5807	species	75.31
report_kraken2_Trim_SRR26320695	Cryptosporidium hominis	237895	species	1.59
report_kraken2_Trim_SRR26320695	Cryptosporidium tyzzeri	756076	species	1.07
report_kraken2_Trim_SRR26320696	Cryptosporidium parvum	5807	species	75.24
report_kraken2_Trim_SRR26320696	Cryptosporidium hominis	237895	species	1.5
report_kraken2_Trim_SRR26320696	Cryptosporidium tyzzeri	756076	species	1.03
report_kraken2_Trim_SRR26320697	Cryptosporidium parvum	5807	species	75.17

report_kraken2_Trim_SRR26320697	Cryptosporidium hominis	237895	species	1.6
report_kraken2_Trim_SRR26320697	Cryptosporidium tyzzeri	756076	species	1.07
report_kraken2_Trim_SRR26320698	Acanthamoeba palestinensis	28015	species	1.49
report_kraken2_Trim_SRR26320698	Acanthamoeba mauritaniensis	196912	species	1.08
report_kraken2_Trim_SRR26320698	Hammondia hammondi	99158	species	0.21
report_kraken2_Trim_SRR26320699	Cryptosporidium parvum	5807	species	76.46
report_kraken2_Trim_SRR26320699	Cryptosporidium hominis	237895	species	1.54
report_kraken2_Trim_SRR26320699	Cryptosporidium tyzzeri	756076	species	1.08
report_kraken2_Trim_SRR26320700	Cryptosporidium parvum	5807	species	75.27
report_kraken2_Trim_SRR26320700	Cryptosporidium hominis	237895	species	1.55
report_kraken2_Trim_SRR26320700	Cryptosporidium tyzzeri	756076	species	1.07
report_kraken2_Trim_SRR26320701	Cryptosporidium parvum	5807	species	75.02
report_kraken2_Trim_SRR26320701	Cryptosporidium hominis	237895	species	1.51
report_kraken2_Trim_SRR26320701	Cryptosporidium tyzzeri	756076	species	1.03
report_kraken2_Trim_SRR26320702	Cryptosporidium parvum	5807	species	75.14
report_kraken2_Trim_SRR26320702	Cryptosporidium hominis	237895	species	1.55
report_kraken2_Trim_SRR26320702	Cryptosporidium tyzzeri	756076	species	1.12
report_kraken2_Trim_SRR26320703	Cryptosporidium parvum	5807	species	0.14
report_kraken2_Trim_SRR26320703	Acanthamoeba mauritaniensis	196912	species	0.01
report_kraken2_Trim_SRR26320703	[Candida] haemuloni	45357	species	0
report_kraken2_Trim_SRR26320704	Cryptosporidium parvum	5807	species	3.53
report_kraken2_Trim_SRR26320704	Acanthamoeba palestinensis	28015	species	1
report_kraken2_Trim_SRR26320704	Acanthamoeba mauritaniensis	196912	species	0.57
report_kraken2_Trim_SRR26320705	Cryptosporidium parvum	5807	species	78.09
report_kraken2_Trim_SRR26320705	Cryptosporidium hominis	237895	species	2.03
report_kraken2_Trim_SRR26320705	Cryptosporidium tyzzeri	756076	species	1.46
report_kraken2_Trim_SRR26320706	Cryptosporidium parvum	5807	species	80.92
report_kraken2_Trim_SRR26320706	Cryptosporidium hominis	237895	species	2.14
report_kraken2_Trim_SRR26320706	Cryptosporidium tyzzeri	756076	species	1.54
report_kraken2_Trim_SRR26320707	Cryptosporidium parvum	5807	species	1.5
report_kraken2_Trim_SRR26320707	Phytophthora palmivora	4796	species	0.08
report_kraken2_Trim_SRR26320707	Cryptosporidium hominis	237895	species	0.05
report_kraken2_Trim_SRR26320708	Cryptosporidium parvum	5807	species	1.79
report_kraken2_Trim_SRR26320708	Cryptosporidium hominis	237895	species	0.06
report_kraken2_Trim_SRR26320708	Histoplasma capsulatum	5037	species	0.05
report_kraken2_Trim_SRR26320709	Cryptosporidium parvum	5807	species	50.69
report_kraken2_Trim_SRR26320709	Giardia intestinalis	5741	species	18.73
report_kraken2_Trim_SRR26320709	Cryptosporidium hominis	237895	species	1.16
report_kraken2_Trim_SRR26320710	Candida albicans	5476	species	0.06
report_kraken2_Trim_SRR26320710	Rhizopus delemar	936053	species	0.04

report_kraken2_Trim_SRR26320710	Tremella mesenterica	5217	species	0.03
report_kraken2_Trim_SRR26320711	Cryptosporidium parvum	5807	species	0.01
report_kraken2_Trim_SRR26320711	[Candida] auris	498019	species	0
report_kraken2_Trim_SRR26320711	Zymoseptoria tritici	1047171	species	0
report_kraken2_Trim_SRR26320712	Cryptosporidium parvum	5807	species	84.84
report_kraken2_Trim_SRR26320712	Cryptosporidium hominis	237895	species	2.45
report_kraken2_Trim_SRR26320712	Cryptosporidium tyzzeri	756076	species	1.74
report_kraken2_Trim_SRR26320713	Cryptosporidium parvum	5807	species	84.9
report_kraken2_Trim_SRR26320713	Cryptosporidium hominis	237895	species	2.42
report_kraken2_Trim_SRR26320713	Cryptosporidium tyzzeri	756076	species	1.69
report_kraken2_Trim_SRR26320714	Cryptosporidium parvum	5807	species	82.58
report_kraken2_Trim_SRR26320714	Cryptosporidium hominis	237895	species	2.24
report_kraken2_Trim_SRR26320714	Cryptosporidium tyzzeri	756076	species	1.59
report_kraken2_Trim_SRR26320715	Acanthamoeba mauritaniensis	196912	species	7.44
report_kraken2_Trim_SRR26320715	Acanthamoeba palestinensis	28015	species	3.3
report_kraken2_Trim_SRR26320715	Hammondia hammondi	99158	species	0.24
report_kraken2_Trim_SRR26320716	Cryptosporidium parvum	5807	species	4.72
report_kraken2_Trim_SRR26320716	Cryptosporidium hominis	237895	species	0.18
report_kraken2_Trim_SRR26320716	Cryptosporidium tyzzeri	756076	species	0.12
report_kraken2_Trim_SRR26320717	Cryptosporidium parvum	5807	species	79.88
report_kraken2_Trim_SRR26320717	Cryptosporidium hominis	237895	species	2.08
report_kraken2_Trim_SRR26320717	Cryptosporidium tyzzeri	756076	species	1.47
report_kraken2_Trim_SRR26320718	Cryptosporidium parvum	5807	species	61.2
report_kraken2_Trim_SRR26320718	Giardia intestinalis	5741	species	13.55
report_kraken2_Trim_SRR26320718	Cryptosporidium hominis	237895	species	1.68
report_kraken2_Trim_SRR26320719	Cryptosporidium parvum	5807	species	84.05
report_kraken2_Trim_SRR26320719	Cryptosporidium hominis	237895	species	2.48
report_kraken2_Trim_SRR26320719	Cryptosporidium tyzzeri	756076	species	1.68
report_kraken2_Trim_SRR26320720	Cryptosporidium parvum	5807	species	45.13
report_kraken2_Trim_SRR26320720	Yarrowia lipolytica	4952	species	11.35
report_kraken2_Trim_SRR26320720	Cryptosporidium hominis	237895	species	0.87
report_kraken2_Trim_SRR26320721	Cryptosporidium parvum	5807	species	69.77
report_kraken2_Trim_SRR26320721	Cryptosporidium hominis	237895	species	1.34
report_kraken2_Trim_SRR26320721	Cryptosporidium tyzzeri	756076	species	0.98
report_kraken2_Trim_SRR26320722	Cryptosporidium parvum	5807	species	73.48
report_kraken2_Trim_SRR26320722	Cryptosporidium hominis	237895	species	1.39
report_kraken2_Trim_SRR26320722	Cryptosporidium tyzzeri	756076	species	0.99
report_kraken2_Trim_SRR26320723	Cryptosporidium parvum	5807	species	58.44
report_kraken2_Trim_SRR26320723	Cryptosporidium hominis	237895	species	1.07
report_kraken2_Trim_SRR26320723	Cryptosporidium tyzzeri	756076	species	0.77

report_kraken2_Trim_SRR26320724	Cryptosporidium parvum	5807	species	11.57
report_kraken2_Trim_SRR26320724	Cryptosporidium hominis	237895	species	0.21
report_kraken2_Trim_SRR26320724	Cryptosporidium tyzzeri	756076	species	0.15
report_kraken2_Trim_SRR26320725	Cryptosporidium parvum	5807	species	3.32
report_kraken2_Trim_SRR26320725	Acanthamoeba palestinensis	28015	species	1.45
report_kraken2_Trim_SRR26320725	Acanthamoeba mauritaniensis	196912	species	0.94
report_kraken2_Trim_SRR26320726	Cryptosporidium parvum	5807	species	73.33
report_kraken2_Trim_SRR26320726	Cryptosporidium hominis	237895	species	1.45
report_kraken2_Trim_SRR26320726	Cryptosporidium tyzzeri	756076	species	1.17
report_kraken2_Trim_SRR26320727	Cryptosporidium parvum	5807	species	74.05
report_kraken2_Trim_SRR26320727	Cryptosporidium hominis	237895	species	1.34
report_kraken2_Trim_SRR26320727	Cryptosporidium tyzzeri	756076	species	0.95
report_kraken2_Trim_SRR26320728	Giardia intestinalis	5741	species	42.84
report_kraken2_Trim_SRR26320728	Cryptosporidium parvum	5807	species	0.46
report_kraken2_Trim_SRR26320728	Cryptosporidium tyzzeri	756076	species	0.01
report_kraken2_Trim_SRR26320729	Cryptosporidium parvum	5807	species	44.13
report_kraken2_Trim_SRR26320729	Giardia intestinalis	5741	species	22.51
report_kraken2_Trim_SRR26320729	Cryptosporidium hominis	237895	species	1.25
report_kraken2_Trim_SRR26320730	Cryptosporidium parvum	5807	species	0.77
report_kraken2_Trim_SRR26320730	Giardia intestinalis	5741	species	0.25
report_kraken2_Trim_SRR26320730	Theileria parva	5875	species	0.07
report_kraken2_Trim_SRR26320731	Cryptosporidium parvum	5807	species	81.68
report_kraken2_Trim_SRR26320731	Cryptosporidium hominis	237895	species	2.61
report_kraken2_Trim_SRR26320731	Cryptosporidium tyzzeri	756076	species	1.79
report_kraken2_Trim_SRR26320732	Cryptosporidium parvum	5807	species	83.73
report_kraken2_Trim_SRR26320732	Cryptosporidium hominis	237895	species	2.56
report_kraken2_Trim_SRR26320732	Cryptosporidium tyzzeri	756076	species	1.8
report_kraken2_Trim_SRR26320733	Cryptosporidium parvum	5807	species	83.31
report_kraken2_Trim_SRR26320733	Cryptosporidium hominis	237895	species	2.4
report_kraken2_Trim_SRR26320733	Cryptosporidium tyzzeri	756076	species	1.64
report_kraken2_Trim_SRR26320734	Cryptosporidium parvum	5807	species	80.19
report_kraken2_Trim_SRR26320734	Giardia intestinalis	5741	species	2.65
report_kraken2_Trim_SRR26320734	Cryptosporidium hominis	237895	species	2.34
report_kraken2_Trim_SRR26320735	Cryptosporidium parvum	5807	species	53.72
report_kraken2_Trim_SRR26320735	Cryptosporidium hominis	237895	species	1.5
report_kraken2_Trim_SRR26320735	Cryptosporidium tyzzeri	756076	species	1.04
report_kraken2_Trim_SRR26320737	Giardia intestinalis	5741	species	33.59
report_kraken2_Trim_SRR26320737	Cryptosporidium parvum	5807	species	30.11
report_kraken2_Trim_SRR26320737	Cryptosporidium hominis	237895	species	0.93
report_kraken2_Trim_SRR26320738	Acanthamoeba palestinensis	28015	species	0.78

report_kraken2_Trim_SRR26320738	Acanthamoeba mauritaniensis	196912	species	0.43
report_kraken2_Trim_SRR26320738	Hammondia hammondi	99158	species	0.22
report_kraken2_Trim_SRR26320739	Cryptosporidium parvum	5807	species	15.43
report_kraken2_Trim_SRR26320739	Acanthamoeba palestinensis	28015	species	0.6
report_kraken2_Trim_SRR26320739	Acanthamoeba mauritaniensis	196912	species	0.4
report_kraken2_Trim_SRR26320740	Cryptosporidium parvum	5807	species	40.31
report_kraken2_Trim_SRR26320740	Cryptosporidium hominis	237895	species	0.9
report_kraken2_Trim_SRR26320740	Cryptosporidium tyzzeri	756076	species	0.64
report_kraken2_Trim_SRR26320741	Cryptosporidium parvum	5807	species	87.25
report_kraken2_Trim_SRR26320741	Cryptosporidium hominis	237895	species	2.09
report_kraken2_Trim_SRR26320741	Cryptosporidium tyzzeri	756076	species	1.63
report_kraken2_Trim_SRR26320742	Cryptosporidium parvum	5807	species	87.44
report_kraken2_Trim_SRR26320742	Cryptosporidium hominis	237895	species	2.03
report_kraken2_Trim_SRR26320742	Cryptosporidium tyzzeri	756076	species	1.43
report_kraken2_Trim_SRR26320743	Cryptosporidium parvum	5807	species	87.52
report_kraken2_Trim_SRR26320743	Cryptosporidium hominis	237895	species	2.01
report_kraken2_Trim_SRR26320743	Cryptosporidium tyzzeri	756076	species	1.39
report_kraken2_Trim_SRR26320744	Cryptosporidium parvum	5807	species	15.21
report_kraken2_Trim_SRR26320744	Cryptosporidium hominis	237895	species	0.39
report_kraken2_Trim_SRR26320744	Kwoniella heveanensis	89924	species	0.27
report_kraken2_Trim_SRR26320745	Cryptosporidium parvum	5807	species	84.4
report_kraken2_Trim_SRR26320745	Cryptosporidium hominis	237895	species	2.12
report_kraken2_Trim_SRR26320745	Cryptosporidium tyzzeri	756076	species	1.49
report_kraken2_Trim_SRR26320746	Cryptosporidium parvum	5807	species	38.65
report_kraken2_Trim_SRR26320746	Cryptosporidium hominis	237895	species	0.82
report_kraken2_Trim_SRR26320746	Cryptosporidium tyzzeri	756076	species	0.59
report_kraken2_Trim_SRR26320747	Cryptosporidium parvum	5807	species	3.31
report_kraken2_Trim_SRR26320747	Kwoniella heveanensis	89924	species	0.38
report_kraken2_Trim_SRR26320747	Kwoniella bestiolae	324769	species	0.2
report_kraken2_Trim_SRR26320748	Cryptosporidium parvum	5807	species	87.63
report_kraken2_Trim_SRR26320748	Cryptosporidium hominis	237895	species	2.14
report_kraken2_Trim_SRR26320748	Cryptosporidium tyzzeri	756076	species	1.49
report_kraken2_Trim_SRR26320749	Cryptosporidium parvum	5807	species	87.47
report_kraken2_Trim_SRR26320749	Cryptosporidium hominis	237895	species	2.14
report_kraken2_Trim_SRR26320749	Cryptosporidium tyzzeri	756076	species	1.5
report_kraken2_Trim_SRR26320750	Cryptosporidium parvum	5807	species	8.91
report_kraken2_Trim_SRR26320750	Kwoniella heveanensis	89924	species	0.39
report_kraken2_Trim_SRR26320750	Cryptosporidium hominis	237895	species	0.28
report_kraken2_Trim_SRR26320751	Cryptosporidium parvum	5807	species	70.56
report_kraken2_Trim_SRR26320751	Cryptosporidium hominis	237895	species	1.43

report_kraken2_Trim_SRR26320751	Cryptosporidium tyzzeri	756076	species	1.02
report_kraken2_Trim_SRR26320752	Acanthamoeba mauritaniensis	196912	species	4.63
report_kraken2_Trim_SRR26320752	Acanthamoeba palestinensis	28015	species	2.31
report_kraken2_Trim_SRR26320752	Hammondia hammondi	99158	species	0.18
report_kraken2_Trim_SRR26320753	Cryptosporidium parvum	5807	species	53.6
report_kraken2_Trim_SRR26320753	Cryptosporidium hominis	237895	species	1.11
report_kraken2_Trim_SRR26320753	Acanthamoeba mauritaniensis	196912	species	1.11
report_kraken2_Trim_SRR3091751	Cryptosporidium parvum	5807	species	58.75
report_kraken2_Trim_SRR3091751	Cryptosporidium hominis	237895	species	0.98
report_kraken2_Trim_SRR3091751	Cryptosporidium tyzzeri	756076	species	0.81
report_kraken2_Trim_SRR3091752	Cryptosporidium parvum	5807	species	88.73
report_kraken2_Trim_SRR3091752	Cryptosporidium hominis	237895	species	2.04
report_kraken2_Trim_SRR3091752	Cryptosporidium tyzzeri	756076	species	1.52
report_kraken2_Trim_SRR3091753	Cryptosporidium parvum	5807	species	77.15
report_kraken2_Trim_SRR3091753	Cryptosporidium tyzzeri	756076	species	1.38
report_kraken2_Trim_SRR3091753	Cryptosporidium hominis	237895	species	1.06
report_kraken2_Trim_SRR3091754	Cryptosporidium parvum	5807	species	87.87
report_kraken2_Trim_SRR3091754	Cryptosporidium hominis	237895	species	1.82
report_kraken2_Trim_SRR3091754	Cryptosporidium tyzzeri	756076	species	1.33
report_kraken2_Trim_SRR3091755	Cryptosporidium parvum	5807	species	82.09
report_kraken2_Trim_SRR3091755	Cryptosporidium hominis	237895	species	2.13
report_kraken2_Trim_SRR3091755	Cryptosporidium tyzzeri	756076	species	1.97
report_kraken2_Trim_SRR3091756	Cryptosporidium parvum	5807	species	85.88
report_kraken2_Trim_SRR3091756	Cryptosporidium hominis	237895	species	2.32
report_kraken2_Trim_SRR3091756	Cryptosporidium tyzzeri	756076	species	1.33
report_kraken2_Trim_SRR3091757	Cryptosporidium parvum	5807	species	85.69
report_kraken2_Trim_SRR3091757	Cryptosporidium hominis	237895	species	1.63
report_kraken2_Trim_SRR3091757	Cryptosporidium tyzzeri	756076	species	1.47
report_kraken2_Trim_SRR3091758	Cryptosporidium parvum	5807	species	79.81
report_kraken2_Trim_SRR3091758	Cryptosporidium hominis	237895	species	2.52
report_kraken2_Trim_SRR3091758	Cryptosporidium tyzzeri	756076	species	2.27
report_kraken2_Trim_SRR3091759	Cryptosporidium parvum	5807	species	77.33
report_kraken2_Trim_SRR3091759	Cryptosporidium hominis	237895	species	1.64
report_kraken2_Trim_SRR3091759	Cryptosporidium tyzzeri	756076	species	1.25
report_kraken2_Trim_SRR3091760	Cryptosporidium parvum	5807	species	81.28
report_kraken2_Trim_SRR3091760	Cryptosporidium hominis	237895	species	2.17
report_kraken2_Trim_SRR3091760	Cryptosporidium tyzzeri	756076	species	1.51
report_kraken2_Trim_SRR3091761	Cryptosporidium parvum	5807	species	84.81
report_kraken2_Trim_SRR3091761	Cryptosporidium hominis	237895	species	2.05
report_kraken2_Trim_SRR3091761	Cryptosporidium tyzzeri	756076	species	1.55

report_kraken2_Trim_SRR3473975	Cryptosporidium parvum	5807	species	77.26
report_kraken2_Trim_SRR3473975	Cryptosporidium hominis	237895	species	2.39
report_kraken2_Trim_SRR3473975	Cryptosporidium tyzzeri	756076	species	1.93
report_kraken2_Trim_SRR6117460	Cryptosporidium parvum	5807	species	69.87
report_kraken2_Trim_SRR6117460	Cryptosporidium hominis	237895	species	1.82
report_kraken2_Trim_SRR6117460	Cryptosporidium tyzzeri	756076	species	1.18
report_kraken2_Trim_SRR6147472	Cryptosporidium parvum	5807	species	78.59
report_kraken2_Trim_SRR6147472	Cryptosporidium hominis	237895	species	1.27
report_kraken2_Trim_SRR6147472	Cryptosporidium tyzzeri	756076	species	0.89
report_kraken2_Trim_SRR6147581	Cryptosporidium parvum	5807	species	80.23
report_kraken2_Trim_SRR6147581	Cryptosporidium hominis	237895	species	1.2
report_kraken2_Trim_SRR6147581	Cryptosporidium tyzzeri	756076	species	0.86
report_kraken2_Trim_SRR6147587	Cryptosporidium parvum	5807	species	81.52
report_kraken2_Trim_SRR6147587	Cryptosporidium hominis	237895	species	1.85
report_kraken2_Trim_SRR6147587	Cryptosporidium tyzzeri	756076	species	1.28
report_kraken2_Trim_SRR6147945	Cryptosporidium parvum	5807	species	76.06
report_kraken2_Trim_SRR6147945	Cryptosporidium hominis	237895	species	1.97
report_kraken2_Trim_SRR6147945	Cryptosporidium tyzzeri	756076	species	1.3
report_kraken2_Trim_SRR6147964	Cryptosporidium parvum	5807	species	80.6
report_kraken2_Trim_SRR6147964	Cryptosporidium hominis	237895	species	1.25
report_kraken2_Trim_SRR6147964	Cryptosporidium tyzzeri	756076	species	0.89
report_kraken2_Trim_SRR6148259	Cryptosporidium parvum	5807	species	72.92
report_kraken2_Trim_SRR6148259	Cryptosporidium hominis	237895	species	1.39
report_kraken2_Trim_SRR6148259	Cryptosporidium tyzzeri	756076	species	1.14
report_kraken2_Trim_SRR6813716	Cryptosporidium parvum	5807	species	61.88
report_kraken2_Trim_SRR6813716	Cryptosporidium hominis	237895	species	1.97
report_kraken2_Trim_SRR6813716	Cryptosporidium tyzzeri	756076	species	1.88
report_kraken2_Trim_SRR6813717	Cryptosporidium parvum	5807	species	61.17
report_kraken2_Trim_SRR6813717	Cryptosporidium hominis	237895	species	2.18
report_kraken2_Trim_SRR6813717	Cryptosporidium tyzzeri	756076	species	2.03
report_kraken2_Trim_SRR6813718	Cryptosporidium parvum	5807	species	73.17
report_kraken2_Trim_SRR6813718	Cryptosporidium hominis	237895	species	2.09
report_kraken2_Trim_SRR6813718	Cryptosporidium tyzzeri	756076	species	1.99
report_kraken2_Trim_SRR6813719	Cryptosporidium parvum	5807	species	60.31
report_kraken2_Trim_SRR6813719	Cryptosporidium hominis	237895	species	2.09
report_kraken2_Trim_SRR6813719	Cryptosporidium tyzzeri	756076	species	1.94
report_kraken2_Trim_SRR6871415	Cryptosporidium parvum	5807	species	70.01
report_kraken2_Trim_SRR6871415	Cryptosporidium hominis	237895	species	1.44
report_kraken2_Trim_SRR6871415	Cryptosporidium tyzzeri	756076	species	1.04
report_kraken2_Trim_SRR7898459	Cryptosporidium parvum	5807	species	64.73

report_kraken2_Trim_SRR7898459	Cryptosporidium hominis	237895	species	1.54
report_kraken2_Trim_SRR7898459	Cryptosporidium tyzzeri	756076	species	1.2
report_kraken2_Trim_SRR9070312	Cryptosporidium parvum	5807	species	92.95
report_kraken2_Trim_SRR9070312	Cryptosporidium hominis	237895	species	1.65
report_kraken2_Trim_SRR9070312	Cryptosporidium tyzzeri	756076	species	0.98
report_kraken2_Trim_SRR9070313	Cryptosporidium parvum	5807	species	93.09
report_kraken2_Trim_SRR9070313	Cryptosporidium hominis	237895	species	1.67
report_kraken2_Trim_SRR9070313	Cryptosporidium tyzzeri	756076	species	1
report_kraken2_Trim_SRR9070314	Cryptosporidium parvum	5807	species	93.08
report_kraken2_Trim_SRR9070314	Cryptosporidium hominis	237895	species	1.74
report_kraken2_Trim_SRR9070314	Cryptosporidium tyzzeri	756076	species	1.03

Supplemental Table 3

Supplemental Figures

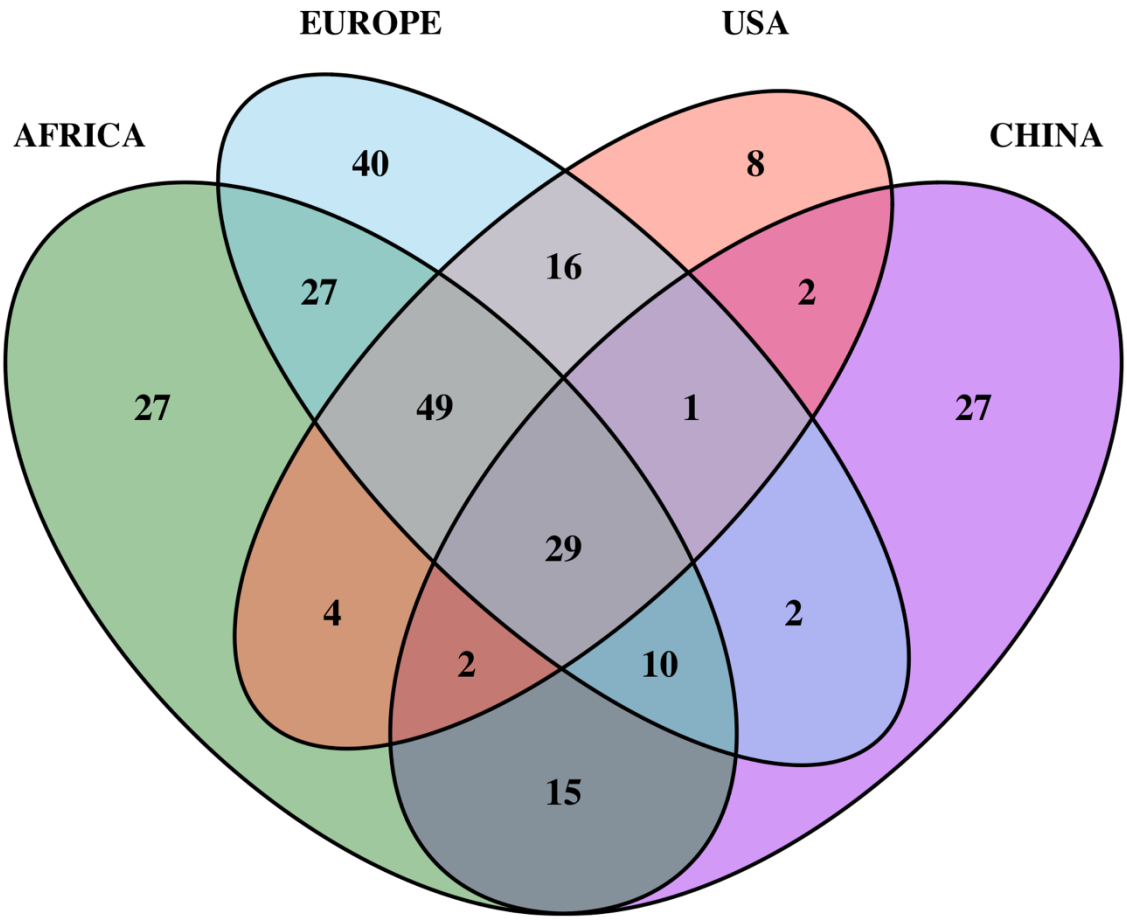


Figure 1. An overlap of candidate genes under positive selection across regions. Venn diagram of genes with pN/pS > 1.5 (putative positive selection) across four regions, Africa, Europe, USA, China. Numbers denote gene counts; 29 genes are shared by all four, while region-specific sets include Africa 27, Europe 40, USA 8, and China 27. Ratios were computed from QC-filtered coding variants.